Section **7.1**   COLLECTING AND GRAPHING DATA
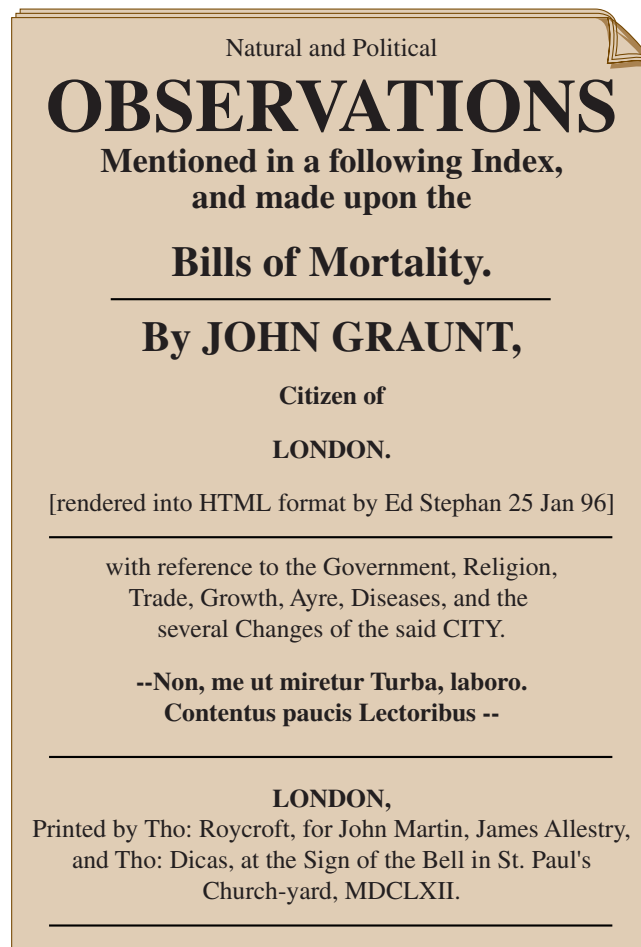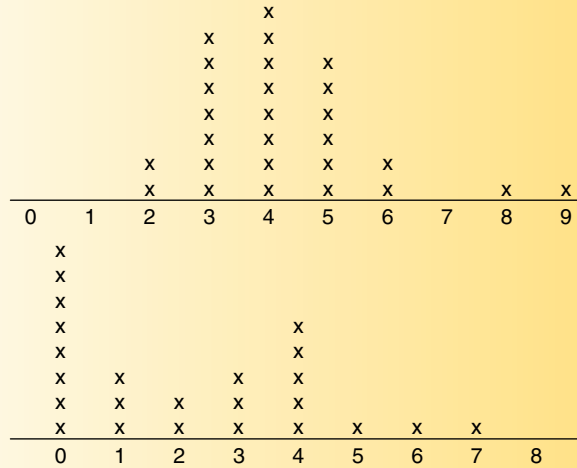
HISTORICAL HIGHLIGHT

Statistics had its beginning in the seventeenth century in the work of Englishman John Graunt (1620–1674). Born in London, Graunt developed human statistical and census methods that later provided the framework for modern demography. He used a publication called *Bills of Mortality,* which listed births, christenings, and deaths. Here are some of his conclusions: The number of male births exceeds the number of female births; there is a higher death rate in urban areas than in rural areas; and more men than women die violent deaths. Graunt used these statistics in his book *Natural and Political Observations of Mortality.* He analyzed the mortality rolls in early London as Charles II and other officials attempted to create a system to warn of the onset and spread of bubonic plague. Graunt summarized great amounts of information to make it understandable (descriptive statistics) and made conjectures about large populations based on small samples (inferential statistics).

Natural and Political

OBSERVATIONS

**Mentioned in a following Index,
and made upon the**

Bills of Mortality.

By JOHN GRAUNT,

Citizen of

LONDON.

[rendered into HTML format by Ed Stephan 25 Jan 96]

with reference to the Government, Religion, Trade, Growth, Ayre, Diseases, and the several Changes of the said CITY.

**--Non, me ut miretur Turba, laboro.
Contentus paucis Lectoribus --**

LONDON,
Printed by Tho: Roycroft, for John Martin, James Allestry, and Tho: Dicas, at the Sign of the Bell in St. Paul's Church-yard, MDCLXII.

PROBLEM OPENER

Two line plots are shown here with each x representing a fourth-grade student. One line plot shows the number of students having a given number of cavities, and the other shows the number of students having a given number of people in their families. Which plot contains the data on cavities?*

