EXERCISES

1. **Histogram.** Find a histogram that shows the distribution of a variable in a newspaper, a magazine, or the Internet.
   a) Does the article identify the W's?
   b) Discuss whether the display is appropriate.
   c) Discuss what the display reveals about the variable and its distribution.
   d) Does the article accurately describe and interpret the data? Explain.

2. **Not a histogram.** Find a graph other than a histogram that shows the distribution of a quantitative variable in a newspaper, a magazine, or the Internet.
   a) Does the article identify the W's?
   b) Discuss whether the display is appropriate for the data.
   c) Discuss what the display reveals about the variable and its distribution.
   d) Does the article accurately describe and interpret the data? Explain.

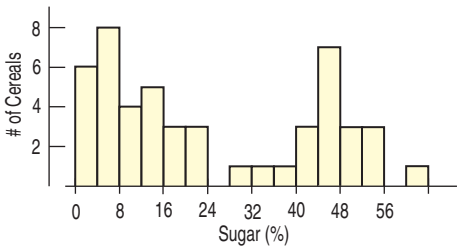
3. **In the news.** Find an article in a newspaper, a magazine, or the Internet that discusses an "average."
   a) Does the article discuss the W's for the data?
   b) What are the units of the variable?
   c) Is the average used the median or the mean? How can you tell?
   d) Is the choice of median or mean appropriate for the situation? Explain.

4. **In the news II.** Find an article in a newspaper, a magazine, or the Internet that discusses a measure of spread.
   a) Does the article discuss the W's for the data?
   b) What are the units of the variable?
   c) Does the article use the range, IQR, or standard deviation?
   d) Is the choice of measure of spread appropriate for the situation? Explain.

5. **Thinking about shape.** Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.
   a) The number of speeding tickets each student in the senior class of a college has ever had.
   b) Players' scores (number of strokes) at the U.S. Open golf tournament in a given year.
   c) Weights of female babies born in a particular hospital over the course of a year.
   d) The length of the average hair on the heads of students in a large class.

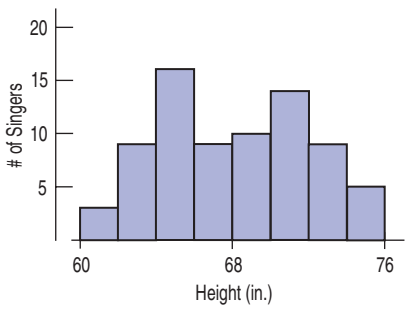
6. **More shapes.** Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.
   a) Ages of people at a Little League game.
   b) Number of siblings of people in your class.
   c) Pulse rates of college-age males.
   d) Number of times each face of a die shows in 100 tosses.

T 7. **Sugar in cereals.** The histogram displays the sugar content (as a percent of weight) of 49 brands of breakfast cereals.



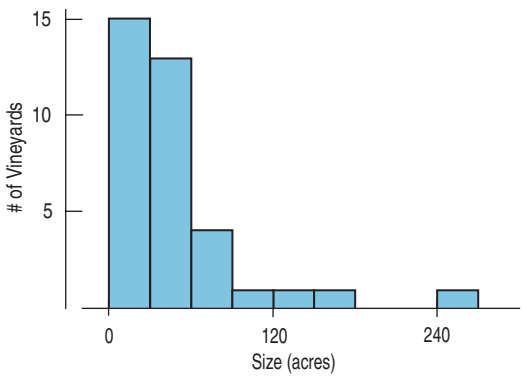
   a) Describe this distribution.
   b) What do you think might account for this shape?

T 8. **Singers.** The display shows the heights of some of the singers in a chorus, collected so that the singers could be positioned on stage with shorter ones in front and taller ones in back.



