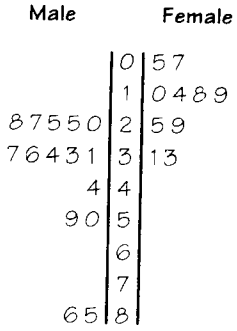
5	7	10	14	18	19	25	29	31	33

We can compare the number of cesarean sections performed by male and fema doctors using a back-to-back stemplot. Figure 1.23 shows the completed graph. As y can see, the stems are listed in the middle and leaves are placed on the left for ma doctors and on the right for female doctors. It is usual to have the leaves increase value as they move away from the stem.

NUMBER OF CESAREAN SECTIONS PERFORMED BY MALE AND FEMALE DOCTORS



Key:
|2| 5 means that a female doctor performed 25 cesarean sections that year
0 |5| means that a male doctor performed 50 cesarean sections that year

FIGURE 1.23 Back-to-back stemplot of the number of cesarean sections performed by male and female Swiss doctors.

The distribution of the number of cesareans performed by female doctors is roughly symmetric. For the male doctors, the distribution is skewed to the right. More than half o the female doctors in the study performed fewer than 20 cesarean sections in a year. The minimum number of cesareans performed by any of the male doctors was 20. Two male physicians performed an unusually high number of cesareans, 85 and 86.

Here are numerical summaries for the two distributions:

	$\bar{x}$	s	Min.	Q_1	M	Q_3	Max.	IQR
Male doctors	41.333	20.607	20	27	34	50	86	23
Female doctors	19.1	10.126	5	10	18.5	29	33	19

The mean and median numbers of cesarean sections performed are higher for the male doctors. Both the standard deviation and the *IQR* for the male doctors are much larger than the corresponding statistics for the female doctors. So there is much greater variability in the number of cesarean sections performed by male physicians. Due to the apparent outliers in the male doctor data and the lack of symmetry of their distribution of cesareans, we should use the medians and *IQR*s in our numerical comparisons.

We have already seen that boxplots can be useful for comparing distributions of quantitative variables. Side-by-side boxplots, like those in the Technology Toolbox on page 47, help us quickly compare shape, center, and spread.

EXERCISES

1.47 GET YOUR HOT DOGS HERE! "Face it. A hot dog isn't a carrot stick." So said *Consumer Reports*, commenting on the low nutritional quality of the all-American frank. Table 1.10 shows the magazine's laboratory test results for calories and milligrams of sodium (mostly due to salt) in a number of major brands of hot dogs. There are three types: beef, "meat" (mainly pork and beef, but government regulations allow up to 15% poultry meat), and poultry. Because people concerned about their health may prefer low-calorie, low-sodium hot dogs, we ask: "Are there any systematic differences among the three types of hot dogs in these two variables?" Use side-by-side boxplots and numerical summaries to help you answer this question. Write a paragraph explaining your findings.

TABLE 1.10 Calories and sodium in three types of hot dogs

Beef hot dogs		Meat hot dogs		Poultry hot dogs	
Calories	Sodium	Calories	Sodium	Calories	Sodium
186	495	173	458	129	430
181	477	191	506	132	375
176	425	182	473	102	396
149	322	190	545	106	383
184	482	172	496	94	387
190	587	147	360	102	542
158	370	146	387	87	359
139	322	139	386	99	357
175	479	175	507	170	528
148	375	136	393	113	513
152	330	179	405	135	426
111	300	153	372	142	513
141	386	107	144	86	358
153	401	195	511	143	581
190	645	135	405	152	588
157	440	140	428	146	522
131	317	138	339	144	545
149	319				
135	298				
132	253				

Source: Consumer Reports, June 1986, pp.366–367

1.48 WHICH AP EXAM IS EASIER: CALCULUS AB OR STATISTICS? The table below gives the distribution of grades earned by students taking the Calculus AB and Statistics exams in 2000.[14]

	5	4	3	2	1
Calculus AB	16.8%	23.2%	23.5%	19.6%	16.8%
Statistics	9.8%	21.5%	22.4%	20.5%	25.8%

(a) Make a graphical display to compare the AP exam grades for Calculus AB and Statistics.

