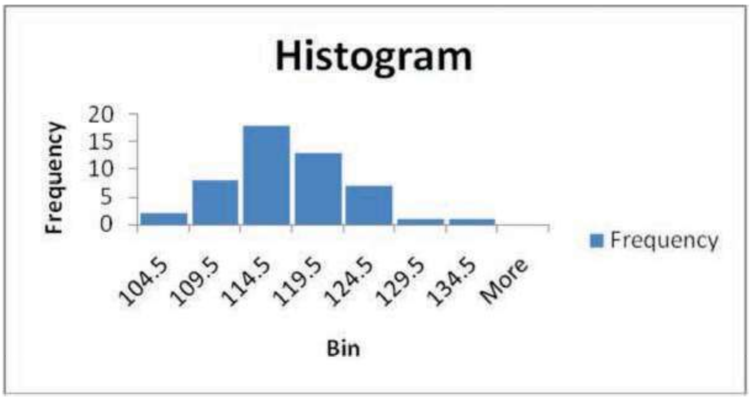
7. Select New Worksheet Ply and Chart Output. Click [OK].



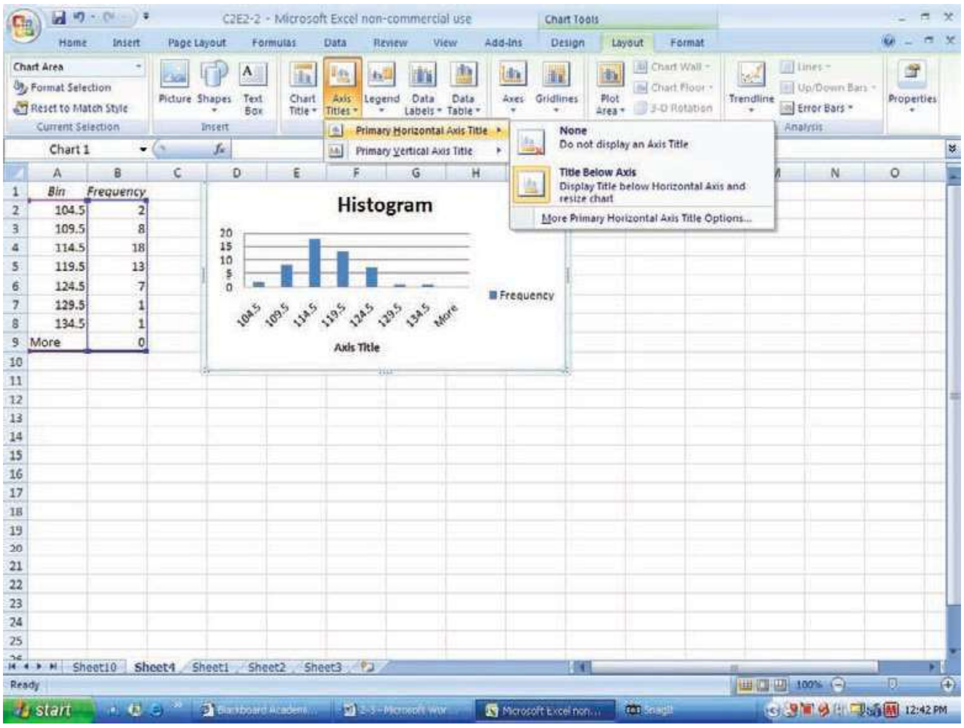
Editing the Histogram

To move the vertical bars of the histogram closer together:

1. Right-click one of the bars of the histogram, and select Format Data Series.
2. Move the Gap Width bar to the left to narrow the distance between bars.

To change the label for the horizontal axis:

1. Left-click the mouse over any region of the histogram.
2. Select the Chart Tools tab from the toolbar.
3. Select the Layout tab, Axis Titles and Primary Horizontal Axis Title.



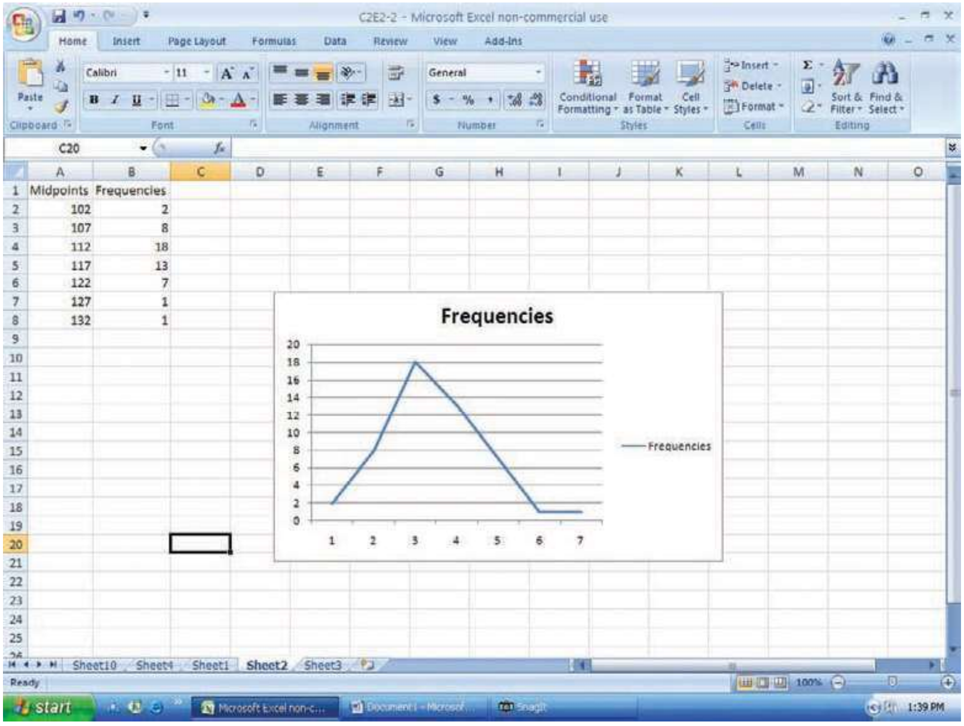
Once the Axis Title text box is selected, you can type in the name of the variable represented on the horizontal axis.

Constructing a Frequency Polygon

1. Press **[Ctrl]-N** for a new workbook.
2. Enter the midpoints of the data from Example 2–2 into column A. Enter the frequencies into column B.

