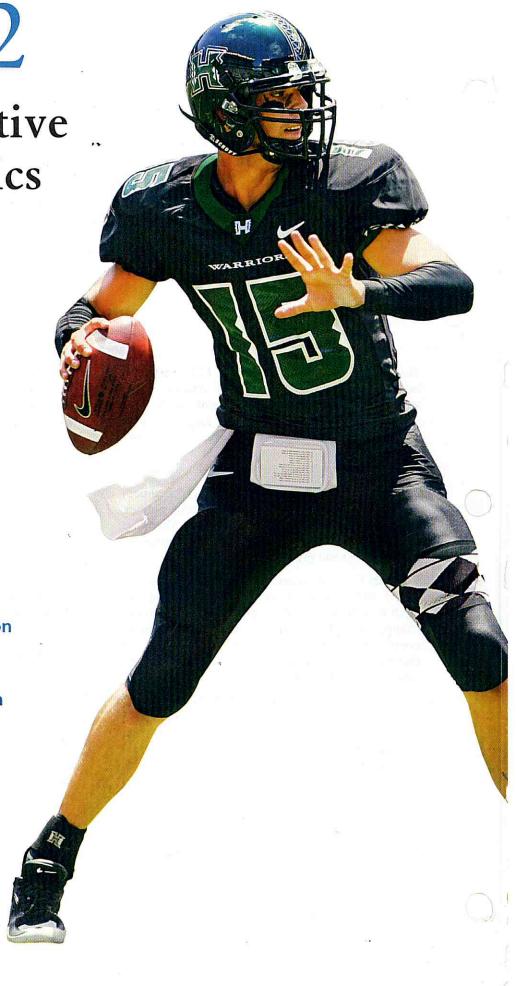
CHAPTER 2

Descriptive Statistics

- 2.1 Frequency
 Distributions and
 Their Graphs
- 2.2 More Graphs and Displays
- 2.3 Measures of Central Tendency
 - **■** ACTIVITY
- 2.4 Measures of Variation
 - ACTIVITY
 - **CASE STUDY**
- 2.5 Measures of Position
 - **USES AND ABUSES**
 - REAL STATISTICS-REAL DECISIONS
 - **TECHNOLOGY**

In 2006, quarterback Colt Brennan of the University of Hawaii set an NCAA record for most touchdown passes in a single season (58).



WHERE YOU'VE BEEN

In Chapter 1, you learned that there are many ways to collect data. Usually, researchers must work with sample data in order to analyze populations, but occasionally it is possible to collect all the data for a given population. For instance, the following represents the number of touchdowns scored by all 119 NCAA Division 1A football teams for the 2006 season.

89, 68, 65, 61, 63, 63, 61, 61, 59, 60, 54, 55, 54, 49, 53, 55, 59, 50, 52, 48, 53, 46, 55, 57, 48, 47, 48, 46, 44, 50, 55, 48, 45, 44, 46, 46, 47, 41, 39, 41, 45, 44, 45, 43, 42, 42, 48, 43, 40, 39, 44, 37, 40, 45, 43, 37, 38, 38, 36, 34, 37, 36, 35, 35, 35, 40, 31, 34, 35, 39, 38, 32, 35, 32, 32, 32, 33, 33, 33, 32, 34, 31, 31, 30, 34, 32, 31, 27, 32, 26, 28, 29, 28, 29, 31, 27, 29, 28, 27, 30, 25, 23, 24, 26, 22, 25, 20, 21, 21, 22, 21, 24, 21, 17, 15, 18, 18, 15, 15

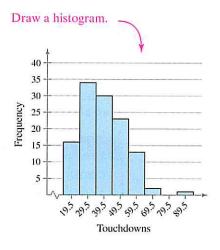
WHERE YOU'RE GOING



In Chapter 2, you will learn ways to organize and describe data sets. The goal is to make the data easier to understand by describing trends, averages, and variations. For instance, in the raw data showing the number of touchdowns scored by all NCAA Division 1A football teams, it is not easy to see any patterns or special characteristics. Here are some ways you can organize and describe the data.

Make a frequency distribution table.

Class	Frequency, f
15-24	16
25-34	34
35-44	30
45-54	23
55-64	13
65-74	2
75-84	0
85-94	1



Mean =
$$\frac{15 + 15 + 15 + 17 + 18 + \dots + 63 + 65 + 68 + 89}{119}$$

= $\frac{4624}{119}$
 $\approx 38.9 \text{ touchdowns}$ Find an average.
Range = $89 - 15$
= 74 touchdowns

2.1 Frequency Distributions and Their Graphs

What You SHOULD LEARN

- How to construct a frequency distribution including limits, midpoints, relative frequencies, cumulative frequencies, and boundaries
- How to construct frequency histograms, frequency polygons, relative frequency histograms, and ogives

Example of a Frequency Distribution

Class	Frequency, f
1–5	5
6-10	8
11-15	6
16-20	8
21-25	5
26-30	4

Study Tip

In a frequency distribution, it is best if each class has the same width. Answers shown will use the minimum data value for the lower limit of the first class. Sometimes it may be more convenient to choose a value that is slightly lower than the minimum value. The frequency distribution produced will vary slightly.

▶ Frequency Distributions

You will learn that there are many ways to organize and describe a data set. Important characteristics to look for when organizing and describing a data set are its **center**, its **variability** (or spread), and its **shape**. Measures of center and shapes of distributions are covered in Lesson 2.3.

Frequency Distributions > Graphs of Frequency Distributions

When a data set has many entries, it can be difficult to see patterns. In this section, you will learn how to organize data sets by grouping the data into intervals called classes and forming a frequency distribution. You will also learn how to use frequency distributions to construct graphs.

DEFINITION

A frequency distribution is a table that shows classes or intervals of data entries with a count of the number of entries in each class. The frequency f of a class is the number of data entries in the class.

In the frequency distribution shown to the left there are six classes. The frequencies for each of the six classes are 5, 8, 6, 8, 5, and 4. Each class has a **lower class limit**, which is the least number that can belong to the class, and an **upper class limit**, which is the greatest number that can belong to the class. In the frequency distribution shown, the lower class limits are 1, 6, 11, 16, 21, and 26, and the upper class limits are 5, 10, 15, 20, 25, and 30. The **class width** is the distance between lower (or upper) limits of consecutive classes. For instance, the class width in the frequency distribution shown is 6 - 1 = 5.

The difference between the maximum and minimum data entries is called the **range.** In the frequency table shown, suppose the maximum data entry is 29, and the minimum data entry is 1. The range then is 29 - 1 = 28. You will learn more about the range in Section 2.4.

GUIDELINES

Constructing a Frequency Distribution from a Data Set

- 1. Decide on the number of classes to include in the frequency distribution. The number of classes should be between 5 and 20; otherwise, it may be difficult to detect any patterns.
- **2.** Find the class width as follows. Determine the range of the data, divide the range by the number of classes, and *round up to the next convenient number*.
- 3. Find the class limits. You can use the minimum data entry as the lower limit of the first class. To find the remaining lower limits, add the class width to the lower limit of the preceding class. Then find the upper limit of the first class. Remember that classes cannot overlap. Find the remaining upper class limits.
- 4. Make a tally mark for each data entry in the row of the appropriate class.
- 5. Count the tally marks to find the total frequency f for each class.