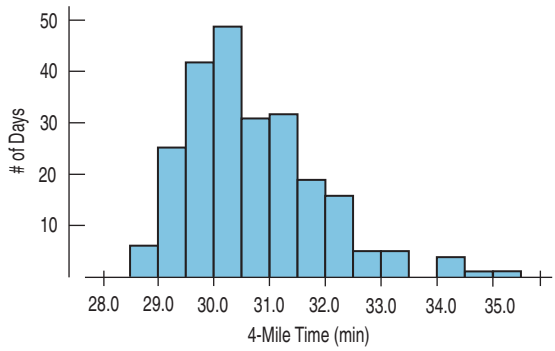
   a) Describe the distribution.
   b) Can you account for the features you see here?

T 9. **Vineyards.** The histogram shows the sizes (in acres) of 36 vineyards in the Finger Lakes region of New York.



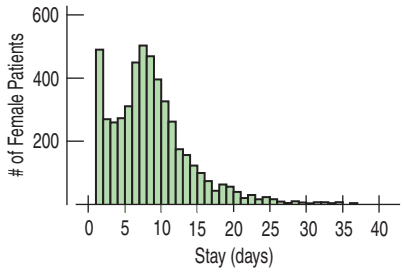
   a) Approximately what percentage of these vineyards are under 60 acres?
   b) Write a brief description of this distribution (shape, center, spread, unusual features).

10. Run times. One of the authors collected the times (in minutes) it took him to run 4 miles on various courses during a 10-year period. Here is a histogram of the times.



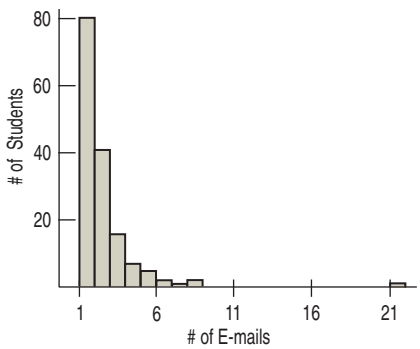
Describe the distribution and summarize the important features. What is it about running that might account for the shape you see?

11. Heart attack stays. The histogram shows the lengths of hospital stays (in days) for all the female patients admitted to hospitals in New York during one year with a primary diagnosis of acute myocardial infarction (heart attack).



a) From the histogram, would you expect the mean or median to be larger? Explain.
b) Write a few sentences describing this distribution (shape, center, spread, unusual features).
c) Which summary statistics would you choose to summarize the center and spread in these data? Why?

12. E-mails. A university teacher saved every e-mail received from students in a large Introductory Statistics class during an entire term. He then counted, for each student who had sent him at least one e-mail, how many e-mails each student had sent.



a) From the histogram, would you expect the mean or the median to be larger? Explain.
b) Write a few sentences describing this distribution (shape, center, spread, unusual features).
c) Which summary statistics would you choose to summarize the center and spread in these data? Why?

13. Super Bowl points. How many points do football teams score in the Super Bowl? Here are the total numbers of points scored by both teams in each of the first 42 Super Bowl games:

45, 47, 23, 30, 29, 27, 21, 31, 22, 38, 46, 37, 66, 50, 37, 47, 44, 47, 54, 56, 59, 52, 36, 65, 39, 61, 69, 43, 75, 44, 56, 55, 53, 39, 41, 37, 69, 61, 45, 31, 46, 31

a) Find the median.
b) Find the quartiles.
c) Write a description based on the 5-number summary.

14. Super Bowl wins. In the Super Bowl, by how many points does the winning team outscore the losers? Here are the winning margins for the first 42 Super Bowl games:

25, 19, 9, 16, 3, 21, 7, 17, 10, 4, 18, 17, 4, 12, 17, 5, 10, 29, 22, 36, 19, 32, 4, 45, 1, 13, 35, 17, 23, 10, 14, 7, 15, 7, 27, 3, 27, 3, 3, 11, 12, 3

a) Find the median.
b) Find the quartiles.
c) Write a description based on the 5-number summary.

