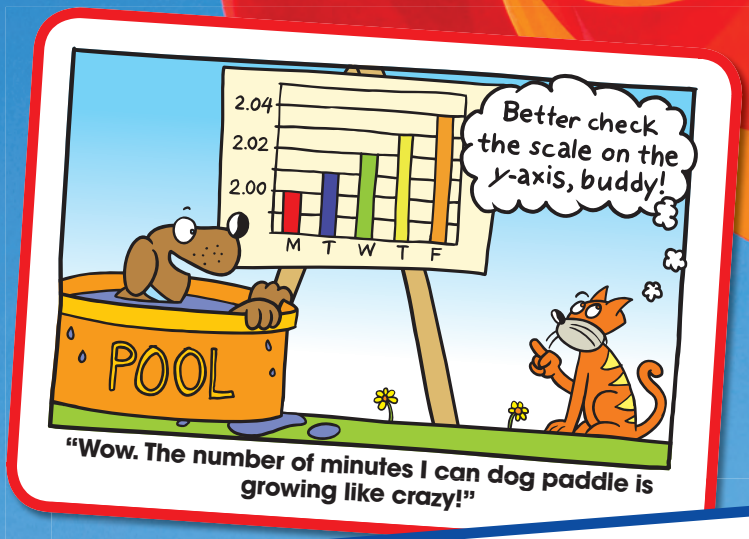
12 Data Analysis and Displays



"Wow. The number of minutes I can dog paddle is growing like crazy!"

"Please hold still. I am trying to find the mean of 6, 8, and 10 by dividing their sum into three equal piles."

Connections to Previous Learning

- Determine measures of central tendency including mean, median, mode, and range.
- Select appropriate measures of central tendency to describe a data set.

- Evaluate the reasonableness of a sample to determine the appropriateness of generalizations made about the population.
- Construct and analyze histograms, stem-and-leaf plots, and circle graphs.

- Construct and analyze dot plots, histograms, box-and-whisker plots, and scatter plots.
- Use the shape of a data distribution to select appropriate measures of central tendency and dispersion, and to account for the effects of outliers in the data.
- Interpret linear models.

Pacing Guide for Chapter 12

Chapter Opener	1 Day
Section 1	1 Day
Section 2	1 Day
Section 3	1 Day
Section 4	1 Day
Study Help / Quiz	1 Day
Section 5	1 Day
Section 6	2 Days
Section 7	1 Day
Section 8	1 Day
Chapter Review / Chapter Tests	2 Days
Total Chapter 12	13 Days
Year-to-Date	164 Days

Chapter Summary

Section	Common Core State Standard	
12.1	Learning	S.ID.2, S.ID.3
12.2	Learning	S.ID.2, S.ID.3
12.3	Learning	S.ID.1, S.ID.2, S.ID.3
12.4	Learning	S.ID.2 ★, S.ID.3 ★
12.5	Learning	8.SP.1 ★, S.ID.6a ★, S.ID.6c ★
12.6	Learning	S.ID.6b ★, S.ID.8 ★, S.ID.9 ★
12.7	Learning	8.SP.4 ★, S.ID.5 ★
12.8	Learning	S.ID.1 ★
★ Teaching is complete. Standard can be assessed.		

Technology for the Teacher

BigIdeasMath.com
Chapter at a Glance
Complete Materials List
Parent Letters: English and Spanish

Common Core State Standards

6.SP.5a Summarize numerical data sets in relation to their context, such as by reporting the number of observations.
6.SP.4 Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

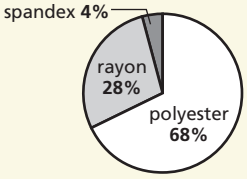
Additional Topics for Review

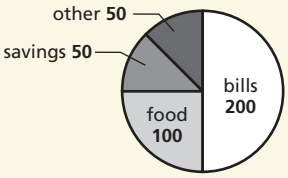
- Operations with real numbers
- Outliers
- Writing equations of lines
- Bar graphs
- Line graphs
- Stem-and-leaf plots

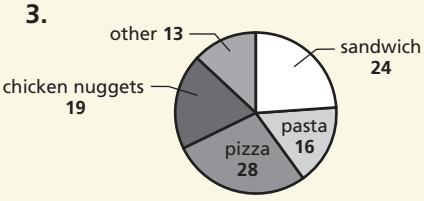
Try It Yourself

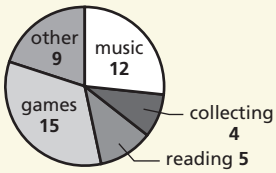
1. Answers will vary.
2. Answers will vary.

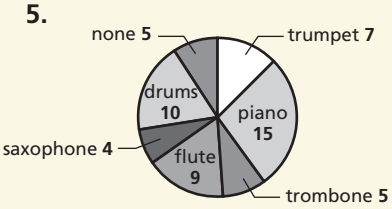
Record and Practice Journal Fair Game Review

1.


2.


3.


4.


5.


6–10. See Additional Answers.

Math Background Notes

Vocabulary Review

- Circle Graph
- Frequency Table
- Histogram

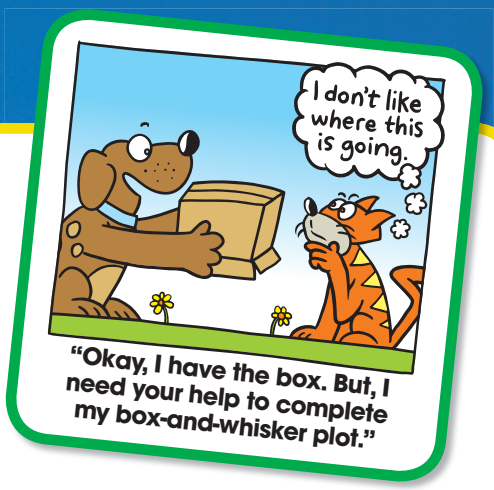
Displaying Data

- Students have collected, analyzed, and displayed data.
- **Teaching Tip:** Example 1 provides a great opportunity for review. Remind students that a circle contains 360°. This is also a good time to review using a protractor.
- Remind students that it is helpful to know the total number of people surveyed before constructing the circle graph. This number will serve as the whole (denominator of the fraction).
- **Teaching Tip:** Example 2 provides an excellent opportunity to explore students' prerequisite knowledge. Students should be familiar with bar graphs, double bar graphs, and histograms. Consider using a Venn diagram to compare and contrast the three types of displays. This will create a nice visual representation to show which characteristics go with which display.
- **Common Error:** Students might forget that a histogram uses intervals rather than individual data values. Remind students that the horizontal axis of the graph will be labeled with intervals and the vertical axis of the graph will be labeled with frequencies.
- **Common Error:** Even when an interval has a frequency of zero, it must appear on the histogram.
- **Teaching Tip:** The exercises in this set require students to take a survey to collect data and then display the data using a circle graph and a histogram. This provides a good opportunity to revisit topics such as population and sample size. What characteristics make for a fair survey? How will students ensure that the data they collect is a fair representation of the population?
- You can adapt the context of the survey to personalize it to your class.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Record and Practice Journal • Fair Game Review Skills Review Handbook Lesson Tutorials	Game Closet at *BigIdeasMath.com* Start the next section

What You Learned Before



Displaying Data (6.SP.5a, 6.SP.4)

Example 1 **The table shows the results of a survey. Display the data in a circle graph.**

Class Trip Location	Water park	Museum	Zoo	Other
Students	25	11	5	4

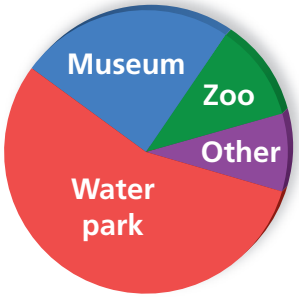
A total of 45 students took the survey.

Water park: $\frac{25}{45} \cdot 360° = 200°$

Museum: $\frac{11}{45} \cdot 360° = 88°$

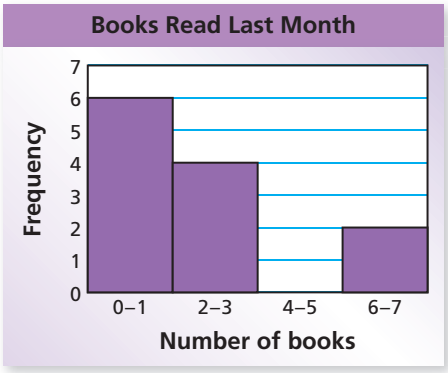
Zoo: $\frac{5}{45} \cdot 360° = 40°$

Other: $\frac{4}{45} \cdot 360° = 32°$



Example 2 **The frequency table shows the numbers of books that 12 people read last month. Display the data in a histogram.**

Books Read Last Month	Frequency
0–1	6
2–3	4
4–5	0
6–7	2



Try It Yourself

1. Conduct a survey to determine the after-school activities of students in your class. Display the results in a circle graph.
2. Conduct a survey to determine the numbers of pets owned by students in your class. Display the results in a histogram.

12.1 Measures of Central Tendency

Essential Question How can you use measures of central tendency to distribute an amount evenly among a group of people?

1 ACTIVITY: Exploring Mean, Median, and Mode

Work with a partner. Forty-five coins are arranged in nine stacks.

a. Record the number of coins in each stack in a table.

Stack	1	2	3	4	5	6	7	8	9
Coins									

b. Find the mean, median, and mode of the data.

c. By moving coins from one stack to another, can you change the mean? the median? the mode? Explain.

d. Is it possible to arrange the coins in stacks so that the median is 6? 8? Explain.

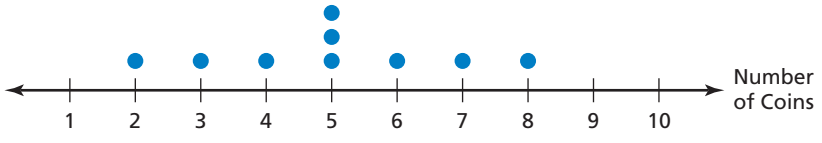
2 EXAMPLE: Drawing a Dot Plot

Work with a partner.

a. Draw a number line. Label the tick marks from 1 to 10.

b. Place each stack of coins in Activity 1 above the number of coins in the stack.

c. Draw a ● to represent each stack. This data display is called a *dot plot*.



Common Core

Measures of Central Tendency

In this lesson, you will

- find the mean, median, and mode of a data set.
- identify and remove outliers.
- explain the effects of changing values in data sets.

Learning Standards
S.ID.2
S.ID.3

Laurie's Notes

Introduction

Standards for Mathematical Practice

- **MP2 Reason Abstractly and Quantitatively:** The physical and graphical models help students make sense of the quantities and their relationships in problem situations. In these activities, students develop an understanding of measures of central tendency.

Motivate

- Explain the work of the U.S. Census Bureau. Share the following U.S. information from the Census Bureau with students. In each case, ask students to interpret what the statistic means.
  - The mean travel time to work for workers age 16 and older is 25.1 minutes.
  - The median household income in 2009 was $50,221.
  - The *average* family size is 3.24 people.

Activity Notes

Activity 1

- For this first activity, students benefit by having some sort of manipulatives, such as coins, circular disks, square tiles, or cubes. If the manipulatives are stackable, the activity will be easier for students to follow.
- **MP2:** The concrete models help students visualize the distribution of the data, and that there could be outliers and/or gaps in the data.
- Review *mean*, *median*, and *mode*.
- Give sufficient time for students to work through each part of this activity.
- Discuss students' explanations for parts (c) and (d).
- **Big Idea:** As the median increases, the distribution becomes more skewed. One way to have a median of 8 coins is to have three stacks of 1 coin, one stack of 2 coins, and five stacks of 8 coins.

Example 2

- This example reviews dot plots.
- Reading the dot plot shown, there are three stacks that have 5 coins in them, no stacks that have 1, 9, or 10 coins in them, and all of the other possibilities have one stack each.
- This plot is reviewed so that it can be used to explore fair distributions in the next activity.

Common Core State Standards

S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Previous Learning

Students should know how to find the mean, median, and mode.

Technology for the Teacher

Dynamic Classroom

Lesson Plans
Complete Materials List

12.1 Record and Practice Journal

Essential Question How can you use measures of central tendency to distribute an amount evenly among a group of people?

1 ACTIVITY: Exploring Mean, Median, and Mode

Work with a partner. Forty-five coins are arranged in nine stacks.

5 4 3 6 2 5 8 7 5

a. Record the number of coins in each stack in a table.

Stack	1	2	3	4	5	6	7	8	9
Coins	5	4	3	6	2	5	8	7	5

b. Find the mean, median, and mode of the data.

mean: 5; median: 5; mode: 5

c. By moving coins from one stack to another, can you change the mean? the median? the mode? Explain.

mean cannot change; median and mode can change

d. Is it possible to arrange the coins in stacks so that the median is 6? 8? Explain. yes; yes;

Sample answer:
median 6: 1, 2, 3, 4, 6, 6, 7, 8, 8
median 8: 1, 1, 1, 2, 8, 8, 8, 8, 8

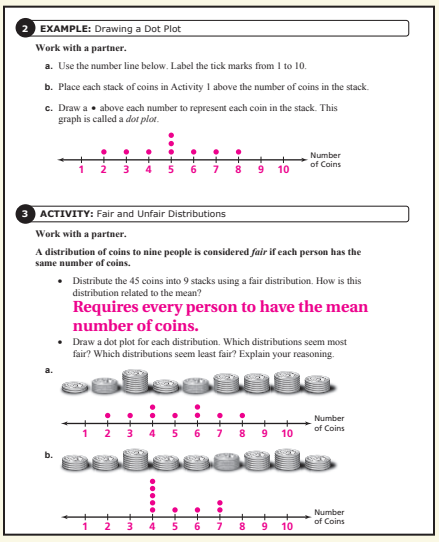
English Language Learners

Graphic Organizer

Provide small groups with different data sets and a copy of the organizer. Have each group find the mean, median, mode, and range of their data set and present their work to the class.

Data set:	
Mean	
Median	
Mode	
Range	

12.1 Record and Practice Journal



2 EXAMPLE: Drawing a Dot Plot

Work with a partner.

a. Use the number line below. Label the tick marks from 1 to 10.

b. Place each stack of coins in Activity 1 above the number of coins in the stack.

c. Draw a • above each number to represent each coin in the stack. This graph is called a *dot plot*.

1 2 3 4 5 6 7 8 9 10 Number of Coins

3 ACTIVITY: Fair and Unfair Distributions

Work with a partner.

A distribution of coins to nine people is considered *fair* if each person has the same number of coins.

- Distribute the 45 coins into 9 stacks using a fair distribution. How is this distribution related to the mean?

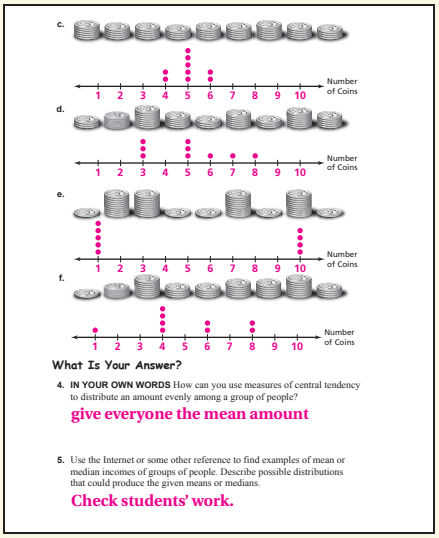
Requires every person to have the mean number of coins.

- Draw a dot plot for each distribution. Which distributions seem most fair? Which distributions seem least fair? Explain your reasoning.

a.

b.

1 2 3 4 5 6 7 8 9 10 Number of Coins



c.

d.

e.

f.

1 2 3 4 5 6 7 8 9 10 Number of Coins

What Is Your Answer?

4. **IN YOUR OWN WORDS** How can you use measures of central tendency to distribute an amount evenly among a group of people?

give everyone the mean amount

5. Use the Internet or some other reference to find examples of mean or median incomes of groups of people. Describe possible distributions that could produce the given means or medians.

Check students' work.

Laurie's Notes

Activity 3

- Read the definition of a fair distribution and discuss what a fair distribution has to do with the mean.
- When students construct the dot plot for each part, they should get the sense that it is similar to constructing a bar graph. The heights of the bars in a bar graph are similar to the heights of the dots in a dot plot.
- Discuss student responses to which distributions seem most and least fair.
- **MP4 Model with Mathematics:** The coins and dot plot visually model the concept of a fair distribution. The models help students understand that mathematics can be used to model everyday concepts such as means, medians, and modes.
- **Extension:** Given the requirement that each stack has to have at least one coin, ask students to make a dot plot for the most fair distribution possible and the least fair distribution possible. The most fair would be 5 coins in each stack, so the dot plot would have nine dots above the 5 and no dots elsewhere. The least fair would likely be eight dots above the 1 and one dot above the 37, meaning the dot plot would need to be extended and there would be a big gap.
- **Big Idea:** You can think of the mean as a sharing process in which you are trying to level out the stacks. If the average amounts of money 3 students have is \$10, it is possible that they all have \$10 (most fair distribution) or one has \$30 and the others have no money (least fair distribution). If all 3 students are known to have some money, pooling it and spreading it out into 3 stacks would level out at \$10 per stack—the mean.

What Is Your Answer?

- For Question 4, the mean is the only measure of central tendency that can be used to distribute an amount evenly.

Closure

- Distribute 50 coins in 10 stacks. Make a dot plot of the distribution.

3 ACTIVITY: Fair and Unfair Distributions

Math Practice 8

Maintain Oversight

What process was used to distribute the coins equally?

Work with a partner.

A distribution of coins to nine people is considered *fair* if each person has the same number of coins.

- Distribute the 45 coins into 9 stacks using a fair distribution. How is this distribution related to the mean?
- Draw a dot plot for each distribution. Which distributions seem most fair? Which distributions seem least fair? Explain your reasoning.

a.

b.

c.

d.

e.

f.

What Is Your Answer?

4. **IN YOUR OWN WORDS** How can you use measures of central tendency to distribute an amount evenly among a group of people?
5. Use the Internet or some other reference to find examples of mean or median incomes of groups of people. Describe possible distributions that could produce the given means or medians.

Practice Use what you learned about measures of central tendency to complete Exercise 4 on page 610.

12.1 Lesson

Check It Out
Lesson Tutorials
BigIdeasMath.com

Key Vocabulary
measure of central tendency, *p. 608*

A **measure of central tendency** is a measure that represents the center of a data set. The *mean*, *median*, and *mode* are measures of central tendency.

Key Ideas

Mean

The *mean* of a data set is the sum of the data divided by the number of data values.

Median

Order the data. For a set with an odd number of values, the *median* is the middle value. For a set with an even number of values, the *median* is the mean of the two middle values.

Mode

The *mode* of a data set is the value or values that occur most often.

Remember

Data can have one mode, more than one mode, or no mode. When each value occurs only once, there is no mode.

EXAMPLE 1 Finding the Mean, Median, and Mode

Students' Hourly Wages	
$3.87	$7.25
$8.75	$8.45
$8.25	$7.25
$6.99	$7.99

An amusement park hires students for the summer. The students' hourly wages are given in the table. Find the mean, median, and mode of the hourly wages.

Mean: $\frac{\text{sum of the data}}{\text{number of values}} \rightarrow \frac{58.8}{8} = 7.35$

Median: 3.87, 6.99, 7.25, 7.25, 7.99, 8.25, 8.45, 8.75 — Order the data.

$\frac{15.24}{2} = 7.62$ — Mean of two middle values

Mode: 3.87, 6.99, 7.25, 7.25, 7.99, 8.25, 8.45, 8.75

The value 7.25 occurs most often.

∴ The mean is \$7.35, the median is \$7.62, and the mode is \$7.25.

On Your Own

Now You're Ready
Exercises 5–8

1. **WHAT IF?** In Example 1, the park hires another student at an hourly wage of \$6.99. How does this additional value affect the mean, median, and mode? Explain.

Laurie's Notes

Introduction

Connect

- **Yesterday:** Students explored the connection between the mean and fair distributions. They used a dot plot to display results. (MP2, MP4)
- **Today:** Students will explore how an outlier affects the three measures of central tendency.

Motivate

- Time to play M & M's! No, it's not the candy; it's mean, median, mode time.
- To help students think about the three measures, ask a series of questions. Give the results and have students come up with the data.
  Examples:
  - Name 3 different numbers with a mean of 10.
  - Name 3 different numbers with a mean of 10 and with a median of 12.
  - Name 5 different numbers with a mean of 10 and with a median of 10.
  - Name 5 numbers with a mean and median of 10, and with a mode of 8.
- Continue to ask questions. Knowing the number of values and the mean tells you the sum of all the data. The median tells you the middle value.

Lesson Notes

Key Ideas

- Write the definition of measure of central tendency, noting that the mean, median, and mode are all measures of central tendency.

Example 1

- "Have any of you had a summer or part-time job?" Answers will vary.
- Have a general discussion of different compensation methods: hourly, hourly plus tips, salaried, and commission.
- "Are there any observations about the wages listed?" Students may mention $3.87 as an outlier. Most wages are around $7 or $8 per hour.
- "What do you need to do to compute the mean?" Add the data and divide by 8.
- You may wish to have students use calculators in this lesson.
- **Common Error:** When finding the median, students forget to sort the data first. In this example, there are an even number of data values, so you need to sort and then find the mean of the middle two values.
- "Why do you think the mean might be less than the median?" Listen for the effect of the outlier, although students may not have a sense of this yet.

On Your Own

- Give time for students to actually compute the three measures.
- Discuss the results.

Goal Today's lesson is determining the effect an outlier has on the **measures of central tendency** for a data set.

Technology for the Teacher

Dynamic Classroom

Lesson Tutorials
Lesson Plans
Answer Presentation Tool

Extra Example 1

An amusement park hires students for the summer. The students' hourly wages are given in the table. Find the mean, median, and mode of the hourly wages.

Students' Hourly Wages	
$3.74	$7.75
$7.30	$8.43
$7.90	$7.83
$8.15	$8.50

mean: $7.45; median: $7.87; no mode

On Your Own

1. mean: $7.31, decreases; median: $7.25, decreases; Because the hourly wage of the student is less than the mean and median, both mean and median decrease. modes: $6.99 and $7.25; The data set now has two modes instead of one mode.

Extra Example 2

Identify the outlier in Extra Example 1. How does the outlier affect the mean, median, and mode? When you remove the outlier, $3.74, the mean increases $7.98 − $7.45 = $0.53, the median increases $7.90 − $7.87 = $0.03, and there still is no mode.

Extra Example 3

In Extra Example 1, the park increases each hourly wage by $0.30. How does this increase affect the mean, median, and mode? The mean and median both increase by $0.30, and there still is no mode.

On Your Own

2. $4\frac{1}{5}$ mi; When you remove the outlier, the mean decreases $1.85 - 1.38 = 0.47$ mile, the median decreases $1.45 - 1.4 = 0.05$ mile, and there is still no mode.
3. The mean and median both increase by $1\frac{1}{2}$ miles. There is still no mode.

Differentiated Instruction

Vocabulary

Have students add a glossary to their math notebook. Key vocabulary words should be added as they are introduced. Vocabulary words for this lesson are *measure of central tendency, outlier, mean, median,* and *mode*. Mean, median, and mode are often confusing to students. Drawing illustrations next to the vocabulary words will help in understanding and reinforcing their meanings.

Laurie's Notes

Example 2

- Remind students of what an outlier is.
- Before the three measures are computed, ask students to predict what they think will happen.
- Work through the computations in the example.
- Remind students to use the correct number of values when computing the mean after removing an outlier.

? "Why did the mean increase in this example?" The data value eliminated was much less than the mean. So, the sum was not affected much and you divide by a lesser number.

? "The median increased in this example. Could it have decreased or stayed the same? Explain." yes; Depending on the middle of the data set, the median can increase or stay the same if the outlier is a low value, *or* the median can decrease or stay the same if the outlier is a high value.

- Discuss students' predictions and the actual results.

Example 3

? "What impact will there be if everyone receives a $0.40 raise?" Answers will vary. Encourage students to reason about the problem.

- Work through the example.
- Discuss students' predictions and the actual results.
- **Big Idea:** When the same amount is added to each data value, the three measures increase by that same amount.
- **MP2 Reason Abstractly and Quantitatively:** If time permits, explore what happens when the hourly wage increases by 10%. Students need to understand the difference between an additive change to all values in the data set and a multiplicative change to all values in the data set.

On Your Own

- These questions integrate a review of fraction operations.
- For Question 3, check to see if students simply add $1\frac{1}{2}$ to each of their previous answers or if they compute the mean, median, and mode again.

Closure

- Explain the effect of an outlier on each of the three measures of central tendency.

EXAMPLE 2 Removing an Outlier

Remember

An *outlier* is a data value that is much greater than or much less than the other values.

Identify the outlier in Example 1. How does the outlier affect the mean, median, and mode?

The value \$3.87 is low compared to the other wages. It is the outlier.

Find the mean, median, and mode without the outlier.

Mean: $\frac{54.93}{7} \approx 7.85$

Median: 6.99, 7.25, 7.25, 7.99, 8.25, 8.45, 8.75 — The middle value, 7.99, is the median.

Mode: 6.99, 7.25, 7.25, 7.99, 8.25, 8.45, 8.75 — The mode is 7.25.

When you remove the outlier, the mean increases \$7.85 − \$7.35 = \$0.50, the median increases \$7.99 − \$7.62 = \$0.37, and the mode is the same.

EXAMPLE 3 Changing the Values of a Data Set

In Example 1, each hourly wage increases by \$0.40. How does this increase affect the mean, median, and mode?

Students' Hourly Wages	
\$4.27	\$7.65
\$9.15	\$8.85
\$8.65	\$7.65
\$7.39	\$8.39

Make a new table by adding \$0.40 to each hourly wage.

Mean: $\frac{62}{8} = 7.75$

Median: 4.27, 7.39, 7.65, 7.65, 8.39, 8.65, 8.85, 9.15 — Order the data.

$\frac{16.04}{2} = 8.02$ — Mean of two middle values

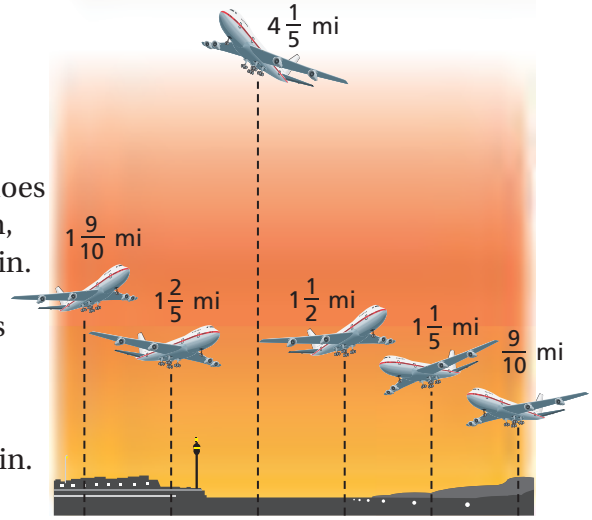
Mode: 4.27, 7.39, 7.65, 7.65, 8.39, 8.65, 8.85, 9.15 — The mode is 7.65.

When each hourly wage increases by \$0.40, the mean, median, and mode all increase by \$0.40.

On Your Own

Now You're Ready
Exercises 14–19

The figure shows the altitudes of several airplanes.



2. Identify the outlier. How does the outlier affect the mean, median, and mode? Explain.

3. Each airplane increases its altitude by $1\frac{1}{2}$ miles. How does this affect the mean, median, and mode? Explain.

12.1 Exercises

Check It Out
Help with Homework
BigIdeasMath.com

Vocabulary and Concept Check

1. **VOCABULARY** Describe the measures of central tendency of a data set.
2. **OPEN-ENDED** Create a data set that has more than one mode.
3. **WRITING** Describe how removing an outlier from a data set affects the mean of the data set.

Practice and Problem Solving

4. Draw a dot plot of the data. Then find the mean, median, and mode of the data.

Bag	1	2	3	4	5	6	7	8	9
Strawberries	10	13	11	15	8	14	7	11	12

Find the mean, median, and mode of the data.

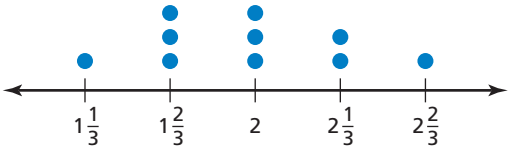
5.

Golf Scores		
3	−2	1
6	4	−1
−3	−1	2

6.

Changes in Stock Value (dollars)			
1.05	2.03	−1.78	−2.41
−2.64	0.67	4.02	1.39
0.66	−0.38	−3.01	2.20

7.



Movie lengths (hours)

8. **Available Memory (megabytes)**

Stem	Leaf
6	5
7	0 5 5
8	0 4 5
9	4

Key: 7|5 = 75 megabytes

Find the value of x.

9. Mean is 6; 2, 8, 9, 7, 6, x
10. Mean is 0; 11.5, 12.5, −10, −7.5, x
11. Median is 14; 9, 10, 12, x, 20, 25
12. Median is 51; 30, 45, x, 100

13. **POLAR BEARS** The table shows the masses of polar bears. Find the value of x when the mean is 410 kilograms.

Masses (kilograms)			
455	262	471	358
364	553	352	x

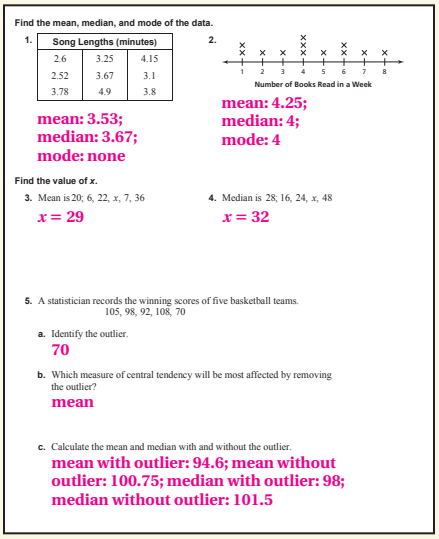
Assignment Guide and Homework Check

Level	Assignment	Homework Check
Average	1–3, 5–19 odd, 14, 22–24	7, 13, 14, 17
Advanced	1–3, 6–20 even, 15, 21, 22–24	12, 14, 20, 21

Common Errors

- **Exercises 5–8** When finding the mean, students may forget to divide by the total number of data values and instead divide by the maximum value. Remind them that the definition of mean is an "average," so they must take into account the total number of items or numbers to get an average. Explain to students that it is as if they are dividing the total evenly among the number of groups.
- **Exercises 5–8** Students may try to identify the median without ordering the data first. Remind them that it is essential to order the data first and then find the median. This also makes finding the mode easier.
- **Exercises 9–12** Students may not know how to find the missing value. Remind them of the definition of mean and median. Encourage students to use these definitions to write an equation to find the value of x.

12.1 Record and Practice Journal



Find the mean, median, and mode of the data.

1.

Song Lengths (minutes)		
2.6	3.25	4.15
2.52	3.67	3.1
3.78	4.9	3.8

mean: 3.53; median: 3.67; mode: none

2. Number of Books Read in a Week

mean: 4.25; median: 4; mode: 4

Find the value of x.

3. Mean is 20; 6, 22, x, 7, 36

$x = 29$

4. Median is 28; 16, 24, x, 48

$x = 32$

5. A statistician records the winning scores of five basketball teams. 105, 98, 92, 108, 70

a. Identify the outlier.

70

b. Which measure of central tendency will be most affected by removing the outlier?

mean

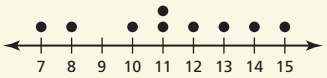
c. Calculate the mean and median with and without the outlier.

mean with outlier: 94.6; mean without outlier: 100.75; median with outlier: 98; median without outlier: 101.5

Vocabulary and Concept Check

1. The mean is the sum of the data divided by the number of data values. The median of an odd number of values is the middle value. The median of an even number of values is the mean of the two middle values. The mode is the value or values that occur most often.
2. *Sample answer:* 1, 2, 2, 3, 6, 8, 8, 9, 12
3. If the outlier is greater than the mean, removing it will decrease the mean. If the outlier is less than the mean, removing it will increase the mean.

Practice and Problem Solving

4. 

mean: $11.\overline{2}$; median: 11
mode: 11

5. mean: 1; median: 1
mode: -1
6. mean: \$0.15; median: \$0.665
mode: none
7. mean: $1\frac{29}{30}$ h
median: 2 h
modes: $1\frac{2}{3}$ h and 2 h
8. mean: 78.5 MB
median: 77.5 MB
mode: 75 MB
9. 4
10. -6.5
11. 16
12. 57
13. 465

Practice and Problem Solving

14. **a.** 6 in rookie season
14 this season

b. mean for both seasons

c. The mean increased by about 5.3; The median increased by 5; the mode increased from 0 and 2 to 4.

15. See *Taking Math Deeper*.

16. All measures increase by 3.

17. All measures increase by k.

18. The mean and median both decrease by \$0.05 and there is still no mode.

19. All measures decrease by k.

20. **a.** mean: 19.37 yr
median: 19 yr
mode: 18 yr

b. 37 yr; The mean decreases about $19.37 - 19.19 = 0.18$ year. The median and mode stay the same.

21. See Additional Answers.

Fair Game Review

22. $-8, -5, -3, 1, 4, 7$

23. $-4.7, -2.8, -\frac{2}{3}, 1.2, \frac{3}{2}, 5.4$

24. B

Mini-Assessment

Find the mean, median, and mode of the data.

1. 10, −4, 3, −1, 12 4, 3, no mode

2. 1.25, 3.80, −0.65, −2.40 0.5, 0.3, no mode

3. 5, 15, 8, 13, 10, 8, 6, 4, 12 9, 8, 8

Find the value of x.

4. Mean is 2; −4, −2, 3, x, 9 4

5. Median is 16.5; 8, 11, x, 18, 24, 26 15

Taking Math Deeper

Exercise 15

The Appalachian Trail is the longest marked trail in the U.S. at 2178 miles. The 11 shelters in this problem are in Massachusetts.

1 Order the data.

0.1, 1.8, 3.3, 5.3, 6.3, 8.8, 8.8, 14, 14.3, 16.7

2 Find the mean, median, and mode.

a. Mean $= \frac{79.4}{10} = 7.94$ miles

Median $= \frac{(6.3 + 8.8)}{2} = 7.55$ miles

The mode is 8.8 miles.

3 A hiker begins the trail at Shelter 2 and therefore skips the 0.1-mile distance of the trail. The mean, median, and mode of the remaining distances are as follows.

1.8, 3.3, 5.3, 6.3, 8.8, 8.8, 14, 14.3, 16.7

Find the mean, median, and mode.

b. Mean $= \frac{79.3}{9} \approx 8.8$ miles

Median $= 8.8$ miles

The mode is 8.8 miles.

The mean increases by about 0.86 mile.
The median increases by 1.25 miles.
The mode does not change.

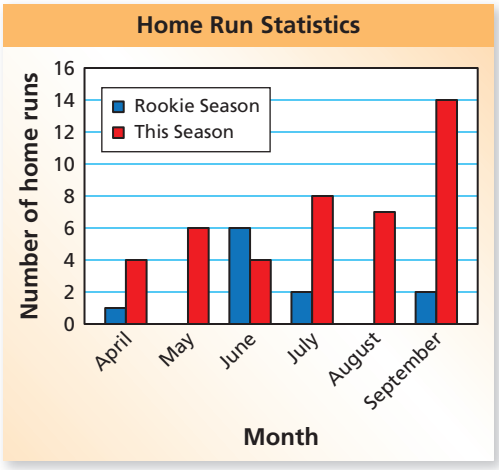
Project

Plan a hiking trip. List the things you need to take with you. Include the approximate amount of time you think it will take.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Resources by Chapter • Practice A and Practice B • Puzzle Time Record and Practice Journal Practice Differentiating the Lesson Lesson Tutorials Skills Review Handbook	Resources by Chapter • Enrichment and Extension Start the next section

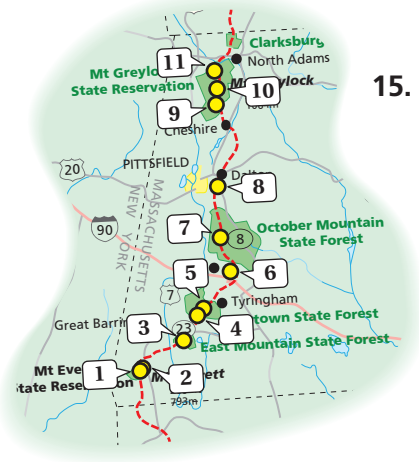
14. BASEBALL The graph shows a player's monthly home run totals in two seasons.



a. Identify the outlier in each season.

b. Which measure of central tendency is most affected by removing the outlier in each season?

c. Compare the means, medians, and modes of the home run totals in the two seasons.



15. TRAIL The map shows the locations of 11 shelters along the Appalachian Trail. The distances (in miles) between these shelters are 0.1, 14.3, 5.3, 1.8, 14, 8.8, 8.8, 16.7, 6.3, and 3.3.

a. Find the mean, median, and mode of the distances.

b. A hiker starts at Shelter 2 and hikes to Shelter 11. How does this affect the mean, median, and mode? Explain.

In Exercises 16–19, explain how the change affects the mean, median, and mode.

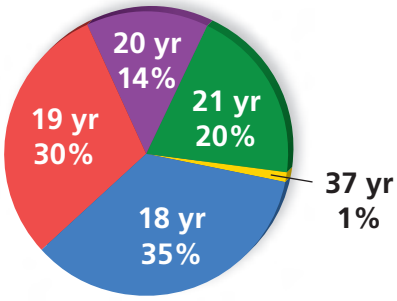
16. In Exercise 4, you add 3 strawberries to each bag.

17. You add a number k to each value in a data set.

18. In Exercise 6, the value of each stock decreases by \$0.05.

19. You subtract a number k from each value in a data set.

College Student Ages



20. COLLEGE The circle graph shows the distribution of the ages of 200 students in a college psychology class.

a. Find the mean, median, and mode of the students' ages.

b. Identify the outliers. How do the outliers affect the mean, median, and mode?

21. Reasoning The mean and median hourly wage at a bagel shop is \$7.20. Hourly wages at the bagel shop increase by 10%. Where are you likely to have a greater hourly wage, at the bagel shop or at the amusement park in Example 1? Explain.

Fair Game Review What you learned in previous grades & lessons

Order the values from least to greatest. *(Skills Review Handbook)*

22. $1, -3, -8, 4, 7, -5$

23. $1.2, -2.8, \frac{3}{2}, 5.4, -4.7, -\frac{2}{3}$

24. MULTIPLE CHOICE Which equation represents a linear function? *(Section 5.5)*

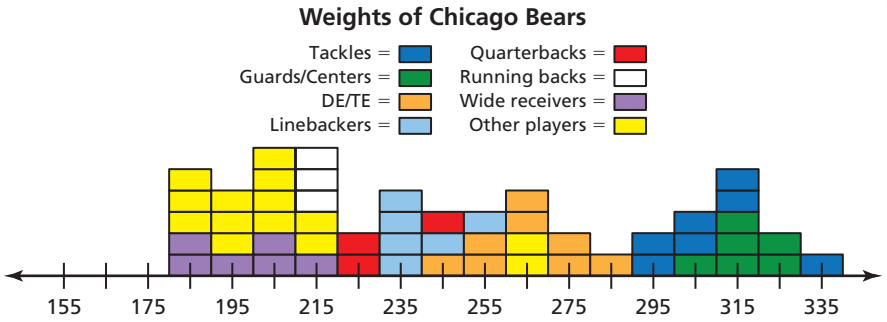
Ⓐ $y = x^2$ Ⓑ $y = 2x$ Ⓒ $y = \frac{2}{x}$ Ⓓ $xy = 2$

12.2 Measures of Dispersion

Essential Question How can you measure the dispersion of a data set?

1 ACTIVITY: Measuring the Dispersion of Data

Work with a partner. The diagram shows the weights of 53 players on the Chicago Bears football team in 2011.



The weights are: 220, 200, 185, 240, 215, 222, 185, 180, 210, 196, 218, 190, 218, 185, 204, 180, 200, 219, 198, 196, 260, 211, 203, 239, 234, 258, 244, 230, 320, 310, 265, 309, 315, 295, 315, 275, 316, 333, 320, 308, 206, 200, 255, 267, 260, 287, 292, 300, 248, 310, 252, 238, 270.

a. Describe the data. How much are the weights dispersed from the mean weight? Explain your reasoning.

> **Definition of Dispersed:** To disperse objects means to spread them over an area. For instance, the population of Texas is much more dispersed than the population of Rhode Island.

b. Does it appear that the weight of a football player is correlated to the position that he plays? Explain your reasoning. Do you think your answer is valid for other types of professional sports, such as basketball, baseball, hockey, and soccer? Explain your reasoning.

Common Core

Measures of Dispersion

In this lesson, you will
- find ranges of data sets.
- compare spreads of data sets.
- find standard deviations of data sets.

Learning Standards
S.ID.2
S.ID.3

Laurie's Notes

Introduction

Standards for Mathematical Practice

- **MP1a Make Sense of Problems** and **MP3 Construct Viable Arguments and Critique the Reasoning of Others:** Having students explain their reasoning and having whole class discussions helps all students make sense of problems. It also helps them develop an ability to form arguments and critique the reasoning of others.

Motivate

- Play a game of "Did You Know?"
  - Walter Camp is called the "Father of American Football."
  - The game originated from the game of rugby.
  - The American Professional Football Association began in 1920 at a car dealership in Canton, Ohio. This association would later become the National Football League (NFL).
  - The Green Bay Packers won the first two Super Bowls in 1967 and 1968. It took them almost 30 years to win the Super Bowl again.
  - A football is 11 inches long and weighs 14 to 15 ounces.

Activity Notes

Activity 1

- Review the differences between a histogram and a bar graph.
- Students should discuss the histogram with their partners. This histogram involves an additional element of color that is important to the problem.
- A calculator may be helpful in finding the mean weight.
- When answering part (a), students may describe only the dispersion of the two extremes values (180 and 333) from the mean. Others may talk about the clusters of numbers above and below the mean. Explain to students that they will learn different ways to measure dispersion.

? "In part (b), what does 'correlated' mean in the context of this problem?"
Answers will vary. Students may use phrases such as "associated with" or "connected to."

- Students with knowledge of football will be able to discuss the physical attributes of different positions. Even if students have little knowledge of football, the colors used in the histogram should provide the information needed to help explain the correlation. The colors are clustered, with no symmetry.
- Students with greater knowledge about professional sports will be able to discuss the second part of this activity.

Common Core State Standards

S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Previous Learning

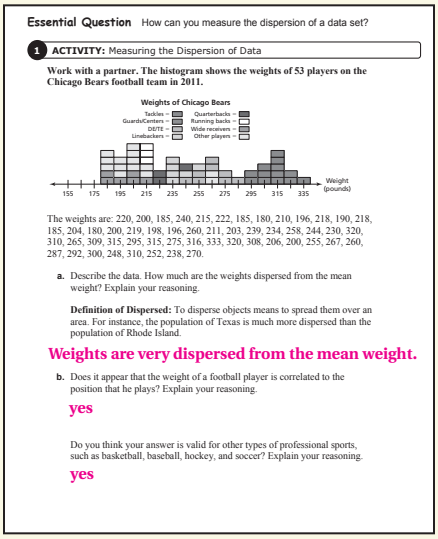
Students should know how to find the mean, median, and mode of a data set.

