

# 2

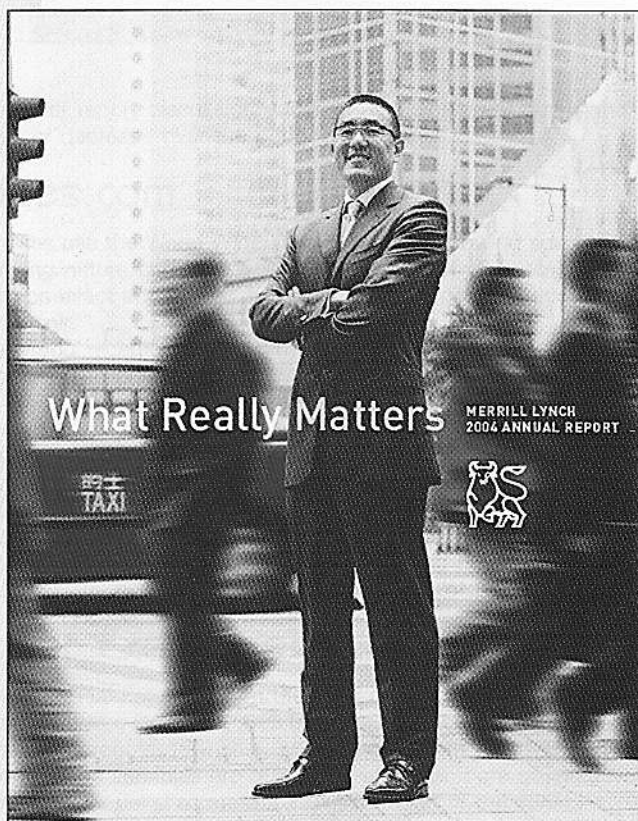
## GOALS

When you have completed this chapter you will be able to:

- 1 Organize qualitative data into a *frequency table*.
- 2 Present a frequency table as a *bar chart* or a *pie chart*.
- 3 Organize quantitative data into a *frequency distribution*.
- 4 Present a frequency distribution for quantitative data using *histograms*, *frequency polygons*, and *cumulative frequency polygons*.

## Describing Data:

### Frequency Tables, Frequency Distributions, and Graphic Presentation



Merrill Lynch recently completed a study of online investment portfolios for a sample of clients. For the 70 participants in the study, draw a histogram for the data. (See Exercise 39 and Goal 4.)

## Introduction

The highly competitive automotive retailing business has changed significantly over the past 5 years, due in part to consolidation by large, publicly owned dealership groups. Traditionally, a local family owned and operated the community dealership, which might have included one or two manufacturers, like Pontiac and GMC Trucks or Chrysler and the popular Jeep line. Recently, however, skillfully managed and well-financed companies have been acquiring local dealerships across large regions of the country. As these groups acquire the local dealerships, they often bring standard selling practices, common software and hardware technology platforms, and management reporting techniques. The goal is to provide an improved buying experience for the consumer, while increasing the profitability of the larger dealership organization. In many cases, in addition to reaping the financial benefits of selling the dealership, the family is asked to continue running the dealership on a daily basis. Today, it is common for these megadealerships to employ over 10,000 people, generate several billion dollars in annual sales, own more than 100 franchises, and be traded on the New York Stock Exchange or NASDAQ.

The consolidation has not come without challenges. With the acquisition of dealerships across the country, AutoUSA, one of the new megadealerships, now sells the inexpensive Korean import brands Kia and Hyundai, the high-line BMW and Mercedes Benz sedans, and a full line of Ford and Chevrolet cars and trucks.

Ms. Kathryn Ball is a member of the senior management team at AutoUSA. She is responsible for tracking and analyzing vehicle selling prices for AutoUSA. Kathryn would like to summarize vehicle selling prices with charts and graphs that she could review monthly. From these tables and charts, she wants to know the typical selling price as well as the lowest and highest prices. She is also interested in describing the demographics of the buyers. What are their ages? How many vehicles do they own? Do they want to buy or lease the vehicle?

Whitner Autoplex, which is located in Raytown, Missouri, is one of the AutoUSA dealerships. Whitner Autoplex includes Pontiac, GMC, and Buick franchises as well as a BMW store. General Motors is actively working with its dealer body to combine at one location several of its franchises, such as Chevrolet, Pontiac, or Cadillac. Combining franchises improves the floor traffic and a dealership has product offerings for all demographics. BMW, with its premium

brand and image, wants to move away from calling its locations dealerships, instead calling them stores. Instead of offering a traditional automobile dealership experience, BMW wants to be more like Nordstrom, a U.S. retail store selling fine quality apparel. Like Nordstrom, BMW wants to offer customers superior service, great products, and a unique personalized shopping experience.