(b) Write a few sentences comparing the two distributions of exam grades. Do you now know which exam is easier? Why or why not?

1.49 WHO MAKES MORE? A manufacturing company is reviewing the salaries of its full-time employees below the executive level at a large plant. The clerical staff is almost entirely female, while a majority of the production workers and technical staff are male. As a result, the distributions of salaries for male and female employees may be quite different. Table 1.11 gives the frequencies and relative frequencies for women and men.

(a) Make histograms for these data, choosing a vertical scale that is most appropriate for comparing the two distributions.

(b) Describe the shape of the overall salary distributions and the chief differences between them.

(c) Explain why the total for women is greater than 100%.

TABLE 1.11 Salary distributions of female and male workers in a large factory

Salary ($1000)	Women		Men	
	Number	%	Number	%
10–15	89	11.8	26	1.1
15–20	192	25.4	221	9.0
20–25	236	31.2	677	27.9
25–30	111	14.7	823	33.6
30–35	86	11.4	365	14.9
35–40	25	3.3	182	7.4
40–45	11	1.5	91	3.7
45–50	3	0.4	33	1.4
50–55	2	0.3	19	0.8
55–60	0	0.0	11	0.4
60–65	0	0.0	0	0.0
65–70	1	0.1	3	0.1
Total	756	100.1	2451	100.0

1.50 BASKETBALL PLAYOFF SCORES Here are the scores of games played in the California Division I-AAA high school basketball playoffs:[15]

71–38 52–47 55–53 76–65 77–63 65–63 68–54 64–62
87–47 64–56 78–64 58–51 91–74 71–41 67–62 106–46

On the same day, the final scores of games in Division V-AA were

98–45 67–44 74–60 96–54 92–72 93–46
98–67 62–37 37–36 69–44 86–66 66–58

(a) Construct a back-to-back stemplot to compare the number of points scored by Division I-AAA and Division V-AA basketball teams.

(b) Compare the shape, center, and spread of the two distributions. Which numerical summaries are most appropriate in this case? Why?

(c) Is there a difference in "margin of victory" in Division I-AAA and Division V-AA playoff games? Provide appropriate graphical and numerical support for your answer.

SUMMARY

A numerical summary of a distribution should report its **center** and its **spread,** or **variability.**

The **mean** $\bar{x}$ and the **median** M describe the center of a distribution in different ways. The mean is the arithmetic average of the observations, and the median is the midpoint of the values.

When you use the median to indicate the center of a distribution, describe its spread by giving the **quartiles.** The **first quartile** Q_1 has one-fourth of the observations below it, and the **third quartile** Q_3 has three-fourths of the observations below it. An extreme observation is an **outlier** if it is smaller than $Q_1 - (1.5 \times IQR)$ or larger than $Q_3 + (1.5 \times IQR)$.

The **five-number summary** consists of the median, the quartiles, and the high and low extremes and provides a quick overall description of a distribution. The median describes the center, and the quartiles and extremes show the spread.

Boxplots based on the five-number summary are useful for comparing two or more distributions. The box spans the quartiles and shows the spread of the central half of the distribution. The median is marked within the box. Lines extend from the box to the smallest and the largest observations that are not outliers. Outliers are plotted as isolated points.

The **variance** s^2 and especially its square root, **the standard deviation** s, are common measures of spread about the mean as center. The standard deviation s is zero when there is no spread and gets larger as the spread increases.

The mean and standard deviation are strongly influenced by outliers or skewness in a distribution. They are good descriptions for symmetric distributions and are most useful for the normal distributions, which will be introduced in the next chapter.

