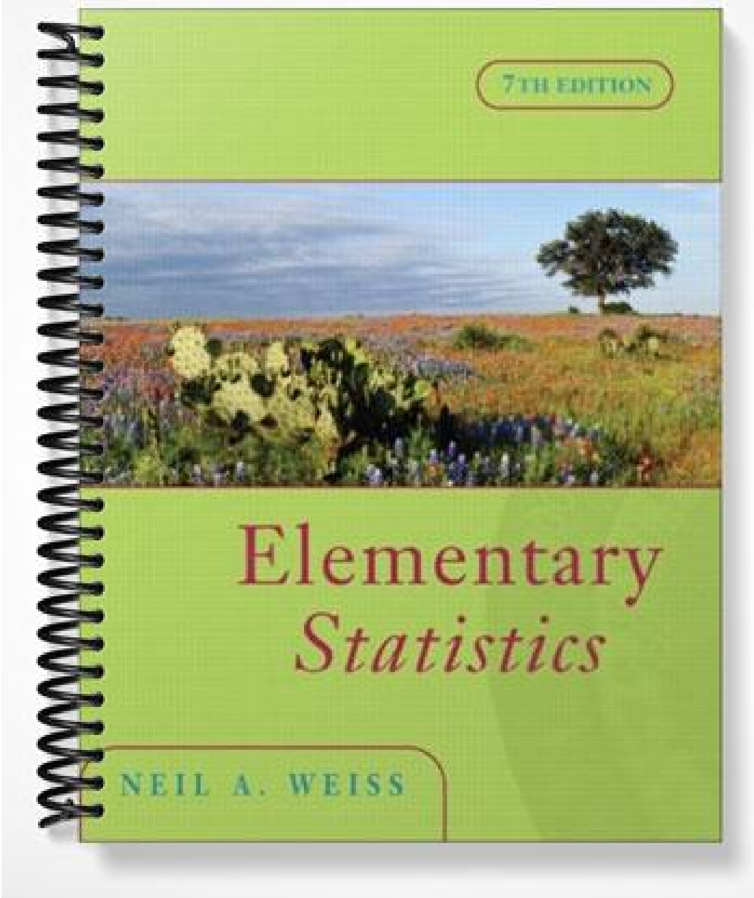
# SOLUTIONS MANUAL



## CHAPTER 2 ANSWERS

**Exercises 2.1**

**2.1** (a) Hair color, model of car, and brand of popcorn are qualitative variables.

(b) Number of eggs in a nest, number of cases of flu, and number of employees are discrete, quantitative variables.

(c) Temperature, weight, and time are quantitative continuous variables.

**2.2** (a) A qualitative variable is a nonnumerically valued variable. Its possible "values" are descriptive (e.g., color, name, gender).

(b) A discrete, quantitative variable is one whose possible values can be listed. It is usually obtained by counting rather than by measuring.

(c) A continuous, quantitative variable is one whose possible values form some interval of numbers. It usually results from measuring.

**2.3** (a) Qualitative data result from observing and recording values of a qualitative variable, such as, color or shape.

(b) Discrete, quantitative data are values of a discrete quantitative variable. Values usually result from counting something.

(c) Continuous, quantitative data are values of a continuous variable. Values are usually the result of measuring something such as temperature that can take on any value in a given interval.

**2.4** The classification of data is important because it will help you choose the correct statistical method for analyzing the data.

**2.5** Of qualitative and quantitative (discrete and continuous) types of data, only qualitative yields nonnumerical data.

**2.6** (a) The first column lists states. Thus, it consists of *qualitative* data.

(b) The second column gives the number of serious doctor disciplinary actions in each state in 2002. These data are integers and therefore are *quantitative, discrete* data.

(c) The third column also consists of *quantitative, discrete* data. The number of doctors in each state is given using whole numbers.

**2.7** (a) The second column consists of *quantitative, discrete* data. This column provides the ranks of the cities with the highest temperatures.

(b) The third column consists of *quantitative, continuous* data since temperatures can take on any value from the interval of numbers found on the temperature scale. This column provides the highest temperature in each of the listed cities.

(c) The information that Phoenix is in Arizona is qualitative data since it is nonnumeric.

**2.8** (a) The first column consists of *quantitative, discrete* data. This column provides the ranks of the deceased celebrities with the top 13 earnings during the period from June 2001 to June 2002.

(b) The third column consists of *quantitative, discrete* data, the earnings of the celebrities. Since money involves discrete units, such as dollars and cents, the data is discrete, although, for all practical purposes, this data might be considered quantitative continuous data.

**2.9** (a) The first column consists of *quantitative, discrete* data. This column provides the ranks of the cities with the highest percentage of internet access connected via broadband in August, 2004. These are whole numbers.

(b) The cities listed in the second column are qualitative data since they are nonnumerical.

(c) The third column contains percentages, which are ratios of whole numbers. Ratios of whole numbers cannot be irrational numbers and therefore there are gaps in the number line that represent values that the ratios cannot assume. These data are therefore *quantitative, discrete.*

**2.10** (a) The first column contains types of products. They are *qualitative* data since they are nonnumerical.

(b) The second column contains money values. Technically, these are *quantitative, discrete* data since there are gaps between possible values at the cent level. For all practical purposes, however, these are *quantitative, continuous* data.

**2.11** The first two columns contain quantitative, discrete data in the form of ranks. These are whole numbers. The third and fourth columns contain qualitative data in the form of names. The last column contains the number of viewers of the programs. Total number of viewers is a whole number and therefore *quantitative, discrete* data.

**2.12** Duration is a measure of time and is therefore *quantitative, continuous*. One might argue that workshops are frequently done in whole numbers of weeks, which would be *quantitative, discrete*. The number of students, the number of each gender, and the number of each ethnicity are whole numbers and are therefore *quantitative, discrete*. The genders and ethnicities themselves are nonnumerical and are therefore *qualitative* data. The number of web reports is a whole number and is *quantitative, discrete* data.

**2.13** The first and fourth columns are nonnumerical and are therefore *qualitative* data. The second, third, and fifth columns are measures of size, time, and weight, all of which are *quantitative, continuous* data.

**2.14** Of the eight items presented, only high school class rank involves ordinal data. The rank is ordinal data.

**Exercises 2.2**

**2.15** One important reason for grouping data is that grouping often makes a large and complicated set of data more compact and easier to understand.

**2.16** For cutpoints and midpoints to make sense, data must be numerical. They do not make sense for qualitative data classes because such data are nonnumerical.

**2.17** The three most important guidelines in choosing the classes for grouping a data set are: (1) the number of classes should be small enough to provide an effective summary, but large enough to display the relevant characteristics of the data; (2) each observation must belong to one, and only one, class; and (3) whenever feasible, all classes should have the same width.

**2.18** (a) The frequency of a class is the number of observations in the class, whereas, the relative frequency of a class is the ratio of the class frequency to the total number of observations.

(b) The percentage of a class is 100 times the relative frequency of the class. Equivalently, the relative frequency of a class is the percentage of the class expressed as a decimal.

**2.19** (a) True. Having identical frequency distributions implies that the total number of observations and the numbers of observations in each class are identical. Thus, the relative frequencies will also be identical.

(b) False. Having identical relative frequency distributions means that the ratio of the count in each class to the total is the same for both frequency distributions. However, one distribution may have twice (or some other multiple) the total number of observations as the other. For example, two distributions with counts of 5, 4, 1 and 10, 8, 2 would be different, but would have the same relative frequency distribution.

(c) If the two data sets have the same number of data values, either a frequency distribution or a relative-frequency distribution is suitable. If, however, the two data sets have different numbers of observations, using relative-frequency distributions is more appropriate because the total of each set of relative frequencies is 1, putting both distributions on the same basis for comparison.

**2.20** The four elements of a grouped-data table are the classes, frequencies, relative frequencies, and midpoints. The classes consist of categories for grouping data values within intervals bounded by lower and upper cutpoints. A frequency for a class is the number of observations that fall within the class. A relative frequency for a class is the ratio of the frequency of a class to the total number of observations. The midpoint of a class is found by averaging the lower and upper cutpoints of the class.

**2.21** In the first method for depicting classes, we used the notation **a –< b** to mean values that are greater than or equal to **a** and up to, but not including **b**, such as 30 –< 40 to mean a range of values greater than or equal to 30, but strictly less than 40. In the alternate method, we used the notation **a-b** to indicate a class that extends from **a** to **b**, including both. For example, 30-39 is a class that includes both 30 and 39. The alternate method is especially appropriate when all of the data values are integers. If the data include values like 39.7 or 39.93, the first method is more advantageous since the cutpoints remain integers; whereas, in the alternate method, the upper limits for each class would have to be expressed in decimal form such as 39.9 or 39.99.

**2.22** (a) For continuous data displayed to one or more decimal places, using the classes a –< b is best since the description of the classes is simpler, regardless of the number of decimal places displayed.

(b) For discrete data with relatively few distinct observations, the single value grouping is best since either of the other two methods would result in combining some of those distinct values into single classes, resulting in too few classes, possibly less than 5.

**2.23** When grouping data using classes that each represents a single possible numerical value, the midpoint of each class would be the same as the value in that class. Thus, listing the midpoints would be redundant.

**2.24** The first class to construct is 40 –< 50. Since all classes are to be of equal width, and the second class begins with 50, we know that the width of all classes is 50 - 40 = 10. All of the classes are presented in column 1. The last class to construct is 150 –< 160, since the largest single data value is 155. Having established the classes, we tally the energy consumption figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 50, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. By averaging the lower and upper class cutpoints for each class, we arrive at the class midpoint for each class. The class midpoints for all classes are presented in column 4.

| Consumption (mil. BTU) | Frequency | Relative frequency | Midpoint |
|---|---|---|---|
| 40 ← 50 | 1 | 0.02 | 45 |
| 50 ← 60 | 7 | 0.14 | 55 |
| 60 ← 70 | 7 | 0.14 | 65 |
| 70 ← 80 | 3 | 0.06 | 75 |
| 80 ← 90 | 6 | 0.12 | 85 |
| 90 ← 100 | 10 | 0.20 | 95 |
| 100 ← 110 | 5 | 0.10 | 105 |
| 110 ← 120 | 4 | 0.08 | 115 |
| 120 ← 130 | 2 | 0.04 | 125 |
| 130 ← 140 | 3 | 0.06 | 135 |
| 140 ← 150 | 0 | 0.00 | 145 |
| 150 ← 160 | 2 | 0.04 | 155 |
| | 50 | 1.00 | |

**2.25** The first class to construct is 52 ← 54. Since all classes are to be of equal width 2, the second class begins with 54. All of the classes are presented in column 1. The last class to construct is 74 ← 76, since the largest single data value is 75.3. Having established the classes, we tally the speed figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 35, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. By averaging the lower and upper class cutpoints for each class, we arrive at the class midpoint for each class. The class midpoints for all classes are presented in column 4.

| Speed (MPH) | Frequency | Relative Frequency | Midpoint |
|---|---|---|---|
| 52 ← 54 | 2 | 0.057 | 53 |
| 54 ← 56 | 5 | 0.143 | 55 |
| 56 ← 58 | 6 | 0.171 | 57 |
| 58 ← 60 | 8 | 0.229 | 59 |
| 60 ← 62 | 7 | 0.200 | 61 |
| 62 ← 64 | 3 | 0.086 | 63 |
| 64 ← 66 | 2 | 0.057 | 65 |
| 66 ← 68 | 1 | 0.029 | 67 |
| 68 ← 70 | 0 | 0.000 | 69 |
| 70 ← 72 | 0 | 0.000 | 71 |
| 72 ← 74 | 0 | 0.000 | 73 |
| 74 ← 76 | 1 | 0.029 | 75 |
| | 35 | 1.001 | |

Note that the relative frequencies sum to 1.001, not 1.00, due to round-off errors in the individual relative frequencies.

**2.26** The first class to construct is 40-49. Since all classes are to be of equal width, and the second class begins with 50, we know that the width of all classes is 50-40=10. All of the classes are presented in column 1. The last class to construct is 150-159 since the largest single data value is 155. Having established the classes, we tally the energy consumption figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 50, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. By averaging the lower limit for each class with the upper limit of the sane class, we arrive at the class mark for each class. The class marks for all classes are presented in column 4.

| Consumption (mil. BTU) | Frequency | Relative Frequency | Class Mark |
|---|---|---|---|
| 40-49 | 1 | 0.02 | 44.5 |
| 50-59 | 7 | 0.14 | 54.5 |
| 60-69 | 7 | 0.14 | 64.5 |
| 70-79 | 3 | 0.06 | 74.5 |
| 80-89 | 6 | 0.12 | 84.5 |
| 90-99 | 10 | 0.20 | 94.5 |
| 100-109 | 5 | 0.10 | 104.5 |
| 110-119 | 4 | 0.08 | 114.5 |
| 120-129 | 2 | 0.04 | 124.5 |
| 130-139 | 3 | 0.06 | 134.5 |
| 140-149 | 0 | 0.00 | 144.5 |
| 150-159 | 2 | 0.04 | 154.5 |
| | 50 | 1.00 | |

**2.27** The first class to construct is 52-53.9. Since all classes are to be of equal width, the second class has limits of 54 and 55.9. All of the classes are presented in column 1. The last class to construct is 74-75.9 since the largest single data value is 75.3. Having established the classes, we tally the speed figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, 35, results in each class's relative frequency, which is presented in column 3. By averaging the lower limit for each class with the upper limit of the same class, we arrive at the class mark for each class. The class marks for all classes are presented in column 4.

| Speed (MPH) | Frequency | Relative Frequency | Class Mark |
|---|---|---|---|
| 52-53.9 | 2 | 0.057 | 52.95 |
| 54-55.9 | 5 | 0.143 | 54.95 |
| 56-57.9 | 6 | 0.171 | 56.95 |
| 58-59.9 | 8 | 0.229 | 58.95 |
| 60-61.9 | 7 | 0.200 | 60.95 |
| 62-63.9 | 3 | 0.086 | 62.95 |
| 64-65.9 | 2 | 0.057 | 64.95 |
| 66-67.9 | 1 | 0.029 | 66.95 |
| 68-69.9 | 0 | 0.000 | 68.95 |
| 70-71.9 | 0 | 0.000 | 70.95 |
| 72-73.9 | 0 | 0.000 | 72.95 |
| 74-75.9 | 1 | 0.029 | 74.95 |
| | 35 | 1.001 | |

Note that the relative frequencies sum to 1.001, not 1.00, due to round-off errors in the individual relative frequencies.

**2.28** The first class to construct is 40 ← 42. Since all classes are to be of equal width 2, the second class begins with 42. All of the classes are presented in column 1. The last class to construct is 48 ← 50, since the largest single data value is 49.1. Having established the classes, we tally the ratings figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 32, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. By averaging the lower and upper class cutpoints for each class, we arrive at the class midpoint for each class. The class midpoints for all classes are presented in column 4.

| Rating | Frequency | Relative Frequency | Midpoint |
|---|---|---|---|
| 40 < 42 | 10 | 0.3125 | 41 |
| 42 < 44 | 6 | 0.1875 | 43 |
| 44 < 46 | 7 | 0.2188 | 45 |
| 46 < 48 | 6 | 0.1875 | 47 |
| 48 < 50 | 3 | 0.0938 | 49 |
| | 32 | 1.0001 | |

Note that the relative frequencies sum to 1.0001, not 1.0000, due to round-off errors in the individual relative frequencies.

**2.29** The first class to construct is 11 < 14. This has width 3 with its midpoint at 12.5. Since all classes are to be of equal width 3, the second class begins with 14. All of the classes are presented in column 1. The last class to construct is 26 < 29, since the largest single data value is 27.0. Having established the classes, we tally the audience sizes into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 35, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. By averaging the lower and upper class cutpoints for each class, we arrive at the class midpoint for each class. The class midpoints for all classes are presented in column 4.

| Audience (Millions) | Frequency | Relative Frequency | Midpoint |
|---|---|---|---|
| 11 < 14 | 5 | 0.25 | 12.5 |
| 14 < 17 | 4 | 0.20 | 15.5 |
| 17 < 20 | 6 | 0.30 | 18.5 |
| 20 < 23 | 2 | 0.10 | 21.5 |
| 23 < 26 | 2 | 0.10 | 24.5 |
| 26 < 29 | 1 | 0.05 | 27.5 |
| | 20 | 1.00 | |

**2.30** Classes are to be based on a single value. Since each data value is one of the integers 1 through 7, inclusive, the classes will be 1 through 7, inclusive. These are presented in column 1. Having established the classes, we tally the household sizes into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, 40, results in each class's relative frequency which is presented in column 3. Since each class is based on a single value, the midpoint is the class itself. Thus, a column of midpoints is not given, since this same information is presented in column 1.

| Number of Persons | Frequency | Relative Frequency |
|---|---|---|
| 1 | 7 | 0.175 |
| 2 | 13 | 0.325 |
| 3 | 9 | 0.225 |
| 4 | 5 | 0.125 |
| 5 | 4 | 0.100 |
| 6 | 1 | 0.025 |
| 7 | 1 | 0.025 |
| | 40 | 1.000 |

**2.31** Since the data values range from 0 to 4, we construct a table with classes based on a single value. The resulting table follows.

| Number of Siblings | Frequency | Relative Frequency |
|---|---|---|
| 0 | 8 | 0.200 |
| 1 | 17 | 0.425 |
| 2 | 11 | 0.275 |
| 3 | 3 | 0.075 |
| 4 | 1 | 0.025 |
| | 40 | 1.000 |

**2.32** The classes are the days of the week and are presented in column 1. The frequency distribution of the networks is presented in column 2. Dividing each frequency by the total number of shows, which is 20, results in each class's relative frequency. The relative frequency distribution is presented in column 3.

| Network | Frequency | Relative Frequency |
|---|---|---|
| ABC | 3 | 0.15 |
| CBS | 9 | 0.45 |
| Fox | 3 | 0.15 |
| NBC | 5 | 0.25 |
| | 20 | 1.00 |

**2.33** The classes are the NCAA wrestling champions and are presented in column 1. The frequency distribution of the champions is presented in column 2. Dividing each frequency by the total number of champions, which is 25, results in each class's relative frequency. The relative frequency distribution is presented in column 3.

| Champion | Frequency | Relative Frequency |
|---|---|---|
| Iowa | 15 | 0.60 |
| Iowa St. | 1 | 0.04 |
| Minnesota | 2 | 0.08 |
| Arizona St. | 1 | 0.04 |
| Oklahoma St. | 6 | 0.24 |
| | 25 | 1.00 |

**2.34** (a) Using Minitab, retrieve the data from the Weiss-Stats-CD. Column 1 contains the names of the states and Colunm 2 contains the names of the region each state is located in. From the tool bar, select **Stat ► Tables ► Tally Individual Variables**, double-click on REGION in the first box so that REGION appears in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The result is

| REGION | Count | Percent |
|---|---|---|
| Midwest | 12 | 24.00 |
| Northeast | 9 | 18.00 |
| South | 16 | 32.00 |
| West | 13 | 26.00 |
| N= | 50 | |

(b) The Midwest region contains 12 states or 24% of the 50 states, the Northeast Region contains 9 states or 18% of the 50 states, the South region contains 16 states or 32%, and the West region contains 13 states or 26%.

**2.35** (a) Using Minitab, retrieve the data from the Weiss-Stats-CD. Column 1 contains the numbers of pups borne in a lifetime for each of 80 female Great White Sharks. From the tool bar, select **Stat ▶ Tables ▶ Tally Individual Variables**, double-click on PUPS in the first box so that PUPS appears in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The result is

| PUPS | Count | Percent |
|---|---|---|
| 3 | 2 | 2.50 |
| 4 | 5 | 6.25 |
| 5 | 10 | 12.50 |
| 6 | 11 | 13.75 |
| 7 | 17 | 21.25 |
| 8 | 17 | 21.25 |
| 9 | 11 | 13.75 |
| 10 | 4 | 5.00 |
| 11 | 2 | 2.50 |
| 12 | 1 | 1.25 |

(b) As the number of pups increases from 3 to 12, the counts increase from 2 to a maximum of 17 at 7 and 8 pups, and then decrease back down to 1 again at 12 pups.

**2.36** (a) Using Minitab, retrieve the data from the Weiss-Stats-CD. Column 1 contains the days of the week for each road rage incident. From the tool bar, select **Stat ▶ Tables ▶ Tally Individual Variables**, double-click on DAY in the first box so that DAY appears in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The result is

| DAY | Count | Percent |
|---|---|---|
| F | 18 | 26.09 |
| M | 5 | 7.25 |
| Sa | 7 | 10.14 |
| Su | 5 | 7.25 |
| Th | 11 | 15.94 |
| Tu | 11 | 15.94 |
| W | 12 | 17.39 |
| N= | 69 | |

(b) The greatest number of road rage incidents occurs on Friday, followed by Wednesday, Tuesday, and Thursday. Saturday, Sunday, and Monday have the fewest incidents.

