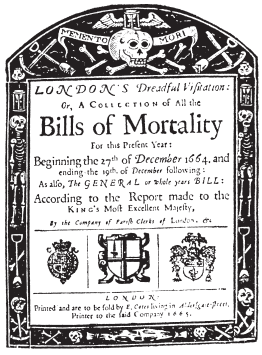
SECTION 7.1

Collecting and Graphing Data

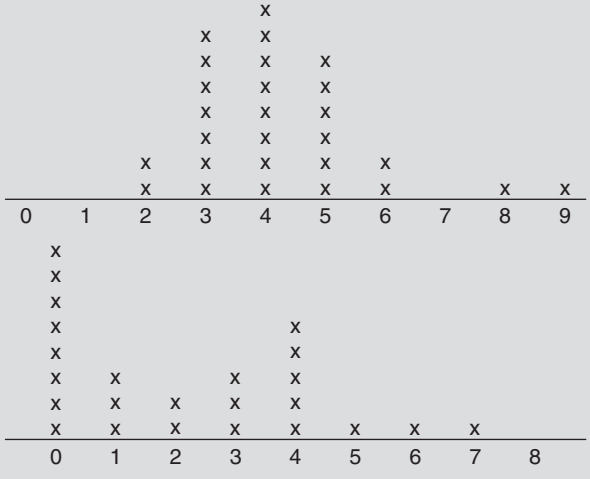
HISTORICAL HIGHLIGHT


LONDON'S Dreadful Visitation:
Or, A Collection of All the
Bills of Mortality
For this Present Year:
Beginning the 27th of December 1664, and ending the 19th of December following:
As also, The GENERAL or Whole years BILL:
According to the Report made to the King's Most Excellent Majesty,
By the Company of Parish Clerks of London, &c.
LONDON:
Printed and are to be sold by E. Cotes living in Aldersgate-Street, Printer to the said Company 1665.

Statistics had its beginning in the seventeenth century in the work of Englishman John Graunt. Graunt 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.* In his work he summarized great amounts of information to make it understandable (descriptive statistics) and made conjectures about large populations based on small samples (inferential statistics).

PROBLEM OPENER

Two line plots are shown below 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?*



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

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

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).

By the end of the second grade, students should be able to organize and display their data through both graphical displays and numerical summaries. They should use counts, tallies, tables, bar graphs, and line plots

Standards, 2000, p. 109

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

Research Statement

Fourth-grade students were successful at literal reading of bar graphs (over 95% success rate), they were less successful at interpreting (52% success rate) and predicting (less than 20% success rate).

Pereira-Mendoza and Mellor 1991

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.

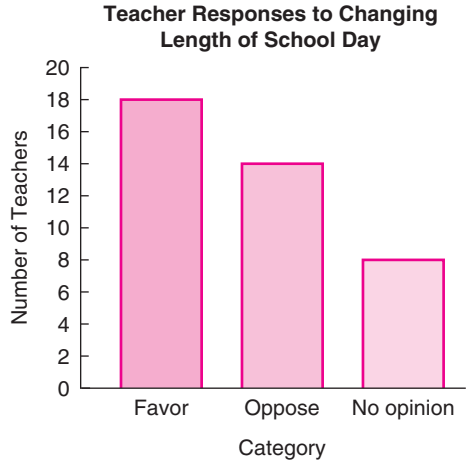


Figure 7.2

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 which 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.*

*U.S. Department of Health and Human Services, *Child Health USA 1998* (Washington, DC: U.S. Government Printing Office, 1998), p. 53.

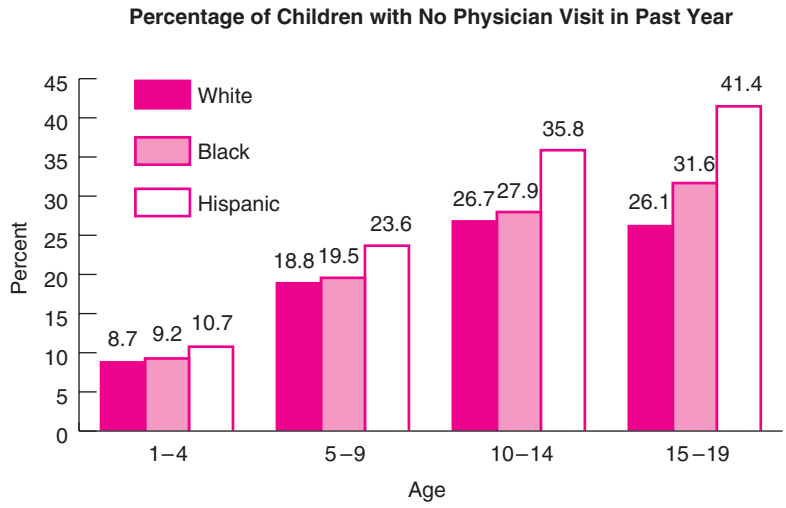


Figure 7.3

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 19 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. 15.3 percent 2. 72.1 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.

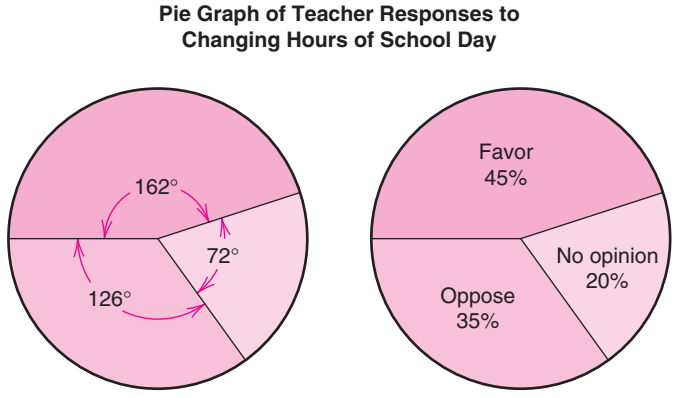


Figure 7.4

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 following pictograph 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.

