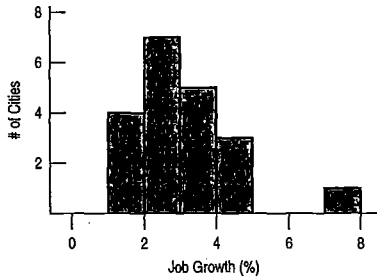
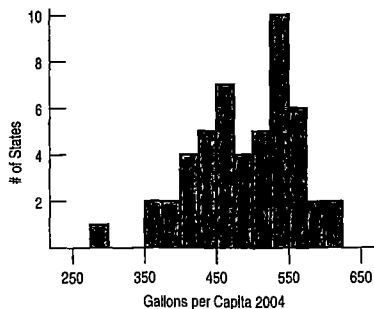


digit gives a rough East-to-West placement in the United States. So, we see that they have almost no customers in the Northeast, but a bar chart by leading digit would be more appropriate.

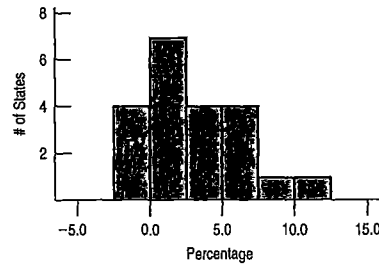
46. Not very much, since zip codes are categorical. However, there is some information in the first digit of zip codes. They indicate a general East (0-1) to West (8-9) direction. So, the distribution shows that a large portion of their sales occur in the West and another in the 32000 area. But a bar chart of the first digits might be a better display of this.
- 47) a) Median 239, IQR 9, Mean 237.6, SD 5.7  
 b) Because it's skewed to the left, probably better to report Median and IQR.  
 c) Skewed to the left; may be bimodal. The center is around 239. The middle 50% of states scored between 233 and 242. Alabama, Mississippi, and New Mexico scores were much lower than other states' scores.
48. a) Here is a histogram of the distribution:



- b) Mean 3.07%, median 2.85%. Mean is higher because of the outlier.  
 c) The median because of the outlier.  
 d) IQR 1.1%, SD 1.37%  
 e) The standard deviation is also influenced by the outlier. The better measure of spread is the IQR.  
 f) Mean and median would be 1.2% lower. SD and IQR would not change.  
 g) Median and IQR won't change very much. The middle value and the two quartiles will shift at most one data value. The mean and SD will decrease.  
 h) The distribution of growth rates for the cities is unimodal and symmetric except for the one outlier, Las Vegas, at 7.5%. The median growth rate for these cities is 2.85%. The middle 50% of the cities had growth rates between 2.25% and 3.35%, for an interquartile range of 1.1%. The median and IQR are the best statistics to report when a distribution has outliers, but if the outlier were omitted, the average growth rate would be 2.84%, with a standard deviation of 0.91%.
49. In the year 2004, per capita gasoline use by state in the United States averaged around 500 gallons per person (mean 488.8, median 500.5). States varied in per capita consumption, with a standard deviation of 68.7 gallons. The only outlier is New York. The IQR of 96.9 gallons shows that 50% of the states had per capita consumption of between 447.5 and 544.4 gallons. The data appear to be bimodal, so the median and IQR are better choices of summary statistics.



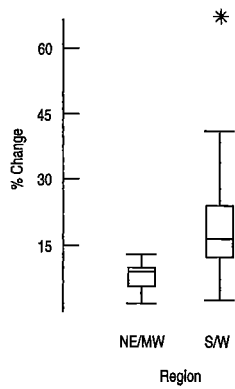
50. In 2005, the median increase in federal prison populations in 20 northeastern and midwestern states was 2.3%; only 4 of the 20 states showed a decrease. The distribution is unimodal and skewed to the right. The large IQR of 4.7% indicates much variability from state to state, with one fourth of these states experiencing prison population increases in excess of 5.5%.



