CHAPTER 2 ANSWERS

Exercises 2.1

2.1 (a) Answers may vary. Eye color and model of car are qualitative variables.

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

(c) Weight and voltage are examples of quantitative continuous variables.

2.3 (a) Qualitative data are obtained by observing the characteristics described by a qualitative variable such as color or shape.

(b) Discrete, quantitative data are numerical data that are obtained by observing values, usually by counting, of a discrete variable whose values form a finite or countably infinite set of numbers.

(c) Continuous, quantitative data are numerical data that are obtained by observing values of a continuous variable. They are usually the result of measuring something such as temperature which can take any value in a given interval.

2.5 Of qualitative and quantitative (discrete and continuous) types of data, only qualitative involves non-numerical data.

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

(b) The third column consists of quantitative, continuous data. This column provides the highest temperature on record in each of the listed cities.

(c) The information that Phoenix is in Arizona is qualitative data since it is non-numeric.

2.9 (a) The third column consists of quantitative, discrete data. Although the data are presented to one decimal point, the data represent the number of albums sold in millions, which can only be whole numbers.

(b) The information that Supernatural was performed by Santana is qualitative data since it is non-numerical.

Exercises 2.2

2.11 One of the main reasons for grouping data is that it often makes a rather complicated set of data easier to understand.

2.13 When grouping data, the three most important guidelines in choosing the classes 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 piece of data must belong to one, and only one, class; and (3) whenever feasible, all classes should have the same width.

2.15 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 data values, relative-frequency distributions should be used because the total of each set of relative frequencies is 1, putting both distributions on the same basis.

2.17 In the first method for depicting classes we used the notation \( a \leq b \) to mean values that are greater than or equal to \( a \) and up to, but not including \( b \). So, for example, \( 30 \leq 40 \) represents 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 endpoints. 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 preferable 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.

2.19 When grouping data using classes that each represent a single possible numerical value, the midpoint of any given class would be the same as the value for that class. Thus listing the midpoints would be redundant.

2.21 The first class is \(52<54\). Since all classes are to be of equal width 2, the classes are presented in column 1. The last class is \(74<76\), since the largest 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 are presented in column 3. By averaging the lower and upper class cutpoints for each class, we arrive at the class midpoints which are presented in column 4.

<table>
<thead>
<tr>
<th>Speed (MPH)</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Midpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>52&lt;54</td>
<td>2</td>
<td>0.057</td>
<td>53</td>
</tr>
<tr>
<td>54&lt;56</td>
<td>5</td>
<td>0.143</td>
<td>55</td>
</tr>
<tr>
<td>56&lt;58</td>
<td>6</td>
<td>0.171</td>
<td>57</td>
</tr>
<tr>
<td>58&lt;60</td>
<td>8</td>
<td>0.229</td>
<td>59</td>
</tr>
<tr>
<td>60&lt;62</td>
<td>7</td>
<td>0.200</td>
<td>61</td>
</tr>
<tr>
<td>62&lt;64</td>
<td>3</td>
<td>0.086</td>
<td>63</td>
</tr>
<tr>
<td>64&lt;66</td>
<td>2</td>
<td>0.057</td>
<td>65</td>
</tr>
<tr>
<td>66&lt;68</td>
<td>1</td>
<td>0.029</td>
<td>67</td>
</tr>
<tr>
<td>68&lt;70</td>
<td>0</td>
<td>0.000</td>
<td>69</td>
</tr>
<tr>
<td>70&lt;72</td>
<td>0</td>
<td>0.000</td>
<td>71</td>
</tr>
<tr>
<td>72&lt;74</td>
<td>0</td>
<td>0.000</td>
<td>73</td>
</tr>
<tr>
<td>74&lt;76</td>
<td>1</td>
<td>0.029</td>
<td>75</td>
</tr>
</tbody>
</table>

Note that the relative frequencies sum to 1.01, not 1.00, due to round-off errors in the individual relative frequencies.

2.23 The first class is \(52-53.9\). Since all classes are to be of equal width, the second class has limits of 54 and 55.9. The classes are presented in column 1. The last class is \(74-75.9\) since the largest 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 are presented in column 4.
Note that the relative frequencies sum to 1.01, not 1.00, due to round-off errors in the individual relative frequencies.