**2.37** (a) In Minitab, retrieve the data from the *WeissStats* CD. Then, assuming that the CD drive is drive D, type in Minitab's session window after the MTB> prompt [You may have to first choose **Enable Commands** from **Editor** on the Tool bar.] the command %D:\Minitab_Macros\Group.mac 'NAEP' and press the ENTER key. We are given three options for specifying the classes. Since we want the first class to have lower cutpoint 234.5 and a class width of 2, we select the third option (3) by entering 3 after the DATA> prompt, press the ENTER key, and then type 234.5 2 when prompted to enter the cutpoint and class width of the first class. Press the ENTER key again. The resulting output is

Grouped-data table for NAEP N = 50

| Row | LowerCut | UpperCut | Freq | RelFreq | Midpoint |
|---|---|---|---|---|---|
| 1 | 234.5 | 236.5 | 3 | 0.06 | 235.5 |
| 2 | 236.5 | 238.5 | 3 | 0.06 | 237.5 |
| 3 | 238.5 | 240.5 | 1 | 0.02 | 239.5 |
| 4 | 240.5 | 242.5 | 3 | 0.06 | 241.5 |
| 5 | 242.5 | 244.5 | 2 | 0.04 | 243.5 |
| 6 | 244.5 | 246.5 | 5 | 0.10 | 245.5 |
| 7 | 246.5 | 248.5 | 4 | 0.08 | 247.5 |
| 8 | 248.5 | 250.5 | 6 | 0.12 | 249.5 |
| 9 | 250.5 | 252.5 | 9 | 0.18 | 251.5 |
| 10 | 252.5 | 254.5 | 9 | 0.18 | 253.5 |
| 11 | 254.5 | 256.5 | 3 | 0.06 | 255.5 |
| 12 | 256.5 | 258.5 | 2 | 0.04 | 257.5 |

Alternatively, enter the NAEP data in Excel with the variable name NAEP in the first row. Use the cursor to highlight the entire data set including the name. Then from the tool bar, select **DDXL ► Tables**; select **Frequency Table** from the **Function type** drop-down list box; specify NAEP in the **Categorical Variable** text box and click **OK**. The result is the following table:

Total Cases 50
Number of Categories 21

| Group | Count | % |
|---|---|---|
| 235 | 1 | 2 |
| 236 | 2 | 4 |
| 238 | 3 | 6 |
| 239 | 1 | 2 |
| 241 | 2 | 4 |
| 242 | 1 | 2 |
| 243 | 1 | 2 |
| 244 | 1 | 2 |
| 245 | 4 | 8 |
| 246 | 1 | 2 |
| 247 | 2 | 4 |
| 248 | 2 | 4 |
| 249 | 3 | 6 |
| 250 | 3 | 6 |
| 251 | 5 | 10 |
| 252 | 4 | 8 |
| 253 | 6 | 12 |
| 254 | 3 | 6 |
| 255 | 1 | 2 |
| 256 | 2 | 4 |
| 257 | 2 | 4 |

(b) Using Minitab, Column 3 contains the RESULT. From the tool bar, select **Stat ▶ Tables ▶ Tally Individual Variables**, double-click on RESULT in the first box so that RESULT appears in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The result is

| RESULT | Count | Percent |
|---|---|---|
| Bush | 30 | 60.00 |
| Gore | 20 | 40.00 |
| N= | 50 | |

**2.38** The first class to construct is 15 < 25. This has width 10 with its midpoint at 20. Since all classes are to be of equal width 10, the second class begins with 25. All of the classes are presented in column 1. The last class to construct does not go beyond 75 < 85, since the largest single data value is 80.33. Having established the classes, we tally the closing prices into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 30, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. By averaging the cutpoints for each class, we arrive at the class midpoint for each class. The class midpoints for all classes are presented in column 4.

| Closing Price (Dollars) | Frequency | Relative Frequency | Midpoint |
|---|---|---|---|
| 15 < 25 | 3 | 0.10 | 20 |
| 25 < 35 | 10 | 0.33 | 30 |
| 35 < 45 | 5 | 0.17 | 40 |
| 45 < 55 | 2 | 0.07 | 50 |
| 55 < 65 | 6 | 0.20 | 60 |
| 65 < 75 | 3 | 0.10 | 70 |
| 75 < 85 | 1 | 0.03 | 80 |
| | 30 | 1.00 | |

**2.39** (a) Since the volumes are given in hundreds, we must first convert the numbers to millions by dividing by 10,000. The new data (in increasing order) are 3.960, 4.873, 5.199, 5.533, 6.105, 6.485, 7.625, 7.640, 7.925, 11.752, 11.761, 11.855, 12.995, 13.177, 13.635, 14.306, 15.095, 16.265, 17.337, 19.193, 19.557, 19.875, 24.198, 25.125, 30.379, 30.937, 40.119, 72.573, 96.678, 191.491.

The first class to construct is 0 < 5. Since all classes are to be of equal width 5, the second class is 5 < 10. All of the classes are presented in column 1. Having established the classes, we tally the volume figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 30, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. By averaging the cutpoints for each class, we arrive at the class midpoint for each class. The class midpoints for all classes are presented in column 4.

| Volume | Frequency | Relative Frequency | Midpoint |
|---|---|---|---|
| 0 < 5 | 2 | 0.07 | 2.5 |
| 5 < 10 | 7 | 0.23 | 7.5 |
| 10 < 15 | 7 | 0.23 | 12.5 |
| 15 < 20 | 6 | 0.20 | 17.5 |
| 20 < 25 | 1 | 0.03 | 22.5 |
| 25 < 30 | 1 | 0.03 | 27.5 |
| 30 < 35 | 2 | 0.07 | 32.5 |
| 35 < 40 | 0 | 0.00 | 37.5 |
| 40 < 45 | 1 | 0.03 | 42.5 |
| 45 & Over | 3 | 0.10 | |
| | 30 | 0.99 | |

(b) Since the last class has no upper cutpoint, the midpoint cannot be computed.

**2.40** Since the data contains values ranging from −0.45 to +1.90, it is appropriate to use classes of uniform width 0.25, starting at −0.625. The first class to construct is −0.625 < −0.375. Since all classes are to be of equal width 0.25, the second class will be -0.375 < -0.125. All of the classes are presented in column 1. Having established the classes, we tally the change figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 30, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. By averaging the cutpoints for each class, we arrive at the class midpoint for each class. The class midpoints for all classes are presented in column 4.The changes range from −0.45 to +1.90. We will choose classes of width 0.25 beginning at a midpoint of -.50. This will result in 11 classes, which is reasonable. Your choice may be different.

| Change (Dollars) | Frequency | Relative Frequency | Midpoint |
|---|---|---|---|
| -0.625 ← -0.375 | 2 | 0.07 | -0.50 |
| -0.375 ← -0.125 | 3 | 0.10 | -0.25 |
| -0.125 ← 0.125 | 5 | 0.17 | 0.00 |
| 0.125 ← 0.375 | 7 | 0.23 | 0.25 |
| 0.375 ← 0.625 | 5 | 0.17 | 0.50 |
| 0.625 ← 0.875 | 5 | 0.17 | 0.75 |
| 0.875 ← 1.125 | 1 | 0.03 | 1.00 |
| 1.125 ← 1.375 | 1 | 0.03 | 1.25 |
| 1.375 ← 1.625 | 0 | 0.00 | 1.50 |
| 1.625 ← 1.875 | 0 | 0.00 | 1.75 |
| 1.875 ← 2.125 | 1 | 0.03 | 2.00 |
| | 30 | 1.00 | |

**2.41** (a) The classes are presented in column 1. With the classes established, we then tally the exam scores into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of exam scores, which is 20, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. The class mark of each class is the average of the lower and upper limits. The class marks for all classes are presented in column 4.

| Score | Frequency | Relative Frequency | Class Mark |
|---|---|---|---|
| 30-39 | 2 | 0.10 | 34.5 |
| 40-49 | 0 | 0.00 | 44.5 |
| 50-59 | 0 | 0.00 | 54.5 |
| 60-69 | 3 | 0.15 | 64.5 |
| 70-79 | 3 | 0.15 | 74.5 |
| 80-89 | 8 | 0.40 | 84.5 |
| 90-100 | 4 | 0.20 | 95.0 |
| | 20 | 1.00 | |

(b) The first six classes have width 10; the seventh class had width 11.

(c) Answers will vary, but one choice is to keep the first six classes the same and make the next two classes 90-99 and 100-109. Another possibility is 31-40, 41-50, …, 91-100.

**2.42** (a) Tally marks for all 50 students, where each student is categorized by age and sex, are presented in the contingency table given in part (b).

(b) Tally marks in each box appearing in the following chart are counted. These counts, or frequencies, replace the tally marks in the contingency table. For each row and each column, the frequencies are added, and their sums are recorded in the proper "Total" box.

Age (yrs)

| Sex | Under 21 | 21 - 25 | Over 25 | Total |
|---|---|---|---|---|
| Male | \|\|\|\|\| \|\|\| | \|\|\|\|\| \|\|\|\|\| \|\| | \|\| | |
| Female | \|\|\|\|\| \|\|\|\|\| \|\| | \|\|\|\|\| \|\|\|\|\| \|\|\| | \|\|\| | |
| Total | | | | |

Age (yrs)

| Sex | Under 21 | 21-25 | Over 25 | Total |
|---|---|---|---|---|
| Male | 8 | 12 | 2 | 22 |
| Female | 12 | 13 | 3 | 28 |
| Total | 20 | 25 | 5 | 50 |

(c) The row and column totals represent the total number of students in each of the corresponding categories. For example, the row total of 22 indicates that 22 of the students in the class are male.

(d) The sum of the row totals is 50, and the sum of the column totals is 50. The sums are equal because they both represent the total number of students in the class.

(e) Dividing each frequency reported in part (b) by the grand total of 50 students results in a contingency table that gives relative frequencies.

Age (yrs)

| Sex | Under 21 | 21-25 | Over 25 | Total |
|---|---|---|---|---|
| Male | 0.16 | 0.24 | 0.04 | 0.44 |
| Female | 0.24 | 0.26 | 0.06 | 0.56 |
| Total | 0.40 | 0.50 | 0.10 | 1.00 |

(f) The 0.16 in the upper left-hand cell indicates that 16% of the students in the class are males *and* under 21. The 0.40 in the lower left-hand cell indicates that 40% of the students in the class are under age 21. A similar interpretation holds for the remaining entries.

**Exercises 2.3**

**2.43** A frequency histogram shows the actual frequencies on the vertical axis; whereas, the relative frequency histogram always shows proportions (between 0 and 1) or percentages (between 0 and 100) on the vertical axis.

**2.44** An advantage of the histogram over a grouped-data table is that it is possible to get an overall view of the data more easily. A disadvantage of the histogram is that it may not be possible to determine exact frequencies for the classes when the number of observations is large.

**2.45** By showing the cutpoints on the horizontal axis, the range of possible data values in each class is immediately known and the midpoint can be quickly determined. This is particularly helpful if it is not convenient to make all classes the same width. The use of the midpoints is appropriate when each class consists of a single value (which is, of course, also the midpoint). Use of the midpoints is not appropriate in other situations since it may be difficult to determine the location of the cutpoints from the values of the midpoints, particularly if the midpoints are not evenly spaced. Midpoints cannot be used if there is an open class.

**2.46** If the classes consist of single values, stem-and-leaf diagrams and frequency histograms are equally useful. If only one diagram is needed and the classes consist of more than one value, the stem-and-leaf diagram allows one to retrieve all of the original data values whereas the frequency histogram does not. If two or more sets of data of different sizes are to be compared, the relative frequency histogram is advantageous because all of the diagrams to be compared will have the same total relative frequency of 1.00. Finally, stem-and-leaf diagrams are not very useful with very large data sets and may present problems with data having many digits in each number.

**2.47** The histogram (especially one using relative frequencies) is generally preferable. Data sets with a large number of observations may result in a stem of the stem-and-leaf diagram having more leaves than will fit on the line. In that case, the histogram would be preferable.

**2.48** You can reconstruct the stem-and-leaf diagram using two lines per stem. For example, instead of listing all of the values from 10 to 19 on a '1' stem, you can make two '1' stems. On the first, you record the values from 10 to 14 and on the second, the values from 15 to 19. If there are still two few stems, you can reconstruct the diagram using five lines per stem, recording 10 and 11 on the first line, 12 and 13 on the second, and so on.

**2.49** Since a bar graph is used for qualitative data, we separate the bars from each other to emphasize that the graph is not a histogram. There is no special ordering of the classes. If the bars were to touch, some viewers might infer an ordering representing quantitative data.

**2.50** Bar graphs are more useful if one is trying to convey the actual frequencies to the viewer. On the other hand, pie charts can always be used to compare the relative frequencies for two or more different sets of data.

**2.51** (a) Each rectangle in the frequency histogram would be of a height equal to the corresponding number of dots in the dotplot.

(b) If the classes for the histogram were based on multiple values, there would not be one rectangle for each column of dots (there would be fewer rectangles than columns of dots). The height of each rectangle would be equal to the total number of dots between each adjacent pair of cutpoints. If the classes were constructed so that only a few columns of dots corresponded to each rectangle, the general impression of the shape of the distribution should remain the same even though the details did not appear to be the same.

**2.52** (a) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in this exercise; i.e., columns 1 and 2. The lower class limits of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical-axis units in the frequency histogram are the even integers 0 through 12, since these are representative of the magnitude and spread of the frequency presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.

(b) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in this exercise; i.e., columns 1 and 3. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 vary in size from 0.00 to 0.20. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.04 (or 4%), from zero to 0.24 (or 24%) to avoid the tallest bar touching the top of the graph. The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.

Figure (a)

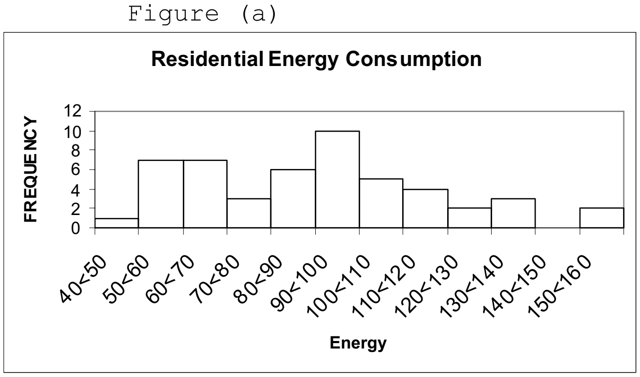
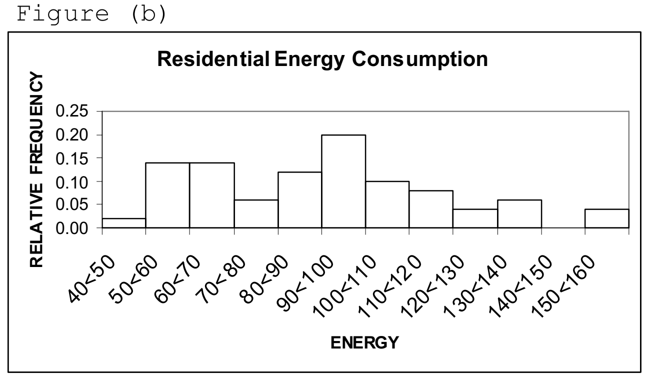


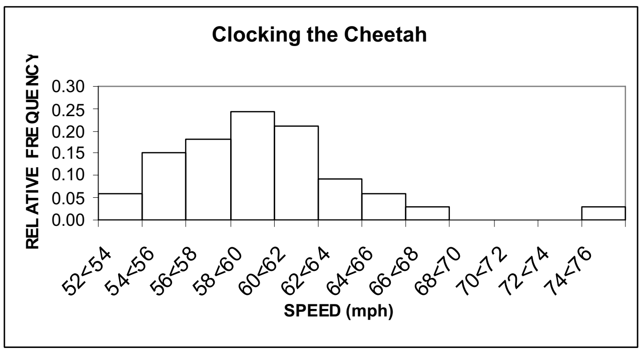

Figure (b)

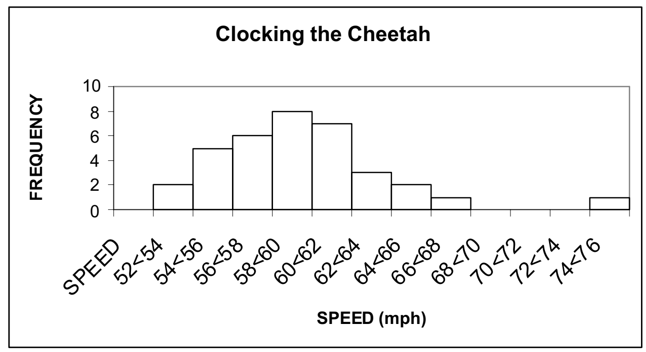



**2.53** (a) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in this exercise; i.e., columns 1 and 2. The lower class limits of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical axis units in the frequency histogram are the even integers 0 through 8, since these are representative of the magnitude and spread of the frequencies presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.

(b) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in this exercise; i.e., columns 1 and 3. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.000 to 0.229. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.25 (or 25%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.

Figure (a) Figure (b)




**2.54** (a) The frequency histogram in Figure (a) is constructed using the frequency distribution obtained in Exercise 2.28; i.e., columns 1 and 2. The lower cutpoints of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical axis units in the frequency histogram are the even integers 0 through 10, since these are representative of the magnitude and spread of the frequencies presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.

(b) The relative-frequency histogram in Figure (b) is constructed using the relative frequency distribution obtained in Exercise 2.28; i.e., columns 1 and 3. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.000 to 0.3125. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.35 (or 35%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.

Figure (a)

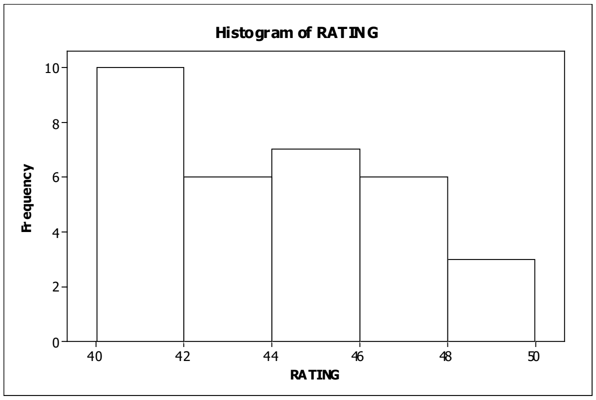
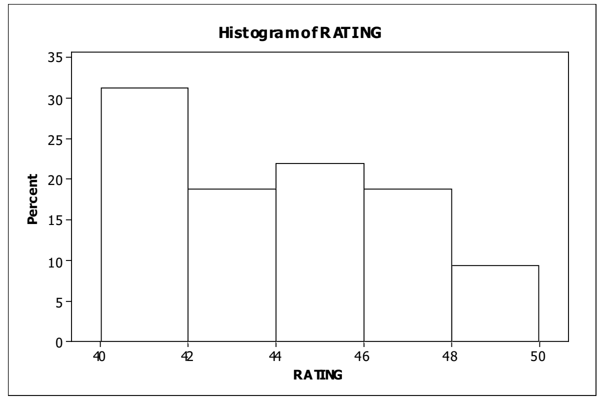



Figure (b)




**2.55** (a) The frequency histogram in Figure (a) is constructed using the frequency distribution obtained in Exercise 2.29; i.e., columns 1 and 2. Column 1 demonstrates that the data are grouped using classes with class widths of 3 starting at the midpoint 12.5. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 0 through 6, since these are representative of the magnitude and spread of the frequencies presented in column 2. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

(b) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution obtained in Exercise 2.29; i.e., columns 1 and 3. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.05 to 0.30. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (5%), from zero to 0.30 (30%). The middle of each histogram bar is placed directly over the class midpoint. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.

(c) Figure (a)

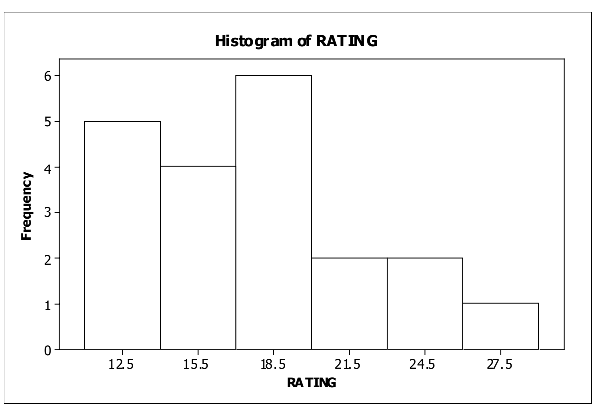
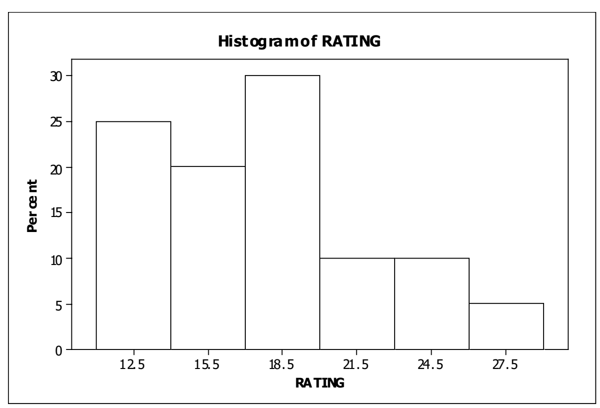



Figure (b)




**2.56** (a) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in Exercise 2.30; i.e., columns 1 and 2. Column 1 demonstrates that the data are grouped using classes based on a single value. These single values in column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical axis units in the frequency histogram are the even integers within the range 0 through 14, since these are representative of the magnitude and spread of the frequencies presented in column 2. When classes are based on a single value, the middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

(b) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in this exercise; i.e., columns 1 and 3. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.025 to 0.325. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.35 (or 35%). The middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.

