2.1 Graphs and Tables for Categorical Data

Objectives:
By the end of this section, I will be able to...

1) Construct and interpret a frequency distribution and a relative frequency distribution for qualitative data.

2) Construct and interpret bar graphs and Pareto charts.

3) Construct and interpret pie charts.

Example 2.1

What careers do teenagers admire?

<table>
<thead>
<tr>
<th>Student</th>
<th>Career</th>
<th>Student</th>
<th>Career</th>
<th>Student</th>
<th>Career</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Doctor</td>
<td>8</td>
<td>Athlete</td>
<td>15</td>
<td>Lawyer</td>
</tr>
<tr>
<td>2</td>
<td>Scientist</td>
<td>9</td>
<td>Doctor</td>
<td>16</td>
<td>Military Officer</td>
</tr>
<tr>
<td>3</td>
<td>Military Officer</td>
<td>10</td>
<td>Scientist</td>
<td>17</td>
<td>Doctor</td>
</tr>
<tr>
<td>4</td>
<td>Military Officer</td>
<td>11</td>
<td>Doctor</td>
<td>18</td>
<td>Scientist</td>
</tr>
<tr>
<td>5</td>
<td>Doctor</td>
<td>12</td>
<td>Military Officer</td>
<td>19</td>
<td>Doctor</td>
</tr>
<tr>
<td>6</td>
<td>Scientist</td>
<td>13</td>
<td>Scientist</td>
<td>20</td>
<td>Lawyer</td>
</tr>
<tr>
<td>7</td>
<td>Military Officer</td>
<td>14</td>
<td>Lawyer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1 Prestigious career survey data set

Example 2.1 continued

Solution

- Count how many students preferred that particular career
- Summarize the data

<table>
<thead>
<tr>
<th>Variable: career</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor</td>
<td>6</td>
</tr>
<tr>
<td>Scientist</td>
<td>5</td>
</tr>
<tr>
<td>Military Officer</td>
<td>5</td>
</tr>
<tr>
<td>Lawyer</td>
<td>3</td>
</tr>
<tr>
<td>Athlete</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.2 Frequency distribution of career
Relative Frequency Distribution

- The **relative frequency** of a particular category of a qualitative variable is its frequency divided by the sample size.

- A **relative frequency distribution** for a qualitative variable is a listing of all values that the variable can take, together with the relative frequencies for each value.

Example 2.2 continued

- Divide each frequency in the frequency distribution by the sample size 20

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>Scientist</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>Military Officer</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>Lawyer</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>Athlete</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 2.3 Relative frequency distribution of career preference by students

Bar Graph

- Graphical equivalent of a frequency distribution or a relative frequency distribution

- It is constructed as follows:
  1. On the horizontal axis, provide a label for each category.
  2. Draw rectangles (bars) of equal width for each category.

- The height of each rectangle represents the frequency or relative frequency for that category.

- Ensure that the bars are not touching each other.

Example 2.3 - Bar graphs of career preferences

Construct a frequency bar graph and a relative frequency bar graph for the career preference distributions in Tables 2.2 and 2.3.

Example 2.3 continued

Solution

![Graphs](image.png)

**FIGURE 2.1** (a) Frequency bar graph; (b) relative frequency bar graph.

Pareto chart

- Bar graph in which the rectangles are presented in decreasing order from left to right

- Figure 2.1a and Figure 2.1b are examples of Pareto charts
Pie Charts
- Graphical device for displaying the relative frequencies of a categorical variable
- Circle divided into sections (that is, slices or wedges)
- Each section representing a particular category
- Size of the section is proportional to the relative frequency

Example 2.4 continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>Career</th>
<th>Relative frequency</th>
<th>Multiply by 360°</th>
<th>Degrees for that section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor</td>
<td>6000</td>
<td>0.30</td>
<td>0.30 x 360° =</td>
<td>108°</td>
</tr>
<tr>
<td>Scientist</td>
<td>5000</td>
<td>0.25</td>
<td>0.25 x 360° =</td>
<td>90°</td>
</tr>
<tr>
<td>Military Officer</td>
<td>5000</td>
<td>0.25</td>
<td>0.25 x 360° =</td>
<td>90°</td>
</tr>
<tr>
<td>Lawyer</td>
<td>3000</td>
<td>0.15</td>
<td>0.15 x 360° =</td>
<td>54°</td>
</tr>
<tr>
<td>Athlete</td>
<td>1000</td>
<td>0.05</td>
<td>0.05 x 360° =</td>
<td>18°</td>
</tr>
<tr>
<td>Total</td>
<td>20000</td>
<td>1.00</td>
<td></td>
<td>360°</td>
</tr>
</tbody>
</table>