EXAMPLE 1

Constructing a Frequency Distribution from a Data Set

The following sample data set lists the number of minutes 50 Internet subscribers spent on the Internet during their most recent session. Construct a frequency distribution that has seven classes.

50	40	41	17	11	7	22	44	28	21	19	23	37	51	54	42	86
41	78	56	72	56	17	7	69	30	80	56	29	33	46	31	39	20
18	29	34	59	73	77	36	39	30	62	54	67	39	31	53	44	

Solution

- 1. The number of classes (7) is stated in the problem.
- 2. The minimum data entry is 7 and the maximum data entry is 86, so the range is 86 7 = 79. Divide the range by the number of classes and round up to find the class width.

Class width =
$$\frac{79}{7}$$
 Range
Number of classes

 ≈ 11.29 Round up to 12.

- 3. The minimum data entry is a convenient lower limit for the first class. To find the lower limits of the remaining six classes, add the class width of 12 to the lower limit of each previous class. The upper limit of the first class is 18, which is one less than the lower limit of the second class. The upper limits of the other classes are 18 + 12 = 30, 30 + 12 = 42, and so on. The lower and upper limits for all seven classes are shown.
- **4.** Make a tally mark for each data entry in the appropriate class. For example, the data entry 51 is in the 43–54 class, so make a tally mark in that class. Continue until you have made a tally mark for each of the 50 data entries.
- 5. The number of tally marks for a class is the frequency for that class.

The frequency distribution is shown in the following table. The first class, 7–18, has six tally marks. So, the frequency for this class is 6. Notice that the sum of the frequencies is 50, which is the number of entries in the sample data set. The sum is denoted by $\sum f$, where \sum is the uppercase Greek letter **sigma**.

Frequency Distribution for

Internet Usage (in minutes)

Minutes online

Class	Tally	Frequency, f
7–18	##1	6
19-30	####	10
31-42	###	13
43-54	## III	8
55-66	##	5
67–78	##1	6
79–90		2
		$\Sigma f = 50$

Number of

subscribers

Check that the sum of the frequencies equals the number in the sample.

Insight

If you obtain a whole number when calculating the class width of a frequency distribution, use the next whole number as the class width.

Doing this ensures that you have enough space in your frequency distribution for all the data values.

Lower limit	Upper limit
7	18 -
19	30
31	42
43	54
55	66
67	78
79	90

Study Tip

The uppercase Greek letter sigma (Σ) is used throughout statistics to indicate a summation of values.

Try It Yourself 1

Construct a frequency distribution using the number of touchdowns data set listed in the Chapter Opener on page 39. Use eight classes.

- a. State the number of classes.
- b. Find the minimum and maximum values and the class width.
- c. Find the class limits.
- d. Tally the data entries.
- e. Write the frequency f for each class.

Answer: Page A32

After constructing a standard frequency distribution such as the one in Example 1, you can include several additional features that will help provide a better understanding of the data. These features, (the midpoint, relative frequency, and cumulative frequency of each class,) can be included as additional columns in your table.

DEFINITION

The **midpoint** of a class is the sum of the lower and upper limits of the class divided by two. The midpoint is sometimes called the *class mark*.

$$Midpoint = \frac{\text{(Lower class limit)} + \text{(Upper class limit)}}{2}$$

The **relative frequency** of a class is the portion or percentage of the data that falls in that class. To find the relative frequency of a class, divide the frequency f by the sample size n.

Relative frequency =
$$\frac{\text{Class frequency}}{\text{Sample size}}$$
$$= \frac{f}{n}$$

The **cumulative frequency** of a class is the sum of the frequency for that class and all previous classes. The cumulative frequency of the last class is equal to the sample size n.

After finding the first midpoint, you can find the remaining midpoints by adding the class width to the previous midpoint. For instance, if the first midpoint is 12.5 and the class width is 12, then the remaining midpoints are

$$12.5 + 12 = 24.5$$

$$24.5 + 12 = 36.5$$

$$36.5 + 12 = 48.5$$

$$48.5 + 12 = 60.5$$

and so on.

You can write the relative frequency as a fraction, decimal, or percent. The sum of the relative frequencies of all the classes must equal 1, or 100%.

EXAMPLE 2

Finding Midpoints, Relative Frequencies, and Cumulative Frequencies

Using the frequency distribution constructed in Example 1, find the midpoint, relative frequency, and cumulative frequency for each class. Identify any patterns.

Solution The midpoint, relative frequency, and cumulative frequency for the first three classes are calculated as follows.

Class	f	Midpoint	Relative frequency	Cumulative frequency
7–18	6	$\frac{7+18}{2} = 12.5$	$\frac{6}{50} = 0.12$	6
19-30	10	$\frac{19+30}{2}=24.5$	$\frac{10}{50} = 0.2$	6 + 10 = 16
31–42	13	$\frac{31 + 42}{2} = 36.5$	$\frac{13}{50} = 0.26$	16 + 13 = 29

The remaining midpoints, relative frequencies, and cumulative frequencies are shown in the following expanded frequency distribution.

Frequency Distribution for Internet Usage (in minutes)

Minutes online_

Number of subscribers -

A	Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
	7–18	6	12.5	0.12	6
	19-30	10	24.5	0.2	16
H	31-42	13	36.5	0.26	29
	43-54	8	48.5	0.16	37
	55-66	5	60.5	0.1	42
	67-78	6	72.5	0.12	48
	79–90	2	84.5	0.04	50
		$\Sigma f = 50$		$\sum_{n=1}^{\infty} \frac{f}{n} = 1$	U

Interpretation There are several patterns in the data set. For instance, the most common time span that users spent online was 31 to 42 minutes.

Try It Yourself 2

Using the frequency distribution constructed in Try It Yourself 1, find the midpoint, relative frequency, and cumulative frequency for each class. Identify any patterns.

- **a.** Use the formulas to find each *midpoint*, *relative frequency*, and *cumulative frequency*.
- **b.** Organize your results in a frequency distribution.
- c. Identify patterns that emerge from the data.

Answer: Page A32

Portion of

subscribers

Graphs of Frequency Distributions

Sometimes it is easier to identify patterns of a data set by looking at a graph of the frequency distribution. One such graph is a frequency histogram.

Study Tip

If data entries are integers, subtract 0.5 from each lower limit to find the lower class boundaries. To find the upper class boundaries, add 0.5 to each upper limit. The upper boundary of a class will equal the lower boundary of the next higher class.



Class	Class boundaries	Frequency,
7–18	6.5–18.5	6
19-30	18.5-30.5	10
31-42	30.5-42.5	13
43-54	42.5-54.5	8
55-66	54.5-66.5	5

66.5-78.5

78.5-90.5

6

2

Insight

67-78

79-90

It is customary in bar graphs to have spaces between the bars, whereas with histograms, it is customary that the bars have no spaces between them.

DEFINITION

A frequency histogram is a bar graph that represents the frequency distribution of a data set. A histogram has the following properties.