Figure (a)

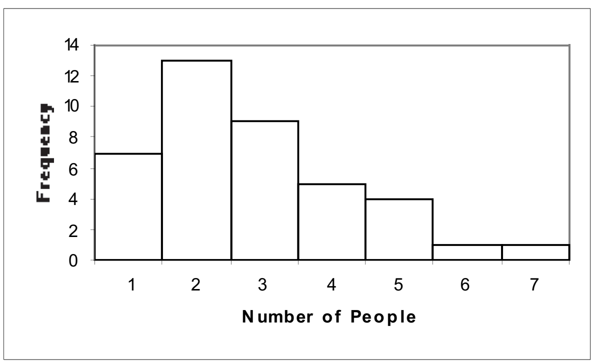
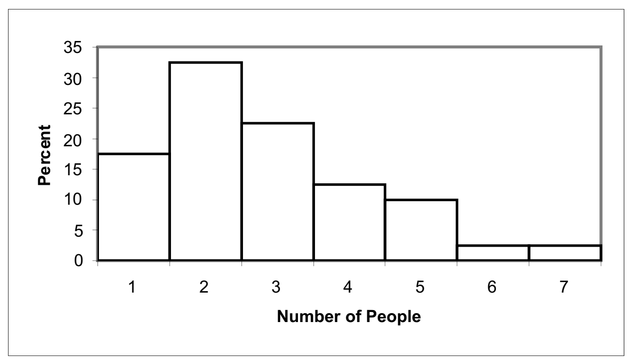



Figure (b)




**2.57** (a) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in Exercise 2.31; i.e., columns 1 and 2. Column 1 demonstrates that the data are grouped using classes based on a single value. These single values in column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 0 through 7, since these are representative of the magnitude and spread of the frequencies presented in column 2. When classes are based on a single value, the middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

(c) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in this exercise; i.e., columns 1 and 3. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.025 to 0.425. Thus, suitable candidates for vertical-axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.45 (or 45%). The middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.

Figure (a)

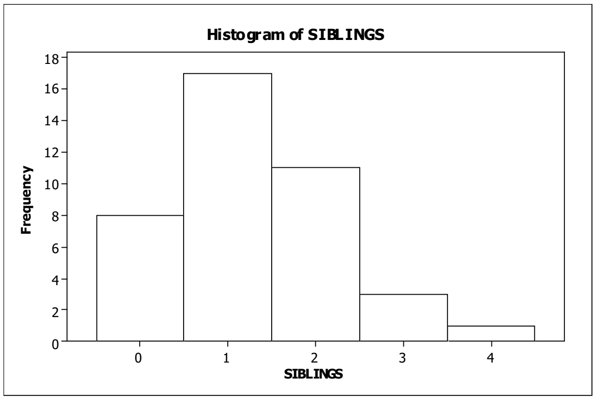
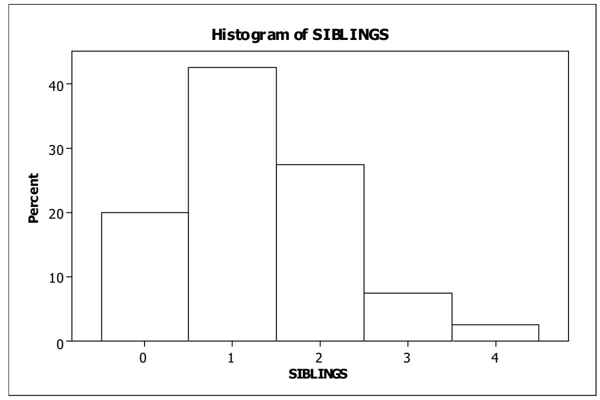
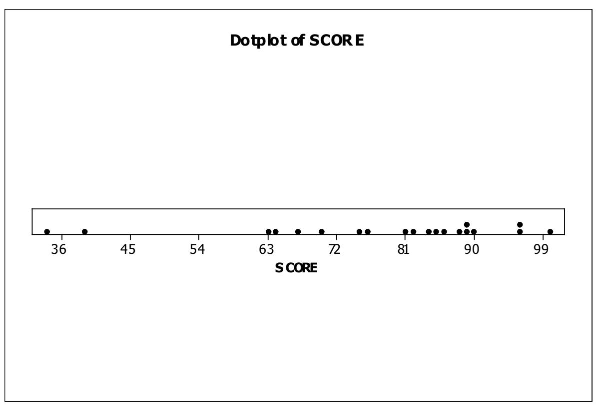


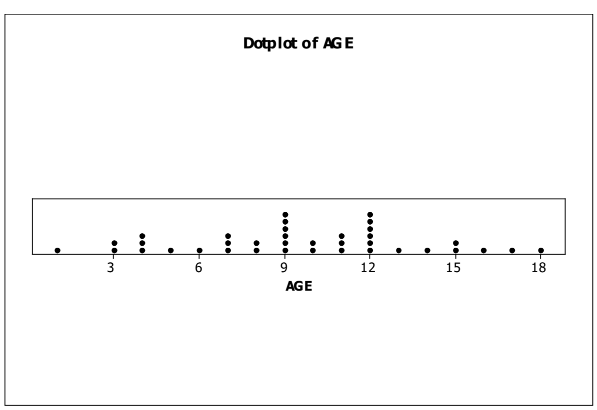

Figure (b)



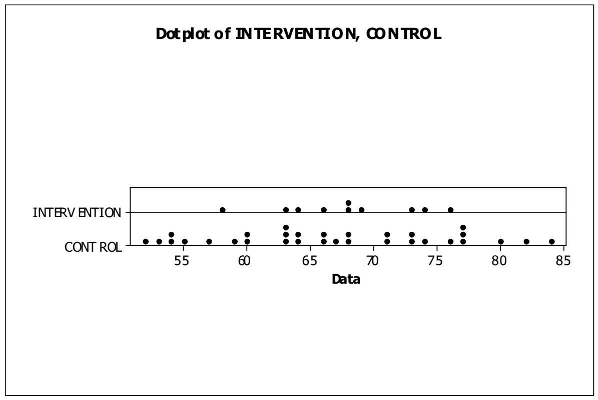

**2.58** The horizontal axis of this dotplot displays a range of possible exam scores. To complete the dotplot, we go through the data set and record each exam score by placing a dot over the appropriate value on the horizontal axis.




**2.59** The horizontal axis of this dotplot displays a range of possible ages. To complete the dotplot, we go through the data set and record each age by placing a dot over the appropriate value on the horizontal axis.

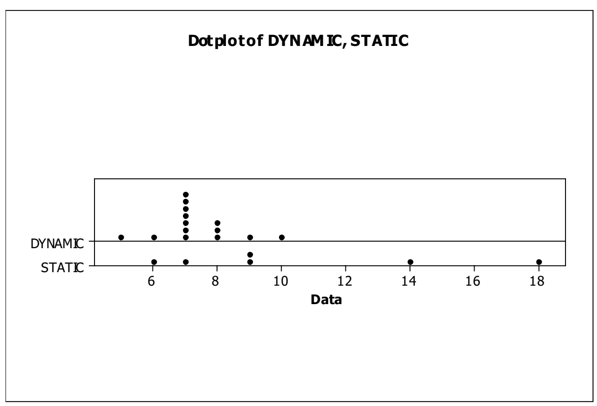


**2.60** (a) The data values range from 52 to 84, so the scale must accommodate those values. We stack dots above each value on two different lines using the same scale for each line. The result is




(b) The two sets of pulse rates are both centered near 68, but the Intervention data are more concentrated around the center than are the Control data.

**2.61** (a) The data values range from 7 to 18, so the scale must accommodate those values. We stack dots above each value on two different lines using the same scale for each line. The result is




(b) The dynamic system does seem to reduce acute postoperative days in the hospital on the average. The Dynamic data are centered at about 7 days, whereas the Static data are centered at about 11 days and are much more spread out than the Dynamic data.

**2.62** Since each data value consists of 3 or 4 digits, we will drop the last digit (ones) and construct the stem-and-leaf diagram with the remaining digits. All of the stems are for data beginning with a 9 or a 10, so we will use five lines per stem. The result is

```
 9  1
 9
 9  455
 9  67777
 9  88888999999
10  01111
10  2233
10
10  6
```

**2.63** Since each data value consists of 2 digits, each beginning with 1, 2, 3, or 4, we will construct the stem-and-leaf diagram with two lines per stem. The result is

```
1  23
1  8
2  1
2  678899
3  344
3  59
4  04
```

**2.64** (a) Since each data value lies between 2 and 93, we will construct the stem-and-leaf diagram with one line per stem. The result is

```
0  2234799
1  11145566689
2  023479
3  004555
4  19
5  5
6  9
7  9
8
9  3
```

(b) Using two lines per stem, the same data result in the following diagram:

```
0  2234
0  799
1  1114
1  5566689
2  0234
2  79
3  004
3  555
4  1
4  9
5
5  5
6
6  9
7
7  9
8
8
9  3
```

(c) The stem with one line per stem is more useful. One gets the same impression regarding the shape of the distribution, but the two lines per stem version has numerous lines with no data, making it take up more space than necessary to interpret the data.

**2.65** (a) Since each data value lies between 26.6 and 34.0, we will construct the stem-and-leaf diagram with one line per stem. The result is

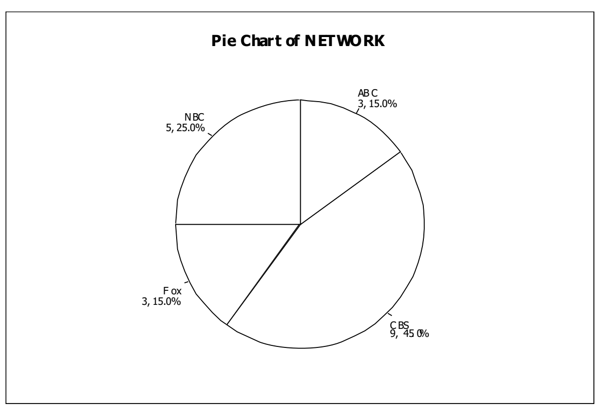
```
26  6
27  7999
28  1456789
29  133447889
30  3579
31  008
32  7
33
34  0
```

(b) Using two lines per stem, the same data result in the following diagram:

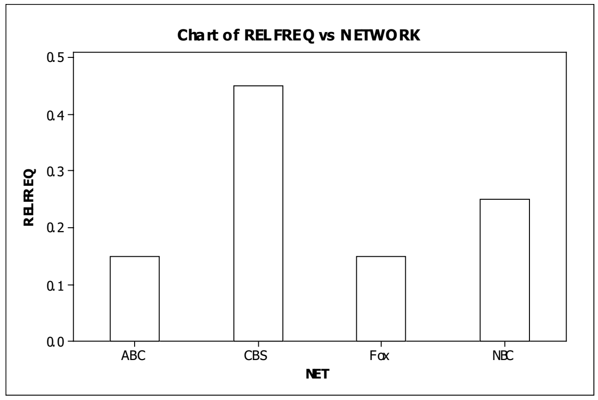
```
26  6
27
27  7999
28  14
28  56789
29  13344
29  7889
30  3
30  579
31  00
31  8
32
32  7
33
33
34  0
```

(c) Both diagrams have an acceptable number of lines. The second diagram makes it a little easier to see how extreme the youngest and oldest teams are, so we find that one slightly more useful.

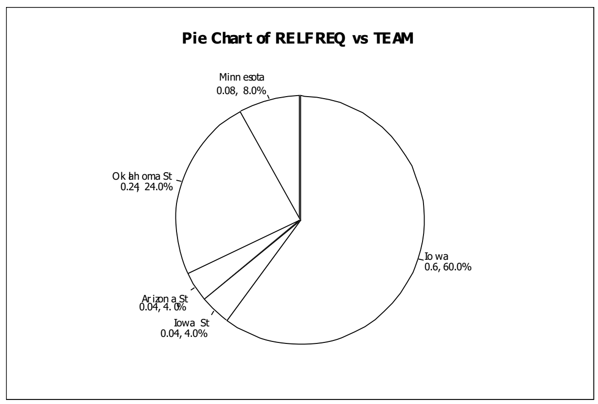
**2.66** (a) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each network. The result is



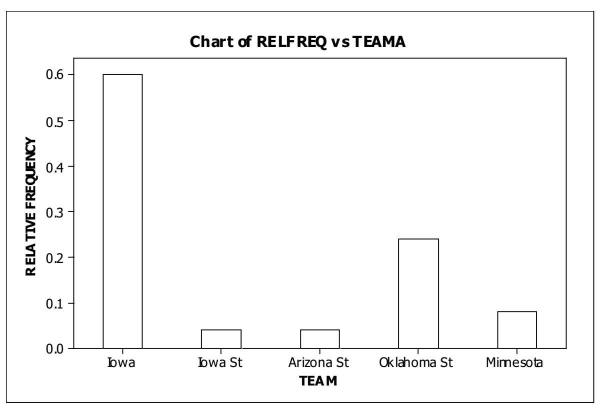

(b) We use the bar chart to show the relative frequency with which each network occurs. The result is



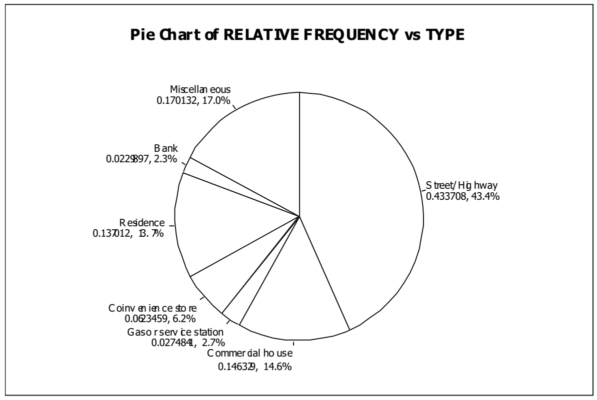

**2.67** (a) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each team. The result is



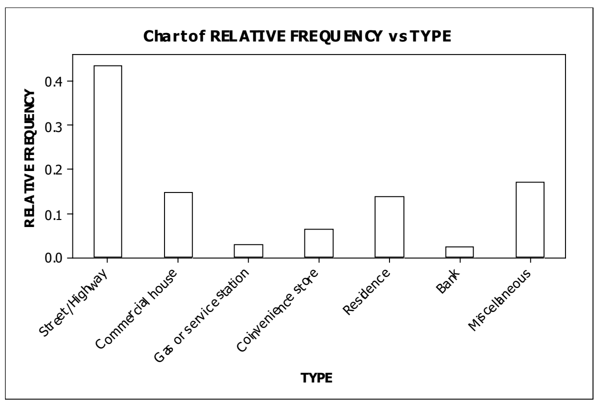

(b) We use the bar chart to show the relative frequency with which each TEAM occurs. The result is



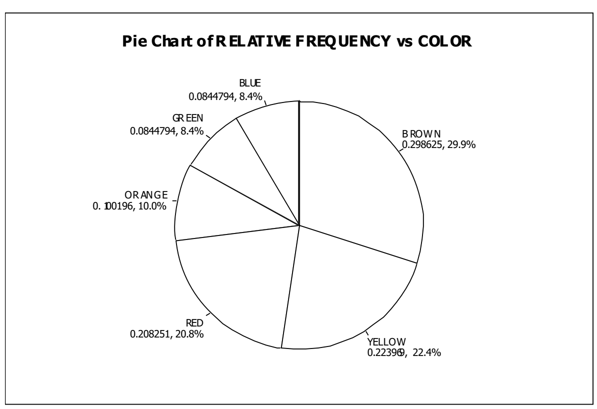
**2.68** (a) We first find each of the relative frequencies by dividing each of the frequencies by the total frequency of 413403. Then we multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each robbery type. The result is



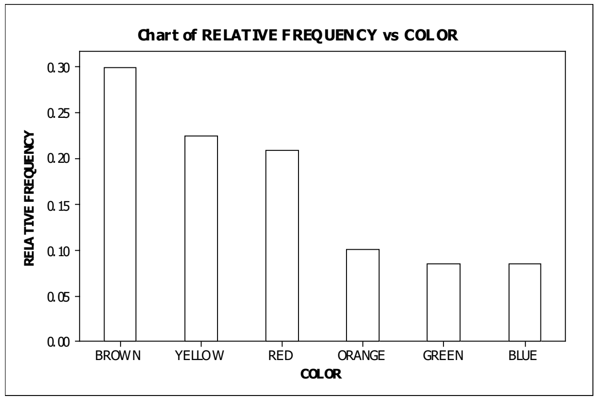

(b) We use the bar chart to show the relative frequency with which each robbery type occurs. The result is




**2.69** (a) We first find each of the relative frequencies by dividing each of the frequencies by the total frequency of 509. Then we multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each color of M&M. The result is



(b) We use the bar chart to show the relative frequency with which each color occurs. The result is




**2.70** The heights of the bars of the relative-frequency histogram indicate that:

(a) About 27.5% of the returns had an adjusted gross income between $10,000 and $19,999, inclusive.

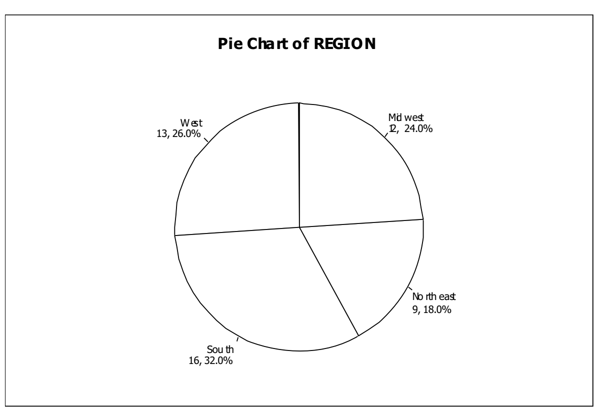
(b) About 37.5% were between $0 and $9,999; 27.5% were between $10,000 and $19,999; and 19% were between $20,000 and $29,999. Thus, about 84% (i.e., 37.5% + 27.5% + 19%) of the returns had an adjusted gross income less than $30,000.

(c) About 11% were between $30,000 and $39,999; and 5% were between $40,000 and $49,999. Thus, about 16% (i.e., 11% + 5%) of the returns had an adjusted gross income between $30,000 and $49,999. With 89,928,000 returns having an adjusted gross income less than $50,000, the number of returns having an adjusted gross income between $30,000 and $49,999 was 14,388,480 (i.e., 0.16 x 89,928,000).

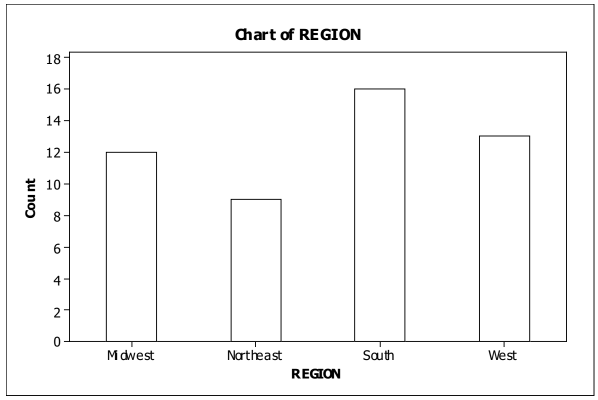
**2.71** The graph indicates that:

(a) 20% of the patients have cholesterol levels between 205 and 209, inclusive.
(b) 20% are between 215 and 219; and 5% are between 220 and 224. Thus, 25% (i.e., 20% + 5%) have cholesterol levels of 215 or higher.
(c) 35% of the patients have cholesterol levels between 210 and 214, inclusive. With 20 patients in total, the number having cholesterol levels between 210 and 214 is 7 (i.e., 0.35 x 20).

**2.72** (a) After retrieving the data from the WeissStats CD, select **Graph ▶ Pie Chart**, click on **Chart raw data**, and enter **Region** in the **Categorical Variables** box. Then click on **Labels**, choose the **Slice Labels** tab and check all four boxes. Click **OK** and **OK**. The result (except for colors and the legend which we deleted) is

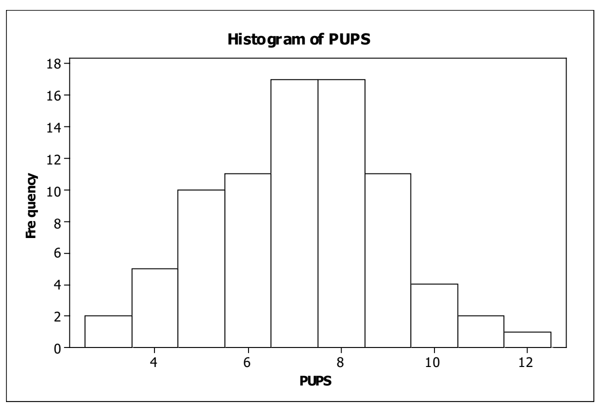


(b) Select **Graph ▶ Bar Chart**, choose **Simple**, and click **OK.** Enter REGION in the **Categorical variables** box and click **OK.** The result is (except for shading)




(c) Both graphs show the counts for the regions to be 12, 9, 16, and 14 for the Midwest, Northeast, South, and West respectively. In addition, the pie chart shows the percentages of the 50 states that are in each region.

