CHAPTER 2
EXERCISE SOLUTIONS

2.1 **a.** 4
b. A state in the United States.
c. $n = 50$.

2.2 **a.** 2
b. A randomly selected person.
c. $n = 620$.

2.3 **a.** Whole population.
b. Sample

2.4 **a.** Whole population.
b. Sample

2.5 **a.** Population parameter.
b. Sample statistic.
c. Sample statistic.

2.6 **a.** Sample statistic.
b. Population parameter.
c. Sample statistic.

2.7 **a.** Sex and self-reported fastest ever driven speed.
b. Students in a statistics class.
c. The answer may vary. If you think the students represent a larger group of individuals, it is sample data. If interest is only in this group of students, or if you think these students do not represent any larger group, it is population data.

2.8 **a.** $n = 2391$.
b. Individuals aged 65 years or older.
c. Frequency of attending religious services and frequency of praying or reading the bible were related to blood pressure,.
d. Sample data. They used the data to make generalizations about a larger population.

2.9 This is a population summary if we restrict our interest only to the fiscal year 1998. (If we were to use this value to represent errors in other years, it could be considered to be a sample summary.)

2.10 **a.** Treatment used (placebo or aspirin) and whether individual died from heart attack or not.
b. Male physicians between 40 and 84 years old.
c. $n = 22{,}071$.
d. Sample data. The used the data to make generalizations about a larger population.

2.11 **a.** Categorical.
b. Quantitative.
c. Quantitative.
d. Categorical.

2.12 **a.** Quantitative.
b. Categorical.
c. Quantitative.
d. Categorical.

2.13 **a.** Not ordinal. It's categorical but the categories are not ordered.
b. Ordinal. Grades are ordered categories.
c. Not ordinal. It's quantitative.

2.14 **a.** Continuous. All weights are possible within an interval of possibilities (although we can't measure accurately enough to observe all possibilities).
b. Not continuous. The number of text messages must be an integer.
c. Not continuous. The number of coins in a pocket would be an integer.

2.15 **a.** Explanatory variable is score on the final exam; response variable is final course grade.
b. Explanatory variable is sex; response variable is opinion about the death penalty.

2.16 **a.** Not ordinal. It's categorical but the categories are not ordered.
b. Ordinal. The ratings are ordered.
c. Not ordinal. It's quantitative.

2.17 **a.** Not continuous. A student could not miss 4.631 classes for example.
b. Continuous. With an accurate enough measuring instrument, any measurement is possible.
c. Continuous. With an accurate enough time piece, any length of time is possible.

2.18 **a.** Explanatory variable is amount person walks or runs per day; response variable is the performance on the lung test.
b. Explanatory variable is age of the respondent; response variable is feeling about religious importance.

2.19 **a.** Whether a person supports the smoking ban or not is a categorical variable.
b. Gains on verbal and math SATs are quantitative variables.

2.20 The explanatory variable is smoker or not. The response variable is Alzheimer sufferer or not. Both variables are categorical.

2.21 **a.** Sex and pulse rate.
b. Sex is categorical, pulse rate is quantitative.
c. Is there a difference between the mean pulse rates of men and women? The sample mean pulse rate for each sex would be useful.

2.22 This will differ for each student. As an example, suppose a survey question about income only allowed the response categories 1 = under $20,000 and 2= $20,000 to $49,999 and 3= more than $49,999. The income categories are ordered, but so little is known about actual income that the mean response is meaningless.

2.23 This will differ for each student. One example where numerical summaries would make sense for an ordinal variable is the response to the question "What grade do you expect in this class? 1=A, 2=B, 3=C, 4=D, 5=F." The mean numerical response is an expected class GPA.

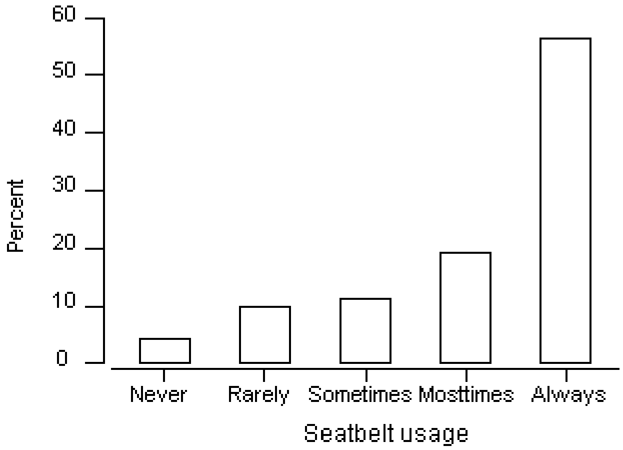
2.24 This will differ for each student.

2.25 **a.** A unit is a person. Dominant hand is a categorical variable and IQ is a quantitative variable. Explanatory variable is dominant hand and response variable is IQ .
b. A unit is a married couple. Eventual divorce status and pet ownership are both categorical variables. Explanatory variable is pet ownership and response variable is eventual divorce status.

2.26 **a.** A unit is a college student. GPA and hours of study each week are both quantitative variables. Explanatory variable is hours of study and response variable is GPA.
b. A unit is a tax-paying individual in the United States. Tax bracket is an ordinal variable and percentage donated to charities is a quantitative variable. Explanatory variable is tax bracket and response variable is percentage donated to charities.

2.27 **a.** 1427/2530 = .564, which is 56.4%.
b. 1 – (1427/2530) = .436, which is 43.6%.
c. Never: 105/2530 = .042 (4.2%); Rarely: 248/2530 = .098 (9.8%); Sometimes: 286/2530 = .113 (11.3%). Most times: 464/2530 = .183 (18.3%); Always: 1427/2530 = .564 (56.4%)
d.

Figure for Exercise 2.27d

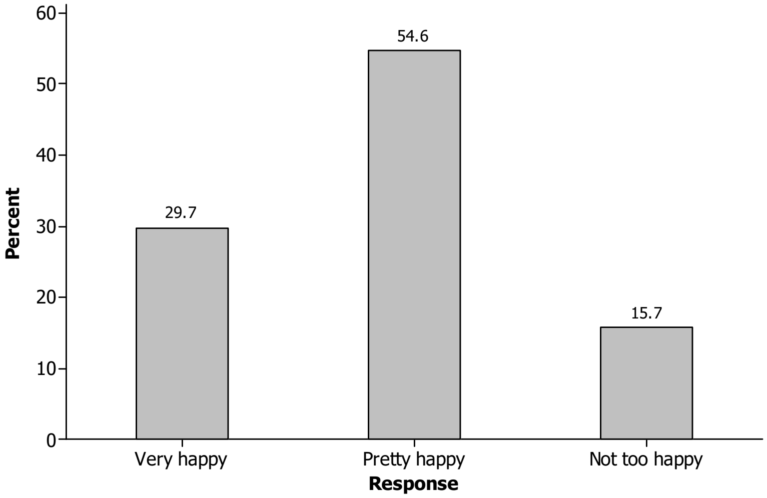


2.28 **a.**

Response	Frequency	Relative frequency
Very happy	599	599/2015 = .297 (29.7%)
Pretty happy	1100	1100/2015 = .546 (54.6%)
Not too happy	316	316/2015= .157 (15.7%)
Total	2015	1 (100%)

b.

Figure for Exercise 2.28b



c. 29.7% + 54.6% = 84.3%

2.29 **a.**

	Preferred use of cell phone		
	To talk	To text	Total
Women	22 (20.8%)	84 (79.2%)	106 (100%)
Men	34 (41.0%)	49 (59.0%)	83 (100%)

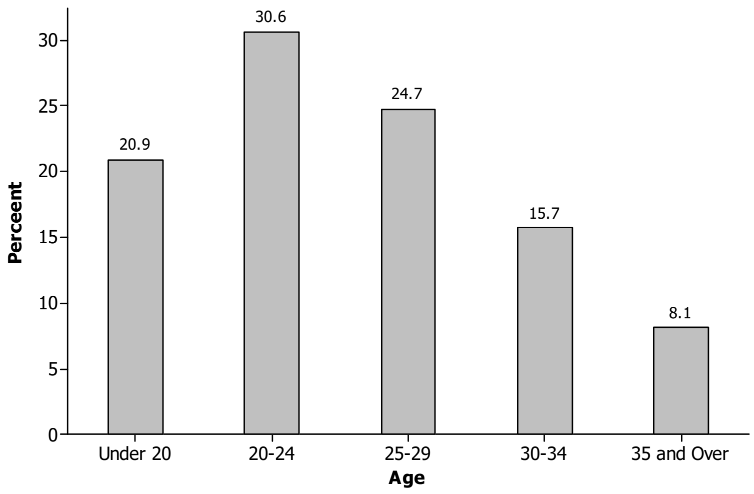
b. Women: 20.8% to talk, 79.2% to text
c. Men: 41.0% to talk, 59.0% to text
d. Women were more likely to say “to text” than men whereas men were more likely to say “to talk.”

2.30 **a.** 1700/2470 = .688, or 68.8%
b. 1056/1700 = .621, or 62.1%
c. 300/657 = .457, or 45.7%
d. 41/113 = .363, or 36.3%

2.31 **a.** Explanatory variable is whether a person smoked or not. Response variable is whether they developed Alzheimer’s or not.
b. Explanatory variable is political party. Response variable is whether a person voted or not.
c. Explanatory variable is income level. Response variable is whether a person has been subjected to a tax audit or not.

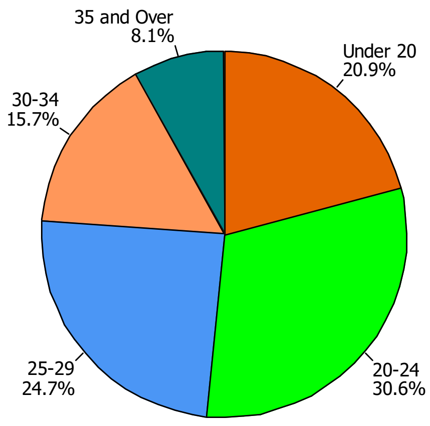
2.32 **a.**

Figure for Exercise 2.32a



b.

Figure for Exercise 2.32b



c. The pie chart may more effectively show that there are three age groups with large percentages, and it may be faster to read these percentages than with the bar chart. One problem, however, is that the age groups are shown in a circular pattern, an unnatural way to view the age. The bar chart gives a better sense of the distribution of ages because the ages are shown along a more natural horizontal number line.

2.33 **a.** The explanatory variable is sex and the response variable is how they feel about their weight.

b.

	Feelings About Weight			
Sex	**Overweight**	**About right**	**Underweight**	**Total**
Female	38 (26.6%)	99 (69.2%)	6 (4.2%)	143
Male	18 (23.1%)	35 (44.9%)	25 (32.1%)	78

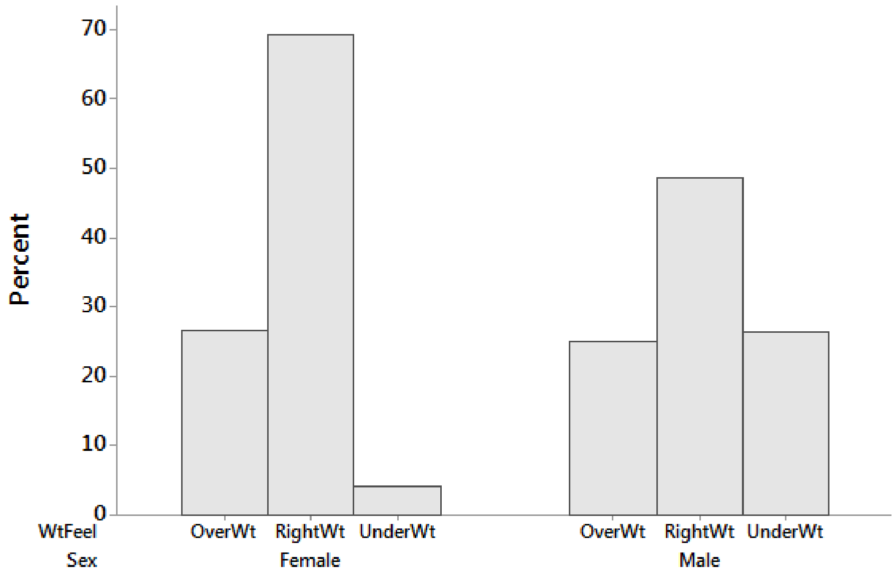
c. Feeling overweight: 38/143 = .266, or 26.6%; right weight: 99/149 = .692, or 69.2%; underweight: 6/149 = .042, or 4.2%.

d. Feeling overweight: 18/78 = .231, or 23.1%; right weight: 35/78 = .449, or 44.9%; underweight: 25/78 = .321 or 32.1%.

e. Males are more likely than females to feel that they are underweight; females are more likely than males to say that their weight is about right.