Since the data values range from 3 to 12, we could construct a table with classes based on a single value or on two values. We will choose classes with a single value because one of the classes based on two values would have contained almost half of the data. The resulting table is shown below.

2.27 (a) The first class is 1<2. Since all classes are to be of equal width 1, the second class is 2<3. 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 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 are presented in column 4.
### SECTION 2.2, GROUPING DATA

<table>
<thead>
<tr>
<th>Volume (100sh)</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Midpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;-2</td>
<td>4</td>
<td>0.13</td>
<td>1.5</td>
</tr>
<tr>
<td>2&lt;-3</td>
<td>4</td>
<td>0.13</td>
<td>2.5</td>
</tr>
<tr>
<td>3&lt;-4</td>
<td>2</td>
<td>0.07</td>
<td>3.5</td>
</tr>
<tr>
<td>4&lt;-5</td>
<td>6</td>
<td>0.20</td>
<td>4.5</td>
</tr>
<tr>
<td>5&lt;-6</td>
<td>3</td>
<td>0.10</td>
<td>5.5</td>
</tr>
<tr>
<td>6&lt;-7</td>
<td>1</td>
<td>0.03</td>
<td>6.5</td>
</tr>
<tr>
<td>7&lt;-8</td>
<td>2</td>
<td>0.07</td>
<td>7.5</td>
</tr>
<tr>
<td>8&lt;-9</td>
<td>3</td>
<td>0.10</td>
<td>8.5</td>
</tr>
<tr>
<td>9&lt;-10</td>
<td>1</td>
<td>0.03</td>
<td>9.5</td>
</tr>
<tr>
<td>10 &amp; Over</td>
<td>4</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>

Note that the relative frequencies sum to 0.99, not 1.00, due to round-off errors in the individual relative frequencies.

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

### 2.31

(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 are presented in column 3. By averaging the lower and upper cutpoints for each class, we arrive at the class mark for each class. The class marks are presented in column 4.

<table>
<thead>
<tr>
<th>Score</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Class Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-39</td>
<td>2</td>
<td>0.10</td>
<td>34.5</td>
</tr>
<tr>
<td>40-49</td>
<td>0</td>
<td>0.00</td>
<td>44.5</td>
</tr>
<tr>
<td>50-59</td>
<td>0</td>
<td>0.00</td>
<td>54.5</td>
</tr>
<tr>
<td>60-69</td>
<td>3</td>
<td>0.15</td>
<td>64.5</td>
</tr>
<tr>
<td>70-79</td>
<td>3</td>
<td>0.15</td>
<td>74.5</td>
</tr>
<tr>
<td>80-89</td>
<td>8</td>
<td>0.40</td>
<td>84.5</td>
</tr>
<tr>
<td>90-100</td>
<td>4</td>
<td>0.20</td>
<td>95.0</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

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

### 2.33

In Minitab, place the cheetah speed data in a column named SPEED and put the WeissStats CD in the CD drive. Assuming that the CD drive is drive D, then type in Minitab’s session window after the MTB> prompt the command

```plaintext
%D:\IS6\Minitab\Macro\group.mac 'SPEED'
```

and press the ENTER key. We are given three options for specifying the classes. Since we want the first class to have lower cutpoint 52 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 52 2 when prompted to enter the cutpoint and class width of the first class.

Press the ENTER key again. The resulting output is
## CHAPTER 2, DESCRIPTIVE STATISTICS

### Grouped-data table for SPEED

<table>
<thead>
<tr>
<th>Row</th>
<th>LowerCut</th>
<th>UpperCut</th>
<th>Freq</th>
<th>RelFreq</th>
<th>Midpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52</td>
<td>54</td>
<td>2</td>
<td>0.057</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>54</td>
<td>56</td>
<td>5</td>
<td>0.143</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>58</td>
<td>6</td>
<td>0.171</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
<td>60</td>
<td>8</td>
<td>0.229</td>
<td>59</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>62</td>
<td>7</td>
<td>0.200</td>
<td>61</td>
</tr>
<tr>
<td>6</td>
<td>62</td>
<td>64</td>
<td>3</td>
<td>0.086</td>
<td>63</td>
</tr>
<tr>
<td>7</td>
<td>64</td>
<td>66</td>
<td>2</td>
<td>0.057</td>
<td>65</td>
</tr>
<tr>
<td>8</td>
<td>66</td>
<td>68</td>
<td>1</td>
<td>0.029</td>
<td>67</td>
</tr>
<tr>
<td>9</td>
<td>68</td>
<td>70</td>
<td>0</td>
<td>0.000</td>
<td>69</td>
</tr>
<tr>
<td>10</td>
<td>70</td>
<td>72</td>
<td>0</td>
<td>0.000</td>
<td>71</td>
</tr>
<tr>
<td>11</td>
<td>72</td>
<td>74</td>
<td>0</td>
<td>0.000</td>
<td>73</td>
</tr>
<tr>
<td>12</td>
<td>74</td>
<td>76</td>
<td>1</td>
<td>0.029</td>
<td>75</td>
</tr>
</tbody>
</table>