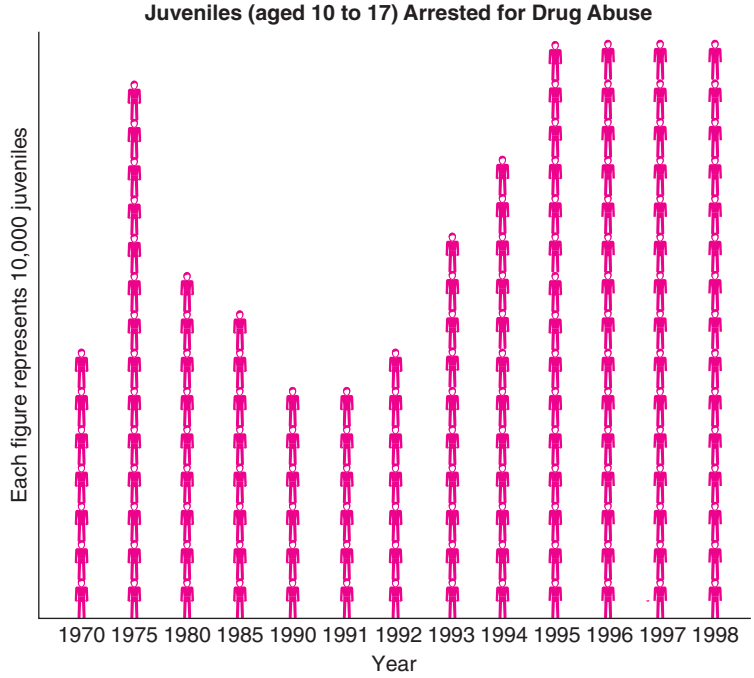


Figure 7.5

**Statistical Abstracts of the United States,* 120th ed. (Washington, DC: Bureau of the Census, 2000), p. 211.

3·1

Objective: Collect, organize, and display data. Make and interpret frequency tables, line plots, and pictographs.

Collect, Organize, and Display Data

Learn

Math Words

data
survey
frequency table
frequency
line plot
gap
cluster
pictograph
key

Do you watch TV on the weekends? On Monday, Carmen asked everyone in her class, "How many hours did you watch TV yesterday?" She recorded their answers in the table.

Carmen collected her data by conducting a survey. **Data** is collected information. A **survey** is a way to gather information by asking questions or observing events. Carmen used a frequency table to organize her data. A **frequency table** shows the number of times each item or number appears. In Carmen's survey, the **frequency** is the number of students who gave each answer.

Hours Spent Watching TV on November 11

Number of Hours	Tally	Number of Students
0	|||	3
1	||||	4
2	𝍸 |||	8
3	𝍸	5
4		0
5 or more	|||	3

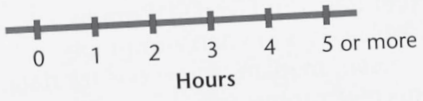
Example 1

You can organize Carmen's data in a line plot. A **line plot** is a vertical graph that shows data in columns of Xs above a number line.

1

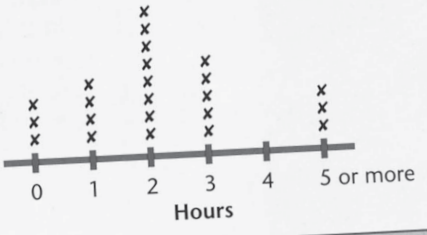
Draw a number line.

Under the number line, write the numbers for the number of hours.



2

Use an X to represent one student. Give the line plot a title.



The **gap**, or empty space, at 4 means none of the students in Carmen's class watched TV for 4 hours. The data form a cluster around 2 hours. A **cluster** is data that are close together in value.

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 5-year period from 1990 through 1994 or the 4-year period from 1995 through 1998?

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

Line Plots

The following table 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.

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
Columbia	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

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.

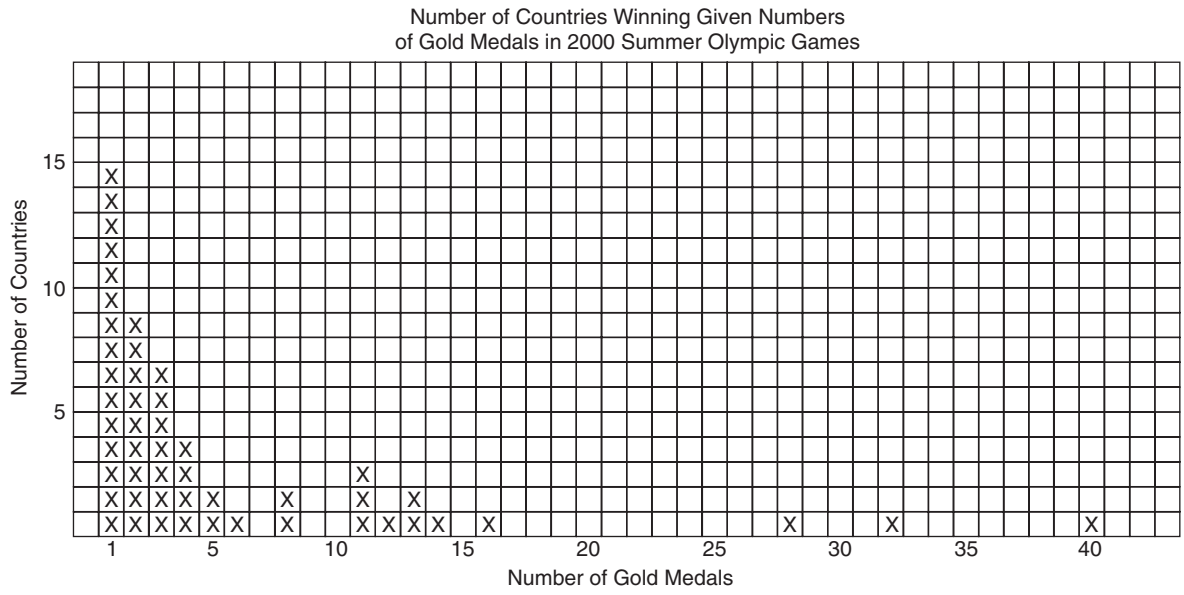


Figure 7.6

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).

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:

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

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
4	
5	
6	
7	
8	
9	

(a)

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

(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, clusters, or outliers 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 on previous page obtains the following scores on a second test:

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

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 427, 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 which 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.

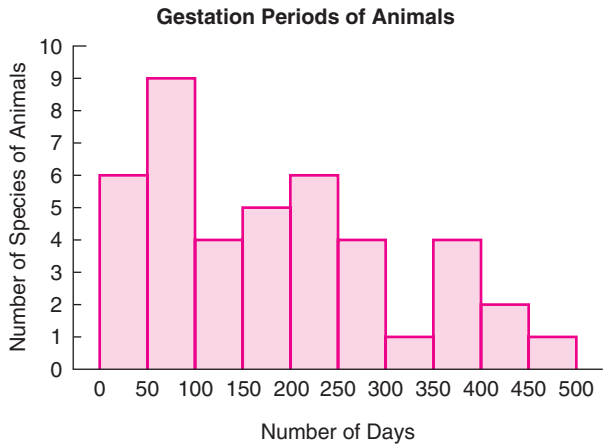


Figure 7.10

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.

