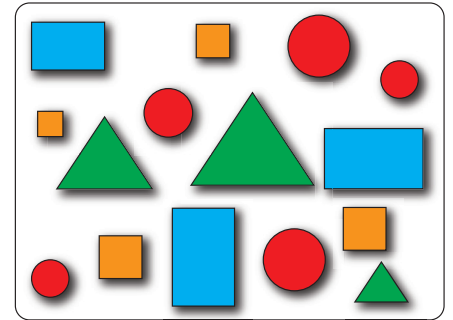
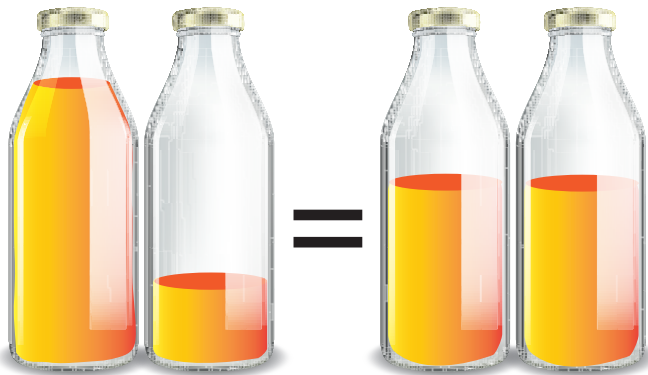


Data Sets and Distributions

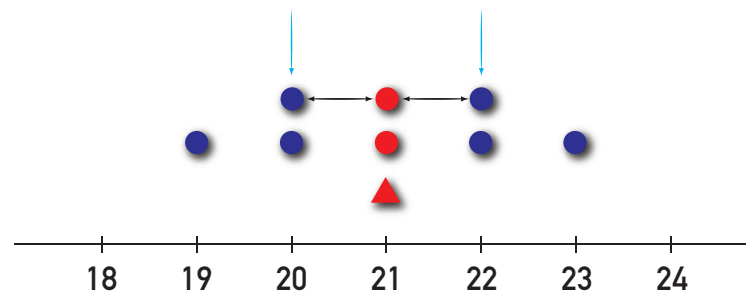


Teacher Guide

Average



Measure of Center



Dogs Weight

Creative Commons Licensing

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.



You are free:

- to Share**—to copy, distribute, and transmit the work
- to Remix**—to adapt the work

Under the following conditions:

Attribution—You must attribute the work in the following manner:
CKMath 6–8 was originally developed by Open Up Resources and authored by Illustrative Mathematics, <https://www.illustrativemathematics.org>, and is copyrighted as 2017–2019 by Open Up Resources. It is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). The Open Up Resources 6–8 Math Curriculum is available at: <https://www.openupresources.org/math-curriculum/>.

Adaptations and updates to the IM 6–8 Math English language learner supports and the additional English assessments marked as "B" are copyright 2019 by Open Up Resources and licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Adaptations and updates to the IM K–8 Math Spanish translation of assessments marked as "B" are copyright 2019 by Illustrative Mathematics. These adaptations and updates are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

This particular work is based on additional work of the Core Knowledge® Foundation (www.coreknowledge.org) made available through licensing under a Creative Commons Attribution-Non Commercial-Share Alike 4.0 International License. This does not in any way imply that the Core Knowledge Foundation endorses this work.

Noncommercial—You may not use this work for commercial purposes.

Share Alike—If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

With the understanding that:

For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page:

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Copyright © 2023 Core Knowledge Foundation
www.coreknowledge.org

All Rights Reserved.

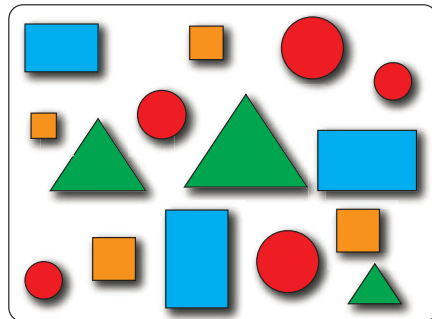
Core Knowledge®, Core Knowledge Curriculum Series™, Core Knowledge Math™ and CKMath™ are trademarks of the Core Knowledge Foundation.

Trademarks and trade names are shown in this book strictly for illustrative and educational purposes and are the property of their respective owners. References herein should not be regarded as affecting the validity of said trademarks and trade names.

Data Sets and Distributions

Table of Contents

Introduction: Unit Narrative	1
Student Learning Targets	3
Terminology	6
Required Materials	7
Lessons Plans and Student Task Statements:	
Section 1: Lessons 1, 2 Data, Variability, and Statistical Questions	9
Section 2: Lessons 3–8 Dot Plots and Histograms	47
Section 3: Lessons 9–12 Measures of Center and Variability .	149
Section 4: Lessons 13–17 Median and IQR	212
Section 5: Lesson 18 Let’s Put it to Work	295
Teacher Resources	313
Family Support Materials	
Unit Assessments	
Assessment Answer Keys	
Cool Downs (Lesson-Level Assessments)	
Instructional Masters	



Data Sets and Distributions
Teacher Guide
Core Knowledge Mathematics™

Data Sets and Distributions

Unit Narrative

In this unit, students learn about populations and study variables associated with a population. They understand and use the terms “numerical data,” “categorical data,” “survey” (as noun and verb), “statistical question,” “variability,” “distribution,” and “frequency.” They make and interpret histograms, bar graphs, tables of frequencies, and box plots. They describe distributions (shown on graphical displays) using terms such as “symmetrical,” “peaks,” “gaps,” and “clusters.” They work with measures of center—understanding and using the terms “mean,” “average,” and “median.” They work with measures of variability—understanding and using the terms “range,” “mean absolute deviation” or MAD, “quartile,” and “interquartile range” or IQR. They interpret measurements of center and variability in contexts.

Progression of Disciplinary Language

In this unit, teachers can anticipate students using language for mathematical purposes such as justifying, representing, and interpreting. Throughout the unit, students will benefit from routines designed to grow robust disciplinary language, both for their own sense-making and for building shared understanding with peers. Teachers can formatively assess how students are using language in these ways, particularly when students are using language to:

Justify

- reasoning for matching data sets to questions (Lesson 2)
- reasoning about dot plots (Lesson 3)
- reasoning about mean and median (Lesson 13)
- reasoning about changes in mean and median (Lesson 14)
- reasoning about which information is needed (Lesson 17)
- which summaries and graphs best represent given data sets (Lesson 18)

Represent

- data using dot plots (Lessons 3 and 4)
- data using histograms (Lesson 7)
- mean using bar graphs (Lesson 9)
- data with five number summaries (Lesson 15)
- data using box plots (Lesson 16)

Interpret

- dot plots (Lessons 4 and 11)
- histograms (Lessons 6 and 18)
- mean of a data set (Lesson 9)

- five number summaries (Lesson 15)
- box plots (Lesson 16)

In addition, students are expected to critique the reasoning of others, describe how quantities are measured, describe and compare features and distributions of data sets, generalize about means and distances in data sets, generalize categories for sorting data sets, and generalize about statistical questions. Students are also expected to use language to compare questions that produce numerical and categorical data, compare dot plots and histograms, and compare histograms and bar graphs.

The table shows lessons where new terminology is first introduced, including when students are expected to understand the word or phrase receptively and when students are expected to produce the word or phrase in their own speaking or writing. Terms from the glossary appear bolded. Teachers should continue to support students' use of a new term in the lessons that follow where it was first introduced.

Learning Targets

Data Sets and Distributions

Lesson 1: Got Data?

- I can collect the correct data to answer a question and use the correct units.
- I can explain the difference between categorical and numerical data.

Lesson 2: Statistical Questions

- I can tell when data has variability.

Lesson 3: Representing Data Graphically

- I can describe the information presented in tables, dot plots, and bar graphs.
- I can use tables, dot plots, and bar graphs to represent distributions of data.

Lesson 4: Dot Plots

- I can describe the center and spread of data from a dot plot.

Lesson 5: Using Dot Plots to Answer Statistical Questions

- I can use a dot plot to represent the distribution of a data set and answer questions about the real-world situation.
- I can use center and spread to describe data sets, including what is typical in a data set.

Lesson 6: Histograms

- I can recognize when a histogram is an appropriate graphical display of a data set.
- I can use a histogram to get information about the distribution of data and explain what it means in a real-world situation.

Lesson 7: Using Histograms to Answer Statistical Questions

- I can draw a histogram from a table of data.
- I can use a histogram to describe the distribution of data and determine a typical value for the data.

Lesson 8: Describing Distributions on Histograms

- I can describe the shape and features of a histogram and explain what they mean in the context of the data.
- I can distinguish histograms and bar graphs.

Lesson 9: Interpreting the Mean as Fair Share

- I can explain how the mean for a data set represents a “fair share.”
- I can find the mean for a numerical data set.

Lesson 10: Finding and Interpreting the Mean as the Balance Point

- I can describe what the mean tells us in the context of the data.
- I can explain how the mean represents a balance point for the data on a dot plot.

Lesson 11: Deviation from the Mean

- I can find the MAD for a set of data.
- I know what the mean absolute deviation (MAD) measures and what information it provides.

Lesson 12: Using Mean and MAD to Make Comparisons

- I can say what the MAD tells us in a given context.
- I can use means and MADs to compare groups.

Lesson 13: The Median of a Data Set

- I can find the median for a set of data.
- I can say what the median represents and what it tells us in a given context.

Lesson 14: Comparing Mean and Median

- I can determine when the mean or the median is more appropriate to describe the center of data.
- I can explain how the distribution of data affects the mean and the median.

Lesson 15: Quartiles and Interquartile Range

- I can use IQR to describe the spread of data.
- I know what quartiles and interquartile range (IQR) measure and what they tell us about the data.
- When given a list of data values or a dot plot, I can find the quartiles and interquartile range (IQR) for data.

Lesson 16: Box Plots

- I can use the five-number summary to draw a box plot.
- I know what information a box plot shows and how it is constructed.

Lesson 17: Using Box Plots

- I can use a box plot to answer questions about a data set.
- I can use medians and IQRs to compare groups.

Lesson 18: Using Data to Solve Problems

- I can decide whether mean and MAD or median and IQR would be more appropriate for describing the center and spread of a data set.
- I can draw an appropriate graphical representation for a set of data.
- I can explain what the mean and MAD or the median and IQR tell us in the context of a situation and use them to answer questions.

lesson	new terminology	
	receptive	productive
6.8.1	numerical data categorical data dot plot	
6.8.2	statistical question variability	
6.8.3	distribution frequency	bar graph
6.8.4	typical	
6.8.5	center spread	variability
6.8.6	histogram bins	distribution center
6.8.7		statistical question spread
6.8.8	symmetrical peak cluster unusual value	numerical data categorical data gap
6.8.9	average mean fair share	
6.8.10	measure of center balance point	
6.8.11	mean absolute deviation (MAD) measure of spread	symmetrical mean
6.8.12		mean absolute deviation (MAD) typical
6.8.13	median	measure of center

lesson	new terminology	
	receptive	productive
6.8.14		peak cluster unusual value
6.8.15	range quartile interquartile range (IQR) five-number summary	measure of spread minimum maximum
6.8.16	box plot whisker	median interquartile range (IQR)
6.8.17		range quartile
6.8.18		dot plot histogram box plot

Required Materials

Blank paper

Decks of playing cards

Dot stickers

Small circular sticker useful for plotting points on a display.

Four-function calculators

Index cards

Measuring tapes

Pre-printed cards, cut from copies of the Instructional master

Pre-printed slips, cut from copies of the Instructional master

Rulers

Rulers marked with centimeters

Snap cubes

Sticky notes

Straightedges

A rigid edge that can be used for drawing line segments. Sometimes a ruler is okay to use as a straightedge, but sometimes it is preferable to use an unruled straightedge, like a blank index card.

Tape

Tools for creating a visual display

Any way for students to create work that can be easily displayed to the class. Examples: chart paper and markers, whiteboard space and markers, shared online drawing tool, access to a document camera.

Section: Data, Variability, and Statistical Questions

Lesson 1: Got Data?

Goals

- Ask survey questions (orally) and record responses (in writing). Include units of measurement when reporting numerical data (orally and in writing).
- Comprehend and use the terms “numerical” and “categorical” to describe data sets (orally and in writing).
- Interpret various representations of data sets and determine whether it is reasonable that a verbal description represents a given numerical data set.

Learning Targets

- I can collect the correct data to answer a question and use the correct units.
- I can explain the difference between categorical and numerical data.

Lesson Narrative

Students begin the unit by interacting closely with data. They collect data about themselves by measuring and answering survey questions, studying the different types of responses collected, and identifying the appropriate variables and units being measured.

Students learn about **categorical** and **numerical** data. They determine whether a particular survey question will produce one type of data or the other. They also get reacquainted with dot plots (often called line plots in earlier grades) as a way to represent data and make sense of what the data points mean in context (MP2).

Alignments

Building On

- 2.MD.D.9: Generate measurement data by measuring lengths of several objects to the nearest whole unit, or by making repeated measurements of the same object. Show the measurements by making a line plot, where the horizontal scale is marked off in whole-number units.
- 4.MD.A.1: Know relative sizes of measurement units within one system of units including km, m, cm; kg, g; lb, oz.; l, ml; hr, min, sec. Within a single system of measurement, express measurements in a larger unit in terms of a smaller unit. Record measurement equivalents in a two-column table. For example, know that 1 ft is 12 times as long as 1 in. Express the length of a 4 ft snake as 48 in. Generate a conversion table for feet and inches listing the number pairs (1, 12), (2, 24), (3, 36), ...
- 5.MD.B: Represent and interpret data.

Addressing

- 6.SP.B: Summarize and describe distributions.

Building Towards

- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

Instructional Routines

- Group Presentations
- MLR2: Collect and Display
- Notice and Wonder

Required Materials

Blank paper
Measuring tapes

Rulers
Sticky notes

Required Preparation

For the activity Surveying the Class:

Choose 4–5 survey questions and measurement activities in advance. Be sure to include questions and activities that would produce both categorical and numerical data. The questions about how and how long it takes students travel to school (the first two prompts) and students' heights in centimeters (the third prompt) will be used in a later lesson, so be sure to include these questions. Provide sticky notes to each student on which they can record their responses. Set up one position in the room for each selected question where students may put their sticky notes to have a visual display of responses.

To collect measurements, prepare measuring stations equipped with the necessary tools (e.g., rulers, measuring tape, etc.), instructions on how to measure, and a way to record the measurements. Students can then rotate through the stations.

Student Learning Goals

Let's explore different kinds of data.

1.1 Dots of Data

Warm Up: 10 minutes

The purpose of this warm-up is to review students' prior knowledge about Representation of numerical data. Students may be familiar with line plot from previous grades but unfamiliar with the term **dot plot**, which is what will be used in this unit and beyond. Students learn that both terms are commonly used for the same type of diagram.

Students examine a dot plot of data and consider which contexts may make sense for the data shown. Then students invent their own context for the data and interpret what additional data in that set would mean in their context.

This is an opportunity to see how students make sense of a data representation in the context of situations (MP2), which is central to the work of the unit.

Building On

- 5.MD.B

Building Towards

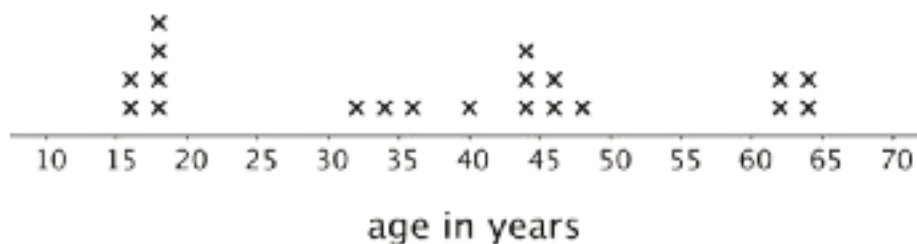
- 6.SP.B.4

Instructional Routines

- Notice and Wonder

Launch

Display the image for all to see.



Ask students what they notice and wonder about the image. Students should notice:

- The numbers represent age based on the label.
- Each X represents one person. For example, it looks like there is 1 40-year old represented but 3 44-year olds.
- There are a total of 20 people represented.

Students may wonder:

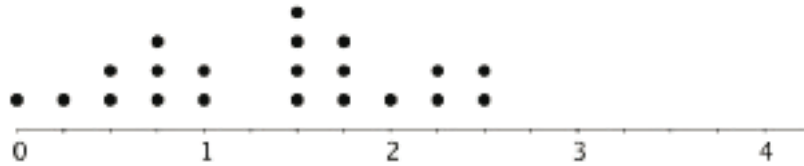
- Why are these people important?
- Why is there such a large age range?
- Why is there nobody in their 20s represented?

If not mentioned by students, remind them that the display is a line plot. Tell students that in the task they will see and use the same type of representation, but it is called a dot plot, and that dots are used instead of Xs.

Give students 2 minutes of quiet work time to complete the task, followed by a whole-class discussion.

Student Task Statement

Here is a dot plot for a data set.



1. Determine if each of the following would be an appropriate label to represent the data in the dot plot? Be prepared to explain your reasoning.
 - a. Number of children per class.
 - b. Distance between home and school, in miles.
 - c. Hours spent watching TV each day.
 - d. Weight of elephants, in pounds.
 - e. Points received on a homework assignment.
2. Think of another label that can be used with the dot plot.
 - a. Write it below the scale of the dot plot. Be sure to include the unit of measurement.
 - b. In your scenario, what does one dot represent?
 - c. In your scenario, what would a data point of 0 mean? What would a data point of $3\frac{1}{4}$ mean?

Student Response

1. Reasoning varies. Sample responses:
 - a. No. We cannot have partial children.
 - b. Yes. The distances are reasonable for the context and can be fractional.
 - c. Yes. The hours are reasonable for the context and can be fractional.
 - d. No. The unit of measurement (pounds) does not work with the context. Elephants weigh much more than several pounds.
 - e. Yes. A teacher could grade a homework assignment on a scale of 1 to 4 and assign partial points.

2. Answers vary. Sample responses:
- Time spent napping, in hours.
 - A dot represents the length of nap (in hours) on one day.
 - A day without a nap.

Activity Synthesis

The purpose of the discussion is for students to understand how to read a dot plot including what dots represent in context.

Poll students on their responses to each scenario presented in the first question and ask a few students to explain their reasoning for each. After each student shares their explanation, ask the class if they agree or disagree and why.

Invite a few students to share another label they think could be used with the set of data and what each dot would represent. Based on each response, ask the class the following the questions:

- “What would a data point of 0 mean in the context mentioned?”
- “What would a data point of $3\frac{1}{4}$ mean in the context mentioned?”

1.2 Surveying the Class

20 minutes

Students begin their statistical explorations with data collection. They answer several survey questions, take some measurements, or do both. This data gathering activity serves several purposes: to give students firsthand experience in gathering and organizing data, to prompt students to notice different types of data that could be collected, and to provide the class with authentic data that can later be analyzed as students gain and expand the skills to do so.

The task statements show a range of survey questions and ideas for collecting measurements. To optimize the data gathering, have several survey questions and measuring activities identified ahead of time and consider how to best collect the responses or measurements.

Building On

- 2.MD.D.9

Addressing

- 6.SP.B

Instructional Routines

- Group Presentations
- MLR2: Collect and Display

Launch

Explain to students that they will gather data to learn more about the students in the class. Tell them which data sets will be collected and give instructions on the collection process. If students are to do a gallery walk or rotate through measuring stations, consider arranging them into groups of 3–4 to facilitate the rotation.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Chunk this task into more manageable parts to support students who benefit from support with organizational skills in problem solving. For example, demonstrate the data collection process and provide graphic organizers for collecting data from the survey questions.

Supports accessibility for: Organization; Attention

Anticipated Misconceptions

When taking measurements, students might not remember to attend to the right units, to start from 0, or to hold the measuring tool so that measurements can be precisely taken. Remind students about these issues as needed.

Student Task Statement

Here are some survey questions. Your teacher will explain which questions can be used to learn more about the students in your class and how the responses will be collected. The data that your class collects will be used in upcoming activities.

1. How long does it usually take you to travel to school? Answer to the nearest minute.
2. How do you travel to school on most days? Choose one.
 - Walk
 - Car
 - Public transport
 - Bike
 - School bus
 - Other
 - Scooter or skateboard
3. How tall are you without your shoes on? Answer to the nearest centimeter.
4. What is the length of your right foot without your shoe on? Answer to the nearest centimeter.
5. What is your arm span? Stretch your arms open, and measure the distance from the tip of your right hand's middle finger to the tip of your left hand's middle finger, across your back. Answer to the nearest centimeter.
6. How important are the following issues to you? Rate each on a scale from 0 (not important) to 10 (very important).

- a. Reducing pollution
- b. Recycling
- c. Conserving water

7. Do you have any siblings? ____ Yes ____ No

8. How many hours of sleep per night do you usually get when you have school the next day? Answer to the nearest half hour.

9. How many hours of sleep per night do you usually get when you do not have school the next day? Answer to the nearest half hour.

10. Other than traveling from school, what do you do right after school on most days?

- Have a snack
- Practice a sport
- Do homework
- Do chores
- Read a book
- Use the computer
- Talk on the phone
- Participate in an extracurricular activity

11. If you could meet one of these celebrities, who would you choose?

- A city or state leader
- A musical artist
- A champion athlete
- A best-selling author
- A movie star

12. Estimate how much time per week you usually spend on each of these activities. Answer to the nearest quarter of an hour.

- a. Playing sports or doing outdoor activities
- b. Using a screen for fun (watching TV, playing computer games, etc.)
- c. Doing homework
- d. Reading

Student Response

Data collected vary.

Activity Synthesis

Tell students to put their responses to the selected questions in the appropriate place in the room. Allow them to do a gallery walk of the data sets discussing with their group things they notice about each question. After students have had a chance to view the data, draw their attention to the twelve survey questions in the task statements. Ask them to think about the types of responses they produce. Students are likely to notice that responses to some questions are numbers and others

are not. Explain that responses that are measurements or quantities are called **numerical data**. For example, the first question about the travel time to school produces numerical data because the responses are quantities, measured in minutes.

Point out that responses to other questions are not quantities but can be sorted into categories. Explain that these types of responses are called **categorical data**. For example, the second question about ways of traveling to school produces categorical data because the responses can be sorted into categories.

Explain that numerical and categorical data will continue to be investigated in upcoming lessons.

The data from the first three questions regarding travel time, travel method, and height will be used in future lessons. Collect this data in an organized way to redistribute and use with students later.

Access for English Language Learners

Representing, Listening: MLR2 Collect and Display. During the gallery walk, listen for words and phrases students use to describe data that is “numerical” (measurements or quantitative values) and data that is “categorical” (where the values represent categories). Display students’ words and phrases grouped into two lists prior to introducing the terms and definitions. This will help students to connect their informal use of language to vocabulary.

Design Principle(s): Optimize output (for describing); Support sense-making

1.3 Numerical and Categorical Data

10 minutes

In the previous activity, students responded to survey questions and collected data. They learned that data can be categorical or numerical. In this activity, students practice distinguishing **categorical** and **numerical** data, using the same survey questions and additional ones. They think about the kind of responses these questions would yield. For numerical responses, they consider the units of measurement. For categorical responses, they identify the characteristic being studied.

Note that some data may have numbers for their values but are categorical rather than numerical. Area codes and zip codes are examples of such categorical data. They are not quantities or measurements, but rather labels that happen to be numbers. We can meaningfully compare quantities or measurements (e.g., 6 minutes is greater than 4 minutes, or 7 years is less than 12 years), but we cannot do the same with the numbers in area codes or zip codes. It would not make sense to say, for example, that the zip code is 19104 is less than 63105.

As students work and discuss, notice students who could succinctly articulate the variables being investigated in both numerical and categorical questions. Also notice any disagreements partners might have about whether a question results in numerical or categorical data. The sixth survey question (rating the importance of environmental issues) and last survey question (hours spent on

out-of-school activities) are likely to generate conversations as they may appear less straightforward than the others.

Building On

- 5.MD.B

Addressing

- 6.SP.B

Launch

Arrange students in groups of 2. Tell students they will need the list of survey questions from the previous activity. Give them 3–4 minutes of quiet work time to complete the first three questions. Ask them to briefly discuss their responses with their partner before completing the last question.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Provide students with a Venn diagram with which to compare the similarities and differences between numerical and categorical data.

Supports accessibility for: Language; Organization

Anticipated Misconceptions

Students may mistake numbers such as area codes, zip codes, or the numbers we use to represent months (e.g., 1 for January) as numerical data. Be sure to discuss this common confusion if arises. See the Activity Narrative section for ideas for addressing it.

Student Task Statement

The list of survey questions in the activity earlier can help you complete these exercises.

1. The first survey question about travel *time* produces **numerical data**. Identify two other questions that produce numerical data. For each, describe what was measured and its unit of measurement.

a. Question #: _____ What was measured:

Unit of measurement:

b. Question #: _____ What was measured:

Unit of measurement:

2. The second survey question about travel *method* produces **categorical data**. Identify two other questions that produce categorical data. For each, describe what characteristic or feature was being studied.

a. Question #: _____ Characteristic being studied:

b. Question #: _____ Characteristic being studied:

3. Think about the responses to these survey questions. Do they produce numerical or categorical data? Be prepared to explain how you know.

a. How many pets do you have?

b. How many years have you lived in this state?

c. What is your favorite band?

d. What kind of music do you like best?

e. What is the area code of your school's phone number?

f. Where were you born?

g. How much does your backpack weigh?

4. Name two characteristics you could investigate to learn more about your classmates. Make sure one would give categorical data and the other would give numerical data.

Student Response

- Answers vary. Questions 3, 4, 5, 6, 8, 9, 12 produce numerical data. Sample response:
 - Question 3. Value being measured: height of student without shoes on. Unit of measurement: centimeters.
 - Question 12. Values being measured: amount of time spent on different activities. Unit of measurement: minutes.
- Answers vary. Questions 7, 10, 11 produce categorical data. Sample response:
 - Question 7. Characteristic studied: whether student is the only child.
 - Question 11. Characteristic studied: first activity after school.
- Numerical
 - Numerical
 - Categorical
 - Categorical
 - Categorical
 - Categorical
 - Numerical

4. Answers vary. Sample response: Numerical: The number of people in a student's household.
Categorical: The name of the street where a student lives.

Are You Ready for More?

Priya and Han collected data on the birth months of students in their class. Here are the lists of their records for the same group of students.

This list shows Priya's records.

Jan	Apr	Jan	Feb	Oct	May	June	July	Aug
Aug	Sep	Jan	Feb	Mar	Apr	Nov	Nov	Dec
Feb	Mar							

This list shows Han's records.

1	4	1	2	10	5	6	7	8
8	9	1	2	3	4	11	11	12
2	3							

1. How are their records alike? How are they different?
2. What kind of data—categorical or numerical—do you think the variable “birth month” produces? Explain how you know.

Student Response

1. Answers vary. Sample response: Alike: They both collected the data of the same students and in the same order. Different: Priya recorded the names of the month (e.g., January, April, etc.), while Han recorded the numbers showing the order in which the months appear in the year (e.g., 5 for May, 7 for July, etc.).
2. Birth month is categorical. Sample explanation: The numbers that Han wrote are labels or names for months, which cannot be measured or compared (i.e., we don't say that month 5 is less than month 7). They appear in a certain order, but they are not quantities.

Activity Synthesis

The purpose of the discussion is for students to be able to identify data collected as either numerical or categorical.

Select a few previously identified students to share their responses to the first two questions. After each student shares, ask the rest of the class if they agree or disagree and discuss any

disagreements. Then, poll the class on their responses to the third set of questions. If not mentioned by students, explain how some categorical data are comprised of numbers, as noted in the Activity Narrative section.

Lesson Synthesis

In this lesson, we collected and explored different types of data. We noticed that certain survey questions produce responses that are quantities or measurements; we call these responses **numerical data**. Other questions produce responses that are not measurements or quantities but can be sorted into categories; we call these **categorical data**.

- “What are some examples of categorical data?”
- “What are some examples of numerical data?”
- “What is a **dot plot**?” (A dot plot is a representation of numerical data.)
- “How does it represent data?” (It represents each data value with a point above a number line.)

Consider creating a permanent display of the vocabulary throughout the unit including numerical data, categorical data, and dot plot for this lesson.

1.4 What’s the Question?

Cool Down: 5 minutes

Building On

- 4.MD.A.1
- 5.MD.B

Addressing

- 6.SP.B

Student Task Statement

1. Would each survey question produce categorical data or numerical data?
 - a. What is your favorite vegetable?
 - b. Have you been to the capital city of your state?
 - c. How old is the youngest person in your family?
 - d. In which zip code do you live?
 - e. What is the first letter of your name?
 - f. How many hours do you spend outdoors each day?
2. Andre collected data measured in centimeters.

8.5 10.5 7.8 9.5 8.1 9.0 10.2 9.6 11.2 10.9 12.7 9.8

What could he be investigating? Select **all** that apply.

- a. The weight of a dozen eggs.
- b. The length of leaves from a tree.
- c. The height of cups and mugs in a cupboard.
- d. The length of songs on a music CD.
- e. The length of colored pencils in a box.

Student Response

1.
 - a. Categorical
 - b. Categorical
 - c. Numerical
 - d. Categorical
 - e. Categorical
 - f. Numerical

2. b, c, e

Student Lesson Summary

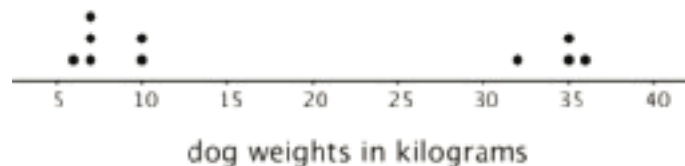
The table contains data about 10 dogs.

dog name	weight (kg)	breed
Duke	36	German shepherd
Coco	6	pug
Pierre	7	pug
Ginger	35	German shepherd
Lucky	10	beagle
Daisy	10	beagle
Buster	35	German shepherd
Pepper	7	pug
Rocky	7	beagle
Lady	32	German shepherd

- The weights of the dogs are an example of **numerical data**, which is data that are numbers, quantities, or measurements. The weights of the dogs are measurements in kilograms.
- The dog breeds are an example of **categorical data**, which is data containing values that can be sorted into categories. In this case, there are three categories for dog breeds: pug, beagle, and German shepherd.

Some data with numbers are categorical because the numbers are *not* quantities or measurements. For example, telephone area codes are categorical data, because the numbers are labels rather than quantities or measurements.

Numerical data can be represented with a **dot plot** (sometimes called a line plot). Here is a dot plot that shows the weights of the dogs.



We can collect and study both kinds of data by doing surveys or taking measurements. When we do, it is important to think about what feature we are studying (for example, breeds of dogs or weights of dogs) and what units of measurement are used.



Glossary

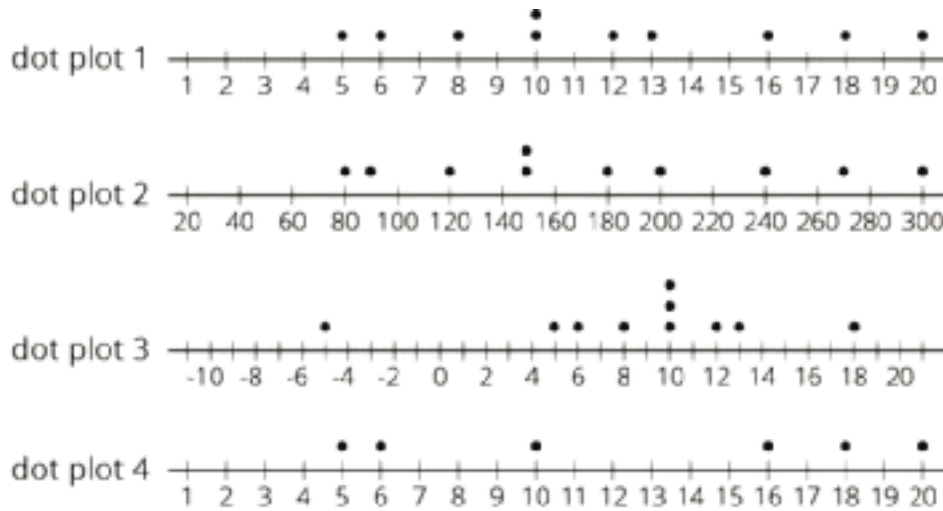
- categorical data
- dot plot
- numerical data

Lesson 1 Practice Problems

Problem 1

Statement

Tyler asked 10 students at his school how much time in minutes it takes them to get from home to school. Determine if each of these dot plots could represent the data Tyler collected. Explain your reasoning for each dot plot.



Solution

Answers vary. Sample responses:

- dot plot 1: This could be a dot plot of the time it takes to get to school in minutes for 10 students. The times range from 5 minutes to 20 minutes, which seems reasonable.
- dot plot 2: This couldn't be a dot plot of the time it takes to get to school in minutes because the values seem too big. The shortest time would be 80 minutes, which is more than an hour. The longest time would be 300 minutes, which is 5 hours. These don't seem like reasonable times that would be responses to the question that Tyler asked.
- dot plot 3: This couldn't be a dot plot of the time it takes to get to school in minutes because there are some negative values represented in the dot plot. The time it takes to get to school can't have a negative value.
- dot plot 4: This couldn't be a dot plot of the time it takes to get to school in minutes for 10 students because there are not 10 data values represented on the dot plot.

Problem 2

Statement

Here is a list of questions. For each question, decide if the responses will produce numerical data or categorical data and give two possible responses.

- a. What is your favorite breakfast food?
- b. How did you get to school this morning?
- c. How many different teachers do you have?
- d. What is the last thing you ate or drank?
- e. How many minutes did it take you to get ready this morning—from waking up to leaving for school?

Solution

- a. Categorical. Sample responses: cereal, toast
- b. Categorical. Sample responses: walked, took the bus
- c. Numerical. Sample responses: 3, 5
- d. Categorical. Sample responses: water, apple
- e. Numerical. Sample responses: 30, 45

Problem 3

Statement

- a. Write two questions that you could ask the students in your class that would result in categorical data. For each question, explain how you know that responses to it would produce categorical data.
- b. Write two questions that you could ask the students in your class that would result in numerical data. For each question, explain how you know that responses to it would produce numerical data.

Solution

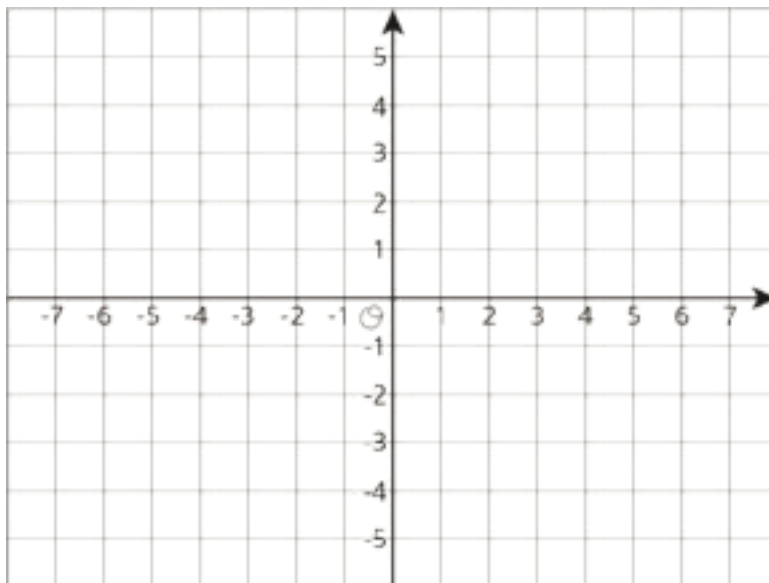
Answers vary. Sample responses:

1. What is your favorite ice cream flavor? What town do you live in? These questions result in categories.
2. How many push-ups can you do in 1 minute? What is the distance from your home to the school? These questions result in quantities or measurements.

Problem 4

Statement

Triangle DEF has vertices $D = (-4, -4)$, $E = (-2, -4)$, and $F = (-3, -1)$.



- Plot the triangle in the coordinate plane and label the vertices.
- Name the coordinates of 3 points that are inside the triangle.
- What is the area of the triangle? Show your reasoning.

Solution

- Three points and the edges connecting them are plotted.
- Answers vary. Sample response: $(-3, -2)$, $(-3, -3)$, $(-3, -2.5)$
- 3 square units; the base is 2 units, the height is 3 units, and $3 \cdot 2 \div 2 = 3$.

(From Unit 7, Lesson 15.)

Lesson 2: Statistical Questions

Goals

- Justify (orally) whether a question is “statistical” based on whether variability is expected in the data that could be collected.
- Match survey questions to data sets representing possible responses and justify (in writing) why they match.

Learning Targets

- I can tell when data has variability.

Lesson Narrative

In this lesson, students continue to analyze questions and the kinds of responses they can expect from those questions. They begin to recognize **variability** in data and learn about **statistical questions** and how they differ from non-statistical questions. In order to define variability, students categorize data sets and name the categories to make use of structure (MP7) because they are seeking mathematically important similarities between the objects.

Alignments

Building On

- 5.MD.B.2: Make a line plot to display a data set of measurements in fractions of a unit ($\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$). Use operations on fractions for this grade to solve problems involving information presented in line plots. For example, given different measurements of liquid in identical beakers, find the amount of liquid each beaker would contain if the total amount in all the beakers were redistributed equally.

Addressing

- 6.SP.A: Develop understanding of statistical variability.
- 6.SP.A.1: Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students' ages.
- 6.SP.B: Summarize and describe distributions.
- 6.SP.B.5.b: Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

Instructional Routines

- MLR3: Clarify, Critique, Correct
- MLR5: Co-Craft Questions

- MLR8: Discussion Supports
- Take Turns
- Think Pair Share

Required Materials

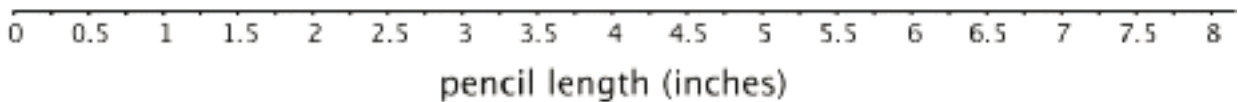
Dot stickers

Small circular sticker useful for plotting points on a display.

Pre-printed cards, cut from copies of the Instructional master
Rulers

Required Preparation

Prior to the warm-up, prepare a large class dot plot for the class to use with a horizontal axis labeled "pencil length (inches)" and intervals of $\frac{1}{4}$ from 0 to 8.



One dot sticker for each student to include in the class dot plot.

Prior to the What Makes a Statistical Question activity, prepare a two-column table to record students' observations and display for all to see.

Print and cut up cards from the Sifting for Statistical Questions Instructional master. Prepare 1 set of cards for every 2 students.

Student Learning Goals

Let's look more closely at data and the questions they can help to answer.

2.1 Pencils on A Plot

Warm Up: 5 minutes

The purpose of this warm-up is for students to review how to represent measurements on a dot plot and how to interpret the data.

Building On

- 5.MD.B.2

Addressing

- 6.SP.B

Launch

Arrange students in groups of 2. Distribute rulers marked in inches to each group, and ensure each student has a pencil.

Display the large class dot plot prepared before class for all to see and access. Tell students to measure the length of their pencil to the nearest $\frac{1}{4}$ inch and record their measurement as a dot on the class dot plot. Give each student a dot sticker as a way to record their measurement.

When the class data is recorded, give students 1 minute of quiet work time. Then, ask partners to briefly share their responses and follow with a whole-class discussion.

Anticipated Misconceptions

Some students may struggle with subtracting the shortest pencil length from the longest. Ask if they could use the horizontal axis to find the difference (e.g., by adding up from the shorter length to the longer one).

Student Task Statement

1. Measure your pencil to the nearest $\frac{1}{4}$ -inch. Then, plot your measurement on the class dot plot.
2. What is the difference between the longest and shortest pencil lengths in the class?
3. What is the most common pencil length?
4. Find the difference in lengths between the most common length and the shortest pencil.

Student Response

Answers vary. There may or may not be one most common pencil length, depending on the distribution of the data.

Activity Synthesis

The purpose of the discussion is for students to recognize the usefulness of the dot plot structure.

Ask a student to share their responses for each of the questions. Record and display their reasoning for all to see. After the student shares, ask the class if they agree or disagree and why. Some discussion may arise about the interpretation of the most common pencil size. It is ok to allow some ambiguity at this time.

To involve more students in the conversation, consider asking some of the following questions:

- Who can restate ___'s reasoning in a different way?
- Did anyone have the same response but would explain it differently?
- Did anyone find the difference between the shortest and longest lengths in a different way?
- Does anyone want to add on to ___'s reasoning?

2.2 What's in the Data?

15 minutes

In this activity, students reason abstractly and quantitatively (MP2) about numerical data sets to match them with questions that are likely to produce the data. Along the way, they categorize data sets based on whether more than one different value is present and make use of this structure (MP7) to define **variability**. They see that some survey questions lead to responses that are expected to vary when posed to different people (e.g., How many books did you read last year?), but others produce responses that are likely to be the same (e.g., What year is it this year?).

As students match questions with data sets, look out for different plausible explanations for their choices. Their matches are reasonable if they can explain why the given data could be responses to a question. Identify one student to share the response to each question. Also notice students who offer different but equally reasonable explanations for the same data set; invite them to share later.

Addressing

- 6.SP.A
- 6.SP.B.5.b

Instructional Routines

- MLR5: Co-Craft Questions
- Think Pair Share

Launch

Tell students that they will be looking at numerical data sets and thinking about what question could produce the responses in each data set. Emphasize that they need to be able to support their matching decisions with reasonable explanations.

If necessary, guide students to understand how to read the table by asking:

- "What values are included in data set A?" (0, 1, 1, 3, 0, 0, 0, 2, 1, 1)
- "To which data set does the number 8 belong?" (data set C)
- "How many people answered '1' to the question that produced the data for data set A?" (4)

Keep students in groups of 2. Give them 5 minutes of quiet work time and 1–2 minutes to share their responses with their partner.

Access for English Language Learners

Representing, Conversing: MLR5 Co-Craft Questions. Display the chart with the five data sets without revealing the questions that follow. Ask pairs of students to create possible survey questions that could lead to these data sets. Then, invite pairs to share their questions with the class. Highlight features of each data set that are described when students share the questions they came up with. This will help students make sense of the data creatively, drawing on their experiences with data, before engaging with the specific language of the questions provided in the task.

Design Principle(s): Support sense-making; Cultivate conversation

Anticipated Misconceptions

Some students may have trouble matching questions and data sets because they do not attend carefully to the range of possible solutions. For example, they may not notice that a data set with “11” as a data value cannot be a response to the first question about flipping a coin 10 times. Ask them to study the questions and data values more closely, and to look for values that seem unlikely or impossible for a given context.

Student Task Statement

Ten sixth-grade students at a school were each asked five survey questions. Their answers to each question are shown here.

data set A	0	1	1	3	0	0	0	2	1	1
data set B	12	12	12	12	12	12	12	12	12	12
data set C	6	5	7	6	4	5	3	4	6	8
data set D	6	6	6	6	6	6	6	6	6	6
data set E	3	7	9	11	6	4	2	16	6	10

1. Here are the five survey questions. Match each question to a data set that could represent the students' answers. Explain your reasoning.

- Question 1: Flip a coin 10 times. How many heads did you get?
- Question 2: How many books did you read in the last year?
- Question 3: What grade are you in?

- Question 4: How many dogs and cats do you have?
 - Question 5: How many inches are in 1 foot?
2. How are survey questions 3 and 5 different from the other questions?

Student Response

1.

Question 1: C, because the numbers in the other data sets are unlikely or impossible (they are too large or too small, are greater than 10, or all the same). When we flip a coin the chance of getting a head is one-half, and the numbers in the data are around 5, which is half of 10.

Question 2: E, because the numbers make sense for the number of books read in a year. It is unlikely that they all read the same number of books or that some read 0 books.

Question 3: D, because all students being interviewed were sixth-grade students.

Question 4: A, because 0–3 make sense for the number of pets people might have.

Question 5: B, because the number of inches in a foot is always 12, regardless of who answers the question.

2. Answers vary. Sample response: Everyone surveyed gave the same answer for these 2 questions.

Activity Synthesis

The purpose of this discussion is for students to define **variability** and recognize when it is present.

Select previously identified students to share their choices and explanations. Briefly poll the class after each explanation to see if others made the same choice for the same reason. If not, invite students with different explanations to share.

Discuss how the question about grade level and the one about number of inches in a foot are different from the others. If not mentioned by students, highlight the idea of variability. Explain that we use the term variability to describe data sets in which not every data value is the same. Data sets B and D are unlike the other sets because they show no variability. In future lessons, we will look deeper into the concept of variability and what it can tell us about the data we have collected.

2.3 What Makes a Statistical Question?

15 minutes

In the previous activities, students made sense of data sets contextually and reasoned about possible questions that could produce them. They also looked at variability in data sets and contrasted data with and without variability. Here they use both the experience of reasoning about questions and the idea of variability to define **statistical questions**.

From their work in earlier grades, students are familiar with the idea that some questions can be answered by collecting data (e.g., “How many students in our class likes ice cream?”). In this activity, students learn that a statistical question is one that can be answered by using data in which variability is expected.

For example, the question, “What is the favorite subject of students in my class?” is a statistical question because we need data about favorite subjects and we can expect students to have different preferences. The question, “What is the counselor’s favorite subject?” is not a statistical question because it can be answered by collecting a single data value. Even if multiple responses were collected, the responses are not expected to show variability.

As students analyze and discuss examples and non-examples of statistical questions, listen for groups who distinguish the two in terms of the data needed to answer the questions. For example, some questions may require collecting data that will probably show some variability while other questions may have only a single response. Invite them to share later.

Addressing

- 6.SP.A.1

Instructional Routines

- MLR3: Clarify, Critique, Correct

Launch

Arrange students in groups of 3–4. Give students 1–2 minutes of quiet time to study the examples and non-examples of statistical questions and then 4–5 minutes to discuss with their group how the two sets are different and generate a rough definition of statistical questions. Pause the class for a discussion about their work and to review the concept of “variability” before having students complete the rest of the activity. Pause the class after the first question.

Set up a two-column table that can be displayed for all to see. Use the two table columns to record students’ observations about characteristics of statistical and non-statistical questions during discussion.

Invite groups to share their observations and record them for all to see. Be sure that the class hears from students who distinguish statistical and non-statistical questions in terms of the data needed to answer them. If not mentioned by students, highlight that answering all three statistical questions requires data, and that each data set will most likely have variability. If not mentioned by students, explain that we use the term variability to describe data sets in which not every data value is the same. In contrast, finding out the color of the principal’s car, whether Elena has a cell phone, and Diego’s reading preference does not require data, or any data collected are not expected to vary.

Access for Students with Disabilities

Representation: Develop Language and Symbols. Display or provide charts with symbols and meanings. Add new examples and non-examples of statistical questions to the visual display.

Supports accessibility for: Conceptual processing; Memory

Anticipated Misconceptions

Students might think that if the response to a question requires counting or some kind of analysis then the question is statistical. Though statistical questions do require analysis, help students see that the starting point for distinguishing a statistical question is to see whether the data used to answer it have variability, which would then determine if analysis is called for. (In other words, a data set that shows no variability—i.e. has the same value for all data points, or has only a single data point—would not require analysis.)

Student Task Statement

These three questions are examples of **statistical questions**:

- What is the most common color of the cars in the school parking lot?
- What percentage of students in the school have a cell phone?
- Which kind of literature—fiction or nonfiction—is more popular among students in the school?

1. Study the examples and non-examples. Discuss with your partner:

- a. How are the three statistical questions alike? What do they have in common?
- b. How are the three non-statistical questions alike? What do they have in common?
- c. How can you find answers to the statistical questions? How about answers to non-statistical questions?
- d. What makes a question a statistical question?

Pause here for a class discussion.

2. Read each question. Think about the data you might collect to answer it and whether you expect to see **variability** in the data. Complete each blank with “Yes” or “No.”

- a. How many cups of water do my classmates drink each day?

- Is variability expected in the data? _____
- Is the question statistical? _____
- b. Where in town does our math teacher live?
 - Is variability expected in the data? _____
 - Is the question statistical? _____
- c. How many minutes does it take students in my class to get ready for school in the morning?
 - Is variability expected in the data? _____
 - Is the question statistical? _____
- d. How many minutes of recess do sixth-grade students have each day?
 - Is variability expected in the data? _____
 - Is the question statistical? _____
- e. Do all students in my class know what month it is?
 - Is variability expected in the data? _____
 - Is the question statistical? _____

Student Response

1. No answer required. Sample explanations:
 - The statistical questions are all about a group (a group of cars, or a group of students). Their answers would require some work to figure out (counting, comparing, calculating, etc.)
 - The non-statistical questions are all about an individual. Their answers seem to be pretty straightforward and can be easily found.
 - To answer statistical questions some research and data collection would be needed (e.g., counting the number of cars of each color; asking each student in the school if they have a cell phone or if they prefer fiction or nonfiction). To answer the non-statistical questions one could simply ask the principal, Elena, or Diego the relevant question.
 - A statistical question requires gathering and studying data to answer.
2.
 - a. Cups of water: Yes. Yes.
 - b. Where math teacher lives: No. No.
 - c. Preparation time: Yes. Yes.
 - d. Recess time: No. No.

e. What month it is: No. No.

Activity Synthesis

Briefly poll students on their responses to the second set of questions. Be sure students understand that a question is statistical if we need data to answer it, and the data are expected to have variability.

Students may have trouble recognizing statistical questions in some cases. Here are two examples to ask students:

- “Who is the tallest person in the class?” may appear to be non-statistical, either because we might be able to visually tell who is tallest (so data seem unnecessary), or because we believe that everyone in the class would give the same answer (so no variability is expected).
- “How long is the longest river in the United States?” may appear to be non-statistical because we could readily look up the answer.

While the tallest person may be obvious in some classrooms, it is helpful to remember that this is not true in all classrooms. The students in a class are often close (but not identical) in height, and finding out who is tallest requires analyzing different heights. So the question, “Who is the tallest person in the class?” is generally a statistical question. In particular, because data may need to be collected in order to answer the question by measuring the heights of students.

Likewise, while the longest river in the country can be easily researched, it is helpful to remember that this was not always the case. The answer may be considered a fact now, but the question was once a statistical question—at some point, lengths of rivers were collected and compared in order to answer it.

To tell statistical questions from non-statistical ones, it is useful to look closely at the context of the questions and what it takes to answer them.

Access for English Language Learners

Reading, Representing, Conversing: MLR3 Clarify, Critique, Correct. To help students make sense of “What makes a question a statistical question?”, offer an incorrect response such as “A statistical question is one in which more than one answer is possible and a non-statistical question has only one possible answer.” Invite students to offer a correct response by asking “What language might you add or change to make this statement more accurate?” Improved responses should include the term “variability” with an explanation of what variability means. This will support student understanding of variability through language production.

Design Principle(s): Maximize meta-awareness; Support sense-making

2.4 Sifting for Statistical Questions

Optional: 15 minutes

This activity provides additional practice in determining what it means for a question to be statistical in nature.

In a previous lesson, students learned about variability in data and about statistical questions. Here they develop a deeper understanding of statistical questions by studying a wider range of examples and non-examples. Students sort a variety of questions and explain why they are or are not statistical. When students explain their reasoning, critique the reasoning of others, and attempt to persuade, they engage in MP3. Students also begin writing statistical questions and think about the data that might be used to answer the questions.

As students discuss and sort the questions in groups, listen for the rationales they give and notice questions that they have trouble classifying so that they could be addressed later.

Addressing

- 6.SP.A.1

Instructional Routines

- MLR8: Discussion Supports
- Take Turns

Launch

Arrange students in groups of 2. Give each group a set of pre-cut cards that contain questions (from the Instructional master). Give partners 4–5 minutes to sort the cards into three piles—Statistical, Non-statistical, and Unsure—and another 3–4 minutes to discuss their decisions with another group and then record the result of their deliberations.

If necessary, demonstrate productive ways for partners to communicate during a sorting activity. For example, partners take turns identifying a category for a card and explaining why they think it is a match. The other partner either accepts their explanation, or explains why they don't think it should be included in that category. Then they change roles for the next card to sort.

The word “typical” appears for the first time in this activity (in one of the questions to be sorted: “What is a typical number of students per class in your school?”). The term is connected to the idea of center and spread later in the unit, but is used informally here. If needed, explain that we can think of “typical” as meaning what is common or what can be expected in a given group.

Access for Students with Disabilities

Action and Expression: Develop Expression and Communication. Maintain a display of important terms and vocabulary. During the launch take time to review the following terms from previous activity that students will need to access for this activity: typical value, examples and non-examples of statistical questions.

Supports accessibility for: Memory; Language

Access for English Language Learners

Conversing: MLR8 Discussion Supports. Students should take turns selecting a card and explaining their reasoning to their partner. Display the following sentence frames for all to see: “___ is a statistical question because . . .”, “___ is not statistical question because . . .”, and “I am unsure about this question because . . .”. Encourage students to challenge each other when they disagree. This will help students clarify their reasoning about what it means for a question to be statistical in nature.

Design Principle(s): Support sense-making; Maximize meta-awareness

Anticipated Misconceptions

As students encounter additional examples of statistical questions, expect to see several areas of confusion.

- Students may confuse statistical questions with survey questions. A survey question is what we use to collect data. A statistical question is what we are trying to answer using collected data. For instance, the question, “How old are you?” is a survey question, because it can be used to gather data about the ages of people in a group being studied. The question, “Are most residents of this building older or younger than 30?” is a statistical question, because answering it requires collecting and analyzing the ages of the residents.
- Related to the potential confusion about statistical and survey questions, students may mistakenly think that the number of possible answers to a question is what defines a statistical question. In other words, they may say that the question, “Which ice cream flavor is most popular in this class?” is not statistical because there is potentially only one answer (e.g., “chocolate is most popular”). Students may need to be reminded that answering the question requires surveying the students on their ice cream preferences, and that the responses are expected to have variability.

Student Task Statement

1. Your teacher will give you and your partner a set of cards with questions. Sort them into three piles: Statistical Questions, Not Statistical Questions, and Unsure.

2. Compare your sorting decisions with another group of students. Start by discussing the two piles that your group sorted into the Statistical Questions and Not Statistical Questions piles. Then, review the cards in the Unsure pile. Discuss the questions until both groups reach an agreement and have no cards left in the Unsure pile. If you get stuck, think about whether the question could be answered by collecting data and if there would be variability in that data.
3. Record the letter names of the questions in each pile.
 - Statistical questions:
 - Non-statistical questions:

Student Response

Statistical questions: A, C, D, E, F, I, J, K

Not statistical questions: B, G, H, L

Are You Ready for More?

Tyler and Han are discussing the question, “Which sixth-grade student lives the farthest from school?”

- Tyler says, “I don’t think the question is a statistical question. There is only one person who lives the farthest from school, so there would not be variability in the data we collect.”
- Han says: “I think it is a statistical question. To answer the question, we wouldn’t actually be asking everyone, ‘Which student lives the farthest from school?’ We would have to ask each student how far away from school they live, and we can expect their responses to have variability.”

Do you agree with either one of them? Explain your reasoning.

Student Response

I agree with Han. While it is true that only one person lives the farthest from school, that information is not likely to be a known fact, so to find out who lives the farthest requires analyzing data that can be expected to vary.

Activity Synthesis

Most of the discussions happen in small groups. Bring the class together to discuss any remaining disagreements or questions. Select previously identified groups that had trouble classifying some of the cards to share their thinking and ask the class to help resolve the issue if possible.

Ask the class:

- “Do you all agree with the list of the statistical questions?”
- “If not, which one(s) are harder to distinguish? Why?”

- “Now that you have seen more examples of statistical questions, what new insights do you have about them?”

Remind students that we might classify some questions differently depending on when the questions are asked. For example, if a statistical question has been previously studied and its answer now considered a fact, the question may no longer be considered statistical later. For example, “How tall is the tallest mountain in the world?” was long ago a statistical question, when not all mountains had not been measured and topographic data had not been assembled and analyzed. We can now find this information easily, without having to collect and study data, because that work has been done and made available. This nuance may help students better distinguish statistical and non-statistical questions, and help them clarify their assumptions when classifying questions as one type or the other (MP3).

Lesson Synthesis

In addition to looking at numerical and categorical data more closely, we also think about whether the data we are studying show **variability**, and the kinds of questions the data sets could help us answer.

- “What does it mean to say that data have variability?” (Not all the data values are the same.)
- “When might we expect data to have variability?” (When the question we are trying to answer is about a feature or a characteristic of a group that has different members, and each member having different features or characteristics.)
- “When might we expect data to have no variability?” (When the question we are trying to answer is about an individual, or about a feature that all group members have in common.)
- “Give some examples of data that would show variability and some that would not show variability.”
- “What is a **statistical question**?” (A question that can be answered using data that are expected to have variability.)
- “Give some examples of statistical and non-statistical questions.”
 - “What kinds of data are needed to answer them?”
 - “How might you collect the data?”
 - “What units of measurement are involved?”

2.5 Questions about Temperature

Cool Down: 5 minutes

Addressing

- 6.SP.A.1

Student Task Statement

Here are two questions:

Question A: Over the past 10 years, what is the warmest temperature recorded, in degrees Fahrenheit, for the month of December in Miami, Florida?

Question B: At what temperature does water freeze in Miami, Florida?

1. Decide if each question is statistical or non-statistical. Explain your reasoning.
2. If you decide that a question is statistical, describe how you would find the answer. What data would you collect?

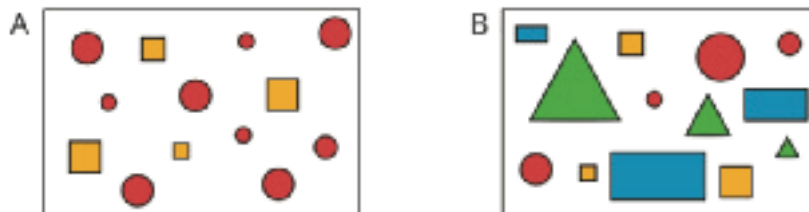
Student Response

1.
 - Question A is a statistical question. Sample reasoning: The temperature in Miami in December changes from day to day and from year to year.
 - Question B is not a statistical question. Water freezes at sea level at 32 degrees Fahrenheit. This is a known fact.
2. To answer Question A (about the warmest temperature), find the temperature records for the past ten years and look for the highest value in degrees Fahrenheit.

Student Lesson Summary

We often collect data to answer questions about something. The data we collect may show **variability**, which means the data values are not all the same.

Some data sets have more variability than others. Here are two sets of figures.



Set A has more figures with the same shape, color, or size. Set B shows more figures with different shapes, colors, or sizes, so set B has greater variability than set A.

Both numerical and categorical data can show variability. Numerical sets can contain different numbers, and categorical sets can contain different categories or types.

When a question can only be answered by using data and we expect that data to have variability, we call it a **statistical question**. Here are some examples.

- Who is the most popular musical artist at your school?
- When do students in your class typically eat dinner?

- Which classroom in your school has the most books?

To answer the question about books, we may need to count all of the books in each classroom of a school. The data we collect would likely show variability because we would expect each classroom to have a different number of books.

In contrast, the question “How many books are in your classroom?” is *not* a statistical question. If we collect data to answer the question (for example, by asking everyone in the class to count books), the data can be expected to show the same value. Likewise, if we ask all of the students at a school where they go to school, that question is not a statistical question because the responses will all be the same.

Glossary

- statistical question
- variability

Lesson 2 Practice Problems

Problem 1

Statement

Sixth-grade students were asked, "What grade are you in?" Explain why this is *not* a statistical question.

Solution

This is not a statistical question, because all of the students are in the sixth grade, so there wouldn't be any variability in the data collected.

Problem 2

Statement

Lin and her friends went out for ice cream after school. The following questions came up during their trip. Select **all** the questions that are statistical questions.

- A. How far are we from the ice cream shop?
- B. What is the most popular ice cream flavor this week?
- C. What does a group of 4 people typically spend on ice cream at this shop?
- D. Do kids usually prefer to get a cup or a cone?
- E. How many toppings are there to choose from?

Solution

["B", "C", "D"]

Problem 3

Statement

Here is a list of questions about the students and teachers at a school. Select **all** the questions that are statistical questions.

- A. What is the most popular lunch choice?
- B. What school do these students attend?
- C. How many math teachers are in the school?
- D. What is a common age for the teachers at the school?
- E. About how many hours of sleep do students generally get on a school night?
- F. How do students usually travel from home to school?

Solution

["A", "D", "E", "F"]

Problem 4

Statement

Here is a list of statistical questions. What data would you collect and analyze to answer each question? For numerical data, include the unit of measurement that you would use.

- a. What is a typical height of female athletes on a team in the most recent international sporting event?
- b. Are most adults in the school football fans?
- c. How long do drivers generally need to wait at a red light in Washington, DC?

Solution

Answers vary. Sample responses:

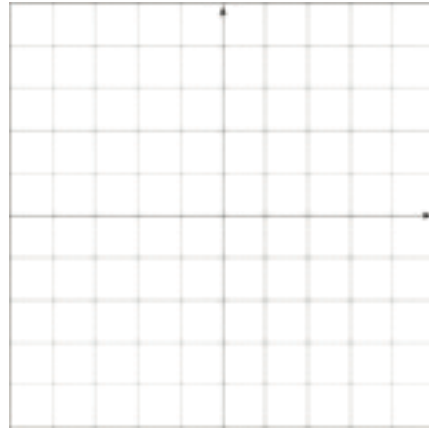
- a. Collect the heights, in inches, of female athletes on the team.
- b. Collect responses from the teachers about whether or not they are football fans.
- c. Determine the number of seconds between the moment a car stops at a red light and the moment it starts moving again for multiple cars at multiple intersections in Washington, DC.

Problem 5

Statement

Describe the scale you would use on the coordinate plane to plot each set of points. What value would you assign to each unit of the grid?

- a. $(1, -6), (-7, -8), (-3, 7), (0, 9)$
- b. $(-20, -30), (-40, 10), (20, -10), (5, -20)$
- c. $(\frac{-1}{3}, -1), (\frac{2}{3}, -1\frac{1}{3}), (\frac{-4}{3}, \frac{2}{3}), (\frac{1}{6}, 0)$



Solution

Answers vary. Sample responses:

- a. Axes labeled by multiples of 2
- b. Axes labeled by multiples of 10
- c. Axes labeled by multiples of $\frac{1}{3}$

(From Unit 7, Lesson 13.)

Problem 6

Statement

Noah's water bottle contains more than 1 quart of water but less than $1\frac{1}{2}$ quarts. Let w be the amount of water in Noah's bottle, in quarts. Select **all** the true statements.

- A. w could be $\frac{3}{4}$.
- B. w could be 1.
- C. $w > 1$
- D. w could be $\frac{4}{3}$.
- E. w could be $\frac{5}{4}$.
- F. w could be $\frac{5}{3}$.
- G. $w > 1.5$

Solution

["C", "D", "E"]

(From Unit 7, Lesson 9.)

Problem 7

Statement

Order these numbers from least to greatest:

$|-17|$ $|-18|$ -18 $|19|$ 20

Solution

$-18, |-17|, |-18|, |19|, 20$

(From Unit 7, Lesson 7.)

Section: Dot Plots and Histograms

Lesson 3: Representing Data Graphically

Goals

- Comprehend the word “frequency” to refer to the number of times a particular value occurs in a data set.
- Create and interpret a dot plot to answer statistical questions about a numerical data set.
- Justify (in writing) whether a dot plot is an appropriate way to display a given data set, paying attention to whether the data set is numerical or categorical.

Learning Targets

- I can describe the information presented in tables, dot plots, and bar graphs.
- I can use tables, dot plots, and bar graphs to represent distributions of data.

Lesson Narrative

In this lesson, students represent distributions of numerical (and optionally categorical) data after organizing them into frequency tables. They construct dot plots for numerical data (and bar graphs for categorical data). Using graphical representations of distributions, they continue to develop a spatial understanding of distributions in preparation for understanding the concepts of “center” and “spread” in future lessons. Students make use of the structure of dot plots (MP7) to describe distributions and draw conclusions about the data.

Alignments

Building On

- 3.MD.B.3: Draw a scaled picture graph and a scaled bar graph to represent a data set with several categories. Solve one- and two-step “how many more” and “how many less” problems using information presented in scaled bar graphs. For example, draw a bar graph in which each square in the bar graph might represent 5 pets.

Addressing

- 6.SP.A.1: Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students' ages.
- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- 6.SP.B.5.a: Reporting the number of observations.

- 6.SP.B.5.b: Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

Building Towards

- 6.SP.A.2: Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.

Instructional Routines

- MLR2: Collect and Display
- MLR8: Discussion Supports

Required Materials

Dot stickers

Small circular sticker useful for plotting points on a display.

Sticky notes

Straightedges

A rigid edge that can be used for drawing line segments. Sometimes a ruler is okay to use as a straightedge, but sometimes it is preferable to use an unruled straightedge, like a blank index card.

Required Preparation

1 sticky note and 1 dot sticker for each student. Straightedges should be made available to create dot plots.

Student Learning Goals

Let's represent data with dot plots and bar graphs.

3.1 Curious about Caps

Warm Up: 5 minutes

The purpose of this warm-up is to reinforce the distinction between statistical and non-statistical questions. Students write a statistical question and articulate why it qualifies as statistical. Students' explanations should focus on the variability in the data used to answer the question. The context will be used again in the next activity, so this also gets students familiar with the object.

Addressing

- 6.SP.A.1

Launch

Arrange students in groups of 3–4. Give students 1 minute of quiet work time. Then, tell students to share their question and reasoning within their group. Tell the groups to decide if each question is statistical or non-statistical. If the group disagrees, discuss the question further and revise it until an agreement is reached.

Student Task Statement

Clare collects bottle caps and keeps them in plastic containers.



Write one statistical question that someone could ask Clare about her collection. Be prepared to explain your reasoning.

Student Response

Answers vary. Sample questions:

- In general, how many caps fit in a container? (This would require counting the number of caps in multiple containers, and the number of caps that can fit in a container will vary).
- What is the most common diameter of the bottle caps in the collection? (The caps do not all have the same diameters, and to find out the most common diameter would require collecting measurements of the caps.)
- What is the most common bottle-cap color in the collection? (This would require tallying up the number of caps of each color.)
- How long did it typically take to fill up a container? (This would require finding out how long it takes to fill each container, and the amount of time would likely vary.)

Activity Synthesis

The purpose of the discussion is to ensure students are comfortable constructing statistical and non-statistical questions.

Ask students to share their statistical questions and reasoning. If possible, ask students to refer back to the image as they share. Record and display their responses for all to see.

To involve more students in the conversation, consider asking some of the following questions:

- “Were there any questions your group had in common?”
- “Were there any questions your group could not agree were statistical? Why?”
- “Were there any questions your group decided were non-statistical? What made that question non-statistical?”

3.2 Estimating Caps

15 minutes

In this activity, students are motivated to create a dot plot by identifying a statistical question and collecting data from the class to answer the question. When the data is initially presented, it is messy and difficult to analyze in an unorganized form. Students are then asked to choose an appropriate representation to use for the data (MP5) in a way that can be understood better by using a dot plot. Once the data is organized, students attempt to find a typical value from the data. In addition, students will be practicing their estimation skills by guessing the number of items in a jar.

Addressing

- 6.SP.B.4

Building Towards

- 6.SP.A.2

Instructional Routines

- MLR8: Discussion Supports

Launch

If any groups in the warm-up asked how many caps were included in the jar, use that as a way to transition into this activity.

Tell students to keep their materials closed for this discussion. Give each student a sticky note and a dot sticker. Display the image and questions for all to see.



- Question 1: How many caps are in the jar?
- Question 2: In this class, what is a common estimate for the number of caps in the jar?

- Question 3: What is the teacher's estimate for the number of caps in the jar?

Give students a minute of quiet think time to determine if each question is a statistical question and their reasoning. Invite students to share their responses. Once the class agrees that Question 2 is the only statistical question and has a good reason, ask students to write down an estimate for the number of caps in the jar on a sticky note and remember their answer. Next, ask them to affix their sticky notes on a wall or a board for all to see, and then use the display to answer the statistical question.

Students should recognize that the data cannot be easily interpreted in this format. Discuss other ways the estimates can be displayed so that they can be easily seen and understood. If no students suggest placing the sticky notes along a number line like a dot plot, suggest this idea to them.

Next, look for the smallest and largest estimates (either by asking students or skimming through the sticky notes). Draw or display a number line that spans those two numbers, large enough so that students could affix their dot stickers along it. Have each student put their dot sticker on the right place along the number line. If more than one person made the same estimate, the second person should put theirs higher up on the board in the same horizontal position. When the dot plot is complete, tell students to open their materials and answer the questions.

Student Task Statement

1. Write down the statistical question your class is trying to answer.
2. Look at the dot plot that shows the data from your class. Write down one thing you notice and one thing you wonder about the dot plot.
3. Use the dot plot to answer the statistical question. Be prepared to explain your reasoning.

Student Response

1. In this class, what is a common estimate for the number of caps in the jar?

2. Answers vary. Sample responses:

I notice that:

- The dots are mostly grouped around a certain number.
- There is a wide range of guesses.
- Nobody guessed fewer than 20.

I wonder:

- Why are the guesses so spread out?
- Where is the middle of the dot plot?
- Where does the real answer fit on the dot plot?

3. Answers vary depending on the class. Sample response: There are a lot of guesses near 60.

Activity Synthesis

The purpose of this discussion is to see why a dot plot can be useful to visualize and quickly understand a large amount of data.

Ask students what they noticed about the display and for any questions they had about the display. Ask students what advantages this display might have over other ways of arranging the sticky notes.

Tell students that the actual number of bottle caps in the jars is 82. Ask students how this number relates to the dot plot created by the class.

In general, the average from a group of individual guesses can be more accurate than any individual guess. This phenomenon is called “wisdom of the crowd” and is relied on for things such as reviews for products sold online.

Access for English Language Learners

Representing, Conversing: MLR8 Discussion Supports. Before the whole-class discussion, give students an opportunity to discuss what advantages a dot plot might have over other ways of arranging the sticky notes. Display sentence frames such as: “This display is useful because...”; “The information that I can see is ...”; “This display helps me visualize the data by ...” Invite students to practice using these sentence frames together before selecting 2–3 students to share with the class. This will help students make sense of how a dot plot can be used to understand a large amount of data.

Design Principle(s): Cultivate conversation; Maximize meta-awareness

3.3 Been There, Done That!

15 minutes

In the previous activity, the class created a dot plot together. In this activity, students create a dot plot on their own. The work of drawing dot plots is not new, but students are asked to describe their analysis of the data broadly and with limited scaffolding. They also learn to use the term *frequency* to describe the number of occurrences associated with each numerical value or category.

As students work and discuss, identify those who are able to describe the *distribution* of the data clearly and succinctly, as well as students who can articulate why a dot plot is an appropriate representation. Ask them to share later.

Addressing

- 6.SP.B.4

Instructional Routines

- MLR2: Collect and Display

Launch

Keep students in groups of 3–4. Explain that the term **frequency** associated with a particular item represents the number of times an item occurs in the data set. For example, in the set of numbers 1, 2, 1, 4, 5, 1, 1, the number 1 has a frequency of 4 since it appears 4 times in the data set. In the same data set, the number 2 has a frequency of 1 and the number 3 has a frequency of 0 since it does not appear in the data set at all. Additionally explain that the **distribution** refers to a description of how the data are arranged (or distributed) in the dot plot.

Give groups 5–6 minutes to work together to organize the given data and draw their dot plots. Then, give them 2–3 minutes of quiet time to analyze the dot plots and answer the last two questions, and time for a brief small-group discussion about their responses. Reconvene as a whole class afterwards.

Access for Students with Disabilities

Representation: Develop Language and Symbols. Create a display of important terms and vocabulary. Include the following terms and maintain the display for reference throughout the unit: spread, distribution and frequency.

Supports accessibility for: Memory; Language

Anticipated Misconceptions

When drawing dot plots, some students might use dots of different sizes or neglect to stack the dots in a straight column. Remind students to use uniform dots and to stack them vertically, using a straightedge as a guide, if needed. For some students, the use of graph paper may be helpful.

Student Task Statement

Priya wants to know if basketball players on a men’s team and a women’s team have had prior experience in international competitions. She gathered data on the number of times the players were on a team before 2016.

men's team

3 0 0 0 0 1 0 0 0 0 0 0

women's team

2 3 3 1 0 2 0 1 1 0 3 1

1. Did Priya collect categorical or numerical data?
2. Organize the information on the two basketball teams into these tables.

Men's Basketball Team Players

number of prior competitions	frequency (number)
0	
1	
2	
3	
4	

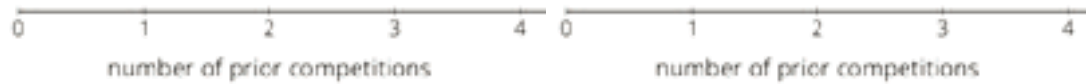
Women's Basketball Team Players

number of prior competitions	frequency (number)
0	
1	
2	
3	
4	

3. Make a dot plot for each table.

Men's Basketball Team Players

Women's Basketball Team Players



4. Study your dot plots. What do they tell you about the competition participation of:

- a. the players on the men's basketball team?
- b. the players on the women's basketball team?

5. Explain why a dot plot is an appropriate representation for Priya's data.

Student Response

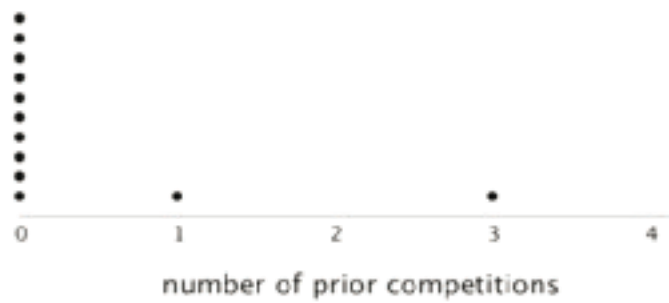
- 1. Priya collected numerical data.
- 2. Men's Basketball Team

number of prior competitions	frequency (number)
0	10
1	1
2	0
3	1
4	0

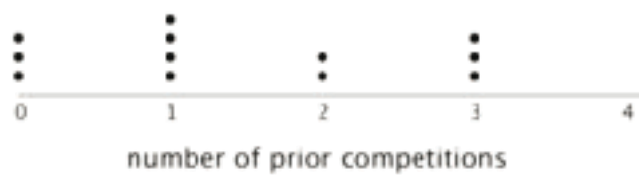
Women's Basketball Team

number of prior competitions	frequency (number)
0	3
1	4
2	2
3	3
4	0

3. Men's Team



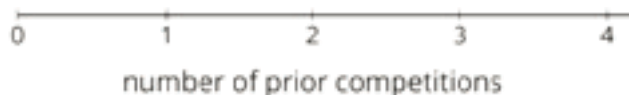
Women's Team



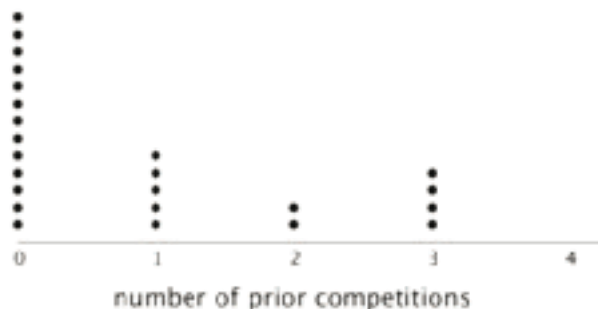
4. Answers vary. Sample responses:
- The vast majority of players on the men's team are athletes for which this is their first competition. Only 2 of the 12 are returning athletes.
 - Three-quarters of the players on the women's team are returning athletes. Only 3 players are at their first competition.
5. Answers vary. Sample response: Priya's data are numerical, so they can be organized along a number line with a dot plot.

Are You Ready for More?

Combine the data for the players on the men's and women's teams and represent it as a single dot plot. What can you say about the repeat participation of the basketball players?



Student Response



Answers vary. Sample response: About half of the basketball players did not play in a previous competition. About 25% of the players had been at one prior competition, and another 25% had 2 or 3 prior competitions.

Activity Synthesis

Select previously identified students to share their comments on the competition participation of the male and female athletes and why a dot plot is an appropriate representation. Then, invite them to compare the features and merits of the representations they used in this lesson. Discuss:

- “In what ways might bar graphs and dot plots be more useful than lists and tables?” (Bar graphs and dot plots both give us a visual snapshot of the data so that we can see patterns or anything unusual.)
- “How are bar graphs and dot plots alike? How are they different?” (They both show frequencies of data vertically—the more frequent a data value occurs, the taller the stack or bar. They are

different in that the horizontal axis of a dot plot shows numbers, while the one for a bar graph shows different labels.)

- “Could a dot plot be used to represent the first letter of the last names of the players represented on these teams? Why or why not?” (No, because a dot plot uses a number line. Each point on the number line represents a number, so where a data point is placed matters. A bar graph could be used where the categories could be shown in any order, and the width of the bars is flexible.)

Explain to students that a frequency table, bar graph, and dot plot all tell us about the distribution of a data set: each of them lists or shows all the possible values or categories in a data set and how often each one occurs. Throughout the unit, we will investigate distributions closely and use them to learn more about data and the groups or situations they represent.

Access for English Language Learners

Reading, Representing: MLR2 Collect and Display. During the discussion about similarities and differences between bar graphs and dot plots, collect students’ responses in a graphic organizer and display it for students to reference later. This will help students recognize language related to dot plots, assisting them with making decisions about appropriate use for different data contexts.

Design Principle(s): Support sense-making; Maximize meta-awareness

3.4 Favorite Summer Sports

Optional: 15 minutes

This activity is marked optional since it goes beyond the expectations of the standards. The activity helps students think about how bar graphs compare to dot plots and are useful in contrast to future lessons working with histograms.

In this activity, students organize categorical data into frequency tables and represent them as a bar graph. Neither task is new to the grade level, but students approach data analysis with a new awareness of data types and they use data to answer questions that are more open-ended. Later, they will contrast the representation for categorical data with those for numerical data. Students also learn to use the term *frequency* to describe the number of times a data value occurs.

As students draw and analyze their bar graphs, listen for the questions they might ask one another about drawing decisions. For example, they might wonder about the order in which the categories are displayed, the width of the bars, or whether certain sports might belong in the same category (e.g., swimming and diving). Also notice the arguments students make about whether a dot plot would be a suitable representation for the data.

Building On

- 3.MD.B.3

Launch

Keep students in groups of 3–4. Provide access to straightedges.

Explain to students that they will organize some data and represent them by drawing a bar graph. Ask a few students to share what they know or remember about bar graphs. Draw or display a sample bar graph for all to see. Solicit some comments about what features bar graphs have, what the bars represent, and what their heights tell us. When discussing the heights of the bars, tell students that we can use the term “frequency” for “the number of occurrences.”

Give groups 5–6 minutes to work together to organize the given data and draw their bar graph. Then, give them 2–3 minutes of quiet time to analyze the bar graph and answer the last two questions, and time for a brief small-group discussion about their responses. Reconvene as a whole class afterwards.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support development of organizational skills, check in with students within the first 2-3 minutes of work time. Check to make sure students have developed a way to keep track of their counting.

Supports accessibility for: Memory; Organization

Anticipated Misconceptions

When determining the frequencies of different sports students might lose track of their counting. If this happens, urge students to check off each sport as they account for them and then double-check their counts afterwards.

Student Task Statement

Kiran wants to know which three summer sports are most popular in his class. He surveyed his classmates on their favorite summer sport. Here are their responses.

swimming	gymnastics	track and field	volleyball
swimming	swimming	diving	track and field
gymnastics	basketball	basketball	volleyball
track and field	track and field	volleyball	gymnastics
diving	gymnastics	volleyball	rowing
track and field	track and field	soccer	swimming
gymnastics	track and field	swimming	rowing

diving

soccer

1. Did Kiran collect categorical or numerical data?
2. Organize the responses in a table to help him find which summer sports are most popular in his class.

sport	frequency

3. Represent the information in the table as a bar graph.



4.
 - a. How can you use the bar graph to find how many classmates Kiran surveyed?
 - b. Which three summer sports are most popular in Kiran's class?
 - c. Use your bar graph to describe at least one observation about Kiran's classmates' preferred summer sports.

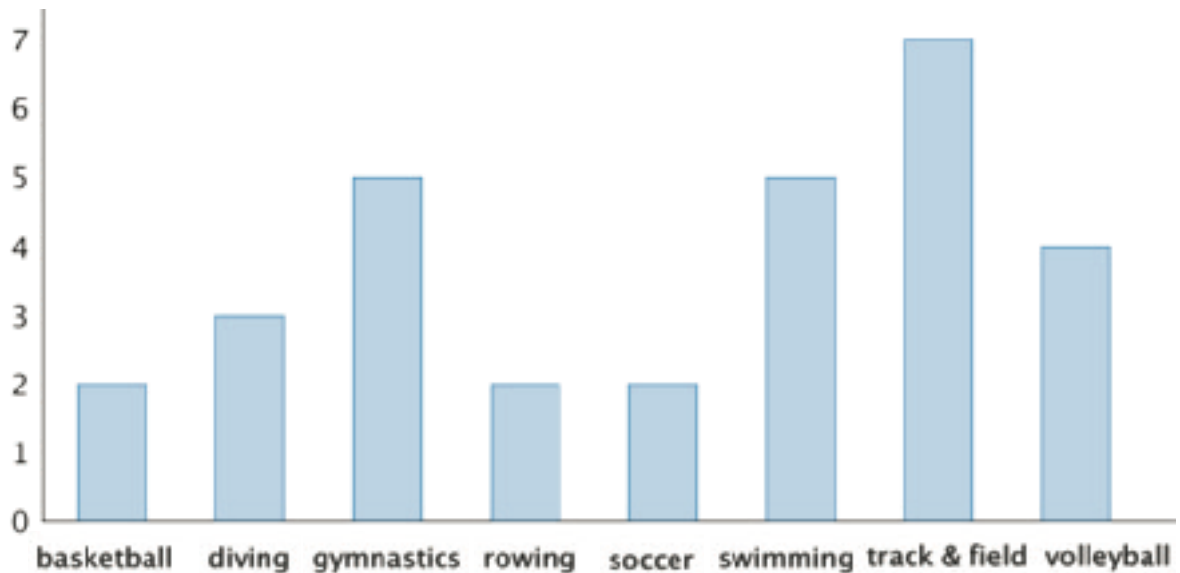
5. Could a dot plot be used to represent Kiran's data? Explain your reasoning.