### Exercises 2.3

2.35 In Minitab, with the data from the Network column in a column named NETWORK,

- Choose Stat >> Tables >> Tally...
- Click in the Variables text box and select NETWORK
- Click in the Counts and Percents boxes under Display
- Click OK. The results are

<table>
<thead>
<tr>
<th>NETWORK</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>5</td>
<td>25.00</td>
</tr>
<tr>
<td>CBS</td>
<td>8</td>
<td>40.00</td>
</tr>
<tr>
<td>NBC</td>
<td>7</td>
<td>35.00</td>
</tr>
<tr>
<td>N=</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

### Exercises 2.37

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.

### Exercises 2.39

Since a bar graph is used for qualitative data, we separate the bars from each other to emphasize that there is no numerical scale and no special ordering of the classes; if the bars were to touch, some viewers might infer an ordering and common values for adjacent bars.

### Exercises 2.41

(a) Each rectangle in the frequency histogram would have a height equal to the number of dots in the dot diagram.

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

### Exercises 2.43

(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 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%), starting with zero and ending at 0.25 (or 25%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.

(a) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in this exercise; 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 20, 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 0.013 to 0.213. Thus, suitable candidates for vertical-axis units in the relative-frequency histogram are increments of 0.05 (5%), starting with zero and ending at 0.25 (25%). 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.
2.47 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.

```
+---------+---------+---------+---------+
| 0.0     | 5.0     | 10.0    | 15.0    | 20.0
|---------+---------+---------+---------+
```

2.49 (a) The pie chart in Figure (a) is used to display the relative-frequency distribution given in columns 1 and 3 of this exercise. The pieces of the pie chart are proportional to the relative frequencies.

(b) The bar graph in Figure (b) displays the same information about the relative frequencies. The height of each bar matches the respective relative frequency.

2.51 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., 35% x 20).

2.53 (a) Consider all three columns of the energy-consumption data given in Exercise 2.42. 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 corresponding cumulative frequency by the total number of pieces of data.)

<table>
<thead>
<tr>
<th>Less than</th>
<th>Cumulative Frequency</th>
<th>Cumulative Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>60</td>
<td>8</td>
<td>0.16</td>
</tr>
<tr>
<td>70</td>
<td>15</td>
<td>0.30</td>
</tr>
<tr>
<td>80</td>
<td>18</td>
<td>0.36</td>
</tr>
<tr>
<td>90</td>
<td>24</td>
<td>0.48</td>
</tr>
<tr>
<td>100</td>
<td>34</td>
<td>0.68</td>
</tr>
<tr>
<td>110</td>
<td>39</td>
<td>0.78</td>
</tr>
<tr>
<td>120</td>
<td>43</td>
<td>0.86</td>
</tr>
<tr>
<td>130</td>
<td>45</td>
<td>0.90</td>
</tr>
<tr>
<td>140</td>
<td>48</td>
<td>0.96</td>
</tr>
<tr>
<td>150</td>
<td>48</td>
<td>0.96</td>
</tr>
<tr>
<td>160</td>
<td>50</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(b) Pair each class limit in the reworked column 1 with its corresponding cumulative relative frequency found in the 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 corresponding cumulative relative frequency. Then join the points with connecting lines. The result, presented in Figure (b), is an ogive based on cumulative relative frequencies. (Note: A similar procedure could be followed using cumulative frequencies.)