The median and quartiles are not affected by outliers, and the two quartiles and two extremes describe the two sides of a distribution separately. The five-number summary is the preferred numerical summary for skewed distributions.

When you add a constant a to all the values in a data set, the mean and median increase by a. Measures of spread do not change. When you multiply all the values in a data set by a constant b, the mean, median, IQR, and standard deviation are multiplied by b. These **linear transformations** are quite useful for changing units of measurement.

Back-to-back stemplots and **side-by-side boxplots** are useful for comparing quantitative distributions.

SECTION 1.2 EXERCISES

1.51 MEAT HOT DOGS Make a stemplot of the calories in meat hot dogs from Exercise 1.47 (page 59). What does this graph reveal that the boxplot of these data did not? *Lesson:* Be aware of the limitations of each graphical display.

1.52 EDUCATIONAL ATTAINMENT Table 1.12 shows the educational level achieved by U.S. adults aged 25 to 34 and by those aged 65 to 74. Compare the distributions of educational attainment graphically. Write a few sentences explaining what your display shows.

TABLE 1.12 Educational attainment by U.S. adults aged 25 to 34 and 65 to 74

	Number of people (thousands)	
	Ages 25–34	Ages 65–74
Less than high school	4474	4695
High school graduate	11,546	6649
Some college	7376	2528
Bachelor's degree	8563	1849
Advanced degree	3374	1266
Total	35,333	16,987

Source: Census Bureau, *Educational Attainment in the United States*, March 2000.

1.53 CASSETTE VERSUS CD SALES Has the increasing popularity of the compact disc (CD) affected sales of cassette tapes? Table 1.13 shows the number of cassettes and CDs sold from 1990 to 1999.

TABLE 1.13 Sales (in millions) of full-length cassettes and CDs, 1990–1999

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Full-length cassettes	54.7	49.8	43.6	38.0	32.1	25.1	19.3	18.2	14.8	8.0
Full-length CDs	31.1	38.9	46.5	51.1	58.4	65.0	68.4	70.2	74.8	83.2

Source: The Recording Industry Association of America, *1999 Consumer Profile.*

Make a graphical display to compare cassette and CD sales. Write a few sentences describing what your graph tells you.

1.54 $\bar{x}$ AND s ARE NOT ENOUGH The mean $\bar{x}$ and standard deviation s measure center and spread but are not a complete description of a distribution. Data sets with different shapes can have the same mean and standard deviation. To demonstrate this fact, use your calculator to find $\bar{x}$ and s for the following two small data sets. Then make a stemplot of each and comment on the shape of each distribution.

Data A:	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74
Data B:	6.58	5.76	7.71	8.84	8.47	7.04	5.25	5.56	7.91	6.89	12.50

1.55 In each of the following settings, give the values of a and b for the linear transformation $x_{new} = a + bx$ that expresses the change in measurement units. Then explain how the transformation will affect the mean, the IQR, the median, and the standard deviation of the original distribution.

(a) You collect data on the power of car engines, measured in horsepower. Your teacher requires you to convert the power to watts. One horsepower is 746 watts.

(b) You measure the temperature (in degrees Fahrenheit) of your school's swimming pool at 20 different locations within the pool. Your swim team coach wants the summary statistics in degrees Celsius ($^\circ$ F = (9/5)$^\circ$ C + 32).

(c) Dr. Data has given a very difficult statistics test and is thinking about "curving" the grades. She decides to add 10 points to each student's score.

1.56 A change of units that multiplies each unit by b, such as the change $x_{new} = 0 + 2.54x$ from inches x to centimeters x_{new}, multiplies our usual measures of spread by b. This is true of the IQR and standard deviation. What happens to the variance when we change units in this way?