2.34

Figure for Exercise 2.34



2.35 **a.**

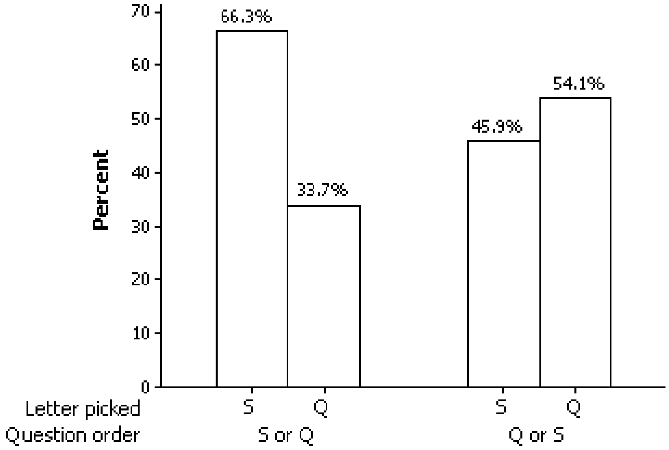
	Picked S	Picked Q	Total
S listed first	61	31	92
Q listed first	45	53	98
Total	106	84	190

b. Picked S $=(61/92)\times 100\% = 66.3\%$; Picked Q $=(31/92)\times 100\% = 33.7\%$;

c. Picked S $=(45/98)\times 100\% = 45.9\%$; Picked Q $=(53/98)\times 100\% = 54.1\%$;

d.

Figure for Exercise 2.35d



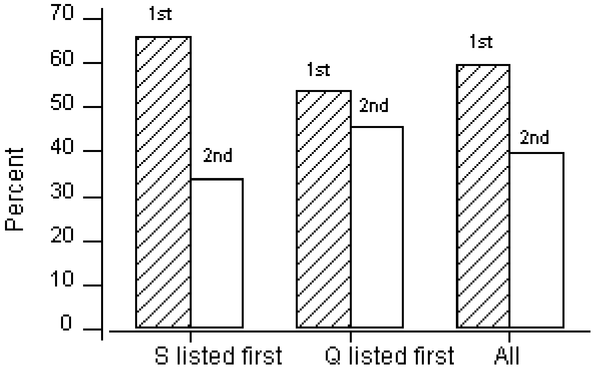
e. Parts (b) and (c) show that the percentage picking S was higher when S was listed first than when Q was listed first. It looks like the letter picked was influenced by the letter listed first.

2.36 **a.** The columns can be labeled "Picked 1st letter" and "Picked 2nd letter." In the "S listed first" row, list the counts in the same order as in the table for Exercise 2.35(a). In the "Q listed first" row, the count for "Q picked" should be in the "Picked 1st letter" column (because Q was the first letter). The table is

	Picked 1st letter	**Picked 2nd letter**	Total
S Listed First	61 (66%)	31 (34%)	92
Q Listed First	53 (54%)	45 (46%)	98
All	114 (60%)	76 (40%)	190

b. A bar chart of percentages picking first and second letters given on each form of the question (row percentages in the table above) along with these percentages for the overall sample follows:

Figure for Exercise 2.36b



c. The variables used in this exercise are more appropriate for illustrating the point of this data set. The question of interest is whether participants might be more likely to pick the fist letter given than the second regardless of whether it was an S or Q. The bar chart given for part (b) illustrates that the fist letter listed was picked more often for both forms.

2.37 **a.** The fastest speed was 150 miles per hour.

b. The slowest speed driven by a male was 55 miles per hour.
c. 1/4 of the females reported having driven at 95 miles per hour or faster. Notice that 95 mph is the *upper quartile* for females. By definition, about 1/4 of the values in a data set are greater than the upper quartile.
d. 1/2 of the females reported having driven 89 mph or faster. Notice that 89 mph is the *median* value.
e. 1/2 of 102 = 51 females have driven 89 mph or faster.

2.38 **a.** The median value is 110 mph for males, compared to 89 mph for females.
b. The spread is about the same for the two sexes. The spread of the extremes is 150− 55 = 95 mph for the males, compared to 130 – 30 = 100 mph for the females. The spread of the quartiles is slightly greater for males (120 – 95 = 25 for males, compared to 95-80 = 15 for females.)

2.39 **a.** The center for the females is at a greater percentage than it is for the males. For females the center looks to be somewhere around 27%. For males, the center looks to be a bit less than 18%.
b. The data are more spread for the females.
c. The greatest two female percentages are set apart from the bulk of the data. The values are about 65% and 72%.

2.40 **a.** Median height = 65 inches.
b. Range = Tallest – Shortest = 71 – 59 = 12 inches.
c. The interval from 59 to 63.5 inches contains the shortest 1/4 of the women.
This interval is from the minimum to the lower quartile.
d. The interval from 63.5 to 67.5 inches contains the middle 1/2 of the women.
This is an interval from the lower quartile to the upper quartile.

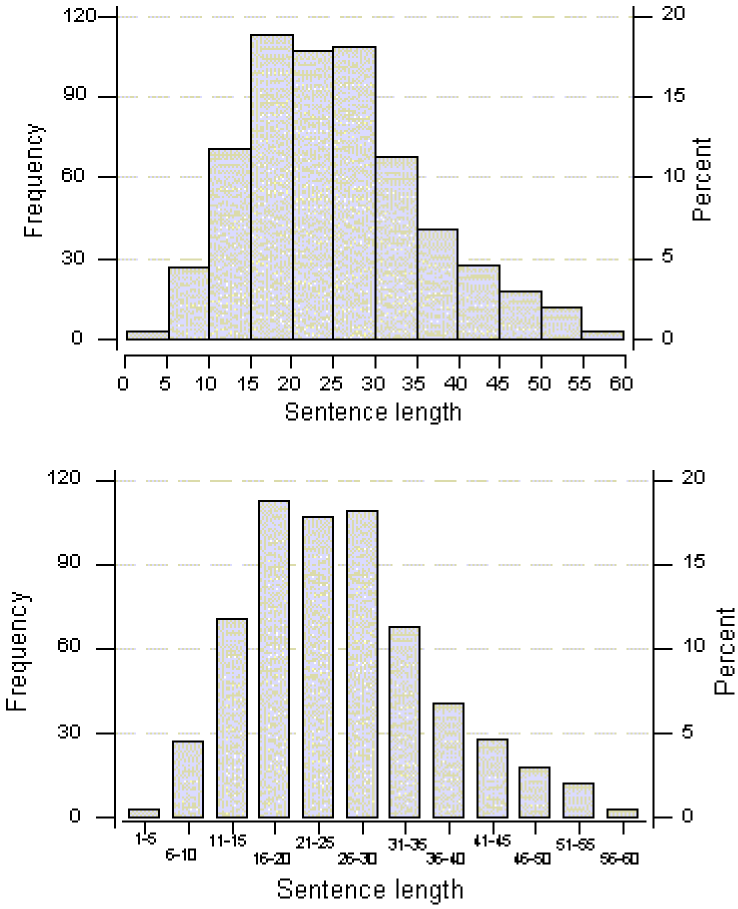
2.41 **a.** The median value, 65 inches, describes the location.
b. The interval described by the extremes, 59 to 71 inches, describes spread. We might also describe spread using the interval 63.5 to 67.5, the spread of the middle 50% of the data.

2.42 **a.** The dataset is skewed to the right (it stretches in that direction).
b. The value 13 looks to be an outlier. It is separate from the bulk of the data.
c. 2 ear pierces was the most reported value. About 44 or so women said they had this many ear pierces.
d. About 32 or so women said they had 4 ear pierces.

2.43 **a.** The dataset looks approximately symmetric and bell-shaped.
b. There are no noticeable outliers.
c. The most frequently reported value for sleep was 7 hours.
d. Roughly 14 or so students said they slept 8 hours the previous night.

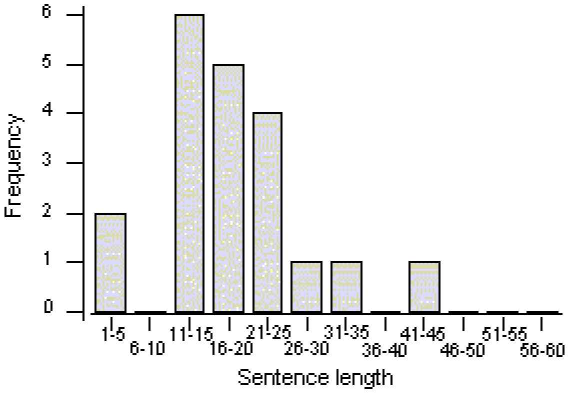
2.44 **a.** Two slightly different versions of the histogram are shown below. In the first, the bars touch each other; in the second, the bars are separated and labeled with the category limits. The first histogram looks a little nicer but there could be confusion about the exact endpoints of intervals. Notice that we have shown a frequency axis on the left and the corresponding relative frequency (percentage) axis on the right.

Figures for Exercise 2.44a



b. A majority of the sentences have between 16 and 30 words. The percentage of sentences with between 16 and 30 words is (113+107+109)/600 ≈55%. With regard to shape, the data are skewed to the right.
c. A stem-and-leaf plot presents all individual values, but we do not have the data in this form. It has already been tallied for specified intervals.
d. The lead-in to the definition of statistics given on page 1 might cause some confusion. If the lead-in (beginning "A more complete definition, and ...") and the definition are counted together as one sentence, that sentence has 42 words. Because the definition is set apart in a box, many may divide these 42 words into two sentences - one with 17 words and one with 25 (the definition itself). The histogram shown below is for the first possibility (one sentence of 42 words).

Figure for Exercise 2.44d



The histogram of number of words in a sentence in the first twenty sentences of Chapter 1 is also skewed to the right. In general, however, the number of words per sentence is less than that in the *Shorter History of England.*

2.45 **a.** The dataset looks approximately symmetric and bell-shaped.
b. The highest temperature is 92°F.
c. The lowest temperature is 64°F.
d. 5/20 = .25, which is 25%.

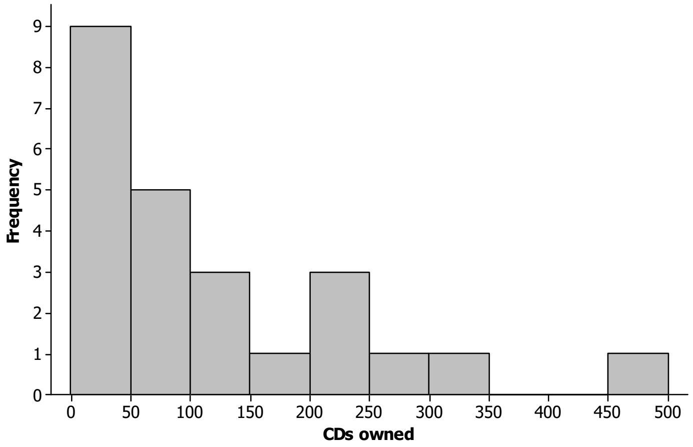
2.46 **a.** The following stem-and-leaf uses "hundreds" for stems (row labels), "tens" for leaves within the rows and truncates the "ones" digit. Each "hundreds" is represented by two separate stems, one for leaves 0-4 and one for leaves 5-9.

Figure for Exercise 2.46a

```
|0| 001222233
|0| 55569
|1| 002
|1| 5
|2| 002
|2| 5
|3| 0
|3|
|4|
|4| 5
```

b. In the following histogram, values tied with the lower value of an interval are counted into that interval. For example, 50 is counted into the interval that spans from 50 to 100.

Figure for Exercise 2.46b



c. The data are skewed to the right with a possible outlier (at 450).

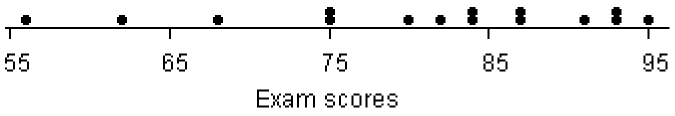
2.47 **a.** The figure below uses separate stems for last digits 0-4 and 5-9. That's not imperative, although doing so give more detail of the shape of the data.

Figure for Exercise 2.47a

```
|5| 6
|6| 2
|6| 8
|7|
|7| 55
|8| 02344
|8| 77
|9| 13
|9| 5
```

b.

Figure for Exercise 2.47b



2.48 **a.** The following histogram uses nine intervals. Others are possible.