Technology for the Teacher

Dynamic Classroom

Lesson Plans
Complete Materials List

12.2 Record and Practice Journal



Essential Question How can you measure the dispersion of a data set?

1 ACTIVITY: Measuring the Dispersion of Data

Work with a partner. The histogram shows the weights of 53 players on the Chicago Bears football team in 2011.

Weights of Chicago Bears

Tackles
Guards/Centers
QB/TE
Linebackers
Quarterbacks
Running backs
Wide receivers
Other players

155 175 195 215 235 255 275 295 315 335 Weight (pounds)

The weights are: 220, 200, 185, 240, 215, 222, 185, 180, 210, 196, 218, 190, 218, 185, 204, 180, 200, 219, 198, 196, 260, 211, 203, 239, 234, 258, 244, 230, 320, 310, 265, 309, 315, 295, 315, 275, 316, 333, 320, 308, 206, 200, 255, 267, 260, 287, 292, 300, 248, 310, 252, 238, 270.

a. Describe the data. How much are the weights dispersed from the mean weight? Explain your reasoning.

Definition of Dispersed: To disperse objects means to spread them over an area. For instance, the population of Texas is much more dispersed than the population of Rhode Island.

Weights are very dispersed from the mean weight.

b. Does it appear that the weight of a football player is correlated to the position that he plays? Explain your reasoning.

yes

Do you think your answer is valid for other types of professional sports, such as basketball, baseball, hockey, and soccer? Explain your reasoning.

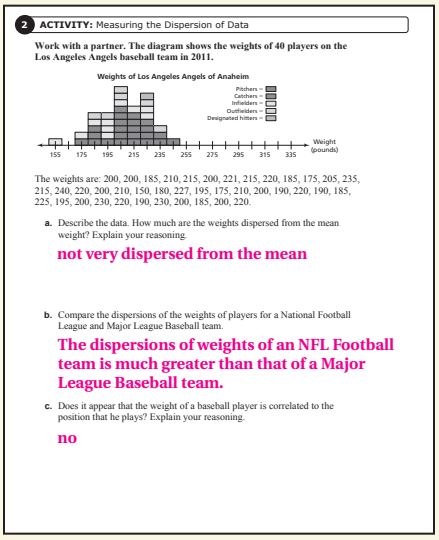
yes

Differentiated Instruction

Advanced

Challenge students to create two data sets with the same mean and median, but with a different range and standard deviation. Have them discuss the differences in the data sets and how measures of dispersion aid in describing a data set.

12.2 Record and Practice Journal



2 ACTIVITY: Measuring the Dispersion of Data

Work with a partner. The diagram shows the weights of 40 players on the Los Angeles Angels baseball team in 2011.

Weights of Los Angeles Angels of Anaheim

Pitchers =
Catchers =
Infielders =
Outfielders =
Designated hitters =

155 175 195 215 235 255 275 295 315 335 Weight (pounds)

The weights are: 200, 200, 185, 210, 215, 200, 221, 215, 220, 185, 175, 205, 235, 215, 240, 220, 200, 210, 150, 180, 227, 195, 175, 210, 200, 190, 220, 190, 185, 225, 195, 200, 230, 220, 190, 230, 200, 185, 200, 220.

a. Describe the data. How much are the weights dispersed from the mean weight? Explain your reasoning.

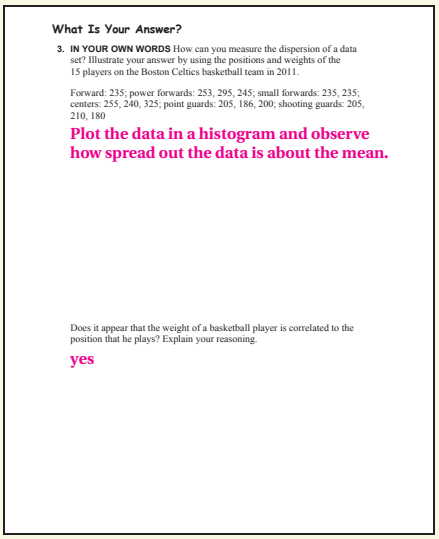
not very dispersed from the mean

b. Compare the dispersions of the weights of players for a National Football League and Major League Baseball team.

The dispersions of weights of an NFL Football team is much greater than that of a Major League Baseball team.

c. Does it appear that the weight of a baseball player is correlated to the position that he plays? Explain your reasoning.

no



What Is Your Answer?

3. **IN YOUR OWN WORDS** How can you measure the dispersion of a data set? Illustrate your answer by using the positions and weights of the 15 players on the Boston Celtics basketball team in 2011.

Forward: 235; power forwards: 253, 295, 245; small forwards: 235, 235; centers: 255, 240, 325; point guards: 205, 186, 200; shooting guards: 205, 210, 180

Plot the data in a histogram and observe how spread out the data is about the mean.

Does it appear that the weight of a basketball player is correlated to the position that he plays? Explain your reasoning.

yes

Laurie's Notes

Activity 2

- This activity is similar to Activity 1 except the sport has changed. The discussion following Activity 1 helps all students to successfully complete this activity.

? "What do you notice about the colors in this histogram?" They are more symmetric, or evenly spread out.

- When comparing the dispersions of data in each activity, students should notice that the baseball data are less spread out.

What Is Your Answer?

- Provide colored pencils so students can construct a histogram of the data. If colored pencils are not available, the positions could be coded using letters.
- When comparing the dispersions of data, students should notice that the basketball data are similar to the football data in that there is a correlation between position and weight.

Closure

- Suppose you have data for the weights of players on a soccer team. When comparing the dispersions of data, which data set do you think it would be similar to? Explain your reasoning. It would most likely be similar to the basketball or baseball data.

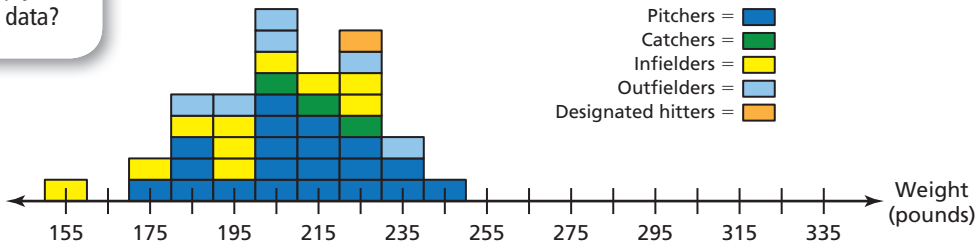
2 ACTIVITY: Measuring the Dispersion of Data

Math Practice 4

Analyze Relationships

How can the diagram help you analyze the data?

Work with a partner. The diagram shows the weights of 40 players on the Los Angeles Angels baseball team in 2011.



The weights are: 200, 200, 185, 210, 215, 200, 221, 215, 220, 185, 175, 205, 235, 215, 240, 220, 200, 210, 150, 180, 227, 195, 175, 210, 200, 190, 220, 190, 185, 225, 195, 200, 230, 220, 190, 230, 200, 185, 200, 220.

a. Describe the data. How much are the weights dispersed from the mean weight? Explain your reasoning.

b. Compare the dispersions of the weights of players for a National Football League team and a Major League Baseball team.

c. Does it appear that the weight of a baseball player is correlated to the position that he plays? Explain your reasoning.

What Is Your Answer?

3. IN YOUR OWN WORDS How can you measure the dispersion of a data set? Illustrate your answer by using the positions and weights of the 15 players on the Boston Celtics basketball team in 2011.

Forward: 235; power forwards: 253, 295, 245; small forwards: 235, 235; centers: 255, 240, 325; point guards: 205, 186, 200; shooting guards: 205, 210, 180

Does it appear that the weight of a basketball player is correlated to the position that he plays? Explain your reasoning.

Practice Use what you learned about measuring the dispersion of data to complete Exercises 3 and 4 on page 616.

12.2 Lesson

Check It Out
Lesson Tutorials
BigIdeasMath.com

Key Vocabulary
measure of dispersion, *p. 614*
range, *p. 614*
standard deviation, *p. 615*

A **measure of dispersion** is a measure that describes the spread of a data set. The simplest measure of dispersion is the range. The **range** of a data set is the difference between the greatest value and the least value.

EXAMPLE 1 Finding the Range

Two reality cooking shows select 12 contestants each. The ages of the contestants are shown in the tables. Find the mean and range of the ages for each show. Compare your results.

Show A	
Ages	
20	29
19	22
25	27
27	29
30	20
21	31

Show B	
Ages	
25	19
20	27
22	25
27	22
48	21
32	24

Show A: $\text{mean} = \frac{300}{12} = 25$

Ordering the data can help you find the least and greatest ages.

19, 20, 20, 21, 22, 25, 27, 27, 29, 29, 30, 31 Order the data.

The least value is 19. The greatest value is 31.

So, the range is $31 - 19$, or 12 years.

Show B: $\text{mean} = \frac{312}{12} = 26$

19, 20, 21, 22, 22, 24, 25, 25, 27, 27, 32, 48 Order the data.

The least value is 19. The greatest value is 48.

So, the range is $48 - 19$, or 29 years.

The mean ages for the shows, 25 and 26, are about the same. The range of the ages for Show A is 12 years and the range for Show B is 29 years. So, the ages for Show B are more spread out.

On Your Own

Now You're Ready
Exercises 5 and 6

1. After the first week, the 25-year-old is voted off Show A. The 48-year-old is voted off Show B. How does this affect the mean and range of the remaining contestants on each show? Explain.

Laurie's Notes

Introduction

Connect

- **Yesterday:** Students explored how to measure the dispersion of a data set. (MP1a, MP3)
- **Today:** Students will find the range and standard deviation of a data set and compare two data sets.

Motivate

- **Review:** Write two data sets on the board and ask students to find the mean, median, and mode of each data set.
  Example: Data Set 1: 85, 89, 90, 91, 95 **Data Set 2:** 0, 50, 100, 100, 200
- Discuss how the data sets are alike (5 data values in each set, mean = 90) and how they are different (median, mode, and spread).

Lesson Notes

Discuss

- Define measure of dispersion and range. Refer back to the data sets in the activity.
- **Common Error:** When finding the range, where the least value is 180 and the greatest value is 333, students may write 180 – 333, meaning 180 to 333. Stress that the range is a single number. It is the *difference* between two numbers.

Example 1

- Ask if any students watch cooking shows. Do any of your students have an interest in cooking?
- If time permits, students could begin by constructing a dot plot or line plot of the data. The shape of the distributions is important in this example.
- ? "How do you find the mean?" Find the sum of the values in the data set and divide by the number of data values.
- Remind students that we use the word "mean" instead of "average" because there are different types of averages—median and mode are both averages.
- Compute the mean and range for each data set.
- The means are about the same but the ranges are not. When concluding that the ages in Show B are more spread out, some students might observe that it is only one piece of data that is significantly different.
- ? "What is an outlier?" A data value that is significantly different than the rest of the data values.
- Students may say that if you ignore the 48-year-old contestant, the ranges would be essentially the same. The effect of the older contestant is mentioned in the On Your Own.

On Your Own

- It is important for students to understand that an outlier affects the range *and* the mean.

Goal Today's lesson is finding the **range** and **standard deviation** of a data set.

Technology for the Teacher

Dynamic Classroom

Lesson Tutorials
Lesson Plans
Answer Presentation Tool

Extra Example 1

The bowling averages for two teams are shown in the tables. Find the mean and range of the bowling averages for each team. Compare your results.

Team A	
185	167
221	205
194	190
208	214

Team B	
182	172
168	170
175	195
190	188

The mean for Team A, 198, is greater than the mean for Team B, 180. The range for Team A, 54, is greater than the range for Team B, 27. The bowling averages for Team A are greater but more spread out than Team B.

On Your Own

1. For Show A, the mean and range remain the same. For Show B, the mean changes from 26 to 24 and the range changes from 29 to 13. The mean and range do not change for Show A because the removed data value is equal to the mean of the data set. The mean and range change for Show B because the removed data value, 48, is an outlier.

Laurie's Notes

Key Idea

- Point out the disadvantage of using the range to describe a data set. It is easily calculated but does not adequately describe how data are spread out. The standard deviation uses all of the data values in its computation.
- Students may be overwhelmed by this formula. Remind them they can use a calculator to find a standard deviation. By looking at its calculation, students will better understand what a standard deviation represents.
- Explain the formula by saying that after the mean is calculated, it is subtracted from each data value. The difference is positive for values greater than the mean and negative for values less than the mean. These differences are squared making them all positive. The sum of all the squared differences is divided by n, finding the 'mean' of the sum of squared differences. The square root of this mean is then calculated.
- Refer to the sets in the Motivate to describe data that are clustered about the mean (Data Set 1) and data that are more spread out (Data Set 2).

Example 2

- **MP1b Persevere in Solving Problems:** Computing the standard deviation by hand requires perseverance.
- **MP5 Use Appropriate Tools Strategically:** Using a table to organize results is an appropriate tool that helps ensure accurate completion of the problem.
- Write $\bar{x} = 25$ and say, "The mean, denoted by x-bar, is 25."
- ? "How do you find the mean of column 4?" Add up the 12 values and divide by 12.
- Interpret the meaning of the answer by saying, "The typical age of a contestant on Show A differs from the mean by about 4.2 years."
- **Big Idea:** A standard deviation of 4.2 years must be viewed in the context of a data set where the mean is 25. A standard deviation of 4.2 years can be looked at differently if the data set has a mean of 8 or 80.

On Your Own

- Discuss the effect of the outlier for Show B.

Closure

- Refer to the sets in the Motivate. Each set has a mean of 90. The standard deviation of Data Set 1 is 3.2 and for Data Set 2, it is 66.3. Give five data values that have a mean of 90 and a standard deviation between 3.2 and 66.3. Explain why you believe you are correct. Answers will vary.

Extra Example 2

Find the standard deviation of the bowling averages for Team A in Extra Example 1. Use a table to organize your work. Interpret your result.
The standard deviation is about 16.4.This means that the typical bowling average on Team A differs from the mean by about 16.4 pins.
See Additional Answers for table.

On Your Own

2. about 7.5; This means that the typical age of a contestant on Show B differs from the mean by about 7.5 years.
3. The standard deviation for Show B is greater than that of Show A. So, the ages of the contestants on Show B are more spread out than the ages of the contestants on Show A.

English Language Learners

Visual Glossary
To help English language learners with the vocabulary and understanding the concepts, create posters for each of the words: *mean*, *median*, *mode*, *range*, and *standard deviation*. Identify whether it is a *measure of central tendency* or a *measure of dispersion*. Include the description, the formula used to find the measure, and an example.

A disadvantage of using the range to describe the spread of a data set is that its calculation uses only two data values. A measure of dispersion that uses all the values of a data set is the *standard deviation*.

Key Idea

Standard Deviation

The **standard deviation** of a data set is a measure of how much a typical value in the data set differs from the mean. It is given by

$$\text{standard deviation} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}$$

where n is the number of values in the data set. The symbol $\bar{x}$ represents the mean. It is read as "x-bar."

A small standard deviation means that the data are clustered around the mean. A large standard deviation means that the data are more spread out.

Remember

The notation consisting of three dots (⋯) indicates that a pattern continues.

EXAMPLE 2 Finding the Standard Deviation

x	$\bar{x}$	$x - \bar{x}$	$(x - \bar{x})^2$
20	25	−5	25
29	25	4	16
19	25	−6	36
22	25	−3	9
25	25	0	0
27	25	2	4
27	25	2	4
29	25	4	16
30	25	5	25
20	25	−5	25
21	25	−4	16
31	25	6	36

Find the standard deviation of the ages for Show A in Example 1. Use a table to organize your work. Interpret your result.

Step 1: Find the mean. From Example 1, the mean is 25.

Step 2: Find the difference between each data value and the mean, $x - \bar{x}$.

Step 3: Square each difference from Step 2, $(x - \bar{x})^2$.

Step 4: Find the mean of the squares from Step 3.

$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} = \frac{25 + 16 + \cdots + 36}{12} \approx 17.7$$

Step 5: Use a calculator to find the square root.

$$\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}} = \sqrt{17.7} \approx 4.2$$

∴ The standard deviation is 4.2. This means that the typical age of a contestant on Show A differs from the mean by about 4.2 years.

On Your Own

Now You're Ready
Exercises 7–12

2. Find the standard deviation of the ages for Show B in Example 1. Interpret your result.
3. Compare the standard deviations for Show A and Show B. What can you conclude?

12.2 Exercises

Check It Out
Help with Homework
BigIdeasMath.com

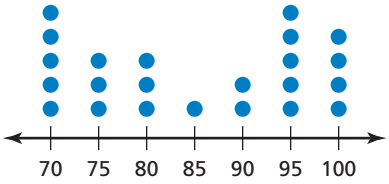
Vocabulary and Concept Check

1. **VOCABULARY** In a data set, what does a measure of central tendency represent? What does a measure of dispersion represent?
2. **REASONING** What is an advantage of using the range to describe a data set? Why do you think the standard deviation is considered a more reliable measure of dispersion than the range?

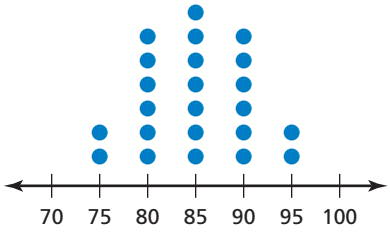
Practice and Problem Solving

Describe the data. How much are the data dispersed from the mean? Explain your reasoning.

3.



4.

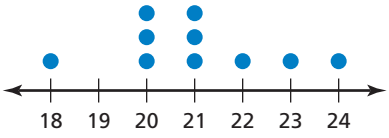


Find the mean and range of each data set. Then compare the data sets.

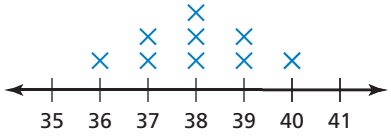
5. Heights (in inches) of two teams
   Tigers: 67, 70, 65, 72, 74, 68, 67, 69
   Centaurs: 74, 71, 68, 63, 75, 63, 65, 73
6. Numbers of fish caught during a week
   Crew A: 120, 100, 75, 112, 135, 80, 106
   Crew B: 104, 140, 159, 135, 158, 165, 140

Find the mean and standard deviation of the data.

7. 4, 2, 7, 3, 6, 5, 5, 8
8. 12, 4, 8, 7, 9, 13, 10
9.



10.

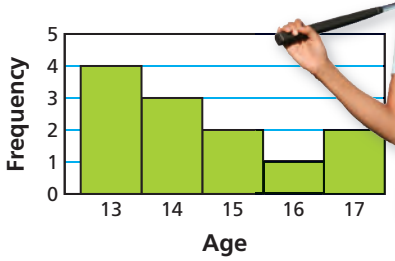


11.

Stem	Leaf
4	0
5	2
6	1 4 5 7
7	3
8	2

Key: 6|1 = 61

12.



13. **GOLF** The scores for two golfers are shown.

Kirsten		Leah	
83	88	89	87
84	95	93	95
91	89	92	94
90	87	88	91
98	95	89	92

   a. Find the mean, range, and standard deviation of the scores for each golfer. Compare your results.
   b. Which golfer do you think is more consistent? Explain.

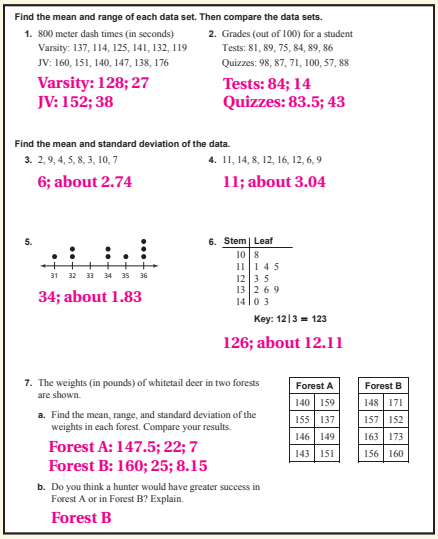
Assignment Guide and Homework Check

Level	Assignment	Homework Check
Average	1–4, 5–17 odd, 14, 18, 22–25	5, 11, 13, 14
Advanced	1–4, 6–16 even, 17–19, 21, 22–25	6, 12, 18, 19

Common Errors

- **Exercises 5 and 6** Students may subtract the first data value from the last data value to find the range. Remind students that the range is the difference between the greatest and least data values. It helps to order the data before identifying the greatest and least data values.
- **Exercises 7–12** Students may forget to square the differences between x_i and the mean when calculating the standard deviation. The standard deviation formula is simply finding the average distance each data point is from the data set mean. Tell the students that squaring the differences is a mathematical way of making all of the distances positive, similar to the distance formula.

12.2 Record and Practice Journal



Find the mean and range of each data set. Then compare the data sets.

1. 800 meter dash times (in seconds)
Varsity: 137, 114, 125, 141, 132, 119
JV: 160, 151, 140, 147, 138, 176
Varsity: 128; 27
JV: 152; 38

2. Grades (out of 100) for a student
Tests: 81, 89, 75, 84, 89, 86
Quizzes: 98, 87, 71, 100, 57, 88
Tests: 84; 14
Quizzes: 83.5; 43

Find the mean and standard deviation of the data.

3. 2, 9, 4, 5, 8, 3, 10, 7
6; about 2.74

4. 11, 14, 8, 12, 16, 12, 6, 9
11; about 3.04

5. 31 32 33 34 35 36
34; about 1.83

6.

Stem	Leaf
10	8
11	1 4 5
12	3 5
13	2 6 9
14	0 3

Key: 12 | 3 = 123

126; about 12.11

7. The weights (in pounds) of whitetail deer in two forests are shown.

Forest A		Forest B	
140	159	148	171
155	137	157	152
146	149	163	173
143	151	156	160

a. Find the mean, range, and standard deviation of the weights in each forest. Compare your results.
Forest A: 147.5; 22; 7
Forest B: 160; 25; 8.15

b. Do you think a hunter would have greater success in Forest A or in Forest B? Explain.
Forest B

Vocabulary and Concept Check

1. A measure of central tendency is a value that represents a typical value in a data set. A measure of dispersion is a value that measures how spread out a data set is.
2. The range is quick and easy to calculate, but its calculation uses only two data values from a data set and it is greatly affected by outliers. The standard deviation uses all of the data values in its calculation and is less affected by outliers compared to the range.

Practice and Problem Solving

3. The mean is 85. The majority of the data are in clusters on each side, far from the mean.
4. The mean is 85. The majority of the data are clustered around the mean.
5–6. See Additional Answers.
7. 5; about 1.9
8. 9; about 2.8
9. 21; about 1.6
10. 38; about 1.2
11. 63; about 11.9
12. 14.5; about 1.4
13. a. Kirsten: 90; 15; about 4.6; Leah: 91; 8; about 2.5; The means are about the same but Kirsten's range and standard deviation are much greater than Leah's range and standard deviation.
    b. Leah; Kirsten's scores are more spread out than Leah's scores.

Practice and Problem Solving

14. See Additional Answers.

15. 6.4; 8.2; about 3.0

16. −1; 8; about 2.7

17. See Additional Answers.

18. Data set (b) has the greatest standard deviation and data set (c) has the least standard deviation. From the dot plots, you can see that the data values of (b) are more spread out than (a) and (c). You can also see that the data values of (c) are clustered closely about the mean of 15.

19. See Additional Answers.

20. See *Taking Math Deeper*.

21. See Additional Answers.

Fair Game Review

22–24. See Additional Answers.

25. B

Mini-Assessment

Find the mean, range, and standard deviation of the data sets. Then compare the data sets.

1. Number of text messages

Student A	24	21	32	16	38	29	22
Student B	20	23	15	27	14	22	19

See Additional Answers.

2. Grade point averages

Team A	3.6	3.2	3.8	3.4	3.5
Team B	3.5	3.9	2.9	4.0	3.2

See Additional Answers.

Taking Math Deeper

Exercise 20

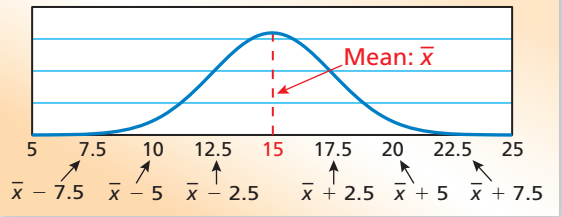
At first glance, students may think the answer is Doctor's Office A, because the mean time is greater than Doctor's Office B. Students need to consider the standard deviation for each data set. Let's take a graphical approach.

1 Sketch an approximate graph of each distribution. Label the mean at the center of each distribution and the standard deviations above and below the mean. Remember the data are evenly distributed.

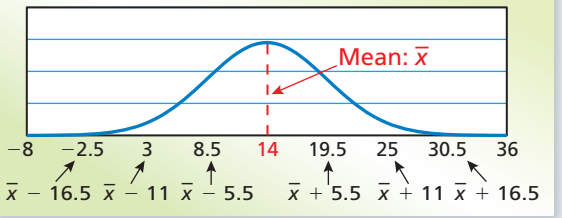
2 Interpret the graphs. From the graph for Office A, you can see that most of the waiting times are between 10 and 20 minutes. For Office B, most of the waiting times are between 3 and 25 minutes.

The standard deviation for Office B is greater, so the data are more spread out.

Waiting Times at Doctor's Office A



Waiting Times at Doctor's Office B



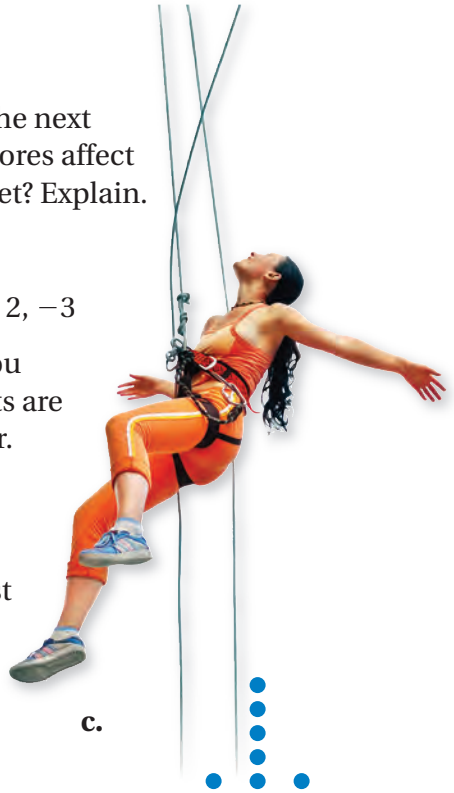
3 At Office A, a waiting time longer than 20 minutes is more than 2 standard deviations from the mean. At Office B, it is a little more than 1 standard deviation from the mean. So, you are more likely to wait longer than 20 minutes at Doctor's Office B.

Note: Explain to students that it is possible to spend more than 20 minutes waiting at either office, but the likelihood is greater at Office B.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Resources by Chapter • Practice A and Practice B • Puzzle Time Record and Practice Journal Practice Differentiating the Lesson Lesson Tutorials Skills Review Handbook	Resources by Chapter • Enrichment and Extension Start the next section

14. INCLUDING A VALUE In Exercise 13, Kirsten's score for the next round is 90, and Leah's is 80. How does each of these scores affect the mean, range, and standard deviation of each data set? Explain.



Find the mean, range, and standard deviation of the data.

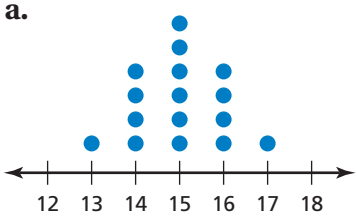
15. 4.1, 2.3, 8.7, 10.5, 6.4

16. $-2, 0, 1, -5, 3, -4, 2, -3$

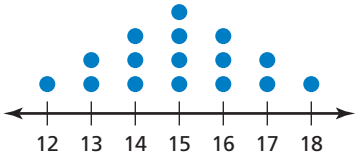
17. REASONING Two data sets have the same range. Can you assume that the standard deviations of the two data sets are about the same? Give an example to justify your answer.

18. ADVENTURE CLUB The dot plots show the ages of members of three different adventure clubs. Without performing calculations, which data set has the greatest standard deviation? Which has the least standard deviation? Explain your reasoning.

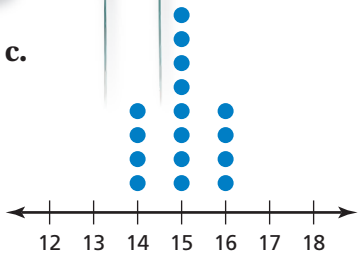
a.



b.



c.



19. PROJECT Measure the heights (in inches) of the students in your class.

a. Use a calculator to find the mean, range, and standard deviation of the heights.

b. A new student who is 7 feet tall joins your class. How would you expect this person's height to affect the mean, range, and standard deviation? Verify your answer.

20. WAITING TIMES The waiting times at two doctors' offices are described below. At which office are you more likely to wait longer than 20 minutes? Explain. Assume the mean is at the center of each distribution and the data are evenly distributed.

Doctor's Office A: mean = 15 minutes, standard deviation = 2.5 minutes

Doctor's Office B: mean = 14 minutes, standard deviation = 5.5 minutes

21. Critical Thinking Can the standard deviation of a data set be 0? Can it be negative? If so, give examples to justify your answers.

Fair Game Review What you learned in previous grades & lessons

Graph the function. Describe the domain and range. *(Section 11.2)*

22. $y = -\frac{3}{x}$

23. $y = \frac{1}{x - 6}$

24. $y = \frac{1}{x + 4} - 5$

25. MULTIPLE CHOICE Find the quotient $(x + 5) \div \frac{x^2 + 4x - 5}{x + 5}$. *(Section 11.4)*

(A) $\frac{x + 5}{x + 1}$ **(B)** $\frac{x + 5}{x - 1}$ **(C)** $x^2 + 4x - 5$ **(D)** $\frac{x + 5}{x - 5}$

12.3 Box-and-Whisker Plots

Essential Question How can you use a box-and-whisker plot to describe a data set?

1 ACTIVITY: Drawing a Box-and-Whisker Plot

Work with a partner.

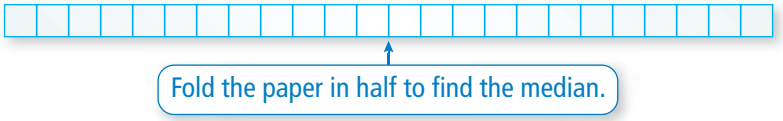
The numbers of first cousins of the students in an eighth-grade class are shown.

A box-and-whisker plot uses a number line to represent the data visually.

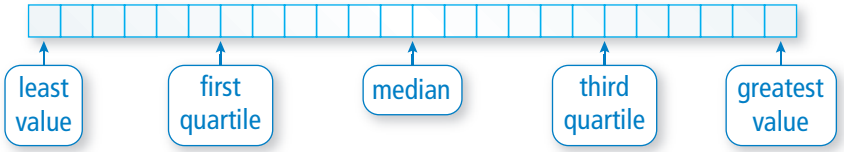
Numbers of First Cousins

3	10	18	8
9	3	0	32
23	19	13	8
6	3	3	10
12	45	1	5
13	24	16	14

a. Order the data set and write it on a strip of grid paper with 24 equally spaced boxes.

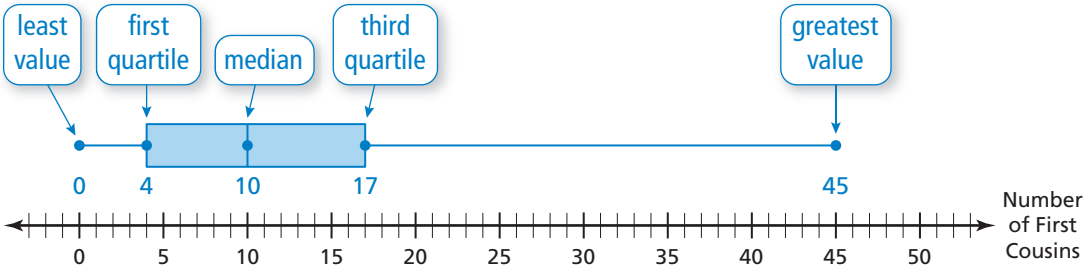


b. Fold the paper in half again to divide the data into four groups. Because there are 24 numbers in the data set, each group should have six numbers.



c. Draw a number line that includes the least value and the greatest value in the data set. Graph the five numbers that you found in part (b).

d. Explain how the box-and-whisker plot shown below represents the data set.



Common Core

Box-and-Whisker Plots

In this lesson, you will
- make and interpret box-and-whisker plots.
- find interquartile ranges of data sets.
- compare box-and-whisker plots.

Learning Standards
S.ID.1
S.ID.2
S.ID.3

Laurie's Notes

Introduction

Standards for Mathematical Practice

- **MP3a Construct Viable Arguments:** Students create box-and-whisker plots to display data sets. Multiple box-and-whisker plots can be displayed on the same number line, therefore students are asked to construct viable arguments in comparing data sets.

Motivate

- Share a story about your commute to school today—perhaps the traffic, something you saw, or a stop you made for coffee. Conclude with how many minutes it took for your commute.
- Collect class data about the numbers of minutes it took your students to commute to school this morning, from the time they left their front doors until they walked into the school. If this is awkward data to collect, change the question. Data can be collected on slips of paper.
- Record the data on the board and leave it for later. You may want to take time to have students make comments about the data set.

Activity Notes

Activity 1

❓ "How many pieces of data are there?" 24

- Explain that today they are going to construct a box-and-whisker plot, a data display that is generally used for very large data sets. For instance, the results of a state test for all 8th graders could be displayed using a box-and-whisker plot.
- The box-and-whisker plot uses a number line to visually represent the data. Specific data values are *not* graphed, but characteristics of the data are still conveyed.

❓ "How many numbers did you graph in making the box-and-whisker plot?" 5

- **Big Idea:** The 5 numbers graphed summarize the entire data set. The least and greatest values are the boundaries. The median separates the data into two parts. The first (or lower) quartile is the median of the lower half. The third (or upper) quartile is the median of the upper half. The box encloses the middle 50% of the data.

❓ "What percent of the data is represented by the upper whisker?" 25%
"How many data values are in the upper whisker?" 6

❓ "What percent of the data is represented by the lower whisker?" 25%
"How many data values are in the lower whisker?" 6

- **MP2 Reason Abstractly and Quantitatively:** These questions provide opportunities for students to reason quantitatively.

Common Core State Standards

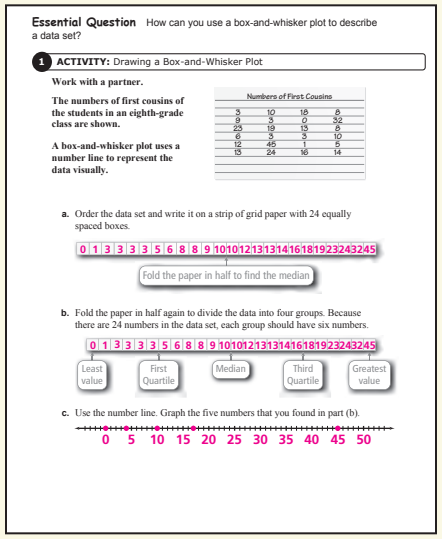
S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).
S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Technology for the Teacher

Dynamic Classroom

Lesson Plans
Complete Materials List

12.3 Record and Practice Journal



Essential Question How can you use a box-and-whisker plot to describe a data set?

1 ACTIVITY: Drawing a Box-and-Whisker Plot

Work with a partner.

The numbers of first cousins of the students in an eighth-grade class are shown.

A box-and-whisker plot uses a number line to represent the data visually.

Numbers of First Cousins

3	10	18	8
9	3	0	32
25	19	15	8
6	3	3	10
12	45	1	5
15	24	16	14

a. Order the data set and write it on a strip of grid paper with 24 equally spaced boxes.

0 1 3 3 3 3 5 6 8 8 9 10 10 12 13 13 14 16 18 19 23 24 32 45

Fold the paper in half to find the median

b. Fold the paper in half again to divide the data into four groups. Because there are 24 numbers in the data set, each group should have six numbers.

0 1 3 3 3 3 5 6 8 8 9 10 10 12 13 13 14 16 18 19 23 24 32 45

Least value | First Quartile | Median | Third Quartile | Greatest value

c. Use the number line. Graph the five numbers that you found in part (b).

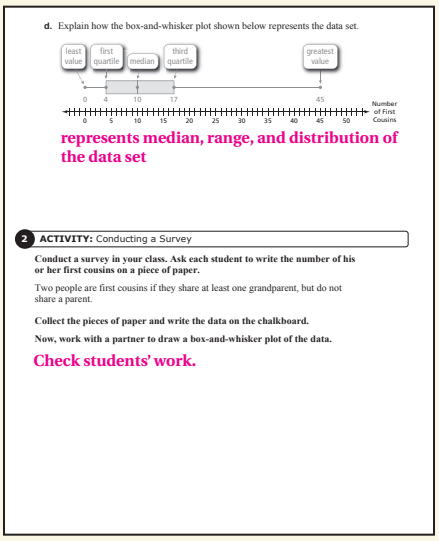
0 5 10 15 20 25 30 35 40 45 50

English Language Learners

Visual

Use a diagram of a generic box-and-whisker plot on an overhead as a visual aid for English learners. Have students identify the parts of the box-and-whisker plot: *median, first quartile, third quartile, least value,* and *greatest value*. Make sure students understand that they can interpret a box-and-whisker plot that does not have a scale.

12.3 Record and Practice Journal



d. Explain how the box-and-whisker plot shown below represents the data set.

represents median, range, and distribution of the data set

2 ACTIVITY: Conducting a Survey

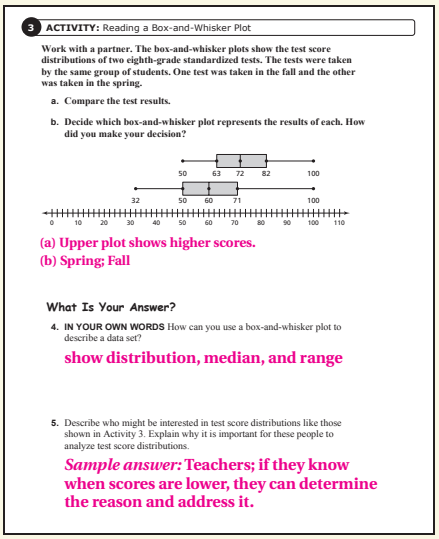
Conduct a survey in your class. Ask each student to write the number of his or her first cousins on a piece of paper.

Two people are first cousins if they share at least one grandparent, but do not share a parent.

Collect the pieces of paper and write the data on the chalkboard.

Now, work with a partner to draw a box-and-whisker plot of the data.

Check students' work.



3 ACTIVITY: Reading a Box-and-Whisker Plot

Work with a partner. The box-and-whisker plots show the test score distributions of two eighth-grade standardized tests. The tests were taken by the same group of students. One test was taken in the fall and the other was taken in the spring.

a. Compare the test results.

b. Decide which box-and-whisker plot represents the results of each. How did you make your decision?

(a) Upper plot shows higher scores.
(b) Spring; Fall

What Is Your Answer?

4. IN YOUR OWN WORDS How can you use a box-and-whisker plot to describe a data set?

show distribution, median, and range

5. Describe who might be interested in test score distributions like those shown in Activity 3. Explain why it is important for these people to analyze test score distributions.

Sample answer: Teachers; if they know when scores are lower, they can determine the reason and address it.

Laurie's Notes

Activity 2

- Explain that you want to practice making a box-and-whisker plot with data collected from students. You will gather and record information about the number of first cousins for each student in your class. If this is awkward, ask a different question.
- Students should follow the steps from Activity 1 to construct this plot. If there is an odd number of students in the class, the median is the middle value of the sorted data. To find the first quartile, exclude the median and find the median of the lower half of the data. To find the third quartile, exclude the median and find the median of the upper half of the data.
- As students are making the plot, make the same plot at the overhead projector for discussion purposes.
- Ask questions about the plot: median, range, number of data values considered versus number of data values graphed.

Activity 3

- One advantage of box-and-whisker plots is that multiple plots can be displayed and analyzed using the same number line. For instance, state test scores for 5 different schools could be displayed on the same number line.
- Read the information given and analyze the two plots.
- **MP3b Critique the Reasoning of Others:** Students should be listening to the analysis offered by their partners or other classmates and critiquing their reasoning.

? "Which test is represented by which plot? Explain." Listen for students discussing the location of the median and the third quartile for each plot. The spring test is the top plot.

? "True or false: 50% of the scores in the top plot are greater than 75% of the scores in the bottom plot." true

What Is Your Answer?

- **Think-Pair-Share:** Students should read each question independently and then work in pairs to answer the questions. When they have answered the questions, the pair should compare their answers with another group and discuss any discrepancies.

Closure

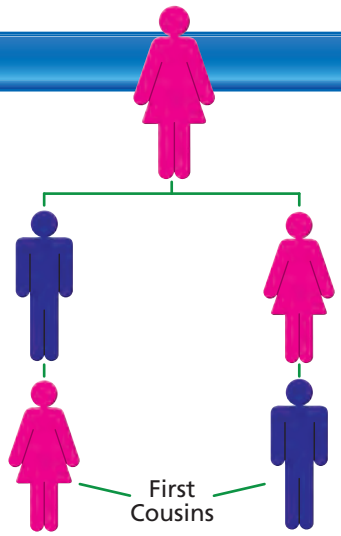
- Make a box-and-whisker plot of the data collected at the beginning of class. Write one or two observations about the plot.

2 ACTIVITY: Conducting a Survey

Conduct a survey in your class. Ask each student to write the number of his or her first cousins on a piece of paper. Collect the pieces of paper and write the data on the chalkboard.

Now, work with a partner to draw a box-and-whisker plot of the data.

Two people are first cousins if they share at least one grandparent, but do not share a parent.



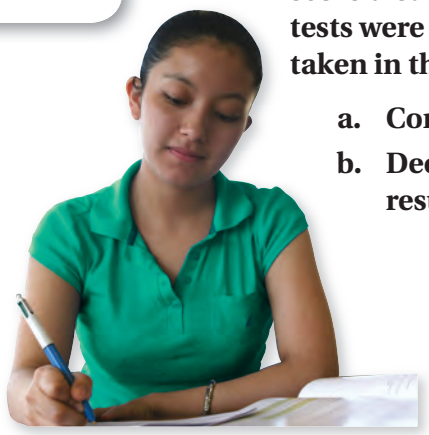
Math Practice 7

View as Components

How do the different components of a box-and-whisker plot help you interpret the values?

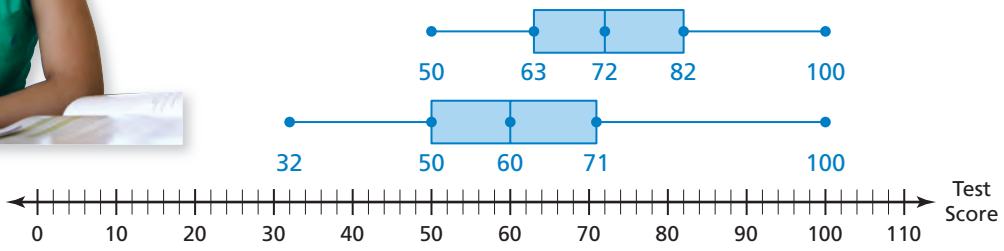
3 ACTIVITY: Reading a Box-and-Whisker Plot

Work with a partner. The box-and-whisker plots show the test score distributions of two eighth-grade standardized tests. The tests were taken by the same group of students. One test was taken in the fall and the other was taken in the spring.



a. Compare the test results.

b. Decide which box-and-whisker plot represents the results of each test. How did you make your decision?