CHAPTER 5

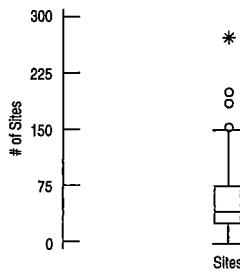
- Answers will vary.
- Answers will vary.
- Answers will vary.
- Answers will vary.
- a) Prices appear to be both higher on average and more variable in Baltimore than in the other three cities. Prices in Chicago may be slightly higher than in Dallas and Denver, but the difference is very small.  
 b) There are outliers on the low end in Baltimore and Chicago and one high outlier in Dallas, but these do not affect the overall conclusions reached in part a).
- a) Coffee, on average, is the most expensive.  
 b) On average, newspapers are more expensive than public transportation rides, but there are cities in which a ride on public transportation costs more than a newspaper in other cities.  
 c) There is one outlying city in each of the three boxplots, but removing this outlier will not affect the conclusions of part a) or b).
- a) Essentially symmetric, very slightly skewed to the right with two high outliers at 36 and 48. Most victims are between the ages of 16 and 24.  
 b) The slight increase between ages 22 and 24 is apparent in the histogram but not in the boxplot. It may be a second mode.  
 c) The median would be the most appropriate measure of center because of the slight skew and the extreme outliers.  
 d) The IQR would be the most appropriate measure of spread because of the slight skew and the extreme outliers.
- a) Both plots show skewness to the high end.  
 b) The histogram shows a bimodal distribution and skewness to the high end. Boxplots can't show multiple modes clearly.  
 c) We'd prefer the median to the mean because of the skewness and possible outliers.  
 d) We'd prefer the IQR to the standard deviation because of the skewness and possible outliers.
- a) About 59%    b) Bimodal  
 c) Some cereals are very sugary; others are healthier low-sugar brands.  
 d) Yes  
 e) Although the ranges appear to be comparable for both groups (about 28%), the IQR is larger for the adult cereals, indicating that there's more variability in the sugar content of the middle 50% of adult cereals.
- a) Unimodal, symmetric    b) 3  
 c) Results from the two procedures were markedly different.  
 d) Biceps transfer  
 e) No, one deltoid transfer was as good as at least 25% of the biceps transfers.  
 f) Although the range in strength is approximately 2 units for both methods, the IQR is much smaller for the deltoid transfers, indicating more consistent results.

11. a)



- b) Growth rates in NE/MW states are tightly clustered near 5%. S/W states are more variable, and bimodal with modes near 14 and 22. The S/W states have an outlier as well. Around all the modes, the distributions are fairly symmetric.

12. a) The distribution is strongly skewed to the right, so use the median and IQR.  
 b) The IQR is 50, so the upper fence is the upper quartile + 1.5 IQRs; that is,  $78 + 75 = 153$ . There appear to be 4 to 5 parks that should be considered as outliers with more than 153 camp sites.  
 c)



- d) The distribution is unimodal with a strong skew to the right. There are several outliers past the  $1.5 \times \text{IQR}$  upper fence of 153 camp sites. The median number of camp sites is 43.5 sites. The mean is 62.8 sites. The mean is larger than the median because it has been influenced by the strong skew and the outliers.

13. a) They should be put on the same scale, from 0 to 20 days.  
 b) Lengths of men's stays appear to vary more than for women. Men have a mode at 1 day and then taper off from there. Women have a mode near 5 days, with a sharp drop afterward.  
 c) A possible reason is childbirth.

14. a) Both are unimodal, skewed to the low end.  
 b) A higher proportion of blacks die between 25 and 74 years old. A higher proportion of whites die at ages older than 75.  
 c) Interval widths are not constant. Most bars are for 10-year intervals, but the first bars are only for ages 0-4, a 5-year span, and the last bars are for 85 and over.

15. a) Both girls have a median score of about 17 points per game, but Scyrine is much more consistent. Her IQR is about 2 points, while Alexandra's is over 10.  
 b) If the coach wants a consistent performer, she should take Scyrine. She'll almost certainly deliver somewhere between 15 and 20 points. But if she wants to take a chance and needs a "big game," she should take Alexandra. Alex scores over 24 points about a quarter of the time. (On the other hand, she scores under 11 points as often.)

16. a) Gas prices increased, on average, over the 3-year period, and the spread increased as well. The prices in 2002 were skewed to the left, with several low outliers. Since then, the distribution has been increasingly skewed to the right. There is a high outlier in 2004.  
 b) 2004 shows both the greatest range and the biggest IQR, so the prices varied a lot.