Table 2.4 Finding the number of degrees for each slice of the pie chart

2.2 Graphs and Tables for Quantitative Data

Objectives:
By the end of this section, I will be able to...

1) Construct and interpret a frequency distribution and a relative frequency distribution for quantitative data.
2) Use histograms and frequency polygons to summarize quantitative data.
3) Construct and interpret stem-and-leaf displays and dotplots.
4) Recognize distribution shape, symmetry, and skewness.

Example 2.5
The National Center for Missing and Exploited Children (www.missingkids.com) keeps an online searchable database of missing children nationwide. Table 2.13 page 47 contains a listing of the 50 children who have gone missing from California and who would have been between 1 and 9 years of age as of March 4, 2007. Suppose we are interested in analyzing the ages of these missing children. Use the data to construct a frequency distribution and a relative frequency distribution of the variable age.

Example 2.5 continued

Solution
- Construct the frequency distribution for the variable age
- Construct a relative frequency distribution by dividing the frequency by the total number of observations, 50.
Example 2.5 continued

<table>
<thead>
<tr>
<th>Age</th>
<th>Tally</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>II</td>
<td>3</td>
<td>0.06</td>
</tr>
<tr>
<td>2</td>
<td>III</td>
<td>3</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>I</td>
<td>2</td>
<td>0.04</td>
</tr>
<tr>
<td>4</td>
<td>III</td>
<td>8</td>
<td>0.16</td>
</tr>
<tr>
<td>5</td>
<td>III</td>
<td>9</td>
<td>0.18</td>
</tr>
<tr>
<td>6</td>
<td>IIII</td>
<td>10</td>
<td>0.20</td>
</tr>
<tr>
<td>7</td>
<td>III</td>
<td>6</td>
<td>0.12</td>
</tr>
<tr>
<td>8</td>
<td>III</td>
<td>6</td>
<td>0.12</td>
</tr>
<tr>
<td>9</td>
<td>II</td>
<td>3</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Total: 59

Table 2.14 Frequency distribution and relative frequency distribution of age

A more concise distribution

- Combine several ages together into "classes"
- Classes represent a range of data values and are used to group the elements in a data set.

Example 2.6

Frequency and relative frequency distributions using classes

Combine the age data from Example 2.5 into three classes, and construct frequency and relative frequency distributions.

Example 2.6 continued

Solution

- Let us define the following classes for the age data: 1–3 years old, 4–6 years old, and 7–9 years old.
- For each class, we group together all the ages in the class.

Example 2.6 continued

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–3</td>
<td>8</td>
<td>0.16</td>
</tr>
<tr>
<td>4–6</td>
<td>17</td>
<td>0.54</td>
</tr>
<tr>
<td>7–9</td>
<td>5</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Total: 30

Table 2.15 Distributions for the variable age, after combining into three classes

Construct Frequency Distributions and Histograms

- The lower class limit of a class equals the smallest value within that class.
- The upper class limit of a class equals the largest value within that class.
- The class width equals the difference between the lower class limits of two successive classes.
Using classes to construct a frequency distribution

- Determine how many classes you will use.
- Determine the class width.
- It is best (though not required) to use the same width for all classes.
- Determine the upper and lower class limits.
- Make sure the classes are nonoverlapping.

Example 2.8 - Constructing a frequency distribution:
the management aptitude test

Twenty management students, in preparation for graduation, took a course to prepare them for a management aptitude test. A simulated test provided the following scores:
77 89 84 83 80 80 83 82 85 92
87 88 87 86 99 93 79 83 81 78
Construct a frequency distribution of these management aptitude test scores.