- 1. The horizontal scale is quantitative and measures the data values.
- 2. The vertical scale measures the frequencies of the classes.
- 3. Consecutive bars must touch.

Because consecutive bars of a histogram must touch, bars must begin and end at class boundaries instead of class limits. **Class boundaries** are the numbers that separate classes *without* forming gaps between them. You can mark the horizontal scale either at the midpoints or at the class boundaries, as shown in Example 3.

EXAMPLE 3

Constructing a Frequency Histogram

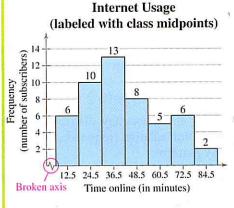
Draw a frequency histogram for the frequency distribution in Example 2. Describe any patterns.

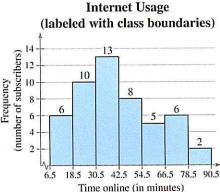
Solution First, find the class boundaries. The distance from the upper limit of the first class to the lower limit of the second class is 19 - 18 = 1. Half this distance is 0.5. So, the lower and upper boundaries of the first class are as follows:

First class lower boundary = 7 - 0.5 = 6.5

First class upper boundary = 18 + 0.5 = 18.5

The boundaries of the remaining classes are shown in the table. Using the class midpoints or class boundaries for the horizontal scale and choosing possible frequency values for the vertical scale, you can construct the histogram.





Interpretation From either histogram, you can see that more than half of the subscribers spent between 19 and 54 minutes on the Internet during their most recent session.

Try It Yourself 3

Use the frequency distribution from Try It Yourself 1 to construct a frequency histogram that represents the number of touchdowns scored by all Division 1A football teams. Describe any patterns.

- a. Find the class boundaries.
- b. Choose appropriate horizontal and vertical scales.
- c. Use the frequency distribution to find the height of each bar.
- d. Describe any patterns for the data.

Answer: Page A33

Another way to graph a frequency distribution is to use a frequency polygon. A **frequency polygon** is a line graph that emphasizes the continuous change in frequencies.

Study Tip

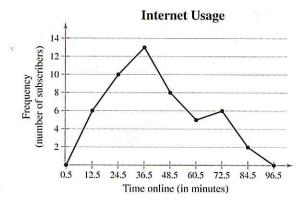
A histogram and its corresponding frequency polygon are often drawn together. If you have not already constructed the histogram, begin constructing the frequency polygon by choosing appropriate horizontal and vertical scales. The horizontal scale should consist of the class midpoints, and the vertical scale should consist of appropriate frequency values.

EXAMPLE 4

Constructing a Frequency Polygon

Draw a frequency polygon for the frequency distribution in Example 2.

Solution To construct the frequency polygon, use the same horizontal and vertical scales that were used in the histogram labeled with class midpoints in Example 3. Then plot points that represent the midpoint and frequency of each class and connect the points in order from left to right. Because the graph should begin and end on the horizontal axis, extend the left side to one class width before the first class midpoint and extend the right side to one class width after the last class midpoint.



Interpretation You can see that the frequency of subscribers increases up to 36.5 minutes and then decreases.

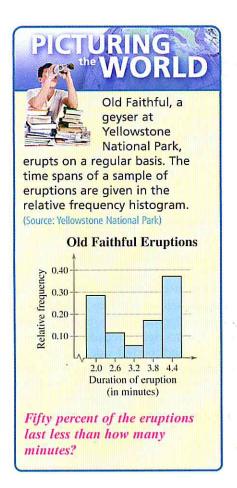
Try It Yourself 4

Use the frequency distribution from Try It Yourself 1 to construct a frequency polygon that represents the number of touchdowns scored by all Division 1A football teams. Describe any patterns.

- a. Choose appropriate horizontal and vertical scales.
- **b.** Plot points that represent the midpoint and frequency for each class.
- c. Connect the points and extend the sides as necessary.
- d. Describe any patterns for the data.

Answer: Page A33

A **relative frequency histogram** has the same shape and the same horizontal scale as the corresponding frequency histogram. The difference is that the vertical scale measures the *relative* frequencies, not frequencies.

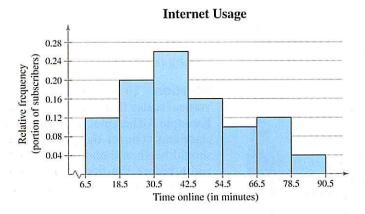


EXAMPLE 5

Constructing a Relative Frequency Histogram

Draw a relative frequency histogram for the frequency distribution in Example 2.

Solution The relative frequency histogram is shown. Notice that the shape of the histogram is the same as the shape of the frequency histogram constructed in Example 3. The only difference is that the vertical scale measures the relative frequencies.



Interpretation From this graph, one piece of information you can quickly see is that 0.20 or 20% of the Internet subscribers spent between 18.5 minutes and 30.5 minutes online, which is not as immediately obvious from the frequency histogram.

Try It Yourself 5

Use the frequency distribution from Try It Yourself 1 to construct a relative frequency histogram that represents the number of touchdowns scored by all Division 1A football teams.

- a. Use the same horizontal scale as used in the frequency histogram.
- **b.** Revise the vertical scale to reflect relative frequencies.
- c. Use the relative frequencies to find the height of each bar.

Answer: Page A33

If you want to describe the number of data entries that are equal to or below a certain value, you can easily do so by constructing a cumulative frequency graph.

DEFINITION

A cumulative frequency graph, or ogive (pronounced o'jive), is a line graph that displays the cumulative frequency of each class at its upper class boundary. The upper boundaries are marked on the horizontal axis, and the cumulative frequencies are marked on the vertical axis.

GUIDELINES

Constructing an Ogive (Cumulative Frequency Graph)

- 1. Construct a frequency distribution that includes cumulative frequencies as one of the columns.
- 2. Specify the horizontal and vertical scales. The horizontal scale consists of upper class boundaries, and the vertical scale measures cumulative frequencies.
- Plot points that represent the upper class boundaries and their corresponding cumulative frequencies.
- 4. Connect the points in order from left to right.
- 5. The graph should start at the lower boundary of the first class (cumulative frequency is zero) and should end at the upper boundary of the last class (cumulative frequency is equal to the sample size).

EXAMPLE 6

Constructing an Ogive

Draw an ogive for the frequency distribution in Example 2. Estimate how many subscribers spent 60 minutes or less online during their last session. Also, use the graph to estimate when the greatest increase in usage occurs.

Solution Using the cumulative frequency distribution, you can construct the ogive shown. The upper class boundaries, frequencies, and cumulative frequencies are shown in the table. Notice that the graph starts at 6.5, where the cumulative frequency is 0, and the graph ends at 90.5, where the cumulative frequency is 50.

	Internet Usage
2 S	50 6.5 18.5 30.5 42.5 54.5 66.5 78.5 90.5
	Time online (in minutes)

Interpretation From the ogive, you can see that about 40 subscribers spent 60 minutes or less online during their last session. It is evident that the greatest increase in usage occurs between 30.5 minutes and 42.5 minutes, as the line segment is steepest between these two class boundaries.

