CHAPTER 1

Answers to Chapter Review Exercises 1.59–1.70

1.59 Answers will vary. Examples include: number of employees, value of company stock, total salaries, total profits, total assets, and potential for growth.

1.60 See the *Teacher's Solutions Manual* for graph. The distribution is roughly symmetric with one value (4.88) that is somewhat low. The center of the distribution is between 5.4 and 5.5. The densities range from 4.88 to 5.85, and there are no outliers. We would estimate the Earth's density to be about 5.45 in these units.

1.61 (a) See the *Teacher's Solutions Manual* for graphs. *H. bihai:* 46.34, 46.69, 47.12, 48.293, 50.26. *Red:* 37.4, 38.07, 39.16, 41.69, 43.09. *Yellow:* 34.57, 34.45, 36.11, 36.82, 38.13. *H. bihai* is the tallest; red is generally taller than yellow. Red is more variable.

(b) *H. bihai:* $\bar{x} = 47.597$; $s = 1.213$. *Red:* $\bar{x} = 39.711$; $s = 1.799$. *Yellow:* $\bar{x} = 36.18$, $s = 0.975$.

(c) See the *Teacher's Solutions Manual* for graph. Skewness would make these distributions unsuitable for x and s.

(d) *H. bihai (inches):* $\bar{x} = 1.874$, $s = 0.048$. *Red (inches):* $\bar{x} = 1.563$, $s = 0.071$. *Yellow (inches):* $\bar{x} = 1.424$, $s = 0.0384$.

Chapter Review Exercises

1.59 Top companies 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 Density of the earth In 1798 the English scientist Henry Cavendish measured the density of the earth by careful work with a torsion balance. The variable recorded was the density of the earth as a multiple of the density of water. Here are Cavendish's 29 measurements:[21]

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85	

Present these measurements graphically in a stemplot. Discuss the shape, center, and spread of the distribution. Are there any outliers? What is your estimate of the density of the earth based on these measurements?

1.61 Hummingbirds and tropical flowers Different varieties of the tropical flower *Heliconia* are fertilized by different species of hummingbirds. Over time, the lengths of the flowers and the forms of the hummingbirds' beaks have evolved to match each other. Here are data on the lengths in millimeters of three varieties of these flowers on the island of Dominica:[22]

H. bihai

47.12	46.75	46.80	47.12	46.67	47.43	46.44	46.64
48.07	48.34	48.15	50.26	50.12	46.34	46.94	48.36

***H. caribaea* red**

41.90	42.01	41.93	43.09	41.47	41.69	39.78	40.57
39.63	42.18	40.66	37.87	39.16	37.40	38.20	38.07
38.10	37.97	38.79	38.23	38.87	37.78	38.01	

***H. caribaea* yellow**

36.78	37.02	36.52	36.11	36.03	35.45	38.13	37.10
35.17	36.82	36.66	35.68	36.03	34.57	34.63	

(a) Make boxplots to compare the three distributions. Report the five-number summaries along with your graphs. What are the most important differences among the three varieties of flower?

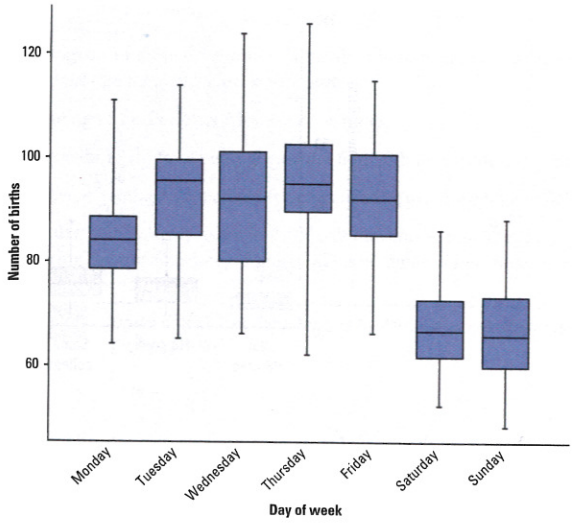
(b) Find $\bar{x}$ and s for each variety.

(c) Make a stemplot of each set of flower lengths. Do the distributions appear suitable for use of $\bar{x}$ and s as summaries?

(d) Starting from the $\bar{x}$ and s-values in millimeters, find the means and standard deviations in inches. (A millimeter is 1/1000 of a meter. A meter is 39.37 inches.)

1.62 Never on Sunday Figure 1.26 shows the distributions of number of births in Toronto, Canada, on each of the 365 days in a year, grouped by day of the week.[23] Based on these plots, give a more detailed description of how births depend on the day of the week.

