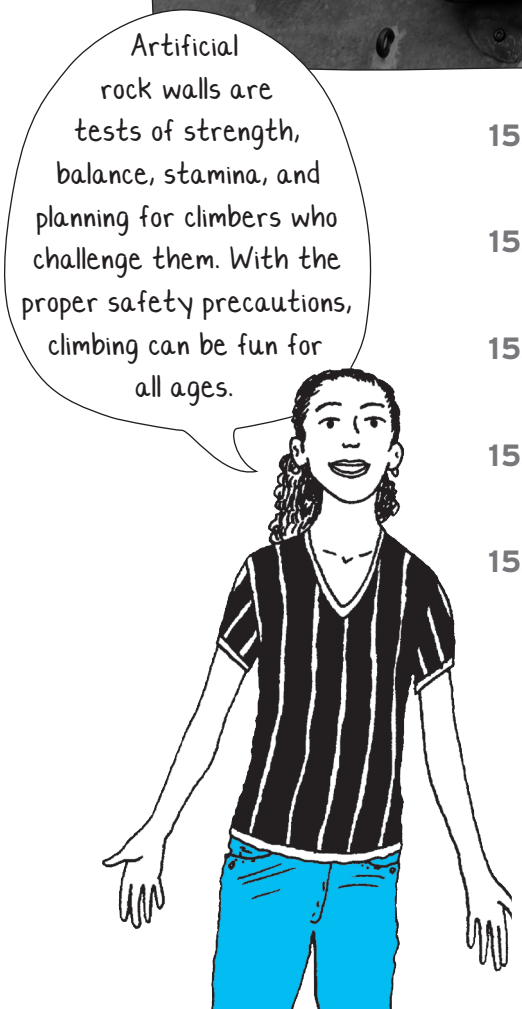
15 COLLECTING AND DISPLAYING DATA



15

15.1 WHY DO WE USE STATISTICS?

Designing Statistical Questions

Learning Goals

In this lesson, you will:

- Learn about statistical questions.
- Recognize statistical questions and anticipate variability in data related to the question.
- Design statistical questions that can be answered by a given data set.
- Differentiate between surveys and experiments.

Key Terms

- data
- statistical question
- population
- sample
- data analysis
- parameter
- statistic
- survey
- experiment

Human interests and interaction seem to go well together. When these interests and interactions meet, clubs are formed. People participate in so many different clubs like video games clubs, dance clubs, computer clubs, collectible toy clubs—and the list can go on and on! Are you part of a club? Are there any clubs in your school? What clubs do you think interest your classmates? Do you think these same clubs interest students in other schools?

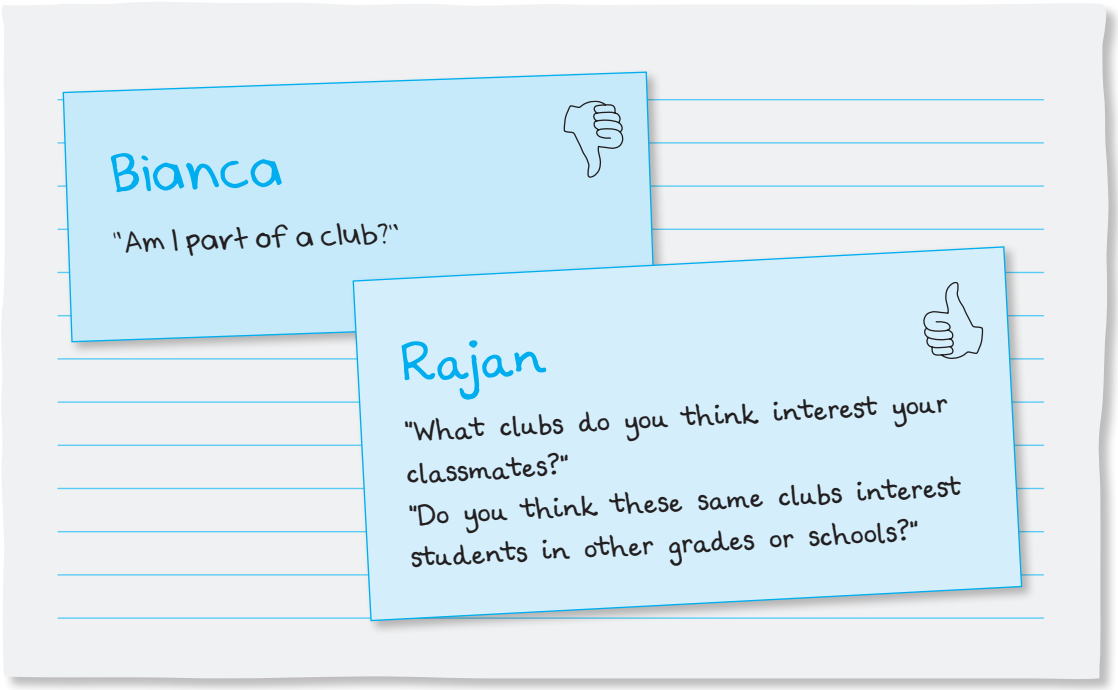
Problem 1 Can We Have the Minutes from the Club Meeting?

15

Data are the facts or numbers that describe the results of an experiment or survey. Data are usually collected from a *statistical question*. A **statistical question** is a question about a *population* or a *sample*.

A **population** is an entire set of items from which data can be collected. A **sample** is a selection from a population.

Analyze each question shown.



Bianca's question is not a statistical question. Bianca knows whether or not she is part of a club. Also this question is not asked of a population—just Bianca.

Rajan's questions are statistical questions because he does not know the answers to these questions. Also these questions can be asked of a population or a sample.

Asking statistical questions and collecting, organizing, and analyzing data to answer those questions is **data analysis**. Through data analysis, you can determine the *parameter* or *statistic* from the data. A **parameter** is the characteristic of a population. A **statistic** is the characteristic of a sample.

1. Analyze the table. Does your school have any of the clubs shown?

Various Clubs	
Art Club	Book Club
Chess Club	Drama Club
Math Club	Science Club
Sports Club	Writers' Club

2. Which club or clubs interest you? Write an "X" in the table by those clubs you are in, or that interest you. You may choose more than one club.

Club	Club(s) That Interest Me	Total Number of Class Votes
Art Club		
Book Club		
Chess Club		
Drama Club		
Math Club		
Science Club		
Sports Club		
Writers' Club		
Other		

3. Record all your classmates' responses in the chart.

4. What clubs appear to interest your classmates?

15

Now, look at the table shown. The data are the results of an online survey that asked sixth graders what clubs in which they were members, or which clubs interested them.

Rank	Club	Number of Online Voters
1	Science Club	47,316
2	Writers' Club	41,615
3	Math Club	39,171
4	Art Club	38,501
5	Sports Club	36,450
6	Drama Club	33,114
7	Chess Club	29,457
8	Book Club	26,003
9	Other	15,879

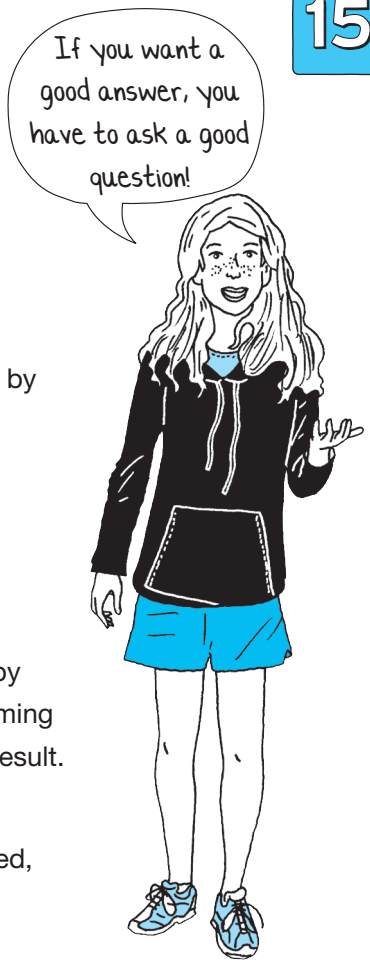
5. What are the similarities between your class list and the online responses?

6. What are the differences between your class list and the online responses?

7. Do you think that these same clubs interest students in other grades or schools?

Problem 2 Surveys and Experiments

When you asked your classmates which clubs they were interested in, you were collecting data. This type of data collection is called a *survey*. A **survey** is one method of collecting information about a certain group of people by asking them a statistical question, or a set of statistical questions.



Let's consider the topic of school lunches. You can ask many statistical questions that can be answered by a survey. For example, what is the most popular lunch served at our school?

1. Write two additional statistical questions that could be answered by conducting a survey about school lunches.

Whenever you have a question and you collect the data to answer it by performing a test for which you decide the conditions, you are performing an **experiment**. Experiments test something to determine a specific result.

Experiments have been a staple of discovering scientific results. Experiments have answered numerous questions scientists have asked, such as:

- How many types of spiders build webs?
- How many fish do dolphins eat per day?

2. Explain why each of the questions is a statistical question?

3. Write two or three statistical questions that you can answer through conducting an experiment.

Talk the Talk

15

Determine if each question shown is a statistical question. If it is not, explain why it is not a statistical question. Then, rewrite it to make it a statistical question.

1. How tall are you to the nearest inch?

2. About how many text messages do sixth graders send each day?

3. What is your school mascot?

4. Do boys or girls spend more time playing video games?

5. How many students bring their lunch in reusable bags?

Be prepared to share your solutions and methods.

15.2 DEALING WITH DATA

Collecting, Displaying, and Analyzing Data

Learning Goals

In this lesson, you will:

- Discuss the different types of data that can be collected, displayed, and analyzed.
- Determine how to select an appropriate graph to display different types of data.
- Organize data into single, double, or stacked bar graphs.
- Analyze data and interpret results from single, double, or stacked bar graphs and circle graphs.

Key Terms

- categorical data
- quantitative data
- discrete data
- continuous data
- bar graph
- frequency
- double bar graph
- key
- stacked bar graph
- circle graph

Once a statistical question has been formulated, data must be collected to answer the question. The type of question that is asked determines the type of data that will be collected. Do you think there are different types of data? Or do you think that all data is the same?

Problem 1 I Want to Learn More About Your Class

Analyze each question shown.

- What is your approximate height in inches?
- What color is your hair?

  Brown _____ Blond _____ Black _____ Red _____ Other_____
- Do you carry a cell phone with you?

  Yes _____ No _____
- About how many text messages do you send in one day?
- About how much time (in minutes) do you spend exercising every weeknight?
- What is your favorite type of television show to watch?

  Comedy _____ Sports _____ Reality _____ Drama _____

  Educational _____ News _____ Music _____ Other _____

1. Complete the survey.

2. What differences do you notice in the types of answers you recorded for the different questions?

When you are collecting, displaying, and analyzing data, the data are one of two types: *categorical data* and *quantitative data*.