1.57 BETTER CORN Corn is an important animal food. Normal corn lacks certain amino acids, which are building blocks for protein. Plant scientists have developed new corn varieties that have more of these amino acids. To test a new corn as an animal food, a group of 20 one-day-old male chicks was fed a ration containing the new corn. A control group of another 20 chicks was fed a ration that was identical except that it contained normal corn. Here are the weight gains (in grams) after 21 days:[16]

Normal corn				New corn			
380	321	366	356	361	447	401	375
283	349	402	462	434	403	393	426
356	410	329	399	406	318	467	407
350	384	316	272	427	420	477	392
345	455	360	431	430	339	410	326

(a) Compute five-number summaries for the weight gains of the two groups of chicks. Then make boxplots to compare the two distributions. What do the data show about the effect of the new corn?

(b) The researchers actually reported means and standard deviations for the two groups of chicks. What are they? How much larger is the mean weight gain of chicks fed the new corn?

(c) The weights are given in grams. There are 28.35 grams in an ounce. Use the results of part (b) to compute the means and standard deviations of the weight gains measured in ounces.

H. COMPARING DISTRIBUTIONS

1. Use side-by-side bar graphs to compare distributions of categorical data.

2. Make back-to-back stemplots and side-by-side boxplots to compare distributions of quantitative variables.

3. Write narrative comparisons of the shape, center, spread, and outliers for two or more quantitative distributions.

CHAPTER 1 REVIEW EXERCISES

1.59 Each year *Fortune* magazine lists the top 500 companies in the United States, ranked according to their total annual sales in dollars. Describe three other variables that could reasonably be used to measure the "size" of a company.

1.60 ATHLETES' SALARIES Here is a small part of a data set that describes major league baseball players as of opening day of the 1998 season:

Player	Team	Position	Age	Salary
⋮				
Perez, Eduardo	Reds	First base	28	300
Perez, Neifi	Rockies	Shortstop	23	210
Pettitte, Andy	Yankees	Pitcher	25	3750
Piazza, Mike	Dodgers	Catcher	29	8000
⋮				

(a) What individuals does this data set describe?

(b) In addition to the player's name, how many variables does the data set contain? Which of these variables are categorical and which are quantitative?

(c) Based on the data in the table, what do you think are the units of measurement for each of the quantitative variables?

1.61 HOW YOUNG PEOPLE DIE The number of deaths among persons aged 15 to 24 years in the United States in 1997 due to the seven leading causes of death for this age group were accidents, 12,958; homicide, 5793; suicide, 4146; cancer, 1583; heart disease, 1013; congenital defects, 383; AIDS, 276.[17]

(a) Make a bar graph to display these data.

(b) What additional information do you need to make a pie chart?

1.62 NEVER ON SUNDAY? The Canadian Province of Ontario carries out statistical studies of the working of Canada's national health care system in the province. The bar graphs in Figure 1.24 come from a study of admissions and discharges from community hospitals in Ontario.[18] They show the number of heart attack patients admitted and discharged on each day of the week during a 2-year period.

(a) Explain why you expect the number of patients admitted with heart attacks to be roughly the same for all days of the week. Do the data show that this is true?

(b) Describe how the distribution of the day on which patients are discharged from the hospital differs from that of the day on which they are admitted. What do you think explains the difference?