Student Response

1. Kiran collected categorical data.

2.

sport	frequency
basketball	2
diving	3
gymnastics	5
rowing	2
soccer	2
swimming	5
track and field	7
volleyball	4



3.

4. a. Add the heights of each bar to find the total number of classmates surveyed. Kiran surveyed 30 classmates.
- b. Track and field is the most popular summer sport. Swimming and gymnastics tied for the second most popular summer sport.
- c. Answers vary. Sample observations: More than half the class chose one of these top three sports as their favorite.

5. No, a dot plot could not be used to represent categorical data. A dot plot is built on a number line and is to be used with numerical data.

Activity Synthesis

The purpose of the discussion is for students to recognize when it is appropriate to use a bar graph and how they compare to dot plots.

Focus the whole-class discussion on the decisions that students made when drawing their bar graph and on the last two questions. Some discussion questions:

- “How did you know which sport to represent first, second, and so on? Does the order in which the sports are listed matter?”
- “Does the width of the bars matter?”
- If some students combined two or more sports into a single category: “How might the bar graph change if we combine two sports, say swimming and diving, into a single bar?”

Select a couple of students to share their observations about favorite summer sports in Kiran's class. Ask the class to see if they agree with those observations. If they don't find those conclusions to be reasonable, ask for their reasoning and alternative conclusions.

End by polling the class about whether a dot plot could be used to represent the data. Ask a student who thinks so to explain their reasoning. Dot plots should only be used to represent numerical data along a number line and not categorical data.

Lesson Synthesis

In this lesson, we look at how to organize and represent data. When working with data, we're often particularly interested in the distribution of data.

- “What is the distribution of data?”
- “How do we tell the frequency of a data value from a data set? What about from a table?”
- “How do we tell the frequencies of different data values from a dot plot?”

If the optional activity was completed:

- “What kind of graphical representation can we use to show the distribution of categorical data? What about for numerical data?”
- “How do we tell the frequencies of different data values from a bar graph?”
- “Both the dot plot and the bar graph are built on a horizontal line. How are the horizontal lines different?”

3.5 Swimmers and Swimming Class

Cool Down: 5 minutes

Addressing

- 6.SP.B.5.a
- 6.SP.B.5.b

Student Task Statement

1. Noah gathered information on the home states of the swimmers on Team USA. He organized the data in a table. Would a dot plot be appropriate to display his data? Explain your reasoning.
2. This dot plot shows the ages of students in a swimming class. How many students are in the class?



3. Based on the dot plot, do you agree with each of the following statements? Explain your reasoning.
 - a. The class is an adult swimming class.
 - b. Half of the students are between 2 and 3 years old.

Student Response

1. No. He could use a bar graph because the home states of swimmers are categorical data.
2. 16 students are in the class.
3.
 - a. Disagree. The unit of measurements is month, and the data shows ages between 18 and 32 months, which means the students are young children, not adults.
 - b. Agree. On the number line, eight of the 16 data points, or half of the class, are placed to the right of 24 months and to the left of 36 months.

Student Lesson Summary

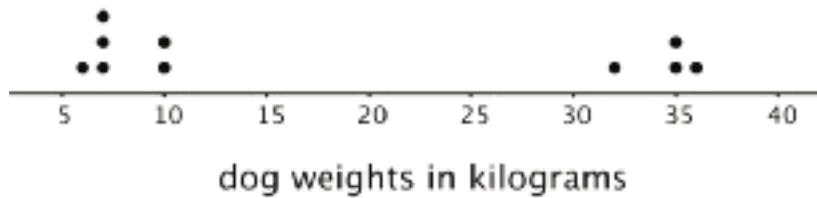
When we analyze data, we are often interested in the **distribution**, which is information that shows all the data values and how often they occur.

In a previous lesson, we saw data about 10 dogs. We can see the distribution of the dog weights in a table such as this one.

weight in kilograms	frequency
6	1
7	3
10	2
32	1
35	2
36	1

The term **frequency** refers to the number of times a data value occurs. In this case, we see that there are three dogs that weigh 7 kilograms, so “3” is the frequency for the value “7 kilograms.”

Recall that dot plots are often used to represent numerical data. Like a frequency table, a dot plot also shows the distribution of a data set. This dot plot, which you saw in an earlier lesson, shows the distribution of dog weights.

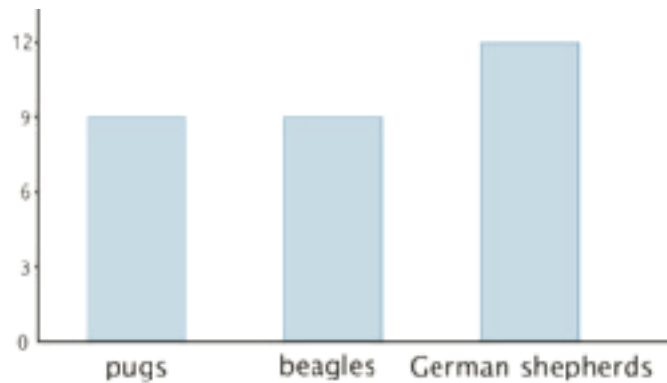


A dot plot uses a horizontal number line. We show the frequency of a value by the number of dots drawn above that value. Here, the two dots above the number 35 tell us that there are two dogs weighing 35 kilograms.

The distribution of categorical data can also be shown in a table. This table shows the distribution of dog breeds.

breed	frequency
pug	9
beagle	9
German shepherd	12

We often represent the distribution of categorical data using a bar graph.



A bar graph also uses a horizontal line. Above it we draw a rectangle (or “bar”) to represent each category in the data set. The height of a bar tells us the frequency of the category. There are 12 German shepherds in the data set, so the bar for this category is 12 units tall. Below the line we write the labels for the categories.

In a dot plot, a data value is placed according to its position on the number line. A weight of 10 kilograms must be shown as a dot above 10 on the number line.

In a bar graph, however, the categories can be listed in any order. The bar that shows the frequency of pugs can be placed anywhere along the horizontal line.

Glossary

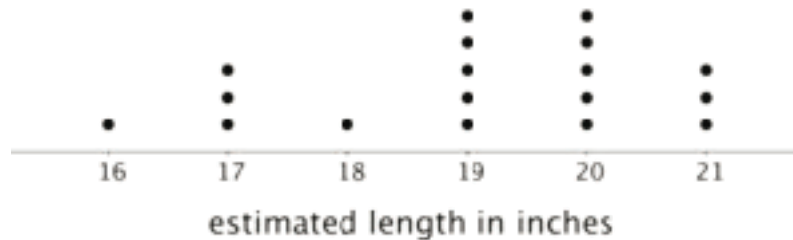
- distribution
- frequency

Lesson 3 Practice Problems

Problem 1

Statement

A teacher drew a line segment that was 20 inches long on the blackboard. She asked each of her students to estimate the length of the segment and used their estimates to draw this dot plot.



- How many students were in the class?
- Were students generally accurate in their estimates of the length of the line? Explain your reasoning.

Solution

- 18
- Responses vary. Sample response: Students tended to underestimate the length of the line, with more than half giving an estimate that was smaller than the actual length of 20. No one overestimated by more than 1 inch, but some students underestimated by as much as 3 or 4 inches.

Problem 2

Statement

Here are descriptions of data sets. Select **all** descriptions of data sets that could be graphed as dot plots.

- Class size for the classes at an elementary school
- Colors of cars in a parking lot
- Favorite sport of each student in a sixth-grade class
- Birth weights for the babies born during October at a hospital
- Number of goals scored in each of 20 games played by a school soccer team

Solution

["A", "D", "E"]

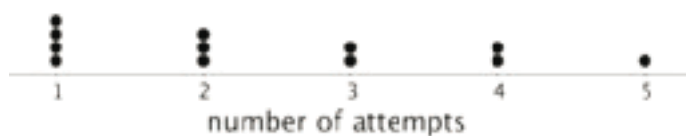
Problem 3

Statement

Priya recorded the number of attempts it took each of 12 of her classmates to successfully throw a ball into a basket. Make a dot plot of Priya's data.

1 2 1 3 1 4 4 3 1 2 5 2

Solution



Problem 4

Statement

Solve each equation.

a. $9v = 1$

b. $1.37w = 0$

c. $1 = \frac{7}{10}x$

d. $12.1 = 12.1 + y$

e. $\frac{3}{5} + z = 1$

Solution

a. $v = \frac{1}{9}$

b. $w = 0$

c. $x = \frac{10}{7}$

d. $y = 0$

e. $z = \frac{2}{5}$

(From Unit 6, Lesson 4.)

Problem 5

Statement

Find the quotients.

a. $\frac{2}{5} \div 2$

b. $\frac{2}{5} \div 5$

c. $2 \div \frac{2}{5}$

d. $5 \div \frac{2}{5}$

Solution

a. $\frac{1}{5}$

b. $\frac{2}{25}$

c. 5

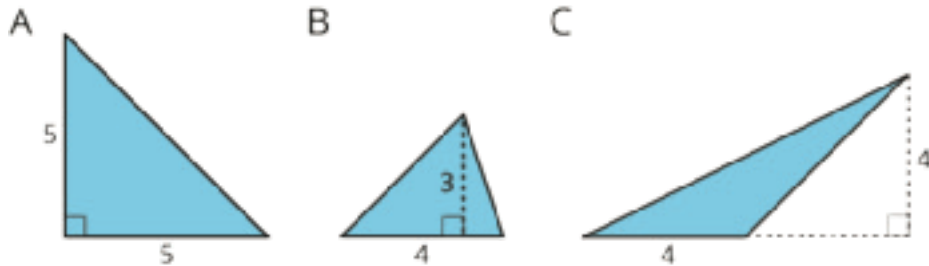
d. $\frac{25}{2}$

(From Unit 4, Lesson 11.)

Problem 6

Statement

Find the area of each triangle.



Solution

A: 12.5 square units ($5 \cdot 5 \div 2 = 12.5$), B: 6 square units ($4 \cdot 3 \div 2 = 6$), C: 8 square units ($4 \cdot 4 \div 2 = 8$)

(From Unit 1, Lesson 9.)

Lesson 4: Dot Plots

Goals

- Describe (orally and in writing) a distribution represented by a dot plot, including informal observations about its center and spread.
- Interpret a dot plot to answer (in writing) statistical questions about a data set and to identify (orally) what values are “typical” for the distribution.

Learning Targets

- I can describe the center and spread of data from a dot plot.

Lesson Narrative

In this lesson, students continue to choose appropriate representation (MP5) to display categorical and numerical data, reason abstractly and quantitatively (MP2) by interpreting the displays in context, and study and comment on features of data distributions they show. Here they begin to use the everyday meaning of the word “typical” to describe a characteristic of a group. They are also introduced to the idea of using center and spread to describe distributions generally. Planted here are seeds for the idea that values near the center of the distribution can be considered “typical” in some sense. These concepts are explored informally at this stage but will be formalized over time, as students gain more experience in describing distributions and more exposure to different kinds of distributions.

Alignments

Building On

- 6.SP.A.1: Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students' ages.

Addressing

- 6.SP.A.2: Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- 6.SP.B: Summarize and describe distributions.
- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- 6.SP.B.5.a: Reporting the number of observations.

Building Towards

- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

Instructional Routines

- MLR7: Compare and Connect
- Think Pair Share

Student Learning Goals

Let's investigate what dot plots and bar graphs can tell us.

4.1 Pizza Toppings (Part 1)

Warm Up: 5 minutes

This warm-up serves two purposes. It prompts students to determine when a dot plot is an appropriate representation and allows students to practice organizing data into a frequency table. The table completed in this activity will be used in the following activity in this lesson.

Addressing

- 6.SP.B.4

Launch

Arrange students in groups of 2. Give students 2 minutes to complete the activity with a partner, and then another 1–2 minutes for them to compare their frequency tables with another group.

Student Task Statement

Fifteen customers in a pizza shop were asked, "How many toppings did you add to your cheese pizza?" Here are their responses:

1	2	1	3	0	1	1	2	0
3	0	0	1	2	2			

1. Could you use a dot plot to represent the data? Explain your reasoning.
2. Complete the table.

number of toppings	frequency (number)
0	
1	
2	
3	

Student Response

1. Yes. The data on the number of toppings are numerical, so a dot plot would be suitable to organize them.

2.

number of toppings	frequency (number)
0	4
1	5
2	4
3	2

Activity Synthesis

Ask a couple of students to share their responses to the first question and poll the class for their agreement or disagreement.

4.2 Pizza Toppings (Part 2)

10 minutes

After organizing the data set into a frequency table in the warm-up, students now represent the information graphically. The drawing of the plot should be fairly straightforward. The emphasis here is on using a graphical representation to study and comment on the data distribution, and to reinforce how it allows us to make observations that are difficult to make by looking at lists and tables.

As students work, identify those who describe distributions not only in terms of individual categories or values (e.g. “four people did not order any toppings”) but also characterize them in broader terms (e.g. “almost all customers in the data set ordered 2 or fewer toppings”). Invite them to share during the discussion following the activity.

Addressing

- 6.SP.B

- 6.SP.B.4

Building Towards

- 6.SP.B.5.c

Instructional Routines

- Think Pair Share

Launch

Keep students in groups of 2. Give students 4–5 minutes of quiet work time and 1–2 minutes to share their responses with a partner.

Access for Students with Disabilities

Action and Expression: Develop Expression and Communication. Invite students to talk about their ideas with a partner before writing them down. Display sentence frames to support students when they explain their ideas. For example, “I notice that . . .,” or “I know ____ because . . .”

Supports accessibility for: Language; Organization

Student Task Statement

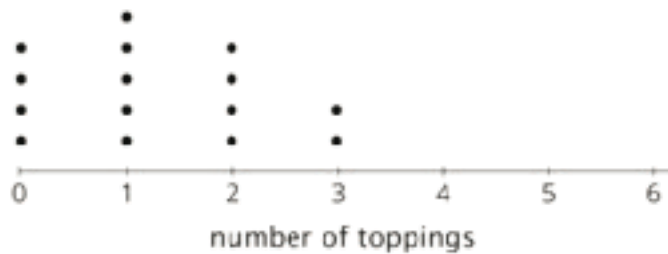
1. Use the tables from the warm-up to display the number of toppings as a dot plot. Label your drawing clearly.



2. Use your dot plot to study the distribution for number of toppings. What do you notice about the number of toppings that this group of customers ordered? Write 2–3 sentences summarizing your observations.

Student Response

- 1.



2. Answers vary. Sample response: Most customers ordered 0, 1, or 2 toppings. Nobody ordered 4 or more toppings. The most common number of toppings was 1.

Are You Ready for More?

Think of a statistical question that can be answered with the data about the number of toppings ordered, as displayed on the dot plot. Then answer this question.

Student Response

Answers vary. Sample responses:

- What is the largest number of toppings ordered by the fifteen customers? (3)
- What percentage of the customers in the data set ordered 3 toppings? (About 13%)

Activity Synthesis

Display a completed dot plot for all to see. Solicit as many observations or comments about the distribution as time permits. The goal of this discussion is to allow students to hear as many ways to describe distributions as possible. Be sure to select previously identified students to make observations that succinctly capture the distributions.

4.3 Homework Time

20 minutes

In the previous activity, students commented on distributions on a dot plot and had opportunities to hear different ways to describe distributions. In this activity, they begin to think about how to characterize a distribution as a whole, in terms of its *center* and *spread*. Students learn that we can give a general description of a distribution by identifying a location that could be the center of the data, and by noting how alike or different the data points are.

Note that at this point “center” and “spread” can only be identified and described informally, via visual observation and intuitive reasoning about how data points are distributed. The lack of precise ways to identify center and spread helps to cultivate the need for more formal measures later.

In this activity, students evaluate and critique another’s reasoning (MP3).

Building On

- 6.SP.A.1

Addressing

- 6.SP.A.2
- 6.SP.B

Building Towards

- 6.SP.A.3

Instructional Routines

- MLR7: Compare and Connect

Launch

Keep students in groups of 2. Ask students to read the question in the stem, “How many hours do you generally spend on homework each week?” Then, ask students to explain why this is a statistical question. This is a statistical question because it can be answered by collecting data, and we can expect variability in the data.

Give them 6–7 minutes of quiet time to work on the first three questions, and then 4–5 minutes to share their responses with their partner and to complete the last two questions together. Be sure to leave at least 5 minutes for a whole-class discussion.

Access for Students with Disabilities

Representation: Internalize Comprehension. Activate or supply background knowledge about calculating percentages. Allow students to use calculators to ensure inclusive participation in the activity.

Supports accessibility for: Memory; Conceptual processing

Anticipated Misconceptions

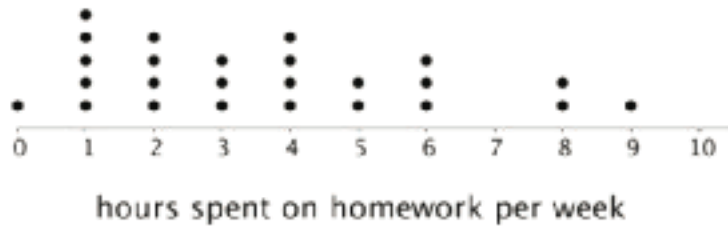
Students may not recall how to find a percentage. Remind them that a percentage can be found if we know the size of a part and that of a whole. If needed, prompt them to determine the size of the entire data set.

Student Task Statement

Twenty-five sixth-grade students answered the question: “How many hours do you generally spend on homework each week?”

This dot plot shows the number of hours per week that these 25 students reported spending on homework.





Use the dot plot to answer the following questions. For each, show or explain your reasoning.

1. What percentage of the students reported spending 1 hour on homework each week?
2. What percentage of the students reported spending 4 or fewer hours on homework each week?
3. Would 6 hours per week be a good description of the number of hours this group of students spends on homework per week? What about 1 hour per week? Explain your reasoning.
4. What value do you think would be a good description of the homework time of the students in this group? Explain your reasoning.
5. Someone said, "In general, these students spend roughly the same number of hours doing homework." Do you agree? Explain your reasoning.

Student Response

1. 20%. 5 of the students reported spending 1 hour on homework each week. 5 out of 25 is 20%.
2. 68%. 17 students reported spending 4 or fewer hours on homework each week. 17 out of 25 is 68%.
3. Answers vary. Sample response: 6 hours would not be a good value for a typical number of hours. Only 3 students in the group reported spending this much time per week doing homework. Most of the group spent a lot less. One hour would also not be a good estimate. Even though 5 students spent this much time on homework, nearly all other students spent much more.
4. Answers vary. Sample response: I would say that a good estimate would be around 3–4 hours, which is the middle of the set, with about the same number of values above and below.
5. Answers vary. Sample response: I disagree. Most of the data spreads out from 0 to 9 hours. There is a student who does not do homework at all, and there are several spending 8–9 hours per week.

Activity Synthesis

Ask a couple of students to share their thinking on what number of hours would be a good description of homework time for this group of students and why. Then, select a couple of other students to discuss how alike or different the lengths of homework time are.

Explain that sometimes it helps to describe a set of data generally, or to characterize it as a whole. One way to do that is by finding a value that describes the whole set reasonably well, or a value whose location can be considered the “center” of a distribution. Ask students:

- “Where on the dot plot would you consider the center of the data? What is the value of that center?”

Another way to characterize a data set is by describing its “spread,” or the variability in the data points. The wider the spread (the more dispersed the data points are on a dot plot), the more variable or different they are. The narrower the spread (the more clustered together), the more alike they are. Ask students:

- “Based on the dot plot, how would you describe the spread of the students’ homework time? Are the amounts of time they spend on homework alike or different?”

Tell students that we will continue to explore center and spread in upcoming lessons.

Access for English Language Learners

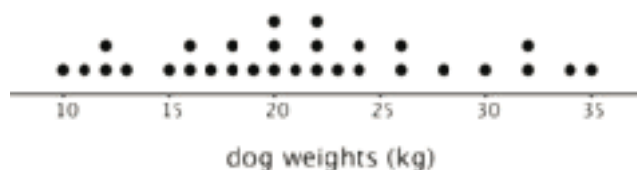
Representing, Writing: MLR7 Compare and Connect. Ask pairs of students to prepare a display for a summary statement that generalizes the the number of hours spent on homework each week using the sentence frame, “Based on this data set, spending ___ hours on homework each week is typical for 6th graders because...”. Note a handful of sentences to highlight in the discussion. Give students opportunities to compare their statements with those shared and agree or disagree with each sentence, providing reasons for their agreement or disagreement supported by the data. Give students 1-2 minutes to revise their statements prior to collecting them for display. This will help students to develop conceptual understanding for “typical.”

Design Principle(s): Maximize meta-awareness, Cultivate conversation

Lesson Synthesis

In this lesson, we look closely at data sets and try to describe what is typical for the data set. We notice that if we study the distribution of data, we can find out what is typical or common.

- What does the term “typical” mean? When someone wants to find out “a typical height of sixth-graders,” what is the information of interest?
- How do tell what is typical from a dot plot such as this one? How might we describe the characteristics of this data set on dog weights?



4.4 Family Size

Cool Down: 5 minutes

Addressing

- 6.SP.B
- 6.SP.B.5.a

Building Towards

- 6.SP.B.5.c

Student Task Statement

A group of students was asked, "How many children are in your family?" The responses are displayed in the dot plot.



1. How many students responded to the questions?
2. What percentage of the students have more than one child in the family?
3. Write a sentence that describes the distribution of the data shown on the dot plot.

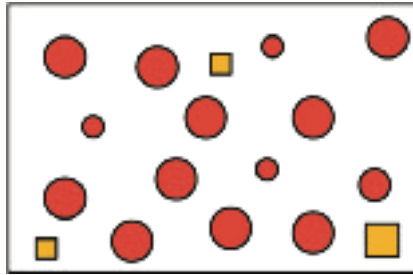
Student Response

1. There are 20 dots and each corresponds to one student in the group.
2. 75%. 15 out of 20 students answered that there are 2 or more children in the family.
3. Answers vary. Sample response: A typical number of children for this group of families is around 2, but some families had many more children than others. There are no students with more than 5 children in the family.

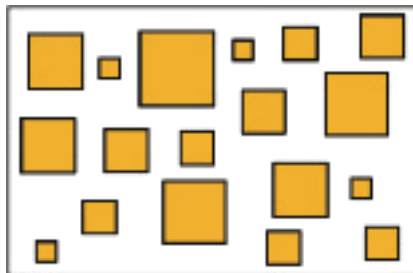
Student Lesson Summary

We often collect and analyze data because we are interested in learning what is "typical," or what is common and can be expected in a group.

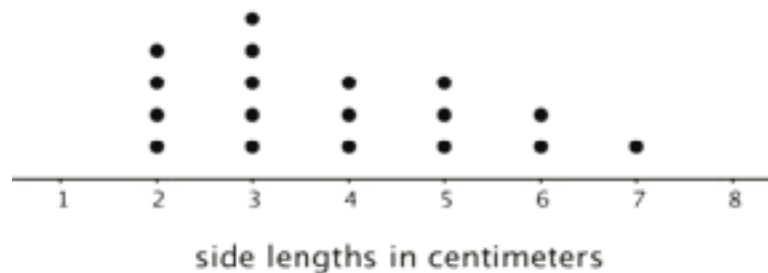
Sometimes it is easy to tell what a typical member of the group is. For example, we can say that a typical shape in this set is a large circle.



Just looking at the members of a group doesn't always tell us what is typical, however. For example, if we are interested in the side length typical of squares in this set, it isn't easy to do so just by studying the set visually.



In a situation like this, it is helpful to gather the side lengths of the squares in the set and look at their distribution, as shown in this dot plot.



We can see that many of the data points are between 2 and 4, so we could say that side lengths between 2 and 4 centimeters or close to these lengths are typical of squares in this set.

Lesson 4 Practice Problems

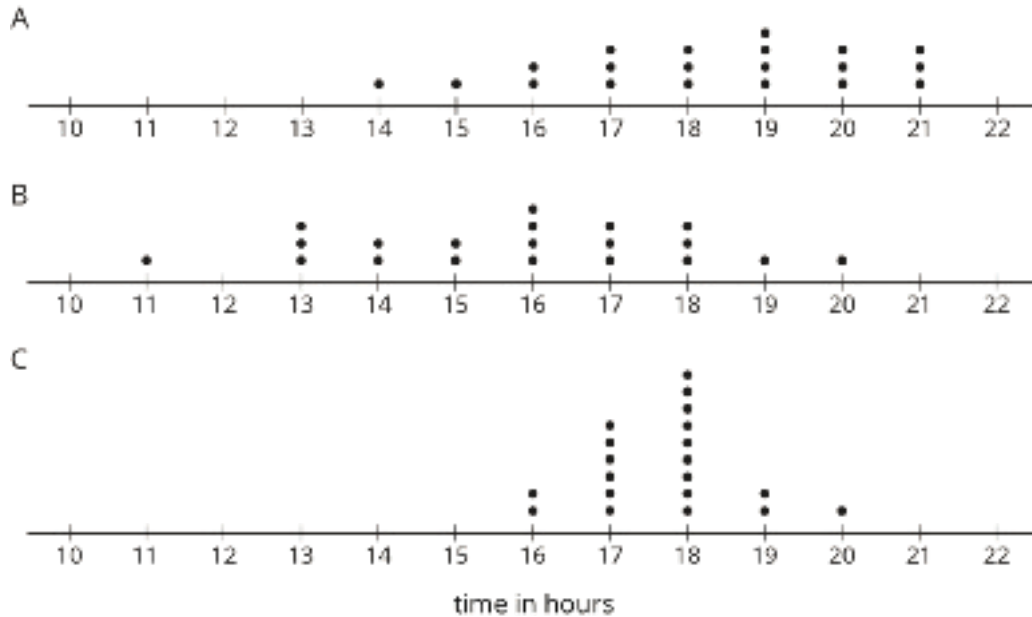
Problem 1

Statement

Clare recorded the amounts of time spent doing homework, in hours per week, by students in sixth, eighth, and tenth grades. She made a dot plot of the data for each grade and provided the following summary.

- Students in sixth grade tend to spend less time on homework than students in eighth and tenth grades.
- The homework times for the tenth-grade students are more alike than the homework times for the eighth-grade students.

Use Clare's summary to match each dot plot to the correct grade (sixth, eighth, or tenth).



Solution

A is the dot plot for eighth grade.

B is the dot plot for sixth grade.

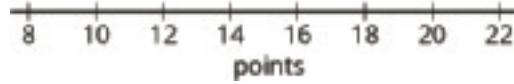
C is the dot plot for tenth grade.

Problem 2

Statement

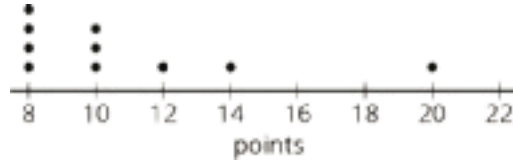
Mai played 10 basketball games. She recorded the number of points she scored and made a dot plot. Mai said that she scored between 8 and 14 points in most of the 10 games, but one

game was exceptional. During that game she scored more than double her typical score of 9 points. Use the number line to make a dot plot that fits the description Mai gave.



Solution

Answers vary. Sample response:

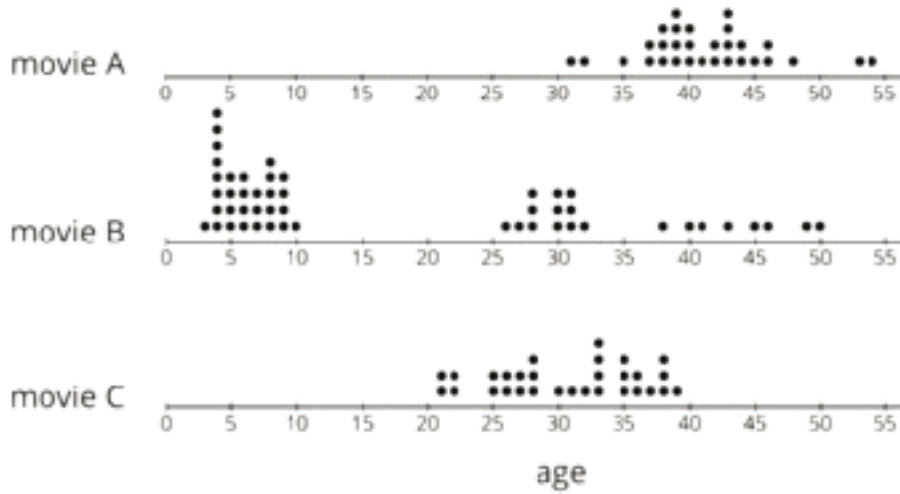


(Any dot plot that has most values between 8 and 14 and one value greater than 18 should be considered correct.)

Problem 3

Statement

A movie theater is showing three different movies. The dot plots represent the ages of the people who were at the Saturday afternoon showing of each of these movies.



- One of these movies was an animated movie rated G for general audiences. Do you think it was movie A, B, or C? Explain your reasoning.
- Which movie has a dot plot with ages that center at about 30 years?
- What is a typical age for the people who were at Movie A?

Solution

- a. Movie B. There are many people with ages between about 4 and 10, and then ages that are between 25 and 50. This movie was attended by kids and the adults that were with the kids, so it is probably the G-rated movie.
- b. Movie C.
- c. A typical age would be around 40 years.

Problem 4

Statement

Find the value of each expression.

- a. $3.727 + 1.384$
- b. $3.727 - 1.384$
- c. $5.01 \cdot 4.8$
- d. $5.01 \div 4.8$

Solution

- a. 5.111
- b. 2.343
- c. 24.048
- d. 1.04375

(From Unit 5, Lesson 13.)

Lesson 5: Using Dot Plots to Answer Statistical Questions

Goals

- Compare and contrast (orally and in writing) dot plots that represent two different data sets measuring the same quantity, paying attention to the “center” and “spread” of each distribution.
- Critique or justify (orally and in writing) claims about the center of a distribution represented on a dot plot.

Learning Targets

- I can use a dot plot to represent the distribution of a data set and answer questions about the real-world situation.
- I can use center and spread to describe data sets, including what is typical in a data set.

Lesson Narrative

In this lesson, students continue to use dot plots to develop their understanding of center and spread—by identifying values of center, describing spread, comparing centers and spreads of different distributions, and making use of the structure of the distributions (MP7) to understand them in the context of situations (MP2). In future lessons, they will make their descriptions and analyses more precise, as they learn about measures of center and spread.

Alignments

Building On

- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

Addressing

- 6.SP.A.2: Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- 6.SP.B: Summarize and describe distributions.
- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- 6.SP.B.5.b: Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

Building Towards

- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

Instructional Routines

- MLR5: Co-Craft Questions
- MLR8: Discussion Supports
- Think Pair Share

Student Learning Goals

Let's use dot plots to describe distributions and answer questions.

5.1 Packs on Backs

Warm Up: 5 minutes

In a previous lesson, students were exposed to the ideas of center and spread. Here, they begin connecting that idea informally to the word “typical” and a value that could be considered typical or characteristic of a data set by thinking about two good options and reasonings. They continue to interpret a dot plot in the context of a situation (MP2).

During the partner discussion, identify two students—one who agrees with Clare and another who agrees with Tyler—to share during the whole-class discussion (MP3).

Addressing

- 6.SP.B.4

Building Towards

- 6.SP.A.3

Instructional Routines

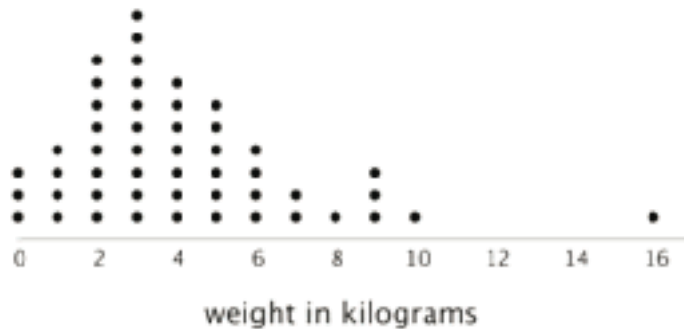
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 2 minutes of quiet work time and a minute to share their responses with a partner. Follow with a whole-class discussion.

Student Task Statement

This dot plot shows the weights of backpacks, in kilograms, of 50 sixth-grade students at a school in New Zealand.



1. The dot plot shows several dots at 0 kilograms. What could a value of 0 mean in this context?
2. Clare and Tyler studied the dot plot.
 - Clare said, "I think we can use 3 kilograms to describe a typical backpack weight of the group because it represents 20%—or the largest portion—of the data."
 - Tyler disagreed and said, "I think 3 kilograms is too low to describe a typical weight. Half of the dots are for backpacks that are heavier than 3 kilograms, so I would use a larger value."

Do you agree with either of them? Explain your reasoning.

Student Response

1. Answers vary. Sample response: A value of 0 could represent students who don't use backpacks.
2. Answers vary. Sample response:
 - I agree with Clare. There are more backpacks that are 3 kilograms than any other weights, and half of the dots are around 3 kilograms (between 2 and 4 kilograms).
 - I agree with Tyler. The middle or half-way point of the data set is between 3 and 4, around 3.5, with the same number of dots above and below, so that value would be a reasonable value to use to describe what is typical for the group.

Activity Synthesis

Ask students to share their response to the first question about data points. Record and display their responses for all to see. Ask the selected students—one who agrees with Clare and another who agrees with Tyler—to share their reasoning. Ask if anyone disagrees with both students, and if so, what value they would consider a better description of the center of the data.

Students should have a reasonable explanation for each argument they favor, but it is not necessary to confirm one way or another at this point. Tell students that we will look more closely at different ways to determine a value that is characteristic of a data set in upcoming activities.

5.2 On the Phone

15 minutes

Earlier, in the backpack example, students saw a distribution described in terms of where data points are clustered on a dot plot and which values have a large number of occurrences. The shape of that distribution was approximately symmetric. In this activity, they continue to analyze distributions in those terms, and try to identify and interpret the **center** and **spread** of a distribution that is not symmetric. The two distributions used here allow students to contrast a narrow spread and a wide spread and develop a deeper understanding of variability.

As students work, notice how students identify a general location for the center of a data set and the descriptions they use to talk about the spread (e.g., “wide,” “narrow,” or “something in between”). Identify students who connect the size of a spread to how different or alike the data points are; ask them to share later. Additionally, identify students who measure spread as the range of the entire data as well as those who use the distance to the center.

Building On

- 6.SP.B.4

Addressing

- 6.SP.A.2
- 6.SP.B

Building Towards

- 6.SP.B.5.c

Instructional Routines

- MLR5: Co-Craft Questions
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 5–6 minutes of quiet work time for the first three sets of questions, and another 4–5 minutes to share their responses and then discuss the last two questions with a partner.

Students are asked to find a percentage. If necessary, briefly review how to find a percentage.

Access for Students with Disabilities

Representation: Internalize Comprehension. Activate or supply background knowledge about calculating percentages. Allow students to use calculators to ensure inclusive participation in the activity.

Supports accessibility for: Memory; Conceptual processing

Access for English Language Learners

Representing, Writing: MLR5 Co-Craft Questions. Invite pairs of students to create one or two mathematical questions that could be answered by the data displayed in the dot plot. Note questions that vary in complexity, making sure to have students share examples that ask about percent, center, or spread, if available. Allow students to ask clarifying questions of their peers regarding how the dot plot could be used to answer the questions, if needed. This will help students to make sense of the data using informal language prior to connecting their existing understanding to formal language.

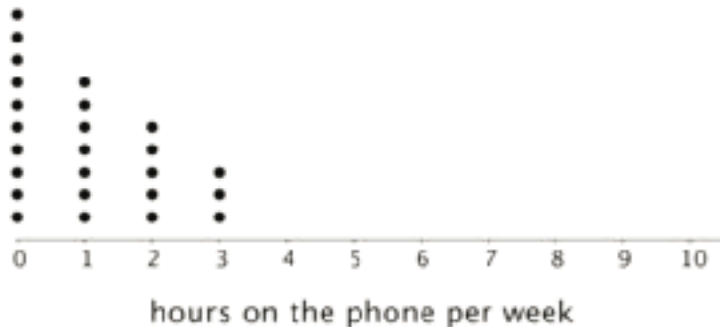
Design Principle(s): Maximize meta-awareness; Support sense-making

Anticipated Misconceptions

Students may neglect to change the rate given (from minutes per week to hours per week, or to minutes per day) and may draw incorrect conclusions as a result. Ask them to think about the unit they are using in their responses.

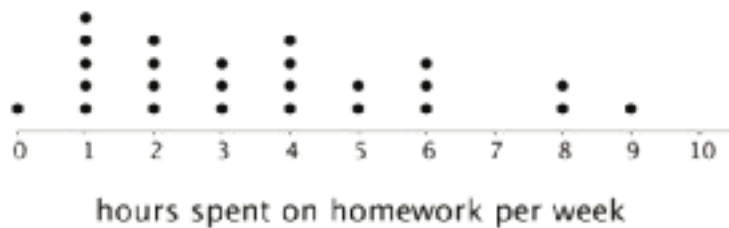
Student Task Statement

Twenty-five sixth-grade students were asked to estimate how many hours a week they spend talking on the phone. This dot plot represents their reported number of hours of phone usage per week.



1. a. How many of the students reported not talking on the phone during the week? Explain how you know.

- b. What percentage of the students reported not talking on the phone?
2.
 - a. What is the largest number of hours a student spent talking on the phone per week?
 - b. What percentage of the group reported talking on the phone for this amount of time?
3.
 - a. How many hours would you say that these students typically spend talking on the phone?
 - b. How many minutes per day would that be?
4.
 - a. How would you describe the **spread** of the data? Would you consider these students' amounts of time on the phone to be alike or different? Explain your reasoning.
 - b. Here is the dot plot from an earlier activity. It shows the number of hours per week the same group of 25 sixth-grade students reported spending on homework.



Overall, are these students more alike in the amount of time they spend talking on the phone or in the amount of time they spend on homework? Explain your reasoning.

5. Suppose someone claimed that these sixth-grade students spend too much time on the phone. Do you agree? Use your analysis of the dot plot to support your answer.

Student Response

1.
 - a. Ten students. I could tell from the number of dots showing a value of 0.
 - b. 40% of the group reported not using the phone.
2.
 - a. The highest usage is 3 hours per week.
 - b. 12% of the students reported talking on the phone for this amount of time.
3. Answers vary. Sample response:
 - a. I would say that these students typically spend about 1 hour on the phone per week.
 - b. About 9 minutes per day, because $60 \div 7$ is about 9.
4. Answers vary. Sample response:

- a. The spread is pretty small. The hours reported span from 0 to 3, with a little more than half of the values being either 0 or 1, and the rest being 2 or 3. I think the phone usage is fairly different between those who don't talk on the phone at all (40% of the group) and those who talk for 2 or 3 hours a week (40% of the group).
- b. The students are more alike in their phone usage than in the time they spend on homework. The spread of the data is much larger on the homework time dot plot, which means there is much more variability in the time they spend doing homework than the time they spend on the phone.

5. Answers vary. Sample response: I disagree. If 8.5 minute a day is typical, it is not a huge amount of time.

Activity Synthesis

The purpose of the discussion is to help students find good ways to describe a distribution based on center and spread.

Select a few students to answer the questions about the first dot plot. Tell students that the “typical” value for the data is generally considered the center. Ask, “What would you consider the center for the two dot plots shown in this activity?”

Ask students how they thought about the spread of the data. If possible, select students who thought of spread as the range of the entire data and those who thought of it as an interval around the center. Ask students to share their interpretation of the what the spread means in the context of using the phone. Make sure to include previously identified students who connect spread to how alike or different the data points are.

Tell students that distributions are generally described using the center and spread. Select a few students to describe the distributions of the two data sets shown in this activity.

5.3 Click-Clack

15 minutes

Previously students analyzed distributions to identify center and spread. In this activity, they continue to practice finding reasonable values for centers of data and describing variability. The focus, however, is on making use of the structure of distributions (MP7) to compare groups in those terms and interpreting their analyses in the context of a situation (MP2).

By comparing distributions, seeing how center and spread for the same population could change, and making sense of what these changes mean, students deepen their understanding of these concepts before learning about more formal measures of center and variability.

Building On

- 6.SP.B.4

Addressing

- 6.SP.A.2
- 6.SP.B

Building Towards

- 6.SP.A.3

Instructional Routines

- MLR8: Discussion Supports
- Think Pair Share

Launch

Give students a brief overview on keyboarding courses. Explain that these are classes designed to help people improve their typing speed and accuracy, which they may need for their jobs. Typing proficiency is usually measured in terms of number of words typed per minute; the more words typed correctly per minute, the faster or more proficient one's typing is.

Keep students in groups of 2. Give them 5–6 minutes of quiet time to work on the first two questions, and then 2–3 minutes to discuss their responses and complete the last question together.

Access for English Language Learners

Representing, Writing, Conversing: MLR8 Discussion Supports. When explaining the context of keyboarding courses, ask students to complete the sentence frame, “A person’s typing speed increases when they type (more / fewer) words per minute because....” This will help students to connect the language of the questions to the meaning of keyboarding improvement.

Design Principle(s): Support sense-making; Cultivate conversation

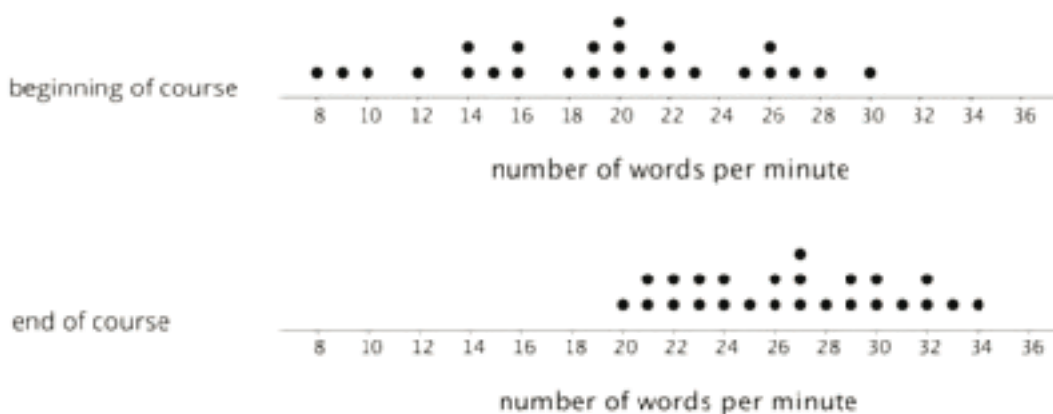
Anticipated Misconceptions

Some students might find it challenging to tell where the center of a distribution could be just by looking at a single dot plot. The idea of center might be more apparent when presented in comparative terms. For example, ask them to describe in their own words how the distribution of the first dot plot differs from that of the second dot plot. Students are likely able to say that, compared to the first dot plot, the aggregation of dots in the second dot plot is overall higher on the number line. Ask them if there is a location on each dot plot around which data points seem to congregate.

Student Task Statement

1. A keyboarding teacher wondered: “Do typing speeds of students improve after taking a keyboarding course?” Explain why her question is a statistical question.

2. The teacher recorded the number of words that her students could type per minute at the beginning of a course and again at the end. The two dot plots show the two data sets.



Based on the dot plots, do you agree with each of the following statements about this group of students? Be prepared to explain your reasoning.

- Overall, the students' typing speed did not improve. They typed at the same speed at the end of the course as they did at the beginning.
 - 20 words per minute is a good estimate for how fast, in general, the students typed at the beginning of the course.
 - 20 words per minute is a good description of the **center** of the data set at the end of the course.
 - There was more variability in the typing speeds at the beginning of the course than at the end, so the students' typing speeds were more alike at the end.
3. Overall, how fast would you say that the students typed after completing the course? What would you consider the center of the end-of-course data?

Student Response

- The question is a statistical question because data are needed to answer it, and we can expect it to have variability.
- Answers vary. Sample response:
 - Disagree. The set of dots were, as a group, placed lower on the number line at the beginning of the course and higher at the end, which means that, as a class, they were typing more words per minute or typing faster.
 - Agree. At the beginning, 20 is more or less in the middle of the set of dots. It is reasonable to use this value to describe the set in general. (Or: Disagree. There are quite a few data points that are much lower or higher than 20. It doesn't seem that 20 is a good estimate of the set.)

- c. Disagree. Almost all students were typing more than 20 words a minute, so 20 would not be a good description of the group's typing speed at the end of the course.
- d. Agree. The spread in first dot plot is wider, which means the speeds of the students were quite different or more variable. The dots on the second dot plot are closer together, or the spread is narrower, which means at the end of the course the typing speeds were less variable.
3. Answers vary. Sample response: I would consider the center of the data to be about 26. In general, the group of students were typing at about 26 words per minute after taking the course.

Are You Ready for More?

Use one of these suggestions (or make up your own). Research to create a dot plot with at least 10 values. Then, describe the center and spread of the distribution.

- Points scored by your favorite sports team in its last 10 games
- Length of your 10 favorite movies (in minutes)
- Ages of your favorite 10 celebrities

Student Response

Answers vary.

Activity Synthesis

The purpose of the discussion is for students to deepen their understanding of distributions and using the descriptions to compare two groups.

Focus the whole-class discussion on two ideas:

- The distinctions between the two distributions: Students should see that, overall, the cluster of data points have both shifted up toward a greater number of words per minute (moving its center up) and become more compressed in its spread by the end of the course. Because the center moved up in location, the value we use to describe that center would also increase.
- What the changes in the center and spread tell us in this situation: Students should recognize that a higher center means that, overall, the group has improved in their typing speed. They should see that a narrower spread at the end of the course suggests that there's now less variability in the typing speeds of different students (compared to a much larger variability initially).

Access for Students with Disabilities

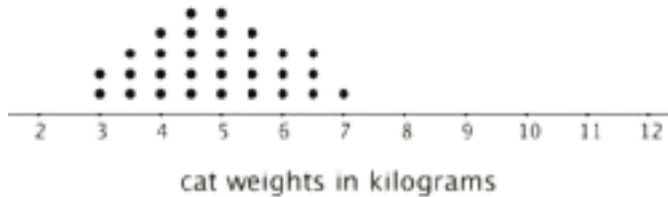
Representation: Internalize Comprehension. Use color coding and annotations to highlight the distinctions between the two distributions. For example, make visible the changes in the center and spread, and note connections to the situation.

Supports accessibility for: Visual-spatial processing

Lesson Synthesis

In this lesson, we talk about using the center and the spread of a distribution to describe a data set.

- “What do we mean by the ‘center’ of a distribution?”
- “What do we mean by the ‘spread’ of a distribution?”
- “How does the center and spread of a distribution relate to a typical value that could represent the group?”
- “How can we see center and spread in a dot plot like this?”



5.4 Packing Tomatoes

Cool Down: 5 minutes

Building On

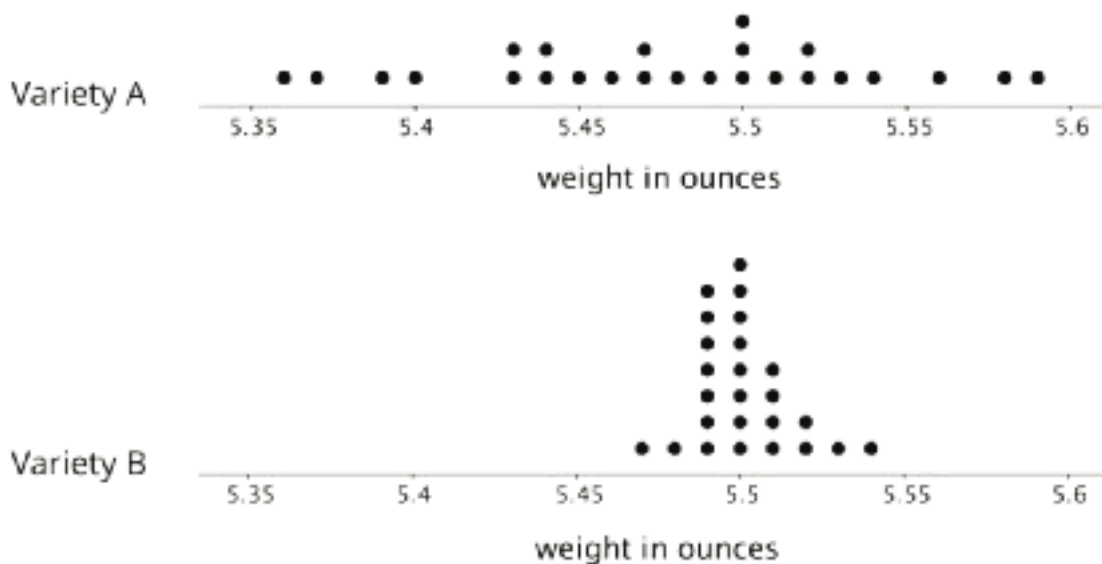
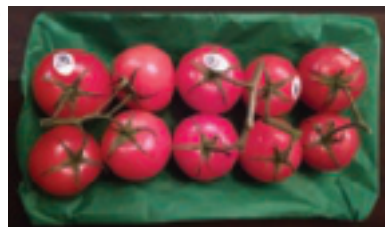
- 6.SP.B.4

Addressing

- 6.SP.A.2
- 6.SP.B.5.b

Student Task Statement

A farmer sells tomatoes in packages of ten. She would like the tomatoes in each package to all be about the same size and close to 5.5 ounces in weight. The farmer is considering two different tomato varieties: Variety A and Variety B. She weighs 25 tomatoes of each variety. These dot plots show her data.



1. What would be a good description for the weight of Variety A tomatoes, in general?
What about for the weight of Variety B tomatoes, in general?
2. Which tomato variety should the farmer choose? Explain your reasoning.

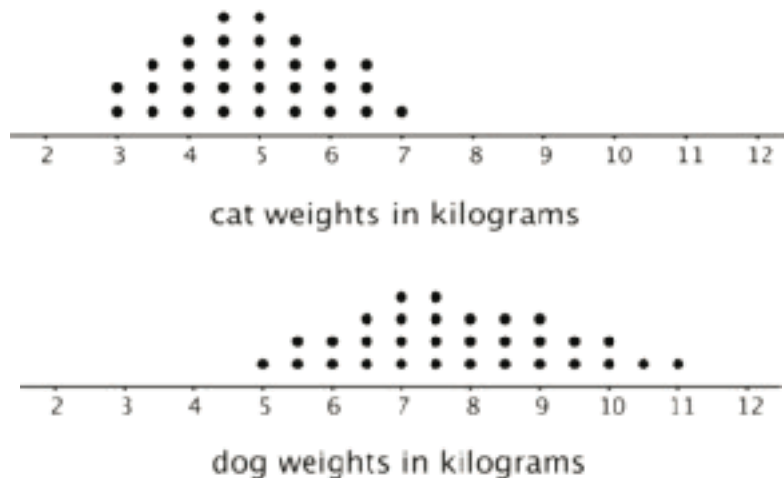
Student Response

1. In general, Variety A tomatoes are about 5.49 ounces and Variety B tomatoes are about 5.5 ounces.
2. She should choose Variety B. Sample reasoning: The two varieties of tomatoes have about the same typical weight, in general, but there was much less variation in Variety B tomato. The weights were much more consistent than the weights for Variety A, so the tomatoes are more likely to be the same size and closer to 5.5 ounces in weight.

Student Lesson Summary

One way to describe what is typical or characteristic for a data set is by looking at the **center** and **spread** of its distribution.

Let's compare the distribution of cat weights and dog weights shown on these dot plots.



The collection of points for the cat data is further to the left on the number line than the dog data. Based on the dot plots, we may describe the center of the distribution for cat weights to be between 4 and 5 kilograms and the center for dog weights to be between 7 and 8 kilograms.

We often say that values at or near the center of a distribution are typical for that group. This means that a weight of 4–5 kilograms is typical for a cat in the data set, and weight of 7–8 kilograms is typical for a dog.

We also see that the dog weights are more spread out than the cat weights. The difference between the heaviest and lightest cats is only 4 kilograms, but the difference between the heaviest and lightest dogs is 6 kilograms.

A distribution with greater spread tells us that the data have greater variability. In this case, we could say that the cats are more similar in their weights than the dogs.

In future lessons, we will discuss how to measure the center and spread of a distribution.

Glossary

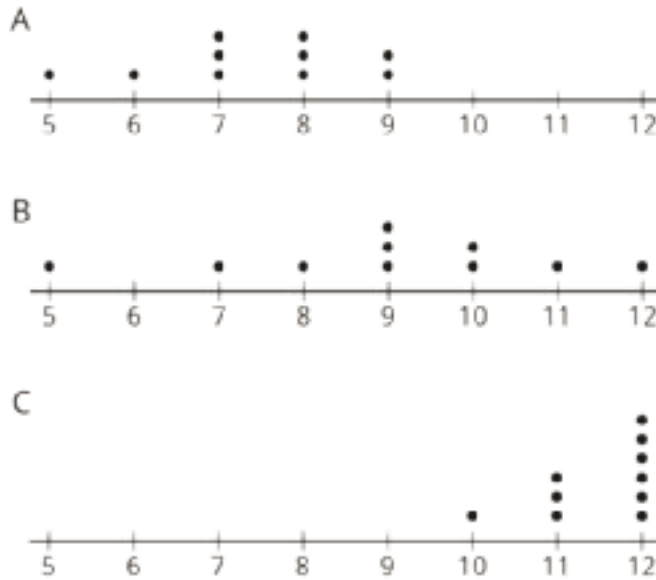
- center
- spread

Lesson 5 Practice Problems

Problem 1

Statement

Three sets of data about ten sixth-grade students were used to make three dot plots. The person who made these dot plots forgot to label them. Match each dot plot with the appropriate label.



- A. Dot Plot A
- B. Dot Plot B
- C. Dot Plot C

1. Ages in years
2. Numbers of hours of sleep on nights before school days
3. Numbers of hours of sleep on nights before non-school days

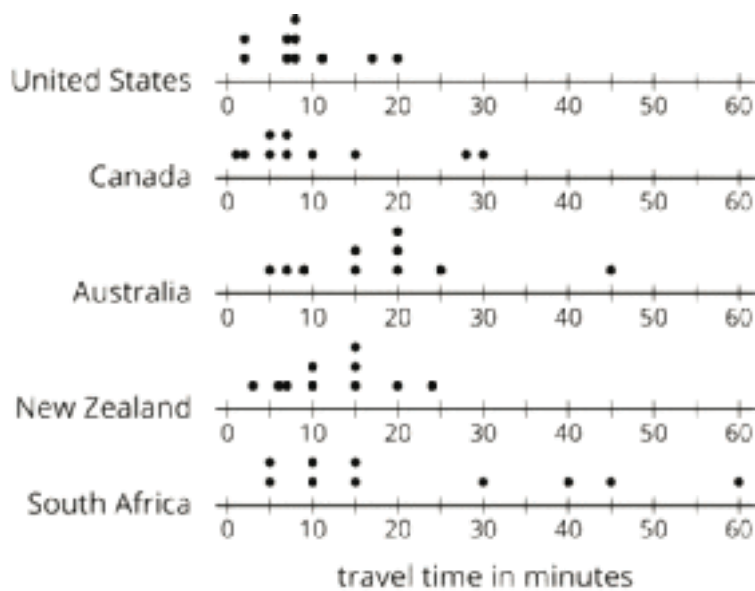
Solution

- A: 2
- B: 3
- C: 1

Problem 2

Statement

The dot plots show the time it takes to get to school for ten sixth-grade students from the United States, Canada, Australia, New Zealand, and South Africa.



- List the countries in order of *typical travel times*, from shortest to longest.
- List the countries in order of *variability in travel times*, from the least variability to the greatest.

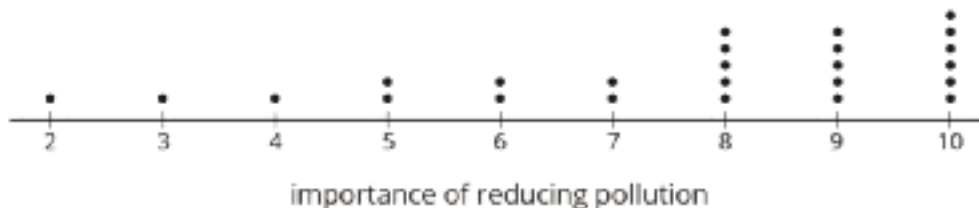
Solution

- U.S., Canada, New Zealand, Australia, South Africa. (The centers for Canada and New Zealand are close, so students may have trouble ordering them. That is acceptable and will be used later to motivate the need for a numerical measure of center, e.g., mean or median.)
- U.S., New Zealand, Australia, Canada, South Africa. (The spread for the U.S. and New Zealand and the spreads for Canada and Australia are close, so students may have trouble ordering them. That is acceptable and related to the need for numerical measures of variability, e.g., mean absolute deviation or interquartile range, which is the topic of later lessons.)

Problem 3

Statement

Twenty-five students were asked to rate—on a scale of 0 to 10—how important it is to reduce pollution. A rating of 0 means “not at all important” and a rating of 10 means “very important.” Here is a dot plot of their responses.



Explain why a rating of 6 is *not* a good description of the center of this data set.

Solution

Responses vary. Sample response: Although 6 is halfway between the largest and the smallest numbers in the data set, most of the values in the data set are larger than 6. Only 7 of the 25 values are less than or equal to 6, but 20 of the data values are greater than or equal to 6.

Problem 4

Statement

Tyler wants to buy some cherries at the farmer's market. He has \$10 and cherries cost \$4 per pound.

- If c is the number of pounds of cherries that Tyler can buy, write one or more inequalities or equations describing c .
- Can 2 be a value of c ? Can 3 be a value of c ? What about -1? Explain your reasoning.
- If m is the amount of money, in dollars, Tyler can spend, write one or more inequalities or equations describing m .
- Can 8 be a value of m ? Can 2 be a value of m ? What about 10.5? Explain your reasoning.

Solution

- Answer varies. Sample response: The inequality $c < 2.5$, or $c = 2.5$, says that Tyler cannot spend more than \$10 for the cherries. The inequality $c > 0$ means that Tyler actually buys some cherries.
- Yes, 2 can be a solution because 2 pounds of cherries cost \$8. No, 3 could not be a solution because Tyler could not buy 3 pounds of cherries as they would cost \$12 and he only has \$10. No, -1 could not be a solution because it does not make sense for Tyler to buy -1 pounds of cherries.
- Answers vary. Sample response: $m < 10$ or $m = 10$ (Tyler can spend \$10 or less), and $m > 0$ (Tyler spends some money).
- Yes, 8 and 2 can both be values of m . The value 8 means Tyler buys 2 pounds of cherries ($2 \cdot 4 = 8$), and 2 means he buys $\frac{1}{2}$ pound of cherries ($\frac{1}{2} \cdot 4 = 2$). No, 10.5 cannot be a value of m in this case. Even though 10.5 is greater than 0, m must also be 10 or less.

(From Unit 7, Lesson 10.)

Lesson 6: Interpreting Histograms

Goals

- Compare and contrast (orally) dot plots and histograms in terms of how useful they are for answering different statistical questions.
- Create a histogram to represent a data set.
- Interpret a histogram to answer (in writing) statistical questions about a data set.

Learning Targets

- I can recognize when a histogram is an appropriate graphical display of a data set.
- I can use a histogram to get information about the distribution of data and explain what it means in a real-world situation.

Lesson Narrative

In this lesson students are introduced to **histograms**. They learn that, like a dot plot, a histogram can be used to show the distribution of a numerical data set, but unlike a dot plot, a histogram shows the frequencies of groups of values, rather than individual values. Students analyze the structures of dot plots and histograms displaying the same data sets and determine what information is easier to understand from each type of display (MP7). Students read and interpret histograms in context (MP2) to prepare them to create a histogram.

Alignments

Addressing

- 6.SP.A.1: Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students' ages.
- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- 6.SP.B.5.b: Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

Instructional Routines

- MLR2: Collect and Display
- MLR8: Discussion Supports

Required Materials

Straightedges

A rigid edge that can be used for drawing line segments. Sometimes a ruler is okay to use as a

straightedge, but sometimes it is preferable to use an unruled straightedge, like a blank index card.

Student Learning Goals

Let's explore how histograms represent data sets.

6.1 Dog Show (Part 1)

Warm Up: 5 minutes

The purpose of this warm-up is to connect the analytical work students have done with dot plots in previous lessons with statistical questions. This activity reminds students that we gather, display, and analyze data in order to answer statistical questions. This work will be helpful as students contrast dot plots and histograms in subsequent activities.

Addressing

- 6.SP.A.1
- 6.SP.A.3

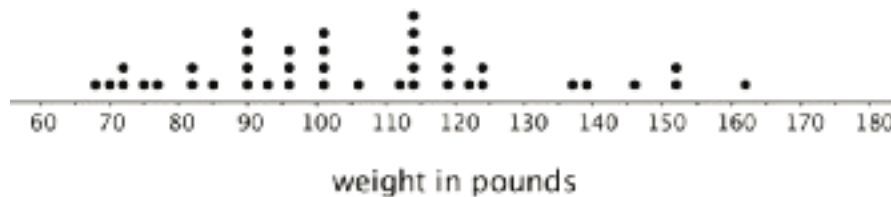
Launch

Arrange students in groups of 2. Give students 1 minute of quiet work time, followed by 2 minutes to share their responses with a partner. Ask students to decide, during partner discussion, if each question proposed by their partner is a statistical question that can be answered using the dot plot. Follow with a whole-class discussion.

If students have trouble getting started, consider giving a sample question that can be answered using the data on the dot plot (e.g., "How many dogs weigh more than 100 pounds?")

Student Task Statement

Here is a dot plot showing the weights, in pounds, of 40 dogs at a dog show.



1. Write two statistical questions that can be answered using the dot plot.
2. What would you consider a typical weight for a dog at this dog show? Explain your reasoning.

Student Response

1. Answers vary. Sample questions:

- How many dogs weigh exactly 70 pounds?
- How many dogs weigh more 80 pounds but less than 150 pounds?
- How much does the heaviest dog at the dog show weigh?
- How many times as heavy as the lightest dog is the heaviest dog?
- How alike or different are the weights of the dog at the show?

2. Answers vary. Sample responses:

- About 114 pounds, because the largest percentage of the dots are at 114, and it seems to be about where the center of the data is.
- About 100 pounds, because about half of the dogs are 100 pounds or lighter, and half are heavier than 100 pounds.

Activity Synthesis

Ask students to share questions they agreed were statistical questions that could be answered using the dot plot. Record and display their responses for all to see. If there is time, consider asking students how they would find the answer to some of the statistical questions.

Ask students to share a typical weight for a dog at this dog show and why they think it is typical. Record and display their responses for all to see. After each student shares, ask the class if they agree or disagree.

6.2 Dog Show (Part 2)

10 minutes

This activity introduces students to **histograms**. By now, students have developed a good sense of dot plots as a tool for representing distributions. They use this understanding to make sense of a different form of data representation. The data set shown on the first histogram is the same one from the preceding warm-up, so students are familiar with its distribution. This allows them to focus on making sense of the features of the new representation and comparing them to the corresponding dot plot.

Note that in all histograms in this unit, the left-end boundary of each bin or bar is included and the right-end boundary is excluded. For example, the number 5 would not be included in the 0–5 bin, but would be included in the 5–10 bin.

Addressing

- 6.SP.B.4
- 6.SP.B.5.b

Instructional Routines

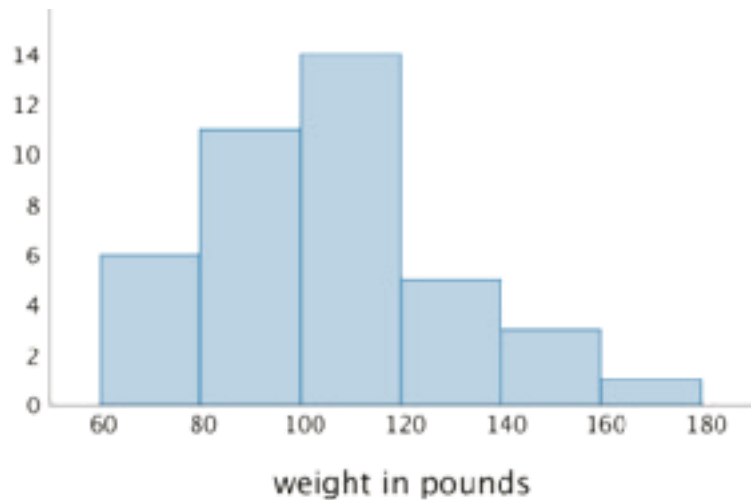
- MLR2: Collect and Display

Launch

Explain to students that they will now explore histograms, another way to represent numerical data. Give students 3–4 minutes of quiet work time, and then 2–3 minutes to share their responses with a partner. Follow with a whole-class discussion.

Student Task Statement

Here is a histogram that shows some dog weights in pounds.



Each bar includes the left-end value but not the right-end value. For example, the first bar includes dogs that weigh 60 pounds and 68 pounds but not 80 pounds.

1. Use the histogram to answer the following questions.
 - a. How many dogs weigh at least 100 pounds?
 - b. How many dogs weigh exactly 70 pounds?
 - c. How many dogs weigh at least 120 and less than 160 pounds?
 - d. How much does the heaviest dog at the show weigh?
 - e. What would you consider a typical weight for a dog at this dog show? Explain your reasoning.
2. Discuss with a partner:
 - If you used the dot plot to answer the same five questions you just answered, how would your answers be different?
 - How are the histogram and the dot plot alike? How are they different?

Student Response

1.
 - a. 23 dogs.
 - b. Unknown. It could be anywhere between 0 and 6.
 - c. 8 dogs.
 - d. The exact weight cannot be determined, but it weighs at least 160 pounds but less than 180 pounds.
 - e. Answers vary. Sample response: Around 100 pounds. The largest percentage (35%) of the weights fall in the third bar (at least 100 pounds and less than 120 pounds), and it is approximately the middle of the data.
2. Answers vary. Sample responses:
 - They are alike in that they are both built on number lines, show the same total number of data values, and show how the values are spread out. They are different in that the dot plot shows individual data points and the histogram groups the data points together.
 - With the dot plot we can see the values of individual points and tell how many there are. With the histogram, we can't tell how many data points have a specific values; we only know how many points fall into a specific range.

Activity Synthesis

Ask a few students to briefly share their responses to the first set of questions to make sure students are able to read and interpret the graph correctly.

Focus the whole-class discussion on the last question. Select a few students or groups to share their observations about whether or how their answers to the statistical questions would change if they were to use a dot plot to answer them, and about how histograms and dot plots compare. If not already mentioned by students, highlight that, in a histogram:

- Data values are grouped into “bins” and represented as vertical bars.
- The height of a bar reflects the combined frequency of the values in that bin.
- A histogram uses a number line.

At this point students do not yet need to see the merits or limits of each graphical display; this work will be done in upcoming lessons. Students should recognize, however, how the structures of the two displays are different (MP7) and start to see that the structural differences affect the insights we are able to glean from the displays.

Access for Students with Disabilities

Representation: Develop Language and Symbols. Create a display of important terms and vocabulary. Include the following terms and maintain the display for reference throughout the unit: histogram.

Supports accessibility for: Memory; Language

Access for English Language Learners

Representing, Listening, Speaking: MLR2 Collect and Display. Record and display the language students use as they discuss how dot plots and histograms are alike and different. Highlight language related to individual specific data values and groups of data values, center, and spread. Keep the display visible for reference in upcoming lessons. This will help students connect unique and shared characteristics of histograms and dot plots.

Design Principle(s): Support sense-making; Maximize meta-awareness

6.3 Population of States

20 minutes

In this activity, students continue to develop their understanding of histograms. They begin to notice that a dot plot may not be best for representing a data set with a lot of variability (or where few values are repeated) or when a data set has a large number of values. Histograms may help us visualize a distribution more clearly in these situations. Students organize a data set into “bins” and draw a histogram to display the distribution.

As students work and discuss, listen for explanations for why certain questions might be easy, hard, or impossible to answer using each graphical display.

Addressing

- 6.SP.B.4

Instructional Routines

- MLR8: Discussion Supports

Launch

Give students a brief overview of census and population data, as some students may not be familiar with them. Refer to the dot plot of the population data and discuss questions such as:

- “How many total dots are there?” (51)

- “What’s the population of the state with the largest population? Do you know what state that is?” (Between 37 and 38 million. It’s California.)
- “Look at the leftmost dot. What state might it represent? Approximately what is its population?” (The leftmost dot represents Wyoming, with a population of around half a million.)
- “Do you know the approximate population of our state? Where do you think we are in the dot plot?”

Explain to students that they will now draw a histogram to represent the population data. Remind them that histograms organize data values into “bins” or groups. In this case, the bins sizes are already decided for them. Then, arrange students in groups of 3–4. Provide access to straightedges. Give students 10–12 minutes to complete the activity. Encourage them to discuss their work within their group as needed.

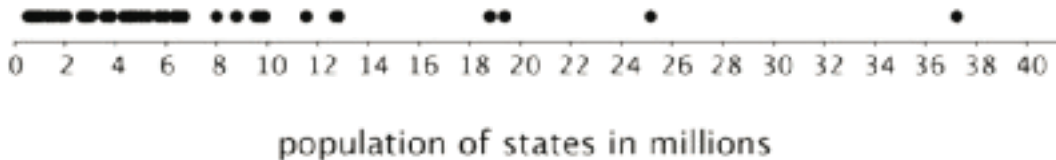
Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support development of organizational skills, check in with students within the first 2-3 minutes of work time. Check to make sure students have developed a way to keep track of counting the 2010 census data.

Supports accessibility for: Organization; Attention

Student Task Statement

Every ten years, the United States conducts a census, which is an effort to count the entire population. The dot plot shows the population data from the 2010 census for each of the fifty states and the District of Columbia (DC).



1. Here are some statistical questions about the population of the fifty states and DC. How difficult would it be to answer the questions using the *dot plot*?

In the middle column, rate each question with an E (easy to answer), H (hard to answer), or I (impossible to answer). Be prepared to explain your reasoning.

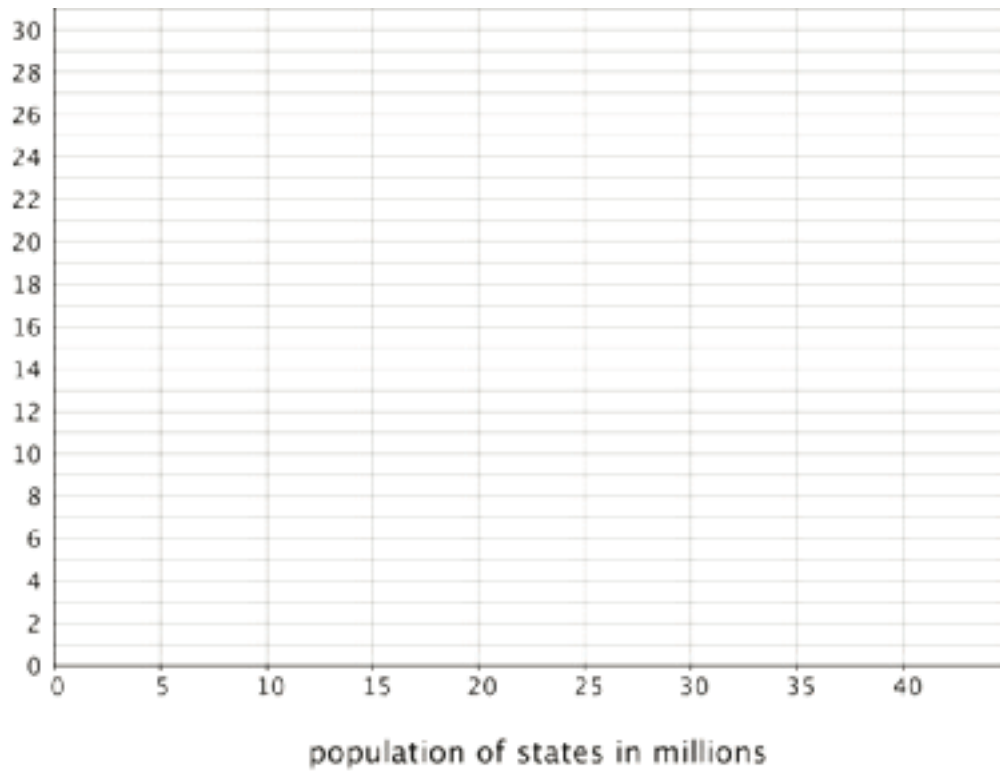
statistical question	using the dot plot	using the histogram
a. How many states have populations greater than 15 million?		
b. Which states have populations greater than 15 million?		
c. How many states have populations less than 5 million?		
d. What is a typical state population?		
e. Are there more states with fewer than 5 million people or more states with between 5 and 10 million people?		
f. How would you describe the distribution of state populations?		

2. Here are the population data for all states and the District of Columbia from the 2010 census. Use the information to complete the table.

Alabama	4.78	Illinois	12.83	Montana	0.99	Rhode Island	1.05
Alaska	0.71	Indiana	6.48	Nebraska	1.83	South Carolina	4.63
Arizona	6.39	Iowa	3.05	Nevada	2.70	South Dakota	0.81
Arkansas	2.92	Kansas	2.85	New Hampshire	1.32	Tennessee	6.35
California	37.25	Kentucky	4.34	New Jersey	8.79	Texas	25.15
Colorado	5.03	Louisiana	4.53	New Mexico	2.06	Utah	2.76
Connecticut	3.57	Maine	1.33	New York	19.38	Vermont	0.63
Delaware	0.90	Maryland	5.77	North Carolina	9.54	Virginia	8.00
District of Columbia	0.60	Massachusetts	6.55	North Dakota	0.67	Washington	6.72
Florida	18.80	Michigan	9.88	Ohio	11.54	West Virginia	1.85
Georgia	9.69	Minnesota	5.30	Oklahoma	3.75	Wisconsin	5.69
Hawaii	1.36	Mississippi	2.97	Oregon	3.83	Wyoming	0.56
Idaho	1.57	Missouri	5.99	Pennsylvania	12.70		

population (millions)	frequency
0-5	
5-10	
10-15	
15-20	
20-25	
25-30	
30-35	
35-40	

3. Use the grid and the information in your table to create a histogram.



4. Return to the statistical questions at the beginning of the activity. Which ones are now easier to answer?

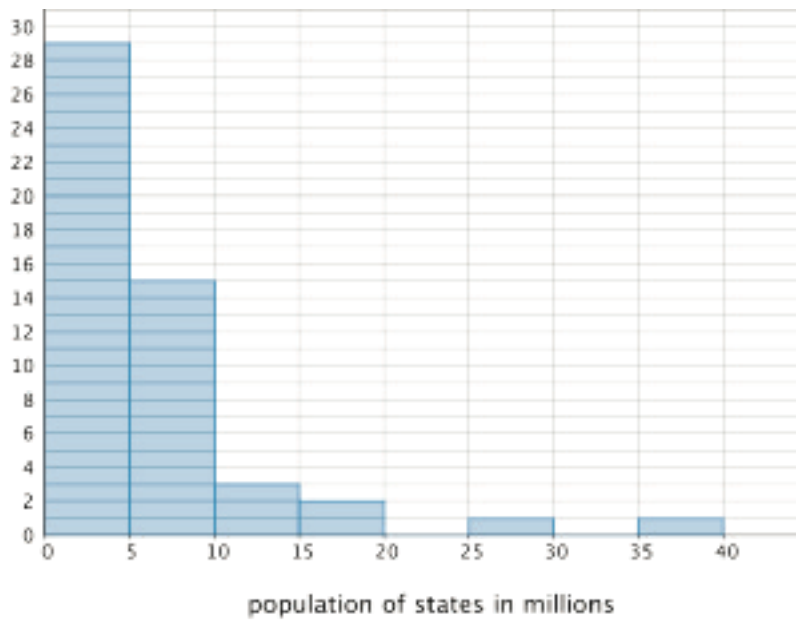
In the last column of the table, rate each question with an E (easy), H (hard), and I (impossible) based on how difficult it is to answer them. Be prepared to explain your reasoning.

Student Response

1.
 - a. Easy (E). Unless some dots are lying directly on top of one another, there are four states with a population greater than 15 million.
 - b. Impossible (I). Since the dots are not labeled, it is impossible to tell which states have a population greater than 15 million.
 - c. Impossible (I). Since the dots are so close together below 5 million, it's impossible to count how many there are.
 - d. Hard (H). Since so many dots are indistinguishable, it's hard to determine a typical state population.
 - e. Hard (H). It appears that there are more dots for populations that are less than 5 million than for those between 5 and 10 million, but we can't be sure because dots might be right on top of each other.
 - f. Hard (H). Since the dots overlap a lot, it is difficult to give a good estimate for the center and spread.

2.

population in millions	frequency
0-5	29
5-10	15
10-15	3
15-20	2
20-25	0
25-30	1
30-35	0
35-40	1



3.

4. Revisiting the questions from the first problem:

- a. Still easy (E).
- b. Still impossible (I), based on the histogram alone.
- c. Now it is easy (E) to tell how many states had a population below 5 million. It was previously impossible (I).
- d. From the histogram, we can estimate that a typical state population has fewer than 10 million, but it is hard (H) to be more precise than that at this point. It was previously hard (H).
- e. Using the histogram it is easy (E) to tell how many states have fewer than 5 million people and how many have between 5 and 10 million (there are more states in the smaller population category). It was previously hard (H).
- f. It is easier (E) to describe the data distribution more precisely because the histogram shows the population sizes in intervals of 5 million people.

Are You Ready for More?

Think of two more statistical questions that can be answered using the data about populations of states. Then, decide whether each question can be answered using the dot plot, the histogram, or both.

Student Response

Answers vary.

Activity Synthesis

Much of the discussion about how to construct histograms should have happened in small groups. Address unresolved questions about drawing histograms if they are relatively simple. Otherwise, consider waiting until students have more opportunities to draw histograms in upcoming lessons.

Focus the discussion on comparing the effectiveness of dot plots and histograms to help us answer statistical questions.

Select a few students or groups to share how their ratings of “easy,” “hard,” and “impossible,” changed when they transitioned from using dot plots to using histograms to answer statistical questions about populations of states. Then, discuss and compare the two displays more generally. Solicit as many ideas and observations as possible to these questions:

- “What are some benefits of histograms?”
- “When might histograms be preferable to dot plots?”
- “When might dot plots be preferable to histograms?”

Students should walk away from the activity recognizing that in cases when a data set has many numerical values, especially if the values do not repeat, a histogram can give us a better visualization of the distribution. In such a case, creating a dot plot can be very difficult to do including finding a scale that can meaningfully display the data while a histogram will be easier to create and display the information in a way that is easier to understand at a glance.

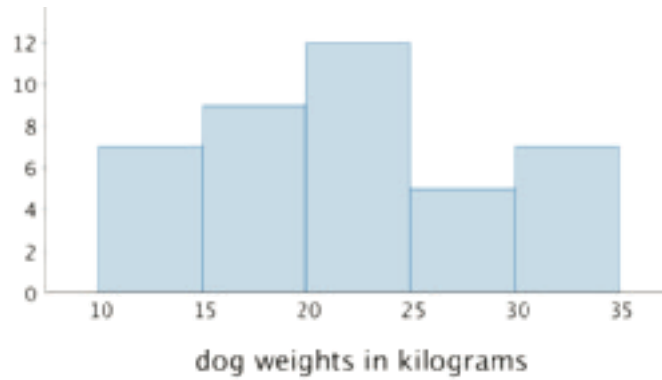
Access for English Language Learners

Representing, Writing, Conversing: MLR8 Discussion Supports. Give students sentence frames during the discussion, such as: “Histograms are easy (or hard or impossible) to use when _____, because . . .”. This will help students make decisions about the type of graph to use to display different types of data sets.

Design Principle(s): Optimize output (for generalization)

Lesson Synthesis

In this lesson, we learn about a different way to represent the distribution of numerical data—using a **histogram**. This histogram, for instance, represents the distribution for the weights of some dogs.



- “What could the smallest dog weigh? The largest?” (10 kilograms up to almost 40 kilograms)
- “What does the bar between 25 and 30 tell you?” (5 dogs weigh between 25 and just under 30 kilograms)
- “What can you say about the dogs who weigh between 10 and 20 kg?” (There are 16 total dogs in this range including 7 dogs between 10 and 15 kg and 9 between 15 and 20 kg)
- “In general, what information does a histogram allow us to see? How is it different from a dot plot?” (A bigger picture of the distribution is shown in the histogram, but some of the detail is lost when compared to a dot plot. For example, this histogram does not show the weight of any individual dogs.)
- “When might it be more useful to use a histogram than a dot plot?” (When the data is very spread out, when there are not very many data points with the same value, or when an overall idea of the distribution is more important than a detailed view.)