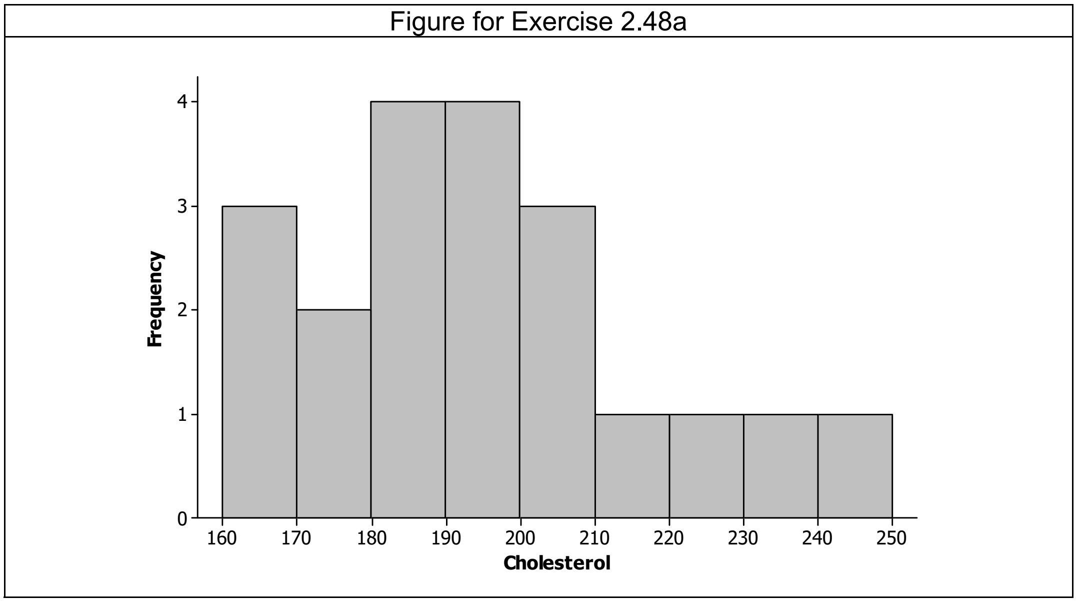


Figure for Exercise 2.48a

b. In the following stem-and-leaf, the first two digits of the values are used for stems (row labels).

Figure for Exercise 2.48b

```
|16| 046
|17| 88
|18| 2248
|19| 6888
|20| 046
|21| 2
|22| 2
|23| 0
|24| 2
```

c. There are no obvious outliers.
d. The shape is difficulty to judge – perhaps a slight skew to the right.

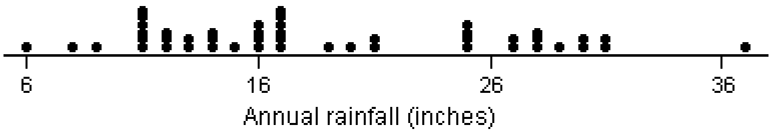
2.49 **a.** In the figure shown here, two stems have been used for each possible "tens" place in the number (values under 10 are an exception because the lowest value is about 6 inches). We rounded the 1995 total of 24.5 inches up to 25.

Figure for Exercise 2.49a

```
|0| 689
|1| 11111122233444
|1| 566667777779
|2| 011
|2| 5555778889
|3| 0011
|3| 7
```

b.

Figure for Exercise 2.49b



c. The data are skewed (but only slightly) to the right.

2.50 *Note*: The data for males are in the dataset **pennstate1M** on the companion website.

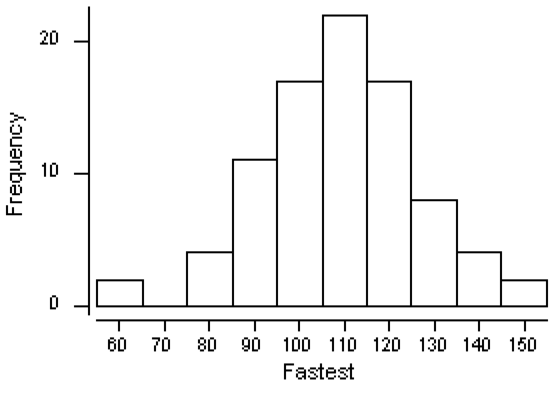
a.

Figure for Exercise 2.50a

```
| 5| 5
| 6| 0
| 7|
| 8| 00005555
| 9| 0000024555555
|10| 00000000012555555559
|11| 0000000000002555555
|12| 00000000004555555
|13| 00
|14| 00005
|15| 0
```

b. The histogram shown here was done with Minitab. Notice that the midpoints of intervals are shown under the bars.

Figure for Exercise 2.50b



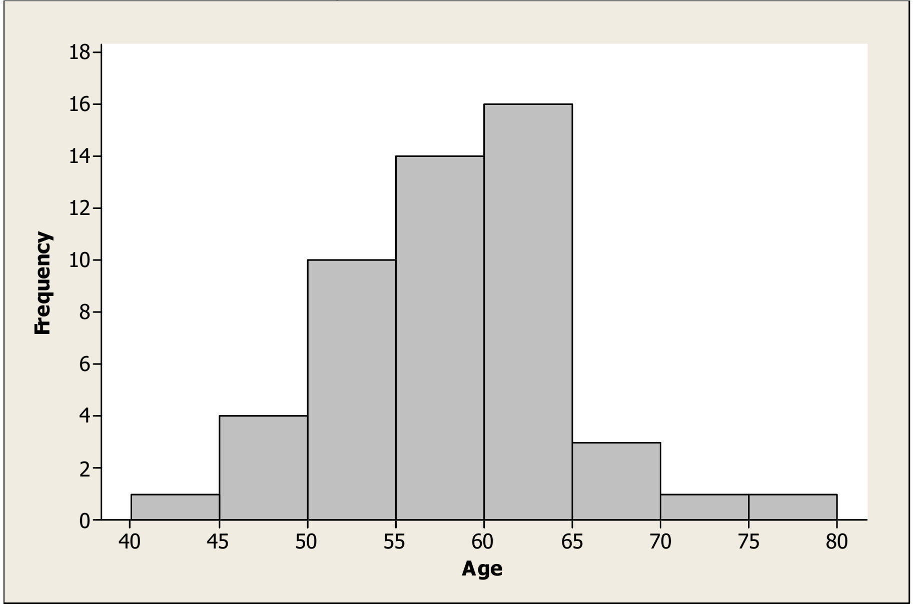
c. The histogram may provide the best view of the overall shape and characteristics of the distribution. The information given in the stem-and-leaf plot is equivalent to that given in the histogram, but is somewhat harder to read. An advantage of the stem-and-leaf is that it's possible to determine individual values. The dotplot gives much information about individual values as well as the range and general location. It is, however, more difficult to judge the shape of the distribution with a dotplot.

d. These data are symmetric in shape (and there may be two outliers).

2.51 Yes, a stem-and-leaf plot provides sufficient information to determine whether a dataset contains an outlier. Because all individual values are shown, it is possible to see whether are any values are inconsistent with the bulk of the data.

2.52 **a.** The answer will vary due to the flexibility possible for deciding on the endpoints of intervals. The histogram shown here is based on 5-year age groups and the age under the left edge of a bar is included in that interval while the age under the right edge is not.

Figure for Exercise 2.52a



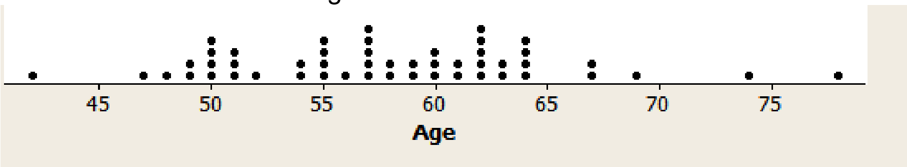
b. Two stems should be used per decade of ages because there should be 6 to 15 stem values. [Note: If only one stem value is used per decade, there would be only 5 values, and if 5 were used per decade there would be 22 (because only 4 would be needed for the 30s and 3 for the 70s).] Notice that if this stem-and-leaf plot were turned on its side, it would have the same shape as the histogram shown for part (a).

Figure for Exercise 2.52b

```
|4| 2
|4| 7899
|5| 0000111244
|5| 55556777778899
|6| 0001122222334444
|6| 779
|7| 4
|7| 8
```

c.

Figure for Exercise 2.52c



d. The CEO ages may be skewed slightly to the left although this is not a pronounced skew.. It may be fair to say that the data are roughly bell-shaped.

e. There do not seem to be any obvious outliers although there is a bit of a gap between the oldest two CEOs and the third oldest and also a small gap between the youngest and the others.
f. This is a conjecture, but it seems more likely that an outlier would occur in the salaries. A person's salary can become very high, and a company might conceivably pay an extraordinary salary to a successful executive.

2.53 Skewed to the left. Along a number line, values stretch more toward the left (small values).

2.54 Answers will differ. One approach is to combine two distinctly different groups into the same dataset. As an example, a histogram of the heights of a sample that includes equal numbers of adult women and six-year old girls will be bimodal.

2.55 **a.** Histogram is better than a boxplot for evaluating shape.
b. A boxplot is useful for identifying outliers, evaluating spread, and for comparing groups.

2.56 The answers will differ for each student.
a. You may be most interested in knowing the average value because it would provide some information about what kind of salary to expect. You may also like to know the spread because the average value would be less important if the annual salary of employees varied widely.
b. Each summary has interest here. The maximum would give information about whether an A is possible with each instructor. For a generally average student, the average might have the most interest. The spread would give information about whether most students performed about the same or whether there was great variability among students.
c. The average may be the most informative about personal life expectancy, although the maximum and the spread would also be useful and interesting general information.

2.57 Generally, females tended to have higher tip percentages. The median is clearly greater for females. The data for the females also shows greater spread than the data for the males.

2.58 **a.** Median = 70. Ordered list of data is 67, 68, 69, 70, 72, 73, 74; median is middle value in this ordered list.
b. Mean = 70.43

2.59 **a.** Mean = 74.33; median = 74.
b. Mean = 25; median = 7.
c. Mean = 27.5; median = 30.

2.60 100 is a large value compared to the rest of the values; it causes the mean to increase, while not affecting the median.

2.61 **a.** 225 −123 = 102 lbs
b. 190 −155 = 45 lbs
c. 50%

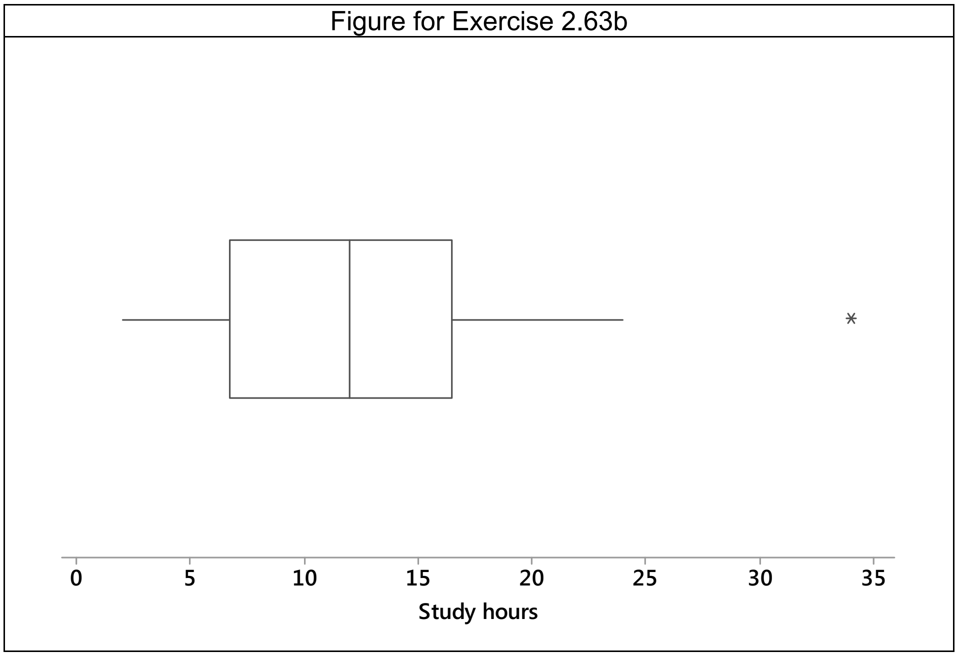
2.62 **a.** 12 letters.
b. 13 letters.
c. The IQR for males is 17−10 = 7 while for females it is 15−10 = 5. The IQR is larger for males.
d. 23 – 6 = 17 letters.
e. 23 – 6 = 17 letters.

2.63 **a.** Min = 2, Q_1 = 7, Median = 12, Q_3 = 16, Max = 34
The ordered data follow. A vertical bar indicates the location of the median. Values of Q_1 and Q_3 are underlined and in bold, They are the medians of the lower and upper halves of the data, respectively.