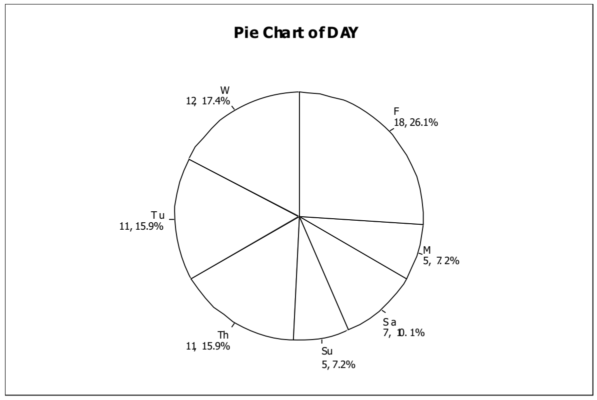
**2.73** (a) After retrieving the data from the WeissStats CD, select **Graph ▶ Histogram**, choose **Simple** and click **OK.** Double click on PUPS in the first box to enter PUPS in the **Graphs variables** box, and click **OK.** The frequency histogram is (except for shading)



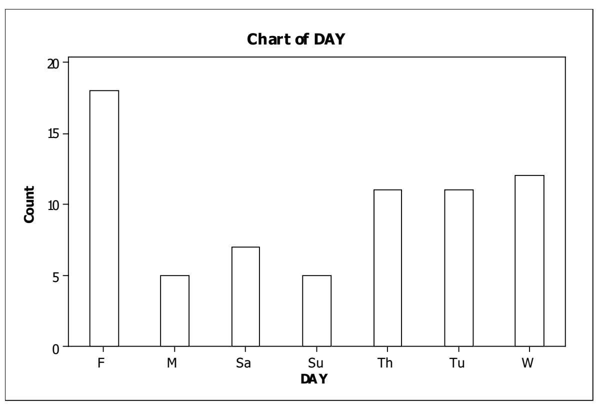

To change to a relative-frequency histogram, before clicking OK the second time, click on the **Scale** button and the **Y-Scale type** tab, and choose **Percent** and click **OK.** The graph will look like the one in part (a), but will have relative frequencies on the vertical scale instead of counts.

(b) The numbers of pups range from 1 to 12 per female with 7 and 8 pups occurring more frequently than any other values.

**2.74** (a) After retrieving the data from the WeissStats CD, select **Graph ▶ Pie Chart**, click on **Chart raw data**, and enter **DAY** in the **Categorical Variables** box. Then click on **Labels**, choose the **Slice Labels** tab and check all four boxes. Click **OK** and **OK.** The result (except for colors and the legend which we deleted) is

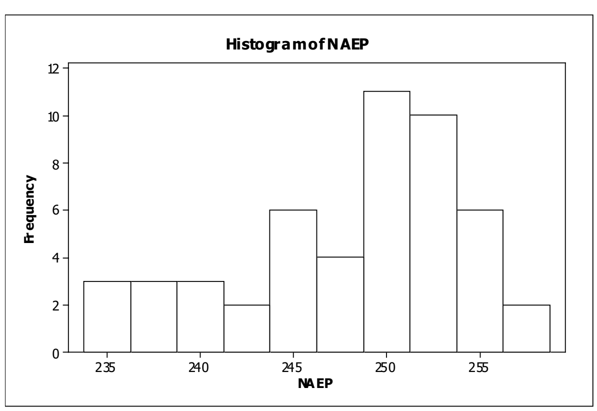



(b) Select **Graph ▶ Bar Chart**, choose **Simple**, and click **OK.** Enter DAY in the **Categorical variables** box and click **OK.** The result is (except for shading)



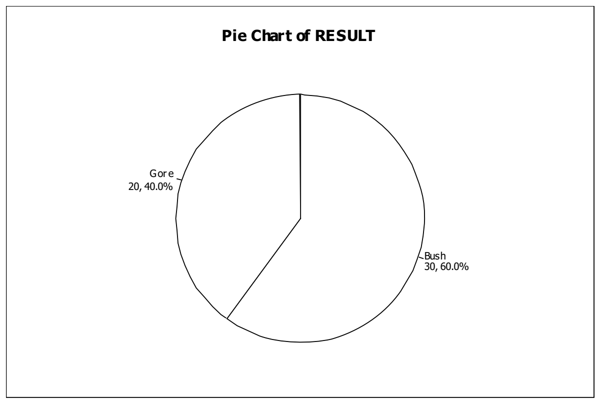

(c) The graphs show that Friday has the most road rage incidents and that Saturday, Sunday, and Monday have the fewest.

**2.75** (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, choose **Simple** and click **OK.** Double click on NAEP to enter NAEP in the **Graph variables** box and click **OK.** The result is

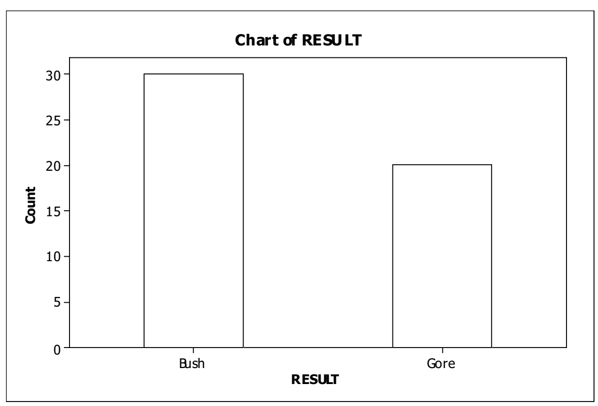


Minitab has chosen the class cutpoints. We see that the scores tend to cluster near the high end of the range of scores with a considerable number of scores somewhat lower than the most common score.

(b) To obtain the pie chart, select **Graph ▶ Pie Chart**, put the cursor in the **Categorical variables** box and double click on RESULT to enter RESULT in the **Categorical variables** box. Click on the **Labels** button and then on the **Slice Labels** tab. Check all four boxes and click **OK** twice. The result (except for colors and legend) is

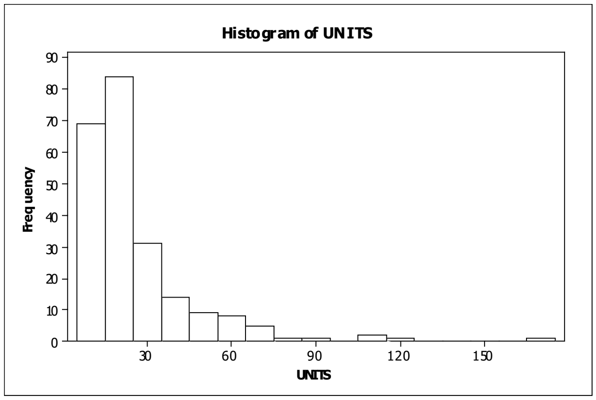



(c) To obtain the bar graph, select **Graph ▶ Bar Chart**, select **Simple**, and click **OK.** Double click on RESULT to enter RESULT in the **Categorical variables** box and click **OK.** The result (except for shading) is



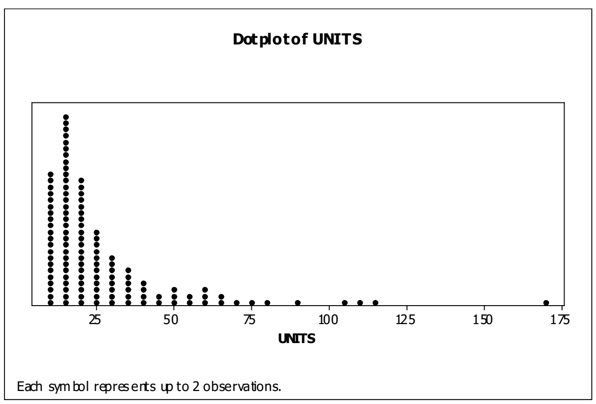

Both charts show that Bush won 30 states and Gore won 20. These graphs do not reflect the proportion of votes cast for Bush and Gore.

**2.76** (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, choose **Simple** and click **OK.** Double click on UNITS to enter UNITS in the **Graph variables** box and click **OK.** The result (except for shading) is




The graph shows that there are only a few artists who sell many units and many artists who sell relatively few units.

(b) To obtain the dotplot, select **Graph** ▶ **Dotplot**, select **Simple** in the **One Y** row, and click **OK**. Double click on UNITS to enter UNITS in the **Graph variables** box and click **OK**. The result (except for shading) is




(c) The graphs are similar, but not identical. This is because Minitab grouped the data values slightly differently for the two graphs. The overall impression, however, remains the same.

**2.77** (a) After entering the data from the WeissStats CD, in Minitab, select **Graph** ▶ **Stem-and-Leaf**, double click on PERCENT to enter PERCENT in the **Graph variables** box and enter a 10 in the **Increment** box, and click **OK**. The result is

```
Stem-and-leaf of PERCENT  N  = 51
Leaf Unit = 1.0


 2    7  79
(41)  8  00111111123445566666667777888888999999999
 8    9  00011122
```

(b) Repeat part (a), but this time enter a 5 in the **Increment** box. The result is

```
Stem-and-leaf of PERCENT  N  = 51
Leaf Unit = 1.0

  2    7  79
 15    8  0011111112344
(28)   8  5566666667777888888999999999
  8    9  00011122
```

(c) Repeat part (a) again, but this time enter a 2 in the **Increment** box. The result is

```
Stem-and-leaf of PERCENT  N  = 51
Leaf Unit = 1.0

  1    7  7
  2    7  9
 11    8  001111111
 13    8  23
 17    8  4455
(11)   8  66666667777
 23    8  888888999999999
  8    9  000111
  2    9  22
```

(d) The last graph is the most useful since it gives a better idea of the shape of the distribution. Typically, we like to have five to fifteen classes and this is the only one of the three graphs that satisfies that condition.

**2.78** (a) After entering the data from the WeissStats CD, in Minitab, select **Graph** ▶ **Stem-and-Leaf**, double click on PERCENT to enter PERCENT in the **Graph variables** box and enter a 10 in the **Increment** box, and click **OK.** The result is

```
Stem-and-leaf of RATE  N  = 51
Leaf Unit = 1.0

  8    2  33467899
 22    3  01112233567789
(19)   4  1111222223333668899
 10    5  02223346
  2    6  1
```

(b) Repeat part (a), but this time enter a 5 in the **Increment** box. The result is

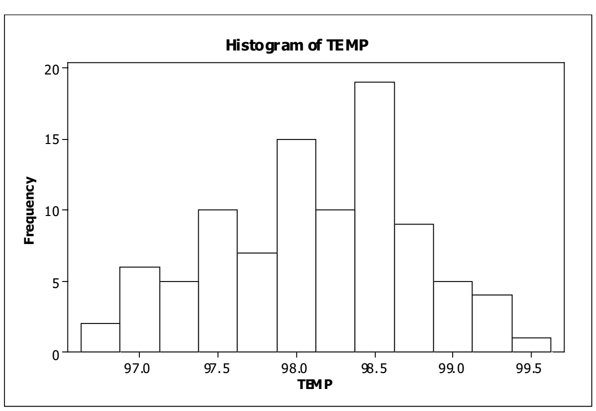
```
Stem-and-leaf of RATE  N  = 51
Leaf Unit = 1.0

  3    2  334
  8    2  67899
 16    3  01112233
 22    3  567789
(13)   4  1111222223333
 16    4  668899
 10    5  0222334
  3    5  6
  2    6  1
  1    6
  1    7
  1    7  7
```

(c) Repeat part (a) again, but this time enter a 2 in the **Increment** box. The result is

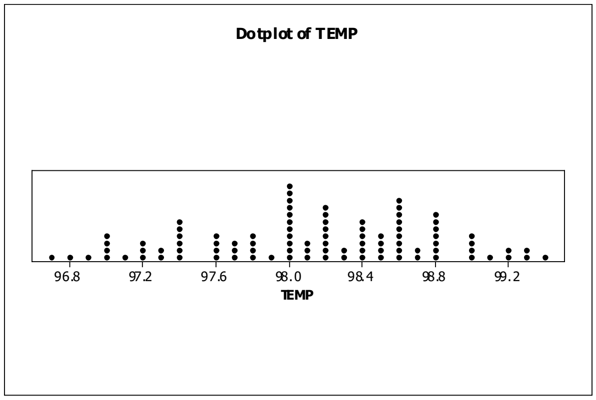
```
Stem-and-leaf of RATE  N  = 51
Leaf Unit = 1.0

 2   2  33
 3   2  4
 5   2  67
 8   2  899
 12  3  0111
 16  3  2233
 17  3  5
 20  3  677
 22  3  89
(4)  4  1111
 25  4  222223333
 16  4
 16  4  66
 14  4  8899
 10  5  0
 9   5  22233
 4   5  4
 3   5  6
 2   5
 2   6  1
 1   6
 1   6
 1   6
 1   6
 1   7
 1   7
 1   7
 1   7  7
```

(d) The second graph is the most useful. The third one has more classes than necessary to comprehend the shape of the distribution and has a number of empty stems. Typically, we like to have five to fifteen classes and the first and second diagrams satisfy that condition, but the second one provides a better idea of the shape of the distribution.

**2.79** (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, select **Simple** and click **OK**. double click on TEMP to enter TEMP in the **Graph variables** box and click **OK**. The result is



(b) Now select **Graph ▶ Dotplot**, select **Simple** in the **One Y** row, and click **OK.** Double click on TEMP to enter TEMP in the **Graph variables** box and click **OK.** The result is

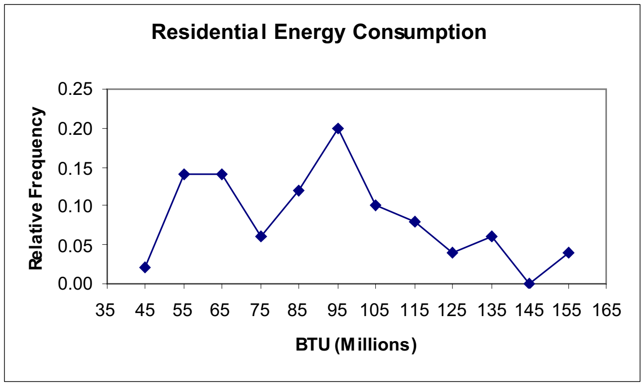



(c) Now select **Graph ▶ Stem-and-Leaf**, double click on TEMP to enter TEMP in the **Graph variables** box and click **OK.** Leave the **Increment** box blank to allow Minitab to choose the number of lines per stem. The result is

```
Stem-and-leaf of TEMP  N  = 93
Leaf Unit = 0.10

 1    96  7
 3    96  89
 8    97  00001
 13   97  22233
 19   97  444444
 26   97  6666777
 31   97  88889
 45   98  00000000000111
(10)  98  2222222233
 38   98  4444445555
 28   98  66666666677
 17   98  8888888
 10   99  00001
 5    99  2233
 1    99  4
```

(d) The dotplot shows all of the individual values. The stem-and-leaf diagram used five lines per stem and therefore each line contains leaves with possibly two values. The histogram chose classes of width 0.25. This resulted in, for example, the class with midpoint 97.0 including all of the values 96.9, 97.0, and 97.1, while the class with midpoint 97.25 includes only the two values 97.2 and 97.3. Thus the 'smoothing' effect is not as good in the histogram as it is in the stem-and-leaf diagram. Overall, the dotplot gives the truest picture of the data and allows recovery of all of the data values.

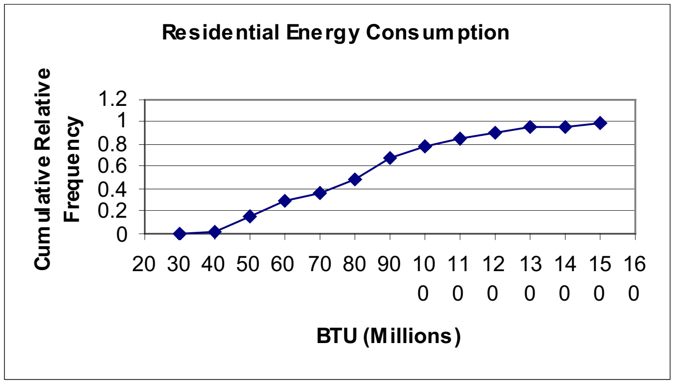
**2.80** Consider columns 1 and 3 of the energy-consumption data given in Exercise 2.24. Compute the midpoint for each class presented in column 1. Pair each midpoint with its corresponding relative frequency found in column 3. Construct a horizontal axis, where the units are in terms of midpoints and a vertical axis where the units are in terms of relative frequencies. For each midpoint on the horizontal axis, plot a point whose height is equal to the relative frequency of the class. Then join the points with connecting lines. The result is a relative-frequency polygon.




**2.81** (a) Consider all three columns of the energy-consumption data given in Exercise 2.24. Column 1 is now reworked to present just the lower cutpoint of each class. Column 2 is reworked to sum the frequencies of all classes representing values less than the specified lower cutpoint. These successive sums are the cumulative frequencies. Column 3 is reworked to sum the relative frequencies of all classes representing values less than the specified cutpoints. These successive sums are the cumulative relative frequencies. (Note: The cumulative relative frequencies can also be found by dividing the each cumulative frequency by the total number of data values.)

| Less than | Cumulative Frequency | Cumulative Relative Frequency |
|---|---|---|
| 40 | 0 | 0.00 |
| 50 | 1 | 0.02 |
| 60 | 8 | 0.16 |
| 70 | 15 | 0.30 |
| 80 | 18 | 0.36 |
| 90 | 24 | 0.48 |
| 100 | 34 | 0.68 |
| 110 | 39 | 0.78 |
| 120 | 43 | 0.86 |
| 130 | 45 | 0.90 |
| 140 | 48 | 0.96 |
| 150 | 48 | 0.96 |
| 160 | 50 | 1.00 |

(b) Pair each cutpoint in reworked column 1 with its corresponding cumulative relative frequency found in reworked column 3. Construct a horizontal axis, where the units are in terms of the cutpoints and a vertical axis where the units are in terms of cumulative relative frequencies. For each cutpoint on the horizontal axis, plot a point whose height is equal to the cumulative relative frequency. Then join the points with connecting lines. The result, presented in Figure (b), is an *ogive* using cumulative relative frequencies. (Note: A similar procedure could be followed using cumulative frequencies.)




**2.82** (a) After rounding each weight to the nearest integer (Any weight ending in .5 is rounded up.), the stem-and-leaf diagram for the rounded weights is

```
12   9
13   27
14   366
15   011599
16   12567
17   003445689
18   235789
19   1
20   9
21   58
22
23
24
25
26
27   9
```

(b) After truncating each weight, the stem-and-leaf diagram for the rounded weights is

```
12   9
13   26
14   2569
15   01588
16   11557
17   002335588
18   225778
19   1
20   9
21   48
22
23
24
25
26
27   8
```

(c) Although there are minor differences between the two diagrams, the overall impression of the distribution of weights is the same for both diagrams.

**2.83** (a) After rounding to the nearest 10 and then dropping the final zero, the stem-and-leaf diagram is

```
 9  1
 9
 9  5
 9  6667
 9  8888999999
10  0000011
10  223333
10
10  6
```

(b) After truncating each observation to the 10s digit, the stem-and-leaf diagram is

```
 9  1
 9
 9  455
 9  67777
 9  88888999999
10  01111
10  2233
10
10  6
```

(c) The overall impression of the shape of the distribution is the same for the diagrams in parts (a) and (b) although there is a slight shift to lower values in part (b). This is due to truncating instead of rounding. The diagram in part (b) is the same as the one in Exercise 2.62.

**2.84** Minitab used truncation. Note that Kobe Bryant's average of 27.6 PPG would have been 28 if it had been rounded. Also the data values 25.7 and 26.1 would both have rounded to 26, but there is only one 26 in the stem-and-leaf diagram.

**Section 2.4**

**2.85** (a) The distribution of a data set is a table, graph, or formula that provides the values of the observations and how often they occur.

(b) Sample data are the values of a variable for a sample of the population.

(c) Population data are the values of a variable for the entire population.

(d) Census data are the same as population data, a complete listing of all data values for the entire population.

(e) A sample distribution is the distribution of sample data.

(f) A population distribution is the distribution of population data.

(g) A distribution of a variable is the same as a population distribution.

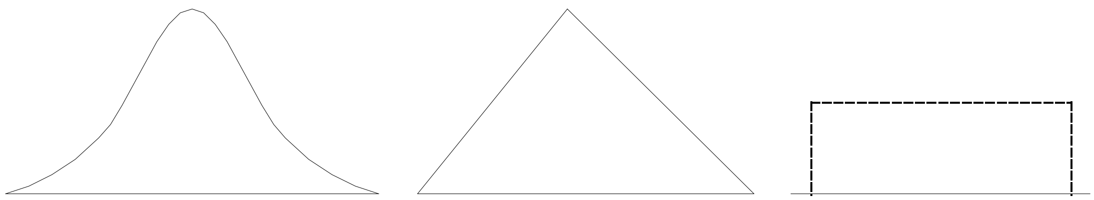
**2.86** A smooth curve makes it a little easier to see the shape of a distribution and to concentrate on the overall pattern without being distracted by minor differences in shape.