Another type of ogive uses percent as the vertical axis instead of frequency (see Example 5 in Section 2.5).

Upper class boundary	f	Cumulative frequency
18.5	6	6
30.5	10	16
42.5	13	29
54.5	8	37
66.5	5	42
78.5	6	48
90.5	2	50

CHAPTER 2

Try It Yourself 6

Use the frequency distribution from Try It Yourself 1 to construct an ogive that represents the number of touchdowns scored by all Division 1A football teams. Estimate the number of teams who scored 44 or fewer touchdowns.

- a. Specify the horizontal and vertical scales.
- **b.** *Plot* the points given by the upper class boundaries and the cumulative frequencies.
- c. Construct the graph.
- d. Estimate the number of teams who scored 44 or fewer touchdowns.
- e. Interpret the results in the context of the data.

Answer: Page A33

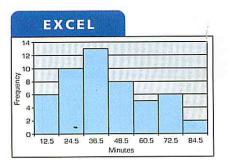
EXAMPLE 7

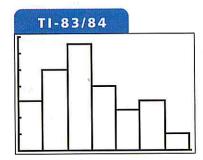
Using Technology to Construct Histograms

Use a calculator or a computer to construct a histogram for the frequency distribution in Example 2.

Solution MINITAB, Excel, and the TI-83/84 each have features for graphing histograms. Try using this technology to draw the histograms as shown.

MINITAB 10 - 12.5 24.5 36.5 48.5 60.5 72.5 84.5 Minutes





Try It Yourself 7

Use a calculator or a computer to construct a frequency histogram that represents the number of touchdowns scored by all Division 1A football teams listed in the Chapter Opener on page 39. Use eight classes.

- a. Enter the data.
- **b.** Construct the histogram.

Answer: Page A33

Study Tip

Detailed instructions for using MINITAB, Excel, and the TI-83/84 are shown in the Technology Guide that accompanies this text. For instance, here are instructions for creating a histogram on a TI-83/84.

STAT ENTER

Enter midpoints in L1. Enter frequencies in L2.

2nd STATPLOT

Turn on Plot 1. Highlight Histogram.

Xlist: L1 Freq: L2

ZOOM 9

WINDOW

Xscl=12 GRAPH



2.11 EXERCISES



Building Basic Skills and Vocabulary

- 1. What are some benefits of representing data sets using frequency distributions?
- 2. What are some benefits of representing data sets using graphs of frequency distributions?
- 3. What is the difference between class limits and class boundaries?
- 4. What is the difference between relative frequency and cumulative frequency?

True or False? In Exercises 5–8, determine whether the statement is true or false. If it is false, rewrite it as a true statement.

- 5. In a frequency distribution, the class width is the distance between the lower and upper limits of a class.
- 6. The midpoint of a class is the sum of its lower and upper limits divided by two.
- 7. An ogive is a graph that displays relative frequency.
- 8. Class boundaries are used to ensure that consecutive bars of a histogram touch.

In Exercises 9–12, use the given minimum and maximum data entries, and the number of classes to find the class width, the lower class limits, and the upper class limits.

9. minimum = 7, maximum = 58, 6 classes

10. minimum = 11, maximum = 94, 8 classes

11. minimum = 15, maximum = 123, 6 classes

12. minimum = 24, maximum = 171, 10 classes

Reading a Frequency Distribution *In Exercises 13 and 14, use the given frequency distribution to find the*

- (a) class width.
- (b) class midpoints.
- (c) class boundaries.
- 13. Cleveland, OH High Temperatures (°F)

Class	Frequency, f
20-30	19
31-41	43
42–52	68
53-63	69
64-74	74
75–85	68
86–96	24

14. Travel Time to Work (in minutes)

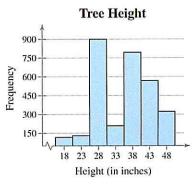
	Class	Frequency, f
	0–9	188
1	10-19	372
	20-29	264
	30-39	205
	40-49	83
	50-59	76
	60-69	32

- **15.** Use the frequency distribution in Exercise 13 to construct an expanded frequency distribution, as shown in Example 2.
- **16.** Use the frequency distribution in Exercise 14 to construct an expanded frequency distribution, as shown in Example 2.

Graphical Analysis In Exercises 17 and 18, use the frequency histogram to

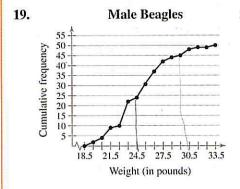
- (a) determine the number of classes.
- (b) estimate the frequency of the class with the least frequency.
- (c) estimate the frequency of the class with the greatest frequency.
- (d) determine the class width.

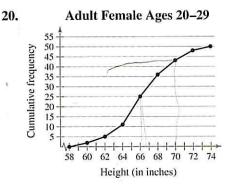




Graphical Analysis In Exercises 19 and 20, use the ogive to approximate

- (a) the number in the sample.
- (b) the location of the greatest increase in frequency.

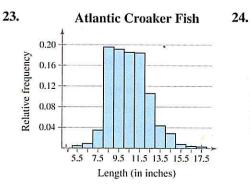


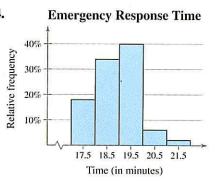


- 21. Use the ogive in Exercise 19 to approximate
 - (a) the cumulative frequency for a weight of 24.5 pounds.
 - (b) the weight for which the cumulative frequency is 45.
- 22. Use the ogive in Exercise 20 to approximate
 - (a) the cumulative frequency for a height of 70 inches.
 - (b) the height for which the cumulative frequency is 25.

Graphical Analysis In Exercises 23 and 24, use the relative frequency histogram to

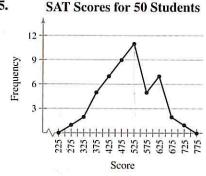
- (a) identify the class with the greatest and the least relative frequency.
- (b) approximate the greatest and least relative frequency.
- (c) approximate the relative frequency of the second class.



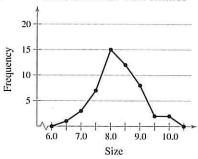


Graphical Analysis In Exercises 25 and 26, use the frequency polygon to identify the class with the greatest and the least frequency.

25.



26. Shoe Sizes for 50 Females



Using and Interpreting Concepts

Constructing a Frequency Distribution In Exercises 27 and 28, construct a frequency distribution for the data set using the indicated number of classes. In the table, include the midpoints, relative frequencies, and cumulative frequencies. Which class has the greatest frequency and which has the least frequency?



27. Newspaper Reading Times

Number of classes: 5

Data set: Time (in minutes) spent reading the newspaper in a day

18 2 30 7 35 12 15 8 6 5 29 0 11 39



28. Book Spending

Number of classes: 6

Data set: Amount (in dollars) spent on books for a semester

530 188 341 376 266 142 273 189 130 489 266 248 101 375 486 398 190 188 269 43 30 127 354 84

indicates that the data set for this exercise is available electronically.