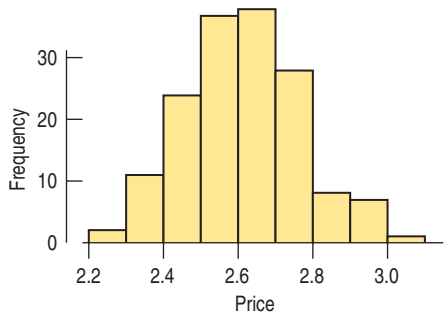
15. Standard deviation I. For each lettered part, a through c, examine the two given sets of numbers. Without doing any calculations, decide which set has the larger standard deviation and explain why. Then check by finding the standard deviations *by hand*.

	Set 1	Set 2
a)	3, 5, 6, 7, 9	2, 4, 6, 8, 10
b)	10, 14, 15, 16, 20	10, 11, 15, 19, 20
c)	2, 6, 6, 9, 11, 14	82, 86, 86, 89, 91, 94

16. Standard deviation II. For each lettered part, a through c, examine the two given sets of numbers. Without doing any calculations, decide which set has the larger standard deviation and explain why. Then check by finding the standard deviations *by hand*.

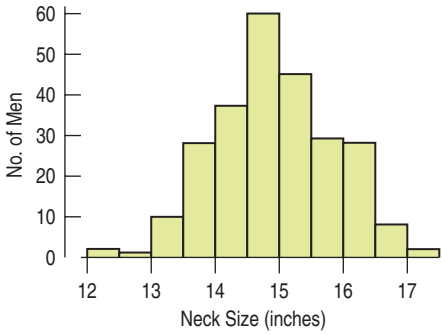
	Set 1	Set 2
a)	4, 7, 7, 7, 10	4, 6, 7, 8, 10
b)	100, 140, 150, 160, 200	10, 50, 60, 70, 110
c)	10, 16, 18, 20, 22, 28	48, 56, 58, 60, 62, 70

17. Pizza prices. The histogram shows the distribution of the prices of plain pizza slices (in $) for 156 weeks in Dallas, TX.



Which summary statistics would you choose to summarize the center and spread in these data? Why?

18. Neck size. The histogram shows the neck sizes (in inches) of 250 men recruited for a health study in Utah.

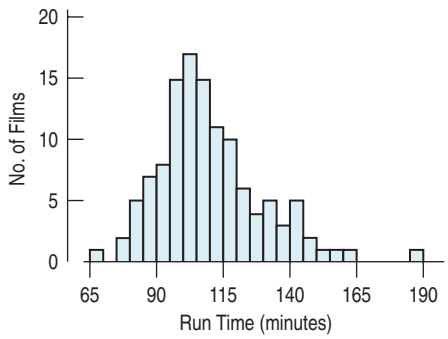


Which summary statistics would you choose to summarize the center and spread in these data? Why?

19. Pizza prices again. Look again at the histogram of the pizza prices in Exercise 17.
a) Is the mean closer to \$2.40, \$2.60, or \$2.80? Why?
b) Is the standard deviation closer to \$0.15, \$0.50, or \$1.00? Explain.

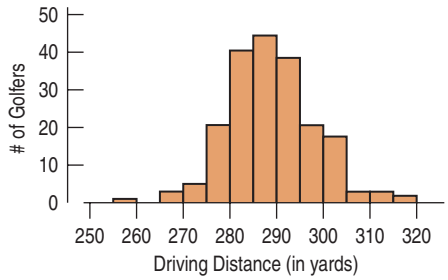
20. Neck sizes again. Look again at the histogram of men's neck sizes in Exercise 18.
a) Is the mean closer to 14, 15, or 16 inches? Why?
b) Is the standard deviation closer to 1 inch, 3 inches, or 5 inches? Explain.

21. Movie lengths. The histogram shows the running times in minutes of 122 feature films released in 2005.



a) You plan to see a movie this weekend. Based on these movies, how long do you expect a typical movie to run?
b) Would you be surprised to find that your movie ran for $2\frac{1}{2}$ hours (150 minutes)?
c) Which would you expect to be higher: the mean or the median run time for all movies? Why?

22. Golf drives. The display shows the average drive distance (in yards) for 202 professional golfers on the men's PGA tour.



a) Describe this distribution.
b) Approximately what proportion of professional male golfers drive, on average, less than 280 yards?
c) Estimate the mean by examining the histogram.
d) Do you expect the mean to be smaller than, approximately equal to, or larger than the median? Why?