2.55 In Minitab, with the raw, ungrouped data from Exercise 2.25 in a column named PUPS,

- Choose **Graph ▶ Histogram...** from the pull-down menu
- Select PUPS for **Graph1** for the **X Variable**
- Click on the **Options** button and select **Frequency** for the **Type of histogram**.
- Select **Midpoint** for the **Type of Intervals**
CHAPTER 2, DESCRIPTIVE STATISTICS

- Click on the Midpoint/Cutpoint positions button and type 3:12/7 in the Midpoint/Cutpoint positions text box
- Click OK
- Click OK

Then repeat the above process selecting Percents instead of Frequency for the Type of Histogram. The resulting histograms follow.

![Histograms](image)

frequencies in a second column labeled FREQ. Then

- Choose Graph ▶ Pie Chart
- Click on Chart Table
- Specify NETWORK in the Categories in: text box
- Specify FREQ in the Frequencies in: text box.
- Type a title for your graph in the Title text box.
- Click OK  The computer output is

All-Time Top TV Programs by Rating

![Pie Chart](image)

Now

- Choose Graph ▶ Chart
- Click on the Function down-arrow and select Sum
- Specify FREQ in the Y column of Graph1 text box
Specify NETWORK in the X column of Graph1 text box.

Click on the Annotations down-arrow, select Title, and then enter your title in the first text box.

Click OK. The computer output is shown below.

Exercises 2.4

2.59 For data sets with many values, a frequency histogram is more suitable for displaying the data since the vertical axis can be scaled appropriately for any number of data values. With very large sets of data, the stem-and-leaf plot would likely be much too large for display purposes unless the font size were made very small, which would render the plot nearly useless.

2.61 Depending on how ‘compact’ the data is, each of the original stems can be divided into either 2 or 5 stems to increase the number of stems and make the diagram more useful.

2.63 (a) Construction of a stem-and-leaf diagram for the heart rate data begins with a vertical listing of the numbers comprising the stems. These numbers are 5, 6, 7, and 8. To the right of this listing is a vertical line which serves as a demarcation between the stems and leaves that are about to be added. Each leaf will be the right-most digit -- the units digit -- of a number presented in the data set. The completed stem-and-leaf diagram is presented below.

```
5  2734459
6  730308466384
7  477637113
8  042
```

(b) The ordered stem-and-leaf is created from the diagram in part (a) by ordering the leaves in each stem numerically as shown below.

```
5  2344579
6  003334466788
7  113346777
8  024
```

(c) With two lines per stem, the leaf digits 0-4 are placed in the first of the two lines and the leaf digits 5-9 are placed in the second line.
CHAPTER 2, DESCRIPTIVE STATISTICS

The completed stem-and-leaf diagram is presented followed by the ordered stem-and-leaf diagram.

5 | 2344
5 | 579
6 | 3030434
6 | 78668
7 | 43113
7 | 7767
8 | 042

Ordered stem-and-leaf

5 | 2344
5 | 579
6 | 0033344
6 | 66788
7 | 11334
7 | 6777
8 | 024

2.65 (a) Construction of a stem-and-leaf diagram for the crime data begins with a vertical listing of the numbers comprising the stems. These numbers are 2, 3, 4, ..., 7. To the right of this listing is a vertical line which serves as a demarcation between the stems and leaves that are about to be added. Each leaf will be the right-most digit -- the units digit -- of a number presented in the data set. The completed stem-and-leaf diagram is shown below with ordered leaves.

2 | 5678
3 | 11247778999
4 | 0113445566778999
5 | 1125555789
6 | 0001349
7 | 23

(b) With two lines per stem, the leaf digits 0-4 are placed in the first of the two lines and the leaf digits 5-9 are placed in the second line. The completed stem-and-leaf diagram follows with the leaves ordered.
(c) With five lines per stem, the leaf digits 0-1 are placed in the first of the five lines, 2-3 are placed in the second line, and so on. The completed stem-and-leaf diagram is presented below as an ordered stem-and-leaf.