Constructing a Frequency Distribution and a Frequency Histogram

In Exercises 29–32, construct a frequency distribution and a frequency histogram for the data set using the indicated number of classes. Describe any patterns.

🕜 29. Sales

Number of classes: 6

Data set: July sales (in dollars) for all sales representatives at a company

3183 1932 1355 3981 1643 1858 1500 4608

🕜 30. Pepper Pungencies

Number of classes: 5

Data set: Pungencies (in 1000s of Scoville units) of 24 tabasco peppers

35 51 44 42 37 38 36 39 44 43 40 40 32 39 41 38 42 39 40 46 37 35 41 39

31. Reaction Times

Number of classes: 8

Data set: Reaction times (in milliseconds) of a sample of 30 adult females to an auditory stimulus

32. Fracture Times

Number of classes: 5

Data set: Amount of pressure (in pounds per square inch) at fracture time for 25 samples of brick mortar

2750 2862 2867 2718 2641 2834 2466 2596 2519

Constructing a Frequency Distribution and a Relative Frequency Histogram In Exercises 33–36, construct a frequency distribution and a relative frequency histogram for the data set using five classes. Which class has the greatest relative frequency and which has the least relative frequency?

33. Bowling Scores

Data set: Bowling scores of a sample of league members

154 257 182 205

34. ATM Withdrawals

Data set: A sample of ATM withdrawals (in dollars)

20 10 20 25 30 50 80

35. Plant Heights

Data set: Heights (in inches) of a sample of tomato plants



36. Years of Service

Data set: Years of service of a sample of New York state troopers

Constructing a Cumulative Frequency Distribution and an Ogive In Exercises 37-40, construct a cumulative frequency distribution and an ogive for the data set using six classes. Then describe the location of the greatest increase in frequency.



37. Retirement Ages

Data set: Retirement ages for a sample of doctors



38. Saturated Fat Intakes

Data set: Daily saturated fat intakes (in grams) of a sample of people



39. Gasoline Purchases

Data set: Gasoline (in gallons) purchased by a sample of drivers during one fill-up



Data set: Lengths (in minutes) of a sample of cellular phone calls

Constructing a Frequency Distribution and a Frequency Polygon In Exercises 41 and 42, construct a frequency distribution and a frequency polygon for the data set. Describe any patterns.



41. Exam Scores

Number of classes: 5

Data set: Exam scores for all students in a statistics class

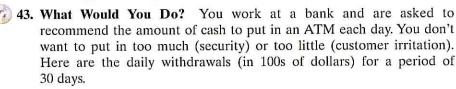
42. Children of the Presidents

Number of classes: 6

Data set: Number of children of the U.S. presidents (Source: presidentschildren.com)

0	5	.6	0	3	4	0	4	10	15	0	6	2	3
0	4	5	4	8	7	3	5	3	2	6	3	3	1
2	2	6	1	2	3	2	2	4	4	4	6	1	2

■ Extending Concepts



```
88
                 104
                           86
72
             76
                       76
                                        89
98
    76
        97
             82
                  84
                       67
                           70
                               81
                               80
            81
                  85
                      78
                           82
                                    91
                                        83
74
    73
        86
```

- (a) Construct a relative frequency histogram for the data using eight classes.
- (b) If you put \$9000 in the ATM each day, what percent of the days in a month should you expect to run out of cash? Explain your reasoning.
- (c) If you are willing to run out of cash for 10% of the days, how much cash, in hundreds of dollars, should you put in the ATM each day? Explain your reasoning.
- 44. What Would You Do? You work in the admissions department for a college and are asked to recommend the minimum SAT scores that the college will accept for a position as a full-time student. Here are the SAT scores for a sample of 50 applicants.

```
982
                 996
                       872 849 785 -706 669 1049
1325 1072
                                       1127 979 1034
                 980 1188 869
                                 1006
 885_ 1367
           935
    1165
           1359
                 667
                      1264 727
                                 -808
                                       955 544 1202
1052
                                 1193
                                       768
                                           812
                                                 -887
-1051
     1173
           410-
                1148
                      1195 1141
1211 1266
           ~830
                 672
                      917
                           988
                                  791 1035
                                                  700
```

- (a) Construct a relative frequency histogram for the data using 10 classes.
- (b) If you set the minimum score at 986, what percent of the applicants will meet this requirement? Explain your reasoning.
- (c) If you want to accept the top 88% of the applicants, what should the minimum score be? Explain your reasoning.
- **45.** Writing What happens when the number of classes is increased for a frequency histogram? Use the data set listed and a technology tool to create frequency histograms with 5, 10, and 20 classes. Which graph displays the data best?

2.2 More Graphs and Displays

What You SHOULD LEARN

- How to graph and interpret quantitative data sets using stem-andleaf plots and dot plots
- How to graph and interpret qualitative data sets using pie charts and Pareto charts
- How to graph and interpret paired data sets using scatter plots and time series charts



Graphing Quantitative Data Sets

In Section 2.1, you learned several traditional ways to display quantitative data graphically. In this section, you will learn a newer way to display quantitative data, called a **stem-and-leaf plot**. Stem-and-leaf plots are examples of **exploratory data analysis (EDA)**, which was developed by John Tukey in 1977.

In a stem-and-leaf plot, each number is separated into a **stem** (for instance, the entry's leftmost digits) and a **leaf** (for instance, the rightmost digit). You should have as many leaves as there are entries in the original data set. A stem-and-leaf plot is similar to a histogram but has the advantage that the graph still contains the original data values. Another advantage of a stem-and-leaf plot is that it provides an easy way to sort data.

EXAMPLE 1

Constructing a Stem-and-Leaf Plot

The following are the numbers of text messages sent last month by the cellular phone users on one floor of a college dormitory.

```
155 159 144 129 105 145 126 116 130 114 122 112 112 142 126 118 118 108 122 121 109 140 126 119 113 117 118 109 109 119 139 139 122 78 133 126 123 145 121 134 124 119 132 133 124 129 112 126 148 147
```

Solution Because the data entries go from a low of 78 to a high of 159, you should use stem values from 7 to 15. To construct the plot, list these stems to the left of a vertical line. For each data entry, list a leaf to the right of its stem. For instance, the entry 155 has a stem of 15 and a leaf of 5. The resulting stem-and-leaf plot will be unordered. To obtain an ordered stem-and-leaf plot, rewrite the plot with the leaves in increasing order from left to right. It is important to include a key for the display to identify the values of the data.



Insight

You can use stem-and-leaf plots to identify unusual data values called outliers. In Example 1, the data value 78 is an outlier. You will learn more about outliers in Section 2.3.

Number of Text Messages Sent

1401	inder of Text Wiessages Sent
7	8 Key: $15 5 = 155$
8	
9	
10	58999
11	6422889378992
12	962621626314496
13	0993423
14	4520587
15	5 9
1000	

Unordered Stem-and-Leaf Plot

Number of Text Messages Sent

7	8 Key: $15 5 = 155$
8	
9	
10	58999
11	2223467888999
12	112223446666699
13	0233499
14	0245578
15	5 9

Ordered Stem-and-Leaf Plot

Interpretation From the ordered stem-and-leaf plot, you can conclude that more than 50% of the cellular phone users sent between 110 and 130 text messages.