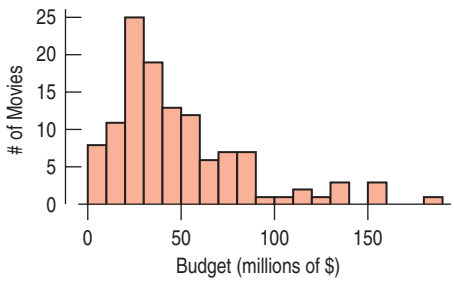
23. Movie lengths II. Exercise 21 looked at the running times of movies released in 2005. The standard deviation of these running times is 19.6 minutes, and the quartiles are $Q_1 = 97$ minutes and $Q_3 = 119$ minutes.
a) Write a sentence or two describing the spread in running times based on
  i) the quartiles.
  ii) the standard deviation.
b) Do you have any concerns about using either of these descriptions of spread? Explain.

24. Golf drives II. Exercise 22 looked at distances PGA golfers can hit the ball. The standard deviation of these average drive distances is 9.3 yards, and the quartiles are $Q_1 = 282$ yards and $Q_3 = 294$ yards.
a) Write a sentence or two describing the spread in distances based on
  i) the quartiles.
  ii) the standard deviation.
b) Do you have any concerns about using either of these descriptions of spread? Explain.

25. Mistake. A clerk entering salary data into a company spreadsheet accidentally put an extra "0" in the boss's salary, listing it as \$2,000,000 instead of \$200,000. Explain how this error will affect these summary statistics for the company payroll:
a) measures of center: median and mean.
b) measures of spread: range, IQR, and standard deviation.

26. Cold weather. A meteorologist preparing a talk about global warming compiled a list of weekly low temperatures (in degrees Fahrenheit) he observed at his southern Florida home last year. The coldest temperature for any week was 36°F, but he inadvertently recorded the Celsius value of 2°. Assuming that he correctly listed all the other temperatures, explain how this error will affect these summary statistics:
a) measures of center: mean and median.
b) measures of spread: range, IQR, and standard deviation.

27. Movie budgets. The histogram shows the budgets (in millions of dollars) of major release movies in 2005.



An industry publication reports that the average movie costs \$35 million to make, but a watchdog group con-

cerned with rising ticket prices says that the average cost is $46.8 million. What statistic do you think each group is using? Explain.

28. Sick days. During contract negotiations, a company seeks to change the number of sick days employees may take, saying that the annual "average" is 7 days of absence per employee. The union negotiators counter that the "average" employee misses only 3 days of work each year. Explain how both sides might be correct, identifying the measure of center you think each side is using and why the difference might exist.

29. Payroll. A small warehouse employs a supervisor at $1200 a week, an inventory manager at $700 a week, six stock boys at $400 a week, and four drivers at $500 a week.

a) Find the mean and median wage.
b) How many employees earn more than the mean wage?
c) Which measure of center best describes a typical wage at this company: the mean or the median?
d) Which measure of spread would best describe the payroll: the range, the IQR, or the standard deviation? Why?

30. Singers. The frequency table shows the heights (in inches) of 130 members of a choir.

Height	Count	Height	Count
60	2	69	5
61	6	70	11
62	9	71	8
63	7	72	9
64	5	73	4
65	20	74	2
66	18	75	4
67	7	76	1
68	12		

a) Find the median and IQR.
b) Find the mean and standard deviation.
c) Display these data with a histogram.
d) Write a few sentences describing the distribution.

31. Gasoline. In March 2006, 16 gas stations in Grand Junction, CO, posted these prices for a gallon of regular gasoline:

2.22	2.21	2.45	2.24
2.27	2.28	2.27	2.23
2.26	2.46	2.29	2.32
2.36	2.38	2.33	2.27

a) Make a stem-and-leaf display of these gas prices. Use split stems; for example, use two 2.2 stems—one for prices between $2.20 and $2.24 and the other for prices from $2.25 to $2.29.
b) Describe the shape, center, and spread of this distribution.
c) What unusual feature do you see?