Example 2.8 continued

Solution
Step 1: Choose the number of classes.
- Use between 5 and 20 classes
- The number of classes increasing with the sample size
- A small data set such as this will do just fine with 5 classes.
- In general, choose the number of classes to be large enough to show the variability in the data set, but not so large that many classes are nearly empty.

Example 2.8 continued

Solution
Step 2: Determine the class widths.
- Find the range of the data
- Divide this range by the number of classes you chose in Step 1.
- This gives an estimate of the class width.
- Here, our largest data value is 99 and our smallest is 77, giving us a range of 99 – 77 = 22.
- In Step 1, we chose 5 classes, so that our estimated class width is 22/5 = 4.4.
- Use a convenient class width of 5.

Example 2.8 continued

Solution
Step 3: Determine the upper and lower class limits.
- Choose limits so that each data point belongs to only one class.
- The classes should not overlap.
- Define classes:
  - 75–79
  - 80–84
  - 85–89
  - 90–94
  - 95–100
- Note that the lower class limit of the first class, 75, is slightly below that of the smallest value in the data set, 77.
- Also note that the class width equals 80 – 75 = 5, as desired.

Example 2.8 continued

<table>
<thead>
<tr>
<th>Class</th>
<th>Tally</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>75–79</td>
<td>H</td>
<td>3</td>
<td>0.15</td>
</tr>
<tr>
<td>80–84</td>
<td>HII</td>
<td>8</td>
<td>0.40</td>
</tr>
<tr>
<td>85–89</td>
<td>HII</td>
<td>6</td>
<td>0.30</td>
</tr>
<tr>
<td>90–94</td>
<td>II</td>
<td>2</td>
<td>0.10</td>
</tr>
<tr>
<td>95–100</td>
<td>I</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>30</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 2.16 Distributions for the management aptitude test scores.
**Histograms**

- Graphical summary for quantitative data
- Constructed using rectangles for each class of data.
- The heights of the rectangles represent the frequencies or relative frequencies of the class.
- The widths of the rectangles represent the class widths of the corresponding frequency distribution.
- The lower class limits are placed on the horizontal axis (along with the upper class limit of the rightmost class), so that the rectangles are touching each other.

**Example 2.9 - Histogram of management aptitude test Scores**

Construct a histogram of the frequency of the management aptitude test scores from Example 2.8.

**Example 2.9 continued**

Solution

Step 1: Find the class limits and draw the horizontal axis.
- The class limits for these data were found in Example 2.8 and are given in Table 2.16.
- The lower class limits are 75, 80, 85, 90, and 95.
- The upper class limit of the rightmost class is 100.
- Draw the horizontal axis, with the numbers 75, 80, 85, 90, 95, and 100 equally spaced.

**Example 2.9 continued**

Solution

Step 2: Determine the frequencies and draw the vertical axis.
- Use the frequencies given in Table 2.16.
- These will indicate the heights of the five rectangles along the vertical axis.
- Find the largest frequency, which is 8.
- Provide a little bit of extra vertical space above the tallest rectangle, so make 9 your highest label along the vertical axis.
- Then provide equally spaced labels along the vertical axis between 0 and 9.

**Example 2.9 continued**

Solution

Step 3: Draw the rectangles
- Draw your first rectangle.
  - Its width is from 75 to 80, and its height is 3, the first frequency.
- Draw the remaining rectangles similarly.
- The resulting frequency histogram is shown in Figure 2.8a.
- The relative frequency histogram is shown in Figure 2.8b.

**FIGURE 2.8**

Note that the two histograms have identical shapes and differ only in the labeling along the vertical axis.
FIGURE 2.8

Histograms are often presented using class midpoints rather than class limits.

Example 2.10

Construct a frequency polygon for the management aptitude test data in Example 2.8.