Try It Yourself 1

Use a stem-and-leaf plot to organize the number of touchdowns data set listed in the Chapter Opener on page 39. What can you conclude?

- a. List all possible stems.
- b. List the leaf of each data entry to the right of its stem and include a key.
- c. Rewrite the stem-and-leaf plot so that the leaves are ordered.
- **d.** Use the plot to make a conclusion.

Answer: Page A33

EXAMPLE 2

Constructing Variations of Stem-and-Leaf Plots

Organize the data given in Example 1 using a stem-and-leaf plot that has two rows for each stem. What can you conclude?

Solution Construct the stem-and-leaf plot as described in Example 1, except now list each stem twice. Use the leaves 0, 1, 2, 3, and 4 in the first stem row and the leaves 5, 6, 7, 8, and 9 in the second stem row. The revised stem-and-leaf plot is shown.

Insight

Compare Examples 1 and 2. Notice that by using two rows per stem, you obtain a more detailed picture of the data.

Number of Text Messages Sent Number of Text Messages Sent

7	Key: $15 5 = 155$	7	Key: $15 5 = 155$
7	8	7	8
8		8	
8		8	
9		9	
9		9	
10	300	10	
10	58999	10	58999
11	42232	11	22234
11	68897899	11	67888999
12	22123144	12	1 1 2 2 2 3 4 4
12	9666696	12	6666699
13	0 3 4 2 3	13	02334
13	99	13	9 9
14	4 2 0	14	0 2 4
14	5587	14	5 5 7 8
15		15	
15	5 9	15	5 9

Unordered Stem-and-Leaf Plot

Ordered Stem-and-Leaf Plot

Interpretation From the display, you can conclude that most of the cellular phone users sent between 105 and 135 text messages.

▶ Try It Yourself 2

Using two rows for each stem, revise the stem-and-leaf plot you constructed in Try It Yourself 1.

- a. List each stem twice.
- **b.** List all leaves using the appropriate stem row.

Answer: Page A34

You can also use a dot plot to graph quantitative data. In a **dot plot**, each data entry is plotted, using a point, above a horizontal axis. Like a stem-and-leaf plot, a dot plot allows you to see how data are distributed, determine specific data entries, and identify unusual data values.

EXAMPLE 3

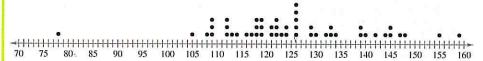
Constructing a Dot Plot

Use a dot plot to organize the text messaging data given in Example 1.

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126 118 118 108 122 121 109 140 126 119 113 117 118 109 109 119 139 139 122 78 133 126 123 145 121 134 124 119 132 133 124 129 112 126 148 147

Solution So that each data entry is included in the dot plot, the horizontal axis should include numbers between 70 and 160. To represent a data entry, plot a point above the entry's position on the axis. If an entry is repeated, plot another point above the previous point.

Number of Text Messages Sent



Interpretation From the dot plot, you can see that most values cluster between 105 and 148 and the value that occurs the most is 126. You can also see that 78 is an unusual data value.

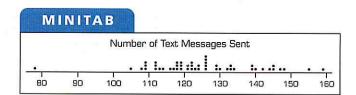
Try It Yourself 3

Use a dot plot to organize the number of touchdowns data set listed in the Chapter Opener on page 39. What can you conclude from the graph?

- a. Choose an appropriate scale for the horizontal axis.
- b. Represent each data entry by plotting a point.
- c. Describe any patterns for the data.

Answer: Page A34

Technology can be used to construct stem-and-leaf plots and dot plots. For instance, a MINITAB dot plot for the text messaging data is shown.



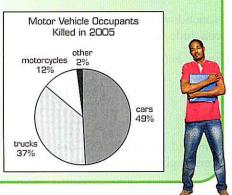
Motor Vehicle Occupants

Killed in 2005

Vehicle type	Killed		
Cars	18,440		
Trucks	13,778		
Motorcycles	4,553		
Other	823		

Study Tip

Here are instructions for constructing a pie chart using Excel. First, enter the vehicle types and their corresponding relative frequencies (written as percents) in two separate columns. Then highlight the two columns, click on the Chart Wizard, and select *Pie* as your chart type. Click *Next* throughout the Chart Wizard while constructing your pie chart.



Graphing Qualitative Data Sets

Pie charts provide a convenient way to present qualitative data graphically as percents of a whole. A **pie chart** is a circle that is divided into sectors that represent categories. The area of each sector is proportional to the frequency of each category. In most cases, you will be interpreting a pie chart or constructing one using technology. Example 4 shows how to construct a pie chart by hand.

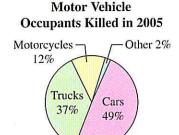
EXAMPLE 4

Constructing a Pie Chart

The numbers of motor vehicle occupants killed in crashes in 2005 are shown in the table. Use a pie chart to organize the data. What can you conclude? (Source: U.S. Department of Transportation, National Highway Traffic Safety Administration)

Solution Begin by finding the relative frequency, or percent, of each category. Then construct the pie chart using the central angle that corresponds to each category. To find the central angle, multiply 360° by the category's relative frequency. For example, the central angle for cars is $360^{\circ}(0.49) \approx 176^{\circ}$. To construct a pie chart in Excel, follow the instructions in the margin.

	ſ	Relative frequency	Angle	
Cars	18,440	0.49	176°	
Trucks	13,778	0.37	133°	
Motorcycles	4,553	0.12	43°	
Other	823	0.02	7°	



Interpretation From the pie chart, you can see that most fatalities in motor vehicle crashes were those involving the occupants of cars.

Try It Yourself 4

The numbers of motor vehicle occupants killed in crashes in 1995 are shown in the table. Use a pie chart to organize the data. Compare the 1995 data with the 2005 data. (Source: U.S. Department of Transportation, National Highway Safety Administration)

Motor Vehicle Occupants Killed in 1995

Vehicle type	Killed
Cars	22,423
Trucks	10,216
Motorcycles	2,227
Other	425

- **a.** Find the *relative frequency* of each category.
- **b.** Use the *central angle* to find the portion that corresponds to each category.
- c. Compare the 1995 data with the 2005 data.

Answer: Page A34

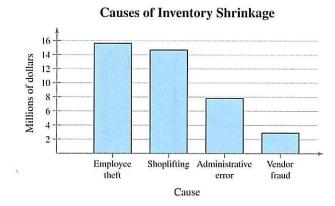
Another way to graph qualitative data is to use a Pareto chart. A **Pareto chart** is a vertical bar graph in which the height of each bar represents frequency or relative frequency. The bars are positioned in order of decreasing height, with the tallest bar positioned at the left. Such positioning helps highlight important data and is used frequently in business.

EXAMPLE 5

Constructing a Pareto Chart

In a recent year, the retail industry lost \$41.0 million in inventory shrinkage. Inventory shrinkage is the loss of inventory through breakage, pilferage, shoplifting, and so on. The causes of the inventory shrinkage are administrative error (\$7.8 million), employee theft (\$15.6 million), shoplifting (\$14.7 million), and vendor fraud (\$2.9 million). If you were a retailer, which causes of inventory shrinkage would you address first? (Source: National Retail Federation and Center for Retailing Education, University of Florida)

Solution Using frequencies for the vertical axis, you can construct the Pareto chart as shown.