32. The Great One. During his 20 seasons in the NHL, Wayne Gretzky scored 50% more points than anyone who ever played professional hockey. He accomplished this amazing feat while playing in 280 fewer games than Gordie Howe, the previous record holder. Here are the number of games Gretzky played during each season:

79, 80, 80, 80, 74, 80, 80, 79, 64, 78, 73, 78, 74, 45, 81, 48, 80, 82, 82, 70

a) Create a stem-and-leaf display for these data, using split stems.
b) Describe the shape of the distribution.
c) Describe the center and spread of this distribution.
d) What unusual feature do you see? What might explain this?

33. States. The stem-and-leaf display shows populations of the 50 states and Washington, DC, in millions of people, according to the 2000 census.

```
3 | 4
2 |
2 | 1
1 | 69
1 | 0122
0 | 5555666667888
0 | 1111111111111222223333333444444
```

State Populations (1| 2 means 12 million)

a) What measures of center and spread are most appropriate?
b) Without doing any calculations, which must be larger: the median or the mean? Explain how you know.
c) From the stem-and-leaf display, find the median and the interquartile range.
d) Write a few sentences describing this distribution.

34. Wayne Gretzky. In Exercise 32, you examined the number of games played by hockey great Wayne Gretzky during his 20-year career in the NHL.

a) Would you use the median or the mean to describe the center of this distribution? Why?
b) Find the median.
c) Without actually finding the mean, would you expect it to be higher or lower than the median? Explain.

35. Home runs. The stem-and-leaf display shows the number of home runs hit by Mark McGwire during the 1986–2001 seasons. Describe the distribution, mentioning its shape and any unusual features.

```
7 | 0
6 | 5
5 | 28
4 | 29
3 | 22399
2 | 29
1 |
0 | 399
```

Home Runs
(7|0 means 70)

36. Bird species. The Cornell Lab of Ornithology holds an annual Christmas Bird Count (www.birdsource.org), in which bird watchers at various locations around the country see how many different species of birds they can spot. Here are some of the counts reported from sites in Texas during the 1999 event:

228	178	186	162	206	166	163
183	181	206	177	175	167	162
160	160	157	156	153	153	152

a) Create a stem-and-leaf display of these data.
b) Write a brief description of the distribution. Be sure to discuss the overall shape as well as any unusual features.

37. Hurricanes 2006. The data below give the number of hurricanes classified as major hurricanes in the Atlantic Ocean each year from 1944 through 2006, as reported by *NOAA* (www.nhc.noaa.gov):

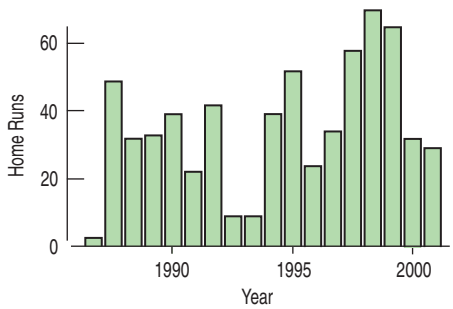
3, 2, 1, 2, 4, 3, 7, 2, 3, 3, 2, 5, 2, 2, 4, 2, 2, 6, 0, 2, 5, 1, 3, 1, 0, 3, 2, 1, 0, 1, 2, 3, 2, 1, 2, 2, 2, 3, 1, 1, 1, 3, 0, 1, 3, 2, 1, 2, 1, 1, 0, 5, 6, 1, 3, 5, 3, 3, 2, 3, 6, 7, 2

a) Create a dotplot of these data.
b) Describe the distribution.