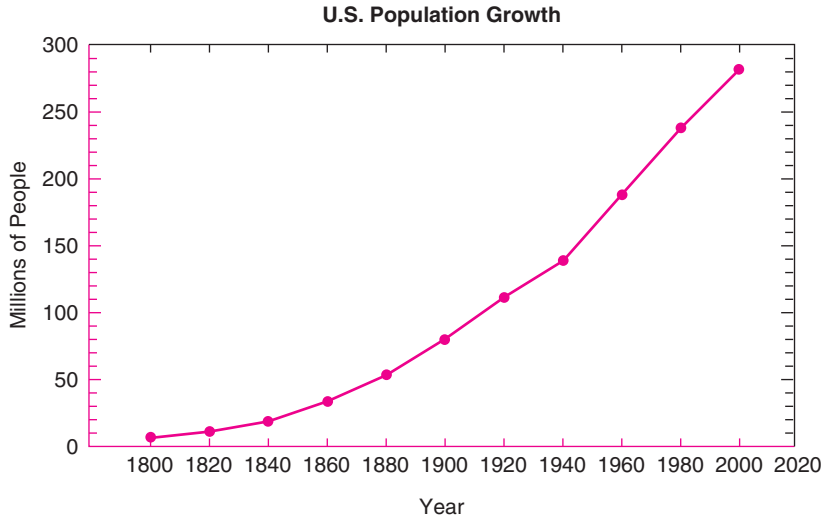


Figure 7.11

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

64	65	62	56	69	61	58	62	63	67	69	64	68	60	58	66	57
6.5	12	6	4.5	9	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.

Students should see a range of examples in which plotting data suggests linear relationships, nonlinear relationships, and no apparent relationship at all. When a scatter plot suggests that a relationship exists, teachers should help students determine the nature of the relationship from the shape and direction of the plot.

Standards 2000, p. 253

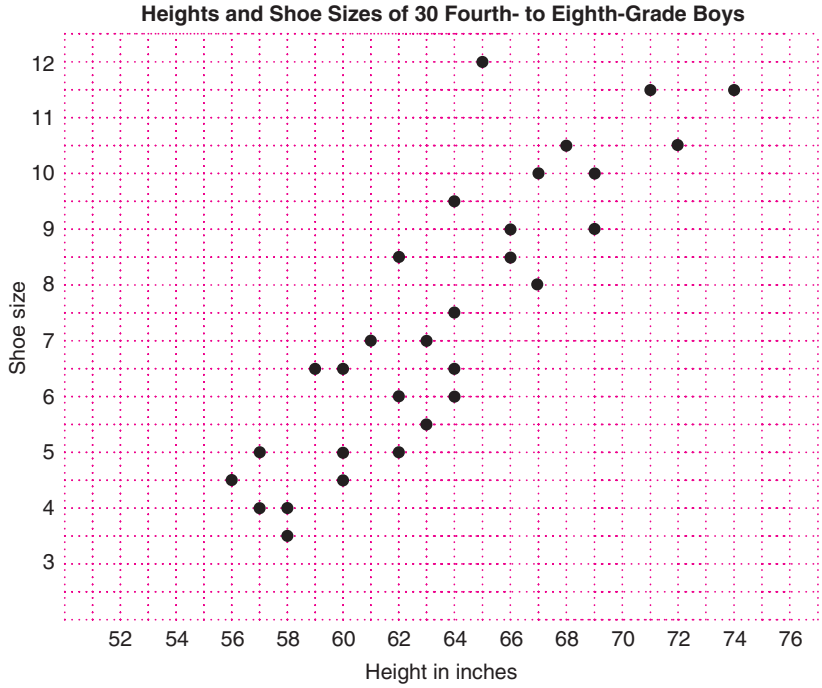


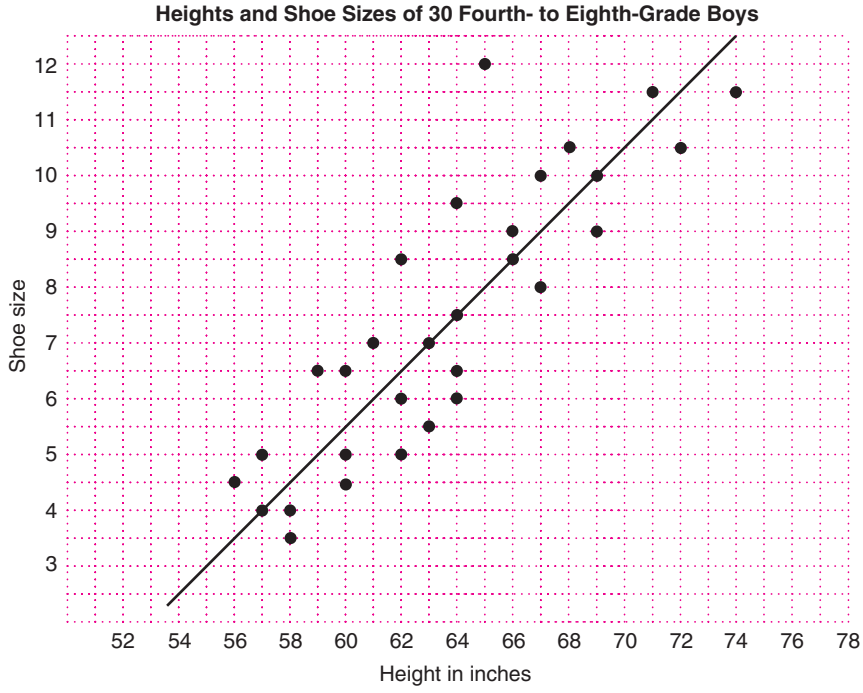
Figure 7.12

TREND LINES A straight line can be drawn from the lower left to the upper right which 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.



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). When data are entered into a graphing calculator or computer, the value of a variable r will be computed which indicates the strength of the association between the data. This number is called a **correlation coefficient** and it varies from $^{-}1$ to 1 ($^{-}1 \le r \le 1$). 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.

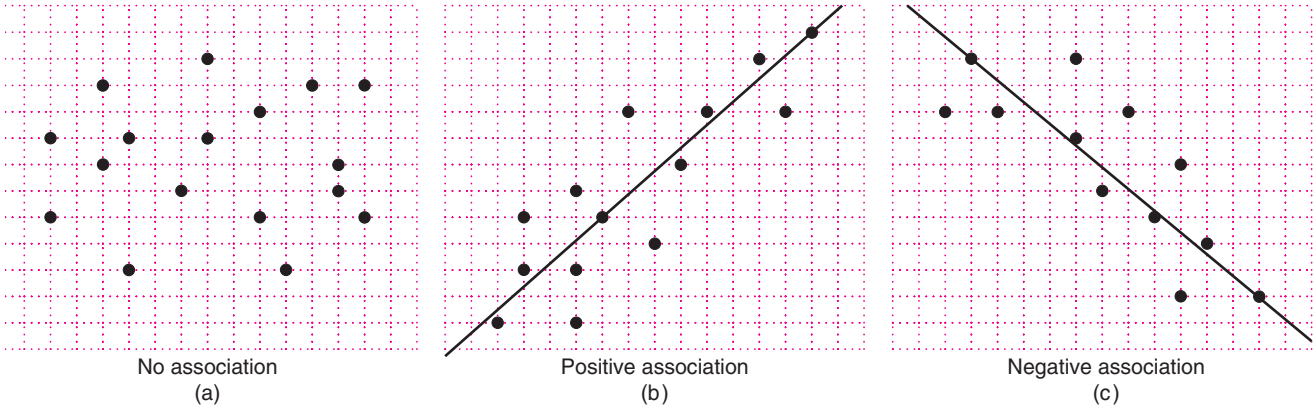


Figure 7.13

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.