Interpretation From the graph, it is easy to see that the causes of inventory shrinkage that should be addressed first are employee theft and shoplifting.

Try It Yourself 5

Every year, the Better Business Bureau (BBB) receives complaints from customers. In a recent year, the BBB received the following complaints.

7792 complaints about home furnishing stores

5733 complaints about computer sales and service stores

14,668 complaints about auto dealers

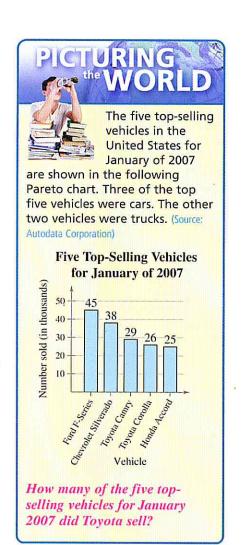
9728 complaints about auto repair shops

4649 complaints about dry cleaning companies

Use a Pareto chart to organize the data. What source is the greatest cause of complaints? (Source: Council of Better Business Bureaus)

- a. Find the frequency or relative frequency for each data entry.
- **b.** Position the bars in decreasing order according to frequency or relative frequency.
- c. Interpret the results in the context of the data.

Answer: Page A34



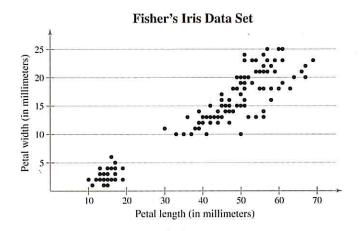
Graphing Paired Data Sets

When each entry in one data set corresponds to one entry in a second data set, the sets are called **paired data sets**. For instance, suppose a data set contains the costs of an item and a second data set contains sales amounts for the item at each cost. Because each cost corresponds to a sales amount, the data sets are paired. One way to graph paired data sets is to use a **scatter plot**, where the ordered pairs are graphed as points in a coordinate plane. A scatter plot is used to show the relationship between two quantitative variables.

EXAMPLE 6

Interpreting a Scatter Plot

The British statistician Ronald Fisher (see page 35) introduced a famous data set called Fisher's Iris data set. This data set describes various physical characteristics, such as petal length and petal width (in millimeters), for three species of iris. In the scatter plot shown, the petal lengths form the first data set and the petal widths form the second data set. As the petal length increases, what tends to happen to the petal width? (Source: Fisher, R. A., 1936)



Solution The horizontal axis represents the petal length, and the vertical axis represents the petal width. Each point in the scatter plot represents the petal length and petal width of one flower.

Interpretation From the scatter plot, you can see that as the petal length increases, the petal width also tends to increase.

Try It Yourself 6

The lengths of employment and the salaries of 10 employees are listed in the table at the left. Graph the data using a scatter plot. What can you conclude?

You will learn more about scatter plots and how to analyze them in Chapter 9.

- a. Label the horizontal and vertical axes.
- **b.** *Plot* the paired data.
- c. Describe any trends.

Answer: Page A34

8 40,000 4 27,350 2 25,000 10 43,000 7 41,650 6 39,225 9 45,100

Salary (in dollars)

32,000

32,500

28,000

Length of employment

(in years)

5

4

3

A data set that is composed of quantitative entries taken at regular intervals over a period of time is a **time series**. For instance, the amount of precipitation measured each day for one month is an example of a time series. You can use a **time series chart** to graph a time series.

EXAMPLE 7

See MINITAB and TI-83/84

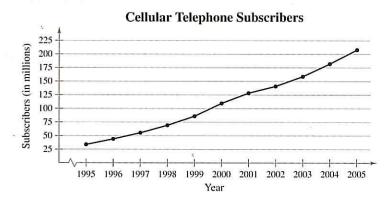
steps on pages 124 and 125.

Constructing a Time Series Chart

The table lists the number of cellular telephone subscribers (in millions) and a subscriber's average local monthly bill for service (in dollars) for the years 1995 through 2005. Construct a time series chart for the number of cellular subscribers. What can you conclude? (Source: Cellular Telecommunications & Internet Association)

Year	Subscribers (in millions)	Average bill (in dollars)
1995	33.8	51.00
1996	44.0	47.70
1997	55.3	42.78
1998	69.2	39.43
1999	86.0	41.24
2000	109.5	45.27
2001	128.4	47.37
2002	140.8	48.40
2003	158.7	49.91
2004	182.1	50.64
2005	207.9	49.98

Solution Let the horizontal axis represent the years and the vertical axis represent the number of subscribers (in millions). Then plot the paired data and connect them with line segments.



Interpretation The graph shows that the number of subscribers has been increasing since 1995, with greater increases recently.

Try It Yourself 7

Use the table in Example 7 to construct a time series chart for a subscriber's average local monthly cellular telephone bill for the years 1995 through 2005. What can you conclude?

- a. Label the horizontal and vertical axes.
- b. Plot the paired data and connect them with line segments.
- c. Describe any patterns you see.

Answer: Page A34

2.2 EXERCISES

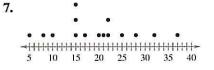
62

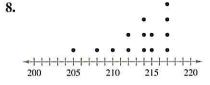


Building Basic Skills and Vocabulary

- 1. Name some ways to display quantitative data graphically. Name some ways to display qualitative data graphically.
- **2.** What is an advantage of using a stem-and-leaf plot instead of a histogram? What is a disadvantage?
- 3. In terms of displaying data, how is a stem-and-leaf plot similar to a dot plot?
- 4. How is a Pareto chart different from a standard vertical bar graph?

Putting Graphs in Context *In Exercises 5–8, match the plot with the description of the sample.*

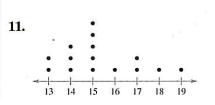


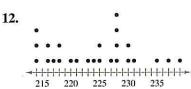


- (a) Time (in minutes) it takes a sample of employees to drive to work
- (b) Grade point averages of a sample of students with finance majors
- (c) Top speeds (in miles per hour) of a sample of high-performance sports cars
- (d) Ages (in years) of a sample of residents of a retirement home

Graphical Analysis In Exercises 9–12, use the stem-and-leaf plot or dot plot to list the actual data entries. What is the maximum data entry? What is the minimum data entry?

10. Key: 12|9 = 12.9

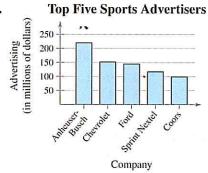




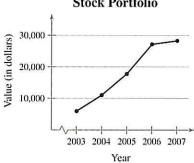
Using and Interpreting Concepts

Graphical Analysis In Exercises 13-16, what can you conclude from the graph?

13.



Stock Portfolio

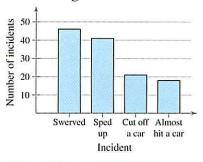


(Source: Nielsen Media Research)

How Other Drivers Irk Us



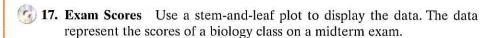
Driving and Cell Phone Use 16.