38. Horsepower. Create a stem-and-leaf display for these horsepowers of autos reviewed by *Consumer Reports* one year, and describe the distribution:

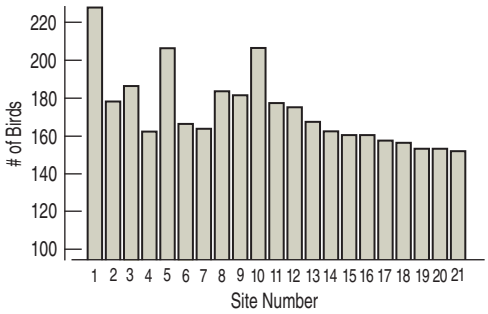
155	103	130	80	65
142	125	129	71	69
125	115	138	68	78
150	133	135	90	97
68	105	88	115	110
95	85	109	115	71
97	110	65	90	
75	120	80	70	

39. Home runs again. Students were asked to make a histogram of the number of home runs hit by Mark McGwire from 1986 to 2001 (see Exercise 35). One student submitted the following display:



a) Comment on this graph.
b) Create your own histogram of the data.

40. Return of the birds. Students were given the assignment to make a histogram of the data on bird counts reported in Exercise 36. One student submitted the following display:



a) Comment on this graph.
b) Create your own histogram of the data.

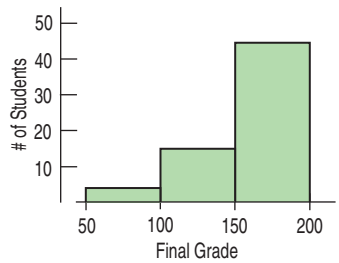
41. Acid rain. Two researchers measured the pH (a scale on which a value of 7 is neutral and values below 7 are acidic) of water collected from rain and snow over a 6-month period in Allegheny County, PA. Describe their data with a graph and a few sentences:

4.57	5.62	4.12	5.29	4.64	4.31	4.30	4.39	4.45
5.67	4.39	4.52	4.26	4.26	4.40	5.78	4.73	4.56
5.08	4.41	4.12	5.51	4.82	4.63	4.29	4.60	

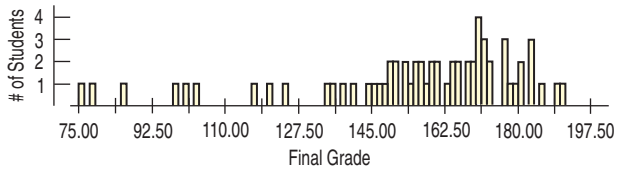
42. Marijuana 2003. In 2003 the Council of Europe published a report entitled *The European School Survey Project on Alcohol and Other Drugs* (www.espad.org). Among other issues, the survey investigated the percentages of 16-year-olds who had used marijuana. Shown here are the results for 20 European countries. Create an appropriate graph of these data, and describe the distribution.

Country	Percentage	Country	Percentage
Austria	21%	Italy	27%
Belgium	32%	Latvia	16%
Bulgaria	21%	Lithuania	13%
Croatia	22%	Malta	10%
Cyprus	4%	Netherlands	28%
Czech Republic	44%	Norway	9%
Denmark	23%	Poland	18%
Estonia	23%	Portugal	15%
Faroe Islands	9%	Romania	3%
Finland	11%	Russia	22%
France	22%	Slovak Republic	27%
Germany	27%	Slovenia	28%
Greece	6%	Sweden	7%
Greenland	27%	Switzerland	40%
Hungary	16%	Turkey	4%
Iceland	13%	Ukraine	21%
Ireland	39%	United Kingdom	38%
Isle of Man	39%		

43. Final grades. A professor (of something other than Statistics!) distributed the following histogram to show the distribution of grades on his 200-point final exam. Comment on the display.

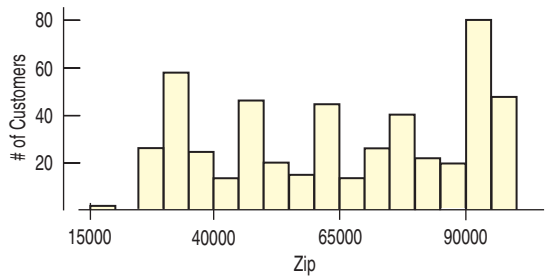


44. Final grades revisited. After receiving many complaints about his final-grade histogram from students currently taking a Statistics course, the professor from Exercise 43 distributed the following revised histogram:



a) Comment on this display.
b) Describe the distribution of grades.