6.4 Rain in Miami

Cool Down: 5 minutes

Addressing

- 6.SP.B.4

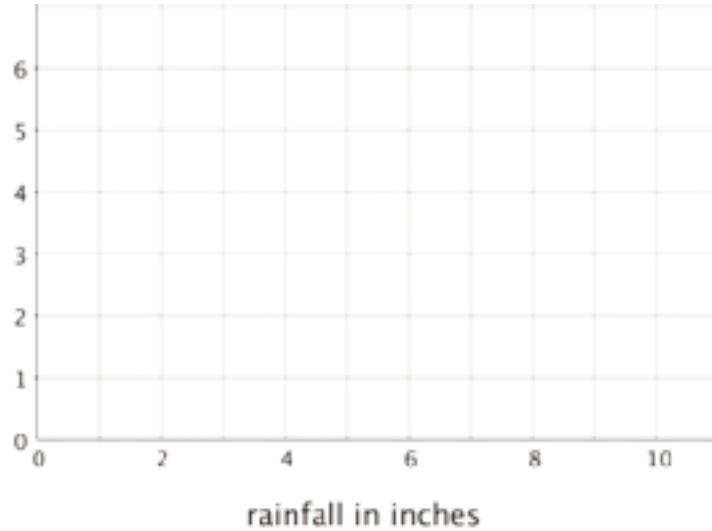
Student Task Statement

Here is the average amount of rainfall, in inches, for each month in Miami, Florida.

month	rainfall (inches)	month	rainfall (inches)
January	1.61	July	6.5
February	2.24	August	8.9
March	2.99	September	9.84
April	3.14	October	6.34
May	5.35	November	3.27
June	9.69	December	2.05

1. Complete the frequency table and use it to make a histogram.

rainfall (inches)	frequency
0-2	
2-4	
4-6	
6-8	
8-10	

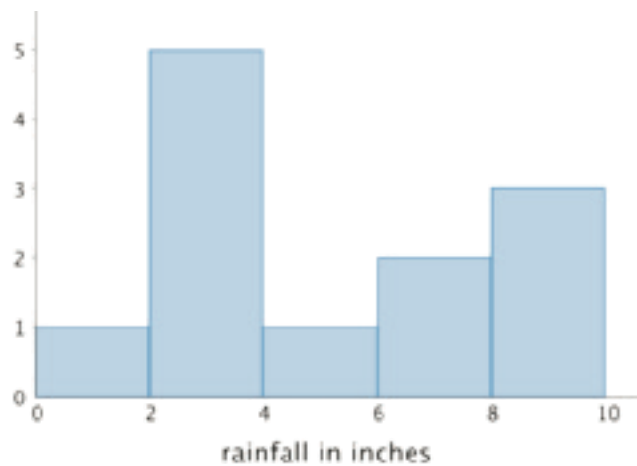


2. What is a typical amount of rainfall in one month in Miami?

Student Response

1.

rainfall (inches)	frequency
0-2	1
2-4	5
4-6	1
6-8	2
8-10	3

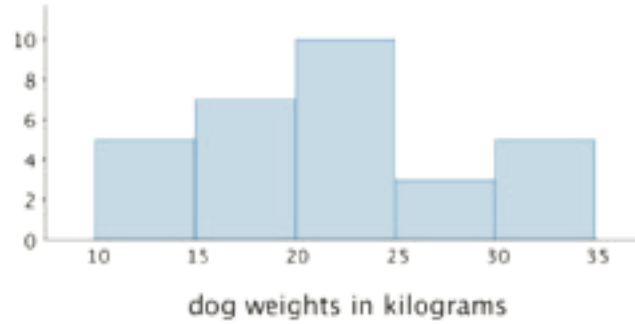
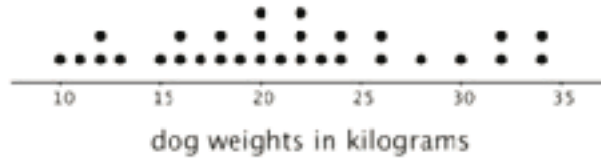


2. Answers vary. Sample response: It is difficult to say what typical amount of rainfall is. A typical month could be between 2 and 4 inches, this happens in 5 months. But for another 5 months, the rainfall is more than 6 inches, between 6 and 10 inches.

Student Lesson Summary

In addition to using dot plots, we can also represent distributions of numerical data using histograms.

Here is a dot plot that shows the weights, in kilograms, of 30 dogs, followed by a histogram that shows the same distribution.

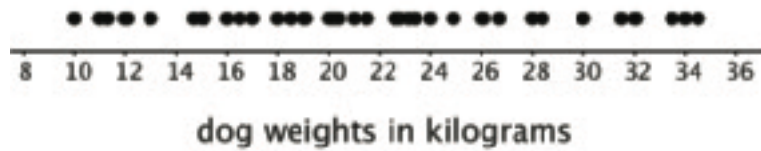


In a histogram, data values are placed in groups or “bins” of a certain size, and each group is represented with a bar. The height of the bar tells us the frequency for that group.

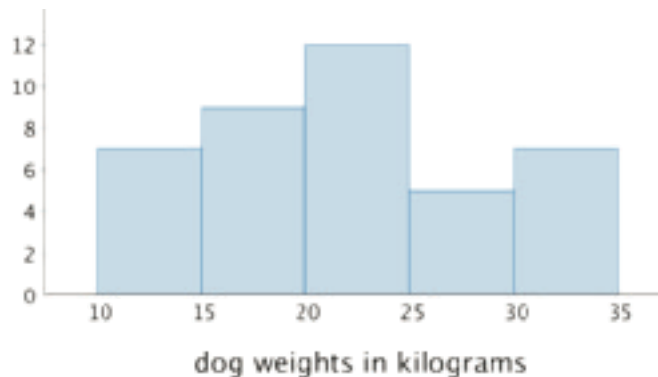
For example, the height of the tallest bar is 10, and the bar represents weights from 20 to less than 25 kilograms, so there are 10 dogs whose weights fall in that group. Similarly, there are 3 dogs that weigh anywhere from 25 to less than 30 kilograms.

Notice that the histogram and the dot plot have a similar shape. The dot plot has the advantage of showing all of the data values, but the histogram is easier to draw and to interpret when there are a lot of values or when the values are all different.

Here is a dot plot showing the weight distribution of 40 dogs. The weights were measured to the nearest 0.1 kilogram instead of the nearest kilogram.



Here is a histogram showing the same distribution.



In this case, it is difficult to make sense of the distribution from the dot plot because the dots are so close together and all in one line. The histogram of the same data set does a much better job showing the distribution of weights, even though we can't see the individual data values.

Glossary

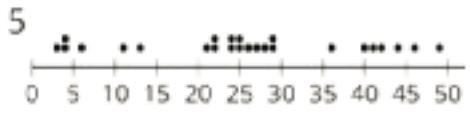
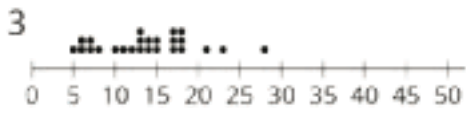
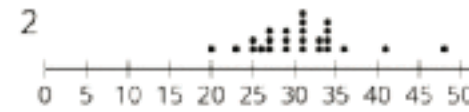
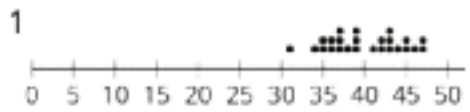
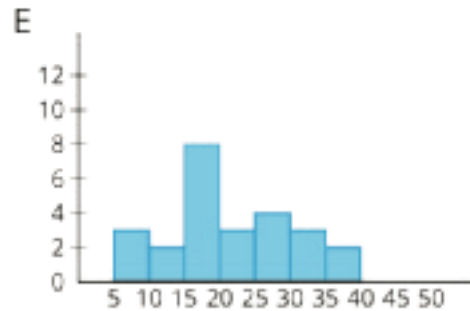
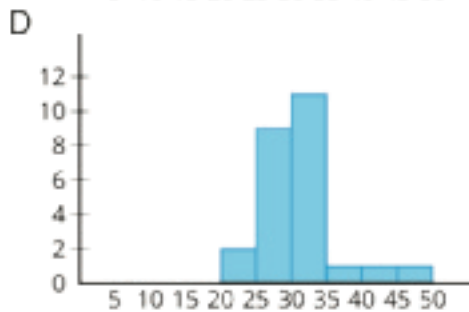
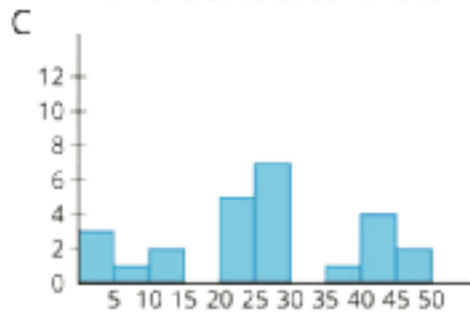
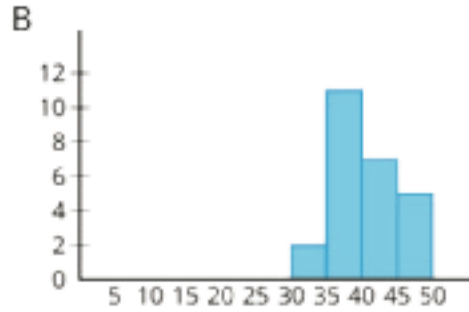
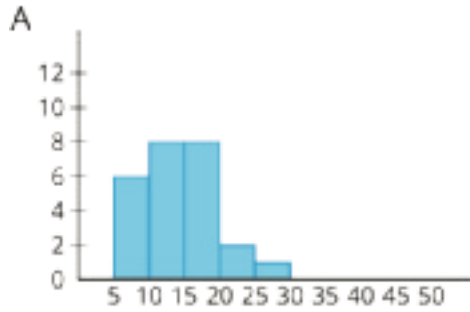
- histogram

Lesson 6 Practice Problems

Problem 1

Statement

Match histograms A through E to dot plots 1 through 5 so that each match represents the same data set.



Solution

1. B
2. D
3. A
4. E
5. C

Problem 2

Statement

$(-2, 3)$ is one vertex of a square on a coordinate plane. Name three points that could be the other vertices.

Solution

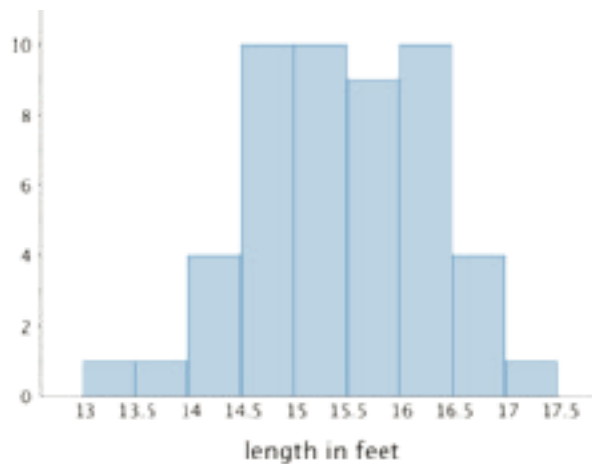
Answers vary. Sample response: $(2, 3)$, $(2, -1)$, $(-2, -1)$

(From Unit 7, Lesson 12.)

Problem 3

Statement

Here is a histogram that summarizes the lengths, in feet, of a group of adult female sharks. Select **all** the statements that are true, according to the histogram.



- A. A total of 9 sharks were measured.
- B. A total of 50 sharks were measured.
- C. The longest shark that was measured was 10 feet long.
- D. Most of the sharks that were measured were over 16 feet long.
- E. Two of the sharks that were measured were less than 14 feet long.

Solution

["B", "E"]

Problem 4

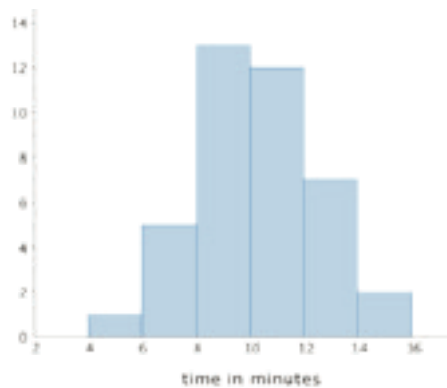
Statement

This table shows the times, in minutes, it took 40 sixth-grade students to run 1 mile.

time (minutes)	frequency
4 to less than 6	1
6 to less than 8	5
8 to less than 10	13
10 to less than 12	12
12 to less than 14	7
14 to less than 16	2

Draw a histogram for the information in the table.

Solution



Lesson 7: Using Histograms to Answer Statistical Questions

Goals

- Compare and contrast (in writing) histograms that represent two different data sets measuring the same quantity.
- Critique (orally) a description of a distribution, recognizing that there are multiple valid ways to describe its center and spread.
- Describe (orally and in writing) the distribution shown on a histogram, including making claims about the center and spread.

Learning Targets

- I can draw a histogram from a table of data.
- I can use a histogram to describe the distribution of data and determine a typical value for the data.

Lesson Narrative

In this lesson, students create, read, and interpret histograms (MP2). They characterize the distribution displayed in a histogram in terms of its shape and spread, and identify a measurement that is typical for the data set by looking for the center in a histogram (MP7). Students also use histograms to make comparisons and to better understand what different spreads and values of center mean in a given context.

Alignments

Addressing

- 6.SP.A.1: Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students' ages.
- 6.SP.A.2: Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- 6.SP.B: Summarize and describe distributions.
- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- 6.SP.B.5.b: Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

Instructional Routines

- MLR1: Stronger and Clearer Each Time
- MLR7: Compare and Connect
- Think Pair Share
- Which One Doesn't Belong?

Required Materials

Rulers marked with centimeters

Student Learning Goals

Let's draw histograms and use them to answer questions.

7.1 Which One Doesn't Belong: Questions

Warm Up: 5 minutes

The purpose of this warm-up is to encourage students to connect the ideas they learned earlier about statistical questions and types of data (categorical and numerical) to the work on describing distributions (center and spread).

There are many ways to interpret the questions and identify how each one is unique. For example, they could say that Question A is about retirement, Question B about jobs, etc. If students begin to depart from thinking in statistical terms, remind them to think about how we might go about answering the questions and what the answers might involve.

Addressing

- 6.SP.A.1
- 6.SP.B

Instructional Routines

- Which One Doesn't Belong?

Launch

Arrange students in groups of 2–4. Display the questions for all to see. Give students 1 minute of quiet think time and ask students to indicate when they have noticed one question that does not belong and can explain why. When the minute is up, give students 2 minutes to share with their group their reasoning on why a question doesn't belong, and then, together, find at least one reason each question doesn't belong.

Student Task Statement

Here are four questions about the population of Alaska. Which question does not belong? Be prepared to explain your reasoning.

1. In general, at what age do Alaska residents retire?
2. At what age can Alaskans vote?
3. What is the age difference between the youngest and oldest Alaska residents with a full-time job?
4. Which age group is the largest part of the population: 18 years or younger, 19–25 years, 25–34 years, 35–44 years, 45–54 years, 55–64 years, or 65 years or older?

Student Response

Answers vary. Sample responses:

1. It is the only question that asks for the center of data (containing the ages at which Alaska residents retire).
2. It is the only non-statistical question.
3. It is the only question that asks for the spread of data.
4. It is the only one that requires categorical data (or requires rearranging numerical data into categories). All the other questions require numerical data.

Activity Synthesis

Ask each group to share one reason why a particular question does not belong. Record and display the responses for all to see. After each response, ask the class if they agree or disagree. Since there is no single correct answer for why a question does not belong, attend to students' explanations and ensure the reasons given are correct.

If students use terms that are essential in this unit (such as center, spread, statistical, non-statistical, numerical data, categorical data, etc.), ask them to explain their meanings in their own words; these are opportunities to reinforce their understanding of the terms and to note any misconceptions. If students give unsubstantiated claims, ask them to substantiate them.

7.2 Measuring Earthworms

20 minutes

In a previous lesson, students had a primer on histograms—how they are drawn, how they differ from dot plots, and what information they can tell us. In this activity, students practice drawing a histogram for a given data set and using it to answer statistical questions. To help students understand the lengths involved in the data set, students are asked to draw various lengths used to group the worms in the first histogram.

As students organize the data set and draw their histogram, notice any challenges or questions students come across. If a question is raised by multiple groups, consider discussing it with the whole class. Also pay attention to how students use the histogram to identify a “typical” length—some might describe it in terms of the size of the bins (e.g., “a typical length is between 20

and 40 mm"); others might choose a value within a bin or a boundary between bins. Invite them to share their reasonings later.

Addressing

- 6.SP.A.2
- 6.SP.B.4

Instructional Routines

- MLR7: Compare and Connect
- Think Pair Share

Launch

Arrange students in groups of 2. Provide access to centimeter rulers.

Consider giving students a brief overview of the context for the problems in the activity. Tell students that there are nearly 6,000 species of earthworms in the world. Some earthworms help the environment, while others (generally not native to the region in which they are found) may harm the environment. Earthworms that are native to a particular region of the world are often raised, by farmers, in terrariums (a container or bin similar to an aquarium but it contains soil and leaves). The terrarium-raised earthworms provide bait for people who fish, provide food for various wildlife, and decompose food waste into soil. Food waste and water are added to the terrariums as food for raising and growing worms. Soil produced by the worms as they eat the food waste is often used as fertilizer.

Explain that the lengths of the worms in the bins provide information about the ages of the worms, which can be useful for the farmer. In this activity, students will organize the lengths of the earthworms in several terrariums or bins.

Give students 8–10 minutes of quiet work time, and then 3–4 minutes to discuss their work and complete the activity with a partner.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Activate or supply background knowledge by demonstrating how to use a ruler to draw a line segment for a given length.

Supports accessibility for: Memory; Conceptual processing

Anticipated Misconceptions

When determining frequencies of data values, students might lose track of their counting. Suggest that they use tally marks to keep track of the number of occurrences for each bin.

When drawing the histogram, students might mistakenly use bar graphs as a reference and leave spaces between the bars. Ask them to look at the bars in other histograms they have seen so far and to think about what the gaps might mean considering that the bars are built on a number line.

Student Task Statement

An earthworm farmer set up several containers of a certain species of earthworms so that he could learn about their lengths. The lengths of the earthworms provide information about their ages. The farmer measured the lengths of 25 earthworms in one of the containers. Each length was measured in millimeters.



1. Using a ruler, draw a line segment for each length:

- 20 millimeters
- 40 millimeters
- 60 millimeters
- 80 millimeters
- 100 millimeters

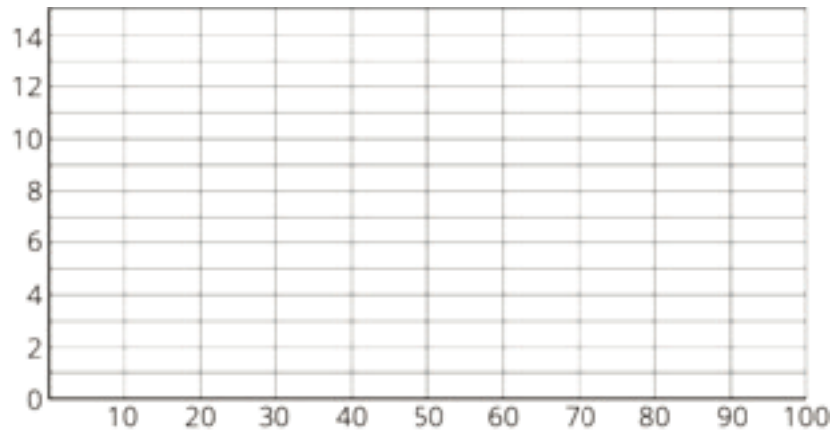
2. Here are the lengths, in millimeters, of the 25 earthworms.

6 11 18 19 20 23 23 25
 25 26 27 27 28 29 32 33
 41 42 48 52 54 59 60 77
 93

Complete the table for the lengths of the 25 earthworms.

length	frequency
0 millimeters to less than 20 millimeters	
20 millimeters to less than 40 millimeters	
40 millimeters to less than 60 millimeters	
60 millimeters to less than 80 millimeters	
80 millimeters to less than 100 millimeters	

3. Use the grid and the information in the table to draw a histogram for the worm length data. Be sure to label the axes of your histogram.



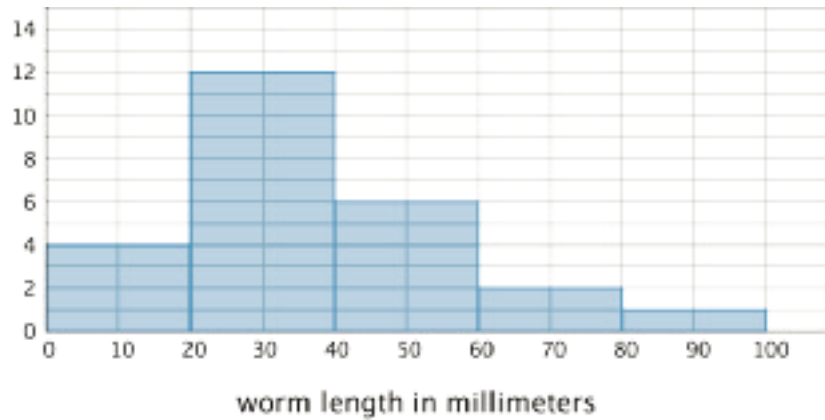
4. Based on the histogram, what is a typical length for these 25 earthworms? Explain how you know.
5. Write 1–2 sentences to describe the spread of the data. Do most of the worms have a length that is close to your estimate of a typical length, or are they very different in length?

Student Response

1. Drawings should show segments of 20 mm, 40 mm, 60 mm, 80 mm, and 100 mm.

2.

length	frequency
0 millimeters to less than 20 millimeters	4
20 millimeters to less than 40 millimeters	12
40 millimeters to less than 60 millimeters	6
60 millimeters to less than 80 millimeters	2
80 millimeters to less than 100 millimeters	1



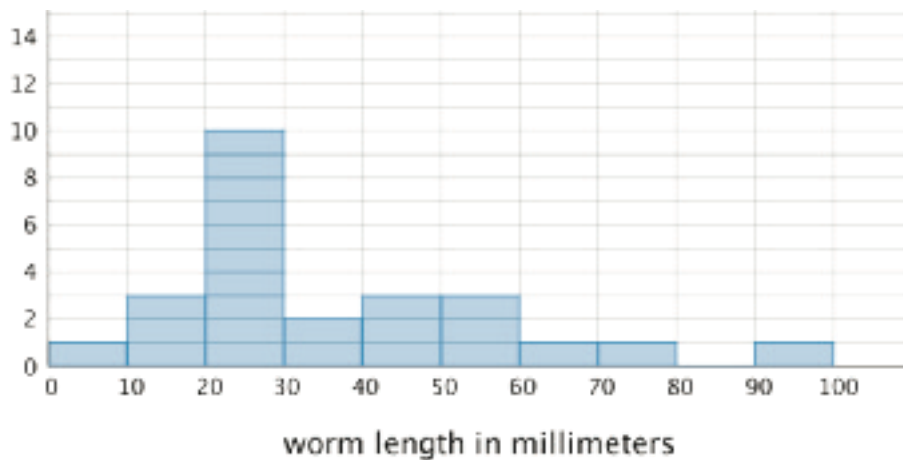
3.

4. Answers vary. Sample responses: Between 20 and 40 mm; about 30 mm; about 35 mm.

5. Answers vary. Sample response: Most of the worms appear to be shorter than my estimate of the typical value, but they are all pretty close to it. There are worms that are longer than my estimate, but not as many.

Are You Ready for More?

Here is another histogram for the earthworm measurement data. In this histogram, the measurements are in different groupings.



1. Based on this histogram, what is your estimate of a typical length for the 25 earthworms?
2. Compare this histogram with the one you drew. How are the distributions of data summarized in the two histograms the same? How are they different?
3. Compare your estimates of a typical earthworm length for the two histograms. Did you reach different conclusions about a typical earthworm length from the two histograms?

Student Response

Answers vary. Sample responses:

1. About 30mm.
2. Both histograms have clusters of data around 20 and 30mm. The one I drew has more data values between 30 and 60mm, and this histogram has data that is more spread out.
3. The typical earthworm lengths are both around 30mm, but the typical length from this histogram is a little smaller.

Activity Synthesis

Ask one or two students to display their completed histograms for all to see and briefly describe the overall distribution. Then, select a few other previously identified students to share their responses and explanations for the last two questions.

Focus the discussion on how identifying center and spread using a histogram is different than doing so using a dot plot. Discuss:

- “In a histogram, are we able to see clusters of values in the distribution?”
- “Can we see the largest and smallest values? Can we tell the overall spread?”
- “How do we identify the center of a distribution?”

From the various estimates that students give for a typical earthworm length (and from earlier exercises), students should begin to see that identifying a typical value of a distribution is not a straightforward or precise process so far. Explain that in upcoming lessons they will explore how to describe a typical value and characterize a distribution more systematically.

Access for English Language Learners

Representing, Conversing: MLR7 Compare and Connect. As the selected students show and describe their histograms, invite pairs to discuss: “What is the same and what is different?” about their own histograms and reasoning. This will help students better understand there are multiple valid ways for describing center and spread.

Design Principle(s): Maximize meta-awareness; Support sense-making

7.3 Tall and Taller Players

10 minutes

Now that students have some experience drawing and interpreting histograms, they use histograms to compare distributions of two populations. In a previous activity, students compared the two dot plots of students in a keyboarding class—one for the typing speeds at the beginning of the course and the other showing the speeds at the end of the course. In this activity, they

recognize that we can compare distributions displayed in histograms in a similar way—by studying shapes, centers, and spreads.

Addressing

- 6.SP.A.2
- 6.SP.B.4
- 6.SP.B.5.b

Instructional Routines

- MLR1: Stronger and Clearer Each Time
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 4–5 minutes of quiet work time and 1–2 minutes to share their responses with a partner.

Access for Students with Disabilities

Engagement: Develop Effort and Persistence. Encourage and support opportunities for peer interactions. Invite students to talk about their ideas with a partner before writing them down. Display sentence frames to support students when they explain their reasoning. For example, “Histogram ___ represents the heights of _____ players because . . .”, and “I agree/disagree because . . .”

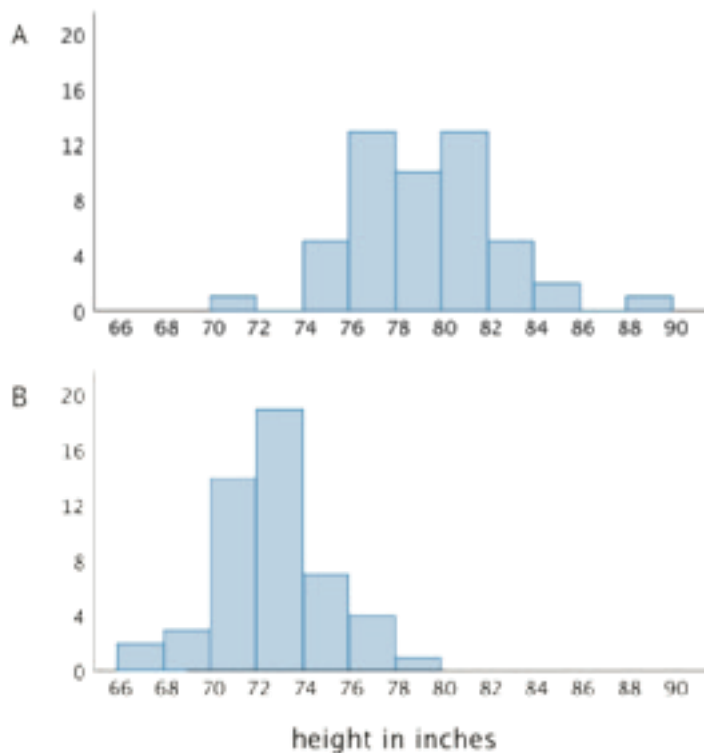
Supports accessibility for: Language; Social-emotional skills

Student Task Statement

Professional basketball players tend to be taller than professional baseball players.

Here are two histograms that show height distributions of 50 male professional baseball players and 50 male professional basketball players.

1. Decide which histogram shows the heights of baseball players and which shows the heights of basketball players. Be prepared to explain your reasoning.



- Write 2–3 sentences that describe the distribution of the heights of the basketball players. Comment on the center and spread of the data.
- Write 2–3 sentences that describe the distribution of the heights of the baseball players. Comment on the center and spread of the data.

Student Response

- Histogram A shows the heights of professional basketball players. Sample explanation: The cluster of bars are located higher on the number line compared to those in Histogram B, which means they represent taller players.
- Answers vary. Sample description: The basketball players' data distribution is centered at around 80–81 inches. Except for two players at the shortest and tallest ends, most players are within 6–7 inches of the center.
- Answers vary. Sample description: The baseball players' data distribution is centered at around 72 inches. The spread is also similar to the basketball players' data. All players are within 6 inches of the center of the data.

Activity Synthesis

Select a few students to share their descriptions about basketball players and baseball players. After each student shares, ask others if they agree with the descriptions and, if not, how they might revise or elaborate on them. In general, students should recognize that the distributions of the two groups of athletes are quite different and be able to describe how they are different.

Highlight the fact that students are using approximations of center and different adjectives to characterize a distribution or a typical height, and that, as a result, there are variations in our descriptions. In some situations, these variations might make it challenging to compare groups more precisely. We will study specific ways to measure center and spread in upcoming lessons.

Access for English Language Learners

Representing, Writing: MLR1 Stronger and Clearer Each Time. Use this routine to give students a structured opportunity to revise and refine their written responses to the questions about describing the distribution of heights of basketball and baseball players. Ask students to meet with 2–3 partners, to share and get feedback on their responses. Provide listeners with prompts for feedback that will clarify the descriptions about the distributions shown in the histograms (e.g., “What details about the histograms are important to describe?”). Finally, give students time to revise their initial written responses to reflect input they received. This will help students write descriptions of the center and spread of distributions displayed in histograms using precise language.

Design Principle(s): Optimize output (for explanation); Maximize meta-awareness

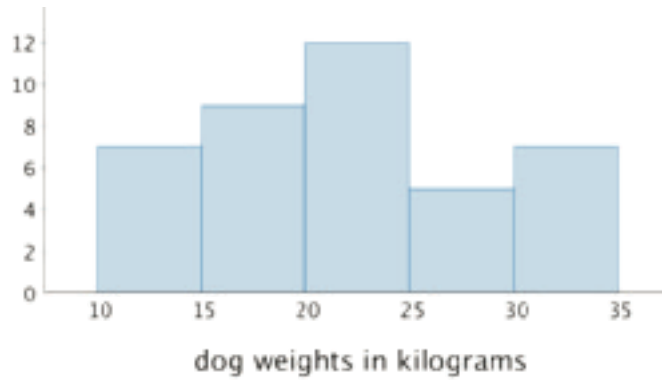
Lesson Synthesis

In this lesson, we learn how to draw a histogram and how to use it to describe characteristics of a data set.

- “What are some decisions we should think about and make before drawing a histogram?”
- “Does the width of each bar have to represent a distance of 5 units, or can it represent other number of units?”
- “What does the horizontal axis of a histogram tell us? What about the vertical axis?”
- “How do we know how tall to make each bar?”

Once we have a histogram drawn, we can use it to answer some questions about a data set.

- “How would you describe a typical weight for this group of dogs?”
- “What can we say about the spread of the dog weights based on this histogram?”



7.4 A Tale of Two Seasons

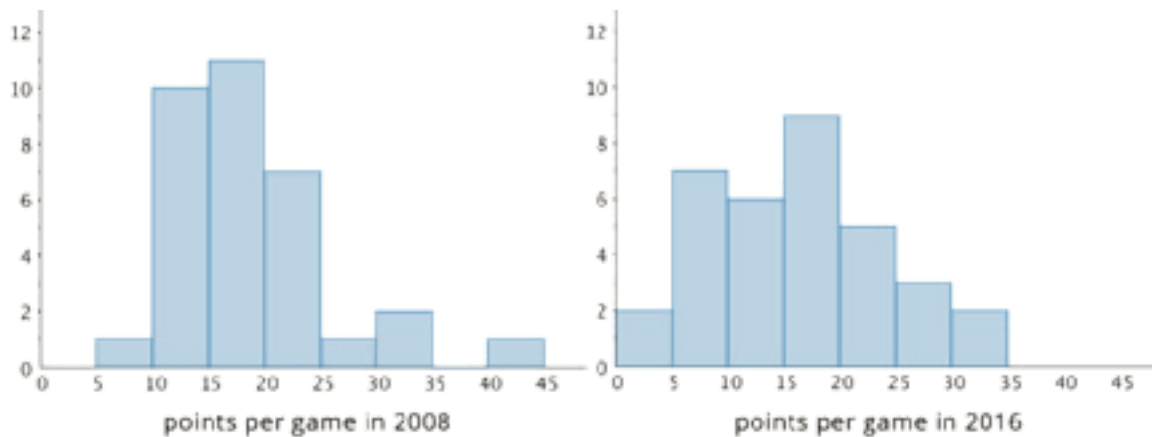
Cool Down: 5 minutes

Addressing

- 6.SP.A.2
- 6.SP.B.4

Student Task Statement

The two histograms show the points scored per game by a basketball player in 2008 and 2016.



1. What is a typical number of points per game scored by this player in 2008? What about in 2016? Explain your reasoning.
2. Write 2–3 sentences that describe the spreads of the two distributions, including what spreads might tell us in this context.

Student Response

1. Answers vary. Sample response: In both seasons, the player typically scored around 15 to 20 points in a game. In each histogram the typical score should be around the center of the spread of data.

2. Answers vary. Sample response: The spread of data was more narrow in 2016 than in 2008. Although the player was more consistent in 2016, she had a really great game in 2008 that changed the spread of the scores.

Student Lesson Summary

Here are the weights, in kilograms, of 30 dogs.

10	11	12	12	13	15	16	16	17
18	18	19	20	20	20	21	22	22
22	23	24	24	26	26	28	30	32
32	34	34						

Before we draw a histogram, let's consider a couple of questions.

- What are the smallest and largest values in our data set? This gives us an idea of the distance on the number line that our histogram will cover. In this case, the minimum is 10 and the maximum is 34, so our number line needs to extend from 10 to 35 at the very least.

(Remember the convention we use to mark off the number line for a histogram: we include the left boundary of a bar but exclude the right boundary. If 34 is the right boundary of the last bar, it won't be included in that bar, so the number line needs to go a little greater than the maximum value.)

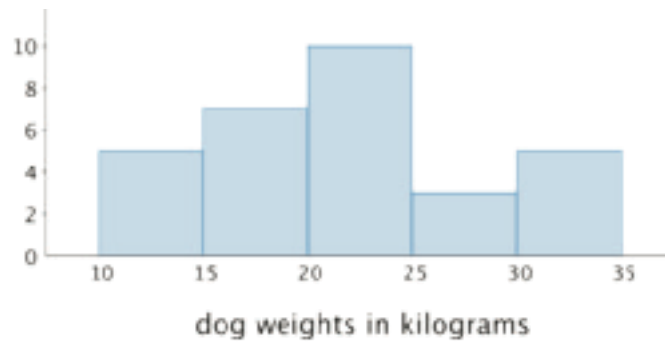
- What group size or bin size seems reasonable here? We could organize the weights into bins of 2 kilograms (10, 12, 14, . . .), 5 kilograms, (10, 15, 20, 25, . . .), 10 kilograms (10, 20, 30, . . .), or any other size. The smaller the bins, the more bars we will have, and vice versa.

Let's use bins of 5 kilograms for the dog weights. The boundaries of our bins will be: 10, 15, 20, 25, 30, 35. We stop at 35 because it is greater than the maximum.

Next, we find the frequency for the values in each group. It is helpful to organize the values in a table.

weights in kilograms	frequency
10 to less than 15	5
15 to less than 20	7
20 to less than 25	10
25 to less than 30	3
30 to less than 35	5

Now we can draw the histogram.



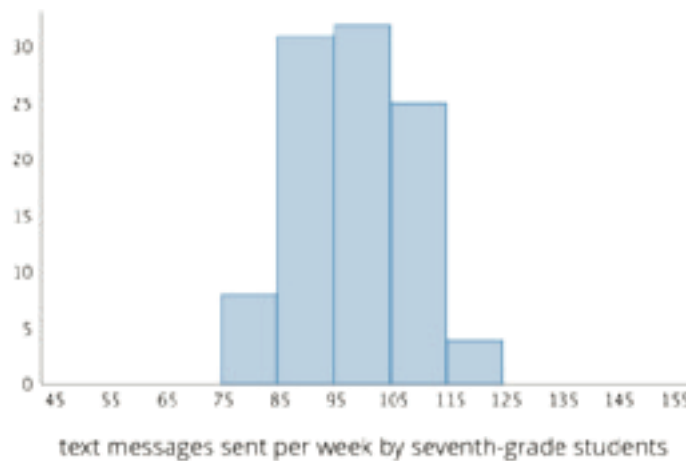
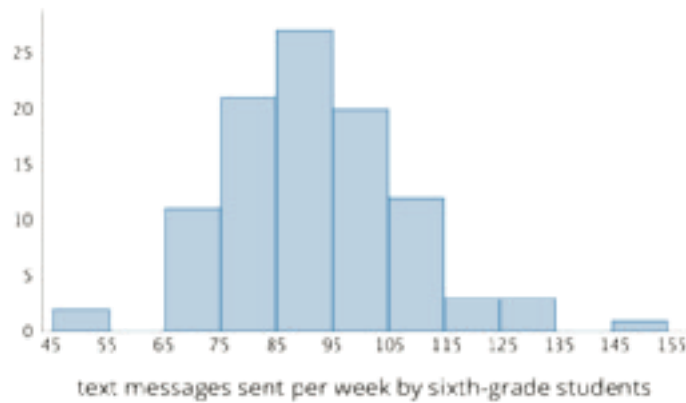
The histogram allows us to learn more about the dog weight distribution and describe its center and spread.

Lesson 7 Practice Problems

Problem 1

Statement

These two histograms show the number of text messages sent in one week by two groups of 100 students. The first histogram summarizes data from sixth-grade students. The second histogram summarizes data from seventh-grade students.



- Do the two data sets have approximately the same center? If so, explain where the center is located. If not, which one has the greater center?
- Which data set has greater spread? Explain your reasoning.
- Overall, which group of students—sixth- or seventh-grade—sent more text messages?

Solution

- Yes, both are centered around 100 text messages.
- The sixth-grade students have a wider spread, much of their data is outside the 75–125 range, while no seventh-grade students are outside this range.

c. Neither. Both send about the same number of text messages, because the center of the two data sets is close.

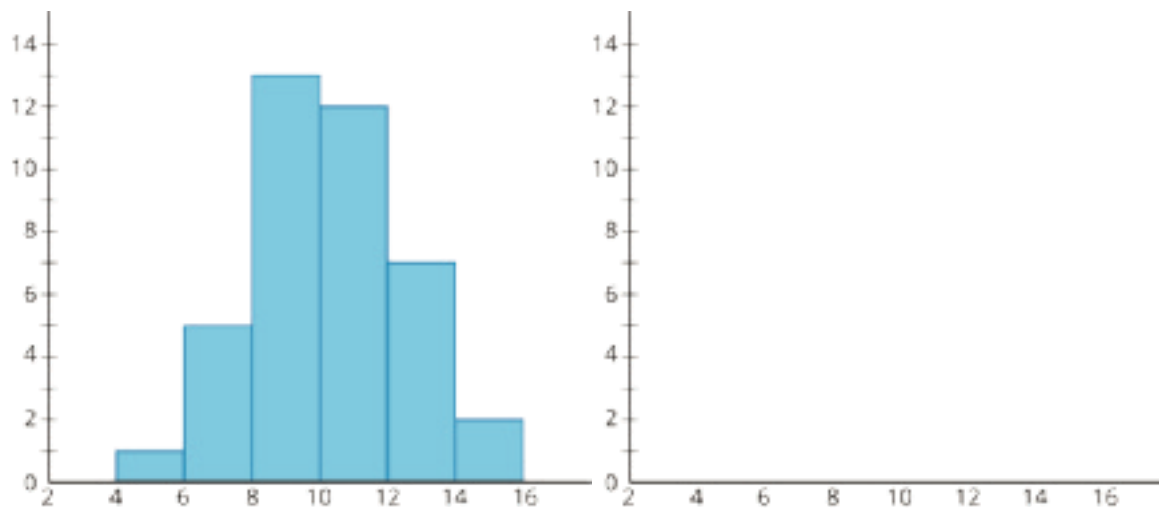
Problem 2

Statement

Forty sixth-grade students ran 1 mile. Here is a histogram that summarizes their times, in minutes. The center of the distribution is approximately 10 minutes.

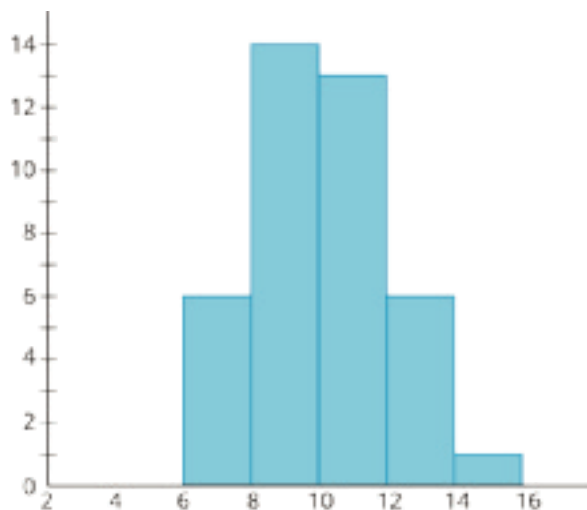
On the blank axes, draw a second histogram that has:

- a distribution of times for a different group of 40 sixth-grade students.
- a center at 10 minutes.
- less variability than the distribution shown in the first histogram.



Solution

Responses vary. Sample response:



Problem 3

Statement

Jada has d dimes. She has more than 30 cents but less than a dollar.

- Write two inequalities that represent how many dimes Jada has.
- Can d be 10?
- How many possible solutions make both inequalities true? If possible, describe or list the solutions.

Solution

- $d > 3$ and $d < 10$ (or $d \geq 4$ and $d \leq 9$)
- No, this does not make $d < 10$ true. 10 dimes is a dollar, which is too much.
- 6 possible solutions: 4, 5, 6, 7, 8, and 9

(From Unit 7, Lesson 9.)

Problem 4

Statement

Order these numbers from greatest to least: -4 , $\frac{1}{4}$, 0 , 4 , $-3\frac{1}{2}$, $\frac{7}{4}$, $-\frac{5}{4}$

Solution

4 , $\frac{7}{4}$, $\frac{1}{4}$, 0 , $-\frac{5}{4}$, $-3\frac{1}{2}$, -4

(From Unit 7, Lesson 4.)

Lesson 8: Describing Distributions on Histograms

Goals

- Compare and contrast (orally) bar graphs and histograms, recognizing that descriptions of shape, center, and spread don't pertain to bar graphs.
- Describe (orally and in writing) the overall shape and features of a distribution represented on a histogram, including peaks, clusters, gaps, and symmetry.
- Identify histograms that display distributions with specific features.

Learning Targets

- I can describe the shape and features of a histogram and explain what they mean in the context of the data.
- I can distinguish histograms and bar graphs.

Lesson Narrative

In this lesson, students explore various shapes and features of a distribution displayed in a histogram. They use the structure (MP7) to look for symmetry, peaks, clusters, gaps, and any unusual values in histograms. Students also begin to consider how these features might affect how we characterize a data set. For example, how might we describe what is typical in a distribution that shows symmetry? What about in a distribution that has one peak that is not symmetrical? This work is informal, but helps to prepare students to better understand measures of center and spread later in the unit. Students also distinguish between the uses and construction of bar graphs and histograms in this lesson.

Alignments

Addressing

- 6.SP.A.2: Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

Building Towards

- 6.SP.B.5.d: Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Instructional Routines

- MLR2: Collect and Display
- MLR5: Co-Craft Questions
- Which One Doesn't Belong?

Required Materials

Pre-printed cards, cut from copies of the **Instructional master**

Required Preparation

Print and cut up cards from the Sorting Histograms Instructional master. Prepare 1 set of cards for every 3–4 students.

The Getting to School activity requires students to use data previously collected on their travel methods and times. Organize the data into the tables in the Instructional master ahead of time or allow time for students to do it themselves. Either make a copy for every 2 students, or display the completed tables for all to see during the activity.

Student Learning Goals

Let's describe distributions displayed in histograms.

8.1 Which One Doesn't Belong: Histograms

Warm Up: 5 minutes

This warm-up encourages students to make sense of histograms in terms of center and spread. It prompts students to hold mathematical conversations and explain their reasoning (MP3), and gives the teacher the opportunity to hear how students compare data sets represented by histograms.

Addressing

- 6.SP.A.2

Instructional Routines

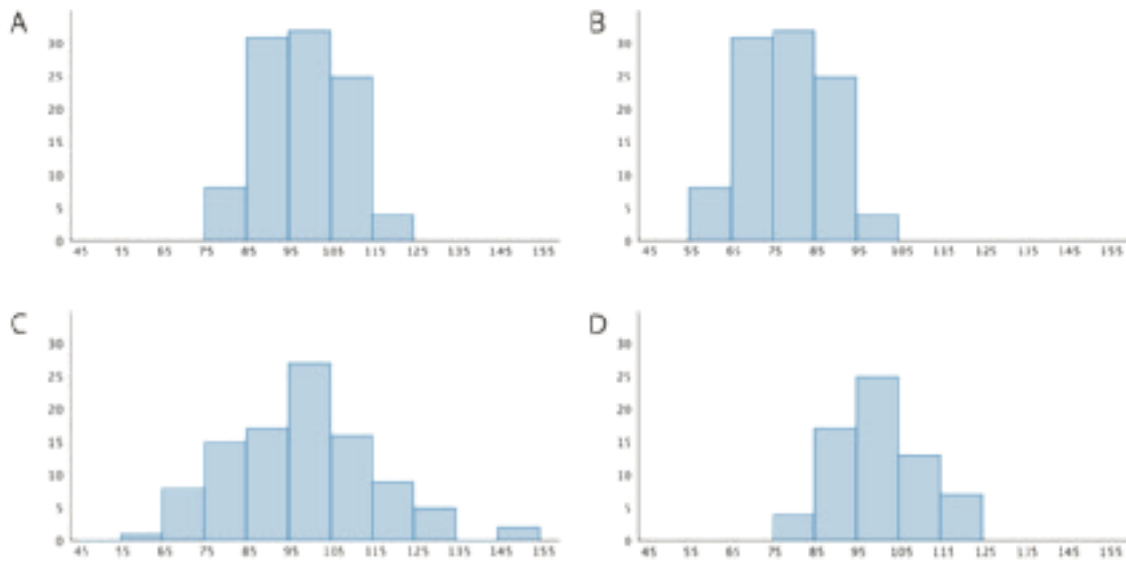
- Which One Doesn't Belong?

Launch

Arrange students in groups of 2–4. Display the images for all to see. Give students 1 minute of quiet think time and ask them to indicate when they have noticed one image that does not belong and can explain why. When the minute is up, give students 2 minutes to share their thinking with their small group, and then, together, find a reason that each image doesn't belong.

Student Task Statement

Which histogram does not belong? Be prepared to explain your reasoning.



Student Response

Answers vary. Sample responses:

- Histogram B does not belong. Unlike the others, its distribution is not centered around 100.
- Histogram C does not belong. The spread of the data is much wider than that of the other histograms.
- Histogram D does not belong. It represents a smaller set of data compared to the others.

Activity Synthesis

Ask students to share one reason their group decided a particular image does not belong. Record and display the responses for all to see. After each response, ask the class if they agree or disagree. Since there is no single correct answer to the question of which one does not belong, attend to students' explanations and ensure the reasons given are reasonable.

If students use terms that are essential in this unit (such as center, spread, distribution, frequency, etc.), ask them to explain their meanings in their own words; these are opportunities to reinforce their understanding of the terms and to note any misconceptions. If students give unsubstantiated claims, ask them to substantiate them.

8.2 Sorting Histograms

20 minutes

This activity is designed to expand both students' exposure to various features of distributions and the language they could use to describe distributions. Students sort histograms based on features such as symmetry, gaps, clusters, and unusual values. In earlier grades, students used the term

“symmetry” to describe geometric figures (4.G.3); here they use it to describe the shape of a distribution.

Addressing

- 6.SP.A.2

Building Towards

- 6.SP.B.5.d

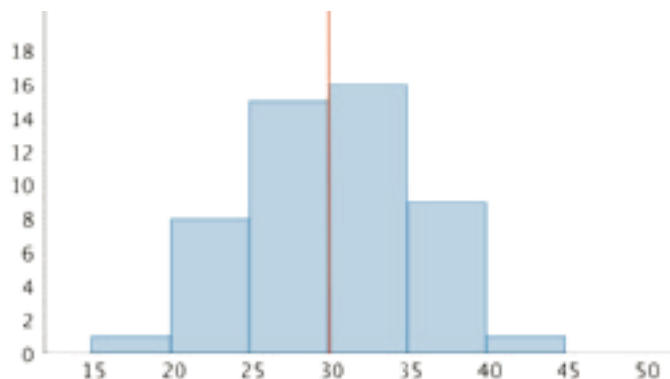
Instructional Routines

- MLR5: Co-Craft Questions

Launch

Display the image of the histogram here for all to see. Explain to students that a diagram of a distribution—a dot plot or a histogram—is described as symmetrical if you can draw a line on the diagram and the parts on one side of the line mirror the parts on the other side. Many distributions are not perfectly symmetrical, but are close to or approximately symmetrical.

The histogram here shows an approximately symmetrical distribution. When a line is drawn at the center (such as the line at 30) the two sides are roughly mirror images. If you were to fold the histogram at the line, the two sides would be close to matching.



Tell students that symmetry is used to describe distributions, and that they will now look for symmetry and other features of distributions in a dozen histograms.

Arrange students in groups of 3–4. Give each group one set of pre-cut cards from the Instructional master. Ask students to study the histograms and identify the features described on their task statement. Give groups 10–12 minutes to complete the activity. Explain that each group will need to discuss their work with another group after the first question and before completing the rest of the activity.

Access for English Language Learners

Representing, Conversing: MLR5 Co-Craft Questions. Begin the launch by inviting students to write mathematical questions about the histogram displayed. Give students 1–2 minutes to share their questions with a partner. Listen for students that use terms center, spread, distribution, or frequency in their questions. Invite students to share their questions with the class, and to compare the language of their questions before continuing with the activity.

Design Principle(s): Optimize output (for generalization); Cultivate conversation

Student Task Statement

1. Your teacher will give your group a set of histogram cards. Sort them into two piles—one for histograms that are approximately symmetrical, and another for those that are not.
2. Discuss your sorting decisions with another group. Do both groups agree which cards should go in each pile? If not, discuss the reasons behind the differences and see if you can reach agreement. Record your final decisions.

- Histograms that are approximately symmetrical:
- Histograms that are not approximately symmetrical:

3. Histograms are also described by how many major peaks they have. Histogram A is an example of a distribution with a single peak that is not symmetrical.

Which other histograms have this feature?

4. Some histograms have a gap, a space between two bars where there are no data points. For example, if some students in a class have 7 or more siblings, but the rest of the students have 0, 1, or 2 siblings, the histogram for this data set would show gaps between the bars because no students have 3, 4, 5, or 6 siblings.

Which histograms do you think show one or more gaps?

5. Sometimes there are a few data points in a data set that are far from the center. Histogram A is an example of a distribution with this feature.

Which other histograms have this feature?

Student Response

Answers vary. Sample responses:

1. No answers required.

2. Approximately symmetrical: Histograms F, I, and J. Students may consider Histograms C, E and K as approximately symmetrical as well. Not approximately symmetrical: Histograms other than the ones previously listed.
3. One peak, not symmetrical: Histograms B. Some may also include Histogram L.
4. With gaps: Histograms A, D, E, G, K and L.
5. With values far from the center: Histograms A, E, G and L. Some may consider including Histogram D, but because there are quite a few data points in the upper group, this really looks more like a distribution that just has a gap.

Activity Synthesis

Students will have had a chance to discuss the different features of a distribution in small groups. Use the whole-class discussion to prompt students to think about what the features might mean, and whether or how they affect the way we characterize a distribution. Remind students that we have been using the center of a distribution to talk about what is typical in a group. Discuss some of these questions:

- “Look at the histograms that you think show symmetry. When a distribution is approximately symmetrical, where might its center be?”
- “Now look at the histograms that you think are not approximately symmetrical. Where might its center be? How might we describe what is typical of a group that has one peak that is not symmetrical (such as that in Histogram B)?”
- “Look at the histograms that show gaps. How might a gap (such as that in Histogram K) affect our description of what is typical in a group?”
- “Look at the histograms that have values that are far away from other values. Do unusual values (such as those in Histogram G) affect our description of center and spread? If those unusual values weren't there, would our description of center and spread change?”

Expect students' answers to be very informal. The goal of the discussion is to raise students' awareness that the shape and features of distributions may affect how we characterize the data. This experience provides a conceptual foundation that would help students make sense of measures of center (mean and median) and measures of spread (mean absolute deviation, interquartile range, and range) later.

8.3 Getting to School

10 minutes

In this activity, students draw a bar graph and histogram, then describe the distributions shown on each display. Although the two visual displays may appear similar at first glance, there are important distinctions between the representations. Students notice differences in how we might characterize distributions in bar graphs and those in histograms, including how we describe typical

values or categories. Along the way, students consolidate their understanding about categorical and numerical data.

Addressing

- 6.SP.A.2
- 6.SP.B.4

Instructional Routines

- MLR2: Collect and Display

Launch

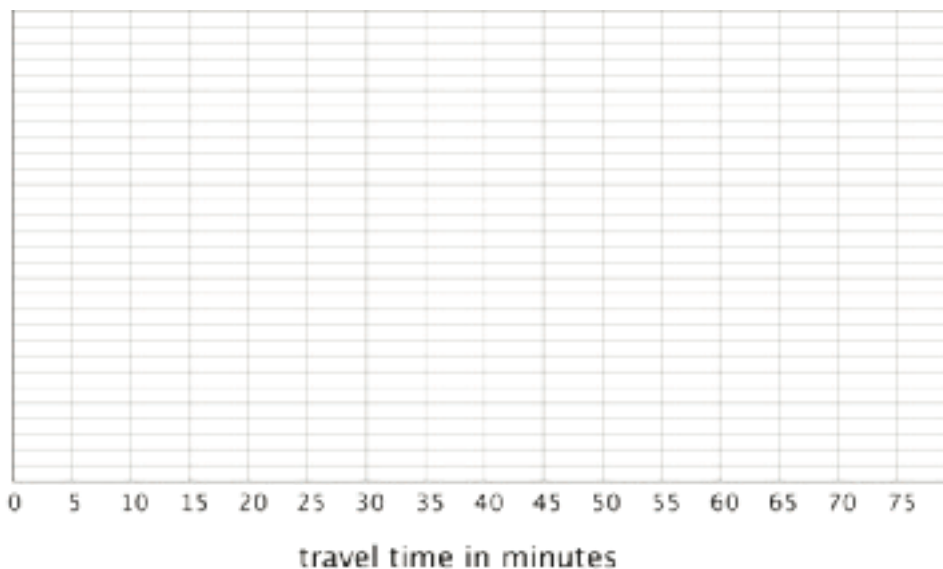
Students will need the data on their travel methods and times, collected at the beginning of the unit. Distribute or display the data collected for these questions from the survey given earlier in the unit. Alternatively, complete the tables in the Instructional master ahead of time.

Arrange students in groups of 2. Give one copy of the Instructional master to each group of students. Display the data from the prior survey or the completed frequency tables for all to see or give a copy to each group of 2 students. Give students 5–6 minutes to complete the activity. Ask one partner to create a bar graph to represent the data on the class’s travel methods and the other to create a histogram to represent the data on travel times, and then answer the questions together.

Student Task Statement

Your teacher will provide you with some data that your class collected the other day.

1. Use the data to draw a histogram that shows your class’s travel times.



2. Describe the distribution of travel times. Comment on the center and spread of the data, as well as the shape and features.

3. Use the data on methods of travel to draw a bar graph. Include labels for the horizontal axis.



4. Describe what you learned about your class's methods of transportation to school. Comment on any patterns you noticed.
5. Compare the histogram and the bar graph that you drew. How are they the same? How are they different?

Student Response

1. Answers and graphs vary based on class data.
2. Answers vary. Students should describe the distribution in terms of center, spread, shape, or other previously discussed features of distribution.
3. Answers and graphs vary based on class data.
4. Answers vary. Students may identify categories that are most or least prevalent. They may also describe the distribution in terms fraction or percentage of data values, or point out both common and unusual characteristics
5. Answers vary. Sample responses:
 - Bar graphs and histograms both use heights of bars to show frequency—the more frequently a value appears, the taller the bar that represents it.
 - The graphs are different in that histograms show numerical data and group data values into “bins” (so we cannot tell how many of a particular value there are). Bar graphs show categorical data, so the horizontal axis shows different categories or labels, rather than a number line. The frequency tells us exactly how many times a particular category appears in the data set.

Are You Ready for More?

Use one of these suggestions (or make up your own). Research data to create a histogram. Then, describe the distribution.

- Heights of 30 athletes from multiple sports
- Heights of 30 athletes from the same sport
- High temperatures for each day of the last month in a city you would like to visit
- Prices for all the menu items at a local restaurant

Student Response

Answers vary.

Activity Synthesis

The purpose of the discussion is for students to recognize the differences between histograms and bar graphs.

Select one student to show a completed histogram and another to show a completed bar graph. Then, solicit several observations about how the two graphical displays compare. Ask questions such as:

- “How are the bar graphs and histograms alike? How are they different?”
- “Can we use a bar graph to display the data on travel times? Why or why not?”
- “Can we use a histogram to display the data on methods of travel? Why or why not?”

Next, select a few other students to share their descriptions of the distributions shown on each type of display. Then, ask questions such as:

- “How are your descriptions of the distribution for travel methods different than those for travel times?”
- “Can you talk about the shape of a distribution shown on a bar graph? Can you talk about its center and spread? Why or why not?”

Students should recognize that only the distribution of numerical data can be described in terms of shape, center, or spread. We cannot analyze these features for a distribution of a categorical data on a bar graph because a bar graph does *not* use a number line. This means the bars can be drawn anywhere, in any order, and with any kind of spacing, so shape, center, and spread would have no meaning.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Provide students with a graphic organizer to record differences between histograms and bar graphs. Consider creating a shared display and keeping it visible throughout the unit.

Supports accessibility for: Language; Organization

Access for English Language Learners

Representing, Conversing: MLR2 Collect and Display. As students discuss the similarities and differences between histograms and bar graphs, record and display phrases you hear in a chart, such as a Venn diagram. Use this as a reference when highlighting or revoicing academic vocabulary from this unit (frequency, bins, numerical data, categorical data). This will help students connect unique and shared characteristics of histograms and bar graphs while reinforcing mathematical language.

Design Principle(s): Maximize meta-awareness

Lesson Synthesis

In this lesson, we look at the shapes and features of distributions that are represented by histograms.

- “What does it mean for a histogram to have symmetry?”
- “What is a ‘peak’ in a distribution? Is it always in the middle, or can it be to one side?”
- “Can a distribution have more than one peak?”
- “What does it mean for a histogram to show a cluster (or more than one clusters)?”
- “What does it mean for a histogram to show a gap?”

We also contrast bar graphs and histograms.

- “When do we use a histogram and when do we use a bar graph?”
- “What are the major differences between how a histogram is drawn and how a bar graph is drawn?”

8.4 Point Spread

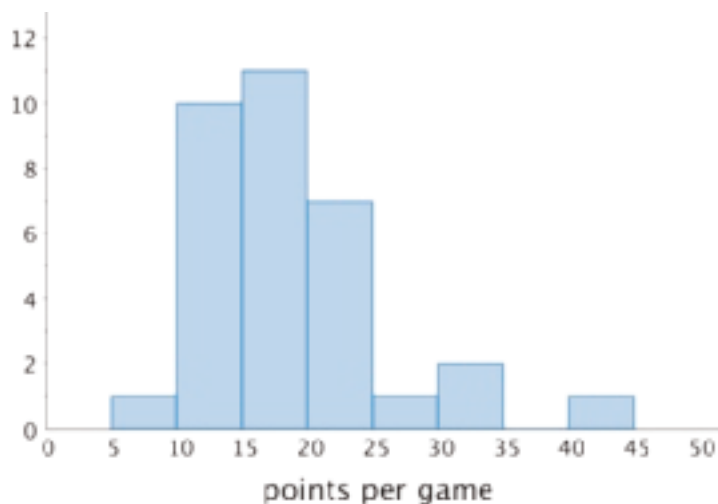
Cool Down: 5 minutes

Addressing

- 6.SP.A.2

Student Task Statement

Here is a histogram that shows the number of points scored by a college basketball player during the 2008 season. Describe the shape and features of the data.

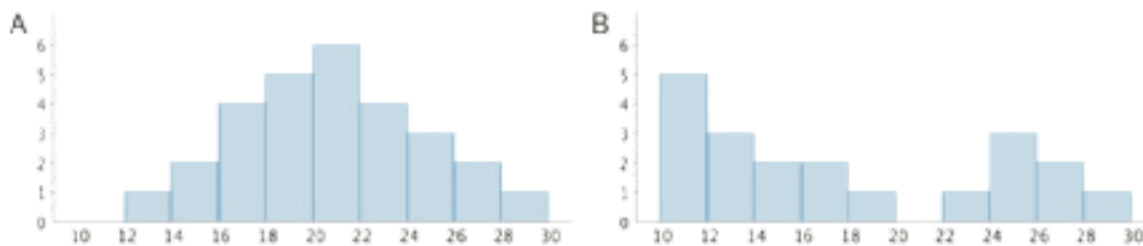


Student Response

Answers vary. Sample response: The data are not symmetrical, there is a peak on the left. The histogram shows a gap between 35 and 40, so there is no game where the player scored 35, 36, 37, 38 or 39 points. There was one game that was unusually high scoring, between 40 and 44 points. The peak is between 15 and 20 points. The center is around 20 points and the data is spread out pretty far above this center.

Student Lesson Summary

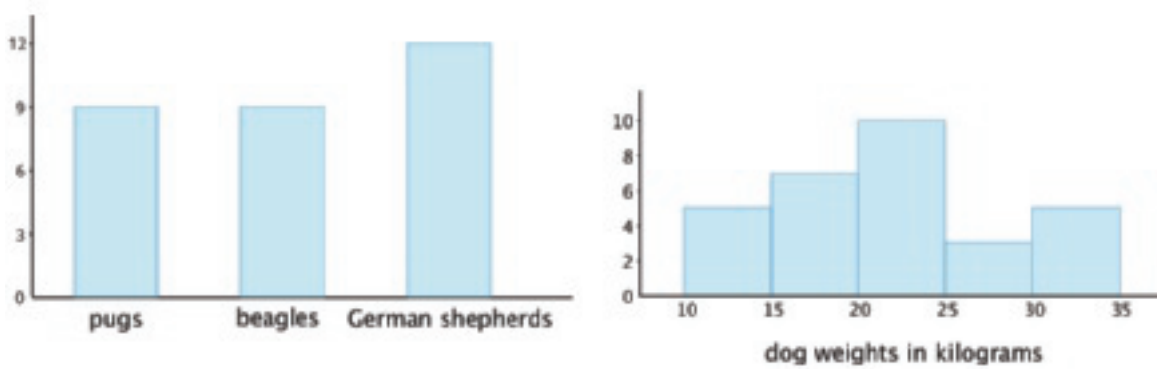
We can describe the shape and features of the distribution shown on a histogram. Here are two distributions with very different shapes and features.



- Histogram A is very symmetrical and has a peak near 21. Histogram B is not symmetrical and has two peaks, one near 11 and one near 25.
- Histogram B has two clusters. A cluster forms when many data points are near a particular value (or a neighborhood of values) on a number line.

- Histogram B also has a gap between 20 and 22. A gap shows a location with no data values.

Here is a bar graph showing the breeds of 30 dogs and a histogram for their weights.



Bar graphs and histograms may seem alike, but they are very different.

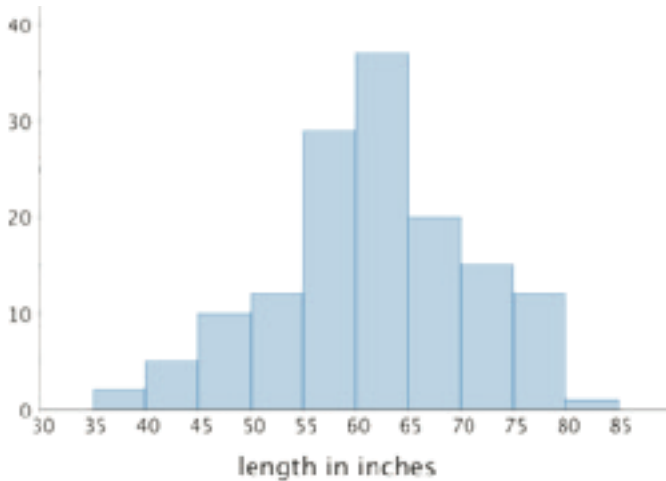
- Bar graphs represent categorical data. Histograms represent numerical data.
- Bar graphs have spaces between the bars. Histograms show a space between bars *only* when no data values fall between the bars.
- Bars in a bar graph can be in any order. Histograms must be in numerical order.
- In a bar graph, the number of bars depends on the number of categories. In a histogram, we choose how many bars to use.

Lesson 8 Practice Problems

Problem 1

Statement

The histogram summarizes the data on the body lengths of 143 wild bears. Describe the distribution of body lengths. Be sure to comment on shape, center, and spread.



Solution

Answers vary. Sample response: The distribution of body lengths is approximately symmetrical. A typical body length for the bears in the group studied is about 60 inches. There is a lot of variability in the body lengths of the bears, with the shortest length being somewhere between 35 and 40 inches and the longest length being somewhere between 80 and 85 inches.

Problem 2

Statement

Which data set is more likely to produce a histogram with a symmetric distribution? Explain your reasoning.

- Data on the number of seconds on a track of music in a pop album.
- Data on the number of seconds spent talking on the phone yesterday by everyone in the school.

Solution

Data on the number of seconds on a track of music in a pop album. Explanations vary. Sample explanation: Most pop songs are around the same amount of time, but most people in the school will not talk much on the phone while a few people will talk a lot so there will be a peak near zero and a few very long times for some people.

Problem 3

Statement

Evaluate the expression $4x^3$ for each value of x .

- a. 1
- b. 2
- c. $\frac{1}{2}$

Solution

- a. 4
- b. 32
- c. $\frac{1}{2}$ or equivalent

(From Unit 6, Lesson 15.)

Problem 4

Statement

Decide if each data set might produce one or more gaps when represented by a histogram. For each data set that you think might produce gaps, briefly describe or give an example of how the values in the data set might do so.

- a. The ages of students in a sixth-grade class.
- b. The ages of people in an elementary school.
- c. The ages of people eating in a family restaurant.
- d. The ages of people who watch football.
- e. The ages of runners in a marathon.

Solution

- a. No.
- b. Yes. Sample reasoning: The data set might show a lot of observations that are between 5 and 12 years old (the students' ages) and older than 20 years (the ages of staff and teachers), but no observations between 12 and 20.
- c. Yes. Sample reasoning: The data set might show a lot of observations of children, then their parents. Observations between 12 and 25 are less likely.
- d. No.

- e. Yes. Sample reasoning: Most of the runners might be adults between 24 and 60 years of age, but there might be a few runners who are in their late teens or older runners in their 80's.

Problem 5

Statement

Jada drank 12 ounces of water from her bottle. This is 60% of the water the bottle holds.

- Write an equation to represent this situation. Explain the meaning of any variables you use.
- How much water does the bottle hold?

Solution

- Answers vary. Sample responses: $12 = \frac{60}{100}b$ or $12 = 0.6b$, where b is the number of ounces of water the bottle holds
- $b = 20$

(From Unit 6, Lesson 7.)

Section: Measures of Center and Variability

Lesson 9: Mean

Goals

- Comprehend the words “mean” and “average” as a measure of center that summarizes the data using a single number.
- Explain (using words and other representations) how to calculate the mean for a numerical data set.
- Interpret diagrams that represent finding the mean as a process of leveling out the data to find a “fair share.”

Learning Targets

- I can explain how the mean for a data set represents a “fair share.”
- I can find the mean for a numerical data set.

Lesson Narrative

In this lesson, students find and interpret the **mean** of a distribution (MP2) as the amount each member of the group would get if everything is distributed equally. This is sometimes called the “leveling out” or the “fair share” interpretation of the mean. For a quantity that cannot actually be redistributed, like the weights of the dogs in a group, this interpretation translates into a thought experiment.

Suppose all of the dogs in a group had different weights and their combined weight was 200 pounds. The mean would be the weight of the dogs if all the dogs were replaced with the same number of identical dogs and the total weight was still 200 pounds.

Here students do not yet make an explicit connection between the mean and the idea of “typical,” or between the mean and the center of a distribution. These connections will be made in upcoming lessons.

Alignments

Building On

- 4.OA.A: Use the four operations with whole numbers to solve problems.

Addressing

- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
- 6.SP.B: Summarize and describe distributions.

- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

Building Towards

- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

Instructional Routines

- Anticipate, Monitor, Select, Sequence, Connect
- MLR7: Compare and Connect
- MLR8: Discussion Supports
- Think Pair Share

Required Materials

Snap cubes

straightedge, but sometimes it is preferable to use an unruled straightedge, like a blank index card.

Straightedges

A rigid edge that can be used for drawing line segments. Sometimes a ruler is okay to use as a

Student Learning Goals

Let's explore the mean of a data set and what it tells us.

9.1 Close to Four

Warm Up: 5 minutes

The purpose of this warm-up is to prepare students to find the mean of a data set. While the goal of the activity is for students to create an expression with a value close to 4, the discussions should focus on the reasoning and strategies students used in creating their expression. Students should notice that they would get the same result if they divided the value of the entire expression in the numerator by 4 as they would if they divided each number in the numerator by 4 because there are 4 numbers in the numerator.

During the partner discussions, identify students with different strategies for creating an expression with a value of 4. Ask them to share during the whole-class discussion.

Building On

- 4.OA.A

Building Towards

- 6.SP.A.3

Instructional Routines

- Think Pair Share

Launch

Arrange students in groups of 2. Give students 1 minute of quiet work time, and then 2 minutes to share their response with a partner. Follow with a whole-class discussion. If students do not interpret the directions of “close to 4” to mean that their expression can have an exact value of 4, tell them it can be exactly 4.

Anticipated Misconceptions

Some students may think they need to use all of the digits from 0 to 9. Tell them that only 4 digits need to be used, although they are welcome to try finding a good solution using 2 digit numbers if they want.

Student Task Statement

Use the digits 0–9 to write an expression with a value as close as possible to 4. Each digit can be used only one time in the expression.

$$\left(\boxed{} + \boxed{} + \boxed{} + \boxed{} \right) \div 4$$

Student Response

Answers vary. Sample response: $(\boxed{2} + \boxed{8} + \boxed{5} + \boxed{1}) \div 4$. Since the expression should be close to 4, the numerator of the fraction should be close to 16.

Activity Synthesis

Poll the class on whether the value of their expression is exactly 4 or is close to 4. Ask selected students to share their strategy for creating an expression with a value of 4. Record and display their responses for all to see.

As students share their reasoning, consider asking some of the following questions:

- “How did you decide on the value of the numerator?”
- “How did the denominator affect your strategy?”
- “How might your strategy change if the denominator was a different number, say, 6 or 10?”
- “How might your strategy change if the numerator had more numbers or fewer numbers?”

9.2 Spread Out and Share

15 minutes (there is a digital version of this activity)

This activity introduces students to the concept of **mean** or **average** in terms of equal distribution or fair share. The two contexts chosen are simple and accessible, and include both discrete and

continuous values. Diagrams are used to help students visualize the distribution of values into equal amounts.

The first set of problems (about cats in crates) can be made even more concrete by providing students with blocks or snap cubes that they can physically distribute into piles or containers.

Students using the digital activities will engage with an applet that allows students to sort cats. For the second set of problems (about hours of work), students are prompted to draw two representations of the number of hours of work before and after they are redistributed, creating a visual representation of fair shares or quantities being leveled out.

As students work, identify those with very different ways of arranging cats into crates to obtain a mean of 6 cats. Also look for students who determine the redistributed work hours differently. For example, some students may do so by moving the number of hours bit by bit, from a server with the most hours to the one with the fewest hours, and continue to adjust until all servers have the same number. Others may add all the hours and divide the sum by the number of servers.

Addressing

- 6.SP.A.3
- 6.SP.B

Instructional Routines

- Anticipate, Monitor, Select, Sequence, Connect
- MLR7: Compare and Connect
- Think Pair Share

Launch

Arrange students in groups of 2. Provide access to straightedges. Also consider providing snap cubes for students who might want to use them to physically show redistribution of data values. If using the digital lesson, students will have access to an applet that will allow them to sort cats, snap cubes may not be necessary but can be provided.

Give students 3–4 minutes of quiet work time to complete the first set of questions and 1–2 minutes to share their responses with a partner. Since there are many possible correct responses to the question about the crates in a second room, consider asking students to convince their partner that the distribution that they came up with indeed has an average of 3 kittens per crate. Then, give students 4–5 min to work on the second set of questions together.

Access for Students with Disabilities

Representation: Develop Language and Symbols. Ensure access to virtual or concrete manipulatives to connect symbols to concrete objects or values. Provide students with blocks or snap cubes that they can physically distribute into piles or containers.

Supports accessibility for: Conceptual processing

Anticipated Misconceptions

In the first room, to get each crate to have the same number of cats, some students might add new cats, not realizing that to “distribute equally” means to rearrange and reallocate existing quantities, rather than adding new quantities. Clarify the meaning of the phrase for these students.

Some students may not recognize that the hours for the servers could be divided so as to not be whole numbers. For example, some may try to give 4 servers 6 hours and 1 server has 7 hours. In this case, the time spent working is still not really divided equally, so ask the student to think of dividing the hours among the servers more evenly if possible.

Student Task Statement

1. The kittens in a room at an animal shelter are placed in 5 crates.



- a. The manager of the shelter wants the kittens distributed equally among the crates. How might that be done? How many kittens will end up in each crate?
 - b. The number of kittens in each crate after they are equally distributed is called the **mean** number of kittens per crate, or the **average** number of kittens per crate. Explain how the expression $10 \div 5$ is related to the average.
 - c. Another room in the shelter has 6 crates. No two crates has the same number of kittens, and there is an average of 3 kittens per crate. Draw or describe at least two different arrangements of kittens that match this description.
2. Five servers were scheduled to work the number of hours shown. They decided to share the workload, so each one would work equal hours.

server A: 3

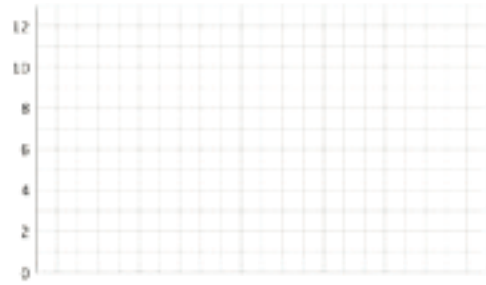
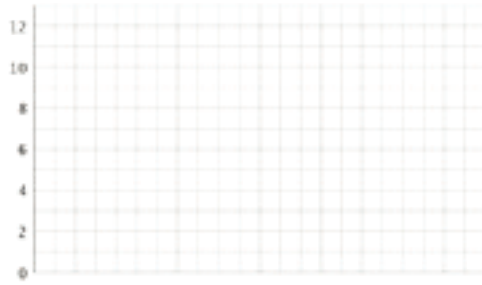
server B: 6

server C: 11

server D: 7

server E: 4

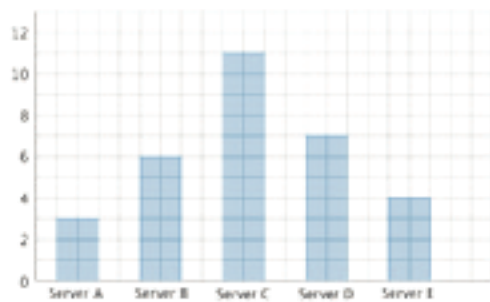
- a. On the grid on the left, draw 5 bars whose heights represent the hours worked by servers A, B, C, D, and E.



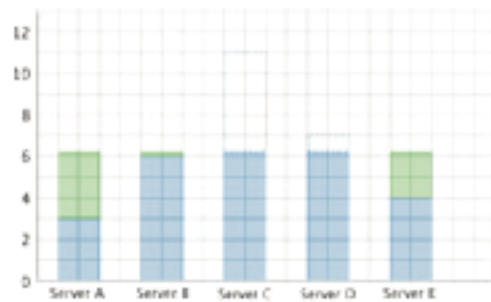
- b. Think about how you would rearrange the hours so that each server gets a fair share. Then, on the grid on the right, draw a new graph to represent the rearranged hours. Be prepared to explain your reasoning.
- c. Based on your second drawing, what is the average or mean number of hours that the servers will work?
- d. Explain why we can also find the mean by finding the value of the expression $31 \div 5$.
- e. Which server will see the biggest change to work hours? Which server will see the least change?

Student Response

1.
 - a. Answers vary. Sample response: Add up the numbers of kittens (a total of 10) and divide that number by 5, which results in 2 kittens per crate.
 - b. The expression is the total number of kittens divided by the number of crates, which is the number of kittens in each crate after they are evenly distributed.
 - c. Answers vary. Sample response: The kittens could be distributed among the 6 crates in the following order: 2, 5, 1, 3, 0, 7, and 6, 2, 4, 5, 0, 1. Since there are 6 crates and an average of 3 kittens in each crates, there are a total of 18 kittens. Any distribution that has a total of 18 kittens will have an average of 3 kittens in each crate.



hours of work before sharing



hours of work after sharing

- a.
 - b. The mean number of hours is 6.2 hours.

- c. The expression shows all the hours being added and divided by 5, which gives us the fair share for each server. $\frac{(3+6+11+7+4)}{5} = 6.2$
- d. Server C will see the biggest change; their work hours will drop by close to 5 hours. Server B will barely see a difference; their work hours will increase only by $\frac{1}{5}$ of an hour or 12 minutes.

Are You Ready for More?

Server F, working 7 hours, offers to join the group of five servers, sharing their workload. If server F joins, will the mean number of hours worked increase or decrease? Explain how you know.

Student Response

Increase. Since the average was 6.2 hours and Server F has 7 hours, it will increase everyone else's hours to even things out again.

Activity Synthesis

Invite several students with different arrangements of cats in the second room with 6 crates to share their solutions and how they know the mean number of cats for their solutions is 3. Make sure everyone understands that their arrangement is correct as long as it had a total of 18 kittens and 6 crates and no two crates have the same number of cats. Show that the correct arrangements could redistribute the 18 cats such that there are 3 cats per crate.

Then, select previously identified students to share how they found the redistributed work hours if the workers were to spread the workload equally. Start with students who reallocated the hours incrementally (from one server to another server) until the hours level out, and then those who added the work hours and dividing the sum by 5.

Students should see that the mean can be interpreted as what each member of the group would get if everything is distributed equally, without changing the sum of values.

Access for English Language Learners

Representing: MLR7 Compare and Connect. After the selected students share their solutions, invite students to discuss "What is the same and what is different?" about the approaches. Call students' attention to the connections between approaches by asking "How is the approach 'divide the sum of work hours by 5' represented visually in the new bar graph?" This will help students make sense of and use different representations of the mean of a set of data.

Design Principle(s): Maximize meta-awareness; Support sense-making

9.3 Getting to School

15 minutes

In this activity, students calculate the mean of a data set and interpret it in the context of the given situation. The first data set students see here has a dozen values, discouraging students from redistributing the values incrementally and encouraging them to use a more efficient method. In the second question, students analyze the values in data sets and use the structure (MP7) to decide whether or not it makes sense that a given mean would match the data set.

As students work and discuss, notice the reasons they give for why the data sets in the second question could or could not be Tyler's data set. Identify students who recognize that the mean of a data set cannot be expected to be higher or lower than most of the values of the data set, and that a fair-share value would have a value that is roughly in the middle of data values.

Addressing

- 6.SP.A.3
- 6.SP.B.5.c

Instructional Routines

- MLR8: Discussion Supports
- Think Pair Share

Launch

Keep students in groups of 2. Give them 6–7 minutes of quiet work time, and then time to discuss their responses with their partner.

Access for Students with Disabilities

Representation: Internalize Comprehension. Activate or supply background knowledge about determining means of data sets. Allow students to use calculators to ensure inclusive participation in the activity.

Supports accessibility for: Memory; Conceptual processing

Student Task Statement

For the past 12 school days, Mai has recorded how long her bus rides to school take in minutes. The times she recorded are shown in the table.

9	8	6	9	10	7	6	12	9	8	10	8
---	---	---	---	----	---	---	----	---	---	----	---

1. Find the mean for Mai's data. Show your reasoning.
2. In this situation, what does the mean tell us about Mai's trip to school?

3. For 5 days, Tyler has recorded how long his walks to school take in minutes. The mean for his data is 11 minutes. Without calculating, predict if each of the data sets shown could be Tyler's. Explain your reasoning.

- data set A: 11, 8, 7, 9, 8
- data set B: 12, 7, 13, 9, 14
- data set C: 11, 20, 6, 9, 10
- data set D: 8, 10, 9, 11, 11

4. Determine which data set is Tyler's. Explain how you know.

Student Response

1. The mean of the data is 8.5 minutes. The mean is found by summing the minutes of travel and dividing by the number of rides:
 $(9 + 8 + 6 + 9 + 10 + 7 + 6 + 12 + 9 + 8 + 10 + 8) \div 12 = 8.5$
2. Answers vary. Sample response: The mean tells us that if the minutes of travel were all leveled out across the 12 days, Mai's trip to school would take 8.5 minutes each day.
3. Answers vary. Sample response: Data set A and data set D could not be Tyler's data, since all of the numbers in each set are either at or below 11. Data set B and data set C could be Tyler's data, since the numbers in each set are distributed around 11.
4. Data set B is Tyler's, since the mean of the numbers is $(11 + 20 + 6 + 9 + 10) \div 5 = 55 \div 5 = 11$.

Activity Synthesis

Select a couple of students to share how they found the mean of Mai's travel times. Poll the class briefly to see if others in the class found the mean the same way.

Then, focus the discussion on the second task and on what values could be reasonably expected of a data set with a particular mean. Ask students how they decided to rule out or keep certain sets of data as potentially belonging to Tyler. If not mentioned by students, highlight that the mean of a data set would be a value in the middle of the range of numbers in order for it to be a fair-share value.

Point out that, unlike hours of work or cats in crates, the times of travel here cannot actually be redistributed. The interpretation of mean translates into a thought experiment:

The mean is the travel time each day if all the travel times in the set were the same such that the combined travel time for this data set and the original data set was the same.

Access for English Language Learners

Speaking, Listening: MLR8 Discussion Supports. Use this routine to support students' explanations of which data set is Tyler's. Invite students to use the sentence frame "Data set ___ is Tyler's because" Invite listeners to use the sentence frame "I agree/disagree, because" This will help students explain their own reasoning and make the sense of the reasoning of others.
Design Principle(s): Support sense-making; Optimize output (for justification)

Lesson Synthesis

In this lesson, we look at finding the **mean** or the **average** of a numerical data set.