Categorical data are data for which each piece of data fits into exactly one of several different groups or categories. Categorical data are also called "qualitative" data. **Quantitative data** are data for which each piece of data can be placed on a numerical scale. Quantitative data are also called "numerical" data.

When quantitative data are a count of how many, the data can be described as *discrete data*. **Discrete data** can only have values that are counting numbers (0, 1, 2, 3,…). When quantitative data are a measurement of something and can have values that are between two counting numbers, the data are called **continuous data**.

3. Determine whether each of the data collected are categorical or quantitative.

   a. height

   b. hair color

   c. carry a cell phone

   d. number of text messages

   e. time you spend exercising

   f. favorite type of TV show

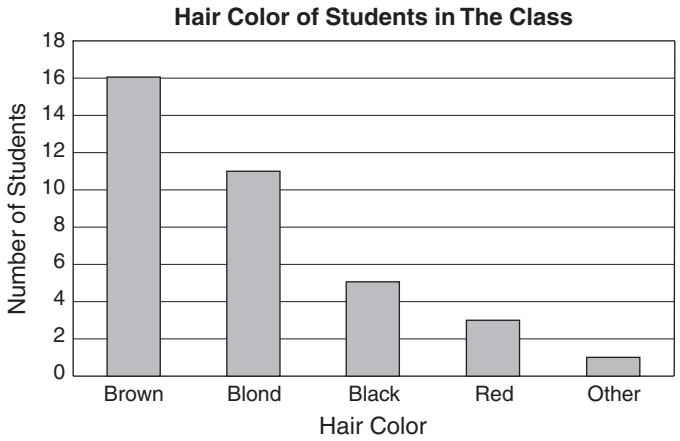
Problem 2 Displaying Data with Bar Graphs

When displaying data, categorical data graphs are different from quantitative data graphs. Categorical data are often displayed in *bar graphs.*

A **bar graph** is a way of displaying categorical data by using either horizontal or vertical bars on a graph. The height or length of each bar indicates the value for that category.

A **frequency** is the number of times an item or number occurs in a data set. Each bar represents a total for the whole category, but still shows how many data pieces make up each group within the entire category.

1. Suppose the bar graph shown displays the results of the student hair color survey question you analyzed.



a. What are the categories for the bar graph?

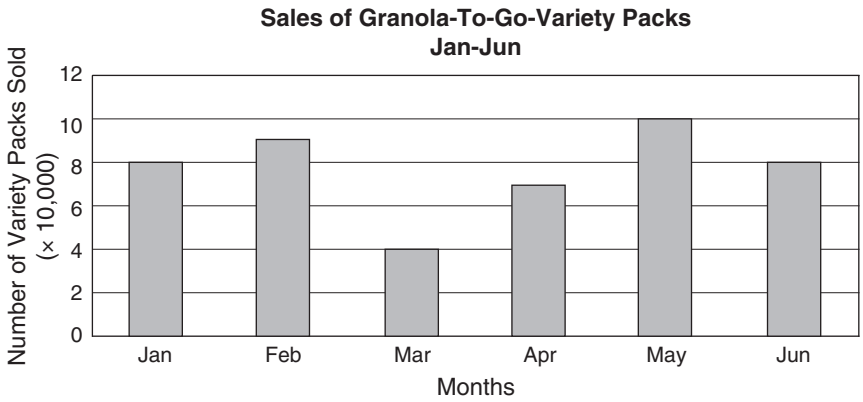
b. What does the vertical axis represent?

c. Name one conclusion you can make about the students' hair color in the class.

d. Would your conclusions change if you changed the order of the categories? Why or why not?

e. For what other data from the class survey would it be appropriate to create a bar graph? Why?

2. The bar graph shows the number of Granola-To-Go Variety Packs sold in the first six months of the year.

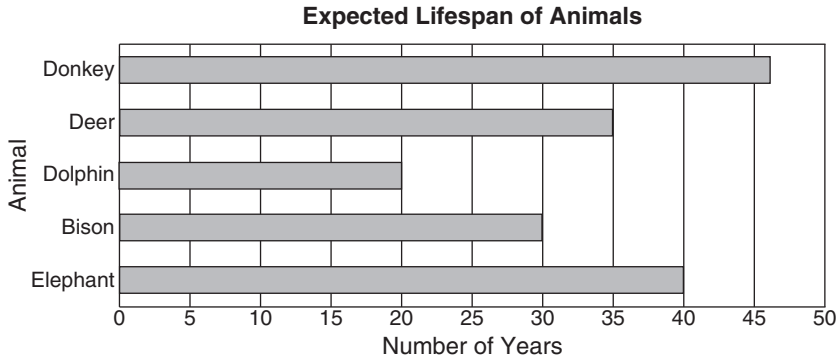


a. What do the bars represent?

b. Which month had the most sales of Granola-To-Go Variety Packs? How many were sold?

c. How many Granola-To-Go Variety Packs were sold in the first half of the year (January through June) according to the data in the bar graph?

3. The graph shown is an example of a bar graph where the bars are horizontal.



a. Which animal has the shortest expected lifespan according to the graph?

b. Which animal has the longest expected lifespan according to the graph?

c. What is the expected lifespan of a typical donkey? How did you calculate your answer from the graph?

d. What interval is used to record the number of years on the bar graph?

4. Compare the "Expected Lifespan of Animals" bar graph with the "Granola-To-Go Variety Pack Sales" bar graph.

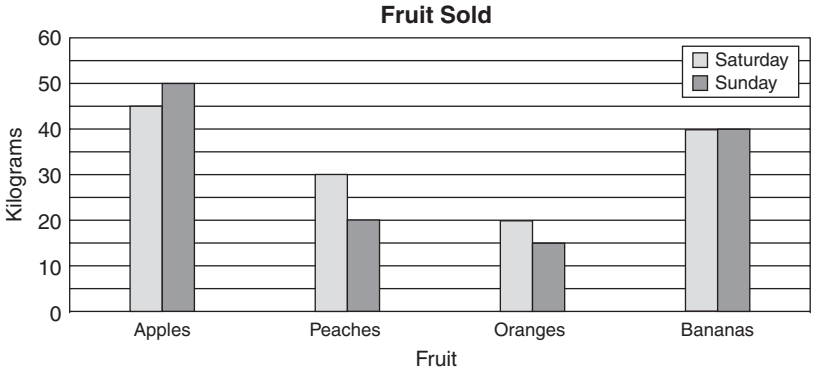
   a. How are the two bar graphs similar?

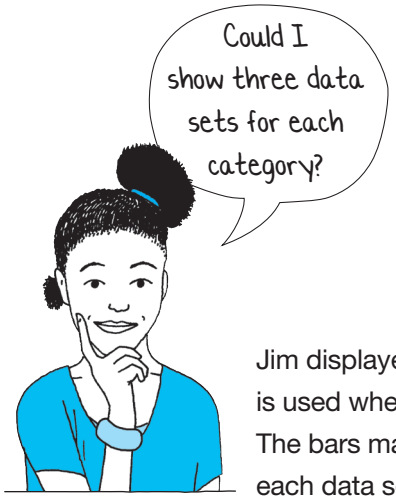
   b. How are the two bar graphs different? State at least two differences in your answer.

Problem 3 Double Your Pleasure with Double Bar Graphs

15

Jim owns a fruit stand at a local farmers market that he only opens on Saturdays and Sundays. Jim recorded the amount of fruit he sold last weekend.





Jim displayed his data in a *double bar graph*. A **double bar graph** is used when each category contains two different data sets. The bars may be vertical or horizontal. A **key** explains how each data set is represented by a color or a pattern in the graph. The two bars representing the same category are side by side. Space is used to separate the categories.

1. Use the double bar graph to answer each question.

a. What does the key tell you?

b. How many kilograms of apples were sold on Sunday?

c. Which fruit showed the greatest difference in sales over the two days?

d. What was the difference between Saturday's sales and Sunday's sales of your answer to part (c)?

e. Which fruit showed the least difference in sales over the two days?

f. What was the difference between Saturday's sales and Sunday's sale of your answer to part (e)?

2. Analyze the table that shows the number of births of boys and girls in the United States between 2000 and 2006. You will complete a double bar graph using the information shown in the table.

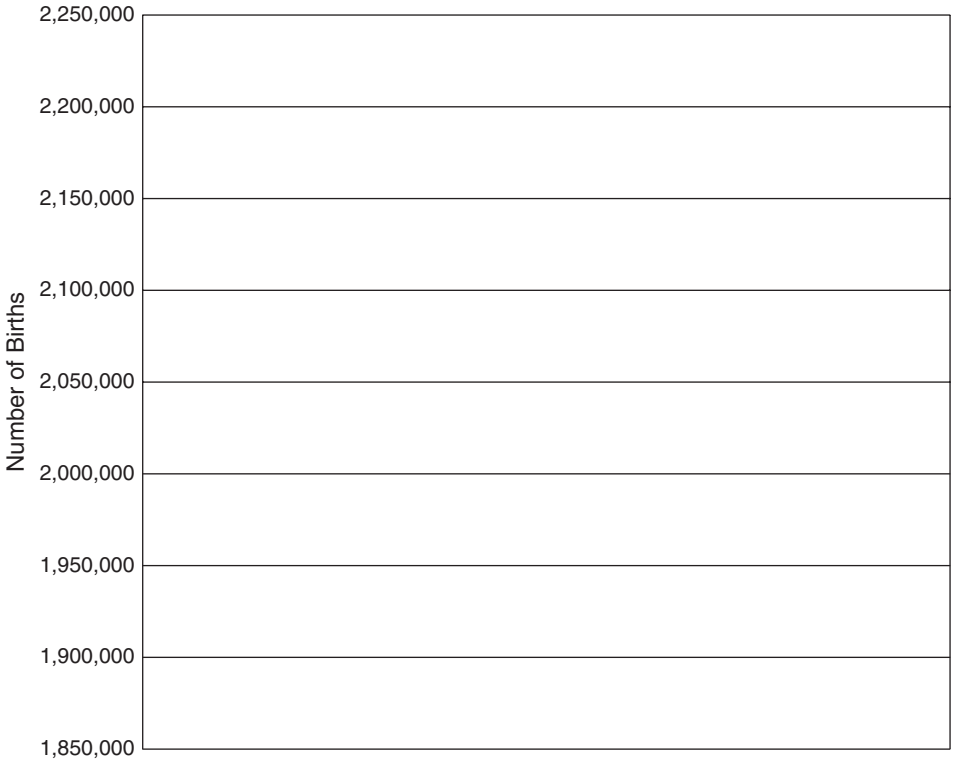
U.S. Male and Female Births 2000-2006		
Year	**Male**	**Female**
2000	2,076,969	1,981,845
2001	2,057,922	1,968,011
2002	2,057,979	1,963,747
2003	2,093,535	1,996,415
2004	2,150,228	1,961,794
2005	2,118,982	2,019,367
2006	2,184,237	2,081,318

a. What will the key show?

b. What will you name the double bar graph?

c. What will the categories represent on the double bar graph?

d. Complete the double bar graph. Include a title and label the horizontal axis.