2 3 4 6 6 **7** 8 9 10 11 12 | 12 13 14 15 15 **16** 18 21 22 24 34

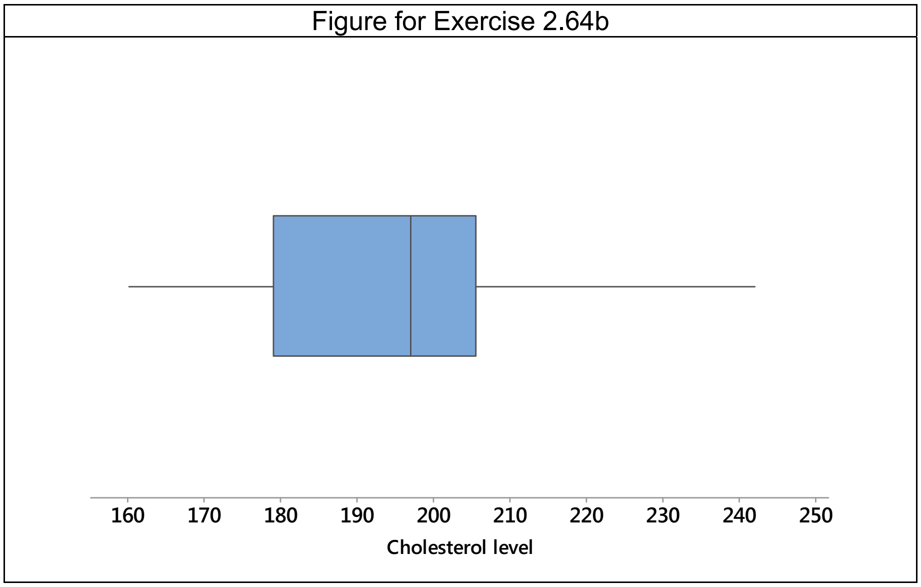
b. The boxplot is below. The lines extend at most 1.5 IQR from the ends of the box, which is 1.5(16 – 7) = 13.5, but stop at the Min and Max if those are reached first. Therefore, the line at the lower end stops at 2, but the line at the upper end extends to (16 + 13.5) = 29.5. Note that the value 34 is marked as an outlier. It exceeds the upper boundary for an outlier, which is Q3 + 1.5 IQR = 16 + 1.5 (16 − 7) = 29.5.

Figure for Exercise 2.63b



2.64 **a.** Min = 160, Q_1 = 180, Median = 197, Q_3 = 205, Max = 242
The ordered data follow. Vertical bars indicates the locations of Q_1, the median, and Q_3, respectively.
160 164 166 178 178 | 182 182 184 188 196 |
198 198 198 200 204 | 206 212 222 230 242
b. The boxplot is below. The lines extend at most 1.5 IQR from the ends of the box, which is 1.5(205 – 180) = 37.5, but stop at the Min and Max if those are reached first. Therefore, the line at the lower end stops at 160 instead of extending to Q1 – 37.5 = 180 – 37.5 = 142.5, and the line at the upper end stops at 242 instead of extending to (205+37.5) = 242.5.

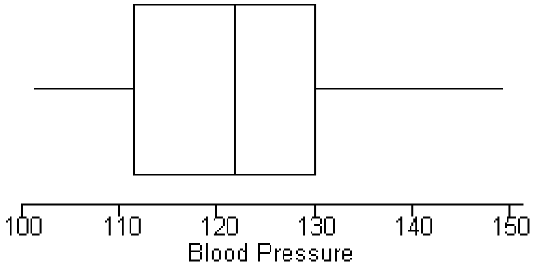
Figure for Exercise 2.64b



2.65 **a.** Min = 109, Q_1 = 180.75, Median = 186, Q_3 = 199.0, Max = 214.0
The data in order are 109.0, 178.5, 183.0, 185.0, 186.0, 188.5, 194.5, 203.5, 214.0.
Median is the middle value (= 186). Lower quartile is median of 109.0, 178.5, 183.0, 185, so equals (178.5 + 183).2 = 180.75. Upper quartile is median of 188.5, 194.5, 203.5, 214, so equals (194.5 + 203.5)/2 = 199.0
b. 109.0 is an outlier. It's below the lower outlier boundary = $Q_1 -$ 1.5 *IQR* = 180.75 – 1.5 (199 – 180.75) = 153.375.
c. The member who weighed 109.

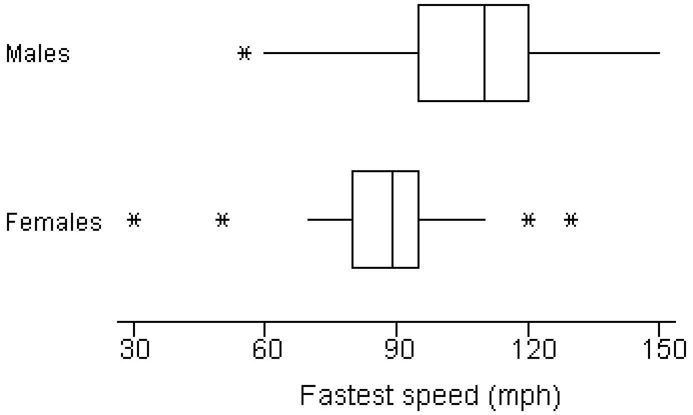
2.66 **a.** (122+123)/2 = 122.5.
b. Q_1 = 114 and Q_3 = 129.5.
c. IQR = $Q_3 - Q_1$ = 129.5–114 = 15.5.
d. 1.5×IQR = 23.25. A number will be considered an outlier if it is either below 114–23.25 = 90.75 or above 129.5+ 23.25 = 152.75. No values fit this criterion, so there are no outliers.
e.

Figure for Exercise 2.66e



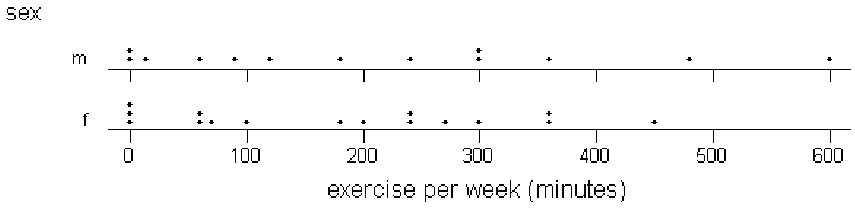
2.67 The boxplots below show that, on average, the fastest speeds ever driven by males tend to be higher than the fastest speeds ever driven by females. It is also seen, if outliers are ignored, that the spread is greater for males than it is for females. A horizontal axis has been used for fastest speeds here, but a vertical axis would be equally appropriate.

Figure for Exercise 2.67



2.68 **a.** The amount of exercise per week is similar for men and women. The dotplot follows.

Figure for Exercise 2.68a



b. *Women*, median = 190 minutes. *Men*, median = 180 minutes.
For *women*, the ordered list of data is:

0, 0, 0, 60, 60, 70, 100, **180**, **200**, 240, 240, 270, 300, 360, 360, 450

The number of women is even (16), so the median is the average of the middle two values in the ordered list. These middle two values, underlined in the above list, are 180 and 200 and their average is 190.
For *men*, the ordered list of data is:

0, 0, 14, 60, 90, 120, **180**, 240, 300, 300, 360, 480, 600

The number of men is odd (13) so the median is the middle value in the ordered data. This value, underlined in the list above is 180.

2.69 **a.**

	Minutes of exercise per week		
Median		180	
Quartiles	42		330
Extremes	0		600

To determine the summary, first write the responses in order from smallest to largest.
The ordered list of data is:

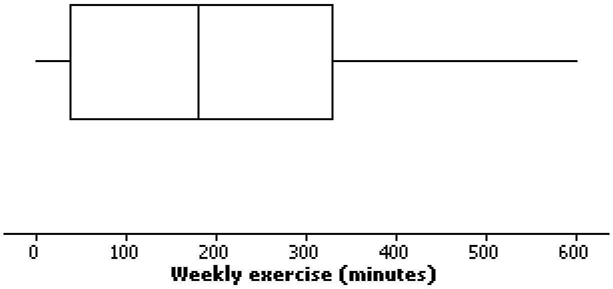
0, 0, 14, 60, 90, 120, 180, 240, 300, 300, 360, 480, 600

M*inimum* = 0 min.
Maximum = 600 min.
Median = 180 min. (middle value in the ordered list)
Lower quartile = 42 min. It is the median of the values smaller than the median.
These are 0, 0, 14, 60, 90, 120.
Median of these six values is (14+70)/2 = 42.
Upper quartile = 330 min. It is the median of the values larger than the median.
Values larger than the median are 240, 300, 300, 360, 480, 600.
Median of these values is (300+360)/2=330.

b. All men in the sample reported exercising between 0 and 600 minutes per week. The median response was 180 min. About 1/2 of the men reported exercising between 42 and 330 minutes per week. About 1/4 said they exercised less than 42 minutes per week while 1/4 said they exercised more than 330 minutes per week.

c.

Figure for Exercise 2.69c



2.70 The median age at death for First Ladies is 71 years old. One woman lived until the age of 97 years and two died in their mid-30s. The five-number summary shows that one-half of the ages at death are between 60 and 83 years (the lower and upper quartiles). The extreme points seem to be the most interesting here. We may wonder what illness or disease caused Martha Jefferson and Hannah Van Buren to die so young.

2.71 The median annual rainfall in Davis, CA is 16.72 inches and the mean is 18.62 inches. The data values vary from 6.14 inches (in 1965) to 37.42 inches (in 1982). Although not an extreme outlier, the 1982 value is separated somewhat from the other values. The 1982 value of 37.42 inches is about 6 inches more than the next highest total, which occurred in 1981. It is interesting to note that the two highest values were in consecutive years.

2.72 The rainfall data are skewed to the right. We therefore expect the mean to be higher than the median. This relationship is true; the mean is 18.69 inches and the median is 16.72 inches.

2.73 There are n =47 values so the median is 24th value in the ordered data (23 will be smaller and 23 will be larger). The lower quartile is the median of the smallest 23 values and the upper quartile is the median of the largest 23 values. These will be, respectively, the 12th lowest and the 12th highest values in the data set. The five-number summary is:

	Rainfall (inches)		
Median		16.72	
Quartiles	12.05		25.37
Extremes	6.14		37.42

The five-number summary above shows that the median annual rainfall for Davis, California is 16.72 inches. The middle ½ of the values are between 12.05 and 25.37 inches. The minimum is 6.14 inches and the maximum is 37.42 inches.

2.74 **a.** Mean = +2.455; Median = −7. To find the median, first order the values, and then determine the middle value. The ordered data are:

−14 −13 −11 −9 −7 −7 −2 4 24 24 38

b. The median point difference is a better summary of the team's season. They lost seven games and only won four, so a negative difference was the more typical experience.

2.75 Although the raw data are not available, we can determine that the median is somewhere between 21 and 25 words. There are n=600 sentences, so the median occurs between the 300th and 301st observations in the ordered data. The total number of sentences with 20 or fewer words is the sum of the frequencies for categories up to and including the 16 to 20 words category. This is (3+27+71+113) =224 sentences. There are 107 sentences in the 21 to 25 words category, so there are 224+107 = 331 sentences with 25 or fewer words, The 300th and 301st values must be somewhere between 21 and 25 words.

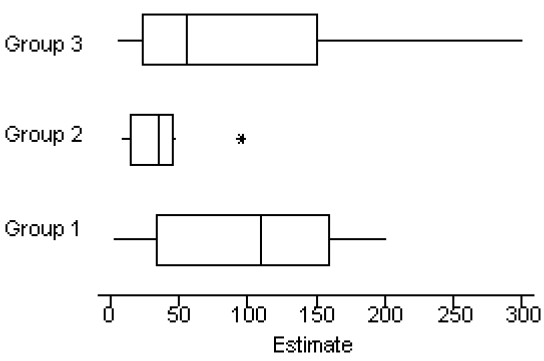
Each quartile is the average of the 150th and 151st value from the appropriate side of the ordered data. Using the same reasoning that we did for the median, the first quartile must be somewhere between 16 and 20 words, and the third quartile is somewhere between 31 and 35 words. As a result, the *IQR* might be as low as 31−20=11 words or as high as 35−16= 19 words.

The exact values of the minimum and maximum aren't given, but the minimum may be about 3 or so and the maximum might be about 57, the range is likely to be somewhere around 57−3=54.

2.76 **a.** (100+120)/2 = 110 million.
b. 35 million.
c. 55 million.

d.