**2.87** A large simple random sample from a bell-shaped distribution would be expected to have roughly a bell-shaped distribution since more sample values should be obtained, on average, from the middle of the distribution.

**2.88** (a) Yes. We would expect both simple random samples to have roughly a reverse J-shaped distribution.

(b) Yes. We would expect some variation in shape between the two sample distributions since it is unlikely that the two samples would produce exactly the same frequency table. It should be noted, however, that as the sample size is increased, the difference in shape for the two samples should become less noticeable.

**2.89** Three distribution shapes that are symmetric are bell-shaped, triangular, and rectangular, shown in that order below. It should be noted that there are others as well.



Bell-shaped Triangular Uniform (or rectangular)

**2.90** (a) The overall shape of the distribution of the number of children of U.S. presidents is right skewed.
(b) The distribution is right skewed.

**2.91** (a) Except for the one data value between 74 and 76, this distribution is close to bell-shaped. That one value makes the distribution slightly right skewed.
(b) The distribution is slightly right skewed.

**2.92** (a) The distribution is approximately bell-shaped.
(b) The distribution is roughly symmetric.

**2.93** (a) The distribution of burrow depths is left skewed.
(b) The distribution is left skewed.

**2.94** (a) The distribution of heights is approximately bell-shaped.
(b) The distribution is nearly symmetric. The lowest one or two values make it slightly left skewed.

**2.95** (a) The distribution of shell thickness is approximately bell-shaped.
(b) The distribution is nearly symmetric.

**2.96** (a) The distribution of adjusted gross incomes is reverse J-shaped.
(a) The distribution is right skewed.

**2.97** (a) The distribution of cholesterol levels appears to be slightly left skewed.
(b) This distribution is nearly symmetric, but is slightly left skewed. Given that the data originated from patients who had high cholesterol levels, one would not expect symmetry. Individuals with low cholesterol levels were not patients and were not included in the testing.

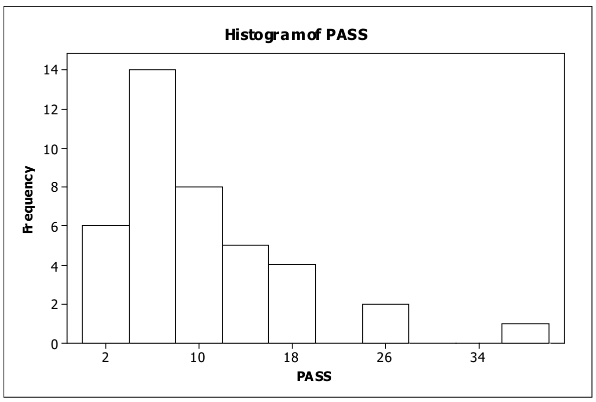
**2.98** (a) The distribution of hemoglobin levels for patients with sickle cell disease is approximately uniform.
(b) This distribution is approximately symmetric.

**2.99** (a) The distribution of length of stay is nearly reverse J-shaped. It is certainly right skewed.
(b) The distribution is right skewed.

**2.100** (a) The frequency distribution for this data is shown in the following table.

| Passengers (Millions) Midpoint | Frequency |
|---|---|
| 2 | 6 |
| 6 | 14 |
| 10 | 8 |
| 145 | 5 |
| 18 | 4 |
| 22 | 0 |
| 26 | 2 |
| 30 | 0 |
| 34 | 0 |
| 38 | 1 |

(b) The histogram for the distribution is shown below.

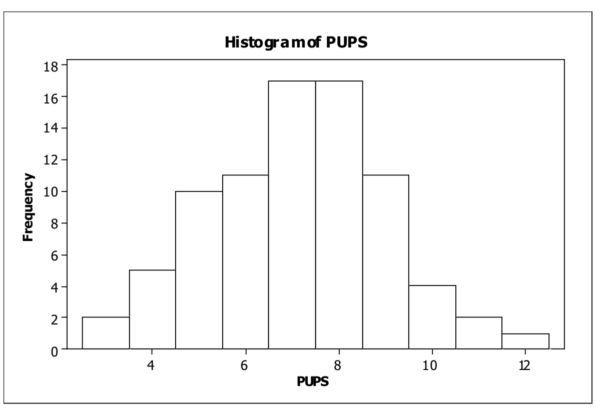



(c) This distribution is very much right skewed.

**2.101** (a) The distribution for 1993 is right skewed and the distribution for 1996 is reverse J shaped.

(b) Both distributions are right skewed.

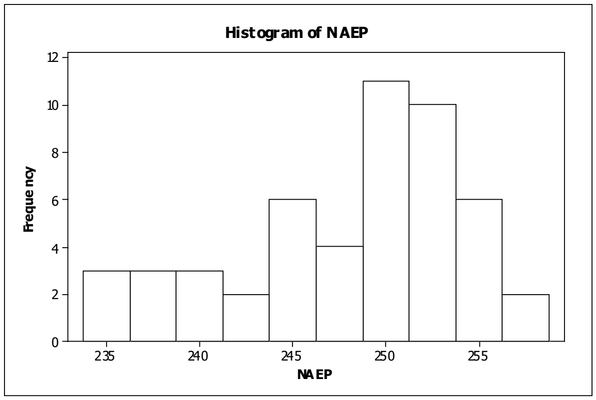
**2.102** (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, select **Simple** and click **OK.** Double click on PUPS to enter PUPS in the **Graph variables** box and click **OK.** The result is




(b) The overall shape of the distribution is bell-shaped.

(c) The distribution is roughly symmetric.

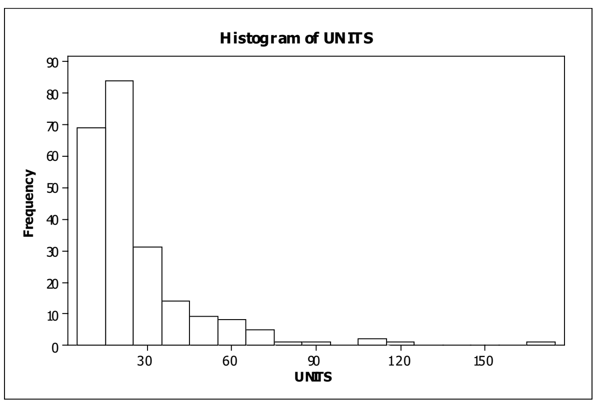
**2.103** (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, select **Simple** and click **OK.** double click on NAEP to enter NAEP in the **Graph variables** box and click **OK.** The result is




(b) The overall distribution of NAEP scores is left skewed.

(c) We classify this distribution as left skewed.

**2.104** (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, select **Simple** and click **OK**. double click on NAEP to enter NAEP in the **Graph variables** box and click **OK**. The result is




(b) The distribution of UNITS is definitely right skewed and comes very close to being reverse J-shaped. If the first class had a higher frequency than the second, we would call it reverse J shaped.

(c) The distribution is right skewed.

**2.105** (a) In Exercise 2.77, we used Minitab to obtain a stem-and-leaf diagram using 5 lines per stem. That diagram is shown below

```
Stem-and-leaf of PERCENT  N  = 51
Leaf Unit = 1.0

  1    7  7
  2    7  9
 11    8  001111111
 13    8  23
 17    8  4455
(11)   8  66666667777
 23    8  888888999999999
  8    9  000111
  2    9  22
```

(b) The overall shape of this distribution is left skewed.

(c) We classify the distribution of PERCENT as left skewed.

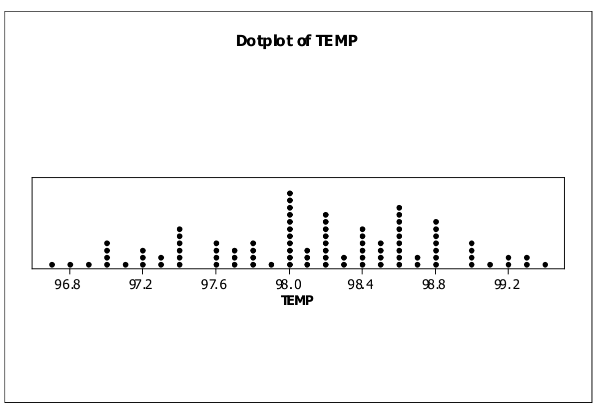
**2.106** (a) In Exercise 2.78, we used Minitab to obtain a stem-and-leaf diagram using 2 lines per stem. That diagram is shown below

```
Stem-and-leaf of RATE  N  = 51
Leaf Unit = 1.0


 3    2  334
 8    2  67899
 16   3  01112233
 22   3  567789
(13)  4  1111222223333
 16   4  668899
 10   5  0222334
 3    5  6
 2    6  1
 1    6
 1    7
 1    7  7
```

(b) The distribution of crime rates is slightly right skewed. Without the largest observation of 77, it would be approximately bell-shaped.

(c) We classify the distribution as right skewed.

**2.107** (a) After entering the data in Minitab, select **Graph ► Dotplot**, select **Simple** in the **One Y** row, and click **OK**. Double click on TEMP to enter TEMP in the **Graph variables** box and click **OK**. The result is




(b) The overall distribution of temperatures is roughly triangular.

(c) The distribution is fairly symmetric.

**2.108** Class Project. The precise answers to this exercise will vary from class to class.

**2.109** The precise answers to this exercise will vary from class to class or individual to individual. Thus your results will likely differ from our results shown below.

(a) We obtained 50 random digits from a table of random numbers. The digits were

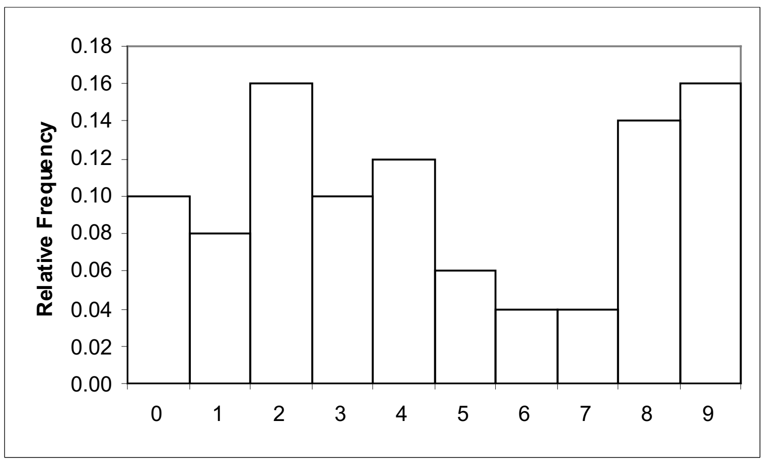
4 5 4 6 8 9 9 7 7 2 2 2 9 3 0 3 4 0 0 8 8 4 4 5 3

9 2 4 8 9 6 3 0 1 1 0 9 2 8 1 3 9 2 5 8 1 8 9 2 2

(b) Since each digit is equally likely in the random number table, we expect that the distribution would look roughly rectangular.

(c) Using single value classes, the frequency distribution is given by the following table. The histogram is shown below.

| Value | Frequency | Relative-Frequency |
|---|---|---|
| 0 | 5 | .10 |
| 1 | 4 | .08 |
| 2 | 8 | .16 |
| 3 | 5 | .10 |
| 4 | 6 | .12 |
| 5 | 3 | .06 |
| 6 | 2 | .04 |
| 7 | 2 | .04 |
| 8 | 7 | .14 |
| 9 | 8 | .16 |




We did not expect to see this much variation.

(d) We would have expected a histogram that was a little more 'even', more like a rectangular distribution, but the relatively small sample size can result in considerable variation from what is expected.

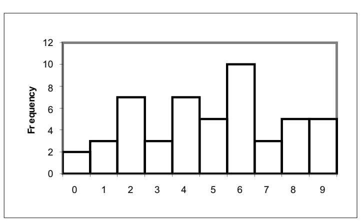
(e) We should be able to get a more evenly distributed set of data if we choose a larger set of data.

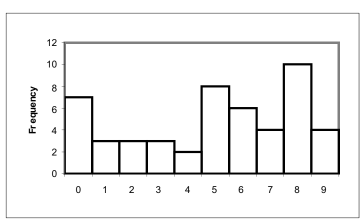
(f) Class project.

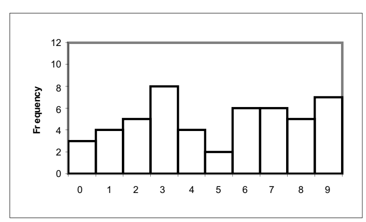
**2.110** (a-c) Your results will differ from the ones below which were obtained using Excel. Enter a name for the data in cell A1, say RANDNO. Click on cell A2 and enter **=RANDBETWEEN(0,9)**. Then copy this cell into cells A3 to A51. There are two ways to produce a histogram of the resulting data in Excel. The easier way is to highlight A!-A51 with the mouse, click on **DDXL** on the toolbar, select **Graphs and Plots**, then choose Histogram in the **Function type** box. Now click on RANDNO in the **Names and Columns** box and drag the name into the **Quantitative Variables** box. Then click OK. A graph and a summary table will be produced. To get five more samples, simply go back to the spreadsheet and press the F9 key. This will generate an entire new sample in Column A and you can repeat the procedure using DDXL. The only disadvantage of this method is that the graphs produced use white lines on a black background.

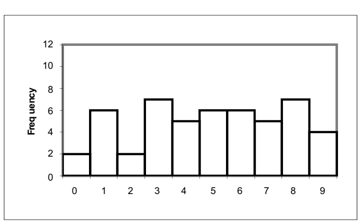
The second method is a bit more cumbersome and does not provide a summary chart, but yields graphs that are better for reproduction and that can be edited. Generate the data in the same way as was done above. In cells B1 to B10 enter the integers 0 to 9. These cells are called the **BIN**. Now click on **Tools, Data Analysis, Histogram.** (If Data Analysis is not in the Tools menu, you will have to add it from the original CD.) Click on the **Input** box and highlight cells A2-A51 with the mouse, then click on the **Bin** box and highlight cells B1-B10. Finally click on the **Output** box and enter **C1.** This will give you a frequency table in columns C and D. Now enter the integers 0 to 9 as text in cells E2 to E11 by entering each digit preceded by a single quote mark, i.e., **'0, '1,** etc. In cell F2, enter **=D2,** and copy this cell into F3 through F11. Now highlight the data in columns E and F with the mouse and click the chart icon, pick the **Column** graph type, pick the first sub-type, click on the **Next** button twice, enter any titles desired, remove the legend, and then click on the **Next** button and then the **Finish** button. The graph will appear on the spreadsheet as a bar chart with spaces between the bars. Use the mouse to point to any one of the bars and click with the right mouse button. Choose **Format Data Series.** Click on the **Options** tab and change the **Gap Width** to zero, and click **OK.** Repeat this sequence to produce additional histograms, but use different cells.

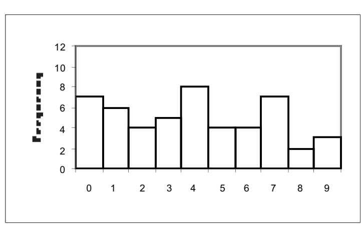
[If you would like to avoid repeating most of the above steps, click near the border of the graph and copy the graph to the Clipboard, then go to Microsoft Word or other word processor, and click on **Edit** on the Toolbar and **Paste Special.** Highlight **Microsoft Excel Chart Object,** and click **OK.** The graphs can be resized in the word processor if necessary. Now go back to Excel and hit the **F9** key. This will produce a completely new set of random numbers. Click on **Tools, Data Analysis, Histogram,** leave all the boxes as they are and click **OK.** Then click **OK** to overwrite existing data. A new table will be created and the existing histogram will be updated automatically. We used this process for the following histograms.]

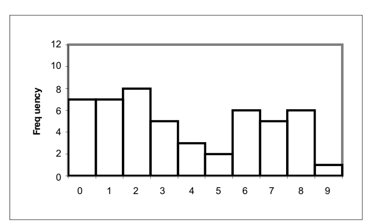


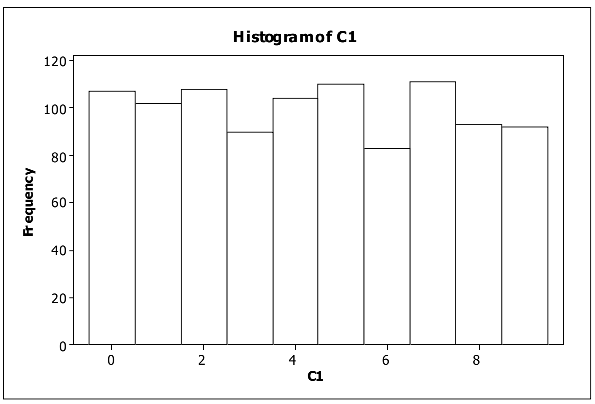



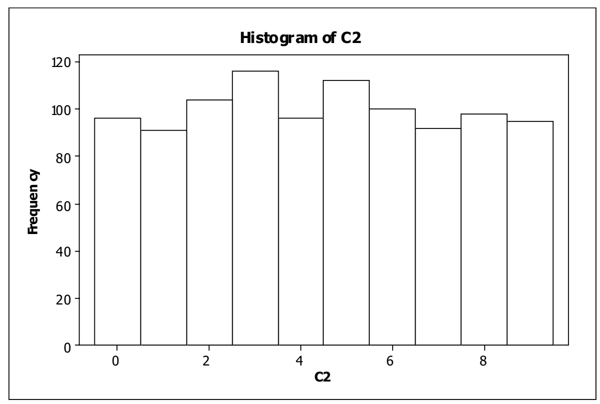



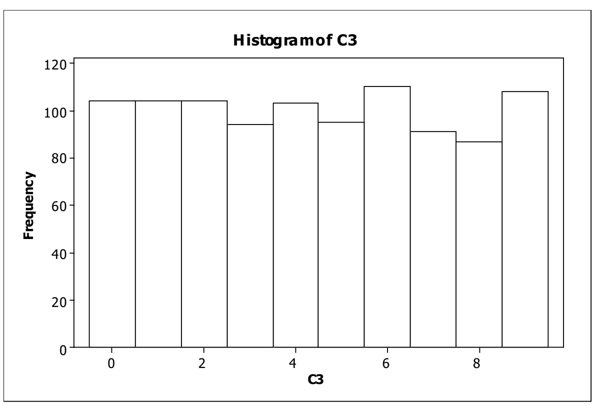

(d) These shapes are about what we expected.

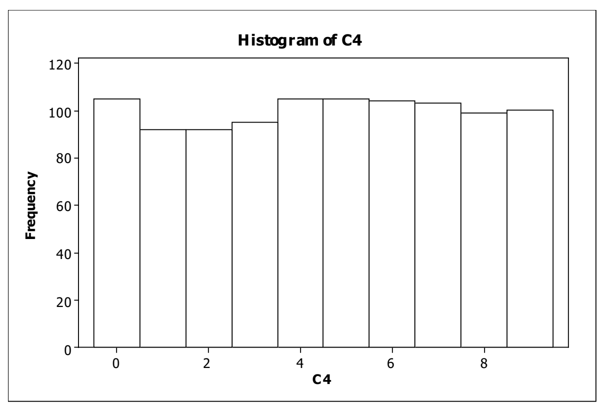
(e) The relative frequency histograms for six samples of digits of size

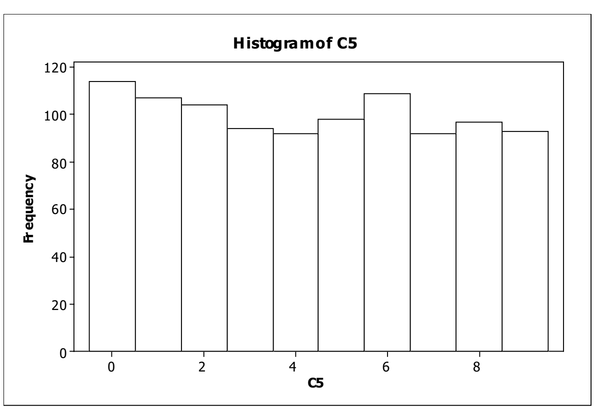
1000 were obtained using Minitab. Choose **Calc** ▶ **Random Data** ▶ **Integer...**, type 1000 in the **Generate rows of data** test box, click in the **Store in column(s)** text box and type C1 C2 C3 C4 C5 C6, click in the **Minimum value** text box and type 0, type in the **Maximum value** text box and type 9 and click **OK**. Then choose **Graph** ▶ **Histogram**, select the **Simple** version and click **OK**, enter C1 C2 C3 C4 C5 C6 in the **Graph variables** text box, C2 in the **Graph 2** text box un **x**, and so on for C3 through C6. Click on the **Multiple Graphs** button. Click on the **On separate graphs** button, and check the boxes for **Same Y** and **Same X, including same bins**. Click **OK** and click **OK**. The following graphs resulted.