	A	B
1	Midpoints	Frequencies
2	102	2
3	107	8
4	112	18
5	117	13
6	122	7
7	127	1
8	132	1

3. Highlight the Frequencies (including the label) from column B.
4. Select the Insert tab from the toolbar and the Line Chart option.
5. Select the 2-D line chart type.



We will need to edit the graph so that the midpoints are on the horizontal axis and the frequencies are on the vertical axis.

1. Right-click the mouse on any region of the graph.
2. Select the Select Data option.

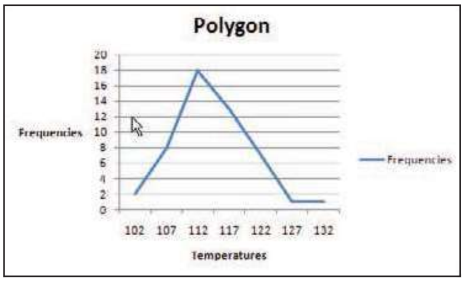
3. Select Edit from the Horizontal Axis Labels and highlight the midpoints from column A, then click [OK].
4. Click [OK] on the Select Data Source box.

Inserting Labels on the Axes

1. Click the mouse on any region of the graph.
2. Select Chart Tools and then Layout on the toolbar.
3. Select Axis Titles to open the horizontal and vertical axis text boxes. Then manually type in labels for the axes.

Changing the Title

1. Select Chart Tools, Layout from the toolbar.
2. Select Chart Title.
3. Choose one of the options from the Chart Title menu and edit.

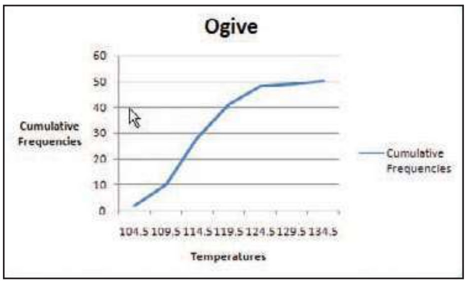


Constructing an Ogive

To create an ogive, you can use the upper class boundaries (horizontal axis) and cumulative frequencies (vertical axis) from the frequency distribution.

1. Type the upper class boundaries and cumulative frequencies into adjacent columns of an Excel worksheet.
2. Highlight the cumulative frequencies (including the label) and select the Insert tab from the toolbar.
3. Select Line Chart, then the 2-D Line option.

As with the frequency polygon, you can insert labels on the axes and a chart title for the ogive.

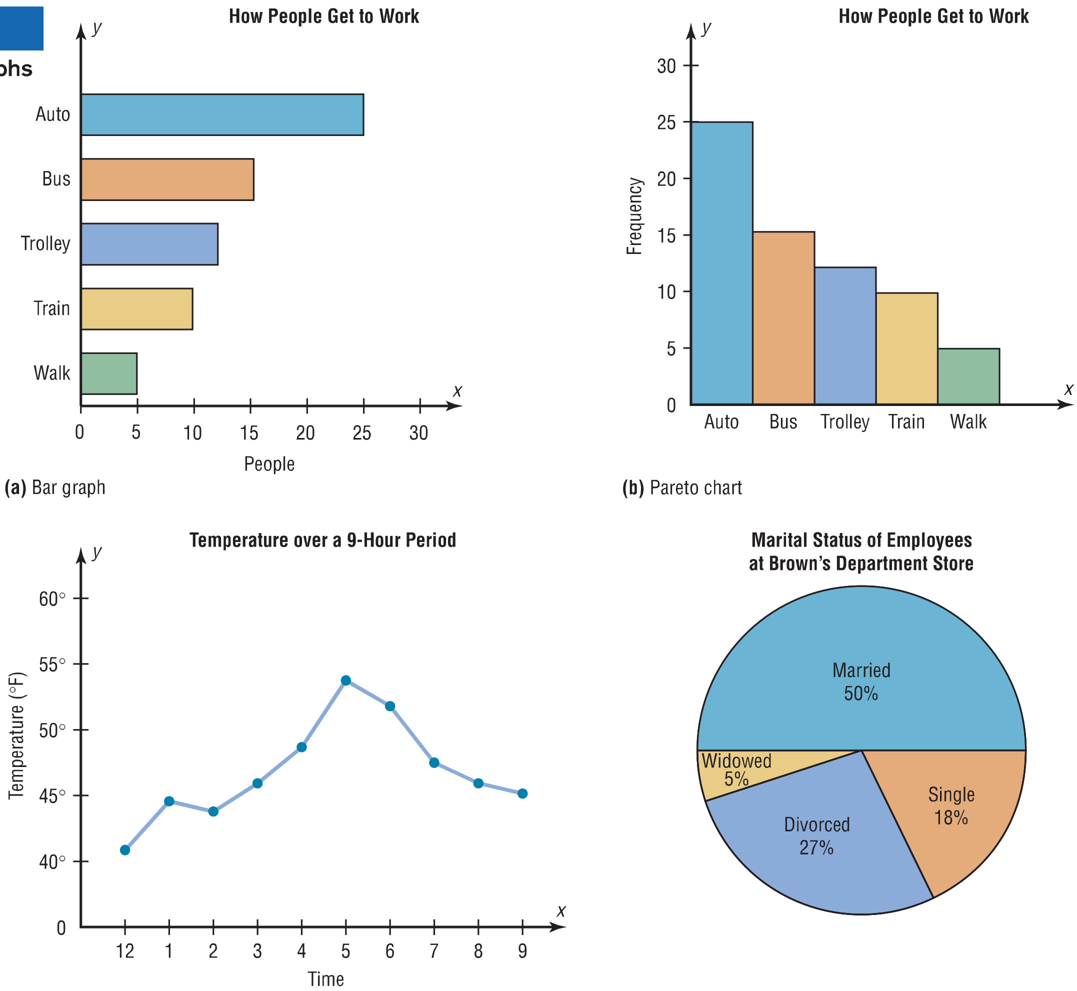


2–3 Other Types of Graphs

In addition to the histogram, the frequency polygon, and the ogive, several other types of graphs are often used in statistics. They are the bar graph, Pareto chart, time series graph, and pie graph. Figure 2–9 shows an example of each type of graph.

Figure 2–9

Other Types of Graphs Used in Statistics



(a) Bar graph

(b) Pareto chart

(c) Time series graph

(d) Pie graph

Objective 3

Represent data using bar graphs, Pareto charts, time series graphs, and pie graphs.