What Is Your Answer?

4. **IN YOUR OWN WORDS** How can you use a box-and-whisker plot to describe a data set?

5. Describe who might be interested in test score distributions like those shown in Activity 3. Explain why it is important for these people to analyze test score distributions.

Practice Use what you learned about box-and-whisker plots to complete Exercise 4 on page 623.

12.3 Lesson

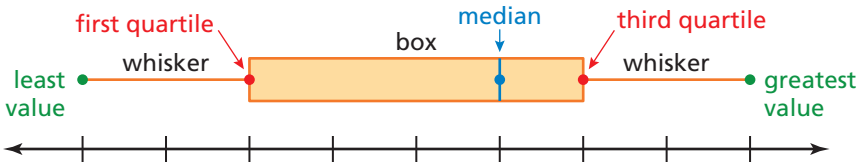
Check It Out
Lesson Tutorials
BigIdeasMath.com

Key Vocabulary
box-and-whisker plot, *p. 620*
quartile, *p. 620*
five-number summary, *p. 620*
interquartile range, *p. 621*

Key Idea

Box-and-Whisker Plot

A **box-and-whisker plot** displays a data set along a number line using medians. **Quartiles** divide the data set into four equal parts. The median (second quartile) divides the data set into two halves. The median of the lower half is the first quartile. The median of the upper half is the third quartile.



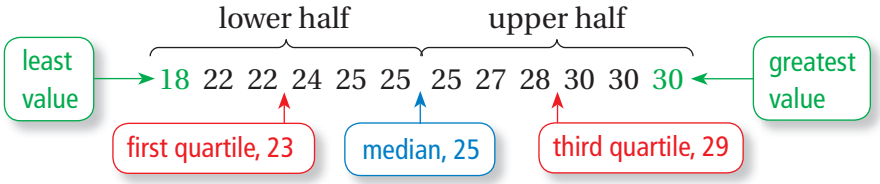
The five numbers that make up the box-and-whisker plot are called the **five-number summary** of the data set.

EXAMPLE 1 Making a Box-and-Whisker Plot

Make a box-and-whisker plot for the ages of the members of the U.S. women's wheelchair basketball team.

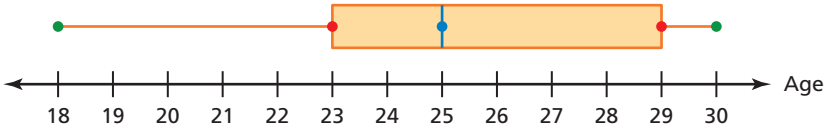
24, 30, 30, 22, 25, 22, 18, 25, 28, 30, 25, 27

Step 1: Order the data. Find the median and the quartiles.



Step 2: Draw a number line that includes the least and greatest values. Graph points above the number line for the five-number summary.

Step 3: Draw a box using the quartiles. Draw a line through the median. Draw whiskers from the box to the least and greatest values.



Study Tip

A box-and-whisker plot shows the *variability* of a data set.

On Your Own

Now You're Ready
Exercises 5–7

1. A basketball player scores 14, 16, 20, 5, 22, 30, 16, and 28 points during a tournament. Make a box-and-whisker plot for the points scored by the player.

Multi-Language Glossary at BigIdeasMath.com

Laurie's Notes

Introduction

Connect

- **Yesterday:** Students gained a general understanding of how a box-and-whisker plot is constructed. (MP2, MP3)
- **Today:** Students will construct and analyze box-and-whisker plots.

Motivate

- The physical involvement of making a human box-and-whisker plot makes a lasting impression on students.
- **Preparation:** Give each student an index card with a number written on it. Include an outlier or two on one end of the data.

? "What is the first step in making a box-and-whisker plot?" sort the data

- Students should stand up and sort themselves. Have the median, the first and third quartiles, and the least and greatest data values take one step forward. If there is an even number of data values, the middle two students must figure out how to represent the mean and take a step forward.
- Make a number line on the floor or on the board. Position the 5 key values. If the plot is done on the floor, use string to form the whiskers. Students will have to visualize the box.
- **MP2 Reason Abstractly and Quantitatively:** Discuss features of the plot. If the plot includes an outlier, the length of the string becomes an instant topic of conversation. Students recognize that the same number of data values (25% of the class) is being represented by each whisker, yet the lengths of string are very different.

Lesson Notes

Key Idea

- Define the box-and-whisker plot constructed by graphing the quartiles. Draw the sample plot and discuss the process and vocabulary.
- **Discuss:** The box-and-whisker plot shows the *variability* of the data. Refer to this idea in each example done today.
- Tell students that box-and-whisker plots are also called *boxplots*.
- Discuss the vocabulary term *five-number summary*.

Example 1

- Remind students that they must sort the data to find the quartiles.
- Notice that the first quartile (23) and the third quartile (29) are not data values from the set. This is fine. The five values are simply giving a marker for how the sorted data is spread out into four groups.

? "How many players are represented in each quartile?" 3

? "What percent of the players were older than 23?" 75%

? "What is the range of ages for the team?" 12

On Your Own

- This is a very small data set to make it manageable for students.

Goal Today's lesson is constructing and analyzing a **box-and-whisker plot**.

Technology for the Teacher

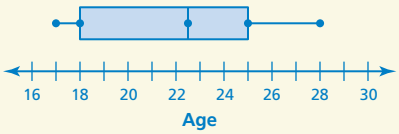
Dynamic Classroom

Lesson Tutorials
Lesson Plans
Answer Presentation Tool

Extra Example 1

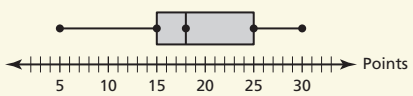
Make a box-and-whisker plot for the ages of the members of a women's basketball team.

25, 22, 18, 23, 27, 20, 18, 25, 28, 17, 23, 18



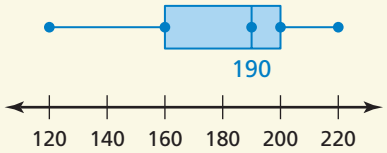
On Your Own

1.



Extra Example 2

You have written several fiction stories about your favorite television character and posted them on the Internet. The box-and-whisker plot represents the number of reader reviews each of your stories received.



Number of Reviews of Each Story

a. Find and interpret the range of the data. range: 100; This means that the number of reviews of each story varies by no more than 100.

b. Describe the distribution of the data. One-quarter of the stories received 160 reviews or less. One half of the stories received between 160 and 200 reviews. One-quarter of the stories received 200 or more reviews.

c. Find and interpret the interquartile range of the data. interquartile range: 40; This means that the middle half of the number of reviews vary by no more than 40.

On Your Own

2. range: 12 yr; This means that the ages vary by no more than 12 years; interquartile range: 6 yr; This means that the middle half of the ages vary by no more than 6 years.

3. See Additional Answers.

Differentiated Instruction

Auditory

Remind students of other words that have the same root as *quartile:* quarter and quart, for example. Define the words as *four* or *fourths*. Mention that one-fourth is the same as 25%.

Laurie's Notes

Discuss

- The box portion of the plot contains 50% of the data. Its length is called the interquartile range, IQR.
- "What portion of the data is in the box and to the left of the median?" $\frac{1}{4}$
- "What portion of the data is in the box and to the right of the median?" $\frac{1}{4}$
- In the sample shown, the lengths of the two parts of the box indicate that the data in the box are less spread out on the right side of the median than on the left side of the median.
- Similar statements about the dispersion of the data can be made by looking at the size (length) of each of the four parts of the plot.

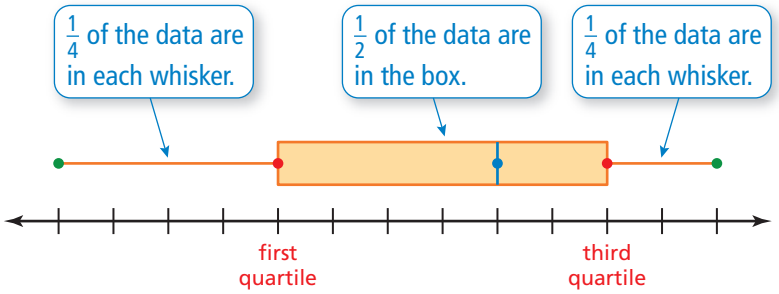
Example 2

- "Do you know from the boxplot how many songs the band played at the concert?" no
- It is important for students to remember that the number of data points cannot be determined solely from a boxplot. Only the five summary points are known, and only two of those (least value and greatest value) *must* be values in the data set.
- "What is the range of the data and what does it mean in the context of the problem?" The range is 300 − 160 = 140. This means that the difference between the shortest song and longest song was 140 seconds, or 2 minutes 20 seconds.
- "What does the length of each part of the boxplot tell you about the distribution of the data?" *Sample answers:* 50% of the songs were between 260 seconds and 300 seconds; 25% of the songs were 220 seconds or less.
- In part (c), the students are focusing on the middle 50% of the data. The IQR is 60 seconds. One way to interpret this result is that the middle half of the song lengths vary by no more than 60 seconds.

On Your Own

- **MP3a Construct Viable Arguments:** Ask different students to explain their answers. Check for correct language and valid arguments.

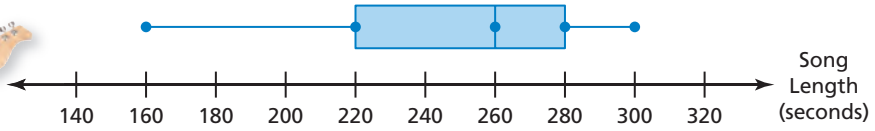
The figure shows how data are distributed in a box-and-whisker plot.



Another measure of dispersion for a data set is the **interquartile range**, which is the difference of the third quartile and the first quartile. It represents the range of the middle half of the data.

EXAMPLE 2 Interpreting a Box-and-Whisker Plot

The box-and-whisker plot represents the lengths of songs (in seconds) played by a rock band at a concert.



a. Find and interpret the range of the data.

The least value is 160. The greatest value is 300.

So, the range is $300 - 160 = 140$ seconds. This means that the song lengths vary by no more than 140 seconds.

b. Describe the distribution of the data.

- 25% of the song lengths are between 160 and 220 seconds.
- 50% of the song lengths are between 220 and 280 seconds.
- 25% of the song lengths are between 280 and 300 seconds.

c. Find and interpret the interquartile range of the data.

$$\begin{aligned}\text{interquartile range} &= \text{third quartile} - \text{first quartile}\\ &= 280 - 220\\ &= 60\end{aligned}$$

So, the interquartile range is 60 seconds. This means that the middle half of the song lengths vary by no more than 60 seconds.

On Your Own

Now You're Ready
Exercises 10 and 11

Use the box-and-whisker plot in Example 1.

2. Find and interpret the range and interquartile range of the data.
3. Describe the distribution of the data.

A box-and-whisker plot shows the shape of a distribution.

Key Ideas

Study Tip

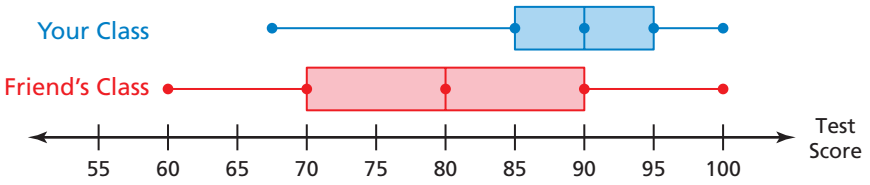
If you can draw a line through the median of a box-and-whisker plot, and each side is a mirror image of the other, then the distribution is symmetric.

Shapes of Box-and-Whisker Plots

Skewed left	*Symmetric*	*Skewed right*
• Left whisker longer than right whisker	• Whiskers about same length	• Right whisker longer than left whisker
• Most data on the right	• Median in the middle of the data	• Most data on the left

EXAMPLE 3 Comparing Box-and-Whisker Plots

The double box-and-whisker plot represents the test scores for your class and your friend's class.



a. Identify the shape of each distribution.

For your class, the left whisker is longer than the right whisker, and most of the data are on the right side of the display. So, the distribution is skewed left.

For your friend's class, the whisker lengths are equal. The median is in the middle of the data. The data appear to be evenly distributed on both sides of the median. So, the distribution is symmetric.

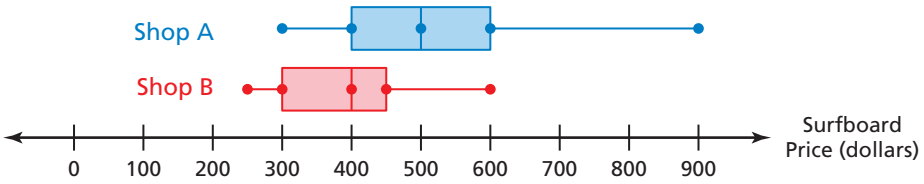
b. Which test scores are more spread out?

The range of the test scores in your friend's class is greater than the range in your class. Also, because the box for your friend's class is longer than the box for your class, the interquartile range is also greater. So, the test scores in your friend's class are more spread out.

On Your Own

Now You're Ready
Exercise 20

4. The double box-and-whisker plot represents the surfboard prices at Shop A and Shop B. Identify the shape of each distribution. Which shop's prices are more spread out? Explain.



Laurie's Notes

Key Ideas

- Sketch each of the box-and-whisker plots. Discuss what the shape of the plot implies about the distribution of the data.

? "If the box-and-whisker plots represent the costs of jeans at three different stores and the range is the same for all three plots, what does the shape tell you about the distribution of the costs at all three stores?" Listen for student understanding of the spread of the data, referencing percentages and relative cost.

Example 3

- One advantage of a box-and-whisker plot is that multiple data sets can be shown in the same display, as with double bar graphs and double line graphs. The data sets can also have different numbers of data values.
- Part (a) asks students to identify the shape. A student might comment that a low test score in "your class" could be the reason for it being skewed. Without the outlier it could be a symmetric graph.
- **MP2:** Supplement the discussion in part (b) with quantitative comparisons, such as 50% of your class scored as well as the top 25% of your friend's class.

On Your Own

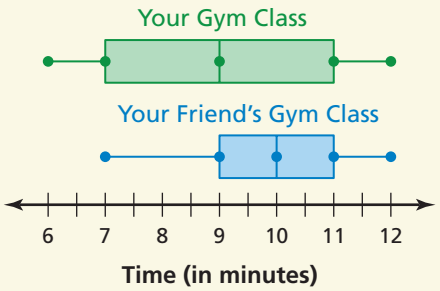
- **Neighbor Check:** Have students work independently and then have their neighbors check their work. Have students discuss any discrepancies.
- Ask other quantitative questions abut the two boxplots.

Closure

- **Exit Ticket:**
  - What are the 5 key values that are graphed in a box-and-whisker plot? least value, first quartile, median, third quartile, greatest value
  - How does an outlier affect a box-and-whisker plot? *Sample answer:* increases the length of one of the whiskers
  - Explain why two data sets of different sizes can be graphed on the same number line. Box-and-whisker plots show the distribution of the data, not individual data points.

Extra Example 3

The double box-and-whisker plot represents the length of time (in minutes) it takes students in your gym class and your friend's gym class to run 1 mile.



a. Identify the shape of each distribution. For your class, the whisker lengths are equal. The median is in the middle of the data. The data appear to be evenly distributed on both sides of the median. So, the distribution is symmetric. For your friend's class, the left whisker is longer than the right whisker, and most of the data are on the right side of the display. So, the distribution is skewed left.

b. Which lengths of time are more spread out? The range of the times in your class is greater than the range in your friend's class. Also, because the box for your class is longer than the box for your friend's class, the interquartile range is also greater. So, the times in your class are more spread out.

On Your Own

4. See Additional Answers.

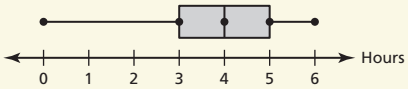
Vocabulary and Concept Check

1. 25%; 50%
2. Find the median of the lower half of an ordered data set.
3. The length gives the range of the data set and it tells how much the data vary.

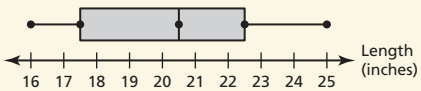
Practice and Problem Solving

4. *Sample answer:* Both sales reps have the same median number of cars sold, but Sales Rep B's numbers are more spread out than Sales Rep A's numbers.

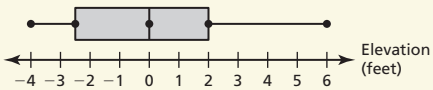
5.



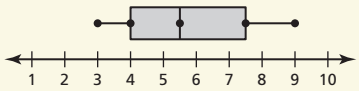
6.



7.



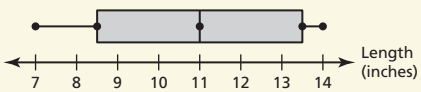
8. The median should be 5.5.



9. **a.** 11 in.

   b. about 8.5 in.; about 13.5 in.

   c.



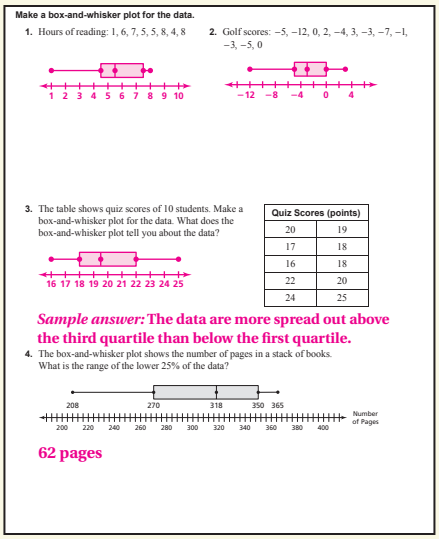
Assignment Guide and Homework Check

Level	Assignment	Homework Check
Average	1–3, 4, 5, 8–10, 12–18 even, 25–29	4, 9, 10, 14
Advanced	1–3, 6–18 even, 19–22, 25–29	10, 14,19, 21

Common Errors

- **Exercise 1** Students may confuse or forget what percent of the data from a data set is represented by each whisker and the box. Remind them of these percents.
- **Exercises 5–9** Students may have difficulty creating the box-and-whisker plots. Remind them of the five-number summary of a data set.

12.3 Record and Practice Journal



Make a box-and-whisker plot for the data.

1. Hours of reading: 1, 6, 7, 5, 5, 8, 4, 8

2. Golf scores: −5, −12, 0, 2, −4, 3, −3, −7, −1, −3, −5, 0

3. The table shows quiz scores of 10 students. Make a box-and-whisker plot for the data. What does the box-and-whisker plot tell you about the data?

Quiz Scores (points)	
20	19
17	18
16	18
22	20
24	25

Sample answer: The data are more spread out above the third quartile than below the first quartile.

4. The box-and-whisker plot shows the number of pages in a stack of books. What is the range of the lower 25% of the data?

62 pages

12.3 Exercises

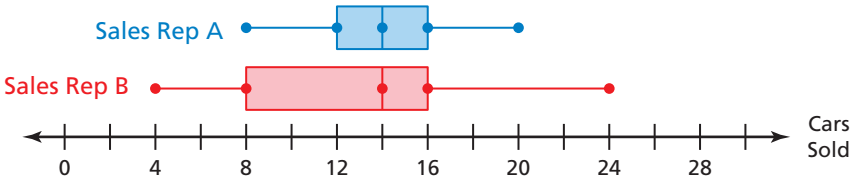
Check It Out
Help with Homework
BigIdeasMath.com

Vocabulary and Concept Check

1. **VOCABULARY** In a box-and-whisker plot, what percent of the data is represented by each whisker? by the box?
2. **WRITING** Describe how to find the first quartile of a data set.
3. **NUMBER SENSE** What does the length of a box-and-whisker plot tell you about the data?

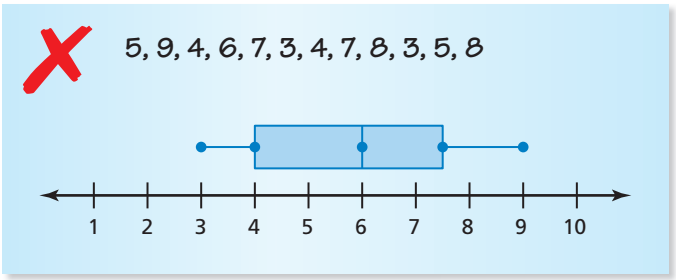
Practice and Problem Solving

4. The box-and-whisker plots show the monthly car sales for a year for two sales representatives. Compare the sales for the two representatives.



Make a box-and-whisker plot for the data.

5. Hours of television watched: 0, 3, 4, 5, 3, 4, 6, 5
6. Lengths (in inches) of cats: 16, 18, 20, 25, 17, 22, 23, 21
7. Elevations (in feet): $-2, 0, 5, -4, 1, -3, 2, 0, 2, -3, 6, -1$



8. **ERROR ANALYSIS** Describe and correct the error in making a box-and-whisker plot for the data.

9. **FISH** The lengths (in inches) of the fish caught on a fishing trip are 9, 10, 12, 8, 13, 10, 12, 14, 7, 14, 8, and 14.
   a. What is the median of the data set?
   b. What are the first and third quartiles of the data set?
   c. Make a box-and-whisker plot for the data.

10. INCHWORM The table shows the lengths of 12 inchworms.

a. Make a box-and-whisker plot for the data.

b. Find and interpret the range of the data.

c. Describe the distribution of the data.

d. Find and interpret the interquartile range of the data.

Length (cm)	2.5	2.4	2.3	2.5	2.7	2.1	2.8	2.6	2.1	2.6	2.9	2.0

Entrée Prices (dollars)			
14.00	17.00	12.50	10.00
11.00	18.25	9.00	8.50
14.75	15.00	14.00	12.00

11. ENTRÉE The table shows the prices of entrées at a restaurant.

a. Make a box-and-whisker plot for the data.

b. Find and interpret the interquartile range of the data.

c. Describe the distribution of the data.

d. Find the standard deviation. Interpret your result.

12. WRITING Given the numbers 36 and 12, identify which number is the range, and which number is the interquartile range, of a set of data. Explain your reasoning.

Determine whether the shape of the box-and-whisker plot is *symmetric*, *skewed left*, or *skewed right*. Explain.

13.

14.

15.

16.

17. ERROR ANALYSIS Describe and correct the error in describing the box-and-whisker plot.

✗ The shape of the distribution is skewed right. So, there are more data values to the right of the median than to the left of the median.

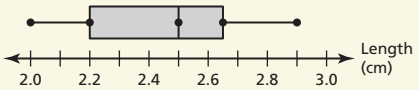
18. LOGIC Give examples of real-life data that are symmetric and real-life data that are not symmetric. Justify your answer.

Common Errors

- **Exercises 10–12, and 20** Students may confuse the range and interquartile range of a data set. Remind them of these definitions.
- **Exercises 13–16** Students may confuse the meanings of skewed left and skewed right. Tell them that the shape is determined by the longer whisker.
- **Exercise 18** Students may have difficulty coming up with examples of real-life data that are symmetric and real-life data that are not symmetric. Have them use this textbook or the Internet to search for examples.

Practice and Problem Solving

10. a.



b. Inchworm lengths vary no more than 0.9 centimeter.

c. 25% of the lengths are between 2.0 and 2.2 centimeters; 50% of the lengths are between 2.2 and 2.65 centimeters; 25% of the lengths are between 2.65 and 2.9 centimeters.

d. The interquartile range is 0.45 centimeter. This means the middle half of the lengths vary by no more than 0.45 centimeter.

11. See Additional Answers.

12. range: 36; The range gives you the amount all the data vary by; interquartile range: 12; The interquartile range gives you the amount the middle half of the data vary by.

13. symmetric; The whiskers are about the same length and the median is in the middle of the data.

14–16. See Additional Answers.

17. The number of data values on each side of the median is the same.

18. *Sample Answer:* symmetric: height of students in a class; not symmetric: exam scores

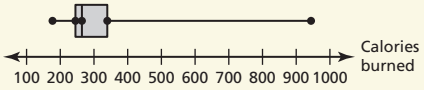
Differentiated Instruction

Kinesthetic

To make ordering data easier, have students write each data value on a sticky note. Using their desk or a piece of notebook paper, students can move the notes to order the data and divide them into quartiles.

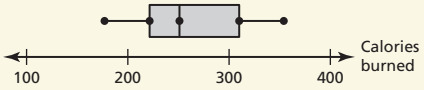
Practice and Problem Solving

19. a.



b. 944 calories

c.



d. The outlier makes the right whisker longer, increases the length of the box, increases the third quartile, and increases the median. In this case, the first quartile and the left whisker were not affected.

20. See Additional Answers.

21. *Sample answer:*
0, 5, 10, 10, 10, 15, 20

22. See *Taking Math Deeper*.

23. *Sample answer:*
1, 7, 9, 10, 11, 11, 12

24. *Sample answer:*
1, 2, 4, 5, 10, 10, 10

Fair Game Review

25. $y = 3x + 2$

26. $y = -\frac{4}{3}x - 1$

27. $y = -\frac{1}{4}x$

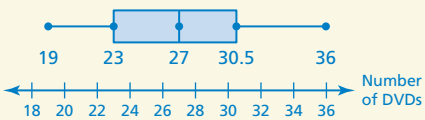
28. $y = \frac{1}{2}x + 4$

29. B

Mini-Assessment

Make a box-and-whisker plot for the number of DVDs your class owns.

25, 31, 27, 36, 19, 22, 20, 24, 30, 32, 29, 27



Taking Math Deeper

Exercise 22

This is a great opportunity for students to *work backwards* and demonstrate their understanding of the *five-number summary*.

1 Interpret the given characteristic:

Both whiskers are the same length as the box.

Express these lengths using the extreme values and the quartiles.

Length of left whisker = first quartile − least value

Length of box = third quartile − first quartile

Length of right whisker = greatest value − third quartile

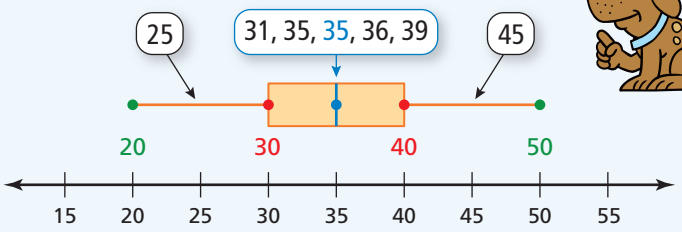
2 Work backwards to write a data set.

The problem does not specify the length of each whisker or the box. So, assign an arbitrary value for each length, such as 10.

Assign an arbitrary value for any one of the extreme values or the quartiles (except the median). Let the greatest value be 50.

Using these values and the equations above, you can determine that the third quartile is 40, the first quartile is 30, and the least value is 20.

3 Choose arbitrary values to complete the data set. Make sure to add an appropriate number of values so that the extreme values and quartiles do not change. Use a box-and-whisker plot to help.



So, one possible data set is 20, 25, 30, 31, 35, 35, 36, 39, 40, 45, 50.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Resources by Chapter • Practice A and Practice B • Puzzle Time Record and Practice Journal Practice Differentiating the Lesson Lesson Tutorials Skills Review Handbook	Resources by Chapter • Enrichment and Extension • School-to-Work Start the next section

19. CALORIES The table shows the numbers of calories burned per hour for nine activities.

a. Make a box-and-whisker plot for the data.

b. Identify the outlier.

c. Make another box-and-whisker plot without the outlier.

d. WRITING Describe how the outlier affects the whiskers, the box, and the quartiles of the box-and-whisker plot.

Calories Burned per Hour	
Fishing	207
Mowing the lawn	325
Canoeing	236
Bowling	177
Hunting	295
Fencing	354
Bike racing	944
Horseback riding	236
Dancing	266

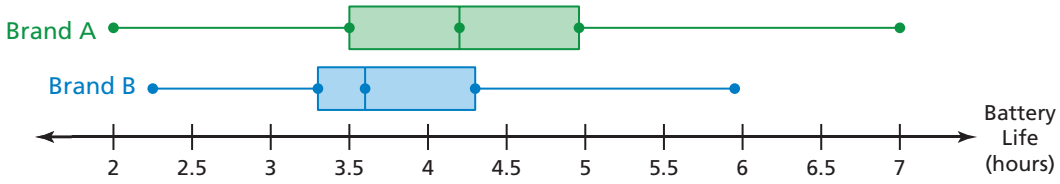
20. CELL PHONES The double box-and-whisker plot compares the battery lives (in hours) of two brands of cell phones.

a. Identify the shape of each distribution.

b. What is the range of the upper 75% of each brand?

c. Compare the interquartile ranges of the two data sets.

d. Which brand do you think has the greater standard deviation? Explain.



Modeling **Create a set of data values for the box-and-whisker plot that has the given characteristic(s).**

21. The least value, greatest value, quartiles, and median are all equally spaced.

22. Both whiskers are the same length as the box.

23. The box between the median and the first quartile is three times as long as the box between the median and the third quartile.

24. There is no right whisker.

Fair Game Review *What you learned in previous grades & lessons*

Write an equation of the line that passes through the points. *(Section 2.6)*

25. $(-4, -10), (2, 8)$

26. $(-3, 3), (0, -1)$

27. $(-4, 1), (4, -1)$

28. $(6, 7), (8, 8)$

29. MULTIPLE CHOICE What is the quotient of $(2z^2 - 13z + 21)$ and $(z - 3)$? *(Section 11.5)*

Ⓐ $2z + 7$ Ⓑ $2z - 7$ Ⓒ $z + 6$ Ⓓ $z - 7$

12.4 Shapes of Distributions

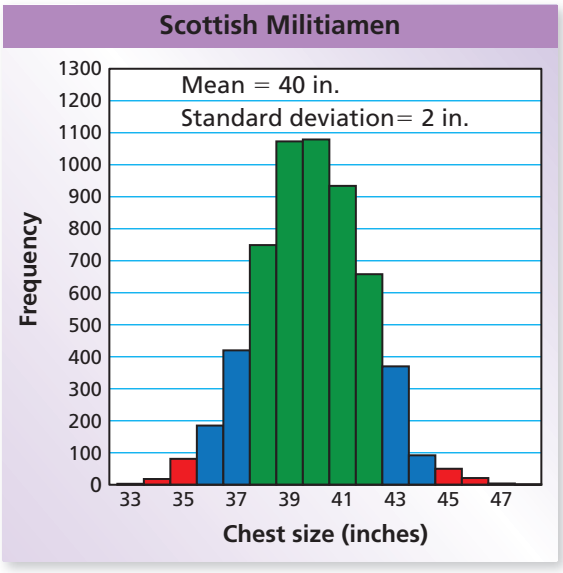
Essential Question How can you use a histogram to characterize the basic shape of a distribution?

1 ACTIVITY: Analyzing a Famous Symmetric Distribution

A famous data set was collected in Scotland in the mid-1800s. It contains the chest sizes, measured in inches, of 5738 men in the Scottish Militia.

The Thin Red Line is a painting by Robert Gibb. It was painted in 1881. Only the left portion of the painting is shown in the photo at the right.

Chest Size	Number of Men
33	3
34	18
35	81
36	185
37	420
38	749
39	1073
40	1079
41	934
42	658
43	370
44	92
45	50
46	21
47	4
48	1



Work with a partner. What percent of the chest sizes lie within (a) 1 standard deviation, (b) 2 standard deviations, and (c) 3 standard deviations of the mean? Explain your reasoning.

COMMON CORE

Data Distributions

In this lesson, you will

- describe shapes of distributions.
- choose appropriate measures of central tendency and dispersion to represent data sets.

Learning Standards
S.ID.2
S.ID.3

Laurie's Notes

Introduction

Standards for Mathematical Practice

- **MP2 Reason Abstractly and Quantitatively:** Students will describe the shape of a distribution and choose appropriate measures of central tendency and spread. The shape of a histogram (visual display) will help students to reason abstractly about the center and spread of the data.

Motivate

- Ask students if they know what a militia is. You could give them the hint that during the American Revolutionary War, the Minutemen were members of a militia. A militia refers to a military force made up of citizens that protect their community/town in times of emergency.
- Tell students about the Scottish Militia Bill, which was passed by the Houses of Parliament in Great Britain in the early 1700s. The purpose of this bill was to help create uniformity in governing the militias across Great Britain. This bill was not passed into law by Queen Anne, because she feared the militia in Scotland would not be loyal. This would mark the last time a British ruler would veto a bill passed by both Houses of Parliament.
- Data about the Scottish Militia is the subject of today's activity.

Activity Notes

Activity 1

- The bar graph in this activity is color-coded according to the frequency table. Discuss the color-coding before students begin the activity with their partner. Students may even note the symmetry of the data about the mean, which can be recognized numerically and graphically.
- ? "In this activity, what does it mean to lie within 1 standard deviation of the mean?" It means a chest size is within 2 inches on either side of the mean of 40 inches. So, it is between 38 inches and 42 inches.
- ? "In this activity, what does it mean to lie within 2 standard deviations of the mean?" It means a chest size is within 4 inches on either side of the mean of 40 inches. So, it is between 36 inches and 44 inches.
- ? "In this activity, what does it mean to lie within 3 standard deviations of the mean?" It means a chest size is within 6 inches on either side of the mean of 40 inches. So, it is between 34 inches and 46 inches.
- A calculator is helpful in calculating these percents. Students who read carefully know that there are 5738 men represented in the frequency table and the graph. The percents are cumulative. So, the men who are within 1 standard deviation of the mean are also within 2 standard deviations of the mean, and so on.
- When students have finished, discuss their results for each part of the activity.
- **Big Idea:** More than 99% of the data are within 3 standard deviations of the mean. In this activity, 0.14% of the men have chest sizes that are more than 3 standard deviations from the mean.

Common Core State Standards

S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Previous Learning

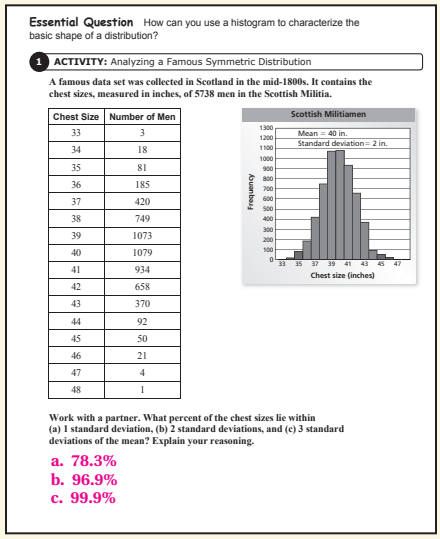
Students should know how to describe the shape of a box-and-whisker plot.

Technology for the Teacher

Dynamic Classroom

Lesson Plans
Complete Materials List

12.4 Record and Practice Journal



Essential Question How can you use a histogram to characterize the basic shape of a distribution?

1 ACTIVITY: Analyzing a Famous Symmetric Distribution

A famous data set was collected in Scotland in the mid-1800s. It contains the chest sizes, measured in inches, of 5738 men in the Scottish Militia.

Chest Size	Number of Men
33	3
34	18
35	81
36	185
37	420
38	749
39	1073
40	1079
41	934
42	658
43	370
44	92
45	50
46	21
47	4
48	1

Scottish Militiamen
Mean = 40 in.
Standard deviation = 2 in.
Frequency
Chest size (inches)

Work with a partner. What percent of the chest sizes lie within (a) 1 standard deviation, (b) 2 standard deviations, and (c) 3 standard deviations of the mean? Explain your reasoning.

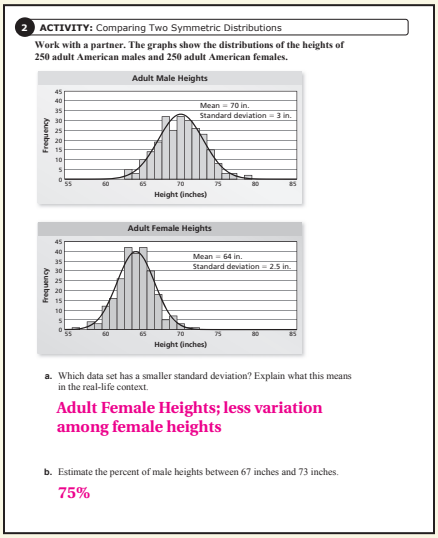
a. 78.3%
b. 96.9%
c. 99.9%

Differentiated Instruction

Inclusion

In this section, students are expected to make many histograms. Provide a resource page of blank axes with horizontal lines and vertical lines. Students add the vertical and horizontal scales, and labels, and color the bars to represent the data.

12.4 Record and Practice Journal



2 ACTIVITY: Comparing Two Symmetric Distributions

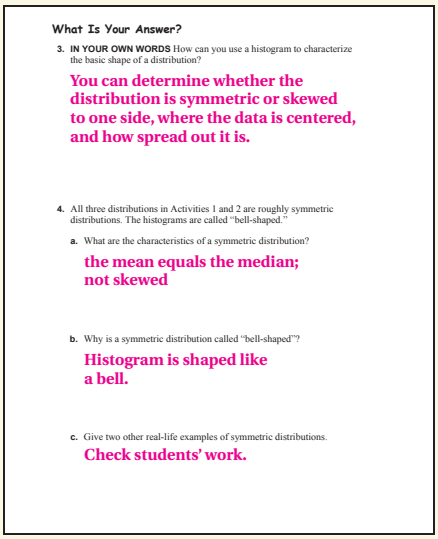
Work with a partner. The graphs show the distributions of the heights of 250 adult American males and 250 adult American females.

a. Which data set has a smaller standard deviation? Explain what this means in the real-life context.

Adult Female Heights; less variation among female heights

b. Estimate the percent of male heights between 67 inches and 73 inches.

75%



What Is Your Answer?

3. **IN YOUR OWN WORDS** How can you use a histogram to characterize the basic shape of a distribution?

You can determine whether the distribution is symmetric or skewed to one side, where the data is centered, and how spread out it is.

4. All three distributions in Activities 1 and 2 are roughly symmetric distributions. The histograms are called "bell-shaped."

a. What are the characteristics of a symmetric distribution?

the mean equals the median; not skewed

b. Why is a symmetric distribution called "bell-shaped"?

Histogram is shaped like a bell.

c. Give two other real-life examples of symmetric distributions.

Check students' work.

Laurie's Notes

Activity 2

- Discuss the two graphs with the class before they begin the activity. Students may ask about the curve that has been plotted on top of the histogram. They may have heard of a "bell-shaped curve." It is fine to suggest that the curve is tracing the basic shape of the distribution of heights. It is smooth versus the line segment outline of the actual graph. The bell-shaped curve will be referred to in Question 4.
- The frequencies on the vertical axis will help students estimate the size of the populations plotted.
- When students have finished, have them share their reasoning for part (a). Students should discuss the symmetry of the distributions and what that means in terms of the distribution of the data.
- Discuss the shape of the distribution and what it means in the real-life context. The mean and standard deviation for males is greater than the mean and standard deviation for females.
- **Connection:** In part (b), the range of the heights is one standard deviation about the mean.
- **MP1a Make Sense of Problems** and **MP3 Construct Viable Arguments and Critique the Reasoning of Others:** Having students explain their reasoning and having a whole class discussion, helps all students make sense of the problem and develop an ability to form arguments and critique the reasoning of others.

What Is Your Answer?

- Make sure students understand the connection between a bell shape and the distribution of the data. In the lesson that follows, students will see that not all distributions are symmetric or bell-shaped.

? "Can a distribution be symmetric and not be bell-shaped?" yes; The distribution could be U-shaped or uniform.

Closure

- Give an example of a real-life data set that may have a symmetric distribution. *Sample answers:* The pulse rates and shoe sizes of students in your class would most likely have symmetric distributions.

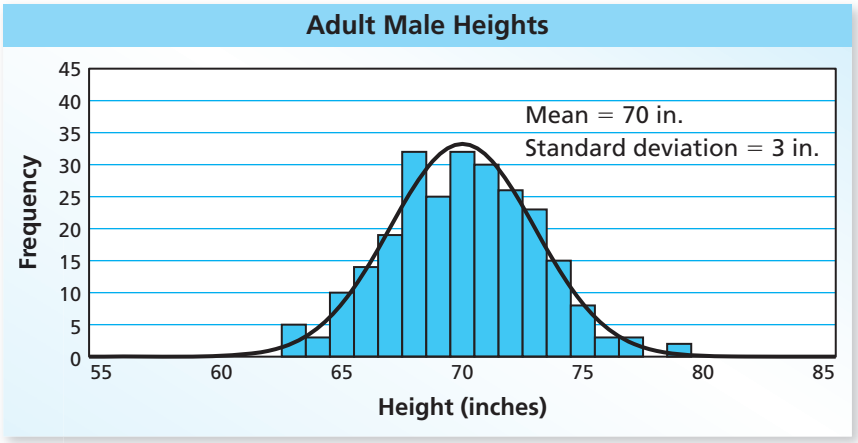
2 ACTIVITY: Comparing Two Symmetric Distributions

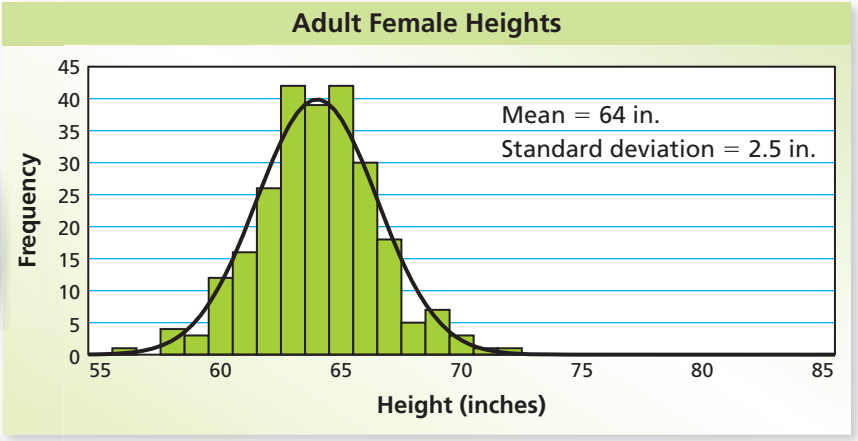
Math Practice 2

Make Sense of Quantities

How can you use a histogram to understand relationships within a data set?

Work with a partner. The graphs show the distributions of the heights of 250 adult American males and 250 adult American females.





a. Which data set has a smaller standard deviation? Explain what this means in the real-life context.

b. Estimate the percent of male heights between 67 inches and 73 inches.

What Is Your Answer?

3. **IN YOUR OWN WORDS** How can you use a histogram to characterize the basic shape of a distribution?

4. All three distributions in Activities 1 and 2 are roughly symmetric distributions. The histograms are called "bell-shaped."

a. What are the characteristics of a symmetric distribution?

b. Why is a symmetric distribution called "bell-shaped"?

c. Give two other real-life examples of symmetric distributions.

Practice Use what you learned about the shapes of distributions to complete Exercises 3 and 4 on page 631.

12.4 Lesson

Check It Out
Lesson Tutorials
BigIdeasMath.com

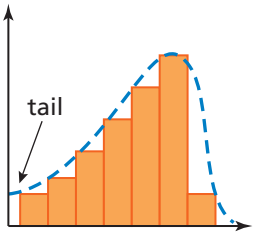
Recall that a histogram is a bar graph that shows the frequency of data values in intervals of the same size. A histogram is another useful data display that shows the shape of a distribution.