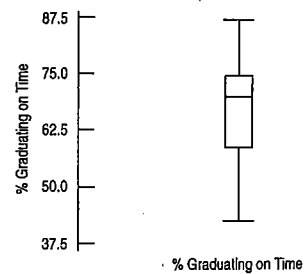
17. Women appear to marry about 3 years younger than men, but the two distributions are very similar in shape and spread.  
 18. Both fuel economy and its spread decrease from 4 to 6 to 8 cylinders (not enough data to compare 5-cylinder cars). The lower 75% of MPG's for the 8-cylinder cars corresponds roughly to the bottom 25% of MPG's for the 6-cylinder cars. All the 8-cylinder cars get less mileage than all the 4-cylinder cars, and their fuel economy is consistently low.  
 19. (Note: Numerical details may vary.) In general, fuel economy is higher in cars than in either SUVs or vans. There are numerous outliers on both ends for cars and a few high outliers for SUVs. The top 50% of cars gets higher fuel economy than 75% of SUVs and nearly all vans. On average, SUVs and vans get about the same fuel economy, although the distribution for vans shows less spread. The range for vans is about 40 mpg, while for SUVs it is nearly 30 mpg.

20. a) April b) February c) August  
 d) The median ozone level in June is slightly higher than in January, but June's readings are much more consistent. June does show two outliers, one low and one high.  
 e) Strong seasonal pattern with low consistent ozone concentrations in later summer/early fall and high variable concentrations in early spring. The medians follow a cyclic pattern, rising from January to April, then falling to October and rising again from October to December.

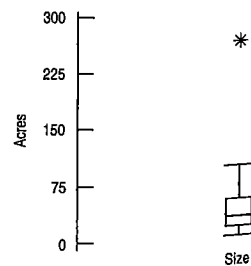
21. The class A is 1, class B is 2, and class C is 3.  
 22. No, boxplots are for quantitative data, and these are categorical, although coded as numbers. The numbers used for hair color and eye color are arbitrary, so the boxplot and any accompanying statistics for eye color make no sense.

23. a) Probably slightly left skewed. The mean is slightly below the median, and the 25th percentile is farther from the median than the 75th percentile.

- b) No, all data are within the fences.  
 c)



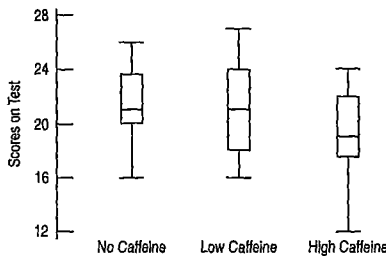
- d) The 48 universities graduate, on average, about 68% of freshmen "on time," with percents ranging from 43% to 87%. The middle 50% of these universities graduate between 59% and 75% of their freshmen in 4 years.  
 24. a) Skewed to the right, since the mean is much larger than the median and the upper quartile is farther from the median than the lower quartile.  
 b) The IQR is about 36, so fences are below 0 and at 109. Since the range is 244 and the minimum is 6, the maximum is 250, which is certainly an outlier. Without knowing the data points, we are not sure of other outliers, but the standard deviation of 47.8 makes us suspect there are others.  
 c) We don't know if there are other outliers above the upper fence, but the boxplot may look something like this:



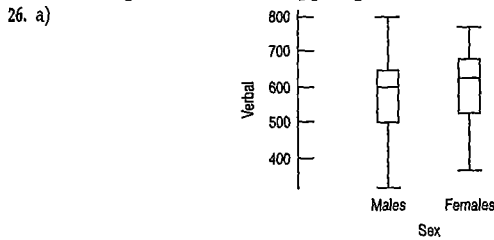
d) The vineyards range in size from 6 to 250 acres. The median size of the 36 vineyards is 33.5 acres, so half are larger and half are smaller. The middle 50% of the vineyards have sizes between 18.5 and 55 acres. The distribution of sizes is skewed to the right, with at least one outlier.

25. a) Who: Student volunteers  
 What: Memory test  
 Where, when: Not specified  
 How: Students took memory test 2 hours after drinking caffeine-free, half-dose caffeine, or high-caffeine soda.  
 Why: To see if caffeine makes you more alert and aids memory retention.

b) Drink: categorical; Test score: quantitative.