Bar Graphs

When the data are qualitative or categorical, bar graphs can be used to represent the data. A bar graph can be drawn using either horizontal or vertical bars.

A **bar graph** represents the data by using vertical or horizontal bars whose heights or lengths represent the frequencies of the data.

Example 2–8

College Spending for First-Year Students

The table shows the average money spent by first-year college students. Draw a horizontal and vertical bar graph for the data.

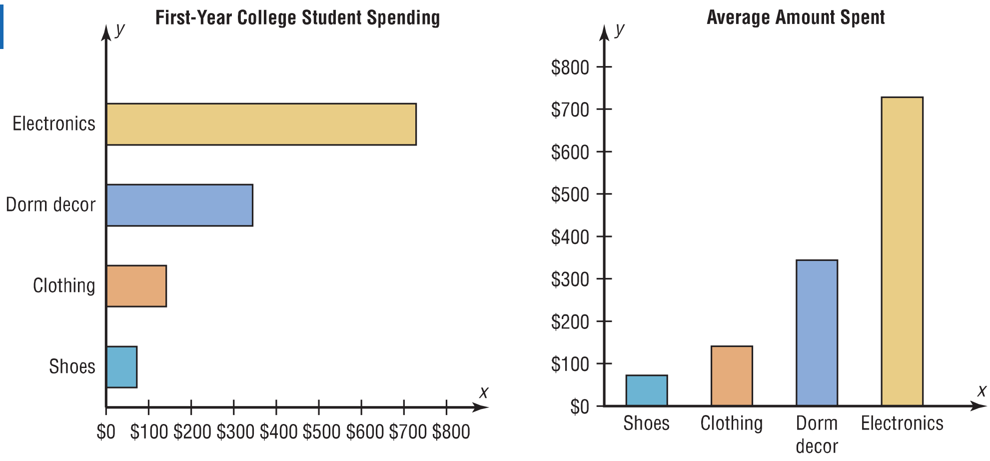
Electronics	$728
Dorm decor	344
Clothing	141
Shoes	72

Source: The National Retail Federation.

Solution

1. Draw and label the *x* and *y* axes. For the horizontal bar graph place the frequency scale on the *x* axis, and for the vertical bar graph place the frequency scale on the *y* axis.
2. Draw the bars corresponding to the frequencies. See Figure 2–10.

Figure 2–10
Bar Graphs for Example 2–8



The graphs show that first-year college students spend the most on electronic equipment including computers.

Pareto Charts

When the variable displayed on the horizontal axis is qualitative or categorical, a *Pareto chart* can also be used to represent the data.

A **Pareto chart** is used to represent a frequency distribution for a categorical variable, and the frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.

Example 2–9 Homeless People

The data shown here consist of the number of homeless people for a sample of selected cities. Construct and analyze a Pareto chart for the data.

City	Number
Atlanta	6832
Baltimore	2904
Chicago	6680
St. Louis	1485
Washington	5518

Source: U.S. Department of Housing and Urban Development.

Historical Note

Vilfredo Pareto (1848–1923) was an Italian scholar who developed theories in economics, statistics, and the social sciences. His contributions to statistics include the development of a mathematical function used in economics. This function has many statistical applications and is called the Pareto distribution. In addition, he researched income distribution, and his findings became known as Pareto's law.

Solution

Step 1 Arrange the data from the largest to smallest according to frequency.

City	Number
Atlanta	6832
Chicago	6680
Washington	5518
Baltimore	2904
St. Louis	1485

Step 2 Draw and label the x and y axes.

Step 3 Draw the bars corresponding to the frequencies. See Figure 2–11.

The graph shows that the number of homeless people is about the same for Atlanta and Chicago and a lot less for Baltimore and St. Louis.

Suggestions for Drawing Pareto Charts

1. Make the bars the same width.
2. Arrange the data from largest to smallest according to frequency.
3. Make the units that are used for the frequency equal in size.

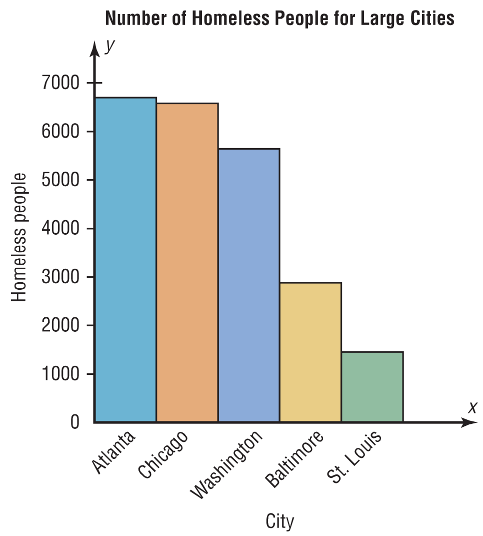
When you analyze a Pareto chart, make comparisons by looking at the heights of the bars.

The Time Series Graph

When data are collected over a period of time, they can be represented by a time series graph.

Figure 2–11

Pareto Chart for Example 2–9



A **time series graph** represents data that occur over a specific period of time.

Example 2–10 shows the procedure for constructing a time series graph.

Example 2–10 Workplace Homicides

The number of homicides that occurred in the workplace for the years 2003 to 2008 is shown. Draw and analyze a time series graph for the data.

Year	’03	’04	’05	’06	’07	’08
Number	632	559	567	540	628	517

Source: Bureau of Labor Statistics.

Solution

Historical Note

Time series graphs are over 1000 years old. The first ones were used to chart the movements of the planets and the sun.

Step 1 Draw and label the x and y axes.

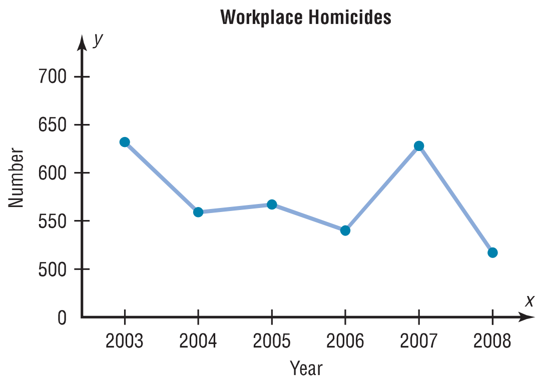
Step 2 Label the x axis for years and the y axis for the number.

Step 3 Plot each point according to the table.