Figure for Exercise 2.76d



d. Group 1, range = (200–2) = 198 million. Group 2, range = (95–8) = 87 million. Group 3 had a range of (300-5) = 295 million. Therefore, Group 3 had the largest range and Group 2 had the smallest range

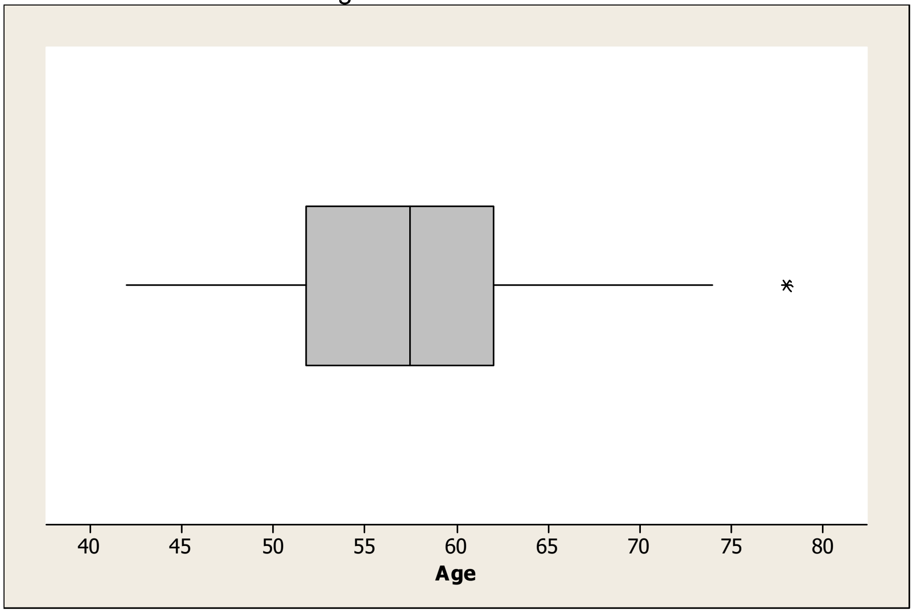
2.77 **a.** The median is the average of the middle two values. There were n=50 ages, so the lower quartile is the median of the 25 lowest ages and the upper quartile is the median of the 25 highest ages.

	CEO ages (years)		
Median		57.5	
Quartiles	52		62
Extremes	42		78

The five-number summary shows that the median age of the 50 CEOs is is 57.5 years. The middle ½ of the CEOs have ages between 52 and 62 years. The youngest CEO is 42 years old. The oldest CEO is 78 years old.

2.78 The boxplot can be made either with a horizontal axis (as shown here) or a vertical axis. Note that the oldest CEO will be marked as an outlier. The boundary for an upper outlier is 62 + 1.5*(62-52) = 77. The oldest CEO is 78.

Figure for Exercise 2.78



2.79 The mean of the CEO ages is 57.84 years. The median is 57.5 years. The mean and the median are similar. This is expected because the data are more or less symmetric in shape.

2.80 **a.** You need the rest of the data, perhaps graphed in one of the ways described in this section. Then you can see whether the value 16 hours is far from the bulk of the data or not.
b. You should determine whether this is a legitimate data value. For instance, check to see if the value was incorrectly entered into the computer. Or, check the student's response form to see if there's evidence that he might have generally been giving false information. The observation should not be removed if it is a legitimate value reflecting the natural variability of the variable.

2.81 This will differ for each student. One example is that a person 80 years old would be an outlier at a traditional college, but would not be an outlier at a retirement home.

2.82 You can not say whether this value is an outlier. You need to know whether it is the right hand-span measurement of the male author or of the female author.

2.83 The rainfall total for 1982 is high (37.42 inches) but not high enough to be classified as an outlier. There are no potential outliers at the lower end.

2.84 Answers will vary.

2.85 Most likely a mistake was made when the data were entered. If possible, the instructor should correct the value (by looking again at the student’s survey form). If the correct height is not available, the value 17 should be deleted from the dataset.

2.86 **a.** Mean =20; Standard deviation = 1.581.
b. Mean =20; Standard deviation = 0.
c. Mean = 20; Standard deviation = 33.09.

2.87 **a.** Mean ± St. Dev is 7 ± 1.7, or 5.3 to 8.7.
b. Mean ± 2 St. Dev is 7 ± (2)(1.7), or 3.6 to 10.4.
c. Mean ± 3 St. Dev is 7 ± (3)(1.7), or 1.9 to 12.1.

2.88 **a.** Mean ± St. Dev is 71 ± 5, or 66 to 76.
b. Mean ± 2 St. Dev is 71 ± (2)(5), or 61 to 81.
c. Mean ± 3 St. Dev is 71 ± (3)(5), or 56 to 86.

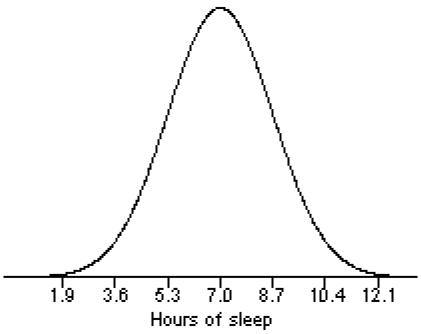
2.89 **a.** Mean = 25, s = 4.24. Calculation of s is is $s = \sqrt{\dfrac{(22-25)^2+(27-25)^2+(30-25)^2+(21-25)^2}{4-1}}$

b. Mean = 30, s = 9.13. Calculation of s is is $s = \sqrt{\dfrac{(25-30)^2+(35-30)^2+(40-30)^2+(20-30)^2}{4-1}}$

2.80 **a.** z = (200–170)/20 = 1.5.
b. z = (140–170)/20 = –1.5.
c. z = (170–170)/20 = 0.
d. z = (230–170)/20 = 3.

2.91

Figure for Exercise 2.91



2.92 All values in List 1 are at the extremes (10 and 30), so the list has the maximum possible overall deviation from the mean for data with this range. List 2 has the same range as List 1, but generally the values tend to be closer to the center than the values in List 1. Standard deviation measures overall deviation from the mean.

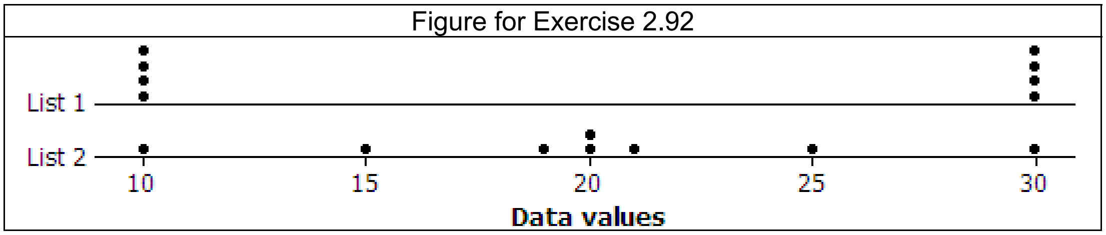


Figure for Exercise 2.92

2.93 **a.** $z = (300–350)/100 = −0.5$.
b. $z = (460–350)/100 = 1.1$.
c. $z = (650–350)/100 = 3.0$.
d. $z = (210–350)/100 = −1.4$..

2.94 **a.** $\bar{x} = 123$, s = 14.452. Calculation of s is $s = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}} =$

$$\sqrt{\frac{(110-113)^2+(123-123)^2+(132-123)^2+(150-123)^2+(127-123)^2+(118-123)^2+(102-123)^2+(122-123)^2}{8-1}}$$

b. $s^2 = 208.857$**.** In the calculation of the standard deviation in part (a), this is the value under the square root sign. It might also be calculated as $s^2 = (14.452)^2$.

2.95 The only possible set of numbers is {50, 50, 50, 50, 50, 50, 50} because a standard deviation of 0 means there is no variability.

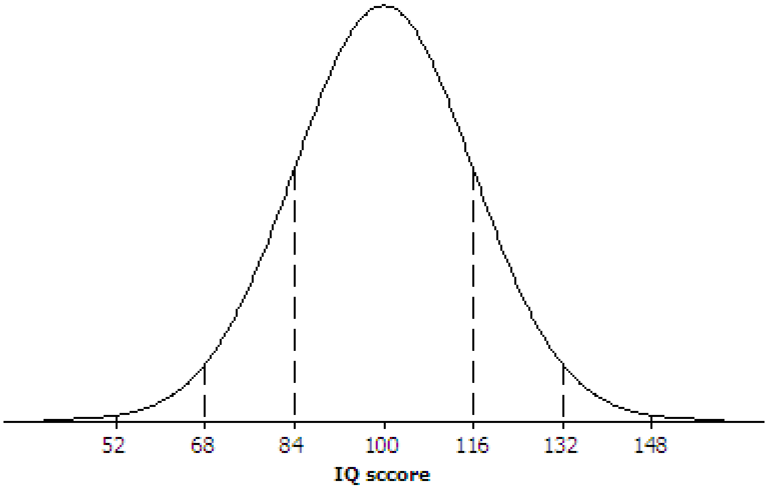
2.96 You should be more satisfied if the standard deviation was 5. This would mean you scored 2 standard deviations above the mean and, if scores are bell-shaped, only about 2.5% of students are expected to score higher.

2.97 **a.** 98–41 = 57.
b. Standard deviation ≈ Range/6 = 57/6 = 9.5.

2.98 **a.** About 68% fall in the interval Mean ± St. Dev, which is 100 ± 16, or 84 to 116.
About 95% fall in the interval Mean ± 2 St. Dev, which is 100 ± (2)(16), or 68 to 132.
About 99.7% fall in the interval Mean ± 3 St. Dev, which is 100 ± (3)(16), or 52 to 148.

b.

Figure for Exercise 2.986b



c. $s^2 = 16^2 = 256$ **(**Variance = squared standard deviation.)

2.99 The Empirical Rule says that 68% of values fall within 1 standard deviation of the mean, 95% fall within 2 standard deviations of the mean, and 99.7% fall within 3 standard deviations of the mean. Of the 103 hand-span measurements for women, 74 or 72% are within 1 standard deviation of the mean (18.2 to 21.8 cm). 100 of the 103 or 97% are within 2 standard deviations. 101 of the 103 or 98% are within 3 standard deviations. This data seems to fit pretty well with the Empirical Rule.

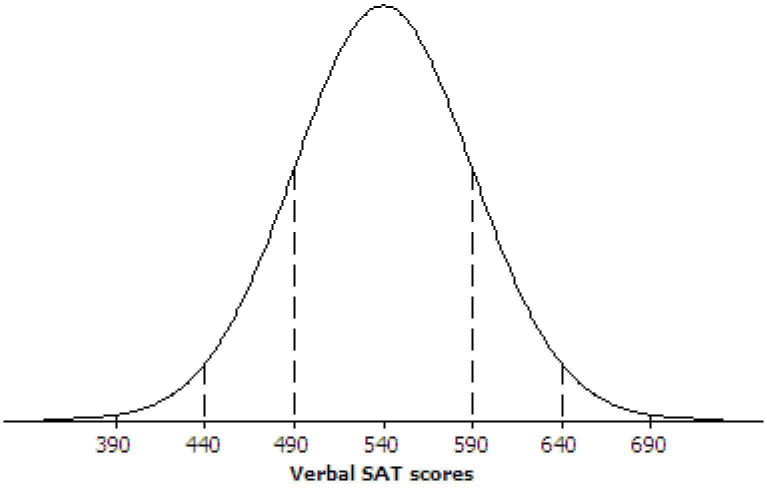
2.100 **a.** If the two lowest values are deleted, the mean will increase and the standard deviation will decrease.
b. The Empirical Rule for mean ± 3 standard deviations says that 99.7% of the values will be between $20.2 \pm 3(1.45)$ or 15.85 and 24.55 cm. All of the data, or 100% of the values, are within this interval.
c. Looking at the figures, it seems like the Empirical Rule should hold when the outliers are removed. The data looks pretty symmetric without those two values. If the outliers are not removed, the Empirical Rule may hold, but not as well, since the data seem more skewed to the left with those two points included.
d. There may be justification for removing the outliers if a convincing argument can be made that they are errors. The value of 12.5 may really be an incorrect entry of 21.5. The value of 13 may really be an incorrect entry of 18 or 23. Assuming the original surveys were available, this could be checked. Or, you could see if either of these women is extremely short or had any other odd measurements.

2.101 **a.** A 52-centimeter head circumference will not occur often, but it will occur. The value 52 is 2 standard deviations below the mean ($z = \frac{\text{value - mean}}{\text{s.d.}} = \frac{52-56}{2} = -2$). This is at the lower end of the interval that describes about 95% of the values. Thus only about 2.5% of male head circumferences are smaller.
a. A 62-centimeter head circumference will be rare. The value 62 is 3 standard deviations above the mean ($z = \frac{\text{value - mean}}{\text{s.d.}} = \frac{62-56}{2} = 3$). This is at the upper end of the interval that describes about 99.7% of the values. Thus only about 0.15% (about 3 in 2000 men) will have a larger circumference.