Example F

The table below 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 of 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.** The trend line sketched below has 2 points on the line and both 8 points above and 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.

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

Curves of Best Fit

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.

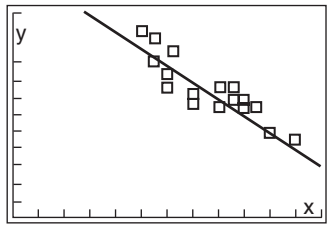


Figure 7.14

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 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.

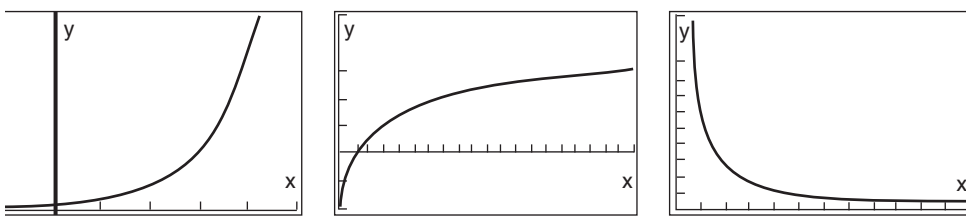


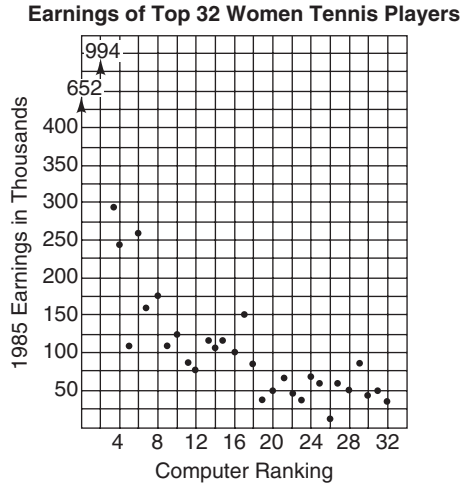
Figure 7.15

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

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.



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 below) are an example of a strong association between data that has generated debate over cause and effect between these measurements.

PROBLEM

The table on the next page 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 the country that consumes 2000 cigarettes per adult per year?

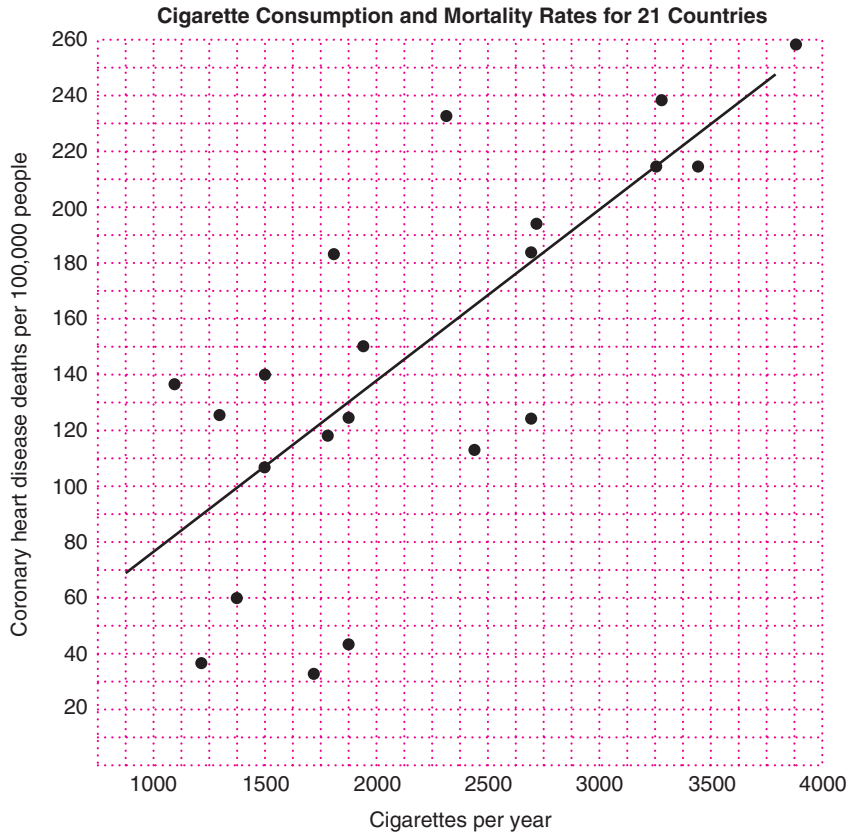
*R. Mulcahy, J. W. McGiluary, and N. Hickey, "Cigarette Smoking Related to Geographic Variations in Coronary Heart Disease Mortality and to Expectation of Life in the Two Sexes," *American Journal of Public Health,* vol. 60, 1970.

Understanding the Problem The cigarette consumption and mortality rates on the next page are given for 21 countries. The problem requires predicting 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?

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
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



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 above 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 cigaretetes per adult per year:

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

**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).

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.

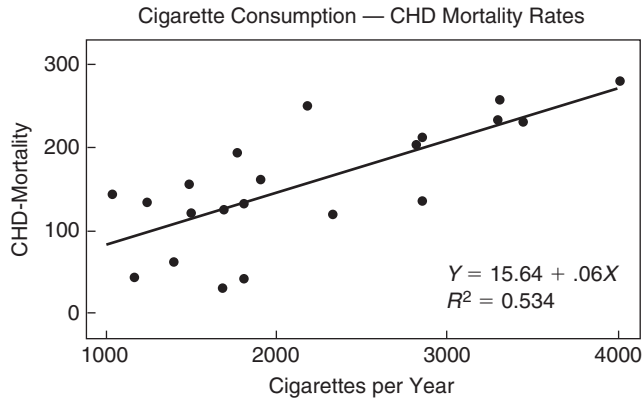


Figure 7.16

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. 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.