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

        x
        x
        x
        x                                     x
        x                                     x
        x       x                     x       x
        x       x       x       x     x
        x       x       x       x     x       x       x       x
        0       1       2       3     4       5       6       7       8
```

Graphs provide quick visual summaries of information and methods of making predictions. Some of the more common graphs are introduced in this section.

BAR GRAPHS

The table in Figure 7.1 lists the responses of 40 teachers to a proposal to begin and end the school day one-half hour earlier. Teachers' responses are classified into one of three categories: favor (F); oppose (O); or no opinion (N).

Teacher	Category	Teacher	Category	Teacher	Category
1	F	14	F	27	N
2	F	15	O	28	O
3	O	16	F	29	F
4	N	17	N	30	O
5	F	18	O	31	F
6	O	19	F	32	F
7	O	20	N	33	N
8	F	21	F	34	F
9	F	22	F	35	O
10	F	23	O	36	F
11	O	24	F	37	N
12	O	25	N	38	F
13	N	26	O	39	O
				40	O

Figure 7.1

*J. Zawojewski, "Polishing a Data Task: Seeking Better Assessment," *Teaching Children Mathematics* 2 (February, 1996): 372–378.

The data from the preceding table are summarized by the **bar graph** in Figure 7.2. The intervals on the horizontal axis represent the three categories, and the vertical axis indicates the number of teachers for each category. Compare the graph to the table and notice that this graph provides a quick summary of the data.
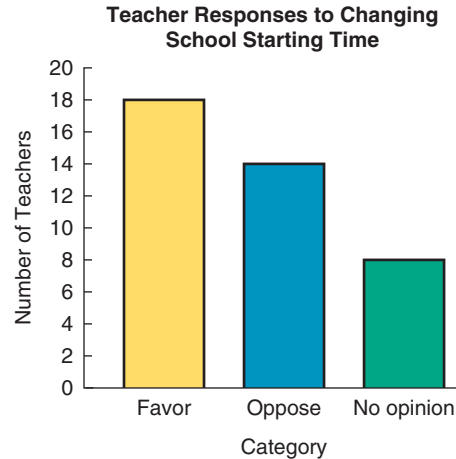

Figure 7.2

Some types of bar graph have two bars for each category and are called a **double-bar graph,** while others have three bars for each category and are called a **triple-bar graph.** Figure 7.3 is a triple-bar graph that has four categories of age groups and compares the percentages of black children, Hispanic children, and white children who have not seen a physician in the past year.*
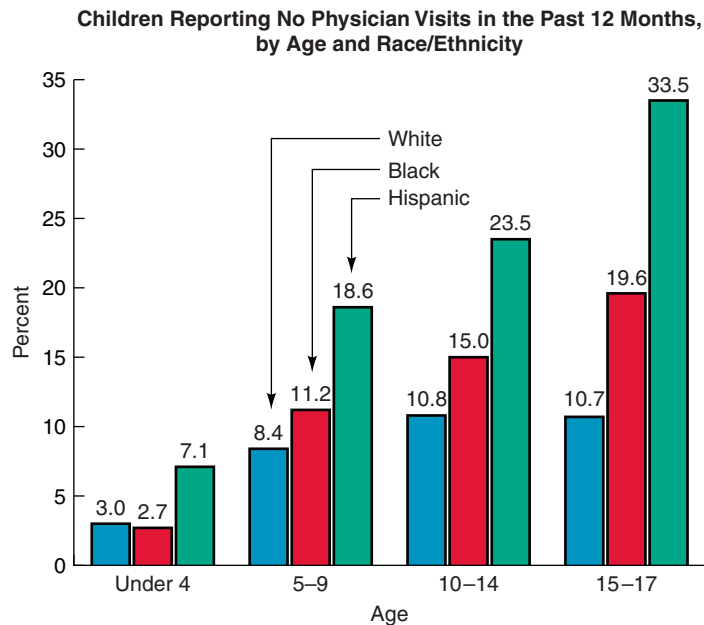

Figure 7.3

*U.S. Department of Health and Human Services, *Child Health USA 2003* (Washington, DC: U.S. Government Printing Office, 2003), p. 58.

EXAMPLE A

Use the information in the graph in Figure 7.3.

1. What is the difference between the percentage of Hispanic children and the percentage of white children aged 15 to 17 who have not seen a physician in the past year?

2. What percentage of black children aged 10 to 14 years visited a physician in the past year?

3. Was there a greater difference in the percentages of black children and white children who did not visit a physician during the past year in the age group of 5 to 9 or the age group of 10 to 14?

Solution   1. 22.8 percent   **2.** 85.0 percent   **3.** Age group of 10 to 14

PIE GRAPHS

A **pie graph** (circle graph) is another way to summarize data visually. A disk (pie) is used to represent the whole, and its pie-shaped sectors represent the parts in proportion to the whole. Consider, for example, the data from Figure 7.1. A total of 40 responses are classified into three categories: 18 in favor, 14 opposed, and 8 with no opinion. These categories represent $\frac{18}{40}$, $\frac{14}{40}$, and $\frac{8}{40}$ of the total responses, respectively. To determine the central angles for the sectors of a pie graph, we multiply these fractions by 360°.

$$\frac{18}{40} \times 360° = 162° \qquad \frac{14}{40} \times 360° = 126° \qquad \frac{8}{40} \times 360° = 72°$$

The pie graph for this data is constructed by first drawing a circle and making three sectors, using the central angles, as in part a of Figure 7.4. Then each sector is labeled so that the viewer can easily interpret the results, as in part b.
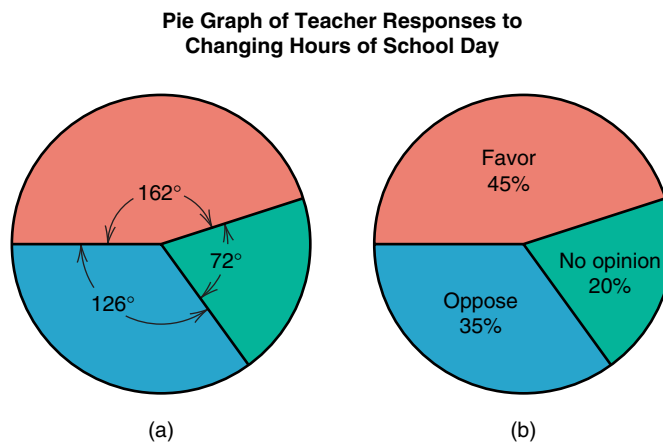
**Pie Graph of Teacher Responses to
Changing Hours of School Day**

Figure 7.4

(a)   (b)

PICTOGRAPHS

A **pictograph** (see Figure 7.5) is similar to a bar graph. The individual figures or icons that are used each represent the same value. For example, each stick figure in the pictograph (on page 442) represents 10,000 juveniles (ages 10 to 17).* Notice how easily you can see increases and decreases in the numbers of juveniles for the given years.

*Statistical Abstracts of the United States, 124th ed. (Washington, DC: Bureau of the Census, 2004), p. 198.

I will explore reading and making line graphs.

6.5 Explore Making Line Graphs

Hands On Activity

You can use **line graphs** to study how things change over time. For example, you can analyze data about the number of passenger planes that flew during a period of 5 years.

VOCABULARY
line graph

Airline Flights

Year	1990	1991	1992	1993	1994
Number of Flights	6,900	6,800	7,000	7,200	7,500

You Will Need

• graph paper

Make a Line Graph

STEP 1

Choose a title and labels. Title the graph "Airline Flights." Label the bottom "Years" and the left side "Number of Flights."

STEP 2

The scale starts at zero and ends at 7,800 on the left side. A zigzag line shows a break in the scale from 0 to 6,600. Mark off the equal spaces in intervals of 200. Write the years on the bottom.

STEP 3

Write ordered pairs for the data (1990, 6,900). Place a point at 1990, between 7,000 and 6,800. Do the same for the remaining data. Connect the points with straight lines.

Technology Link

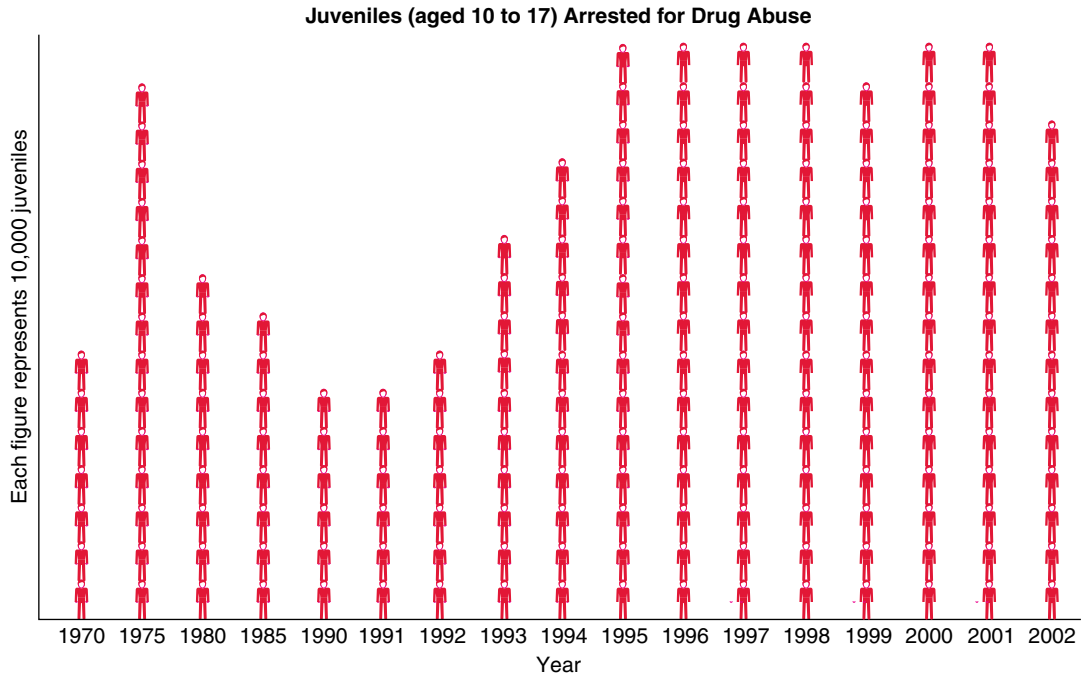Use the Graphs tool in the **Math Tool Chest** to help you make a line graph.

Airline Flights

Juveniles (aged 10 to 17) Arrested for Drug Abuse
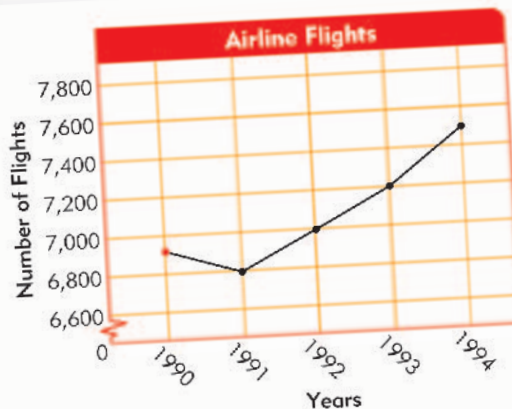
Figure 7.5

EXAMPLE B

The number of juveniles arrested for drug abuse in each of the years given in Figure 7.5 is rounded to the nearest 10,000.

1. How many fewer juveniles, to the nearest 10,000, were arrested for drug abuse in 1980 than in 1975?

2. To the nearest 10,000, what was the total number of juveniles arrested for drug abuse in 1990 through 1995?

3. Were there more juveniles arrested in the 4-year period from 1995 through 1998 or the 4-year period from 1999 through 2002?

Solution   **1.** 50,000   **2.** 560,000   **3.** More in the period from 1995 through 1998

LINE PLOTS

The table on the next page shows the countries that won one or more gold medals at the 2000 Summer Olympics in Sydney, Australia. Some information can be spotted quickly from the table, such as determining the countries that won large numbers of gold medals, but details such as comparing the numbers of countries that won one, two, or three gold medals are more time-consuming.

To assist in analyzing and viewing the data in the table, a line plot has been drawn in Figure 7.6. A **line plot** is formed by drawing a line, marking categories and recording data by placing a mark such as a dot or an X above the line for each value of the data.

A line plot is easy to construct and interpret, and it gives a clear graphical picture of the data. Also certain features of the data become more apparent from a line plot than from a table. Such features include **gaps** (large spaces in the data) and **clusters** (isolated groups of data).

Countries That Won at Least One Gold Medal in the 2000 Summer Olympics

|					
Algeria	1	Ethiopia	4	New Zealand	1
Australia	16	Finland	2	Norway	4
Austria	2	France	13	Poland	6
Azerbaijan	2	Germany	14	Romania	11
Bahamas	1	Greece	4	Russia	32
Belarus	3	Hungary	8	Slovakia	1
Britain	11	Indonesia	1	Slovenia	2
Bulgaria	5	Iran	3	South Korea	8
Cameroon	1	Italy	13	Spain	3
Canada	3	Japan	5	Sweden	4
China	28	Kazakhstan	3	Switzerland	1
Colombia	1	Kenya	2	Thailand	1
Croatia	2	Latvia	1	Turkey	3
Cuba	11	Lithuania	2	Ukraine	3
Czech Republic	2	Mexico	1	United States	40
Denmark	2	Mozambique	1	Uzbekistan	1
Estonia	1	Netherlands	12	Yugoslavia	1

Number of Countries Winning Given Numbers of Gold Medals in 2000 Summer Olympic Games
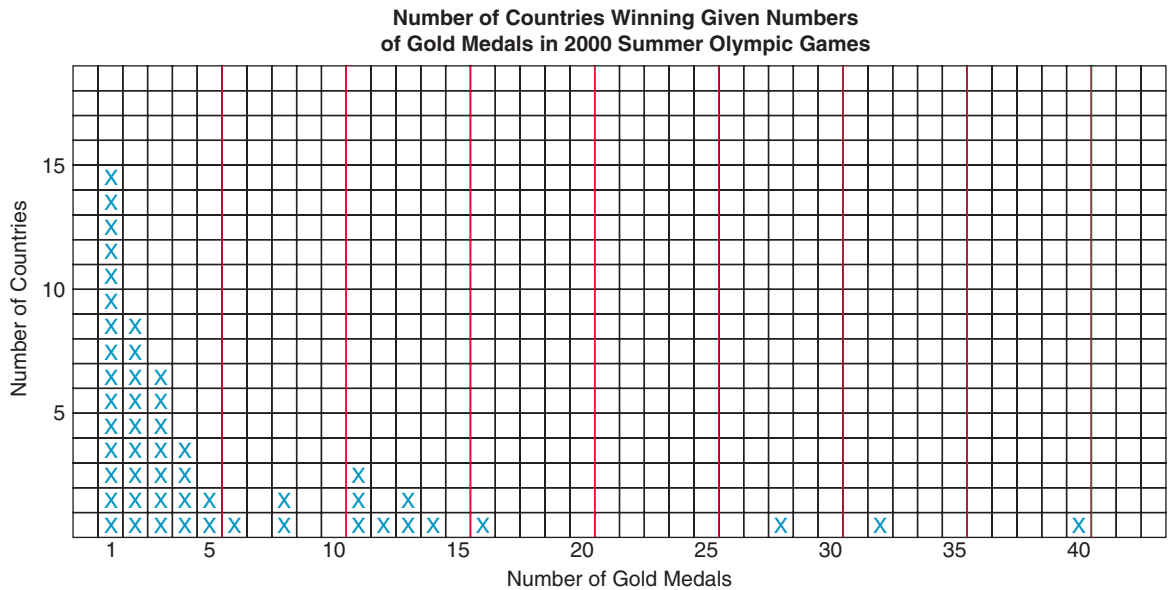
Figure 7.6

EXAMPLE C   Use the line plot in Figure 7.6 to answer these questions.

> **1.** There is one large cluster of data for the countries that won six or fewer gold medals. How many countries are represented in this cluster?
>
> **2.** The largest gap in the data occurs between which two numbers? How large is this gap?

Solution   **1.** 38   **2.** Between 16 and 28. The size of the gap is 11.

STEM-AND-LEAF PLOTS

A **stem-and-leaf plot** is a quick numerical method of providing a visual summary of data. As the name indicates, this method suggests the stems of plants and their leaves. Consider the following test scores for a class of 26 students:

<div align="center">

82   66   70   77   94   67   73   78   82   74   90   45   62
85   57   72   94   83   85   70   95   71   89   87   75   74

</div>

Since the scores in the preceding list range from the 40s to the 90s, the tens digits of 4, 5, 6, 7, 8, and 9 are chosen as the stems, and the unit digits of the numbers will represent the leaves (Figure 7.7). The first step in forming a stem-and-leaf plot is to list the stem values in increasing order in a column (see part a). Next, each leaf value is written in the row corresponding to that number's stem (part b). Here the leaf values have been recorded in the order in which they appear, but they could be listed in increasing order. For example, the leaves for stem 6 can be recorded as 2, 6, 7 rather than 6, 7, 2, as shown in Figure 7.7b. The stem-and-leaf plot shows at a glance the lowest and highest test scores and that the 70s interval has the greatest number of scores.

Stem	Leaf		Stem	Leaf
4			4	5
5			5	7
6			6	6 7 2
7			7	0 7 3 8 4 2 0 1 5 4
8			8	2 2 5 3 5 9 7
9			9	4 0 4 5
	(a)			(b)

Figure 7.7

A stem-and-leaf plot shows where the data are concentrated and the extreme values. You may have noticed that this method of portraying data is like a bar graph turned on its side (rotate this page 90° counterclockwise). Although a stem-and-leaf plot is not as attractive as a bar graph, it has the advantage of showing all the original data. Furthermore, unlike a bar graph, it shows any gaps or clusters in the data.

A stem-and-leaf plot that compares two sets of data can be created by forming a central stem and plotting the leaves for the first set of data on one side of the stem and the leaves for the second set on the other side (Figure 7.8). This is called a **back-to-back stem-and-leaf plot.** Suppose the same class of students whose test scores are shown above obtains the following scores on a second test:

<div align="center">

85   89   70   76   49   66   71   71   75   82   73   77   68
79   55   91   52   63   64   84   81   68   73   67   66   72

</div>

A stem-and-leaf plot of the scores on both tests is shown in Figure 7.8. In this plot the leaves for both sets of scores have been arranged in order to aid in comparing the test scores. It appears that overall performance was better on the first test. For example, the first test has almost twice as many scores above 80 and one-half as many scores below 70 as the second test.
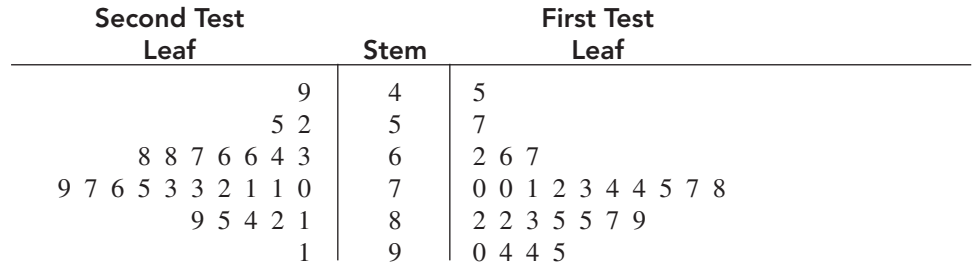
	Second Test Leaf	Stem	First Test Leaf
	9	4	5
	5 2	5	7
	8 8 7 6 6 4 3	6	2 6 7
	9 7 6 5 3 3 2 1 1 0	7	0 0 1 2 3 4 4 5 7 8
	9 5 4 2 1	8	2 2 3 5 5 7 9
	1	9	0 4 4 5

Figure 7.8

HISTOGRAMS

When data fall naturally into a few categories, as in Figure 7.1 on page 438, they can be illustrated by bar graphs or pie graphs. However, data are often spread over a wide range with many different values. In this case it is convenient to group the data in intervals.

The following list shows the gestation periods in days for 42 species of animals.

Ass 365	Deer 201	Moose 240
Baboon 187	Dog 61	Mouse 21
Badger 60	Elk 250	Opossum 15
Bat 50	Fox 52	Pig 112
Black bear 219	Giraffe 425	Puma 90
Grizzly bear 225	Goat, domestic 151	Rabbit 37
Polar bear 240	Goat, mountain 184	Rhinoceros 498
Beaver 122	Gorilla 257	Sea lion 350
Buffalo 278	Guinea pig 68	Sheep 154
Camel 406	Horse 330	Squirrel 44
Cat 63	Kangaroo 42	Tiger 105
Chimpanzee 231	Leopard 98	Whale 365
Chipmunk 31	Lion 100	Wolf 63
Cow 284	Monkey 165	Zebra 365

Since there are many different gestation periods, we group them in intervals. The intervals should be nonoverlapping, and their number is arbitrary but usually a number from 5 to 15. One method of determining the length of each interval is to first compute the difference between the highest and lowest values, which is $498 - 15 = 483$. Then select the desired number of intervals and determine the length of the interval. If we select 10 as the number of intervals, then

$$483 \div 10 = 48.3$$

and we may choose 49 (because of its convenience) as the width of each interval. Figure 7.9 lists the number of animals in each interval and is called a **frequency table.**

Frequency Table

Interval	0–49	50–99	100–149	150–199	200–249	250–299	300–349	350–399	400–449	450–499
Frequency	6	9	4	5	6	4	1	4	2	1

Figure 7.9

The graph for the grouped data in the frequency table is shown in Figure 7.10. This graph, which is similar to a bar graph, is called a **histogram.** A histogram is made up of adjoining bars that have the same width, and the bars are centered above the midpoints of the intervals or categories. The vertical axis shows the frequency of the data for each interval or category on the horizontal axis. We can see from this histogram that the greatest number of gestation periods occurs in the interval from 50 to 99 days, and there are only a few animals with gestation periods over 400 days.

Gestation Periods of Animals
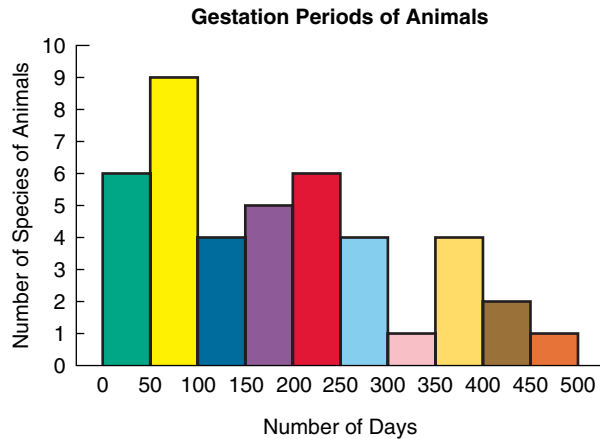
Figure 7.10

LINE GRAPHS

Another method of presenting data visually is the **line graph.** This type of graph is often used to show changes over a period of time. For example, the line graph in Figure 7.11 shows the increase in population from 1800 to 2000 at 20-year intervals.

U.S. Population Growth
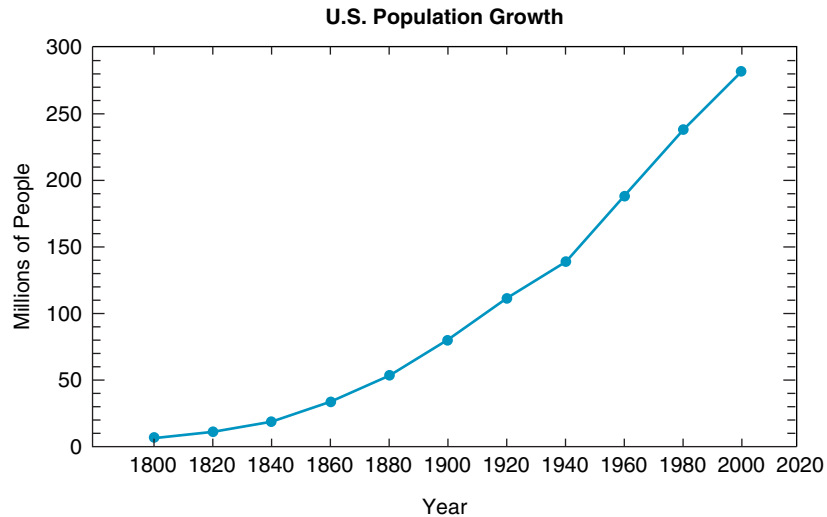
Figure 7.11

EXAMPLE D   Use the line graph in Figure 7.11 to answer these questions.

1. What was the approximate population increase from 1880 to 1920?

2. Compare the population change for the period from 1800 to 1900 to the population change from 1960 to 2000. Which period had the greater increase in population?

Solution   **1.** 55 million   **2.** 1960 to 2000

SCATTER PLOTS

Consider the following table, which records the heights and corresponding shoe sizes of 30 fourth-grade to eighth-grade boys. It is difficult to see any patterns or relationships between the heights and shoe sizes from this information.

Height (inches)	59	71	57	72	64	60	64	62	66	63	74	60	67
Shoe size	6.5	11.5	4	10.5	9.5	5	7.5	8.5	9	7	11.5	4.5	8

Height (inches)	64	65	62	56	69	61	58	62	63	67	69	64	68	60	58	66	57
Shoe size	6.5	12	6	4.5	10	7	4	5	5.5	10	9	6	10.5	6.5	3.5	8.5	5

The pairs of numbers in the table have been graphed in Figure 7.12, where the first coordinate of each point on the graph is a height and the second coordinate of the point is the corresponding shoe size. Such a graph is called a **scatter plot.** The scatter plot enables us to see if there are any patterns or trends in the data. Although there are boys who have larger shoe sizes than some of the boys who are taller, in general it appears that taller boys have larger shoe sizes.
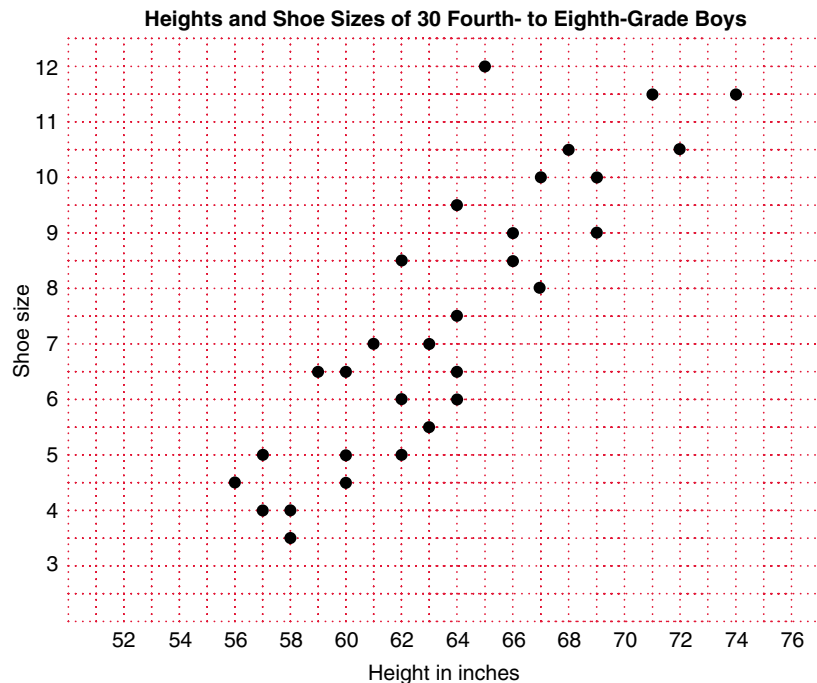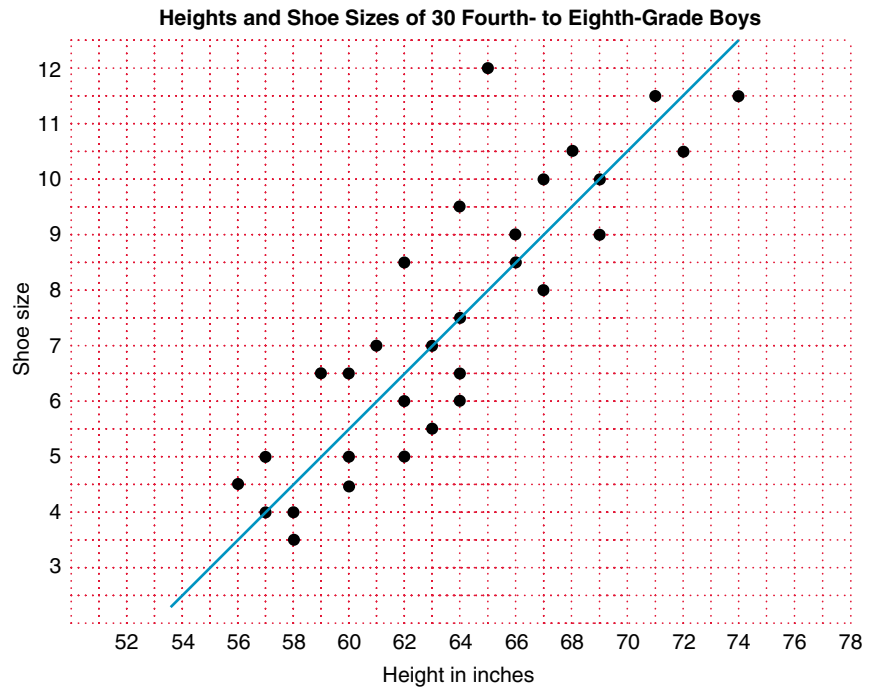
Figure 7.12

Trend Lines   A straight line can be drawn from the lower left to the upper right that approximates the points of the graph in Figure 7.12. Such a line is called a **trend line.** One method of locating a trend line is to place a line so that it approximates the location of the points and there are about the same number of points of the graph above the line as below.

EXAMPLE E

1. Draw a trend line for the scatter plot in Figure 7.12.

2. Use your trend line to predict the shoe size for a boy of height 68 inches and the height of a boy with a shoe size of 8.

Solution   1. While different people may select different locations for a trend line, these lines will be fairly close to the line shown on the scatter plot below.   **2.** Approximately 9.5; approximately 65 inches.

Heights and Shoe Sizes of 30 Fourth- to Eighth-Grade Boys

Some scatter plots, such as the one in Figure 7.13a, may show *no association* between the data. Or, if the trend line goes from lower left to upper right, as for Figure 7.13b, there is a positive association, slope of line is positive, and if the trend line goes from upper left to lower right, as for Figure 7.13c, there is a negative association, slope of line is negative (see pages 85–86 for positive and negative slopes). When data are entered into a graphing calculator or computer, the value of a variable r will be computed that indicates the strength of the association between the data. This number is called a **correlation coefficient** and it varies from $^-1$ to 1 ($^-1 \leq r \leq 1$). If r is close to 0, there is little or no association. If r is close to 1, there is a strong positive association; and if r is close to $^-1$, there is a strong negative association.