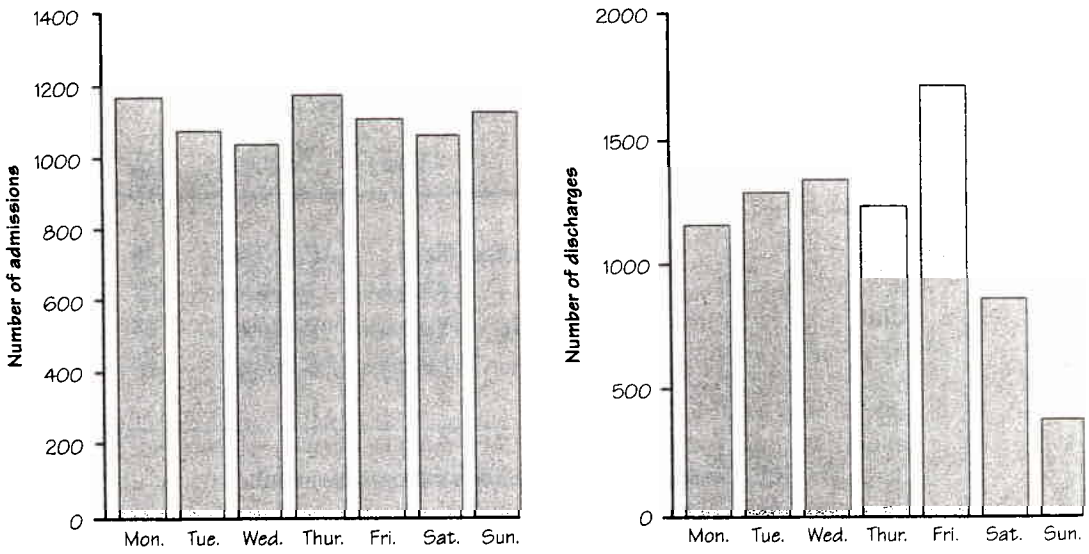


FIGURE 1.24 Bar graphs of the number of heart attack victims admitted and discharged on each day of the week by hospitals in Ontario, Canada.

1.63 PRESIDENTIAL ELECTIONS Here are the percents of the popular vote won by the successful candidate in each of the presidential elections from 1948 to 2000:

Year:	1948	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992	1996	2000
Percent:	49.6	55.1	57.4	49.7	61.1	43.4	60.7	50.1	50.7	58.8	53.9	43.2	49.2	47.9

(a) Make a stemplot of the winners' percents. (Round to whole numbers and use split stems.)

(b) What is the median percent of the vote won by the successful candidate in presidential elections? (Work with the unrounded data.)

(c) Call an election a landslide if the winner's percent falls at or above the third quartile. Find the third quartile. Which elections were landslides?

1.64 HURRICANES The histogram in Figure 1.25 (next page) shows the number of hurricanes reaching the east coast of the United States each year over a 70-year period.[19] Give a brief description of the overall shape of this distribution. About where does the center of the distribution lie?

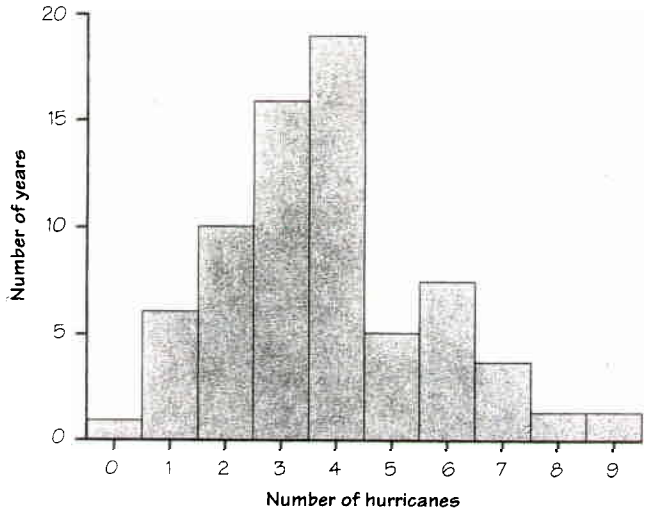


FIGURE 1.25 The distribution of the annual number of hurricanes on the U.S. east coast over a 70-year period, for Exercise 1.64.

1.65 DO SUVS WASTE GAS? Table 1.3 (page 17) gives the highway fuel consumption (in miles per gallon) for 32 model year 2000 midsize cars. We constructed a dotplot for these data in Exercise 1.8. Table 1.14 shows the highway mileages for 26 four-wheel-drive model year 2000 sport utility vehicles.

(a) Give a graphical and numerical description of highway fuel consumption for SUVs. What are the main features of the distribution?