45. Zip codes. Holes-R-Us, an Internet company that sells piercing jewelry, keeps transaction records on its sales. At a recent sales meeting, one of the staff presented a histogram of the zip codes of the last 500 customers, so that the staff might understand where sales are coming from. Comment on the usefulness and appropriateness of the display.



46. Zip codes revisited. Here are some summary statistics to go with the histogram of the zip codes of 500 customers from the Holes-R-Us Internet Jewelry Salon that we saw in Exercise 45:

Count	500
Mean	64,970.0
StdDev	23,523.0
Median	64,871
IQR	44,183
Q1	46,050
Q3	90,233

What can these statistics tell you about the company's sales?

47. Math scores 2005. The National Center for Education Statistics (http://nces.ed.gov/nationsreportcard/) reported 2005 average mathematics achievement scores for eighth graders in all 50 states:

State	Score	State	Score
Alabama	225	Montana	241
Alaska	236	Nebraska	238
Arizona	230	Nevada	230
Arkansas	236	New Hampshire	246
California	230	New Jersey	244
Colorado	239	New Mexico	224
Connecticut	242	New York	238
Delaware	240	North Carolina	241
Florida	239	North Dakota	243
Georgia	234	Ohio	242
Hawaii	230	Oklahoma	234
Idaho	242	Oregon	238
Illinois	233	Pennsylvania	241
Indiana	240	Rhode Island	233
Iowa	240	South Carolina	238
Kansas	246	South Dakota	242
Kentucky	231	Tennessee	232
Louisiana	230	Texas	242
Maine	241	Utah	239
Maryland	238	Vermont	244
Massachusetts	247	Virginia	240
Michigan	238	Washington	242
Minnesota	246	West Virginia	231
Mississippi	227	Wisconsin	241
Missouri	235	Wyoming	243

a) Find the median, the IQR, the mean, and the standard deviation of these state averages.
b) Which summary statistics would you report for these data? Why?
c) Write a brief summary of the performance of eighth graders nationwide.

48. Boomtowns. In 2006, *Inc.* magazine (www.inc.com) listed its choice of "boomtowns" in the United States—larger cities that are growing rapidly. Here is the magazine's top 20, along with their job growth percentages:

City	1-Year Job Growth (%)
Las Vegas, NV	7.5
Fort Lauderdale, FL	4.2
Orlando, FL	4.5
West Palm Beach-Boca Raton, FL	3.4
San Bernadino-Riverside, CA	1.9
Phoenix, AZ	4.4
Northern Virginia, VA	3.1
Washington, DC-Arlington-Alexandria, VA	3.2
Tampa-St. Petersburg, FL	2.6
Camden-Burlington counties, NJ	2.6

(*continued*)

City	1-Year Job Growth (%)
Jacksonville, FL	2.6
Charlotte, NC	3.3
Raleigh-Cary, NC	2.8
Richmond, VA	2.9
Salt Lake City, UT	3.3
Putnam-Rockland-Westchester counties, New York	2.3
Santa Ana-Anaheim-Irvine, CA	1.7
Miami-Miami Beach, FL	2.2
Sacramento, CA	1.5
San Diego, CA	1.4

a) Make a suitable display of the growth rates.
b) Summarize the typical growth rate among these cities with a median and mean. Why do they differ?
c) Given what you know about the distribution, which of the measures in b) does the better job of summarizing the growth rates? Why?
d) Summarize the spread of the growth rate distribution with a standard deviation and with an IQR.
e) Given what you know about the distribution, which of the measures in d) does the better job of summarizing the growth rates? Why?
f) Suppose we subtract from each of the preceding growth rates the predicted U.S. average growth rate of 1.20%, so that we can look at how much these growth rates exceed the U.S. rate. How would this change the values of the summary statistics you calculated above? (*Hint:* You need not recompute any of the summary statistics from scratch.)
g) If we were to omit Las Vegas from the data, how would you expect the mean, median, standard deviation, and IQR to change? Explain your expectations for each.
h) Write a brief report about all of these growth rates.