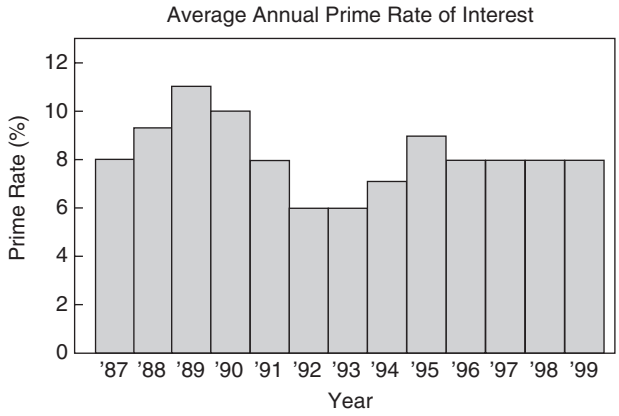
EXERCISES AND PROBLEMS 7.1

"We understand you tore the little tag off your mattress."

The numbers of troops (to the nearest 100,000) in active service in 1997 in the four major branches of the service are as follows: Army, 5; 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 1987 to 1999.** Use this graph in exercises 3 and 4.

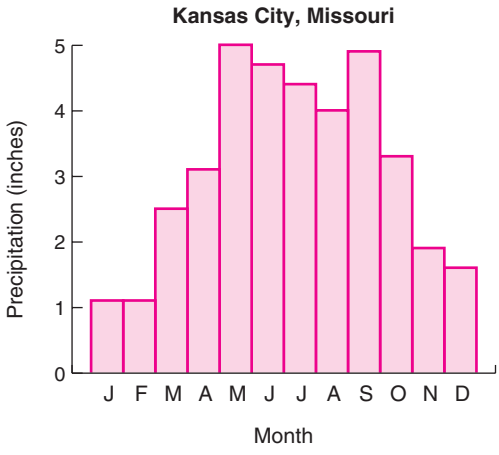


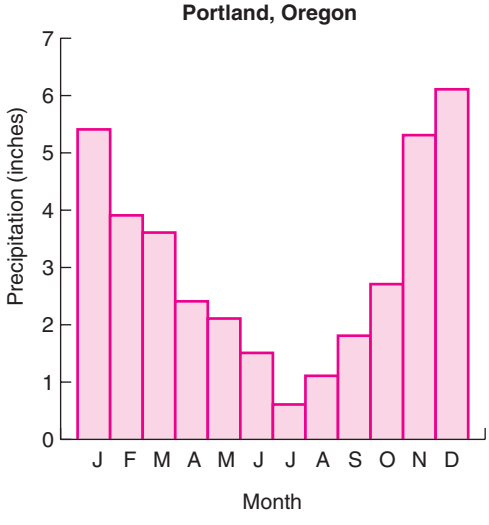
3. a. In which year was the prime rate the highest, and what was the rate?
   b. In which years did the prime rate increase, and how much was the increase?
4. a. In which year was the prime rate the lowest, and what was the rate?
   b. In which years did the prime rate decrease, and how much was the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

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

**Statistical Abstract of the United States,* 120th ed. (Washington, DC: U.S. Department of Defense, 2000), p. 366.
†Copy the rectangular grid from the inside cover of the book or from the website for the bar graphs and histograms on these pages.
***Statistical Abstract of the United States,* 118th ed. (Washington, DC: Bureau of the Census, 1998), p. 526.

**Statistical Abstract of the United States,* 118th ed. (Washington, DC: Bureau of the Census, 1998), p. 253.





6. a. In Portland, which two months have the greatest amounts of precipitation?

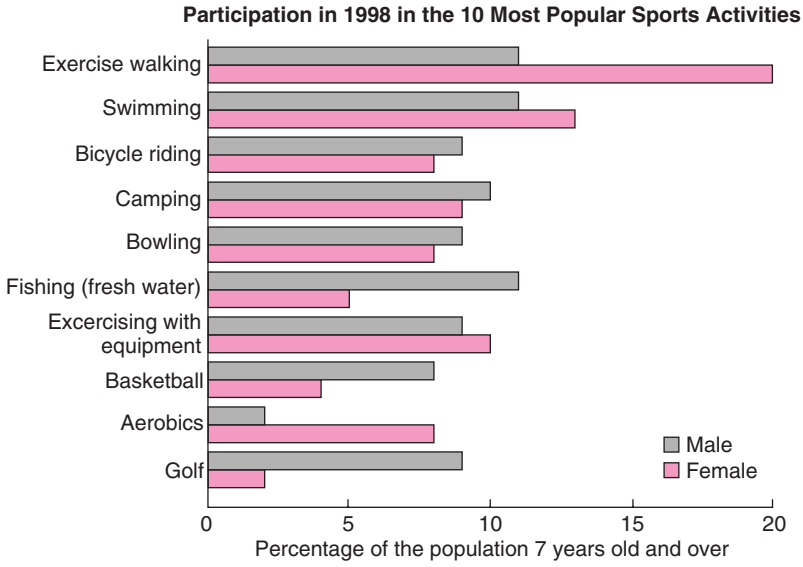
b. In Kansas City, which month has the least amount of precipitation?

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 double-bar graph on the next page 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.

**Statistical Abstract of the United States,* 120th ed. (Washington, DC: Bureau of the Census, 2000), p. 262.



8. a. In which sports activities do the females have the greater percentage of participation?

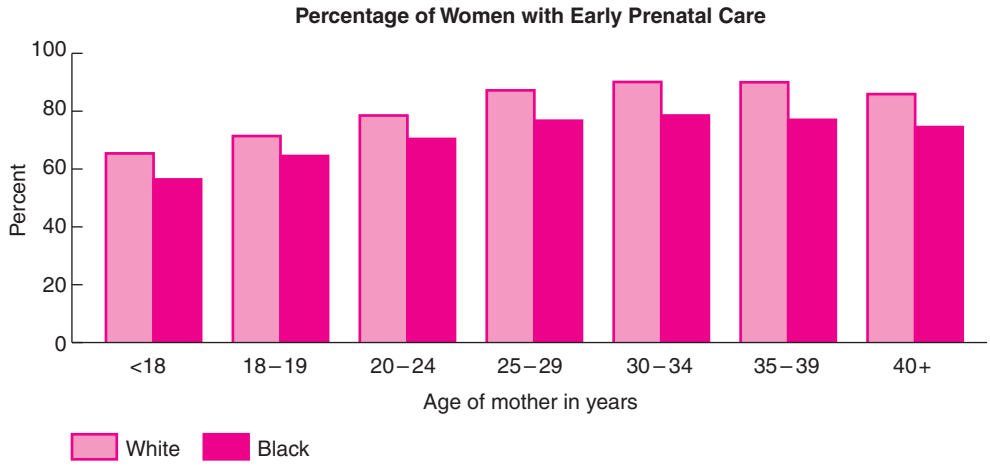
b. In which sports activities is the percentage of participation by females 4 times the percentage of participation by males?

c. In which sports activities is the participation by males about 1 percent greater than the participation by females?