Step 4 Draw line segments connecting adjacent points. Do not try to fit a smooth curve through the data points. See Figure 2–12.

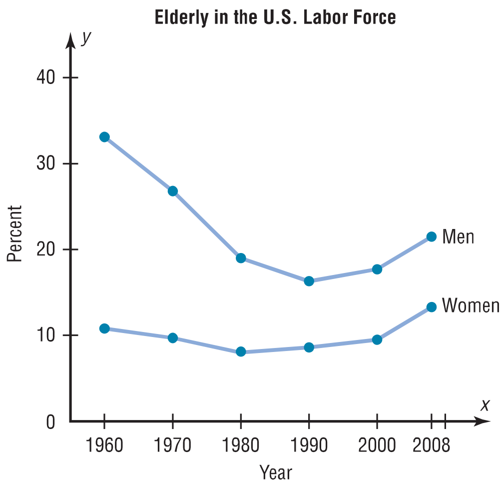
There was a slight decrease in the years ’04, ’05, and ’06, compared to ’03, and again an increase in ’07. The largest decrease occurred in ’08.

Figure 2–12
Time Series Graph for Example 2–10



When you analyze a time series graph, look for a trend or pattern that occurs over the time period. For example, is the line ascending (indicating an increase over time) or descending (indicating a decrease over time)? Another thing to look for is the slope, or steepness, of the line. A line that is steep over a specific time period indicates a rapid increase or decrease over that period.

Figure 2–13
Two Time Series Graphs for Comparison



Source: Bureau of Census, U.S. Department of Commerce.

Two or more data sets can be compared on the same graph called a *compound time series graph* if two or more lines are used, as shown in Figure 2–13. This graph shows the percentage of elderly males and females in the United States labor force from 1960 to 2008. It shows that the percent of elderly men decreased significantly from 1960 to 1990 and then increased slightly after that. For the elderly females, the percent decreased slightly from 1960 to 1980 and then increased from 1980 to 2008.

The Pie Graph

Pie graphs are used extensively in statistics. The purpose of the pie graph is to show the relationship of the parts to the whole by visually comparing the sizes of the sections. Percentages or proportions can be used. The variable is nominal or categorical.

A **pie graph** is a circle that is divided into sections or wedges according to the percentage of frequencies in each category of the distribution.

Example 2–11 shows the procedure for constructing a pie graph.

Example 2–11 **Super Bowl Snack Foods**

This frequency distribution shows the number of pounds of each snack food eaten during the Super Bowl. Construct a pie graph for the data.

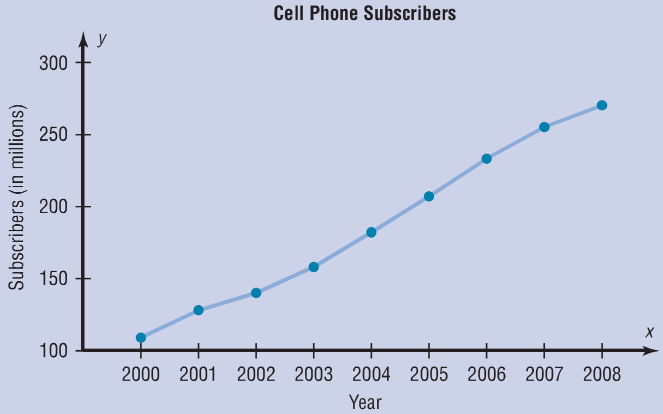
Snack	Pounds (frequency)
Potato chips	11.2 million
Tortilla chips	8.2 million
Pretzels	4.3 million
Popcorn	3.8 million
Snack nuts	2.5 million
Total	n = 30.0 million

Source: *USA TODAY Weekend.*

Speaking of Statistics

Cell Phone Usage

The graph shows the estimated number (in millions) of cell phone subscribers since 2000. How do you think the growth will affect our way of living? For example, emergencies can be handled faster since people are using their cell phones to call 911.



Source: The World Almanac and Book of Facts 2010.

Solution

Step 1 Since there are 360° in a circle, the frequency for each class must be converted into a proportional part of the circle. This conversion is done by using the formula

$$\text{Degrees} = \frac{f}{n} \cdot 360°$$

where f = frequency for each class and n = sum of the frequencies. Hence, the following conversions are obtained. The degrees should sum to 360°.*

Potato chips	$\frac{11.2}{30} \cdot 360° =$	134°
Tortilla chips	$\frac{8.2}{30} \cdot 360° =$	98°
Pretzels	$\frac{4.3}{30} \cdot 360° =$	52°
Popcorn	$\frac{3.8}{30} \cdot 360° =$	46°
Snack nuts	$\frac{2.5}{30} \cdot 360° =$	30°
Total		360°

Step 2 Each frequency must also be converted to a percentage. Recall from Example 2–1 that this conversion is done by using the formula

$$\% = \frac{f}{n} \cdot 100$$

Hence, the following percentages are obtained. The percentages should sum to 100%.†

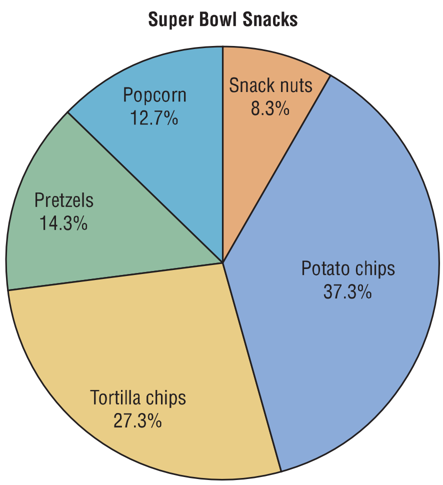
Potato chips	$\frac{11.2}{30} \cdot 100 = 37.3\%$
Tortilla chips	$\frac{8.2}{30} \cdot 100 = 27.3\%$

*Note: The degrees column does not always sum to 360° due to rounding.
†Note: The percent column does not always sum to 100% due to rounding.

Pretzels	$\frac{4.3}{30} \cdot 100 =$	14.3%
Popcorn	$\frac{3.8}{30} \cdot 100 =$	12.7%
Snack nuts	$\frac{2.5}{30} \cdot 100 =$	8.3%
Total		99.9%

Step 3 Next, using a protractor and a compass, draw the graph using the appropriate degree measures found in step 1, and label each section with the name and percentages, as shown in Figure 2–14.