- "Suppose that a data set contains the amounts of money in five piggy banks. What would the mean of this data set tell us?"
- "Why might it make sense to think of the mean as a 'fair share?'"
- "How do we find the mean of a data set?"
- "How can we interpret the mean of the heights of students in a class?"

9.4 Finding Means

Cool Down: 5 minutes

Addressing

- 6.SP.B.5.c

Student Task Statement

1. Last week, the daily low temperatures for a city, in degrees Celsius, were 5, 8, 6, 5, 10, 7, and 1. What was the average low temperature? Show your reasoning.
2. The mean of four numbers is 7. Three of the numbers are 5, 7, and 7. What is the fourth number? Explain your reasoning.

Student Response

1. The mean low temperature was 6 degrees Celsius. The sum of the temperatures divided by the total number of recorded temperatures is $(5 + 8 + 6 + 5 + 10 + 7 + 1) \div 7 = 6$.
2. The fourth number is 9. The 4 numbers must be distributed evenly around 7. Since 2 of the numbers are 7, and the third number is two less than 7, the fourth number must be 2 more than 7.

Student Lesson Summary

Sometimes a general description of a distribution does not give enough information, and a more precise way to talk about center or spread would be more useful. The **mean**, or **average**, is a number we can use to summarize a distribution.

We can think about the mean in terms of “fair share” or “leveling out.” That is, a mean can be thought of as a number that each member of a group would have if all the data values were combined and distributed equally among the members.

For example, suppose there are 5 bottles which have the following amounts of water: 1 liter, 4 liters, 2 liters, 3 liters, and 0 liters.



To find the mean, first we add up all of the values. We can think of this as putting all of the water together: $1 + 4 + 2 + 3 + 0 = 10$.



To find the “fair share,” we divide the 10 liters equally into the 5 containers: $10 \div 5 = 2$.



Suppose the quiz scores of a student are 70, 90, 86, and 94. We can find the mean (or average) score by finding the sum of the scores ($70 + 90 + 86 + 94 = 340$) and dividing the sum by four ($340 \div 4 = 85$). We can then say that the student scored, on average, 85 points on the quizzes.

In general, to find the mean of a data set with n values, we add all of the values and divide the sum by n .

Glossary

- average
- mean

Lesson 9 Practice Problems

Problem 1

Statement

A preschool teacher is rearranging four boxes of playing blocks so that each box contains an equal number of blocks. Currently Box 1 has 32 blocks, Box 2 has 18, Box 3 has 41, and Box 4 has 9.

Select **all** the ways he could make each box have the same number of blocks.

- A. Remove all the blocks and make four equal piles of 25, then put each pile in one of the boxes.
- B. Remove 7 blocks from Box 1 and place them in Box 2.
- C. Remove 21 blocks from Box 3 and place them in Box 4.
- D. Remove 7 blocks from Box 1 and place them in Box 2, and remove 21 blocks from Box 3 and place them in Box 4.
- E. Remove 7 blocks from Box 1 and place them in Box 2, and remove 16 blocks from Box 3 and place them in Box 4.

Solution

["A", "E"]

Problem 2

Statement

In a round of mini-golf, Clare records the number of strokes it takes to hit the ball into the hole of each green.

2 3 1 4 5 2 3 4 3

She said that, if she redistributed the strokes on different greens, she could tell that her average number of strokes per hole is 3. Explain how Clare is correct.

Solution

Answers vary. Sample explanation: For both of the greens where she got 4 strokes, moving 1 stroke to the two greens where she got 2 strokes means that all 4 four of those greens now take 3 strokes. Likewise, moving 2 strokes from the green where it took her 5 strokes to the green where she got 1 stroke would also mean 3 strokes for each green.

Problem 3

Statement

Three sixth-grade classes raised \$25.50, \$49.75, and \$37.25 for their classroom libraries. They agreed to share the money raised equally. What is each class's equal share? Explain or show your reasoning.

Solution

\$37.50. Explanations vary. Sample explanation: The total raised is \$112.50, and one-third of that is \$37.50.

Problem 4

Statement

In her English class, Mai's teacher gives 4 quizzes each worth 5 points. After 3 quizzes, she has the scores 4, 3, and 4. What does she need to get on the last quiz to have a mean score of 4? Explain or show your reasoning.

Solution

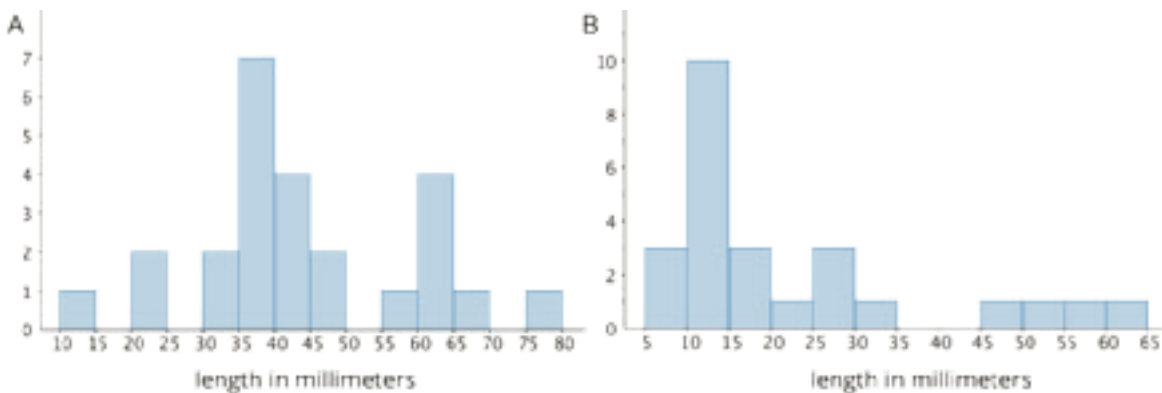
5. Explanations vary. Sample explanation: To get a mean of 4, one point needs to be redistributed to the score of 3, so the last quiz must be a 5 so that it can share one point and still be at the mean itself.

Problem 5

Statement

An earthworm farmer examined two containers of a certain species of earthworms so that he could learn about their lengths. He measured 25 earthworms in each container and recorded their lengths in millimeters.

Here are histograms of the lengths for each container.



a. Which container tends to have longer worms than the other container?

- b. For which container would 15 millimeters be a reasonable description of a typical length of the worms in the container?
- c. If length is related to age, which container had the most young worms?

Solution

- a. Container A
- b. Container B
- c. Container B

(From Unit 8, Lesson 7.)

Problem 6

Statement

Diego thinks that $x = 3$ is a solution to the equation $x^2 = 16$. Do you agree? Explain or show your reasoning.

Solution

No. Explanations vary. Sample explanation: I disagree with Diego. I tried using 3 for x in the equation, but $3^2 = 9$, not 16. Another sample explanation: I disagree with Diego. I know that $4^2 = 16$, so it cannot be true that $3^2 = 16$.

(From Unit 6, Lesson 15.)

Lesson 10: Finding and Interpreting the Mean as the Balance Point

Goals

- Calculate and interpret (orally and in writing) distances between data points and the mean of the data set.
- Interpret diagrams that represent the mean as a “balance point” for both symmetrical and non-symmetrical distributions.
- Represent the mean of a data set on a dot plot and interpret it in the context of the situation.

Learning Targets

- I can describe what the mean tells us in the context of the data.
- I can explain how the mean represents a balance point for the data on a dot plot.

Lesson Narrative

In the previous lesson, students interpreted the mean as a fair-share value—i.e., what each group member would have if all the values are distributed such that all members have the same amount. In this lesson, students use the structure of the data (MP7) to interpret the mean as the balance point of a numerical distribution. They calculate how far away each data point is from the mean and study how the distances on either side of the mean compare.

Students connect this interpretation to why we call the mean a **measure of the center** of a distribution and, through this interpretation, begin to see how the mean is useful in characterizing a “typical” value for the group. Students continue to practice calculating the mean of a data set (MP8) and interpreting it in context (MP2).

Alignments

Addressing

- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

Instructional Routines

- MLR1: Stronger and Clearer Each Time
- MLR5: Co-Craft Questions

- Which One Doesn't Belong?

Student Learning Goals

Let's look at another way to understand the mean of a data set.

10.1 Which One Doesn't Belong: Division

Warm Up: 5 minutes

This warm-up encourages students to analyze the structure and value of expressions, and to connect them to the process of calculating a mean. Each expression has one obvious reason it does not belong, however, there is not one single correct answer.

As students discuss in small groups, listen for ideas related to finding the mean of a data set. Highlight these ideas during whole-class discussion.

Addressing

- 6.SP.A.3

Instructional Routines

- Which One Doesn't Belong?

Launch

Arrange students in groups of 2–4. Display the expressions for all to see. Give students 1 minute of quiet think time, and ask them to indicate when they have noticed one expression that does not belong and can explain why. When the minute is up, ask them to share their thinking with their small group, and then, together, find at least one reason each expression doesn't belong.

Student Task Statement

Which expression does not belong? Be prepared to explain your reasoning.

$$\frac{8 + 8 + 4 + 4}{4}$$

$$\frac{10 + 10 + 4}{4}$$

$$\frac{9 + 9 + 5 + 5}{4}$$

$$\frac{6 + 6 + 6 + 6 + 6}{5}$$

Student Response

Answers vary. Sample responses:

$\frac{8+8+4+4}{4}$ doesn't belong because it is the only expression where each number in the numerator is a multiple of the denominator.

$\frac{9+9+5+5}{4}$ doesn't belong because it is the only expression with a value of 7.

$\frac{10+10+4}{4}$ doesn't belong because it is the only expression where the number of terms in the numerator is not the same as the value of the denominator.

$\frac{6+6+6+6+6}{5}$ doesn't belong because it is the only expression with 5 as the denominator.

Activity Synthesis

Ask each group to share one reason why a particular expression does not belong. Record and display the responses for all to see. After each response, ask the class if they agree or disagree. Since there is no single correct answer to the question, attend to students' explanations and ensure the reasons given are reasonable. If students give unsubstantiated claims, ask them to substantiate them.

At the end of the discussion, ask students which expression or expressions most represent how they would find the mean of a data set. There is a reason each expression (other than C) could represent how they would find the mean of a data set, however, highlight reasoning about the number of terms in the numerator being the same as the value of the denominator (e.g., there are 5 terms in the numerator and the denominator is 5).

10.2 Travel Times (Part 1)

15 minutes

In this activity, students explore the idea of the mean as a **measure of center** of all the values in the data, using a dot plot to help them visualize this idea. Students determine the distance between each data point and the mean, and notice that the sum of distances to the left is equal to the sum of distances to the right. In this sense, the mean can be seen as “balancing” the sets of points with smaller values than it and those with larger values. They make use of the structure (MP7) to calculate the distance between each data point and another point that is *not* the mean to see that the sums on the two sides are not equal. The idea of the mean as a *measure of center* of a distribution is introduced in this context.

As students work and discuss, identify those who could articulate why the mean can be considered a balancing point of a data set.

Addressing

- 6.SP.B.5.c

Instructional Routines

- MLR1: Stronger and Clearer Each Time

Launch

Arrange students in groups of 2. Give students 5 minutes to complete the first two questions with a partner, and then 5 minutes of quiet work time to complete the last two questions. Follow with a whole-class discussion.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Chunk this task into more manageable parts to support students who benefit from support with organizational skills. Check in with students within the first 3-5 minutes of work time to see how they calculate the distance between each point and 11. If necessary, remind students that distances are positive.

Supports accessibility for: Organization; Attention

Anticipated Misconceptions

Some students might write negative values for distances between the mean and points to the left of the mean. They might recall looking at distances between 0 and numbers to the left of it in a previous unit and mistakenly think that numbers to the left of the mean would have a negative distance from the mean. Remind students that distances are always positive; the answer to “How far away?” or “How many units away?” cannot be a negative number.

Student Task Statement

Here is the data set from an earlier lesson showing how long it takes for Diego to walk to school, in minutes, over 5 days. The mean number of minutes is 11.

12 7 13 9 14

1. Represent Diego’s data on a dot plot. Mark the location of the mean with a triangle.
2. The mean can also be seen as a **measure of center** that balances the points in a data set. If we find the distance between every point and the mean, add the distances on each side of the mean, and compare the two sums, we can see this balancing.
 - a. Record the distance between each point and 11 and its location relative to 11.

time in minutes	distance from 11	left of 11 or right of 11?
12		
7		
13		
9		
14		

- b. Sum of distances left of 11: _____ Sum of distances right of 11: _____

What do you notice about the two sums?

3. Can another point that is *not* the mean produce similar sums of distances?
Let's investigate whether 10 can produce similar sums as those of 11.

a. Complete the table with the distance of each data point from 10.

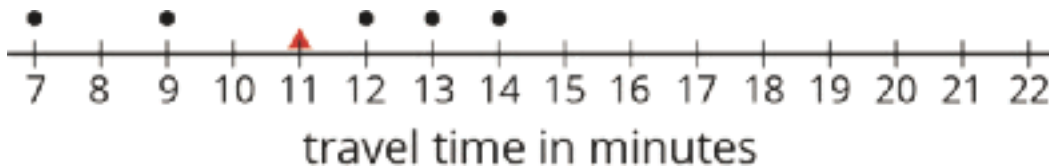
time in minutes	distance from 10	left of 10 or right of 10?
12		
7		
13		
9		
14		

b. Sum of distances left of 10: _____ Sum of distances right of 10: _____

What do you notice about the two sums?

4. Based on your work so far, explain why the mean can be considered a balance point for the data set.

Student Response



- 1.
- 2.

a.

time in minutes	distance from 11	left of 11 or right of 11?
12	1	right
7	4	left
13	2	right
9	2	left
14	3	right

b. The sum of the distances to the left of 11 is $4 + 2 = 6$. The sum of the distances to the right of 11 is $1 + 2 + 3 = 6$. The two sums are equal.

3. a.

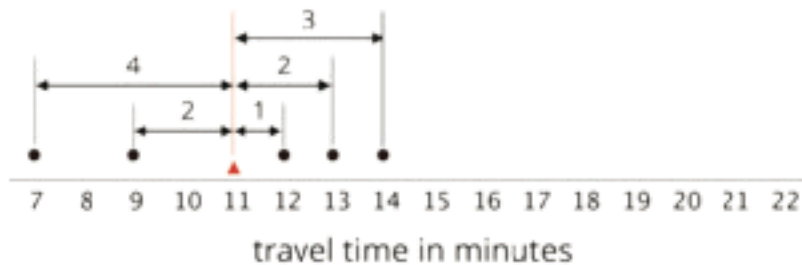
time in minutes	distance from 10	left of 11 or right of 10?
12	2	right
7	3	left
13	3	right
9	1	left
14	4	right

b. The sum of the distances to the left of 10 is $3 + 1 = 4$. The sum of the distances to the right of 10 is $2 + 3 + 4 = 9$. The two sums are not equal.

4. The sum of distances to the left of the mean is equal to the sum of distances to the right of the mean, so the mean balances the data values that are larger and those that are smaller. If a number is not the mean of the data set, then the sum of distances to the left and the sum of the distances to the right of it are not equal.

Activity Synthesis

Select a couple of students to share their observations on the distances between Diego's mean travel time and other points. To facilitate discussion, display this dot plot (with the distances labeled) for all to see. Discuss how the sums of distances change when different points are chosen as a reference from which deviations are measured.



Ask:

- “How are the sums of distances to mean (11 minutes) and the sums of distances to another point other than the mean (e.g., 10 minutes) different?”
- “If you choose another point or location on the number line, would it produce equal sums of distances to the left and to the right?”

Highlight the idea that only the mean could produce an equal sum of distances. Remind students that they have previously described centers of data sets. Explain that the mean is used as a *measure of center* of a distribution because it balances the values in a data set. Because data points that are greater than the mean balance with those that are less than the mean, the mean is used to describe what is typical for a data set.

Access for English Language Learners

Writing, Speaking, Listening: MLR1 Stronger and Clearer Each Time. Use this routine to give students a structured opportunity to revise and refine their initial response to “Based on your work so far, explain why the mean can be considered a balance point for the data set.” Display prompts for feedback that will help students strengthen their ideas and clarify their language. For example, “Can you include details about the sum of the distances from the balancing point that are important?” and “What did you mean when you said . . . ?” Invite students to build on to their initial response by incorporating some of the feedback. This will help students refine their interpretation of the mean as a balancing point of a data set before the whole-group discussion.

Design Principle(s): Support sense-making

10.3 Travel Times (Part 2)

15 minutes

This activity serves two key purposes: to reinforce the idea of the mean as a balance point and a measure of center of a distribution, and to introduce the idea that distances of data points from the mean can help us describe variability in data, which prepares students to think about mean absolute deviation in the next lesson. Students also practice calculating mean of a distribution and interpreting it in context.

Unlike in previous activities, students are given less scaffolding for finding both the mean and the sums of distances from the mean. As students work, notice those who may need additional prompts to perform these tasks. Also listen for students’ explanations on what a larger mean tells us in this context. Identify those who can clearly distinguish how the mean differs from deviations from the mean.

Addressing

- 6.SP.B.5.c

Instructional Routines

- MLR5: Co-Craft Questions

Launch

Keep students in groups of 2. Give students 5 minutes of quiet work time to complete the first set of questions and then 2–3 minutes to discuss their responses with their partner before working on the second set of questions together.

The term “variation” is used in student text for the first time. If needed, explain that it has a similar meaning as “variability” and refers to how different or alike the data values are.

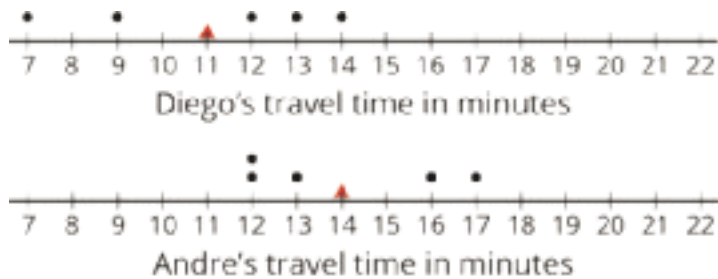
Access for English Language Learners

Representing, Conversing: MLR5 Co-craft Questions. Before revealing the task, display the image of Diego's and Andres's dot plots, and only the first sentence of the problem statement. Ask students to write a list of mathematical questions that could be asked about what they see. Invite students to share their questions with a partner before selecting 2–3 to share with the class. Listen for questions that use the terms ‘mean’, ‘spread’, or ‘center’ and highlight where these are represented in the dot point. This helps students use mathematical language related to representing distributions of data sets and to understand the context of this problem prior to be asked to reason about the different quantities in the situation.

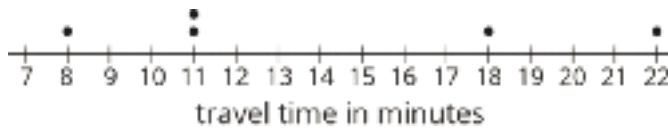
Design Principle(s): Cultivate conversation; Maximize meta-awareness

Student Task Statement

1. Here are dot plots showing how long Diego's trips to school took in minutes—which you studied earlier—and how long Andre's trips to school took in minutes. The dot plots include the means for each data set, marked by triangles.



- a. Which of the two data sets has a larger mean? In this context, what does a larger mean tell us?
 - b. Which of the two data sets has larger sums of distances to the left and right of the mean? What do these sums tell us about the variation in Diego's and Andre's travel times?
2. Here is a dot plot showing lengths of Lin's trips to school.



- a. Calculate the mean of Lin's travel times.
- b. Complete the table with the distance between each point and the mean as well whether the point is to the left or right of the mean.

time in minutes	distance from the mean	left or right of the mean?
22		
18		
11		
8		
11		

- c. Find the sum of distances to the left of the mean and the sum of distances to the right of the mean.
- d. Use your work to compare Lin's travel times to Andre's. What can you say about their average travel times? What about the variability in their travel times?

Student Response

1.
 - a. Andre's data set has a larger mean, since Andre's mean is 14 minutes, and Diego's mean is 11 minutes. In the context of this problem, this implies Andre's average travel time is longer than Diego's average.
 - b. The sum of distances to the left and right of the mean for Diego's data set is $(11 - 7) + (11 - 9) + (12 - 11) + (13 - 11) + (14 - 11) = 4 + 2 + 1 + 2 + 3 = 12$. The sum of distances for Andre's data set is $2(14 - 12) + (14 - 13) + (16 - 14) + (17 - 14) = 2 \cdot 2 + 1 + 2 + 3 = 10$. Diego's sum is greater than Andre's implying that the variability in of Diego's travel times is greater.
2.
 - a. The mean is $\frac{8+11+11+18+22}{5} = \frac{70}{5} = 14$

b.

time in minutes	distance from the mean	left or right of the mean?
22	8	right
18	4	right
11	3	left
8	6	left
11	3	left

c. Sum of distances to the left: $3 + 3 + 6 = 12$. Sum of distances to the right: $8 + 4 = 12$.

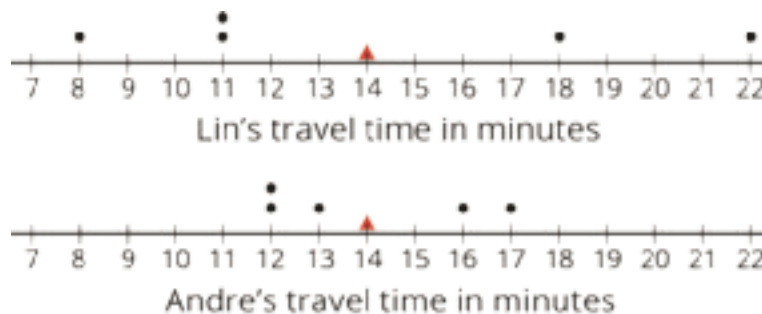
d. Answers vary. Sample response: Lin's average travel time is the same as Andre's (both have a mean of 14 minutes), and these are both greater than Diego's. The sum of the distances of data points from Lin's mean is much larger than Andre's and Diego's sums. I think this implies that Lin's travel times are much more varied than Andre's and Diego's travel times.

Activity Synthesis

The two big ideas to emphasize during discussion are: what the means tell us in this context, and what the sums of distances to either side of each mean tell us about the travel times.

Select a couple of students to share their analyses of Diego and Andre's travel times. After each student explains, briefly poll the class for agreement or disagreement. If one or more students disagree with an analysis, ask for their reasoning and alternative explanations.

Then, focus the conversation how Lin and Andre's travel times compare. Display the dot plots of their travel times for all to see.



Discuss:

- "How do the data points in Lin's dot plot compare to those in Andre's?"
- "How do their means compare? How do their sums of distances from the mean compare?"
- "What do the sums of distances tell us about the travel times?"

- “If more than half of Lin’s data points are far from the mean of 14 minutes, is the mean still a good description of her typical travel time? Why or why not?”

Students should see that larger distances from the mean suggest greater variability in the travel times. Even though both students have the same average travel time (both 14 minutes), Lin’s travel times are much more varied than Andre’s. A couple of Lin’s travel times are a lot longer or shorter than 14 minutes. Overall, her data points are within 6–8 minutes of the mean. For Andre, all of his data points are within 3 minutes of the mean.

The last discussion question prepares students to think about a different way to measure the center of a distribution in upcoming lessons.

Lesson Synthesis

In this lesson, we learn that the mean can be interpreted as the balance point of a distribution.

- “How does the mean balance the distribution of a data set?”
- “How can a dot plot help us make sense of this interpretation?”
- “Could another value—besides the mean—balance a data distribution? How can we tell?”

We also learn that the mean is used as a **measure of center** of a distribution, or a number that summarizes the center of a distribution.

- “Why might it make sense for the mean to be a number that describes the center of a distribution?”
- “In earlier lessons, we had used an estimate of the center of a distribution to describe what is typical or characteristic of a group. Why might it make sense to use the mean to describe a typical feature of a group?”

10.4 Text Messages

Cool Down: 5 minutes

Addressing

- 6.SP.B.5.c

Student Task Statement

The three data sets show the number of text messages sent by Jada, Diego, and Lin over 6 days. One of the data sets has a mean of 4, one has a mean of 5, and one has a mean of 6.

Jada	4	4	4	6	6	6
------	---	---	---	---	---	---

Diego 4 5 5 6 8 8

Lin 1 1 2 2 9 9

1. Which data set has which mean? What does this tell you about the text messages sent by the three students?
2. Which data set has the greatest variability? Explain your reasoning.

Student Response

1. Jada's mean is 5, since $\frac{4+4+4+6+6+6}{6} = \frac{30}{6} = 5$. Diego's mean is 6, since $\frac{4+5+5+6+8+8}{6} = \frac{36}{6} = 6$. Lin's mean is 4, since $\frac{1+1+2+2+9+9}{6} = \frac{24}{6} = 4$. On average, Diego sent the most text messages per day, and Lin sent the fewest text messages per day.
2. Answers vary. Sample response: Lin's data had the highest variability. All data points lie far away from the mean; the sum of the differences is the largest.

Student Lesson Summary

The mean is often used as a **measure of center** of a distribution. This is because the mean of a distribution can be seen as the "balance point" for the distribution. Why is this a good way to think about the mean? Let's look at a very simple set of data on the number of cookies that each of eight friends baked:

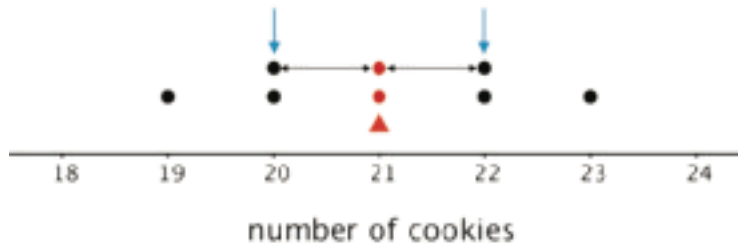
19 20 20 21 21 22 22 23

Here is a dot plot showing the data set.

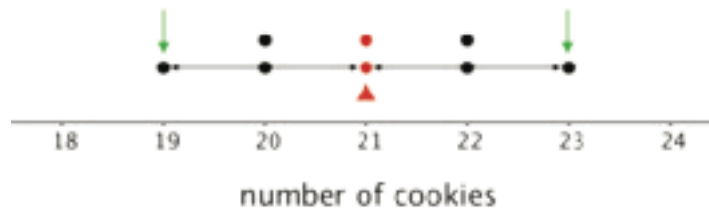


The distribution shown is completely symmetrical. The mean number of cookies is 21, because $(19 + 20 + 20 + 21 + 21 + 22 + 22 + 23) \div 8 = 21$. If we mark the location of the mean on the dot plot, we can see that the data points could balance at 21.

In this plot, each point on either side of the mean has a mirror image. For example, the two points at 20 and the two at 22 are the same distance from 21, but each pair is located on either side of 21. We can think of them as balancing each other around 21.



Similarly, the points at 19 and 23 are the same distance from 21 but are on either side of it. They, too, can be seen as balancing each other around 21.



We can say that the distribution of the cookies has a center at 21 because that is its balance point, and that the eight friends, on average, baked 21 cookies.

Even when a distribution is not completely symmetrical, the distances of values below the mean, on the whole, balance the distances of values above the mean.

Glossary

- measure of center

Lesson 10 Practice Problems

Problem 1

Statement

On school days, Kiran walks to school. Here are the lengths of time, in minutes, for Kiran's walks on 5 school days:

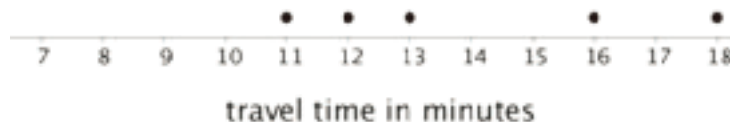
16 11 18 12 13

- Create a dot plot for Kiran's data.
- Without calculating, decide if 15 minutes would be a good estimate of the mean. If you think it is a good estimate, explain your reasoning. If not, give a better estimate and explain your reasoning.
- Calculate the mean for Kiran's data.
- In the table, record the distance of each data point from the mean and its location relative to the mean.

time in minutes	distance from the mean	left or right of the mean?
16		
11		
18		
12		
13		

- Calculate the sum of all distances to the left of the mean, then calculate the sum of distances to the right of the mean. Explain how these sums show that the mean is a balance point for the values in the data set.

Solution



- Answers vary. Sample response: 15 minutes seems to be a little too high an estimate for the mean, because (looking at the dot plot) the sum of the distances between the mean and the points to its left seems to be greater than the sum of the distances between it and the points

to its right. A better estimate would be 14 minutes, because the sums of distances to its left and the right would be more balanced.

c. $\frac{11+12+13+16+18}{5} = 14.$

d.

time in minutes	distance from mean	direction: left or right of mean?
16	2	right
11	3	left
18	4	right
12	2	left
13	1	left

e. The sum of the distances to the left: $3 + 2 + 1 = 6$. The sum of the distances to the right: $4 + 2 = 6$. The sum of distances on the left of the mean is equal to the sum of distances to the right of the mean, which tells us that the data values are balanced on the mean.

Problem 2

Statement

Noah scored 20 points in a game. Mai's score was 30 points. The mean score for Noah, Mai, and Clare was 40 points. What was Clare's score? Explain or show your reasoning.

Solution

70 points. Reasoning varies. Sample reasoning:

- Clare would need to have a score that would be 30 points to the right of the mean score of 40. This score would balance the 30 points at the left of the mean.
- If the mean score was 40 points, Noah's score was 20 points short and Mai's was 10 points short. Clare's score must be 30 points above the mean so that when the points are distributed each person's share is 40 points.

Problem 3

Statement

Compare the numbers using $>$, $<$, or $=$.

a. -2 ____ 3

a. 3 ____ -4

a. 7 ____ -11

b. $|-12|$ ____ $|15|$

b. $|15|$ ____ $|-12|$

b. -4 ____ $|5|$

Solution

- a. <
- b. <
- c. >
- d. >
- e. >
- f. <

(From Unit 7, Lesson 7.)

Problem 4

Statement

- a. Plot $\frac{2}{3}$ and $\frac{3}{4}$ on a number line.
- b. Is $\frac{2}{3} < \frac{3}{4}$, or is $\frac{3}{4} < \frac{2}{3}$? Explain how you know.

Solution

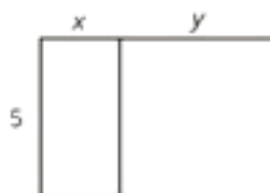
- a. The number line should show $\frac{2}{3}$ to the left of $\frac{3}{4}$, and both closer to 1 than to 0.
- b. $\frac{2}{3} < \frac{3}{4}$ because $\frac{2}{3}$ is to the left of $\frac{3}{4}$ on the number line.

(From Unit 7, Lesson 3.)

Problem 5

Statement

Select all the expressions that represent the total area of the large rectangle.



- A. $5(x + y)$
- B. $5 + xy$
- C. $5x + 5y$
- D. $2(5 + x + y)$
- E. $5xy$

Solution

["A", "C"]

(From Unit 6, Lesson 10.)

Lesson 11: Variability and MAD

Goals

- Calculate the mean absolute deviation (MAD) for a data set and interpret what it tells us about the situation.
- Compare and contrast (in writing) distributions that have the same mean, but different amounts of variability.
- Comprehend that “mean absolute deviation (MAD)” is a measure of variability, i.e., a single number summarizing how spread out the data set is.

Learning Targets

- I can find the MAD for a set of data.
- I know what the mean absolute deviation (MAD) measures and what information it provides.

Lesson Narrative

In a previous lesson, students computed and interpreted distances of data points from the mean. In this lesson, they take that experience to make sense of the formal idea of **mean absolute deviation (MAD)**. Students learn that the MAD is the average distance of data points from the mean. They use their knowledge of how to calculate and interpret the mean to calculate (MP8) and interpret (MP2) the MAD.

Students also learn that we think of the MAD as a *measure of variability* or a *measure of spread* of a distribution. They compare distributions with the same mean but different MADs, and recognize that the centers are the same but the distribution with the larger MAD has greater variability or spread.

Alignments

Addressing

- 6.SP.A.2: Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.
- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

Instructional Routines

- MLR1: Stronger and Clearer Each Time

- MLR2: Collect and Display
- MLR8: Discussion Supports
- Poll the Class

Required Materials

Decks of playing cards

Required Preparation

If students are playing the optional Game of 22, prepare one standard deck of 52 playing cards for every 2–3 students.

Student Learning Goals

Let's study distances between data points and the mean and see what they tell us.

11.1 Shooting Hoops (Part 1)

Warm Up: 5 minutes

The purpose of this warm-up is for students to first reason about the mean of a data set without calculating and then practice calculating mean. The context will be used in an upcoming activity in this lesson so this warm-up familiarizes students with the context for talking about deviation from the mean.

In their predictions, students may think that Elena will have the highest mean, because she has a few very high scores (7, 8, and 9 points). They may also think that Lin and Jada will have very close means because they each have 5 higher scores than one another and their other scores are the same. Even though each player has the same mean, all of these ideas are reasonable things for students to consider when looking at the data. Record and display their predictions without further questions until they have calculated and compared the mean of their individual data sets.

Addressing

- 6.SP.A.3
- 6.SP.B.5.c

Instructional Routines

- Poll the Class

Launch

Arrange students in groups of 3. Display the data sets for all to see. Ask students to predict which player has the largest mean and which has the smallest mean. Give students 1 minute of quiet think time and then poll students on the player who they think has the largest and smallest mean. Ask a few students to share their reasoning.

Tell each group member to calculate the mean of the data set for one player in the task, share their work in the small group, and complete the remaining questions.

Student Task Statement

Elena, Jada, and Lin enjoy playing basketball during recess. Lately, they have been practicing free throws. They record the number of baskets they make out of 10 attempts. Here are their data sets for 12 school days.

Elena

4 5 1 6 9 7 2 8 3 3 5 7

Jada

2 4 5 4 6 6 4 7 3 4 8 7

Lin

3 6 6 4 5 5 3 5 4 6 6 7

1. Calculate the mean number of baskets each player made, and compare the means. What do you notice?
2. What do the means tell us in this context?

Student Response

1. Elena's mean score is $\frac{4+5+1+6+9+7+2+8+3+3+5+7}{12} = 5$. Jada's mean score is $\frac{2+4+5+4+6+6+4+7+3+4+8+7}{12} = 5$. Lin's mean score is $\frac{3+6+6+4+5+5+3+5+4+6+6+7}{12} = 5$. I noticed that all three players have the same mean score.
2. Answers vary. Sample explanation: The means show that all three students make, on average, half of the 10 attempts to get the basketball in the hoop.

Activity Synthesis

Ask students to share the mean for each player's data set. Record and display their responses for all to see. After each student shares, ask the class if they agree or disagree and what the mean tells us in this context. If the idea that the means show that all three students make, on average, half of the 10 attempts to get the basketball in the hoop does not arise, make that idea explicit.

If there is time, consider revisiting the predictions and asking how the mean of Elena's data set can be the same as the others when she more high scores?

11.2 Shooting Hoops (Part 2)

15 minutes

In this activity, students continue to develop their understanding of what could be considered typical for a group as well as variability in a data set. Students compare distributions with the same mean but different spreads and interpret them in the context of a situation. The context given here (basketball score) prompts them to connect the mean to the notion of how “well” a player plays in general, and deviations from the mean to how “consistently” that player plays.

They encounter the idea of calculating the *average absolute deviation from the mean* as a way to describe variability in data.

Addressing

- 6.SP.A.2
- 6.SP.B.5.c

Instructional Routines

- MLR2: Collect and Display

Launch

Arrange students in groups of 3–4. Give groups 6–7 minutes to answer the questions and follow with a whole-class discussion.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Begin with a small-group or whole-class demonstration and think aloud to remind students how calculate mean using data from a dot plot. Keep the worked-out calculations on display for students to reference as they work.

Supports accessibility for: Memory; Conceptual processing

Access for English Language Learners

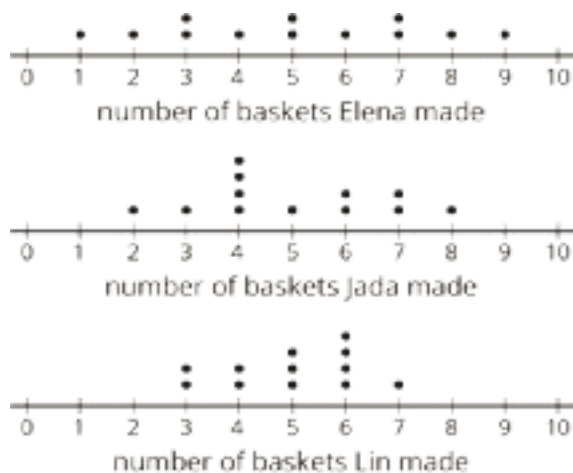
Listening, Speaking: MLR2 Collect and Display. Use this routine to collect student ideas and language about playing “well” and “consistently” as they discuss the players’ data in the three dot plots. Record students’ language on a display for all to see. This will help students connect their reasoning about the visual representations of data to statistical terms such as mean, deviation from the mean, and variability in data.

Design Principle(s): Support sense-making

Student Task Statement

Here are the dot plots showing the number of baskets Elena, Jada, and Lin each made over 12 school days.

1. On each dot plot, mark the location of the mean with a triangle (Δ). Then, contrast the dot plot distributions. Write 2–3 sentences to describe the shape and spread of each distribution.



2. Discuss the following questions with your group. Explain your reasoning.
 - a. Would you say that all three students play equally well?
 - b. Would you say that all three students play equally consistently?
 - c. If you could choose one player to be on your basketball team based on their records, who would you choose?

Student Response

1. Each dot plot should show a triangle at 5. Descriptions vary. Sample response: The data distributions for all players are centered at 5. Elena's data set is symmetric and very spread out; she has scores between 1 and 9 points. Jada's data set is not symmetric. Her data is less spread out than Elena's; they span between 2 and 8. Lin's data set has the narrowest spread, spanning from 3 to 7.
2. Answers vary.

Activity Synthesis

The purpose of the discussion is to highlight that the center of the distribution is not always the only consideration when discussing data. The variability or spread can also influence how we understand the data.

There are many ways to answer the second set of questions. Invite students or groups who have different interpretations of "playing well" and "playing consistently" to share their thinking. Allow as many interpretations to be shared as time permits. Discuss:

- "How might we use the given data to quantify 'playing well' and 'playing consistently?'"

- “Is there a way to describe variability and consistency in playing precisely and in an objective way (rather than using broad, verbal descriptions)?”

Explain that we can describe variability more formally and precisely—using a number to sum it up; we will look at how to do so in the next activity.

11.3 Shooting Hoops (Part 3)

15 minutes

In the previous activity, students evaluated the performance of three students based on the mean and variability. Here they learn the term *mean absolute deviation (MAD)* as a way to quantify variability and calculate it by finding distances between the mean and each data value. Students compare data sets with the same mean but different MADs and interpret the variations in context.

While this process of calculating MAD involves taking the absolute value of the difference between each data point and the mean, this formal language is downplayed here. Instead, the idea of “finding the distance,” which is always positive, is used. This is done for a couple of reasons. One reason is to focus students' attention on the statistical work rather than on terminology or symbolic work. Another reason is that finding these differences may involve operations with signed numbers, which are not expected in grade 6.

Addressing

- 6.SP.B.5.c

Instructional Routines

- MLR8: Discussion Supports

Launch

Remind students in the last lesson they found distance between each data point and the mean, and found that the sum of those distances on the left and the sum on the right were equal, which allows us to think of the mean as the balancing point or the center of the data. Explain that the distance between each point and the mean can tell us something else about a distribution.

Arrange students in groups of 2. Give students 4–5 minutes to complete the first set of questions with their partner, and then 4–5 minutes of quiet time to complete the remaining questions. Follow with a whole-class discussion.

Anticipated Misconceptions

Students may recall the previous lesson about thinking of the mean as a balance point and think that the MAD should always be zero since the left and right distances should be equal. Remind them that distances are always positive, so finding the average of these distances to the mean can only be zero if all the data points are exactly at the mean.

Student Task Statement

The tables show Elena, Jada, and Lin's basketball data from an earlier activity. Recall that the mean of Elena's data, as well as that of Jada and Lin's data, was 5.

1. Record the distance between each of Elena's scores and the mean.

Elena	4	5	1	6	9	7	2	8	3	3	5	7
distance from 5	1			1								

Now find *the average of the distances* in the table. Show your reasoning and round your answer to the nearest tenth.

This value is the **mean absolute deviation (MAD)** of Elena's data.

Elena's MAD: _____

2. Find the mean absolute deviation of Jada's data. Round it to the nearest tenth.

Jada	2	4	5	4	6	6	4	7	3	4	8	7
distance from 5												

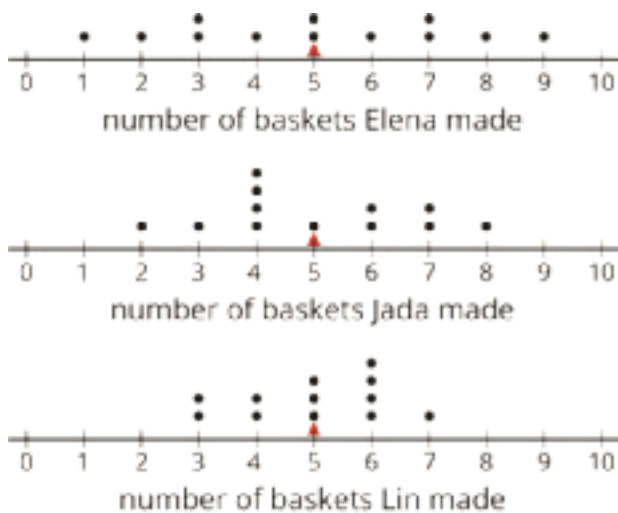
Jada's MAD: _____

3. Find the mean absolute deviation of Lin's data. Round it to the nearest tenth.

Lin	3	6	6	4	5	5	3	5	4	6	6	7
distance from 5												

Lin's MAD: _____

4. Compare the MADs and dot plots of the three students' data. Do you see a relationship between each student's MAD and the distribution on her dot plot? Explain your reasoning.



Student Response

1.

Elena	4	5	1	6	9	7	2	8	3	3	5	7
distance from 5	1	0	4	1	4	2	3	3	2	2	0	2

Elena's MAD: $2. \frac{1+0+4+1+4+2+3+3+2+2+0+2}{12} = \frac{24}{12} = 2$

2.

Jada	2	4	5	4	6	6	4	7	3	4	8	7
distance from 5	3	1	0	1	1	1	1	2	2	1	3	2

Jada's MAD: $1.5. \frac{3+1+0+1+1+1+1+2+2+1+3+2}{12} = 1.5$

3.

Lin	3	6	6	4	5	5	3	5	4	6	6	7
distance from 5	2	1	1	1	0	0	2	0	1	1	1	2

Lin's MAD: $1. \frac{2+1+1+1+0+0+2+0+1+1+1+2}{12} = 1$

4. Answers vary. Sample response: Yes, I see a relationship between the MAD and the distribution of data. The largest MAD value corresponds to the dot plot with the widest spread. The smallest MAD value corresponds to the dot plot with the narrowest spread.

Are You Ready for More?

Invent another data set that also has a mean of 5 but has a MAD greater than 2. Remember, the values in the data set must be whole numbers from 0 to 10.

Student Response

Answers vary. Sample response: 0, 0, 0, 0, 0, 0, 10, 10, 10, 10, 10, 10

Activity Synthesis

During discussion, highlight that finding how far away, on average, the data points are from the mean is a way to describe the variability of a distribution. Discuss:

- “What can we say about a data set whose data points have very small distances from the mean?”
- “What about a data set with points that show large distances from the mean?”
- “Does a data set with smaller distances (and therefore smaller average distances) show less or more variability?”
- “What do MAD values of 2, 1.5, and 1 mean in this context?”

Access for Students with Disabilities

Representation: Develop Language and Symbols. Create a display of important terms and vocabulary. Include the following term and maintain the display for reference throughout the unit: mean absolute deviation (MAD). Invite students to suggest language or diagrams to include on the display that will support their understanding of this term.

Supports accessibility for: Memory; Language

Access for English Language Learners

Listening, Speaking, Representing: MLR8 Discussion Supports. To provide a support as students explain their reasoning, invite students to use the sentence frame: “The largest (or smallest) mean absolute deviation value matches with the dot plot with the ____ spread because . . .”. This will help students produce explanations connecting visual and numeric representations of data distribution.

Design Principle(s): Support sense-making

11.4 Game of 22

Optional: 15 minutes

This optional activity uses a game to help students develop the idea of variability. Use this activity if students could benefit from more concrete experiences around the idea of distance from the mean. Students will draw from a standard deck of playing cards and find the sum. The player with the least mean distance from 22 wins the round.

Addressing

- 6.SP.B.5.c

Instructional Routines

- MLR1: Stronger and Clearer Each Time

Launch

Students in groups of 2–3. A deck of playing cards per group. Play 4–6 rounds.

Access for Students with Disabilities

Representation: Internalize Comprehension. Begin with a demonstration of the steps for the activity to support understanding. Invite a group of students to play an example round while the class observes.

Supports accessibility for: Conceptual processing; Visual-spatial processing

Student Task Statement

Your teacher will give your group a deck of cards. Shuffle the cards, and put the deck face down on the playing surface.

- To play: Draw 3 cards and add up the values. An ace is a 1. A jack, queen, and king are each worth 10. Cards 2–10 are each worth their face value. If your sum is anything other than 22 (either above or below 22), say: “My sum deviated from 22 by ____,” or “My sum was off from 22 by ____.”
- To keep score: Record each sum and each distance from 22 in the table. After five rounds, calculate the average of the distances. The player with the lowest average distance from 22 wins the game.

player A	round 1	round 2	round 3	round 4	round 5
sum of cards					
distance from 22					

Average distance from 22: _____

player B	round 1	round 2	round 3	round 4	round 5
sum of cards					
distance from 22					

Average distance from 22: _____

player C	round 1	round 2	round 3	round 4	round 5
sum of cards					
distance from 22					

Average distance from 22: _____

Whose average distance from 22 is the smallest? Who won the game?

Student Response

Playing results vary.

Activity Synthesis

Ask students to think about how *average distance from a number* can be used to summarize variability and invite a couple of students to share their thinking.

In the game, we can think of the player with the least average distance from 22 as having cards that are, on the whole, closest to 22 or the “least different” from 22. By the same token, a player with the greatest average distance from 22 can be seen as having cards that are, on the whole, farthest away from 22 or the “most different” from 22. Connect this to the idea that a data set with a large MAD means it has many values that vary from what we could consider a typical member of the group.

Access for English Language Learners

Writing, Conversing: MLR1 Stronger and Clearer Each Time. Use this routine with 2–3 successive pair shares to give students a structured opportunity to revise and refine their response to “How can the average distance from a number to the mean be used to summarize variability?” Provide students with prompts for feedback (e.g., “Could you give a specific example from a round of Game of 22?” or, “Could you use the term ‘mean absolute deviation’ to explain your example further?”). Students can borrow ideas and language from each other to strengthen their final product. This will help students strengthen their ideas and clarify their language.
Design Principle(s): Optimize output (for explanation); Cultivate conversation

Lesson Synthesis

In an earlier lesson, we learned that the mean can tell us about what is typical for (or what is characteristic of) a data set because it measures the center of a distribution. In this lesson, we learn that we can also use a measure of spread to tell us about how much the values in a data set vary.

- “How did we measure spread?”
- “We learned about a measure called the mean absolute deviation (or MAD). What is the meaning of this term?”
- “What does the MAD tell us?” (The average distance between data points and the mean. It tells us how spread out the data values are.)
- “How do we find the MAD?”
- “How is MAD related to the variability of a data set?”

11.5 Text Messages, Again

Cool Down: 5 minutes

Addressing

- 6.SP.B.5.c

Student Task Statement

These three data sets show the number of text messages sent by Jada, Diego, and Lin over 6 days as well as the mean number of text messages sent by each student per day.

Jada	4	4	4	6	6	6
------	---	---	---	---	---	---

mean: 5

Diego	4	5	5	6	8	8
-------	---	---	---	---	---	---

mean: 6

Lin	1	1	2	2	9	9
-----	---	---	---	---	---	---

mean: 4

1. Predict which data set has the largest MAD and which has the smallest MAD.
2. Compute the MAD for each data set to check your prediction.

Student Response

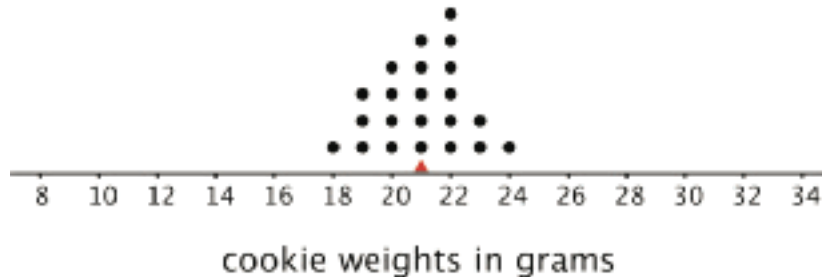
1. Lin's data set will have the largest MAD, because the data has the most variability. Jada's data set will have the smallest MAD, because the data has the least variability.

2. Jada's MAD is $\frac{1+1+1+1+1}{6} = 1$; Diego's MAD is $\frac{2+1+1+0+2+2}{6} = 1.33$; Lin's MAD is $\frac{3+3+2+2+5+5}{6} = 3.33$.

Student Lesson Summary

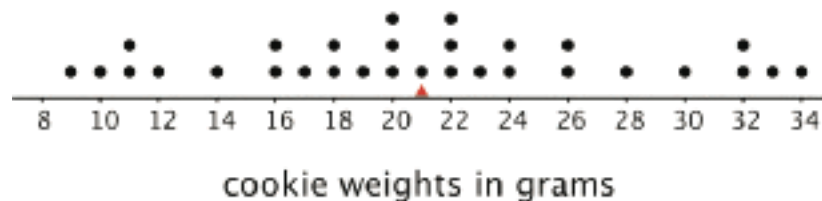
We use the mean of a data set as a measure of center of its distribution, but two data sets with the same mean could have very different distributions.

This dot plot shows the weights, in grams, of 22 cookies.



The mean weight is 21 grams. All the weights are within 3 grams of the mean, and most of them are even closer. These cookies are all fairly close in weight.

This dot plot shows the weights, in grams, of a different set of 30 cookies.

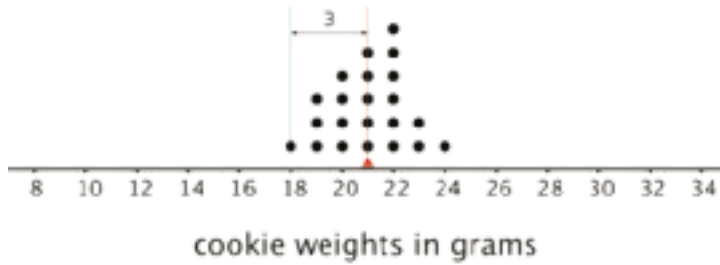


The mean weight for this set of cookies is also 21 grams, but some cookies are half that weight and others are one-and-a-half times that weight. There is a lot more variability in the weight.

There is a number we can use to describe how far away, or how spread out, data points generally are from the mean. This *measure of spread* is called the **mean absolute deviation (MAD)**.

Here the MAD tells us how far cookie weights typically are from 21 grams. To find the MAD, we find the distance between each data value and the mean, and then calculate the mean of those distances.

For instance, the point that represents 18 grams is 3 units away from the mean of 21 grams. We can find the distance between each point and the mean of 21 grams and organize the distances into a table, as shown.



weight in grams	18	19	19	19	20	20	20	20	21	21	21	21	21	21	22	22	22	22	22	22	23	23	24
distance from mean	3	2	2	2	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	2	2	3

The values in the first row of the table are the cookie weights for the first set of cookies. Their mean, 21 grams, is the *mean of the cookie weights*.

The values in the second row of the table are the *distances* between the values in the first row and 21. The mean of these distances is the *MAD of the cookie weights*.

What can we learn from the averages of these distances once they are calculated?

- In the first set of cookies, the distances are all between 0 and 3. The MAD is 1.2 grams, which tells us that the cookie weights are typically within 1.2 grams of 21 grams. We could say that a typical cookie weighs between 19.8 and 22.2 grams.
- In the second set of cookies, the distances are all between 0 and 13. The MAD is 5.6 grams, which tells us that the cookie weights are typically within 5.6 grams of 21 grams. We could say a typical cookie weighs between 15.4 and 26.6 grams.

The MAD is also called a *measure of the variability* of the distribution. In these examples, it is easy to see that a higher MAD suggests a distribution that is more spread out, showing more variability.

Glossary

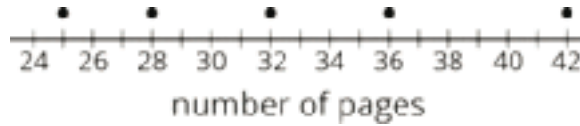
- mean absolute deviation (MAD)

Lesson 11 Practice Problems

Problem 1

Statement

Han recorded the number of pages that he read each day for five days. The dot plot shows his data.



- Is 30 pages a good estimate of the mean number of pages that Han read each day? Explain your reasoning.
- Find the mean number of pages that Han read during the five days. Draw a triangle to mark the mean on the dot plot.
- Use the dot plot and the mean to complete the table.

number of pages	distance from mean	left or right of mean
25		left
28		
32		
36		
42		

- Calculate the mean absolute deviation (MAD) of the data. Explain or show your reasoning.

Solution

- The value of 30 is not a good estimate of the mean as it would not balance the other values.
- The mean is 32.6 pages.
-

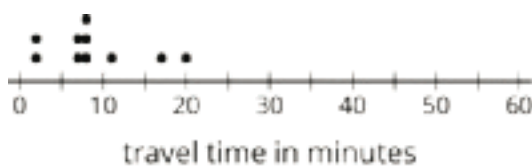
number of pages	distance from mean	left of mean or right of mean
25	7.6	left
28	4.6	left
32	0.6	left
36	3.4	right
42	9.4	right

d. The MAD is 5.12 pages: $\frac{7.6+4.6+0.6+3.4+9.4}{5} = 5.12$

Problem 2

Statement

Ten sixth-grade students recorded the amounts of time each took to travel to school. The dot plot shows their travel times.



The mean travel time for these students is approximately 9 minutes. The MAD is approximately 4.2 minutes.

- Which number of minutes—9 or 4.2—is a typical amount of time for the ten sixth-grade students to travel to school? Explain your reasoning.
- Based on the mean and MAD, Jada believes that travel times between 5 and 13 minutes are common for this group. Do you agree? Explain your reasoning.
- A different group of ten sixth-grade students also recorded their travel times to school. Their mean travel time was also 9 minutes, but the MAD was about 7 minutes. What could the dot plot of this second data set be? Describe or draw how it might look.

Solution

- 9 minutes. Explanations vary. Sample reasoning: On the dot plot, the center or balance point of the data set is located at or near 9, so that value is a good description of a typical travel time.
- Agree. Explanations vary. Sample reasoning: The MAD tells us that, on average, the travel times of the students in this group are 4.2 minutes below the mean (about 5 minutes) or 4.2 minutes above the mean (about 13 minutes), so travel times between 5 and 13 minutes are common.

- c. Answers vary. Sample response: The data points on the second dot plot would be more spread out, with more points farther from 9, because the MAD is larger. Travel times between 2 minutes and 16 minutes would be typical for this group.

Problem 3

Statement

In an archery competition, scores for each round are calculated by averaging the distance of 3 arrows from the center of the target.

An archer has a mean distance of 1.6 inches and a MAD distance of 1.3 inches in the first round. In the second round, the archer's arrows are farther from the center but are more consistent. What values for the mean and MAD would fit this description for the second round? Explain your reasoning.

Solution

Answers vary. Correct responses should have a mean greater than 1.6 inches and a MAD less than 1.3 inches.

Lesson 12: Using Mean and MAD to Make Comparisons

Goals

- Compare (orally and in writing) the means and mean absolute deviations of different distributions, specifically those with the same MAD but different means.
- Interpret the mean and mean absolute deviation (MAD) in the context of the data.

Learning Targets

- I can say what the MAD tells us in a given context.
- I can use means and MADs to compare groups.

Lesson Narrative

In this lesson, students continue to develop their understanding of the mean and MAD as measures of center and spread as well as interpret these values in context. They practice computing the mean and the MAD for distributions; compare distributions with the same MAD but different means; and interpret the mean and MAD in the context of the data (MP2).

Alignments

Addressing

- 6.NS.B.3: Fluently add, subtract, multiply, and divide multi-digit decimals using the standard algorithm for each operation.
- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
- 6.SP.B.5.d: Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Building Towards

- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

Instructional Routines

- MLR8: Discussion Supports
- Notice and Wonder
- Number Talk

Student Learning Goals

Let's use mean and MAD to describe and compare distributions.

12.1 Number Talk: Decimal Division

Warm Up: 5 minutes

This number talk helps students to review strategies for dividing a decimal number by a whole number and to build their fluency. It also prepares them to calculate mean and MAD more efficiently. While four problems are given, it may not be possible to share every strategy. Consider gathering only two or three different strategies per problem, saving most of the time for the final question.

Addressing

- 6.NS.B.3

Building Towards

- 6.SP.A.3

Instructional Routines

- MLR8: Discussion Supports
- Number Talk

Launch

Reveal one problem at a time. Give students 30 seconds of quiet think time for each problem and ask them to give a signal when they have an answer and a strategy. Keep all previous problems displayed throughout the task. Follow with a whole-class discussion.

Access for Students with Disabilities

Representation: Internalize Comprehension. To support working memory, provide students with sticky notes or mini whiteboards.

Supports accessibility for: Memory; Organization

Student Task Statement

Find the value of each expression mentally.

$$42 \div 12$$

$$2.4 \div 12$$

$$44.4 \div 12$$

$$46.8 \div 12$$

Student Response

- 3.5. Possible strategy: $36 \div 12 + 6 \div 12 = 3.5$
- 0.2. Possible strategy: $2.4 \cdot 10 \div 12 = 2$ and $2 \div 10 = 0.2$
- 3.7. Possible strategy: $42 \div 12 + 2.4 \div 12 = 3.7$
- 3.9. Possible strategy: $42 \div 12 + 2.4 \div 12 + 2.4 \div 12 = 3.9$

Activity Synthesis

Ask students to share their strategies for each problem. Record and display their explanations for all to see. To involve more students in the conversation, consider asking:

- "Who can restate ___'s reasoning in a different way?"
- "Did anyone have the same strategy but would explain it differently?"
- "Did anyone solve the problem in a different way?"
- "Does anyone want to add on to ____'s strategy?"
- "Do you agree or disagree? Why?"

Access for English Language Learners

Speaking: MLR8 Discussion Supports.: Display sentence frames to support students when they explain their strategy. For example, "First, I ____ because . . ." or "I noticed ____ so I" Some students may benefit from the opportunity to rehearse what they will say with a partner before they share with the whole class.

Design Principle(s): Optimize output (for explanation)

12.2 Which Player Would You Choose?

20 minutes

This activity allows students to practice calculating MAD and build a better understanding of what it tells us. Students continue to compare data sets with the same mean but different MADs and interpret what these differences imply in the context of the situation.

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

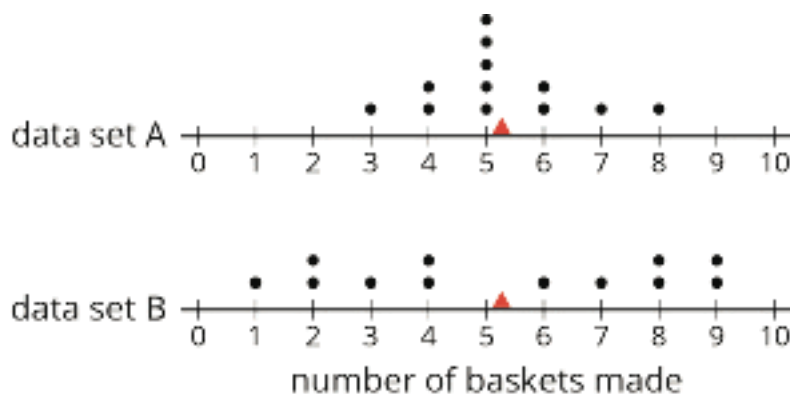
Instructional Routines

- MLR8: Discussion Supports

- Notice and Wonder

Launch

Arrange students in groups of 3–4. Before students read the task statements, display the two dot plots in the task for all to see. Give students up to 1 minute to study the dot plots and share with their group what they notice and wonder about the plots.



Next, select a few students to share their observations and questions; it is not necessary to confirm or correct students' observations or answer their questions at this point. If no one noticed or wondered about what variable the data sets show, ask students what they think the context for the data sets might be or what quantities they show. Explain to students that they will find more information in the task statement to help them compare and interpret the dot plots.

Give students 3–4 minutes of quiet work time to complete the first set of questions, and then 8–10 minutes to complete the second set with their group. Allow at least a few minutes for a whole-class discussion.

Access for Students with Disabilities

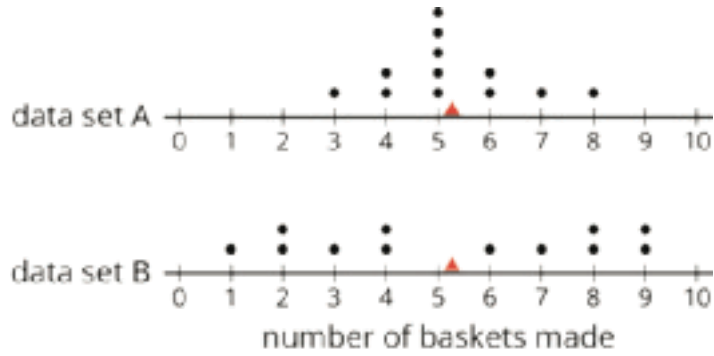
Representation: Internalize Comprehension. Activate or supply background knowledge about comparing and interpreting data sets in dot plots. Allow students to use calculators to ensure inclusive participation in the activity.

Supports accessibility for: Memory; Conceptual processing

Student Task Statement

1. Andre and Noah joined Elena, Jada, and Lin in recording their basketball scores. They all recorded their scores in the same way: the number of baskets made out of 10 attempts. Each collected 12 data points.
 - Andre's mean number of baskets was 5.25, and his MAD was 2.6.
 - Noah's mean number of baskets was also 5.25, but his MAD was 1.

Here are two dot plots that represent the two data sets. The triangle indicates the location of the mean.



- Without calculating, decide which dot plot represents Andre's data and which represents Noah's. Explain how you know.
 - If you were the captain of a basketball team and could use one more player on your team, would you choose Andre or Noah? Explain your reasoning.
2. An eighth-grade student decided to join Andre and Noah and kept track of his scores. His data set is shown here. The mean number of baskets he made is 6.

eighth-grade student	6	5	4	7	6	5	7	8	5	6	5	8
distance from 6												

- Calculate the MAD. Show your reasoning.
- Draw a dot plot to represent his data and mark the location of the mean with a triangle (Δ).
- Compare the eighth-grade student's mean and MAD to Noah's mean and MAD. What do you notice?
- Compare their dot plots. What do you notice about the distributions?
- What can you say about the two players' shooting accuracy and consistency?

Student Response

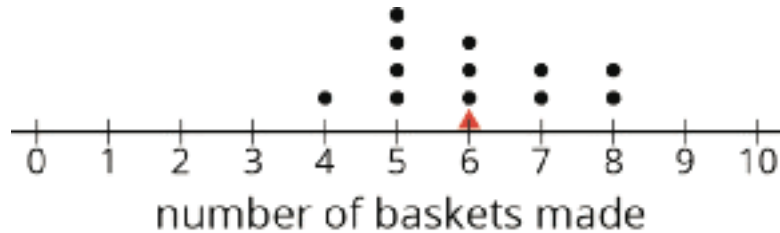
- Dot plot A represents Noah's data, and dot plot B represents Andre's data. Sample explanation: I know because Andre's data has a larger MAD than Noah's, so the data would be more spread out than Noah's.
 - Answers vary. Sample response: I would choose Noah. Both players made, on average, about 5 out of 10 baskets, but Noah is more consistent, so he's less likely to miss more than 6 out of 10 shots. Andre scored some high points a few times, but he also scored some very low ones.

2.

eighth-grade student	6	5	4	7	6	5	7	8	5	6	5	8
distance from 6	0	1	2	1	0	1	1	2	1	0	1	2

a. MAD: $\frac{0+1+2+1+0+1+1+2+1+0+1+2}{12} = 1$

b. Dot plot for the eighth-grade student:



c. Answers vary. Sample response: The eighth-grade student has a mean of 6, which is larger than Noah's mean of 5.25. Both Noah and the eighth-grade student have the same value for their MAD. Although their means are different, their data is similarly spread around their means.

d. Answers vary. Sample response: Both dot plots show roughly the same spread; the points are clustered toward the middle and are within 2–3 units away from the mean.

e. Answers vary. Sample response: The eighth-grade student has a higher mean, so on average, he shoots more accurately than Noah does. The same MAD value suggests that the two players are equally consistent. Each person typically scores within 1 point of their mean, rather than scoring some very high scores and some very low ones.

Are You Ready for More?

Invent a data set with a mean of 7 and a MAD of 1.

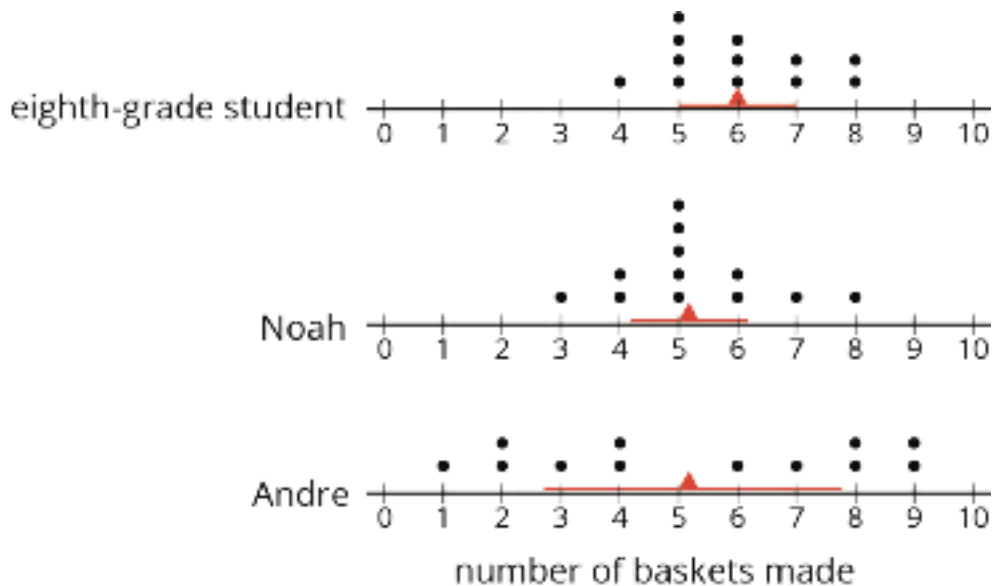
Student Response

Answers vary. Sample response: take the eighth-grade student's results and add 1 to each value.

Activity Synthesis

Select a couple of students to share their responses to the first set of questions about how they matched the dot plots to the players (Andre and Noah) and how they knew.

Then, display a completed table and the MAD for the second set of questions. Give students a moment to check their work. To facilitate discussion, help students connect MAD and the spread of data, and enable them to make comparison, consider displaying all three dot plots at the same scale and using a line segment to represent the MAD on each dot plot, as shown here.



Invite a few students to share their observations about how the means and MADs of Noah and the eighth-grade student compare. Discuss:

- “How are the distributions of points related to the mean? How are they related to the MAD?”
- “Which might be more desirable for a basketball team: a lower mean or a higher mean? Why?”
- “Which might be more desirable: a lower MAD or a higher MAD? Why?”
- “Of the three students, which one would you want on your team? Why?”

Expect students to choose different players to be on their team, but be sure they support their preferences with a reasonable explanation (MP3). Students should walk away understanding that, in this context, a higher MAD indicates more variability and less consistency in the number of shots made.

Access for English Language Learners

Speaking: MLR8 Discussion Supports. Use this routine to support whole-class discussion. For each observation that is shared, ask students to restate and/or revoice what they heard using mathematical language. Consider providing students time to restate what they hear to a partner, before selecting one or two students to share with the class. Ask the original speaker if their peer was accurately able to restate their thinking. Call students' attention to any words or phrases that helped to clarify the original statement. This will provide more students with an opportunity to produce language and build a better understanding of MAD.

Design Principle(s): Support sense-making

12.3 Swimmers Over the Years

10 minutes

In this activity, students continue to practice interpreting the mean and the MAD and to use them to answer statistical questions. A new context is introduced, but students should continue to consider both the center and variability of the distribution as ways of thinking about what is typical for a set of data and how consistent the data tends to be.

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Instructional Routines

- MLR8: Discussion Supports

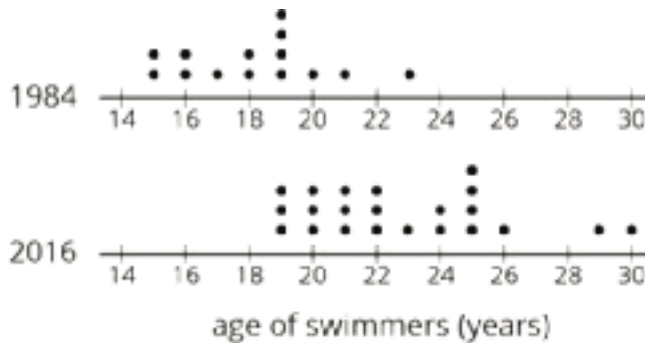
Launch

Give students 5–7 minutes of quiet work time. Ask students to consider drawing a triangle and a line segment on each dot plot in the last question to represent the mean and MAD for each data set (as was done in an earlier lesson).

Student Task Statement

In 1984, the mean age of swimmers on the U.S. women’s swimming team was 18.2 years and the MAD was 2.2 years. In 2016, the mean age of the swimmers was 22.8 years, and the MAD was 3 years.

1. How has the average age of the women on the U.S. swimming team changed from 1984 to 2016? Explain your reasoning.
2. Are the swimmers on the 1984 team closer in age to one another than the swimmers on the 2016 team are to one another? Explain your reasoning.
3. Here are dot plots showing the ages of the women on the U.S. swimming team in 1984 and in 2016. Use them to make two other comments about how the women’s swimming team has changed over the years.



Student Response

1. Answers vary. Sample response: The average age of the women's swimming team has increased by about 4.6 years over the past three decades, from 18.2 to 22.8 years old. The swimmers representing the U.S. in 2016 are older, on average, than those in 1984.
2. Yes. Explanations vary. Sample response: In 2016, the average distance from the mean age was 3 years. In 1984, the mean absolute deviation was 2.2 years, which means that the swimmers were closer to one another in age.
3. Answers vary. Students could comment on the number of observations, the range, shape, or distribution of data, or other features of the dot plots. Sample responses:
 - The 2016 team was much older than the 1984 team. In 2016, about half the swimmers were 22 or younger, while in 1984, all but one swimmer was 22 or younger.
 - The youngest swimmers in 2016 were 19 years, 4 years older than the youngest swimmers in 1984.
 - The oldest swimmer in 2016 was 7 years older than the oldest swimmer in 1984.

Activity Synthesis

Have a student display dot plots with the means and MADs marked on them. Invite several students to share their comments about how the composition of the swimming team has changed over the three decades. Some discussion questions:

- "What can you say about the size of the team? Has it changed?"
- "Overall, has the team gotten older, younger, or stayed about the same? How do you know?"
- "Has the age of the youngest swimmer changed? What about the age of the oldest swimmer?"
- "Has the team become more diverse in ages, in general? Or have the swimmers become more alike in their age? How do you know?"

Students should recognize that the higher mean and MAD for the 2016 swimming team tell us that the team, on the whole, has gotten older and more diverse in ages. In 1984, 18.2 years was a typical age for the swimmers. A typical age for the 2016 swimmers was 22.8 years and there was a wider range of ages represented.

Access for English Language Learners

Speaking, Listening: MLR8 Discussion Supports. Use this routine to support students as they produce statements about interpreting the mean and the MAD. Display sentence frames for students to use such as: "I know the average age changed ___ because..."; "Over the three decades, the _____ of the swimming team has changed by _____"; "I know the swimmers' ages in the year ___ are closer to one another because...." This will help students to explain their interpretations of the data sets.

Design Principle(s): Support sense-making; Optimize output (for explanation)

Lesson Synthesis

In this lesson, we look at what different means and MADs tell us in situations.

- "Suppose the mean height of the students in a class is 60 inches and the MAD is 2.5 inches. How do the mean and the MAD tell us about what is typical for the students' heights?"
- "How do two distributions compare if they have the same means but different MADs?" (Same center, different variability or spread.)
- "How do two distributions compare if they have the same MADs but different means?" (Same variability or spread, different centers.)

12.4 Travel Times Across the World

Cool Down: 5 minutes

Addressing

- 6.SP.B.5.c

Student Task Statement

Ten sixth-grade students in five different countries were asked about their travel times to school. Their responses were organized into five data sets. The mean and MAD of each data set is shown in the table.

	mean (minutes)	MAD (minutes)
United States	9	4.2
Australia	18.1	7.9
South Africa	23.5	16.2
Canada	11	8
New Zealand	12.3	5.5

1. Which group of students has the greatest variability in their travel times? Explain your reasoning.
2.
 - a. The mean of the data set for New Zealand is close to that of Canada. What does this tell us about the travel times of students in those two data sets?
 - b. The MAD of the data set for New Zealand is quite different than that of Canada. What does this tell us about the travel times of students in those two data sets?
3. The data sets for Australia and Canada have very different means (18.1 and 11 minutes) but very similar MADs. What can you say about the travel times of the students in those two data sets?

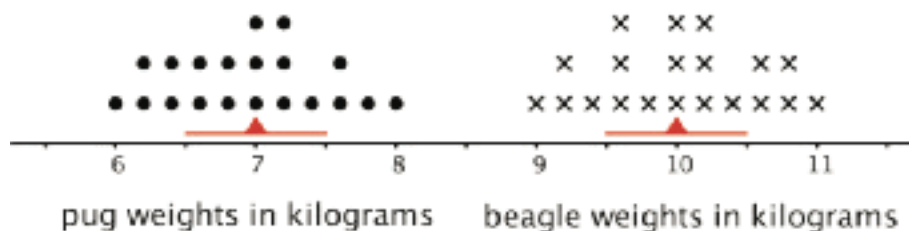
Student Response

1. South Africa, because it has the largest MAD.
2. Answers vary. Sample responses:
 - a. The average travel times for the students in New Zealand and Canada are similar. The students in New Zealand travel only about 1 minute longer than those in Canada.
 - b. The travel times of the students in New Zealand are less variable than those of the Canadian students. On average, the travel times of the students in Canada are 8 minutes longer or shorter than the mean. For New Zealand students, the travel times differ from the mean by an average of 5.5 minutes.
3. Answers vary. Sample response: On average, the students in Australia have a longer commute to school than students in Canada, but the travel times of students in both countries have the same variability. The data points are, on average, about 8 minutes from the mean.


Student Lesson Summary

Sometimes two distributions have different means but the same MAD.

Pugs and beagles are two different dog breeds. The dot plot shows two sets of weight data—one for pugs and the other for beagles.



- The mean weight for pugs is 7 kilograms, and the MAD is 0.5 kilogram.
- The mean weight for beagles is 10 kilograms, and the MAD is 0.5 kilogram.



We can say that, in general, the beagles are heavier than the pugs. A typical weight for the beagles in this group is about 3 kilograms heavier than a typical weight for the pugs.

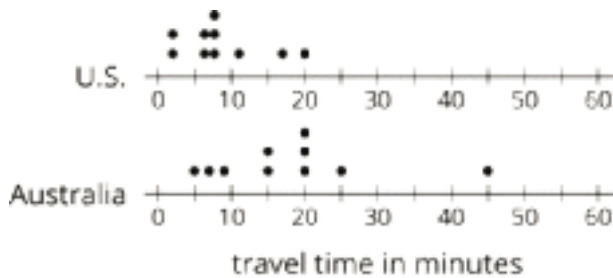
The variability of pug weights, however, is about the same as the variability of beagle weights. In other words, the weights of pugs and the weights of beagles are equally spread out.

Lesson 12 Practice Problems

Problem 1

Statement

The dot plots show the amounts of time that ten U.S. students and ten Australian students took to get to school.



Which statement is true about the MAD of the Australian data set?

- A. It is significantly less than the MAD of the U.S. data set.
- B. It is exactly equal to the MAD of the U.S. data set.
- C. It is approximately equal to the MAD of the U.S. data set.
- D. It is significantly greater than the MAD of the U.S. data set.

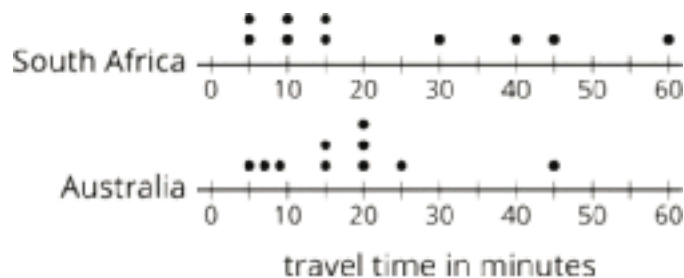
Solution

C

Problem 2

Statement

The dot plots show the amounts of time that ten South African students and ten Australian students took to get to school. Without calculating, answer the questions.



- a. Which data set has the smaller mean? Explain your reasoning.
- b. Which data set has the smaller MAD? Explain your reasoning.
- c. What does a smaller mean tell us in this context?
- d. What does a smaller MAD tell us in this context?

Solution

Responses vary. Sample responses:

- The mean of the Australian data set is smaller. The balance point of the Australian data set is less than 20 minutes. The balance point for the South African set is probably larger because four of the points are much larger than 20.
- The MAD for the Australian data set is smaller. The data points are closer together and closer to the center of the distribution.
- A smaller mean tells us that the travel times of the students in the group are shorter overall.
- A smaller MAD tells us the travel times of the students in the group are closer to each other, closer to the mean, and more alike overall.

Problem 3

Statement

Two high school basketball teams have identical records of 15 wins and 2 losses. Sunnyside High School's mean score is 50 points and its MAD is 4 points. Shadyside High School's mean score is 60 points and its MAD is 15 points.

Lin read the records of each team's score. She likes the team that had nearly the same score for every game it played. Which team do you think Lin likes? Explain your reasoning.

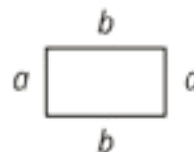
Solution

Sunnyside High School. Explanations vary. Sample explanation: The smaller MAD indicates that most of the scores for the games were close to the mean of 50 points.

Problem 4

Statement

Jada thinks the perimeter of this rectangle can be represented with the expression $a + a + b + b$. Andre thinks it can be represented with $2a + 2b$. Do you agree with either of them? Explain your reasoning.



Solution

They are both correct. $2a + 2b$ and $a + a + b + b$ are equivalent expressions because $2a$ means $a + a$ and $2b$ means $b + b$.

(From Unit 6, Lesson 8.)

Problem 5

Statement

Draw a number line.

- Plot and label three numbers between -2 and -8 (not including -2 and -8).
- Use the numbers you plotted and the symbols $<$ and $>$ to write three inequality statements.

Solution

- Answers vary. Sample response: any three numbers between -2 and -8.
- Answers vary. Sample response: $-2 > -8$, $-2 > -7$, and $-6 < -3$.

(From Unit 7, Lesson 3.)

Problem 6

Statement

Adult elephant seals generally weigh about 5,500 pounds. If you weighed 5 elephant seals, would you expect each seal to weigh exactly 5,500 pounds? Explain your reasoning.

Solution

No. Not every seal will weigh the same amount. There will be variability in the data.

(From Unit 8, Lesson 2.)

Section: Median and IQR

Lesson 13: Median

Goals

- Comprehend that the “median” is another measure of center, which uses the middle of all the values in an ordered list to summarize the data.
- Identify and interpret the median of a data set given in a table or on a dot plot.
- Informally estimate the center of a data set and then compare (orally and in writing) the mean and median with this estimate.

Learning Targets

- I can find the median for a set of data.
- I can say what the median represents and what it tells us in a given context.

Lesson Narrative

In this lesson, students consider another measure of center, the **median**, which divides the data into two groups with half of the data greater and half of the data less than the median. To find the median, they learn that the data are to be arranged in order, from least to greatest. They make use of the structure of the data set (MP7) to see that the median partitions the data into two halves: one half of the values in the data set has that value or smaller values, and the other half has that value or larger. Students learn how to find the median for data sets with both even and odd number of values.

Students engage in MP2 as they find the median of a numerical data set and interpret it in context. They begin to see that, just like the mean, the median can be used to describe what is typical in a distribution, but that it is interpreted differently than is the mean.

Alignments

Addressing

- 6.SP.B: Summarize and describe distributions.
- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

Building Towards

- 6.SP.B.5.d: Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Instructional Routines

- MLR1: Stronger and Clearer Each Time
- MLR3: Clarify, Critique, Correct
- Think Pair Share

Required Materials

Index cards

Required Preparation

For the Finding the Middle activity, each student will need an index card.

Student Learning Goals

Let's explore the median of a data set and what it tells us.

13.1 The Plot of the Story

Warm Up: 5 minutes

This warm-up reinforces students' understanding about the relationship between the mean absolute deviation (MAD) and the spread of data. In the given scenarios, the number of people attending the two events and their mean age are the same, but the MADs are different. In the first question, students interpret these measures in the context of the situations. In the second, they draw a dot plot that could represent an age distribution with the same mean and yet another MAD.

As students work and discuss, identify several students who drew dot plots that correctly meet the criteria in the second question. Ask students with different dot plots to share during whole-class discussion.

Students may need more time to make sense of how to generate their own dot plot for the second question. If it is not possible to give students additional time, consider presenting the second question at a different time.