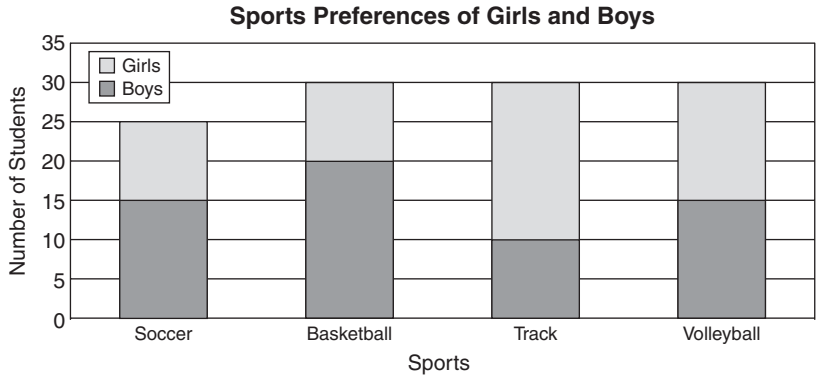


3. When is a double bar graph a better graph choice than a single bar graph?

Problem 4 It *Is* All It's Stacked up to Be!

15

Boys and girls in an after-school program were asked which sport they preferred. Their answers are recorded in the graph shown.



The "Sports Preferences of Girls and Boys" display is a *stacked bar graph*. A **stacked bar graph** is a graph that stacks the frequencies of two different groups for a given category on top of one another so that you can compare the parts to the whole.

1. Use the stacked bar graph to answer each question.

a. What does the key indicate?

b. How many total students prefer basketball over the other choices?

c. How many girls and how many boys prefer basketball?

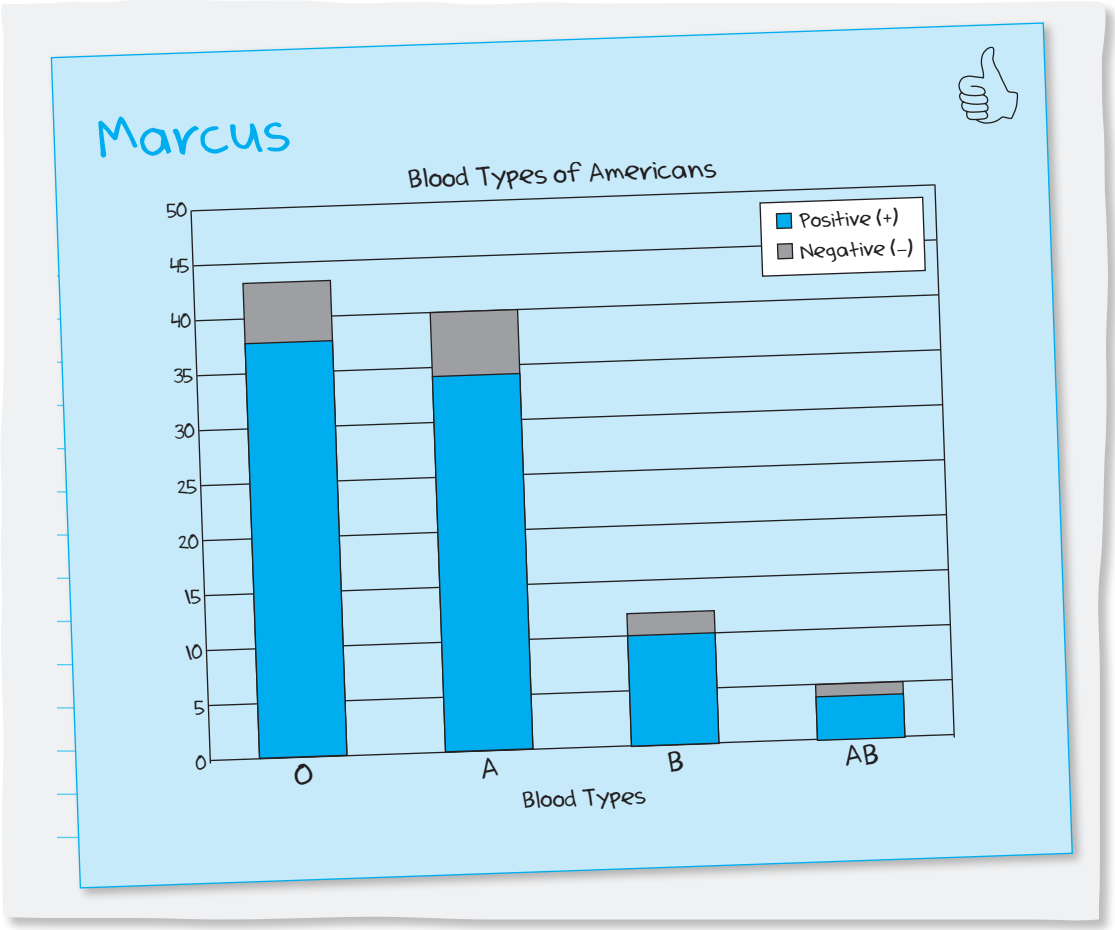
d. Which sport is preferred equally by the boys and the girls in the after-school program? How can you tell?

15

Do you know what blood type you have? Do you know anyone in your class who has your same blood type? There are eight different human blood types and four blood groups (O, A, B, and AB). These groups indicate what types of blood a person can receive if that person needs a blood transfusion. Some blood types are more common than others.

Blood Type	Percent of Americans with Blood Type
O+	37
O−	6
A+	34
A−	6
B+	10
B−	2
AB+	4
AB−	1

Marcus created the stacked bar graph shown from the data in the table.



2. Use Marcus's stacked bar graph to answer each question.

a. What is the most common blood type found in Americans?

b. What is the least common blood group, when its positive and negative carriers are combined?

c. What percent of Americans have Group B blood with either a positive or negative factor?

3. How is a stacked bar graph the same as other bar graphs?

4. How is a stacked bar graph different from other bar graphs?

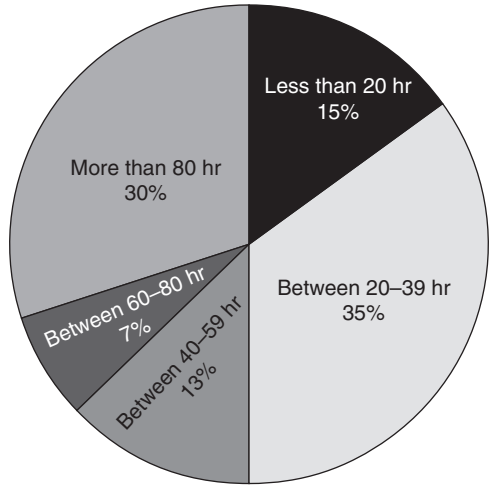
5. When would you use a stacked bar graph?

Problem 5 How Much is One Piece of the Pie?

Another type of graph used to display categorical data is a *circle* or *pie graph*. A **circle graph** is a graph that shows how parts of the whole relate to the whole, and how parts of the whole relate to other parts. The area of each sector, or "wedge," indicates the percentage of the part in relation to the whole.

In 2005, the *Volunteer Spotlights Magazine* surveyed over 20,000 students in Grades 6 through 12. One question on the survey asked: How much time do you spend volunteering during the year? The results are shown in the circle graph.

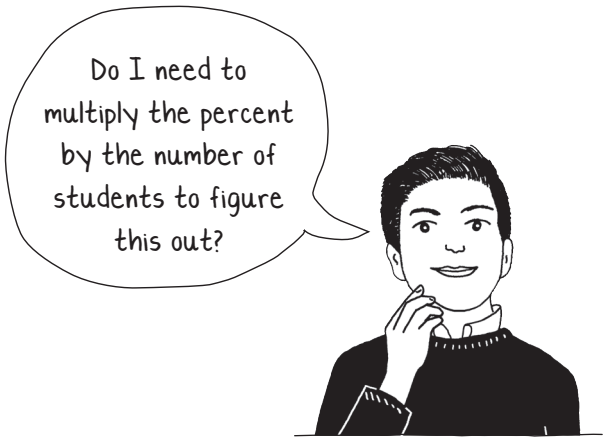
Hours Spent Volunteering in a Year



1. What percent of the students volunteer more than 80 hours a year?

2. How many hours per year do half as many students volunteer as those who volunteer over 80 hours per year?

3. If 20,000 students responded to the survey, how many students volunteer between 20 and 39 hours per year?



4. Another question on the survey asked students, "What kind of volunteering do you do?" The results are:

   Volunteer through school: 64%
   Volunteer through religious affiliation: 50%
   Volunteer through community groups: 47%
   Other: 45%
   I don't volunteer: 4%

   Would a circle graph be an appropriate graph for the data? Why or why not?

Talk the Talk

Categorical data are a data set for which each piece of data fits into exactly one of several different groups or categories. Categorical data are also called "qualitative" data. Some examples of categorical data are gender, favorite color, and birth month.

1. Describe how a bar graph displays categorical data.

2. Describe how a circle graph displays categorical data.

Be prepared to share your solutions and methods.

15.3 AT THE OLYMPICS

Line Plots and Stem-and-Leaf Plots

Learning Goals

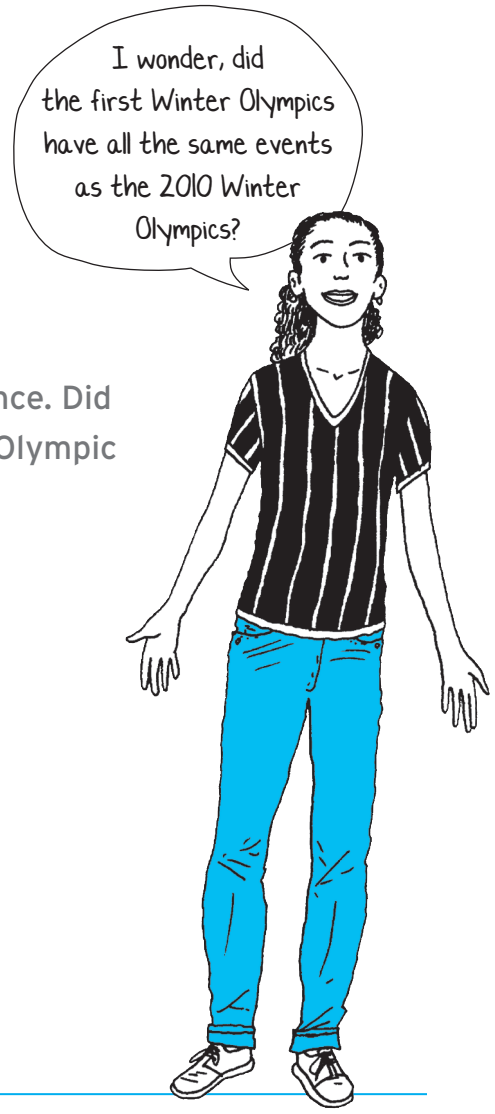
In this lesson, you will:

- Organize data and interpret dot plots.
- Organize data and interpret stem-and-leaf plots.