Figure 2–14
Pie Graph for Example 2–11



Example 2–12

Construct a pie graph showing the blood types of the army inductees described in Example 2–1. The frequency distribution is repeated here.

Class	Frequency	Percent
A	5	20
B	7	28
O	9	36
AB	4	16
	25	100

Solution

Step 1 Find the number of degrees for each class, using the formula

$$\text{Degrees} = \frac{f}{n} \cdot 360^\circ$$

For each class, then, the following results are obtained.

A $\quad \frac{5}{25} \cdot 360^\circ = 72^\circ$

B $\quad \frac{7}{25} \cdot 360^\circ = 100.8^\circ$

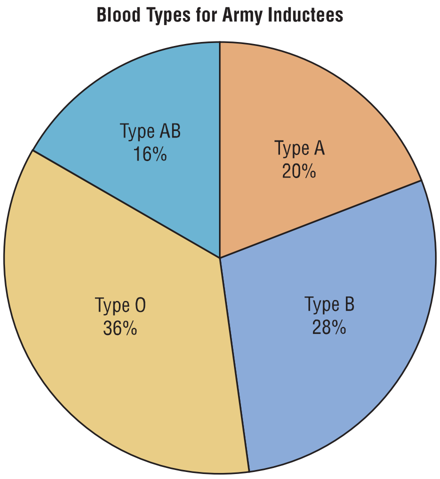
O $\frac{9}{25} \cdot 360° = 129.6°$

AB $\frac{4}{25} \cdot 360° = 57.6°$

Step 2 Find the percentages. (This was already done in Example 2–1.)

Step 3 Using a protractor, graph each section and write its name and corresponding percentage, as shown in Figure 2–15.

Figure 2–15
Pie Graph for Example 2–12



The graph in Figure 2–15 shows that in this case the most common blood type is type O.

To analyze the nature of the data shown in the pie graph, look at the size of the sections in the pie graph. For example, are any sections relatively large compared to the rest?

Figure 2–15 shows that among the inductees, type O blood is more prevalent than any other type. People who have type AB blood are in the minority. More than twice as many people have type O blood as type AB.

Misleading Graphs

Graphs give a visual representation that enables readers to analyze and interpret data more easily than they could simply by looking at numbers. However, inappropriately drawn graphs can misrepresent the data and lead the reader to false conclusions. For example, a car manufacturer's ad stated that 98% of the vehicles it had sold in the past 10 years were still on the road. The ad then showed a graph similar to the one in Figure 2–16. The graph shows the percentage of the manufacturer's automobiles still on the road and the percentage of its competitors' automobiles still on the road. Is there a large difference? Not necessarily.

Notice the scale on the vertical axis in Figure 2–16. It has been cut off (or truncated) and starts at 95%. When the graph is redrawn using a scale that goes from 0 to 100%, as in Figure 2–17, there is hardly a noticeable difference in the percentages. Thus, changing the units at the starting point on the y axis can convey a very different visual representation of the data.

Figure 2–16

Graph of Automaker's Claim Using a Scale from 95 to 100%

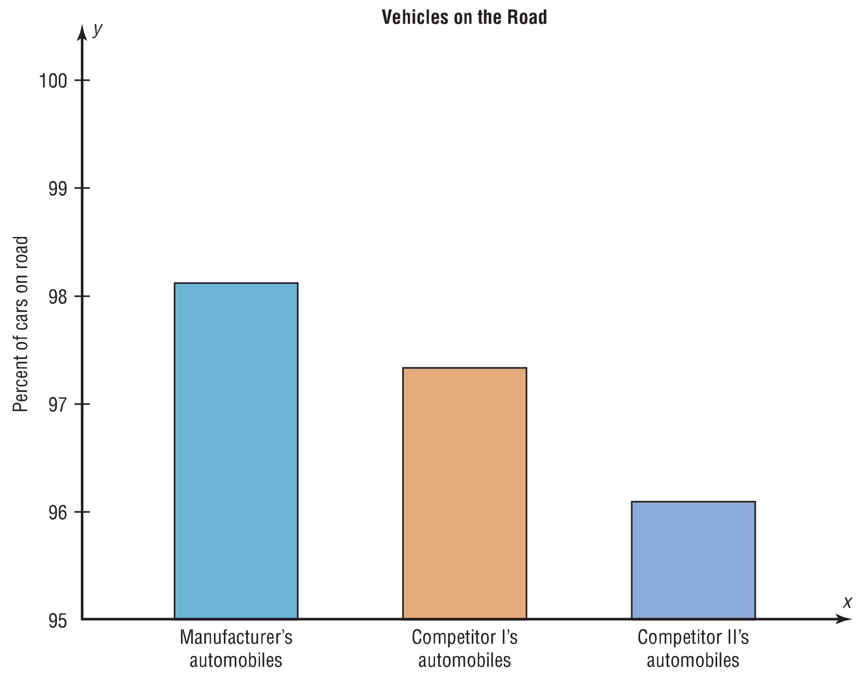
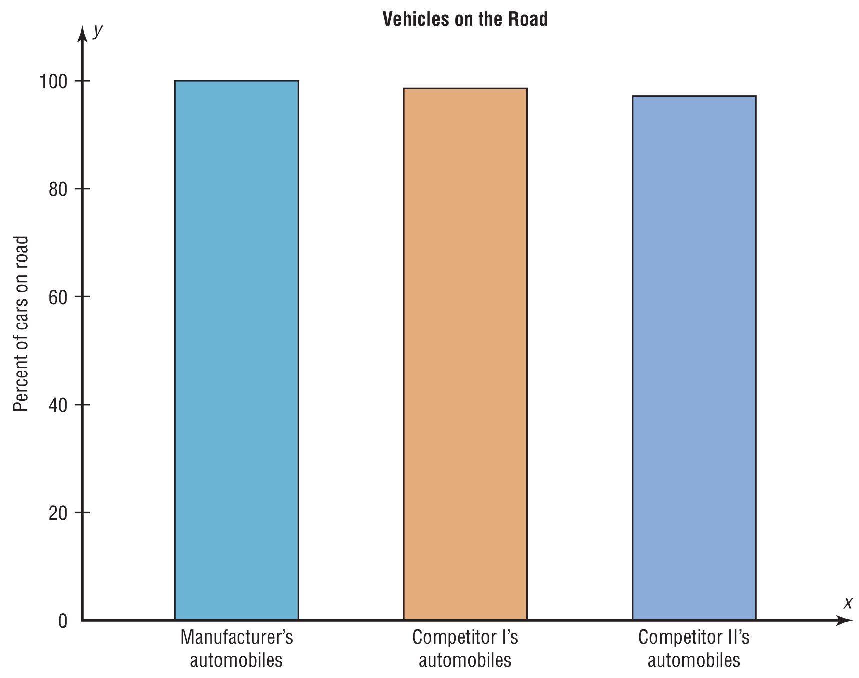


Figure 2–17

Graph in Figure 2–16 Redrawn Using a Scale from 0 to 100%



It is not wrong to truncate an axis of the graph; many times it is necessary to do so. However, the reader should be aware of this fact and interpret the graph accordingly. Do not be misled if an inappropriate impression is given.