(Adapted from Reuters/Zogby)

(Adapted from USA TODAY)

Graphing Data Sets In Exercises 17–32, organize the data using the indicated type of graph. What can you conclude about the data?



75 85 90 80 87 67 82 88 95 91 73 80 83 92 94 68 75 91 79 95 87 76 91 85

18. World's Richest People Use a stem-and-leaf plot to display the data. The data represent the ages of the top 25 wealthiest people in the world. (Source: Forbes)

> 51 76 67 80 56 73 58 71 78 49 62 84 49 87 40 59 47 54 84 61 79 59 52 63

19. Ice Thickness Use a stem-and-leaf plot to display the data. The data represent the thicknesses (in centimeters) of ice measured at 20 different locations on a frozen lake.

> 5.8 6.4 6.9 7.2 5.1 4.9 4.3 5.8 7.0 6.8 8.1 7.5 7.2 6.9 5.8 7.2 8.0 7.0 6.9 5.9

20. Apple Prices Use a stem-and-leaf plot to display the data. The data represent the price (in cents per pound) paid to 28 farmers for apples.

19.2 19.6 16.4 17.1 19.0 17.4 17.3 20.1 19.0 17.5

17.6 18.6 18.4 17.7 19.5 18.4 18.9 17.5 19.3 20.8 19.3 18.6 18.6 18.3 17.1 18.1 16.8 17.9

21. Advertisements Use a dot plot to display the data. The data represent the number of advertisements seen or heard in one week by a sample of 30 people from the United States.

598 494 441 595 728 690 684 486 735 808 734 590 673 545 702 481 298 135 846 764 317 649 732 582 637 588 540 727 486 703

22. Life Spans of House Flies Use a dot plot to display the data. The data represent the life spans (in days) of 40 house flies.

·8 11 10 13 9 8 14 10 6 10 10 8 14 11 13 10 14 11 14 8 13 14 10

23. United Nations Use a pie chart to display the data. The data represent the number of countries in the United Nations by continent. (*Source: United Nations*)

North America 23 Europe 43 Africa 53 South America 12 Oceania 14 Asia 47

24. NASA Budget Use a pie chart to display the data. The data represent the 2007 NASA budget (in millions of dollars) divided among three categories. (Source: NASA)

Science, aeronautics, and exploration 10,651 Exploration capabilities 6108 Inspector General 34

25. Airline Baggage Use a Pareto chart to display the data. The data represent the results of a 2005 worldwide study of all airlines on the causes for baggage delay. (Source: Société International de Télécommunications Aéronautiques)

26. UV Index Use a Pareto chart to display the data. The data represent the ultraviolet index for five cities at noon on a recent date. (Source: National Oceanic and Atmospheric Administration)

Atlanta, GA Boise, ID Concord, NH Denver, CO Miami, FL 9 7 8 7 10

27. Hourly Wages Use a scatter plot to display the data in the table. The data represent the number of hours worked and the hourly wages (in dollars) for a sample of 12 production workers. Describe any trends shown.

Hours	Hourly wage
33	12.16
37	9.98
34	10.79
40	11.71
35	11.80
33	11.51
40	13.65
33	12.05
28	10.54
45	10.33
37	11.57
28	10.17

Number of students per teacher	Average teacher's salary
17.1	28.7
17.5	47.5
18.9	31.8
17.1	28.1
20.0	40.3
18.6	33.8
14.4	49.8
16.5	37.5
13.3	42.5
18.4	31.9

TABLE FOR EXERCISE 28

- 28. Salaries Use a scatter plot to display the data shown in the table. The data represent the number of students per teacher and the average teacher salary (in thousands of dollars) for a sample of 10 school districts. Describe any trends shown.
- 29. UV Index Use a time series chart to display the data. The data represent the ultraviolet index for Memphis, TN, on June 14–23 during a recent year. (Source: Weather Services International)

June 14	June 15	June 16	June 17	June 18
9	4	10	10	10
June 19 10	June 20 10	June 21 10	June 22 9	June 23 9

30. Daily High Temperatures Use a time series chart to display the data. The data represent the daily high temperatures for a city for a period of 12 days.

May 1	May 2	May 3	May 4	May 5	May 6
77°	77°	79°	81°	82°	82°
May 7	May 8	May 9	May 10	May 11	May 12
85°	87°	90°	88°	89°	82°

31. Egg Prices Use a time series chart to display the data. The data represent the prices of Grade A eggs (in dollars per dozen) for the indicated years. (Source: U.S. Bureau of Labor Statistics)

1994	1995	1996	1997	1998	1999
0.87	1.16	1.31	1.17	1.09	0.92
2000	2001	2002	2003	2004	2005
0.96	0.93	1.18	1.56	1.20	1.35

32. Ground Beef Prices Use a time series chart to display the data. The data represent the prices of 100% ground beef (in dollars per pound) for the indicated years. (Source: U.S. Bureau of Labor Statistics)

1994	1995	1996	1997	1998	1999
1.38	1.40	1.42	1.39	1.39	1.53
2000	2001	2002	2003	2004	2005
1.63	1.71	1.69	2.23	2.14	2.30

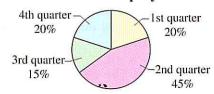
■ Extending Concepts

A Misleading Graph? In Exercises 33 and 34,

- (a) explain why the graph is misleading.
- (b) redraw the graph so that it is not misleading.



34. Sales for Company B



1st	2nd	3rd	4th
quarter	quarter	quarter	quarter
20%	15%	45%	20%

35. Law Firm Salaries A back-to-back stem-and-leaf plot compares two data sets by using the same stems for each data set. Leaves for the first data set are on one side while leaves for the second data set are on the other side. The back-to-back stem-and-leaf plot shows the salaries (in thousands) of all lawyers at two small law firms.

Law Fir	m A		Law Firm B
	50	9	0 3
8 5 2	2 2	10	5 7
9 9 7	0 0	11	0 0 5
	1 1	12	0 3 3 5
		13	2 2 5 9
	1	14	1 3 3 3 9
		15	5 5 5 6
		16	4 9 9
9 9 5	10	17	1 2 5
5 5 5	2 1	18	9
9 9 8		19	0
	3	20	

Key: 5|19|0 = \$195,000 for Law Firm A and \$190,000 for Law Firm B

- (a) What are the lowest and highest salaries at Law Firm A? at Law Firm B?
- (b) How many lawyers are in each firm?
- (c) Compare the distribution of salaries at each law firm. What do you notice?
- 36. Yoga Classes The back-to-back stem-and-leaf plot shows the ages of all participants in two yoga classes.

3:00 P.M. Class

Key: 5|3|1 = 35-year-old in 3:00 P.M. class and 31-year-old in 8:00 P.M. class

- (a) What are the lowest and highest ages of participants in the 3:00 P.M. -class? in the 8:00 P.M. class?
- (b) How many participants are in each class?
- (c) Compare the distribution of ages in each class. What conclusion(s) can you make based on your observations?