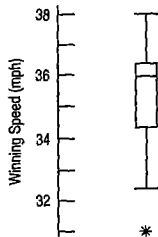


d) The participants scored about the same with no caffeine and low caffeine. The medians for both were 21 points, with slightly more variation for the low-caffeine group. The high-caffeine group generally scored lower than the other two groups on all measures of the 5-number summary: min, lower quartile, median, upper quartile, and max.

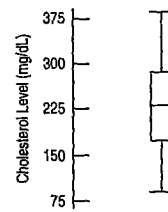


b) Median score by females, at 625 points, is 25 points higher than that by males. Female mean is higher by 12. The middle 50% of females scored between 530 and 680, while the middle 50% of males scored between 515 and 650. The males did have a larger range, from 310 to 800, the highest score for both genders. Both distributions are slightly skewed to the left.

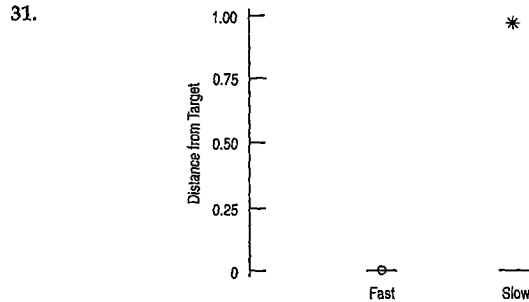
27. a) About 36 mph  
 b)  $Q_1$  about 35 mph and  $Q_3$  about 37 mph  
 c) The range appears to be about 7 mph, from about 31 to 38 mph. The IQR is about 2 mph.  
 d) We can't know exactly, but the boxplot may look something like this:



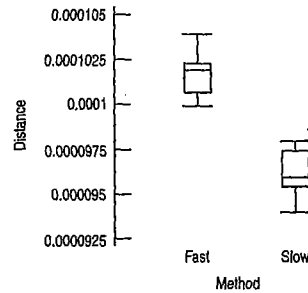
e) The median winning speed has been about 36 mph, with a max of about 38 and a min of about 31 mph. Half have run between about 35 and 37 mph, for an IQR of 2 mph.  
 28. Distribution is essentially symmetric, with median near 225. The IQR is about 75 points. Extremes are about 80 and 380.



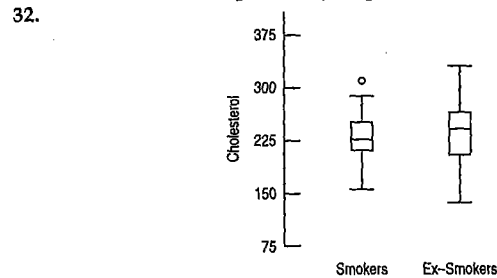
29. a) Boys    b) Boys    c) Girls  
 d) The boys appeared to have more skew, as their scores were less symmetric between quartiles. The girls' quartiles are the same distance from the median, although the left tail stretches a bit farther to the left.  
 e) Girls. Their median and upper quartiles are larger. The lower quartile is slightly lower, but close.  
 f)  $[14(4.2) + 11(4.6)]/25 = 4.38$   
 30. a) The median and IQR, because the means are much larger than the median and the SDs are much larger than the IQR, indicating either right skewness and/or outliers.  
 b) Since the median rainfall for seeded clouds is more than 4 times that for unseeded clouds, it appears that seeding clouds may be effective.



There appears to be an outlier! This point should be investigated. We'll proceed by redoing the plots with the outlier omitted:

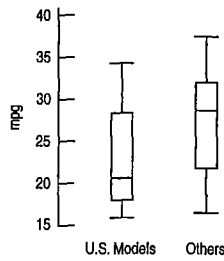


It appears that slow speed provides much greater accuracy. But the outlier should be investigated. It is possible that slow speed can induce an infrequent very large distance.



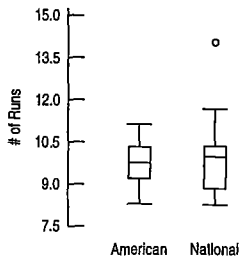
Smoker's distribution has a lower median and is less spread out. There is an outlier in the smoker's group at 351. (Other answers are possible too. Could possibly say that ex-smoker's distribution is more symmetric if histogram or stem-and-leaf provided.)

33. a)



b) Mileage for U.S. models is typically lower, although the variability is about the same as for cars made elsewhere. The median for U.S. models is around 21 mpg, compared to 28 for the others. Half of U.S. models fall below the first quartile of others. (Other answers possible.)