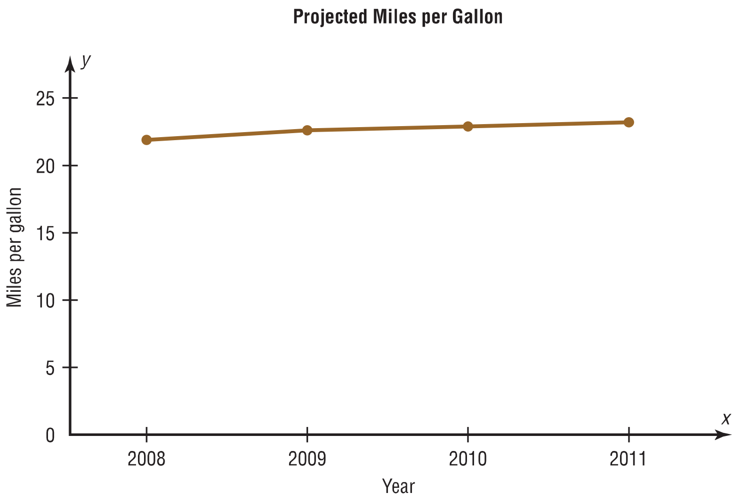
Let us consider another example. The projected required fuel economy in miles per gallon for General Motors vehicles is shown. In this case, an increase from 21.9 to 23.2 miles per gallon is projected.

Year	2008	2009	2010	2011
MPG	21.9	22.6	22.9	23.2

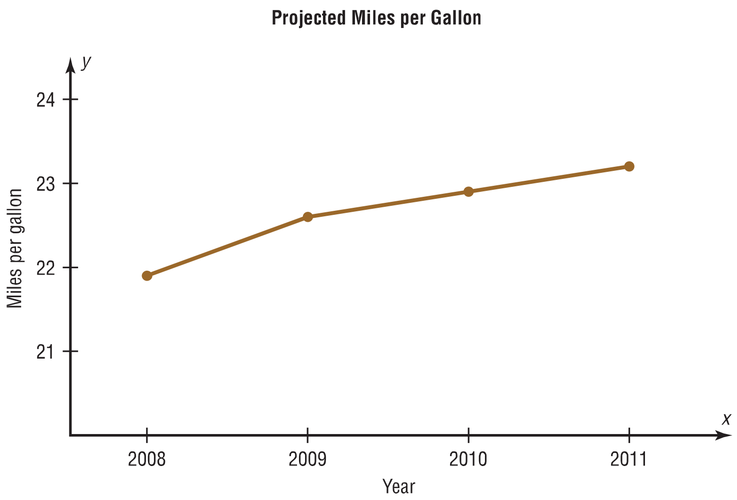
Source: National Highway Traffic Safety Administration.

When you examine the graph shown in Figure 2–18(a) using a scale of 0 to 25 miles per gallon, the graph shows a slight increase. However, when the scale is changed to 21

Figure 2–18
Projected Miles per Gallon



(a)



(b)

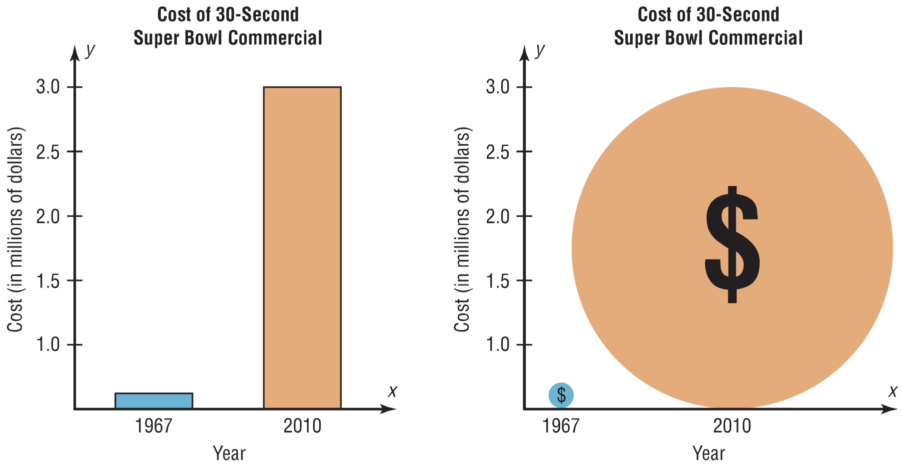
to 24 miles per gallon, the graph shows a much larger increase even though the data remain the same. See Figure 2–18(b). Again, by changing the units or starting point on the *y* axis, one can change the visual representation.

Another misleading graphing technique sometimes used involves exaggerating a one-dimensional increase by showing it in two dimensions. For example, the average cost of a 30-second Super Bowl commercial has increased from $42,000 in 1967 to $3 million in 2010 (Source: *USA TODAY*).

The increase shown by the graph in Figure 2–19(a) represents the change by a comparison of the heights of the two bars in one dimension. The same data are shown two-dimensionally with circles in Figure 2–19(b). Notice that the difference seems much larger because the eye is comparing the areas of the circles rather than the lengths of the diameters.

Note that it is not wrong to use the graphing techniques of truncating the scales or representing data by two-dimensional pictures. But when these techniques are used, the reader should be cautious of the conclusion drawn on the basis of the graphs.