Addressing

- 6.SP.B

Instructional Routines

- Think Pair Share

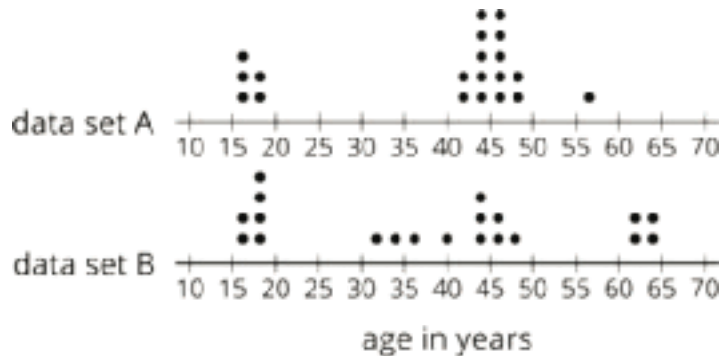
Launch

Arrange students in groups of 2. Give students 1 minute of quiet think time for the first question, and then 2–3 minutes to work on the second question with a partner. Display the following questions for all to see. Ask students to think about and discuss them before drawing their dot plots:

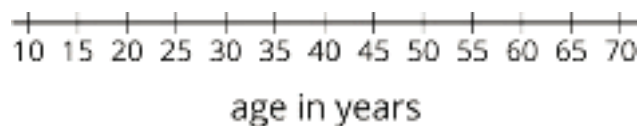
- “How many data points should be on the dot plot?”
- “How would the mean help us place the data points?”
- “How would the MAD help us place the data points?”
- “How should our dot plot compare to the dot plots of data sets A and B?”

Student Task Statement

1. Here are two dot plots and two stories. Match each story with a dot plot that could represent it. Be prepared to explain your reasoning.

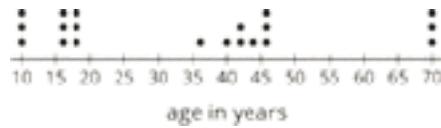


- Twenty people—high school students, teachers, and invited guests—attended a rehearsal for a high school musical. The mean age was 38.5 years and the MAD was 16.5 years.
 - High school soccer team practice is usually watched by supporters of the players. One evening, twenty people watched the team practice. The mean age was 38.5 years and the MAD was 12.7 years.
2. Another evening, twenty people watched the soccer team practice. The mean age was similar to that from the first evening, but the MAD was greater (about 20 years). Make a dot plot that could illustrate the distribution of ages in this story.



Student Response

1. Data set A goes with the high-school soccer practice story. Data set B goes with the musical performance story. Sample explanation: The data points in data set B are more spread out, so the MAD for that data set would be larger.
2. Answers vary, but should show a center of between 38 and 40 and a wider spread than data set B. Sample dot plot:



Activity Synthesis

Poll students on their response to the first question. Ask a student to explain how they matched one context to its dot plot and another student to explain the second matching context and dot plot. Record and display their responses for all to see. If possible, record their responses directly on the dot plots.

Ask selected students to share their dot plots for the second question and their reasoning. To involve more students in the conversation, consider asking some of the following questions:

- “What was the first piece of information you used to draw your dot plot? Why?”
- “How did you decide where to place your dots?”
- “How is your dot plot the same or different than the first evening of soccer practice?”
- “Do you agree or disagree with this representation of the context? Why?”
- “Do you have any questions to ask the student who drew the dot plot?”

13.2 Siblings in the House

15 minutes

The aim of this activity is to expose the limits of the mean in summarizing a data set that has gaps and values far from the center, and to motivate a need to have another measure of center.

Students first use a table of values and a dot plot to estimate a “typical” value for a data set. Then, they calculate the mean and notice that it does not match their estimate of a typical value. A closer look helps them see that when a data set contains values that are far away from the bulk of the data, or when there are gaps in the data set, the mean can be a little or a lot higher or lower than what we would consider typical for the data.

In the next activity, the median will be introduced as another measure of center of a data set.

Addressing

- 6.SP.B.5.c

Building Towards

- 6.SP.B.5.d

Instructional Routines

- MLR1: Stronger and Clearer Each Time
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 7–8 minutes of quiet work time and then 3–4 minutes to discuss their responses with a partner. If there are disagreements, ask them to discuss them until they reach agreement. Follow with a whole-class discussion.

Access for Students with Disabilities

Representation: Internalize Comprehension. Activate or supply background knowledge about determining means of data sets. Allow students to use calculators to ensure inclusive participation in the activity.

Supports accessibility for: Memory; Conceptual processing

Anticipated Misconceptions

Since previous lessons have used the mean as the best way to find a typical value, some students may go directly to that method from the beginning. Although this is valid at this stage, encourage them to look at the dot plot and think about what a typical value *should* be.

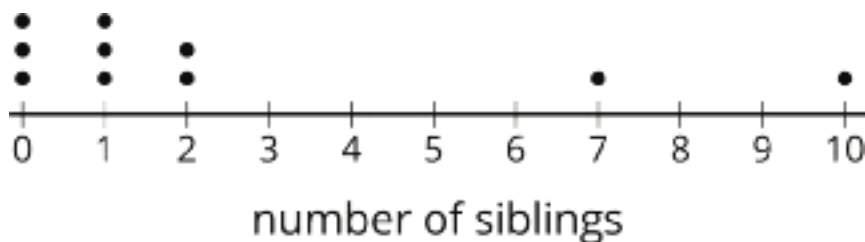
Student Task Statement

Here is data that shows the numbers of siblings of ten students in Tyler’s class.

1 0 2 1 7 0 2 0 1 10

1. Represent the data shown with a dot plot.
2. Without making any calculations, estimate the center of the data based on your dot plot. What is a typical number of siblings for these sixth-grade students? Mark the location of that number on your dot plot.
3. Find the mean. Show your reasoning.
4.
 - a. How does the mean compare to the value that you marked on the dot plot as a typical number of siblings? (Is it a little larger, a lot larger, exactly the same, a little smaller, or a lot smaller than your estimate?)
 - b. Do you think the mean summarizes the data set well? Explain your reasoning.

Student Response



2. Answers vary. Sample response: I'd estimate the center to be between 1 and 2 siblings.
3. The mean is 2.4 siblings. $\frac{1+0+2+1+7+0+2+0+1+10}{10} = \frac{24}{10} = 2.4$.
4. a. Answers vary. Sample response: The mean is a little larger than my estimate.
- b. Answers vary. Sample response: I don't think the mean summarizes the data very well. Eight out of 10 of the data points are below the mean, and more than half of the students have either no siblings or only 1 sibling, so to say that 2.4 is a typical number of siblings is not accurate.

Are You Ready for More?

Invent a data set with a mean that is significantly lower than what you would consider a typical value for the data set.

Student Response

Answers vary. Sample response: The data could have most values close to or equal to 8 and a small number of much lower values: 0, 0, 1, 7, 8, 8, 8, 9, 9, 10. The mean of this data set is 6, while we might say that 8 is typical.

Activity Synthesis

Select a few students to share their estimate for a typical number of siblings. Consider asking students:

- "When you looked at the table of values, what was your sense of a typical number of siblings for the ten students in Tyler's class?"
- "When you looked at the dot plot, did your estimate change?"

Then, discuss how the calculated mean compared to their estimates. Draw students' attention to the idea that the mean may not always represent a typical value for a data set. Discuss:

- "We have learned that the mean is a way to measure the center of a distribution. How did your calculated mean compare to your estimate of what was typical for the data set?"
- "Why do you think the mean was higher than your estimate?" (Only two of the points are above the mean of 2.4 and both are quite far above it, and seven points are below 2.4, so the mean might not paint an accurate picture of what is typical in this situation.)
- "If the mean does not always reflect what is typical in a data set, should we always rely on it as the best way to describe the center? If not, in what other ways might we measure the center of a data set?"

Explain to students that in the next activity we will look at a different measure of center.

Access for English Language Learners

Writing, Conversing: MLR1 Stronger and Clearer Each Time. Use this routine to give students a structured opportunity to refine their response to “Do you think the mean summarizes the data set well?” Give students time to meet with 2–3 partners, to share and get feedback on their responses. Provide listeners with prompts for feedback that will help their partner strengthen their ideas and clarify their language. For example, “Can you explain what a typical value should be?” and “How can you expand on the using the mean to find a typical value?” Give students 2–3 minutes to revise their initial draft based on feedback from their peers. This helps students clarify their reasoning about how the mean may or may not summarize the data well.

Design Principle(s): Optimize output (for generalization); Cultivate conversation

13.3 Finding the Middle

15 minutes

This activity introduces students to the term **median**. They learn that the median describes the middle value in an ordered list of data, and that it can capture what we consider typical for the data in some cases.

Students learn about the median through a kinesthetic activity. They line up in order of the number of letters in their name. Then, those at both ends of the line count off and sit down simultaneously until one or two people in the middle remain standing. If one person remains standing, that person has the median number of letters. If two people remain standing, the median is the mean or the average of their two values.

Students then practice identifying the median of other data sets, by analyzing both tables of values and dot plots.

Addressing

- 6.SP.B.5.c

Instructional Routines

- MLR3: Clarify, Critique, Correct

Launch

Explain to students that, instead of using the mean, sometimes we use the “middle” value in an ordered list of data set as a measure of center. We call this the **median**. Guide students through the activity:

- Give each student an index card. Ask them to write their first and last names on the card and record the total number of letters in their name. Display an example for all to see.

- Ask students to stand up, holding their index cards in front of them, and arrange themselves in order based on the number of letters in their name. (Consider asking students to do so without speaking at all.) Look for the student whose name has the fewest letters and ask him or her to be the left end of the line. Ask the student with the longest full name to be the right end of the line. Students who have the same number of letters should stand side-by-side.
- Tell students that, to find the median or the middle number, we will count off from both ends at the same time. Ask the students at the two ends of the line say “1” at the same time and sit on the floor, and the students next to them to say “2” and then sit down, and so on. Have students count off in this fashion until only one or two students are standing.
- If the class has an odd number of students, one student will remain standing. Tell the class that this student’s number is the median. Give this student a sign that says “median.” If the class has an even number of students, two students will remain standing. The median will be the mean or average of their numbers. Ask both students to hold the sign that says “median.” Explain that the median is also called the “50th percentile,” because half of the data values are the same size or less than it and fall to the left of it on the number line, and half are the same size or greater than it and fall to the right.
- Ask students to find the median a couple more times by changing the data set (e.g., asking a few students to leave the line or adding new people who are not members of the class with extremely long or short names). Make sure that students have a chance to work with both odd and even numbers of values.
- Collect the index cards and save them; they will be used again in the lesson on box plots.

Ask students to complete the rest of the questions on the task statement.

Access for Students with Disabilities

Representation: Develop Language and Symbols. Create a display of important terms and vocabulary. Include the following term and maintain the display for reference throughout the unit: median. Invite students to suggest language or diagrams to include on the display that will support their understanding of this term.

Supports accessibility for: Memory; Language

Anticipated Misconceptions

When determining the median, students might group multiple data points that have the same value and treat it as a single point, instead of counting each one separately. Remind them that when they lined up to find the median number of letters in their names, every student counted off, even their name has the same number of letters as their neighbors.

Student Task Statement

1. Your teacher will give you an index card. Write your first and last names on the card. Then record the total number of letters in your name. After that, pause for additional instructions from your teacher.

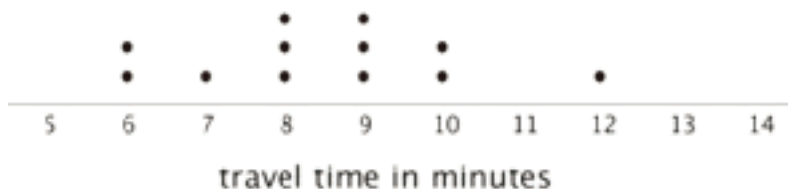
2. Here is the data set on numbers of siblings from an earlier activity.

1 0 2 1 7 0 2 0 1 10

a. Sort the data from least to greatest, and then find the **median**.

b. In this situation, do you think the median is a good measure of a typical number of siblings for this group? Explain your reasoning.

3. Here is the dot plot showing the travel time, in minutes, of Elena's bus rides to school.



a. Find the median travel time. Be prepared to explain your reasoning.

b. What does the median tell us in this context?

Student Response

1. No answer required.

2. a. The numbers of siblings in order: 0, 0, 0, 1, 1, 1, 2, 2, 7, 10 has median 1. The median is the average of the fifth and sixth values.

b. Answers vary. Sample response: Yes, I think the median is a good measure, because 1 sibling is in the middle of where most of the points in the data set are. Almost a third of the students have 1 sibling.

3. a. The median length of travel is 8.5 minutes. It is the average of the sixth and seventh data points, which are 8 and 9 minutes.

b. Answers vary. Sample response: The median tells that half of her trips to school took 8.5 minutes or less and the other half took 8.5 minutes or more.

Activity Synthesis

Select a few students to share their responses to the questions about number of siblings and Elena's travel times. Focus the discussion on the median as another measure of the center of a data set and whether it captures what students would estimate to be a typical value for each data set.

Emphasize to students that the median is a *value* and not an *individual*. For example, if the last person standing in the class has 5 letters in their first name, the median is the number 5 and not the person standing. If there is another student who had 5 letters in their name, they might have switched places with the last person standing when lining up initially. Although the person standing changed, the median remains the same value of 5.

At this point, it is unnecessary to compare the mean and the median. Students will have many more opportunities to explore the median and think about how it differs from the mean in the upcoming activities.

Access for English Language Learners

Writing, Conversing: MLR3 Clarify, Critique, Correct. Before students share their responses to the questions about number of siblings and Elena's travel times, present an incorrect response. For example, display the statement: "There is no median for this data set because two students remained standing." Demonstrate the process of interpreting the statements to uncover possible errors by thinking aloud. Voice the questions you ask yourself to call students' attention to the possible types of errors. Invite students to work with a partner to clarify the reasoning, and create an improved statement. Select 2–3 groups to share with the whole class. This helps students evaluate, and improve upon the written mathematical arguments of others.

Design Principle(s): Cultivate conversation; Maximize meta-awareness

Lesson Synthesis

In this lesson, we learn about another measure of center called the **median**. The discussion should focus on what the median is, how to find it, and why it is useful.

- "What is the median?" (The number in the middle of an ordered list of data.)
- "How can we find it?" (We order the data values from least to greatest and find the value in the middle.)
- "Is the median always one of the values in the data set? If not, when is it not?" (No. When the number of values in a data set is even, there will be two middle values. The median is the number exactly between them which may not be a value in the data set.)
- "What does the median tell you about a data set? Why is it used as a measure of the center of a distribution?" (It tells us where to divide a data set so that half of the data points have that value or smaller values and the other half have that value or larger.)
- "Why do we need another measure of center other than the mean?" (Sometimes the mean is not a good indication of what is typical for the data set.)

13.4 Practicing the Piano

Cool Down: 5 minutes

Addressing

- 6.SP.B.5.c

Student Task Statement

Jada and Diego are practicing the piano for an upcoming rehearsal. The tables list the number of minutes each of them practiced in the past few weeks.

Jada's practice times:

10	10	20	15	25	25	8	15	20
20	35	25	40					

Diego's practice times:

25	10	15	30	15	20	20	25	30
45								

1. Find the median of each data set.
2. Explain what the medians tell you about Jada's and Diego's piano practice.

Student Response

1. Jada's times in order: 8, 10, 10, 15, 15, 20, 20, 20, 25, 25, 25, 35, 40. The median is 20 minutes.
Diego's times in order: 10, 15, 15, 20, 20, 25, 25, 30, 30, 45. The median is 22.5 minutes.
2. Answers vary. Sample response: One half of Jada's practices were 20 minutes or shorter; the other half of her practices were 20 minutes or longer. Half of Diego's practices were 22.5 minutes or shorter, the other half were 22.5 minutes or longer.

Student Lesson Summary

The **median** is another measure of center of a distribution. It is the middle value in a data set when values are listed in order. Half of the values in a data set are less than or equal to the median, and half of the values are greater than or equal to the median.

To find the median, we order the data values from least to greatest and find the number in the middle.

Suppose we have 5 dogs whose weights, in pounds, are shown in the table. The median weight for this group of dogs is 32 pounds because three dogs weigh less than or equal to 32 pounds and three dogs weigh greater than or equal to 32 pounds.

20 25 32 40 55

Now suppose we have 6 cats whose weights, in pounds, are as shown in the table. Notice that there are *two* values in the middle: 7 and 8.

4 6 7 8 10 10

The median weight must be between 7 and 8 pounds, because half of the cats weigh less or equal to 7 pounds and half of the cats weigh greater than or equal to 8 pounds.

In general, when we have an even number of values, we take the number exactly in between the two middle values. In this case, the median cat weight is 7.5 pounds because $(7 + 8) \div 2 = 7.5$.

Glossary

- median

Lesson 13 Practice Problems

Problem 1

Statement

Here is data that shows a student's scores for 10 rounds of a video game.

130 150 120 170 130 120 160 160 190 140

What is the median score?

- A. 125
- B. 145
- C. 147
- D. 150

Solution

B

Problem 2

Statement

When he sorts the class's scores on the last test, the teacher notices that exactly 12 students scored better than Clare and exactly 12 students scored worse than Clare. Does this mean that Clare's score on the test is the median? Explain your reasoning.

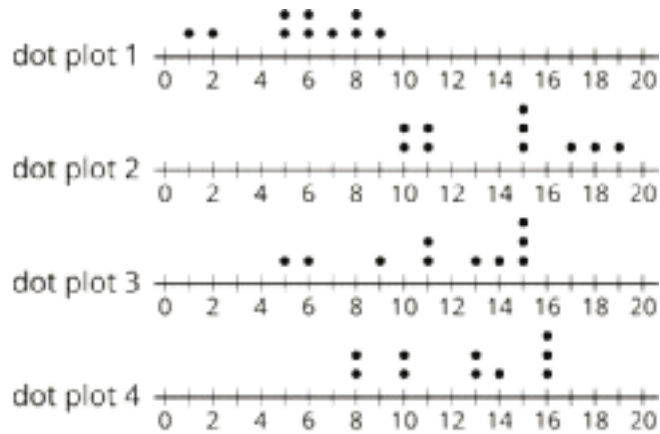
Solution

Yes. There are 25 students in the class, and Clare's score was exactly in the middle, so her score is the median.

Problem 3

Statement

The medians of the following dot plots are 6, 12, 13, and 15, but not in that order. Match each dot plot with its median.



Solution

- 6: dot plot 1
- 12: dot plot 3
- 13: dot plot 4
- 15: dot plot 2

Problem 4

Statement

Invent a data set with five numbers that has a mean of 10 and a median of 12.

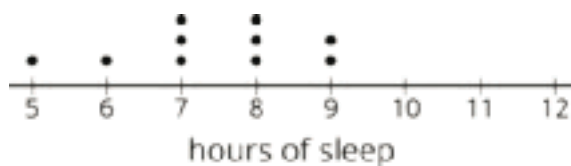
Solution

Answers vary. Sample response: 1, 11, 12, 13, 13.

Problem 5

Statement

Ten sixth-grade students reported the hours of sleep they get on nights before a school day. Their responses are recorded in the dot plot.



Looking at the dot plot, Lin estimated the mean number of hours of sleep to be 8.5 hours. Noah's estimate was 7.5 hours. Diego's estimate was 6.5 hours.

Which estimate do you think is best? Explain how you know.

Solution

Noah's estimate of 7.5 hours is best. Explanations vary. Sample explanation: It is a better balance point for the data, balancing the distances to the left and to the right.

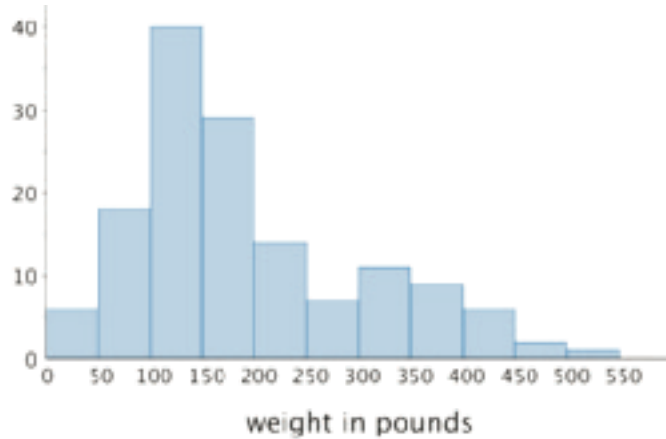
(From Unit 8, Lesson 10.)

Problem 6

Statement

In one study of wild bears, researchers measured the weights, in pounds, of 143 wild bears that ranged in age from newborn to 15 years old. The data were used to make this histogram.

- What can you say about the heaviest bear in this group?
- What is a typical weight for the bears in this group?
- Do more than half of the bears in this group weigh less than 250 pounds?
- If weight is related to age, with older bears tending to have greater body weights, would you say that there were more old bears or more young bears in the group? Explain your reasoning.



Solution

- The heaviest bear had a weight between 500 and 550 pounds.
- About 200 pounds
- Yes
- More young bears. There are more bears with smaller weights than large weights, so there are probably more young bears in the group.

(From Unit 8, Lesson 8.)

Lesson 14: Comparing Mean and Median

Goals

- Choose which measure of center to use to describe a given data set and justify (orally and in writing) the choice.
- Explain (orally) that the median is a better estimate of a typical value than the mean for distributions that are not symmetric or contain values far from the center.
- Generalize how the shape of the distribution affects the mean and median of a data set.

Learning Targets

- I can determine when the mean or the median is more appropriate to describe the center of data.
- I can explain how the distribution of data affects the mean and the median.

Lesson Narrative

In this lesson, students investigate whether the mean or the median is a more appropriate measure of the center of a distribution in a given context. They learn that when the distribution is symmetrical, the mean and median have similar values. When a distribution is not symmetrical, however, the mean is often greatly influenced by values that are far from the majority of the data points (even if there is only one unusual value). In this case, the median may be a better choice.

At this point, students may not yet fully understand that the choice of measures of center is not entirely black and white, or that the choice should always be interpreted in the context of the problem (MP2) and should hinge on what insights we seek or questions we would like to answer. This is acceptable at this stage. In upcoming lessons, they will have more opportunities to include these considerations into their decisions about measures of center.

Alignments

Addressing

- 6.SP.B.5.b: Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
- 6.SP.B.5.d: Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Instructional Routines

- MLR7: Compare and Connect

- MLR8: Discussion Supports
- Take Turns

Required Materials

Four-function calculators

Instructional master

Pre-printed cards, cut from copies of the

Required Preparation

For The Tallest and Smallest in the World activity, students will need the data on their heights (collected in the first lesson). Consider preparing a class dot plot that shows this data set to facilitate discussions.

For the Mean or Median activity, one copy of the Instructional master for each group of 3--4 students cut into cards for sorting and examining.

Student Learning Goals

Let's compare the mean and median of data sets.

14.1 Heights of Presidents

Warm Up: 5 minutes

In this warm-up, students review ways to interpret and compare data shown on a dot plot. The discussion on each given statement gives the teacher an opportunity to hear how students reason about the median, mean, typical value, spread, balance point, and MAD. This discussion will be helpful in upcoming activities, as students compare median and mean values for different data sets.

Addressing

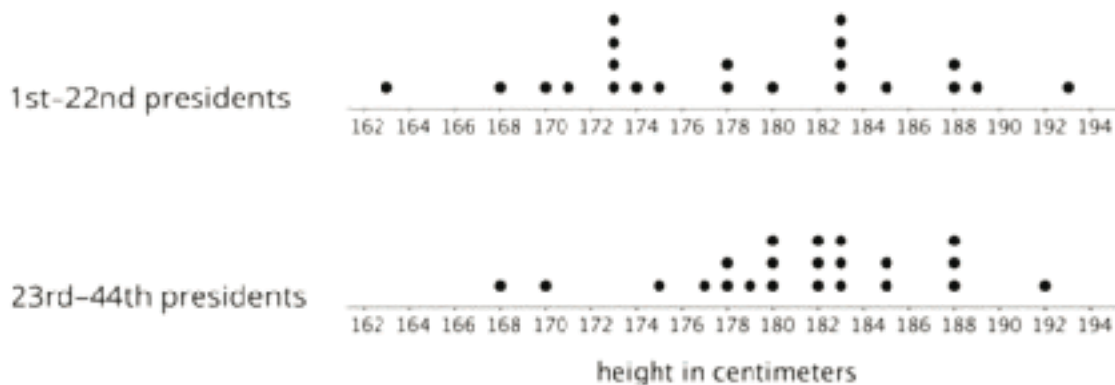
- 6.SP.B.5.b
- 6.SP.B.5.c

Launch

Give students 2 minutes of quiet work time. Follow with a whole-class discussion.

Student Task Statement

Here are two dot plots. The first dot plot shows the heights of the first 22 U.S. presidents. The second dot plot shows the heights of the next 22 presidents.



Based on the two dot plots, decide if you agree or disagree with each of the following statements. Be prepared to explain your reasoning.

1. The median height of the first 22 presidents is 178 centimeters.
2. The mean height of the first 22 presidents is about 183 centimeters.
3. A typical height for a president in the second group is about 182 centimeters.
4. U.S. presidents have become taller over time.
5. The heights of the first 22 presidents are more alike than the heights of the second 22 presidents.
6. The MAD of the second data set is greater than the MAD of the first set.

Student Response

1. Agree. The median is the average of 178 and 178 centimeters, which are the 11th and 12th data points.
2. Disagree. Though there are 4 presidents who are 183 centimeters tall, 183 is not the balance point of the data. There are many more presidents who are shorter than 183 centimeters than are taller than 183 centimeters.
3. Agree. The center of the data appears to be about 182 centimeters.
4. Agree. The center of the data in the second group is higher than in the first group.
5. Disagree. The spread of the data for the first 22 presidents is wider than that for the other group, so overall their heights are more different.
6. Disagree. Compared to the first group, the data points for the second group are clustered closer together, so their average distance from the mean is likely smaller, not greater.

Activity Synthesis

For each statement, ask students to indicate if they agree or disagree. If all students agree or all students disagree, ask a couple of students to explain their reasoning. If the class is divided on a statement, ask students on both sides to share their reasoning until the class comes to an agreement. As students share, record and display their responses for all to see. If possible, record their reasoning on the dot plots to highlight important terms students use.

To help facilitate the discussion, consider asking:

- “Who can restate ___’s reasoning in a different way?”
- “Did anyone have the same reasoning but would explain it differently?”
- “Did anyone reason about the statement in a different way?”
- “Does anyone want to add on to ____’s reasoning?”
- “Do you agree or disagree? Why?”

14.2 The Tallest and the Smallest in the World

15 minutes

In this lesson, students begin to notice how the distribution of data affects the mean and median of a data set. Using their class height data, they examine how both measures of center are affected when a value that is far from the center is added to a data set. Students find that adding these unusually large or small values pulls the mean up or down while having little or no effect on the median. They begin to see that for data sets with some far-off values, the median might be a better choice for describing a typical value because the sizes of those extreme values (whether very large or very small) do not affect the median as much as they do the mean.

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Instructional Routines

- MLR7: Compare and Connect

Launch

Consider preparing a large-scale dot plot that shows student height data and displaying it for all to see.

Arrange students in groups of 3–4. Provide each student with the data on students' heights (collected in the first lesson of the unit) and access to calculators. Give groups 8–10 minutes to complete the first three questions, and then ask them to pause for a brief class discussion before moving on to the last set of questions.

During this discussion, compare the mean that students calculated and the median they found, and solicit students' ideas on why the mean changed more than the median if the tallest person in the world joined the class. (If a class dot plot that shows student heights is made, add the point that represents the tallest person in the world.)

At this time students should begin to see that the value of the additional data point greatly affects the mean because finding the mean entails redistributing data values so that they are all the same (or moving the balance point toward the new data point to keep the distribution balanced). The new data point does not affect the median as much in this case because finding the median entails finding the point in a distribution that divides the data in half, and the numbers around the middle of the distribution are likely very close.

Give groups another few minutes to complete the rest of the task.

Access for Students with Disabilities

Representation: Internalize Comprehension. Represent the same information through different modalities by using physical objects to represent abstract concepts. Some students may benefit by starting with a physical demonstration of student's actual heights in the class.

Supports accessibility for: Conceptual processing; Visual-spatial processing

Access for English Language Learners

Representing, Conversing: MLR7 Compare and Connect. Use this routine to support small-group discussion while students work on the task. Ask students, "What changes and what stays the same for each set of data when the heights of the world's tallest, and world's smallest adults are included?" Listen for students who generalize how the mean and median change when including the largest height. Amplify students' use of the targeted language such as, mean, median, spread of data, high/low/typical values, etc. This will help students use mathematical language to describe the effect of extreme data points on the mean and median.

Design Principle(s): Maximize meta-awareness

Anticipated Misconceptions

When calculating the mean after a new person joined the class, some students might enter individual heights into a calculator once again, rather than using the sum from their original mean calculation to save time (e.g. by adding 251 to the sum of heights of students and then dividing by a class size that includes one additional student). Urge them to think about how they might use the previous calculation to make the process more efficient.

Student Task Statement

Your teacher will provide the height data for your class. Use the data to complete the following questions.

1. Find the mean height of your class in centimeters.
2. Find the median height in centimeters. Show your reasoning.
3. Suppose that the world's tallest adult, who is 251 centimeters tall, joined your class.
 - a. Discuss the following questions with your group and explain your reasoning.
 - How would the mean height of the class change?
 - How would the median height change?
 - b. Find the new mean.
 - c. Find the new median.
 - d. Which measure of center—the mean or the median—changed more when this new person joined the class? Explain why the value of one measure changed more than the other.
4. The world's smallest adult is 63 centimeters tall. Suppose that the world's tallest and smallest adults both joined your class.
 - a. Discuss the following questions with your group and explain your reasoning.
 - How would the mean height of the class change from the original mean?
 - How would the median height change from the original median?
 - b. Find the new mean.
 - c. Find the new median.
 - d. How did the measures of center—the mean and the median—change when these two people joined the class? Explain why the values of the mean and median changed the way they did.

Student Response

Answers vary for all questions.

3. d. Sample response: The mean changed more than the median when the new student is added. This is because the very large value of 251 is far from the center. It pulled the average and the balance point up by quite a bit. The median did not change very much because the middle value is simply the next point in the data set (or the average of the two middle data points).

4. d. Sample response: The new mean is a little higher than the original mean, and the median is unchanged. This is because the height difference between the world's tallest person and the original mean is larger than the difference between the world's smallest person and the original mean, so the new center or balance point is pulled toward the higher end. The median is not

changed because a new data point is added on each end, so the original middle value is also the new middle value.

Activity Synthesis

Use the whole-class discussion to further explore how unusually high or low values affect the mean and the median. Invite several students to share the new mean and median, and to explain how these measures changed if the world's shortest person joined the class. (If a class dot plot that shows student heights is made, add the point that represents the smallest person in the world.) Discuss:

- “What affect does the smallest person in the world have on the mean? Why?”
- “Which would affect the mean more: the height of the tallest person, or the height of the smallest person? Why?”
- “Suppose a new student who has a height close to the mean joined the class. Would her height affect the mean? Why or why not?”
- “Does adding two values—one unusually high and one unusually low—affect the median? Why or why not?”

Students should see that adding a data point to each end does not change the median much, if at all; the original middle value would still be the middle value as the halfway point is unchanged. Conversely, the mean will change greatly due to values that are extremely greater or less than most of the data.

14.3 Mean or Median?

15 minutes

In the previous activity, students analyzed the effects of unusually high or low values on the mean and median. Here they study distributions (displayed using dot plots and a histogram) for which the mean and median can be the same, close, or far apart, and make conjectures about how the distributions affect the mean and median (MP7). Along the way, students recognize that the mean and median are equal or close when the distribution is roughly symmetrical and are farther apart when the distribution is non-symmetrical.

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Instructional Routines

- MLR8: Discussion Supports
- Take Turns

Launch

Arrange students in groups of 3–4. Provide each group with a cut-up set of cards from the Instructional master. Give groups 4–5 minutes to take turns sorting the cards and completing the first two problems. Then, pause the activity to discuss the sorting decisions and observations of the class.

Ask a few groups how they sorted the cards. If not mentioned by students, highlight that in three of the distributions, the mean and median of the data are approximately equal. In the other three distributions, the mean and median are quite different. Discuss:

- “What do you notice about the shape and features of distributions that have roughly equal mean and median?” (They are roughly symmetrical and each have one peak in the middle, with roughly the same number of values to the left and right. They may have gaps, but the gaps are somewhat evenly spaced out.)
- “What about the shape and features of a distribution that has very different mean and median?” (They are not at all symmetrical. They may have one peak, but it is off to one side, or they don't really show any peaks. They may have gaps or data values that are unusually high or low. There is more variability in these data sets.)
- “In the second group, why might the mean and the median be so different?” (The mean is pulled toward the direction of unusually large or small values. The median simply tells us where the middle of the data lies when sorted, so it is not as affected by these values that are far from where most data points are.)

Afterwards, give students another 3–4 minutes to answer and discuss the remaining questions with their group.

Access for Students with Disabilities

Action and Expression: Develop Expression and Communication. Invite students to talk about their ideas with a partner before writing them down. Display sentence frames to support students when they explain their ideas. For example, “Cars in this category all have . . .” or “The differences between these categories are . . .”

Supports accessibility for: Language; Organization

Access for English Language Learners

Speaking: MLR8 Discussion Supports. Use this routine to support small-group discussion as students decide which measure of center to use for the dot plots. Display the sentence frame: “The measure of center that best represents card ___ is ___ because . . .” Listen for ways students relate the choice of measures of center and variability to the shape of the data distribution.

Design Principle(s): Support sense-making

Student Task Statement

1. Your teacher will give you six cards. Each has either a dot plot or a histogram. Sort the cards into *two* piles based on the distributions shown. Be prepared to explain your reasoning.
2. Discuss your sorting decisions with another group. Did you have the same cards in each pile? If so, did you use the same sorting categories? If not, how are your categories different?

Pause here for a class discussion.

3. Use the information on the cards to answer the following questions.
 - a. Card A: What is a typical age of the dogs being treated at the animal clinic?
 - b. Card B: What is a typical number of people in the Irish households?
 - c. Card C: What is a typical travel time for the New Zealand students?
 - d. Card D: Would 15 years old be a good description of a typical age of the people who attended the birthday party?
 - e. Card E: Is 15 minutes or 24 minutes a better description of a typical time it takes the students in South Africa to get to school?
 - f. Card F: Would 21.3 years old be a good description of a typical age of the people who went on a field trip to Washington, D.C.?
4. How did you decide which measure of center to use for the dot plots on Cards A–C? What about for those on Cards D–F?

Student Response

1. Answers vary. Sample responses:
 - The data points are more or less clustered together and their distribution is roughly symmetric. There are no values far from the center.

- The distributions are either clumped up on one side, have several clusters, or have values that are very far away from the rest of the data. They are not symmetric.
 - The data points with significantly larger or smaller values affected the mean, which moves the balance point away from the data's middle value and toward one end or the other.
2. No answer required.
- 3.
- a. 8 years old.
 - b. 5 people per household.
 - c. 12.5 minutes.
 - d. No, 15 years old would not be a good description of a typical age. The vast majority of the partygoers are 8 years old.
 - e. 15 would be a better description of a typical travel time because it is the middle value of the data set.
 - f. No, 21.3 years old would not be a good description of a typical age of the people on the field trip. Three-quarters of the people on the trip are 15 or 16 years old.
4. Answers vary. Sample responses:
- For Cards A–C, I could use either one, since the two measures of center are either identical or very close.
 - For Cards D–F, I chose the median because they represent the center of the data set better than the mean.

Are You Ready for More?

Most teachers use the mean to calculate a student's final grade, based on that student's scores on tests, quizzes, homework, projects, and other graded assignments.

Diego thinks that the median might be a better way to measure how well a student did in a course. Do you agree with Diego? Explain your reasoning.

Student Response

Answers vary. Sample responses:

- I think that mean makes the most sense. Each assignment affects the final grade, so every assignment reflects the student's learning on all of the material the best.
- I think that median makes the most sense. When the mean is used, a grade of 0 (or 100) can greatly influence the final grade in one direction if that is far away from most of a student's grades. With the median, a few very low (or very high) grades will not influence the center as much, so the final grade will better reflect the general understanding of the student.

Activity Synthesis

Use the whole-class discussion to reinforce the idea that the distribution of a data set can tell us which measure of center best summarizes what is typical for the data set. Briefly review the answers to the statistical questions, and then focus the conversation on the last questions (how students knew which measure of center to use in each situation). Select a couple of students to share their responses. Discuss:

- “For data sets with non-symmetrical distributions, why does the median turn out to be a better measure of center for non-symmetrical data sets?” (Non-symmetrical data sets often have unusual values that pull the mean away from the center of data. The median is less influenced by these values.)
- “Does it matter which measure we choose to describe a typical value? For example, in Card F, would it matter if we said that a typical age for the people who went on the field trip to D.C. was about 21 years old?” (Yes, it does matter in some cases. In that example, it wouldn’t really make sense to say that 21 years is a typical age because the vast majority of the people on the trip were teenagers.)

Lesson Synthesis

We see in the lesson that sometimes the two measures of center could be the same or very close, but other times they could be very different.

- “When are the mean and median likely to be close together?” (When the distribution is approximately symmetrical.)
- “When are they likely to be different?” (When the distribution is not roughly symmetrical or has unusually high or low values that are far from others.)
- “Why might the median be a more useful measure of center when the distribution is not symmetrical?” (Values far from the middle tend to have a greater influence on the mean than the median, so individual values can have a greater impact.)
- “In the situations we saw today, did it matter which measure we choose to describe a typical value?” (Yes, it did matter in some cases. For the 8-year-old’s birthday party, it would not make sense to say that 15 years is a typical age for the partygoers.)
- “A student reports that 7 is a typical number of pets that students in her class has. Do you think she used the mean number of pets or the median? How do you know?” (The mean. That number seems too high to be a typical number of pets. The data may have included one or more students who have a tank of fish or other small animals that get counted individually, which would pull the mean up.)
- “Can you think of other real-world situations where reporting the mean or median can be misleading?” (Example: Salaries at a company with many low-level workers and one executive who gets paid a lot more than anyone else would be better reported with a median.)

If they had a job opening and reported the mean, it might make people think they will make more than they probably will.)

14.4 Which Measure of Center to Use?

Cool Down: 5 minutes

Addressing

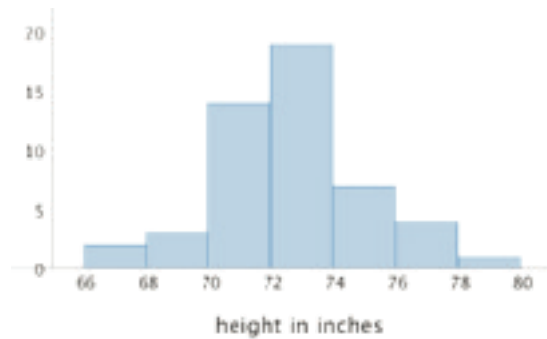
- 6.SP.B.5.c
- 6.SP.B.5.d

Student Task Statement

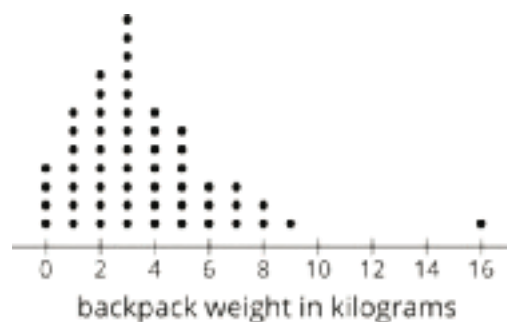
For each dot plot or histogram:

- Predict if the mean is greater than, less than, or approximately equal to the median. Explain your reasoning.
- Which measure of center—the mean or the median—better describes a typical value for the following distributions?

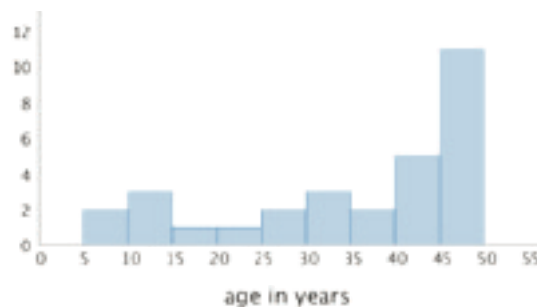
- Heights of 50 NBA basketball players



- Backpack weights of 55 sixth-grade students



3. Ages of 30 people at a family dinner party



Student Response

Answers vary. Sample responses:

1.
 - a. The mean would be approximately equal to the median, because the data are roughly symmetric.
 - b. Since I think the values would be pretty close, either the mean or the median would describe a typical height pretty well.
2.
 - a. The mean would be higher than the median. The value of 16 kilograms would bring the mean up and move it away from the center of the data.
 - b. The median would better describe a typical backpack weight, since that value would lie in the center of the large cluster of data points.
3.
 - a. The mean would be lower than the median, because even though a large fraction of the people at the dinner party are 40 or older, the ages of the people that span from 5 to 40 would bring the average age down.
 - b. The median would better describe the center of the distribution of around 40–45 years old.

Student Lesson Summary

Both the mean and the median are ways of measuring the center of a distribution. They tell us slightly different things, however.

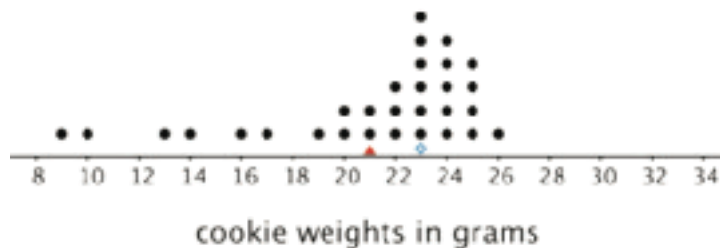
The dot plot shows the weights of 30 cookies. The mean weight is 21 grams (marked with a triangle). The median weight is 20.5 grams (marked with a diamond).



The mean tells us that if the weights of all cookies were distributed so that each one weighed the same, that weight would be 21 grams. We could also think of 21 grams as a balance point for the weights of all of the cookies in the set.

The median tells us that half of the cookies weigh more than 20.5 grams and half weigh less than 20.5 grams. In this case, both the mean and the median could describe a typical cookie weight because they are fairly close to each other and to most of the data points.

Here is a different set of 30 cookies. It has the same mean weight as the first set, but the median weight is 23 grams.



In this case, the median is closer to where most of the data points are clustered and is therefore a better measure of center for this distribution. That is, it is a better description of a typical cookie weight. The mean weight is influenced (in this case, pulled down) by a handful of much smaller cookies, so it is farther away from most data points.

In general, when a distribution is symmetrical or approximately symmetrical, the mean and median values are close. But when a distribution is not roughly symmetrical, the two values tend to be farther apart.

Lesson 14 Practice Problems

Problem 1

Statement

Here is a dot plot that shows the ages of teachers at a school.

Which of these statements is true of the data set shown in the dot plot?



- A. The mean is less than the median.
- B. The mean is approximately equal to the median.
- C. The mean is greater than the median.
- D. The mean cannot be determined.

Solution

A

Problem 2

Statement

Priya asked each of five friends to attempt to throw a ball in a trash can until they succeeded. She recorded the number of unsuccessful attempts made by each friend as: 1, 8, 6, 2, 4. Priya made a mistake: The 8 in the data set should have been 18.

How would changing the 8 to 18 affect the mean and median of the data set?

- A. The mean would decrease and the median would not change.
- B. The mean would increase and the median would not change.
- C. The mean would decrease and the median would increase.
- D. The mean would increase and the median would increase.

Solution

B

Problem 3

Statement

In his history class, Han's homework scores are:

100 100 100 100 95 100 90 100 0

The history teacher uses the mean to calculate the grade for homework. Write an argument for Han to explain why median would be a better measure to use for his homework grades.

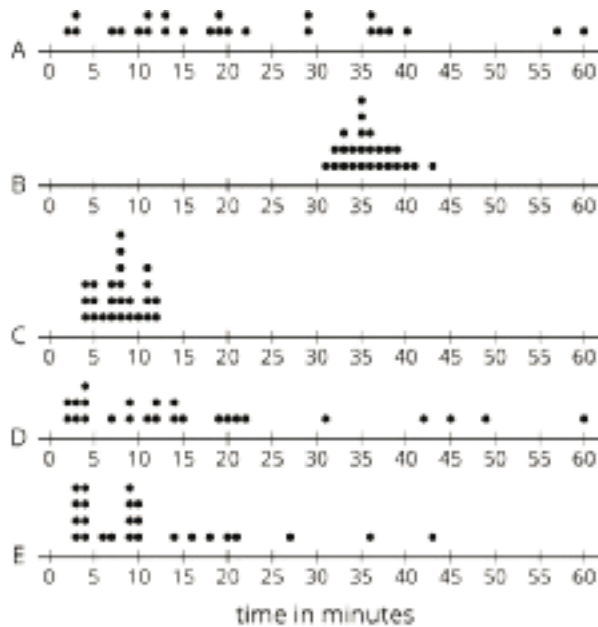
Solution

Answers vary. Sample response: The zero grade affects the mean much more than the median for these scores since it is so much lower than the others. Han does well on most of his homework and should not be punished so severely for the zero. The median value represents his typical understanding of the material more than the average value does.

Problem 4

Statement

The dot plots show how much time, in minutes, students in a class took to complete each of five different tasks. Select **all** the dot plots of tasks for which the mean time is approximately equal to the median time.



Solution

["B", "C"]

Problem 5

Statement

Zookeepers recorded the ages, weights, genders, and heights of the 10 pandas at their zoo. Write two statistical questions that could be answered using these data sets.

Solution

Answers vary. Sample responses:

- What is a typical age for the pandas at this zoo?
- What is a typical weight for the pandas at this zoo?
- Do most of the pandas weigh more than 200 pounds?
- Are a majority of the pandas female?
- What is a typical height of the pandas at this zoo?
- Do the female pandas tend to weigh more than male pandas?

(From Unit 8, Lesson 2.)

Problem 6

Statement

Here is a set of coordinates. Draw and label an appropriate pair of axes and plot the points.
 $A = (1, 0)$, $B = (0, 0.5)$, $C = (4, 3.5)$, $D = (1.5, 0.5)$

Solution

Answers vary. Check student work to ensure they made reasonable choices about axes and scale that allowed them to clearly plot all the points.

(From Unit 7, Lesson 12.)

Lesson 15: Quartiles and Interquartile Range

Goals

- Calculate the range and interquartile range (IQR) of a data set and interpret (orally and in writing) what they tell us about the situation.
- Comprehend that “interquartile range (IQR)” is another measure of variability that describes the span of the middle half of the data.
- Identify and interpret (in writing) the numbers in the five-number summary for a data set, i.e., the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum.

Learning Targets

- I can use IQR to describe the spread of data.
- I know what quartiles and interquartile range (IQR) measure and what they tell us about the data.
- When given a list of data values or a dot plot, I can find the quartiles and interquartile range (IQR) for data.

Lesson Narrative

Previously, students learned about decomposing a data set into two halves and using the halfway point, the median, as a measure of center of the distribution. In this lesson, they learn that they could further decompose a data set—into quarters—and use the **quartiles** to describe a distribution. They learn that the three quartiles—marking the 25th, 50th, and 75th percentiles—plus the maximum and minimum values of the data set make up a five-number summary.

Students also explore the **range** and **interquartile range (IQR)** of a distribution as two ways to measure its spread. Students reason abstractly and quantitatively (MP2) as they find and interpret the IQR as describing the distribution of the middle half of the data. This lesson prepares students to construct box plots in a future lesson.

Alignments

Addressing

- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
- 6.SP.B.5.d: Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Building Towards

- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

Instructional Routines

- MLR2: Collect and Display
- MLR8: Discussion Supports
- Notice and Wonder
- Think Pair Share

Student Learning Goals

Let's look at other measures for describing distributions.

15.1 Notice and Wonder: Two Parties

Warm Up: 5 minutes

In earlier lessons, students learned that the mean absolute deviation (MAD) is a measure of variability. In this warm-up, they study two distributions that appear very different but turn out to have the same MAD. Students notice that the MAD may not fully tell us about the variability of a data set. The work here motivates the need to have a different way to quantify variability, which is the focus of this lesson. While students may notice and wonder many things about these images, highlight ideas related to the variability of the data sets.

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Instructional Routines

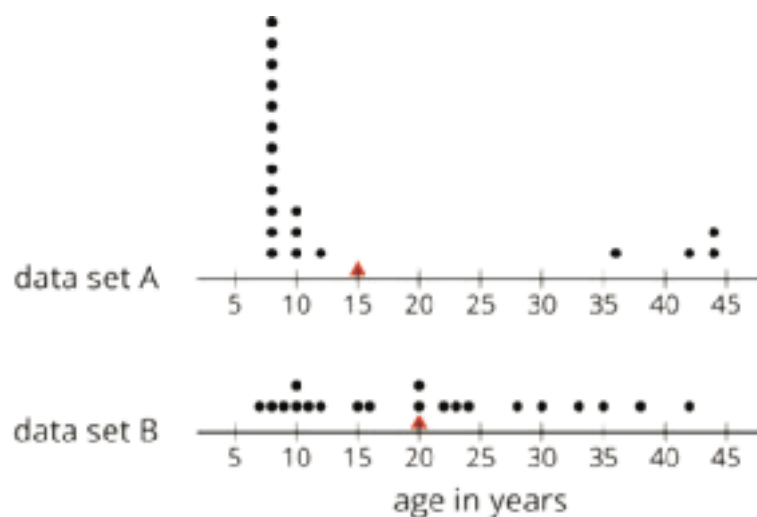
- Notice and Wonder

Launch

Arrange students in groups of 2. Display the dot plots for all to see. Ask students to identify at least one thing they notice and at least one thing they wonder about the dot plots, and to give a signal when they have both. Give students 1 minute of quiet think time, and then 1 minute to discuss their observation and question with their partner. Follow with a whole-class discussion.

Student Task Statement

Here are dot plots that show the ages of people at two different parties. The mean of each distribution is marked with a triangle.



What do you notice and what do you wonder about the distributions in the two dot plots?

Student Response

Answers vary. Sample responses:

Students may notice:

- The mean of the two data sets are different—the mean for the second data set is 5 years higher than that for the first.
- The range in values of the two data sets are about the same.
- Most points in the first data set are clustered around 8 and 10; only a few are much higher.
- The mean for the first data set is located where there are no observations.
- The points in the second data set are not clustered anywhere; they are distributed along the dot plot, between 5 and 42 years.
- The MAD values could be close for the two data sets.

Students may wonder:

- Why the data distributions look so different.
- If the MAD values are the same or close.
- If there could be other distributions that look very different than these two but also have the same MAD.

Activity Synthesis

Invite students to share what they noticed and wondered. Record and display their responses for all to see. If possible, record the relevant reasoning on or near the image. After each response, ask the class if they agree or disagree and to explain alternative ways of thinking, referring back to the dot plots each time. Discuss:

- “Do you think the ages of the people at the first party are alike or different? What about the ages of the people at the second party?”
- “The MAD for both data sets is approximately 10.5 years. What does a MAD of 10.5 years tell us in this context?”
- “Is the MAD a useful description of variability in the first data set? What about in the second data set?”

Two key ideas to uncover here are:

- The MAD is a way to summarize variation from the mean, but the single number does not always tell us how the data are distributed.
- The same MAD could result from very different distributions.

If the key ideas above are not uncovered during discussion, be sure to highlight them.

15.2 The Five-Number Summary

15 minutes

This activity introduces students to the *five-number summary* and the process of identifying the five numbers. Students learn how to partition the data into four sets: using the median to decompose the data into upper and lower halves, and then finding the middle of each half to further decompose it into quarters. They learn that each value that decomposes the data into four parts is called a **quartile**, and the three quartiles are the first quartile (Q1), second quartile (Q2, or the median), and third quartile (Q3). Together with the minimum and maximum values of the data set, the quartiles provide a five-number summary that can be used to describe a data set without listing or showing each data value.

Students reason abstractly and quantitatively (MP2) as they identify and interpret the quartiles in the context of the situation given.

Addressing

- 6.SP.B.5.c

Building Towards

- 6.SP.B.4

Instructional Routines

- MLR2: Collect and Display

Launch

Explain to students that they previously summarized variability by finding the MAD, which involves calculating the distance of each data point from the mean and then finding the average of those distances. Explain that we will now explore another way to describe variability and summarize the

distribution of data. Instead of measuring how far away data points are from the mean, we will decompose a data set into four equal parts and use the markers that partition the data into quarters to summarize the spread of data.

Remind students that when there is an even number of values, the median is the average of the middle two values.

Arrange students in groups of 2. Give groups 8–10 minutes to complete the activity. Follow with a whole-class discussion.

Student Task Statement

Here are the ages of the people at one party, listed from least to greatest.

7	8	9	10	10	11	12	15	16
20	20	22	23	24	28	30	33	35
38	42							

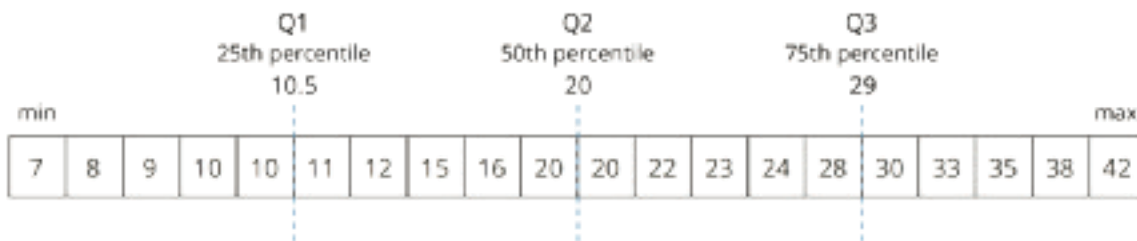
- Find the median of the data set and label it “50th percentile.” This splits the data into an upper half and a lower half.
 - Find the middle value of the *lower* half of the data, without including the median. Label this value “25th percentile.”
 - Find the middle value of the *upper* half of the data, without including the median. Label this value “75th percentile.”
- You have split the data set into four pieces. Each of the three values that split the data is called a **quartile**.
 - We call the 25th percentile the *first quartile*. Write “Q1” next to that number.
 - The median can be called the *second quartile*. Write “Q2” next to that number.
 - We call the 75th percentile the *third quartile*. Write “Q3” next to that number.
- Label the lowest value in the set “minimum” and the greatest value “maximum.”
- The values you have identified make up the *five-number summary* for the data set. Record them here.

minimum: ____ Q1: ____ Q2: ____ Q3: ____ maximum: ____
- The median of this data set is 20. This tells us that half of the people at the party were 20 years old or younger, and the other half were 20 or older. What do each of these other values tell us about the ages of the people at the party?
 - the third quartile

b. the minimum

c. the maximum

Student Response



- 1.
2. See above.
3. See above.
4. Minimum: 7 years; Q1: 10.5 years; Q2: 20 years; Q3: 29 years; Maximum: 42 years.
5.
 - a. Q3 tells us that a quarter of the party goes are over the age of 29 years old, and the rest are younger.
 - b. The minimum tells us that the youngest person at the party is 7 years old.
 - c. The maximum tells us that the oldest person at the party is 42 years old.

Are You Ready for More?

There was another party where 21 people attended. Here is the five-number summary of their ages.

minimum: 5 Q1: 6 Q2: 27 Q3: 32 maximum: 60

1. Do you think this party had more children or fewer children than the earlier one? Explain your reasoning.
2. Were there more children or adults at this party? Explain your reasoning.

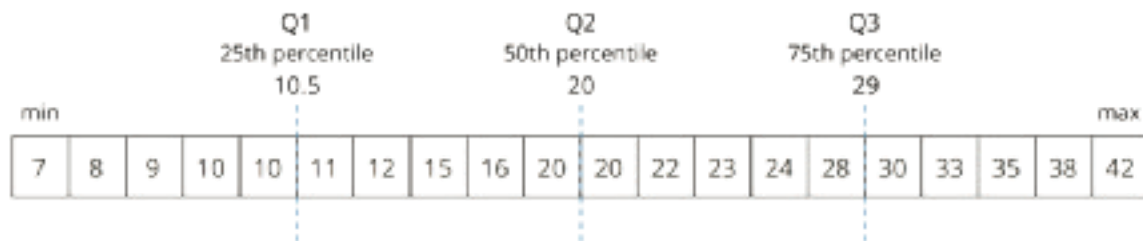
Student Response

Answers vary. Sample responses:

1. There are about the same, or possibly more, kids. Since the first quartile (Q1) is 6 years old, there are at least 6 and as many as 10 children at this party.
2. There are more adults at this party. The median age is 27 years old, an adult. Besides this adult, half of the other guests are adults aged 27 or older.

Activity Synthesis

Ask a student to display the data set they have decomposed and labeled, or display the following image for all to see.



Focus the conversation on students' interpretation of the five numbers. Discuss:

- “In this context, what do the minimum and maximum values tell us?” (The ages of the youngest and oldest partygoers.)
- “Why are Q1 called 25th percentile, Q2 50th percentile, and Q3 75th percentile?” (Each quartile tells us how many quarters of the ordered data values are accounted for up to that point. The first quartile tells us that one quarter, or 25 percent, of data values are less than or equal to that value. The second quartile tells us that two quarters, or 50 percent, of data values are less than or equal to that value, and so on.)
- “In this context, what does Q1 (10.5) tell us?” (That a quarter of the partygoers are 10.5 years old or younger.)
- “What does Q3 (29) tell us?” (That three quarters of the partygoers are 29 years old or younger.)
- “How do the five numbers help us to see the distribution of the data?” (It divides the values in the data into sections containing one fourth of the values each. This gives us an idea about the distribution of the data by looking at how varied each section is.)

Access for Students with Disabilities

Representation: Internalize Comprehension. Use color coding and annotations to identify each value of the five-number summary on a display. For example, label Q1 with the meaning in context (a quarter of the partygoers are 10.5 years old or younger).

Supports accessibility for: Visual-spatial processing

Access for English Language Learners

Representing, Conversing: MLR2 Collect and Display. Use this routine to collect student responses to the question: “How do the five numbers help us see the distribution of the data?” Pay close attention to language that amplifies the idea of variability and spread in each quartile as well as in the entire data set. Display the captured language for the whole class to see. This will help students to reference appropriate mathematical language related to quartiles and the five number summary of data.

Design Principle(s): Support sense-making

15.3 Range and Interquartile Range

15 minutes

In the previous activity, students learned about the five-number summary and how it could be used to summarize a data set. Here, students extend their work to finding the **range** and **interquartile range (IQR)** of a data set. They learn that both values provide information about a distribution of data: the range is the difference between the maximum and minimum values in the data, while the IQR is the difference between the third and first quartiles. While the range tells us how spread out (or close together) the overall data values are, the IQR tells us how spread out (or close together) the middle half of the data values are.

Students identify the range and IQR of a data set and analyze distributions with different IQRs. They reason abstractly and quantitatively (MP2) as they use the IQR to describe the variability of data.

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Instructional Routines

- MLR8: Discussion Supports
- Think Pair Share

Launch

Tell students that they will write the five-number summary of a distribution shown on a dot plot. Give students a moment of quiet time to look at the dot plot in the first question and think about how they might identify the quartiles. Then, ask students to share their ideas. Students might suggest the following strategies.

- List the values of all the data points, put them in order, and then count off the values to find the median and the other two quartiles.

- Count the points by 3's (because the data set is to be decomposed into 4 equal parts and $12 \div 4 = 3$) and mark the end of the first set with Q1, the end of the second set with Q2, etc.
- Divide the points into two halves (by counting 6 points from the left or from the right), and then dividing each half into two halves.

It is not necessary that all of these ideas are brought up at this point, but if no students mentioned the first approach (listing all values), mention it. The concrete process of writing out all the values, in order, is likely to be accessible to most students. The list of values would also be familiar, as it would resemble the one in the preceding activity.

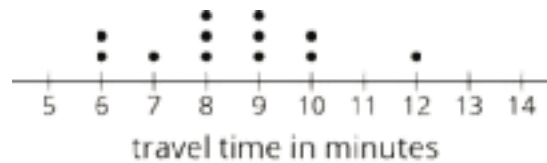
Arrange students in groups of 2. Give students 3–4 minutes of quiet work time for the first question, and 5–7 minutes to discuss their work with their partner and to complete the rest of the activity. Follow with a whole-class discussion.

Anticipated Misconceptions

When finding the IQR of the dot plots in the last question, students might neglect to divide the data set into four parts. Or they might instead divide the distance between the maximum and minimum into four parts (rather than dividing the data points into four parts). Remind students about the conversation at the start of the task about listing all the values or counting off the data points in order to find the quartiles.

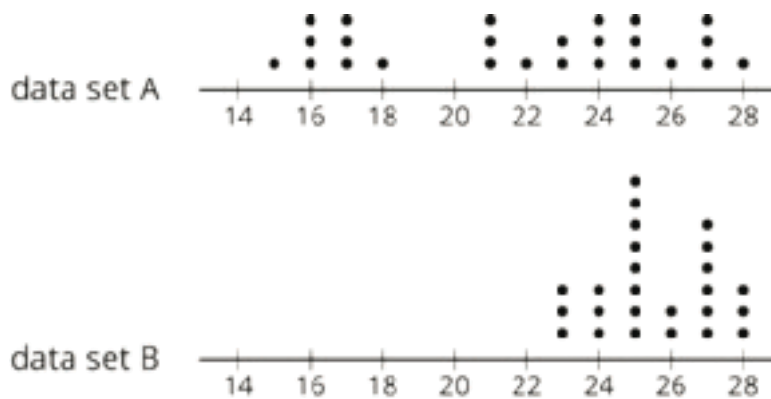
Student Task Statement

1. Here is a dot plot that shows the lengths of Elena's bus rides to school, over 12 days.



Write the five-number summary for this data set. Show your reasoning.

2. The **range** is one way to describe the *spread* of values in a data set. It is the difference between the maximum and minimum. What is the range of Elena's travel times?
3. Another way to describe the spread of values in a data set is the **interquartile range (IQR)**. It is the difference between the upper quartile and the lower quartile.
 - a. What is the interquartile range (IQR) of Elena's travel times?
 - b. What fraction of the data values are between the lower and upper quartiles?
4. Here are two more dot plots.



Without doing any calculations, predict:

- a. Which data set has the smaller range?
- b. Which data set has the smaller IQR?

5. Check your predictions by calculating the range and IQR for the data in each dot plot.

Student Response

1. The five-number summary is minimum: 6; Q1: 7.5; Q2: 8.5; Q3: 9.5; maximum: 12. This is found by ordering the travel times in a list. The Q1 is the average of 7 and 8, the 3rd and 4th values. The Q2 is the average of 8 and 9, the 6th and 7th values. The Q3 is the average of 9 and 10, the 9th and 10th values.
2. The range of Elena's data set is 6. The smallest data point is 6, and the largest data point is 12; the difference between these is 6.
3.
 - a. The 1st and 3rd quartiles are 7.5 and 9.5. The difference between these values, or the IQR, is 2.
 - b. $\frac{1}{2}$ of the data set is between the upper and lower quartiles.
4.
 - a. Data set B has the smaller range. The minimum and maximum are closer together in data set B than in data set A.
 - b. Data set B has the smaller IQR. The data points in data set B are closer together than in data set A, so the distance between Q1 and Q3 should be smaller.
5. The range for data set A is 13 since the maximum value is 28 and the minimum value is 15. The range for data set B is 5 since the maximum value is 28 and the minimum value is 23. The IQR of data set A is 8, since the Q1 is 17, and the Q3 is 25. The IQR of data set B is 2.5, since the Q1 is 24.5, and the Q3 is 27.

Activity Synthesis

Ensure that students know how to find the range and IQR, and then focus the discussion on interpreting these two measures and how they provide information about a distribution.

Select a couple of students to share the range and IQR of Elena's data. Ask:

- “What does a range of 6 minutes tell us about Elena's travel time?” (Elena's travel times vary by 6 minutes at most, or that the difference between the shortest commute and the longest one is 6 minutes.)
- “What does an IQR of 2 minutes tell us about her travel time?” (The middle half of Elena's travel times vary by 2 minutes.)

Then, select a few other students to explain their response to the third question. Discuss:

- “Without calculating, how did you determine which data set had the smaller range?” (The dot plot that has the narrower spread would have the smaller range because the distance between the greatest and least values is smaller.)
- “How did you determine which one had the smaller IQR?” (The dot plot whose middle half of the points seem more clustered together would have the smaller IQR.)
- “In general, what does a larger range tell us?” (A wider spread in the data, more variability in the data set.)
- “What does a larger IQR tell us?” (A wider spread around the center of data, more variability in the middle half of the data set.)
- “Can a data set have a large range and a small IQR?” (It is possible, if the data set has most of its points very close together but there are a few points that are far away from the cluster.)

If not mentioned by students, explain that the IQR plays a similar role as the mean absolute deviation (MAD): it tells us how different and spread out the data values are. But instead of measuring the average distance of data values from the mean, it measures the span of the middle half of the data.

Access for Students with Disabilities

Representation: Develop Language and Symbols. Create a display of important terms and vocabulary. Include the following terms and maintain the display for reference throughout the unit: quartile, range, and interquartile range (IQR). Invite students to suggest language or diagrams to include on the display that will support their understanding of this term.

Supports accessibility for: Memory; Language

Access for English Language Learners

Speaking: MLR8 Discussion Supports. Use this routine to support whole-class discussion. For each response that is shared, ask students to restate and/or revoice what they heard using mathematical language. Consider providing students time to restate what they hear to a partner, before selecting one or two students to share with the class. Ask the original speaker if their peer was accurately able to restate their thinking. Call students' attention to any words or phrases that helped to clarify the original statement. This will provide more students with an opportunity to produce language and support their understanding of distributions with different IQRs.

Design Principle(s): Support sense-making

Lesson Synthesis

- “What are the **quartiles** for a numerical data set?” (Numbers that show where we split the data up so it is in quarters.)
- “What is the relationship between the quartiles and the median?” (The second quartile is also the median).
- “What is the **Interquartile range (IQR)**? What does it mean?” (The IQR is the difference between the third and first quartile. It is a measure of the variability or spread of the data. It tells us how much “space” the middle half of the data occupies.)
- “Compare MAD and IQR. How are they alike? How are they different?” (They both provide information on the distribution of a set of data. MAD works with the mean while IQR works with the median. MAD considers all the data values and tells us the average distance between each data value and the mean, while IQR focuses on the middle half of the data and tells us how widely distributed it is.)

15.4 How Far Can You Throw?

Cool Down: 5 minutes

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Student Task Statement

Diego wondered how far sixth-grade students could throw a ball. He decided to collect data to find out. He asked 10 friends to throw a ball as far as they could and measured the distance from the starting line to where the ball landed. The data shows the distances he recorded in feet.

40 76 40 63 47 57 49 55 50 53

1. Find the median and IQR of the data set.
2. On a later day, he asked the same group of 10 friends to throw a ball again and collected another set of data. The median of the second data set is 49 feet, and the IQR is 6 feet.
 - a. Did the 10 friends, as a group, perform better (throw farther) in the second round compared to the first round? Explain how you know.
 - b. Were the distances in the second data set more variable or less variable compared to those in the first round? Explain how you know.

Student Response

1. The median is 51.5 feet. $(50 + 53) \div 2 = 51.5$. The IQR is 10, because Q1 is 47, Q3 is 57, and $57 - 47 = 10$.
2.
 - a. Worse. Sample reasoning: The median of the second data set is 49 feet, which is 2.5 feet lower than in the first round.
 - b. Less variable. Sample reasoning: The IQR of the second data set is smaller, so the values are less spread out.

Student Lesson Summary

Earlier we learned that the mean is a measure of the center of a distribution and the MAD is a measure of the variability (or spread) that goes with the mean. There is also a measure of spread that goes with the median. It is called the interquartile range (IQR).

Finding the IQR involves splitting a data set into fourths. Each of the three values that splits the data into fourths is called a **quartile**.

- The median, or second quartile (Q2), splits the data into two halves.
- The first quartile (Q1) is the middle value of the lower half of the data.
- The third quartile (Q3) is the middle value of the upper half of the data.

For example, here is a data set with 11 values.

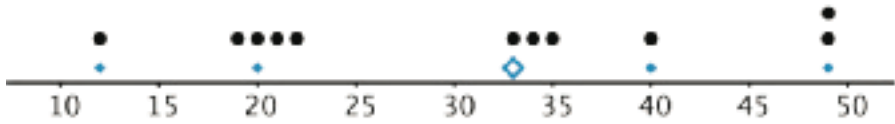
12	19	20	21	22	33	34	35	40	40	49
		Q1			Q2			Q3		

- The median is 33.
- The first quartile is 20. It is the median of the numbers less than 33.
- The third quartile 40. It is the median of the numbers greater than 33.

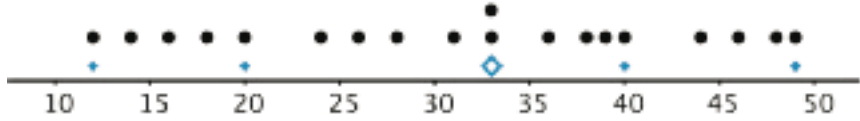
The difference between the maximum and minimum values of a data set is the **range**. The difference between Q3 and Q1 is the **interquartile range (IQR)**. Because the distance between Q1 and Q3 includes the middle two-fourths of the distribution, the values between those two quartiles are sometimes called the *middle half of the data*.

The bigger the IQR, the more spread out the middle half of the data values are. The smaller the IQR, the closer together the middle half of the data values are. This is why we can use the IQR as a measure of spread.

A *five-number summary* can be used to summarize a distribution. It includes the minimum, first quartile, median, third quartile, and maximum of the data set. For the previous example, the five-number summary is 12, 20, 33, 40, and 49. These numbers are marked with diamonds on the dot plot.



Different data sets can have the same five-number summary. For instance, here is another data set with the same minimum, maximum, and quartiles as the previous example.



Glossary

- interquartile range (IQR)
- quartile
- range

Lesson 15 Practice Problems

Problem 1

Statement

Suppose that there are 20 numbers in a data set and that they are all different.

- How many of the values in this data set are between the first quartile and the third quartile?
- How many of the values in this data set are between the first quartile and the median?

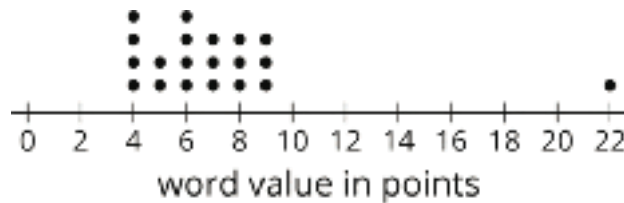
Solution

10. There are 5 numbers in each quartile, and there are two quartiles in between the first and third quartiles.
5. The median is the second quartile. The first quartile to the second comprises one quartile.

Problem 2

Statement

In a word game, 1 letter is worth 1 point. This dot plot shows the scores for 20 common words.



- What is the median score?
- What is the first quartile (Q1)?
- What is the third quartile (Q3)?
- What is the interquartile range (IQR)?

Solution

- 6.5 points
- 5
- 8
- 3

Problem 3

Statement

Mai and Priya each played 10 games of bowling and recorded the scores. Mai's median score was 120, and her IQR was 5. Priya's median score was 118, and her IQR was 15. Whose scores probably had less variability? Explain how you know.

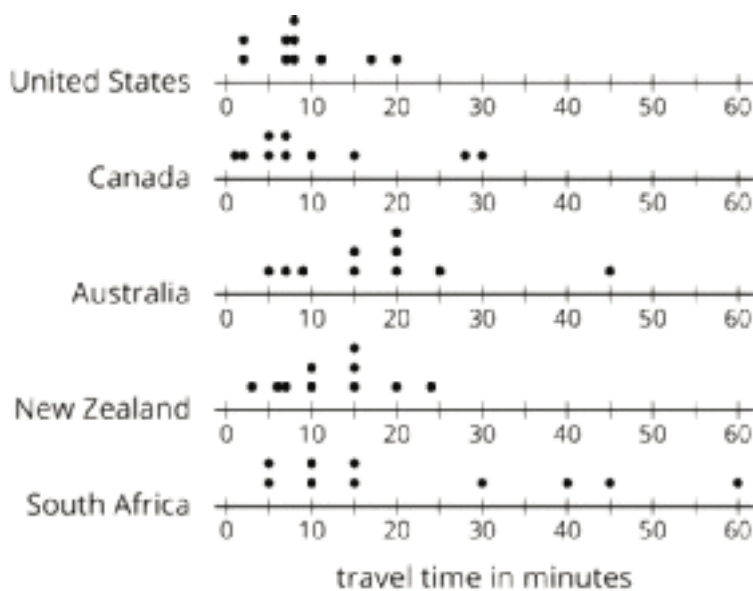
Solution

Answers vary. Sample explanation: Mai's IQR was smaller, so her scores probably varied less than Priya's scores.

Problem 4

Statement

Here are five dot plots that show the amounts of time that ten sixth-grade students in five countries took to get to school. Match each dot plot with the appropriate median and IQR.



- a. Median: 17.5, IQR: 11
- b. Median: 15, IQR: 30
- c. Median: 8, IQR: 4
- d. Median: 7, IQR: 10
- e. Median: 12.5, IQR: 8

Solution

United States: 3

Canada: 4

Australia: 1

New Zealand: 5

South Africa: 2

Problem 5

Statement

Draw and label an appropriate pair of axes and plot the points. $A = (10, 50)$, $B = (30, 25)$, $C = (0, 30)$, $D = (20, 35)$

Solution

Answers vary. Check student work to ensure they made reasonable choices about axes and scale that allowed them to clearly plot all the points.

(From Unit 7, Lesson 12.)

Problem 6

Statement

There are 20 pennies in a jar. If 16% of the coins in the jar are pennies, how many coins are there in the jar?

Solution

125, because $20 \div 0.16 = 125$.

(From Unit 6, Lesson 7.)

Lesson 16: Box Plots

Goals

- Compare and contrast (orally) a dot plot and a box plot that represent the same data set.
- Create a box plot to represent a data set.
- Describe (orally) the parts of a box plot that correspond with each number in the five-number summary, the range, and the IQR of a data set.

Learning Targets

- I can use the five-number summary to draw a box plot.
- I know what information a box plot shows and how it is constructed.

Lesson Narrative

In this lesson, students use the five-number summary to construct a new type of data display: a **box plot**. Similar to their first encounter with the median, students are introduced to the structure of a box plot through a kinesthetic activity. Using the class data set that contains the numbers of letters in their names (from an earlier lesson), they first identify the numbers that make up the five-number summary. Then, they use their numbers to position themselves on a number line on the ground, and are guided through how a box plot would be constructed with them as the data points.

Later, students draw and make sense of the structure of a box plot on paper (MP7). They notice that, unlike the dot plot, it is not possible to know all the data points from a box plot. They understand that the box plot summarizes a data set by showing the range of the data, where the middle half of the data set is located, and how the values are divided into quarters by the quartiles.

Alignments

Addressing

- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
- 6.SP.B.5.d: Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Instructional Routines

- MLR2: Collect and Display

- MLR3: Clarify, Critique, Correct
- Notice and Wonder

Required Materials

Index cards

Tape

Required Preparation

For the Human Box Plot activity:

- Each student will need the index card that shows their name and the number of letters in their name (used for the Finding the Middle activity), as well as a class data set.
- Compile the numbers on the cards into a single list or table. Prepare one copy of the data set for each student.
- Have some extra index cards available for students who might have been absent in that earlier lesson.
- Prepare five index cards that are labeled with "minimum," "maximum," "Q1," "Q2," and "Q3."
- Make a number line on the ground using thin masking tape (0.5 inch). It should show whole number intervals and span at least from the lowest data value to the highest. The intervals should be at least a student's shoulder's width.
- Prepare a roll of wide masking tape (2- or 3-inch wide) to create a box and two whiskers on the ground.

Student Learning Goals

Let's explore how box plots can help us summarize distributions.

16.1 Notice and Wonder: Puppy Weights

Warm Up: 5 minutes

This warm-up encourages students to analyze a data set carefully and reason about its distribution. It also gives them a chance to use the language they have acquired in this unit in an open-ended setting.

As students discuss their observations and questions with their partners, identify a few who notice or wonder about the features, center, and spread of the distribution.

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Instructional Routines

- Notice and Wonder

Launch

Arrange students in groups of 2. Display the table for all to see. Give students 1 minute of quiet time to look at the data set and identify at least one thing they notice and at least one thing they wonder about the distribution of the data. Ask students to give a signal when they have noticed or wondered about something. When the minute is up, give students 1 minute to discuss their observations and questions with their partner. Follow with a whole-class discussion.

Student Task Statement

Here are the birth weights, in ounces, of all the puppies born at a kennel in the past month.

13	14	15	15	16	16	16	16	17
17	17	17	17	17	17	18	18	18
18	18	18	18	18	19	20		

What do you notice and wonder about the distribution of the puppy weights?

Student Response

Answers vary. Sample responses:

Students may notice:

- The median birth weight is 17 ounces.
- The median looks like it's the same as the mean.
- Almost a third of the puppies weigh 18 ounces.
- There are no gaps in the data or values that are very far from the center.
- There are 25 puppies.

Students may wonder:

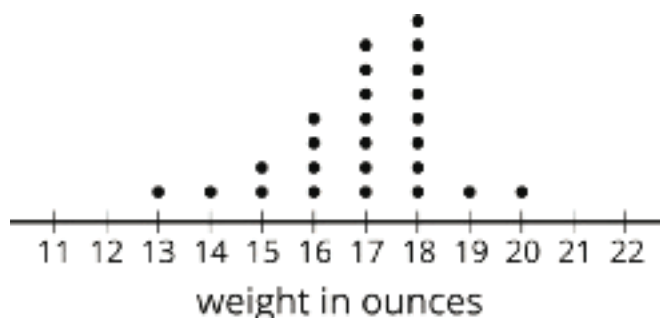
- If the IQR is less than 1.
- If a box plot could be used to represent the data.
- If a dot plot could be used to represent this data.

Activity Synthesis

While students may notice and wonder about many aspects of the data set, focus the discussion on observations and questions about the features, center, and spread of the distribution. Select previously identified students to share their responses. After each person shares, ask if others noticed or wondered about the same thing.

If students commented on or wondered about the shape of the distribution, consider displaying the following dot plot for all to see and discussing questions such as:

- “Can you use the dot plot to answer what you wondered about?”
- “Did you anticipate the distribution to have this shape?”
- “What do you notice about the distribution here that you did not notice in the table?”
- “Which representation of data—the table or the dot plot—allows you see the center and spread of the distribution more easily? Why?”
- “Which representation allows you to find the mean more easily? What about the median?”



16.2 Human Box Plot

15 minutes

Previously, students learned to identify the median, quartiles, and five-number summary of data sets. They also calculated the range and interquartile range of distributions. In this activity, students rely on those experiences to make sense of **box plots**. They explore this new representation of data kinesthetically: by creating a human box plot to represent class data on the lengths of student names, which they collected in the Finding the Middle activity in an earlier lesson.

Addressing

- 6.SP.B.4

Instructional Routines

- MLR2: Collect and Display

Launch

Before the lesson, use thin painter’s tape to make a number line on the ground. If the floor is tiled with equal-sized tiles, consider using the tiles for the intervals of the number line. Otherwise, mark off equal intervals on the tape. The number line should cover at least the distance between least data value (the fewest number of letters in a student’s name) to the greatest (the most number of letters).

Provide each student with a copy of the data on the lengths of students' names from the Finding the Middle activity. If any students were absent then, add their names and numbers of letters to the data set.

Give students 4–5 minutes to find the quartiles and write the five-number summary of the data. Then, invite several students to share their findings and come to an agreement on the five numbers. Record and display the summary for all to see.

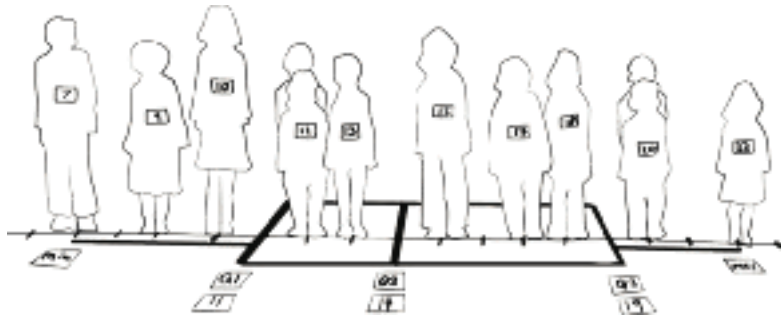
Explain to students that the five-number summary can be used to make another visual representation of a data set called a box plot. Tell students that they will create a human box plot in a similar fashion as when they were finding the median.

- Return to students the index cards from the lesson on finding the median. If any students were absent when the cards were made, give them each an index card and ask them to record on the card their full name and the number of letters in their name. If any student who made a card is absent, have another student with the same number of letters in their name hold the card of the absent student.
- Ask students to stand up, holding their index card in front of them, and place themselves on the point on the number line that corresponds to their number. (Consider asking students to do so without speaking at all.) Students who have the same number of letters should stand one in front of the other.
- Hold up the index cards that have been labeled with “minimum.” Ask students who should claim the card, then hand the card to the appropriate student. Do the same for the other labels of a five-number summary. If any of the quartiles falls between two students' numbers, write that number of the index card and have both students hold that card together.

Now that the five numbers are identified and each associated with one or more students, use wide painter's tape to construct a box plot.

- Form a rectangle on the ground by affixing the tape around the group of students between Q1 and Q3. If a quartile is between two people, put the tape down between them. If a quartile has the value of a student's number, put the tape down at that value and have the student stand on it.
- Put a tape segment at Q2, from the top side of the rectangle to the bottom side, to subdivide the rectangle into two smaller rectangles. If Q2 is a student's number, have the student stand on the tape.
- For the left whisker, affix the end of tape to the Q1 end of the rectangle; extend it to where the student holding the “minimum” card is standing. Do the same for the right whisker, from Q3 to the maximum.
- Tape the five-number summary cards and students' cards that correspond to them in the right locations.

This image shows an example of a completed human box plot.



Explain to students that they have made a human box plot. Consider taking a picture of the box plot for reference and discussion later.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Provide students with a graphic organizer to collect the class data.

Supports accessibility for: Language; Organization

Student Task Statement

Your teacher will give you the data on the lengths of names of students in your class. Write the five-number summary by finding the data set's minimum, Q1, Q2, Q3, and the maximum.