The scatter plot in Example E shows a positive association between the heights of the boys and their shoe sizes: as heights increase, shoe sizes increase. In Example F, there is a negative association between the two types of data.
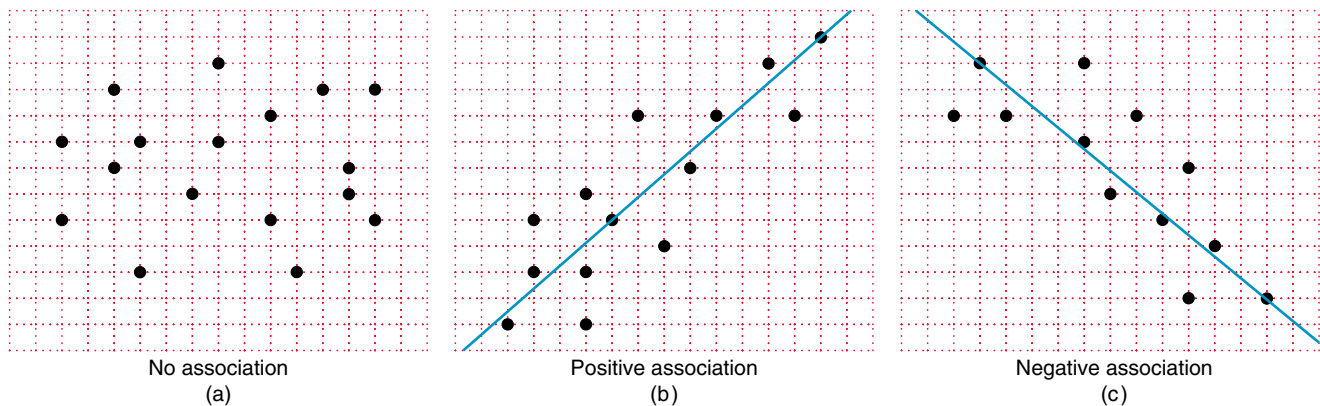

| No association | Positive association | Negative association |
| (a) | (b) | (c) |

Figure 7.13

EXAMPLE F

The following table contains data on one aspect of child development—the time required to hop a given distance.* The age of each child is rounded to the nearest half-year.

Age (years)	5	5	5.5	5.5	6	6	6.5	7	7	8	8	8.5	8.5	9	9	9.5	10	11
Time (seconds) to hop 50 feet	10.8	10.8	10.5	9.0	8.4	7.5	9.0	7.1	6.7	7.5	6.3	7.5	6.8	6.7	6.3	6.3	4.8	4.4

1. Form a scatter plot of these data. Mark intervals for ages on the horizontal axis and intervals for time on the vertical axis.

2. Locate a trend line.

3. Use your line to predict the time for a $7\frac{1}{2}$-year-old girl to hop 50 feet.

4. The negative (downward) slope of your line shows an association between the age of the girl and the time required to hop 50 feet. Describe this association.

Solution   1. The "2" in the following scatter plot indicates that two children of the same age required the same time to hop 50 feet.   2. Counting the point labelled "2" as two points, the trend line sketched below has 2 points on the line (marked by red dots), 8 points above the line, and 8 points below the line.   3. Approximately 8 seconds   4. The older a 5- to 11-year-old girl becomes, the less time is required to hop 50 feet.

Times of 18 Girls to Hop 50 Feet
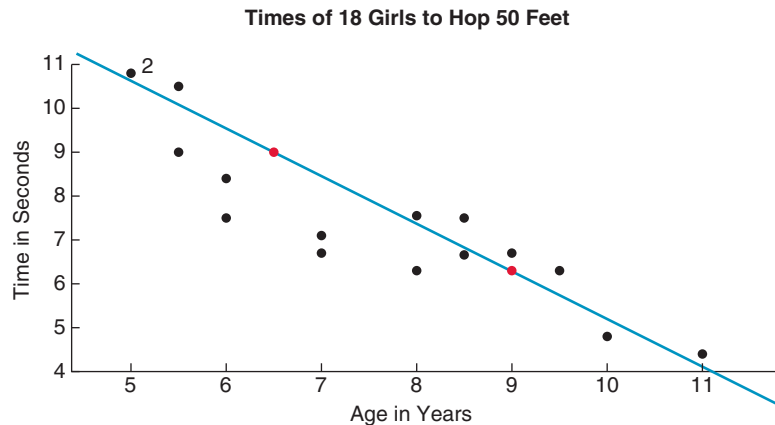
CURVES OF BEST FIT

Technology Connection

Most graphing calculators and some computer software such as Excel and Minitab have graphing features that include scatter plots and trend lines or curves of best fit. A graphing calculator screen is shown in Figure 7.14 with the scatter plot and trend line for the data in Example F.
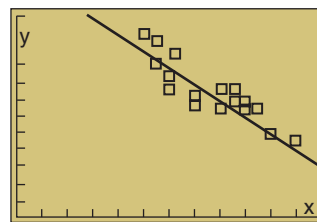

Figure 7.14

If you have a graphing calculator or a computer with suitable software, you may wish to enter the data from some of the preceding examples and obtain the trend lines. The

*Adapted from Kenneth S. Holt, *Child Development* (Boston: Butterworth-Heinmann, 1991), p. 143.

calculator or computer will print the equation of a trend line $y = ax + b$, where a is the slope of the line and b is the y intercept. You may find it interesting to use such equations to obtain new predictions. For example, the equation of the trend line in Figure 7.14 with the slope and y intercept rounded to the nearest hundredth is $y = {}^-.92x + 14.45$. Using this equation with $x = 7.5$, approximately how much time is required for a $7\frac{1}{2}$-year-old girl to hop 50 feet? Calculate this time and compare it to the time of 8 seconds obtained from the trend line in Example F.

Sometimes the *curve of best fit* for a scatter plot is not a straight line. Graphing calculators and some computer software have several types of curves of best fit. The equations for curves and trend lines are called **regression equations,** and such equations are algebraic models for approximating the location of points in a scatter plot. In addition to a straight line, the three types of curves shown in Figure 7.15 are common curves of best fit.
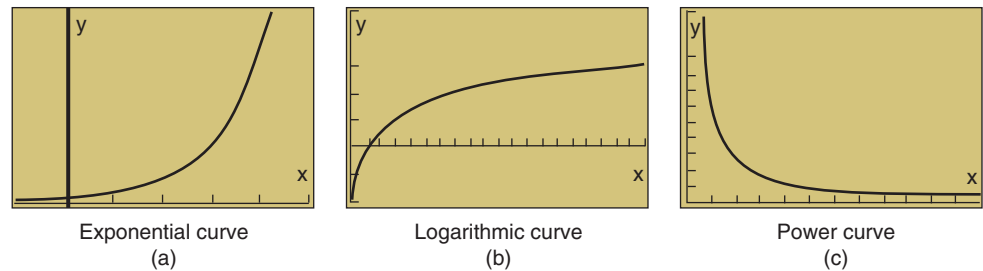

Exponential curve
(a)

Logarithmic curve
(b)

Power curve
(c)

Figure 7.15

A curve of best fit often can be obtained by visualizing a curve that approximates the location of points on a graph, as in Example G.

EXAMPLE G

The scatter plot shows the earnings of the top 32 women tennis players. Notice the notation which indicates that the earnings for the number 1 player ($652,000) and the number 2 player ($994,000) are off the graph.

1. Which type of curve from Figure 7.15 best fits the points of this graph?

2. Visualize this curve and use it to predict the earnings for the woman tennis player who ranks 35th.

Earnings of Top 32 Women Tennis Players
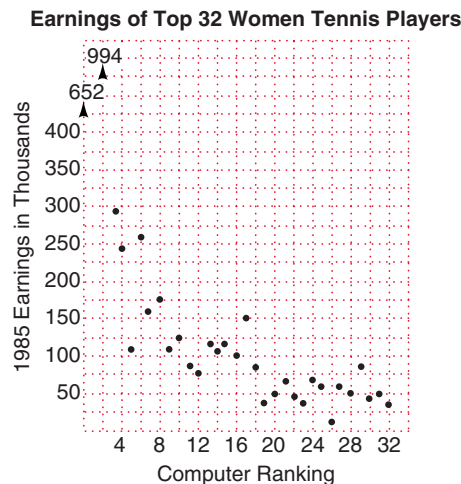
Solution   1. The power curve   **2.** Between $20,000 and $40,000

PROBLEM-SOLVING APPLICATION

A strong association between data does not necessarily imply that one type of measurement causes the other. Cigarette consumption and coronary heart disease mortality rates (see the problem-solving application that follows) are an example of a strong association between data that has generated debate over cause and effect between these measurements.

Problem

The table for the following problem lists the number of deaths for each 100,000 people aged 35 to 64 due to coronary heart disease and the average number of cigarettes consumed per adult per year for 21 countries.* Given this information, what is the number of deaths per 100,000 people due to coronary heart disease for a country that consumes 2000 cigarettes per adult per year?

Understanding the Problem   The cigarette consumption and mortality rates that are given for 21 countries in the table below and on the following page do not have a cigarette consumption of 2000 per adult per year, in the graph are given for 21 countries. The problem requires using the graph on page 452 to predict the mortality rate due to coronary heart disease for a country with a consumption of 2000 cigarettes per adult per year. **Question 1:** Which country in the table has a cigarette consumption that is closest to 2000?

Devising a Plan   One approach is to use the data in the table for West Germany and Finland, the two countries whose cigarette consumption is closest to 2000. **Question 2:** How can the data for these countries be used, and what approximation do you obtain by your method? Another approach is to use a scatter plot of the data in the table and a trend line to make a prediction.

Carrying Out the Plan   The scatter plot and trend line on the next page are for the data in the table. **Question 3:** Given this trend line, what is the coronary heart disease mortality rate for a country that consumes an average of 2000 cigarettes per adult per year?

Technology Connection

Trend Lines

Try to predict the curve of best fit for the points plotted below. By entering the coordinates of the points for this graph, (1, 1), (2, 2), (4, 1), (4, 2), etc., into a graphing calculator and using the statistical menu, you can obtain the curve of best fit and its correlation coefficient. This is one of the graphs you will explore using your calculator in this investigation.

Mathematics Investigation
Chapter 7, Section 1
www.mhhe.com/bennett-nelson
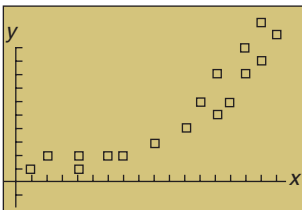
Country	Cigarette Consumption per Adult per Year	Mortality Rate per 100,000 People
Australia	3220	238
Austria	1770	182
Belgium	1700	118
Canada	3350	212
Denmark	1500	145
Finland	2160	233
France	1410	60
Greece	1800	41
Iceland	2770	111
Ireland	2770	187

(Continued)

*R. Mulcahy, J. W. McGiluary, and N. Hickey, "Cigarette Smoking Related to Geographic Variations in Coronary Heart Disease Mortality and to Expectation of Life in the Two Sexes," *American Journal of Public Health,* vol. 60, 1970.

Country	Cigarette Consumption per Adult per Year	Mortality Rate per 100,000 People
Italy	1510	114
Mexico	1680	32
Netherlands	1810	125
New Zealand	3220	212
Norway	1090	136
Spain	1200	44
Sweden	1270	127
Switzerland	2780	125
United Kingdom	2790	194
United States	3900	257
West Germany	1890	150

Cigarette Consumption and Mortality Rates for 21 Countries
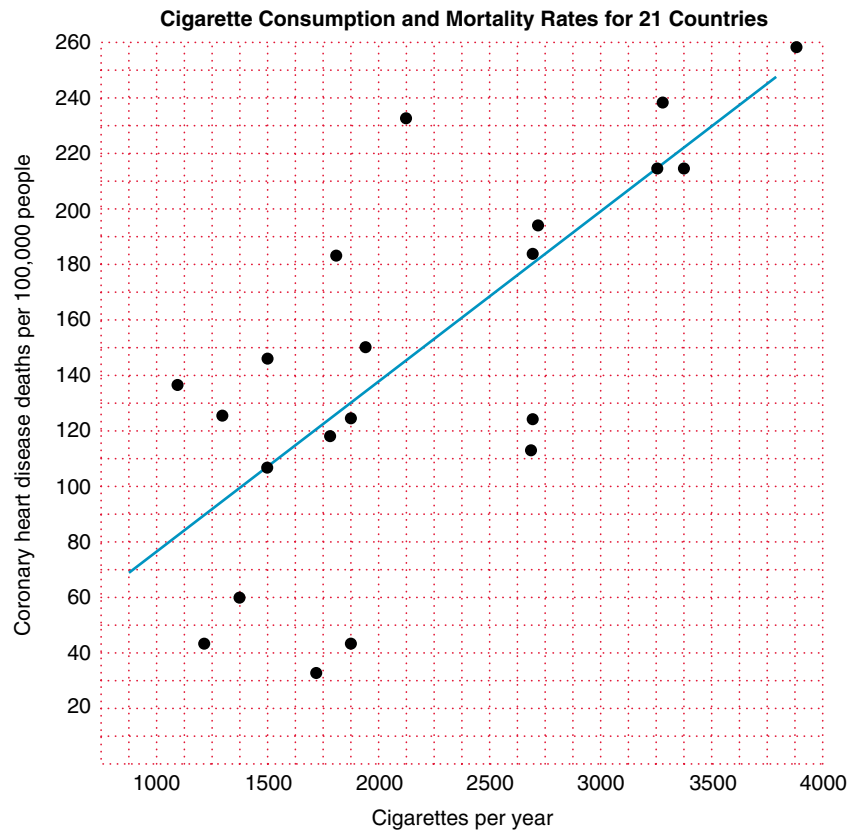
Looking Back   The trend line also enables predictions regarding the cigarette consumption for a country, if the coronary heart disease mortality rate is known. **Question 4:** What is the average number of cigarettes consumed per adult per year by a given country whose coronary heart disease mortality rate is 240 for every 100,000 people?

Answers to Questions 1–4   1. West Germany **2.** One possibility is 191.5, the mean of 150 and 233. Another possibility is 184, since 2000 is approximately 41 percent of the distance between 1890 and 2160 and 184 is approximately 41 percent of the distance between 150 and 233. **3.** Approximately 140 **4.** Approximately 3600*