Ms. Ball decided to collect data on three variables at Whitner Autoplex: selling price (\$000), buyer's age, and car type (domestic, coded as 1, or foreign, coded as 0). A portion of the data set is shown in the adjacent Excel output. The entire data set is available on the student CD (included with the book), at the McGraw-Hill website, and in Appendix A.5 at the end of the text.



Microsoft Excel - WHITNER-2005

1	Price	Price(\$000)	Age	Type
2	23187	23.187	48	0
3	23372	23.372	48	0
4	20454	20.454	40	1
5	23591	23.591	40	0
6	26551	26.551	48	1
7	27453	27.453	37	1
8	17266	17.266	32	1
9	18021	18.021	28	1
10	28583	28.583	38	1
11	30872	30.872	43	0
12	19507	19.507	32	0
13	23169	23.169	47	0
14	35851	35.851	56	0
15	19251	19.251	42	1
16	20947	20.947	28	1
17	24285	24.285	56	0
18	24324	24.324	50	1
19	24603	24.603	31	1
20	28870	28.87	51	1
21	15546	15.546	26	1
22	15335	15.335	25	1
23	19073	19.073	45	1

Ready

## Constructing a Frequency Table

Recall from Chapter 1 that techniques used to describe a set of data are called descriptive statistics. To put it another way, descriptive statistics organize data to show the general shape of the data and where values tend to concentrate and to expose extreme or unusual data values. The first procedure we discuss to organize and summarize a set of data is a **frequency table**.

**FREQUENCY TABLE** A grouping of qualitative data into mutually exclusive classes showing the number of observations in each class.

In Chapter 1, we distinguished between qualitative and quantitative variables. To review, a qualitative variable is nonnumeric, that is, it can only be classified into distinct categories. There is no particular order to these categories. Examples of qualitative data include political affiliation (Republican, Democrat, Independent), state of birth (Alabama, . . . , Wyoming), and method of payment for a purchase at Barnes and Noble (cash, check, or charge). On the other hand, quantitative variables are numerical in nature. Examples of quantitative data relating to college students include the price of their textbooks, their age, and hours they spend studying each week.

In the Whitner Autoplex data, Ms. Ball observed three variables for each vehicle sale: selling price, age of the buyer, and car type. Selling price and age are quantitative variables, but vehicle type is a qualitative measure with two values, domestic or foreign. Suppose that Ms. Ball wanted to summarize last month's sales using vehicle type.

To summarize this qualitative data, we classify the vehicles as either domestic (coded as 1) or foreign (coded as 0) and count the number in each class. We use vehicle type to develop a frequency table with two mutually exclusive (distinctive) classes. This means that a particular vehicle cannot belong to both classes. The vehicle is either domestic or foreign and can never be both domestic and foreign. This frequency table is shown in Table 2-1. The number of observations in each class is called the **class frequency**. So the class frequency for domestic vehicles sold is 50.

TABLE 2-1 Frequency Table for Vehicles Sold at Whitner Autoplex Last Month

Car Type	Number of Cars
Domestic	50
Foreign	30

## Relative Class Frequencies

You can convert class frequencies to **relative class frequencies** to show the fraction of the total number of observations in each class. So, a relative frequency captures the relationship between a class total and the total number of observations. In the vehicle sales example, we may want to know the percent of total cars sold that were domestic or foreign.

To convert a frequency distribution to a *relative* frequency distribution, each of the class frequencies is divided by the total number of observations. For example, 0.625, found by 50 divided by 80, is the fraction of domestic vehicles sold last month. The relative frequency distribution is shown in Table 2-2.