(b) Make boxplots to compare the highway fuel consumption of midsize cars and SUVs. What are the most important differences between the two distributions?

TABLE 1.14 Highway gas mileages for model year 2000 four-wheel-drive SUVs

Model	MPG	Model	MPG
BMW X5	17	Kia Sportage	22
Chevrolet Blazer	20	Land Rover	17
Chevrolet Tahoe	18	Lexus LX470	16
Dodge Durango	18	Lincoln Navigator	17
Ford Expedition	18	Mazda MPV	19
Ford Explorer	20	Mercedes-Benz ML320	20
Honda Passport	20	Mitsubishi Montero	20
Infinity QX4	18	Nissan Pathfinder	19
Isuzu Amigo	19	Nissan Xterra	19
Isuzu Trooper	19	Subaru Forester	27
Jeep Cherokee	20	Suzuki Grand Vitara	20
Jeep Grand Cherokee	18	Toyota RAV4	26
Jeep Wrangler	19	Toyota 4Runner	21

1.66 DR. DATA RETURNS! Dr. Data asked her students how much time they spent using a computer during the previous week. Figure 1.26 is an ogive of her students' responses.

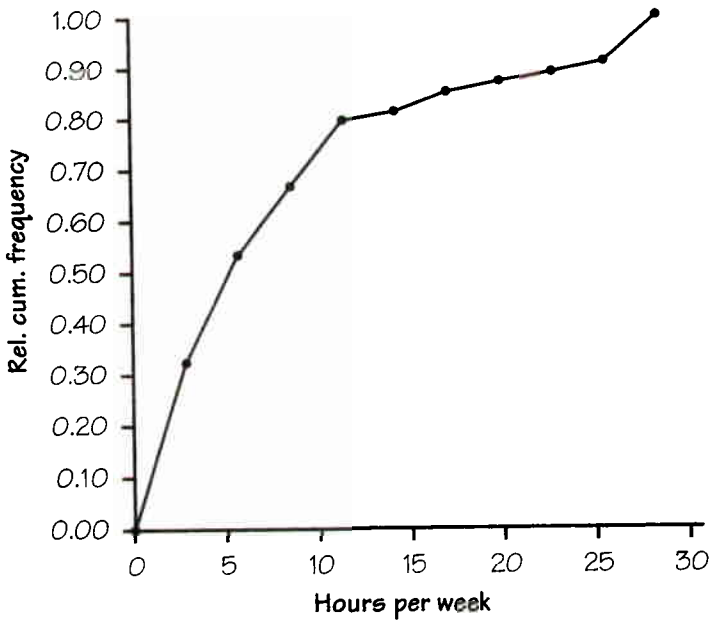


FIGURE 1.26 Ogive of weekly computer use by Dr. Data's statistics students.

(a) Construct a relative frequency table based on the ogive. Then make a histogram.

(b) Estimate the median, Q_1, and Q_3 from the ogive. Then make a boxplot. Are there any outliers?

(c) At what percentile does a student who used her computer for 10 hours last week fall?

1.67 WAL-MART STOCK The rate of return on a stock is its change in price plus any dividends paid. Rate of return is usually measured in percent of the starting value. We have data on the monthly rates of return for the stock of Wal-Mart stores for the years 1973 to 1991, the first 19 years Wal-Mart was listed on the New York Stock Exchange. There are 228 observations.

Figure 1.27 (next page) displays output from statistical software that describes the distribution of these data. The stems in the stemplot are the tens digits of the percent returns. The leaves are the ones digits. The stemplot uses split stems to give a better display. The software gives high and low outliers separately from the stemplot rather than spreading out the stemplot to include them.

(a) Give the five-number summary for monthly returns on Wal-Mart stock.

(b) Describe in words the main features of the distribution.