T **49. Gasoline usage 2004.** The California Energy Commission (www.energy.ca.gov/gasoline/) collects data on the amount of gasoline sold in each state. The following data show the per capita (gallons used per person) consumption in the year 2004. Using appropriate graphical displays and summary statistics, write a report on the gasoline use by state in the year 2004.

State	Gallons per Capita	State	Gallons per Capita
Alabama	529.4	Hawaii	358.7
Alaska	461.7	Idaho	454.8
Arizona	381.9	Illinois	408.3
Arkansas	512.0	Indiana	491.7
California	414.4	Iowa	555.1
Colorado	435.7	Kansas	511.8
Connecticut	435.7	Kentucky	526.6
Delaware	541.6	Louisiana	507.8
Florida	496.0	Maine	576.3
Georgia	537.1	Maryland	447.5
Massachusetts	458.5	Oklahoma	614.2
Michigan	482.0	Oregon	418.4
Minnesota	527.7	Pennsylvania	386.8
Mississippi	558.5	Rhode Island	454.6
Missouri	550.5	South Carolina	578.6
Montana	544.4	South Dakota	564.4
Nebraska	470.1	Tennessee	552.5
Nevada	367.9	Texas	532.7
New Hampshire	544.4	Utah	460.6
New Jersey	488.2	Vermont	545.5
New Mexico	508.8	Virginia	526.9
New York	293.4	Washington	423.6
North Carolina	505.0	West Virginia	426.7
North Dakota	553.7	Wisconsin	449.8
Ohio	451.1	Wyoming	615.0

T **50. Prisons 2005.** A report from the U.S. Department of Justice (www.ojp.usdoj.gov/bjs/) reported the percent changes in federal prison populations in 21 northeastern and midwestern states during 2005. Using appropriate graphical displays and summary statistics, write a report on the changes in prison populations.

State	Percent Change	State	Percent Change
Connecticut	−0.3	Iowa	2.5
Maine	0.0	Kansas	1.1
Massachusetts	5.5	Michigan	1.4
New Hampshire	3.3	Minnesota	6.0
New Jersey	2.2	Missouri	−0.8
New York	−1.6	Nebraska	7.9
Pennsylvania	3.5	North Dakota	4.4
Rhode Island	6.5	Ohio	2.3
Vermont	5.6	South Dakota	11.9
Illinois	2.0	Wisconsin	−1.0
Indiana	1.9		

JUST CHECKING
Answers

(Thoughts will vary.)

1. Roughly symmetric, slightly skewed to the right. Center around 3 miles? Few over 10 miles.
2. Bimodal. Center between 1 and 2 hours? Many people watch no football; others watch most of one or more games. Probably only a few values over 5 hours.
3. Strongly skewed to the right, with almost everyone at $0; a few small prizes, with the winner an outlier.
4. Fairly symmetric, somewhat uniform, perhaps slightly skewed to the right. Center in the 40s? Few ages below 25 or above 70.
5. Uniform, symmetric. Center near 5. Roughly equal counts for each digit 0–9.
6. Incomes are probably skewed to the right and not symmetric, making the median the more appropriate measure of center. The mean will be influenced by the high end of family incomes and not reflect the "typical" family income as well as the median would. It will give the impression that the typical income is higher than it is.
7. An IQR of 30 mpg would mean that only 50% of the cars get gas mileages in an interval 30 mpg wide. Fuel economy doesn't vary that much. 3 mpg is reasonable. It seems plausible that 50% of the cars will be within about 3 mpg of each other. An IQR of 0.3 mpg would mean that the gas mileage of half the cars varies little from the estimate. It's unlikely that cars, drivers, and driving conditions are that consistent.
8. We'd prefer a standard deviation of 2 months. Making a consistent product is important for quality. Customers want to be able to count on the MP3 player lasting somewhere close to 5 years, and a standard deviation of 2 years would mean that lifespans were highly variable.