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>67</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>777</td>
</tr>
<tr>
<td>3</td>
<td>8999</td>
</tr>
<tr>
<td>4</td>
<td>011</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4455</td>
</tr>
<tr>
<td>4</td>
<td>6677</td>
</tr>
<tr>
<td>4</td>
<td>8999</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>5555</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>89</td>
</tr>
<tr>
<td>6</td>
<td>0001</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
</tr>
</tbody>
</table>

(d) Two lines per stem seems to be the most useful for visualizing the shape of this set of data. Five lines per stem produces a graph that is too spread out with several "holes" and numerous lines with only one leaf. One line per stem is acceptable, but concentrates about two-thirds of the data in the second, third, and fourth lines. All of the diagrams show the near symmetry of the data.
CHAPTER 2, DESCRIPTIVE STATISTICS

2.67 (a) The data rounded to the nearest 10 ml with the terminal 0 dropped are shown at the left with a 5-line stem-and-leaf plot at the right (with data ending in '5' rounded to the nearest even 10 ml).

| 102 | 98 | 102 | 98 | 98 | 9 | 1 |
| 99  | 96 | 96  | 103 | 96 | 9 |
| 99  | 91 | 101 | 99 | 103 | 9 | 5 |
| 99  | 100| 98  | 97 | 102 | 9 | 6 | 6 | 6 | 7 |
| 106 | 103| 99  | 100| 100 | 9 | 8 | 8 | 8 | 9 | 9 | 9 | 8 | 9 | 9 |
| 100 | 101| 95  | 100| 99  | 10 | 1 | 0 | 0 | 0 | 1 | 0 |
|      |    |     |    |    | 10 | 2 | 2 | 3 | 3 | 2 | 3 |
|      |    |     |    |    | 10 | 6 |

(b) The data with the units digits truncated are shown at the left in the following table and the stem-and-leaf plot is on the right.

| 102 | 97 | 101 | 97 | 97 | 9 | 1 |
| 99  | 95 | 95  | 103 | 96 | 9 |
| 98  | 91 | 101 | 98 | 102 | 9 | 5 | 5 | 4 |
| 98  | 100| 98  | 97 | 101 | 9 | 7 | 7 | 6 | 7 |
| 106 | 103| 99  | 99 | 99  | 9 | 9 | 8 | 8 | 8 | 9 | 9 | 9 | 9 | 8 |
| 99  | 101| 94  | 99 | 98  | 10 | 1 | 1 | 0 | 1 | 1 |
|      |    |     |    |    | 10 | 2 | 3 | 3 |
|      |    |     |    |    | 10 | 4 |
|      |    |     |    |    | 10 | 6 |

(c) While the values of a number of the leaves are different in the two diagrams and some of them change from one stem to another, the general shape of the data is virtually the same in both diagrams, with the diagram in (b) looking slightly more symmetric. Comparing with the diagram in Exercise 2.62, we see that both of the current diagrams are missing the "hole" that appeared in the 100 stem previously. Otherwise, the general shape is the same.

2.69 (a) With the data in a column named CRIME,

- Choose **Graph ▶ Stem and Leaf...**
- Select CRIME in the **Variables** text box
- Click on the **Increment** text box and type **10** to produce one line per stem.
- Click **OK**.

The result is

Stem-and-leaf of CRIME     N  = 50
Leaf Unit = 1.0

```
15 3 11247778999
(16) 4 0113445566778999
19 5 1125555789
9 6 0001349
2 7 23
```

(b) Follow the same procedure used in part (a), except type **5** for the interval.
The result is

Stem-and-leaf of CRIME  N = 50
Leaf Unit = 1.0

4 2 5678
8 3 1124
15 3 7778999
21 4 011344
(10) 4 5566778999
19 5 112
16 5 5555789
9 6 000134
3 6 9
2 7 23

(c) Follow the same procedure used in part (a), except type 2 for the interval. The result is

Stem-and-leaf of CRIME  N = 50
Leaf Unit = 1.0

1 2 5
3 2 67
4 2 8
6 3 11
7 3 2
8 3 4
11 3 777
15 3 8999
18 4 011
19 4 3
23 4 4455
(4) 4 6677
(4) 4 8999
19 5 11
17 5 2
16 5 5555
12 5 7
11 5 89
9 6 0001
5 6 3
4 6 4
3 6
3 6 9
2 7
2 7 23

Exercises 2.5

2.71 (a) The distribution of a data set is a table, graph, or formula that gives the values of the observations and how often each one occurs.

(b) Sample data is data obtained by observing the values of a variable for a sample of the population.

(c) Population data is data obtained by observing the values of a variable for all of the members of a population.