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?



This double-bar graph shows the percentage of women with early prenatal care by age and race for 1998.* 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 18 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 18 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	14.8%	24.0%
Public assistance	24.9%	63.4%
Private insurance	66.3%	18.8%

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

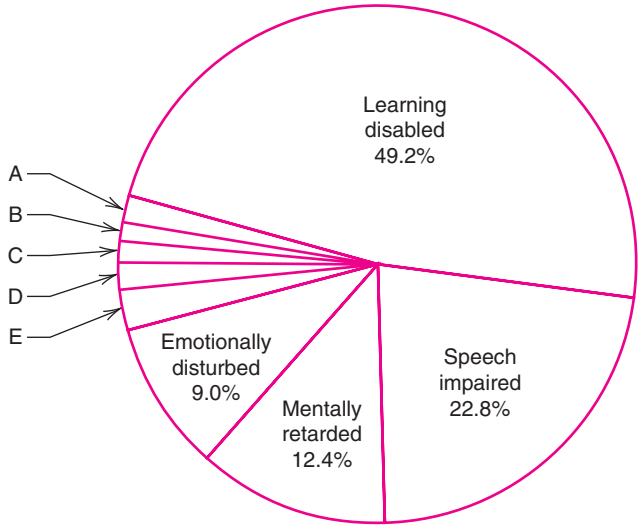
12. a. Form a double-bar graph for the following three categories: no coverage, public assistance, and private insurance on the horizontal axis.
b. What is the difference between the percentage of private insurance for all children under age 18 and the percentage of private insurance for children in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

*U.S. Department of Health and Human Services, *Child Health USA, 1998* (Washington, DC: U.S. Government Printing Office, 1998), p. 58.
†U.S. Department of Health and Human Services, *Child Health USA, 1998* (Washington, DC: U.S. Government Printing Office, 1998), p. 49.

The pie graph below 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 in 1997–1998 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	12
300 to 399	15
400 to 499	16
500 to 599	14
600 to 699	10
700 to 799	7
800 to 899	3
900 to 999	3
1000 or more	4

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 in 1999–2000 is contained in the following table.*

SCHOOL LEVEL	NUMBER OF COMPUTERS
Elementary	4,384,000
Middle/Junior High	1,783,000
Senior High	3,047,000
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?

**Statistical Abstract of the United States,* 118th ed. (Washington, DC: Bureau of the Census, 1998), p. 172.

**Statistical Abstract of the United States,* 120th ed. (Washington, DC: Bureau of the Census, 2000), p. 173.

19. The following table shows the number of people to the nearest 1000 in different age categories who were involved in automobile crashes, as reported by the police in 1998.*

AGE LEVELS	NUMBER OF PEOPLE
16 to 20	13,001,000
21 to 24	12,481,000
25 to 34	37,265,000
35 to 44	41,857,000
45 to 54	33,662,000
55 to 64	21,337,000
65 and older	25,814,000

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	542	Missouri	651
Alaska	97	Montana	110
Arizona	623	Nebraska	200
Arkansas	319	Nevada	229
California	4270	New Hampshire	147
Colorado	501	New Jersey	936
Connecticut	399	New Mexico	232
Delaware	80	New York	2028
District of Columbia	57	North Carolina	921
Florida	1704	North Dakota	77
Georgia	1029	Ohio	1301
Hawaii	135	Oklahoma	448
Idaho	169	Oregon	380
Illinois	1452	Pennsylvania	1267
Indiana	693	Rhode Island	112
Iowa	337	South Carolina	478
Kansas	327	South Dakota	91
Kentucky	465	Tennessee	665
Louisiana	558	Texas	2868
Maine	151	Utah	329
Maryland	602	Vermont	73
Massachusetts	705	Virginia	815
Michigan	245	Washington	696
Minnesota	586	West Virginia	206
Mississippi	365	Wisconsin	601
		Wyoming	64

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

b. What is the interval that contains the median of the numbers represented in the line plot?

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

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

**Statistical Abstract of the United States,* 120th ed. (Washington, DC: Bureau of the Census, 2000), p. 638.
†*Statistical Abstract of the United States,* 120th ed. (Washington, DC: Bureau of the Census, 2000), p. 167.

21. The following table shows the average salaries of classroom teachers in 1999–2000.*

Salaries (in thousands) of K–12 classroom teachers, 1999–2000

STATE	AVERAGE SALARY	STATE	AVERAGE SALARY
Alabama	$35.8	Montana	$31.4
Alaska	46.8	Nebraska	32.9
Arizona	35.0	Nevada	38.9
Arkansas	32.4	New Hampshire	37.4
California	45.4	New Jersey	51.2
Colorado	38.0	New Mexico	32.4
Connecticut	51.6	New York	49.4
Delaware	43.2	North Carolina	36.1
District of Columbia	47.2	North Dakota	29.0
Florida	35.9	Ohio	40.6
Georgia	39.7	Oklahoma	31.1
Hawaii	40.4	Oregon	42.8
Idaho	34.1	Pennsylvania	48.5
Illinois	45.6	Rhode Island	45.7
Indiana	41.2	South Carolina	34.5
Iowa	34.9	South Dakota	28.6
Kansas	37.4	Tennessee	36.5
Kentucky	35.5	Texas	35.0
Louisiana	32.5	Utah	33.0
Maine	34.9	Vermont	36.8
Maryland	42.5	Virginia	37.5
Massachusetts	45.1	Washington	38.7
Michigan	48.2	West Virginia	34.2
Minnesota	39.5	Wisconsin	40.7
Mississippi	29.5	Wyoming	33.5
Missouri	34.7		

a. Form a line plot for the 51 teachers' salaries by using intervals of $1000 on the horizontal axis. (Round up each tenth which 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 $40,600. What percentage of the salaries, to the nearest .1 percent, represented in the line plot is less than $41,000?

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 below.