Key Terms

- dot plot
- distribution
- symmetric
- skewed right
- skewed left
- clusters
- gaps
- stem-and-leaf plot
- side-by-side stem-and-leaf plot

What do the years 2014, 2018, 2022, and 2026 all have in common? If you said that these years mark when the Winter Olympics are scheduled to take place, you would be correct! The Olympics first took place in ancient Greece. Many experts believe the first games occurred around 775 B.C. However, the modern day Olympics first took place in 1896, with the Winter Olympics starting roughly 30 years later in France. Did you watch the 2010 Winter Olympics? Were there any Olympic events that interested you?



Problem 1 Victory at the Olympics—Pure Gold!

The 2010 Winter Olympic Games were held in Vancouver, British Columbia, Canada. The Games featured 85 medal events over 15 sports. The table shown lists the number of gold medals, and total medals won by various countries.

2010 Winter Olympics Top 20 Countries with Medal Wins					
Country	**Gold Medals**	**Total Medals**	**Country**	**Gold Medals**	**Total Medals**
United States	9	37	Switzerland	6	9
Germany	10	30	Netherlands	4	8
Canada	14	26	Czech Republic	2	6
Norway	9	23	Poland	1	6
Austria	4	16	Italy	1	5
Russian Federation	3	15	Japan	0	5
South Korea	6	14	Finland	0	5
China	5	11	Australia	2	3
Sweden	5	11	Belarus	1	3
France	2	11	Slovakia	1	3

1. Analyze the data shown in the table.

a. What conclusions can you make about the numbers of total medals won at the 2010 Winter Olympics?

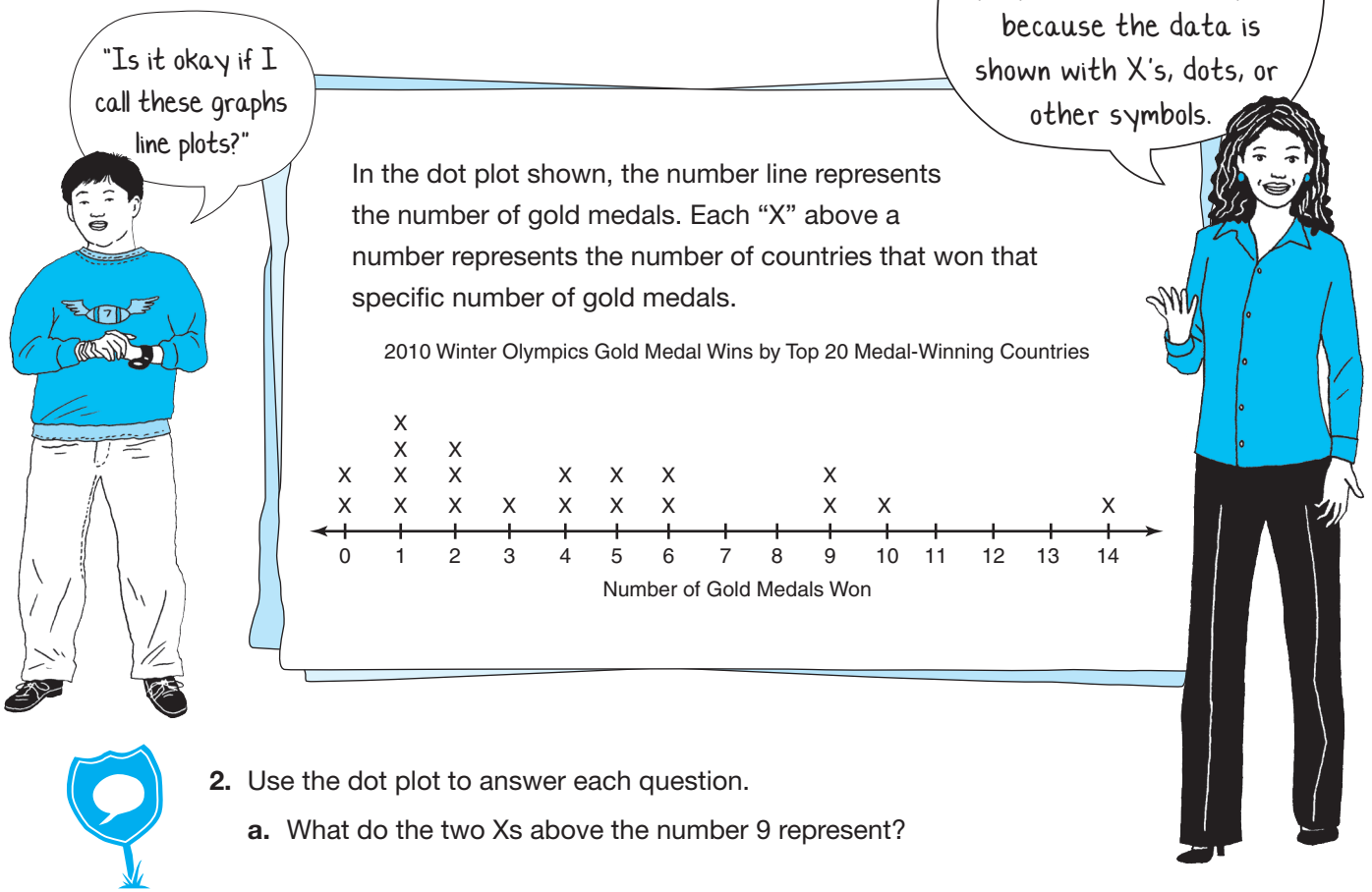
b. Are the data in the table categorical or quantitative? Explain your reasoning.

c. Are the data in the table discrete or continuous? Explain your reasoning.

15

One way to describe a set of quantitative data is by drawing a graphical display of the data.

A **dot plot** (sometimes called a line plot) is a graph that shows how the discrete data is graphed using a number line. Dot plots help organize and display a small number of data points.



2. Use the dot plot to answer each question.

a. What do the two Xs above the number 9 represent?

b. What do the two Xs above the 0 represent?

c. Why are there no Xs above the 8?

3. Let's create a dot plot to display the total number of medals won as listed in the 2010 Winter Olympics data table.

a. What will you name your dot plot?

b. What numbers will begin and end your number line? How did you select the numbers?

c. What interval will you use on your number line? How did you decide on the interval?

d. Create a dot plot displaying the data from the 2010 Winter Olympics data table.

Problem 2 The Shape of It All

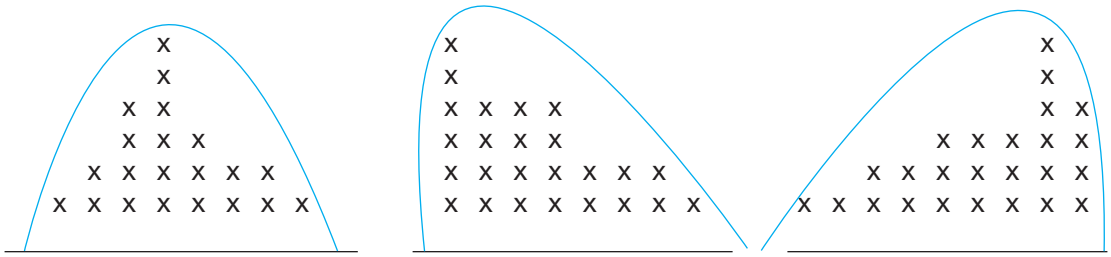
When you analyze a graphical display of data, you can look at several characteristics of the graph to draw conclusions. By doing this, you can ask yourself some of the questions shown, such as:

1. What is the overall shape of the graph? Does it have any interesting patterns?
2. Where is the approximate middle, or center, of the graph?
3. What does the graph tell me about how spread out the data values are?

The overall shape of a graph is called the *distribution* of data. A **distribution** is the way in which the data are distributed.

The shape of the distribution can reveal a lot of information about data. There are many different distributions, but the most common are *symmetric*, *skewed right*, or *skewed left*.

Typical Distributions of Graphical Displays of Data



symmetric

- The left and right halves of the graph are mirror images of each other.
- There is often a "peak" in the middle of the graph indicating there are many data values in the center of the graph.

skewed right

- The peak of the data is to the left side of the graph.
- There are only a few data points to the right side of the graph.

skewed left

- The peak of the data is to the right side of the graph.
- There are only a few data points to the left side of the graph.

1. Miko says that the dot plot shown in Problem 1 for the number of gold medals won is skewed right. Do you agree with her statement? Explain your reasoning.

2. Examine the dot plot you created for the total number of medals won by the top 20 medal-winning countries. What is the distribution of the dot plot? Explain what this means in terms of the total number of medals won.

When analyzing a graphical display of data, you can also look for any interesting patterns. Some of these patterns include:

- **clusters**—areas of the graph where data are grouped close together
- **gaps**—areas of the graph where there are no data

3. Examine the dot plot you analyzed for the number of gold medals won by the top 20 medal-winning countries. Identify any clusters or gaps. Explain what this means in terms of the number of gold medals won.

4. Examine the dot plot you created for the total number of medals won by the top 20 medal-winning countries. Identify any clusters or gaps. Explain what this means in terms of the total number of medals won.

Problem 3 Let the Games Begin!

At the 2010 Winter Olympics, 82 countries competed in the events, but only 26 won medals. By contrast, the 2008 Summer Olympics in Beijing, China, had 205 countries compete in the events, but only 43 won at least 5 medals in various competitions.

Total Number of Medals Won by Various Countries, 2008 Summer Olympics			
Country	**Medals Won**	**Country**	**Medals Won**
United States	110	Norway	10
China	100	New Zealand	9
Russian Federation	72	Romania	8
Great Britain	47	Turkey	8
Australia	46	Ethiopia	7
Germany	41	Denmark	7
France	40	Azerbaijan	7
South Korea	31	Czech Republic	6
Italy	28	Slovakia	6
Ukraine	27	Georgia	6
Japan	25	North Korea	6
Cuba	24	Argentina	6
Belarus	19	Switzerland	6
Spain	18	Uzbekistan	6
Canada	18	Armenia	6
Netherlands	16	Slovenia	5
Brazil	15	Bulgaria	5
Kenya	14	Indonesia	5
Kazakhstan	13	Sweden	5
Jamaica	11	Croatia	5
Poland	10	Lithuania	5
Hungary	10		

1. What conclusion can you make between the number of medals won at the 2008 Summer Olympics, and the number of medals won at the 2010 Winter Olympics?