2.102 $s^2 = 2^2 = 4$ **(**Variance = squared standard deviation.)

2.103 **a.** About 68% fall in the interval 540 ± 50, which is 490 to 590.
About 95% fall in the interval 540 ± (2)(50), which is 440 to 640.
About 99.7% fall in the interval 540 ± (3)(50), which is 390 to 690.

Figure for Exercise 2.103a



b. $s^2 = 50^2 = 2500$ (Variance = squared standard deviation.)

2.104 **a.** 52 cm, which is 2 standard deviations below the mean.
b. 60 cm, which is 2 standard deviations above the mean.
c. 58 cm, which is 1 standard deviation above the mean.

2.105 **a.** 590, which is 1 standard deviation above the mean.
b. 640, which is 2 standard deviations above the mean.
c. 490, which 1 standard deviation below the mean.

2.106 **a.** $z = \dfrac{\text{value} - \text{mean}}{\text{standard deviation}} = \dfrac{450 - 500}{100} = -0.5$, and the proportion below is .3085.

b. $z = \dfrac{\text{value} - \text{mean}}{\text{standard deviation}} = \dfrac{36.5 - 34}{1} = 2.5$, and the proportion below is .9938.

c. $z = \dfrac{79 - 75}{8} = 0.5$, and the proportion below is .6915.

d. $z = \dfrac{79 - 75}{4} = 1$, and the proportion below is .8413.

2.107 A categorical variable cannot have a bell-shaped distribution. A variable must be quantitative for it to be possible to have a distribution with any particular shape. For a categorical variable, the raw data are category labels without a meaningful numerical ordering.

2.108 No, the standard deviation is not a resistant statistic. An outlier inflates the standard deviation because an outlier creates additional deviation from the mean in a data set. As an example, suppose that a set of 5 pulse rates is 71, 73, 74, 75, 77. For these data, the mean is $\bar{x}$=74 and the standard deviation is $s = \sqrt{5} = 2.24$. Add the value 50 to the data as a sixth observation. For the new data set, $\bar{x}$=70 and s = 10.
Note: Software like Minitab or Excel, or possibly a calculator, could be used to determine the standard deviation(s) for the example. The "by hand" calculation would be done with the formula $\sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n-1}}$.
For the example given here, if the outlier included, the mean is $\bar{x} = 70$. The following table shows the squared difference between each data value and the mean.

x_i	50	71	73	74	75	77
$x_i - 70$	–20	1	3	4	5	7
$(x_i - 70)^2$	400	1	9	16	25	49

Using the entries in the third row of the table, $s = \sqrt{\dfrac{400+1+9+16+25+49}{6-1}} = \sqrt{\dfrac{500}{5}} = 10$.

Notice how much the squared difference between 50 and 70 contributes to the size of the answer.

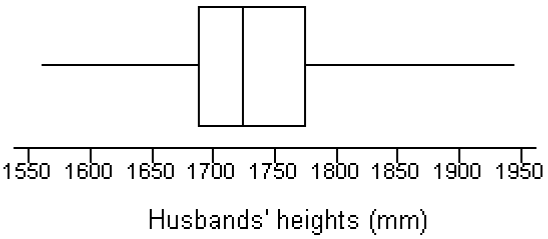
2.109 **a.** The lower quartile Q1 = 0 hours. By definition, 25% of the values are at or below the lower quartile.
b. 0 to 2 hours (Minimum to Median).
c. 2 to 70 (Median to Maximum).
d. Yes, 70 hours would be marked as an outlier.
The boundary defining an outlier on the high side is Q3 + 1.5×IQR = 5 + 1.5 (5-0) = 12.5 hours.
e. Range/6 = (70-0)/6 = 11.67.This is notably greater than the standard deviation. The outlier (70 hours) and skewness in the data cause Range/6 to differ from the standard deviation.
f. The mean is greater than the median. The outlier and skewness to the right causes this to occur.

2.110 The answers will differ for each student.
a. Most may prefer the data value to be near the average. If the data value were an outlier, the number of children would be excessive, not something most would want (although some may prefer this).
b. An outlier on the high side seems preferable. Most would want to make more money than everyone else does.
c. An outlier on the high side would be preferable because that would be great gas mileage. But, an outlier on the low side is not desirable.
d. An outlier on the low side would be desirable because most would like to live in a town with a really low crime rate. In reality, there may not be outliers on the low side because many small towns have low crime rates, so a single low rate may not stand apart from other values. In that case, the average would be desirable.

2.111 **a.** Telephone exchange is a categorical variable.
b. Number of telephones is a quantitative variable.
c. Dollar amount of last month's phone bill is a quantitative variable
d. Long distance phone company used is a categorical variable.

2.112 **a.**

Figure for Exercise 2.116



b. *Range = maximum–minimum* = 1949–1559 = 390 mm.
The range often spans a distance of about 6 standard deviations, so the estimated standard deviation is *Range*/6 =390/6 = 65 mm.
c. The approximation is appropriate if it is assumed that the heights follow a bell-shape.
d. The interval *mean* ± 3 *standard deviations* should cover about 99.7% of the values. This interval is 1732.5 mm ± (3 × 68.8 mm), which is 1732.5 mm ± 206.4 mm. The interval spans from 1526.1 mm to 1938.9 mm so it does not cover the maximum value of 1949 mm.

2.113 **a.** Yes, a variable can be both explanatory and categorical. The phrase "explanatory variable" means that the variable might influence a response variable, and there is no restriction concerning whether the explanatory variable is categorical or quantitative (or ordinal). For an example, consider Example 2.2 in which type of nighttime lighting is a categorical explanatory variable.

b. No, a variable cannot be both continuous and ordinal. The term "ordinal variable" is used when the raw data are ordered categories while "continuous" means that all values in an interval are possible.

Supplemental note: Some might ask whether the term "ordinal" could apply to a continuous variable because the raw data can be used to order the sample observations. A restriction of ordinal numbers, however, is that all possible values can be counted. For a continuous variable, however, all possible values in an interval cannot be counted. There always are an infinite number of values between any two points in the interval so it's impossible to determine what "exact" value is second, third, and so on.

c. Yes, a variable can be both quantitative and a response variable. Generally, there is no restriction concerning whether a response variable is quantitative, categorical, or ordinal. For an example, consider Example 2.5 in which the quantitative response variable is right handspan (and the explanatory variable is sex).

d. Yes, a variable can be both bell-shaped and a response variable. An example is verbal SAT, which is designed to have bell-shaped distribution, and would be a response variable in a comparison of the verbal SAT scores of females and males.

2.114 **a.** This will differ for each student.

b. Kind of coin (penny, nickel, etc.) is an ordinal variable because the kinds can be ordered by their monetary value.

c. The total monetary value of the coins is a quantitative variable.

d. “Kind of coin” is the data of interest when answering the question “Which coin occurred most often?” “Total monetary value of the coins” is the data of interest in answering the question “What is the average amount of change per student?”

2.115 **a.** The mean will be larger than the median. While most households may have between 0 and 4 or so children, there will be some households with large numbers of children, so the distribution will be skewed to the right.

b. The mean will be larger than the median. People like Bill Gates will create large outliers. And, generally income data tends to be skewed to the right because high incomes can become quite high but incomes can't be any lower than 0.

c. If all of the high school students are included, the mean will be higher than the median. This is because many high school students are too young to work or do not want to work, resulting in many students with $0 income earned in a job outside the home. There is even a chance the median could be 0!

d. The mean is 10.33 cents. Calculate this assuming there is one of each type of coin. The calculation is (1+5+25)/3 = 31/3 = 10.33. The exact number of each type of coin doesn't matter. As long as there are equal numbers of each type, the mean will be 10.33 cents. The median is the middle amount so it will be 5 cents. The mean is higher than the median because the monetary amounts are skewed to the right.

2.116 This will differ for each student.

2.117 The interval is −6.1 to 22.7 hours. The interval includes negative values, which are impossible times. Thus, the interval based an assumption of a bell-shaped curve would not reflect reality.

2.118 This will differ for each student.

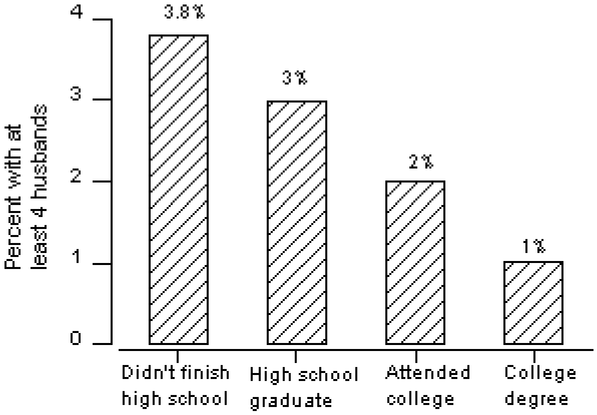
2.119 **a.** The First Ladies may constitute a population rather than a sample. They lived in unique circumstances, so it is hard to view these women as a representative sample from any larger population. And, they can't be considered to be a sample from a larger population of First Ladies because future First Ladies will have different circumstances affecting life expectancy.

b. If the First Ladies are viewed as a population, the population standard deviation is σ= 15.37 years. In Excel, this can be found with the command "=STDEVP( )" and many calculators have a key for the population standard deviation. If the argument is made in part (a) that the First Ladies constitute a sample, the correct answer here is that the sample standard deviation is s 15.57 years.

2.120 For each part of this problem, a bar chart comparing percentages provides the clearest visual display. Notice that each part involves a comparison of groups (or years) with regard to the percentage with a specified trait. The percentage not having the trait in each group could be calculated and included in a graph, but we think the comparisons are clearer if this is not done in parts (a) and (b). In part (c), there may be merit to showing the percentages that do and do not take part in regular activity.

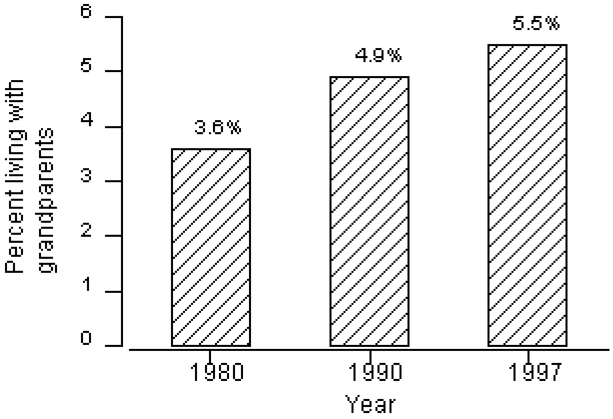
a.

Figure for Exercise 2.120a



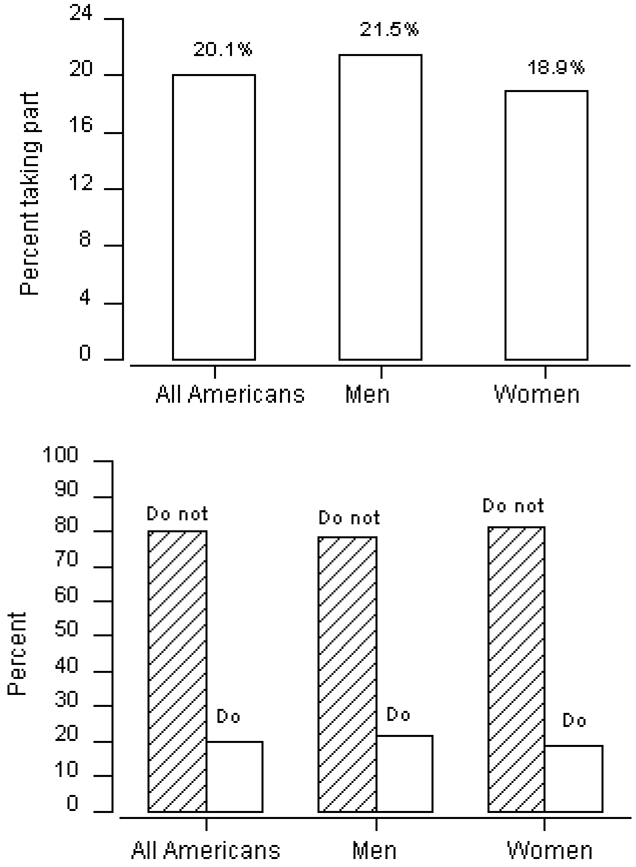
b.

Figure for Exercise 2.120b



c. The first bar chart shown below displays only the given information. The second bar chart displays the percentages that do and do not take part in regular activity.

Figure for Exercise 2.120c



2.121 **a.** Amount of beer consumed is the explanatory variable. Systolic blood pressure is the response variable.
b. Daily caloric intake of protein is the explanatory variable. Presence of colon cancer is the response variable.