Solution:

Example 2.11 – Explain stem-and-leaf display

Construct a stem-and-leaf display for the final-exam scores of 20 psychology students, given below:

75 81 82 70 60 59 94 77 68 98
96 68 85 72 70 91 78 86 51 67

Example 2.11 continued

Solution
- First, find the leading digits of the numbers.
- Each number has one of the following as its leading digit: 5, 6, 7, 8, 9.
- Place these five numbers, called the stems, in a column:

  5
  6
  7
  8
  9

- Now consider the ones place of each data value.
- First score, 75, has 5 in the ones place
- Place this number, called the leaf, next to its stem:
Example 2.11 continued

Solution
- Find leaf for each remaining number and place next to stem
- For each stem, order the leaves from left to right in increasing order

<table>
<thead>
<tr>
<th>5</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0788</td>
</tr>
<tr>
<td>7</td>
<td>002578</td>
</tr>
<tr>
<td>8</td>
<td>12566</td>
</tr>
<tr>
<td>9</td>
<td>148</td>
</tr>
</tbody>
</table>

- Contains all the information that a histogram turned on its side does

Dotplots

- Simple but effective graphical display
- Each data point is represented by a dot above the number line
- Useful for comparing two variables
- Each dot may represent more than one data point for large sample sizes

Dotplots

![Figure 2.10 Dotplot of the managerial aptitude test scores. The two dots above 87 indicate that two tests had the same score of 87. Which test score was the most common?](image)

Distribution Shape, Symmetry, and Skewness

- The distribution of a variable is a table, graph, or formula that identifies the variable values and frequencies for all elements in the data set.
- The shape of a distribution is the overall form of a graphical summary.
- Shape is symmetric if there is a line (axis of symmetry) that splits the image in half so that one side is the mirror image of the other.

The Bell-Shaped Curve

![Figure 2.12 The bell-shaped curve superimposed on a histogram.](image)

Analyzing the Shape of a Distribution

- Due to random variation, data from the real world rarely exhibit perfect symmetry.
- The chi-square distribution, which is not symmetric but is skewed.

![Figure 2.14](image)
Analyzing the Shape of a Distribution

FIGURE 2.15 Some distributions are left-skewed.

2.3 Further Graphs and Tables for Quantitative Data

Objectives:
By the end of this section, I will be able to...
1) Build cumulative frequency distributions and cumulative relative frequency distributions.
2) Create frequency ogives and relative frequency ogives.
3) Construct and interpret time series graphs.

Cumulative Frequency Distributions and Cumulative Relative Frequency Distributions

- For a discrete variable, a cumulative frequency distribution shows the total number of observations less than or equal to the category value.
- For a continuous variable, a cumulative frequency distribution shows the total number of observations less than or equal to the upper class limit.
- A cumulative relative frequency distribution shows the proportion of observations less than or equal to the category value (for a discrete variable) or the proportion of observations less than or equal to the upper class limit (for a continuous variable).

Example 2.14 - Constructing cumulative frequency and relative frequency distributions

The first three columns in Table 2.21 contain the frequency distribution and relative frequency distribution for the total 2007 attendance for 25 Major League Baseball teams.

Construct a cumulative frequency distribution and a cumulative relative frequency distribution for the attendance figures.

Example 2.14 continued

Solution

<table>
<thead>
<tr>
<th>Interval (millions)</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Cumulative Frequency</th>
<th>Cumulative Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>150-220</td>
<td>5</td>
<td>0.20</td>
<td>5</td>
<td>0.20</td>
</tr>
<tr>
<td>220-300</td>
<td>6</td>
<td>0.24</td>
<td>5 x 6 = 30</td>
<td>30 / 25 = 1.20</td>
</tr>
<tr>
<td>300-380</td>
<td>4</td>
<td>0.16</td>
<td>5 x 6 = 30 + 6 = 36</td>
<td>36 / 25 = 1.44</td>
</tr>
<tr>
<td>380-460</td>
<td>3</td>
<td>0.12</td>
<td>5 x 6 = 36 + 6 = 42</td>
<td>42 / 25 = 1.68</td>
</tr>
<tr>
<td>460-500</td>
<td>1</td>
<td>0.04</td>
<td>5 x 6 = 42 + 6 = 48</td>
<td>48 / 25 = 1.92</td>
</tr>
</tbody>
</table>

Total 25 100

Table 2.21 Cumulative frequency distribution and cumulative relative frequency distribution

Ogives

- An ogive (pronounced "oh jive") is the graphical equivalent of a cumulative frequency distribution or a cumulative relative frequency distribution.

Like a frequency polygon, an ogive consists of a set of plotted points connected by line segments.

- The x coordinates of these points are the upper class limits; the y coordinates are the cumulative frequencies or cumulative relative frequencies.