Key Ideas

Remember

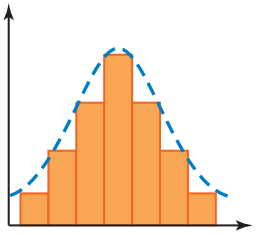
If all the bars of a histogram are about the same height, then the distribution is a *flat*, or *uniform*, distribution. A uniform distribution is also symmetric.

Symmetric and Skewed Distributions



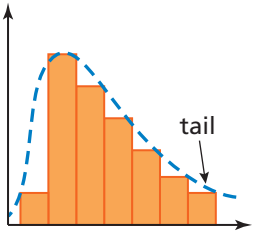
Skewed left

- The "tail" of the graph extends to the left.
- Most data are on the right.



Symmetric

- The data are evenly distributed on each side of the highest bar.



Skewed right

- The "tail" of the graph extends to the right.
- Most data are on the left.

EXAMPLE 1 Describing the Shape of a Distribution

Number of Tickets Sold	Frequency
1–8	5
9–16	9
17–24	16
25–32	25
33–40	20
41–48	8
49–56	7

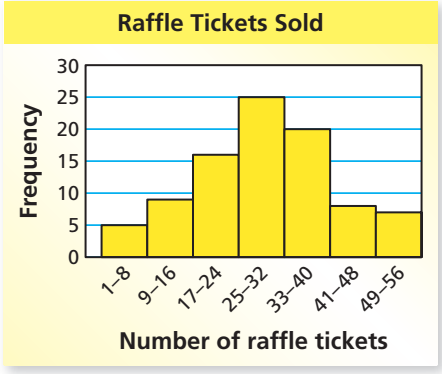
The frequency table shows the numbers of raffle tickets sold by students in your grade. Display the data in a histogram. Describe the shape of the distribution.

Step 1: Draw and label the axes.

Step 2: Draw a bar to represent the frequency of each interval.

The graph is high in the middle, and the data are about evenly distributed on each side of the highest bar.

∴ So, the distribution is symmetric.



On Your Own

Now You're Ready
Exercises 5 and 6

1. The frequency table shows the numbers of pounds of aluminum cans collected by students for a fundraiser. Display the data in a histogram. Describe the shape of the distribution.

Number of Pounds	Frequency
1–10	7
11–20	8
21–30	10
31–40	16
41–50	34
51–60	15

Laurie's Notes

Introduction

Connect

- **Yesterday:** Students used a histogram to describe the basic shape of a distribution. (MP1a, MP2, MP3)
- **Today:** Students will describe the shape of a distribution and choose measures of central tendency and spread to describe the data.

Motivate

- **Story time:** Tell the students that yesterday in the teacher's lounge, three teachers were describing the results of a recent test they gave.
  - **Teacher A:** Lots of low scores; Ds and Fs, some Cs, a few Bs, and one A
  - **Teacher B:** Lots of high scores; As and Bs, some Cs, a few Ds, and one F
  - **Teacher C:** Lots of average scores; Cs, some Bs and Ds, and a few Fs and As
- Ask students to quickly sketch a histogram for each of the three sets of test scores. They should sketch something similar to the three distributions in the Key Ideas.
- Discuss the attributes of a histogram: similar to a bar graph, intervals are all the same size, and frequencies of data values are displayed in the vertical direction by the bar heights.

Lesson Notes

Key Ideas

- Draw a sketch of each type of distribution and label it: skewed left, symmetric, and skewed right.
- Connect the distributions to the test score descriptions above.
- Explain that a dotted line can be drawn to help identify the greatest frequency and to help highlight the basic shape of a distribution.

? "Do all distributions fall into one of these three categories?" no; There are many different types of distributions. A distribution can be flat or uniform. It can have clusters of data or there could be gaps between data values. In short, there are distributions that have no discernible shape at all.

Example 1

- This is a good review of drawing a histogram. The intervals have already been determined, which is often the difficult part for students.
- There is no space between the bars of a histogram. When one of the frequencies is 0, the bar height for that interval is 0.
- **Extension:** Even though the actual data values are not known, have students estimate the mean of the distribution. You could also ask about the maximum range ($56 - 1 = 55$).

Goal Today's lesson is describing the shape of a distribution and determining the best way to represent it.

Technology for the Teacher

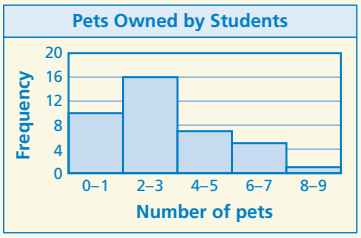
Dynamic Classroom

Lesson Tutorials
Lesson Plans
Answer Presentation Tool

Extra Example 1

The frequency table shows the numbers of pets owned by a group of students. Display the data in a histogram. Describe the shape of the distribution.

Number of Pets	Frequency
0–1	10
2–3	16
4–5	7
6–7	5
8–9	1



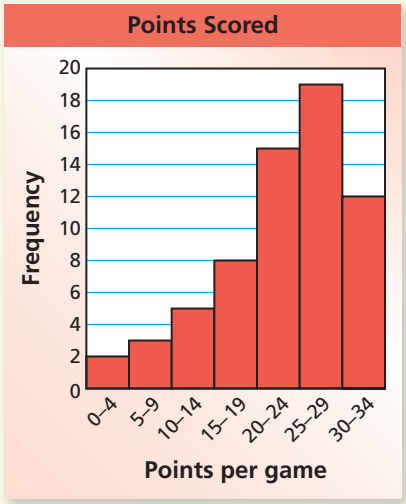
skewed right

On Your Own

1. See Additional Answers.

Extra Example 2

Which measure of central tendency best represents the data? Explain your reasoning.



Because the distribution is high on the right and the tail of the graph extends left, the distribution is skewed left. So, the median best represents the data.

On Your Own

2. Because the distribution is high on the right and the tail of the graph extends left, the distribution is skewed left. So, the median best represents the data.

English Language Learners

Notebook Development

Provide English language learners with a handout of a simplified version of the notes from this lesson. This will allow the student to focus on the concepts being taught. Students should highlight information and add additional notes to the handout during the classroom discussion. They can refer to the page as they work through the exercises.

Laurie's Notes

On Your Own

- Students may want to answer the question without drawing the histogram. Certainly students could reason about the shape of the distribution from the frequencies given in the table, but without making a histogram, the degree of skewness might be difficult to judge.
- **Extension:** A graphing calculator can be used to construct the histogram.

Discuss

- Explain that the shape of a distribution can help you decide the most appropriate measure of central tendency when describing a data set.

? "When a distribution is symmetric, what do you know about the mean and median?" They are about the same and at the center of the distribution.

? "When a distribution is skewed left, what do you know about the mean and median?" The median is greater than the mean.

? "When a distribution is skewed right, what do you know about the mean and median?" The median is less than the mean.

? "When there is an outlier, which is affected more, the mean or the median?" mean

Example 2

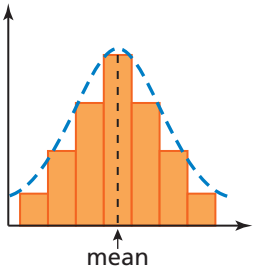
? "What type of distribution is given in part (a)?" skewed right

? "Which measure of central tendency would best represent the data and why?" median; The mean is in the direction the distribution is skewed, making it greater than the median and much of the data.

- **MP3 Construct Viable Arguments and Critique the Reasoning of Others:** The explanations that students offer provide additional practice in constructing viable arguments.
- Discuss the histogram in part (b) in a similar manner.

The shape of a distribution can be used to choose the most appropriate measure of central tendency that describes a data set.

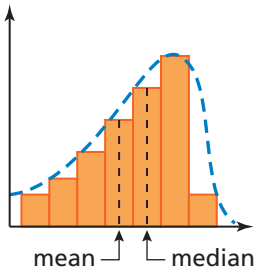
For a symmetric distribution, the mean and median are about the same, although the mean should be used to describe the center.

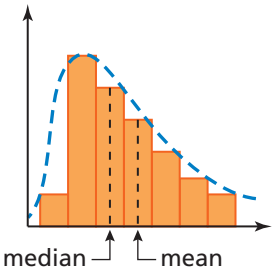


Remember

Outliers can affect the mean of a data set more than they affect the median.

When the distribution is skewed, the mean will be in the direction in which the distribution is skewed while the median will be less affected. So, when the data are skewed, use the median to describe the center.





EXAMPLE 2 Choosing an Appropriate Measure of Central Tendency

Which measure of central tendency best represents the data? Explain your reasoning.

a.

b.

a. Because the distribution is high on the left and the tail of the graph extends to the right, the distribution is skewed right. So, the median best represents the data.

b. Because the distribution is high in the middle and the data are about evenly distributed on both sides, the distribution is symmetric. So, the mean best represents the data.

On Your Own

2. Which measure of central tendency best represents the data in On Your Own Question 1? Explain your reasoning.

When a distribution is symmetric, use the standard deviation to describe the spread of the data set. When a distribution is skewed, use the five-number summary to describe the spread of the data set.

EXAMPLE 3 Choosing Appropriate Measures

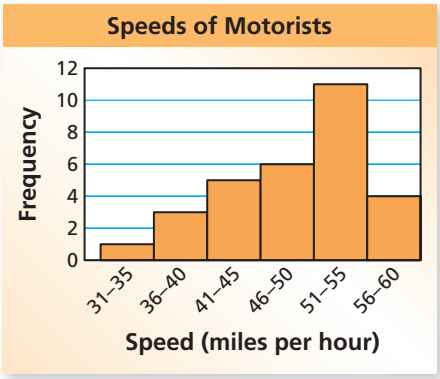
Speeds (mi/h)		
32	44	39
53	38	48
56	41	42
50	50	55
55	45	49
51	53	52
54	60	55
52	50	52
55	40	60
45	58	47

A police officer measures the speeds (in miles per hour) of 30 motorists. The results are shown in the table at the left.

a. Display the data in a histogram using six intervals beginning with 31–35.

Make a frequency table using the described intervals. Then use the frequency table to make a histogram.

Speed (mi/h)	Frequency
31–35	1
36–40	3
41–45	5
46–50	6
51–55	11
56–60	4



b. Which measures of central tendency and dispersion best represent the data?

Because the distribution is high on the right and the tail of the graph extends to the left, the distribution is skewed left. So, use the median to describe the center and the five-number summary to describe the spread.

c. The speed limit is 45 miles per hour. How would you interpret these results?

Because the distribution is skewed left, most of the speeds are more than 45 miles per hour. This shows that most of the motorists were speeding.

On Your Own

Now You're Ready
Exercises 8–11

3. You record the numbers of email attachments sent by 30 employees of a company in one week. Your results are shown in the table.

Email Attachments Sent				
74	105	98	68	64
85	75	60	48	51
65	55	58	45	38
64	52	65	30	70
72	5	45	77	83
42	25	95	16	120

a. Display the data in a histogram using six intervals beginning with 1–20.

b. Which measures of central tendency and dispersion best represent the data? Why?

Laurie's Notes

Discuss

- When choosing appropriate measures of center and spread, tell students that the standard deviation is a measure of dispersion that is related to the mean because the mean is used in its formula. The quartiles of a five-number summary are measures of dispersion that are related to the median.

Example 3

- Ask for a volunteer to read the problem as you write the frequency table. Have students describe the process for making the histogram.

? "How would you describe the shape of the distribution?" The distribution is skewed left.

- **Connection:** When students constructed box-and-whisker plots using the five-number summary, it gave a visual representation of the spread of the data.
- **Teaching Tip:** In part (c), it may be helpful to draw a dotted vertical line along the right side of the 41–45 bar. This helps students see that much of the data is to the right of the dotted line, so most of the motorists were speeding.

On Your Own

- This problem takes additional time because students must make the frequency table before making the histogram. Circulate to check that students are using the correct intervals.
- Ask a few students to do their work on transparencies if a document camera is not available.

Closure

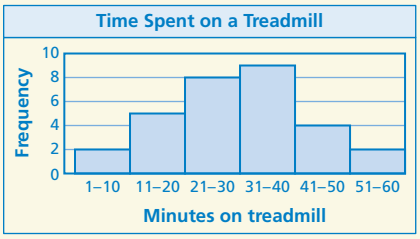
- Refer to the three data sets in the Motivate. Ask students questions about the most appropriate measures of central tendency and spread for each data set.

Extra Example 3

A personal trainer records the numbers of minutes that 30 clients run on a treadmill. The results are shown in the table.

Minutes on Treadmill		
25	20	35
33	50	22
28	15	34
60	50	8
25	20	37
6	39	24
30	45	15
40	25	45
18	60	30
40	35	40

a. Display the data in a histogram using six intervals beginning with 1–10.



b. Which measures of central tendency and dispersion best represent the data? mean and standard deviation

On Your Own

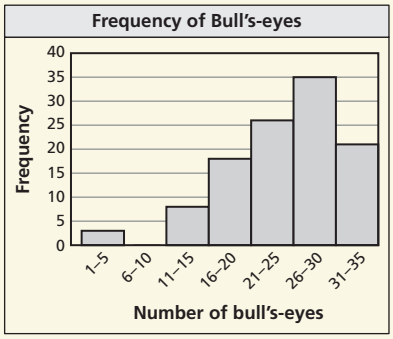
3. a. See Additional Answers.

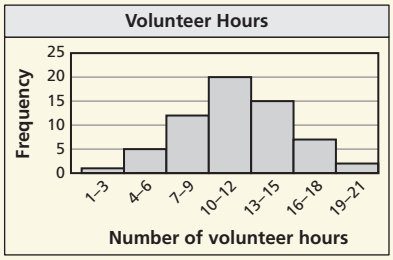
b. The distribution is symmetric. So, use the mean to describe the center and the standard deviation to describe the spread.

Vocabulary and Concept Check

1. The shape of a skewed distribution will have a tail on one side. The shape of a symmetric distribution is even, or symmetrical, with respect to the mean.
2. The mean should be used to describe the center of a symmetric distribution. The median should be used to describe the center of a skewed distribution.

Practice and Problem Solving

3. About 95%
4. About 95%
5. 

Most of the data are on the right. So, the distribution is skewed left.
6. 

Most of the data are in the middle and evenly distributed. So, the distribution is symmetric.
7. See Additional Answers.

Assignment Guide and Homework Check

Level	Assignment	Homework Check
Average	1, 2, 3–11 odd, 12–15, 17, 19–22	7, 11, 15, 17
Advanced	1, 2, 4–10 even, 16–18, 19–22	6, 10, 16, 18

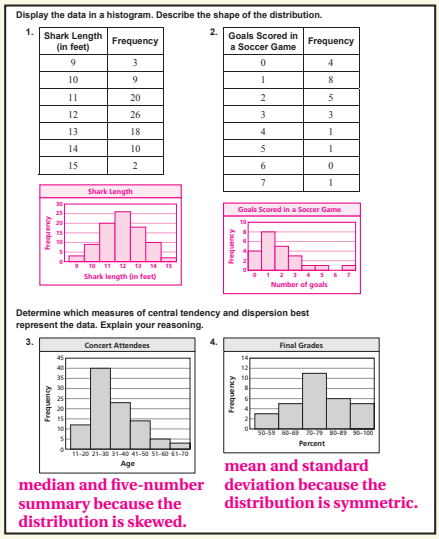
For Your Information

- **Exercise 16** Note in part (c) that as students add more IQ scores to the data set, the number of intervals will increase to the left. Students should visualize the distribution shape gradually changing from skewed right to symmetric.

Common Errors

- **Exercises 5–7** Students may create histograms with gaps between the bars. Remind students that the bars in a histogram should be touching to show that they represent an entire range of values.

12.4 Record and Practice Journal



12.4 Exercises

Check It Out
Help with Homework
BigIdeasMath.com

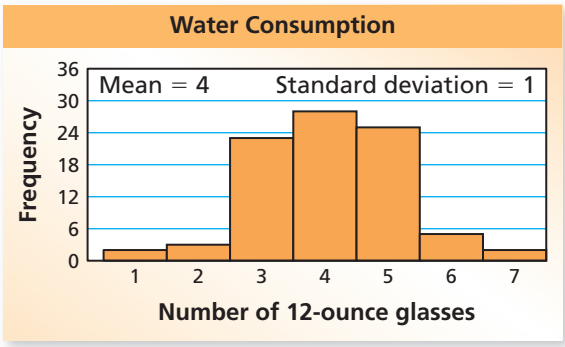
Vocabulary and Concept Check

1. **VOCABULARY** How does the shape of a symmetric distribution differ from the shape of a skewed distribution?
2. **WRITING** How does the shape of a distribution help you decide which measure of central tendency best describes the data?

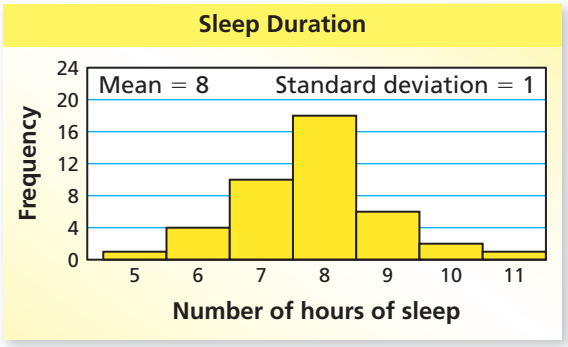
Practice and Problem Solving

Estimate the percent of data within 2 standard deviations of the mean.

3.



4.



Display the data in a histogram. Describe the shape of the distribution.

1

5.

Number of Bull's-eyes	Frequency
1–5	3
6–10	0
11–15	8
16–20	18
21–25	26
26–30	35
31–35	21

6.

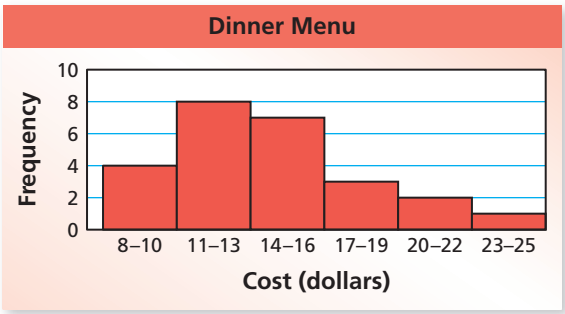
Number of Volunteer Hours	Frequency
1–3	1
4–6	5
7–9	12
10–12	20
13–15	15
16–18	7
19–21	2

7. **ONLINE** A survey asks people how many hours they spend online per day. The results are shown in the table. Display the data in a histogram. Describe the shape of the distribution.

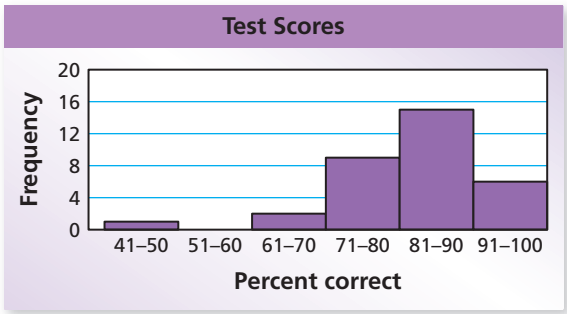
Hours Online	0–2	3–5	6–8	9–11	12–14
Frequency	33	45	12	4	2

Determine which measures of central tendency and dispersion best represent the data. Explain your reasoning.

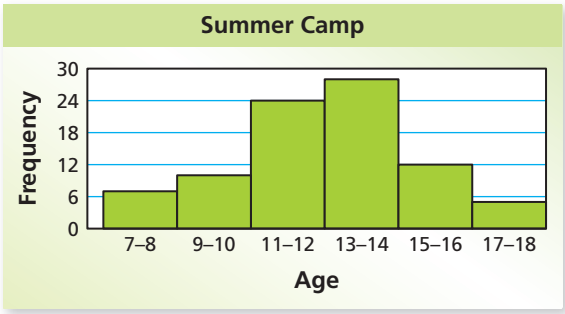
2 3 **8.**



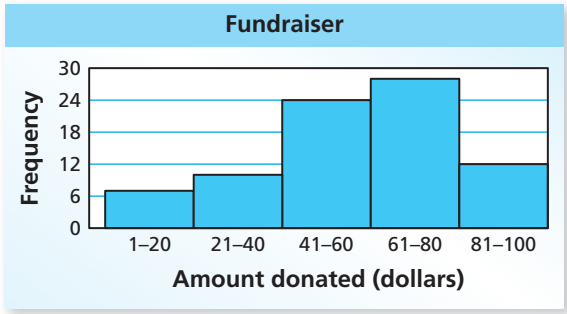
9.



10.

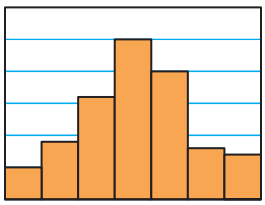


11.

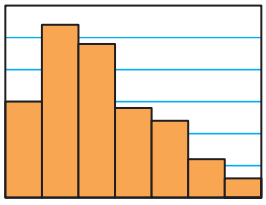


MATCHING **Match the distribution with the corresponding box-and-whisker plot.**

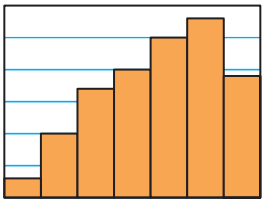
12.



13.



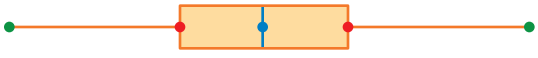
14.



A.

B.

C.



15. **CHOOSE TOOLS** A stem-and-leaf plot is another data display that shows the distribution of data. For a large data set, would you use a stem-and-leaf plot or a histogram to show the distribution of the data? Explain.

Common Errors

- **Exercises 16–18** Students may try to create a histogram without constructing a frequency table. Remind students to organize the data in a frequency table before attempting to create a histogram.
- **Exercises 16–18** Students may create histograms with gaps between the bars. Remind students that the bars in a histogram should be touching to show that they represent an entire range of values.

Practice and Problem Solving

8. Because the distribution is high on the left and the tail of the graph extends to the right, the distribution is skewed right. So, the median best represents the center of the data and the five-number summary best represents the spread of the data.

9. Because the distribution is high on the right and the tail of the graph extends to the left, the distribution is skewed left. So, the median best represents the center of the data and the five-number summary best represents the spread of the data.

10. Because the distribution is high in the middle and the data is about evenly distributed on both sides, the distribution is symmetric. So, the mean best represents the center of the data and the standard deviation best represents the spread of the data.

11. See Additional Answers.

12. C

13. A

14. B

15–16. See Additional Answers.

English Language Learners

Pair Activity

Pair English language learners with English speakers. Assign each person a problem. When they have completed their problems, they explain their solution to their partner who follows along. This will engage English language learners in conversation and help with understanding math concepts.

Practice and Problem Solving

17. See *Taking Math Deeper.*

18. See Additional Answers.

Fair Game Review

19. $y = -x + 4$

20. $y = -\frac{1}{3}x + 3$

21. $y = 2x - 7$ **22.** C

Mini-Assessment

1. The frequency table shows the family sizes of students in a classroom. Display the data in a histogram. Describe the shape of the distribution.

Family Size	Frequency
2–3	5
4–5	13
6–7	7
8–9	2
10–11	1

2. You record the numbers of hours that 30 students spent on homework during the past week. The results are shown in the table.

Hours of Homework		
18	22	8
20	23	20
24	19	25
13	5	21
21	24	12
27	15	4
16	17	29
18	21	14
14	25	8
23	10	18

a. Display the data in a histogram using six intervals beginning with 1–5.

b. Which measures of central tendency and dispersion best represent the data? Explain.

1–2. See Additional Answers.

Taking Math Deeper

Exercise 17

The challenge in this problem is finding the number of intervals for the histogram and the width of each interval.

1 Examine the data.

- There are 20 data values.
- All data values are multiples of 5 and 10.
- Largest withdrawal is \$100, smallest is \$10. So, the range is \$90.

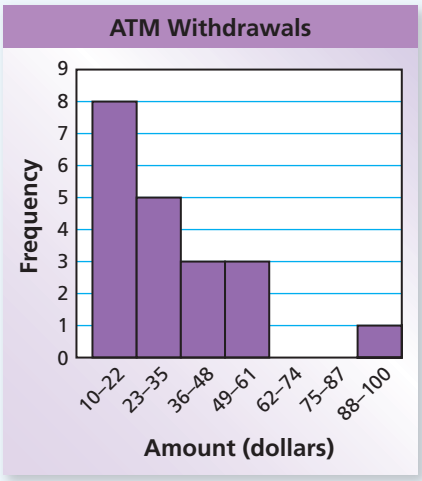
2 Find the number of intervals and the width of each interval.

Using too few or too many intervals may make it difficult to determine the shape of a distribution or find a pattern. You should have at least 5 intervals and because there are only 20 data values, you want to have no more than 7 or 8 intervals. So let's use 7 intervals.

Divide the range by the number of intervals to find each interval width. You get $90 \div 7 \approx 13$.

3 Make a frequency table and draw the histogram. Use the least value, 10, as the first number in the first interval. Then add 13 to find the first number of the second interval and so on.

Withdrawal Amount	Frequency
10–22	8
23–35	5
36–48	3
49–61	3
62–74	0
75–87	0
88–100	1



Reteaching and Enrichment Strategies

If students need help . . .	If students got it . . .
Resources by Chapter • Practice A and Practice B • Puzzle Time Record and Practice Journal Practice Differentiating the Lesson Lesson Tutorials Skills Review Handbook	Resources by Chapter • Enrichment and Extension • School-to-Work Start the next section

16. **MODELING** Measuring an IQ is an inexact science. However, IQ scores have been around for years in an attempt to measure human intelligence. The greatest known IQ scores are shown in the table.

IQ Scores		
170	190	180
160	180	210
154	170	180
195	230	160
170	186	180
225	190	170

a. Display the data in a histogram using five intervals beginning with 151–166.

b. Which measures of central tendency and dispersion best represent the data?

c. The distribution of IQ scores for the human population is symmetric. What happens to the shape of the distribution in part (a) as you include more and more IQ scores from the population in the data set?

17. **ATM** The table shows your last 20 ATM withdrawals. What intervals would you use to display the data in a histogram? Explain your reasoning. Then display the data in a histogram.

ATM Withdrawals (dollars)									
20	25	30	10	60	10	45	20	50	25
50	20	45	100	20	10	30	25	40	20

18. **Reasoning** You record the following waiting times at a restaurant.

Waiting Times (minutes)														
26	38	15	8	22	42	25	10	17	26	58	35	24	31	12
29	25	0	34	44	32	20	18	7	40	42	19	32	13	21

a. Display the data in a histogram using six intervals beginning with 0–9.

b. Display the data in a histogram using twelve intervals beginning with 0–4.

c. What happens when the number of intervals is increased?

d. Which histogram best represents the data? Explain your reasoning.

Fair Game Review What you learned in previous grades & lessons

Write an equation of the line that passes through the given point and is perpendicular to the given line. *(Section 2.6)*

19. $(2, 2); y = x + 3$ **20.** $(-3, 4); y = 3x - 1$ **21.** $(1, -5); y = -\frac{1}{2}x + 4$

22. **MULTIPLE CHOICE** Which equation represents the line that passes through $(0, 0)$ and is parallel to the line passing through $(5, -2)$ and $(1, -3)$? *(Section 2.6)*

A $y = -4x$ **B** $y = -\frac{1}{4}x$ **C** $y = \frac{1}{4}x$ **D** $y = 4x$

12 Study Help

Check It Out
Graphic Organizer
BigIdeasMath.com

You can use a **concept circle** to organize information about a concept. Here is an example of a concept circle for measures of central tendency.

Measures of Central Tendency

Concept
Measures that represent the center of a data set

Example
1, 3, 3, 5, 6, 18
Mean: $\frac{36}{6} = 6$
Median: 4
Mode: 3

Example
Remove the outlier.
1, 3, 3, 5, 6
Mean: $\frac{18}{5} = 3.6$
Median: 3
Mode: 3

Vocabulary
mean
median
mode

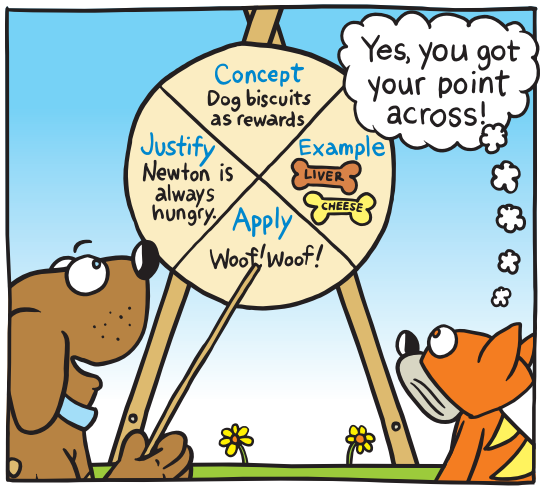
On Your Own

Make concept circles to help you study these topics.

1. measures of dispersion
2. box-and-whisker plots
3. shapes of distributions

After you complete this chapter, make concept circles for the following topics.

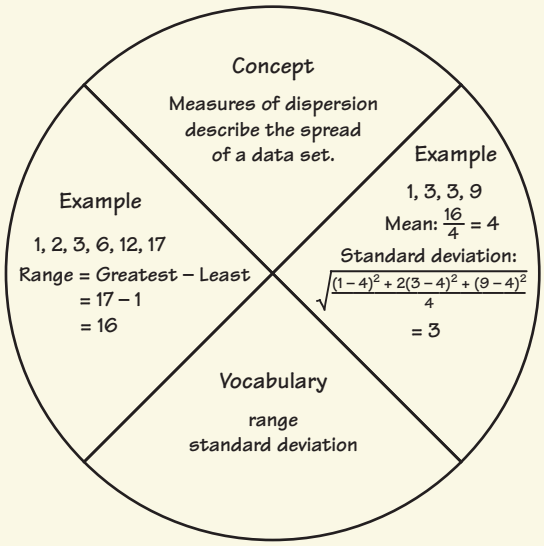
4. scatter plots
5. lines of fit
6. two-way tables
7. choosing a data display



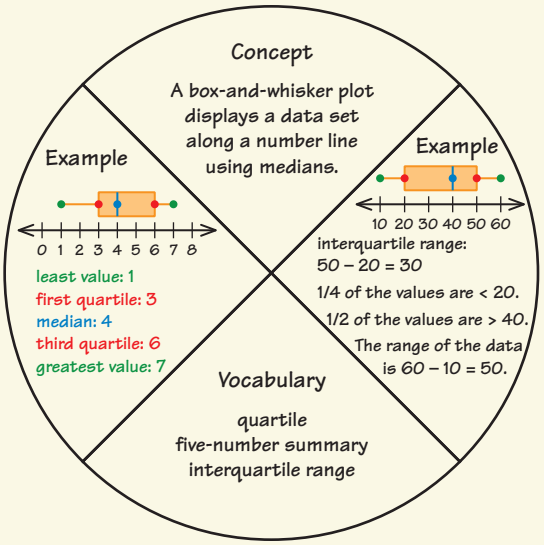
"Do you think this concept circle will help my owner understand that 'Speak' and 'Sit' need motivation?"

Sample Answers

1. Measures of Dispersion



2. Box-and-Whisker Plots



3. Available at *BigIdeasMath.com*.

List of Organizers

Available at *BigIdeasMath.com*

Comparison Chart
Concept Circle
Definition (Idea) and Example Chart
Example and Non-Example Chart
Formula Triangle
Four Square
Information Frame
Information Wheel
Notetaking Organizer
Process Diagram
Summary Triangle
Word Magnet
Y Chart

About this Organizer

A **Concept Circle** can be used to organize information about a concept. Students write the concept above the circle. Then students write associated information in the sectors of the circle. Associated information can include (an explanation of the) *Concept, Apply, Solve, Check, Example,* and *Justify*. Concept circles can have more or fewer than four sectors. Students can place their concept circles on note cards to use as a quick study reference.

Technology for the Teacher

Editable Graphic Organizer

Answers

1. mean: 20
   median: 30
   mode: 40

2. mean: 3

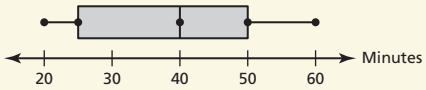
   median: $3\frac{1}{4}$

   mode: $3\frac{1}{2}$

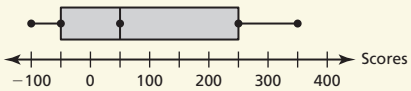
3. Girls: 3.8; 4; about 1.3
   Boys: 5.8; 6; about 1.9
   There tend to be 2 more boys absent each day than girls, and the number tends to differ from the mean slightly more for boys than for girls.

4. Juniors: 16.5; 11; about 3.4
   Seniors: 24; 17; about 5.7
   The seniors tend to score about 8 points per game more than the juniors and the point total tends to differ from the mean slightly more for seniors than for juniors.

5.



6.



7–9. See Additional Answers.

Alternative Quiz Ideas

100% Quiz	Math Log
Error Notebook	Notebook Quiz
Group Quiz	Partner Quiz
Homework Quiz	Pass the Paper

Group Quiz

Students work in groups. Give each group a large index card. Each group writes five questions that they feel evaluate the material they have been studying. On a separate piece of paper, students solve the problems. When they are finished, they exchange cards with another group. The new groups work through the questions on the card.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Resources by Chapter • Study Help • Practice A and Practice B • Puzzle Time Lesson Tutorials *BigIdeasMath.com*	Resources by Chapter • Enrichment and Extension • School-to-Work Game Closet at *BigIdeasMath.com* Start the next section

Technology for the Teacher

Online Assessment
Assessment Book
ExamView® Assessment Suite

12.1–12.4 Quiz

Check It Out
Progress Check
BigIdeasMath.com

Find the mean, median, and mode of the data. *(Section 12.1)*

1.

Checkbook Balances (dollars)		
40	10	−20
0	−10	40
30	40	50

2.

Hours Spent on Project		
$3\frac{1}{2}$	5	$2\frac{1}{2}$
3	$3\frac{1}{2}$	$\frac{1}{2}$

Find the mean, range, and standard deviation of each data set. Then compare the data sets. *(Section 12.2)*

3. Absent students during a week
Girls: 6, 2, 4, 3, 4
Boys: 5, 3, 6, 6, 9

4. Numbers of points scored
Juniors: 19, 15, 20, 10, 14, 21, 18, 15
Seniors: 22, 19, 29, 32, 15, 26, 30, 19

Make a box-and-whisker plot for the data. *(Section 12.3)*

5. Minutes of violin practice: 20, 50, 60, 40, 40, 30, 60, 40, 50, 20, 20, 35

6. Players' scores at end of first round: 200, −100, 100, 350, −50, 0, −50, 300

7. Display the data in a histogram. Describe the shape of the distribution. *(Section 12.4)*

Bowling Scores	51–100	101–150	151–200	201–250	251–300
Frequency	12	21	9	4	2

8. ANOLES The table shows the lengths of 12 green anoles. *(Section 12.1 and Section 12.3)*

a. Find the mean, median, and mode of the data.

b. Make a box-and-whisker plot for the data.

c. Find and interpret the interquartile range of the data.

d. Describe the distribution of the data.

e. How does including 8.0 in the data set affect the mean, median, and mode?

Length (cm)	17.5	17.3	16.5	16.8	17.0	16.5	17.0	16.7	16.5	17.0	17.4	17.1

9. PRESENTATIONS The times of 20 presentations are shown in the table. *(Section 12.2 and Section 12.4)*

a. Display the data in a histogram using five intervals beginning with 3–5.

b. Determine and calculate the measures of central tendency and dispersion that best represent the data.

c. The presentations are supposed to be 10 minutes long. How would you interpret these results?

Time (minutes)			
9	7	10	12
10	11	8	10
10	17	11	5
9	10	4	12
6	14	8	10

12.5 Scatter Plots and Lines of Fit

Essential Question How can you use data to predict an event?

1 ACTIVITY: Representing Data by a Linear Equation

Work with a partner. You have been working on a science project for 8 months. Each month, you have measured the length of a baby alligator.

The table shows your measurements.

	September							April
Month, x	0	1	2	3	4	5	6	7
Length (in.), y	22.0	22.5	23.5	25.0	26.0	27.5	28.5	29.5

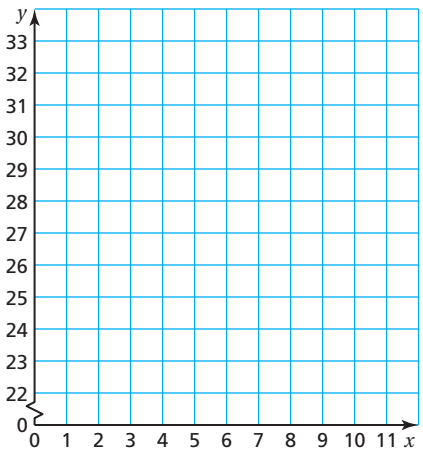
Use the following steps to predict the baby alligator's length next September.

a. Graph the data in the table.

b. Draw the straight line that you think best approximates the points.

c. Write an equation of the line you drew.

d. Use the equation to predict the baby alligator's length next September.



Common Core

Scatter Plots

In this lesson, you will
- interpret scatter plots.
- identify relationships from scatter plots.
- find lines of fit.
- solve real-life problems.

Learning Standards
8.SP.1
S.ID.6a
S.ID.6c

Laurie's Notes

Introduction

Standards for Mathematical Practice

- **MP1a Make Sense of Problems:** A scatter plot shows the relationship between two data sets. Students will find a line of fit for data and interpret what the equation means in the context of the problem.

Motivate

- Solicit information about what students know about alligators. Share alligator facts with students as a warm-up. (See the next page.)
- That should be enough information to set the context for this first activity.

Activity Notes

Activity 1

- "Look at the table of values. What do the ordered pairs represent?" (month, length of alligator)
- "Do the data represent the first 7 months of growth of a baby alligator? Explain." No, it does not suggest that this is from birth to age 7 months.
- "Are there any observations about the data in the table?" Months are increasing by 1. Lengths are increasing by about one-half to an inch each month.
- **MP6 Attend to Precision:** Students will ask what drawing a line "that best approximates the points" means. You should explain that it is a line that passes as closely as possible to all the points. Use a straightedge to lightly draw the line.
- "What does the jagged symbol at the bottom of the *y*-axis mean?" broken axis
- "Do you think everyone in class drew the exact same line? Explain." no; They will be close, but they do not have to be exactly the same.
- "How did you write the equation for the line?" Listen for an approximation of the slope (rise over run) and the *y*-intercept (close to 22). Write the equation in slope-intercept form.
- "Does everyone have the same slope?" no; They should be relatively close, however, and should match the observations made about the data when looking at the table.
- **MP1a:** Have students interpret the slope and *y*-intercept in the context of the problem.
- "How does the equation help you answer part (d)?" Substitute 12 for *x*, and find *y*.
- "Without the equation, can you predict the length of the alligator next September?" yes; You need to extend the graph and use eyesight to approximate the ordered pair.

Common Core State Standards

8.SP.1 Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data.
S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

Previous Learning

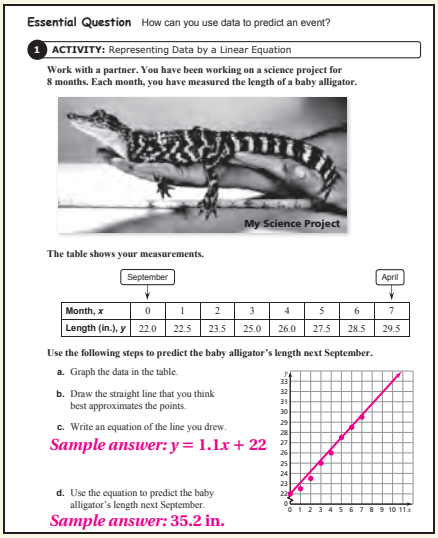
Students should know how to plot ordered pairs and write equations in slope-intercept form.

Technology for the Teacher

Dynamic Classroom

Lesson Plans
Complete Materials List

12.5 Record and Practice Journal



Essential Question How can you use data to predict an event?

1 ACTIVITY: Representing Data by a Linear Equation

Work with a partner. You have been working on a science project for 8 months. Each month, you have measured the length of a baby alligator.

My Science Project

The table shows your measurements.

Month, x	0	1	2	3	4	5	6	7
Length (in.), y	22.0	22.5	23.5	25.0	26.0	27.5	28.5	29.5

(September ↓ … April ↓)

Use the following steps to predict the baby alligator's length next September.

a. Graph the data in the table.

b. Draw the straight line that you think best approximates the points.

c. Write an equation of the line you drew.

Sample answer: $y = 1.1x + 22$

d. Use the equation to predict the baby alligator's length next September.

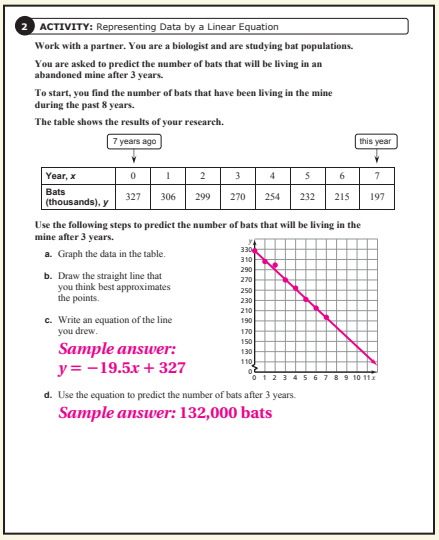
Sample answer: 35.2 in.

English Language Learners

Class Activity

Provide English learners with an opportunity to interact while learning the concept. Draw a coordinate plane on poster board. Label the horizontal axis *shoe size* and the vertical axis *height*. Have students place a sticker on the ordered pair that represents their shoe size and height. Then have the class fit a line to the data and write an equation of the line.

12.5 Record and Practice Journal



2 ACTIVITY: Representing Data by a Linear Equation

Work with a partner. You are a biologist and are studying bat populations.

You are asked to predict the number of bats that will be living in an abandoned mine after 3 years.

To start, you find the number of bats that have been living in the mine during the past 8 years.

The table shows the results of your research.

7 years ago → this year

Year, x	0	1	2	3	4	5	6	7
Bats (thousands), y	327	306	299	270	254	232	215	197

Use the following steps to predict the number of bats that will be living in the mine after 3 years.

a. Graph the data in the table.

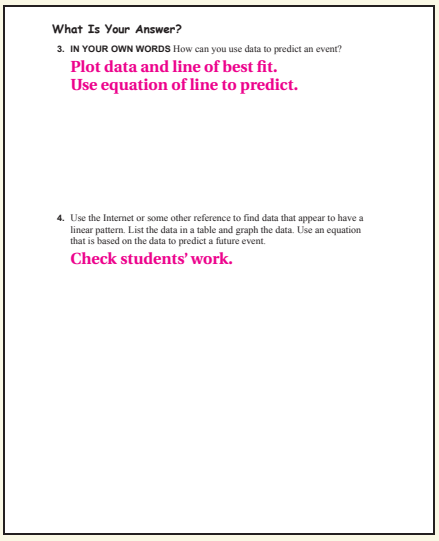
b. Draw the straight line that you think best approximates the points.

c. Write an equation of the line you drew.

Sample answer: $y = -19.5x + 327$

d. Use the equation to predict the number of bats after 3 years.

Sample answer: 132,000 bats



What Is Your Answer?

3. IN YOUR OWN WORDS How can you use data to predict an event?

Plot data and line of best fit.
Use equation of line to predict.