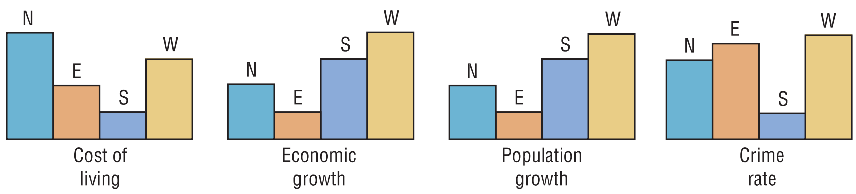
Figure 2–19
Comparison of Costs for a 30-Second Super Bowl Commercial



(a) Graph using bars
(b) Graph using circles

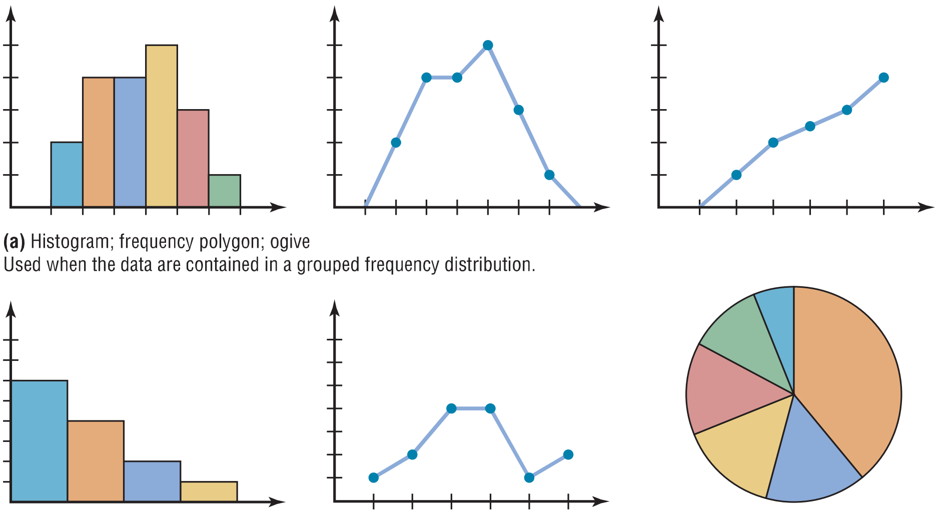
Another way to misrepresent data on a graph is by omitting labels or units on the axes of the graph. The graph shown in Figure 2–20 compares the cost of living, economic growth, population growth, etc., of four main geographic areas in the United States. However, since there are no numbers on the *y* axis, very little information can be gained from this graph, except a crude ranking of each factor. There is no way to decide the actual magnitude of the differences.

Figure 2–20
A Graph with No Units on the *y* Axis



Finally, all graphs should contain a source for the information presented. The inclusion of a source for the data will enable you to check the reliability of the organization presenting the data. A summary of the types of graphs and their uses is shown in Figure 2–21.

Figure 2–21
Summary of Graphs and Uses of Each



(a) Histogram; frequency polygon; ogive
Used when the data are contained in a grouped frequency distribution.

(b) Pareto chart
Used to show frequencies for nominal or qualitative variables.

(c) Time series graph
Used to show a pattern or trend that occurs over a period of time.

(d) Pie graph
Used to show the relationship between the parts and the whole. (Most often uses percentages.)

Stem and Leaf Plots

The stem and leaf plot is a method of organizing data and is a combination of sorting and graphing. It has the advantage over a grouped frequency distribution of retaining the actual data while showing them in graphical form.

Objective 4
Draw and interpret a stem and leaf plot.

> A **stem and leaf plot** is a data plot that uses part of the data value as the stem and part of the data value as the leaf to form groups or classes.

Example 2–13 shows the procedure for constructing a stem and leaf plot.

Example 2–13

At an outpatient testing center, the number of cardiograms performed each day for 20 days is shown. Construct a stem and leaf plot for the data.

25	31	20	32	13
14	43	02	57	23
36	32	33	32	44
32	52	44	51	45

Speaking of **Statistics**

How Much Paper Money Is in Circulation Today?

The Federal Reserve estimated that during a recent year, there were 22 billion bills in circulation. About 35% of them were \$1 bills, 3% were \$2 bills, 8% were \$5 bills, 7% were \$10 bills, 23% were \$20 bills, 5% were \$50 bills, and 19% were \$100 bills. It costs about 3¢ to print each \$1 bill.

The average life of a \$1 bill is 22 months, a \$10 bill 3 years, a \$20 bill 4 years, a \$50 bill 9 years, and a \$100 bill 9 years. What type of graph would you use to represent the average lifetimes of the bills?

Solution

Step 1 Arrange the data in order:

02, 13, 14, 20, 23, 25, 31, 32, 32, 32,
32, 33, 36, 43, 44, 44, 45, 51, 52, 57

Note: Arranging the data in order is not essential and can be cumbersome when the data set is large; however, it is helpful in constructing a stem and leaf plot. The leaves in the final stem and leaf plot should be arranged in order.

Step 2 Separate the data according to the first digit, as shown.

02 13, 14 20, 23, 25 31, 32, 32, 32, 32, 33, 36
43, 44, 44, 45 51, 52, 57

Step 3 A display can be made by using the leading digit as the *stem* and the trailing digit as the *leaf.* For example, for the value 32, the leading digit, 3, is the stem and the trailing digit, 2, is the leaf. For the value 14, the 1 is the stem and the 4 is the leaf. Now a plot can be constructed as shown in Figure 2–22.

Leading digit (stem)	Trailing digit (leaf)
0	2
1	3 4
2	0 3 5
3	1 2 2 2 2 3 6
4	3 4 4 5
5	1 2 7

Figure 2–22

Stem and Leaf Plot for Example 2–13

0	2
1	3 4
2	0 3 5
3	1 2 2 2 2 3 6
4	3 4 4 5
5	1 2 7

Figure 2–22 shows that the distribution peaks in the center and that there are no gaps in the data. For 7 of the 20 days, the number of patients receiving cardiograms was between 31 and 36. The plot also shows that the testing center treated from a minimum of 2 patients to a maximum of 57 patients in any one day.

If there are no data values in a class, you should write the stem number and leave the leaf row blank. Do not put a zero in the leaf row.

Example 2–14

An insurance company researcher conducted a survey on the number of car thefts in a large city for a period of 30 days last summer. The raw data are shown. Construct a stem and leaf plot by using classes 50–54, 55–59, 60–64, 65–69, 70–74, and 75–79.

52	62	51	50	69
58	77	66	53	57
75	56	55	67	73
79	59	68	65	72
57	51	63	69	75
65	53	78	66	55

Solution

Step 1 Arrange the data in order.