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

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

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

(g) A distribution of a variable is the same as a population distribution, a distribution of population data.
2.73 A large sample from a bell-shaped distribution would be expected to have roughly a bell shape.

2.75 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.

[Diagram showing bell-shaped, triangular, and uniform (or rectangular) distributions]

2.77 (a) The overall shape of the distribution of the number of white shark pups is roughly bell-shaped.

(b) The distribution is roughly symmetric.

2.79 (a) The distribution of cholesterol levels of high-level patients is left skewed. Note: The answer bell-shaped is also acceptable.

(b) The shape of the distribution of cholesterol levels of high-level patients is left skewed. Note: The answer symmetric is also acceptable.

2.81 (a) The distribution of the lengths of stay in Europe and the Mediterranean of the 36 U.S. residents is right skewed.

(b) The shape of the distribution of the lengths of stay in Europe and the Mediterranean of the 36 U.S. residents is right skewed.

2.83 The precise answers to this exercise will vary from class to class or individual to individual. Thus your results are likely to 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 2 2 2 9 3 0 3 4 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 be roughly rectangular.

(c) Using single value classes, the frequency distribution is given by the following table. The histogram is shown below.
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 when the sample size is so small, there can be 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.

2.85 (a) Your results will differ from the ones below which were obtained using Minitab. Choose Calc $\rightarrow$ Random Data $\rightarrow$ Normal..., type 3000 in the Generate rows of data text box, click in the Store in column(s) text box and type STDNORM, and click OK. Then choose Graph $\rightarrow$ Histogram, enter STDNORM in the Graph 1 text box under X, and click OK.

(b) The histogram in part (b) is bell-shaped. The sample of 3000 is
representative of the population from which the sample was taken. This suggests that the standard normal distribution is bell-shaped.

Exercises 2.6

2.87  (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 graph 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.

2.89  (a) A good 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 1993, the district average is much greater relative to the national average than it actually is.

2.91  (a) The problem with the bar graph is that it is truncated. That is, the vertical axis, which should start at $0$ (trillions), starts with $3.05$ (trillions) instead. The part of the graph from $0$ (trillions) to $3.05$ (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 a nontruncated and unmodified vertical axis is presented in Figure (a). Notice that the vertical axis starts at $0.00$ (trillions). Increments are in halves of a trillion dollars. In contrast to the original bar graph, this one illustrates that the changes in money supply from week to week are very small. 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 indicate 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 that part of the vertical axis between $0.00$ (trillions) and $3.05$ (trillions) has been removed.
2.93 (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. Pictorially, the "new" ball dwarfs the "old" ball. From the perspective of measurement, if the "new" ball has twice the radius of the "old" ball, it will have eight times the volume of the "old" ball (since the volume of a sphere is proportional to the cube of its radius). Thus, the scaling is improper because it gives the impression that the "new" ball lasts roughly eight times as long as the "old" ball.

(b) One possible way for the manufacturer to illustrate that the "new" ball lasts twice as long as the "old" ball is to present a picture of two balls, side by side, each of the same magnitude as the "old" ball and to
CHAPTER 2, DESCRIPTIVE STATISTICS

label this set of two balls "new ball". (See below.) This will illustrate that a purchaser will be getting twice as much for his or her money.

Old Ball

New Ball

REVIEW TEST FOR CHAPTER 2

1. (a) A variable is a characteristic that varies from one person or thing to another.
   (b) Variables can be quantitative or qualitative.
   (c) Quantitative variables can be discrete or continuous.
   (d) Data is information obtained by observing 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 the data do not take numerical values.

4. (a) The midpoint is halfway between the cutpoints. Since the class width is 8, the cutpoints are 6 and 14.
    (b) The class width is also the distance between consecutive midpoints. Therefore the second midpoint is $10 + 8 = 18$.
    (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 is discrete with relatively few distinct observations.

7. (a) The vertical edges of the bars will be aligned with the cutpoints.
    (b) Each bar is centered over its midpoint.

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 shaped  Uniform

11. (a) Slightly skewed to the right. Assuming that the most typical heights are around 5'10", there are likely to be more heights above 6'4" than below 5'4". An answer of roughly bell-shaped is also acceptable.
(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 before they become seniors.