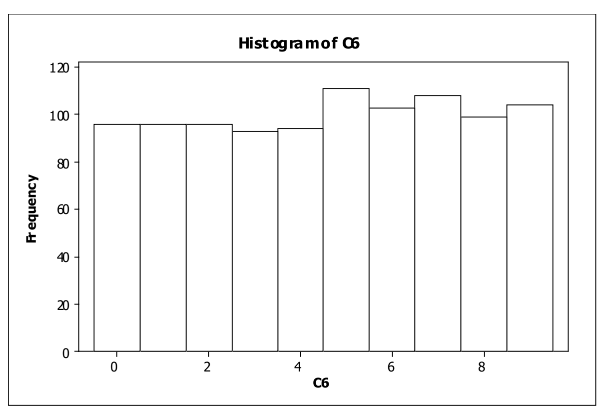


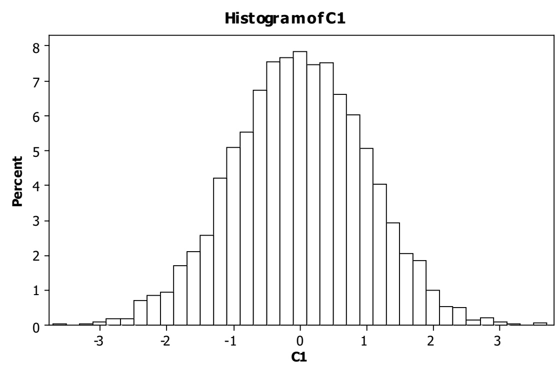



The histograms for samples of size 1000 are much closer to the rectangular distribution we expected than are the ones for samples of size 50.

**2.111** (a) Your result will differ from, but be similar to, the one below which was obtained using Minitab. Choose **Calc ► Random Data ► Normal...**, type 3000 in the **Generate rows of data** text box, click in the **Store in column(s)** text box and type C1, and click **OK.**

(b) Then choose **Graph ► Histogram**, choose the **Simple** version, click **OK**, enter C1 in the **Graph variables** text box, click on the **Scale** button and then on the **Y-Scale Type** tab. Check the **Percent** box and click **OK** twice.




(c) The histogram in part (b) has the shape of a standard normal distribution. The sample of 3000 is representative of the population from which the sample was taken.

**Section 2.5**

**2.112** Graphs are sometimes constructed in ways that cause them to be misleading.

**2.113** (a) A truncated graph is one for which the vertical axis starts at a value other than its natural starting point, usually zero.

(b) A legitimate motivation for truncating the axis of a graph is to place the emphasis on the ups and downs of the distribution rather than on the actual height of the graph.

(c) To truncate a graph and avoid the possibility of misinterpretation, one should start the axis at zero and put slashes in the axis to indicate that part of the axis is missing.

**2.114** Answers will vary.

**2.115** (a) A large lower portion of the graph is eliminated. When this is done, differences between district and national averages appear greater than in the original figure.

(b) Even more of the graph is eliminated. Differences between district and national averages appear even greater than in part (a).

(c) The truncated graphs give the misleading impression that, in 2005, the district average is much greater relative to the national average than it actually is.

**2.116** (a) A break is shown in the first bar on the left to warn the reader that part of the first bar itself has been removed.

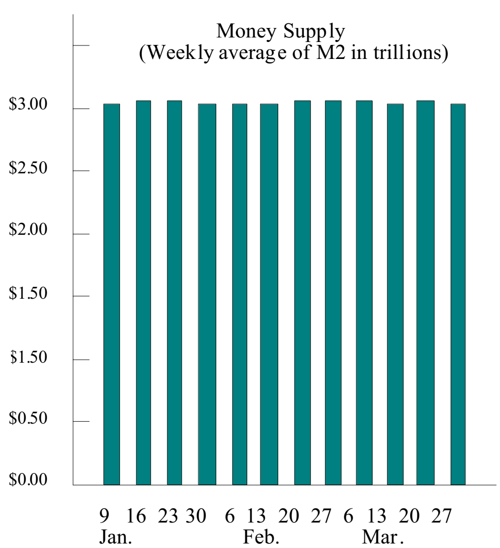
(b) It was necessary to construct the graph with a broken bar to let the reader know that the first bar is actually much taller than it appears. If the true height of the first bar were presented, but without the break, the height would span most of an entire page. This would have used up, perhaps, more room than that desired by the person reporting the graph.

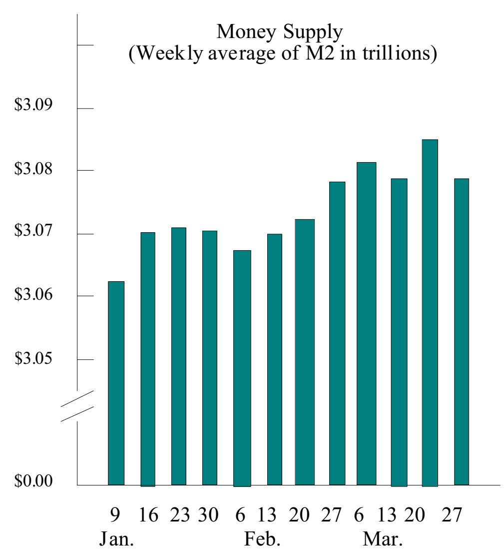
(c) This bar graph is potentially misleading if the reader does not pay attention to the *true* magnitude of the first bar relative to the other three bars. This is precisely the reason for the break in the first bar, however. It is meant to alert the reader that special treatment is to be applied to the first bar. It is actually much taller than it appears. Supplying the numbers for each bar of the graph makes it clear that there was no intention to mislead the reader. This was necessary also because there is no scale on the vertical axis.

**2.117** (a) The problem with the bar graph is that it is truncated. That is, the vertical axis, which should start at $0 (in trillions), starts with $3.05 (in trillions) instead. The part of the graph from $0 (in trillions) to $3.05 (in trillions) has been cut off. This truncation causes the bars to be out of correct proportion and hence creates the misleading impression that the money supply is changing more than it actually is.

(b) A version of the bar graph with an untruncated and unmodified vertical axis is presented in Figure (a). Notice that the vertical axis starts at $0.00 (in trillions). Increments are in halves of trillion dollars. In contrast to the original bar graph, this one illustrates that the changes in money supply from week to week are not that different. However, the "ups" and "downs" are not as easy to spot as in the original, truncated bar graph.

(c) A version of the bar graph in which the vertical axis is modified in an acceptable manner is presented in Figure (b). Notice that the special symbol "//" is used near the base of the vertical axis to signify that the vertical axis has been modified. Thus, with this version of the bar graph, not only are the "ups" and "downs" easy to spot but the reader is also aptly warned by the slashes that part of the vertical axis between $0.00 (in trillions) and $3.05 (in trillions) has been removed.




**2.118** (a) 1) The scale on the left begins at 140 million with the graph for the number of licensed drivers beginning at 150.2 million. 2) The scale on the right for the number of Drunk Driving Fatalities begins at 9000 with the graph ending at 13,041. Thus both graphs are truncated, making it look like the number of licensed drivers has increased more dramatically than it really has and making it look like the number of drunk driving fatalities has decreased proportionately more than it really has. 3) The difference between two tic marks on the left-hand scale is 10 million licensed drivers while the difference between two tic marks on the right-hand scale is 2500 fatalities. The graphs don't really cross. If a single scale were used, the fatalities graph would be far below the licensed drivers graph.

(b) All of that being said, to display both sets of data on one graph requires some accommodation due to the relative sizes of the numbers in the two sets of data.

(c) If both graphs are to be presented in one diagram, the axes should both have a broken line between the lowest two tic marks, with the lowest tic mark labeled with a zero. Another option is to produce two separate correctly prepared graphs since the scales are not related anyway.

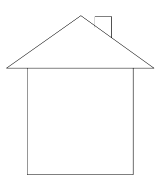
**2.119** (b) Without the vertical scale, it would appear that oil prices dropped about 75% from the peak price to the last day shown on the graph.

(c) The actual drop was from about \$70 per barrel to \$63.05 per barrel, a drop of about \$7 dollars per barrel. This is a drop of 7/70 = 0.10 or about 10%.

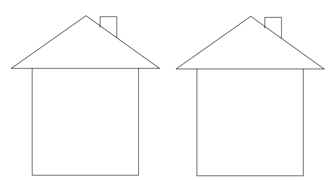
(d) The graph is potentially misleading because if the reader doesn't pay attention to the vertical scale, he or she may be led to conclude that the oil price was much more volatile than it actually was.

(e) The graph could be made less potentially misleading by either making the vertical scale range from zero to 70 or by starting at zero and putting a break in the vertical axis to call attention to the fact that part of the vertical axis is missing.

**2.120** A correct way in which the developer can illustrate the fact that twice as many homes will be built in the area this year as last year is as follows:



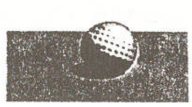
**Last Year**



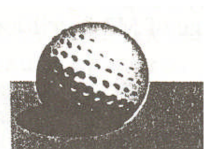
**This Year**

**2.121** (a) The brochure shows a "new" ball with twice the radius of the "old" ball. The intent is to give the impression that the "new" ball lasts roughly twice as long as the "old" ball. However, if the "new" ball has twice the radius of the "old" ball, the "new" ball will have eight times the volume of the "old" ball (since the volume of a sphere is proportional to the cube of its radius, or the radius $2^3 = 8$). Thus, the scaling is improper because it gives the impression that the "new" ball lasts eight times as long as the "old" ball rather than merely two times as long.

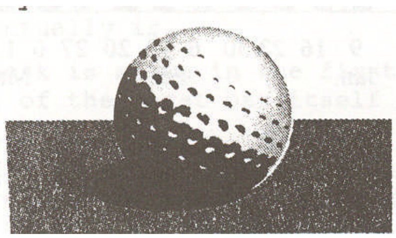
Old Ball



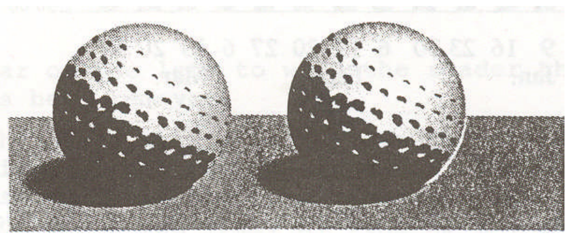
New Ball



(b) One possible way for the manufacturer to illustrate the fact that the "new" ball lasts twice as long as the "old" ball is to present pictures of two balls, side by side, each of the same magnitude as the picture of the "old" ball and to label this set of two balls "new ball". This will illustrate the point that a purchaser will be getting twice as much for the money.



Old Ball



New Ball

**Review Problems For Chapter 2**

**1.** (a) A variable is a characteristic that varies from one person or thing to another.

(b) Variables are quantitative or qualitative.

(c) Quantitative variables can be discrete or continuous.

(d) Data are values of a variable.

(e) The data type is determined by the type of variable being observed.

**2.** It is important to group data in order to make large data sets more compact and easier to understand.

**3.** The concepts of midpoints and cutpoints do not apply to qualitative data since no numerical values are involved in the data.

**4.** (a) The midpoint is halfway between the cutpoints. Since the class width is 8, 10 is halfway between 6 and 14.

(b) The class width is also the distance between consecutive midpoints. Therefore, the second midpoint is at 10 + 8 = 18.

(c) The sequence of cutpoints is 6, 14, 22, 30, 38, ... Therefore the lower and upper cutpoints of the third class are 22 and 30.

(d) An observation of 22 would go into the third class since that class contains data greater than or equal to 22 and strictly less than 30.

**5.** (a) The common class width is the distance between consecutive cutpoints, which is 15 - 5 = 10.

(b) The midpoint of the second class is halfway between the cutpoints 15 and 25, and is therefore 20.

(c) The sequence of cutpoints is 5, 15, 25, 35, 45, ... Therefore, the lower and upper cutpoints of the third class are 25 and 35.

**6.** Single value grouping is appropriate when the data are discrete with relatively few distinct observations.

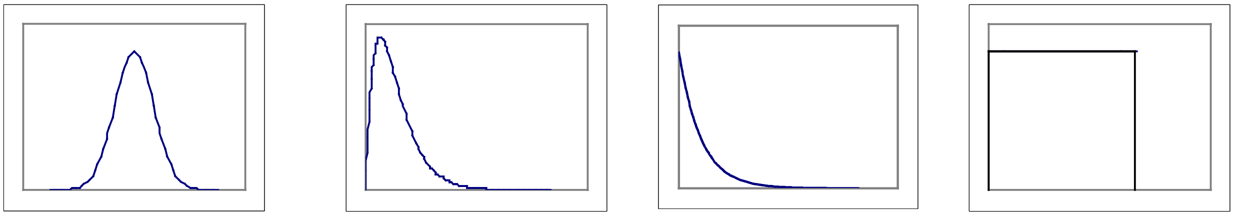
**7.** (a) The vertical edges of the bars will be aligned with the cutpoints.

(b) The horizontal centers of the bars will be aligned with the midpoints.

**8.** The two main types of graphical displays for qualitative data are the bar chart and the pie chart.

**9.** A histogram is better than a stem-and-leaf diagram for displaying large quantitative data sets since it can always be scaled appropriately and the individual values are of less interest than the overall picture of the data.

**10.** Bell-shaped Right skewed Reverse J shape Uniform



**11.** (a) Slightly skewed to the right. Assuming that the most typical heights are around 5'10", most heights below that figure would still be above 5'4", whereas heights above 5'10" extend to around 7'.

(b) Skewed to the right. High incomes extend much further above the mean income than low incomes extend below the mean.

(c) Skewed to the right. While most full-time college students are in the 17-22 age range, there are very few below 17 while there are many above 22.

(d) Skewed to the right. The main reason for the skewness to the right is that those students with GPAs below fixed cutoff points have been suspended by the time they would have been seniors.

**12.** (a) The distribution of the large simple random sample will reflect the distribution of the population, so it would be left-skewed as well.

(b) No. The randomness in the samples will certainly produce different sets of observations resulting in shapes that are not identical.

(c) Yes. We would expect both of the simple random samples to reflect the shape of the population and be left-skewed.

**13.** (a) The first column ranks the hydroelectric plants. Thus, it consists of *quantitative, discrete* data.

(b) The fourth column provides measurements of capacity. Thus, it consists of *quantitative, continuous* data.

(c) The third column provides nonnumerical information. Thus, it consists of *qualitative* data.

**14.** (a) The first class to construct is 40-44. Since all classes are to be of equal width, and the second class begins with 45, we know that the width of all classes is 45 - 40 = 5. All of the classes are presented in column 1 of the grouped-data table in the figure below. The last class to construct does not go beyond 65-69, since the largest single data value is 69. Having established the classes, we tally the ages into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 43, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3.

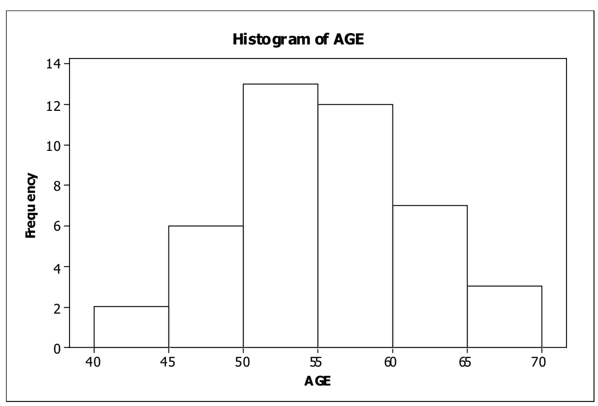
By averaging the lower and upper limits for each class, we arrive at the class mark for each class. The class marks for all classes are presented in column 4.

| Age at Inauguration | Frequency | Relative Frequency | Class Mark |
|---|---|---|---|
| 40-44 | 2 | 0.047 | 42 |
| 45-49 | 6 | 0.140 | 47 |
| 50-54 | 13 | 0.302 | 52 |
| 55-59 | 12 | 0.279 | 57 |
| 60-64 | 7 | 0.163 | 62 |
| 65-69 | 3 | 0.070 | 67 |
| | 43 | 1.001 | |

(b) The lower cutpoint for the first class is 40. The upper cutpoint for the first class is 45 since that is the smallest value that can go into the second class.

(c) The common class width is 45 - 40 = 5.

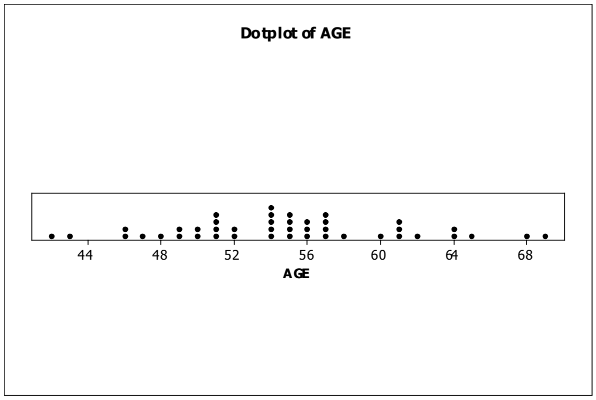
(d) The frequency histogram presented below is constructed using the frequency distribution presented above; i.e., columns 1 and 2. Notice that the lower cutpoints of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical-axis units in the frequency histogram are the even integers within the range 0 through 14, since these are representative of the magnitude and spread of the frequencies presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.




(e) The overall shape of the inauguration ages is somewhere between triangular and bell-shaped.

(f) The distribution is roughly symmetric.

**15.** The horizontal axis of this dotplot displays a range of possible ages for the 43 Presidents of the United States. To complete the dotplot, we go through the data set and record each age by placing a dot over the appropriate value on the horizontal axis.




**16.** (a) Using *one* line per stem in constructing the ordered stem-and-leaf diagram means vertically listing the numbers comprising the stems *once*. The leaves are then placed with their respective stems in order. The ordered stem-and-leaf diagram using one line per stem is presented in Figure (a).

(b) Using *two* lines per stem in constructing the ordered stem-and-leaf diagram means vertically listing the numbers comprising the stems *twice*. In turn, if the leaf is one of the digits 0 through 4, it is ordered and placed with the first of the two stem lines. If the leaf is one of the digits 5 through 9, it is ordered and placed with the second of the two stem lines. The ordered stem-and-leaf diagram using two lines per stem is presented in Figure (b).

(a)

| | |
|---|---|
| 4 | 2 3 6 6 7 8 9 9 |
| 5 | 0 0 1 1 1 1 2 2 4 4 4 4 4 5 5 5 5 6 6 6 7 7 7 7 8 |
| 6 | 0 1 1 1 2 4 4 5 8 9 |

(b)

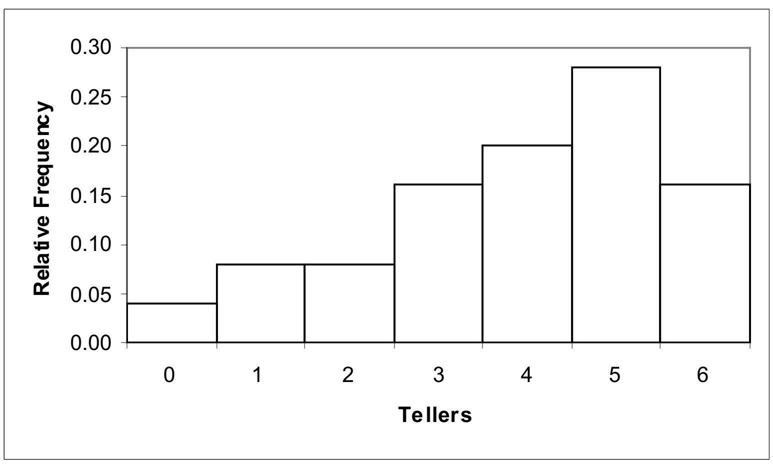
| | |
|---|---|
| 4 | 2 3 |
| 4 | 6 6 7 8 9 9 |
| 5 | 0 0 1 1 1 1 2 2 4 4 4 4 4 |
| 5 | 5 5 5 5 6 6 6 7 7 7 7 8 |
| 6 | 0 1 1 1 2 4 4 |
| 6 | 5 8 9 |

(c) Two lines per stem corresponds to the frequency distribution.

**17.** (a) The grouped-data table presented below is constructed using classes based on a single value. Since each data value is one of the integers 0 through 6, inclusive, the classes will be 0 through 6, inclusive. These are presented in column 1. Having established the classes, we tally the number of busy tellers into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 25, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. Since each class is based on a single value, the class mark is the class itself. Thus, a column of midpoints is not given, since this same information is presented in column 1.

| Number Busy | Frequency | Relative Frequency |
|---|---|---|
| 0 | 1 | 0.04 |
| 1 | 2 | 0.08 |
| 2 | 2 | 0.08 |
| 3 | 4 | 0.16 |
| 4 | 5 | 0.20 |
| 5 | 7 | 0.28 |
| 6 | 4 | 0.16 |
| | 25 | 1.00 |

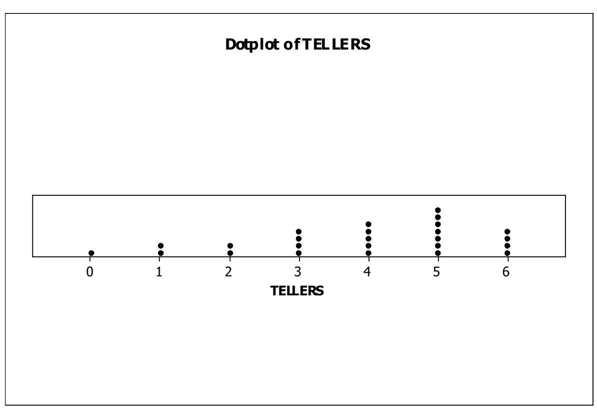
(b) The following relative-frequency histogram is constructed using the relative-frequency distribution presented in part (a); i.e., columns 1 and 3. Column 1 demonstrates that the data are grouped using classes based on a single value. These single values in column 1 are used to label the horizontal axis of the relative-frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.04 to 0.28 (4% to 28%). Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05, starting with zero and ending at 0.30. The middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.