2. Do you think using a dot plot would be a good way to organize and analyze the data in the Summer Olympics data table? Explain your reasoning.

Because there are many data points in the 2008 Summer Olympics data table, a dot plot would probably not be a good graphic display of the data. A graphic display that can easily display all the data points is needed. A **stem-and-leaf plot** is a graphical method used to represent ordered numerical data. Once the data is ordered, the stem and leaves are determined. Typically, the stem is all the digits in a number except the right-most digit, which is the leaf.

A stem-and-leaf plot displaying the number of medals won in the 2008 Summer Olympics is shown.

Total Medals Won by Various Countries
2008 Summer Olympics

Stem	Leaves
0	5 5 5 5 5 5 6 6 6 6 6 6 6 6 7 7 7 8 8 9
1	0 0 0 1 3 4 5 6 8 8 9
2	4 5 7 8
3	1
4	0 1 6 7
5	
6	
7	2
8	
9	
10	0
11	0

Key: 3|1 = 31 medals won.

3. Use the stem-and-leaf plot to answer each question.

a. Describe what you notice about the stem-and-leaf plot?

b. What does 4|0 mean in the stem-and-leaf plot for the 2008 Summer Olympics?

c. What does 0|5 mean?

d. How many stems are in the stem-and-leaf plot?

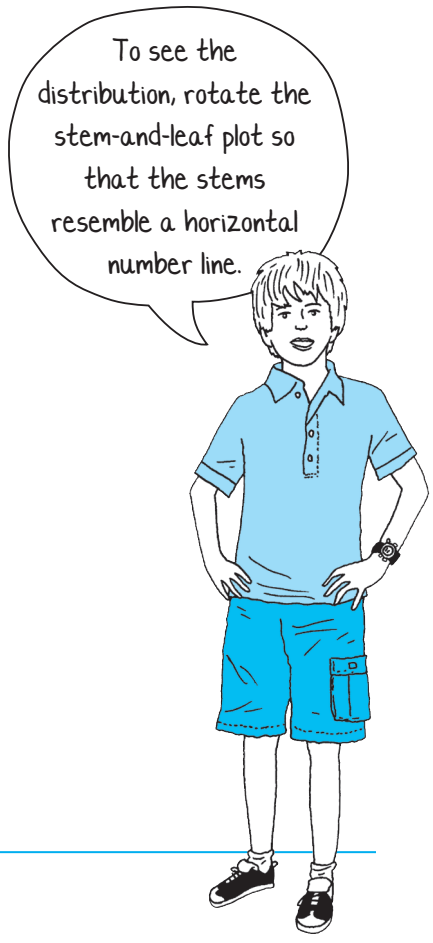
e. How many leaves are in the stem-and-leaf plot? Why are there that many leaves?

f. Why do you think a stem has more than one leaf?

g. Why do some stems have no leaves?

h. Carlos claims that he should write 0s as leaves after the stems 5, 6, 8, and 9 to show that there are no countries having medals numbering the 50s, 60s, 80s and 90s. Is Carlos correct? Explain your reasoning.

i. Describe the distribution and any interesting patterns you notice in the stem-and-leaf plot. Interpret your findings in terms of the number of medals won in the 2008 Olympics.



Problem 4 Interesting Presidential Facts

1. The table shown lists the ages when 38 former Presidents of the United States were first inaugurated president, and the ages when they died.

President	Age at Inauguration	Age at Death	President	Age at Inauguration	Age at Death
Washington	57	67	Garfield	49	49
Adams	61	90	Arthur	51	57
Jefferson	57	83	Cleveland	42	71
Madison	57	85	Harrison	55	67
Monroe	58	73	McKinley	54	58
Adams	57	80	Roosevelt	42	60
Jackson	61	78	Taft	51	72
Van Buren	54	79	Wilson	56	67
Harrison	68	68	Harding	55	57
Tyler	51	71	Coolidge	51	60
Polk	49	53	Hoover	54	90
Taylor	64	65	Roosevelt	51	63
Fillmore	50	74	Truman	60	88
Pierce	48	64	Eisenhower	62	78
Buchanan	65	77	Kennedy	43	46
Lincoln	52	56	Johnson	55	64
Johnson	56	66	Nixon	56	81
Grant	46	63	Ford	61	93
Hayes	54	70	Reagan	69	93

a. What do you notice about the presidents and ages listed in the table?

b. Sonya says, "To organize the data, I will use a stem-and-leaf plot because there are a lot of data points." James says, "I don't think that it is possible because there are two groups of data points." Who's correct? Explain your reasoning.

2. Let's display the data of the age at which each president died in a stem-and-leaf plot.

a. What will you choose for your stems? Why did you choose those numbers?

b. How many leaves will you have in your stem-and-leaf plot? Explain your reasoning.

c. How many stems will you have in your stem-and-leaf plot?

d. Create a key for your stem-and-leaf plot. Why is this needed?

e. Complete the stem-and-leaf plot to show the age at which each president died.

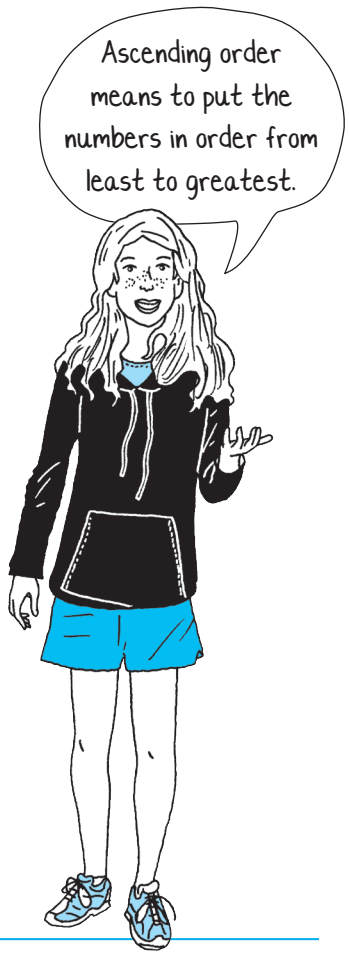
The Ages Former U.S. Presidents Died	
4	
5	
6	
7	
8	
9	

Key:

f. If the values in your leaf portion from part (e) are not in ascending order, rewrite the leaves in ascending order.

The Ages Former U.S. Presidents Died	
4	
5	
6	
7	
8	
9	

Key:



g. Why do you think it is important to put the values in the leaf portion in ascending order?

15

h. Use what you know about distributions and patterns of graphical displays to write a description of the observations you can make from the information you recorded.

The table you used to complete your stem-and-leaf plot also had information about each president's age at inauguration. A **side-by-side stem-and-leaf plot** allows a comparison of the two data sets in two columns. In a side-by-side stem-and-leaf plot, two sets of leaves share the same stem. The leaves are to the left *and* the right of the stems.

3. Look at the side-by-side stem-and-leaf plot shown.

First 38 U.S. Presidents' Ages at Inauguration Compared to Ages at Death

Ages at Inauguration		Ages at Death
9 9 8 6 3 2 2	4	6 9
8 7 7 7 7 6 6 6 5 5 5 4 4 4 4 2 1 1 1 1 1 0	5	3 6 7 7 8
9 8 5 4 2 1 1 1 0	6	0 0 3 3 4 4 5 6 7 7 7 8
	7	0 1 1 2 3 4 7 8 8 9
	8	0 1 3 5 8
	9	0 0 3 3

Key: 7|1 = 71 years of age

a. What do the leaves in the left-hand column represent?

b. What do you notice about the order of the leaves on the right and left of the stems?

c. What observations can you make from the side-by-side stem-and-leaf plot? Describe your observations in your own words. Use what you know about the distributions and patterns of graphical displays for the side-by-side stem-and-leaf plot.

d. What questions does this side-by-side stem-and-leaf plot bring up for you?

Problem 5 Take Your Mark, Get Set . . .

15

Routinely, track team runners have coaches record their practice and competitive times in races. For the John Glenn Middle School track team, the following times were recorded for one practice run and one competitive run of the 100 meter dash.

Practice Runs: 13.1, 14.2, 13.3, 13.3, 13.0, 14.6, 12.9, 13.9, 12.8, 15.1, 15.0, 15.2, 14.3, 14.3, 16.1, 16.4

Competitive Runs: 12.4, 14.0, 12.6, 13.8, 14.2, 12.9, 14.2, 12.9, 14.3, 13.6, 12.9, 13.7, 13.4, 12.6, 13.3, 12.7

1. Let's create a side-by-side stem-and-leaf plot to organize the practice and competition runs.

a. What stems will you use in your side-by-side stem-and-leaf plots?

b. What key will you use? What will the key mean?

c. Create the side-by-side stem-and-leaf plot.

d. What does this side-by-side stem-and-leaf plot tell you about the practice and competitive runs? Use what you know about the distribution and patterns of graphical displays for your side-by-side stem-and-leaf plot.

2. Why do you think the times were scattered in the practice run times, but more clustered in the competitive run times?

Talk the Talk

1. Summarize the uses for a stem-and-leaf plot.

2. Summarize the uses for a side-by-side stem-and-leaf plot.

Be prepared to share your solutions and methods.

15

15.4 BUILDING UP
Using Histograms

Learning Goals

In this lesson, you will:

- Organize data into a histogram.
- Analyze data presented in a histogram.

Key Terms

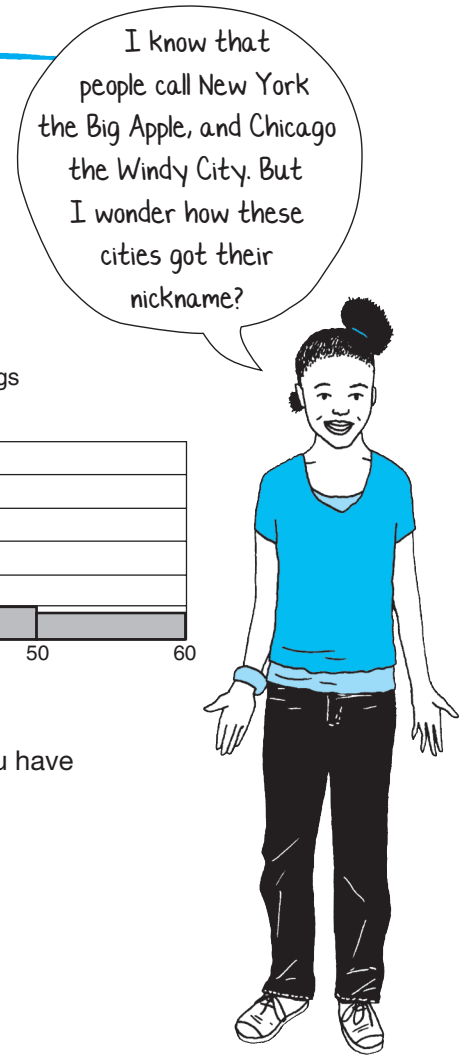
- histogram
- frequency table

Are there any skyscrapers in your city? How about in your state? Chances are that in the largest city in your state, there is a cluster of tall buildings. In the early to mid-twentieth century, cities began to build up instead of spreading out. Skyscrapers are home for various businesses, television and radio station headquarters, and even people! In fact, the book you're reading right now was created in a skyscraper in Pittsburgh, Pennsylvania. What is the largest city in your state? Does it have a cluster of tall buildings?