12. (a) The distribution of the sample will reflect the distribution of the population, so it should be left-skewed as well.
(b) No. The randomness in the samples will almost certainly produce different sets of observations resulting in nonidentical shapes.
(c) Yes. We would expect both of the samples to reflect the shape of the population and to be left-skewed if the samples are reasonably large.

13. (a) The first column ranks the hydroelectric plants. Thus, it consists of quantitative, discrete data.
(b) The fourth column provides measurements of capacity in megawatts. Thus, it consists of quantitative, continuous data.
(c) The third column provides non-numerical information. Thus, it consists of qualitative data.

14. (a) The first class 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. The classes are presented in column 1 of the grouped-data table below. The last class is 65-69, since the largest 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 42, results in each class's relative frequency. The relative frequencies for 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 are presented in column 4.
(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 following frequency histogram 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 are the even integers in the range 0 through 12, since these are representative of the magnitude and spread of the frequencies. The height of each bar in the frequency histogram matches the respective frequency in column 2.
16. (a) Using one line per stem in constructing the ordered stem-and-leaf diagram means vertically listing the numbers comprising the stems once only. The ordered leaves are then placed with their respective stems. 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. If a leaf is one of the digits 0 through 4, it is ordered and placed with the first of the two stem lines. If a 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)     (b)

<table>
<thead>
<tr>
<th></th>
<th>2 3 6 6 7 8 9 9</th>
<th>4</th>
<th>2 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
<td>4</td>
<td>6 6 7 8 9 9</td>
</tr>
<tr>
<td>5</td>
<td>0 0 1 1 1 2 2 4 4 4 4 5 5 5 5 6 6 6 7 7 7 7 8</td>
<td>5</td>
<td>0 0 1 1 1 2 2 4 4 4 4</td>
</tr>
<tr>
<td>6</td>
<td>0 1 1 2 4 4 5 8 9</td>
<td>6</td>
<td>0 1 1 2 4 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>5 8 9</td>
</tr>
</tbody>
</table>

(c) The stem-and-leaf diagram with two lines per stem corresponds to the frequency distribution in Problem 14(a) as it groups the data in the same classes, names 40-44, 45-49, ... 65-69.

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 are presented in column 3. Since each class is based on a single value, it is not necessary to give midpoints.

<table>
<thead>
<tr>
<th>Number busy</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.08</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.08</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.16</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.20</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>0.28</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(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 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 are increments of 0.05, starting with zero and ending at 0.30. Each histogram bar is centered over the single numerical value represented by its class. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.

\[
\begin{array}{ccc}
\text{BUSY BANK TELLERS} \\
\hline
\text{TELLERS} & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
\text{Percent} & 0 & 10 & 20 & 30 & & & \\
\end{array}
\]

18. (a) The table below shows both the frequency distribution and the relative frequency distribution. If each frequency is divided by the total number of students, which is 40, we obtain the relative frequency (or percentage) of the class.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fr</td>
<td>6</td>
<td>0.150</td>
</tr>
<tr>
<td>So</td>
<td>15</td>
<td>0.375</td>
</tr>
<tr>
<td>Ju</td>
<td>12</td>
<td>0.300</td>
</tr>
<tr>
<td>Se</td>
<td>7</td>
<td>0.175</td>
</tr>
</tbody>
</table>

(b) The following pie chart displays the percentage of students at each class level.

\[
\begin{array}{ccc}
\text{CLASS LEVELS} \\
\hline
\text{So (15, 37.5%)} & \text{Fr (6, 15.0%)} & \text{Se (7, 17.5%)} \\
\text{Ju (12, 30.0%)} & \\
\end{array}
\]
(c) The following bar graph also displays the relative frequencies of each class.

![Student Class Levels Graph]

19. (a) The first class is 0<1000. Since all classes are to be of equal width, we know that the width of all classes is 1000 - 0 = 1000. The classes are presented in column 1 of Figure (a) below. The last class is 11,000<12,000, since the largest data value is 11,568.80. 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 are presented in column 3. By averaging the lower and upper cutpoints for each class, we obtain the midpoint for each class. The midpoints are presented in column 4.