2.122 **a.** If a *z*-score is 0, the value must equal the mean.
b. Begin by setting the formula for a *z*-score equal to 1.

$$\frac{\text{observed value} - \text{mean}}{\text{standard deviation}} = 1$$

Two steps of algebra lead to *observed value* = *mean* + 1 *standard deviation*.
Another strategy is to make observed value = *mean* + 1 *standard deviation* in the *z*-score formula. Algebraic simplification leads to $z = 1$.

2.123 **a.** If the two possible outliers are ignored, the data appear to be more or less bell-shaped so the Empirical Rule may hold.
b. The Empirical Rule implies that the range should span about 4 to 6 standard deviations. About 95% of the data will be within 2 standard deviations (plus or minus) of the mean and about 99.7% of a data set should be within 3 standard deviations (plus or minus) of the mean. Here, *range* = *maximum* − *minimum* = 23.25 – 12.5 = 10.75 cm. This span is equal to 10.75/1.8 = 5.97 standard deviations so it is consistent with the Empirical Rule.

2.124 **a.** The heights of all of the children in an elementary school will have a larger standard deviation because there will be a wide variety of different aged students included. This will lead to a wide variety of heights.

b. The systolic blood pressure for 30 people who visit a health clinic in one day will have a larger standard deviation because there is more variability from person to person as far as blood pressure is concerned than for measurements made on the same person from day to day.
c. The SAT scores for the students in an honors class will have a larger standard deviation. Even though it is an honors class and most students will do well on the SAT, the range of SAT scores is likely to be much larger than the range of scores on an English final examination, which at can be at most 0 to 100. This larger range will lead to a larger standard deviation.

2.125 **a.** The mean is 57.84 years and the sample standard deviation is $s = 6.997$ years.
Note: If this batch of data is viewed as a population rather than a sample, the standard deviation is 6.926 years. See the "technical note" in section 2.7 for an explanation of the difference between sample and population standard deviations.
b. *Range* = *maximum* – *minimum* = (78 – 42) = 36 years. This is a span of $42/6.997 = 5.14$ standard deviations, so the stated relationship between range and standard deviation does hold for the data.
c. For the youngest CEO, $z = \dfrac{\text{observed value} - \text{mean}}{\text{standard deviation}} = \dfrac{42 - 57.84}{6.997} = -2.264$.
For the oldest CEO, $z = \dfrac{74 - 51.47}{6.997} = 2.881$.
Remember that a z-score measures the number of standard deviations a value is from the mean. These values are about what you would expect because the Empirical Rule states that about 95% of the values fall within 2 standard deviations of the mean, and 99.7% of the values fall within 3 standard deviations of the mean. So, z-scores for the lowest and highest values are often somewhere between 2 and 3 in absolute magnitude.

2.126 **a.** For a bell-shaped data set, the median equals the mean so the z-score = 0.
b. About 99.7% of the data set will be within 3 standard deviations of the mean, so the lowest value corresponds to a z-score of about –3. Put another way, the lowest value is likely to be 3 standard deviations below the mean. *Note*: If the sample size is very large, the lowest value might be farther than 3 standard deviations from the mean because about 100%–99.7%=0.3% (which is 3 in 1,000) of a list of values might not be within three standard deviations of the mean. Also, when the sample size is relatively small, the lowest z-score tends to be closer to –2.
c. About 99.7% of the data set will be within 3 standard deviations of the mean, so the highest value corresponds to a z-score of about +3. Put another way, the highest value is likely to be 3 standard deviations above the mean. This will generally be true unless the sample size is very large (or very small).
d. The mean corresponds to a z-score of 0.

2.127 What percentage of kindergarten children lives with their mother only? Their father only? One or both grandparents?

2.128 Do the students who studied the most last week tend to be the students with the highest grade point averages?

2.129 Is the average amount of coffee consumed per day the same for married people as it is for single people?

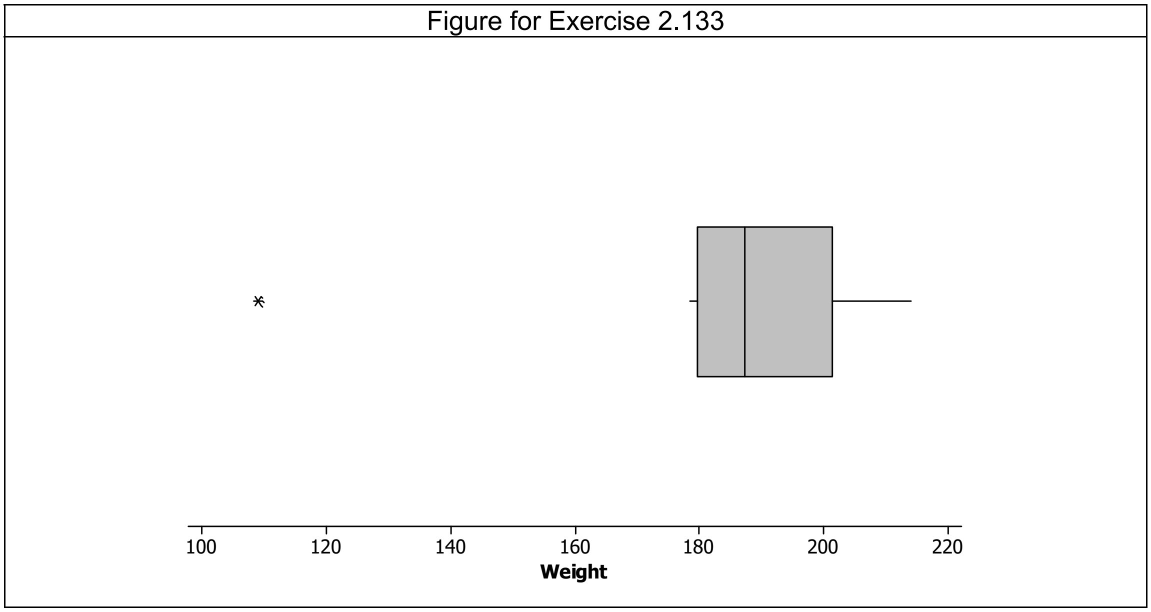
2.130 Are females more likely to dream in color than males?

2.131 **a.** Low = 0, Q_1= 22.5, median = 55, Q_3 = 175, High = 450. To find these values, first write the data in order. The median is the average of the middle two values. The lower quartile is the median of the smallest 12 values and the upper quartile is the median of the larger 12 values.
b. 450 would be marked as an outlier. The boundary for upper outliers is 175 + 1.5 (175 – 22.5) = 403.75.

2.132 **a.** Eye color is the explanatory variable and whether or not they wear corrective lens is the response variable.
b. Time between HIV infection and pregnancy is the explanatory variable. Whether or not they transmitted HIV to their infant is the response variable.

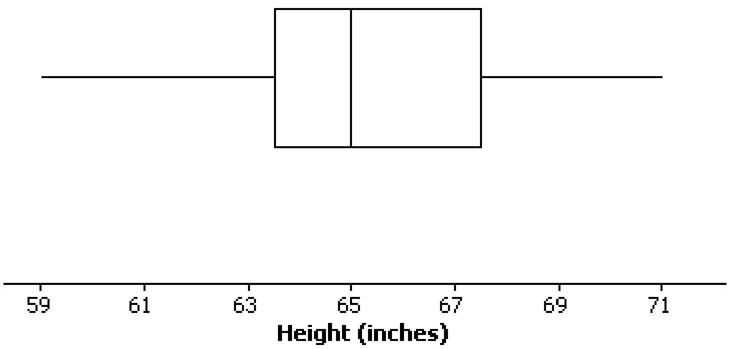
2.133 Low = 109, Q_1 =180.75, Median = 187.25, Q3 = 199, High = 214; 10; 109 is an outlier

Figure for Exercise 2.133



2.134

Figure for Exercise 2.134

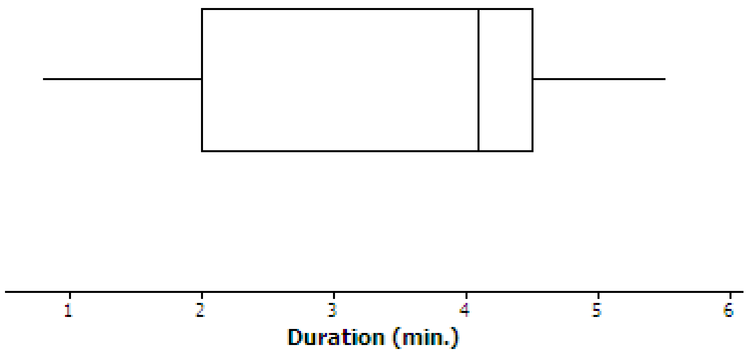


2.135 Outliers affect the standard deviation. This happens because the calculation uses the deviation from the mean for every value. An outlier has a large deviation from the mean, so it inflates the standard deviation. Extreme values generally do not affect the quartiles, and consequently they generally don't affect the interquartile range. Remember that a quartile is determined by counting through the ordered data to a particular location, so the exact size of the largest or smallest observations doesn't matter.

2.136 You expect women's heights to have a bell-shape curve because it is more common for a woman to have a height close to the mean than far from the mean. Generally, the further a height is from the mean (in either direction), the fewer the number of women with that height. The ages at marriage for women will probably not follow a bell-curve. Most of the ages will be in the 20's, but the data will not be symmetric. The ages can only be as low as law permits—15, maybe. The other direction extends much farther from the mean—some women do not get married until they are 40 or 50.

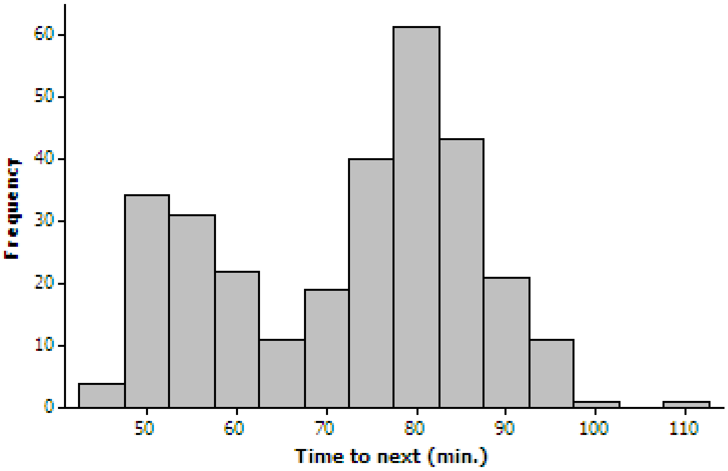
2.137 **a.** The boxplot does not show the bimodal nature of the distribution.

Figure for Exercise 2.137a



b. The distribution is bimodal.

Figure for Exercise 2.137b



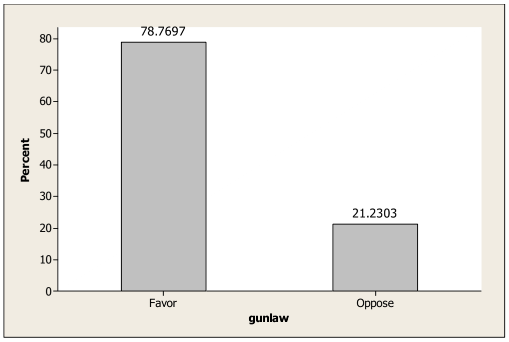
2.138 **a.** Of the 916 respondents who answered this question, 1050/1333=78.77% said they favored stronger gun control laws and 283/1333=21.23% said they did not. The following output was created using **Stat>Tables>Tally Individual Variables** in Minitab.

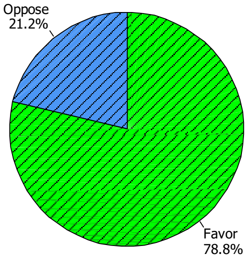
Output for Exercise 2.138a

```
gunlaw  Count  Percent
 Favor   1050    78.77
Oppose    283    21.23
    N=   1333
    *=    690
```

b. Either a bar chart or a pie chart could be done. With only two response categories, the pie chart might be a more effective display.

Figures for Exercise 2.138b





c. When groups have different sample sizes, it may be difficult to determine from the counts whether a relationship exists. The table is:

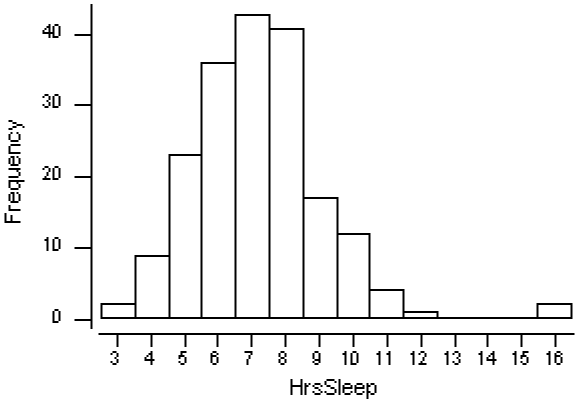
	Favor	Oppose	All
Female	622	106	728
Male	428	177	605
All	1050	253	1333

d. The percentage of females favoring stronger gun control is 622/728 = 85.44%. The percentage of males in favor is 428/605=70.74%.

e. The percentage favoring stronger gun control laws is higher for females than for males, so in this sample there is a relationship between sex and opinion about gun control.