34. a) (Other displays possible.)



b) The National League scores slightly fewer runs in general, but the distribution is more spread out. The median for both leagues is about 9.5, but the American League's is slightly higher. There is a high outlier in the National League at about 14 runs.

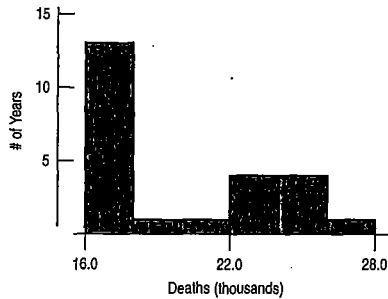
c) Yes, it looks like an outlier. It is nominated by the boxplot rule and is well separated from all the other parks.

35. a) Day 16 (but any estimate near 20 is okay).

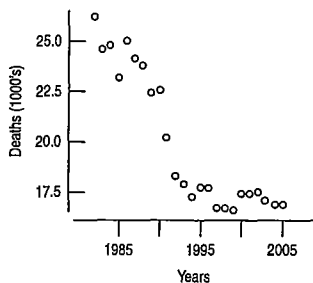
b) Day 65 (but anything around 60 is okay).

c) Around day 50

36. a)



b)



c) The histogram shows a bimodal distribution. The timeplot shows that drunk-driving deaths fell between 1980 and 1995, but after 1995, they have remained at about the same level.

37. a) Most of the data are found in the far left of this histogram. The distribution is very skewed to the right.

b) Re-expressing the data by, for example, logs or square roots might help make the distribution more nearly symmetric.

38. a) The data are extremely skewed to the right. That makes it difficult to determine a center. Our perception of spread will be affected by the very long tail.

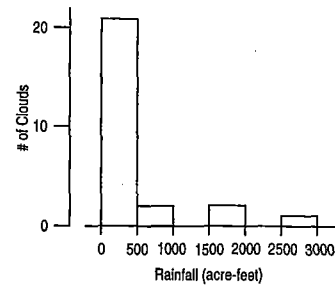
b) Re-expressing the data by, for example, logs or square roots might help make the distribution more nearly symmetric.

39. a) The logarithm makes the histogram more symmetric. It is easy to see that the center is around 3.5 in log assets.

b) That has a value of around 2,500 million dollars.

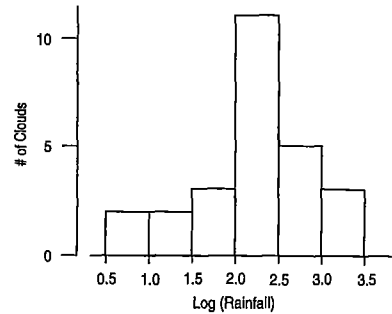
c) That has a value of around 1,000 million dollars.

40. a) The numbers are manageable (each acre-foot is about 320,000 gallons).



b) The distribution is very skewed to the right.

c)



d)  $\log_{10}$  is the exponent of 10, which results in the original value. For example,  $10^3 = 1000$ , so  $\log_{10}$  of 1000 is 3.

41. a) Fusion time and group.

b) Fusion time is quantitative (units = seconds). Group is categorical.

c) Both distributions are skewed to the right with high outliers. The boxplot indicates that visual information may reduce fusion time. The median for the Verbal/Visual group seems to be about the same as the lower quartile of the No/Verbal group.

42. The analysis would probably be improved by using the log-transformed data. The distributions are more symmetric, and it is easier to compare the groups. The outliers are eliminated.

### CHAPTER 6

1. a) 72 oz., 40 oz.

b) 4.5 lb, 2.5 lb

2. a) 264 sec, 138 sec

b) 240 sec, 138 sec

3. a) Skewed to the right; mean is higher than median.

b) \$350 and \$950.

c) Minimum \$350. Mean \$750. Median \$550. Range \$1200. IQR \$600. Q1 \$400. SD \$400.

d) Minimum \$330. Mean \$770. Median \$550. Range \$1320. IQR \$660. Q1 \$385. SD \$440.

4. a) Range 3.30 pounds. IQR 0.95 pounds.

b) Slightly skewed to the left because the mean is lower than the median and the first quartile is farther from the median than the third quartile.