Sometimes the *looking-back* part of solving a problem involves using a different approach. Figure 7.16 shows a computer printout of the scatter plot for the data in the preceding table using the software Minitab.[†] Notice that the trend line is in about the same position as the one shown in the preceding scatter plot. The equation for this line, $y = .06x + 15.64$, is printed on the screen by the computer. Using this equation, we can obtain another prediction of the number of coronary heart disease mortalities for a country that consumes an average of 2000 cigarettes per adult per year:

$$y = .06(2000) + 15.64 = 135.64$$

To the nearest whole number, this is 136 deaths for each 100,000 people. Compare this to the prediction we obtained from drawing the trend line in the preceding scatter plot.
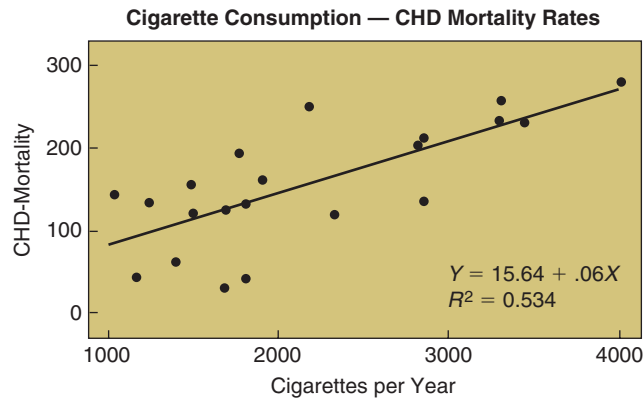

Cigarette Consumption — CHD Mortality Rates

$Y = 15.64 + .06X$
$R^2 = 0.534$

Figure 7.16

Summary   Bar and pie graphs, pictographs, line plots, stem-and-leaf plots, and histograms provide visual descriptions for interpreting data that involve **one variable,** that is, data with one type of measurement. For example, the pictograph in Figure 7.5 records the number of juveniles arrested for drug abuse, and there is one variable—*number of arrests.* Often we wish to compare two or more sets of data that involve one type of measurement (one variable) and double-bar or triple-bar graphs, and back-to-back stem-and-leaf plots are used for this purpose. As examples, in Figure 7.3 a triple-bar graph compares data on three sets of children, and in Figure 7.8 a back-to-back stem-and-leaf plot compares two sets of test scores.

A line graph and a scatter plot, on the other hand, provide a visual description of data that involve **two variables,** that is, data with two different types of measurement. As examples, the line graph in Figure 7.11 plots population for given years and the two variables are *numbers of people* and *years,* and in Figure 7.12 the heights and shoe sizes of boys were graphed on a scatter plot, and the two variables are *height* and *shoe size.*

Usually any one of several graphical methods can be chosen for one-variable sets of data, but there are some general guidelines. Bar graphs, pie graphs, and pictographs are best chosen when there are relatively small numbers of categories, such as 3 to 10. A histogram is often used for grouped data and 10 to 12 is a convenient number of groups.

Note: Entering the data from the table on cigarette consumption into a graphing calculator shows there is a positive correlation with $r \approx .71$ and the equation of the trend line is approximately $y = .06x + 20.4$.
[†] Minitab Release 11 (State College, Pennsylvania: Minitab Incorporated, 1996).

A line plot is used for plotting intermediate numbers of data, such as 25 to 50. Stem-and-leaf plots accommodate a greater number of data, such as 20 to 100. To compare two sets of data with a back-to-back stem-and-leaf plot, there should be approximately the same number of values on both sides of the stem. Line plots and stem-and-leaf plots have an advantage over bar graphs and histograms in showing individual values of data, gaps, clusters, and outliers.

Exercises and Problems 7.1
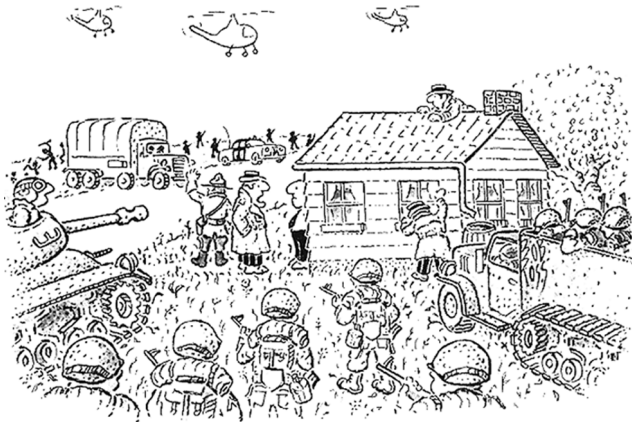
"We understand you tore the little tag off your mattress."

The numbers of troops (to the nearest 100,000) in active service in 2003 in the four major branches of the service are as follows: Army, 14; Navy, 4; Air Force, 4; and Marine Corps, 2.* Use this information in exercises 1 and 2.

1. Draw a bar graph for the data.[†]
   a. The number of people in the Army is how many times the number of people in the Marine Corps?
   b. How many more people are in the Navy than in the Marine Corps?

2. Draw a pie graph for the data.
   a. What is the measure of the central angle for each of the four regions of the pie graph?
   b. What is the total number of people in the four branches of the service?

The following graph shows the average annual prime rate of interest (to the nearest whole percent) charged by banks for each of the years from 1991 to 2003.* Use this graph in exercises 3 and 4.

Average Annual Prime Rate of Interest
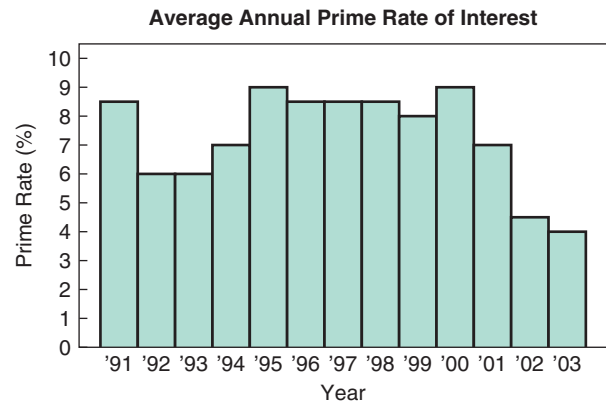
3. a. In which year was the prime rate the highest, and what was the rate?
   b. In which years did the prime rate increase, and how much was the increase?

4. a. In which year was the prime rate the lowest, and what was the rate?
   b. In which years did the prime rate decrease, and how much was the decrease?

5. A family's monthly budget is divided as follows: rent, 32 percent; food, 30 percent; utilities, 15 percent; insurance, 4 percent, medical, 5 percent; entertainment, 8 percent; other, 6 percent.
   a. Draw a pie graph of the data.
   b. What is the measure of the central angle (to the nearest degree) in each of the seven regions of the graph?

*Statistical Abstract of the United States, 124th ed. (Washington, DC: U.S. Department of Defense, 2004), p. 330.

[†]Copy the rectangular grid from the website for the bar graphs and histograms on these pages.

*Statistical Abstract of the United States, 124th ed. (Washington, DC: Bureau of the Census, 2004), p. 749.

The following graphs show the average monthly amounts of precipitation for Kansas City, Missouri, and Portland, Oregon.* Use these graphs in exercises 6 and 7.

Kansas City, Missouri

Portland, Oregon
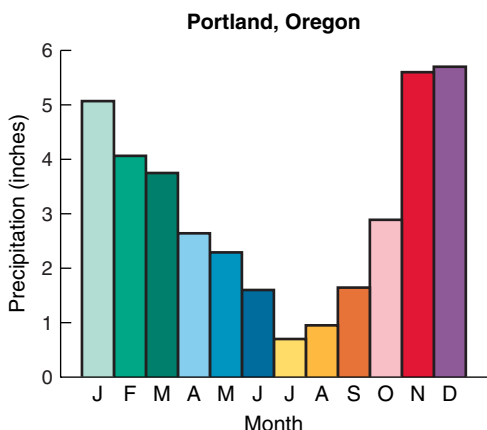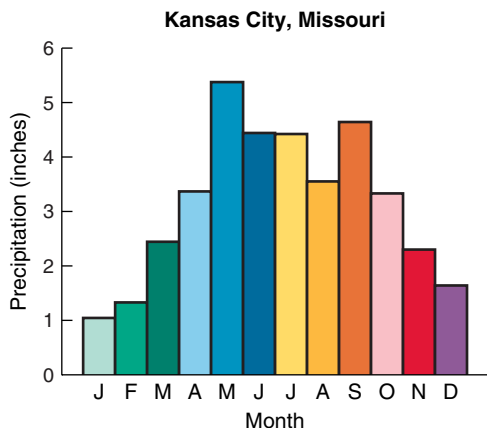
6. a. In Portland, which two months have the greatest amounts of precipitation?
   b. In Kansas City, which month has the least amount of precipitation?

7. a. Compare the amounts of precipitation during the summer months (June, July, and August). Which city has the most precipitation during the summer?
   b. Rounding the amount of precipitaion for each month to the nearest whole number, determine the approximate amount of precipitation for each city for the year.

Sometimes the bars on a graph are placed horizontally rather than vertically. The following double-bar graph compares the percentages of participation of males and females over 7 years of age in the 10 most popular sports activities.* Use this graph in exercises 8 and 9.

8. a. In which sports activities do the females have the greater percentage of participation?
   b. In which sports activities is the percentage of participation by females 4 times the percentage of participation by males?
   c. In which sports activities is the participation by males about 1 percent greater than the participation by females?
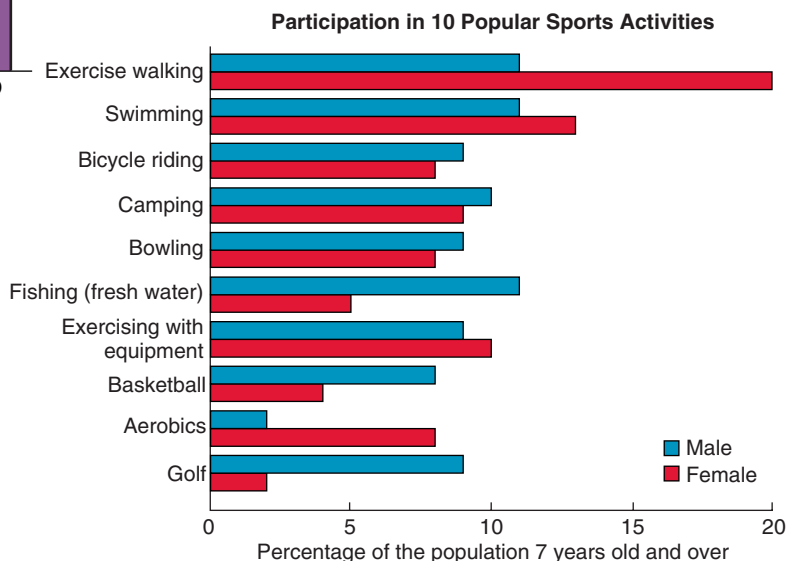
Participation in 10 Popular Sports Activities

9. **a.** In which sports activities do the males have the greater percentage of participation?
   b. In which sports activities is the percentage of participation by males more than 4 times the percentage of participation by females?
   c. In which sports activity is the percentage of participation by females almost twice the percentage of participation by males?

The following double-bar graph shows the percentage of women with early prenatal care by age and race.* Use this information in exercises 10 and 11.

10. **a.** Which age groups have a difference of over 10 percent between black women with prenatal care and white women with prenatal care?
    b. What percentage of white women of ages 15 to 19 does not have early prenatal care?
    c. The percentage of black women with prenatal care increases for the first few age groups and then decreases. For which age groups was this percentage decreased from the previous age group?

11. **a.** Which age group has the least difference between the percentage of black women with early prenatal care and the percentage of white women with early prenatal care?
    b. What percentage of the black women aged 15 to 19 does not have early prenatal care?
    c. The percentage of white women with early prenatal care increases for the first few age groups and then decreases. For which age groups was this percentage increased from the previous age group?

The following table shows the percentage of health-care coverage for children under 18 years of age and for children in poverty under 18 years of age.* Use this information in exercises 12 and 13.

	Children under Age 18	Children in Poverty under Age 18
No coverage	11.6%	22.0%
Public assistance	26.8%	62.1%
Private insurance	67.5%	22.0%