Example: Construct a relative frequency ogive for the attendance data in Table 2.21
Example 2.15 continued
Solution
- For the x coordinates, we use the upper class limits for attendance, and for the y coordinates we use the cumulative relative frequencies.

![Cumulative relative frequencies graph](image)

**FIGURE 2.15** Cumulative relative frequencies.

Time Series Graphs
- A graph of time series data is called a **time series plot**.
- The horizontal axis of a time series plot represents time (for example, hours, days, months, years).
- The values of the time series data are plotted on the vertical axis, and line segments are drawn to connect the points.

Example 2.16 - Constructing a time series plot

Table 2.22 contains the amount of carbon dioxide in parts per million (ppm) found in the atmosphere above Mauna Loa in Hawaii, measured monthly from October 2006 to September 2007.

Construct a time series plot of these data.

<table>
<thead>
<tr>
<th>Month</th>
<th>Carbon dioxide (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct</td>
<td>379.03</td>
</tr>
<tr>
<td>Nov</td>
<td>380.37</td>
</tr>
<tr>
<td>Dec</td>
<td>381.86</td>
</tr>
<tr>
<td>Jan</td>
<td>382.94</td>
</tr>
<tr>
<td>Feb</td>
<td>383.17</td>
</tr>
<tr>
<td>Mar</td>
<td>381.86</td>
</tr>
</tbody>
</table>

Table 2.22 Atmospheric carbon dioxide at Mauna Loa, October 2006 to September 2007

Example 2.16 continued
Solution
- Twelve months October through September on the horizontal axis
- For each month, we plot the amount of carbon dioxide
- Join the points using line segments

![Time series plot](image)

**FIGURE 2.16** Time series plot. Carbon dioxide levels at Mauna Loa, Hawaii.

2.4 Graphical Misrepresentations of Data

**Objective:**
By the end of this section, I will be able to...

1) Understand what can make a graph misleading, confusing, or deceptive.
Eight Common Methods for Making a Graph Misleading
1. Graphing/Selecting an inappropriate statistic.
2. Omitting the zero on the relevant scale.
3. Manipulating the scale.
4. Using two dimensions (area) to emphasize a one-dimensional difference.
5. Careless combination of categories in a bar graph.
6. Inaccuracy in relative lengths of bars in a bar graph.
7. Biased distortion or embellishment.
8. Unclear labeling.

Example 2.18 - Inappropriate choice of statistic
The United Nations Office on Drugs and Crime reports the statistics, given in Table 2.26, on the top 5 nations in the world ranked by numbers of cars stolen in 2000.

<table>
<thead>
<tr>
<th>Country</th>
<th>Cars stolen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. United States</td>
<td>1,147,300</td>
</tr>
<tr>
<td>2. United Kingdom</td>
<td>338,706</td>
</tr>
<tr>
<td>3. Japan</td>
<td>309,638</td>
</tr>
<tr>
<td>4. France</td>
<td>303,539</td>
</tr>
<tr>
<td>5. Italy</td>
<td>243,859</td>
</tr>
</tbody>
</table>

Table 2.26 Top five nations for total number of cars stolen in 2000

Example 2.18 continued
The car thieves seem to be preying on cars in the United States, which has endured nearly as many cars stolen as the next four highest countries put together.

Example 2.18 continued
Is it possible that, per capita (per person), the car theft rate in the United States is not so bad?

Example 2.18 continued
Solution
- Total number of cars stolen is an inappropriate statistic since the population of the United States is greater than the populations of the other countries.
- To find the per capita car theft rate, divide the number of cars stolen in a country by that country's population.

Example 2.18 - Solution
- Top 5 countries for per capita car theft contains a few surprises
- The US is now in 9th place

<table>
<thead>
<tr>
<th>Country</th>
<th>Cars stolen per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Australia</td>
<td>0.00712</td>
</tr>
<tr>
<td>2. Denmark</td>
<td>0.00609</td>
</tr>
<tr>
<td>3. United Kingdom</td>
<td>0.00652</td>
</tr>
<tr>
<td>4. New Zealand</td>
<td>0.00653</td>
</tr>
<tr>
<td>5. Norway</td>
<td>0.00616</td>
</tr>
</tbody>
</table>

Table 2.27 Top five nations for total number of cars stolen per capita in 2000
Example 2.19 - Omitting the zero

MediaMatters.com reported that CNN.com used a misleading graph, reproduced here as Figure 2.32, to exaggerate the difference between the percentages of Democrats and Republicans who agreed with the Florida court's decision to remove the feeding tube from Terri Schiavo in 2005. Explain how Figure 2.32 is misleading.