(c) If you had \$1000 worth of Wal-Mart stock at the beginning of the best month during these 19 years, how much would your stock be worth at the end of the month? If you had \$1000 worth of stock at the beginning of the worst month, how much would your stock be worth at the end of the month?

(d) Find the interquartile range (IQR) for the Wal-Mart data. Are there any outliers according to the $1.5 \times$ IQR criterion? Does it appear to you that the software uses this criterion in choosing which observations to report separately as outliers?

```
Mean   =   3.064
Standard deviation   =   11.49

N   =  228    Median   =   3.4691
Quartiles   =   -2.950258,   8.4511

Decimal point is 1 place to the right of the colon

Low:   -34.04255   -31.25000   -27.06271   -26.61290

 -1 : 985
 -1 : 444443322222110000
 -0 : 999988777666666665555
 -0 : 444444443333333222222222221111111100
  0 : 000001111111111112222223333333344444444
  0 : 555555555555555555556666666666677777778888888888899999
  1 : 000000000111111111222333344444
  1 : 55566667889
  2 : 011334

High:   32.01923   41.80531   42.05607   57.89474   58.67769
```

FIGURE 1.27 Output from software describing the distribution of monthly returns from Wal-Mart stock.

1.68 A study of the size of jury awards in civil cases (such as injury, product liability and medical malpractice) in Chicago showed that the median award was about $8000. But the mean award was about $69,000. Explain how this great difference between the two measures of center can occur.

1.69 You want to measure the average speed of vehicles on the interstate highway on which you are driving. You adjust your speed until the number of vehicles passing you equals the number you are passing. Have you found the mean speed or the median speed of vehicles on the highway?

TABLE 1.15 Data on education in the United States for Exercises 1.70 to 1.73

State	Region	Population (1000)	SAT Verbal	SAT Math	Percent taking	Percent no HS diploma	Teachers' pay ($1000)
AL	ESC	4,447	561	555	9	33.1	32.8
AK	PAC	627	516	514	50	13.4	51.7
AZ	MTN	5,131	524	525	34	21.3	34.4
AR	WSC	2,673	563	556	6	33.7	30.6
CA	PAC	33,871	497	514	49	23.8	43.7

TABLE 1.15 Data on education in the United States, for Exercises 1.70 to 1.73 (*continued*)

State	Region	Population (1000)	SAT Verbal	SAT Math	Percent taking	Percent no HS diploma	Teachers' pay ($1000)
CO	MTN	4,301	536	540	32	15.6	37.1
CT	NE	3,406	510	509	80	20.8	50.7
DE	SA	784	503	497	67	22.5	42.4
DC	SA	572	494	478	77	26.9	46.4
FL	SA	15,982	499	498	53	25.6	34.5
GA	SA	8,186	487	482	63	29.1	37.4
HI	PAC	1,212	482	513	52	19.9	38.4
ID	MTN	1,294	542	540	16	20.3	32.8
IL	ENC	12,419	569	585	12	23.8	43.9
IN	ENC	6,080	496	498	60	24.4	39.7
IA	WNC	2,926	594	598	5	19.9	34.0
KS	WNC	2,688	578	576	9	18.7	36.8
KY	ESC	4,042	547	547	12	35.4	34.5
LA	WSC	4,469	561	558	8	31.7	29.7
ME	NE	1,275	507	503	68	21.2	34.3
MD	SA	5,296	507	507	65	21.6	41.7
MA	NE	6,349	511	511	78	20.0	43.9
MI	ENC	9,938	557	565	11	23.2	49.3
MN	WNC	4,919	586	598	9	17.6	39.1
MS	ESC	2,845	563	548	4	35.7	29.5
MO	WNC	5,595	572	572	8	26.1	34.0
MT	MTN	902	545	546	21	19.0	30.6
NE	WNC	1,711	568	571	8	18.2	32.7
NV	MTN	1,998	512	517	34	21.2	37.1
NH	NE	1,236	520	518	72	17.8	36.6
NJ	MA	8,414	498	510	80	23.3	50.4
NM	MTN	1,819	549	542	12	24.9	30.2
NY	MA	18,976	495	502	76	25.2	49.0
NC	SA	8,049	493	493	61	30.0	33.3
ND	WNC	642	594	605	5	23.3	28.2
OH	ENC	11,353	534	568	25	24.3	39.0
OK	WSC	3,451	567	560	8	25.4	30.6
OR	PAC	3,421	525	525	53	18.5	42.2
PA	MA	12,281	498	495	70	25.3	47.7
RI	NE	1,048	504	499	70	28.0	44.3
SC	SA	4,012	479	475	61	31.7	33.6
SD	WNC	755	585	588	4	22.9	27.3
TN	ESC	5,689	559	553	13	32.9	35.3
TX	WSC	20,852	494	499	50	27.9	33.6
UT	MTN	2,233	570	568	5	14.9	33.0
VT	NE	609	514	506	70	19.2	36.3
VA	SA	7,079	508	499	65	24.8	36.7
WA	PAC	5,894	525	526	52	16.2	38.8
WV	SA	1,808	527	512	18	34.0	33.4
WI	ENC	5,364	584	595	7	21.4	39.9
WY	MTN	494	546	551	10	17.0	32.0