50, 51, 51, 52, 53, 53, 55, 55, 56, 57, 57, 58, 59, 62, 63, 65, 65, 66, 66, 67, 68, 69, 69, 72, 73, 75, 75, 77, 78, 79

Step 2 Separate the data according to the classes.

50, 51, 51, 52, 53, 53　　55, 55, 56, 57, 57, 58, 59
62, 63　　65, 65, 66, 66, 67, 68, 69, 69　　72, 73
75, 75, 77, 78, 79

Step 3 Plot the data as shown here.

Leading digit (stem)	Trailing digit (leaf)
5	0 1 1 2 3 3
5	5 5 6 7 7 8 9
6	2 3
6	5 5 6 6 7 8 9 9
7	2 3
7	5 5 7 8 9

The graph for this plot is shown in Figure 2–23.

Figure 2–23
Stem and Leaf Plot for Example 2–14

5	0 1 1 2 3 3
5	5 5 6 7 7 8 9
6	2 3
6	5 5 6 6 7 8 9 9
7	2 3
7	5 5 7 8 9

When the data values are in the hundreds, such as 325, the stem is 32 and the leaf is 5. For example, the stem and leaf plot for the data values 325, 327, 330, 332, 335, 341, 345, and 347 looks like this.

32	5 7
33	0 2 5
34	1 5 7

When you analyze a stem and leaf plot, look for peaks and gaps in the distribution. See if the distribution is symmetric or skewed. Check the variability of the data by looking at the spread.

Interesting Fact

The average number of pencils and index cards David Letterman tosses over his shoulder during one show is 4.

Related distributions can be compared by using a back-to-back stem and leaf plot. The back-to-back stem and leaf plot uses the same digits for the stems of both distributions, but the digits that are used for the leaves are arranged in order out from the stems on both sides. Example 2–15 shows a back-to-back stem and leaf plot.

Example 2–15

The number of stories in two selected samples of tall buildings in Atlanta and Philadelphia is shown. Construct a back-to-back stem and leaf plot, and compare the distributions.

Atlanta					Philadelphia				
55	70	44	36	40	61	40	38	32	30
63	40	44	34	38	58	40	40	25	30
60	47	52	32	32	54	40	36	30	30
50	53	32	28	31	53	39	36	34	33
52	32	34	32	50	50	38	36	39	32
26	29								

Source: *The World Almanac and Book of Facts.*

Solution

Step 1 Arrange the data for both data sets in order.

Step 2 Construct a stem and leaf plot using the same digits as stems. Place the digits for the leaves for Atlanta on the left side of the stem and the digits for the leaves for Philadelphia on the right side, as shown. See Figure 2–24.

Figure 2–24
Back-to-Back Stem and Leaf Plot for Example 2–15

Atlanta		Philadelphia
9 8 6	2	5
8 6 4 4 2 2 2 2 2 1	3	0 0 0 0 2 2 3 4 6 6 6 8 8 9 9
7 4 4 0 0	4	0 0 0 0
5 3 2 2 0 0	5	0 3 4 8
3 0	6	1
0	7	

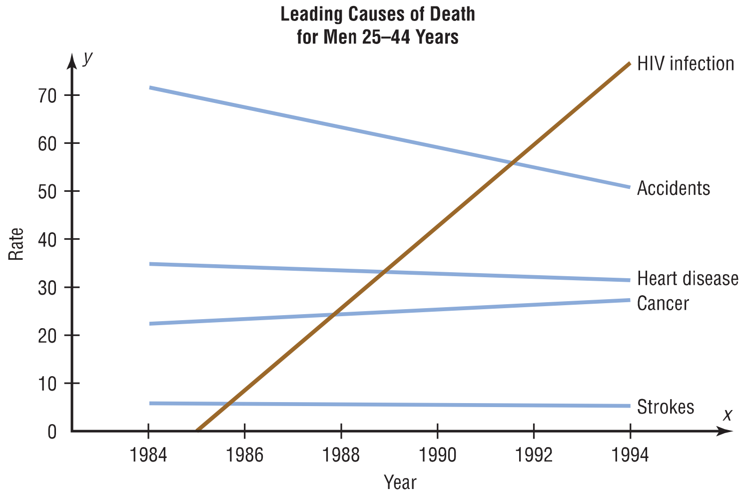
Step 3 Compare the distributions. The buildings in Atlanta have a large variation in the number of stories per building. Although both distributions are peaked in the 30- to 39-story class, Philadelphia has more buildings in this class. Atlanta has more buildings that have 40 or more stories than Philadelphia does.

Stem and leaf plots are part of the techniques called *exploratory data analysis.* More information on this topic is presented in Chapter 3.

Applying the Concepts 2–3

Leading Cause of Death

The following shows approximations of the leading causes of death among men ages 25–44 years. The rates are per 100,000 men. Answer the following questions about the graph.



1. What are the variables in the graph?
2. Are the variables qualitative or quantitative?
3. Are the variables discrete or continuous?
4. What type of graph was used to display the data?
5. Could a Pareto chart be used to display the data?
6. Could a pie chart be used to display the data?
7. List some typical uses for the Pareto chart.
8. List some typical uses for the time series chart.

See page 101 for the answers.

Exercises 2–3

1. Number of Hurricanes Construct a vertical bar chart for the total number of hurricanes by month from 1851 to 2008.

May	18
June	79
July	101
August	344
September	459
October	280
November	61

Source: National Hurricane Center.

2. Worldwide Sales of Fast Foods The worldwide sales (in billions of dollars) for several fast-food franchises for a specific year are shown. Construct a horizontal bar graph and a Pareto chart for the data.

Wendy's	$ 8.7
KFC	14.2
Pizza Hut	9.3
Burger King	12.7
Subway	10.0

Source: *Franchise Times.*

3. Calories Burned While Exercising Construct a Pareto chart for the following data on exercise.

	Calories burned per minute
Walking, 2 mph	2.8
Bicycling, 5.5 mph	3.2
Golfing	5.0
Tennis playing	7.1
Skiing, 3 mph	9.0
Running, 7 mph	14.5

Source: Physiology of Exercise.

4. Roller Coaster Mania The World Roller Coaster Census Report lists the following number of roller coasters on each continent. Represent the data graphically, using a Pareto chart and a horizontal bar graph.

Africa	17
Asia	315
Australia	22
Europe	413
North America	643
South America	45

Source: www.rcdb.com