**TABLE 2-2** Relative Frequency Table of Vehicles Sold By Type At Whitner Autoplex Last Month

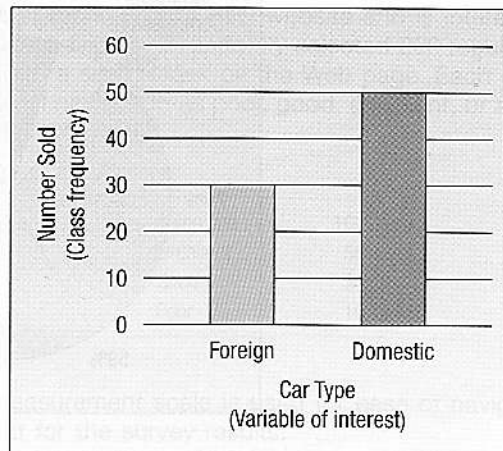
Vehicle Type	Number Sold	Relative Frequency
Domestic	50	0.625
Foreign	30	0.375
Total	80	1.000

## Graphic Presentation of Qualitative Data

The most common device to present a qualitative variable in graphic form is a **bar chart**. In most cases the horizontal axis shows the variable of interest and the vertical axis the amount, number, or fraction of each of the possible outcomes. A distinguishing characteristic of a bar chart is there is a distance or a gap between the bars. That is, because the variable of interest is qualitative the bars are not adjacent to each other. Thus, a bar chart graphically describes a frequency table using a series of uniformly wide rectangles, where the height of each rectangle is the class frequency.

**BAR CHART** A graph in which the classes are reported on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are proportional to the heights of the bars.

We use the Whitner Autoplex data as an example (Chart 2-1). The variable of interest is the vehicle type and the number of each type sold is the class frequency. We scale the vehicle type (domestic or foreign) on the horizontal axis and the number of each item on the vertical axis. The height of the bars, or rectangles, corresponds to the number of vehicles of each type sold. So for the number of foreign vehicles sold the height of the bar is 30. The order of foreign or domestic on the X-axis does not matter because the values of car type are qualitative.

**CHART 2-1** Vehicle Sold by Type Last Month at Whitner Autoplex

Another useful type of chart for depicting qualitative information is a **pie chart**.

**PIE CHART** A chart that shows the proportion or percent that each class represents of the total number of frequencies.

We explain the details of constructing a pie chart using the information in Table 2-3, which shows a breakdown of the expenses for the Ohio State Lottery in 2004.

TABLE 2-3 Ohio State Lottery Expenses in 2004

Use of Sales	Amount (\$ million)	Percent of Share
Prizes	1,276.0	59
Payments to Education	648.1	30
Bonuses/Commissions	132.8	6
Operating Expenses	97.7	5
Total	2,154.6	100

The first step to develop a pie chart is to record the percentages 0, 5, 10, 15, and so on evenly around the circumference of a circle (see Chart 2-2). To plot the 59 percent share awarded for prizes, draw a line from the center of the circle to 0 and another line from the center of the circle to 59 percent. The area in this "slice" represents the lottery proceeds that were awarded in prizes. Next, add the 59 percent of expenses awarded in prizes to the 30 percent payments to education; the result is 89 percent. Draw a line from the center of the circle to 89 percent, so the area between 59 percent and 89 percent depicts the payments made to education. Continuing, add the 6 percent for bonuses and commissions, which gives us a total of 95 percent. Draw a line from the center of the circle to 95, so the "slice" between 89 percent and 95 percent represents the payment of bonuses and commissions. The remaining 5 percent is for operating expenses.