Note: Percentages add to more than 100 because some individuals receive coverage from more than one source.

12. **a.** Form a double-bar graph for the following three categories: no coverage, public assistance, and private insurance on the horizontal axis.
    b. What is the difference between the percentage of private insurance for all children under age 18 and the percentage of private insurance for children in poverty under age 18?

13. **a.** Form a double-bar graph for the following three categories: coverage, no public assistance, and no private insurance on the horizontal axis.
    b. What is the difference between the percentage of no public assistance for children in poverty under 18 years of age and the percentage of no public assistance for all children under age 18?

Mothers Beginning Prenatal Care in the First Trimester, by Age and Race
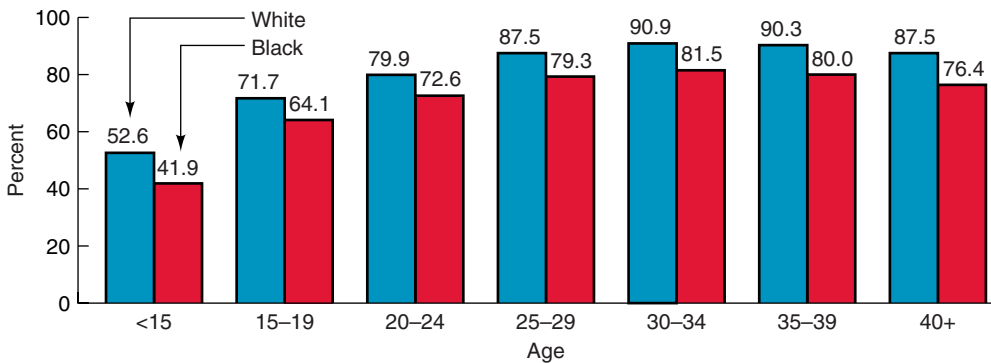
This pie graph represents the percentages of federal funds spent on programs for the handicapped in public schools. Use this graph in exercises 14 and 15.

(Region *A*, hearing-impaired, 1.3 percent; region *B*, orthopedically handicapped, 1.1 percent; region *C*, other health impaired, 1.3 percent; region *D*, visually handicapped, .7 percent; and region *E*, multihandicapped, 2.2 percent.)
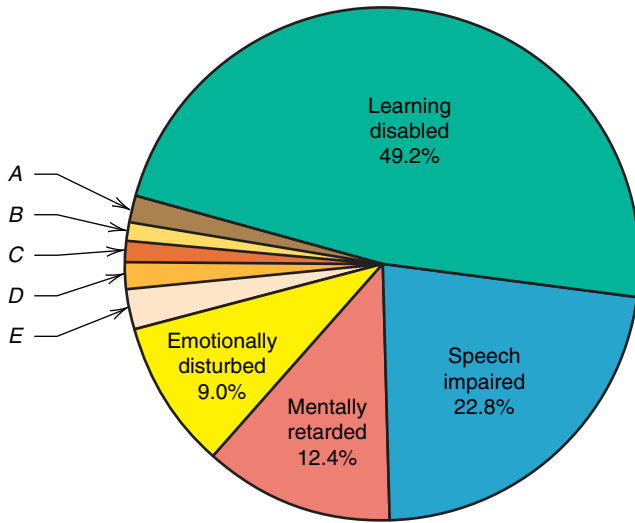
14. **a.** What was the total percentage spent on the learning-disabled and speech-impaired?
   b. The amount of money spent on programs for the speech-impaired was how many times to the nearest .1 the amount spent on programs for the emotionally disturbed?

15. **a.** What was the total percentage spent on the hearing-impaired and visually handicapped?
   b. The amount of money spent on programs for the mentally retarded was how many times to the nearest .1 the amount spent on programs for the hearing-impaired?

The following table shows the percentage to the nearest whole number of elementary schools, in the United States for various size categories.* Use this table in exercises 16 and 17.

Size of School (No. of Students)	Percentage of Schools
Under 100	6
100 to 199	9
200 to 299	13
300 to 399	16
400 to 499	16
500 to 599	13
600 to 699	9
700 to 799	6
800 to 899	7
900 to 999	0
1000 or more	5

16. **a.** Draw a pie graph of these data. Label the measure to the nearest degree of the central angle for each region of the graph.
   b. What percentage of elementary schools has 500 or more students?

17. **a.** Draw a bar graph of these data.
   b. What percentage of elementary schools has from 200 to less than 500 students?

18. The number of microcomputers in public schools for student instruction is contained in the following table.*

School Level	Number of Computers
Elementary	5,627,000
Middle/Junior High	2,468,000
Senior High	4,152,000
K–12 other	582,000

   a. Form a pictograph of the data in the table by choosing an icon and selecting the number of computers that the icon represents. Label the categories on the horizontal axis, and select an informative title for the graph.
   b. Discuss the reasons for your choice of value for the icon. In general, what disadvantages result if the value of the icon is too large or too small?

19. The following table shows the number of people to the nearest 1000 in different age categories who were involved in automobile crashes.*

Age Levels	Number of People
16 to 20	12,765,000
21 to 24	13,463,000
25 to 34	33,734,000
35 to 44	41,040,000
45 to 54	38,711,000
55 to 64	25,609,000
65 and older	27,930,000

a. Form a pictograph of the data in the table by choosing an icon and selecting the number of people that the icon represents. Label the categories on the horizontal axis, and select an informative title for the graph.

b. What does the pictograph show about the total number of people aged 16 to 24 years who had crashes compared to the number of people aged 25 to 34 years?

20. The numbers of students in thousands in public schools in grades K–8 are shown by states in the following table.[†]

State	Students (1000)	State	Students (1000)
Alabama	536	Florida	1797
Alaska	95	Georgia	1075
Arizona	672	Hawaii	132
Arkansas	318	Idaho	171
California	4480	Illinois	1484
Colorado	529	Indiana	711
Connecticut	410	Iowa	330
Delaware	81	Kansas	322
District of Columbia	58	Kentucky	473
		Louisiana	537

(Continued)

State	Students (1000)	State	Students (1000)
Maine	144	Ohio	1287
Maryland	611	Oklahoma	446
Massachusetts	699	Oregon	382
Michigan	1223	Pennsylvania	1255
Minnesota	573	Rhode Island	113
Mississippi	362	South Carolina	500
Missouri	643	South Dakota	87
Montana	103	Tennessee	675
Nebraska	195	Texas	3016
Nevada	262	Utah	338
New Hampshire	144	Vermont	69
New Jersey	972	Virginia	826
New Mexico	225	Washington	696
New York	2017	West Virginia	200
North Carolina	956	Wisconsin	592
North Dakota	70	Wyoming	59

a. Form a line plot for the numbers of students in each state by marking off the horizontal axis in intervals of 100 students, that is, 0–99, 100–199, 200–299, etc. (*Note:* To accommodate all the intervals, you may find it convenient to place breaks in the horizontal axis.)

b. What percentage of the states to the nearest whole percent have less than 700,000 students?

c. What percentage of the states to the nearest whole percent have more than 2,000,000 students?

21. The following table shows the average salaries of classroom teachers.*

Salaries (in thousands) of K–12 Classroom Teachers

State	Salary	State	Salary
Alabama	$38.2	California	$56.3
Alaska	49.7	Colorado	42.7
Arizona	40.9	Connecticut	55.4
Arkansas	37.8	Delaware	50.8

(Continued)

*Statistical Abstract of the United States, 124th ed. (Washington, DC: Bureau of the Census, 2004), p. 702.

[†]Statistical Abstract of the United States, 124th ed. (Washington, DC: Bureau of the Census, 2004), p. 152.

*Statistical Abstract of the United States, 124th ed. (Washington, DC: Bureau of the Census, 2004), p.155.

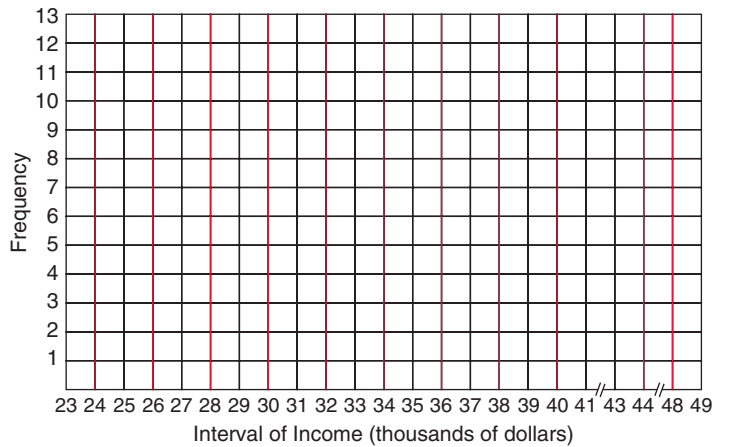| State	Salary	State	Salary
District of Columbia	$50.8	New Jersey	$54.2
Florida	40.3	New Mexico	37.0
Georgia	45.5	New York	53.0
Hawaii	44.5	North Carolina	43.1
Idaho	40.1	North Dakota	33.9
Illinois	51.5	Ohio	45.5
Indiana	45.0	Oklahoma	34.9
Iowa	39.1	Oregon	47.6
Kansas	37.8	Pennsylvania	51.4
Kentucky	39.0	Rhode Island	51.1
Louisiana	37.2	South Carolina	40.4
Maine	38.5	South Dakota	32.4
Maryland	49.7	Tennessee	39.7
Massachusetts	50.8	Texas	40.0
Michigan	53.6	Utah	38.3
Minnesota	44.7	Vermont	41.5
Mississippi	34.6	Virginia	43.2
Missouri	37.7	Washington	45.0
Montana	35.8	West Virginia	38.5
Nebraska	37.9	Wisconsin	42.8
Nevada	41.8	Wyoming	38.8
New Hampshire	41.9		

a. Form a line plot for the 51 teachers' salaries by using intervals of $1000 on the horizontal axis, that is, 0–999, 1000–1999, 2000–2999, etc. (Round up each tenth that is greater than or equal to .5.)

b. What is the interval with the most salaries represented in the line plot?

c. The average U.S. salary of the K–12 teachers is $45,900. What percentage of the salaries, to the nearest .1 percent, represented in the line plot is less than $46,000?

22. The following 40 scores are from a college mathematics test for elementary school teachers.

92, 75, 78, 90, 73, 67, 85, 80, 58, 87, 62, 74, 74, 76, 89, 95, 72, 86, 80, 57, 89, 97, 65, 77, 91, 83, 71, 75, 67, 68, 57, 86, 62, 65, 72, 75, 81, 72, 76, 69

a. Form a stem-and-leaf plot for these test scores.

b. How many scores are below 70?

c. What percentage of the scores is greater than or equal to 80?

23. The life spans in years of the 36 U.S. presidents from George Washington to Richard Nixon are listed here.

67, 90, 83, 85, 73, 80, 78, 79, 68, 71, 53, 65, 74, 64, 77, 56, 66, 63, 70, 49, 57, 71, 67, 58, 60, 72, 67, 57, 60, 90, 63, 88, 78, 46, 64, 81

a. Form a stem-and-leaf plot of these data.

b. What percentage, to the nearest .1 percent, of the 36 presidents lived 80 years or more?

c. What percentage, to the nearest .1 percent, of the 36 presidents did not live 60 years?

24. The following test scores are for two classes that took the same test. (The highest possible score on the test was 60.)

Class 1 (24 scores): 34, 44, 53, 57, 19, 50, 41, 56, 38, 27, 56, 49, 39, 24, 41, 50, 45, 47, 35, 51, 40, 44, 48, 43

Class 2 (25 scores): 51, 40, 45, 28, 44, 56, 31, 33, 41, 34, 34, 39, 50, 36, 37, 32, 50, 22, 35, 43, 40, 50, 45, 33, 48