Pause for additional instructions from your teacher.

Student Response

Answers vary.

Activity Synthesis

Tell students that a box plot is a representation of a data set that shows the five-number summary. Discuss:

- “Where can the median be seen in the box plot?” (It is the line inside the box.) What about the first and third quartiles? (The left and right sides of the box.)
- “Where can the IQR be seen in the box plot?” (It is the length of the box.)
- “The two segments of tape on the two ends are called ‘whiskers.’ What do they represent?” (The lower one-fourth of the data and the upper one-fourth of data.)
- “How many people are part of the box, between Q1 and Q3? Approximately what fraction of the data set is that number?” (About half. Note that the number of people that are part of the box may not be exactly one half of the total number of people, depending on whether the number of data points is odd or even, and depending on how the values are distributed.)

- “Why might it be helpful to summarize a data set with a box plot?” (It could help us see how close together or spread out the values are, and where they are concentrated.)

Explain to students that we will draw and analyze box plots in upcoming activities and further explore why they might be useful.

Access for English Language Learners

Representing, Speaking: MLR2 Collect and Display. Use this to collect student language for the question: “Why might it be helpful to summarize a data set with a box plot?” Create a visual reference that highlights words such as median, quartiles, IQR, and distribution of data. This will help students use mathematical language and make sense of using a box plot to represent data.

Design Principle(s): Support sense-making; Maximize meta-awareness

16.3 Studying Blinks

15 minutes

In the last activity, students constructed a box plot based on the five-number summary of their name length data. In this activity, they learn to draw a box plot and they explore the connections between a dot plot and a box plot of the same data set.

Addressing

- 6.SP.B.4

Instructional Routines

- MLR3: Clarify, Critique, Correct

Launch

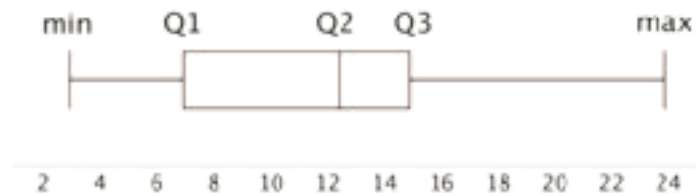
Tell students that they will now draw a box plot to represent another set of data. For their background information, explain that scientists believe people blink their eyes to keep the surface of the eye moist and also to give the brain a brief rest. On average, people blink between 15 and 20 times a minute; some blink less and others blink much more.

Arrange students in groups of 2. Give 4–5 minutes of quiet work time for the first set of questions and a minute to discuss their work with their partner. Ask students to pause afterwards.

Display the box plot for all to see. Reiterate that a box plot is a way to represent the five-number summary and the overall distribution. Explain:

- “The left and right sides of the box are drawn at the first and third quartiles (Q1 and Q3).”
- “A vertical line inside the box is drawn at the median (Q2).”

- “The two horizontal lines (or ‘whiskers’) extend from the first quartile to the minimum and from the third quartile to the maximum.”
- “The height of the box does not give additional information about the data, but should be tall enough to distinguish the box from the whiskers.”



Ask students to now draw a box plot on the same grid, above their dot plot. Give students 4–5 minutes to complete the questions. Follow with a whole-class discussion.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Chunk this task into more manageable parts to support students who benefit from support with organizational skills in problem solving. Check in with students within the first 3-5 minutes of work time to see if they were able to find Q1, Q2, and Q3. If necessary, remind students how to find each quartile.

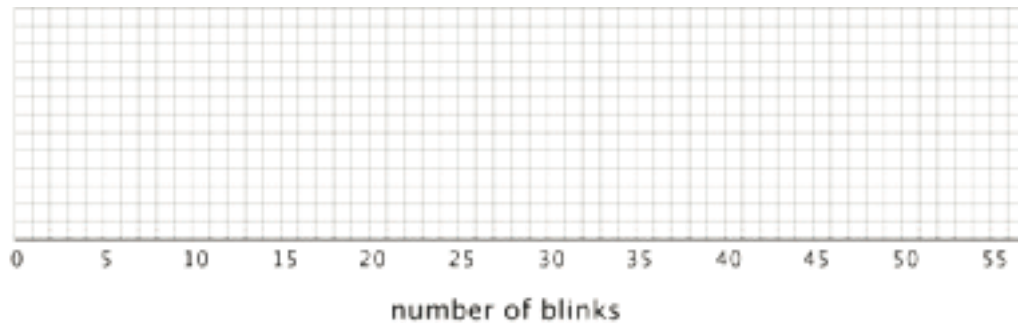
Supports accessibility for: Organization; Attention

Student Task Statement

Twenty people participated in a study about blinking. The number of times each person blinked while watching a video for one minute was recorded. The data values are shown here, in order from smallest to largest.

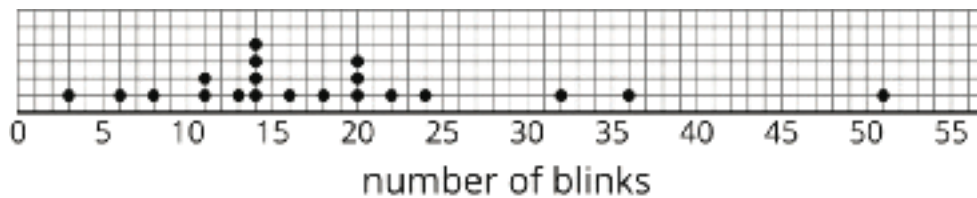
3	6	8	11	11	13	14	14	14
14	16	18	20	20	20	22	24	32
36	51							

1. a. Use the grid and axis to make a dot plot of this data set.



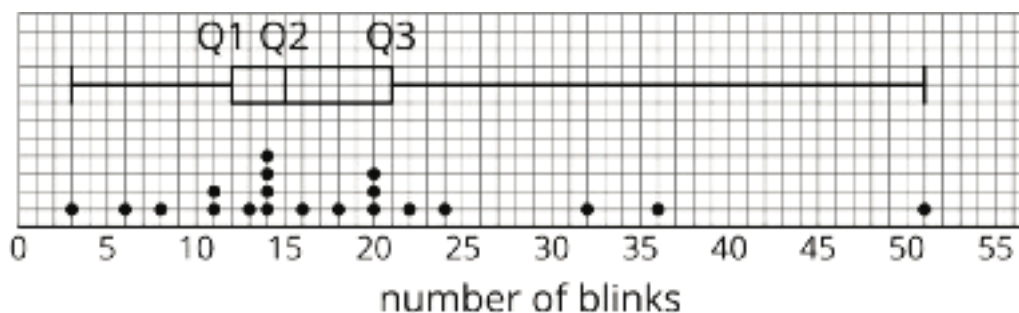
- b. Find the median (Q2) and mark its location on the dot plot.
 - c. Find the first quartile (Q1) and the third quartile (Q3). Mark their locations on the dot plot.
 - d. What are the minimum and maximum values?
2. A **box plot** can be used to represent the five-number summary graphically. Let's draw a box plot for the number-of-blinks data. On the grid, *above* the dot plot:
- a. Draw a box that extends from the first quartile (Q1) to the third quartile (Q3). Label the quartiles.
 - b. At the median (Q2), draw a vertical line from the top of the box to the bottom of the box. Label the median.
 - c. From the left side of the box (Q1), draw a horizontal line (a whisker) that extends to the minimum of the data set. On the right side of the box (Q3), draw a similar line that extends to the maximum of the data set.
3. You have now created a box plot to represent the number of blinks data. What fraction of the data values are represented by each of these elements of the box plot?
- a. The left whisker
 - b. The box
 - c. The right whisker

Student Response



- a.
1. b. 15. The median is 15 blinks, which is the average of the 10th and 11th values.

- c. 12. The first quartile is 12 blinks, which is the average of the 5th and 6th values. The third quartile is 21 blinks, which is the average of the 25th and 26th values.
- d. 3, 51. The minimum data point is 3 blinks, and the maximum data point is 51 blinks.



- 2.
3. a. $\frac{1}{4}$. The left whisker contains all the data values from the minimum to Q1. There are 5 points, which represents $\frac{1}{4}$ of the data.
- b. $\frac{1}{2}$. The box contains the values in between the 1st and 3rd quartile. There are 10 values, which comprises $\frac{1}{2}$ of the data.
- c. $\frac{1}{4}$. The right whisker contains all the data values from Q3 to the maximum. There are 5 points, which represents $\frac{1}{4}$ of the data.

Are You Ready for More?

Suppose there were some errors in the data set: the smallest value should have been 6 instead of 3, and the largest value should have been 41 instead of 51. Determine if any part of the five-number summary would change. If you think so, describe how it would change. If not, explain how you know.

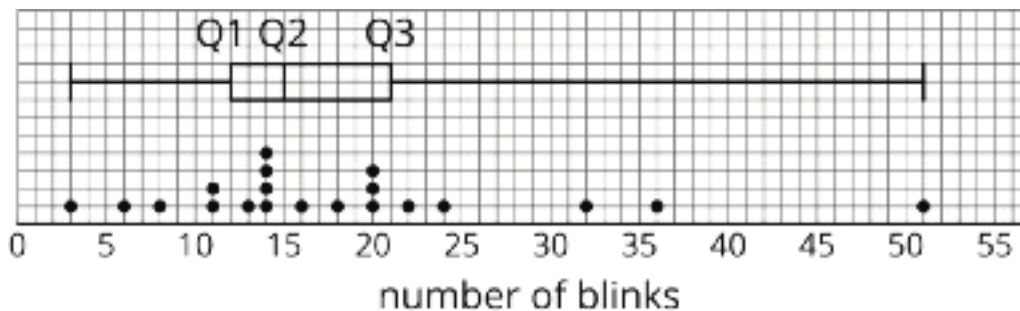
Student Response

Reasonings vary. Sample reasonings:

- Minimum: Yes. The minimum would change from 3 to 6.
- First quartile (Q1): No. The lower quartile would still be the average of 11 and 13, because the number of data points in the set has not changed.
- Median (Q2): No. The median would still be the average of 14 and 16, because the number of data points in the set has not changed.
- Third Quartile (Q3): No. The upper quartile would still be the average of 20 and 22, because the number of data points in the set has not changed.
- Maximum: Yes. The maximum would change from 51 to 41.

Activity Synthesis

Display the dot plot and the box plot for all to see.



Discuss:

- “How many data values are included in each part of the box plot?” (5 data values in each part.)
- “If you just look at the box plot, can you tell what any of the data values are?” (Only the minimum and the maximum values.)
- “If you just look at the dot plot, can you tell where the median is? Can you tell which values of the data make up the middle half of the data? Can you tell where each quarter of the data values begin and end?” (It is possible to tell, but it is not straightforward; it requires some counting.)

The focus of this activity is on constructing a box plot and understanding its parts, rather than on interpreting it in context. If students seem to have a good grasp of the drawing process and what the parts entail and mean, consider asking them to interpret the plots in the context of the research study. Ask: “Suppose you are the scientist who conducted the research and are writing an article about it. Write 2–3 sentences that summarize your findings, based on your analyses of the dot plot and the box plot.”

Access for English Language Learners

Representing, Conversing: MLR3 Clarify, Critique, Correct. Use this routine to help students clarify a common misconception when creating a box plot from a dot plot. Display a box plot with the correct median of 15, but use 11 for Q1 and 22 for Q2, which are the values students might calculate if they did not average the two middle numbers. Invite students to work with a partner to identify the error and write a correct response. Look for and amplify mathematical uses of language involving the shape, center and spread of the distribution. This will help students affirm their understanding of how to use box plots and dot plots to graphically represent a numerical data set.

Design Principle(s): Maximize meta-awareness; Cultivate conversation

Lesson Synthesis

In the lesson we see another way to graphically represent a numerical data set. Review with students:

- “How is a **box plot** made?” (The box is a rectangle with the left side at Q1 and the right side at Q3. The line inside the box is the median. The “whiskers” on the sides extend to the minimum and maximum values of the data set.)
- “What does a box plot tell you about the shape, center, and spread of a distribution?” (The median is the line in the middle, which tells you about the center. The IQR is the width of the box in the middle, which tells you about the spread. You can also tell if the distribution is roughly symmetrical.)
- “Why is it useful to use a dot plot and a box plot together?” (The dot plot shows the actual data values while the box plot tells the story of the data in fourths.)
- “How can the box plot be helpful in comparing two data sets?” (You could compare the minimum and maximum values, where the median falls, and how the data is distributed among the four quarters.)

16.4 Boxes and Dots

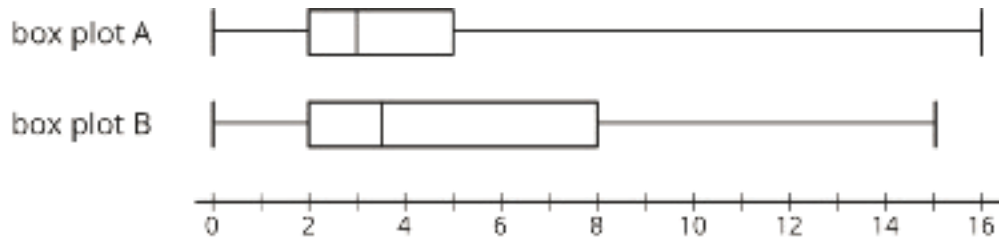
Cool Down: 5 minutes

Addressing

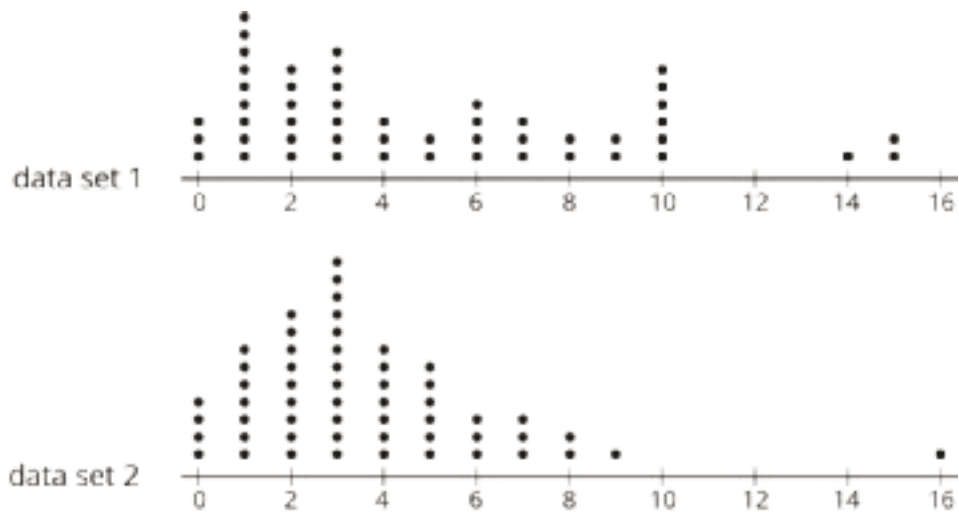
- 6.SP.B.4

Student Task Statement

1. Here are two box plots that summarize two data sets. Do you agree with each of the following statements?



- a. Both data sets have the same range.
 - b. Both data sets have the same minimum value.
 - c. The IQR shown in box plot B is twice the IQR shown in box plot A.
 - d. Box plot A shows a data set that has a quarter of its values between 2 and 5.
2. These dot plots show the same data sets as those represented by the box plots. Decide which box plot goes with each dot plot. Explain your reasoning.



Student Response

1.
 - a. Disagree
 - b. Agree
 - c. Agree.
 - d. Disagree
2. Box plot A goes with data set 2. Box plot B goes with data set 1. Sample reasonings:
 - The maximum values tell which box plot goes with which dot plot.
 - The middle half of the points in data set 1 are more spread out compared to those in data set 2, so box plot B, which has a longer box, goes with data set 1.
 - Three quarters of the points in data set 2 are between 0 and 5, which matches box and left whisker in box plot A.

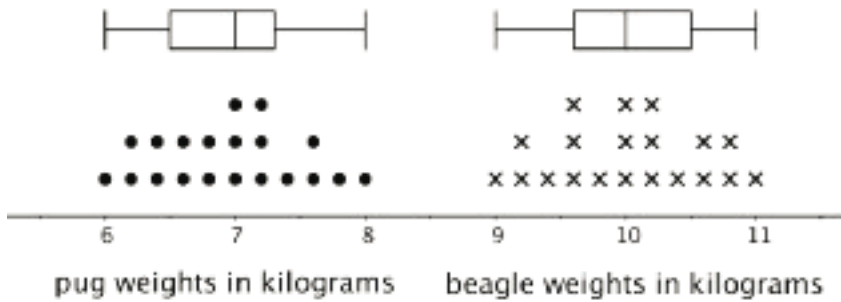
Student Lesson Summary

A **box plot** represents the five-number summary of a data set.

It shows the first quartile (Q1) and the third quartile (Q3) as the left and right sides of a rectangle or a box. The median (Q2) is shown as a vertical segment inside the box. On the left side, a horizontal line segment—a “whisker”—extends from Q1 to the minimum value. On the right, a whisker extends from Q3 to the maximum value.

The rectangle in the middle represents the middle half of the data. Its width is the IQR. The whiskers represent the bottom quarter and top quarter of the data set.

Earlier we saw dot plots representing the weights of pugs and beagles. The box plots for these data sets are shown above the corresponding dot plots.



We can tell from the box plots that, in general, the pugs in the group are lighter than the beagles: the median weight of pugs is 7 kilograms and the median weight of beagles is 10 kilograms. Because the two box plots are on the same scale and the rectangles have similar widths, we can also tell that the IQRs for the two breeds are very similar. This suggests that the variability in the beagle weights is very similar to the variability in the pug weights.

Glossary

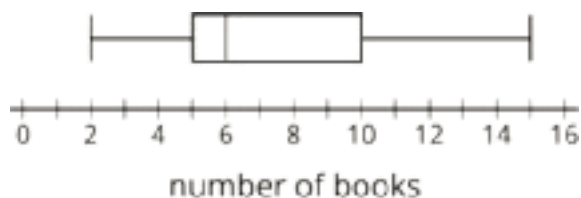
- box plot

Lesson 16 Practice Problems

Problem 1

Statement

Each student in a class recorded how many books they read during the summer. Here is a box plot that summarizes their data.



- What is the greatest number of books read by a student in this group?
- What is the median number of books read by the students?
- What is the interquartile range (IQR)?

Solution

- 15
- 6
- 5

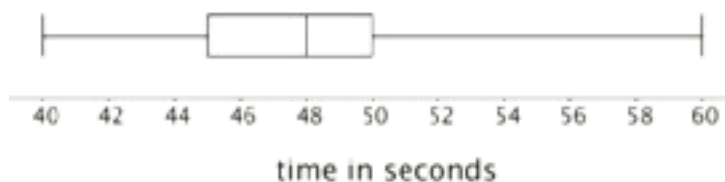
Problem 2

Statement

Use this five-number summary to draw a box plot. All values are in seconds.

- Minimum: 40
- First quartile (Q1): 45
- Median: 48
- Third quartile (Q3): 50
- Maximum: 60

Solution



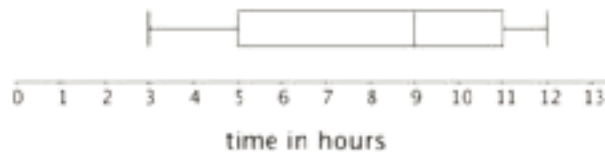
Problem 3

Statement

The data shows the number of hours per week that each of 13 seventh-grade students spent doing homework. Create a box plot to summarize the data.

3	10	12	4	7	9	5	5	11
11	5	12	11					

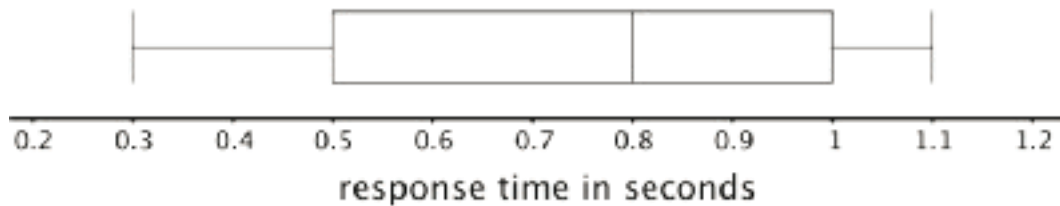
Solution



Problem 4

Statement

The box plot displays the data on the response times of 100 mice to seeing a flash of light. How many mice are represented by the rectangle between 0.5 and 1 second?



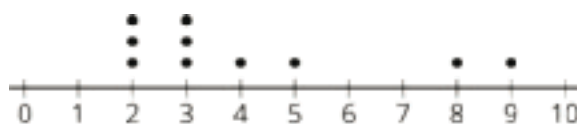
Solution

50

Problem 5

Statement

Here is a dot plot that represents a data set. Explain why the mean of the data set is greater than its median.



Solution

Explanations vary. Sample explanation: The dot plot has a peak on the left and the median is near that peak. The values 8 and 9 are much larger than the other values. These large values make the mean larger than the median.

(From Unit 8, Lesson 14.)

Problem 6

Statement

Jada earns money from babysitting, walking her neighbor's dogs, and running errands for her aunt. Every four weeks, she combines her earnings and divides them into three equal parts—one for spending, one for saving, and one for donating to a charity. Jada donated \$26.00 of her earnings from the past four weeks to charity.

How much could she have earned from each job? Make two lists of how much she could have earned from the three jobs during the past four weeks.

Solution

Answers vary. Sample response:

- \$15, \$16, \$19, \$28
- \$17, \$23, \$12.50, \$25.50

(Any correct response will have a total of \$78.)

(From Unit 8, Lesson 9.)

Lesson 17: Using Box Plots

Goals

- Compare and contrast (orally and in writing) box plots that represent different data sets, including ones with the same median but very different IQRs and vice versa.
- Determine what information is needed to solve problems about comparing box plots. Ask questions to elicit that information.
- Interpret a box plot to answer (orally) statistical questions about a data set.

Learning Targets

- I can use a box plot to answer questions about a data set.
- I can use medians and IQRs to compare groups.

Lesson Narrative

In the previous lesson, students analyzed a dot plot and a box plot in order to study the distribution of a data set. They saw that, while the box plot summarizes the distribution of the data and highlights some key measures, it was not possible to know all the data values of the distribution from the dot plot alone. In this lesson, students use box plots to make sense of the data in context (MP2), compare distributions, and answer statistical questions about them.

Students compare box plots for distributions that have the same median but different IQRs, as well as box plots with the same IQRs but different medians. They recognize and articulate that the centers are the same but the spreads are different in the first case, and the centers are different but the spreads are the same in the second case. They use this understanding to compare typical members of different groups in terms of the context of the problem (MP2).

Alignments

Addressing

- 6.SP.A.1: Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students' ages.
- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- 6.SP.B.5: Summarize numerical data sets in relation to their context, such as by:

Instructional Routines

- Anticipate, Monitor, Select, Sequence, Connect
- MLR4: Information Gap Cards

- MLR7: Compare and Connect
- Think Pair Share

Required Materials

Pre-printed slips, cut from copies of the Instructional master Straightedges

A rigid edge that can be used for drawing line segments. Sometimes a ruler is okay to use as a

straightedge, but sometimes it is preferable to use an unruled straightedge, like a blank index card.

Required Preparation

Print and cut up slips from the Sea Turtles Info Gap Instructional master. Prepare 1 set for every 2 students. Provide access to straightedges for drawing box plots. Consider creating a few paper planes of different sizes or styles to fly for the Paper Planes activity.

Student Learning Goals

Let's use box plots to make comparisons.

17.1 Hours of Slumber

Warm Up: 5 minutes

This warm-up allows students to practice creating a box plot from a five-number summary and think about the types of questions that can be answered using the box plot. To develop questions based on the box plot prompts students to put the numbers of the five-number summary into context (MP2).

As students work, identify a student who has clearly and correctly drawn the box plot to share during the whole-class discussion.

For the second question, some students may write decontextualized questions that are simply about parts of the box plot (e.g., "What is the IQR?" or "What is the range?"). Others might write contextualized questions that the box plot could help to answer (e.g., "What is the least amount of sleep in this data set?" or "What is the median number of hours of sleep for this group?"). Identify a few students from each group so that they can share later.

Addressing

- 6.SP.A.1
- 6.SP.B.4

Instructional Routines

- Anticipate, Monitor, Select, Sequence, Connect
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 2 minutes of quiet work time, followed by a whole-class discussion.

Student Task Statement

Ten sixth-grade students were asked how much sleep, in hours, they usually get on a school night. Here is the five-number summary of their responses.

- Minimum: 5 hours
- Median: 7.5 hours
- Maximum: 9 hours
- First quartile: 7 hours
- Third quartile: 8 hours

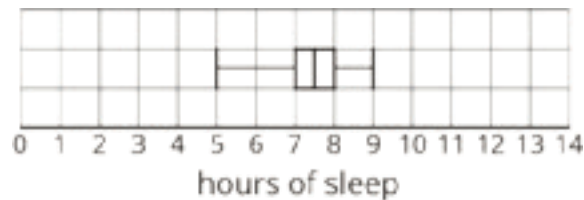
1. On the grid, draw a box plot for this five-number summary.



2. What questions could be answered by looking at this box plot?

Student Response

1.



2. Answers vary. Possible responses:

- What is the least amount of sleep a student usually gets on a school night?
- How many students usually get at least 7 hours of sleep on a school night?

Activity Synthesis

Select the previously identified student with a correct box plot to display it for all to see. If that is not possible, ask them to share how they drew the box plot and record and display the drawing based on their directions.

Select other previously identified students to share questions that could be answered by looking at the box plot first those that can be answered without the context followed by questions that rely on the context. Record and display these questions for all to see. After each question, ask the rest of the class if they agree or disagree that the answer can be found using the box plot. If time permits, ask students for the answer to each shared question.

Point out questions that are contextualized versus those that are not. Explain that a box plot can help to make sense of a data set in context and answer questions about a group or a characteristic of a group in which we are interested. The different measures that we learned to identify or calculate help to make sense of data distribution in context.

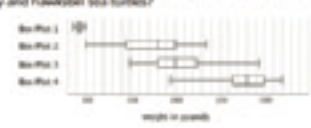

17.2 Info Gap: Sea Turtles

15 minutes

In this info gap activity, students practice analyzing box plots and thinking carefully about what questions they could help to answer. They connect features of box plots to information about IQR, range, median, and minimum and maximum values of several data sets on sea turtles (MP7). Along the way, they see the usefulness of box plots in comparing distributions and characteristics of different populations.

The info gap structure requires students to make sense of problems by determining what information is necessary, and then to ask for information they need to solve it. This may take several rounds of discussion if their first requests do not yield the information they need (MP1). It also allows them to refine the language they use and ask increasingly precise questions until they get the information they need (MP6).

Here is the text of the cards for reference and planning:

<p>Info Gap: Sea Turtles Problem Card 1</p> <p>The box plots represent data on the weights of several species of sea turtles that nest at the Outer Banks of North Carolina. Two of the species are the olive ridley and the hawksbill.</p> <p>Which two box plots represent the data for the weights of olive ridley and hawksbill sea turtles?</p> 	<p>Info Gap: Sea Turtles Data Card 1</p> <ul style="list-style-type: none"> • The median for the hawksbill data is about twice the median for the olive ridley data. • The range for the hawksbill data is 135 pounds. • The IQR for the olive ridley data is 5 pounds and the range is 15 pounds. • The heaviest weight in the olive ridley data is 100 pounds.
<p>Info Gap: Sea Turtles Problem Card 2</p> <p>Loggerhead and green sea turtles are two species that also nest at the Outer Banks. Here is the box plot for a set of data on weights of green sea turtles.</p> <ol style="list-style-type: none"> Which of the two species has a heavier typical weight? Explain how you know. Which species shows more variation in weight? Explain how you know. 	<p>Info Gap: Sea Turtles Data Card 2</p> <ul style="list-style-type: none"> • The IQR for the loggerhead data is 45 pounds. • The lightest weight in the loggerhead data is 150 pounds. • The median for the loggerhead data is 130 pounds less than the median for the green turtle data. • The heaviest weight in the loggerhead data is 295 pounds.

Addressing

- 6.SP.B.5

Instructional Routines

- MLR4: Information Gap Cards

Launch

Tell students that they will now practice using box plots to answer questions about populations of sea turtles. Provide students with the following background information: Sea turtles are air-breathing amphibians that spend most of their time floating in seaweed beds. Females come on shore to lay their eggs. One nesting place for several species of sea turtles is the Outer Banks of North Carolina. All sea turtles are considered to be endangered and only about 1 in 1,000 newly hatched sea turtles survives, which is why their nesting areas are protected.

Arrange students in groups of 2. Give each group a pair of cut-up cards from the Instructional master—a Problem Card for one partner and a Data Card for the other. Remind students of the protocol for asking and answering questions, as shown on the task statement. Give students 10 minutes to complete the activity. If any groups finish early, offer a second pair of cards and ask the partners to switch roles.

Access for Students with Disabilities

Engagement: Develop Effort and Persistence. Display or provide students with a physical copy of the written directions. Check for understanding by inviting students to rephrase directions in their own words. Keep the display of directions visible throughout the activity.

Supports accessibility for: Memory; Organization

Access for English Language Learners

Conversing: This activity uses *MLR4 Information Gap* to give students a purpose for discussing information necessary to connect features of a box plot to information about the five-number summary. Display questions or question starters for students who need a starting point such as: “Can you tell me . . . (specific piece of information)”, and “Why do you need to know . . . (that piece of information)?”

Design Principle(s): Cultivate Conversation

Student Task Statement

Your teacher will give you either a Problem Card or a Data Card about sea turtles that nest on the Outer Banks of North Carolina. Do not show or read your card to your partner.



If your teacher gives you the *problem card*:

1. Silently read your card and think about what information you need to be able to answer the question.
2. Ask your partner for the specific information that you need.
3. Explain how you are using the information to solve the problem.

Continue to ask questions until you have enough information to solve the problem.

4. Share the *problem card* and solve the problem independently.
5. Read the *data card* and discuss your reasoning.

If your teacher gives you the *data card*:

1. Silently read your card.
2. Ask your partner “*What specific information do you need?*” and wait for them to *ask* for information.

If your partner asks for information that is not on the card, do not do the calculations for them. Tell them you don’t have that information.

3. Before sharing the information, ask “*Why do you need that information?*” Listen to your partner’s reasoning and ask clarifying questions.
4. Read the *problem card* and solve the problem independently.
5. Share the *data card* and discuss your reasoning.

Pause here so your teacher can review your work. Ask your teacher for a new set of cards and repeat the activity, trading roles with your partner.

Student Response

Problem Card 1: Box Plot 1 represents the data for the olive ridley sea turtles and Box Plot 2 represents the data for the hawksbill.

Problem Card 2:

1. The green sea turtles have a heavier typical weight because their median weight is higher.
2. The loggerhead sea turtles show greater variability. Their IQRs are about the same (50 pounds for the green sea turtles and 45 for the loggerhead), but the range of weights for the loggerhead sea turtles is much larger (145 pounds, compared to 85 pounds for the green sea turtles).

Activity Synthesis

Invite students to share their experiences with the Info Gap activity. Consider discussing some of the following questions.

- For students who had a Problem Card:
 - “How did you decide what information to ask for? How did the information on your card help?”

- “How easy or difficult was it to explain why you needed the information you were asking for?”
- “Give an example of a question that you asked, the clue you received, and how you made use of it.”
- “How many questions did it take for you to be able to solve the problem? What were those questions?”
- “Was anyone able to solve the problem with a different set of questions?”
- For students who had a Data Card:
 - “When you asked your partner why they needed a specific piece of information, what kind of explanations did you consider acceptable?”
 - “Were you able to tell from their questions what statistical question they were trying to answer? If so, how? If not, why might that be?”

17.3 Paper Planes

15 minutes

In this lesson, students continue to create box plots from data sets. They compare and interpret box plots for distributions with the same median but very different IQRs, and use the plots to answer questions.

As students work, make sure that they correctly identify the five-number summary of each data set. If students have trouble making comparisons, prompt them to study the medians, IQRs, and ranges the data sets. Then, notice how they compare the box plots and whether they interpret the different measures in the context of the given situation. If they make comparisons only in abstract terms (e.g., “The median for both data sets are the same”), push them to specify what the comparisons mean in this situation (e.g., “What does the equal median tell us in this context?”). Identify students who made sense of these numbers in terms of typical distances and consistency of the flights of each person's plane. Ask them to share later.

Addressing

- 6.SP.B.5

Instructional Routines

- MLR7: Compare and Connect

Launch

Tell students that they will analyze data sets about flight distances of paper airplanes. To familiarize students with the context of this activity, consider preparing a few different styles or sizes of paper airplanes. Before students begin working, fly each paper plane a couple of times and ask students to observe their flight distances.

Arrange students in groups of 3–4. Provide access to straightedges. Give groups 8–10 minutes to complete the activity. Ask each group member to find the five-number summary and draw the box plot for *one* student (Andre, Lin, or Noah) and then share their summaries and drawings. Ask them to pause and have their summaries and drawings reviewed before answering the last two questions. Consider posting somewhere in the classroom the five-number summaries and the box plots so that students can check their answers. Ask students to be prepared to explain how Andre, Lin, and Noah's flight distances are alike or different.

Access for Students with Disabilities

Engagement: Develop Effort and Persistence. Display or provide students with a physical copy of the written directions. Check for understanding by inviting students to rephrase directions in their own words. Keep the display of directions visible throughout the activity.

Supports accessibility for: Memory; Organization

Student Task Statement

Andre, Lin, and Noah each designed and built a paper airplane. They launched each plane several times and recorded the distance of each flight in yards.

Andre

25 26 27 27 27 28 28 28 29 30 30

Lin

20 20 21 24 26 28 28 29 29 30 32

Noah

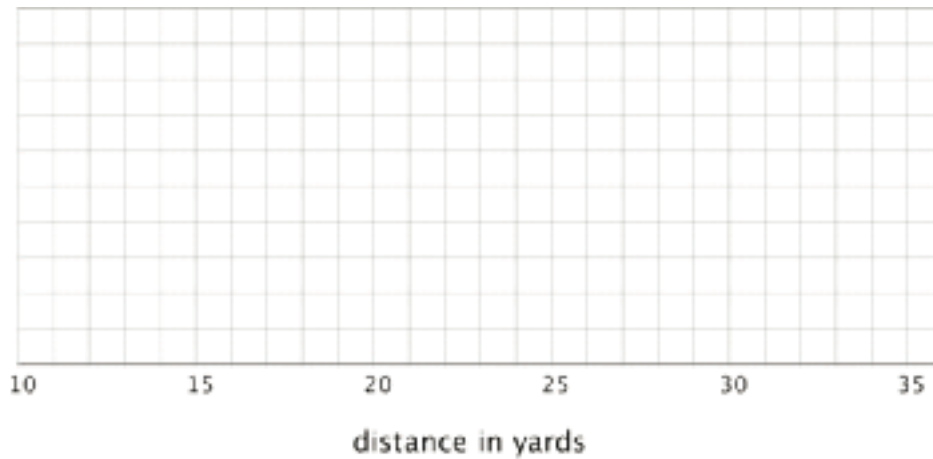
13 14 15 18 19 20 21 23 23 24 25

Work with your group to summarize the data sets with numbers and box plots.

1. Write the five-number summary for the data for each airplane. Then, calculate the interquartile range for each data set.

min	Q1	median	Q3	max	IQR

2. Draw three box plots, one for each paper airplane. Label the box plots clearly.

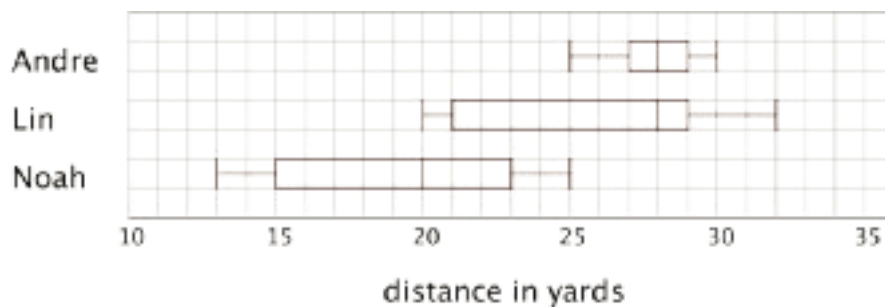


3. How are the results for Andre and Lin's planes the same? How are they different?
4. How are the results for Lin and Noah's planes the same? How are they different?

Student Response

1.

min	Q1	Q2 (median)	Q3	max	IQR
25	27	28	29	30	2
20	21	28	29	32	8
13	15	20	23	25	8



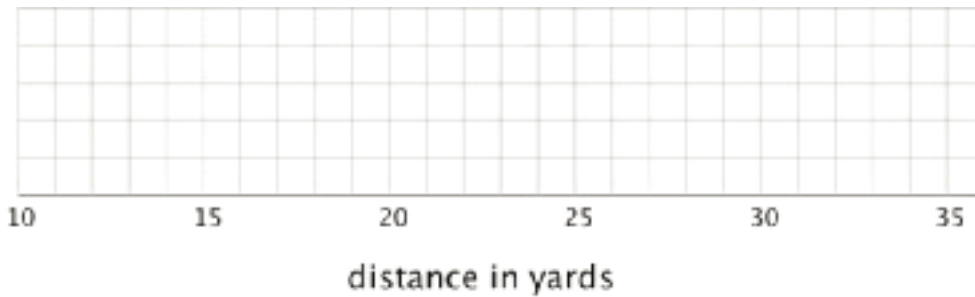
2.

3. Andre and Lin's results were similar in that they had the same median flight length of 28 yards. Their outcomes were different because Andre's plane traveled more consistent distances than Lin's plane, which traveled more variable distances, as shown by her plane's larger IQR.
4. Lin and Noah's planes had the same IQR of 8 yards, meaning that their planes had a similar amount of variability in flight distance. Noah's median was much lower than Lin's, however, so generally speaking, his planes fly a much shorter distance.

Are You Ready for More?

Priya joined in the paper-plane experiments. She launched her plane eleven times and recorded the lengths of each flight. She found that her maximum and minimum were equal to Lin's. Her IQR was equal to Andre's.

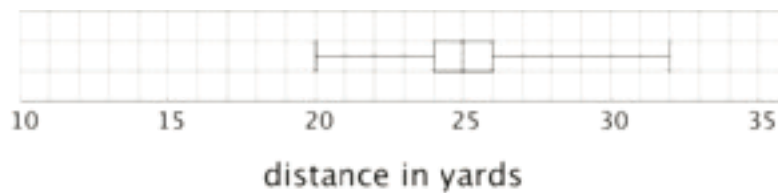
Draw a box plot that could represent Priya's data.



With the information given, can you estimate the median for Priya's data? Explain your reasoning.

Student Response

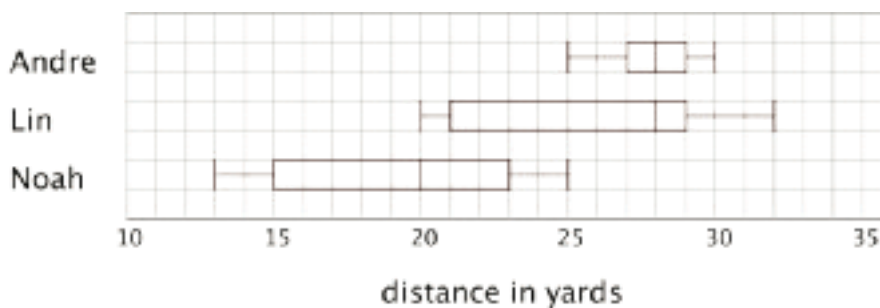
Box plots vary. Sample box plot:



Answers vary. Sample response: No, I cannot estimate the median. I know the IQR, but I don't have enough information to tell where the first, second, or third quartiles are. The box could be anywhere between 20 and 32, and the median could be anywhere in that box.

Activity Synthesis

Focus the whole-class discussion on students' analyses and interpretations of the box plots. Display the box plots for all to see.



Select a few students or groups to share their responses comparing Andre's and Lin's data. Be sure to discuss what it means when two data sets have the same median but different IQRs, as in

Andre's and Lin's cases. If no students connect these values to the center and spread and data, ask them to do so.

- "What can you say about the center of Andre's data and that of Lin's data?" (They have the same center of 28 yards.)
- "What does the same center tell us in this context?" (The same number could be used to describe a typical flight distance for both Andre's and Lin's flight distances.)
- "What can you say about the spread of Andre's data and that of Lin's data?" (Andre's data are far more concentrated than Lin's, which are quite spread out. The range of her data is 12 yards, which tells us there is much more variability in her flight distances. For Andre the range is only 5 yards, less than half of Lin's, and his IQR is a quarter of Lin's. Overall, there is much less variability in his data.)
- "Which of the two planes—Andre's or Lin's—flies a more consistent distance? How do you know?" (Andre's, because the spread of his data is much smaller.)

Then, select a few other students or groups to compare Lin's and Noah's data. Be sure to discuss what it means when two data sets have the same spread (IQR) but different medians.

- "What do the two very different centers tell us in this context?" (Generally speaking, a typical flight distance for Lin's plane is quite different than that for Noah's.)
- "What can you say about the spreads of Lin's and Noah's data?" (Both sets have the same range and the same IQR, though the values of the quartiles are different for the two sets of data.)
- "What does the same range tell us in this case?" (The difference between the shortest flight distance and the longest one is the same for both data sets.)
- "What does the same IQR tell us in this case?" (The middle half of the two sets of data cover the same distance.)
- "Whose plane—Lin's or Noah's—flies a more consistent distance?" (Their planes fly with very similar consistency, or inconsistency. The identical IQR and range tell us that their data have very similar variability.)

Access for English Language Learners

Representing, Listening: MLR7 Compare and Connect. Use this routine to support students as they interpret the box plot representations of the data. Ask students, “What is the same and what is different among the box plots?” Invite students to use the language of median, IQR, distribution, center, spread as they discuss comparisons. Ask students to make connections in the box plot diagrams to the values in the 5-number summary table. For example, “Andre’s IQR is 2 in the table which is the length of the rectangle in Andre’s box plot.” These exchanges strengthen students’ mathematical language related to representations of data sets used and help them connect what the box plots tell us about the center and spread of data sets.

Design Principle(s): Optimize output (for comparison); Maximize meta-awareness

Lesson Synthesis

In this lesson, we see that box plots can tell us stories about the center and spread of data sets.

- “What are some questions you can ask to match box plots to data?” (Questions about the 3 quartiles, maximum, or minimum will help distinguish the different box plots.)
- “What does it mean when two box plots show the same median but different IQRs?” (The data they summarize have the same center but different spreads.)
 - “How can we see this in a box plot?” (The lines inside the box will be at the same place, but the widths of the boxes will be different.)
 - “What does the same median, different IQRs’ mean in context?” (It means that we can use the same number to describe what is typical for each group, but the variability in the data is different.)
- “What does it mean when two box plots show the same IQR but different medians?” (The data they summarize have different centers but the same spread.)
 - “How can we see this in a box plot?” (The Q2 segments in the boxes are located in different positions along the number line, but the boxes have the same width.)
 - “What does ‘the same IQR, different medians’ tell us in context?” (It means that what is typical for each group is different, but the variability is similar for the two groups.)

17.4 Humpback Whales

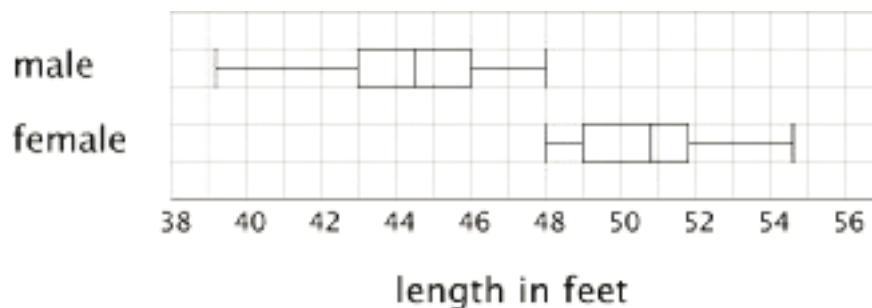
Cool Down: 5 minutes

Addressing

- 6.SP.B.5

Student Task Statement

Researchers measured the lengths, in feet, of 20 male humpback whales and 20 female humpback whales. Here are two box plots that summarize their data.



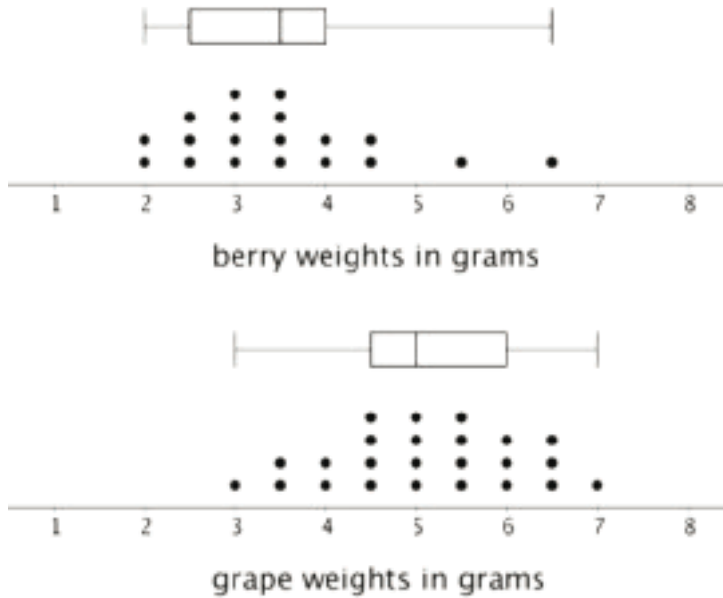
1. How long was the longest whale measured? Was this whale male or female?
2. What was a typical length for the male humpback whales that were measured?
3. Do you agree with each of these statements about the whales that were measured? Explain your reasoning.
 - a. More than half of male humpback whales measured were longer than 46 feet.
 - b. The male humpback whales tended to be longer than female humpback whales.
 - c. The lengths of the male humpback whales tended to vary more than the lengths of the female humpback whales.

Student Response

1. The longest whale was about 55 feet long and was a female.
2. A typical male humpback whale was about 44.5 feet long.
3.
 - a. Disagree. Sample explanation: The upper quartile of the data for the male humpbacks is 46 feet, which means a quarter of the whales are longer than 46 feet.
 - b. Disagree. Sample explanation: The entire distribution for the lengths of female humpbacks is greater than that for male humpbacks, so female humpbacks tend to be longer than their male counterparts.
 - c. Agree. Sample explanation: The IQR of the data for male humpbacks is slightly greater than that for female humpbacks, and the range of the data for the males is larger than that for females, so the lengths of male humpbacks tend to vary more.

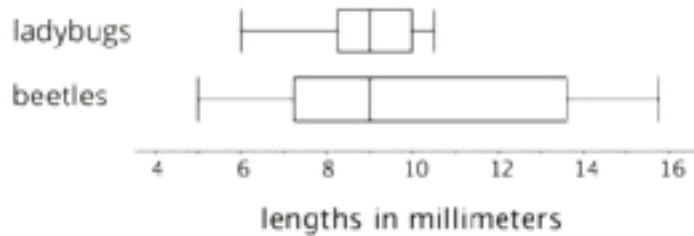
Student Lesson Summary

Box plots are useful for comparing different groups. Here are two sets of plots that show the weights of some berries and some grapes.



Notice that the median berry weight is 3.5 grams and the median grape weight is 5 grams. In both cases, the IQR is 1.5 grams. Because the grapes in this group have a higher median weight than the berries, we can say a grape in the group is typically heavier than a berry. Because both groups have the same IQR, we can say that they have a similar variability in their weights.

These box plots represent the length data for a collection of ladybugs and a collection of beetles.



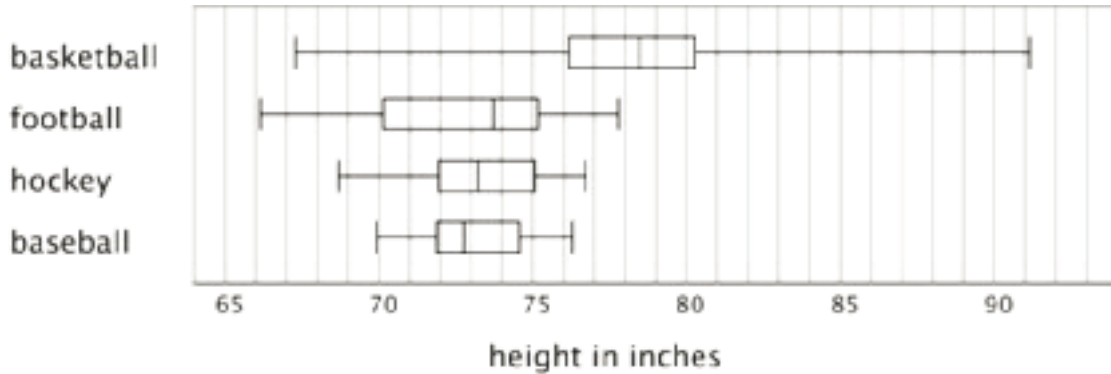
The medians of the two collections are the same, but the IQR of the ladybugs is much smaller. This tells us that a typical ladybug length is similar to a typical beetle length, but the ladybugs are more alike in their length than the beetles are in their length.

Lesson 17 Practice Problems

Problem 1

Statement

Here are box plots that summarize the heights of 20 professional male athletes in basketball, football, hockey, and baseball.



- In which two sports are the players' height distributions most alike? Explain your reasoning.
- Which sport shows the greatest variability in players' heights? Which sport shows the least variability?

Solution

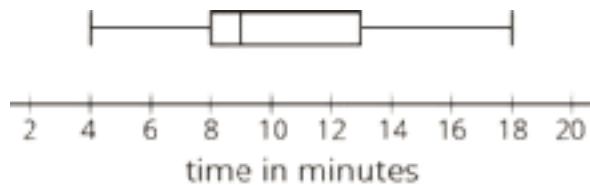
- Hockey and baseball players are most alike. Sample explanation: The two medians are very close (around 73 inches each), their IQRs differ by only about $\frac{1}{2}$ inch.
- Overall, basketball players show the greatest variability in height (indicated by the largest range). Variability for the middle half of data is the greatest for football players (shown by the largest IQR). Baseball players show the least variability in height (shown by the smallest range and IQR).

Problem 2

Statement

Here is a box plot that summarizes data for the time, in minutes, that a fire department took to respond to 100 emergency calls.

Select **all** the statements that are true, according to the dot plot.



- A. Most of the response times were under 13 minutes.
- B. Fewer than 30 of the response times were over 13 minutes.
- C. More than half of the response times were 11 minutes or greater.
- D. There were more response times that were greater than 13 minutes than those that were less than 9 minutes.
- E. About 75% of the response times were 13 minutes or less.

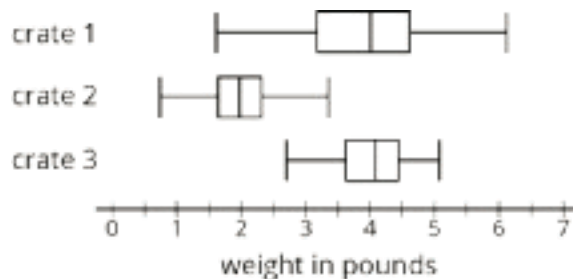
Solution

["A", "B", "E"]

Problem 3

Statement

Pineapples were packed in three large crates. For each crate, the weight of every pineapple in the crate was recorded. Here are three box plots that summarize the weights in each crate.



Select **all** of the statements that are true, according to the box plots.

- A. The weights of the pineapples in Crate 1 were the most variable.
- B. The heaviest pineapple was in Crate 1.
- C. The lightest pineapple was in Crate 1.
- D. Crate 3 had the greatest median weight and the greatest IQR.
- E. More than half the pineapples in Crate 1 and Crate 3 were heavier than the heaviest pineapple in Crate 2.

Solution

["A", "B", "E"]

Problem 4

Statement

Two TV shows each asked 100 viewers for their ages. For one show, the mean age of the viewers was 35 years and the MAD was 20 years. For the other show, the mean age of the viewers was 30 years and the MAD was 5 years.

A sixth-grade student says he watches one of the shows. Which show do you think he watches? Explain your reasoning.

Solution

The first show. Explanations vary. Sample explanation: Even though the second show has a lower mean, the much higher MAD of the first show means its viewers have a wider age range.

(From Unit 8, Lesson 12.)

Section: Let's Put it to Work

Lesson 18: Using Data to Solve Problems

Goals

- Recognize that different graphical displays offer different insights into a distribution. Choose an appropriate graphical display to represent a data set, and justify the choice (orally and in writing).
- Recognize that different measures of center and variability offer different insights into a data set. Choose an appropriate measure of center and variability to describe a data set, and justify the choice (orally and in writing).

Learning Targets

- I can decide whether mean and MAD or median and IQR would be more appropriate for describing the center and spread of a data set.
- I can draw an appropriate graphical representation for a set of data.
- I can explain what the mean and MAD or the median and IQR tell us in the context of a situation and use them to answer questions.

Lesson Narrative

This lesson is a good opportunity for students to use the information they have learned in the unit and apply it to different situations, but may be shortened to fit time constraints.

In this lesson, students compare the center and spread of different distributions. They determine what these different measures (mean and MAD or median and IQR) represent in context. They select an appropriate representation for the distribution based on the structure of the data, an appropriate set of measures of center and spread, and interpret their meaning in the context (MP4).

For students who are curious why we are asking them to compute measures of center and variation by hand when computers would be more efficient and accurate, tell them that understanding the meaning of the values and knowing what questions to ask are skills computers have not yet mastered. By practicing with calculations on small data sets, students are becoming familiar with these measures as well as questioning skills so they can correctly interpret results from computers in the future. If students do not raise the question themselves, this point may be left until the topics are revisited in later grades.

Alignments

Addressing

- 6.SP.A.2: Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.

- 6.SP.B: Summarize and describe distributions.
- 6.SP.B.5.c: Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
- 6.SP.B.5.d: Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Instructional Routines

- Group Presentations
- MLR2: Collect and Display
- MLR8: Discussion Supports
- Notice and Wonder
- Think Pair Share

Required Materials

Straightedges

A rigid edge that can be used for drawing line segments. Sometimes a ruler is okay to use as a straightedge, but sometimes it is preferable to use an unruled straightedge, like a blank index card.

Tools for creating a visual display

Any way for students to create work that can be easily displayed to the class. Examples: chart paper and markers, whiteboard space and markers, shared online drawing tool, access to a document camera.

Required Preparation

Provide access to straightedges for students to use when drawing box plots. Preview the background information about the yellow perch fish for the main activity. Prepare tools for creating a visual display, one set for every 3–4 students.

Student Learning Goals

Let's compare data sets using visual displays.

18.1 Wild Bears

Warm Up: 5 minutes

This warm-up allows students to review two important ideas of this unit: interpreting data in a box plot and writing statistical questions based on a data set. Students write statistical questions based on given box plots, then trade questions to answer questions written by another student.

Addressing

- 6.SP.A.2

Instructional Routines

- Think Pair Share

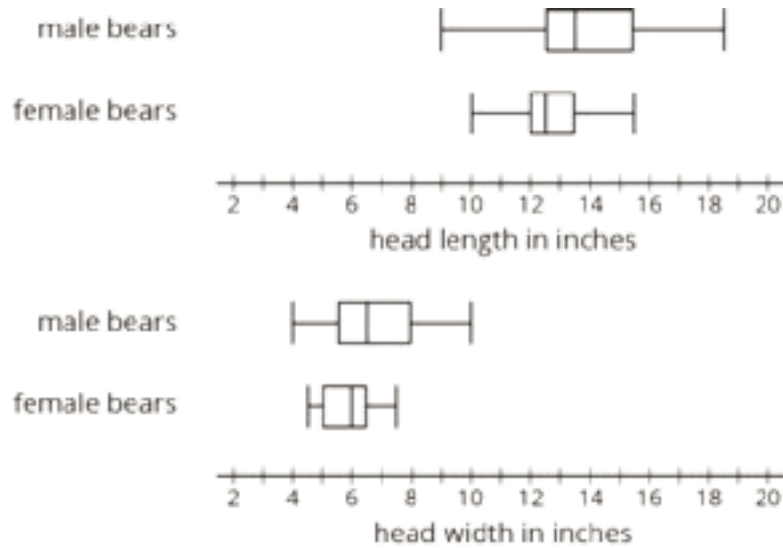
Launch

Arrange students in groups of 2. Tell students that, for the first question, one partner should write two questions about the head lengths and the other partner should write two questions about the head widths. For the second question, they should exchange and review each other's questions. If their partner's question does not seem to be a statistical question, suggest a revision so that it becomes a statistical question, and then answer the question. Remind students to consider units of measurement.

Give students 2 minutes of quiet work time for the first question and 2 minutes for collaboration afterwards.

Student Task Statement

In one study on wild bears, researchers measured the head lengths and head widths, in inches, of 143 wild bears. The ages of the bears ranged from newborns (0 years) to 15 years. The box plots summarize the data from the study.



1. Write four statistical questions that could be answered using the box plots: two questions about the head length and two questions about the head width.
2. Trade questions with your partner.
 - a. Decide if each question is a statistical question.
 - b. Use the box plots to answer each question.

Student Response

1. Answers vary. Sample statistical questions:
 - What is a typical head length, in inches, for male bears in the data set?

- Do female bears generally have longer heads than male bears?
 - Which data set shows more variability in head widths: male bears or female bears?
 - What is the widest head width, in inches, for male bears?
 - How do the ranges of head widths compare? Which group—male bears or female bears—has a larger range, and by how much?
2. Answers vary. Sample responses to the questions above:
- A typical head length for male bears is about 13.5 inches.
 - No, female bears generally have shorter heads than male bears.
 - The data for male bears show more variability in head width.
 - The widest head width for male bears is 10 inches.
 - Male bears have a larger range of head widths; it is nearly twice as large as female bears' range of head widths. The range for male bears is about 6 inches, and the range for female bears is a little over 3 inches.

Activity Synthesis

Ask several students to share their questions about the head width and head length. Record and display their responses for all to see. After each student shares, ask the class if they agree or disagree it is a statistical question. If they agree, ask how they would find the answer, or for the answer itself. If they disagree, ask how they could rewrite the question so it is a statistical question.

18.2 Math Homework (Part 1)

Optional: 15 minutes (there is a digital version of this activity)

In this activity, students compare and contrast different measures of center and variability for data sets that have gaps and are not symmetrical. They interpret mean, MAD, median, and IQR in the context of a situation. Unlike many of the data sets students have seen so far, this one shows values that could roughly divide into three parts: the days when there is little or no homework, the days when there is a moderate number of homework problems, and the days when the assignment is relatively large. Because of this distribution, finding a typical number of homework problems (or whether it would be helpful to identify a typical number) is not obvious, prompting students to interpret measures of center and spread more carefully (MP2).

As students work and discuss, identify at least one student or group that decides that the mean and MAD are appropriate measures of center and spread and can explain their reasoning, and another that decides to go with the median and IQR and could support their choice. Invite them to share during whole-class discussion.

Addressing

- 6.SP.B.5.c

- 6.SP.B.5.d

Instructional Routines

- Notice and Wonder

Launch

Keep students in groups of 2. Give students a moment of quiet time to look at the data on homework problems and identify at least one thing they notice and one thing they wonder. Give them another brief moment to share their observation and question with their partner. Then, ask a few students to share their responses with the class.

Students are likely to notice that the data values are quite different, that there are some days with no homework and others with quite a few problems, that there is not an obvious cluster, or that the number of problems could be roughly grouped into three kinds (a little, moderate, and a lot). They are likely to wonder why the numbers are so spread out and varied.

Briefly discuss the following questions to encourage students to think about the data contextually:

- “Why might the homework assignment data show this distribution? What are some possible explanations?” (When only one problem was assigned, the problem might be particularly challenging or might require considerable work or collaboration. Another possibility: there might be an upcoming exam, so the homework load was reduced. When many problems were assigned, the problems might be quick exercises with short answers, or the assignment might be review materials for an entire chapter.)
- “How might we describe ‘a typical number of homework problems’ in this case?”
- “Which do you predict would be higher: the mean or the median number of problems? Why?”

Next, give students 8–10 minutes to complete the task, either independently or collaboratively. Ask students to think quietly about the last question before discussing their response with their partner.

If students are using the digital activities, they will need to enter the data points in the column “A” for the applet to “list”, “sort”, etc. The applet allows for students to populate their own mean, Q1 and Q3 values.

Student Task Statement

Over a two-week period, Mai recorded the number of math homework problems she had each school day.

2 15 20 0 5 25 1 0 10 12

1. Calculate the following. Show your reasoning.

- a. The mean number of math homework problems

- b. The mean absolute deviation (MAD)
2. Interpret the mean and MAD. What do they tell you about the number of homework problems Mai had over these two weeks?
3. Find or calculate the following values and show your reasoning.
 - a. The median, quartiles, maximum, and minimum of Mai's data
 - b. The interquartile range (IQR)
4. Which pair of measures of center and variability—mean and MAD, or median and IQR—do you think summarizes the distribution of Mai's math homework assignments better? Explain your reasoning.

Student Response

1.
 - a. The mean is 9 homework problems per day. $\frac{2+15+20+0+5+25+1+0+10+12}{10} = \frac{90}{10} = 9$.
 - b. The MAD is 7.4 homework problems per day. This is computed using the absolute deviations from the mean: 7, 6, 11, 9, 4, 16, 8, 9, 1, 3. The average of these deviations is $\frac{7+6+11+9+4+16+8+9+1+3}{10} = \frac{74}{10}$.
2. The mean tells us that a typical number of homework problems given in a day is 9. Since the MAD is 7.4, which is almost as large as the mean, the data is widely spread out. There is a lot of variation in the number of homework problems given each day.
3.
 - a. The data listed in order is: 0, 0, 1, 2, 5, 10, 12, 15, 20, 25. This means that Q1 is 1, Q2, or the median, is 7.5 (the average of the 5th and 6th data points), and Q3 is 15. The minimum is 0, and the maximum is 25.
 - b. The IQR is 14, which is the difference between Q3 and Q1.
4. Answers vary. Sample response: Both the MAD and IQR show the large variability in the data, and the mean and median are pretty close in value. I think the median and IQR better summarize the center and spread of the data. There are more days where not much homework was given, and the median is lower than the mean.

Activity Synthesis

Briefly discuss students' interpretations of the measures they just calculated:

- "What do the mean of 9 and MAD of 7.4 tell us? How can we interpret them in this context?"
- "What do the median of 7.5 and IQR of 14 tell us?"

Then, select two or more previously identified students to share their responses about which measures of center and spread are appropriate for summarizing the data set. After each person shares, briefly poll the class to see if others reasoned about the measures the same way. Sum up by asking:

- “Now that you have two pairs of measures of center and spread, how would you respond if someone asked you, ‘What is a typical number of homework problems for Mai’s class?’ Is the question easier to answer now?”

Students should walk away with increased awareness that, in some cases, measures of center and spread do not always paint a full picture of what the actual data set entails, and that the measures should be interpreted with care.

18.3 Math Homework (Part 2)

Optional: 15 minutes (there is a digital version of this activity)

In the previous activity, students considered appropriate measures of center and spread for describing distributions. Here, they show the same data set using three different kinds of graphical representations—a dot plot, a box plot, and histograms using different bin sizes—and decide which are more useful or more appropriate for communicating the distribution.

As students work and discuss, identify those who draw clear graphical displays, those who noticed that the different displays offer different insights about the data distribution, and those who advocate for using different representations to display Jada's data. Ask them to share with the class later.

Addressing

- 6.SP.B

Instructional Routines

- MLR2: Collect and Display

Launch

Arrange students in groups of 3–4. Provide access to straightedges.

Explain to students that they will now represent Jada's homework data graphically and think about which representation(s) might appropriately communicate the distribution of her data. Give students 4–5 quiet minutes to draw a dot plot and a box plot (the first two questions), and then another 4–5 minutes to collaborate on drawing histograms with different bin sizes. Ask each student in a group to be in charge of one histogram with a particular bin size. After all representations are drawn, students should analyze them and discuss the last question in their group.

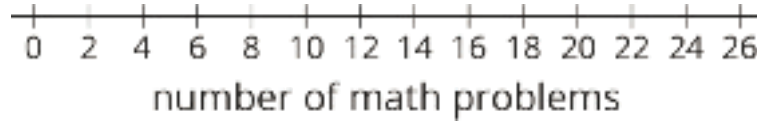
Classes using the digital version have an applet to create the statistical graphs. Data must be entered as a list, in curved brackets, separated by commas. Choices for histogram settings appear when that graph is selected.

Student Task Statement

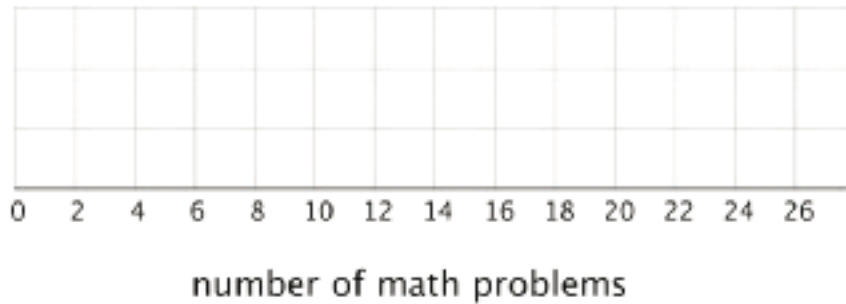
Jada wanted to know whether a dot plot, a histogram, or a box plot would best summarize the center, variability, and other aspects of her homework data.

2 15 20 0 5 25 1 0 10 12

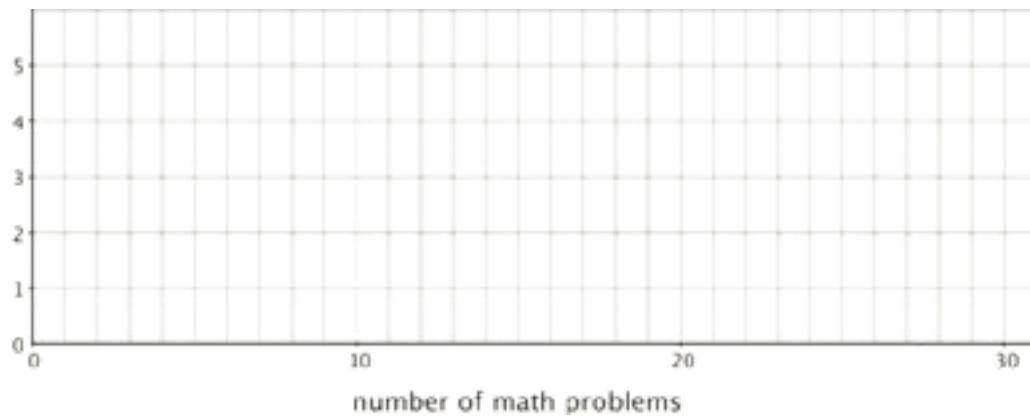
1. Use the axis to make a dot plot to represent the data. Mark the position of the mean, which you calculated earlier, on the dot plot using a triangle (Δ). From the triangle, draw a horizontal line segment to the left and right sides to represent the MAD.



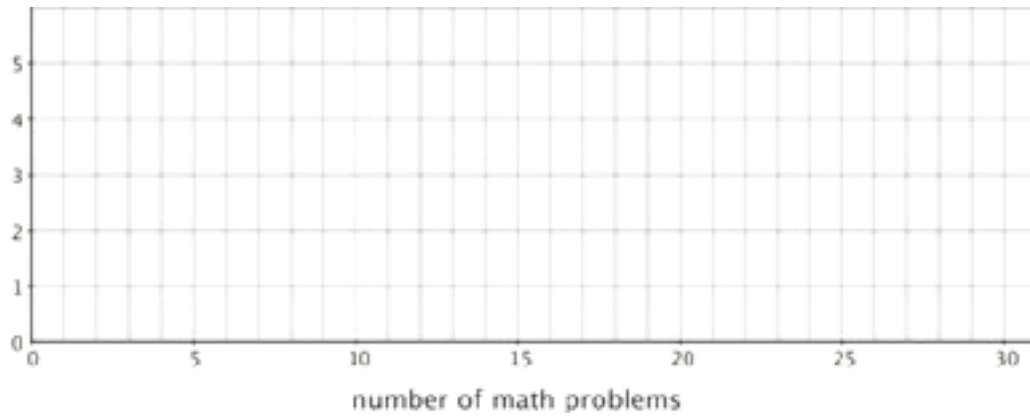
2. Draw a box plot that represents Jada's homework data.



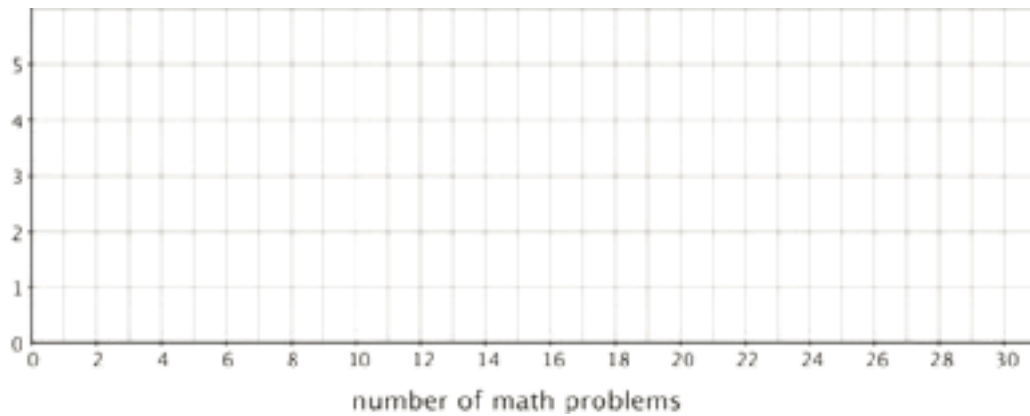
3. Work with your group to draw three histograms to represent Jada's homework data. The width of the bars in each histogram should represent a different number of homework problems, which are specified as follows.
 - a. The width of one bar represents 10 problems.



- b. The width of one bar represents 5 problems.

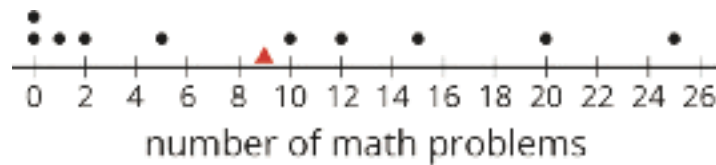


c. The width of one bar represents 2 problems.

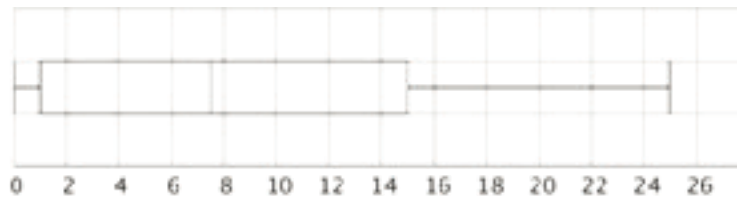


4. Which of the five representations should Jada use to summarize her data? Should she use a dot plot, box plot, or one of the histograms? Explain your reasoning.

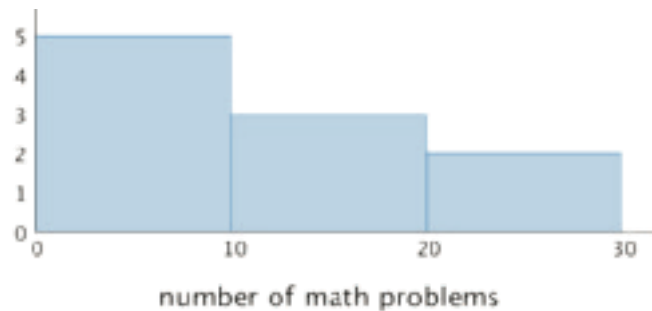
Student Response



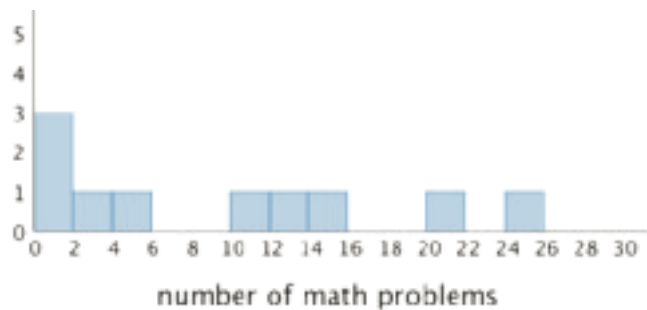
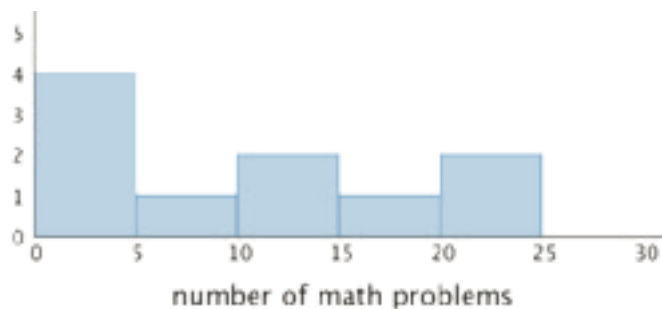
1.



2.



3.



4. Answers vary. Sample responses:

- Jada should use the dot plot. Because it shows all the data points, we could use it to find the mean and MAD, or the median and IQR, to summarize the data.
- Jada should use the box plot. It shows that for about three-quarters of the days Jada's class had 15 or fewer problems, so we know that days with more than 15 problems are *not* typical.
- Jada should use the histogram with a bin size of 10 or 5. All the histograms more or less communicate the distribution, but the last one (where one bar represents 2 problems) is not much different from the dot plot. The first histogram (where one bar represents 10 problems) is helpful because tells us that on most days Jada's class had fewer than 20 problems. The second histogram (where one bar represents 5 problems) gives us more detail but is at the same time harder to summarize or describe. We can see that one half of the days the class had fewer than 10 problems, and the other half they had 10 or more problems. (A sample counter-argument: With only 10 values in the data set, the individual values should be used to gain accuracy rather than combining them into groups and losing that information.)

Activity Synthesis

Invite previously identified students to share their dot plot, box plot, and histograms. Display their drawings for all to see. Then, select several students or groups to share their response to the last question (which representation should Jada choose?) and their explanation. If not already mentioned by students, discuss the different insights that each display offers, or different challenges it poses. (Some possible observations are listed under Student Response section.) For instance, consider asking the following questions about each data display:

- “What information can we get from this display?”
- “Does it give us a meaningful snapshot of the distribution?”
- “What characteristics of a different data set would make this representation more useful?”

Help students see that, in this case, none of the representations here are ill-suited to represent the data set, but a couple of them (e.g. the box plot, or the first histogram with a bin size of 10) allow us to describe the distribution the data set more easily because of how they summarize the data values in some ways.

Access for Students with Disabilities

Representation: Internalize Comprehension. Use color coding and annotations to highlight differences between how dot plots, box plots, and histograms represent the data.

Supports accessibility for: Visual-spatial processing

Access for English Language Learners

Representing, Listening: MLR2 Collect and Display. As students discuss which representation Jada should choose, collect students' responses in a graphic organizer, such as a Venn diagram, and display for all to see. Throughout the remainder of the lesson, continue to update collected student language and remind students to borrow language from the display as needed. Chart language related to dot plot, box plot, and histogram representations. This will help students to use mathematical language during paired and group discussions.

Design Principle(s): Support sense-making; Maximize meta-awareness

18.4 Will the Yellow Perch Survive?

Optional: 30 minutes

In this culminating activity, students use what they have learned in the unit to answer statistical questions about a species of fish in the Great Lakes region. They use a histogram to represent the given data distribution, decide on appropriate measures of center and variability, and use their analyses to draw conclusions about a certain fish population.

Addressing

- 6.SP.B

Instructional Routines

- Group Presentations
- MLR8: Discussion Supports

Launch

Tell students that they will now look at an example in which data analysis could be used to help conservation efforts. Provide students with the following background information.

The yellow perch is a freshwater fish that is a popular food for people in the Great Lakes region (Minnesota, Wisconsin, Michigan, Illinois, Indiana, Ohio, Pennsylvania, and New York). In past research, samples of yellow perch taken from the Great Lakes seemed to be mostly male and mostly old. People worried that yellow perch might not survive and efforts were made to limit commercial and individual fishing in order to try to increase in the number of younger fish. An important part of these efforts is to periodically check the typical age of the fish in the Great Lakes.

The Wisconsin Department of Natural Resources and the Great Lakes Water Institute collected data from samples of yellow perch in Lake Michigan. Students at Rufus King High School in Milwaukee, Wisconsin participated in the research. They evaluated the data and presented their findings in a student-conducted press conference. Explain to students that, in this task, they will investigate some of the same questions that these students addressed in their research.

Arrange students in groups of 3–4. Provide access to straightedges. Give students 7–8 minutes of quiet work time for the first three questions, and then 10–12 minutes to discuss their responses, complete the remainder of the task, and prepare a brief presentation on their response to the last set of questions.

Give each group access to tools for creating a visual display. Ask them to support their conclusions with specific pieces of evidence, such as their histogram, their analysis of the distribution, measures of center and spread, etc.

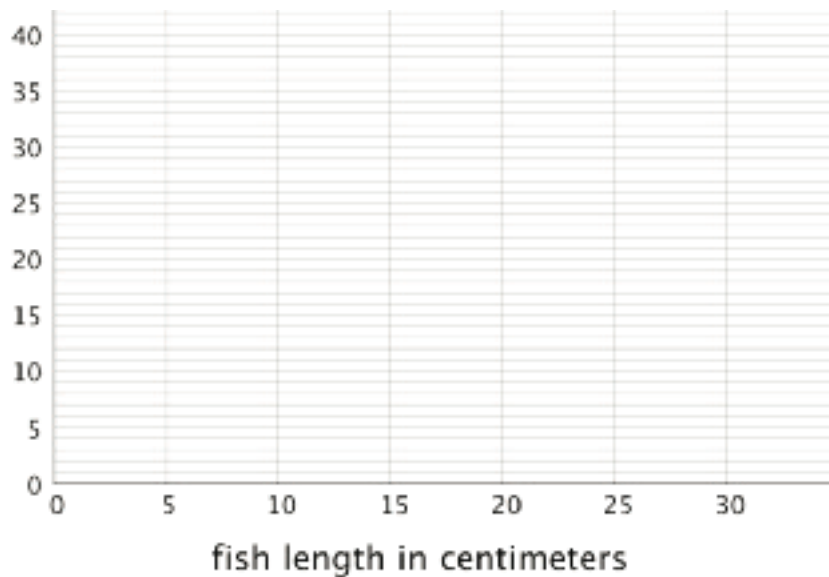
Student Task Statement

Scientists studying the yellow perch, a species of fish, believe that the length of a fish is related to its age. This means that the longer the fish, the older it is. Adult yellow perch vary in size, but they are usually between 10 and 25 centimeters.

Scientists at the Great Lakes Water Institute caught, measured, and released yellow perch at several locations in Lake Michigan. The following summary is based on a sample of yellow perch from one of these locations.

length of fish in centimeters	number of fish
0 to less than 5	5
5 to less than 10	7
10 to less than 15	14
15 to less than 20	20
20 to less than 25	24
25 to less than 30	30

- Use the data to make a histogram that shows the lengths of the captured yellow perch. Each bar should contain the lengths shown in each row in the table.

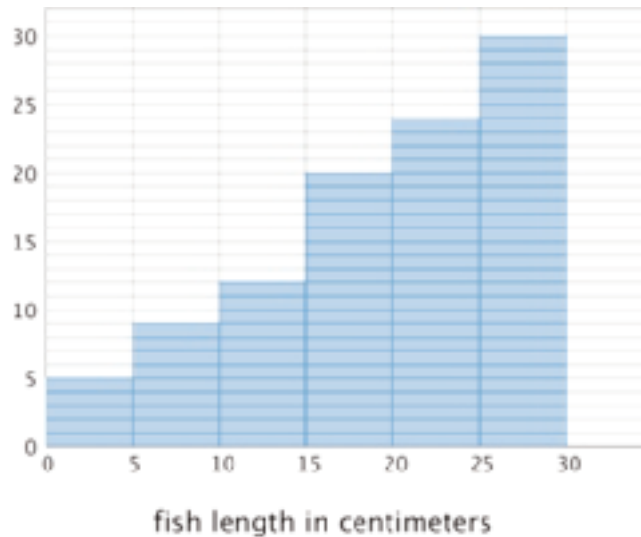


- How many fish were measured? How do you know?
- Use the histogram to answer the following questions.
 - How would you describe the shape of the distribution?
 - Estimate the median length for this sample. Describe how you made this estimate.
 - Predict whether the mean length of this sample is greater than, less than, or nearly equal to the median length for this sample of fish? Explain your prediction.
 - Would you use the mean or the median to describe a typical length of the fish being studied? Explain your reasoning.

4. Based on your work so far:

- a. Would you describe a typical age for the yellow perch in this sample as: “young,” “adult,” or “old”? Explain your reasoning.
- b. Some researchers are concerned about the survival of the yellow perch. Do you think the lengths (or the ages) of the fish in this sample are something to worry about? Explain your reasoning.

Student Response



- 1.
2. 100 fish were measured. The numbers of fish in all length groups add up to 100.
3. Answers vary. Sample responses:
 - a. The data is not symmetrical and has a peak in the range 25–30 cm.
 - b. I estimate the median to be 22–23 cm. I look in the table for where the 50th and 51st value would be and see that it is in the “20 to less than 25 cm” group.
 - c. I predict the mean to be less than the median. Because the data has a peak on the right, the average should be closer to the right as well, but the values that are to the left of the median (or less than the median) would pull the average down.
 - d. I would use the median, because it would better describe where the center of the data is (the data is not symmetric).
4.
 - a. I would describe a typical fish as “old” because its length is on the higher end of the range of adult sizes.
 - b. The lengths seem to be something to worry about. If the yellow perch in the Great Lakes tend to be old, and there are not many young fish around, the species might not survive.