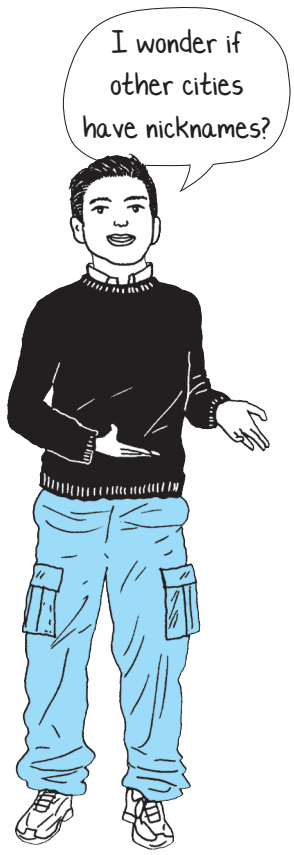
15

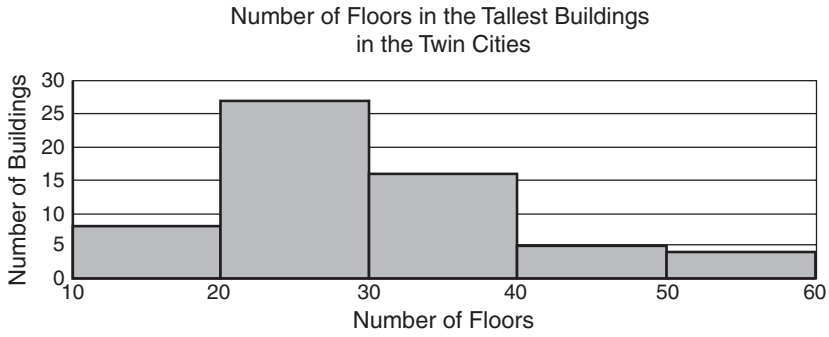
Problem 1 Onward and Upward!

Minneapolis and St. Paul are known as the Twin Cities because they are close to each other in Minnesota. Both cities are home to flourishing downtowns with tall buildings.



1. Look at the graph shown.





a. How is this graph different from the bar graphs you have used previously?

The graph shown is a *histogram*. A **histogram** is a graphical way to display quantitative or numerical data using vertical bars. The width of a bar in a histogram represents an interval of data and is often referred to as a bin. The height of the bar indicates the frequency, or the number of data values included in any given bin. Bins are represented by intervals of data instead of showing individual data values.

b. What information does the histogram display? Describe the data represented in the histogram shown. Look at the title and the labels on the axes.

c. Are the data represented in the histogram discrete or continuous? Explain your reasoning.

2. Let's think about how the bars are displayed in the histogram.

a. How many bins are shown?

b. Are all the bins the same size?

c. What does the height of each bar represent?

The first vertical bar on the histogram represents eight buildings that have at least 10 floors but less than 20 floors.

3. Describe the range of floors included in each of the remaining bins shown on the horizontal axis.

- 2nd bin: interval 20 – 30
- 3rd bin: interval 30 – 40
- 4th bin: interval 40 – 50
- 5th bin: interval 50 – 60

4. If a new building was constructed in one of the Twin Cities that had 20 floors, which bin would change? How would it change?

15

5. Bella says, "There are 5 buildings represented in the histogram since there are 5 bars." Do you agree or disagree with Bella's statement? If you do not agree with Bella, determine how many buildings are represented in the histogram.

6. Can you determine how many buildings have 31 floors? Explain your reasoning.

7. Is it possible to determine the number of buildings that have more than 35 floors from the histogram? Why or why not?

8. Is it possible to determine the range of the data set from the histogram? Why or why not?

To create a histogram, data is usually organized into a *frequency table*. A **frequency table** is a table used to organize data according to how many times a data value occurs. The frequency table for the number of floors in the Twin Cities' 60 tallest buildings is shown.

Floor Intervals	Frequency (f)
10–20	8
20–30	27
30–40	16
40–50	5
50–60	4

9. Compare the histogram to the frequency table.

Problem 2 Building Up, Not Spreading Out

Did you know that New York City has over 5800 tall buildings and is home to the Empire State Building, which is 381 meters tall? In fact, it is the third tallest building in the United States. Not to be outdone, Chicago is home to the Willis Tower, also known as the Sears Tower, which is the tallest building in the United States. It stands an impressive 442 meters tall. So how do these big cities stack up to each other? Are there any similarities or differences in the number of floors each city's 20 tallest buildings have? Look at the table listing each city's 20 tallest buildings.

15

New York City's 20 Tallest Buildings	
Name of Building	**Number of Floors**
12521 Avenue of the Americas	54
20 Exchange Place	57
60 Wall Street	55
American International Building	66
AXA Center	54
Carnegie Hall Tower	60
Chrysler Building	77
Citigroup Center	59
CitySpire Center	75
Empire State Building	102
GE Building (Rockefeller Center)	69
MetLife Building	59
One Chase Manhattan Plaza	60
One Penn Plaza	57
The Beekman	76
The Trump Building	70
Time Warner Center North	55
Time Warner Center South	55
Trump World Tower	72
Woolworth Building	57

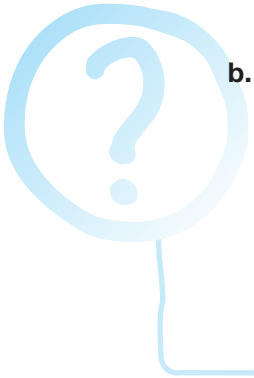
Chicago's 20 Tallest Buildings	
Name of Building	**Number of Floors**
111 South Wacker	51
300 North LaSalle	60
311 South Wacker	65
330 North Wabash	52
900 North Michigan	66
Aon Center	83
Aqua (Lakeshore East)	82
AT&T Corporate Center	60
Blue Cross-Blue Shield Tower	54
Chase Tower	60
Chicago Title 7 Trust Building	50
John Hancock Center	100
Olympia Centre	63
One Museum Park (Central Station)	62
Park Tower	67
Three First National Plaza	57
Trump International Hotel & Tower	96
Two Prudential Plaza	64
Water Tower Place	74
Willis Tower	108

1. How is the table organized? Is this organization helpful to you in answering your question?

To organize the data, you will make two frequency tables. You will then make two histograms to compare the different number of floors for each city's 20 tallest buildings.

2. Let's make a frequency table for New York's tallest buildings.

   a. What is the least number of floors in the table for New York's 20 tallest buildings?



   b. Riki suggests having each data bin represent the intervals: 50–60, 60–70, 70–80, 80–90, 90–100, 100–110. Do you think these are appropriate data bins to use? Why or why not?

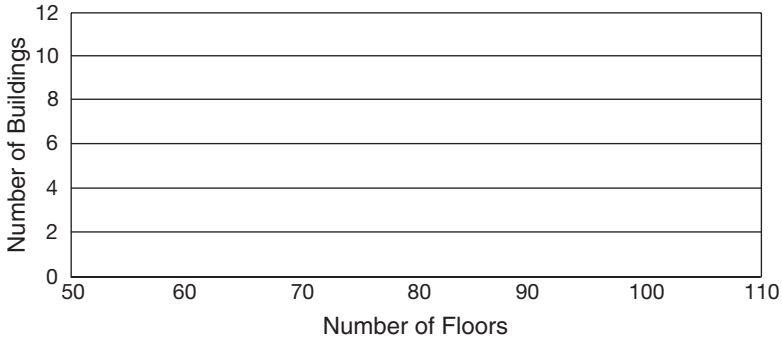
   c. Complete the frequency table for the number of floors in New York's 20 tallest buildings. Make sure that you name your frequency table.

Number of Floors	Frequency (f)

3. Now that you have created a frequency table, let's create a histogram.

a. What will you name your histogram?

b. Complete the histogram.



c. Use what you know about the distributions and patterns of a graphical display to describe what the histogram says about the number of floors in New York's 20 tallest buildings.

d. Andrew claims that 60–70 of the buildings have 4 floors. Do you agree or disagree with Andrew's claim? Explain your reasoning.

4. Follow the same procedure you used to create a histogram for New York's 20 tallest buildings to create a histogram for Chicago's 20 tallest buildings.

a. Create a frequency table.

Number of Floors	Frequency (f)

b. Create a histogram using the Chicago building data in your frequency table.

c. Use what you know about the distributions and patterns of graphical displays to describe what the histogram says about the number of floors in Chicago's 20 tallest buildings.

5. What is similar about the histograms you created for New York and Chicago's 20 tallest buildings?

6. What are the differences between the two histograms?

7. Why is a histogram an appropriate graph to use in comparing the number of floors in a city's tallest buildings?

8. What are the limitations of using a histogram to display the number of floors in the city's tallest buildings?

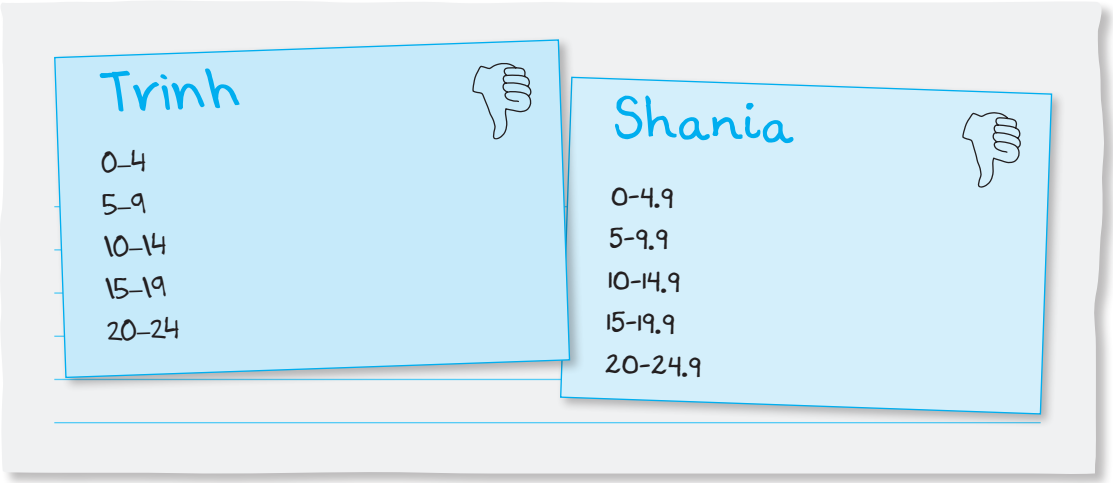
Problem 3 Continuing with Continuous Data

Trinh and Shania want to know how much time their classmates spend playing video games on weekends (Friday night, Saturday, and Sunday). So, they asked each classmate to keep track of how much time they spent playing video games last weekend. The results of their class survey are shown.