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 for 1999 are shown in the following table.*

**Statistical Abstract of the United States,* 120th ed. (Washington, DC: Bureau of the Census, 2000), p.169.

**Statistical Abstract of the United States,* 120th ed. (Washington, DC: Bureau of the Census, 2000), p. 460.

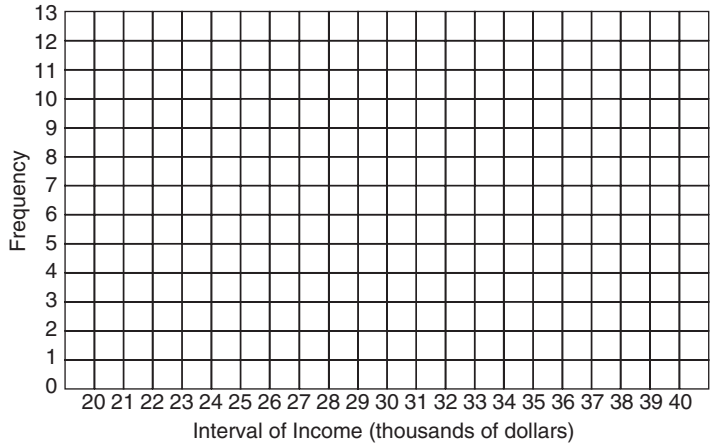
Average annual per capita income by states for 1999

STATE	INCOME	STATE	INCOME
AL	$22,946	MT	$22,314
AK	28,523	NE	27,437
AZ	25,307	NV	30,351
AR	22,114	NH	30,905
CA	29,819	NJ	36,106
CO	31,678	NM	22,063
CT	39,167	NY	33,946
DE	30,685	NC	26,220
DC	28,228	ND	23,518
FL	28,023	OH	27,081
GA	27,198	OK	22,801
HI	27,842	OR	27,135
ID	23,445	PA	28,676
IL	31,278	RI	29,720
IN	26,092	SC	23,496
IA	25,727	SD	25,107
KS	26,633	TN	25,581
KY	23,161	TX	26,525
LA	22,792	UT	23,356
ME	24,960	VT	25,892
MD	32,166	VA	29,484
MA	35,733	WA	30,295
MI	27,844	WV	20,888
MN	30,622	WI	27,412
MS	20,506	WY	26,003
MO	26,187		

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

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 capital 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 1998 for selected cities.*

**Statistical Abstract of the United States,* 120th ed. (Washington, DC: Bureau of the Census, 2000), p. 249.

CITY	SNOWFALL	CITY	SNOWFALL
Juneau	99	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	91
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

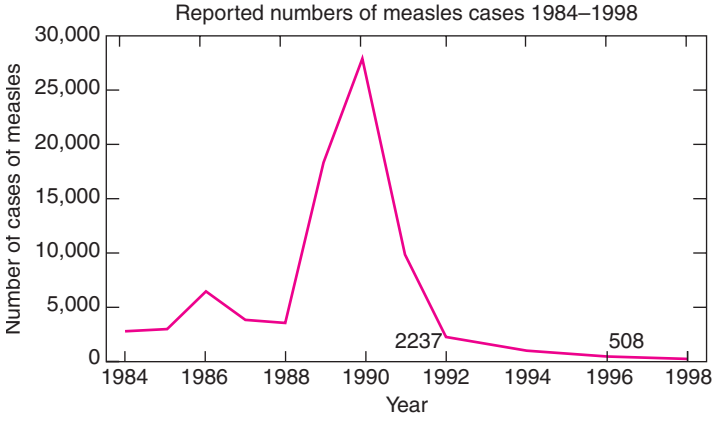
a. Form a frequency table for the snowfall data, using the following intervals: 0–15; 16–30; 31–45; 46–60; 61–75; 76–90; 91–105; 106–120.

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.



*U.S. Department of Health and Human Services, *Child Health USA 1998* (Washington, DC: U.S. Government Printing Office, 1998), p. 50.

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?

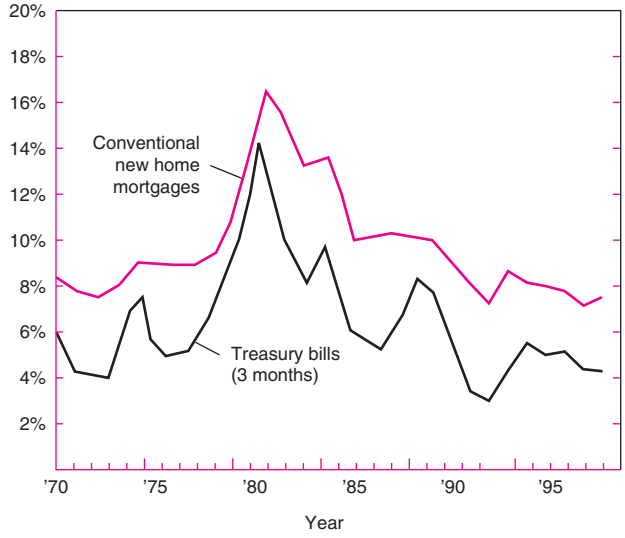
29. a. Was the number of cases of measles increasing or decreasing from 1986 to 1988?

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 red line in the next figure shows the average annual mortgage rates for new homes from 1970 to 1999, and the black 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.

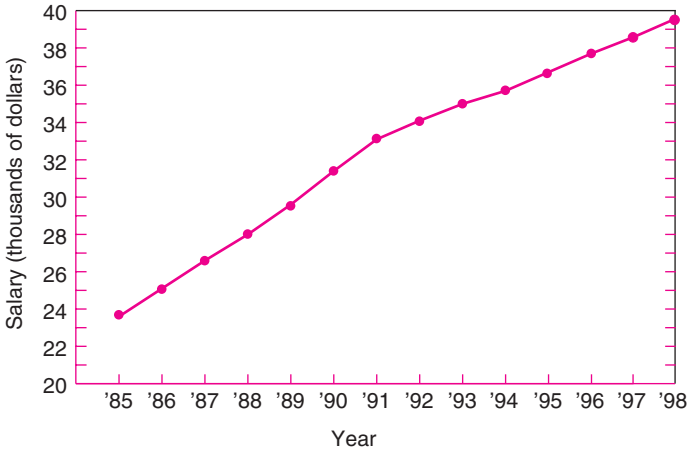


Statistical Abstract of the United States, 120th ed. (Washington, DC: Bureau of the Census, 2000), p. 521.

30. **a.** What was the highest annual mortgage rate between 1970 and 1999 and in what year did it occur?
b. What was the annual mortgage rate for 1999?
c. What was the lowest annual interest rate for Treasury bills between 1970 and 1999?
d. What was the greatest difference between the annual mortgage rate and the Treasury bill interest rate from 1970 to 1999, and what was the year?

31. **a.** What was the lowest annual mortgage rate between 1970 and 1999, and in what years did it occur?
b. What was the annual mortgage rate for 1980?
c. What was the highest annual interest rate for Treasury bills between 1970 and 1999?
d. What was the smallest difference between the annual mortgage rate and the Treasury bill interest rate from 1970 to 1999, and what was the year?

32. This line graph shows the average salaries for public school teachers between 1985 and 1999.*



a. What is the approximate average salary for public school teachers in 1999 if the increase from 1998 to 1999 is the same as in the preceding year?
b. Which 5-year period had the greater increase in salaries, 1985–1990 or 1993–1998?
c. What is the total increase in salaries from 1985 to 1998?

Statistical Abstract of the United States, 120th ed. (Washington, DC: Bureau of the Census, 2000), p. 169.