(c) The overall shape of this distribution is left skewed.

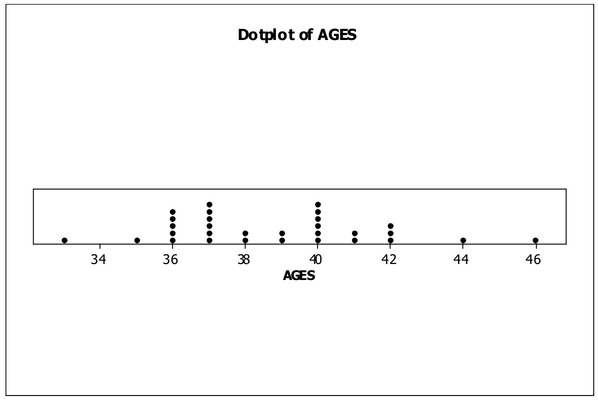
(d) The distribution is left skewed.

(e)



(f) Since both the histogram and the dotplot are based on single value grouping, they both convey exactly the same information.

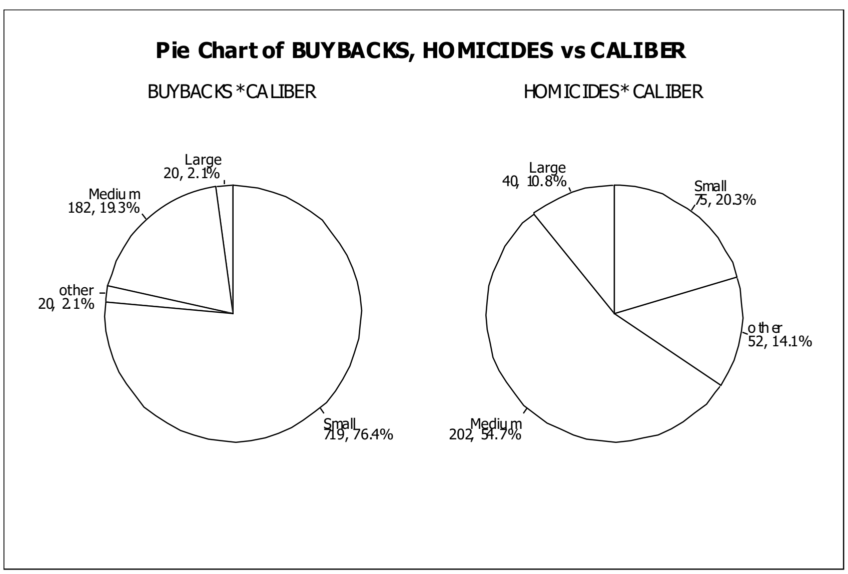
**18.** (a) The dotplot of the ages of the oldest player on each major league baseball team is




(b) The overall shape of the distribution of ages is bimodal.

(c) The distribution is roughly symmetric.

**19.** (a)-(b) The two pie-charts are shown below. In each case, the proportion of the circle for each category is found by multiplying 360 degrees by the category frequency and dividing by the total frequency (941 for Buybacks and 369 for Homicides).

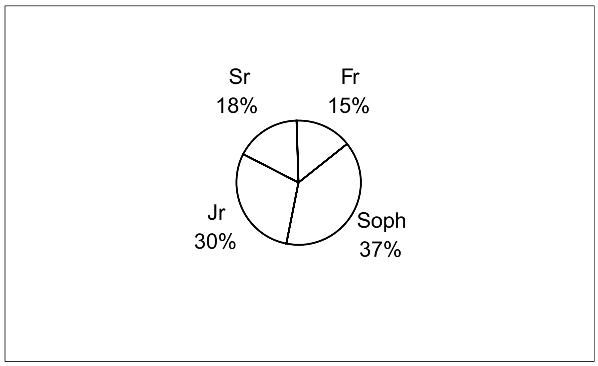



(c) The most striking characteristic of the two charts is that most of the buybacks are of small caliber guns while most of the homicides are committed with medium caliber guns. It should also be noted that small and medium caliber guns are the two largest categories of both buybacks and homicides, accounting for 95.7% of the buybacks and 75.0% of the homicides.

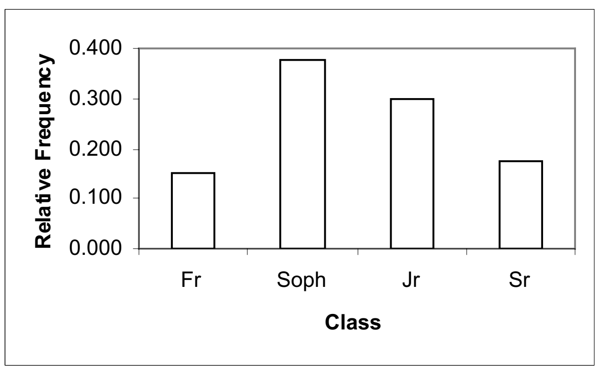
**20.** (a) The table below shows both the frequency distribution and the relative frequency distribution. If each frequency reported in the table is divided by the total number of students, which is 40, we arrive at the relative frequency (or percentage) of each class.

| **Class** | **Frequency** | **Relative Frequency** |
|---|---|---|
| **Fr** | 6 | 0.150 |
| **So** | 15 | 0.375 |
| **Jr** | 12 | 0.300 |
| **Sr** | 7 | 0.175 |

(b) The following pie chart is used to display the percentage of students in each of the four classes.



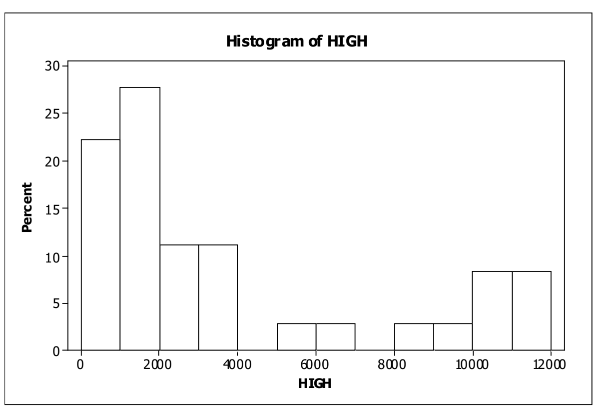

(c) The following bar graph also displays the relative frequencies of each class.



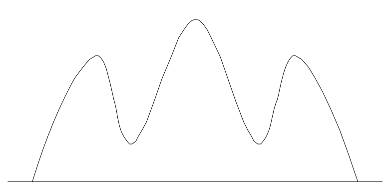

**21.** (a) The first class to construct is 0 < 1000. Since all classes are to be of equal width, we know that the width of all classes is 1000 - 0 = 1000. All of the classes are presented in column 1 of Figure (a) below. The last class to construct is 11000 < 12000, since the largest single data value is 11722.98. Having established the classes, we tally the highs into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 36, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. By averaging the numbers that appear as the lower and upper cutpoints for each class, we arrive at the midpoint for each class. The midpoints for all classes are presented in column 4.

| High | Freq. | Relative Frequency | Midpoint |
|---|---|---|---|
| 0 < 1000 | 8 | 0.222 | 500 |
| 1000 < 2000 | 10 | 0.278 | 1500 |
| 2000 < 3000 | 4 | 0.111 | 2500 |
| 3000 < 4000 | 4 | 0.111 | 3500 |
| 4000 < 5000 | 0 | 0.000 | 4500 |
| 5000 < 6000 | 1 | 0.028 | 5500 |
| 6000 < 7000 | 1 | 0.028 | 6500 |
| 7000 < 8000 | 0 | 0.000 | 7500 |
| 8000 < 9000 | 1 | 0.028 | 8500 |
| 9000 < 10000 | 1 | 0.028 | 9500 |
| 10000 < 11000 | 3 | 0.083 | 10500 |
| 11000 < 12000 | 3 | 0.083 | 11500 |
| | 36 | 1.000 | |

(b) The following relative-frequency histogram is constructed using the relative-frequency distribution presented above; i.e., columns 1 and 3. The lower cutpoints of column 1 are at the left edges of each rectangle in the relative-frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.000 to 0.278. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 5%, from 0,00 (0%) to 0.30 (30%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.




**22.** Answers will vary, but here is one possibility:



**23.** (a) The break in the third bar is to emphasize that the bar as shown is not as tall as it should be.

(b) The bar for the space available in coal mines is at a height of about 30 billion tonnes. To accurately represent the space available in saline aquifers (10,000 billion tonnes), the third bar would have to be over 300 times as high as the first bar and more than 10 times as high as the second bar. If the first two bars were kept at the sizes shown, there wouldn't be enough room on the page for the third bar to be shown at its correct height. If a reasonable height were chosen for the third bar, the first bar wouldn't be visible. The only apparent solution is to present the third bar as a broken bar.

**24.** (a) Covering up the numbers on the vertical axis totally obscures the percentages.

(b) Having followed the directions in part (a), we might conclude that the percentage of women in the labor force for 2000 is about three and one-half times that for 1960.

(c) Not covering up the vertical axis, we find that the percentage of women in the labor force for 2000 is about 1.8 times that for 1960.

(d) The graph is potentially misleading because it is truncated. Notice that vertical axis units begin at 30 rather than at zero.

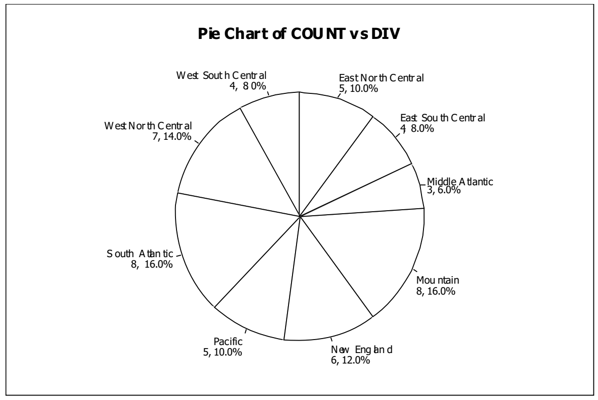
(e) To make the graph less potentially misleading, we can start it at zero instead of 30.

**25.** (a) The population consists of all of the states of the U.S. The variable under consideration is the division to which the U.S. Census Bureau assigns each state.

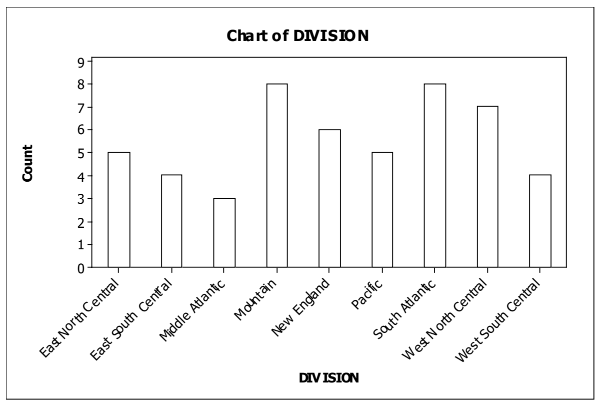
(b) After entering the data in Minitab from the WeissStats CD, choose **Stat ▶ Tables ▶ Tally Individual Variables.** Double-click on DIVISION to enter DIVISION in the **Variables** box. Put check marks by **Counts** and **Percents** under **Display** and click **OK.** The result is shown in the Session Window as below:

| DIVISION | Count | Percent |
|---:|---:|---:|
| East North Central | 5 | 10.00 |
| East South Central | 4 | 8.00 |
| Middle Atlantic | 3 | 6.00 |
| Mountain | 8 | 16.00 |
| New England | 6 | 12.00 |
| Pacific | 5 | 10.00 |
| South Atlantic | 8 | 16.00 |
| West North Central | 7 | 14.00 |
| West South Central | 4 | 8.00 |
| N= | 50 | |

(c) To obtain a pie chart, first enter the nine divisions in another column of Minitab, say C3, and enter the counts from the table in part (a) in the next column C4. Name the columns C3 and C4 DIV and COUNT respectively. Choose **Graph ▶ Pie Chart**, enter DIV in the **Categorical Variable** box and COUNT in the S**ummary variables** box. Click on the **LABELS** button and then on the **Slice Labels** tab. Put check marks in all four boxes and click **OK** twice. The result is (except for colors and legend)



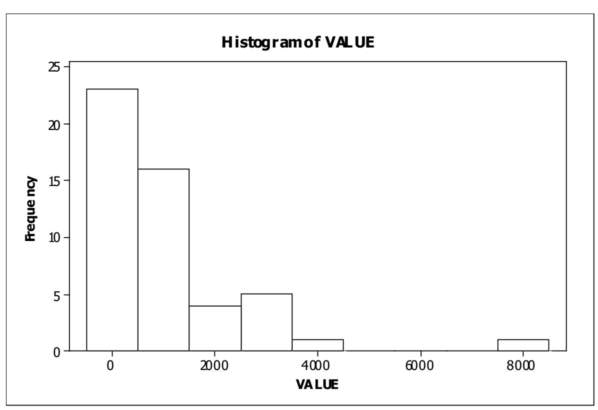

(d) Still in Minitab, choose **Graph ▶ Bar Chart**, make sure that Bars represents counts of unique values, click on **Simple** and on **OK.** Now enter DIVISION in the **Graph variables** box and click **OK.** The result is



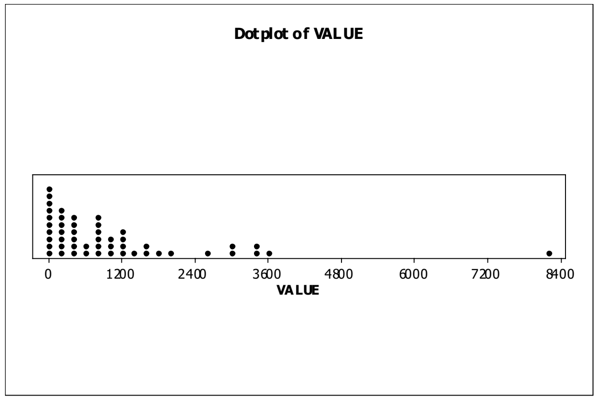

(e) Both charts indicate that Mountain and South Atlantic divisions have the most states (8) while the Middle Atlantic division has the fewest (3).

**26.** (a) The population consists of the states of the U.S. and the variable under consideration is the value of the exports of each state.

(b) Using Minitab, we enter the data from the WeissStats CD, choose **Graph ▶ Histogram**, click on **Simple** and click **OK.** Then double click on VALUE to enter it in the **Graph variables** box and click **OK.** The result is



(c) For the dotplot, we choose **Graph ▶ Dotplot**, click on **Simple** from the **One Y** row and click **OK.** Then double click on VALUE to enter it in the **Graph variables** box and click **OK.** The result is




(d) For the stem-and-leaf plot, we choose **Graph ▶ Stem-and-Leaf**, double click on VALUE to enter it in the **Graph variables** box and click **OK.** The result is

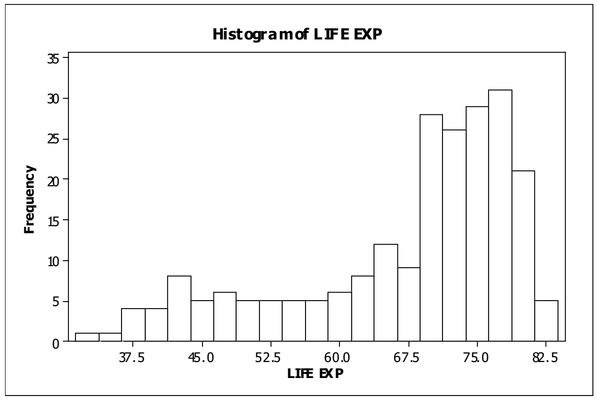
```
Stem-and-leaf of VALUE  N  = 50
Leaf Unit = 100


 23   0  00000000001112222344444
(10)  0  5677888899
 17   1  012224
 11   1  5679
 7    2
 7    2  69
 5    3  034
 2    3  6
 1    4
 1    4
 1    5
 1    5
 1    6
 1    6
 1    7
 1    7
 1    8  2
```

(e) The overall shape of the distribution is reverse J shaped.

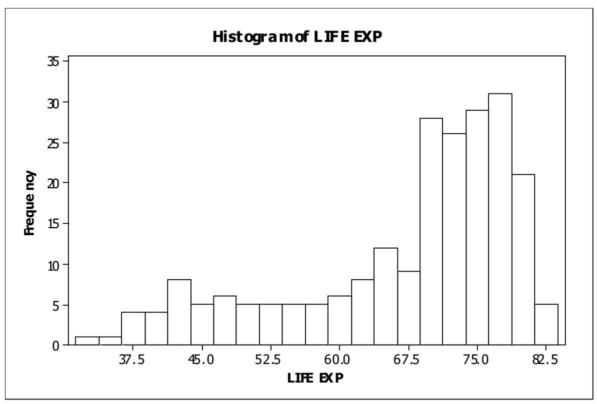
(f) The distribution is right skewed.

**27.** (a) The population consists of countries of the world, and the variable under consideration is the expected life in years for people in those countries.

(b) Using Minitab, we enter the data from the WeissStats CD, choose **Graph ▶ Histogram**, click on **Simple** and click **OK.** Then double click on LIFE EXP to enter it in the **Graph variables** box and click **OK.** The result is




(c) For the dotplot, we choose **Graph ▶ Dotplot**, click on **Simple** from the **One Y** row and click **OK.** Then double click on LIFE EXP to enter it in the **Graph variables** box and click **OK.** The result is




(d) For the stem-and-leaf plot, we choose **Graph ▶ Stem-and-Leaf**, double click on LIFE EXP to enter it in the **Graph variables** box and click **OK.** The result is

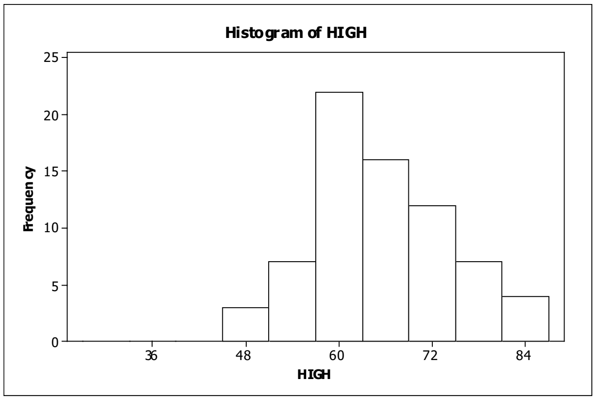
```
Stem-and-leaf of LIFE EXP  N  = 224
Leaf Unit = 1.0


 1     3  2
 9     3  56788999
 21    4  112222233444
 31    4  5567788889
 42    5  01111333444
 50    5  56667779
 70    6  00111112233333444444
 99    6  55556666777788888899999999999
(52)   7  0000000001111111111111122222222223333333333344444444
 73    7  55555555555555666666666667777777777788888888888888888999999999
 11    8  00000011113
```

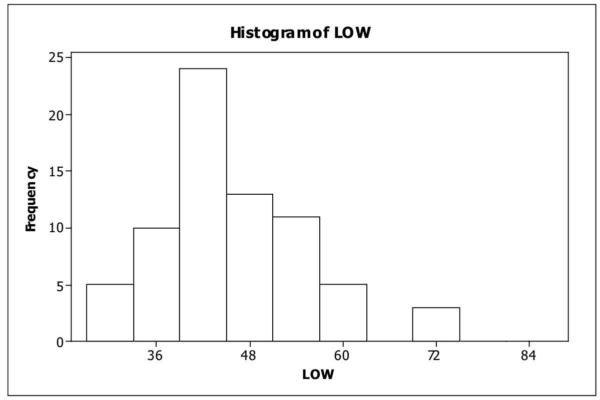
(e) The overall shape of the distribution is left skewed.

(f) This distribution is classified as left skewed.

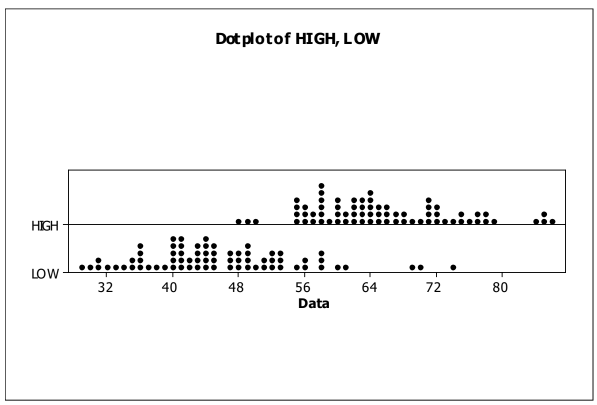
**28.** (a) The population consists of cities in the U.S., and the variables under consideration are their annual average maximum and minimum temperatures.