4. Use the Internet or some other reference to find data that appear to have a linear pattern. List the data in a table and graph the data. Use an equation that is based on the data to predict a future event.

Check students' work.

Laurie's Notes

Activity 2

- In this activity, the data have not been collected from an experiment. The data have been collected from a documented source and recorded in the table.
- Read the introduction. The purpose of making a scatter plot is stated. You want to make a prediction about the future by examining known data.

? "Are there any observations about the data in the table?" Students may recognize that as the years increase, the number of bats is decreasing by about 15–20 (in thousands) per year.

- Discuss equations written by students. Record students' results on the board. There will likely be a bit more variation of results than in the first activity.
- **MP1a:** Have students interpret what the slope and y-intercept mean in the context of the problem.
- **MP4 Model with Mathematics:** Discuss how the equation allows us to make predictions about the future.

What Is Your Answer?

- Question 4 can become a project due at the conclusion of the chapter.

Closure

- **Exit Ticket:** Describe the difference in the source of data for Activity 1 versus Activity 2. The data in Activity 1 are the result of gathering actual data from an experiment. The data in Activity 2 have been collected from a documented source and recorded in a table.

More about Alligators

- The American alligator (Alligator mississippiensis) is the largest reptile in North America. The first reptiles appeared 300 million years ago. Ancestors of the American alligator appeared 200 million years ago.
- The name alligator comes from early Spanish explorers who called them "El legarto" or "big lizard" when they first saw these giant reptiles.
- Louisiana and Florida have the most alligators. There are over one million wild alligators in each state with over a quarter million more on alligator farms.
- Alligators are about 10–12 inches in length when they are hatched from eggs. Growth rates vary from 2 inches per year to 12 inches per year, depending on the habitat, sex, size, and age of the alligator.
- Females can grow to about 9 feet in length and over 200 pounds. Males can grow to about 13 feet in length and over 500 pounds.
- The largest alligator was taken in Louisiana and measured 19 feet 2 inches.
- Alligators live about as long as humans, an average of 70 years.

2 ACTIVITY: Representing Data by a Linear Equation

Work with a partner. You are a biologist and are studying bat populations.

You are asked to predict the number of bats that will be living in an abandoned mine after 3 years.

To start, you find the number of bats that have been living in the mine during the past 8 years.

The table shows the results of your research.

7 years ago — this year

Year, x	0	1	2	3	4	5	6	7
Bats (thousands), y	327	306	299	270	254	232	215	197

Math Practice 6

Label Axes

When labeling axes of a coordinate plane, what information do you need to label the axes correctly? Why?

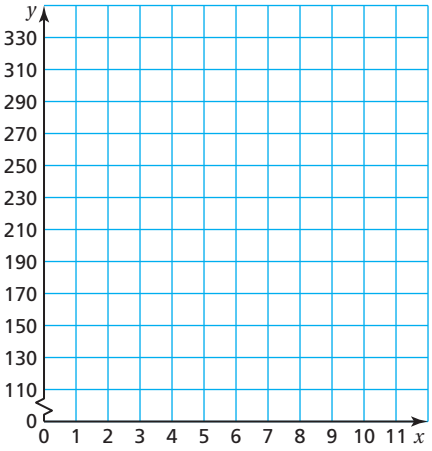
Use the following steps to predict the number of bats that will be living in the mine after 3 years.

a. Graph the data in the table.

b. Draw the straight line that you think best approximates the points.

c. Write an equation of the line you drew.

d. Use the equation to predict the number of bats after 3 years.



What Is Your Answer?

3. **IN YOUR OWN WORDS** How can you use data to predict an event?

4. Use the Internet or some other reference to find data that appear to have a linear pattern. List the data in a table and graph the data. Use an equation that is based on the data to predict a future event.

Practice Use what you learned about scatter plots and lines of fit to complete Exercise 3 on page 641.

12.5 Lesson

Check It Out
Lesson Tutorials
BigIdeasMath.com

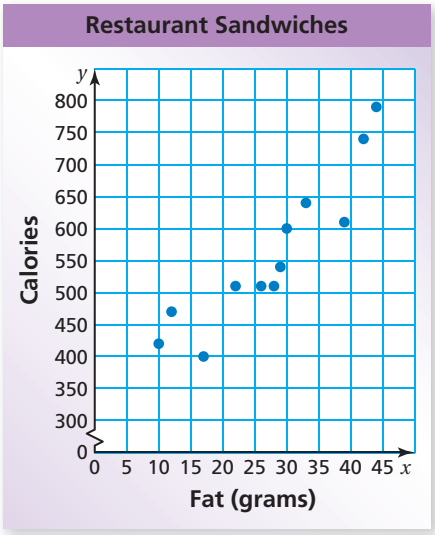
Key Vocabulary
scatter plot, *p. 638*
line of fit, *p. 640*

Key Idea

Scatter Plot

A **scatter plot** is a graph that shows the relationship between two data sets. The two sets of data are graphed as ordered pairs in a coordinate plane.

EXAMPLE 1 Interpreting a Scatter Plot

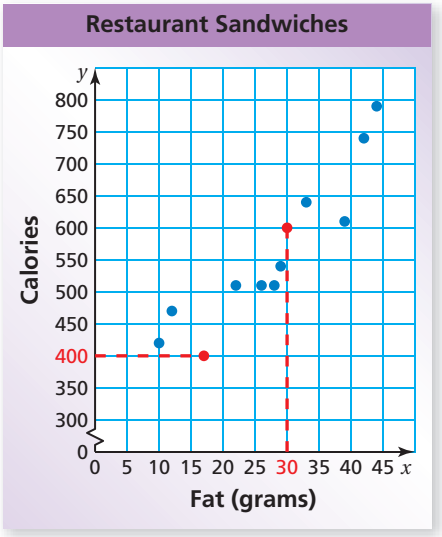


The scatter plot at the left shows the amounts of fat (in grams) and the numbers of calories in 12 restaurant sandwiches.

a. How many calories are in the sandwich that contains 17 grams of fat?

Draw a horizontal line from the point that has an x-value of 17. It crosses the y-axis at 400.

∴ So, the sandwich has 400 calories.



b. How many grams of fat are in the sandwich that contains 600 calories?

Draw a vertical line from the point that has a y-value of 600. It crosses the x-axis at 30.

∴ So, the sandwich has 30 grams of fat.

c. What tends to happen to the number of calories as the number of grams of fat increases?

Looking at the graph, the plotted points go up from left to right.

∴ So, as the number of grams of fat increases, the number of calories increases.

On Your Own

Now You're Ready
Exercises 4 and 5

1. **WHAT IF?** A sandwich has 650 calories. Based on the scatter plot in Example 1, how many grams of fat would you expect the sandwich to have? Explain your reasoning.

Multi-Language Glossary at BigIdeasMath.com

Laurie's Notes

Goal Today's lesson is making a **scatter plot** and identifying a **line of fit**.

Introduction

Connect

- **Yesterday:** Students gained an intuitive understanding of how to construct scatter plots and write an equation a line of fit. (MP1a, MP4, MP6)
- **Today:** Students will construct scatter plots, draw the line of fit, and analyze the equation.

Motivate

- **Preparation:** Stop by any fast food restaurant to pick up a pamphlet, or go online to find nutritional information about the menu items.
- ? "Do you think there is a relationship between the grams of fat and number of calories in the sandwich?" yes
- Share the information about a few of the sandwiches from your pamphlet or printout to confirm students' opinions.

Lesson Notes

Key Idea

- Explain that the plot they are going to make today displays the relationship, if any, between two variables, such as grams of fat and calories.
- Define scatter plot.
- Discuss the two scatter plots made in the activity. In Activity 1, the two sets of data were months and alligator length. In Activity 2, the two sets of data were years and number of bats.
- Point out that a scatter plot differs from previous data displays in that it is bivariate (paired data).

Example 1

- **MP1a Make Sense of Problems:** This example helps students understand how a scatter plot is read and interpreted. Discuss the labels on the axes and what an ordered pair represents: (grams of fat, number of calories). There are 12 different sandwiches that are represented.
- To read information from the plot, move horizontally to the x-value, find the ordered pair, and then move to the y-axis to read the y-value. It is helpful to use your hands to demonstrate the motion.
- A scatter plot allows you to see trends in the data. You read a scatter plot from left to right. As the x-coordinate increases, is the y-coordinate increasing, decreasing, staying the same, *or* is there no pattern?

On Your Own

- This question implies that because you can see a particular trend in the data, you are able to make estimates about points which are not part of the data set but would fall within the trend in the data. Although it is possible that the 650 calorie sandwich has 10 grams of fat, you would not predict it based upon this scatter plot.

Technology for the Teacher

Dynamic Classroom

Lesson Tutorials
Lesson Plans
Answer Presentation Tool

Extra Example 1

Use the scatter plot in Example 1.

a. How many grams of fat are in a sandwich that contains 740 calories? about 42 g

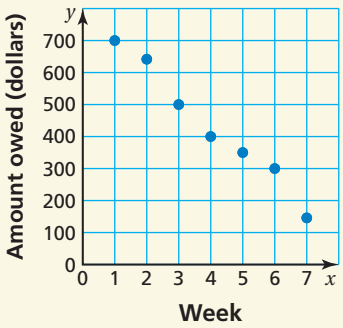
b. How many calories are in a sandwich that contains 33 grams of fat? about 640 calories

On Your Own

1. about 35 g; The point just below $y = 650$ has an x-value just below $x = 35$.

Extra Example 2

Tell whether the data show a *positive*, a *negative*, or *no* relationship.

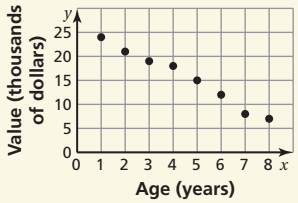


negative relationship

On Your Own

2. See Additional Answers.

3.



negative relationship

Differentiated Instruction

Visual

Some students may find it easier to draw a line of fit before determining if the data have a positive relationship, a negative relationship, or no relationship. A line with a *positive* slope means the data have a *positive* relationship. A line with a *negative* slope means the data have a *negative* relationship. If a line cannot be drawn, the data have *no* relationship.

Laurie's Notes

Discuss

- There are three general cases that describe the relationship between two data sets. Draw a quick example of each case.
- The alligator data was an example of a positive relationship and the bat data was an example of a negative relationship.
- Discuss the study tip and provide examples to show how these features can be seen in a scatter plot.

Example 2

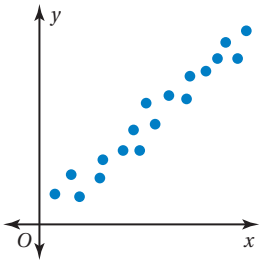
- Have students review the two scatter plots shown.
- Ask students to complete this sentence. As the size of the television increases, the price increases. This is an example of a positive relationship.
- **Connection:** By this point, some students have made the connection between the slope of a line of fit from the activities and the relationship between the two data sets. A positive relationship is related to a positive slope.
- ? "Should there be a relationship between a person's age and the number of pets they own?" no
- Part (b) makes sense to students. There should be no trend in the data.

On Your Own

- Give time for students to complete these two scatter plots.
- **MP6 Attend to Precision:** A common difficulty for students is deciding how to scale the axes. Students should look at the range of numbers that need to be displayed, and then decide if it is necessary to start their axes at 0 or if another starting point (broken axes) makes sense.
- Have transparency grids available so that results can be shared quickly as a class.

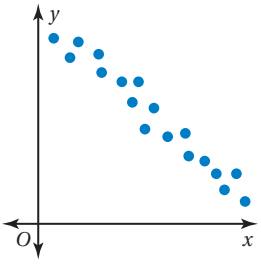
A scatter plot can show that a relationship exists between two data sets.

Positive Relationship



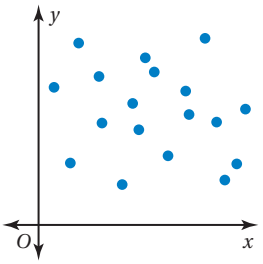
As x increases, y increases.

Negative Relationship



As x increases, y decreases.

No Relationship



The points show no pattern.

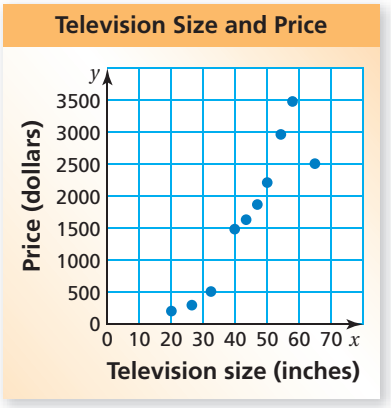
Study Tip

Scatter plots can also show unusual features of a data set, such as outliers, or gaps and clusters in the data.

EXAMPLE 2 Identifying a Relationship

Tell whether the data show a *positive*, a *negative*, or *no* relationship.

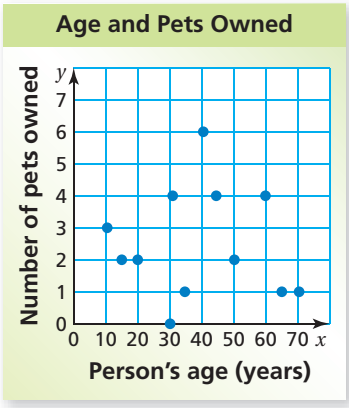
a. Television size and price



As the size of the television increases, the price increases.

So, the scatter plot shows a positive relationship.

b. Age and number of pets owned



The number of pets owned does not depend on a person's age.

So, the scatter plot shows no relationship.

On Your Own

Now You're Ready
Exercises 6–8

Make a scatter plot of the data. Tell whether the data show a *positive*, a *negative*, or *no* relationship.

2.

Study Time (min), x	30	20	60	90	45	10	30	75	120	80
Test Score, y	87	74	92	97	85	62	83	90	95	91

3.

Age of a Car (years), x	1	2	3	4	5	6	7	8
Value (thousands), y	\$24	\$21	\$19	\$18	\$15	\$12	\$8	\$7

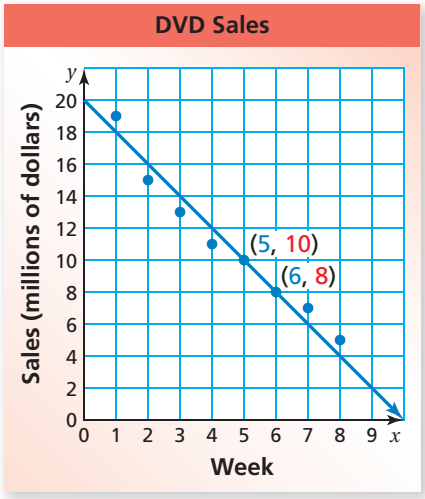
A **line of fit** is a line drawn on a scatter plot close to most of the data points. It can be used to estimate data on a graph.

EXAMPLE 3 Finding a Line of Fit

Week, x	Sales (millions), y
1	\$19
2	\$15
3	\$13
4	\$11
5	\$10
6	\$8
7	\$7
8	\$5

The table shows the weekly sales of a DVD and the number of weeks since its release. (a) Make a scatter plot of the data and draw a line of fit. (b) Write an equation of the line of fit. (c) Interpret the slope of the line of fit. (d) Predict the sales in week 9.

a. Plot the points in a coordinate plane. The scatter plot shows a negative relationship. Draw a line that is close to the data points. Try to have as many points above the line as below it.



b. The line passes through (5, 10) and (6, 8).

$$\text{slope} = \frac{\text{rise}}{\text{run}} = \frac{-2}{1} = -2$$

Because the line crosses the y-axis at (0, 20), the y-intercept is 20.

∴ So, an equation of the line of fit is $y = -2x + 20$.

c. The slope of the line of fit is -2. This means that the sales are decreasing by about \$2 million each week.

d. To predict the sales in week 9, substitute 9 for x in the equation of the line of fit.

$$y = -2x + 20 = -2(9) + 20 = 2$$

∴ The sales in week 9 should be about \$2 million.

Study Tip

A line of fit does not need to pass through any of the data points.

On Your Own

Now You're Ready
Exercise 14

4. The table shows the numbers of people who have attended a neighborhood festival over an 8-year period.

Year, x	1	2	3	4	5	6	7	8
Attendance, y	420	500	650	900	1100	1500	1750	2400

a. Make a scatter plot of the data and draw a line of fit.

b. Write an equation of the line of fit.

c. Interpret the slope of the line of fit.

d. Predict the number of people who will attend the festival in year 10.

Laurie's Notes

Discuss

- **MP5 Use Appropriate Tools Strategically:** Define and discuss a line of fit. It is helpful to model this with a piece of spaghetti. Use a scatter plot from the previous page that was completed on the transparency. Model how the spaghetti can approximate the trend of the data.
- Move the spaghetti so that it does *not* represent the data, and then move the spaghetti so that it does. You will use your eyesight when judging where to draw the line.

Example 3

? "What observations can you make about the sales of the DVD as the weeks go on?" Sales are decreasing.

- Carefully plot the ordered pairs on a transparency grid.
- When drawing a line of fit, try to put as many points above the line as below it.
- Students may draw different lines of fit and still get a reasonable answer.
- In this example, the line passes through two actual data points, (5, 10) and (6, 8). As noted in the *Study Tip*, a line of fit does not need to pass through any of the data points.
- Finish working the problem as shown.
- **MP4 Model with Mathematics:** The purpose of writing the equation of the line of fit is to make predictions. The equation becomes a model for the data, describing its behavior.

On Your Own

- This is a nice summary problem. Students should quickly observe the positive relationship just from the table of values.
- Share results of this problem as a whole class.

Closure

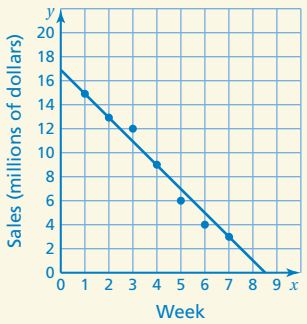
- **Exit Ticket:** In Example 3, interpret the y-intercept in the context of the problem. Does it make sense in this problem? The y-intercept, 20, represents the sales in millions of dollars for week 0. It does not make sense in this problem, because there would not have been any sales before the first week.

Extra Example 3

The table shows the weekly sales of a DVD and the number of weeks since its release.

Week, x	Sales (millions), y
1	\$15
2	\$13
3	\$12
4	\$9
5	\$6
6	\$4
7	\$3

a. Make a scatter plot of the data and draw a line of fit.



b. Write an equation of the line of fit.
$y = -2x + 17$

c. Interpret the slope of the line of fit.
The slope of the line of fit is -2. This means the sales are decreasing by about \$2 million each week.

d. Predict the sales in week 8.
about \$1 million

On Your Own

4. a. See Additional Answers.

b. *Sample answer:* $y = 270x$

c. *Sample answer:* The slope of the line of fit is 270. This means the number of people attending is increasing by about 270 people each year.

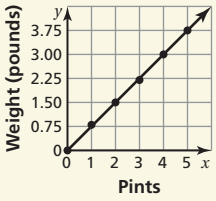
d. *Sample answer:* about 2700 people

Vocabulary and Concept Check

1. They must be ordered pairs so there are equal amounts of x- and y-values.
2. You can estimate and predict values.

Practice and Problem Solving

3. a–b.



c. *Sample answer:* $y = 0.75x$

d. *Sample answer:* 7.5 lb

e. *Sample answer:* $16.88

4. a. 2007

b. about 875 SUVs

c. There is a negative relationship between year and number of SUVs sold.

5. a. 3.5 h

b. $85

c. There is a positive relationship between hours worked and earnings.

6. negative relationship

7. positive relationship

8. no relationship

9. positive relationship

10. *Sample answer:* bank account balance after a shopping spree

11. *Sample answer:* not a good representation; Too many points in the data set lie below the line.

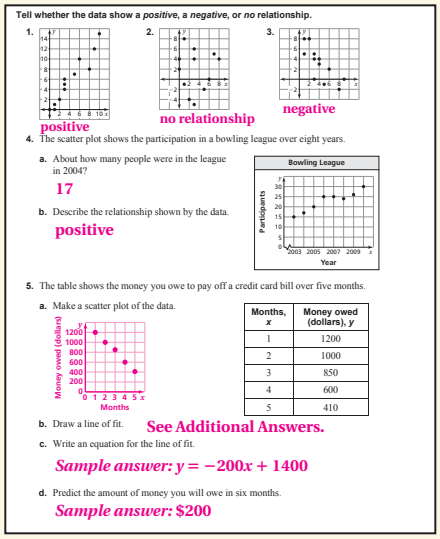
Assignment Guide and Homework Check

Level	Assignment	Homework Check
Average	1, 2, 4–18 even, 15, 17, 23–26	4, 6, 12, 14, 17
Advanced	1, 2, 4–22 even, 17, 23–26	4, 6, 14, 17, 18

Common Errors

- **Exercise 3** Students may use inconsistent increments or forget to label their graphs. Students should use consistent increments to represent the data. Remind them to label the axes so that information can be read from the graph.
- **Exercises 4 and 5** When finding values from the graph, students may accidentally shift over or up too far and get an answer that is off by an increment. Encourage them to start at the given value and trace the graph to where the point or line of fit is, and then trace down or left to the other axis for the answer.

12.5 Record and Practice Journal



Tell whether the data show a *positive*, a *negative*, or *no* relationship.

1. positive
2. no relationship
3. negative

4. The scatter plot shows the participation in a bowling league over eight years.

a. About how many people were in the league in 2004?
17

b. Describe the relationship shown by the data.
positive

Bowling League
Participants
Year

5. The table shows the money you owe to pay off a credit card bill over five months.

a. Make a scatter plot of the data.

Months, x	Money owed (dollars), y
1	1200
2	1000
3	850
4	600
5	410

Money owed (dollars)
Months

b. Draw a line of fit. See Additional Answers.

c. Write an equation for the line of fit.
Sample answer: $y = -200x + 1400$

d. Predict the amount of money you will owe in six months.
Sample answer: $200

12.5 Exercises

Check It Out
Help with Homework
BigIdeasMath.com

Vocabulary and Concept Check

1. **VOCABULARY** What type of data are needed to make a scatter plot? Explain.
2. **WRITING** Explain why a line of fit is helpful when analyzing data.

Practice and Problem Solving

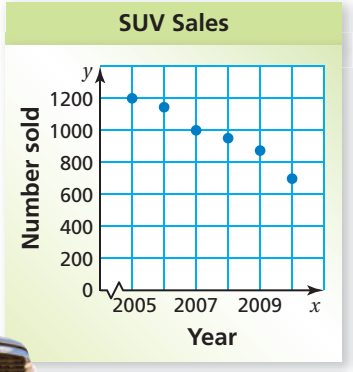
3. **BLUEBERRIES** The table shows the weights y of x pints of blueberries.

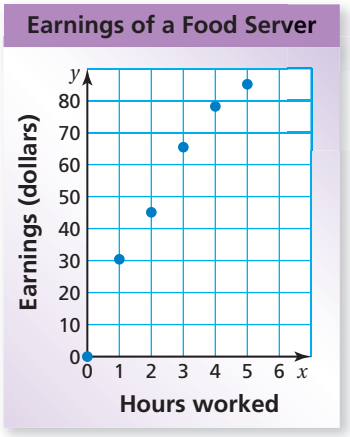
Number of Pints, x	0	1	2	3	4	5
Weight (pounds), y	0	0.8	1.50	2.20	3.0	3.75

   a. Graph the data in the table.
   b. Draw the straight line that you think best approximates the points.
   c. Write an equation of the line you drew.
   d. Use the equation to predict the weight of 10 pints of blueberries.
   e. Blueberries cost \$2.25 per pound. How much do 10 pints of blueberries cost?

4. **SUVS** The scatter plot shows the numbers of sport utility vehicles sold in a city from 2005 to 2010.

   a. In what year were 1000 SUVs sold?
   b. About how many SUVs were sold in 2009?
   c. Describe the relationship shown by the data.



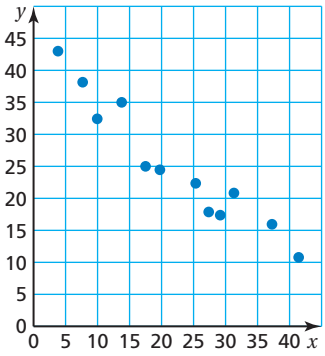


5. **EARNINGS** The scatter plot shows the total earnings (wages and tips) of a food server during 1 day.

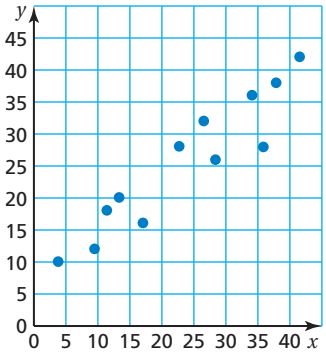
   a. About how many hours must the server work to earn \$70?
   b. About how much did the server earn for 5 hours of work?
   c. Describe the relationship shown by the data.

Tell whether the data show a *positive*, a *negative*, or *no* relationship.

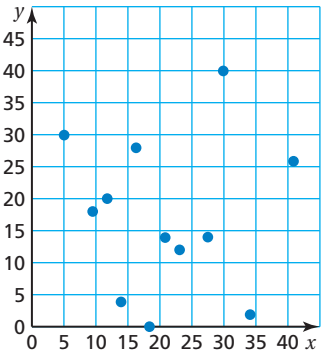
2 **6.**



7.



8.



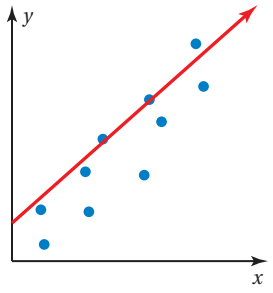
9. HONEY The table shows the average price per pound for honey in the United States from 2007 to 2010. What type of relationship do the data show?

Year, *x*	2007	2008	2009	2010
Average Price per Pound, *y*	$1.08	$1.42	$1.47	$1.60

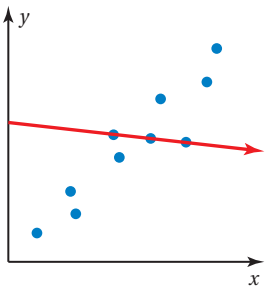
10. OPEN-ENDED Describe a set of real-life data that has a negative relationship.

Tell whether the line drawn on the graph is a good fit for the data. Explain your reasoning.

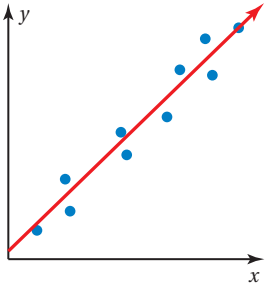
11.



12.



13.



3 **14. VACATION** The table shows the distance you travel over a 6-hour period.

Hours, *x*	Distance (miles), *y*
1	62
2	123
3	188
4	228
5	280
6	344

a. Make a scatter plot of the data and draw a line of fit.

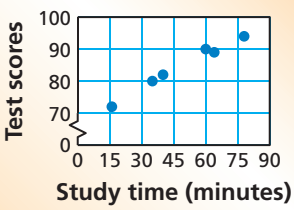
b. Write an equation of the line of fit.

c. Interpret the slope of the line of fit.

d. Predict the distance you will travel in 7 hours.

15. TEST SCORES The scatter plot shows the relationship between numbers of minutes spent studying and test scores for a science class.

Study Time and Test Scores

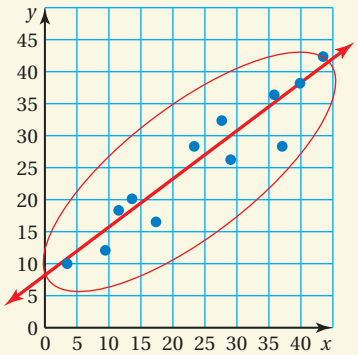


a. What type of relationship do the data show?

b. Interpret the relationship.

Common Errors

- **Exercises 6–8** Students may mix up positive and negative relationships. Remind them about slope. The slope is positive when the line rises from left to right and negative when it falls from left to right. The same is true for relationships in a scatter plot. If the data rises from left to right, it is a positive relationship. If it falls from left to right, it is a negative relationship.
- **Exercise 14** Students may draw a line of fit that does not accurately reflect the data trend. Remind them that the line does not have to go through any of the data points. Also remind them that the line should go through the middle of the data so that about half of the data points are above the line and half are below. One strategy is to draw an oval around the data and then draw a line through the middle of the oval. For example:



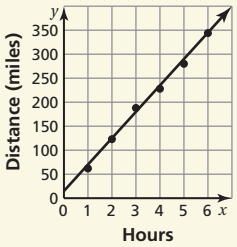
- **Exercise 14** Students may struggle writing an equation for a line of fit. When drawing the line, encourage them to try to make the line go through a lattice point. Also, students can use lattice points that are very close to the line to help them find the slope.

Practice and Problem Solving

12. *Sample answer:* not a good representation; Even though the line passes through several points, it does not indicate the overall relationship of the data set.

13. *Sample answer:* good representation; The same number of points in the data set lie above and below the line.

14. a.



b. *Sample answer:* $y = 55x + 15$

c. *Sample answer:* The slope of the line of fit is 55. This means the number of miles driven each hour is about 55.

d. *Sample answer:* 400 mi

15. a. positive relationship

b. The more time spent studying, the better the test score.

Differentiated Instruction

Kinesthetic

Form groups of 8 to 10 students who will create life-size models of a positive relationship, a negative relationship, and no relationship. Give two pairs of students 10-foot lengths of string and have them form the *x*- and *y*-axes of a coordinate plane. The remaining students will be the data points. Have these students represent a *positive relationship* in the coordinate plane. After students have had a few minutes, check their positions. Continue by having students represent a *negative relationship* and *no relationship*. Extend the activity by having two students hold a third string to show a line of fit.

Practice and Problem Solving

16. no; There is no line that lies close to most of the points.

17. See *Taking Math Deeper*.

18. a data point that is far removed from the other points in a data set

19–22. See Additional Answers.

Fair Game Review

23. 2 **24.** 8

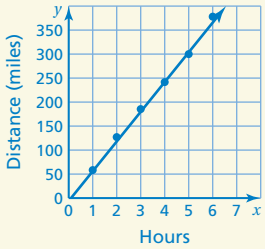
25. −4 **26.** B

Mini-Assessment

The table shows the distance you travel over a 6-hour period.

Hours, x	Distance (miles), y
1	60
2	130
3	186
4	244
5	300
6	378

a. Make a scatter plot of the data and draw a line of fit.



b. Write an equation of the line of fit. *Sample answer:* $y = 60x$

c. Interpret the slope of the line of fit. The slope of the line of fit is 60. This means the number of miles traveled each hour is about 60.

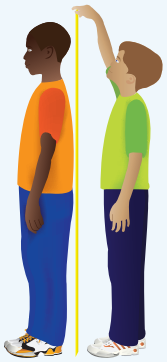
d. Predict the distance traveled after 7 hours. *Sample answer:* about 420 mi

Taking Math Deeper

Exercise 17

The project in the student text is described so that it can be assigned as homework. Another way to assign the project is to ask students to do the project in class.

1. Gather the data by having students measure each other's height and arm span.



2. Plot the data for the entire class in a coordinate plane. Scale the x-axis and y-axis so that the measurements of your students (in inches) will fit.

3. 

Ask your students to describe the relationship between height and arm span. The slope of the line of fit should be approximately equal to 1.

Project

Research Leonardo da Vinci's drawing of the *Vitruvian Man*. Explain the concept behind the drawing.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Resources by Chapter • Practice A and Practice B • Puzzle Time Record and Practice Journal Practice Differentiating the Lesson Lesson Tutorials Skills Review Handbook	Resources by Chapter • Enrichment and Extension • School-to-Work • Financial Literacy Start the next section

16. **REASONING** A data set has no relationship. Is it possible to find a line of fit for the data? Explain.

17. **CHOOSE TOOLS** Use a ruler or a yardstick to find the heights and arm spans of three people.

a. Make a scatter plot using the data you collected. Then draw a line of fit for the data.

b. Use your height and the line of fit to predict your arm span.

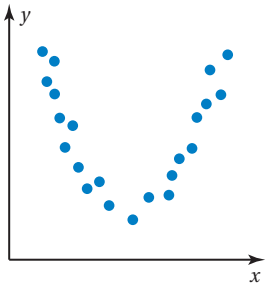
c. Measure your arm span. Compare the result with your prediction in part (b).

d. Is there a relationship between a person's height x and arm span y? Explain.

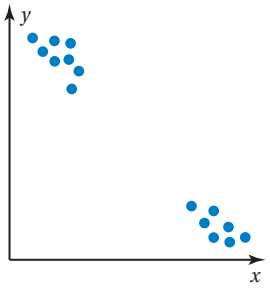
18. **REASONING** How can an outlier be identified in a scatter plot?

Describe the scatter plot and any relationship between the variables.

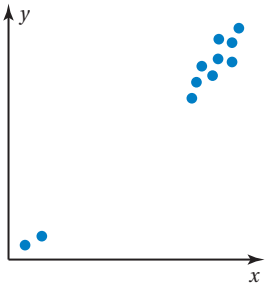
19.



20.



21.



Price of Admission (dollars), x	Yearly Attendance, y
19.50	50,000
21.95	48,000
23.95	47,500
24.00	40,000
24.50	45,000
25.00	43,500

22. Critical Thinking The table shows the prices of admission to a local theater and the attendances for several years.

a. Identify the outlier.

b. How does the outlier affect the line of fit? Explain.

c. Make a scatter plot of the data and draw the line of fit.

d. Use the line of fit to predict the attendance when the admission cost is \$27.

Fair Game Review What you learned in previous grades & lessons

Use a graph to solve the equation. Check your solution. *(Section 4.4)*

23. $5x = 2x + 6$

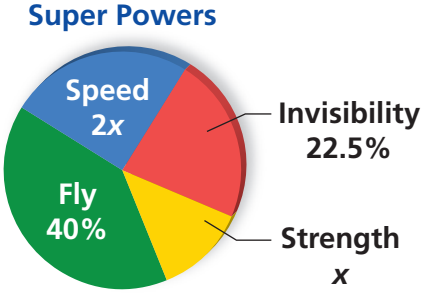
24. $7x + 3 = 9x - 13$

25. $\frac{2}{3}x = -\frac{1}{3}x - 4$

26. **MULTIPLE CHOICE** The circle graph shows the super powers chosen by a class. What percent of the students want strength as their super power? *(Skills Review Handbook)*

Ⓐ 10.5% Ⓑ 12.5%

Ⓒ 15% Ⓓ 25%



12.6 Analyzing Lines of Fit

Essential Question How can you find a line that best models a data set?

1 ACTIVITY: Comparing Lines of Fit

Work with a partner. You are researching the prices of liquid crystal display (LCD) televisions. The tables show the sizes and prices of several LCD televisions.

TV Size (in.), x	19	19	22	24	32
Price (dollars), y	170	180	170	250	320

TV Size (in.), x	32	37	40	40	46
Price (dollars), y	300	400	480	500	600

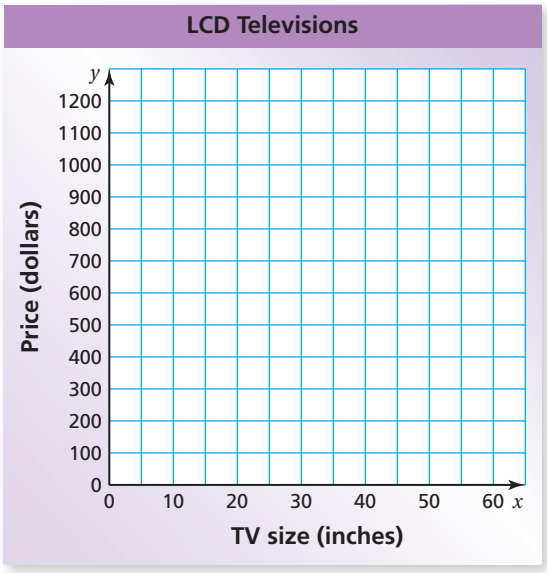
TV Size (in.), x	46	47	52	55	55
Price (dollars), y	850	800	950	1000	1150

a. Make a scatter plot of the data. Describe the pattern.

b. Draw a line of fit. Then have your partner draw a different line of fit.

c. Write an equation for each line of fit.

d. Compare your line of fit with your partner's line of fit. Are they similar? Which line of fit seems to model the data better? Why?



Common Core

Lines of Fit

In this lesson, you will

- use residuals to determine whether models are a good fit.
- fine lines of best fit using technology.
- identify correlations and causations.

Learning Standards
S.ID.6b
S.ID.8
S.ID.9

2 ACTIVITY: Choosing a Line of Fit

Compare your line of fit with the lines of fit of the other students in your class. Which line of fit do you think best models the data? What criteria did you use when choosing the line of fit? Explain your reasoning.

Laurie's Notes

Introduction

Standards for Mathematical Practice

- **MP5 Use Appropriate Tools Strategically:** To understand lines of best fit, it is important for students to fit a line by hand first. Once this is done, the regression capabilities of a graphing calculator can be used to find the line of best fit. It is important to explore the correlation coefficient to gauge how closely the equation models the data.

Motivate

- Write the following table on the board.

Number of weeks	1	2	3	4	6	7
Cumulative snowfall (in.)	11	18	32	52	73	82

- Use the data in the table to review the process of making a scatter plot using a graphing calculator. The important skills are entering the data, selecting the scatter plot option, setting an appropriate viewing window, and making the plot.

❓ "Do you think you could draw a line that models the data?" yes

❓ "Could your line help me estimate the cumulative snowfall in week 5? Explain." yes; Use the equation to find the y-value when $x = 5$.

Activity Notes

Activity 1

- Non-contextual scatter plots are often quick and easy to graph because they have *nice* domains and ranges. This contextual problem has a wide range. Students need to be as accurate as possible when plotting points.
- **Teaching Tip:** Use spaghetti to visualize lines of fit.
- A line of fit does not have to pass through two of the data points, however, students at this level may find it easier if it does.
- Discuss the results when students have finished the activity.

❓ "What does the slope mean in the context of the problem?" It is the average increase in price for each 1-inch increase in TV size.

Activity 2

- To help facilitate this activity, write the equations from many different students on the board.
- Give ample time for students to view all of the equations.
- While students are discussing, enter several of the equations in a graphing calculator, along with the ordered pairs.

❓ "Which line do you think best models the data?" Answers will vary.

❓ "What criteria did you use?" Student explanations should include a discussion of points being above and below the line.

Common Core State Standards

S.ID.6b Informally assess the fit of a function by plotting and analyzing residuals.
S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.
S.ID.9 Distinguish between correlation and causation.

Previous Learning

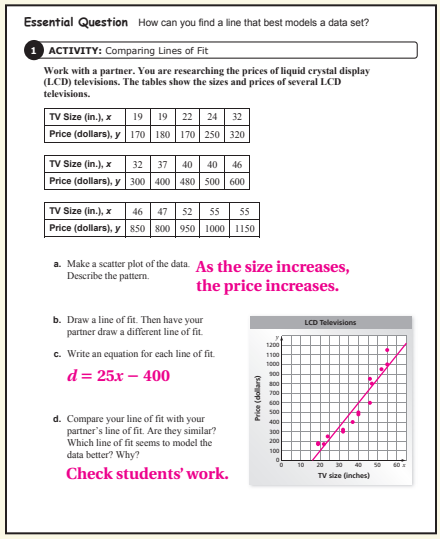
Students should know how to draw and write equations of lines of fit.

Technology for the Teacher

Dynamic Classroom

Lesson Plans
Complete Materials List

12.6 Record and Practice Journal



Essential Question How can you find a line that best models a data set?

1 ACTIVITY: Comparing Lines of Fit

Work with a partner. You are researching the prices of liquid crystal display (LCD) televisions. The tables show the sizes and prices of several LCD televisions.

TV Size (in.), x	19	19	22	24	32
Price (dollars), y	170	180	170	250	320

TV Size (in.), x	32	37	40	40	46
Price (dollars), y	300	400	480	500	600

TV Size (in.), x	46	47	52	55	55
Price (dollars), y	850	800	950	1000	1150

a. Make a scatter plot of the data. Describe the pattern. As the size increases, the price increases.

b. Draw a line of fit. Then have your partner draw a different line of fit.

c. Write an equation for each line of fit. $d = 25x - 400$

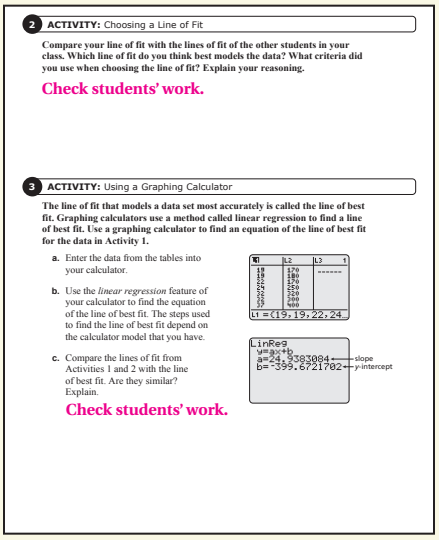
d. Compare your line of fit with your partner's line of fit. Are they similar? Which line of fit seems to model the data better? Why? Check students' work.

Differentiated Instruction

Kinesthetic

Give students a piece of a transparency sheet. Have them use a ruler and permanent marker to draw a straight line about 4 inches long. Use a hole punch to make a hole at each end of the line. Students can move the transparency over a scatter plot until they find a line of fit, make pencil marks in the holes, and use the marks to draw the line on the scatter plot.

12.6 Record and Practice Journal



2 ACTIVITY: Choosing a Line of Fit

Compare your line of fit with the lines of fit of the other students in your class. Which line of fit do you think best models the data? What criteria did you use when choosing the line of fit? Explain your reasoning.

Check students' work.

3 ACTIVITY: Using a Graphing Calculator

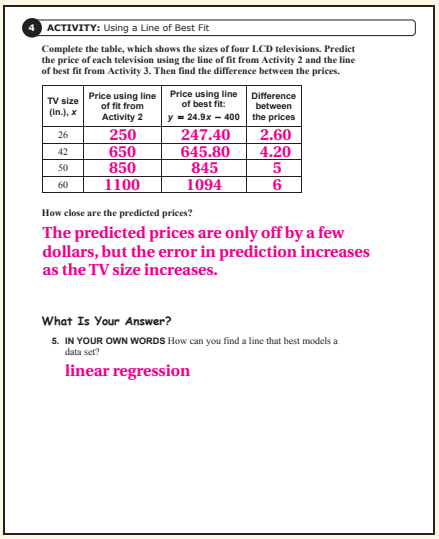
The line of fit that models a data set most accurately is called the line of best fit. Graphing calculators use a method called linear regression to find a line of best fit. Use a graphing calculator to find an equation of the line of best fit for the data in Activity 1.

a. Enter the data from the tables into your calculator.

b. Use the *linear regression* feature of your calculator to find the equation of the line of best fit. The steps used to find the line of best fit depend on the calculator model that you have.

c. Compare the lines of fit from Activities 1 and 2 with the line of best fit. Are they similar? Explain.

Check students' work.



4 ACTIVITY: Using a Line of Best Fit

Complete the table, which shows the sizes of four LCD televisions. Predict the price of each television using the line of fit from Activity 2 and the line of best fit from Activity 3. Then find the difference between the prices.

TV size (in.), x	Price using line of fit from Activity 2	Price using line of best fit: $y = 24.9x - 400$	Difference between the prices
26	250	247.40	2.60
42	650	645.80	4.20
50	850	845	5
60	1100	1094	6

How close are the predicted prices?

The predicted prices are only off by a few dollars, but the error in prediction increases as the TV size increases.