Example 2.19 continued

Solution
- Figure 2.32 is misleading because the vertical scale does not begin at zero.
- MediaMatters.com published an amended graphic, reproduced here as Figure 2.33, which includes the zero on the vertical axis and much reduces the difference among the political parties.

Example 2.20 - Manipulating the scale

Figure 2.34 shows a Minitab relative frequency bar graph of the majors chosen by 25 business school students. Explain how we could manipulate the scale to de-emphasize the differences.

Example 2.20 continued

Solution
- If we wanted to de-emphasize the differences, we could extend the vertical scale up to its maximum, 1.0 - 100%, to produce the graph in Figure 2.35.

Example 2.21 - Using two dimensions for a one-dimensional difference and unclear labeling

Figure 2.36 compares the leaders in career points scored in the NBA All-Star Game among players active in 2007. Explain how this graphic may be misleading.
Example 2.21 continued

Solution
The height of the players is supposed to represent the total points, but this is not clearly labeled. Points should be indicated using a vertical axis, but there is no vertical axis at all.

Example 2.21 continued

• Note that Shaquille O'Neal dominates the graphic, because his body area is larger than the body areas of the other players.
• All four players should have the same body width, just as all bars in a bar graph have the same width.
• This graph uses two dimensions (height and width) to emphasize a one-dimensional (points) difference.

Example 2.22 - Careless combination of categories in a bar graph and biased embellishment

Figure 2.37 shows a graphic of how often people have observed drivers running red lights. Explain how this graphic may be considered both confusing and biased.

Example 2.22 continued

Solution
• Categories of seldom and never have been combined.
• What is "seldom" to one person may not be "seldom" to someone else.
• Red light of the "Seldom/never" category is lit up, which may be evidence of bias.

Example 2.22 continued

How often we see red lights run

Dial: 24%
Few times a week: 33%
Seldom/never: 33%

FIGURE 2.37

Example 2.23 - Inaccuracy in relative lengths of bars in a bar graph and unclear labeling

Figure 2.38 is a horizontal bar graph of the three teams with the most World Series victories in baseball history.

Explain what is unclear or misleading about this graph.
Example 2.23 continued
Solution
- Note that 127 is more than twice as many as 52, and so the 'yankees' bar should be more than twice as long as the Cardinals' bar, which it is not.

Example 2.24 - Presenting the same data set as both symmetric and left-skewed
The National Center for Education Statistics sponsors the Trends in International Mathematics and Science Study (TIMSS). In 2003, science tests were administered to eighth-grade students in countries around the world. Construct two different histograms, one that shows the data as almost symmetric and one that shows the data as left-skewed.

Example 2.24 continued

<table>
<thead>
<tr>
<th>Country</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td>573</td>
</tr>
<tr>
<td>Taiwan</td>
<td>571</td>
</tr>
<tr>
<td>South Korea</td>
<td>559</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>555</td>
</tr>
<tr>
<td>Japan</td>
<td>553</td>
</tr>
<tr>
<td>Hungary</td>
<td>545</td>
</tr>
<tr>
<td>Netherlands</td>
<td>531</td>
</tr>
<tr>
<td>United States</td>
<td>523</td>
</tr>
<tr>
<td>Australia</td>
<td>527</td>
</tr>
<tr>
<td>Sweden</td>
<td>524</td>
</tr>
<tr>
<td>Switzerland</td>
<td>526</td>
</tr>
</tbody>
</table>

Table 2.28 Science test scores

Example 2.24 continued
Solution
- Figure 2.39 is nearly symmetric.
- But Figure 2.40 is clearly left-skewed. Both figures are histograms of the very same data set.
- Clever choices for the number of classes and the class limits can affect how a histogram presents the data.
- Reader beware!