(b) Using Minitab, we enter the data from the WeissStats CD, choose **Graph ▶ Histogram**, click on **Simple** and click **OK.** Double click on HIGH to enter it in the **Graph variables** box, and double click on LOW to enter it in the **Graph variables** box. Now click on the **Multiple graphs** button and click to **Show Graph Variables on separate graphs** and also check both boxes under **Same Scales for Graphs**, and click **OK** twice. The result is




(c) For the dotplot, we choose **Graph ▶ Dotplot**, click on **Simple** from the **Multiple Y's** row and click **OK.** Then double click on HIGH and then LOW to enter them in the **Graph variables** box and click **OK.** The result is




(d) For the stem-and-leaf diagram, we choose **Graph ▶ Stem-and-Leaf**, double click on HIGH and then on LOW to enter then in the **Graph variables** box and click **OK.** The result is

```
Stem-and-leaf of HIGH  N  = 71
Leaf Unit = 1.0

 3    4  789
 6    5  444
 21   5  555677777888999
(17)  6  00001122222333444
 33   6  55556677779
 22   7  00001122234
 11   7  5577889
 4    8  444
 1    8  5
```

```
Stem-and-leaf of LOW  N  = 71
Leaf Unit = 1.0


 1    2  9
 7    3  001234
 19   3  555556779999
(20)  4  00011111223333344444
 32   4  5567777888899
 19   5  11122223
 11   5  5667889
 4    6  1
 3    6  9
 2    7  04
```

(e) Both variables have distributions that are slightly right skewed.
(f) Both distributions are close to symmetric, but are slightly right skewed. It would take only a few changes in the data to make the distributions approximately symmetric.

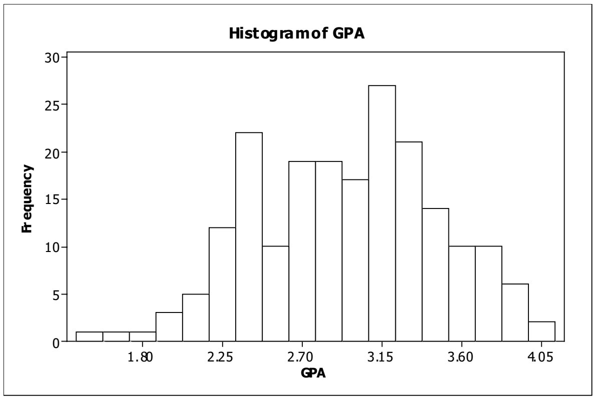
## Using the FOCUS Database: Chapter 2

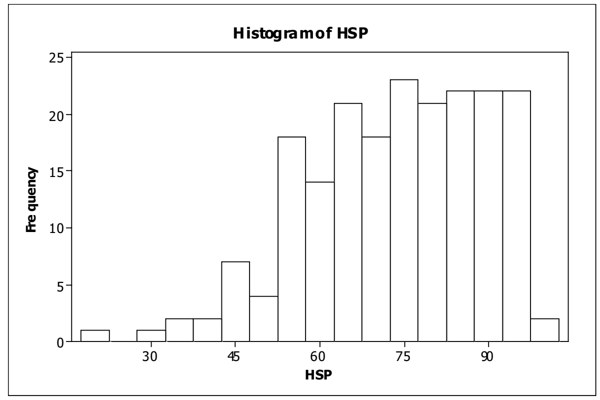
We use the *Menu commands* in Minitab to complete parts (a)-(e). The data sets in the Focus database and their names have already been stored in the file FOCUS.MTW on the WeissStats CD supplied with the text, and, assuming that that disk is in Drive D, all information can be recovered if you

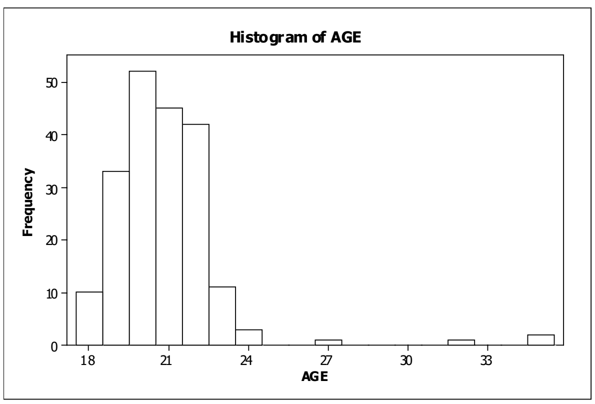
- Choose **File ▶ Open Worksheet...** and **Look in: d:\FocusSample** or **d:\Focus** (depending on which part of this exercise you are working on) and Click **OK**

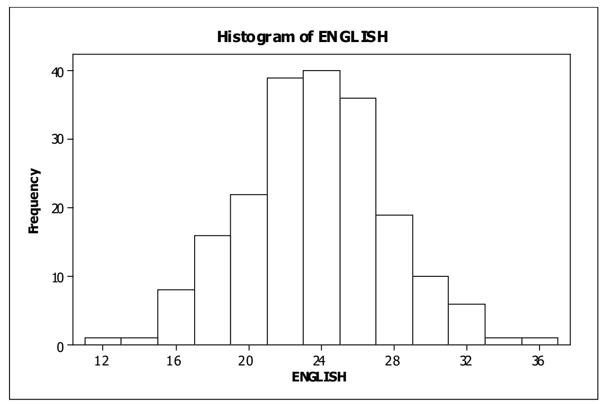
(a) UWEC is a school that attracts good students. HSP reflects pre-college experience and will tend to be left-skewed since fewer students with lower high school percentile scores will have been admitted, but exceptions are made for older students whose high school experience is no longer relevant. GPA will probably show left skewness tendencies since many, but not all, students with lower cumulative GPAs (below 2.0 on a 4-point scale) will likely have been suspended and will not appear in the database, but there are also upper limits on these scores, so the scores will tend to bunch up nearer to the high end than to the low end. AGE will be right skewed because there are few students below the typical 17-22 ages, but many above that range. ENGLISH, MATH, and COMP will be closer to bell-shaped. The ACT typically is taken only by high school students intending to go to college, and the scores are designed to roughly follow a bell-shaped curve. Individual colleges may, however, have a different profile that reflects their admission policies.

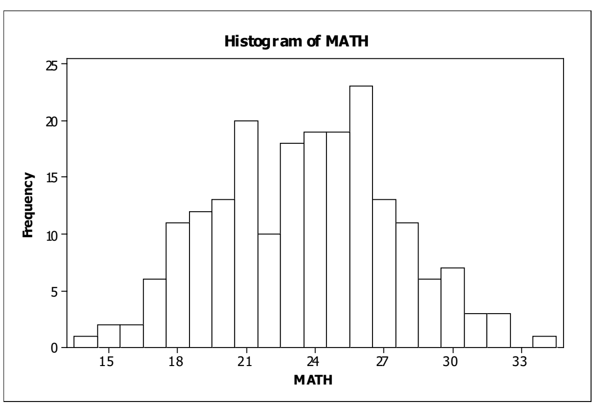
(b) Using Minitab with FocusSample, choose **Graph ▶ Histogram...**, select the **Simple** version, and Click **OK.** Then specify HSP GPA AGE ENGLISH MATH COMP in the **Graph variables** text box and click on the button for **Multiple Graphs.** Click on the button for **On separate graphs** and click **OK** twice. The results are

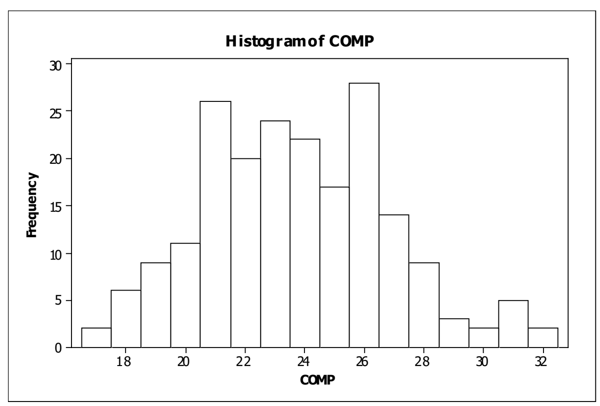


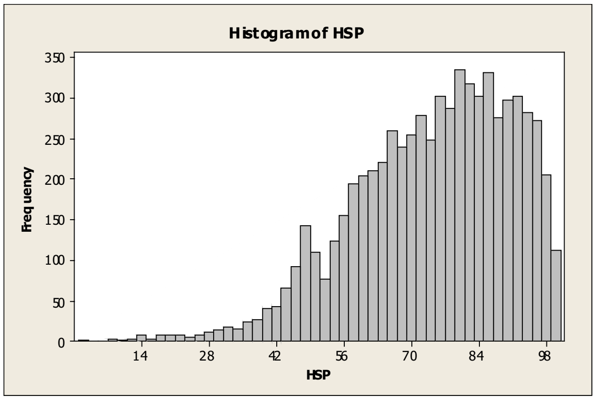



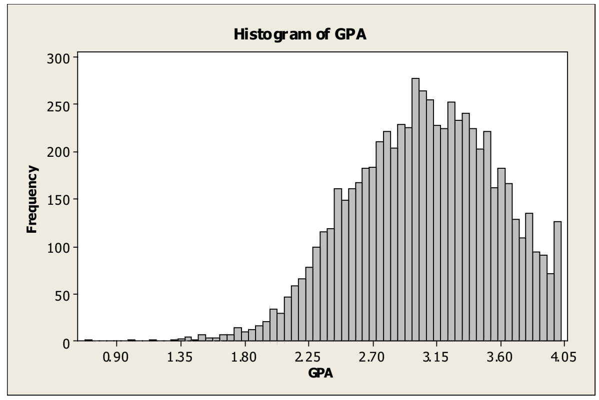



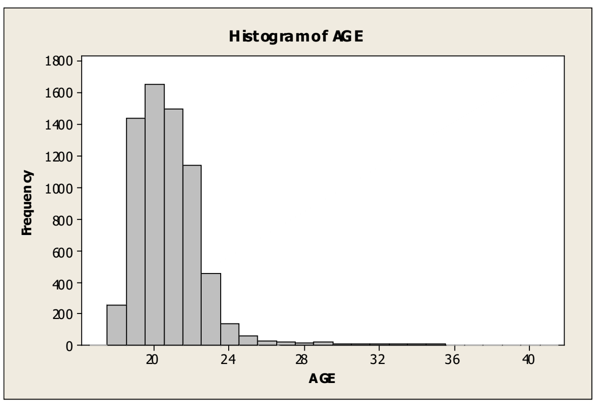



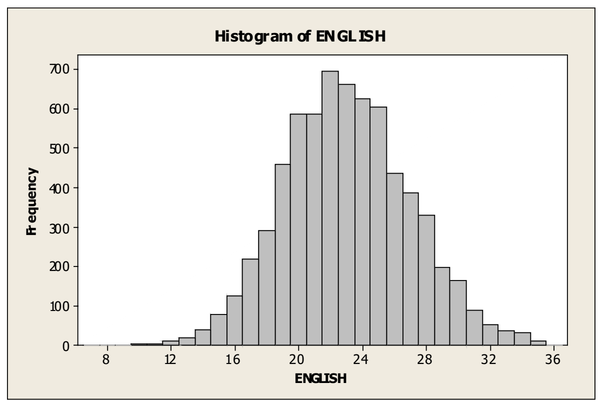

The graphs compare quite well with the educated guesses for all six variables.

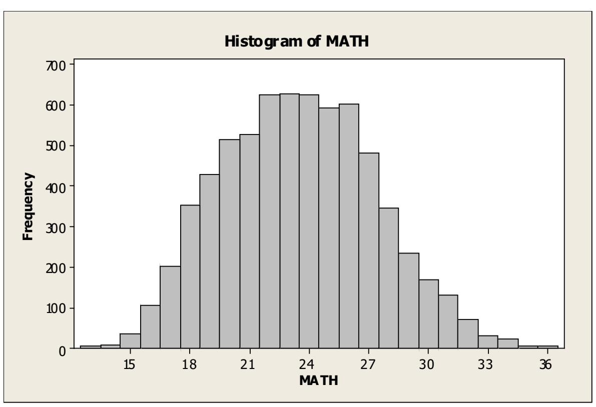
(c) Using Minitab with Focus, choose **Graph ► Histogram...**, select the **Simple** version, and Click **OK.** Then specify HSP GPA AGE ENGLISH MATH COMP in the **Graph variables** text box and click on the button for **Multiple Graphs.** Click on the button for **On separate graphs** and click **OK** twice. The results are

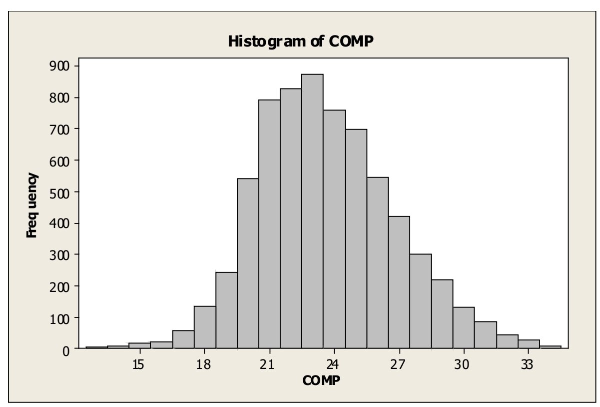


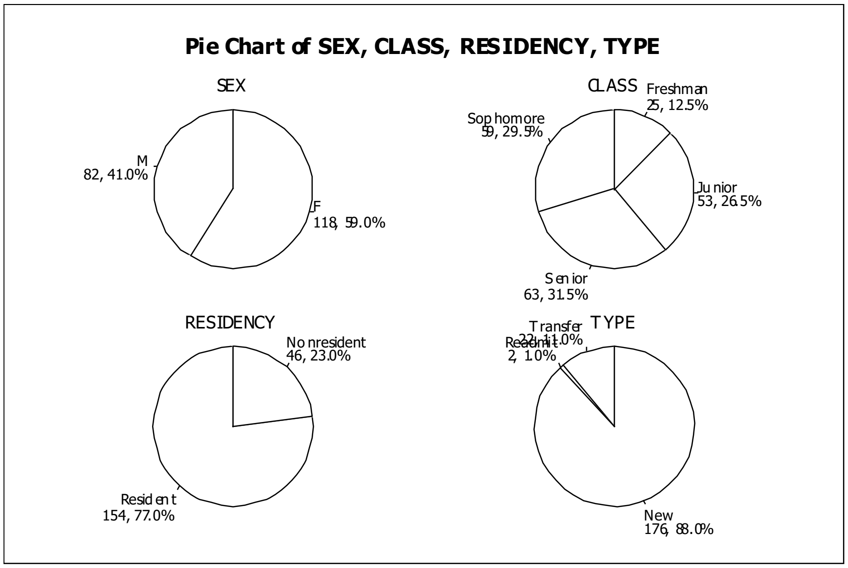




We were correct on the first five variables: HSP and GPA are left skewed, AGE is right skewed, ENGLISH and MATH are fairly symmetric. COMP is close to symmetric, but is slightly right skewed. Comparing the graphs for the sample with those for the entire population, we see similarities between each pair of graphs, but the outline of the histogram for the entire population is much smoother than that of the histogram for the sample.

(d) Using Minitab and the FocusSample file, choose **Graph ▶ Piechart**, click on the **Chart raw data** button, specify SEX CLASS RESIDENCY TYPE in the **Graph variables** text box and click on the **Labels** button. Now click on the tab for **Slice Labels**, check all four boxes and click **OK**, click on the button for

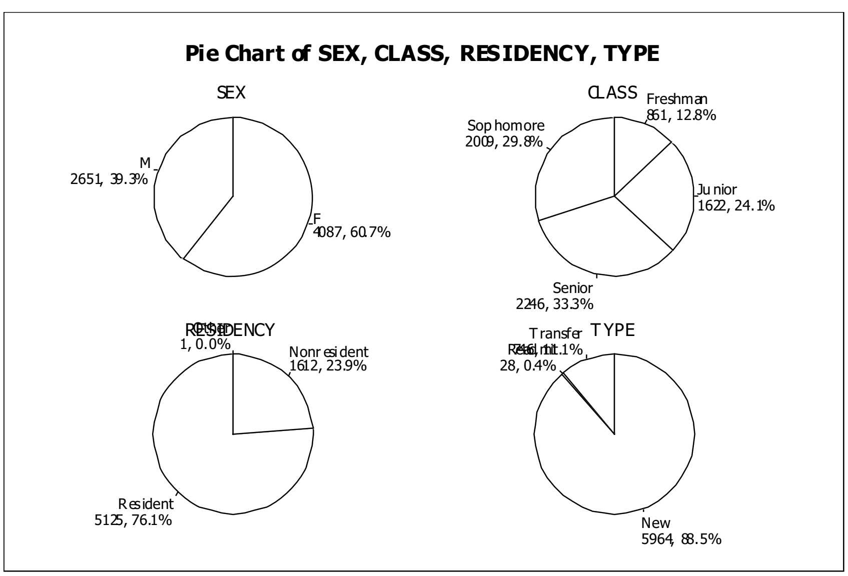
**Multiple Graphs** and ensure that the button for **On the same graph** is checked, and click **OK** twice.

Once the graphs are displayed, we right clicked on the legend that was shown and selected **Delete** since we already had provided for each slice of the graphs to be labeled.




From the graph of SEX, we see that about 59% of the students are females. From the graph of CLASS, we see that the student sample is about 12.5% Freshmen, 29.5% Sophomores, 26.5% Juniors, and 31.5% Seniors. From the graph of RESIDENCY, we see that about 77% of the students are Wisconsin residents and 23$ are nonresidents. From the graph of TYPE, we see that 88.0% of the students were admitted initially as new students, 11.0% were admitted initially as transfer students, and 1.0% are readmits, that is, students who were initially new or transfer students, left the university, and were later readmitted.

(e) Now repeat part (d) using the entire Focus file. The results are



From the graph of SEX, we see that about 61% of the students are females. From the graph of CLASS, we see that the student population is about 13% Freshmen, 30% Sophomores, 24% Juniors, and 33% Seniors. From the graph of RESIDENCY, we see that about 76% of the students are Wisconsin residents and 24$ are nonresidents. From the graph of TYPE, we see that 85.5% of the students were admitted initially as new students, 11.1% were admitted initially as transfer students, and 0.4% are readmits, that is, students who were initially new or transfer students, left the university, and were later readmitted. We would expect that the two sets of graphs would be approximately the same, but not identical since the sample contains only 200 students out of a population of 6738. This is, in fact, the case. The percentages in each sample graph are very close to the percentages in the corresponding population graph.

**Case Study: Preventing Infant Mortality**

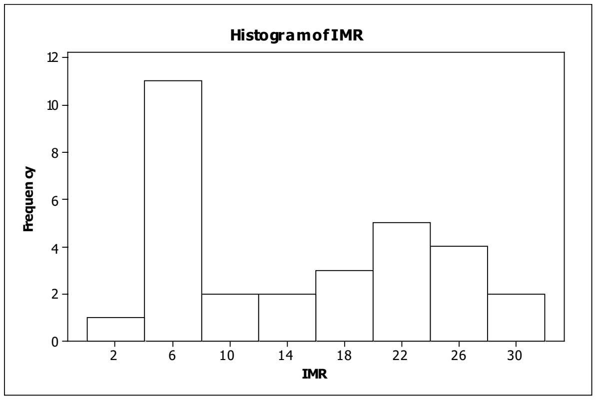
(a) The second column provides nonnumerical information. Thus, it consists of *qualitative* data.

(b) Stating that the U.S. ranks 11th in infant mortality among the developed nations provides information concerning rank. Thus, the type of data provided by this statement is *quantitative, discrete*.

(c) The first class to construct is "4-under 6." Since the upper class limit "under 5" does not include the number 6, the number 6 becomes the lower class limit of the next class. Since all classes are to be of equal width, and the second class begins with 6, we know that the width of all classes is 6-4 = 2. All of these classes are presented in column 1. The last class to construct is "18-under 20," since the largest data value is 18.4. Having established the classes, we tally the infant mortality data into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 24, results in the relative frequencies for each class which are presented in column 3. By averaging the lower and upper cutpoints for each class, we arrive at the class midpoint for each class. The midpoints for all classes are presented in column 4.

| **Infant Mortality (deaths per 1000)** | **Frequency** | **Relative Frequency** | **Midpoints** |
|---|---|---|---|
| 0 ⋖ 4 | 1 | 0.033 | 2 |
| 4 ⋖ 8 | 11 | 0.367 | 6 |
| 8 ⋖ 12 | 2 | 0.067 | 10 |
| 12 ⋖ 16 | 2 | 0.067 | 14 |
| 16 ⋖ 20 | 3 | 0.100 | 18 |
| 20 ⋖ 24 | 5 | 0.167 | 22 |
| 24 ⋖ 28 | 4 | 0.133 | 26 |
| 28 ⋖ 32 | 2 | 0.067 | 30 |
| | 30 | 1.001 | |

(d) The frequency histogram for infant mortality rates is constructed using the frequency distribution presented in part (c); i.e., columns 1 and 2. The lower class cutpoints of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical axis units in the frequency histogram are the integers 0 through 11, since these are representative of the magnitude and spread of the frequencies presented in column 2. The height of each bar matches the respective frequency in column 2.




(e) We will drop the tenths digit and use five lines per stem for the remaining whole numbers. Using *five* lines per stem in constructing the stem-and-leaf diagram means vertically listing the numbers comprising the stems *five times*. This is presented as follows:

```
0  3
0  4444455
0  6667
0  99
1
1
1  45
1  67
1  9
2  555
2  8
3  0
```

(f) The dotplot for the IMRs is