Number of Hours My Classmates Spent Playing Video Games on Weekends				
9.95	10	2.25	16.75	0
19.5	22.5	7.5	15	12.75
5.5	11	10	20.75	17.5
23	21.9	24	23.75	18
20	15	22.9	18.8	20.5

1. Is the data in the table continuous or discrete?

Shania and Trinh decide to make a histogram for the data set. The intervals they each want to use for the histogram are shown.



2. Explain why both Trinh's and Shania's intervals are incorrect. Use a data point from the table as an example why each person's intervals are incorrect.

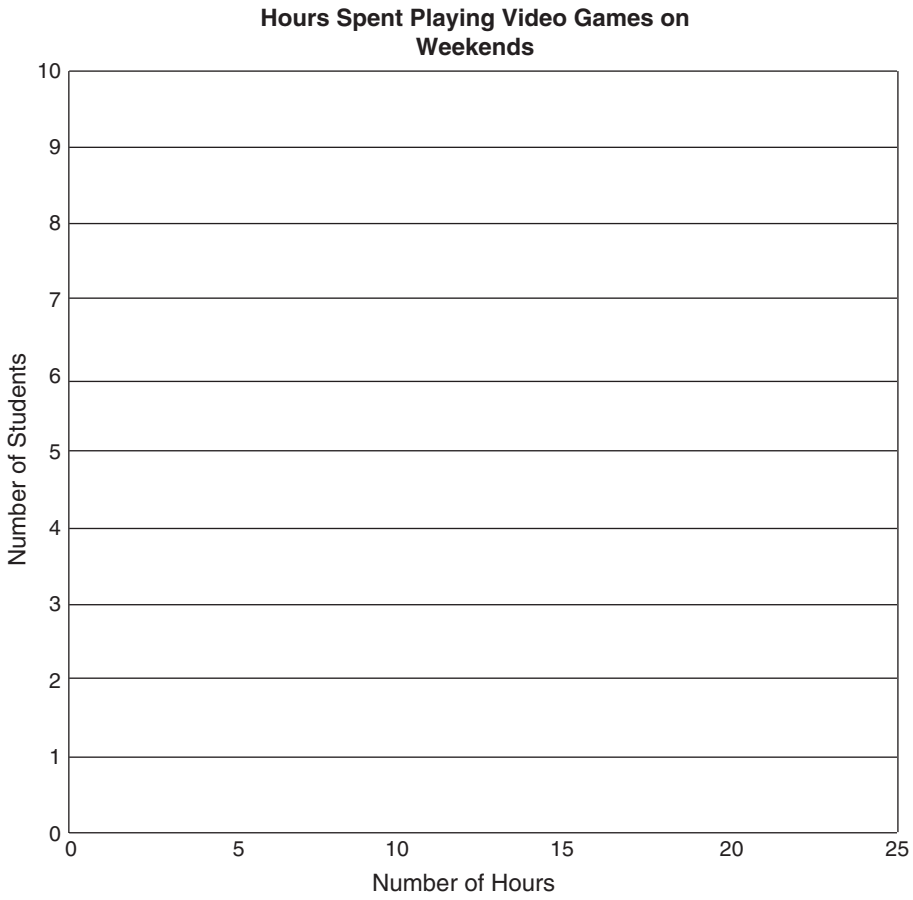
3. Create a frequency table for the number of hours the classmates spend playing video games on weekends.

Number of hours	Data Values from the Table	Frequency
0–5		
5–10		
10–15		
15–20		
20–25		

If a data value lies on one of the bounds, it should be included in the interval to the right of that bound. For example, the interval between 5–10 includes the value of 5, while the interval 0–5 would not include the value of 5.

4. Complete the frequency table with the data values and the frequencies for each interval.

a. Use the frequency table you completed to complete the histogram.



b. What conclusions can you make about the number of hours Trinh and Shania's classmates spend playing video games on weekends? Use what you know about distributions and patterns of graphical displays.

15

5. The city of Hometown recorded the number of inches of rain that fell each month for a two-year period. The results are displayed in the stem-and-leaf plot shown:

Monthly Rainfall in Hometown over a Two-Year Period

0	2 4 5
1	0 4 5 8 9
2	2 2 3 4 7 7 8 9
3	1 2 4 4 6 8
4	0 1

Key: 2 | 4 = 2.4 inches

a. Are the data represented in the stem-and-leaf plot discrete or continuous? Explain your reasoning.

b. What is the distribution of the stem-and-leaf plot, and what does that tell you about the data?

c. Create a histogram of the rainfall data.

d. What conclusions can you draw about the rainfall in Hometown?

e. How are the stem-and-leaf plot and the histogram similar in what they tell you about the data? How are they different?

Be prepared to share your solutions and methods.

15

15.5 ANALYZE THIS!

Designing and Implementing an Experiment

Learning Goal

In this lesson, you will:

- Analyze data recorded during an experiment and answer questions that were posed at the beginning of the experiment.

Rock climbing has been a sport for quite some time, but climbing walls or rock climbing simulations are a very young sport.

Some experts believe that the idea of climbing walls was created in the United Kingdom in the mid 1960s. A physical education teacher by the name of Don Robinson is credited with creating the first climbing wall. Since then, climbing walls have exploded in popularity and have been a good source of exercise. Have you ever tried rock climbing or indoor climbing? What muscles do you use when rock climbing?

Problem 1 Conducting an Experiment to Answer a Question

15

Ms. Nicholson, the physical education teacher, has just had a new rock-climbing simulation wall installed in the gymnasium. Wearing a safety harness attached to a rope, students can climb the rock wall, ring a bell at the top, and then make their way back to the ground by using the rope. You want to conduct an experiment to determine whether the students in Ms. Nicholson sixth grade physical education class are better rock climbers than the students in her seventh grade physical education class.

"Which class, between Ms. Nicholson's sixth-grade physical education class and her seventh-grade physical education class, has the fastest average rock-climbing time if each student is given one attempt?"

1. Is the question shown a statistical question? Why or why not?

You conduct the rock-climbing experiment and record these times displayed in the table.

Sixth Grade Completion Times (seconds)	54, 89, 70, 48, 55, 52, 58, 48, 59, 72, 68, 61, 67, 74, 44, 42, 46
Seventh Grade Completion Times (seconds)	63, 87, 61, 76, 73, 48, 69, 81, 66, 51, 57, 51, 65, 75, 75, 53, 42, 68, 82

2. Complete the side-by-side stem-and-leaf plot of the data. Be sure to include a key.

6th-Grade Completion Times (seconds)		7th-Grade Completion Times (seconds)

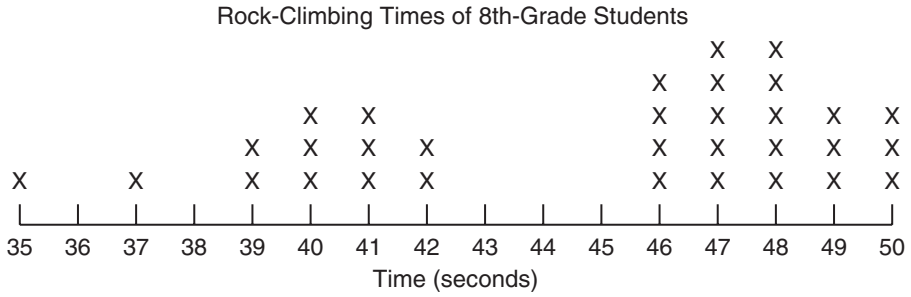
3. Using what you know about distributions and patterns of graphical displays, what conclusions can you draw about the class that has the fastest rock-climbing times?

4. Construct histograms for the rock-climbing times of each class. Do not forget to name your histograms and label the axes correctly.

5. Using what you know about distributions and patterns of graphical displays, what conclusions can you draw about the class that has the fastest rock-climbing time?

Problem 2 Don't Forget the 8th Graders! They Can Rock-Climb Too!

Mr. Garcia, the 8th-grade physical education teacher, also timed his students on the rock-climbing wall. The students' results are shown on a dot plot.



1. How many students are in Mr. Garcia's class? (Assume every student in the class was timed on the rock wall.) How do you know?

2. Describe the distribution and any patterns in the dot plot.

3. What conclusions can you draw about the rock-climbing times of Mr. Garcia's students?

Be prepared to share your solutions and methods.

Chapter 15 Summary

Key Terms

- data (15.1)
- statistical question (15.1)
- population (15.1)
- sample (15.1)
- data analysis (15.1)
- parameter (15.1)
- statistic (15.1)
- survey (15.1)
- experiment (15.1)
- categorical data (15.2)
- quantitative data (15.2)
- discrete data (15.2)
- continuous data (15.2)
- bar graph (15.2)
- frequency (15.2)
- double bar graph (15.2)
- key (15.2)
- stacked bar graph (15.2)
- circle graph (15.2)
- dot plot (15.3)
- distribution (15.3)
- symmetric (15.3)
- skewed right (15.3)
- skewed left (15.3)
- clusters (15.3)
- gaps (15.3)
- stem-and-leaf plot (15.3)
- side-by-side stem-and-leaf plot (15.3)
- histogram (15.4)
- frequency table (15.4)

15.1 Recognizing and Designing Statistical Questions

Data are the facts or numbers that describe the results of an experiment or survey. Data are often collected using a statistical question. A statistical question is a question about a population or a sample. A population is an entire set of items from which data can be collected. A sample is a selection from a population.

Example

"Is soccer my favorite sport?"

The question is not a statistical question, because the person asking the question would know the answer to it.

"What percentage of the students in my physical education class say soccer is their favorite sport?"

This question is a statistical question because the person asking the question would not know the answer to the question without conducting a survey of the population of students in the class.

15

15.1 Differentiating between a Parameter and a Statistic

Collecting, organizing, and analyzing data to answer statistical questions is called data analysis. Through data analysis, you can determine the parameter or statistic from the data. A parameter is the characteristic of a population. A statistic is the characteristic of a sample.