<table>
<thead>
<tr>
<th>High</th>
<th>Freq.</th>
<th>Relative Frequency</th>
<th>Midpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>0&lt;1000</td>
<td>13</td>
<td>0.361</td>
<td>500</td>
</tr>
<tr>
<td>1000&lt;2000</td>
<td>10</td>
<td>0.278</td>
<td>1500</td>
</tr>
<tr>
<td>2000&lt;3000</td>
<td>4</td>
<td>0.111</td>
<td>2500</td>
</tr>
<tr>
<td>3000&lt;4000</td>
<td>4</td>
<td>0.111</td>
<td>3500</td>
</tr>
<tr>
<td>4000&lt;5000</td>
<td>0</td>
<td>0.000</td>
<td>4500</td>
</tr>
<tr>
<td>5000&lt;6000</td>
<td>1</td>
<td>0.028</td>
<td>5500</td>
</tr>
<tr>
<td>6000&lt;7000</td>
<td>1</td>
<td>0.028</td>
<td>6500</td>
</tr>
<tr>
<td>7000&lt;8000</td>
<td>0</td>
<td>0.000</td>
<td>7500</td>
</tr>
<tr>
<td>8000&lt;9000</td>
<td>1</td>
<td>0.028</td>
<td>8500</td>
</tr>
<tr>
<td>9,000&lt;10,000</td>
<td>1</td>
<td>0.028</td>
<td>9500</td>
</tr>
<tr>
<td>10,000&lt;11,000</td>
<td>0</td>
<td>0.000</td>
<td>10,500</td>
</tr>
<tr>
<td>11,000&lt;12,000</td>
<td>1</td>
<td>0.028</td>
<td>11,500</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>1.001</td>
<td></td>
</tr>
</tbody>
</table>

(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 used to label the horizontal axis. We notice that the relative frequencies presented in column 3 range in size from 0.000 to 0.361 (0 to 36.1%). Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (5%), starting with 0.00 (0%) and ending at 0.40 (40%).
32 CHAPTER 2, DESCRIPTIVE STATISTICS

The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.

20. (a) The shape of the distribution of the inauguration ages of the first 42 presidents of the United States is roughly bell-shaped.

(b) The shape of the distribution of the number of tellers busy with customers at Prescott National Bank during 25 spot checks is left skewed.

21. Answers will vary, but here is one possibility:

22. (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-third times that for 1960.

(c) Using the vertical scale, 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 at 30.

23. (a) In Minitab, first store the data in a column named AGE. Then

- Choose Graph ▶ Histogram...
CHAPTER 2 REVIEW TEST

- Specify AGE in the X text box for Graph 1.
- Click the Options... button
- Select the Cutpoint option button from the Type of Intervals field
- Select the Midpoint/cutpoint positions text box and type 40:70/5
- Click OK
- Click OK

To print the result at the right from the Graph window, choose

File ▶ Print Window... .

(b) With the data already stored in the column name AGE,

- Choose Graph ▶ Character Graphs ▶ Dotplot...
- Specify AGE in the Variables text box.
- Click OK

The computer output shown in the Session window is:

Character Dotplot

```
.:  :  :  :  :  :
.:  :  :  :  :  :  :
```

-------------------+---------+---------+---------+---------+---------
AGE                45.0      50.0      55.0      60.0      65.0      70.0

(c) With the data already stored in the column named AGES,

- Choose Graph ▶ Stem-and-Leaf...
- Specify AGE in the Variables text box.
- Click on the Intervals text box and type 5
- Click OK
CHAPTER 2, DESCRIPTIVE STATISTICS

The computer output is:

Stem-and-leaf of AGE  N = 42
Leaf Unit = 1.0
2  4  23
8  4  667899
20  5  001111224444
(12)  5  555566677778
10  6  011244
3  6  589

24. (a) Using Minitab to create the pie chart and bar chart, enter the student class level in a column named CLASS and the frequency in a column named FREQ. Then

- Choose Graph ▶ Pie Chart...
- Click on Chart Table, enter CLASS in Categories in text box and enter FREQ in the Frequencies in text box. Enter 'Pie Chart of Classes' in the Titles text box.
- Click OK. The chart follows after part (b).

(b) To create the bar chart, enter the class abbreviations in a column named CLASS and the relative frequencies from Problem 18 in a column named REL FREQ, choose Graph ▶ Chart..., select REL FREQ for the Y variable in Graph 1, and CLASS for the X variable. Click OK.