Figure 1.26 *Side-by-side boxplots of the distributions of numbers of births in Toronto, Canada, for each day of the week during a year, for Exercise 1.62.*



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 2004:

Year:	1948	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992	1996	2000	2004
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	50.7

(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 Income and education level Each March, the Bureau of Labor Statistics compiles an Annual Demographic Supplement to its monthly Current Population Survey.[24] Data on about 71,067 individuals between the ages of 25 and 64 who were employed full-time in 2001 were collected in one of these surveys. The boxplots in Figure 1.27 compare the distributions of income for people with five levels of education. This figure is a variation

CHAPTER 1

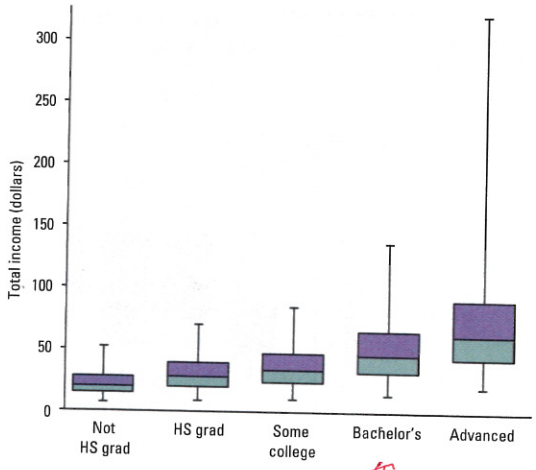
1.62 Answers will vary. There are fewer births on Saturday and Sunday than on other days. Monday has fewer births than the other weekdays.

1.63 (a) See the *Teacher's Solutions Manual* for graph.
(b) 50.7%.
(c) $Q_3 = 57.4\%$. Landslides occurred in 1956, 1964, 1972, and 1984.

CHAPTER 1

1.64 (a) Estimates will vary. The median is at position 7480 in the list, which is about $45,000.
(b) Positions 3740 and 11,220. Their values are about $32,000 and $65,000.
(c) Omitting these observations should have *no* effect on the median and quartiles.
(d) Positions 748 and 14,211.
(e) The "whiskers" on the box extend to approximately $13,000 and $137,000.
(f) *All* five income distributions are skewed to the right. As highest education level rises, the median, quartiles, and extremes rise. The distributions become more and more spread out.

Figure 1.27 *Boxplots comparing the distributions of income for employed people aged 25 to 64 years with five different levels of education, for Exercise 1.64. The lines extend from the quartiles to the 5th and 95th percentiles.*



of the boxplot idea: because large data sets often contain very extreme observations, the lines extend from the central box only to the 5th and 95th percentiles. The data include 14,959 people whose highest level of education is a bachelor's degree.

(a) What is the position of the median in the ordered list of incomes (1 to 14,959) of people with a bachelor's degree? From the boxplot, about what is the median income?

(b) What is the position of the first and third quartiles in the ordered list of incomes for these people? About what are the numerical values of Q_1 and Q_3?

(c) You answered (a) and (b) from a boxplot that omits the lowest 5% and the highest 5% of incomes. Explain why leaving out these values has only a very small effect on the median and quartiles.

(d) About what are the positions of the 5th and 95th percentiles in the ordered list of incomes of the 14,959 people with a bachelor's degree? Incomes outside this range do not appear in the boxplot.

(e) About what are the numerical values of the 5th and 95th percentiles of income? (For comparison, the largest income among all 14,959 people was $481,720. That one person made this much tells us less about the group than does the 95th percentile.)

(f) Write a brief description of how the distribution of income changes with the highest level of education reached. Be sure to discuss center, spread, and skewness. Give some specifics read from the graphs to back up your statements.

1.65 Drive time Professor Moore, who lives a few miles outside a college town, records the time he takes to drive to the college each morning. Here are the times (in minutes) for 42 consecutive weekdays, with the dates in order along the rows:

8.25	7.83	8.30	8.42	8.50	8.67	8.17	9.00	9.00	8.17	7.92
9.00	8.50	9.00	7.75	7.92	8.00	8.08	8.42	8.75	8.08	9.75
8.33	7.83	7.92	8.58	7.83	8.42	7.75	7.42	6.75	7.42	8.50
8.67	10.17	8.75	8.58	8.67	9.17	9.08	8.83	8.67		

(a) Make a histogram of these drive times. Is the distribution roughly symmetric, clearly skewed, or neither? Are there any clear outliers?

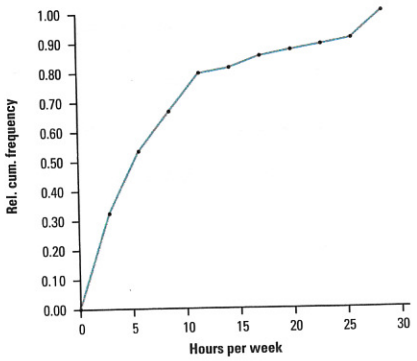
(b) Construct an ogive for Professor Moore's drive times.