a. Form a stem-and-leaf plot with one stem. Put the leaves for one class on the right side of the stem and the leaves for the other class on the left side. Record the leaves in increasing order.

b. Which class appears to have better performance? Support your answer.

25. The following data are the weights in kilograms of 53 third-graders.

19.3, 20.2, 22.3, 17.0, 23.8, 24.6, 20.5, 20.3, 21.8, 16.6, 23.4, 25.1, 20.1, 21.6, 22.5, 19.7, 19.0, 18.2, 20.6, 21.5, 27.7, 21.6, 21.0, 20.4, 18.2, 17.2, 20.0, 22.7, 23.1, 24.6, 18.1, 20.8, 24.6, 17.3, 19.9, 20.1, 22.0, 23.2, 18.6, 25.3, 19.7, 20.6, 21.4, 21.2, 23.0, 21.2, 19.8, 22.1, 23.0, 19.1, 25.0, 22.0, 24.2

a. Form a stem-and-leaf plot of these data, using 16 through 27 as stems and the tenths digits as the leaves. (*Note:* It is not necessary to write decimal points.)

b. Which stem value has the greatest number of leaves?

c. What are the highest and lowest weights?

26. The average annual per capita incomes by states are shown in the following table.*

Average Annual per Capita Income by States

State	Income	State	Income
AL	$26,338	MT	$ 25,920
AK	33,568	NE	30,758
AZ	26,838	NV	31,266
AR	24,289	NH	34,702
CA	33,749	NJ	40,427
CO	34,283	NM	25,541
CT	43,173	NY	36,574
DE	32,810	NC	28,235
DC	48,342	ND	29,204
FL	30,446	OH	29,944
GA	29,442	OK	26,656
HI	30,913	OR	29,340
ID	25,911	PA	31,998
IL	33,690	RI	31,916
IN	28,783	SC	26,132
IA	29,043	SD	29,234
KS	29,935	TN	28,455
KY	26,252	TX	29,372
LA	26,100	UT	24,977
ME	28,831	VT	30,740
MD	37,331	VA	33,671
MA	39,815	WA	33,332
MI	30,439	WV	24,379
MN	34,443	WI	30,898
MS	23,448	WY	32,808
MO	29,252		

a. Complete the frequency table by recording a tally mark for each income.

*Statistical Abstract of the United States, 124th ed. (Washington, DC: Bureau of the Census, 2004), p. 434.

Interval ($)	Tallies
20,000–20,999	_____
21,000–21,999	_____
22,000–22,999	_____
23,000–23,999	_____
24,000–24,999	_____
25,000–25,999	_____
26,000–26,999	_____
27,000–27,999	_____
28,000–28,999	_____
29,000–29,999	_____
30,000–30,999	_____
31,000–31,999	_____
32,000–32,999	_____
33,000–33,999	_____
34,000–34,999	_____
35,000–35,999	_____
36,000–36,999	_____
37,000–37,999	_____
38,000–38,999	_____
39,000–39,999	_____

b. Construct a histogram for the given intervals.

c. Which interval has the greatest frequency of incomes?

d. In how many states was the per capita income greater than $32,000?

**e.** In how many states was the per capita income less than $22,000?

27. The following table records the amounts of snowfall (to the nearest inch) for selected cities.*

City	Snowfall	City	Snowfall
Juneau	97	St. Louis	20
Denver	60	Great Falls	58
Hartford	49	Omaha	30
Wilmington	21	Reno	24
Washington	17	Concord	64
Boise	21	Atlantic City	16
Chicago	38	Albany	64
Peoria	25	Buffalo	94
Indianapolis	23	New York	28
Des Moines	33	Bismarck	44
Wichita	16	Cincinnati	24
Louisville	17	Cleveland	56
Portland, ME	71	Pittsburgh	44
Baltimore	21	Providence	36
Boston	42	Salt Lake City	59
Detroit	41	Burlington	78
Sault Ste. Marie	118	Seattle-Tacoma	11
Duluth	81	Spokane	49
Minneapolis	50	Charleston	34

a. Form a frequency table for the snowfall data, using the following intervals: 0–15; 16–30; 31–45; 46–60; 61–75; 76–90; 91–105; 106–120.
b. Draw a histogram for the snowfall data for the intervals in part a.
c. Which interval contains the greatest number of cities?
d. How many cities had snowfalls of more than 60 inches?

Vaccines for measles became available for use in 1963. The line graph below shows the approximate number of children with measles from 1984 to 1998.* Use this information in exercises 28 and 29.
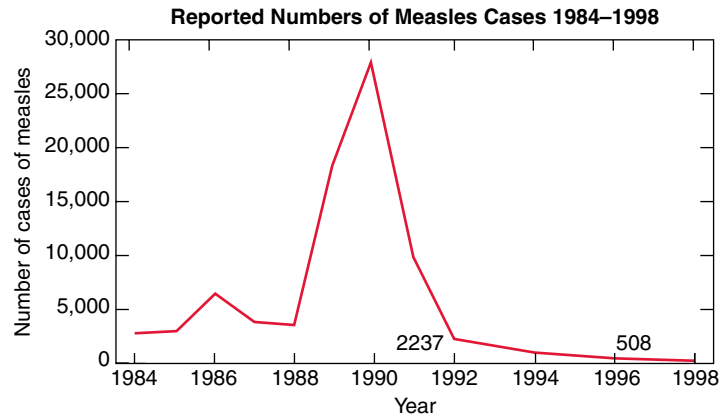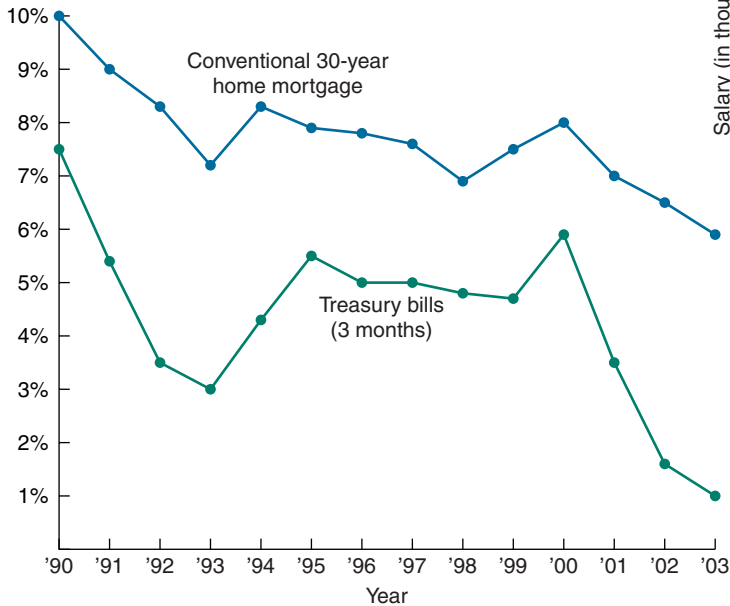

28. a. Was the number of cases of measles increasing or decreasing from 1984 to 1985?
   b. To the nearest 1000, how many more cases of measles were reported in 1988 than in 1998?
   c. In which year were the fewest cases of measles reported?
   d. There was a 2-year period in which the number of cases of measles reported rose sharply, due in part to the failure to immunize children at the recommended ages of 12 to 15 months. What were these two years, and what was the increase to the nearest 1000?
29. a. Was the number of cases of measles increasing or decreasing from 1986 to 1988?
   b. To the nearest 1000 how many more cases of measles were reported in 1991 than in 1992?
   c. In which year was the greatest number of cases of measles reported?
   d. In which 2-year period was the greatest decrease in the number of measles cases reported? What was this decrease to the nearest 1000?

The blue line in the next figure shows the average annual mortgage rates for new homes from 1990 to 2003, and the green line shows the average annual interest rates on Treasury bills during the same period.* Use these line graphs to determine approximate answers in exercises 30 and 31.

a. What will be the average salary for public school teachers in 2005 if the increase from 2000 to 2005 is the same as in the preceding five-year period?

b. Which 5-year period had the greater increase in salaries?

c. What is the total increase in salaries from 1980 to 2000?

30. a. What was the highest annual mortgage rate from 1990 to 2003 and in what year did it occur?

b. What was the annual mortgage rate for 2002?

c. What was the lowest annual interest rate for Treasury bills from 1990 to 2003?

d. What was the greatest difference between the annual mortgage rate and the Treasury bill interest rate from 1990 to 2003, and what was the year?

31. a. What was the lowest annual mortgage rate from 1990 to 2003, and in what year did it occur?

b. What was the annual mortgage rate for 1999?

c. What was the highest annual interest rate for Treasury bills from 1990 to 2003?

d. What was the smallest difference between the annual mortgage rate and the Treasury bill interest rate from 1990 to 2003, and what was the year?

32. This line graph shows the average salaries for public school teachers in five-year intervals from 1980 to 2000.[†]

The percentages of public elementary schools with Internet access grew rapidly from 1994 to 2002 are as follows: 1994, 30 percent; 1995, 46 percent; 1996, 61 percent; 1997, 75 percent; 1998, 88 percent; 1999, 95 percent; 2000, 98 percent; and 2002, 99 percent.* Use this information in exercises 33 and 34.

33. a. Draw a line graph with the years from 1994 to 2002 represented on the horizontal axis.

b. During which two-year period did the percentage of elementary schools with Internet access approximately double?

34. a. Draw a bar graph with the years from 1994 to 2002 represented on the horizontal axis.

b. During which group of years did the percentage of elementary schools with Internet access approximately triple?

Inheritance factors in physical growth have been studied to compare the mother's height to the daughter's and son's heights and the father's height to the daughter's and son's heights. Some researchers have found that the **midparent height,** which is the number halfway between the height of

*Statistical Abstract of the United States, 124th ed. (Washington, DC: Bureau of the Census, 2004), p. 749.

†Statistical Abstract of the United States, 124th ed. (Washington, DC: Bureau of the Census, 2004), p. 155.

*Statistical Abstract of the United States, 124th ed. (Washington, DC: Bureau of the Census, 2004), p. 159.

each parent, is more closely related to the heights of their children. The scatter plots in exercises 35 and 36 compare midparent heights to the daughters' heights and midparent heights to the sons' heights.*

35. a. Locate a trend line for the following scatter plot. Briefly explain your method of determining this line.
  b. Use your line to predict the heights of daughters for midparent heights of 160 and 174 centimeters.
  c. Use your line to predict the midparent heights for daughters' heights of 163 and 170 centimeters.
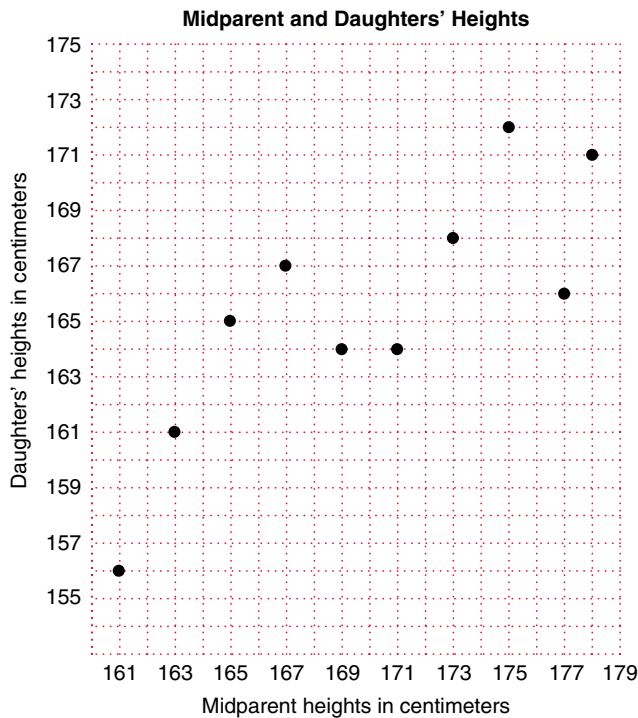
36. a. Locate a trend line for the scatter plot shown below. Briefly explain your method of determining this line.
  b. Use your line to predict the sons' heights for the midparent heights of 170 and 180 centimeters.
  c. Use your line to predict the midparent heights for the sons' heights of 179 and 182 centimeters.
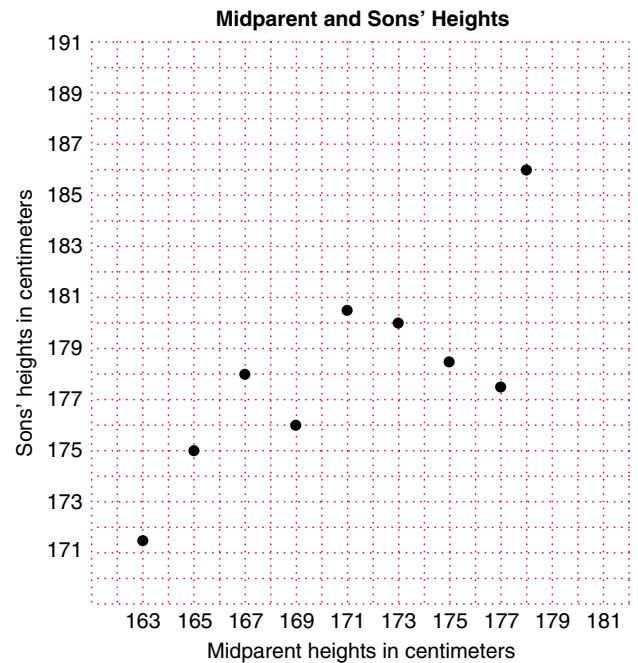
Midparent and Daughters' Heights

Midparent and Sons' Heights

The following table contains the percentages of adolescents (aged 18 to 24) who have not completed high school and are not enrolled.*