2.139 **a.** Generally, the distribution is bell-shaped, although there are two outliers at 16 hours of sleep. Disregarding the outliers, the center of the distribution is somewhere near 7 hours.

Exercise for Exercise 2.139a



b. The five-number summary is:

	Hours of sleep		
Median		7	
Quartiles	6		8
Extremes	3		16

c. *Range* = *maximum–minimum* = 16–3 = 13 hours. *IQR* = $Q_3 - Q_1$ = 8–6 = 2 hours.

2.140 **a.** The stem-and-leaf plot produced by Minitab is shown below. To interpret the entries, think of each number of CDs owned as a three digit number, even if the value is less than 100. For example, think of 50 CDs as 050 CDs. A stem label is the first digit, a leaf is the second digit, and the third digit has been dropped (truncated). Notice the "+" symbols at the ends of the first two rows. This means there are more values not shown. Minitab ran out of room along these display lines.

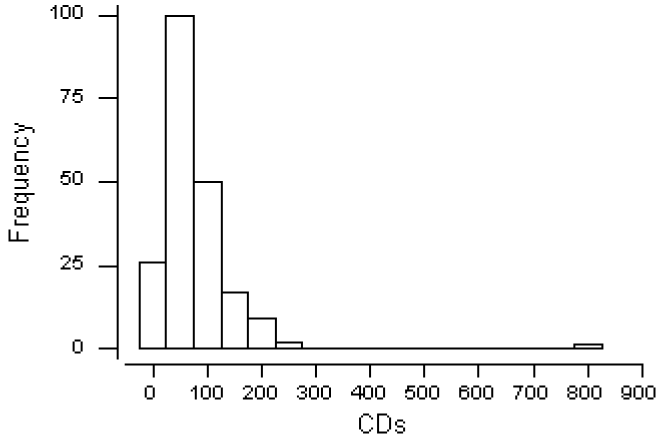
Figure for Exercise 2.140a

```
Stem-and-leaf of CDs       N  = 205
Leaf Unit = 10

0 000000111111111222222222222222222223333333333333333333333333334444+
0 55555555555555555555555555555555566666666666666777777777777777888+
1 000000000000000000000000112222333
1 555555555556677
2 00000001
2 55
3
3
4
4
5
5
6
6
7
7
8 0
```

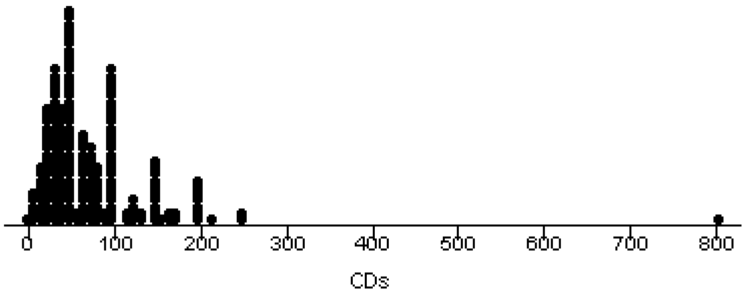
b.

Figure for Exercise 2.140b



c.

Figure for Exercise 2.140c



d. There is an outlier (800 CDs) and the data are skewed to the right. Many reported owning 100 or fewer CDs.
e. The mean is 72.84 CDs and the median is 50 CDs. The mean is higher.
f. The median is a better description of the locations. The mean has been inflated by the outlier and by the general skewness of the data.

2.141 **a.** Of 1,308 respondents who answered this question, 1263/1902=66.4% said they favor capital punishment and 639/1902 = 33.60% said they oppose it. The following output was created using **Stat>Tables>Tally individual variables** in Minitab version 14.

Output for Exercise 2.141a

```
cappun  Count  Percent
 Favor   1263    66.40
Oppose    639    33.60
    N=   1902
    *=    121
```

b. The sample size is different for this part than for part (a) because nine people who answered the question about capital punishment did not give a political party preference. Useful conditional percentages are the percentages within the various political preferences. The counts and percentages determined using **Stat>Tables>Cross Tabulation in** Minitab, are:

	Favor	**Oppose**	**All**
Democrat	372 (53.9%)	318 (46.1%)	690
Independent	457 (67.3%)	222 (32.7%)	679
Other	22 (61.1%)	14 (38.9%)	36
Republican	406 (83.2%)	82 (16.8%)	488
All	1257 (66.4%)	636 (33.6%)	1893

c. The variables are related. The percentage in favor of capital punishment is noticeably less for Democrats than it is for Republicans. For Independents, the percentage in favor is between what it is for Democrats and Republicans.

2.142 **a.** The relevant statistics, and for part (b) can be found in Minitab using **Stat>Basic Statistics>Display Descriptive Statistics**. Other software may give slightly different answers for the quartiles.

The mean is 193.13 and the standard deviation is 22.30.
The five-number summary is:

	Cholesterol (Control)		
Median		187	
Quartiles	178		204.5
Extremes	160		242

b. The mean is 253.93 and the standard deviation is 47.71.
The five-number summary is:

	Cholesterol (Heart attack)		
Median		268	
Quartiles	224.5		282
Extremes	142		360

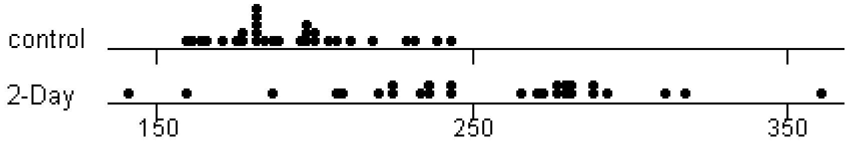
c. The cholesterol levels are generally higher for the heart attack patients. The difference in means is 253.93−193.13 = 60.8 and the difference in medians is 268−187 = 81.

d. The measurements in the heart attack group have greater spread than the measurements in the control group. A comparison of the three measures of spread is:

	Std. deviation	*Range*	*IQR*
Heart attack	47.71	360–142 = 218	282–224.5 = 57.5
Control	22.30	242–160 = 82	204.5–178 = 26.5

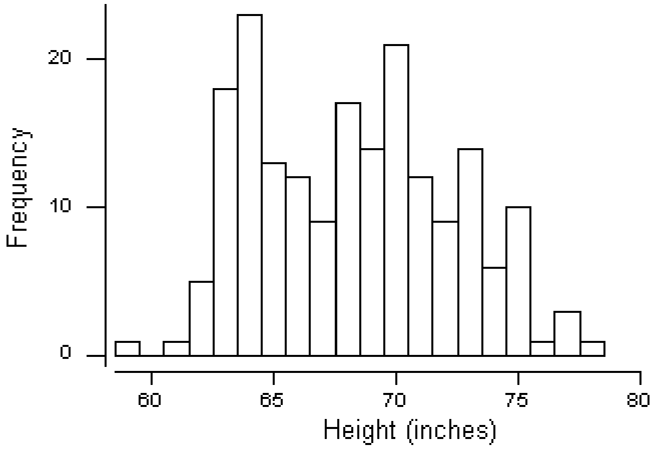
e. The dotplot illustrates the differences found in parts (c) and (d). The heart attack patients have generally higher cholesterol levels, and also exhibit more variability (spread) among their measurements.

Figure for Exercise 2.142e



2.143 **a.**

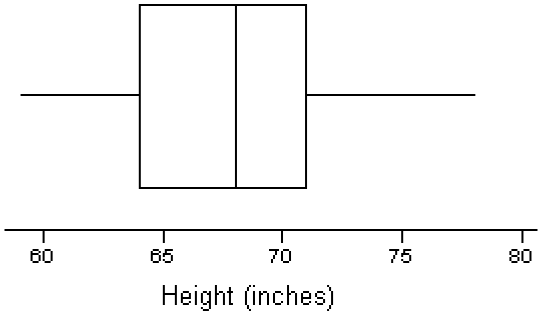
Figure for Exercise 2.143a



b. It's difficult to describe the shape of the histogram, but it's easy to say what it is not. It's not bell-shaped, and it's not skewed. Theoretically, it is bimodal, which means that there are two different peaks. In this case one peak occurs near the mean height of women and the other occurs near the mean height for men.

c.

Figure for Exercise 2.143c



d. The histogram is more informative because it gives more detail about the pattern between 64 inches and 71 inches. We're able to see the two distinct peaks in the histogram.

2.144 **a.** The degree variable is ordinal. The categories are ordered with respect to amount of education.
b. The results are presented in alphabetical order of degree categories:

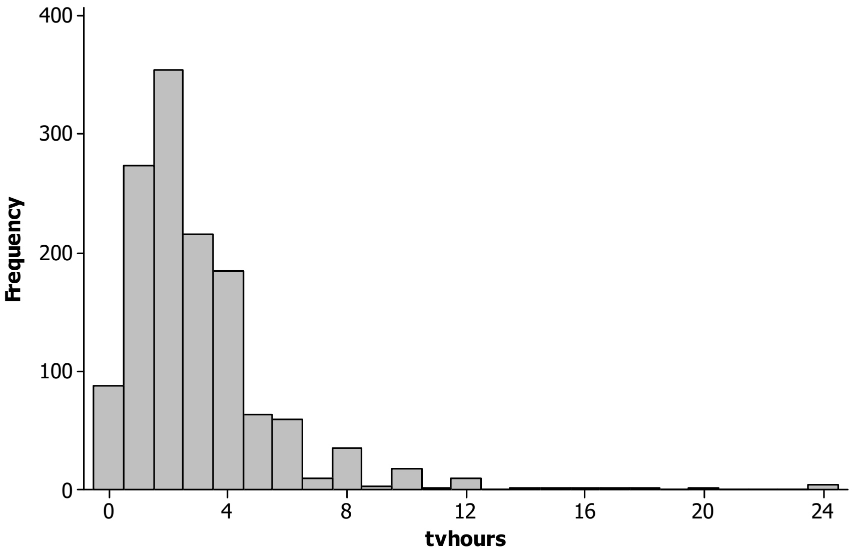
degree	Count	Percentage
Bachelor	355	17.56
Graduate	194	9.59
High School	1003	49.60
Junior College	173	8.56
Not high school	297	14.69
All	2022	100.00
* (missing)	1	

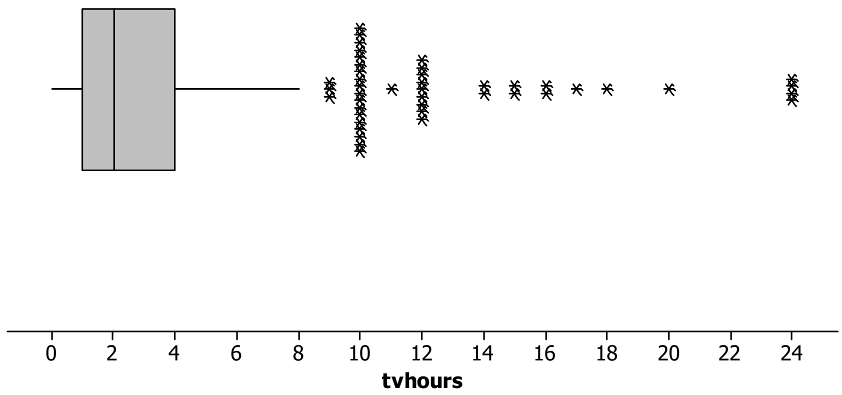
c. Sum of bachelor, graduate and junior college = 17.56% + 9.59% + 8.56% = 35.71%
d. In Minitab, this can be done using **Stat>Basic Stats>Display Descriptive Statistics** and requesting statistics "By" ***degree***. The means, rounded to two decimal places and presented in order of amount of education, are:

degree	Mean (hrs)
Not high school	4.61
High school	3.07
Junior college	2.55
Bachelor	2.12
Graduate	1.99

e. Yes, there is a relationship. Mean hours of self-reported television watching decreases as years of education increases.
f. The visual display might be a histogram, boxplot, dotplot or a stemplot. A histogram and a boxplot are shown below. The data are skewed to the right. Interestingly, a few respondents claimed watching television more than 20 hours per day.

Figures for Exercise 2.144f




0
2
4
6
8
10
12
14
16
18
20
22
24
tvhours