What Is Your Answer?

5. **IN YOUR OWN WORDS** How can you find a line that best models a data set?

linear regression

Laurie's Notes

Activity 3

- **MP5:** Explain that the calculator has a built-in feature that gives an equation for the *line of best fit*.
- The mathematics of how the calculator actually does the calculation may be studied in a future course. What students need to understand is that the process takes into account all of the data points.
- Different calculators have different keystrokes. It would be helpful to have students bring calculator manuals.
- Circulate around the room to answer questions as they arise. Pairing students with similar calculators will help.
- When students have finished, have a discussion about the equation generated by the calculator.
- **MP6 Attend to Precision:** The calculator will give a slope and y-intercept, typically with many decimal places. Ask whether students think it is necessary to carry all of the digits to the right of the decimal point.
- **Common Error:** As students observe the equation computed by the calculator, it is likely that someone will say, "That's not what my calculator says." If this happens, then the student did not enter the data correctly.

Activity 4

- **Big Idea:** One of the reasons we write lines of best fit is to make predictions. The equation can be used to interpolate or extrapolate unknown data values.
- **Note:** The equation for the line of best fit from Activity 3 is given in the table to avoid rounding errors and discrepancies in answers.
- When students finish, discuss the differences between the prices using their lines of fit and the calculator's line of best fit. The differences should be small.
- **Extension:** If time permits, add a new data value that is an outlier, such as (36, 950). Ask how it affects the calculator's line of best fit.

What Is Your Answer?

- Ask volunteers to share their answers.

Closure

- **Exit Ticket:** Find the line of best fit for the data in the Motivate. Use the line to estimate the cumulative snowfall in week 5. $y = 12.5x - 3$; 59.5 in.

3 ACTIVITY: Using a Graphing Calculator

Math Practice 5

Recognize Usefulness of Tools

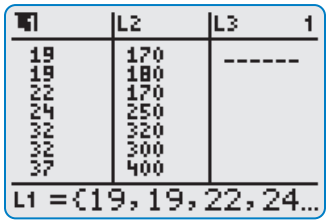
When is it useful to use a graphing utility to find the line of best fit? What are the advantages and disadvantages?

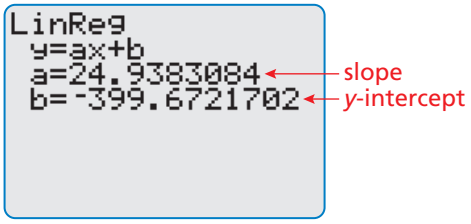
The line of fit that models a data set most accurately is called the line of best fit. Graphing calculators use a method called linear regression to find a line of best fit. Use a graphing calculator to find an equation of the line of best fit for the data in Activity 1.

a. Enter the data from the tables into your calculator.

b. Use the *linear regression* feature of your calculator to find the equation of the line of best fit. The steps used to find the line of best fit depend on the calculator model that you have.

c. Compare the lines of fit from Activities 1 and 2 with the line of best fit. Are they similar? Explain.





4 ACTIVITY: Using a Line of Best Fit

Copy and complete the table, which shows the sizes of four LCD televisions. Predict the price of each television using the line of fit from Activity 2 and the line of best fit from Activity 3. Then find the difference between the prices.

TV Size (in.), x	Price Using Line of Fit from Activity 2	Price Using Line of Best Fit: $y = 24.9x - 400$	Difference Between the Prices
26			
42			
50			
60			

How close are the predicted prices?

What Is Your Answer?

5. **IN YOUR OWN WORDS** How can you find a line that best models a data set?

Practice Use what you learned about analyzing lines of fit to complete Exercise 4 on page 649.

12.6 Lesson

Check It Out
Lesson Tutorials
BigIdeasMath.com

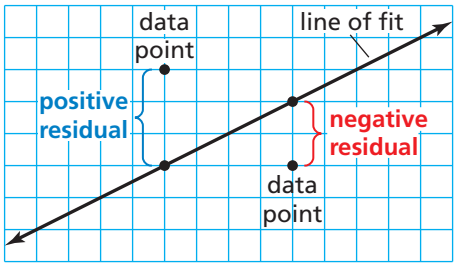
Key Vocabulary
residual, *p. 646*
linear regression, *p. 647*
line of best fit, *p. 647*
correlation coefficient, *p. 647*
causation, *p. 648*

One way to determine how well a line of fit models a data set is to analyze *residuals*.

Key Idea

Residuals

A **residual** is the difference between the y-value of a data point and the corresponding y-value found using the line of fit. A residual can be positive, negative, or zero.



A scatter plot of the residuals shows how well a model fits a data set. If the model is a good fit, then the residual points will be randomly dispersed about the horizontal axis. If the model is not a good fit, then the residual points will form some type of pattern.

EXAMPLE 1 Using Residuals

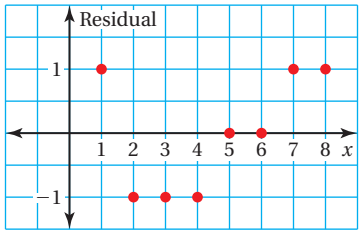
Week, x	Sales (millions), y
1	\$19
2	\$15
3	\$13
4	\$11
5	\$10
6	\$8
7	\$7
8	\$5

In Example 3 in Section 12.5, the equation $y = -2x + 20$ models the data in the table at the left. Is the model a good fit?

Step 1: Calculate the residuals and organize your results in a table.

Step 2: Use the points (x, residual) to make a scatter plot.

x	y	y-Value from Model	Residual
1	19	18	$19 - 18 = 1$
2	15	16	$15 - 16 = -1$
3	13	14	$13 - 14 = -1$
4	11	12	$11 - 12 = -1$
5	10	10	$10 - 10 = 0$
6	8	8	$8 - 8 = 0$
7	7	6	$7 - 6 = 1$
8	5	4	$5 - 4 = 1$



∴ The points are randomly dispersed about the horizontal axis. So, the equation $y = -2x + 20$ is a good fit.

Multi-Language Glossary at BigIdeasMath.com

Laurie's Notes

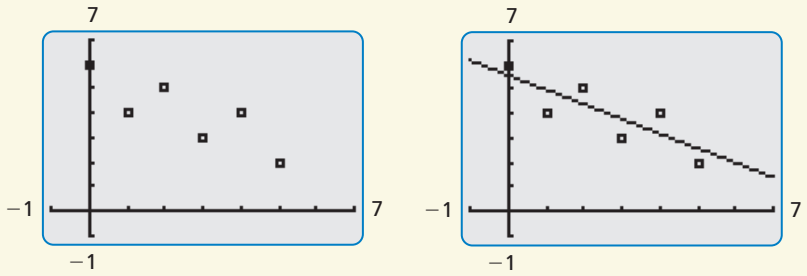
Introduction

Connect

- **Yesterday:** Students compared a line of fit to the line of best fit. (MP5, MP6)
- **Today:** Students will use residuals and find lines of best fit.

Motivate

- Display a scatter plot and ask students to draw a line of fit.
- Anticipate that it doesn't pass through any of the points.



- Ask for observations about the line. Hopefully they will say that about half of the points are above the line and about half are below the line.
- You want students to focus on the difference between the *y*-values of the line and the actual *y*-values.

Lesson Notes

Key Idea

- Define residual. Use a sketch to demonstrate residuals that are positive, negative, and zero.
- Explain that each *x*-value has a residual.
- When a model is a good fit, the residual points form a random pattern. That is, some of the residuals will be above and some will be below the horizontal axis. They show no discernible pattern.
- When a model is not a good fit, the residuals form a pattern, such as a linear or U-shaped pattern.
- I like to draw a scatter plot with a U-shape. When you try to fit a line to the curved pattern, the left side and right side of the points will be above the line, and the middle points will be below the line.
- **Note:** You may wish to tell students that calculators use residuals to calculate lines of best fit.

Example 1

- **MP4 Model with Mathematics:** Review Example 3 from Section 12.5. The problem involved sales of DVDs over a period of 8 weeks.
- Say, "We want to compare the actual *y*-values and the *y*-values from the model."
- Begin by listing the actual ordered pairs in a table.

Goal Today's lesson is analyzing lines of fit.

Technology for the Teacher

Dynamic Classroom

Lesson Tutorials
Lesson Plans
Answer Presentation Tool

Extra Example 1

The equation $y = x - 1$ models the data in the table. Is the model a good fit?

x	*y*
2	1
3	2
4	4
5	2
6	4
7	5
8	8
9	8

The points (*x*, residual) are randomly dispersed about the horizontal axis. So, the equation $y = x - 1$ is a good fit.

Extra Example 2

The table shows the ounces y in a bottle of water after x minutes of exercising. The equation $y = -0.2x + 16$ models the data. Is the model a good fit?

Minutes, x	Ounces, y
5	18
10	15
15	15
20	14
25	12
30	9
35	6
40	2

The points (x, residual) form a linear pattern with a negative slope. So, the equation $y = -0.2x + 16$ does not model the data well.

On Your Own

1. The points (x, residual) are randomly dispersed about the horizontal axis. So, the model $y = -9.8x + 850$ is a good fit.

English Language Learners

Vocabulary

Help English language learners with the meaning of the key vocabulary word *correlation coefficient*. The word *correlate* can be broken into two parts, *co-* and *relate*. The prefix *co-* means "with" or "together" as in *copilot*. The word *relate* means "to show or make a connection."

Laurie's Notes

Example 1 (continued)

- Then find the y-values from the model. Use the language, "actual y-value" and "y-value from the model" to help students differentiate between the two.
- Finally, compute the residuals.
- The table of residuals is often plotted on a horizontal line and called a residual plot. Each x-value has a residual, even if it is 0. If time permits, explore the residual plot feature on a graphing calculator.

? "Are the residuals randomly dispersed about the x-axis? Explain." yes; They show no discernible pattern.

- **MP3a Construct Viable Arguments:** A residual of 1 for a y-value of 19 is relatively small, but a residual of 1 for a y-value of 2 is not small. A residual of 100 for a y-value of 35,000 is relatively small, but a residual of 100 for a y-value of 250 is not small. The residuals in this example are relatively small.

Example 2

- Read the problem and then create a table as in Example 1. Make a residual plot.

? "Are the residuals both positive and negative?" yes "Is there a pattern or are the points randomly dispersed?" There is a pattern; The residuals are negative at the beginning and end of the data set, and positive in the middle.

- Conclude that because the residual plot forms a ∩ -shape, the model does not fit the data well.

On Your Own

- Students may want to make a scatter plot and graph the model provided using a graphing calculator.

Discuss

- Remind students of the activity in which they used calculators to determine the line of best fit. It was called linear regression.
- Explain that in addition to calculating the line of best fit, the calculator also gives a value called the correlation coefficient. This is a measure of how well the line fits the data.
- Describe positive, negative, and no correlation.
- Say, "The correlation coefficient is a number between -1 and 1." Draw and label the graphic shown.
- **Common Misconception:** Students may think that a strong negative correlation is *bad*. Stress that the "negative" merely refers to the downward trend of the data, or slope.

EXAMPLE 2 Using Residuals

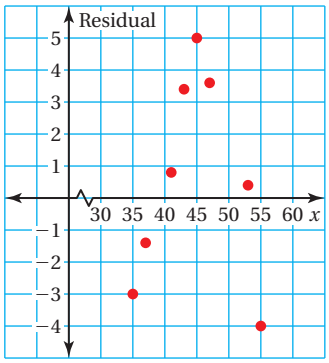
Age, x	Salary, y
35	42
37	44
41	47
43	50
45	52
47	51
53	49
55	45

The table at the left shows the ages x and salaries y (in thousands of dollars) of eight employees at a company. The equation $y = 0.2x + 38$ models the data. Is the model a good fit?

Step 1: Calculate the residuals and organize your results in a table.

Step 2: Use the points (x, residual) to make a scatter plot.

x	y	y-Value from Model	Residual
35	42	45.0	$42 - 45.0 = -3.0$
37	44	45.4	$44 - 45.4 = -1.4$
41	47	46.2	$47 - 46.2 = 0.8$
43	50	46.6	$50 - 46.6 = 3.4$
45	52	47.0	$52 - 47.0 = 5.0$
47	51	47.4	$51 - 47.4 = 3.6$
53	49	48.6	$49 - 48.6 = 0.4$
55	45	49.0	$45 - 49.0 = -4.0$



∴ The points form a ∩-shaped pattern. So, the equation $y = 0.2x + 38$ does not model the data well.

On Your Own

Now You're Ready
Exercises 5 and 6

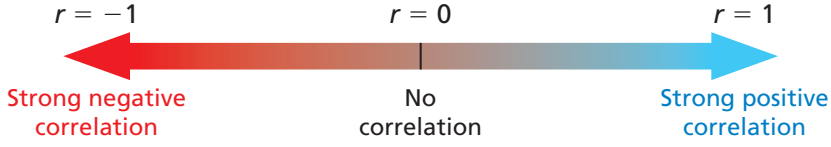
1. The table shows the attendance y (in thousands) at an amusement park from 2000 to 2009, where $x = 0$ represents the year 2000. The equation $y = -9.8x + 850$ models the data. Is the model a good fit?

Year, x	0	1	2	3	4	5	6	7	8	9
Attendance, y	850	845	828	798	800	792	785	781	775	760

Study Tip

You know how to use two points to find an equation of a line of fit. When finding an equation of the line of best fit, every point in the data set is used.

Graphing calculators use a method called **linear regression** to find a precise line of fit called a **line of best fit.** This line best models a set of data. A calculator often gives a value r called the **correlation coefficient.** This value tells whether the correlation is positive or negative, and how closely the equation models the data. Values of r range from -1 to 1. When r is close to 1 or -1, there is a strong correlation between the variables. As r gets closer to 0, the correlation becomes weaker.

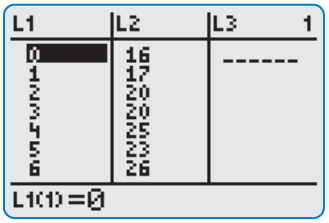


EXAMPLE 3 Finding a Line of Best Fit Using Technology

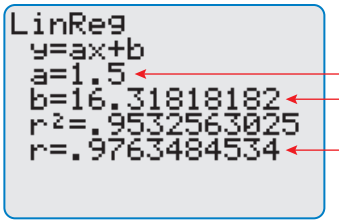
The table shows the worldwide movie ticket sales y (in billions of dollars) from 2000 to 2010, where $x = 0$ represents the year 2000. Use a graphing calculator to find an equation of the line of best fit. Identify and interpret the correlation coefficient.

Year, x	0	1	2	3	4	5	6	7	8	9	10
Ticket Sales, y	16	17	20	20	25	23	26	26	28	29	32

Step 1: Enter the data from the table into your calculator.



Step 2: Use the *linear regression* feature.



An equation of the line of best fit is $y = 1.5x + 16$. The correlation coefficient is about 0.976. This means that the relationship between the years and ticket sales is a strong positive correlation and the equation closely models the data.

Study Tip

The slope of 1.5 indicates that sales are increasing by about \$1.5 billion each year. The y-intercept of 16 represents the ticket sales of \$16 billion for 2000.

When a change in one variable x results in a change in another variable y, it is called **causation.** Causation produces a strong correlation between the two variables. The converse of the statement is not true. In other words, correlation does not imply causation.

EXAMPLE 4 Identifying Correlation and Causation

Tell whether a correlation is likely in the situation. If so, tell whether there is a causal relationship. Explain your reasoning.

a. time spent exercising and the number of calories burned

There is a positive correlation and a causal relationship because the more time you spend exercising, the more calories you burn.

b. the number of banks and the population of a city

There may be a positive correlation but no causal relationship. Building more banks will not cause the population to increase.

Reading

A causal relationship exists when one variable causes a change in another variable.

On Your Own

2. Use a graphing calculator to find an equation of the line of best fit for the data in On Your Own Question 1. Identify and interpret the correlation coefficient.

3. Is there a correlation between time spent playing video games and grade point average? If so, is there a causal relationship? Explain.

Now You're Ready
Exercises 7, 8, and 13–16

Laurie's Notes

Example 3

- **MP1a Make Sense of Problems:** Ask a student to read the problem. Make sure students understand that $x = 0$ represents the year 2000.
- On many calculators, the correlation coefficient is a feature that can be turned on and off. Instruct students how to turn this feature on so that it is displayed when they perform the regression. An r^2-value may also be displayed, which students may learn about in a future course.

? "What is the line of best fit?" $y = 1.5x + 16$

? "What do the slope and y-intercept mean?" A slope of 1.5 means that ticket sales are increasing by about \$1.5 billion each year. A y-intercept of 16 means that in 2000, the ticket sales were about \$16 billion.

? "What is the correlation coefficient and what does it mean?" $r \approx 0.976$; It implies a strong positive correlation between years and ticket sales.

Discuss

? "In Example 3, would the year increasing to 2012 *cause* the sales to increase also?" Students may be unsure of this, however, they should recognize that it is not the year that causes an increase in sales. Many other factors are the reasons behind the increases in ticket sales.

- Define causation. It is important for students to understand the difference between correlation and causation. Correlation between two variables does not imply that one causes the other.

Example 4

- Read each statement in parts (a) and (b). Discuss the explanation of causation for each.

On Your Own

- **Think-Pair-Share:** Students should read each question independently and then work in pairs to answer the questions. When they have answered the questions, the pair should compare their answers with another group and discuss any discrepancies.

Closure

- **Exit Ticket:** Give an example of a situation in which there is correlation between the variables, but no causation. *Sample answer:* The number of teachers in a school and the number of students in the school

Extra Example 3

Use a graphing calculator to find an equation of the line of best fit for the data in Extra Example 2. Identify and interpret the correlation coefficient.
$y = -0.4x + 21$; The correlation coefficient is about -0.966. This means that the relationship between minutes and ounces is a strong negative correlation and the equation closely models the data.

Extra Example 4

Is there a correlation between the number of doctors at a hospital and the number of patients in the hospital? If so, is there a causal relationship? Explain.
There may be a positive correlation but there is no causal relationship. Increasing the number of doctors will not cause the number of patients to increase.

On Your Own

2. $y = -9.6x + 845$; The correlation coefficient is about -0.964. The relationship between the years and attendance have a strong negative correlation and the equation closely models the data.
3. There may be a negative correlation. A causal relationship is possible because the more time you spend playing video games, the less time you spend studying.

Vocabulary and Concept Check

1. when actual value > value from model; when value from model > actual value
2. Plot the points (x, residual). If the points form a pattern, the line does not fit the data well.
3. -0.98, because it is closer to -1 than 0.91 is to 1. $(|-0.98| > |0.91|)$

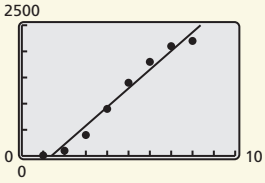
Practice and Problem Solving

4. a. *Sample answer:* Using (1, 6) and (4, 3.9), $y = -0.7x + 6.7$.

   b. $y = -0.7x + 6.8$

   c. *Sample answer:* 2.5 in.; 2.6 in.

5. The points (x, residual) are all above the horizontal axis. So, the equation does not model the data well.
6. The points (x, residual) are randomly dispersed about the horizontal axis. So, the equation is a good fit.
7. $y = 3.5$; $r = 0$; There is no correlation between x and y. The equation does not fit the data.
8. $y = -1.2x + 7$; $r \approx -0.883$; The relationship between x and y is a fairly strong negative correlation and the equation fits the data fairly well.
9. $y = 357.5x - 495$



10. B
11. A

Assignment Guide and Homework Check

Level	Assignment	Homework Check
Average	1–3, 5, 7, 9, 13–19 odd, 22–24	5, 13, 17, 19
Advanced	1–3, 6, 8, 9, 14–20 even, 19, 22–24	6, 14, 19, 20

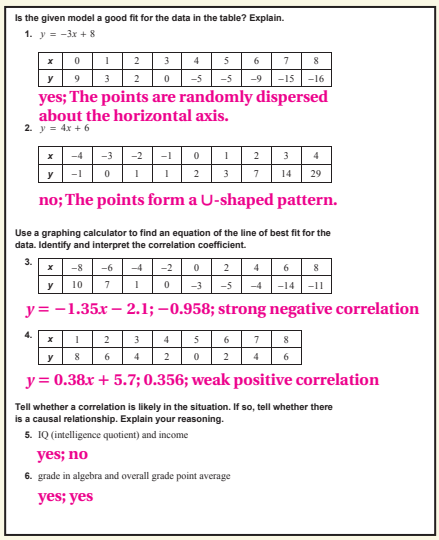
For Your Information

- **Exercise 16** This will be a good exercise for discussion. Overall, there is no correlation. There are many different types and sizes of dogs with varying tail lengths. However, an argument could be made that there is a correlation depending on the circumstance. For example, consider a growing puppy.

Common Errors

- **Exercises 5 and 6** Students may have trouble calculating the residuals. Remind them that a residual is the y-value from the data minus the y-value from the model.
- **Exercises 7–9** Students may enter the x-values into the y list of the graphing calculator, and vice versa. Encourage them to double check their data entries and understand which lists represent the x- and y-values.

12.6 Record and Practice Journal



Is the given model a good fit for the data in the table? Explain.

1. $y = -3x + 8$

x	0	1	2	3	4	5	6	7	8
y	9	3	2	0	−5	−5	−9	−15	−16

yes; The points are randomly dispersed about the horizontal axis.

2. $y = 4x + 6$

x	−4	−3	−2	−1	0	1	2	3	4
y	−1	0	1	1	2	3	7	14	29

no; The points form a ∪-shaped pattern.

Use a graphing calculator to find an equation of the line of best fit for the data. Identify and interpret the correlation coefficient.

3.

x	−8	−6	−4	−2	0	2	4	6	8
y	10	7	1	0	−3	−5	−4	−14	−11

$y = -1.35x - 2.1$; -0.958; strong negative correlation

4.

x	1	2	3	4	5	6	7	8
y	8	6	4	2	0	2	4	6

$y = 0.38x + 5.7$; 0.356; weak positive correlation

Tell whether a correlation is likely in the situation. If so, tell whether there is a causal relationship. Explain your reasoning.

5. IQ (intelligence quotient) and income

yes; no

6. grade in algebra and overall grade point average

yes; yes

12.6 Exercises

Check It Out
Help with Homework
BigIdeasMath.com

Vocabulary and Concept Check

1. **VOCABULARY** When is a residual positive? When is it negative?
2. **WRITING** Explain how you can use residuals to determine how well a line of fit models a data set.
3. **NUMBER SENSE** Which correlation coefficient indicates a stronger relationship, -0.98 or 0.91? Explain.

Practice and Problem Solving

4. **ANTLERS** The table shows the weekly growth y (in inches) of an elk's antlers.

Week, x	1	2	3	4	5
Growth, y	6.0	5.5	4.7	3.9	3.3

   a. Find a line of fit for the data.
   b. Use a graphing calculator to find an equation of the line of best fit.
   c. Use each model to predict the antler growth in week 6.

Is the given model a good fit for the data in the table? Explain.

1 2 5. $y = 4x - 5$

x	-4	-3	-2	-1	0	1	2	3	4
y	-18	-13	-10	-7	-2	0	6	10	15

6. $y = -1.3x + 1$

x	-8	-6	-4	-2	0	2	4	6	8
y	9	10	5	8	-1	1	-4	-12	-7

Use a graphing calculator to find an equation of the line of best fit for the data. Identify and interpret the correlation coefficient.

3 7.

x	0	1	2	3	4	5	6	7
y	8	5	2	-1	-1	2	5	8

8.

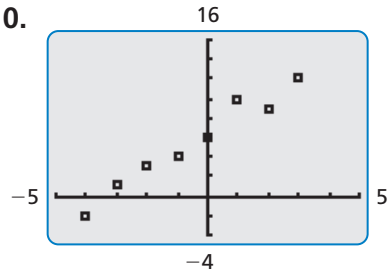
x	-8	-6	-4	-2	0	2	4	6	8	10
y	20	8	17	7	8	1	5	-2	2	-8

9. **EARTHQUAKE** The table shows the total number y of people reporting an earthquake x minutes after it ended. Use a graphing calculator to find an equation of the line of best fit. In the same viewing window, graph the line and plot the data.

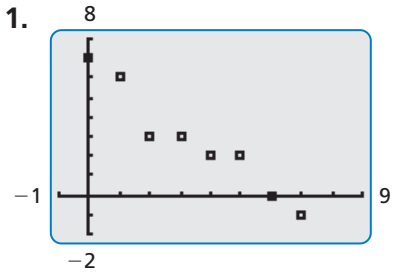
Minutes, x	1	2	3	4	5	6	7	8
People, y	10	100	400	900	1400	1800	2100	2200

MATCHING Match the graph of the data with its corresponding linear regression screen.

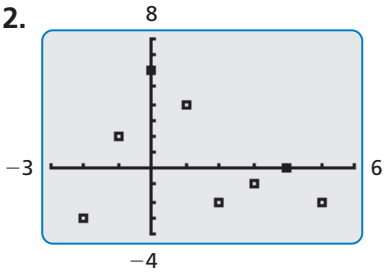
10.



11.



12.



A.

```
LinReg
 y=ax+b
 a=-1.071428571
 b=6.5
 r²=.9361997226
 r=-.9675741432
```

B.

```
LinReg
 y=ax+b
 a=1.916666667
 b=6.333333333
 r²=.9650768883
 r=.9823832668
```

C.

```
LinReg
 y=ax+b
 a=-.3571428571
 b=1.035714286
 r²=.0744047619
 r=-.2727723628
```

Tell whether a correlation is likely in the situation. If so, tell whether there is a causal relationship. Explain your reasoning.

13. the amount of time spent talking on a cell phone and the remaining battery life

14. the height of a toddler and the size of the toddler's vocabulary

15. the number of hats you own and the size of your head

16. the weight of a dog and the length of its tail

17. **FUEL MILEAGE** The table shows the prices x (in thousands of dollars) and fuel economies y (in miles per gallon) of several automobiles.

Price (thousands of dollars), x	24	32	30	28	35	20	22	26
Fuel Economy (miles per gallon), y	30	30	34	35	28	25	28	36

a. Use a graphing calculator to find an equation of the line of best fit. Identify and interpret the correlation coefficient.

b. Calculate the residuals. Then make a scatter plot of the residuals and interpret the results.

18. **TEXTING** The table shows the numbers y (in billions) of text messages sent from 2006 to 2011, where $x = 6$ represents the year 2006.

Year, x	Text Messages (billions), y
6	113
7	241
8	601
9	1360
10	1806
11	2206

a. Use a graphing calculator to find an equation of the line of best fit. Identify and interpret the correlation coefficient.

b. Interpret the slope of the line of best fit.

c. Calculate the residuals. Then make a scatter plot of the residuals and interpret the results.

d. Predict the number of text messages sent in 2015.

Common Errors

- **Exercises 13–16** Students may have a difficult time determining causation. Remind them that a correlation does not imply causation.
- **Exercises 17–21** Students may enter the x-values into the y list of the graphing calculator, and vice versa. Encourage them to double check their data entries and understand which lists represent the x- and y-values.

Practice and Problem Solving

12. C

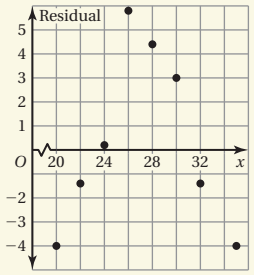
13. yes; yes; Talking longer causes the battery life to decrease.

14. yes; no; Taller toddlers are likely older and likely to know more words, but being taller does not cause an increase in vocabulary.

15. no **16.** no

17. **a.** $y = 0.2x + 25$; $r \approx 0.283$; The relationship between x and y is a weak positive correlation and the equation does not fit the data well.

b.



The points (x, residual) form a ∩–shaped pattern. So, the equation does not model the data well.

18–19. See Additional Answers.

English Language Learners

Pair Activity

Pair English language learners with English speakers to solve problems using a graphing calculator. English language learners will benefit by conversing in English and all students will benefit by learning how to operate the graphing calculator.

Practice and Problem Solving

20. The correlation coefficient changes from about -0.965 to about -0.667 which weakens the correlation. This happens because the new data point is an outlier.

21. See *Taking Math Deeper*.

Fair Game Review

22. 67; 19; about 7.0

23. 15; 8; 2

24. C

Mini-Assessment

1. The equation $y = 2x + 1$ models the data in the table. Is the model a good fit?

x	1	2	3	4	5	6
y	4	6	10	11	14	14

The points (x, residual) are all above the horizontal axis. So, the equation $y = 2x + 1$ does not model the data well.

2. The table shows the forecasted temperature x (in °F) and the actual temperature y (in °F) during one week. Use a graphing calculator to find an equation of the line of best fit. Identify and interpret the correlation coefficient.

x (°F)	72	75	73	82	83	78	77
y (°F)	68	74	79	85	83	81	76

$y = 1.1x - 10$; $r \approx 0.826$; The relationship between x and y is a fairly strong positive correlation and the equation fits the data fairly well.

3. In Exercise 2, is there a causal relationship between x and y? Explain. no; While there is a correlation, forecasting a temperature does not cause the actual temperature to change. The forecast is just a prediction.

Taking Math Deeper

Exercise 21

This exercise challenges students to combine previous knowledge about quadratic functions with knowledge about regression to solve a problem they have not been taught how to solve.

1 **a.** Copy and complete the table.
Each entry in the y-row must show the number of *new* responses each minute.

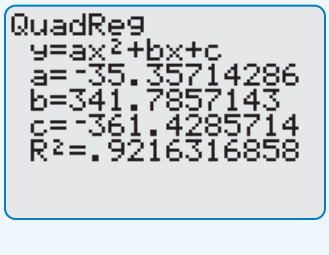
x	1	2	3	4	5	6	7	8
y	10	90	300	500	500	400	300	100

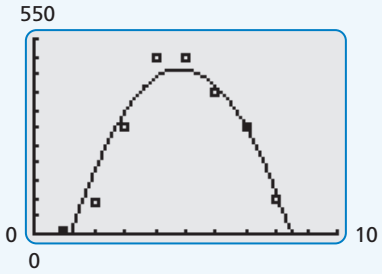
2 **b.** As x increases, the y-values increase, and then decrease.

A linear function will not fit the data well because the data do not show an approximately constant rate of change.

Because the y-values increase and then decrease, a quadratic function may fit the data well.

3 **c.** For linear functions, you used the *linear regression* feature. So, find and use the *quadratic regression* feature.





The data can be modeled by $y = -35.36x^2 + 341.8x - 361$. Challenge students to make a conjecture about the meaning of the R^2-value.

Reteaching and Enrichment Strategies

If students need help . . .	If students got it . . .
Resources by Chapter • Practice A and Practice B • Puzzle Time Record and Practice Journal Practice Differentiating the Lesson Lesson Tutorials Skills Review Handbook	Resources by Chapter • Enrichment and Extension • School-to-Work • Financial Literacy • Technology Connection Start the next section

19. GRADES The table shows the numbers x of hours spent watching television each week, and the grade point averages y of several students.

Hours, x	Grade Point Average, y
10	3.0
5	3.4
3	3.5
12	2.7
20	2.1
15	2.8
8	3.0
4	3.7
16	2.5

a. Use a graphing calculator to find an equation of the line of best fit. Identify and interpret the correlation coefficient.

b. Interpret the slope and y-intercept of the line of best fit.

c. Another student watches about 14 hours of television each week. Predict the student's grade point average.

d. Do you think watching more television each week may cause a lower grade point average? Explain.

20. REASONING A student spends 2 hours each week watching television and has a grade point average of 2.4. Include this information in the data set in Exercise 19. How does including this value affect the correlation coefficient? Explain.

21. Modeling Consider the earthquake data in Exercise 9.

a. Copy and complete the table to show the number y of people reporting the earthquake in the xth minute after the earthquake ended.

x	1	2	3	4	5	6	7	8
y	10	90	300					

b. Describe how the y-values change as x increases. Do you think a linear function will fit the data well? If not, what type of function do you think will fit the data well? Explain.

c. Use a graphing calculator to find the model in part (b).

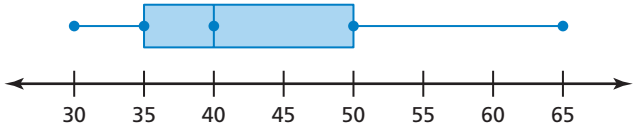
Fair Game Review What you learned in previous grades & lessons

Find the mean, range, and standard deviation of the data. *(Section 12.2)*

22. 59, 70, 62, 68, 75, 77, 58

23. 15, 14, 11, 15, 16, 19, 14, 16, 15

24. MULTIPLE CHOICE What is the interquartile range of the box-and-whisker plot? *(Section 12.3)*



(A) 5 (B) 10 (C) 15 (D) 35

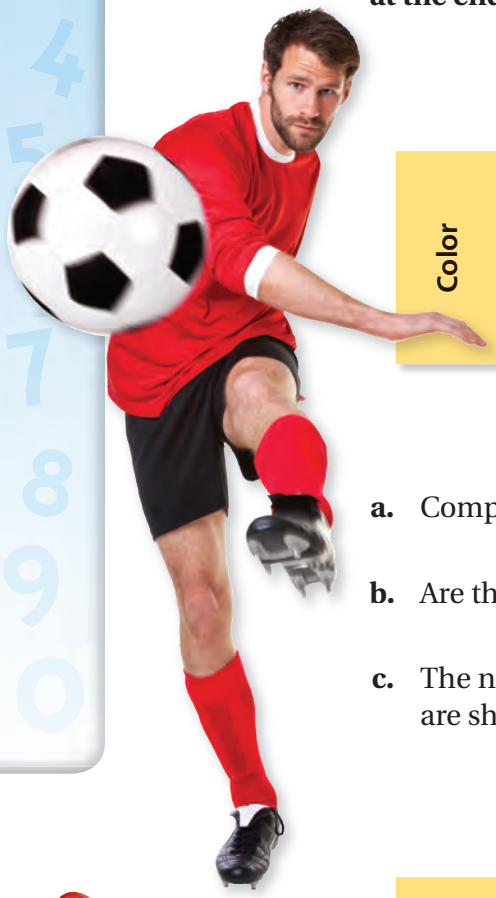
12.7 Two-Way Tables

Essential Question How can you read and make a two-way table?

Two categories of data can be displayed in a **two-way table.**

1 ACTIVITY: Reading a Two-Way Table

Work with a partner. You are the manager of a sports shop. The two-way table shows the numbers of soccer T-shirts that your shop has left in stock at the end of the season.



		T-Shirt Size					
		S	M	L	XL	XXL	Total
Color	Blue/White	5	4	1	0	2	
	Blue/Gold	3	6	5	2	0	
	Red/White	4	2	4	1	3	
	Black/White	3	4	1	2	1	
	Black/Gold	5	2	3	0	2	
	Total						65

a. Complete the totals for the rows and columns.

b. Are there any black and gold XL T-shirts in stock? Justify your answer.

c. The numbers of T-shirts you ordered at the beginning of the season are shown below. Complete the two-way table.

		T-Shirt Size					
		S	M	L	XL	XXL	Total
Color	Blue/White	5	6	7	6	5	
	Blue/Gold	5	6	7	6	5	
	Red/White	5	6	7	6	5	
	Black/White	5	6	7	6	5	
	Black/Gold	5	6	7	6	5	
	Total						

d. How would you alter the numbers of T-shirts you order for next season? Explain your reasoning.

COMMON CORE

Two-Way Tables

In this lesson, you will
- read two-way tables.
- find marginal frequencies.
- make two-way tables.
- find relationships in two-way tables.

Learning Standards
8.SP.4
S.ID.5

Laurie's Notes

Introduction

Standards for Mathematical Practice

- **MP2 Reason Abstractly and Quantitatively:** In this section, students are translating information into an organized table to make sense of the problem and to make observations and reason about the information. The goal is not to simply construct a table or read information from the table, but to reason about any relationships that may exist between categories in the table.

Motivate

- **Story Time:** Tell students about a few of the sessions and workshops you attended at a 3-day math conference. When you returned, you submitted your expenses.

	Day 1	Day 2	Day 3	Totals
Meals				A
Lodging				
Taxi				
Totals		B		C

- The school district does not want to share your expenses publicly so they are blacked out.
- ? "What do the numbers in A, B, and C represent?" A is the total amount spent on meals for 3 days; B is the total expenses for day 2; C is the total expenses for all 3 days.
- ? "How do you find C?" Find the sum of the last column or the sum of the last row.
- If students do not know the vocabulary—column and row, be sure to clarify.

Activity Notes

Activity 1

- Students should find that reading a two-way table is relatively easy. The term "two-way table" is new to students, yet they do not need a formal definition to make sense of the problem and what it is asking. In fact, some students will jump in and start adding the entries in the rows and columns without reading the introduction.
- When students have finished, discuss their responses to part (d).
- ? "What size(s) of shirts sold well?" XL and perhaps XXL "What size(s) of shirts did not sell well?" S and M
- ? "What color(s) of shirts sold well?" Black/White "What color(s) of shirts did not sell well?" Blue/Gold
- ? "Is there a way to quantify or rank the popular sizes and colors?" yes; You can compute the percent of each size or color that was sold.
- ? "Do you think merchants keep track of inventory in this manner?" Answers will vary. Successful merchants do track inventory to see what is selling.

Common Core State Standards

8.SP.4 Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table . . .. Use relative frequencies calculated for rows or columns to describe possible association between the two variables.
S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data . . .. Recognize possible associations and trends in the data.

Previous Learning

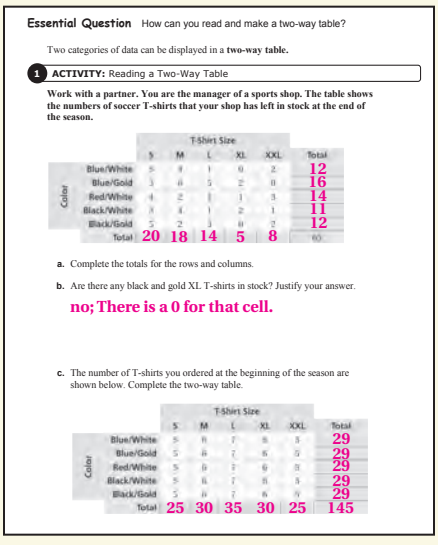
Students should know how to display data using different types of displays, such as histograms, box-and-whisker plots, and scatter plots.

Technology for the Teacher

Dynamic Classroom

Lesson Plans
Complete Materials List

12.7 Record and Practice Journal



Essential Question How can you read and make a two-way table?

Two categories of data can be displayed in a **two-way table.**

1 ACTIVITY: Reading a Two-Way Table

Work with a partner. You are the manager of a sports shop. The table shows the numbers of soccer T-shirts that your shop has left in stock at the end of the season.

T-Shirt Size

Color	S	M	L	XL	XXL	Total
Blue/White	5	4	1	0	2	12
Blue/Gold	3	6	5	2	0	16
Red/White	4	2	3	1	4	14
Black/White	3	4	1	2	1	11
Black/Gold	5	2	4	0	1	12
Total	20	18	14	5	8	65

a. Complete the totals for the rows and columns.

b. Are there any black and gold XL T-shirts in stock? Justify your answer.

no; There is a 0 for that cell.

c. The number of T-shirts you ordered at the beginning of the season are shown below. Complete the two-way table.

T-Shirt Size

Color	S	M	L	XL	XXL	Total
Blue/White	5	6	7	6	5	29
Blue/Gold	5	6	7	6	5	29
Red/White	5	6	7	6	5	29
Black/White	5	6	7	6	5	29
Black/Gold	5	6	7	6	5	29
Total	25	30	35	30	25	145

Differentiated Instruction

Advanced

Have students enter the data from a two-way table into a spreadsheet. Use the *chart* feature to create a 3-D column chart (as shown on page 653), a clustered column chart, a stacked column chart, and a 100% stacked column chart. Students should describe the relationship between the data values in each chart and give an example of what information could be gathered from reading the chart.

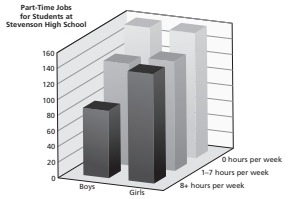
12.7 Record and Practice Journal

d. How would you alter the numbers of T-shirts you order for next season? Explain your reasoning.

Check students' work.

2 ACTIVITY: Analyzing Data

Work with a partner. The three-dimensional two-way table shows information about the numbers of hours students at a high school work at part-time jobs during the school year.



a. Make a two-way table showing the data. Use estimation to find the entries in your table.

		Gender		
		Boys	Girls	Total
Hours	0 hours per week	160	160	320
	1–7 hours per week	130	140	270
	8+ hours per week	90	140	230
	Total	380	440	820

b. Write two observations you can make that summarize the data in your table.

Check students' work.

c. A newspaper article claims that more boys than girls drop out of high school to work full-time. Do the data support this claim? Explain your reasoning.

no; The data is not about students dropping out or working full-time jobs.

What Is Your Answer?

3. **IN YOUR OWN WORDS** How can you read and make a two-way table?

Check students' work.

4. Find a real-life data set that can be represented by a two-way table. Then make a two-way table for the data set.

Check students' work.

Laurie's Notes

Activity 2

- "Have you ever seen a three-dimensional two-way table?" Students who are familiar with spreadsheets may recognize this type of display. Students may have seen them on the Internet, or in newspapers or magazines.
- Ask students to explain what one of the prisms represents.
- "What information is represented by the taller red prism?" It represents the number of girls that work 8 or more hours per week at a part-time job.
- "How many girls are in that category?" about 120
- Check that students have the correct labels for the rows and columns of the two-way table: gender (boys, girls) and hours (3 intervals shown).
- Ask a volunteer to display the two-way table.
- "Can you determine how many students attend Stevenson High School? Explain." yes; Find the sum of the last column or the last row in the table. They should be equal.
- **MP2:** In part (b), discuss student observations and listen for evidence of their statements. For instance, instead of saying that more girls have part-time jobs than boys, it is a stronger comparison to say that about 40 more girls have part-time jobs than boys. Encourage students to give quantitative evidence in their reasoning.
- Discuss part (c). Students may mention that the school should have about the same number of boys and girls. Because there are fewer boys represented in the table, you might infer that more boys have dropped out. Students may offer a viable argument as to *why* more boys might have dropped out of high school than girls, and it may not be for the purpose of working full-time.
- **MP2** and **MP3a Construct Viable Arguments:** This type of discussion is important in developing reasoning habits and constructing a logical argument.

What Is Your Answer?

- Question 4 can become a project due at the conclusion of the chapter.

Closure

- Refer back to the information about your expenses at the math conference. What would an entry in the blacked-out area represent? type of expense and day of expense How do you know if your total amount of expenses is correct? The sum of the last column should be equal to the sum of the last row.