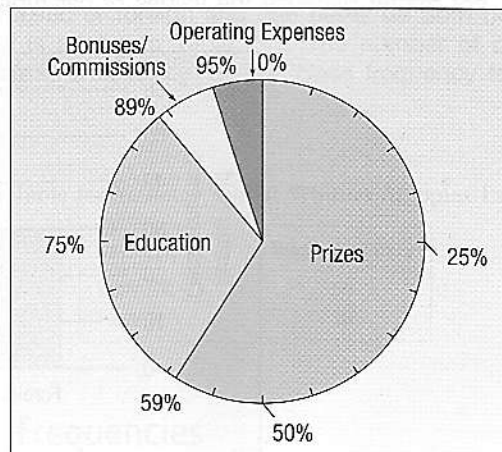
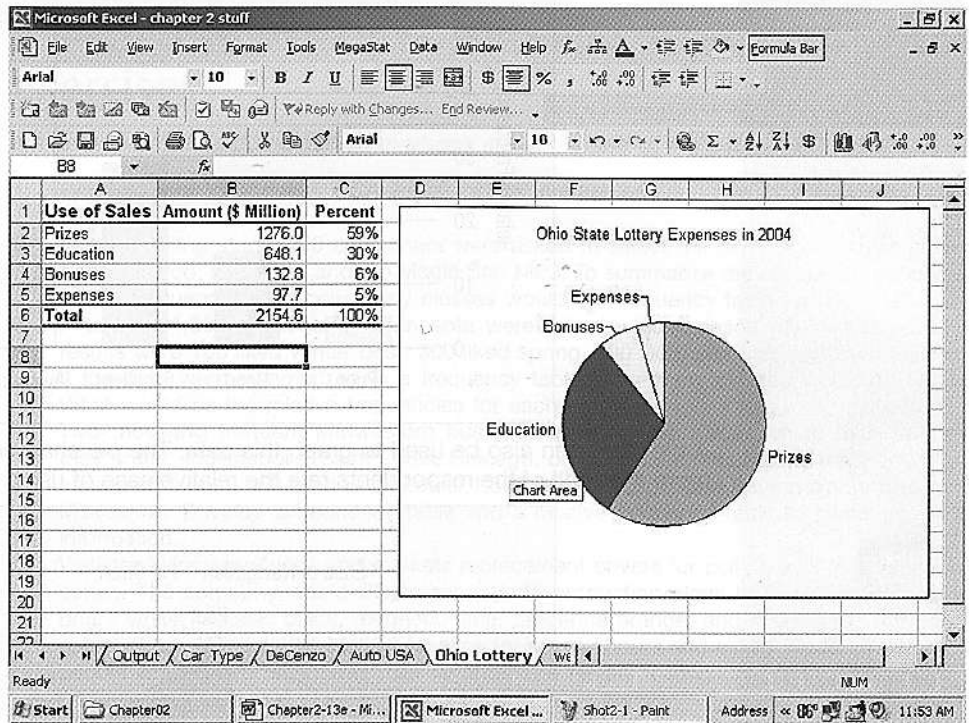


CHART 2-2 Pie Chart of Ohio Lottery Expenses in 2004

Because each slice of the pie represents the relative share of each component, we can easily compare them:

- The largest expense of the Ohio Lottery is for prizes.
- About one-third of the proceeds are transferred to education.
- Operating expenses account for only 5 percent of the proceeds.

The Excel system will develop a pie chart. See the following chart for the information in Table 2–3.



## Example

Skilodges.com is test marketing its new website and is interested in how easy its Web page design is to navigate. It randomly selected 200 regular Internet users and asked them to perform a search task on the Web page. Each person was asked to rate the relative ease of navigation as poor, good, excellent, or awesome. The results are shown in the following table:

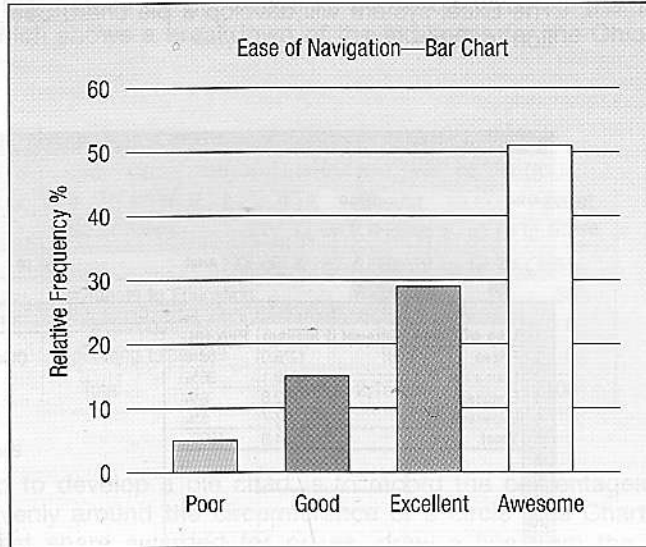
Awesome	102
Excellent	58
Good	30
Poor	10

1. What type of measurement scale is used for ease of navigation?
2. Draw a bar chart for the survey results.
3. Draw a pie chart for the survey results.

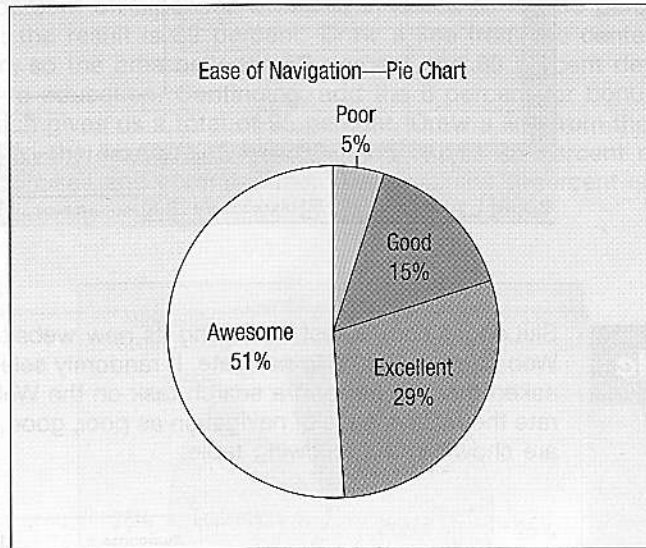
## Solution

The data are measured on an ordinal scale. That is, the scale is ranked in relative ease when moving from "poor" to "awesome." Also, the interval between each rating is unknown so it is impossible, for example, to conclude that a rating of good is twice the value of a poor rating.