Activity Synthesis

To allow all groups a chance to present, consider putting 2–3 groups together and asking them to present their work to each other. Groups that are not the first to present should focus on sharing new insights that have not been mentioned by the preceding groups. Invite students who are not presenting to attend carefully to the reasoning of the presenting group and to ask clarifying questions.

If time permits, highlight some conclusions that students drew about whether the fish in the sample were young, adult age, or old, and whether researchers should be worried.

Tell students that several years after the students at Rufus King High School participated in the research, newer samples of yellow perch showed more favorable length-age distributions: more of the fish were smaller or younger.

Access for English Language Learners

Speaking: MLR8 Discussion Supports. During the group presentations, provide students with sentence frames such as: “Based on the histogram, I think ____ because . . .” or “I think ____ is/ is not something to worry about because . . .”. This will help students articulate their ideas, and use mathematical language such as mean, median, mean absolute deviation, and interquartile range as supporting evidence during their presentation.

Design Principle(s): Support sense-making; Maximize meta-awareness

Lesson Synthesis

In this lesson we practice finding measures of center and variability (mean, MAD, median, and IQR) and making sense of them in the context of the given situation. We notice that they give us different insights into the distribution of a data set.

- “What do the mean and MAD tell us?” (The mean tells us the fair share or balance point of the distribution and the MAD tells us the average distance a value is from the mean.)
- “How do we interpret this statement: ‘Noah’s mean number of homework problems per day is 10 and the MAD is 6.’?” (If we were to distribute Noah’s assignments so that the number of problems he has each day is the same, he would have 10 per day. The MAD of 6 tells us that there is some variability in the number of problems assigned, so not all days have exactly 10 problems assigned. The average distance between the number of problems assigned and the mean of 10 is 6.)
- “What do the median and IQR tell us?” (The median tells us the value for which half the data set is equal to or greater and half the data set is equal to or less and the IQR tells us the range for the middle half of the data set.)
- “How do we interpret this statement: ‘Lin’s median number of homework problems per day is 10 and the IQR is 6.’?” (One half of Lin’s assignments involve 10 or fewer problems, and the

other half involve 10 or more problems. The IQR tells us that half of Lin's assignments are between 7 and 13 problems.)

We also looked at different ways to graphically represent a numerical distribution.

- "What are the ways we can represent a data set?" (Dot plot, histogram, box plot.)
- "Which representations are helpful for summarizing a distribution?" (It varies depending on the distribution we're studying and what information we want to know.)

18.5 Time Spent on Chores

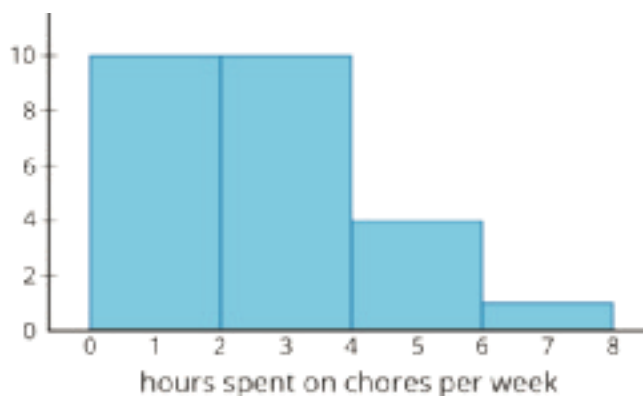
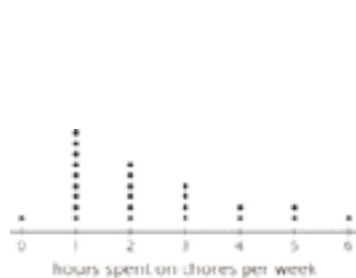
Cool Down: 5 minutes

Addressing

- 6.SP.B.5.c
- 6.SP.B.5.d

Student Task Statement

Lin surveyed her classmates on the number of hours they spend doing chores each week. She represented her data with a dot plot and a histogram.



1. Lin thinks that she could find the median, the minimum, and the maximum of the data set using both the dot plot and the histogram. Do you agree? Explain your reasoning.
2. Should Lin use the mean and MAD or the median and IQR to summarize her data? Explain your reasoning.

Student Response

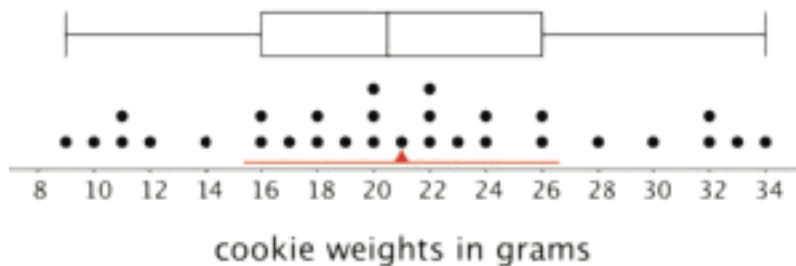
Answers vary. Samples responses.

1. Disagree. The dot plot makes it possible to find the median, the minimum, and the maximum fairly easily since it shows each data value individually. The histogram makes it possible to estimate these values, but it is impossible to tell the exact values because the data points are grouped together.

2. Lin should use the median and IQR as the data is not approximately symmetrical and has values far from the center. There are a few larger values that are not similar to most of the other values.

Student Lesson Summary

The dot plot shows the distribution of 30 cookie weights in grams.

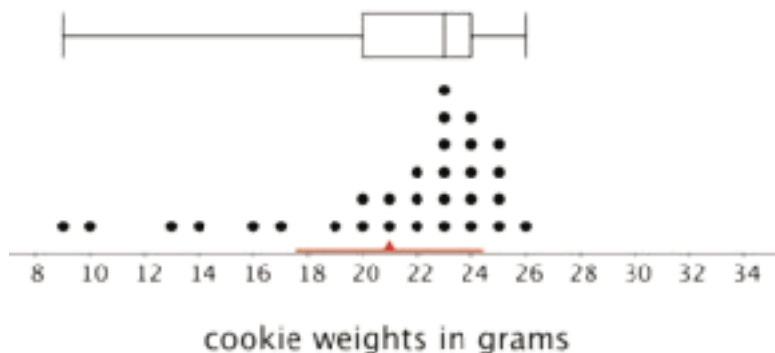


The mean cookie weight, marked by the triangle, is 21 grams. This tells us that if the weights of all of the cookies were redistributed so they all had the same weight, each cookie would weigh 21 grams. The MAD is 5.6 grams, which suggests that a cookie typically weighs between 15.4 grams and 26.6 grams.


The box plot for the same data set is shown above the dot plot. The median shows that half of the weights are greater than or equal to 20.5 grams, and half are less than or equal to 20.5 grams. The box shows that the IQR is 10 and that the middle half of the cookies weigh between 16 and 26 grams.

In this case, the median weight is very close to the mean weight, and the IQR is about twice the MAD. This tells us that the two pairs of measures of center and spread are very similar.

Now let's look at another example of 30 different cookies.



Here the mean is 21 grams, and the MAD is 3.4 grams. This suggests that a cookie typically weighs between 17.6 and 24.4 grams. The median cookie weight is 23 grams, and the box plot shows that the middle half of the data are between 20 and 24 grams. These two pairs of measures paint very different pictures of the variability of the cookie weights.



The median (23 grams) is closer to the middle of the big cluster of values. If we were to ignore the smaller cookies, the median and IQR would give a more accurate picture of how much a cookie typically weighs.

When a distribution is not symmetrical, the median and IQR are often better measures of center and spread than the mean and MAD. However the decision on which pair of measures to use depends on what we want to know about the group we are investigating.

CKMath™
Core Knowledge **MATHEMATICS™**



Family Support
Materials

Family Support Materials

Data Sets and Distributions

Here are the video lesson summaries for Grade 6, Unit 8: Data Sets and Distributions. Each video highlights key concepts and vocabulary that students learn across one or more lessons in the unit. The content of these video lesson summaries is based on the written Lesson Summaries found at the end of lessons in the curriculum. The goal of these videos is to support students in reviewing and checking their understanding of important concepts and vocabulary. Here are some possible ways families can use these videos:

- Keep informed on concepts and vocabulary students are learning about in class.
- Watch with their student and pause at key points to predict what comes next or think up other examples of vocabulary terms (the bolded words).
- Consider following the Connecting to Other Units links to review the math concepts that led up to this unit or to preview where the concepts in this unit lead to in future units.

Grade 6, Unit 8: Data Sets and Distributions	Vimeo	YouTube
Video 1: Data and Variability (Lessons 1–3)	Link	Link
Video 2: Distributions and Histograms (Lessons 4–8)	Link	Link
Video 3: Mean (Lessons 9–10)	Link	Link
Video 4: Variability and MAD (Lessons 11–12)	Link	Link
Video 5: Median (Lessons 13–14)	Link	Link
Video 6: Five Number Summary and Box Plots (Lessons 15–17)	Link	Link

Video 1

Video 'VLS G6U8V1 Data and Variability (Lessons 1–3)' available here:
<https://player.vimeo.com/video/523867530>.

Video 2

Video 'VLS G6U8V2 Distributions and Histograms (Lessons 4–8)' available here:
<https://player.vimeo.com/video/523871227>.

Video 3

Video 'VLS G6U8V3 Mean (Lessons 9–10)' available here: <https://player.vimeo.com/video/524348492>.

Video 4

Video 'VLS G6U8V4 Variability and MAD (Lessons 11–12)' available here:
<https://player.vimeo.com/video/524349628>.

Video 5

Video 'VLS G6U8V5 Median (Lessons 13–14)' available here: <https://player.vimeo.com/video/526589069>.

Video 6

Video 'VLS G6U8V6 Five Number Summary and Box Plots (Lessons 15–17)' available here:
<https://player.vimeo.com/video/529045240>.

Connecting to Other Units

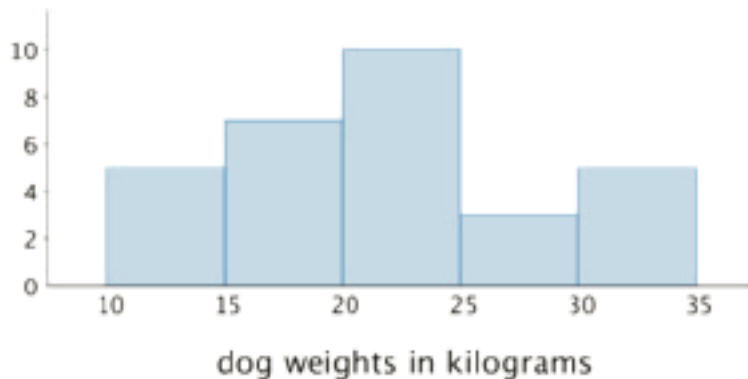
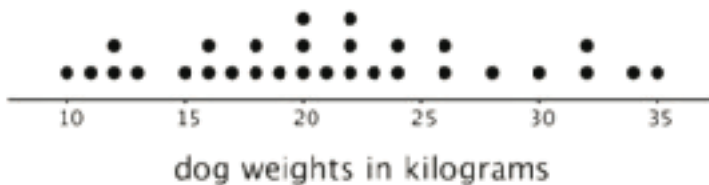
- *Coming soon*

Data, Variability, and Statistical Questions

Family Support Materials 1

This week, your student will work with data and use data to answer **statistical questions**. Questions such as “Which band is the most popular among students in sixth grade?” or “What is the most common number of siblings among students in sixth grade?” are statistical questions. They can be answered using data, and the data are expected to vary (i.e. the students do not all have the same musical preference or the same number of siblings).

Students have used bar graphs and line plots, or **dot plots**, to display and interpret data. Now they learn to use **histograms** to make sense of numerical data. The following dot plot and histogram display the distribution of the weights of 30 dogs.

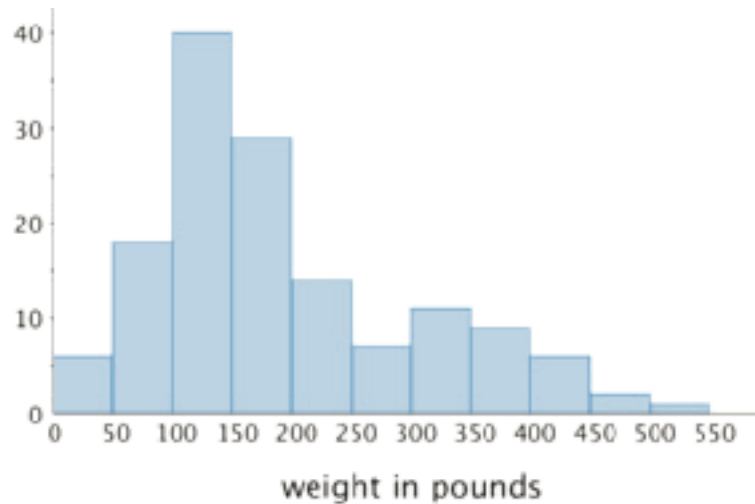


A dot plot shows individual data values as points. In a histogram, the data values are grouped. Each group is represented as a vertical bar. The height of the bar shows how many values are in that group. The tallest bar in this histogram shows that there are 10 dogs that weigh between 20 and 25 kilograms.

The shape of a histogram can tell us about how the data are distributed. For example, we can see that more than half of the dogs weigh less than 25 kilograms, and that a dog weighing between 25 and 30 kilograms is not typical.

Here is a task to try with your student:

This histogram shows the weights of 143 bears.



1. About how many bears weigh between 100 and 150 pounds?
2. About how many bears weigh less than 100 pounds?
3. Noah says that because almost all the bears weigh between 0 and 500 pounds, we can say that a weight of 250 pounds is typical for the bears in this group. Using the histogram, explain why this is incorrect.

Solution:

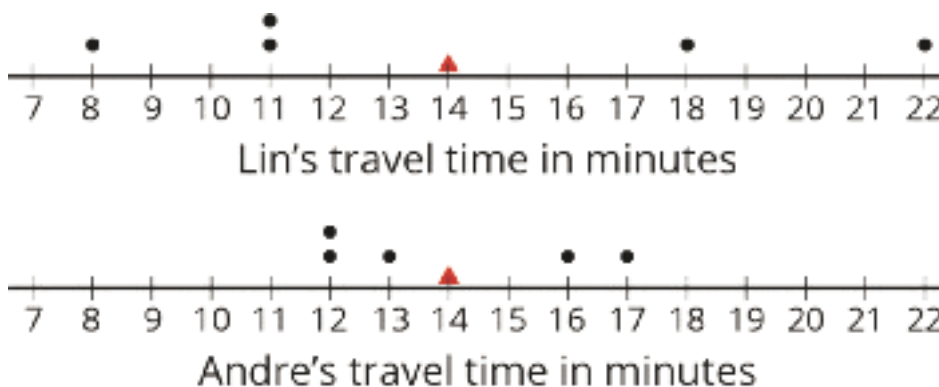
1. About 40 bears. This is the height of the tallest bar of the histogram.
2. About 24 bears. The two leftmost bars represent the bears that weigh less than 100 pounds. Add the heights of these two bars.
3. We can visually tell from the histogram that most bears weigh less than 250 pounds: the bars to the left of 250 are taller than those to the right. If we add the heights of bars, fewer than 40 bears weigh more than 250 pounds, while over 100 bears weigh less than 250 pounds, so it is not accurate to say that 250 pounds is a typical weight.

Measures of Center and Variability

Family Support Materials 2

This week, your student will learn to calculate and interpret the **mean**, or the average, of a data set. We can think of the mean of a data set as a fair share—what would happen if the numbers in the data set were distributed evenly. Suppose a runner ran 3, 4, 3, 1, and 5 miles over five days. If the total number of miles she ran, 16 miles, was distributed evenly across five days, the distance run per day, 3.2 miles, would be the mean. To calculate the mean, we can add the data values and then divide the sum by how many there are.

If we think of data points as weights along a number line, the mean can also be interpreted as the balance point of the data. The dots show the travel times, in minutes, of Lin and Andre. The triangles show each mean travel time. Notice that the data points are “balanced” on either side of each triangle.



Your student will also learn to find and interpret the **mean absolute deviation** or the **MAD** of data. The MAD tells you the distance, on average, of a data point from the mean. When the data points are close to the mean, the distances between them and the mean are small, so the average distance—the MAD—will also be small. When data points are more spread out, the MAD will be greater.

We use mean and MAD values to help us summarize data. The mean is a way to describe the center of a data set. The MAD is a way to describe how spread out the data set is.

Here is a task to try with your student:

1. Use the data on Lin's and Andre's dot plots to verify that the mean travel time for each student is 14 minutes.

2. Andre says that the mean for his data should be 13 minutes, because there are two numbers to the left of 13 and two to the right. Explain why 13 minutes cannot be the mean.
3. Which data set, Lin's or Andre's, has a higher MAD (mean absolute deviation)? Explain how you know.

Solution:

1. For Lin's data, the mean is $\frac{8+11+11+18+22}{5} = \frac{70}{5}$, which equals 14. For Andre's data, the mean is $\frac{12+12+13+16+17}{5} = \frac{70}{5}$, which also equals 14.
2. Explanations vary. Sample explanations:
- The mean cannot be 13 minutes because it does not represent a fair share.
 - The mean cannot be 13 minutes because the data would be unbalanced. The two data values to the right of 13 (16 and 17) are much further away from the two that are to the left (12 and 12).
3. Lin's data has a higher MAD. Explanations vary. Sample explanations:
- In Lin's data, the points are 6, 3, 3, 4, and 8 units away from the mean of 14. In Andre's data, the points are 2, 2, 1, 2, and 3 units away from the mean of 14. The average distance of Lin's data will be higher because those distances are greater.
 - The MAD of Lin's data is 4.8 minutes, and the MAD of Andre's data is 2 minutes.
 - Compared to Andre's data points, Lin's data points are farther away from the mean.

Median and IQR

Family Support Materials 3

This week, your student will learn to use the **median** and **interquartile range** or **IQR** to summarize the distribution of data.

The median is the middle value of a data set whose values are listed in order. To find the median, arrange the data in order from least to greatest, and look at the middle of the list.

Suppose nine students reported the following numbers of hours of sleep on a weeknight.

6 7 7 8 9 9 10 11 12

The middle number is 9, so the median number of hours of sleep is 9 hours. This means that half of the students slept for less than or equal to 9 hours, and the other half slept for greater than or equal to 9 hours.

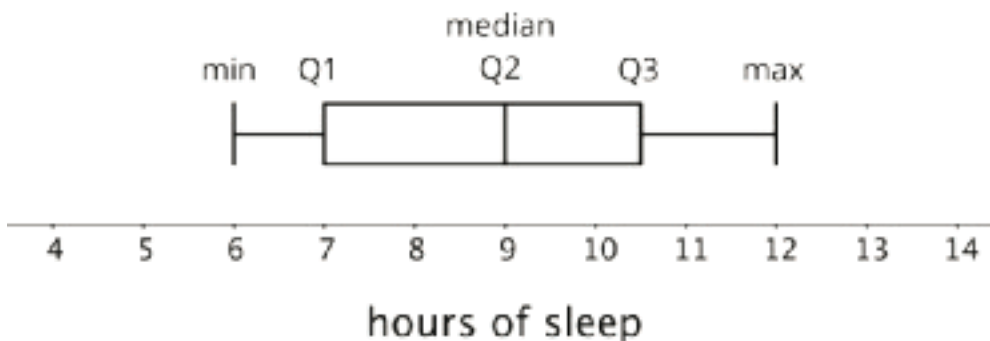
Suppose eight teachers reported these numbers of hours of sleep on a weeknight.

5 6 6 6 7 7 7 8

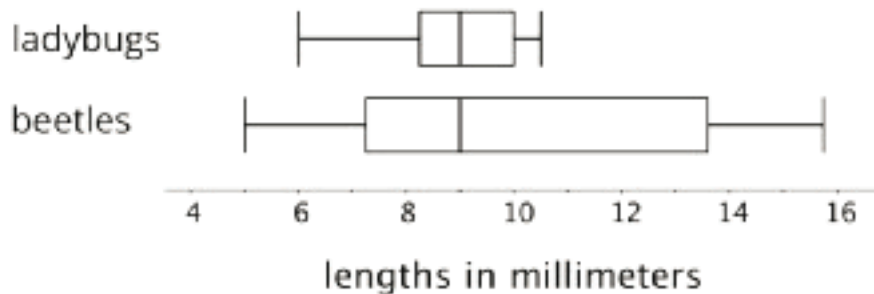
This data set has an even number of values, so there are two numbers in the middle—6 and 7. The median is the number exactly in between them: 6.5. In other words, if there are two numbers in the middle of a data set, the median is the average of those two numbers.

The median marks the 50th percentile of sorted data. It breaks a data set into two halves. Each half can be further broken down into two parts so that we can see the 25th and 75th percentiles. The 25th, 50th, and 75th percentiles are called the first, second, and third **quartiles** (or Q1, Q2, and Q3).

A **box plot** is a way to represent the three quartiles of a data set, along with its maximum and minimum. This box plot shows those five numbers for the data on the students' hours of sleep.



The distance between the first and third quartiles is the **interquartile range** or the **IQR** of data. It tells us about the middle half of the data and is represented by the “width” of the box of the box plot. We can use it to describe how alike or different the data values are. Box plots are especially useful for comparing the distributions of two or more data sets.



The box plots show that the smallest measured beetle is 5 millimeters long, and that half of the beetles are between approximately 7 and 14 millimeters long.

Here is a task to try with your student:

1. Look at the box plots for the ladybugs and beetles.
 - a. Which group has a greater IQR: ladybugs or beetles? Explain how you know.

 - b. Which group shows more variation in lengths: ladybugs or beetles? Explain how you know.

2. Here is data showing the number of points Jada scored in 10 basketball games.

10 14 6 12 38 12 8 7 10 23

What is her median score?

Solution:

1.
 - a. Beetles have a greater IQR. For ladybugs, the IQR (the distance from the first quartile to the third quartile) is about 1.7 millimeters. For beetles, the IQR is about 6.3 millimeters.
 - b. Beetles show more variation in lengths. Ladybugs are much more alike in their lengths. The IQR for ladybugs is a smaller number and the box in the plot is narrower, which mean that their lengths are fairly close to one another.
2. 11 points. First, sort the data: 6, 7, 8, 10, 10, 12, 12, 14, 23, 38. Then look at the middle of the list: the numbers 10 and 12 are the fifth and sixth numbers in the list. The median is the average of these numbers: $\frac{10+12}{2} = 11$.

Unit Assessments

Check Your Readiness A and B
Mid-Unit Assessment A and B
End-of-Unit Assessment A and B

Data Sets and Distributions: Check Your Readiness (A)

Do not use a calculator.

1. Calculate each quotient.

a. $125 \div 5$

b. $800 \div 16$

c. $3.5 \div 2$

d. $21.4 \div 10$

2. Evaluate each expression.

a. $16 - 2.2$

b. $3 - 0.26$

c. $11.26 + 5.34$

d. $2.06 + 3.94$

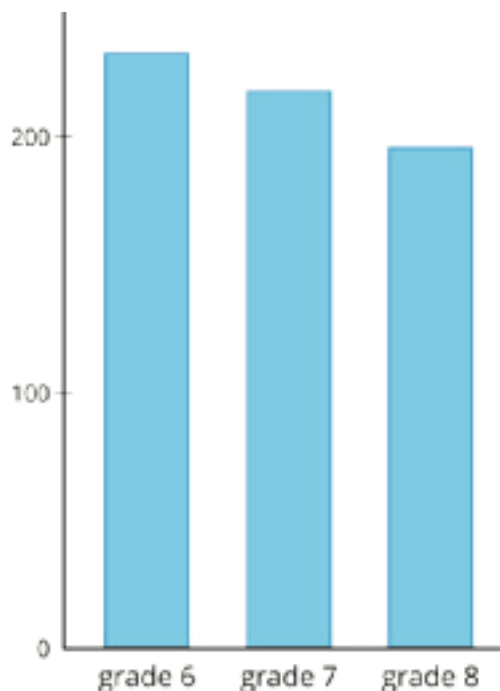
e. $51 - 0.07$

3. The bar graph shows how many students are in grades 6, 7, and 8 at a school.

a. About how many total students are in these three grades?

b. Which grade has the most students?

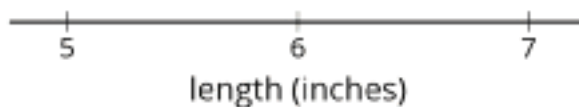
c. About how many more sixth graders are there than eighth graders?



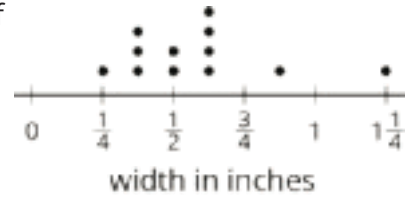
4. A pet store has 11 lizards, with these lengths in inches:

$5\frac{1}{2}$ $5\frac{3}{4}$ $5\frac{3}{4}$ $5\frac{7}{8}$ $5\frac{7}{8}$ 6 $6\frac{1}{8}$ $6\frac{1}{8}$
 $6\frac{1}{8}$ $6\frac{1}{4}$ $6\frac{3}{8}$

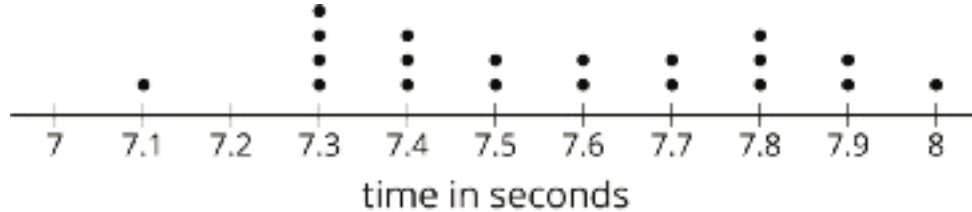
Draw a line plot for this data.



5. Diego drew this line plot showing the width, in inches, of all the bolts in a storage bin. Select **all** the true statements.



- A. There are 12 bolts in the storage bin.
- B. The widest bolt is $\frac{7}{8}$ inches wide.
- C. Half of the bolts are less than $\frac{1}{2}$ inch wide.
- D. None of the bolts is exactly $\frac{2}{8}$ inches wide.
- E. None of the bolts is exactly $\frac{6}{8}$ inches wide.
- F. The difference between the maximum and minimum width of the bolts is 1 inch.
6. This line plot shows the amount of time, in seconds, that it took 20 sixth graders to run a 50-meter dash.



- Select **all** the true statements.
- A. The fastest time was 7.0 seconds.
- B. No runner recorded a time of 7.2 seconds.
- C. The fastest 5 students' total time was 36.3 seconds.
- D. Exactly half of the students were faster than 7.7 seconds.
- E. The difference between the fastest and slowest times was 0.9 seconds.

7. Calculate each percentage.

a. 25% of 50

b. 25% of 60

c. 50% of 60

d. 75% of 60

e. 75% of 30

f. 100% of 22.5

g. 10% of 22.5

h. 50% of 45.7

Data Sets and Distributions: Check Your Readiness (B)

Do not use a calculator.

1. Calculate each quotient.

a. $180 \div 5$

b. $600 \div 15$

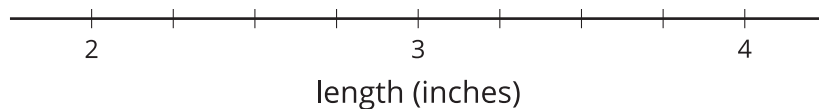
c. $7.5 \div 2$

d. $51.6 \div 10$

2. A toy set has 10 mini toy dinosaurs, with these lengths in inches:

length (inches)	$2\frac{1}{4}$	$2\frac{1}{4}$	$2\frac{1}{4}$	$2\frac{1}{8}$	$2\frac{1}{2}$	$2\frac{5}{8}$	$2\frac{5}{8}$	3	3	$3\frac{3}{4}$
--------------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	---	---	----------------

Draw a line plot for this data.



3. Evaluate each expression.

a. $11 - 4.9$

b. $7 - 0.18$

c. $13.23 + 2.67$

d. $6.08 + 2.02$

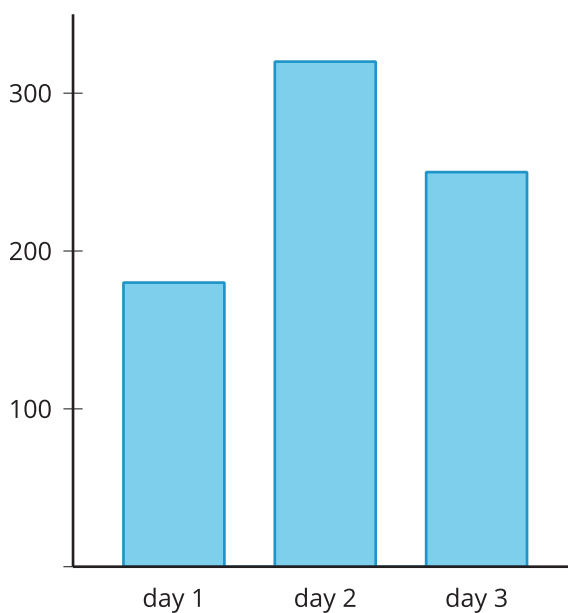
e. $35 - 0.02$

4. The bar graph shows how many people donated blood on 3 different days of the blood drive.

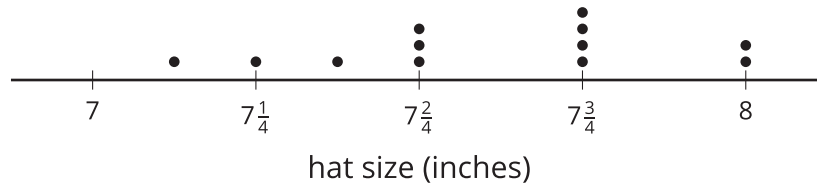
a. About how many total people donated blood on these three days?

b. Which day had the least number of people?

c. About how many more people donated blood on day 2 than on day 3?

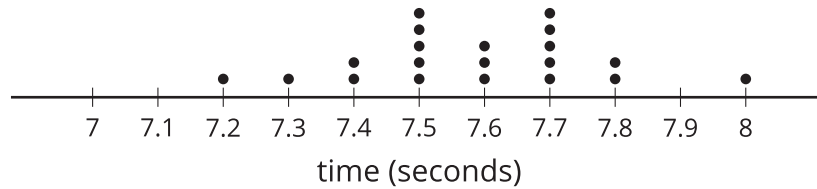


5. Jada drew this line plot showing the hat sizes, in inches, of all the players on her softball team.



Select **all** the true statements.

- A. There are six players on the team.
 - B. The largest hat size is 8 inches.
 - C. Half of the hat sizes are less than $7\frac{5}{8}$ inches.
 - D. None of the players has a hat size of $7\frac{6}{8}$ inches.
 - E. The difference between the maximum and minimum hat size is $\frac{7}{8}$ inch.
6. This line plot shows the amount of time, in seconds, that it took 20 sixth graders to complete a set of 10 sit ups in gym class.



Select **all** the true statements.

- A. The difference between the fastest and slowest times was 0.9 seconds.
- B. The fastest time was 7.0 seconds.
- C. The fastest 3 students' total time was 23.6 seconds.
- D. No student recorded a time of 7.1 seconds.
- E. Exactly half of the students were faster than 7.6 seconds.

7. Calculate each percentage.

a. 25% of 30

b. 25% of 80

c. 50% of 80

d. 75% of 80

e. 75% of 50

f. 100% of 46.5

g. 10% of 46.5

h. 50% of 25.7

Data Sets and Distributions: Mid-Unit Assessment (A)

You may use a four-function or scientific calculator, but not a graphing calculator.

1. Select **all** the statistical questions.

- A. How many nickels does it take to make a dollar?
- B. In what year was the first penny made in the United States?
- C. Among all the pennies at a bank, what is the most frequent year the pennies were made?
- D. Which coin (penny, nickel, dime, or quarter) is used most frequently in transactions at a bank?
- E. On average, how many pennies do people receive in change when they make a purchase at a store?

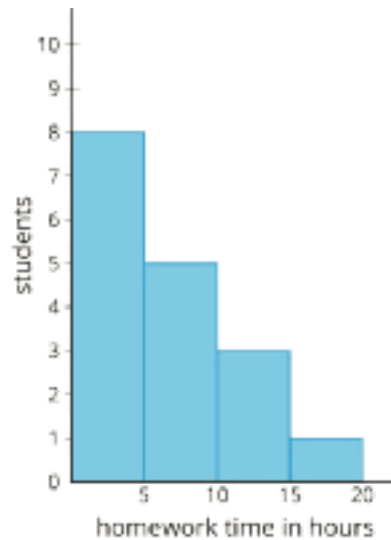
2. This dot plot shows the number of hours 20 sixth grade students slept on a Saturday night.



Select **all** the true statements about the data used to build the dot plot.

- A. Six students slept for at least 8 hours.
- B. The mean amount of sleep was 6.25 hours.
- C. More than half of the students slept 7 hours or less.
- D. Only 1% of the students slept more than 8 hours.
- E. The difference between the most hours of sleep and the least for these students was 2.5 hours.

3. Students responded to a survey. The survey asked students to report the amount of time they spent doing homework during a week, to the nearest hour. This histogram displays the data. Which of these statements *must* be true?

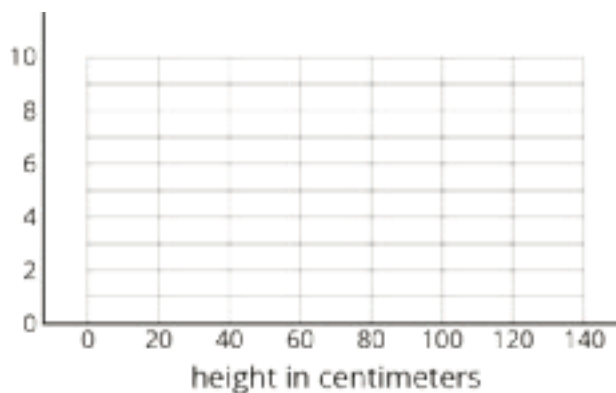


- A. A total of 4 students participated in the survey.
- B. Every student spent at least 1 hour doing homework.
- C. Every student spent less than 20 hours doing homework.
- D. More students spent 7 hours doing homework than spent 12 hours.

4. Here is the height of 20 flowers in the school garden, in centimeters.

5	5	10	10	15	25	25	25
25	30	30	35	45	45	50	55
65	70	105	110				

a. Draw a histogram to display the data.

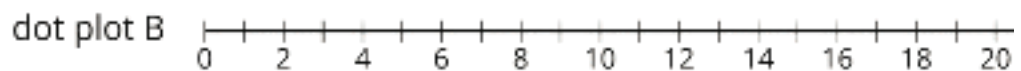
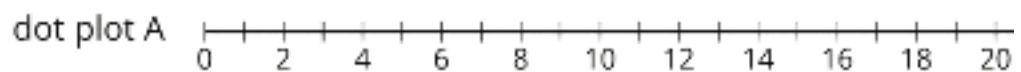


b. Based on the histogram, what is a typical length for these 20 flowers?

5. The mean of four numbers is 25. Three of the four numbers are 20, 26, and 26. What is the fourth number?

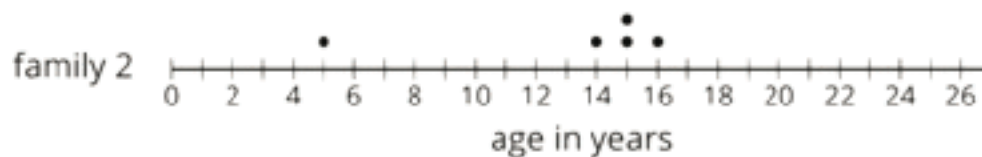
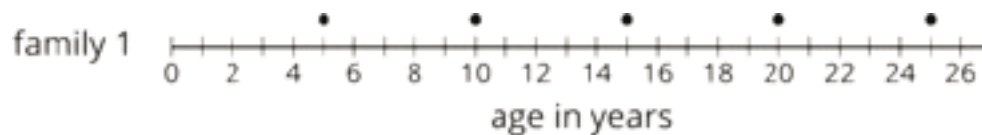
6. a. Draw two dot plots, each with 7 or fewer data points, so that:

- both dot plots display data with approximately the same mean
- the data displayed in Dot Plot A has a much larger MAD (mean absolute deviation) than the data displayed in Dot Plot B.



b. How can you tell, visually, that one dot plot displays data with a larger MAD than another?

7. Here are two dot plots, the ages of the five children in each of two families.

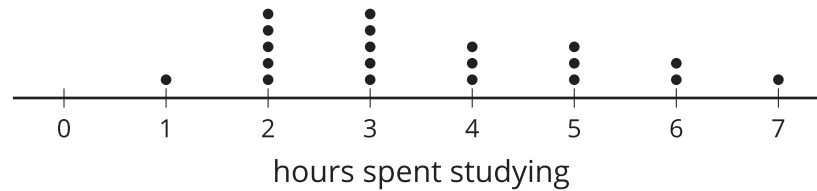


- a. For each family, identify anyone whose age is unusual compared to the rest of the family.
- b. Calculate the MAD (mean absolute deviation) of each data set. Which family has the wider spread of ages?

Data Sets and Distributions: Mid-Unit Assessment (B)

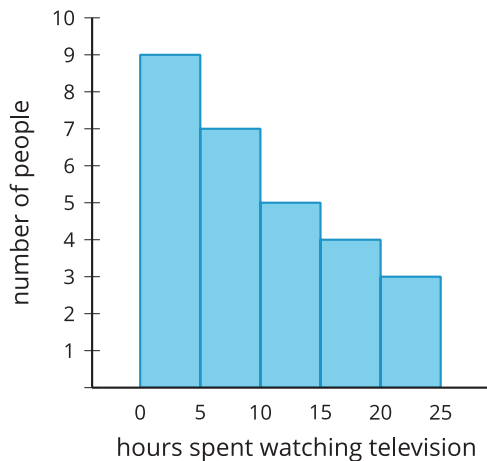
You may use a four-function or scientific calculator, but not a graphing calculator.

1. Select **all** the statistical questions.
 - A. How many cups are in a gallon?
 - B. Which measuring spoon (tablespoon, teaspoon, half-teaspoon, or quarter-teaspoon) do home bakers use most frequently?
 - C. On average, how much salt do professional pastry chefs use in a pie crust?
 - D. Which is larger, 17 ounces or 2 cups?
 - E. Did a specific store sell more measuring cups or measuring spoons on a specific day?
2. The dot plot shows the number of hours 20 sixth grade students spent studying in a week.



- Select **all** true statements about the data used to build the dot plot.
- A. The difference between the most hours and the least hours spent studying was 6 hours.
 - B. The mean amount of study time was 5 hours.
 - C. Nine students studied for at least 4 hours.
 - D. Most students studied less than 4 hours.
 - E. Only 1% of the students studied more than 6 hours.

3. A survey asked people how many hours they spend watching television during a week, to the nearest hour. The histogram displays the data.

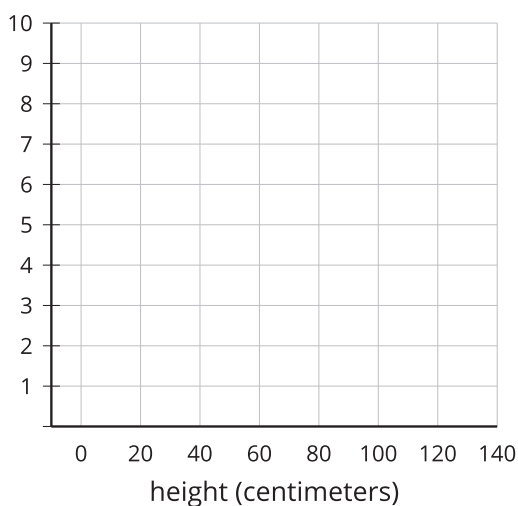


Which of these statements *must* be true?

- A. Every person in the survey spent less than 10 hours watching television.
 - B. Every person in the survey watched some television during the week.
 - C. There are more people in the survey who watched 12 hours of television than people who watched 17 hours.
 - D. More than 20 people participated in the survey.
4. Here is the height of 20 flowers in the school garden, in centimeters.

5	5	10	10	15	25	25	30	35	45	45	45	45	50	50	55	65	70	105	110
---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

a. Draw a histogram to display the data.

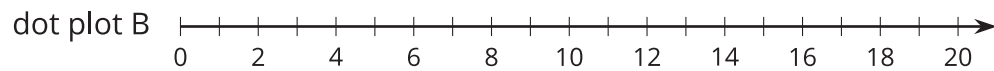
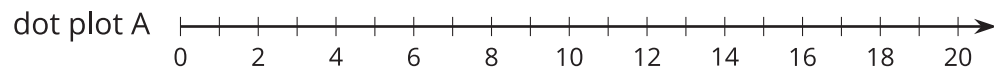


b. Based on the histogram, what is a typical height for these 20 flowers?

5. The mean of four numbers is 40. Three of the four numbers are 35, 41, and 41. What is the fourth number?

6. Draw two dot plots, each with 7 or fewer data points, so that:

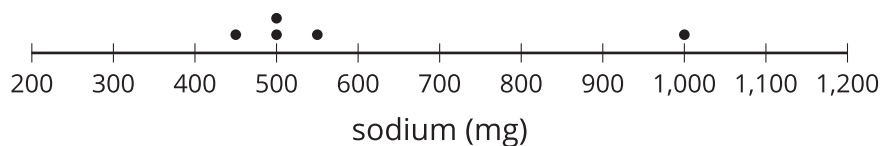
- both dot plots display data with the same median
- the data displayed in Dot Plot B has a much larger IQR (interquartile range) than the data displayed in Dot Plot A.



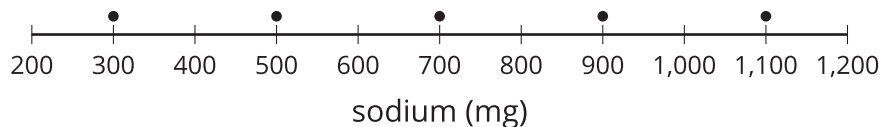
How can you tell, visually, that one dot plot displays data with a larger IQR than another?

7. Here are two dot plots showing the milligrams of sodium in the 5 most popular menu items at two different fast food restaurants.

Restaurant 1



Restaurant 2



- a. For each restaurant, identify any menu item whose sodium is unusual compared to the other items from the restaurant.
- b. Calculate the MAD (mean absolute deviation) of each data set. Which restaurant has menu items with a wider spread of sodium content?

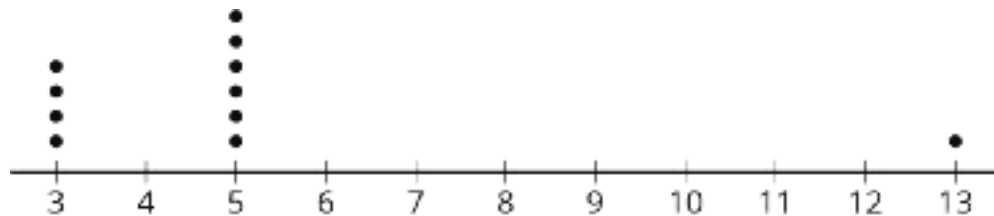
Data Sets and Distributions: End-of-Unit Assessment (A)

You may use a four-function or scientific calculator, but not a graphing calculator.

1. Select **all** the true statements.

- A. Given a box plot, it is always possible to calculate the mean of the data.
- B. Given a box plot, it is always possible to calculate the median of the data.
- C. Given a box plot, it is always possible to construct a corresponding dot plot.
- D. Given a dot plot, it is always possible to construct a corresponding box plot.
- E. Given a histogram, it is always possible to construct a corresponding box plot.

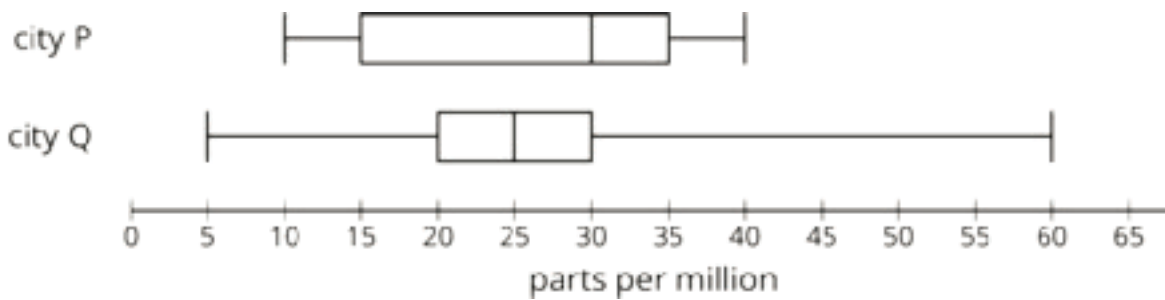
2. Here's a dot plot of a data set.



Which statement is true about the mean of the data set?

- A. The mean is less than 5.
- B. The mean is equal to 5.
- C. The mean is greater than 5.
- D. There is not enough information to determine the mean.

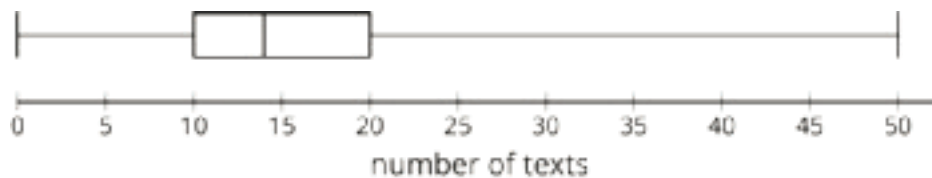
3. The air quality was tested in many office buildings in two cities. The results of the testing are shown in these box plots.



A level of less than 50 parts per million is considered healthy. A level of 50 or more parts per million is considered unhealthy.

Select **all** the statements that must be true.

- A. The lowest recorded measurement was in city Q.
 - B. All buildings tested in city P are in the healthy range.
 - C. The mean for city P is greater than the mean for city Q.
 - D. The range for city Q is greater than the range for city P.
 - E. The median for city P is greater than the median for city Q.
4. This box plot displays information about the number of text messages some students sent one day.

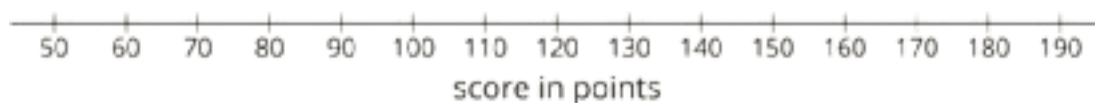


- a. What is the median number of texts sent by students?
- b. What is the IQR (interquartile range)?
- c. Is this data set symmetric? Explain how you know.

5. Two groups went bowling. Here are the scores from each group.

Group A	80	100	190	110	70	90	130
Group B	50	110	100	120	107	140	150

a. Draw two box plots, one for the data in each group.



b. Which group shows greater variability?

6. Ten students each attempted 10 free throws. This list shows how many free throws each student made.

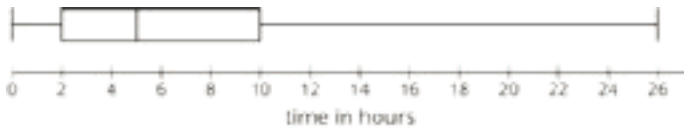
8 5 6 6 4 9 7 6 5 9

a. What is the median number of free throws made?

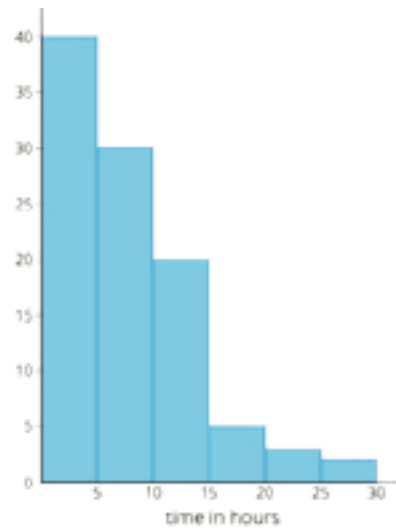
b. What is the IQR (interquartile range)?

7. Jada asked some students at her school how many hours they spent watching television last week, to the nearest hour. Here are a box plot and a histogram for the data she collected.

Box plot:



Histogram:



- About how many students did Jada ask? Explain how you know.
- Is the mean or the median a more appropriate measure of center for this data set? Explain your reasoning.
- Can Jada use these data displays to find the exact median? Explain how you know.
- Can Jada use these data displays to find the exact mean? Explain how you know.
- What would be an appropriate measure of variability for this data set? Find or estimate its value.

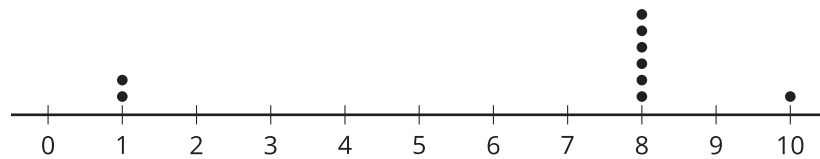
Data Sets and Distributions: End-of-Unit Assessment (B)

You may use a four-function or scientific calculator, but not a graphing calculator.

1. Select **all** the true statements.

- A. Given a dot plot, it is always possible to construct a corresponding histogram.
- B. Given a dot plot, it is always possible to calculate the mean of the data.
- C. Given a box plot, it is always possible to calculate the IQR of the data.
- D. Given a histogram, it is always possible to calculate the mean of the data.
- E. Given a histogram, it is always possible to calculate the median of the data.

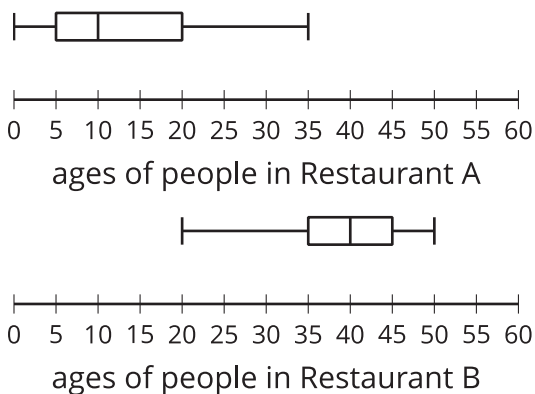
2. Here is a dot plot of a data set.



Which statement is true about the mean of the data set?

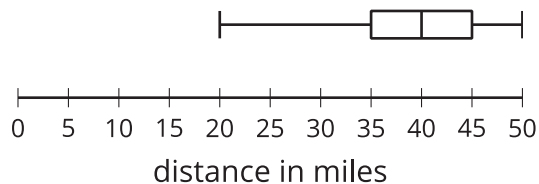
- A. The mean is less than 8.
- B. The mean is equal to 8.
- C. The mean is greater than 8.
- D. There is not enough information to determine the mean.

3. The ages of people dining in two restaurants are shown in the following box plots.



Select **all** the statements that must be true.

- A. The median age of people dining in Restaurant B is greater than the median age of people dining in Restaurant A.
 - B. The MAD (mean absolute deviation) for Restaurant B is greater than the MAD (mean absolute deviation) for Restaurant A.
 - C. The youngest person was dining in Restaurant A.
 - D. The IQR (interquartile range) for Restaurant A is equal to the IQR (interquartile range) for Restaurant B.
 - E. Every person dining in Restaurant B is older than everyone dining in Restaurant A.
4. This box plot displays information about the distance in miles teachers drive to school each day.

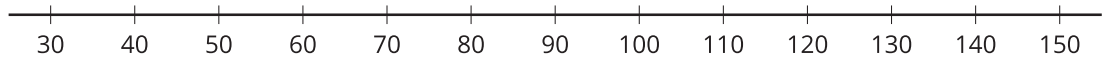


- a. What is the IQR (interquartile range)?
- b. What is the median distance in miles driven by teachers?
- c. Is this data set symmetric? Explain how you know.

5. Two groups of students had a car wash fundraiser every day for a week in the summer. Here are the funds raised from each group.

Group P	80	60	65	80	70	85	70
Group Q	60	140	80	40	50	90	50

a. Draw two box plots, one for the data in each group.



b. Which group shows greater variability? Explain your reasoning.

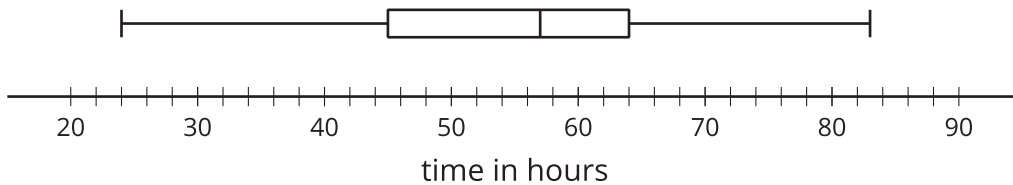
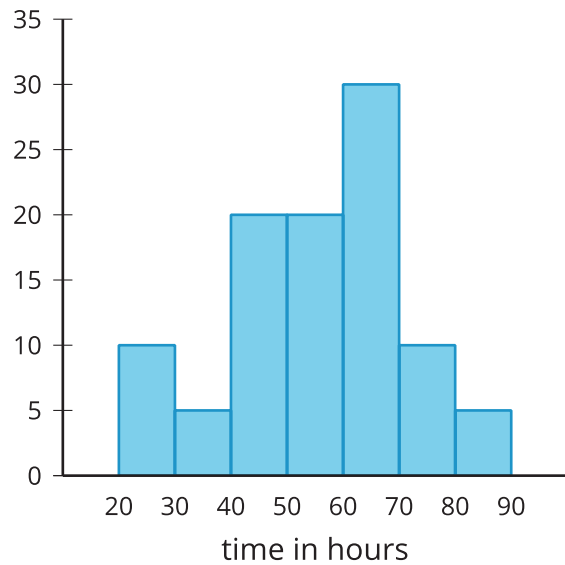
6. Ten students took a history quiz. This list shows how many questions each student answered correctly.

9 6 8 10 5 8 10 7 9 8

a. What is the IQR (interquartile range)?

b. What is the mean?

7. Priya asked some student athletes at her school how many hours each month they spent practicing their sport. Here are a box plot and histogram for the data she collected.



- How many students did Priya ask? Explain how you know.
- Is the mean or the median a more appropriate measure of center for this data set? Explain your reasoning.
- Can Priya use these data displays to find the exact median? Explain how you know.
- Can Priya use these data displays to find the exact mean? Explain how you know.
- What would be an appropriate measure of variability for this data set? Find or estimate its value.

Assessment Answer Keys

Check Your Readiness A and B
Mid-Unit Assessment A and B
End-of-Unit Assessment A and B

Assessment Answer Keys

Assessment : Check Your Readiness (A)

Teacher Instructions

Calculators should not be used. Because of several multi-part problems, this assessment may take longer to administer than a typical pre-unit assessment.

Student Instructions

Do not use a calculator.

Problem 1

The content assessed in this problem is first encountered in Lesson 11: Variability and MAD.

This item is about general number sense and understanding of division. Students will need to make calculations like this when they calculate the mean or MAD of a data set, and will frequently need to efficiently divide by 2 when calculating the median.

If most students struggle with this item, plan to add extra practice in the weeks before Lesson 11. There are practice problems in Lesson 4 that can be used as well as in Unit 5. Those problems and the problems here can also be used as a quick Number Talk.

Statement

Calculate each quotient.

1. $125 \div 5$
2. $800 \div 16$
3. $3.5 \div 2$
4. $21.4 \div 10$

Solution

1. 25
2. 50
3. 1.75
4. 2.14

Aligned Standards

5.NBT.B.6, 5.NBT.B.7

Problem 2

The content assessed in this problem is first encountered in Lesson 11: Variability and MAD.

Students will need to make calculations like the following in order to calculate mean absolute deviation (MAD).

If most students struggle with this item, plan to add extra practice in the weeks before Lesson 11. There are practice problems in Lesson 4 that can be used as well as in Unit 5. Those problems and the problems here can also be used as a quick Number Talk.

Statement

Evaluate each expression.

1. $16 - 2.2$

2. $3 - 0.26$

3. $11.26 + 5.34$

4. $2.06 + 3.94$

5. $51 - 0.07$

Solution

1. 13.8

2. 2.74

3. 16.6

4. 6

5. 50.93

Aligned Standards

5.NBT.B.7

Problem 3

The content assessed in this problem is first encountered in Lesson 3: Representing Data Graphically.

Students interpret a bar graph. Students are not required to judge the exact values assigned to each bar, so the problem is more about general understanding, interpretation, and estimation. If

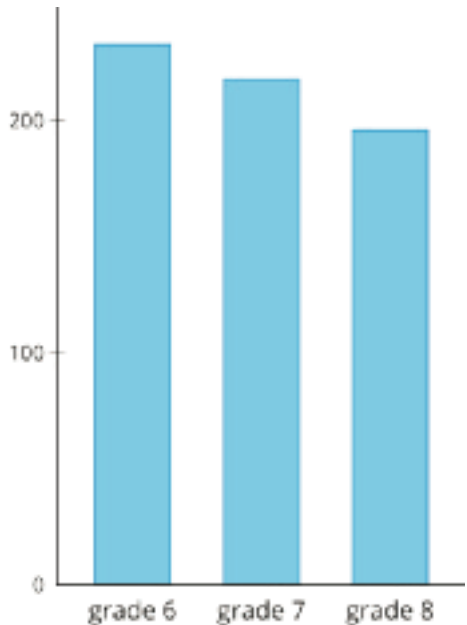
students answer incorrectly, it is possible that they have made arithmetic errors, but it is more likely they are having difficulty interpreting the bar graph.

If most students struggle with this item, plan to do the optional activity in Lesson 3, Favorite Summer Sports. This is an opportunity to think about bar graphs and compare them to dot plots and histograms. Students can also plan their own data collection and bar graph development as needed.

Statement

The bar graph shows how many students are in grades 6, 7, and 8 at a school.

1. About how many total students are in these three grades?
2. Which grade has the most students?
3. About how many more sixth graders are there than eighth graders?



Solution

1. About 650
2. Sixth grade
3. Answers vary. Sample response: between 30 and 40

Aligned Standards

3.MD.B.3

Problem 4

The content assessed in this problem is first encountered in Lesson 2: Statistical Questions.

Students create a line plot showing fractional values. They will need to find a common denominator for the fractions and label the values on the number line correctly.

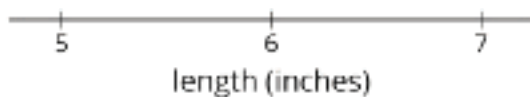
If most students struggle with this item, plan to support their understanding by creating the dot plot with the students in Activity 1, Pencils on a Plot. Amplify the steps as it is created, including partitioning equally and labeling.

Statement

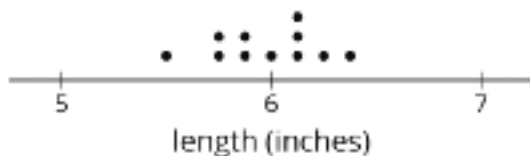
A pet store has 11 lizards, with these lengths in inches:

$$5\frac{1}{2} \quad 5\frac{3}{4} \quad 5\frac{3}{4} \quad 5\frac{7}{8} \quad 5\frac{7}{8} \quad 6 \quad 6\frac{1}{8} \quad 6\frac{1}{8} \quad 6\frac{1}{8}$$
$$6\frac{1}{4} \quad 6\frac{3}{8}$$

Draw a line plot for this data.



Solution



Aligned Standards

5.MD.B.2

Problem 5

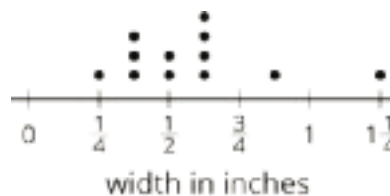
The content assessed in this problem is first encountered in Lesson 2: Statistical Questions.

Students failing to select A may not recall that multiple marks in the same column refer to multiple instances. Students selecting C are probably misinterpreting the meaning of “less than,” since half of the bolts are less than *or equal to* $\frac{1}{2}$ inch wide. Students selecting D probably failed to recognize that $\frac{2}{8}$ is equivalent to $\frac{1}{4}$.

If most students struggle with this item, plan to do Activity 1, Pencils on a Plot, allowing for extra time for students to share their understandings. For extra practice, make a class number line showing zero to 2 and add fractions to it.

Statement

Diego drew this line plot showing the width, in inches, of all the bolts in a storage bin. Select **all** the true statements.



- A. There are 12 bolts in the storage bin.
- B. The widest bolt is $\frac{7}{8}$ inches wide.
- C. Half of the bolts are less than $\frac{1}{2}$ inch wide.
- D. None of the bolts is exactly $\frac{2}{8}$ inches wide.
- E. None of the bolts is exactly $\frac{6}{8}$ inches wide.
- F. The difference between the maximum and minimum width of the bolts is 1 inch.

Solution

["A", "E", "F"]

Aligned Standards

5.MD.B.2, 5.NF.A.1

Problem 6

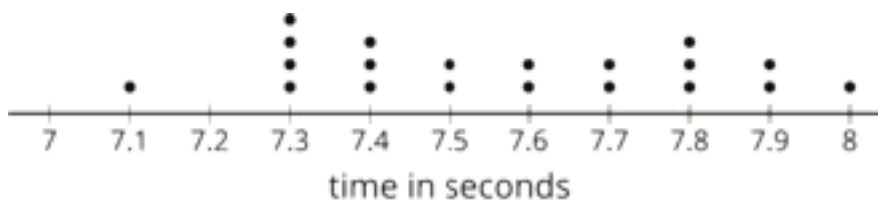
The content assessed in this problem is first encountered in Lesson 5: Using Dot Plots to Answer Statistical Questions.

In addition to reading the line plot, students must also add and subtract decimals in context.

If most students struggle with this item, plan to adapt Lesson 5 by using the data collected in Lesson 1 Activity 2 question 12 to create and interpret a line plot. This data will allow for questions that require adding and subtracting decimals in context.

Statement

This line plot shows the amount of time, in seconds, that it took 20 sixth graders to run a 50-meter dash.



Select **all** the true statements.

- A. The fastest time was 7.0 seconds.
- B. No runner recorded a time of 7.2 seconds.
- C. The fastest 5 students' total time was 36.3 seconds.
- D. Exactly half of the students were faster than 7.7 seconds.
- E. The difference between the fastest and slowest times was 0.9 seconds.

Solution

["B", "C", "E"]

Aligned Standards

5.MD.B, 5.NBT.B.7

Problem 7

The content assessed in this problem is first encountered in Lesson 4: Dot Plots.

Students calculate several benchmark percentages, notably 25%, 50%, and 75%. This skill will be essential when they find quartiles. Whole number values and decimal values are both included, as they require both skills in this unit.

If most students struggle with this item, plan to explicitly share the strategies used by students to determine percentages in Lesson 4 Activity 3. In addition, here are two percentage questions that could be added to the lesson synthesis: What percentage of dogs weighed 15kg or more? What percentage of dogs weighed less than 15kg?

Statement

Calculate each percentage.

1. 25% of 50
2. 25% of 60
3. 50% of 60
4. 75% of 60
5. 75% of 30
6. 100% of 22.5
7. 10% of 22.5
8. 50% of 45.7

Solution

1. 12.5
2. 15
3. 30
4. 45
5. 22.5
6. 22.5
7. 2.25
8. 22.85

Aligned Standards

6.NS.B.3, 6.RP.A.3.c

Assessment : Check Your Readiness (B)

Teacher Instructions

Calculators should not be used. This pre-assessment includes more problems than is typical, so you may want to make decisions about which problems to assign to students.

Student Instructions

Do not use a calculator.

Problem 1

The content assessed in this problem is first encountered in Lesson 11: Variability and MAD.

This item is about general number sense and understanding of division. Students will need to make calculations like this when they calculate the mean or MAD of a data set, and will frequently need to efficiently divide by 2 when calculating the median.

If most students struggle with this item, plan to add extra practice in the weeks before Lesson 11. There are practice problems in Lesson 4 that can be used as well as in Unit 5. Those problems and the problems here can also be used as a quick Number Talk.

Statement

Calculate each quotient.

1. $180 \div 5$

2. $600 \div 15$

3. $7.5 \div 2$

4. $51.6 \div 10$

Solution

1. 36

2. 40

3. 3.75

4. 5.16

Aligned Standards

5.NBT.B.6, 5.NBT.B.7

Problem 2

The content assessed in this problem is first encountered in Lesson 2: Statistical Questions.

Students create a line plot showing fractional values. They will need to find a common denominator for the fractions and label the values on the number line correctly.

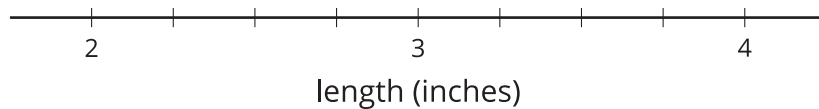
If most students struggle with this item, plan to support their understanding by creating the dot plot with the students in Activity 1, Pencils on a Plot. Amplify the steps as it is created, including partitioning equally and labeling.

Statement

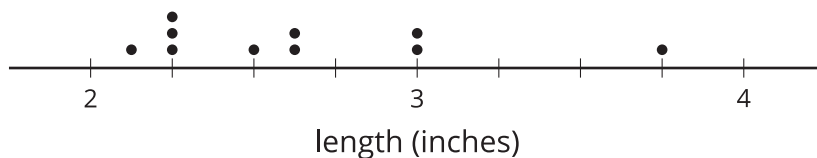
A toy set has 10 mini toy dinosaurs, with these lengths in inches:

length (inches)	$2\frac{1}{4}$	$2\frac{1}{4}$	$2\frac{1}{4}$	$2\frac{1}{8}$	$2\frac{1}{2}$	$2\frac{5}{8}$	$2\frac{5}{8}$	3	3	$3\frac{3}{4}$
-----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	---	---	----------------

Draw a line plot for this data.



Solution



Aligned Standards

5.MD.B.2

Problem 3


The content assessed in this problem is first encountered in Lesson 11: Variability and MAD.

Students will need to make calculations like the following in order to calculate mean absolute deviation (MAD).

If most students struggle with this item, plan to add extra practice in the weeks before Lesson 11. There are practice problems in Lesson 4 that can be used as well as in Unit 5. Those problems and the problems here can also be used as a quick Number Talk.

Statement

Evaluate each expression.



1. $11 - 4.9$

2. $7 - 0.18$

3. $13.23 + 2.67$

4. $6.08 + 2.02$

5. $35 - 0.02$

Solution

1. 6.1

2. 6.82

3. 15.9

4. 8.1

5. 34.98

Aligned Standards

5.NBT.B.7

Problem 4

The content assessed in this problem is first encountered in Lesson 3: Representing Data Graphically.

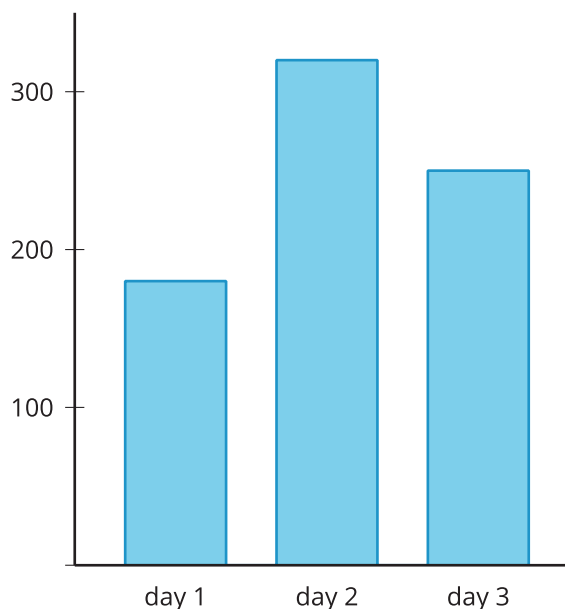
Students interpret a bar graph. Students are not required to judge the exact values assigned to each bar, so the problem is more about general understanding, interpretation, and estimation. If students answer incorrectly, it is possible that they have made arithmetic errors, but it is more likely they are having difficulty interpreting the bar graph.

If most students struggle with this item, plan to do the optional activity in Lesson 3, Favorite Summer Sports. This is an opportunity to think about bar graphs and compare them to dot plots and histograms. Students can also plan their own data collection and bar graph development as needed.

Statement

The bar graph shows how many people donated blood on 3 different days of the blood drive.

1. About how many total people donated blood on these three days?
2. Which day had the least number of people?
3. About how many more people donated blood on day 2 than on day 3?



Solution

1. About 750
2. Day 1
3. Answers vary. Sample response: between 60 and 80.

Aligned Standards

3.MD.B.3

Problem 5

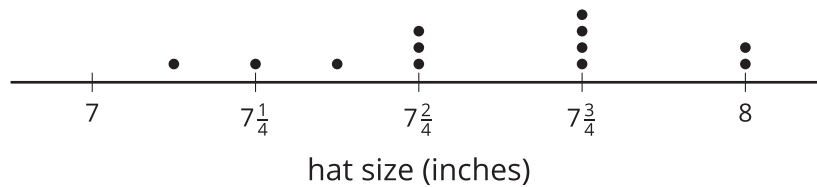
The content assessed in this problem is first encountered in Lesson 2: Statistical Questions.

Students selecting A may not recall that multiple marks in the same column refer to multiple instances. Students selecting C can divide the line into 8ths appropriately and read the graph correctly. Students selecting D did not recognize that $7\frac{6}{8}$ inches is equivalent to $7\frac{3}{4}$.

If most students struggle with this item, plan to do Activity 1, Pencils on a Plot, allowing for extra time for students to share their understandings. For extra practice, make a class number line showing zero to 2 and add fractions to it.

Statement

Jada drew this line plot showing the hat sizes, in inches, of all the players on her softball team.



Select **all** the true statements.

- A. There are six players on the team.
- B. The largest hat size is 8 inches.
- C. Half of the hat sizes are less than $7\frac{5}{8}$ inches.
- D. None of the players has a hat size of $7\frac{6}{8}$ inches.
- E. The difference between the maximum and minimum hat size is $\frac{7}{8}$ inch.

Solution

["B", "C", "E"]

Aligned Standards

5.MD.B.2, 5.NF.A.1

Problem 6

The content assessed in this problem is first encountered in Lesson 5: Using Dot Plots to Answer Statistical Questions.

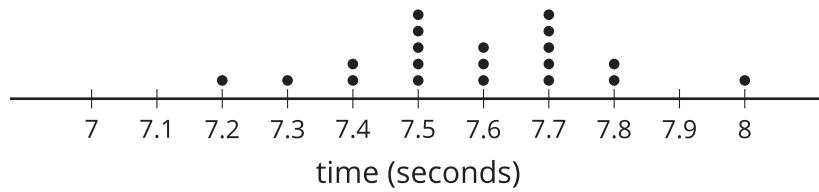
In addition to reading the line plot, students must also add and subtract decimals in context.

Students that fail to select A are using the lowest and highest tick marks to calculate the difference. Students selecting B and D have a misunderstanding of dot plots and their values. Students failing to select C may have made a calculation error.

If most students struggle with this item, plan to adapt Lesson 5 by using the data collected in Lesson 1 Activity 2 question 12 to create and interpret a line plot. This data will allow for questions that require adding and subtracting decimals in context.

Statement

This line plot shows the amount of time, in seconds, that it took 20 sixth graders to complete a set of 10 sit ups in gym class.



Select **all** the true statements.

- A. The difference between the fastest and slowest times was 0.9 seconds.
- B. The fastest time was 7.0 seconds.
- C. The fastest 3 students' total time was 23.6 seconds.
- D. No student recorded a time of 7.1 seconds.
- E. Exactly half of the students were faster than 7.6 seconds.

Solution

["C", "D"]

Aligned Standards

5.MD.B, 5.NBT.B.7

Problem 7

The content assessed in this problem is first encountered in Lesson 4: Dot Plots.


Students calculate several benchmark percentages, notably 25%, 50%, and 75%. This skill will be essential when they find quartiles. Whole number values and decimal values are both included as they require both skills in this unit.

If most students struggle with this item, plan to explicitly share the strategies used by students to determine percentages in Lesson 4 Activity 3. In addition, here are two percentage questions that could be added to the lesson synthesis: What percentage of dogs weighed 15kg or more? What percentage of dogs weighed less than 15kg?

Statement

Calculate each percentage.

1. 25% of 30
2. 25% of 80
3. 50% of 80
4. 75% of 80
5. 75% of 50



6. 100% of 46.5

7. 10% of 46.5

8. 50% of 25.7

Solution

1. 7.5

2. 20

3. 40

4. 60

5. 37.5

6. 46.5

7. 4.65

8. 12.85

Aligned Standards

6.NS.B.3, 6.RP.A.3.c

Assessment : Mid-Unit Assessment (A)

Teacher Instructions

Use of a four-function or scientific calculator is acceptable, but should not provide a significant advantage. Do not allow use of more advanced calculators that may include functions to draw data displays, or to calculate the mean or MAD directly.

This assessment is designed to be administered after lesson 13.

Student Instructions

You may use a four-function or scientific calculator, but not a graphing calculator.

Problem 1

Students selecting A did not recognize this requires only a calculation. Students selecting B did not understand that a question requiring research is not necessarily a statistical question, because there is no variability in the answer. Students failing to select C may not have understood that different pennies have different years. Students failing to select D may think that because the answer is the name of a coin, it is not statistical. Students failing to select E most likely have a deep misunderstanding of the nature of statistical questions.

Statement

Select all the statistical questions.

- A. How many nickels does it take to make a dollar?
- B. In what year was the first penny made in the United States?
- C. Among all the pennies at a bank, what is the most frequent year the pennies were made?
- D. Which coin (penny, nickel, dime, or quarter) is used most frequently in transactions at a bank?
- E. On average, how many pennies do people receive in change when they make a purchase at a store?

Solution

["C", "D", "E"]

Aligned Standards

6.SP.A.1

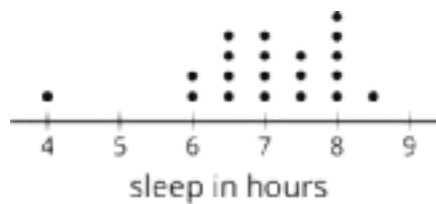
Problem 2

Students analyze a dot plot, including simple questions about spread and distribution. Students are asked to calculate percentages: there are 20 data points to make calculations simpler.

Students failing to select A may have mistaken the group of 5 as one, may have miscounted, or may have a misunderstanding about the difference between “at least” and “more than.” Students selecting B have confused mean with midrange (the average of the largest and smallest values). Students failing to select C may not be accounting for the frequencies when determining information in a dot plot. Students selecting D have a misunderstanding about 1%, compared to the data of one student. Students selecting E likely excluded 4 hours as an outlier.

Statement

This dot plot shows the number of hours 20 sixth grade students slept on a Saturday night.



Select **all** the true statements about the data used to build the dot plot.

- A. Six students slept for at least 8 hours.
- B. The mean amount of sleep was 6.25 hours.
- C. More than half of the students slept 7 hours or less.
- D. Only 1% of the students slept more than 8 hours.
- E. The difference between the most hours of sleep and the least for these students was 2.5 hours.

Solution

["A", "C"]

Aligned Standards

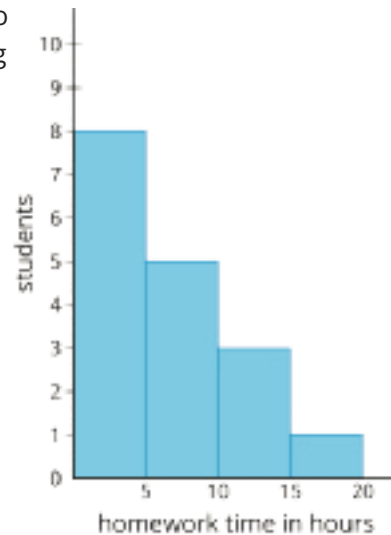
6.RP.A.3.c, 6.SP.B.4

Problem 3

Students selecting A misread the number of bins for the number of observations. Students selecting B do not understand that a histogram’s bin includes its left edge value, in this case 0. Students selecting D are looking at the size of the bin and drawing a conclusion that a histogram is not capable of making.

Statement

Students responded to a survey. The survey asked students to report the amount of time they spent doing homework during a week, to the nearest hour. This histogram displays the data. Which of these statements *must* be true?



- A. A total of 4 students participated in the survey.
- B. Every student spent at least 1 hour doing homework.
- C. Every student spent less than 20 hours doing homework.
- D. More students spent 7 hours doing homework than spent 12 hours.

Solution

C

Aligned Standards

6.SP.B.4, 6.SP.B.5.a

Problem 4

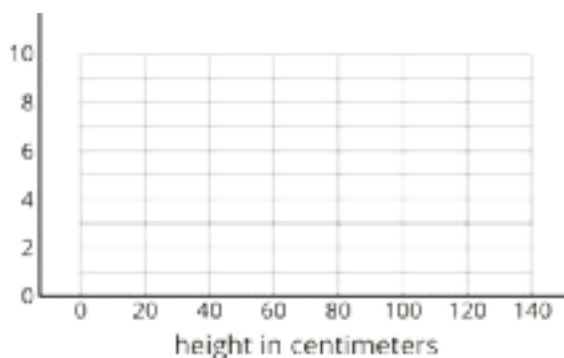
For the second question, accept any answer from 20 to 40 centimeters. The mean of the data is 39.25 centimeters, but students should not need this calculation to answer the question. Watch for students making minor errors in building the histogram. The problem does not specify the intervals students must use, so accept alternate correct histograms; nearly all students will use the intervals provided.

Statement

Here is the height of 20 flowers in the school garden, in centimeters.

5	5	10	10	15	25	25	25	25
30	30	35	45	45	50	55	65	70
105	110							

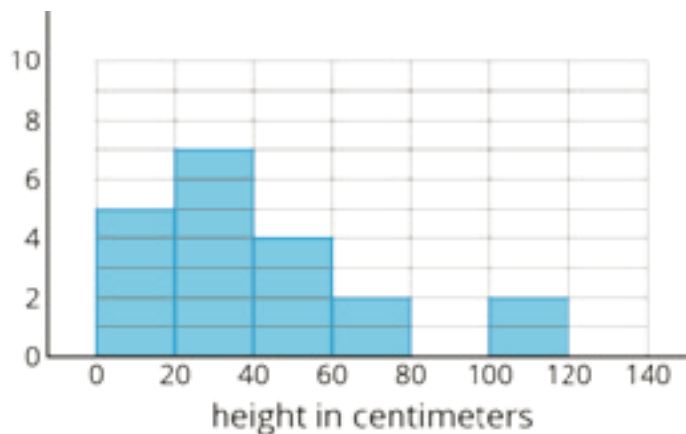
1. Draw a histogram to display the data.



2. Based on the histogram, what is a typical length for these 20 flowers?

Solution

1.



2. Answers vary. Sample response: About 30–40 centimeters. Most of the flowers are close to 30 centimeters, with very few flowers more than 60 centimeters.

Tier 1 response:

- Accurate, correct work.
- All histogram bar heights are correct, and the typical length given is in the 20–40 cm range.

Tier 2 response:

- Work shows general conceptual understanding and mastery, with some errors.
- Sample errors: one or two mistakes in histogram bar heights; correct histogram but a typical length above 40 cm given.

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.

- Sample errors: major mistakes in histogram bar heights; attempt to draw a different type of plot; mistakes in histogram and an incorrect typical length; empty or nonsensical answer to typical length question.

Aligned Standards

6.SP.A.3, 6.SP.B.4

Problem 5

There is more than one way to do this problem. Thinking of the mean as a fair share is the intended method. Students may also recognize that if the mean of four numbers is 25, the sum of the four numbers is 100.

Statement

The mean of four numbers is 25. Three of the four numbers are 20, 26, and 26. What is the fourth number?

Solution

28. (One number is 5 less than the mean, and two numbers are each 1 more than the mean. Since the four numbers must be evenly distributed around 25, the last number must be 3 more than the mean.)

Aligned Standards

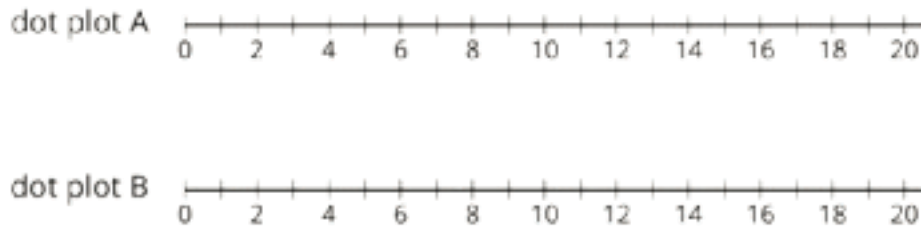
6.SP.B.5.c

Problem 6

Some students may have trouble here constructing the dot plots, because the problem is quite open-ended. The guideline of 7 or fewer data points is intended only to limit the amount of time students spend on this problem. Also, note that students do not have to make the means equal, just approximately the same.

Statement

1. Draw two dot plots, each with 7 or fewer data points, so that:
 - both dot plots display data with approximately the same mean
 - the data displayed in Dot Plot A has a much larger MAD (mean absolute deviation) than the data displayed in Dot Plot B.



2. How can you tell, visually, that one dot plot displays data with a larger MAD than another?

Solution

- Answers vary. The dot plots should show roughly the same center, with Dot Plot B showing a much tighter clustering than Dot Plot A.
- Answers vary. Sample response: since MAD is a measure of spread, the data in a dot plot with a larger MAD will have the wider spread.

Tier 1 response:

- Accurate, correct work.
- Dot plots show roughly same center; Dot Plot A has visibly larger MAD; appropriate description of spread given. Acceptable errors: dot plots include more than 7 points.

Tier 2 response:

- Work shows general conceptual understanding and mastery, with some errors.
- Sample errors: dot plots show significantly different center but correct MAD; dot plots show correct center but similar MAD; dot plots are correct but backwards; no reasonable description given.

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.
- Sample errors: two or more error types from Tier 2 response; failure to draw two dot plots.