## Percentages of Adolescent High School Dropouts

Year	White (Non-Hispanic)	Black (Non-Hispanic)	Hispanic Origin
1975	13.9	27.3	34.9
1980	14.4	23.5	40.3
1985	13.5	17.6	31.5
1990	13.5	15.1	37.3
1995	13.6	14.4	34.7
1996	12.5	16	34.5
1997	12.4	16.7	30.6
1998	13.7	17.1	34.4
1999	12.8	16	33.9
2000	12.2	15.3	32.3
2001	13.4	13.8	31.7
2002	12.2	14.6	30.1

37. a. What was the percentage decrease in black non-Hispanic school dropouts from 1975 to 2002?

b. The percentage decrease of black non-Hispanic dropouts from 1975 to 2002 was how many times the percentage decrease of Hispanic dropouts (to the nearest whole number)?

c. Form a scatter plot to compare the Hispanic dropouts to the black non-Hispanic dropouts by forming intervals from 30 to 42 percent on the vertical axis for white non-Hispanic and from 12 to 28 percent on the horizontal axis for black non-Hispanic. Is there a positive or negative association?

d. Locate a trend line for your scatter plot, and use it to predict the black non-Hispanic dropout percentage for an Hispanic dropout rate of 36 percent.

38. a. What was the percentage decrease in white non-Hispanic dropouts from 1975 to 2002?

b. The percentage decrease of black non-Hispanic dropouts from 1975 to 2002 was how many times the percentage decrease of white non-Hispanic dropouts (to the nearest whole number)?

c. Form a scatter plot to compare the black non-Hispanic dropouts to the white non-Hispanic dropouts by forming intervals from 12 to 28 percent on the horizontal axis for black non-Hispanic dropouts and from 12 to 14.5 percent on the vertical axis for white non-Hispanic dropouts. Is there a positive or negative association?

d. Locate a trend line for your scatter plot, and use it to predict the white non-Hispanic dropout percentage to the nearest percent for a black non-Hispanic dropout rate of 17 percent.

39. This scatter plot shows the ages of 27 trees and their corresponding diameters.*
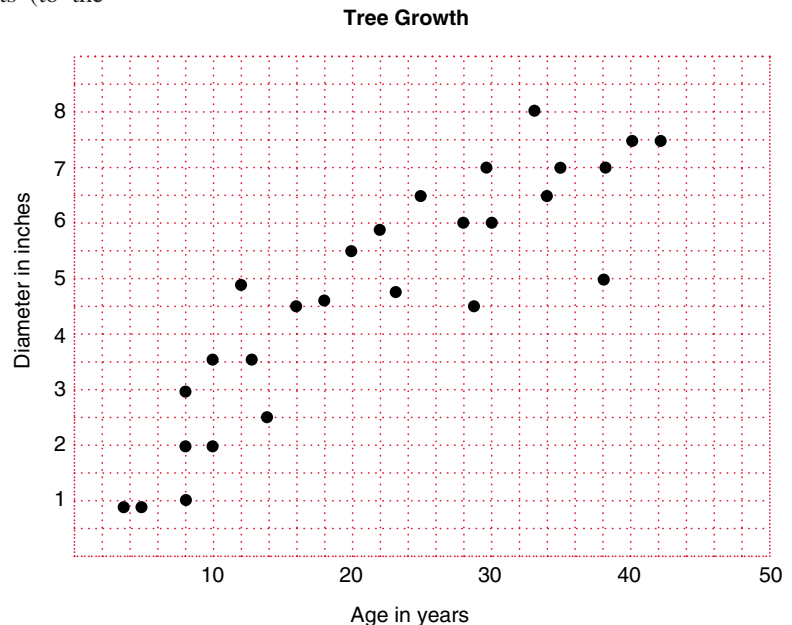
Tree Growth

a. What are the greatest diameter and the oldest age of the trees represented in this graph?

b. Is there a positive or negative association?

c. Locate a trend line and use it to predict the diameters of a 26-year-old tree and a 32-year-old tree.

d. Use your trend line to predict the approximate age of a tree, if its diameter is 9 inches.

Reasoning and Problem Solving

40. Featured Strategy: Drawing a Graph   Two ardent baseball fans were comparing the numbers of home runs hit by the American and National Leagues' home run leaders and posed the following question: Is there an association from year to year between the numbers of these runs? That is, in general, if the number of home runs hit by one league's home run leader for a given year is low (or high), will the number of home runs hit by the other league's home run leader be low (or high)? Use the tables on pages 484 and 485.

a. **Understanding the Problem**   Consider the two years with the smallest numbers of home runs by the leaders and the two years with the largest numbers of home runs by the leaders. What were these years and numbers?

b. **Devising a Plan**   One possibility for considering an association between the data is to form a scatter plot. To form a scatter plot, first mark off axes for numbers of home runs by each league's home run leaders. For each league, what is the difference between the smallest number of home runs by the leaders and the greatest number?

c. **Carrying Out the Plan**   Form a scatter plot by plotting the number of home runs hit by each pair of leaders for each year to see if there appears to be an association between the data. If so, is the association positive or negative? Are there points of the plot that might be considered outliers? Use your graph to determine the year with the greatest difference in the number of home runs hit by each league's home run leaders.

d. **Looking Back**   Draw a trend line and use your line to predict the number of home runs by the National League's home run leader for a given year if the number of home runs by the American League's leader is 54.

41. A company's record of amounts invested in advertisements and the corresponding amounts of sales produced are listed by the following pairs of numbers. The first number is the amount for advertisements to the nearest tenth of a million dollars, and the second number is the amount of sales to the nearest million dollars.

(3.8, 17), (1.4, 3), (2.8, 7), (4.9, 26), (2.3, 5), (1.8, 3), (3.3, 10), (5.3, 39), (4.4, 23), (5.1, 31), (2.6, 6), (1.2, 2)

a. Form a scatter plot with the amounts for advertisements on the horizontal axis and the corresponding amounts for sales on the vertical axis.

b. Which type of curve from Figure 7.15 best fits the points of the scatter plot?

c. Sketch the curve of best fit from part b and use your curve to predict the amount of sales for $3.6 million in advertisements.

d. Use your sketch in part c to predict the amount invested in advertisements, if the total resulting sales was $20 million.

42. For a math project one middle school student recorded the number of words that her friend Amy was able to memorize in different amounts of time. In the following pairs of numbers, the first number is the amount of time in minutes, and the second is the number of words memorized for the given time: (.5, 5), (1, 9), (1.5, 11), (2, 12), (2.5, 13), (3, 14), (3.5, 15), (4, 15), (4.5, 13), (5.5, 16), (5.5, 18), (6, 17), (7, 18), (8, 18)

a. Form a scatter plot for this data with the times on the horizontal axis and the corresponding numbers of words on the vertical axis.

b. Which type of curve from Figure 7.15 best fits the points of the scatter plot?

c. Sketch the type of curve from part b to approximate the location of points on the scatter plot and use your curve to predict the number of words that Amy could memorize in 9 minutes.

d. Use your curve in part c to predict the time period, if Amy memorized 10 words.

Writing and Discussion

1. The Spotlight on Teaching at the beginning of this chapter shows a bar graph and a line plot involving 21 students. Suppose this were your class. Explain how you would have your class transfer the bar graph data to a line plot. Then describe the different information each graph displays.

2. The scatter plot on page 447 compares the heights of boys to their shoe sizes. Use this data to form stem-and-leaf plots for these two types of measurements. How would you label each plot and explain what information each plot contains?

3. What is the purpose of the "zigzag" line on the vertical axis of the graph on the Elementary School Text example on page 441? Explain how you would introduce

this idea to a fourth-grade class by listing a series of questions that would lead them to understand and be able to scale similar data.

4. In an elementary school class that was graphing data obtained from student surveys, a student wanted to know if a histogram was the same as a bar graph. How would you answer this question?

Making Connections

1. Eight different ways of graphing data are presented in Section 7.1. Refer to the *Data Analysis and Probability* **Standards for PreK–2, Grades 3–5, and 6–8** (see inside covers) to record the different graphical representations that are recommended and their grade levels.

2. The **Research** Statement on page 439 notes two of the difficulties fourth-graders have with bar graphs. Read this statement, and then use the bar graph on that page as an example to write a question involving the interpretation of the data and another question involving predictions.

3. The **Standards** quote on page 447 mentions three types of relationships that can exist for data when forming a scatter plot. Create an example for each of these types of relationships that would be meaningful to school students.

4. The **PreK–2 Standards,** *Data Analysis and Probability,* states that students should be able to represent data using concrete objects as well as pictures and graphs. Describe three examples of data that students at this level can collect and represent using concrete objects.

5. On page 441 the example from the **Elementary School Text** uses a line graph to represent the data given in its table. Describe and illustrate another way to represent the data or explain why the line graph is the only effective way to represent the data.