Example

After conducting a survey of the students in her class, Jill concludes that 40 percent of the students in her class play at least one school sport. The students in Jill's class represent the population. The survey results are a parameter because they describe the population.

15.2 Distinguishing between Categorical and Quantitative Data

Categorical data are data for which each piece of data fits into exactly one of several different groups or categories. Categorical data are also called "qualitative" data. Quantitative data are data for which each piece of data can be placed on a numerical scale. Quantitative data are also called "numerical" data. When quantitative data are a count of how many, the data can be described as discrete data. Discrete data can only have values that are counting numbers (0, 1, 2, 3, . . .). When quantitative data are a measurement of something and can have values that are between two counting numbers, the data are called continuous data.

Example

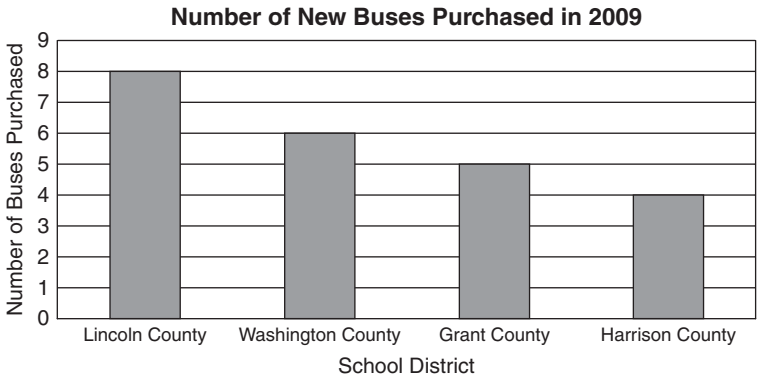
Your teacher records the number of attempts each student takes to shoot a basketball free throw shot. The data are quantitative because the data count the number of free throws each student attempts. The data is discrete because the values can only be counting numbers.

15.2 Creating and Analyzing Single Bar Graphs

Categorical data are often displayed in bar graphs. On a bar graph, categorical data are displayed using horizontal or vertical bars so that the height or length of the bars indicates the value for that category.

Example

Several school districts purchased new buses in 2009. Lincoln County purchased 8 new buses, Washington County purchased 6 new buses, Grant County purchased 5 new buses, and Harrison County purchased 4 new buses. The data are displayed in a bar graph.

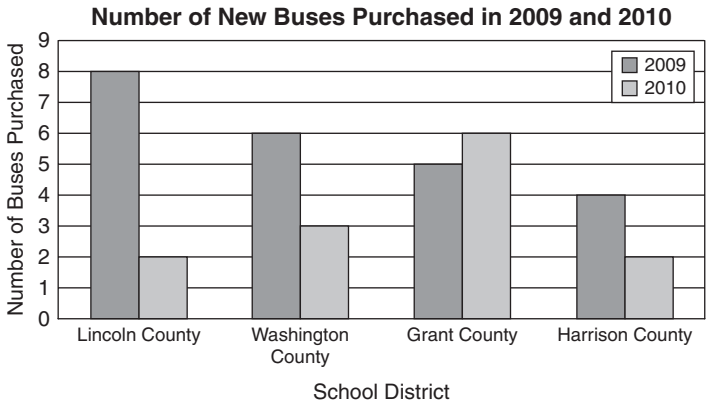


15.2 Creating and Analyzing Double Bar Graphs

Double bar graphs can be used to display categorical data when each category contains two different data sets. A key explains how each data set is represented by a color or a pattern in the graph. The two bars representing the same category are side by side and space is used to separate the categories.

Example

In 2009, four school districts purchased new buses. In 2010, each school district purchased more new buses. Lincoln County purchased 2 more new buses, Washington County purchased 3 more new buses, Grant County purchased 6 more new buses, and Harrison County purchased 2 more new buses.

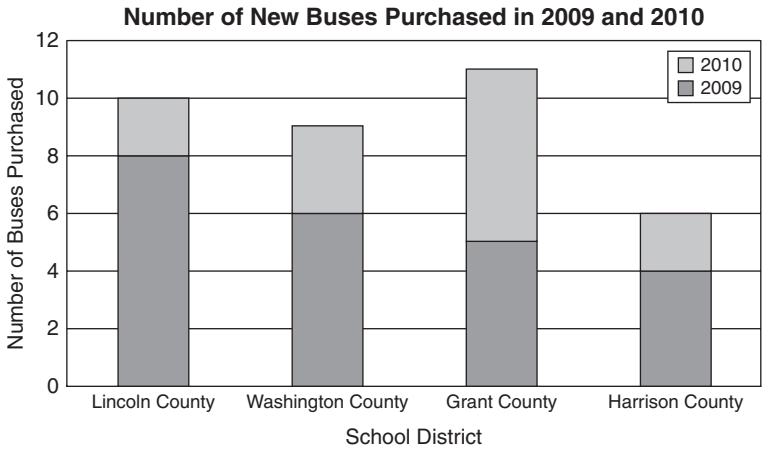


15.2 Creating and Analyzing Stacked Bar Graphs

A stacked bar graph is a graph that stacks the frequencies of two different groups for a given category. The bars are stacked so that the parts can be compared to the whole. A frequency is the number of times an item or number occurs in a data set. Each bar represents a total for the whole category, but still shows how many data pieces make up each group within the entire category.

Example

To determine which school district purchased the most new buses over the 2-year period, observe which district has the tallest stacked bar. According to the stacked bar graph, the Grant County school district purchased the most new buses during 2009 and 2010. They purchased a total of 11 new buses.

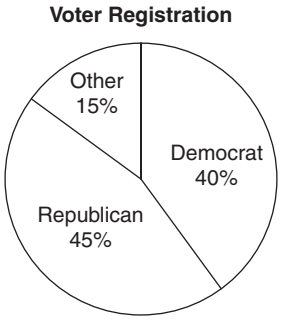


15.2 Analyzing Circle Graphs

Another type of graph used to display categorical data is a circle or pie graph. A circle graph is a graph which shows how parts of the whole relate to the whole, and how parts relate to other parts. The area of each sector, or "wedge," indicates the percentage of the part in relation to the whole.

Example

The circle graph displays the results of a survey given to 300 registered voters at a polling location on Election Day. Voters were asked with which party they were registered.



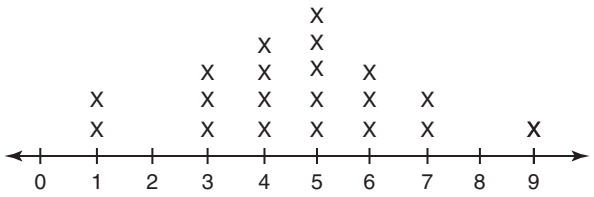
To determine the number of respondents who were Democrats, multiply 300 by 40%. So, $300 \times 0.40 = 120$. Therefore, 120 of the respondents were Democrats.

15.3 Creating and Analyzing Line Plots

A line plot is a graph that shows how discrete data is graphed using a number line. Line plots help organize and display a small number of data points.

Example

The data distribution is symmetric. There is a large cluster from 3 to 7. There are gaps between 1 and 3 and between 7 and 9.



15.3 Creating and Analyzing Stem-and-Leaf Plots

A stem-and-leaf plot is a graphical method used to represent ordered numerical data. Once the data are ordered, the stems and leaves are determined. Typically, the stem is all the digits in a number except the right-most digit, which is the leaf.

Example

The data 32, 11, 28, 15, 18, 20, 10, 12, 26, 44, 51, 15, 23, 56, and 34 are shown in the stem-and-leaf plot.

1	0	1	2	5	5	8
2	0	3	6	8		
3	2	4				
4	4					
5	1	6				

Key: 1|0 = 10

The data distribution is skewed right.

15.3 Creating and Analyzing Side-by-Side Stem-and-Leaf Plots

A side-by-side stem-and-leaf plot allows a comparison of two data sets in two columns. In a side-by-side stem-and-leaf plot two sets of leaves share the same stem. The leaves are to the left and right of the stem.

Example

The two data sets are graphed in the side-by-side stem-and-leaf plot.

Data set one is 2, 5, 8, 12, 12, 13, 17, 19, 20, 22, 25, 25, 30, 32 and 33.

Data set two is 3, 3, 4, 7, 8, 9, 10, 11, 16, 19, 24, 26, and 35.

Data Set One	Stem	Data Set Two
8 5 2	0	3 3 4 7 8 9
9 7 3 2 2	1	0 1 6 9
5 5 2 0	2	4 6
3 2 0	3	5

Key: 1|0 = 10

The data distribution is symmetric in data set one. The data distribution is skewed right in data set two.

15.4 Creating and Analyzing Histograms

A histogram is a way of displaying quantitative data using vertical and horizontal bars so that the height or the length of the bars indicates the frequency. The number of data values in an interval is the frequency. The width of the bar represents the interval, while the height indicates the frequency of values in that interval. To create a histogram, data are usually organized into a frequency table. A frequency table is a table used to organize data according to how many times a data value occurs.

Example

Jeremy's scores for the first 17 times he played the card game, *Clubs and Swords*, are displayed in the frequency table.

Jeremy's Scores for *Clubs and Swords*	
Score	**Frequency (*f*)**
0–100	3
100–200	7
200–300	4
300–400	3

To determine the number of scores that were 200 or greater, add the number of scores in the 200-300 interval and the number of scores in the 300-400 interval. There were 4 scores in the 200-300 interval, and 3 scores in the 300-400 interval. Therefore, there were 7 scores that were 200 or greater.

15.5 Analyzing Data

After conducting a survey or experiment, it is important to correctly analyze the collected data in order to make the proper conclusions.

Example

Your principal randomly selects 20 students and 20 teachers to compete in a contest. Each contestant stands on one leg as long as they can and their times are recorded to the nearest second.

Student Times (seconds)	4, 8, 10, 15, 15, 22, 25, 27, 27, 28, 30, 33, 34, 34, 36, 37, 38, 38, 39, 40
Teacher Times (seconds)	2, 3, 5, 6, 6, 7, 9, 11, 12, 12, 13, 15, 16, 17, 19, 20, 21, 26, 30, 40

The data are recorded in the side-by-side stem-and-leaf plot.

Student Times (seconds)		Teacher Times (seconds)
8 4	0	2 3 5 6 6 7 9
5 5 0	1	1 2 2 3 5 6 7 9
8 7 7 5 2	2	0 1 6
9 8 8 7 6 4 4 3 0	3	0
0	4	0

Key: 2|0 = 20 seconds

The student data is skewed left and the teacher data is skewed right. The students are clearly better at standing on one leg than the teachers. Most of the student times are above 20 seconds, while most of the teacher times are below 20 seconds.

15