(c) Use your ogive from (b) to estimate the center and 90th percentile of the distribution.

(d) Use your ogive to estimate the percentile corresponding to a drive time of 8.00 minutes.

1.66 Computer use Mrs. Causey asked her students how much time they had spent using a computer during the previous week. Figure 1.28 is an ogive of her students' responses.

Figure 1.28 *Ogive of weekly computer use by Mrs. Causey's students, for Exercise 1.66.*



(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

CHAPTER 1

1.65 (a) See the *Teacher's Solutions Manual* for graph. Roughly symmetric with no outliers.
(b) See the *Teacher's Solutions Manual* for graph.
(c) The median is about 8.4 min; the 90th percentile is about 8.8 min.
(d) About the 38th percentile.

1.66 (a) See the *Teacher's Solutions Manual* for graphs.
(b) Median: about 5; Q_1: about 2.5; Q_3: about 11. There are outliers.
(c) At about the 70th percentile.

CHAPTER 1

1.67 (a) −34.04%, −2.95%, 3.47%, 8.45%, 58.68%.
(b) The distribution is roughly symmetric, with high and low outliers.
(c) *Best:* $1586.78. *Worst:* $659.57.
(d) *IQR* = 11.401; there are four "low" and five "high" outliers.

1.68 The distribution of awards is skewed sharply to the right.

first 19 years Wal-Mart was listed on the New York Stock Exchange. There are 228 observations. Figure 1.29 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.

Figure 1.29 *Output from Minitab software describing the distribution of monthly returns from Wal-Mart stock, for Exercise 1.67.*

```
Session
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 : 444444443333332222222222221111111100
  0 : 000001111111111112222223333333344444444
  0 : 55555555555555555555566666666667777777888888888899999
  1 : 000000000111111112223333444 4
  1 : 55566667889
  2 : 011334

High:  32.01923  41.80531  42.05607  57.89474  58.67769
```

(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 × *IQR* criterion? Does it appear to you that the software uses this criterion in choosing which observations to report separately as outliers?

1.68 Jury awards 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 a difference this big between the two measures of center can occur.

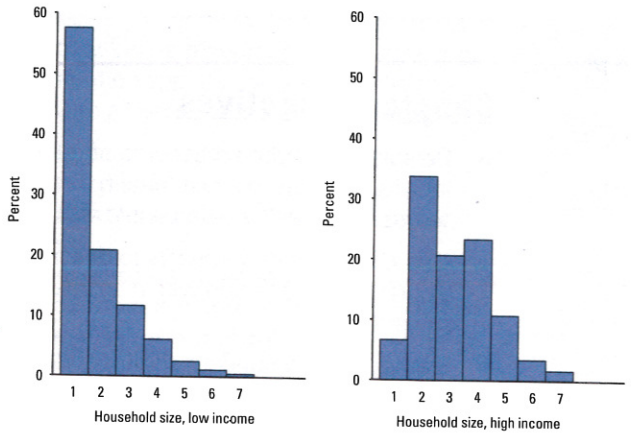
1.69 Women runners Women were allowed to enter the Boston Marathon in 1972. Here are the times (in minutes, rounded to the nearest minute) for the winning woman from 1972 to 2003:

Year	Time	Year	Time	Year	Time	Year	Time
1972	190	1980	154	1988	145	1996	147
1973	186	1981	147	1989	144	1997	146
1974	167	1982	150	1990	145	1998	143
1975	162	1983	143	1991	144	1999	143
1976	167	1984	149	1992	144	2000	146
1977	168	1985	154	1993	145	2001	144
1978	165	1986	145	1994	142	2002	141
1979	155	1987	146	1995	145	2003	145

Make a graph that shows change over time. What overall pattern do you see? Have times stopped improving in recent years? If so, when did improvement end?

1.70 Household incomes Rich and poor households differ in ways that go beyond income. Figure 1.30 displays histograms that compare the distributions of household size (number of people) for low-income and high-income households in 2002.[25] Low-income households had incomes less than $15,000, and high-income households had incomes of at least $100,000.

Figure 1.30 *The distributions of household size for households with incomes less than $15,000 and households with incomes of at least $100,000, for Exercise 1.70.*



(a) About what percent of each group of households consisted of two people?

(b) What are the important differences between these two distributions? What do you think explains these differences?

CHAPTER 1

1.69 See the *Teacher's Solutions Manual* for graph. Times decreased quite rapidly from 1972 until the mid-1980s. Since that time, they have been fairly consistent.

1.70 (a) About 20% of low-income and 33% of high-income households consisted of two people.

(b) The majority of low-income households, but only about 7% of high-income households, consists of one person. One-person households often have less income.