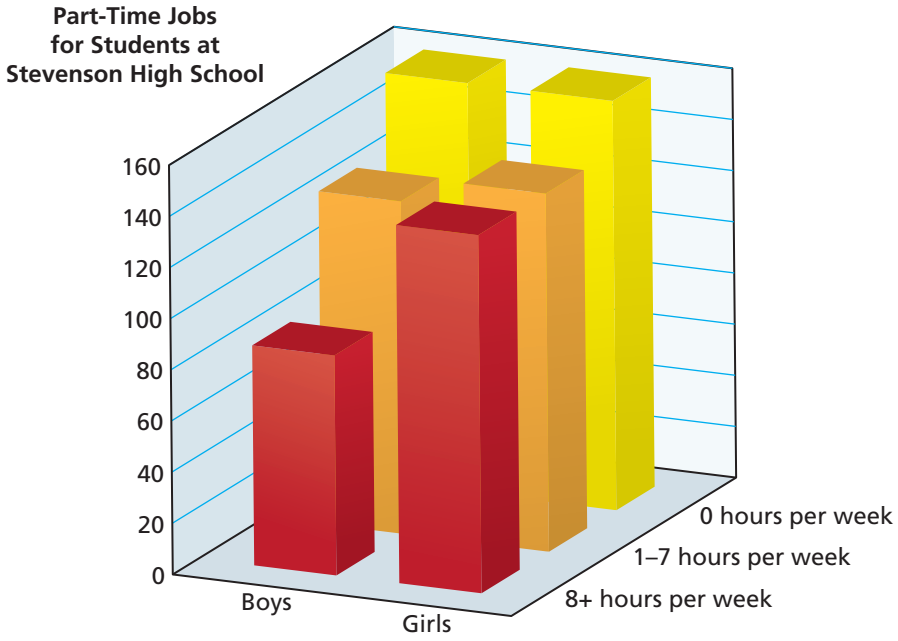
2 ACTIVITY: Analyzing Data

Math Practice 3

Construct Arguments

What are the advantages of using a table instead of a graph to analyze data?

Work with a partner. The three-dimensional two-way table shows information about the numbers of hours students at a high school work at part-time jobs during the school year.



a. Make a two-way table showing the data. Use estimation to find the entries in your table.

b. Write two observations you can make that summarize the data in your table.

c. A newspaper article claims that more boys than girls drop out of high school to work full-time. Do the data support this claim? Explain your reasoning.

What Is Your Answer?

3. IN YOUR OWN WORDS How can you read and make a two-way table?

4. Find a real-life data set that can be represented by a two-way table. Then make a two-way table for the data set.

Practice Use what you learned about two-way tables to complete Exercises 5 and 6 on page 656.

12.7 Lesson

Check It Out
Lesson Tutorials
BigIdeasMath.com

Key Vocabulary
two-way table, *p. 654*
joint frequency, *p. 654*
marginal frequency, *p. 654*

A **two-way table** displays two categories of data collected from the same source.

You randomly survey students in your school about their grades on the last test and whether they studied for the test. The two-way table shows your results. Each entry in the table is called a **joint frequency.**

		Student	
		Studied	Did Not Study
Grade	Passed	21	2
	Failed	1	6

joint frequency

EXAMPLE 1 Reading a Two-Way Table

How many of the students in the survey above studied for the test and passed?

The entry in the "Studied" column and "Passed" row is 21.

∴ So, 21 of the students in the survey studied for the test and passed.

The sums of the rows and columns in a two-way table are called **marginal frequencies.**

EXAMPLE 2 Finding Marginal Frequencies

Find and interpret the marginal frequencies for the survey above.

Create a new column and row for the sums. Then add the entries.

		Student		
		Studied	Did Not Study	Total
Grade	Passed	21	2	23
	Failed	1	6	7
	Total	22	8	30

23 students passed.
7 students failed.
30 students were surveyed.
22 students studied.
8 students did not study.

On Your Own

Now You're Ready
Exercises 3–6

1. You randomly survey students in a cafeteria about their plans for a football game and a school dance. The two-way table shows your results.

		Football Game	
		Attend	Not Attend
Dance	Attend	35	5
	Not Attend	16	20

   a. How many students will attend the dance but not the football game?

   b. Find and interpret the marginal frequencies for the survey.

Laurie's Notes

Introduction

Connect

- **Yesterday:** Students read two-way tables. (MP2, MP3a)
- **Today:** Students will construct two-way tables and identify relationships between categories of a two-way table.

Motivate

- Tell students about data you have collected from the faculty. You asked your coworkers who own both a computer and a cell phone if they use a Mac or a PC, and if they use a smartphone or a basic cell phone.

? "How can you represent this data?" Answers will vary.

- Explain that today they will study a way in which this data can be displayed and analyzed.

Lesson Notes

Discuss

- Define two-way table. Emphasize that information is known about two categories from the same source. The focus in this lesson is drawing conclusions from the data in a two-way table.
- Refer to examples from yesterday in explaining "information about two categories from the same source," such as soccer shirts–size and color.
- Define joint-frequency. Each entry in the two-way table is a frequency for two categories, hence the name joint-frequency.

Example 1

? "What category do the rows represent?" test grade: passed or failed

? "What category do the columns represent?" preparation of the student: studied or did not study

Example 2

- Define marginal frequencies. The sums of the rows and columns appear on the *margins* of the two-way table.
- Expand the two-way table. Label the new row and new column "Total."
- Add the rows and columns. Identify the sums using the labels shown.
- Ask students general questions about the row and column totals.
- Make sure students understand that 30 students were surveyed, not 60. Because each student is tallied twice, once for each category, you do not add $22 + 8 + 23 + 7$ to find the number surveyed. The sum of the rows and the sum of the columns should be equal.

? "What can you conclude about the data?" *Sample answer:* Of the 30 students, all but one of those who studied for the test passed.

On Your Own

- **Extension:** Ask students percent questions such as what percent of the students in the survey are not planning to attend either event.

Goal Today's lesson is constructing and analyzing **two-way tables**.

Technology for the Teacher

Dynamic Classroom

Lesson Tutorials
Lesson Plans
Answer Presentation Tool

Extra Example 1

You randomly survey students about whether they like orange juice. The two-way table shows your results. How many female students in the survey like orange juice? 29

		Gender	
		Male	Female
Orange juice	No	12	22
	Yes	37	29

Extra Example 2

Find and interpret the marginal frequencies for the survey above.

A total of 34 students do not like orange juice. A total of 66 students like orange juice. A total of 49 male students participated in the survey. A total of 51 female students participated in the survey.

On Your Own

1. **a.** 5 students

   b. A total of 51 students will attend the game. A total of 25 students will not attend the game. A total of 40 students will attend the dance. A total of 36 students will not attend the dance.

Extra Example 3

You randomly survey students in 6th, 7th, and 8th grade about whether they are going to try to join student council. The results are shown in the tally sheets. Make a two-way table that includes the marginal frequencies. See Additional Answers.

Join

Grade	Tally
6	𝍸 𝍸 |
7	𝍸 𝍸 |||
8	𝍸 ||

Not Join

Grade	Tally
6	𝍸 𝍸
7	𝍸 𝍸 ||||
8	𝍸 𝍸 𝍸

Extra Example 4

Use the two-way table in Extra Example 3.

a. For each grade, what percent of the students in the survey are going to try to join student council? not try to join student council? Organize the results in a two-way table. Explain what one of the entries represents. See Additional Answers.

b. Does the table in part (a) show a relationship between grade and whether students are going to try to join student council? Explain. See Additional Answers.

On Your Own

2. See Additional Answers.

English Language Learners

Class Activity

Form groups of 2 to 4 students with at least one English language learner and one English speaker. Have them work together to create and conduct a survey that includes two categories. The results of the survey should be presented to the class in a two-way table that includes the marginal frequencies.

Laurie's Notes

Example 3

- Guide students through the construction of the table and ask them to explain what several of the values represent.
- The amount of data in the two-way table may be overwhelming to some students. Make sure to talk through the problem, giving students time to stop and think about what each entry represents.
- **Teaching Tip:** Use two colors in the table, one for the joint frequencies and one for the marginal frequencies. This makes it easier to read.

? "What is an advantage of the two-way table over the tally marks on the sheets of paper?" Answers will vary. Students might say the table is more organized, easier to read, and more condensed.

Example 4

- Have a student read the problem. Explain that the problem is asking about percents within each *age group*—the data represented in the columns.

? "How many 12- to 13-year-olds ride the bus?" 24 "How many 12- to 13-year-olds are in the survey?" 40 "What percent of the 12- to 13-year-olds ride the bus?" $24/40 = 60\%$

- Guide students through the construction of the two-way table. Ask them to explain what each entry represents.
- **MP2 Reason Abstractly and Quantitatively:** Ask students to make an observation about the percents. As age increases, students are less likely to ride the bus to school. Ask students why this might be the case, encouraging them to reason about data.

? "In part (b), the sums of the columns are each 100%. Why don't the rows add up to 100%?" The base used to compute the percents referred to each of the age groups, not whether the student rides the bus.

? "Can percents in the table be found using the row totals?" yes "What would be the first entry in the table and what would it represent?" $24/50 = 48\%$; 48% of the students who ride the bus are 12–13 years old.

On Your Own

- **Common Error:** Students may find the percent of students who pack a lunch out of the total number of students who pack a lunch instead of the percent for each grade level.

Closure

- In the first example, is it likely that if you study for a test you will pass? Explain. yes; The table shows that the majority of students who studied for the test passed and the majority of students who did not study for the test failed.

EXAMPLE 3 Making a Two-Way Table

You randomly survey students between the ages of 12 and 17 about whether they ride the bus to school. The results are shown in the tally sheets. Make a two-way table that includes the marginal frequencies.

The two categories for the table are the ages and whether or not they ride the bus. Use the tally sheets to calculate each joint frequency. Then add to find each marginal frequency.

Rides bus

Age	Tally
12-13	𝍸 𝍸 𝍸 𝍸 𝍸 ||||
14-15	𝍸 𝍸 ||
16-17	𝍸 𝍸 ||||

Does not ride bus

Age	Tally
12-13	𝍸 𝍸 𝍸 |
14-15	𝍸 𝍸 |||
16-17	𝍸 𝍸 𝍸 𝍸 |

		Age			
		12–13	14–15	16–17	Total
Student	Rides Bus	24	12	14	50
	Does Not Ride Bus	16	13	21	50
	Total	40	25	35	100

EXAMPLE 4 Finding a Relationship in a Two-Way Table

Use the two-way table in Example 3.

a. For each age group, what percent of the students in the survey ride the bus to school? do not ride the bus to school? Organize the results in a two-way table. Explain what one of the entries represents.

		Age		
		12–13	14–15	16–17
Student	Rides Bus	60%	48%	40%
	Does Not Ride Bus	40%	52%	60%

$\frac{14}{35} = 0.4$
So, 40% of the 16- and 17-year-old students in the survey ride the bus to school.

b. Does the table in part (a) show a relationship between age and whether students ride the bus to school? Explain.

∴ The table shows that as age increases, students are less likely to ride the bus to school.

On Your Own

Now You're Ready
Exercises 10 and 11

2. You randomly survey students in a school about whether they buy a school lunch or pack a lunch. Your results are shown.

Grade 6 students
11 pack lunch, 9 buy school lunch
Grade 7 students
23 pack lunch, 27 buy school lunch
Grade 8 students
16 pack lunch, 14 buy school lunch

a. Make a two-way table that includes the marginal frequencies.

b. For each grade level, what percent of the students in the survey pack a lunch? buy a school lunch? Organize the results in a two-way table. Explain what one of the entries represents.

c. Does the table in part (b) show a relationship between grade level and lunch choice? Explain.

12.7 Exercises

Check It Out
Help with Homework
BigIdeasMath.com

Vocabulary and Concept Check

1. **VOCABULARY** Explain the relationship between joint frequencies and marginal frequencies.
2. **OPEN-ENDED** Describe how you can use a two-way table to organize data you collect from a survey.

Practice and Problem Solving

You randomly survey students about participating in their class's yearly fundraiser. You display the two categories of data in the two-way table.

3. Find the total of each row.
4. Find the total of each column.
5. How many female students will be participating in the fundraiser?
6. How many male students will *not* be participating in the fundraiser?

		Fundraiser	
		No	Yes
Gender	Female	22	51
	Male	30	29

Find and interpret the marginal frequencies.

7.

		School Play	
		Attend	Not Attend
Class	Junior	41	30
	Senior	52	23

8.

		Cell Phone Minutes	
		Limited	Unlimited
Text Plan	Limited	78	0
	Unlimited	175	15

9. **GOALS** You randomly survey students in your school. You ask whether grades, popularity, or sports is most important to them. You display your results in the two-way table.

   a. How many 10th graders chose sports? How many 11th graders chose grades?

   b. Find and interpret the marginal frequencies for the survey.

   c. What percent of students in the survey are 9th graders who chose popularity?

		Goal		
		Grades	Popularity	Sports
Grade	9th	31	18	23
	10th	39	16	19
	11th	42	6	17

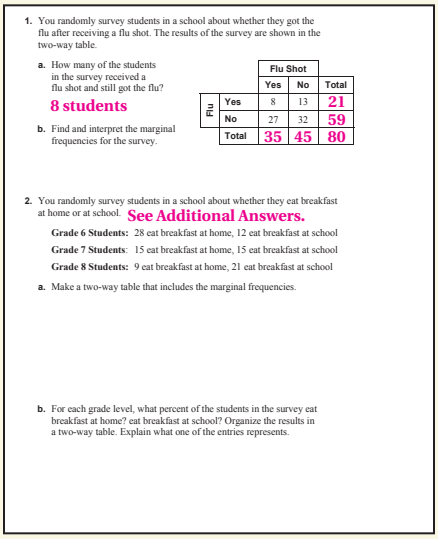
Assignment Guide and Homework Check

Level	Assignment	Homework Check
Average	1, 2, 3–7, 9, 11, 14–17	2, 5, 9, 11
Advanced	1, 2, 8–13, 14–17	2, 9, 11, 12

Common Errors

- **Exercises 1, 7–11** Students may incorrectly identify joint frequencies as marginal frequencies or marginal frequencies as joint frequencies. Remind them of these definitions.
- **Exercise 11c** Students may find the percents based on all students surveyed, not just the students from each eye color group. Encourage them to read carefully.

12.7 Record and Practice Journal



1. You randomly survey students in a school about whether they got the flu after receiving a flu shot. The results of the survey are shown in the two-way table.

   a. How many of the students in the survey received a flu shot and still got the flu?

   8 students

   b. Find and interpret the marginal frequencies for the survey.

		Flu Shot		
		Yes	No	Total
Flu	Yes	8	13	21
	No	27	32	59
	Total	35	45	80

2. You randomly survey students in a school about whether they eat breakfast at home or at school. **See Additional Answers.**

   Grade 6 Students: 28 eat breakfast at home, 12 eat breakfast at school
   Grade 7 Students: 15 eat breakfast at home, 15 eat breakfast at school
   Grade 8 Students: 9 eat breakfast at home, 21 eat breakfast at school

   a. Make a two-way table that includes the marginal frequencies.

   b. For each grade level, what percent of the students in the survey eat breakfast at home? eat breakfast at school? Organize the results in a two-way table. Explain what one of the entries represents.

Vocabulary and Concept Check

1. The joint frequencies are the entries in the two-way table that differentiate the two categories of data collected; The marginal frequencies are the sums of the rows and columns of the two-way table.
2. displays two categories of data collected from the same source

Practice and Problem Solving

3. total of females surveyed: 73; total of males surveyed: 59
4. total of "no" participants: 52; total of "yes" participants: 80
5. 51
6. 30
7. 71 students are juniors; 75 students are seniors; 93 students are attending the school play; 53 students are not attending the school play.
8. 78 people have limited cell phone texting plans; 190 people have unlimited cell phone texting plans; 253 people have limited cell phone minutes; 15 people have unlimited cell phone minutes.
9. **a.** 19; 42

   b. number of students surveyed: 9th grade: 72
   10th grade: 74
   11th grade: 65; 112 students chose grades, 40 students chose popularity, 59 students chose sports.

   c. about 8.5%

Practice and Problem Solving

10–12. See Additional Answers.

13. See *Taking Math Deeper*.

Fair Game Review

14. $-1; x = 1$

15. $\frac{x+2}{x+4}; x = 3$ and $x = -4$

16. $\frac{5x-2}{7x+1}; x = 0$ and $x = -\frac{1}{7}$

17. B

Mini-Assessment

1. You randomly survey students about whether they are involved in school sports.

Grade 5: 12 involved, 26 not involved
Grade 8: 23 involved, 19 not involved

a. Make a two-way table that includes the marginal frequencies See Additional Answers.

b. For each grade level, what percent of the students in the survey are involved in school sports? are not involved in school sports? Organize the results in a two-way table. Explain what one of the entries represents. See Additional Answers.

c. Does the table in part (b) show a relationship between grade level and involvement in school sports? Explain. See Additional Answers.

Taking Math Deeper

Exercise 13

This problem can help students see the benefits of a two-way table and the many different questions that can be asked regarding the entries. Encourage students to pay close attention to the wording of a question.

1 When finding the percent of students that are either female or have green eyes, students may mistakenly count the number of females with green eyes twice.

		Eye Color			
		Green	Blue	Brown	Total
Gender	Male	5	16	27	48
	Female	3	19	18	40
	Total	8	35	45	88

$$\frac{5}{88} + \frac{3}{88} + \frac{19}{88} + \frac{18}{88} = \frac{45}{88} \approx 51.1\%$$

2 To find the percent of students that are males that do not have green eyes, divide the sum of the remaining two joint frequencies by the number of students in the survey.

$$\frac{16}{88} + \frac{27}{88} = \frac{43}{88} \approx 48.9\%$$

3 The sum of the percents is $51.1\% + 48.9\% = 100\%$.

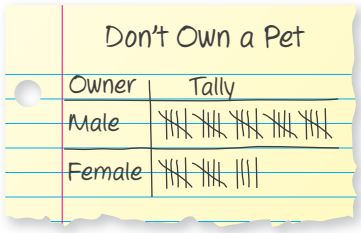
These two percents account for everyone in the survey.

Note: You may want to challenge students by asking what type of data display could be used to display the information in a two-way table, such as a double bar graph.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Resources by Chapter • Practice A and Practice B • Puzzle Time Record and Practice Journal Practice Differentiating the Lesson Lesson Tutorials Skills Review Handbook	Resources by Chapter • Enrichment and Extension • School-to-Work • Financial Literacy • Technology Connection • Life Connections Start the next section

3 **10. PETS** You randomly survey students in your school about whether they own a pet. The results are shown in the tally sheets. Make a two-way table that includes the marginal frequencies.



4 **11. EYE COLOR** You randomly survey students in your school about the color of their eyes. The results are shown in the tables.

a. Make a two-way table.

b. Find and interpret the marginal frequencies for the survey.

c. For each eye color, what percent of the students in the survey are male? female? Organize the results in a two-way table. Explain what two of the entries represent.

Eye Color of Males Surveyed		
Green	Blue	Brown
5	16	27

Eye Color of Females Surveyed		
Green	Blue	Brown
3	19	18

12. REASONING Use the information from Exercise 11. For each gender, what percent of the students in the survey have green eyes? blue eyes? brown eyes? Organize the results in a two-way table. Explain what two of the entries represent.

13. Precision What percent of students in the survey in Exercise 11 are either female or have green eyes? What percent of students in the survey are males that do not have green eyes? Find and explain the sum of these two percents.

Fair Game Review What you learned in previous grades & lessons

Simplify the rational expression, if possible. State the excluded value(s). *(Section 11.3)*

14. $\dfrac{1-x}{x-1}$

15. $\dfrac{x^2-x-6}{x^2+x-12}$

16. $\dfrac{15x^3-6x^2}{21x^3+3x^2}$

17. MULTIPLE CHOICE What is the solution of $\dfrac{1}{x}=\dfrac{3}{x+4}$? *(Section 11.7)*

A $x=-2$ **B** $x=2$ **C** $x=3$ **D** $x=4$

12.8 Choosing a Data Display

Essential Question How can you display data in a way that helps you make decisions?

1 ACTIVITY: Displaying Data

Work with a partner. Analyze and display each data set in a way that best describes the data. Explain your choice of display.

a. ROAD KILL A group of schools in New England participated in a 2-month study and reported 3962 dead animals.

Birds 307	Mammals 2746
Amphibians 145	Reptiles 75
Unknown 689	

b. BLACK BEAR ROAD KILL The data below show the numbers of black bears killed on a state's roads from 1993 to 2012.

1993	30	2000	47	2007	99
1994	37	2001	49	2008	129
1995	46	2002	61	2009	111
1996	33	2003	74	2010	127
1997	43	2004	88	2011	141
1998	35	2005	82	2012	135
1999	43	2006	109		

c. RACCOON ROAD KILL A 1-week study along a 4-mile section of road found the following weights (in pounds) of raccoons that had been killed by vehicles.

13.4	14.8	17.0	12.9
21.3	21.5	16.8	14.8
15.2	18.7	18.6	17.2
18.5	9.4	19.4	15.7
14.5	9.5	25.4	21.5
17.3	19.1	11.0	12.4
20.4	13.6	17.5	18.5
21.5	14.0	13.9	19.0

d. What do you think can be done to minimize the number of animals killed by vehicles?

Common Core

Data Displays

In this lesson, you will

- choose appropriate data displays.
- identify and analyze misleading data displays.

Learning Standard S.ID.1

Laurie's Notes

Introduction

Standards for Mathematical Practice

- **MP3 Construct Viable Arguments and Critique the Reasoning of Others:** In this section, students make decisions about how to display data. They will need to explain their reasoning for selecting a particular display. If two students select different data displays, it is important that they discuss the reasoning behind their choices.

Motivate

- The theme for the first activity is road kill. While students may giggle at the thought, automobile accidents involving large animals can be serious. I had my first and only accident with a deer 5 years ago. I was 2 miles from home and I was traveling 40 miles per hour. The deer was killed, my daughter and I were not injured, and repairs to my car were about $1400.
- Allow time for students to share personal stories.
- Use the Internet to research and share vehicular data with students, such as the number of miles of roads in the U.S., the number of registered vehicles, the number of accidents, and animal related accidents.

Activity Notes

Discuss

- Discuss the data displays with which students are familiar: pictograph, bar graph, line graph, circle graph, stem-and-leaf plot, histogram, dot plot, box-and-whisker plot, and scatter plot. Have students describe the feature(s) of each display.
- Discuss the different numerical tools they have for describing data: mean, median, mode, range, standard deviation, quartile, and interquartile range.

Activity 1

- Students need to decide what display makes sense for the type of data that they have. There may be more than one appropriate answer.
- **MP3:** Discuss students' choices and their explanations.
- Possible data displays:
  - Part (a): a circle graph (what part of the whole set is each animal) or a bar graph (compare the different categories, although there is a large difference in bar heights: 75 to 2746)
  - Part (b): a scatter plot and line of best fit (pair data, show trend over time, and make predictions for the future) or a line graph
  - Part (c): a stem-and-leaf plot (spread of data), along with calculating the mean (about 16.7) and median (17.1)
  - Part (d): As a class, discuss students' ideas for minimizing the number of animals killed by vehicles.

Common Core State Standards

S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

Previous Learning

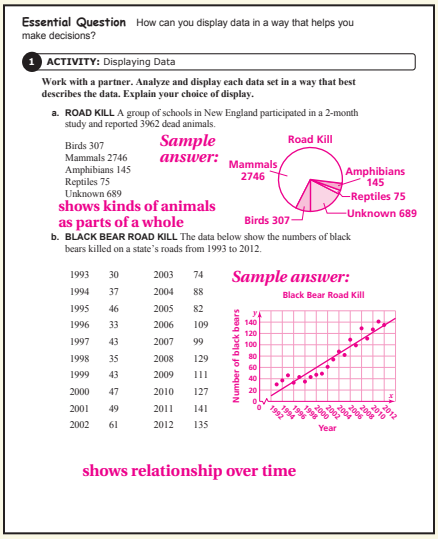
Students should know how to construct a variety of data displays from this year and past years.

Technology for the Teacher

Dynamic Classroom

Lesson Plans
Complete Materials List

12.8 Record and Practice Journal



Essential Question How can you display data in a way that helps you make decisions?

1 ACTIVITY: Displaying Data

Work with a partner. Analyze and display each data set in a way that best describes the data. Explain your choice of display.

a. ROAD KILL A group of schools in New England participated in a 2-month study and reported 3962 dead animals.

Birds 307
Mammals 2746
Amphibians 145
Reptiles 75
Unknown 689

Sample answer:

shows kinds of animals as parts of a whole

b. BLACK BEAR ROAD KILL The data below show the numbers of black bears killed on a state's roads from 1993 to 2012.

1993	30	2003	74
1994	37	2004	88
1995	46	2005	82
1996	33	2006	109
1997	43	2007	99
1998	35	2008	129
1999	43	2009	111
2000	47	2010	127
2001	49	2011	141
2002	61	2012	135

Sample answer:

shows relationship over time

English Language Learners

Vocabulary

English learners may need help understanding the word *scale*. There are several meanings in the English language. Some of the common meanings are:

a series of musical notes,
the covering of a reptile,
a device for weighing,
a ratio,
to climb.

In bar graphs, the scale is a series of markings used for measuring. Most scales start at 0 and go to (at least) the greatest value of the data.

12.8 Record and Practice Journal

c. **RACCOON ROAD KILL** A 1-week study along a 4-mile section of road found the following weights (in pounds) of raccoons that had been killed by vehicles.

13.4	14.8	17.0	12.9	21.3	21.5	16.8	14.8
15.2	18.7	18.6	17.2	18.5	9.4	19.4	15.7
14.5	9.5	25.4	21.5	17.3	19.1	11.0	12.4
20.4	13.6	17.5	18.5	21.5	14.0	13.9	19.0

See Additional Answers.

d. What do you think can be done to minimize the number of animals killed by vehicles?

Check students' work.

2 ACTIVITY: Statistics Project

ENDANGERED SPECIES PROJECT Use the Internet or some other reference to write a report about an animal species that is (or has been) endangered. Include graphical displays of the data you have gathered.

Sample: Florida Key Deer
In 1939, Florida banned the hunting of Key deer. The numbers of Key deer fell to about 100 in the 1940s.

About half of Key deer deaths are due to vehicles.

In 1947, public sentiment was stirred by 11-year-old Glenn Allen from Miami. Allen organized Boy Scouts and others in a letter-writing campaign that led to the establishment of the National Key Deer Refuge in 1957. The approximately 8600-acre refuge includes 2280 acres of designated wilderness.

One of two Key deer wildlife underpasses on Big Pine Key.

The Key Deer Refuge has increased the population of Key deer. A recent study estimated the total Key deer population to be approximately 800.

Check students' work.

What Is Your Answer?

3. **IN YOUR OWN WORDS** How can you display data in a way that helps you make decisions? Use the Internet or some other reference to find examples of the following types of data displays.

- Bar graph
- Circle graph
- Scatter plot
- Stem-and-leaf plot
- Box-and-whisker plot

Data displays make it easy to interpret data and make conclusions.

Laurie's Notes

Activity 2

- Ask a volunteer to read the information presented about Key deer. Discuss how the actions of one person can often make a big difference.
- It would be ideal if the library or computer room is available. If not, you or your students could bring in newspapers and magazines that contain graphical displays.
- If you assign this project, students will need several days.

What Is Your Answer?

- **MP4 Model with Mathematics:** Many students can make the displays, if they are told which display to use. It is equally important that students be able to select the display based upon the data and the question you hope to answer from making the display.
- The information gathered by students can be made into classroom posters.

Closure

- **Class Discussion:** Have students present their answers to Question 3. Then have students discuss features of each display, and what types of data lend itself to each data display.

2 ACTIVITY: Statistics Project

Math Practice 4

Use a Graph

How can you use a graph to represent the data you have gathered for your report? What does the graph tell you about the data?

ENDANGERED SPECIES PROJECT **Use the Internet or some other reference to write a report about an animal species that is (or has been) endangered. Include graphical displays of the data you have gathered.**

Sample: Florida Key Deer

In 1939, Florida banned the hunting of Key deer. The numbers of Key deer fell to about 100 in the 1940s.

In 1947, public sentiment was stirred by 11-year-old Glenn Allen from Miami. Allen organized Boy Scouts and others in a letter-writing campaign that led to the establishment of the National Key Deer Refuge in 1957. The approximately 8600-acre refuge includes 2280 acres of designated wilderness.

The Key Deer Refuge has increased the population of Key deer. A recent study estimated the total Key deer population to be approximately 800.

About half of Key deer deaths are due to vehicles.

One of two Key deer wildlife underpasses on Big Pine Key

What Is Your Answer?

3. **IN YOUR OWN WORDS** How can you display data in a way that helps you make decisions? Use the Internet or some other reference to find examples of the following types of data displays.

- Bar graph
- Circle graph
- Scatter plot
- Stem-and-leaf plot
- Box-and-whisker plot

Practice Use what you learned about choosing data displays to complete Exercise 3 on page 662.

12.8 Lesson

Check It Out
Lesson Tutorials
BigIdeasMath.com

Key Idea

Data Display	What does it do?
Pictograph	shows data using pictures 
Bar Graph	shows data in specific categories 
Circle Graph	shows data as parts of a whole 
Line Graph	shows how data change over time 
Histogram	shows frequencies of data values in intervals of the same size 
Stem-and-Leaf Plot	orders numerical data and shows how they are distributed 
Box-and-Whisker Plot	shows the variability of a data set using quartiles 
Dot Plot	shows the number of times each value occurs in a data set 
Scatter Plot	shows the relationship between two data sets using ordered pairs in a coordinate plane 

EXAMPLE 1 Choosing an Appropriate Data Display

Choose an appropriate data display for the situation. Explain your reasoning.

a. the number of students in a marching band each year

A line graph shows change over time. So, a line graph is an appropriate data display.

b. comparison of people's shoe sizes and their heights

You want to compare two different data sets. So, a scatter plot is an appropriate data display.

On Your Own

Now You're Ready
Exercises 4–7

Choose an appropriate data display for the situation. Explain your reasoning.

1. the population of the United States divided into age groups
2. the percents of students in your school who speak Spanish, French, or Haitian Creole

Laurie's Notes

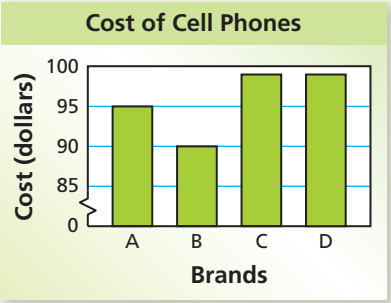
Introduction

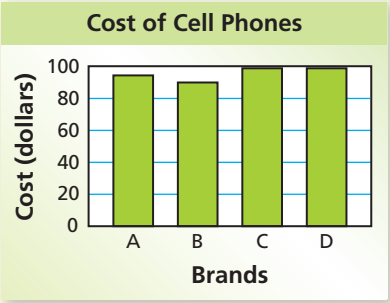
Connect

- **Yesterday:** Students reviewed data displays. (MP3, MP4)
- **Today:** Students will choose and construct an appropriate data display.

Motivate

- Make a quick sketch of the two bar graphs shown and ask students to comment on each.





Lesson Notes

Key Idea

- Write the Key Idea. This is a terrific summary of data displays that students have learned to make.
- Emphasize that *choosing an appropriate display* is more poetry than science. On the other hand, it is clearly possible to use any of the graphs in misleading ways. This is science.
- **MP3 Construct Viable Arguments and Critique the Reasoning of Others:** Students should be able to state their reasons for selecting a particular data display, *and* why they did not select a different data display. If another student selected a different data display, students should compare their reasoning.
- There may be examples of each of these displays around your room.

Example 1

- Read each problem. Students should not have difficulty determining the appropriate data display for each problem.

On Your Own

- **Think-Pair-Share:** Students should read each question independently and then work in pairs to answer the questions. When they have answered the questions, the pair should compare their answers with another group and discuss any discrepancies.

Goal Today's lesson is choosing and constructing an appropriate data display.

Technology for the Teacher

Dynamic Classroom

Lesson Tutorials
Lesson Plans
Answer Presentation Tool

Differentiated Instruction

Auditory

Ask students what data display would best represent the given data.

- the number of baseball cards each boy in the class has box-and-whisker plot
- the number of hours studying for a test and the test scores of students in a class scatter plot

Extra Example 1

You conduct a survey at your school about insects that students fear the most. Choose an appropriate data display. Explain your reasoning.

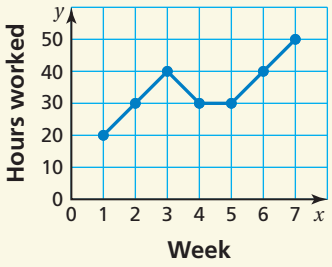
Sample answers: Circle graph: shows data as parts of a whole; Bar graph: shows data in specific categories; Pictograph: shows data using pictures.

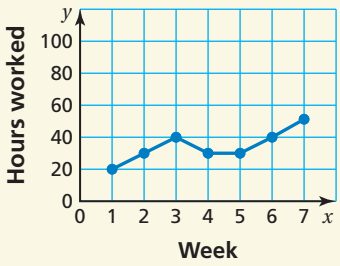
On Your Own

1. *Sample answer:* histogram; Shows frequencies of ages (data values) in intervals of the same size.
2. *Sample answer:* bar graph; Shows data in specific categories.

Extra Example 2

Which line graph is misleading? Explain.



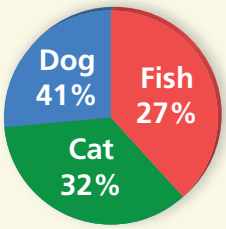


the second graph; The *y*-scale makes the change from week to week appear smaller.

Extra Example 3

Explain why the data display is misleading.

Favorite Pets



The size of each part of the circle is not proportional to the percent each choice represents.

On Your Own

3. The tickets vary in width and the break in the vertical axis makes the difference in ticket prices appear to be greater.
4. The bars become wider as the years progress, making the increase in profit appear greater.

Laurie's Notes

Discuss

- I have a collection of misleading data displays. When you find a data display in the newspaper or magazine that is misleading, cut it out and save it for later use. Ask colleagues in your school to do the same.
- **MP6 Attend to Precision:** Often what makes a graph misleading is the scale selected for one, or both, of the axes. By spreading out the scale, or condensing it, the graph becomes misleading.
- As I always tell my students, the person who makes the data display influences how we will view it. They control the extent to which we can see, or not see, features of the data.

Example 2

- ? "The same data are displayed in each line graph. How do the graphs differ?" The vertical scale is different.
- ? "Which graph is misleading and why?" first graph; It makes it appear that there has been a rapid growth in box office receipts.
- **Extension:** Have students pretend that both graphs appear in the newspaper with an article, and ask them what they would use for a headline for each article. What story does the author want readers to see when they look at each graph?

Example 3

- Have students "read" the pictograph and ask them to summarize what information it describes.
- Many students will conclude that the amount of cans and the amount of boxes is about the same due to the horizontal distance each set of icons takes up. They are mistakenly reading it more like a bar graph.
- ? "Approximately how many cans of food and boxes of food have been donated?" 11 cans $\times$ 20 = 220 cans; 6 boxes $\times$ 20 = 120 boxes
- Almost twice as many cans of food have been donated as boxes, so this is misleading. The box icon is too large. It should be the same width as the can.

On Your Own

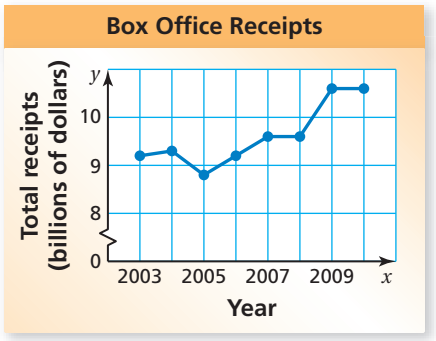
- **Think-Pair-Share:** Students should read each question independently and then work in pairs to answer the questions. When they have answered the questions, the pair should compare their answers with another group and discuss any discrepancies.

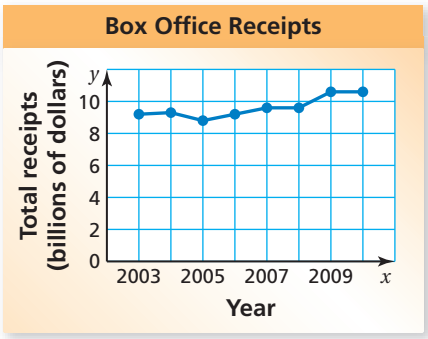
Closure

- **Exit Ticket:** Make a pictograph for the data in Example 3 that would not be misleading.

EXAMPLE 2 Identifying a Misleading Data Display

Which line graph is misleading? Explain.

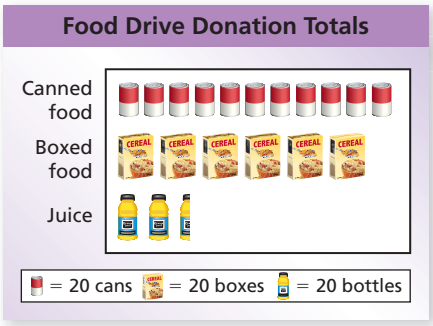




The vertical axis of the line graph on the left has a break (≒) and begins at 8. This graph makes it appear that the total receipts increased rapidly from 2005 to 2010. The graph on the right has an unbroken axis. It is more honest and shows that the total receipts increased slowly.

∴ So, the graph on the left is misleading.

EXAMPLE 3 Analyzing a Misleading Data Display



A volunteer concludes that the numbers of cans of food and boxes of food donated were about the same. Is this conclusion accurate? Explain.

Each icon represents the same number of items. Because the box icon is larger than the can icon, it looks like the number of boxes is about the same as the number of cans, but the number of boxes is actually about half of the number of cans.

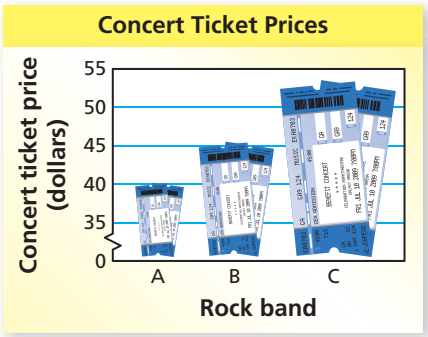
∴ So, the conclusion is not accurate.

On Your Own

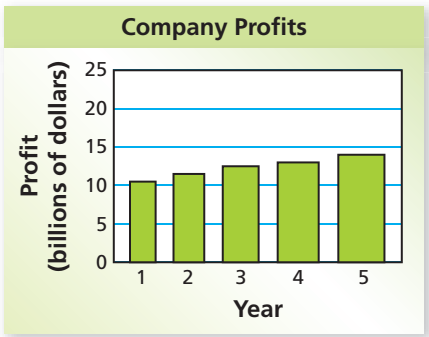
Now You're Ready
Exercises 9–12

Explain why the data display is misleading.

3.



4.



12.8 Exercises

Check It Out
Help with Homework
BigIdeasMath.com

Vocabulary and Concept Check

1. **REASONING** Can more than one display be appropriate for a data set? Explain.
2. **OPEN-ENDED** Describe how a histogram can be misleading.

Practice and Problem Solving

3. Analyze and display the data in a way that best describes the data. Explain your choice of display.

Notebooks Sold in One Week				
192 red	170 green	203 black	183 pink	230 blue
165 yellow	210 purple	250 orange	179 white	218 other

Choose an appropriate data display for the situation. Explain your reasoning.

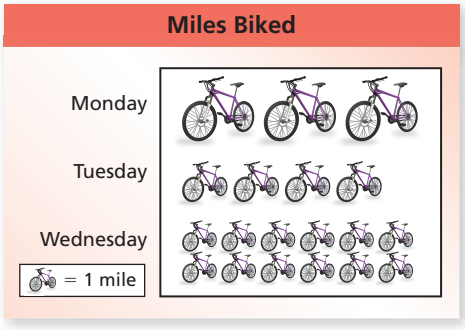
1 4. a student's test scores and how the scores are spread out

5. the distance a person drives each month
6. the outcome of rolling a number cube
7. homework problems assigned each day

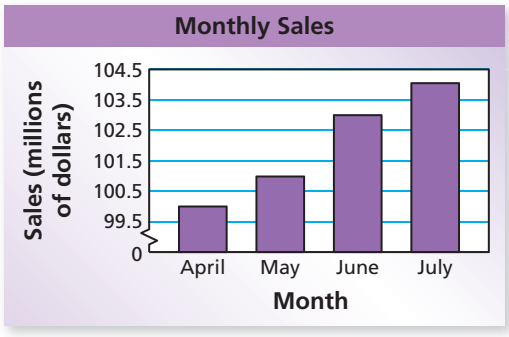
8. **WRITING** When would you choose a histogram instead of a bar graph to display data?

Explain why the data display is misleading.

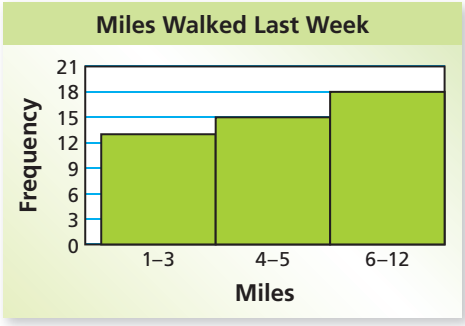
2 3 9.



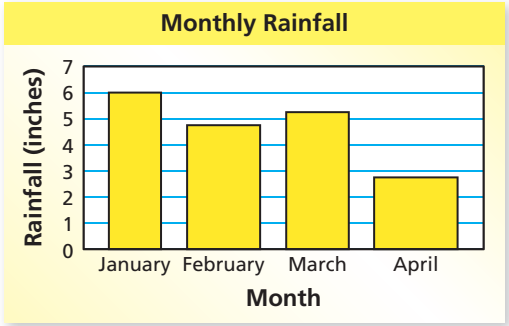
10.



11.



12.



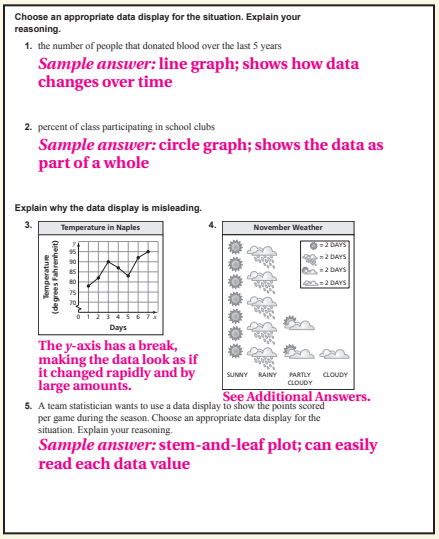
Assignment Guide and Homework Check

Level	Assignment	Homework Check
Average	1, 2, 4–14, 18–20	4, 8, 10, 14
Advanced	1, 2, 4–12 even, 13–17, 18–20	8, 10, 14, 16

Common Errors

- **Exercises 9–12** Students may not be able to recognize why the data display is misleading. As a class, make a list of things to examine when analyzing a data display. For example, check the increments or intervals for the axes.
- **Exercise 15** Students may say that the best data display for showing the mode is a stem-and-leaf plot because the leaves that have more repeated data will be wider. This display, however, could have other data in the leaf. Remind them that a dot plot isolates each data value and shows the frequency of each individual number, so this is the best data display.

12.8 Record and Practice Journal



Choose an appropriate data display for the situation. Explain your reasoning.

1. the number of people that donated blood over the last 5 years

Sample answer: line graph; shows how data changes over time

2. percent of class participating in school clubs

Sample answer: circle graph; shows the data as part of a whole

Explain why the data display is misleading.

3. Temperature in Naples

The y-axis has a break, making the data look as if it changed rapidly and by large amounts.

4. November Weather

SUNNY RAINY PARTLY CLOUDY CLOUDY

See Additional Answers.

5. A team statistician wants to use a data display to show the points scored per game during the season. Choose an appropriate data display for the situation. Explain your reasoning.