The percentages of public elementary schools with Internet access for instruction from 1994 to 1998 are as follows: 1994, 30 percent; 1995, 46 percent; 1996, 61 percent; 1997, 75 percent; and 1998, 88 percent.* Use this information in exercises 33 and 34.

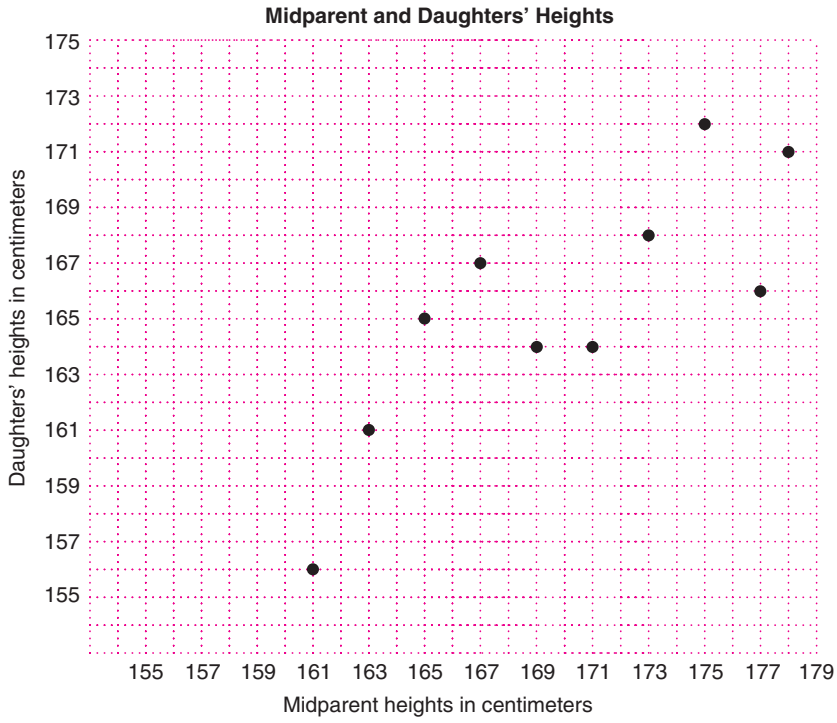
33. **a.** Draw a line graph with the years from 1994 to 1998 represented on the horizontal axis.
b. During which two-year period did the percentage of elementary schools with Internet access approximately double?

34. **a.** Draw a bar graph with the years from 1994 to 1998 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 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.

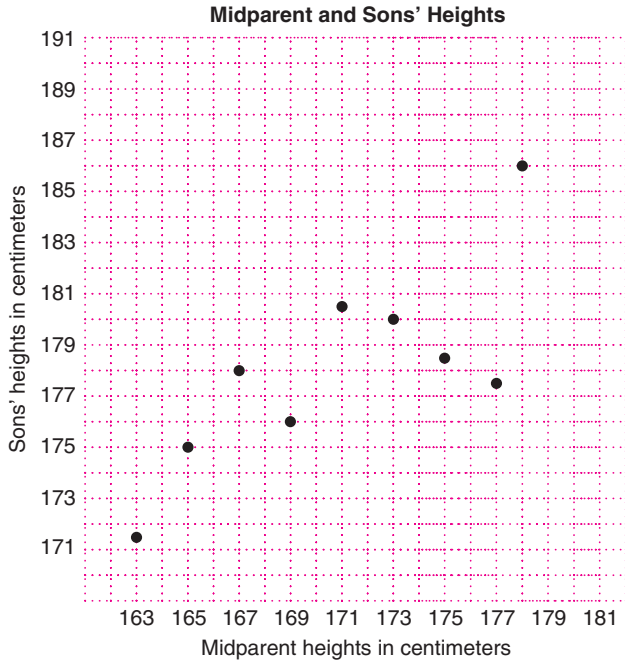
Statistical Abstract of the United States, 118th ed. (Washington, DC: Bureau of the Census, 1998), p. 179.
†W. M. Krogman, *Child Growth* (Ann Arbor: The University of Michigan Press, 1972), p. 157.



36. **a.** Locate a trend line for the scatter plot at the right. 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.



As of October 1998, approximately 479,000 youths ages 16–24 had dropped out of high school in the previous 12 months. Those most likely to drop out of school in 1998 were those living in western states, boys, students aged 19 and older, and students living in low-income families. The table below contains the percentages of adolescents (aged 16 to 24) who were high school dropouts for the odd-numbered years 1971 to 1999.* Use this information in exercises 37 and 38.

Percentages of adolescent high school dropouts

YEAR	WHITE, NON-HISPANIC	BLACK, NON-HISPANIC	HISPANIC ORIGIN
1971	13.4	23.7	34.1
1973	11.6	22.2	33.5
1975	11.4	22.9	29.2
1977	11.9	19.8	33.0
1979	12.0	21.1	33.8
1981	11.4	18.4	33.2
1983	11.2	18.0	31.6
1985	10.4	15.2	27.6
1987	10.4	14.1	28.6
1989	9.4	13.9	33.0
1991	8.9	13.6	35.3
1993	7.9	13.6	27.5
1995	8.6	12.1	30.3
1997	7.1	13.0	29.1
1999	7.9	14.0	30.1

37. a. What was the percentage decrease in black non-Hispanic school dropouts from 1971 to 1997?

b. The percentage decrease of black non-Hispanic dropouts from 1971 to 1997 was how many times the percentage decrease of Hispanic dropouts (to the nearest whole number)?

c. Form a scatter plot to compare the white non-Hispanic dropouts to the black non-Hispanic dropouts by forming intervals from 6 to 14 percent on the horizontal axis for white non-Hispanic and from 12 to 24 percent on the vertical 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 a white non-Hispanic dropout rate of 12 percent.

38. a. What was the percentage decrease in white non-Hispanic dropouts from 1971 to 1997?

b. The percentage decrease of black non-Hispanic dropouts from 1971 to 1977 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 Hispanic dropouts by forming intervals from 12 to 24 percent on the horizontal axis for black non-Hispanic dropouts and from 26 to 36 percent on the vertical axis for Hispanic dropouts. Is there a positive or negative association?

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

*U.S. Department of Health and Human Services, *Child Health USA 2000* (Washington, DC: U.S. Government Printing Office, 2000).

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

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 474 and 475.

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 47.

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)

**Data Analysis and Statistics across the Curriculum* (Reston, VA: National Council of Teachers of Mathematics, 1992), p. 43.

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 fits 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 names 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, 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.

ONLINE LEARNING CENTER www.mhhe.com/bennett-nelson

• Math Investigation 7.1 *Trend Lines*
Section-Related: • Links • Writing/Discussion Problems • Bibliography