Aligned Standards

6.SP.B.4, 6.SP.B.5.c

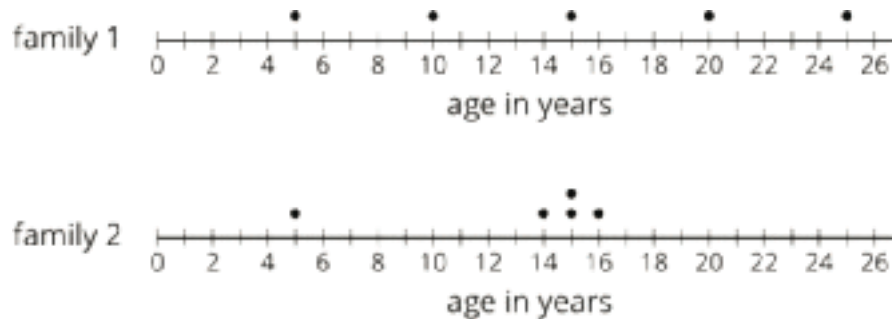
Problem 7

Students analyze the MAD of two data sets. One data set has an unusually low value, while the other is evenly spread. The first set has a larger MAD because “on average” it is more spread out.

Some students may decide 5 and 25 are unusual in Family 1. Because formal definitions for outliers have not yet been learned, this is an acceptable response.

Statement

Here are two dot plots, the ages of the five children in each of two families.



1. For each family, identify anyone whose age is unusual compared to the rest of the family.
2. Calculate the MAD (mean absolute deviation) of each data set. Which family has the wider spread of ages?

Solution

1. Family 1: no unusual values, or 5 and 25 (both acceptable). Family 2: one unusual value, 5.
2. Family 1 has the larger spread. Family 1: 6 years, Family 2: 3.2 years.

Tier 1 response:

- Accurate, correct work.
- Correct list of unusual data, correct calculations of MAD, correct selection of Family 1 as having larger spread.

Tier 2 response:

- Work shows good conceptual understanding and mastery, with minor errors.
- Sample errors: a calculation error causes one mean or MAD to be incorrect; incorrect list of unusual data; failure to correctly compare MADs with otherwise accurate work. Acceptable errors: an error in calculating causes an incorrect conclusion about which family has the larger spread.

Tier 3 response:

- Work shows a developing but incomplete conceptual understanding, with significant errors.
- Sample errors: two or more error types from Tier 2 response; multiple calculation errors; an incorrect MAD with no work shown; minor errors in chosen method of determining mean or MAD, including failure to use absolute value of deviations.

Tier 4 response:

- Work includes major errors or omissions that demonstrate a lack of conceptual understanding and mastery.
- Sample errors: badly incorrect algorithm for calculating mean or MAD; no calculation of MAD.

Aligned Standards

6.SP.A.3, 6.SP.B.5.c

Assessment : Mid-Unit Assessment (B)

Teacher Instructions

Use of a four-function or scientific calculator is acceptable, but should not provide a significant advantage. Do not allow use of more advanced calculators that may include functions to draw data displays, or to calculate the mean or MAD directly.

This assessment is designed to be administered after lesson 13.

Student Instructions

You may use a four-function or scientific calculator, but not a graphing calculator.

Problem 1

Statement

Select **all** the statistical questions.

- A. How many cups are in a gallon?
- B. Which measuring spoon (tablespoon, teaspoon, half-teaspoon, or quarter-teaspoon) do home bakers use most frequently?
- C. On average, how much salt do professional pastry chefs use in a pie crust?
- D. Which is larger, 17 ounces or 2 cups?
- E. Did a specific store sell more measuring cups or measuring spoons on a specific day?

Solution

["B", "C"]

Aligned Standards

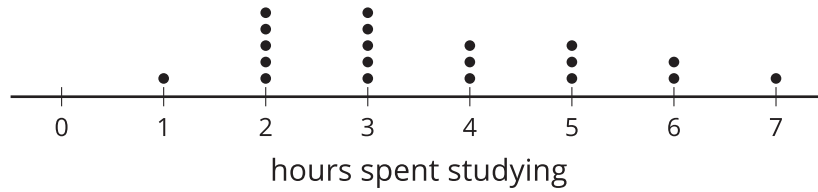
6.SP.A.1

Problem 2

Students analyze a dot plot, including simple questions about spread and distribution. Students are asked to calculate percentages: there are 20 data points to make calculations simpler. Students selecting B were visually looking at the dot plot giving an approximate value for the mean instead of calculating the mean. Students failing to select A may not have noticed 0 did not have a tick mark and took that as the lowest value. Students failing to select C may have misinterpreted the term “at least”. Students selecting E have a misunderstanding about 1% versus the data of one student.

Statement

The dot plot shows the number of hours 20 sixth grade students spent studying in a week.



Select **all** true statements about the data used to build the dot plot.

- A. The difference between the most hours and the least hours spent studying was 6 hours.
- B. The mean amount of study time was 5 hours.
- C. Nine students studied for at least 4 hours.
- D. Most students studied less than 4 hours.
- E. Only 1% of the students studied more than 6 hours.

Solution

["A", "C", "D"]

Aligned Standards

6.RP.A.3.c, 6.SP.B.4

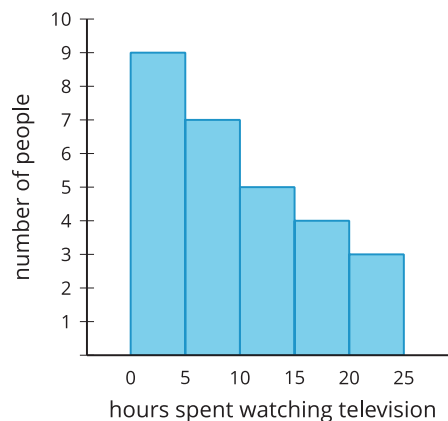
Problem 3

Students selecting A may have confused the meaning of the two axes. Students selecting B do not understand that a histogram's bin includes its left edge value, in this case 0. Students selecting C are looking at the size of the bin and drawing a conclusion that a histogram is not capable of making.

Statement

A survey asked people how many hours they spend watching television during a week, to the nearest hour. The histogram displays the data.

Which of these statements *must* be true?



- A. Every person in the survey spent less than 10 hours watching television.
- B. Every person in the survey watched some television during the week.
- C. There are more people in the survey who watched 12 hours of television than people who watched 17 hours.
- D. More than 20 people participated in the survey.

Solution

D

Aligned Standards

6.SP.B.4, 6.SP.B.5.a

Problem 4

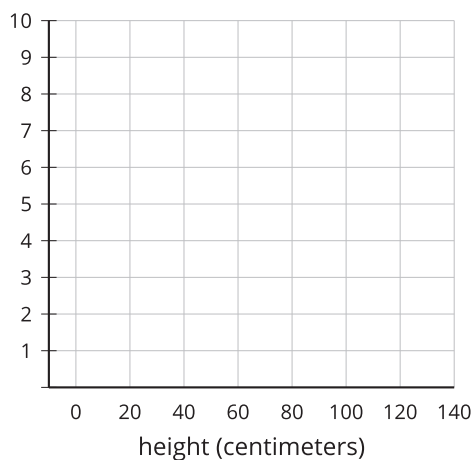
For the second question, accept any answer from 30 to 55 centimeters. The mean of the data is 42.25 centimeters, but students should not need this calculation to answer the question. The problem does not specify the intervals students must use, so accept alternate correct histograms; students are likely to use the intervals provided.

Statement

Here is the height of 20 flowers in the school garden, in centimeters.

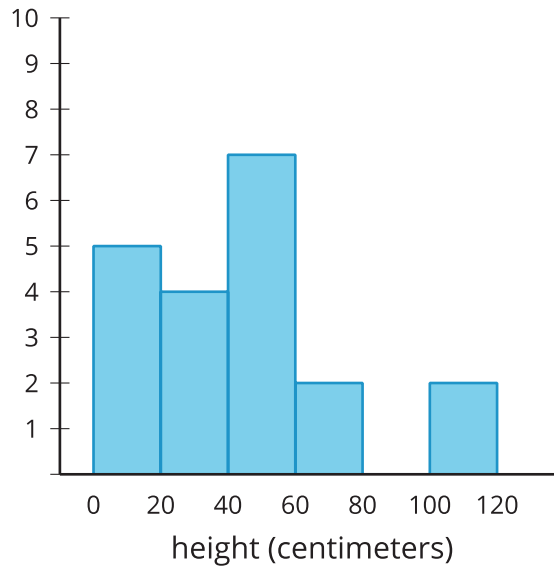
5	5	10	10	15	25	25	30	35	45	45	45	45	50	50	55	65	70	105	110
---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

- a. Draw a histogram to display the data.



- b. Based on the histogram, what is a typical height for these 20 flowers?

Solution



Answers vary. Sample response: About 40 centimeters. Most of the flowers are 40–50 centimeters, with very few flowers more than 60 centimeters.

Tier 1 response:

- Accurate, correct work.
- All histogram bar heights are correct, and the typical length given is in the 30–55 cm range.

Tier 2 response:

- Work shows general conceptual understanding and mastery, with some errors.
- Sample errors: one or two mistakes in histogram bar heights; correct histogram but a typical length above 55 cm given.

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.
- Sample errors: major mistakes in histogram bar heights; attempt to draw a different type of plot; mistakes in histogram and an incorrect typical length; empty or nonsensical answer to typical length question.

Aligned Standards

6.SP.A.3, 6.SP.B.4

Problem 5

There is more than one way to do this problem. Thinking of the mean as a fair share is the intended method. (One number is 5 less than the mean, and two numbers are each 1 more than the mean. Since the four numbers must be evenly distributed around 40, the last number must be 3 more

than the mean.) Students may also recognize that if the mean of four numbers is 40, the sum of the four numbers is 160.

Statement

The mean of four numbers is 40. Three of the four numbers are 35, 41, and 41. What is the fourth number?

Solution

43

Aligned Standards

6.SP.B.5.c

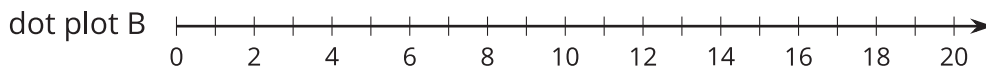
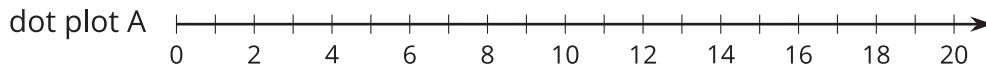
Problem 6

Some students may have trouble here constructing the dot plots, because the problem is quite open-ended. The guideline of 7 or fewer data points is intended only to limit the amount of time students spend on this problem.

Statement

Draw two dot plots, each with 7 or fewer data points, so that:

- both dot plots display data with the same median
- the data displayed in Dot Plot B has a much larger IQR (interquartile range) than the data displayed in Dot Plot A.



How can you tell, visually, that one dot plot displays data with a larger IQR than another?

Solution

1. Answers vary. The dot plots should show the same median, with Dot Plot A showing a much tighter clustering than Dot Plot B.
2. Answers vary. Sample response: Since IQR is a measure of spread, the data in a dot plot with a larger IQR will have the wider spread.

Tier 1 response:

- Accurate, correct work.
- Dot plots show the same median; Dot Plot B has visibly larger IQR; appropriate description of spread given. Acceptable errors: dot plots include more than 7 points.

Tier 2 response:

- Work shows general conceptual understanding and mastery, with some errors.
- Sample errors: dot plots show significantly different center but correct IQR; dot plots show correct center but similar IQR; dot plots are correct but backwards; no reasonable description given.

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.
- Sample errors: two or more error types from Tier 2 response; failure to draw two dot plots.

Aligned Standards

6.SP.B.4, 6.SP.B.5.c

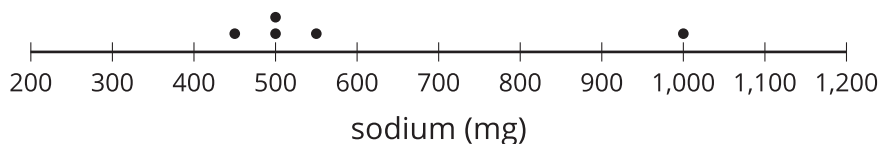
Problem 7

Students analyze the MAD of two data sets. One data set has an unusually high value, while the other is evenly spread. The first set has a lesser MAD because on average it is less spread out.

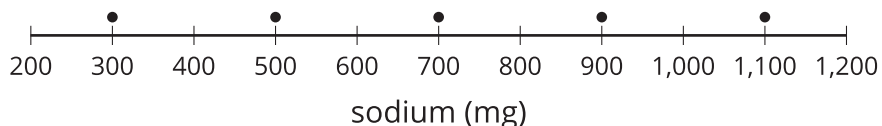
Statement

Here are two dot plots showing the milligrams of sodium in the 5 most popular menu items at two different fast food restaurants.

Restaurant 1



Restaurant 2



1. For each restaurant, identify any menu item whose sodium is unusual compared to the other items from the restaurant.
2. Calculate the MAD (mean absolute deviation) of each data set. Which restaurant has menu items with a wider spread of sodium content?

Solution

1. Restaurant 1: one unusual value of 1,000. Restaurant 2: no unusual values.
2. Restaurant 1 has a mean absolute deviation of 160, Restaurant 2 has a mean absolute deviation of 240. Restaurant 2 has the larger spread.

Tier 1 response:

- Accurate, correct work.
- Correct list of unusual data, correct calculations of MAD, correct selection of Restaurant 2 as having the larger spread.

Tier 2 response:

- Work shows good conceptual understanding and mastery, with minor errors.
- Sample errors: a calculation error causes one mean or MAD to be incorrect; incorrect list of unusual data; failure to correctly compare MADs with otherwise accurate work. Acceptable errors: an error in calculating causes an incorrect conclusion about which restaurant has the larger spread.

Tier 3 response:

- Work shows a developing but incomplete conceptual understanding, with significant errors.
- Sample errors: two or more error types from Tier 2 response; multiple calculation errors; an incorrect MAD with no work shown; minor errors in chosen method of determining mean or MAD, including failure to use absolute value of deviations.

Tier 4 response:

- Work includes major errors or omissions that demonstrate a lack of conceptual understanding and mastery.
- Sample errors: badly incorrect algorithm for calculating mean or MAD; no calculation of MAD.

Aligned Standards

6.SP.A.3, 6.SP.B.5.c

Assessment : End-of-Unit Assessment (A)

Teacher Instructions

Use of a four-function or scientific calculator is acceptable, but should not provide a significant advantage. Do not allow use of more advanced calculators that may include functions to draw data displays, or to calculate the mean, median, MAD, or IQR directly.

Student Instructions

You may use a four-function or scientific calculator, but not a graphing calculator.

Problem 1

A vital fact to know here is that only the dot plot can be used to reconstruct the entire data set. Given a histogram or box plot, only certain information is readily known.

Students selecting A and not B may be confused about median versus mean as it applies to the construction of a box plot. Students selecting C may think the corresponding dot plot only needs to have the five number summary, instead of the full original data set. Students failing to select D may not be thinking about how a box plot is constructed; even though the dot plot does not directly include the information in a box plot, it can still be constructed. Students selecting E may have a significant misconception about what a histogram represents, notably that it indicates only ranges of information, not specific values.

Statement

Select **all** the true statements.

- A. Given a box plot, it is always possible to calculate the mean of the data.
- B. Given a box plot, it is always possible to calculate the median of the data.
- C. Given a box plot, it is always possible to construct a corresponding dot plot.
- D. Given a dot plot, it is always possible to construct a corresponding box plot.
- E. Given a histogram, it is always possible to construct a corresponding box plot.

Solution

["B", "D"]

Aligned Standards

6.SP.B.4, 6.SP.B.5

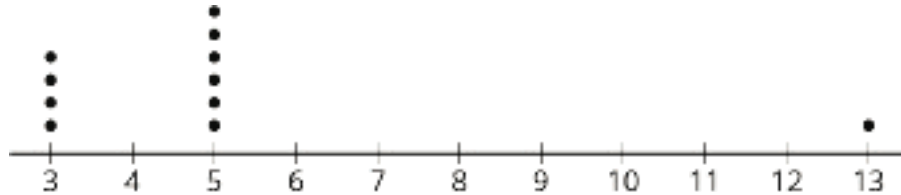
Problem 2

The “fair share” interpretation of mean is very useful here. There are four data points each that are 2 less than 5, and one data point 8 more than 5.

Students selecting A may think that because there are more values less than 5, the mean must be less than 5. Students selecting C may think the mean is larger because 13 is significantly larger than 5. Students selecting D may think the dot plot does not contain full information about the data, but it does.

Statement

Here's a dot plot of a data set.



Which statement is true about the mean of the data set?

- A. The mean is less than 5.
- B. The mean is equal to 5.
- C. The mean is greater than 5.
- D. There is not enough information to determine the mean.

Solution

B

Aligned Standards

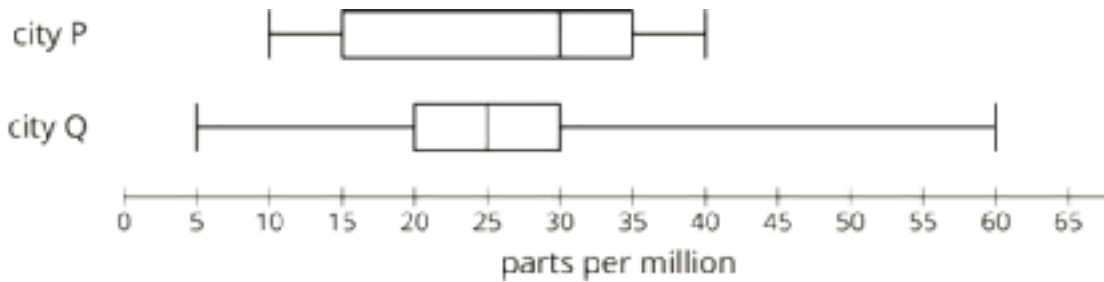
6.SP.B.5.c

Problem 3

Students failing to select A are looking at the box instead of the whiskers. Students failing to select B may have misinterpreted “50 parts per million” as a division, ignoring the label of the box plot. Students selecting C may be attempting to judge the mean by the overall location of the box plots, but no such judgment can be made. Students failing to select D are looking at the IQR instead of the range. Students selecting C instead of E may think box plots indicate means instead of medians.

Statement

The air quality was tested in many office buildings in two cities. The results of the testing are shown in these box plots.



A level of less than 50 parts per million is considered healthy. A level of 50 or more parts per million is considered unhealthy.

Select **all** the statements that must be true.

- A. The lowest recorded measurement was in city Q.
- B. All buildings tested in city P are in the healthy range.
- C. The mean for city P is greater than the mean for city Q.
- D. The range for city Q is greater than the range for city P.
- E. The median for city P is greater than the median for city Q.

Solution

["A", "B", "D", "E"]

Aligned Standards

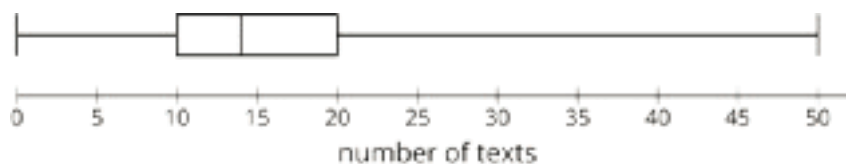
6.SP.B.4, 6.SP.B.5.c

Problem 4

Students use a box plot to make conclusions about a data set, including the median, IQR, and the shape of the distribution. Students may struggle to identify the shape of the distribution without the actual data, but the box plot gives enough information to answer the last question.

Statement

This box plot displays information about the number of text messages some students sent one day.



1. What is the median number of texts sent by students?
2. What is the IQR (interquartile range)?
3. Is this data set symmetric? Explain how you know.

Solution

1. 14 text messages (also accept 13 and 13.5).
2. 10 text messages (the IQR is $20 - 10$, which is 10).
3. No. The top quartile (or top whisker) is much wider than the bottom quartile.

Tier 1 response:

- Accurate, correct work.
- Correct answers to all three questions, including a correct explanation of why the data set is not symmetric. Acceptable errors: claim that the data set is not symmetric because the right side of the box is wider than the left side, without reference to whiskers.

Tier 2 response:

- Work shows general conceptual understanding and mastery, with some errors.
- Sample errors: incorrect median; incorrect IQR; incorrect answer or explanation on data symmetry question, including a general statement that the box plot is not symmetric (not specific enough).

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.
- Sample errors: two or more error types from Tier 2 response.

Aligned Standards

6.SP.A.2, 6.SP.B.5.c

Problem 5

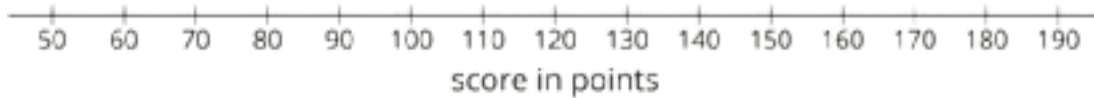
Because box plots are being constructed, students are very likely to use the IQR or range as measures of variability, but Group A also has the higher MAD if students somehow decide to compute it.

Statement

Two groups went bowling. Here are the scores from each group.

Group A	80	100	190	110	70	90	130
Group B	50	110	100	120	107	140	150

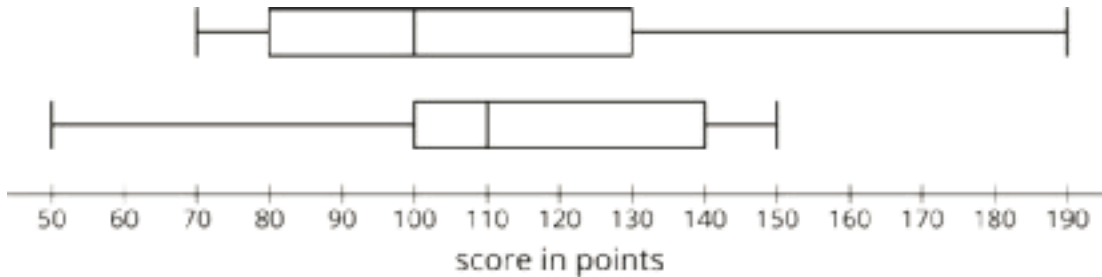
1. Draw two box plots, one for the data in each group.



2. Which group shows greater variability?

Solution

1.



2. Group A shows greater variability. It has a wider range (120 to Group B's 100), and a wider IQR (50 to Group B's 40).

Tier 1 response:

- Accurate, correct work.
- Both box plots drawn correctly, correctly stating that Group A shows greater variability.

Tier 2 response:

- Work shows general conceptual understanding and mastery, with some errors.
- Sample errors: 1 or 2 *types* of minor errors in creating box plots (incorrect placement of median, quartiles, max or min, badly drawn box); incorrectly stating Group B shows greater variability or omitting question.

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.
- Sample errors: more than 2 *types* of minor errors in creating box plots; major errors in creating box plots, such as not using 5 numbers to generate box plot; creating only one box plot.

Aligned Standards

6.SP.B.4, 6.SP.B.5

Problem 6

Watch for students attempting to answer the question without first sorting the data (those students will give a median of 6.5). Also watch for students excluding the center 6 values from the quartile calculation, which leads to an incorrect (larger) IQR.

Statement

Ten students each attempted 10 free throws. This list shows how many free throws each student made.

8 5 6 6 4 9 7 6 5 9

1. What is the median number of free throws made?
2. What is the IQR (interquartile range)?

Solution

1. 6 free throws. (The ordered list is 4, 5, 5, 6, 6, 6, 7, 8, 9, 9. The two middle terms in the ordered list are both 6.)
2. 3 free throws. (The first half of the data is 4, 5, 5, 6, 6; its median is 5. The second half of the data is 6, 7, 8, 9, 9; its median is 8. The IQR is 3, since $8 - 5 = 3$.)

Aligned Standards

6.SP.A.3, 6.SP.B.5.c

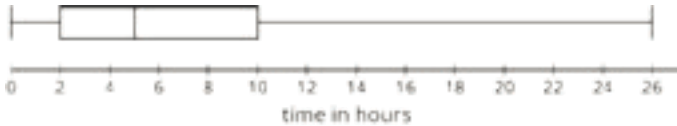
Problem 7

This question is about the limitations of the histogram and box plot, which provide only partial information about a distribution. Notably, one display may be more useful than another depending on the question asked about the data.

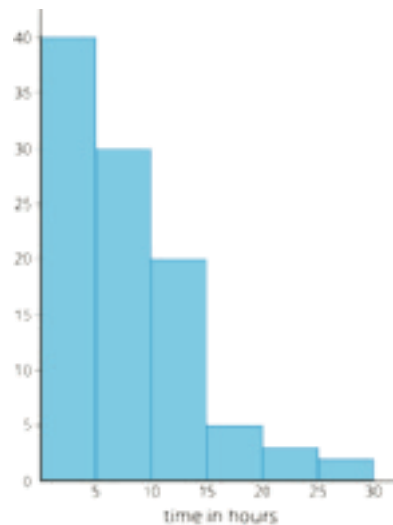
Statement

Jada asked some students at her school how many hours they spent watching television last week, to the nearest hour. Here are a box plot and a histogram for the data she collected.

Box plot:



Histogram:



1. About how many students did Jada ask? Explain how you know.
2. Is the mean or the median a more appropriate measure of center for this data set? Explain your reasoning.
3. Can Jada use these data displays to find the exact median? Explain how you know.
4. Can Jada use these data displays to find the exact mean? Explain how you know.
5. What would be an appropriate measure of variability for this data set? Find or estimate its value.

Solution

1. Jada asked about 100 students, found by adding the heights of the histogram bars.
2. The median is more appropriate because the data is not symmetric.
3. Yes, the box plot gives the exact median, 5 hours.
4. No, the histogram and box plot only give ranges where data falls, not the actual data. The mean can be estimated, but cannot be found precisely.
5. Answers vary. Sample response: The IQR (interquartile range) is appropriate because the median is being used as a measure of center. The box plot gives the IQR of 8 hours because $10 - 2 = 8$.

Tier 1 response:

- Accurate, correct work.

- Correct answer to each question, description of why IQR is appropriate measure of spread, correct IQR. Acceptable errors: mistake in determining median or IQR caused by a misreading of the box plot.

Tier 2 response:

- Work shows good conceptual understanding and mastery, with minor errors.
- Sample errors: incorrect response for histogram total, larger than 6; stating that the data is symmetric; attempt to calculate precise mean; incorrect or missing IQR calculation. Acceptable errors: incorrect MAD estimation, given (incorrect) statement that data is symmetric.

Tier 3 response:

- Work shows a developing but incomplete conceptual understanding, with significant errors.
- Sample errors: two or more error types from Tier 2 response; incorrect response for histogram total, 6 or fewer; incorrect median; invalid use of box plot to determine mean.

Tier 4 response:

- Work includes major errors or omissions that demonstrate a lack of conceptual understanding and mastery.
- Sample errors: three or more error types from Tier 2 response; two or more error types from Tier 3 response; multiple omitted parts.

Aligned Standards

6.SP.B.4, 6.SP.B.5.a, 6.SP.B.5.c, 6.SP.B.5.d

Assessment : End-of-Unit Assessment (B)

Teacher Instructions

Use of a four-function or scientific calculator is acceptable, but should not provide a significant advantage. Do not allow use of more advanced calculators that may include functions to draw data displays, or to calculate the mean, median, MAD, or IQR directly.

Student Instructions

You may use a four-function or scientific calculator, but not a graphing calculator.

Problem 1

An important fact is that only a dot plot can be used to reconstruct the entire data set. Given a histogram or box plot, only certain information is readily known. Students selecting D and E may have a significant misconception about what a histogram represents, notably, that it indicates only ranges of information, not specific values.

Statement

Select **all** the true statements.

- A. Given a dot plot, it is always possible to construct a corresponding histogram.
- B. Given a dot plot, it is always possible to calculate the mean of the data.
- C. Given a box plot, it is always possible to calculate the IQR of the data.
- D. Given a histogram, it is always possible to calculate the mean of the data.
- E. Given a histogram, it is always possible to calculate the median of the data.

Solution

["A", "B", "C"]

Aligned Standards

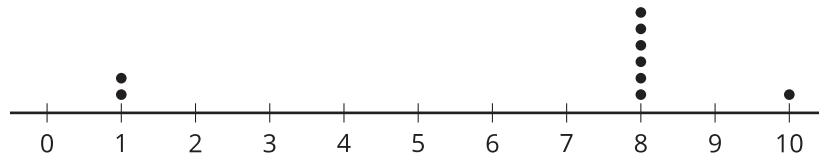
6.SP.B.4, 6.SP.B.5

Problem 2

The “fair share” interpretation of mean is very useful here. There are three data points; two that are 7 less than 8, and one data point 2 more than 8. Students selecting B may think that because there are more values of 8 than the values 1 or 10. Students selecting C may think the mean is larger because 10 is significantly larger than 1. Students selecting D may think the dot plot does not contain full information about the data, but it does.

Statement

Here is a dot plot of a data set.



Which statement is true about the mean of the data set?

- A. The mean is less than 8.
- B. The mean is equal to 8.
- C. The mean is greater than 8.
- D. There is not enough information to determine the mean.

Solution

A

Aligned Standards

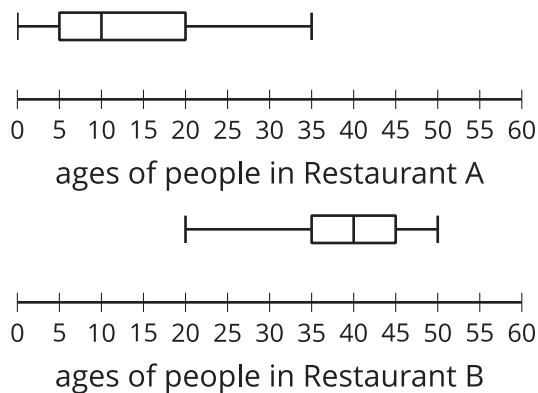
6.SP.B.5.c

Problem 3

Students selecting B are attempting to judge the mean absolute deviation by visually assessing the box plots, but no such judgement can be made. Students selecting D may have calculated the IQR incorrectly or made an incorrect visual assessment of the IQR. Students selecting E did not realize that the oldest person dining in Restaurant A is older than the youngest person dining in Restaurant B.

Statement

The ages of people dining in two restaurants are shown in the following box plots.



Select **all** the statements that must be true.

- A. The median age of people dining in Restaurant B is greater than the median age of people dining in Restaurant A.
- B. The MAD (mean absolute deviation) for Restaurant B is greater than the MAD (mean absolute deviation) for Restaurant A.
- C. The youngest person was dining in Restaurant A.
- D. The IQR (interquartile range) for Restaurant A is equal to the IQR (interquartile range) for Restaurant B.
- E. Every person dining in Restaurant B is older than everyone dining in Restaurant A.

Solution

["A", "C"]

Aligned Standards

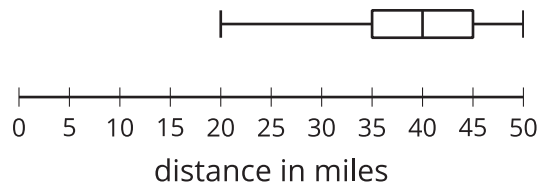
6.SP.B.4, 6.SP.B.5.c

Problem 4

Students use a box plot to make conclusions about a data set, including the median, IQR, and the shape of the distribution.

Statement

This box plot displays information about the distance in miles teachers drive to school each day.



1. What is the IQR (interquartile range)?
2. What is the median distance in miles driven by teachers?
3. Is this data set symmetric? Explain how you know.

Solution

1. 10, because $45 - 35 = 10$.
2. 40
3. No. If a vertical line is drawn at the median, the left and right sides are not mirror images.

Tier 1 response:

- Accurate, correct work.

- Correct answers to all three questions, including a correct explanation for why the data set is not symmetric. Acceptable errors: arithmetic error computing the IQR using correct values for quartiles 1 and 3; different wording for the last question that conveys the same idea.

Tier 2 response:

- Work shows general conceptual understanding and mastery, with some errors.
- Sample errors: incorrect median; incorrect IQR; incorrect answer or explanation on data symmetry question, including a general statement that the box plot is not symmetric with no explanation.

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.
- Sample errors: two or more error types from Tier 2 response.

Aligned Standards

6.SP.A.2, 6.SP.B.5.c

Problem 5

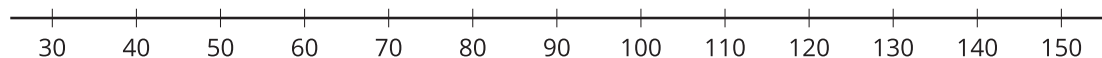
Because box plots are being constructed, students are very likely to use the IQR or range as measures of variability, but Group A also has the higher MAD if students decide to compute it.

Statement

Two groups of students had a car wash fundraiser every day for a week in the summer. Here are the funds raised from each group.

Group P	80	60	65	80	70	85	70
Group Q	60	140	80	40	50	90	50

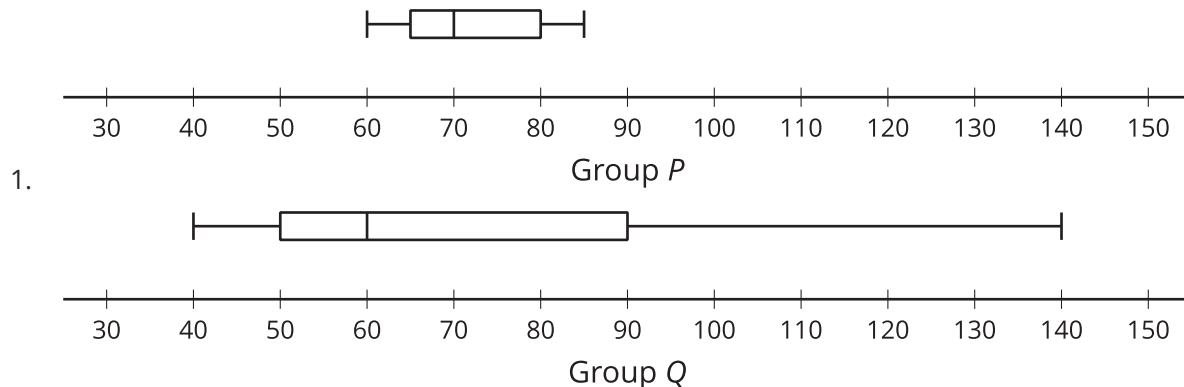
1. Draw two box plots, one for the data in each group.



2. Which group shows greater variability? Explain your reasoning.

Solution

Solution



2. Group Q shows greater variability. It has a wider range (100 to Group Q's 25), and a wider IQR (40, compared to Group P's 15).

Tier 1 response:

- Accurate, correct work.
- Both box plots drawn correctly, correctly stating that Group Q shows greater variability and using the range or IQR as evidence.

Tier 2 response:

- Work shows general conceptual understanding and mastery, with some errors.
- Sample errors: 1 or 2 types of minor errors in creating box plots (incorrect placement of median, quartiles, max or min, badly drawn box); incorrectly stating Group P shows greater variability or omitting the second question.

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.
- Sample errors: more than 2 types of minor errors in creating box plots; major errors in creating box plots, such as not using 7 numbers to generate box plot; creating only one box plot.

Aligned Standards

6.SP.B.4, 6.SP.B.5

Problem 6

Watch for students attempting to answer the question without first sorting the data. Also watch for students excluding the center values from the quartile calculation, which leads to an incorrect (larger) IQR.

Statement

Ten students took a history quiz. This list shows how many questions each student answered correctly.

9 6 8 10 5 8 10 7 9 8

1. What is the IQR (interquartile range)?
2. What is the mean?

Solution

1. 2 questions. (The ordered list is 5, 6, 7, 8, 8, 8, 9, 9, 10, 10. The first half of the data is 5, 6, 7, 8, 8; its median is 7; The second half of the data is 8, 9, 9, 10, 10; its median is 9. The IQR is 2, since $9 - 7 = 2$.)
2. 8 questions. ($9 + 6 + 8 + 10 + 5 + 8 + 7 + 10 + 9 + 8 = 80$ and $80 \div 10 = 8$.)

Aligned Standards

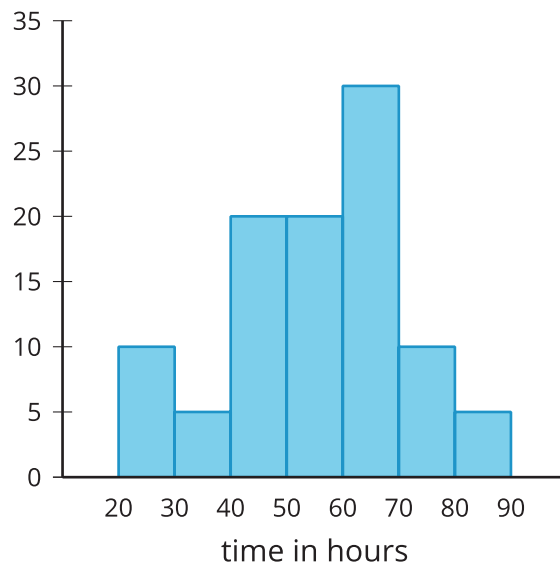
6.SP.A.3, 6.SP.B.5.c

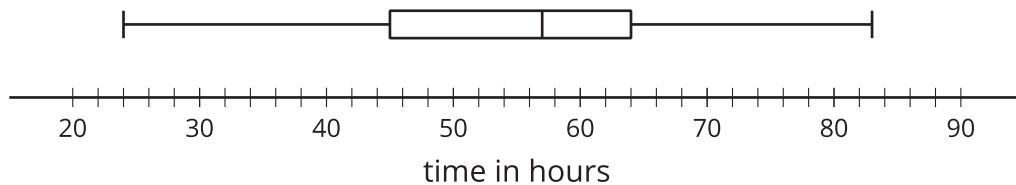
Problem 7

This question is about the limitations of the histogram and box plot, which provide only partial information about a distribution. Notably, one display may be more useful than another, depending on the question asked about the data.

Statement

Priya asked some student athletes at her school how many hours each month they spent practicing their sport. Here are a box plot and histogram for the data she collected.





1. How many students did Priya ask? Explain how you know.
2. Is the mean or the median a more appropriate measure of center for this data set? Explain your reasoning.
3. Can Priya use these data displays to find the exact median? Explain how you know.
4. Can Priya use these data displays to find the exact mean? Explain how you know.
5. What would be an appropriate measure of variability for this data set? Find or estimate its value.

Solution

Solution

1. Priya asked 100 students, found by adding the heights of the histogram bars.
2. The median is more appropriate because the data is not symmetric.
3. Yes, the box plot gives the median, which is 55 hours.
4. No, the histogram and box plot only give ranges where data falls, not the actual data. The mean can be estimated, but cannot be found precisely.
5. Answers vary. Sample response: The IQR (interquartile range) is appropriate, because the median is being used as a measure of center. The box plot gives the IQR of 20 hours because $60 - 40 = 20$.

Tier 1 response:

- Accurate, correct work.
- Correct answer to each question, description of why IQR is an appropriate measure of spread, correct IQR. Acceptable errors: mistake in determining median or IQR caused by a misreading of the box plot.

Tier 2 response:

- Work shows good conceptual understanding and mastery, with minor errors.
- Sample errors: incorrect response for total number of students; stating that the data is symmetric; attempt to calculate precise mean; incorrect or missing IQR calculation. Acceptable errors: incorrect MAD estimation, given (incorrect) statement that data is symmetric.

Tier 3 response:

- Work shows a developing but incomplete conceptual understanding, with significant errors.
- Sample errors: two or more error types from Tier 2 response; incorrect response for histogram total, 6 or fewer; incorrect median; invalid use of box plot to determine mean.

Tier 4 response:

- Work includes major errors or omissions that demonstrate a lack of conceptual understanding and mastery.
- Sample errors: three or more error types from Tier 2 response; two or more error types from Tier 3 response; multiple omitted parts.

Aligned Standards

6.SP.B.4, 6.SP.B.5.a, 6.SP.B.5.c, 6.SP.B.5.d

CKMath™
Core Knowledge **MATHEMATICS™**



Lesson
Cool Downs

Lesson 1: Got Data?

Cool Down: What's the Question?

1. Would each survey question produce categorical data or numerical data?

- a. What is your favorite vegetable?
- b. Have you been to the capital city of your state?
- c. How old is the youngest person in your family?
- d. In which zip code do you live?
- e. What is the first letter of your name?
- f. How many hours do you spend outdoors each day?

2. Andre collected data measured in centimeters.

8.5 10.5 7.8 9.5 8.1 9.0 10.2 9.6 11.2 10.9 12.7 9.8

What could he be investigating? Select **all** that apply.

- a. The weight of a dozen eggs.
- b. The length of leaves from a tree.
- c. The height of cups and mugs in a cupboard.
- d. The length of songs on a music CD.
- e. The length of colored pencils in a box.

Lesson 2: Statistical Questions

Cool Down: Questions about Temperature

Here are two questions:

Question A: Over the past 10 years, what is the warmest temperature recorded, in degrees Fahrenheit, for the month of December in Miami, Florida?

Question B: At what temperature does water freeze in Miami, Florida?

1. Decide if each question is statistical or non-statistical. Explain your reasoning.

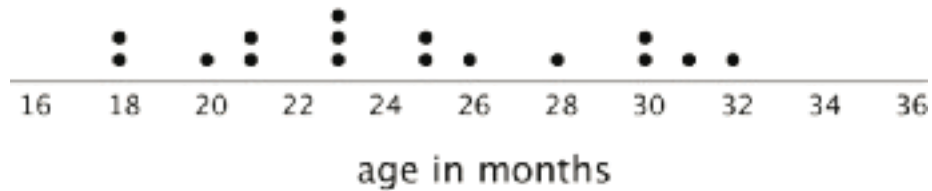
2. If you decide that a question is statistical, describe how you would find the answer. What data would you collect?

Lesson 3: Representing Data Graphically

Cool Down: Swimmers and Swimming Class

1. Noah gathered information on the home states of the swimmers on Team USA. He organized the data in a table. Would a dot plot be appropriate to display his data? Explain your reasoning.

2. This dot plot shows the ages of students in a swimming class. How many students are in the class?



3. Based on the dot plot, do you agree with each of the following statements? Explain your reasoning.

a. The class is an adult swimming class.

b. Half of the students are between 2 and 3 years old.

Lesson 4: Dot Plots

Cool Down: Family Size

A group of students was asked, "How many children are in your family?" The responses are displayed in the dot plot.

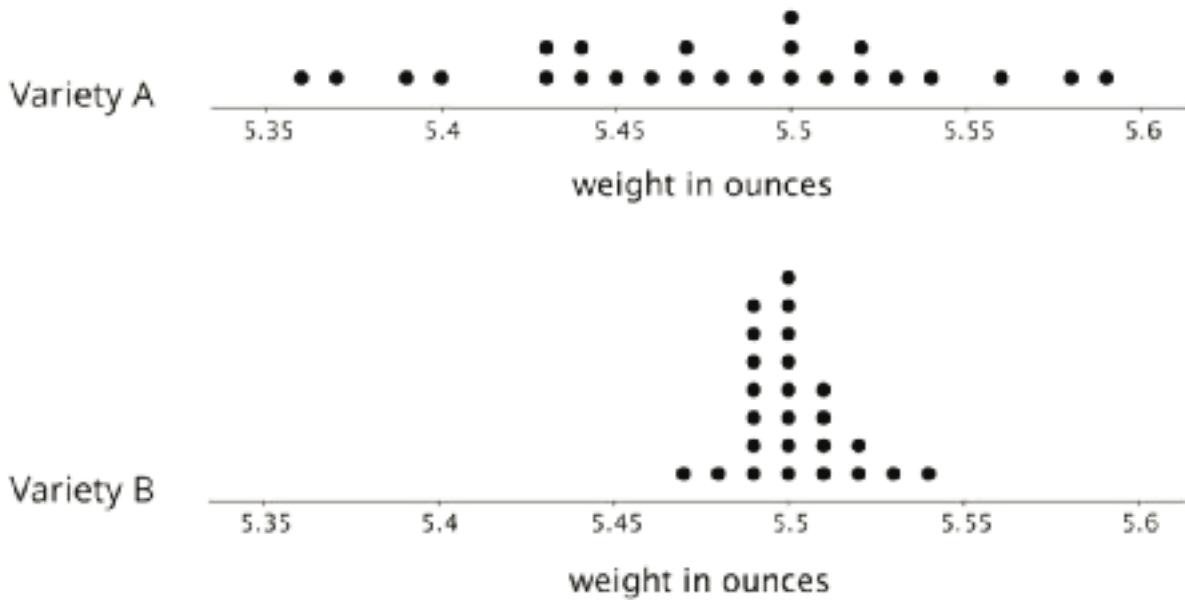


1. How many students responded to the questions?
2. What percentage of the students have more than one child in the family?
3. Write a sentence that describes the distribution of the data shown on the dot plot.

Lesson 5: Using Dot Plots to Answer Statistical Questions

Cool Down: Packing Tomatoes

A farmer sells tomatoes in packages of ten. She would like the tomatoes in each package to all be about the same size and close to 5.5 ounces in weight. The farmer is considering two different tomato varieties: Variety A and Variety B. She weighs 25 tomatoes of each variety. These dot plots show her data.



1. What would be a good description for the weight of Variety A tomatoes, in general? What about for the weight of Variety B tomatoes, in general?

2. Which tomato variety should the farmer choose? Explain your reasoning.

Lesson 6: Interpreting Histograms

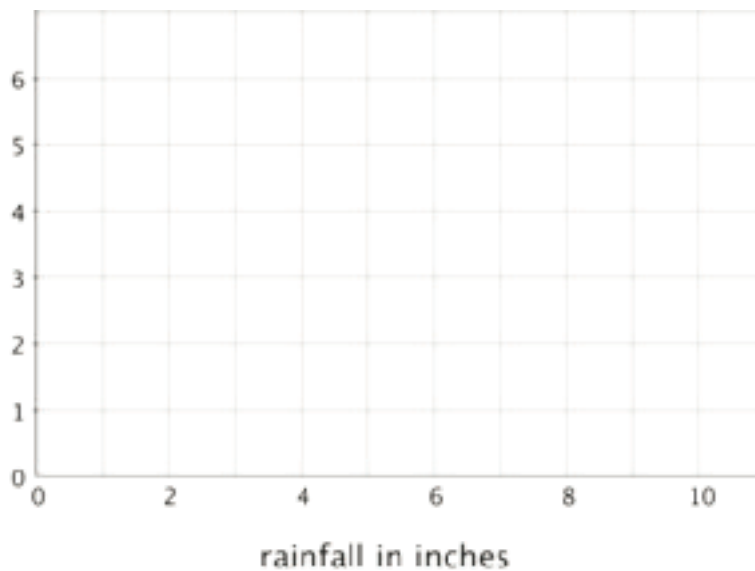
Cool Down: Rain in Miami

Here is the average amount of rainfall, in inches, for each month in Miami, Florida.

month	rainfall (inches)	month	rainfall (inches)
January	1.61	July	6.5
February	2.24	August	8.9
March	2.99	September	9.84
April	3.14	October	6.34
May	5.35	November	3.27
June	9.69	December	2.05

1. Complete the frequency table and use it to make a histogram.

rainfall (inches)	frequency
0-2	
2-4	
4-6	
6-8	
8-10	

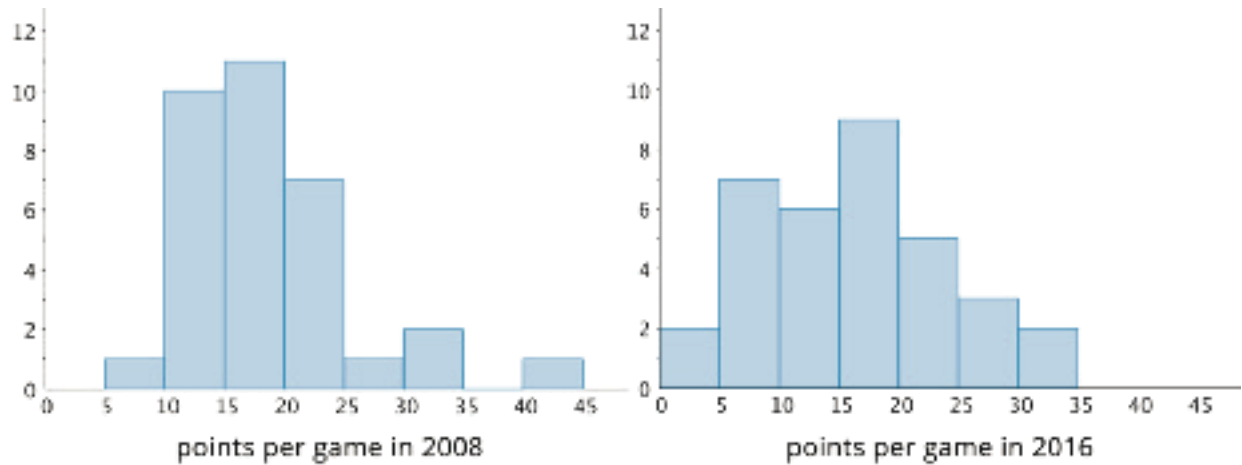


2. What is a typical amount of rainfall in one month in Miami?

Lesson 7: Using Histograms to Answer Statistical Questions

Cool Down: A Tale of Two Seasons

The two histograms show the points scored per game by a basketball player in 2008 and 2016.



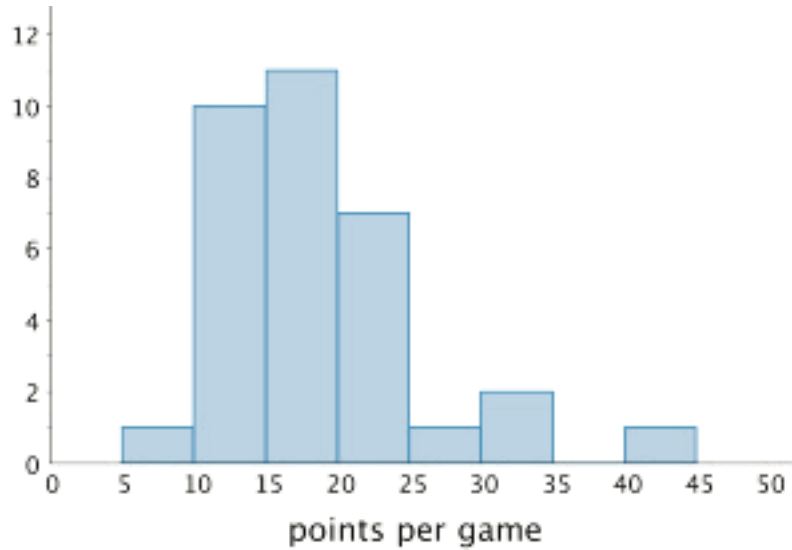
1. What is a typical number of points per game scored by this player in 2008? What about in 2016? Explain your reasoning.

2. Write 2–3 sentences that describe the spreads of the two distributions, including what spreads might tell us in this context.

Lesson 8: Describing Distributions on Histograms

Cool Down: Point Spread

Here is a histogram that shows the number of points scored by a college basketball player during the 2008 season. Describe the shape and features of the data.



Lesson 10: Finding and Interpreting the Mean as the Balance Point

Cool Down: Text Messages

The three data sets show the number of text messages sent by Jada, Diego, and Lin over 6 days. One of the data sets has a mean of 4, one has a mean of 5, and one has a mean of 6.

Jada	4	4	4	6	6	6
Diego	4	5	5	6	8	8
Lin	1	1	2	2	9	9

1. Which data set has which mean? What does this tell you about the text messages sent by the three students?

2. Which data set has the greatest variability? Explain your reasoning.

Lesson 11: Variability and MAD

Cool Down: Text Messages, Again

These three data sets show the number of text messages sent by Jada, Diego, and Lin over 6 days as well as the mean number of text messages sent by each student per day.

Jada 4 4 4 6 6 6
mean: 5

Diego 4 5 5 6 8 8
mean: 6

Lin 1 1 2 2 9 9
mean: 4

1. Predict which data set has the largest MAD and which has the smallest MAD.

2. Compute the MAD for each data set to check your prediction.

Lesson 12: Using Mean and MAD to Make Comparisons

Cool Down: Travel Times Across the World

Ten sixth-grade students in five different countries were asked about their travel times to school. Their responses were organized into five data sets. The mean and MAD of each data set is shown in the table.

	mean (minutes)	MAD (minutes)
United States	9	4.2
Australia	18.1	7.9
South Africa	23.5	16.2
Canada	11	8
New Zealand	12.3	5.5

1. Which group of students has the greatest variability in their travel times? Explain your reasoning.
2.
 - a. The mean of the data set for New Zealand is close to that of Canada. What does this tell us about the travel times of students in those two data sets?
 - b. The MAD of the data set for New Zealand is quite different than that of Canada. What does this tell us about the travel times of students in those two data sets?
3. The data sets for Australia and Canada have very different means (18.1 and 11 minutes) but very similar MADs. What can you say about the travel times of the students in those two data sets?

Lesson 13: Median

Cool Down: Practicing the Piano

Jada and Diego are practicing the piano for an upcoming rehearsal. The tables list the number of minutes each of them practiced in the past few weeks.

Jada's practice times:

10	10	20	15	25	25	8	15
20	20	35	25	40			

Diego's practice times:

25	10	15	30	15	20	20	25
30	45						

1. Find the median of each data set.

2. Explain what the medians tell you about Jada's and Diego's piano practice.

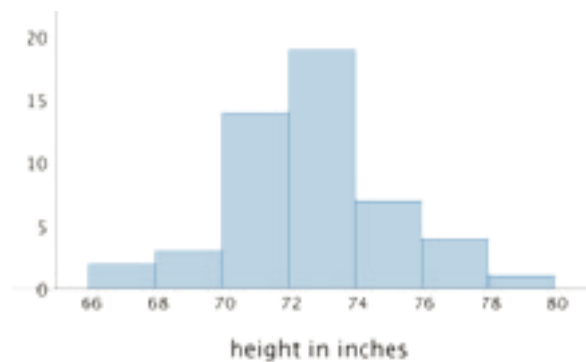
Lesson 14: Comparing Mean and Median

Cool Down: Which Measure of Center to Use?

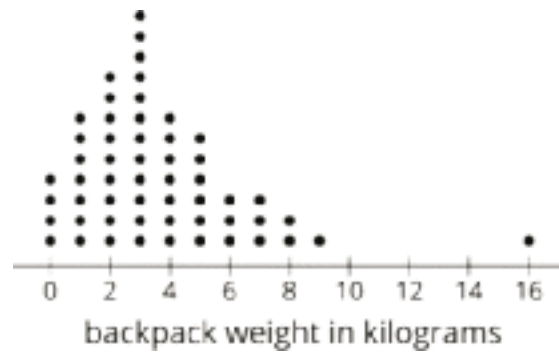
For each dot plot or histogram:

- Predict if the mean is greater than, less than, or approximately equal to the median. Explain your reasoning.
- Which measure of center—the mean or the median—better describes a typical value for the following distributions?

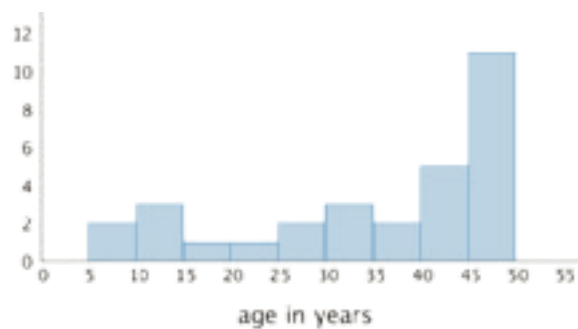
1. Heights of 50 NBA basketball players



2. Backpack weights of 55 sixth-grade students



3. Ages of 30 people at a family dinner party



Lesson 15: Quartiles and Interquartile Range

Cool Down: How Far Can You Throw?

Diego wondered how far sixth-grade students could throw a ball. He decided to collect data to find out. He asked 10 friends to throw a ball as far as they could and measured the distance from the starting line to where the ball landed. The data shows the distances he recorded in feet.

40 76 40 63 47 57 49 55 50 53

1. Find the median and IQR of the data set.

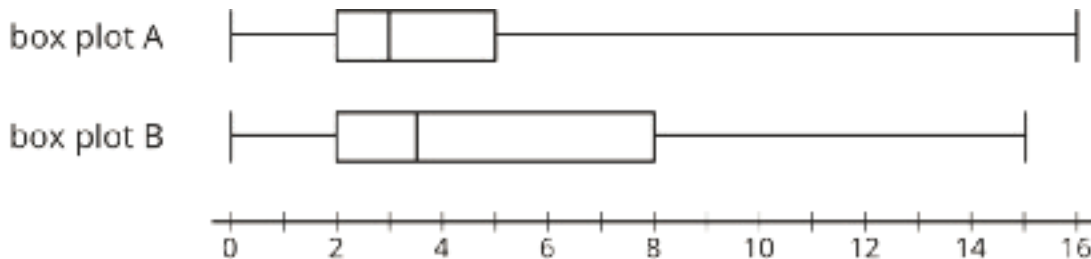
2. On a later day, he asked the same group of 10 friends to throw a ball again and collected another set of data. The median of the second data set is 49 feet, and the IQR is 6 feet.
 - a. Did the 10 friends, as a group, perform better (throw farther) in the second round compared to the first round? Explain how you know.

 - b. Were the distances in the second data set more variable or less variable compared to those in the first round? Explain how you know.

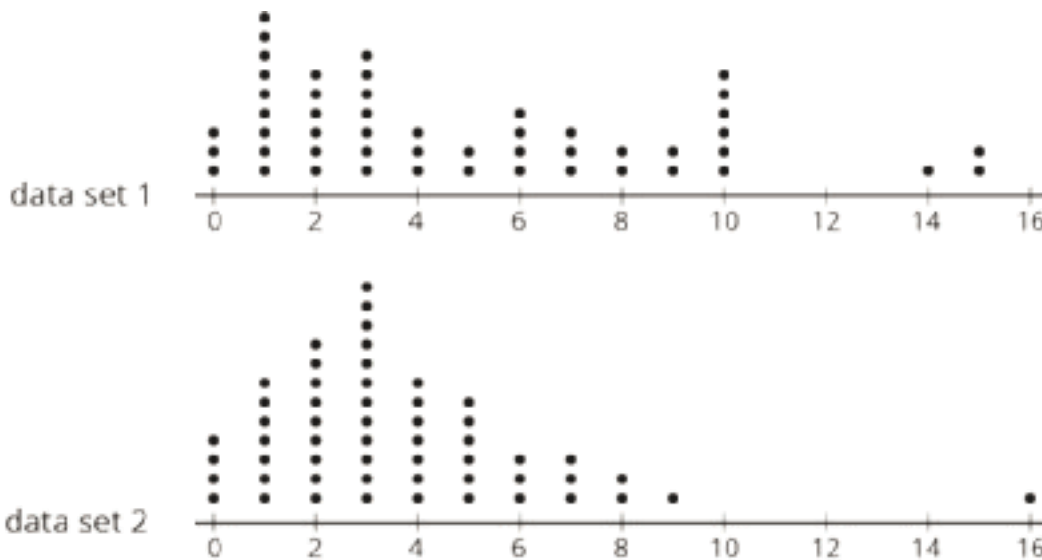
Lesson 16: Box Plots

Cool Down: Boxes and Dots

1. Here are two box plots that summarize two data sets. Do you agree with each of the following statements?



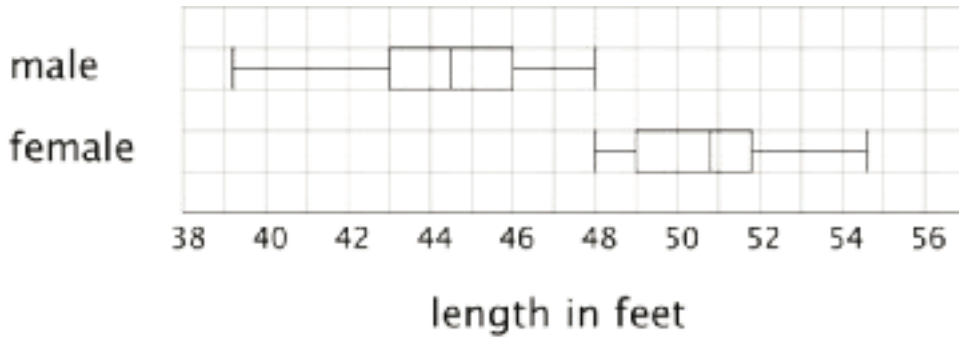
- Both data sets have the same range.
 - Both data sets have the same minimum value.
 - The IQR shown in box plot B is twice the IQR shown in box plot A.
 - Box plot A shows a data set that has a quarter of its values between 2 and 5.
2. These dot plots show the same data sets as those represented by the box plots. Decide which box plot goes with each dot plot. Explain your reasoning.



Lesson 17: Using Box Plots

Cool Down: Humpback Whales

Researchers measured the lengths, in feet, of 20 male humpback whales and 20 female humpback whales. Here are two box plots that summarize their data.

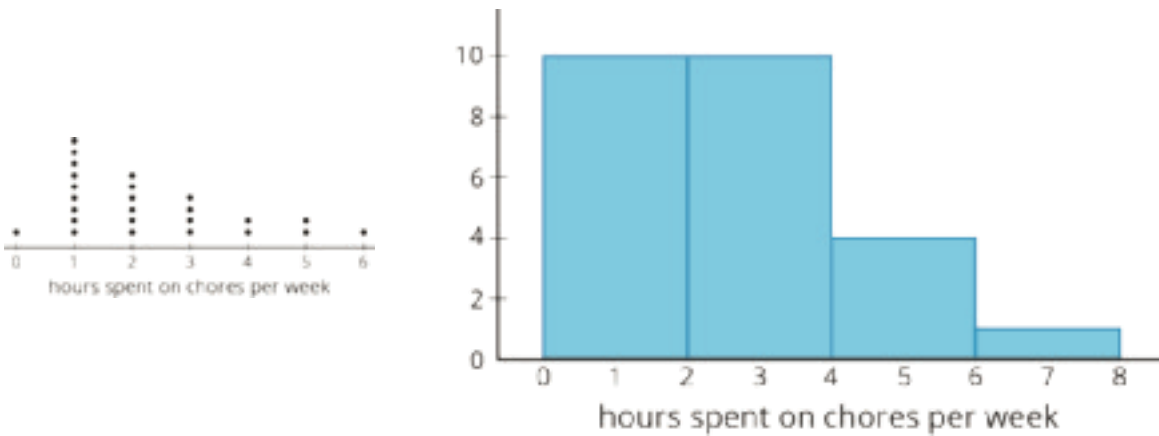


1. How long was the longest whale measured? Was this whale male or female?
2. What was a typical length for the male humpback whales that were measured?
3. Do you agree with each of these statements about the whales that were measured? Explain your reasoning.
 - a. More than half of male humpback whales measured were longer than 46 feet.
 - b. The male humpback whales tended to be longer than female humpback whales.
 - c. The lengths of the male humpback whales tended to vary more than the lengths of the female humpback whales.

Lesson 18: Using Data to Solve Problems

Cool Down: Time Spent on Chores

Lin surveyed her classmates on the number of hours they spend doing chores each week. She represented her data with a dot plot and a histogram.



1. Lin thinks that she could find the median, the minimum, and the maximum of the data set using both the dot plot and the histogram. Do you agree? Explain your reasoning.

2. Should Lin use the mean and MAD or the median and IQR to summarize her data? Explain your reasoning.

CKMath[™]
Core Knowledge **MATHEMATICS**[™]



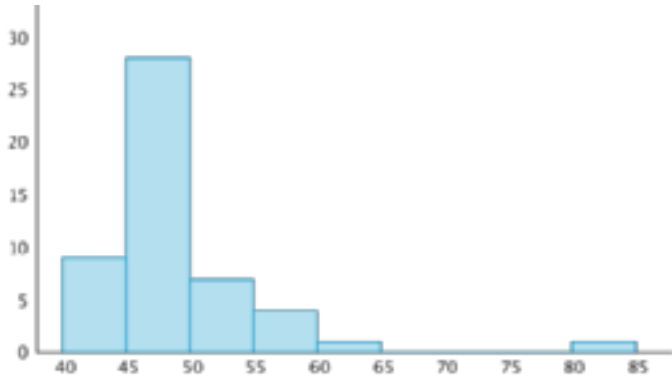
Instructional Masters

Instructional Masters for Data Sets and Distributions

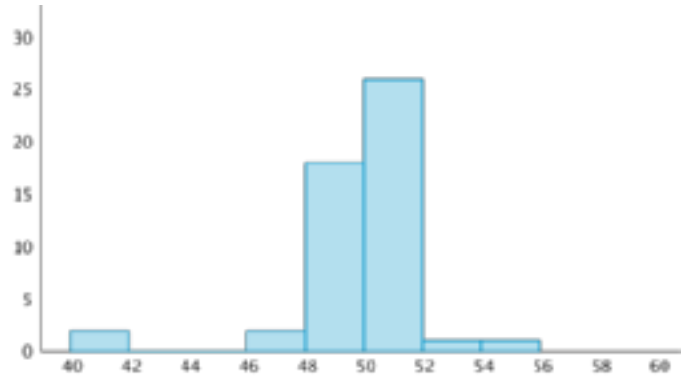
address	title	students per copy	written on?	requires cutting?	card stock recommended?	color paper recommended?
Activity Grade6.8.8.2	Sorting Histograms	1	no	yes	yes	no
Activity Grade6.8.14.3	Mean or Median?	1	no	yes	no	no
Activity Grade6.8.17.2	Info Gap: Sea Turtles	1	no	yes	no	no
Activity Grade6.8.2.4	Sifting for Statistical Questions	1	no	no	no	no
Activity Grade6.8.8.3	Getting to School	1	no	no	no	no

<p>Sifting for Statistical Questions</p> <p>A. What fraction of the people in your school wear glasses?</p>	<p>Sifting for Statistical Questions</p> <p>G. Who is the current Vice President of the United States?</p>
<p>Sifting for Statistical Questions</p> <p>B. How many centimeters tall is the door of your classroom?</p>	<p>Sifting for Statistical Questions</p> <p>H. How old is the principal at your school?</p>
<p>Sifting for Statistical Questions</p> <p>C. What percentage of students in your class say they like dogs better than cats?</p>	<p>Sifting for Statistical Questions</p> <p>I. What is the most common favorite color for the students in your class?</p>
<p>Sifting for Statistical Questions</p> <p>D. What is the average age of the teachers at your school?</p>	<p>Sifting for Statistical Questions</p> <p>J. Do students at your school generally get more sleep on school nights than on weekends and holidays?</p>
<p>Sifting for Statistical Questions</p> <p>E. What do students in your class prefer to have on a hot dog: ketchup, mustard, both ketchup and mustard, or neither?</p>	<p>Sifting for Statistical Questions</p> <p>K. Who is is the oldest staff member at the school?</p>
<p>Sifting for Statistical Questions</p> <p>F. What is a typical number of students per class in your school?</p>	<p>Sifting for Statistical Questions</p> <p>L. What day of the week is today?</p>

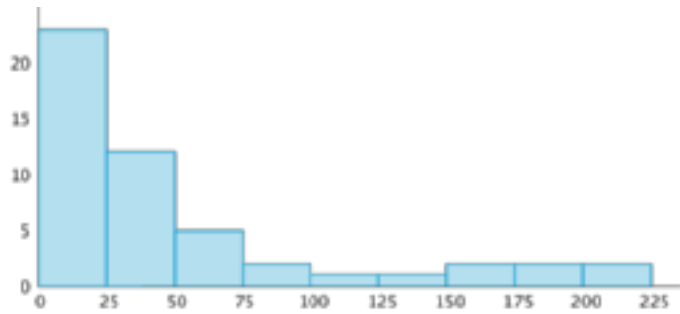
Sorting Histograms
Histogram A



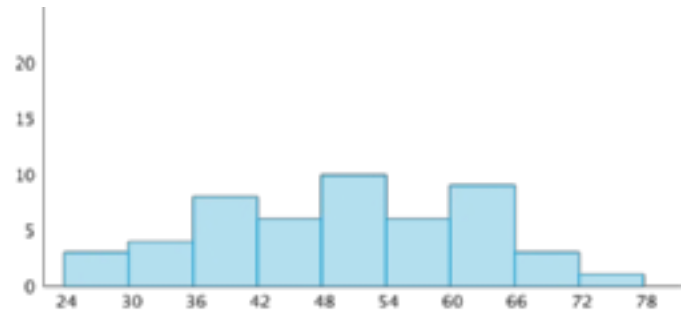
Sorting Histograms
Histogram E



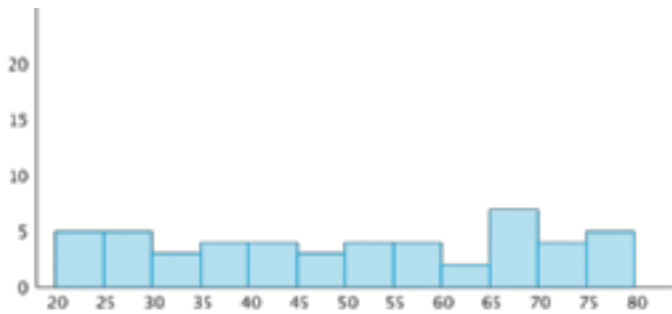
Sorting Histograms
Histogram B



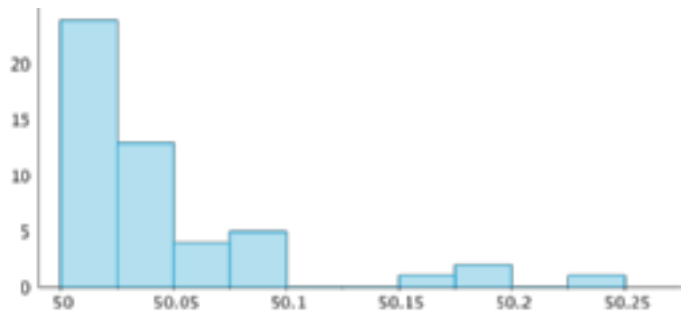
Sorting Histograms
Histogram F



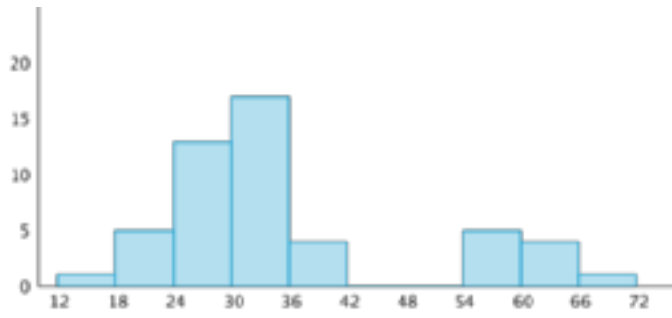
Sorting Histograms
Histogram C



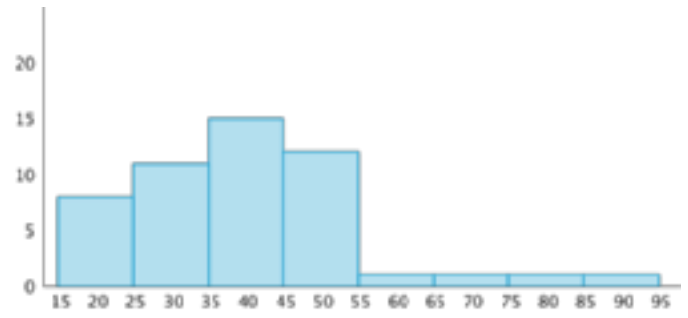
Sorting Histograms
Histogram G



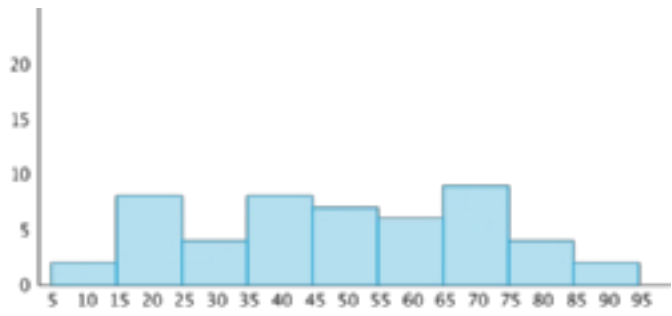
Sorting Histograms
Histogram D



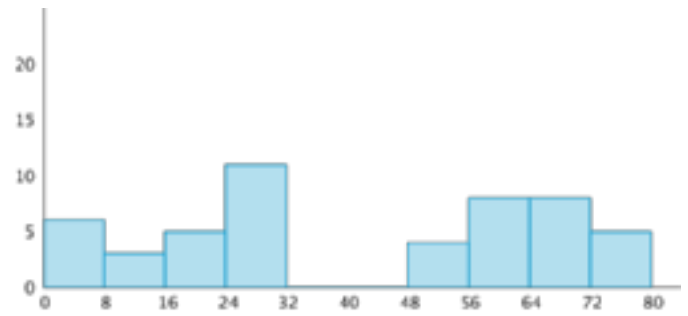
Sorting Histograms
Histogram H



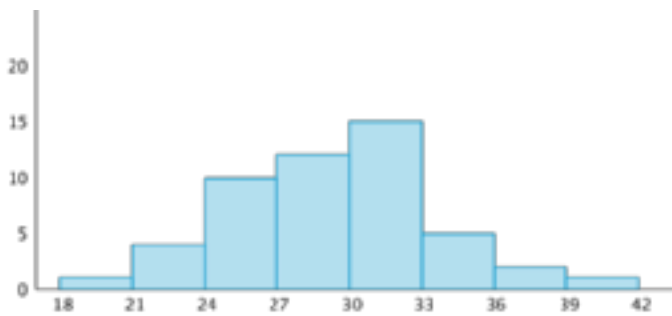
Sorting Histograms
Histogram I



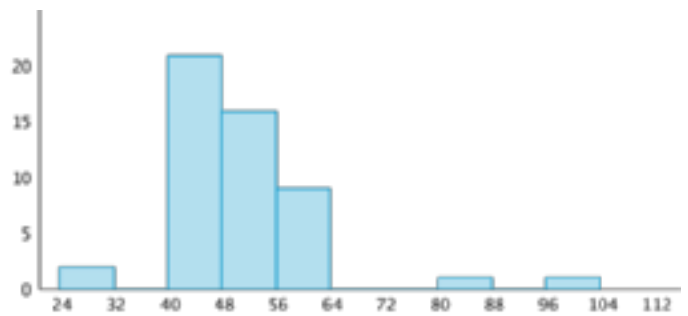
Sorting Histograms
Histogram K



Sorting Histograms
Histogram J



Sorting Histograms
Histogram L



6.8.8.3 Getting to School.

methods of travel	frequency
walk	
bike	
scooter or skateboard	
school bus	
car	
public transportation	
other	

minutes of travel time	frequency
0 to less than 5 minutes	
5 to less than 10 minutes	
10 to less than 15 minutes	
15 to less than 20 minutes	
20 to less than 25 minutes	
25 to less than 30 minutes	
30 to less than 35 minutes	
35 to less than 40 minutes	
40 to less than 45 minutes	
45 to less than 50 minutes	

Mean or Median?

Card A

Ages of 20 dogs being treated at an animal hospital



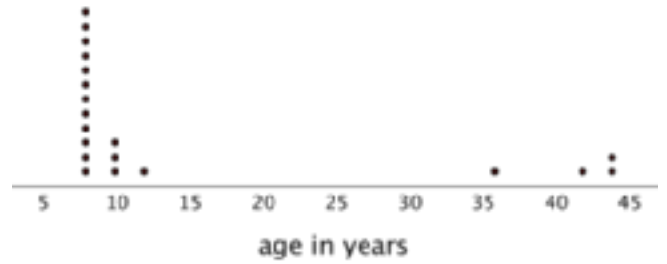
Mean: Approximately 7.8 years old

Median: 8 years old

Mean or Median?

Card D

Ages of 20 people at an 8-year-old's birthday party



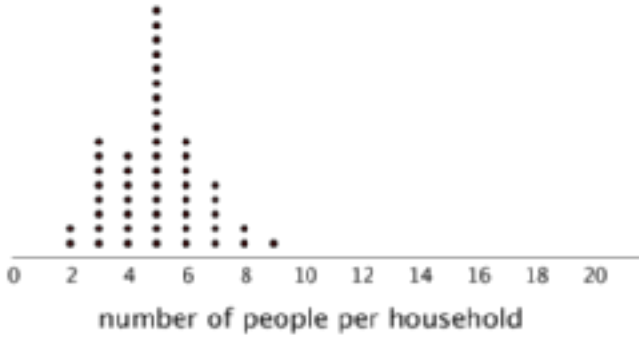
Mean: Approximately 15 years old

Median: 8 years old

Mean or Median?

Card B

Size of household of 50 people in Ireland



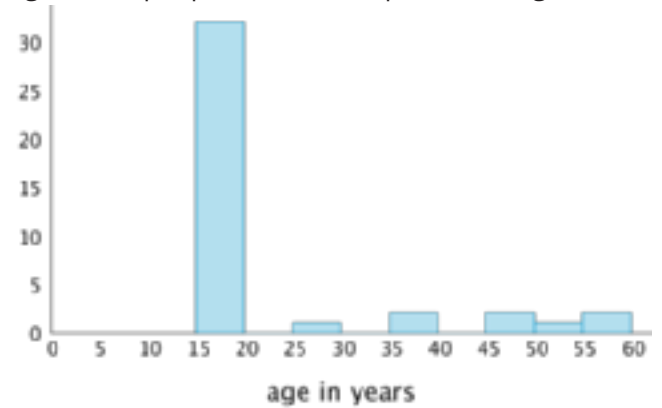
Mean: Approximately 5 people per household

Median: 5 people per household

Mean or Median?

Card F

Ages of 40 people on a field trip to Washington D.C.



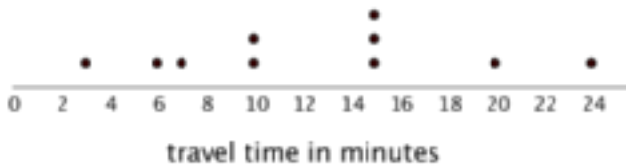
Mean: Approximately 21.3 years old

Median: 15 years old

Mean or Median?

Card C

Time it took ten sixth-grade students in New Zealand to travel to school



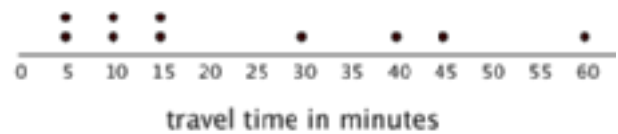
Mean: Approximately 12.5 minutes

Median: 12.5 minutes

Mean or Median?

Card E

Time it took ten sixth-grade students in South Africa to travel to school



Mean: Approximately 24 minutes

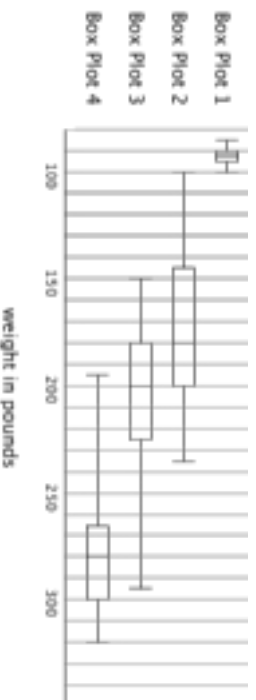
Median: 15 minutes

Info Gap: Sea Turtles

Problem Card 1

The box plots represent data on the weights of several species of sea turtles that nest at the Outer Banks of North Carolina. Two of the species are the olive ridley and the hawksbill.

Which two box plots represent the data for the weights of olive ridley and hawksbill sea turtles?



Info Gap: Sea Turtles

Data Card 1

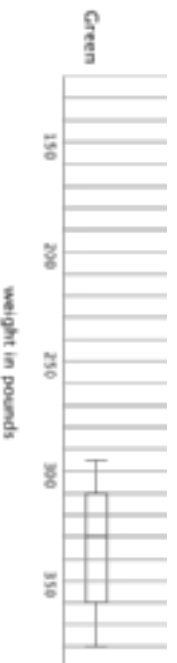
- The median for the hawksbill data is about twice the median for the olive ridley data.
- The range for the hawksbill data is 135 pounds.
- The IQR for the olive ridley data is 5 pounds and the range is 15 pounds.
- The heaviest weight in the olive ridley data is 100 pounds.

Info Gap: Sea Turtles

Problem Card 2

Loggerhead and green sea turtles are two species that also nest at the Outer Banks. Here is the box plot for a set of data on weights of green sea turtles.

- Which of the two species has a heavier typical weight? Explain how you know.
- Which species shows more variation in weight? Explain how you know.



Info Gap: Sea Turtles

Data Card 2

- The IQR for the loggerhead data is 45 pounds.
- The lightest weight in the loggerhead data is 150 pounds.
- The median for the loggerhead data is 130 pounds less than the median for the green turtle data.
- The heaviest weight in the loggerhead data is 295 pounds.

Credits

CKMath K–8 was originally developed by Open Up Resources and authored by Illustrative Mathematics, <https://www.illustrativemathematics.org>, and is copyrighted as 2017–2019 by Open Up Resources. It is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). The Open Up Resources K–8 Math Curriculum is available at: <https://www.openupresources.org/math-curriculum/>.

Adaptations and updates to the IM K–8 Math English language learner supports are copyright 2019 by Open Up Resources and licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0),

Adaptations and updates to IM K–8 Math are copyright 2019 by Illustrative Mathematics, including the additional English assessments marked as "B", and the Spanish translation of assessments marked as "B". These adaptations and updates are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

This particular work is based on additional work of the Core Knowledge® Foundation (www.coreknowledge.org) made available through licensing under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Illustration and Photo Credits

Ivan Pestic / Cover Illustrations

Illustrative Math K–8 / Cover Image, all interior illustrations, diagrams, and pictures / Copyright 2019 / Licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

These materials include public domain images or openly licensed images that are copyrighted by their respective owners, unless otherwise noted/credited. Openly licensed images remain under the terms of their respective licenses.



CKMath™
Core Knowledge **MATHEMATICS™**

CKMath™
Core Knowledge MATHEMATICS™

A comprehensive program for mathematical skills and concepts
as specified in the **Core Knowledge Sequence**
(content and skill guidelines for Grades K–8).

Core Knowledge MATHEMATICS™

units at this level include:

Area and Surface Area
Introducing Ratios
Unit Rates and Percentages
Dividing Fractions
Arithmetic in Base Ten
Expressions and Equations
Rational Numbers
Data Sets and Distributions
Putting it All Together

www.coreknowledge.org

Core Knowledge Curriculum Series™