Source: U.S. Census Bureau Web site, http://www.census.gov, 2001.

Table 1.15 presents data about the individual states that relate to education. Study of a data set with many variables begins by examining each variable by itself. Exercises 1.70 to 1.73 concern the data in Table 1.15.

1.70 POPULATION OF THE STATES Make a graphical display of the population of the states. Briefly describe the shape, center, and spread of the distribution of population. Explain why the shape of the distribution is not surprising. Are there any states that you consider outliers?

1.71 HOW MANY STUDENTS TAKE THE SAT? Make a stemplot of the distribution of the percent of high school seniors who take the SAT in the various states. Briefly describe the overall shape of the distribution. Find the midpoint of the data and mark this value on your stemplot. Explain why describing the center is not very useful for a distribution with this shape.

1.72 HOW MUCH ARE TEACHERS PAID? Make a graph to display the distribution of average teachers' salaries for the states. Is there a clear overall pattern? Are there any outliers or other notable deviations from the pattern?

1.73 PEOPLE WITHOUT HIGH SCHOOL EDUCATIONS The "Percent no HS" column gives the percent of the adult population in each state who did not graduate from high school. We want to compare the percents of people without a high school education in the northeastern and the southern states. Take the northeastern states to be those in the MA (Mid-Atlantic) and NE (New England) regions. The southern states are those in the SA (South Atlantic) and ESC (East South Central) regions. Leave out the District of Columbia, which is a city rather than a state.

(a) List the percents without high school for the northeastern and for the southern states from Table 1.15. These are the two data sets we want to compare.

(b) Make numerical summaries and graphs to compare the two distributions. Write a brief statement of what you find.

NOTES AND DATA SOURCES

1. Data from *Beverage Digest*, February 18, 2000.
2. Seat-belt data from the National Highway and Traffic Safety Administration, *NOPUS Survey*, 1998.
3. Data from the 1997 *Statistical Abstract of the United States*.
4. Data on accidental deaths from the Centers for Disease Control Web site, www.cdc.gov.
5. Data from the *Los Angeles Times*, February 16, 2001.
6. Based on experiments performed by G. T. Lloyd and E. H. Ramshaw of the CSIRO Division of Food Research, Victoria, Australia, 1982–83.
7. Maribeth Cassidy Schmitt, from her Ph.D. dissertation, "The effects of an elaborated directed reading activity on the metacomprehension skills of third graders," Purdue University, 1987.
8. Data from "America's best small companies," *Forbes*, November 8, 1993.
9. The Shakespeare data appear in C. B. Williams, *Style and Vocabulary: Numerological Studies*, Griffin, London, 1970.