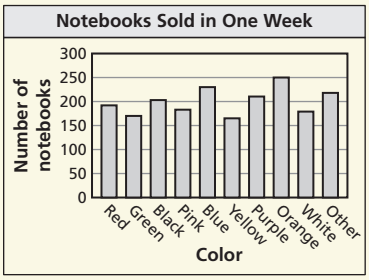
Sample answer: stem-and-leaf plot; can easily read each data value

Vocabulary and Concept Check

1. yes; Different displays may show different aspects of the data.
2. *Sample answer:* The scale of the vertical axis could be too small or too large.

Practice and Problem Solving

3. *Sample answer:*



A bar graph shows the data in different color categories.

4. *Sample answer:* Stem-and-leaf plot: shows how data is distributed.
5. *Sample answer:* Dot graph: shows changes over time.
6. *Sample answer:* Line plot: shows the number of times each outcome occurs.
7. *Sample answer:* Line graph: shows changes over time.
8. when the data is in terms of intervals of one category, as opposed to multiple categories
9. The pictures of the bikes are larger on Monday, which makes it seem like the distance is the same each day.
10. The break in the scale for the vertical axis makes it appear as though there is a greater difference in sales between months.
11. The intervals are not the same size.

Practice and Problem Solving

12. The width of the bars are different, so it looks like some months have more rainfall.

13. *Sample answer:* bar graph; Each bar can represent a different vegetable.

14. yes; The vertical axis has a scale that increases by powers of 10, which makes the data appear to have a linear relationship.

15. *Sample answer:* dot plot

16. **a.** The percents do not add up to 100%.

b. *Sample answer:* bar graph; It would show the frequency of each sport.

17. See *Taking Math Deeper*.

Fair Game Review

18. $x + 3 = 5$

19. $8x = 24$

20. A

Mini-Assessment

Choose an appropriate data display for the situation. Explain your reasoning.

1. the outcome of flipping a coin
Sample answers: Pictograph: shows number of times heads or tails appears using picture of coins; Bar graph: shows number of times you get heads or tails; Dot plot: shows number of times you get a heads or tails.

2. comparison of student's test scores and how long students studied
Sample answer: Scatter plot: you want to compare two data sets.

3. the number of students participating in after-school sports each year
Sample answer: Line graph: shows how data change over time.

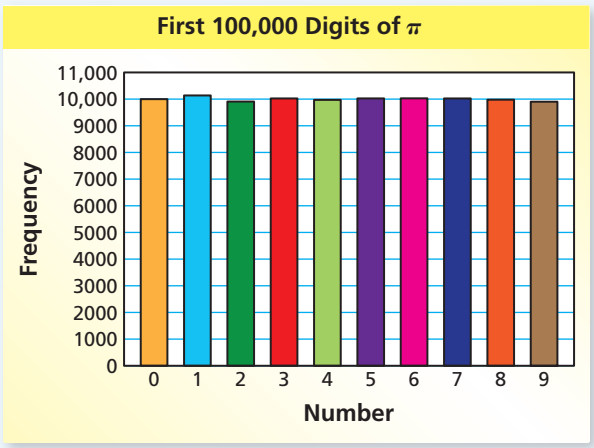
Taking Math Deeper

Exercise 17

This exercise introduces students to an amazing property of the number pi. Pi is an irrational number and therefore its decimal representation is not repeating. Even so, the ten digits from 0 to 9 each occur about 10 percent of the time, when one considers thousands of digits.

1 Display the data in a bar graph.

a.



2 Display the data in a circle graph.

b.

3 **c. and d.** Compare the two displays.

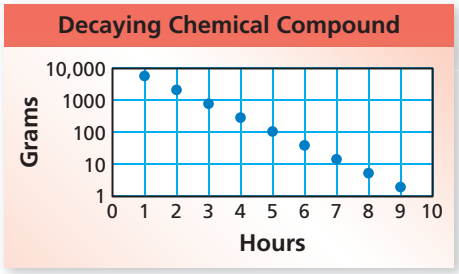
Both graphs show that each digit occurs about 10% of the time. The bar graph has a slight advantage because it shows that some digits occur slightly more than others.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Resources by Chapter • Practice A and Practice B • Puzzle Time Record and Practice Journal Practice Differentiating the Lesson Lesson Tutorials Skills Review Handbook	Resources by Chapter • Enrichment and Extension • School-to-Work • Financial Literacy • Technology Connection • Life Connections • Stories in History Start the next section

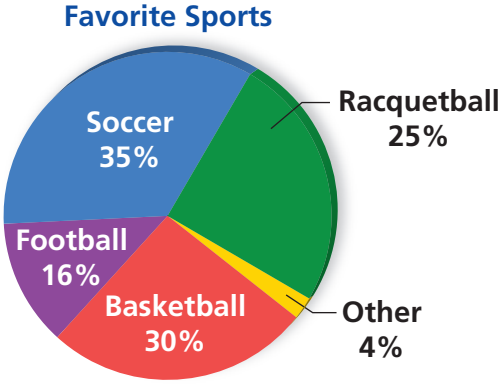
13. VEGETABLES A nutritionist wants to use a data display to show the favorite vegetables of the students at a school. Choose an appropriate data display for the situation. Explain your reasoning.

14. CHEMICALS A scientist gathers data about a decaying chemical compound. The results are shown in the scatter plot. Is the data display misleading? Explain.



15. REASONING What type of data display is appropriate for showing the mode of a data set?

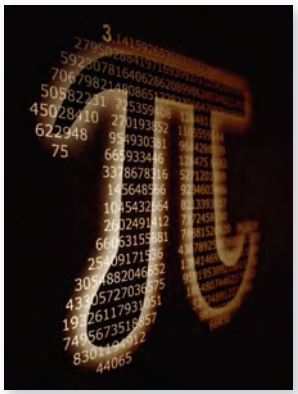
16. SPORTS A survey asked 100 students to choose their favorite sports. The results are shown in the circle graph.



a. Explain why the graph is misleading.

b. What type of data display would be more appropriate for the data? Explain.

17. Structure With the help of computers, mathematicians have computed and analyzed billions of digits of the irrational number π. One of the things they analyze is the frequency of each of the numbers 0 through 9. The table shows the frequency of each number in the first 100,000 digits of π.



a. Display the data in a bar graph.

b. Display the data in a circle graph.

c. Which data display is more appropriate? Explain.

d. Describe the distribution.

Number	0	1	2	3	4	5	6	7	8	9
Frequency	9999	10,137	9908	10,025	9971	10,026	10,029	10,025	9978	9902

Fair Game Review What you learned in previous grades & lessons

Write the verbal statement as an equation. *(Skills Review Handbook)*

18. A number plus 3 is 5.

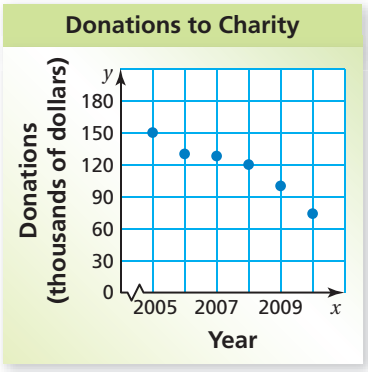
19. 8 times a number is 24.

20. MULTIPLE CHOICE What is 20% of 25% of 400? *(Skills Review Handbook)*

Ⓐ 20 Ⓑ 200 Ⓒ 240 Ⓓ 380

12.5–12.8 Quiz

Check It Out
Progress Check
BigIdeasMath.com



1. The scatter plot shows the amounts of money donated to a charity from 2005 to 2010. *(Section 12.5)*

a. In what year did the charity receive $150,000?

b. How much did the charity receive in 2008?

c. Describe the relationship shown by the data.

2. Use a graphing calculator to find the equation of the line of best fit for the data in the table below. Identify and interpret the correlation coefficient. Make a scatter plot of the residuals and interpret the results. *(Section 12.6)*

x	0	1	2	3	4	5	6	7
y	12	16	15	14	18	22	20	25

3. The results of a recycling survey are shown in the two-way table. Find and interpret the marginal frequencies. *(Section 12.7)*

		Recycle	
		Yes	No
Gender	Female	28	9
	Male	24	14

Choose an appropriate data display for the situation. Explain your reasoning. *(Section 12.8)*

4. percent of band students in each section of instruments

5. company's profit for each week

6. CATS The table shows the number of cats adopted from an animal shelter each month. *(Section 12.5)*

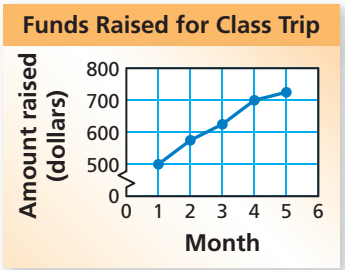
Month	1	2	3	4	5	6	7	8	9
Cats	3	6	7	11	13	14	15	18	19

a. Make a scatter plot of the data and draw a line of fit.

b. Write an equation of the line that fits the data.

c. Interpret the slope of the line.

d. Predict how many cats will be adopted in month 10.



7. FUNDRAISER The line graph shows the amount of money that the eighth-grade students at a school raised each month to pay for a class trip. Is the graph misleading? Explain. *(Section 12.8)*

Alternative Assessment Options

Math Chat Student Reflective Focus Question
Structured Interview Writing Prompt

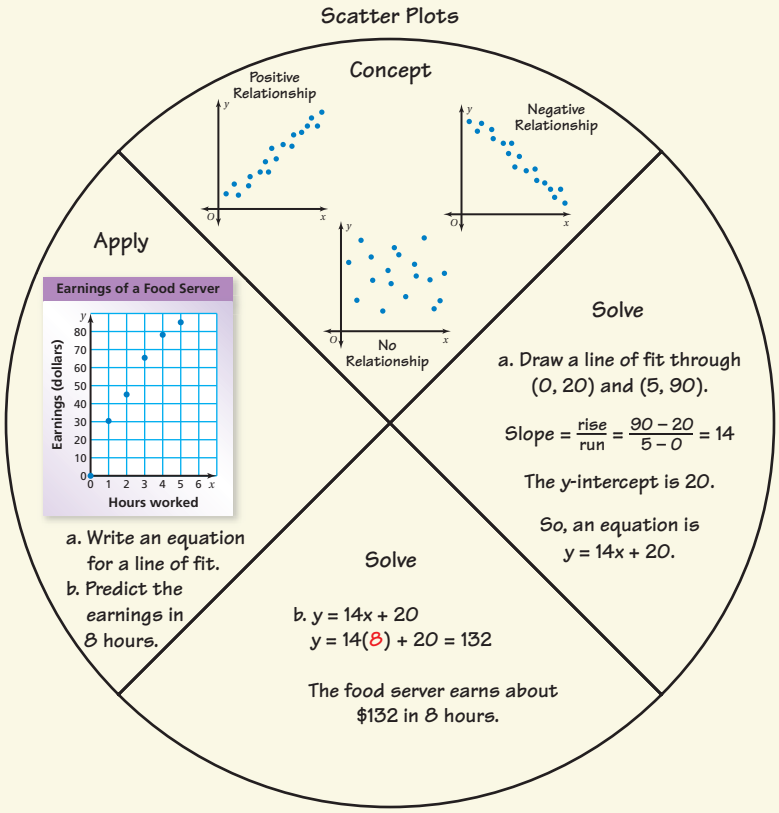
Math Chat

Ask students to use their own words to summarize how they would choose an appropriate data display. Be sure that they include examples. Select students at random to present to the class.

Study Help Sample Answers

Remind students to complete Graphic Organizers for the rest of the chapter.

4.



5–7. Available at *BigIdeasMath.com*.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Resources by Chapter • Study Help • Practice A and Practice B • Puzzle Time Lesson Tutorials *BigIdeasMath.com*	Resources by Chapter • Enrichment and Extension • School-to-Work Game Closet at *BigIdeasMath.com* Start the Chapter Review

Answers

1. **a.** 2005
   b. $120,000
   c. There is a negative relationship between year and amount of donations.
2. $y = 1.6x + 12$; $r \approx 0.908$; This means that the relationship between x and y is a fairly strong positive correlation and the equation fits the data fairly well; The residuals are random values from −2.8 to 2.4, so the equation is a good fit.
3. 52 people recycle; 23 people do not recycle; 37 females were surveyed; 38 males were surveyed.
4. *Sample answer:* Circle graph: shows data as parts of a whole.
5. *Sample answer:* Line graph: shows changes over time.
6. See Additional Answers.
7. yes; The break in the vertical axis makes it appear that the amount of money raised increased very rapidly from month to month.

Technology for the Teacher

Online Assessment
Assessment Book
ExamView® Assessment Suite

For the Teacher
Additional Review Options

- *BigIdeasMath.com*
- Online Assessment
- Game Closet at *BigIdeasMath.com*
- Vocabulary Help
- Resources by Chapter

Answers

1. The mean stays the same, the median decreases $4.2 - 4.1 = 0.1$ kilometer, and the mode changes to 4.0 and 4.3.
2. mean: 1.7
   median: 1
   mode: 1
3. mean: 4
   median: 3
   mode: 10

Review of Common Errors

- **Exercises 1–3** When finding the mean, students may forget to divide by the total number of data values and instead divide by the maximum value. Remind them that the definition of mean is an "average," so they must take into account the total number of items or numbers to get an average. Explain to students that it is as if they are dividing the total evenly among the number of groups.
- **Exercises 1–3** Students may try to identify the median without ordering the data first. Remind them that it is essential to order the data first and then find the median. This also makes finding the mode easier.

12 Chapter Review

Check It Out
Vocabulary Help
BigIdeasMath.com

Review Key Vocabulary

measure of central tendency, *p. 608*
measure of dispersion, *p. 614*
range, *p. 614*
standard deviation, *p. 615*
box-and-whisker plot, *p. 620*
quartile, *p. 620*
five-number summary, *p. 620*
interquartile range, *p. 621*
scatter plot, *p. 638*
line of fit, *p. 640*
residual, *p. 646*
linear regression, *p. 647*
line of best fit, *p. 647*
correlation coefficient, *p. 647*
causation, *p. 648*
two-way table, *p. 654*
joint frequency, *p. 654*
marginal frequency, *p. 654*

Review Examples and Exercises

12.1 Measures of Central Tendency *(pp. 606–611)*

The table shows the number of kilometers you ran each day for the past 10 days. Find the mean, median, and mode of the distances.

Kilometers Run	
3.5	4.1
4.0	4.3
4.4	4.5
3.9	2.0
4.3	5.0

Mean: sum of the data → $\frac{40}{10} = 4$ ← number of values

Median: 2.0, 3.5, 3.9, 4.0, 4.1, 4.3, 4.3, 4.4, 4.5, 5.0 — Order the data.

$\frac{8.4}{2} = 4.2$ — Mean of two middle values

Mode: 2.0, 3.5, 3.9, 4.0, 4.1, 4.3, 4.3, 4.4, 4.5, 5.0

The value 4.3 occurs most often.

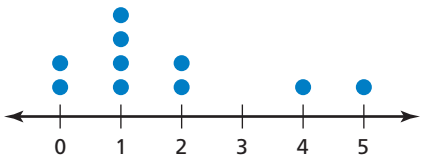
∴ The mean is 4 kilometers, the median is 4.2 kilometers, and the mode is 4.3 kilometers.

Exercises

1. Use the data in the example above. You run 4.0 kilometers on day 11. How does this additional value affect the mean, median, and mode? Explain.

Find the mean, median, and mode of the data.

2.



Goals per game

3.

Ski Resort Temperatures (°F)		
11	3	3
0	−9	−2
10	10	10

12.2 Measures of Dispersion *(pp. 612–617)*

Find the mean, range, and standard deviation of the bowling scores for each person. Then compare the data sets.

Ryan	
205	190
185	200
210	219
174	203
194	230

Emma	
228	205
172	181
154	240
235	235
168	192

Ryan: $\text{mean} = \frac{2010}{10} = 201$

174, 185, 190, 194, 200, 203, 205, 210, 219, 230 Order the data.

The range is $230 - 174 = 56$.

$$\sqrt{\frac{(205 - 201)^2 + (185 - 201)^2 + \cdots + (230 - 201)^2}{10}} = \sqrt{242.2} \approx 15.6$$

The standard deviation is 15.6.

Emma: $\text{mean} = \frac{2010}{10} = 201$

154, 168, 172, 181, 192, 205, 228, 235, 235, 240 Order the data.

The range is $240 - 154 = 86$.

$$\sqrt{\frac{(228 - 201)^2 + (172 - 201)^2 + \cdots + (192 - 201)^2}{10}} = \sqrt{919.8} \approx 30.3$$

The standard deviation is 30.3.

The mean, 201, is the same for each data set. The range for Ryan's scores is 56 and the standard deviation is 15.6. The range for Emma's scores is 86 and the standard deviation is 30.3. So, Emma's scores are more spread out than Ryan's scores.

Exercises

4. Find the mean, range, and standard deviation of the prices (in dollars) of portable keyboards at each store. Then compare the data sets.

Store A	
130	180
200	250
150	190
250	160

Store B	
225	310
260	190
200	285
210	230

Review of Common Errors (continued)

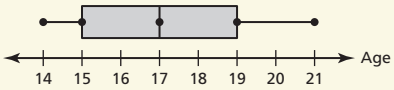
- **Exercise 4** Students may subtract the first data value from the last data value to find the range. Remind students that the range is the difference between the greatest and least data values. It helps to order the data before identifying the greatest and least data values.

Answers

4. Store A: 188.75; 120; about 41.1
Store B: 238.75; 120; about 39.7
The mean at Store B is greater, the ranges are the same, and the standard deviations are about the same. So, the portable keyboard's prices at Store B tend to be higher and they differ from the mean about as much as at Store A.

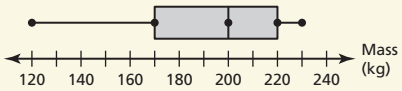
Answers

5.



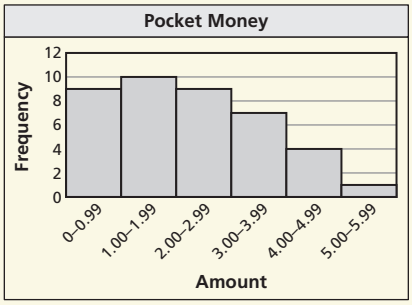
The distribution is skewed right.

6.



The distribution is skewed left.

7.



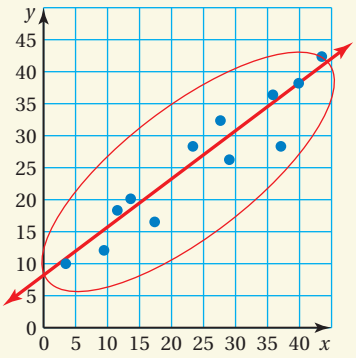
The distribution is skewed right.

8. Use the median to describe the center and the five-number summary to describe the spread.

9. *Sample answer:* $y = -0.02x + 8$; \$2.50

Review of Common Errors (continued)

- **Exercises 5–6** Students may have difficulty creating the box-and-whisker plots. Remind them of the five-number summary of a data set.
- **Exercises 5–6** Students may confuse the meanings of skewed left and skewed right. Remind them that the shape is determined by the longer whisker.
- **Exercise 7** Students may create a histogram with gaps between the bars. Remind students that the bars in a histogram should be touching to show that they represent an entire range of values.
- **Exercise 9** Students may draw a line of best fit that does not accurately reflect the data trend. Remind them that the line does not have to go through any of the data points. Also remind them that the line should go through the middle of the data so that about half of the data points are above the line and half are below. One strategy is to draw an oval around the data and then draw a line through the middle of the oval. For example:



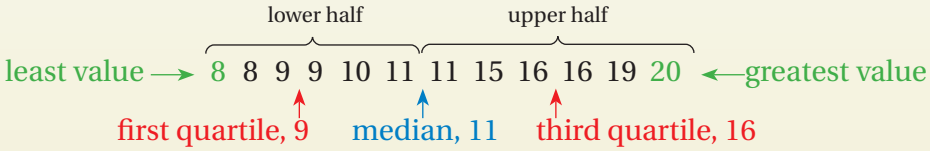
- **Exercise 9** Students may struggle writing an equation for a line of fit. When drawing the line, encourage them to try to make the line go through a lattice point. Also, students can use lattice points that are very close to the line to help them find the slope.
- **Exercise 10** Students may enter the x-values into the y list of the graphing calculator, and vice versa. Encourage them to double check their data entries and understand which lists represent the x- and y-values.
- **Exercise 11** Students may incorrectly identify joint frequencies as marginal frequencies or marginal frequencies as joint frequencies.

12.3 Box-and-Whisker Plots *(pp. 618–625)*

Make a box-and-whisker plot for the weights (in pounds) of pumpkins sold at a market. Identify the shape of the distribution.

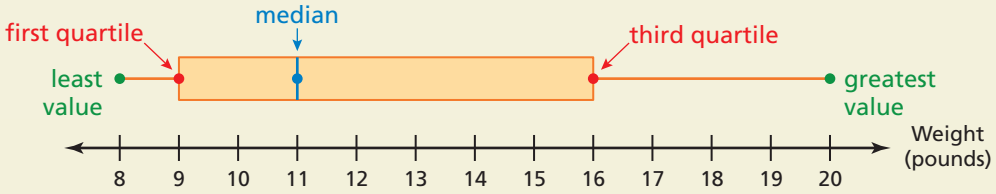
16, 20, 11, 15, 10, 8, 8, 19, 11, 9, 9, 16

Step 1: Order the data. Find the median and the quartiles.



Step 2: Draw a number line that includes the least and greatest values. Graph points above the number line for the five-number summary.

Step 3: Draw a box using the quartiles. Draw a line through the median. Draw whiskers from the box to the least and greatest values.



The right whisker is longer than the left whisker, and most of the data are on the left side of the display. So, the distribution is skewed right.

Exercises

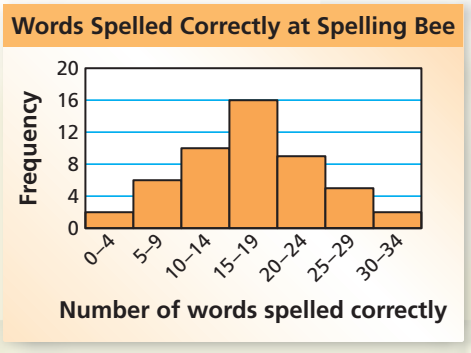
Make a box-and-whisker plot for the data. Identify the shape of the distribution.

5. Ages of volunteers at a hospital: 14, 17, 20, 16, 17, 14, 21, 18

6. Masses (in kilograms) of lions: 120, 230, 180, 210, 200, 200, 230, 160

12.4 Shapes of Distributions *(pp. 626–633)*

The histogram shows the numbers of words spelled correctly by students at a spelling bee. Describe the shape of the distribution. Which measures of central tendency and dispersion would best represent the data?



The distribution is symmetric. So, use the mean to describe the center and the standard deviation to describe the spread.

Exercises

The frequency table shows the amounts of money the students in a class have in their pockets.

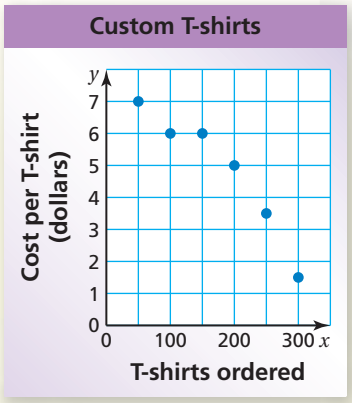
Amount	Frequency
0–0.99	9
1.00–1.99	10
2.00–2.99	9
3.00–3.99	7
4.00–4.99	4
5–5.99	1

7. Display the data in a histogram. Describe the shape of the distribution.

8. Which measures of central tendency and dispersion best represent the data?

12.5 Scatter Plots and Lines of Fit *(pp. 636–643)*

Your school is ordering custom T-shirts. The scatter plot shows the costs per T-shirt for various numbers of T-shirts ordered. What type of relationship do the data show?



The plotted points go down from left to right. As the number of T-shirts ordered increases, the cost per T-shirt decreases.

∴ So, the scatter plot shows a negative relationship.

Exercises

9. Use the scatter plot above. Write an equation of a line of fit. Predict the cost per T-shirt when you order 275 T-shirts.

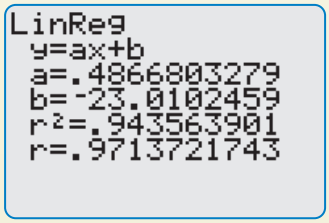
12.6 Analyzing Lines of Fit *(pp. 644–651)*

The table shows the heights x (in inches) and shoe sizes y of several students. Use a graphing calculator to find an equation of the line of best fit. Identify and interpret the correlation coefficient.

x	65	62	70	72	68	67	70	67	64	63
y	8.5	7	10.5	12	10	10.5	11	10	8	7.5

Step 1: Enter the data from the table into your calculator.

Step 2: Use the *linear regression* feature of your calculator to find an equation of the line of best fit and the correlation coefficient.



∴ An equation of the line of best fit is $y = 0.49x - 23.0$. The correlation coefficient is about 0.971. This means that the relationship between the heights and shoe sizes is a strong positive correlation and the equation closely models the data.

Review Game

Rolling for Data

Materials per Pair

- two number cubes
- paper
- pencil

Directions

Students should work in pairs. Students in each pair take turns rolling the two number cubes one at a time. The first number they roll represents the tens digit and the second number represents the ones digit of a whole number. For example, If a 1 is rolled and then a 6, the whole number is 16. Students record the whole numbers in a stem-and-leaf plot and keep rolling until they have 10 leaves for any one stem. Once a pair acquires the 10 leaves, they race to find the mean, median, and mode of all their whole numbers and make a box-and-whisker plot to display the data.

Who Wins?

The first pair to finish all tasks wins 10 points, the second 9 points, the third 8 points, and so on. The game can be repeated as many times as desired. The pair with the most points after a predetermined number of rounds or amount of time wins.

For the Student
Additional Practice

- Lesson Tutorials
- Multi-Language Glossary
- Self-Grading Progress Check
- *BigIdeasMath.com*
  Dynamic Student Edition
  Student Resources

Answers

10. $y = 0.49x - 23.2$; yes; The correlation coefficient changed from about 0.971 to about 0.954. There is still a strong positive correlation.

11.

		Food Court		
		Like	Dislike	Total
Age	Adults	21	79	100
	Students	96	4	100
	Total	117	83	200

79%

12. *Sample answer:* Bar graph: shows data in specific categories.

My Thoughts on the Chapter

What worked. . .

What did not work. . .

What I would do differently. . .

Teacher Tip

Not allowed to write in your teaching edition? Use sticky notes to record your thoughts.

Exercises

10. Use the data in the example. You take height and shoe size measurements of three more students: (64, 7), (65, 9), and (71, 11). Find a new equation of the line of best fit. Did the correlation coefficient change? Explain.

12.7 Two-Way Tables *(pp. 652–657)*

You randomly survey students in your school about whether they liked a recent school play. The results are shown. Make a two-way table that includes the marginal frequencies. What percent of the students surveyed liked the play?

Male students
48 likes, 12 dislikes

Female students
56 likes, 14 dislikes

		Student		
		Liked the Play	Did Not Like the Play	Total
Gender	Male	48	12	60
	Female	56	14	70
	Total	104	26	130

Of the 130 students surveyed, 104 students liked the play.

∴ Because $\frac{104}{130} = 0.8$, 80% of the students in the survey liked the play.

Exercises

11. You randomly survey people at a mall about whether they like the new food court. The results are shown. Make a two-way table that includes the marginal frequencies. What percent of the adults surveyed dislike the new food court?

Adults
21 likes, 79 dislikes

Teenagers
96 likes, 4 dislikes

12.8 Choosing a Data Display *(pp. 658–663)*

Choose an appropriate data display for the situation. Explain your reasoning.

a. the percent of votes that each candidate received in an election

∴ A circle graph shows data as parts of a whole. So, a circle graph is an appropriate data display.

b. the distribution of the ages of U.S. presidents at their inauguration(s)

∴ A stem-and-leaf plot orders numerical data and shows how they are distributed. So, a stem-and-leaf plot is an appropriate data display.

Exercises

12. A principal wants to use a data display to compare the number of cans of food donated by each eighth-grade class. Choose an appropriate data display for the situation. Explain your reasoning.

12 Chapter Test

Check It Out
Test Practice
BigIdeasMath.com

Find the mean, median, and mode of the data.

1.

Distances (feet) Above or Below Water Level in Pool		
−3	0	−3
3	10	0
11	−6	−3

2. **Cooking Times (minutes)**

Stem	Leaf
3	5 8
4	0 1 8
5	0 4 4 4 5 9
6	0

Key: 4|1 = 41 minutes

3. TURTLES The tables show the weights (in pounds) of turtles caught in two ponds. Find the mean, range, and standard deviation of the weights of the turtles in each pond. Then compare the data sets.

Pond A			
12	13	15	6
7	8	12	7

Pond B			
9	12	5	8
12	15	16	19

4. Which type of data display would you use for the information in Exercise 3? Explain.

5. SWIMMING The table shows the numbers of hours you swam for several weeks.

Hours Swimming			
7	3.5	8	3.5
6	7	7	2
5.5	7.5	7.5	7.5

a. Make a box-and-whisker plot for the data.

b. Find the range and interquartile range.

c. Which measures of central tendency and dispersion best represent the data?

6. NEWBORNS The table shows the lengths and weights of several newborn babies.

Length (inches)	Weight (pounds)
19	6
19.5	7
20	7.75
20.25	8.5
20.5	8.5
22.5	11

a. Write an equation of a line that fits the data.

b. Use a graphing calculator to find an equation of the line of best fit. Identify and interpret the correlation coefficient.

c. Predict the weight of a newborn that is 21 inches long using the equations from parts (a) and (b). Compare the results.

7. SAT The table shows the numbers y of students (in thousands) who took the SAT from 2003 to 2010, where $x = 3$ represents the year 2003. Use a graphing calculator to find an equation of the line of best fit. Then make a scatter plot of the residuals to tell whether the line of best fit models the data well.

x	3	4	5	6	7	8	9	10
y	1406	1419	1476	1466	1495	1519	1530	1548

8. RECYCLING You randomly survey shoppers at a supermarket about whether they use reusable bags. Of 60 male shoppers, 15 use reusable bags. Of 110 female shoppers, 60 use reusable bags. Organize your results in a two-way table. Include the marginal frequencies.

Test Item References

Chapter Test Questions	Section to Review	Common Core State Standards
1, 2	12.1	S.ID.2, S.ID.3
3	12.2	S.ID.2, S.ID.3
5a, 5b	12.3	S.ID.1, S.ID.2, S.ID.3
5c	12.4	S.ID.2, S.ID.3
6a, 6c	12.5	8.SP.1, S.ID.6a, S.ID.6c
6b, 6c, 7	12.6	S.ID.6b, S.ID.8, S.ID.9
8	12.7	8.SP.4, S.ID.5
4	12.8	S.ID.1

Test-Taking Strategies

Remind students to quickly look over the entire test before they start so that they can budget their time. When they receive their test, students should list the different types of data displays. Have students use the **Stop** and **Think** strategy before they answer each question.

Common Errors

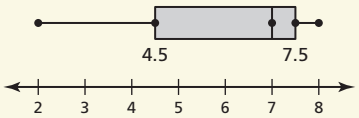
- **Exercises 1–2** When finding the mean, students may forget to divide by the total number of data values and instead divide by the maximum value. Remind them that the definition of mean is an "average," so they must take into account the total number of items or numbers to get an average. Explain to students that it is as if they are dividing the total evenly among the number of groups.
- **Exercise 3** Students may subtract the first data value from the last data value to find the range. Remind students that the range is the difference between the greatest and least data values.
- **Exercise 5** Students may have difficulty creating the box-and-whisker plot. Remind them of the five-number summary of a data set.

Reteaching and Enrichment Strategies

If students need help. . .	If students got it. . .
Resources by Chapter • Practice A and Practice B • Puzzle Time Record and Practice Journal Practice Differentiating the Lesson Lesson Tutorials *BigIdeasMath.com* Skills Review Handbook	Resources by Chapter • Enrichment and Extension • School-to-Work • Financial Literacy • Technology Connection • Life Connections • Stories in History Game Closet at *BigIdeasMath.com* Start Standards Assessment

Answers

1. mean: 1; median: 0; mode: -3
2. mean: 49; median: 52; mode: 54
3. Pond A: 10; 9; about 3.2
   Pond B: 12; 14; about 4.3
   The turtles in Pond B tend to be heavier and their weights tend to differ more from the mean than the turtles in Pond A.
4. *Sample answer:*
   Box-and whisker plot: shows the mean and variability of the data.
5. **a.**



   b. 6; 3

   c. median for central tendency, five-number summary for spread
6. See Additional Answers.
7. $y = 20.3x + 1350$
8. See Additional Answers.

Technology for the Teacher

Online Assessment
Assessment Book
ExamView® Assessment Suite

Test-Taking Strategies

Available at *BigIdeasMath.com*

After Answering Easy Questions, Relax
Answer Easy Questions First
Estimate the Answer
Read All Choices before Answering
Read Question before Answering
Solve Directly or Eliminate Choices
Solve Problem before Looking at Choices
Use Intelligent Guessing
Work Backwards

About this Strategy

When taking a multiple choice test, be sure to read each question carefully and thoroughly. Look closely for words that change the meaning of the question, such as *not*, *never*, *all*, *every*, and *always*.

Answers

1. B
2. G
3. 16
4. C

Item Analysis

1. **A.** The student assumes the correct answer is the least of the answer alternatives, because 4 of the 7 scores in the table are greater than 45.5.

   B. Correct answer

   C. The student uses the value of the mean given in the problem statement.

   D. The student finds the mean of the seven given scores.

2. **F.** The student confuses the direction of the vertical translation and the direction in which the parabola opens.

   G. Correct answer

   H. The student confuses the direction of the vertical translation.

   I. The student confuses the direction in which the parabola opens.

3. **Gridded Response:** Correct answer: 16

   Common Error: Instead of squaring 4, the student takes the square root of 4, yielding an answer of 2.

4. **A.** The student multiplies coefficients to get the coefficient of z and adds the constant terms.

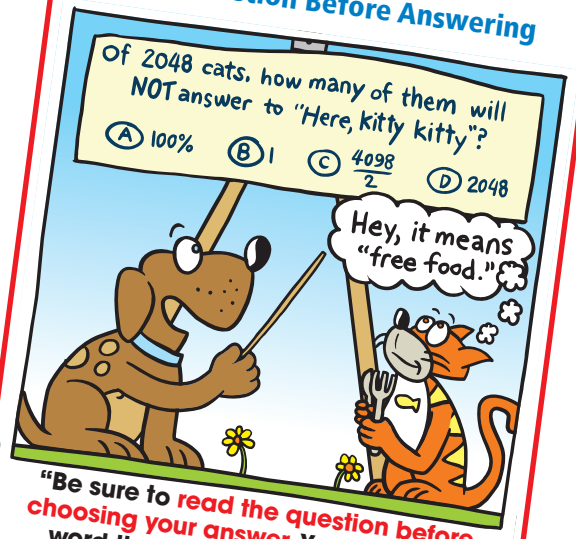
   B. The student adds $4z$ instead of subtracting $4z$.

   C. Correct answer

   D. The student forgets the product of the outside terms.

12 Standards Assessment

Test-Taking Strategy
Read Question Before Answering



"Be sure to read the question before choosing your answer. You may find a word that changes the meaning."

1. What is the value of x when the mean of the video game scores is 45.5? *(S.ID.2)*

Video Game Scores			
36	28	x	48
42	57	63	52

A. 35

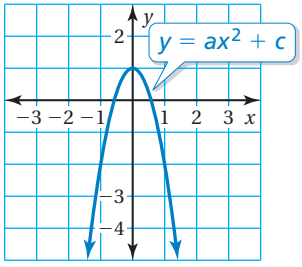
B. 38

C. 45.5

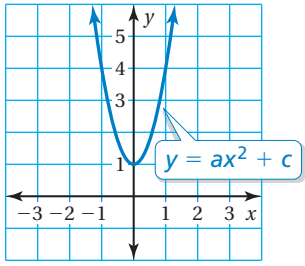
D. 46.57

2. Which graph represents $y = 3x^2 - 1$? *(F.BF.3)*

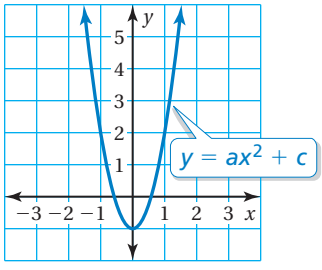
F.



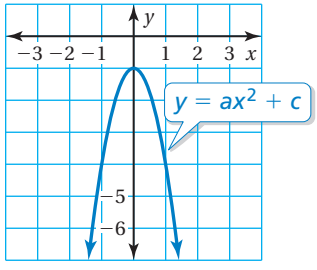
H.



G.



I.



3. What is the solution of the equation $-6 + \sqrt{x} = -2$? *(N.RN.2)*

4. What is the product $(3z - 2)(2z + 4)$? *(A.APR.1)*

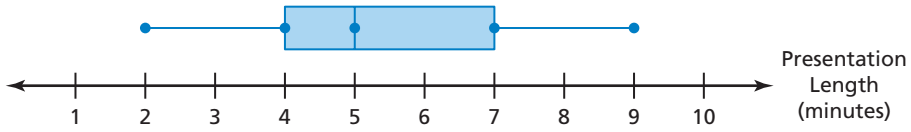
A. $6z + 2$

B. $6z^2 + 16z + 8$

C. $6z^2 + 8z - 8$

D. $6z^2 - 4z - 8$

5. The box-and-whisker plot represents the lengths of project presentations (in minutes) at a science fair. Find the interquartile range of the data. What does this represent in the context of the situation? *(S.ID.2)*



F. 7; The middle half of the presentation lengths vary by no more than 7 minutes.

G. 3; The presentation lengths vary by no more than 3 minutes.

H. 3; The middle half of the presentation lengths vary by no more than 3 minutes.

I. 7; The presentation lengths vary by no more than 7 minutes.

6. What is the simplified form of the expression? *(N.RN.2)*

$$\left(\frac{4}{5x}\right)^{-2}$$

A. $\frac{25}{16x^2}$

B. $\frac{25x^2}{16}$

C. $\frac{16}{25x^2}$

D. $\frac{16x^2}{25}$

7. Which equation shows inverse variation? *(A.REI.10)*

F. $y = -3x + 7$

G. $2y = x$

H. $y = \frac{1}{5}x$

I. $y = \frac{4}{x}$

8. You randomly survey students in your school. You ask whether they have jobs. You display your results in the two-way table. How many male students do *not* have a job? *(S.ID.5)*

		Job	
		Yes	No
Gender	Male	27	12
	Female	31	17

Item Analysis (continued)

5. **F.** The student finds the range instead of the interquartile range.

G. The student finds the correct interquartile range, but confuses it with the range when describing what it represents.

H. Correct answer

I. The student confuses the range with the interquartile range.

6. **A.** The student applies the definition of negative exponent correctly to the numerical values but not to the variable.

B. Correct answer

C. The student applies an exponent of 2 instead of -2.

D. The student applies the definition of negative exponent correctly to the variable, but not to the numerical values.

7. **F.** The student confuses the concept of negative slope with the concept of inverse variation.

G. The student confuses $2y = x$ with the inverse variation form $k = xy$.

H. The student confuses $y = \frac{1}{5}x$ with the inverse variation form $y = \frac{k}{x}$.

I. Correct answer

8. **Gridded Response:** Correct answer: 12

Common Error: The student adds the entries in the "No" column for an answer of 29.

Answers

5. H

6. B

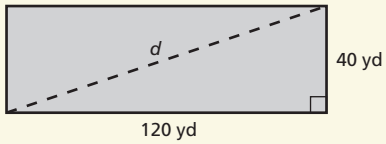
7. I

8. 12

Answers

9. D

10.



$d^2 = 40^2 + 120^2$

$d = \sqrt{40^2 + 120^2}$

$= \sqrt{1600 + 14{,}400}$

$= \sqrt{16{,}000}$

≈ 126.5 yd

The distance between opposite corners is the length of the hypotenuse of a right triangle, so you can use the Pythagorean Theorem to find the distance.

11. I

12. C

Answer for Extra Example

1. **A.** Correct answer

B. The student confuses the words "quarterly" and "quartiles."

C. The student assumes quarterly profits can be shown as a circle graph divided into four quarters.

D. The student misunderstands the use of a dot plot.

Item Analysis (continued)

9. **A.** The student uses the incorrect form $a_n = a_1 n + d$.

B. The student uses the incorrect form $a_n = a_1 n - (a_1 - d)$.

C. The student uses the incorrect form $a_n = a_1 + (n + 1)d$.

D. Correct answer

10. **2 points** The student demonstrates a thorough understanding of the Pythagorean Theorem and its application to this problem. The student correctly calculates the distance between the opposite corners of the field and explains how the Pythagorean Theorem is used.

1 point The students demonstrates a partial understanding of the Pythagorean Theorem. The student understands that the Pythagorean Theorem applies and attempts to use it, but fails to calculate the distance correctly.

0 points The student provides no response, a completely incorrect or incomprehensible response, or a response that demonstrates insufficient understanding of the Pythagorean Theorem and its application to this problem.

11. **F.** The student confuses the characteristics that show positive and negative relationships in scatter plots.

G. The student confuses the characteristics that show no relationship and a negative relationship in scatter plots.

H. The student confuses the characteristics that show constant and negative relationships in scatter plots.

I. Correct answer

12. **A.** The student makes an error placing a decimal point.

B. The student makes an error applying the Addition Property of Equality and makes an error placing a decimal point.

C. Correct answer

D. The student makes an error applying the Addition Property of Equality.

Extra Example

1. Which data display is the most appropriate to show the trend of quarterly profits for a company? *(S.ID.1)*

A. a line graph

B. a box-and-whisker plot

C. a circle graph

D. a dot plot

9. What is an equation for the nth term of the arithmetic sequence? *(F.LE.2)*

$$-\frac{3}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4}, \ldots$$

A. $a_n = -\frac{3}{4}n + \frac{1}{2}$

B. $a_n = -\frac{3}{4}n + \frac{5}{4}$

C. $a_n = \frac{1}{2}n - \frac{1}{4}$

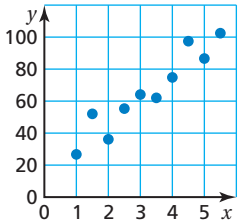
D. $a_n = \frac{1}{2}n - \frac{5}{4}$

10. Think Solve Explain

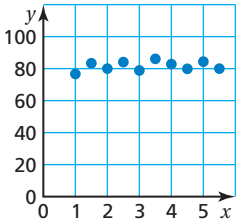
A football field is 40 yards wide and 120 yards long. Find the distance between opposite corners of the football field. Show your work and explain your reasoning. *(8.G.7)*

11. Which scatter plot shows a negative relationship between x and y? *(8.SP.1)*

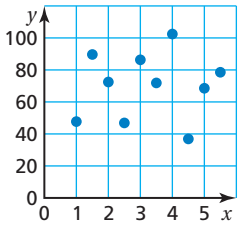
F.



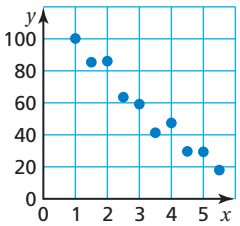
H.



G.



I.



12. What is the solution of the equation? *(A.REI.3)*

$$0.22(x + 6) = 0.2x + 1.8$$

A. $x = 2.4$

B. $x = 15.6$

C. $x = 24$

D. $x = 156$