We can use a bar chart to graph the data. The vertical scale shows the relative frequency and the horizontal shows the values of the ease of navigation scale.



A pie chart can also be used to graph this data. The pie chart emphasizes the fact the more than half of the respondents rate the relative ease of using the website awesome.



**Self-Review 2-1**

*The answers are at the end of the chapter.*



DeCenzo Specialty Food and Beverage Company has been serving a cola drink with an additional flavoring, Cola-Plus, that is very popular among its customers. The company is interested in customer preferences for Cola-Plus versus Coca-Cola, Pepsi, and a lemon-lime beverage. They ask 100 randomly sampled customers to take a taste test and select the beverage they preferred most. The results are shown in the following table:

Beverage	Number
Cola-Plus	40
Coca-Cola	25
Pepsi	20
Lemon-Lime	15
Total	100

- (a) Is the data qualitative or quantitative? Why?
- (b) What is the table called? What does it show?
- (c) Develop a bar chart to depict the information.
- (d) Develop a pie chart using the relative frequencies.

## Exercises

The answers to the odd-numbered exercises are at the end of the book.

1. Refer to your local paper, *USA Today*, or the Internet and find two examples of qualitative variables.
2. In a marketing study, 100 consumers were asked to select the best digital music player from the iPod, the iRiver, and the Magic Star MP3. To summarize the consumer responses with a frequency table, how many classes would the frequency table have?
3. A total of 1,000 residents in Minnesota were asked which season they preferred. The results were 100 liked winter best, 300 liked spring, 400 liked summer, and 200 liked fall. If the data were summarized in a frequency table, how many classes would be used? What would be the relative frequencies for each class?
4. Two thousand frequent Midwestern business travelers are asked which Midwest city they prefer: Indianapolis, Saint Louis, Chicago, or Milwaukee. The results were 100 liked Indianapolis best, 450 liked Saint Louis, 1,300 liked Chicago, and the remainder preferred Milwaukee. Develop a frequency table and a relative frequency table to summarize this information.
5. Wellstone, Inc., produces and markets replacement covers for cell phones in a variety of colors. The company would like to allocate its production plans to five different colors: bright white, metallic black, magnetic lime, tangerine orange, and fusion red. The company set up a kiosk in the Mall of America for several hours and asked randomly selected people which cover color was their favorite. The results follow:

Bright white	130
Metallic black	104
Magnetic lime	325
Tangerine orange	455
Fusion red	286

- a. What is the table called?
  - b. Draw a bar chart for the table.
  - c. Draw a pie chart.
  - d. If Wellstone, Inc., plans to produce 1 million cell phone covers, how many of each color should it produce?
6. A small business consultant is investigating the performance of several companies. The fourth quarter sales for last year (in thousands of dollars) for the selected companies were:

Corporation	Fourth-Quarter Sales (\$ thousands)
Hoden Building Products	\$ 1,645.2
J & R Printing, Inc.	4,757.0
Long Bay Concrete Construction	8,913.0
Mancell Electric and Plumbing	627.1
Maxwell Heating and Air Conditioning	24,612.0
Mizelle Roofing & Sheet Metals	191.9

The consultant wants to include a chart in his report comparing the sales of the six companies. Use a bar chart to compare the fourth-quarter sales of these corporations and write a brief report summarizing the bar chart.



## Constructing Frequency Distributions: Quantitative Data

In Chapter 1 and earlier in this chapter, we distinguished between qualitative and quantitative data. In the previous section using the Whitner Autoplex data, we summarized the qualitative variable, vehicle type, using a frequency table, a relative frequency table, a bar chart, and a pie chart.

The Whitner Autoplex data also includes two quantitative variables: selling price and age of the buyer. Suppose Ms. Ball wants to summarize last month's sales using selling price. We can describe selling price using a **frequency distribution**.

**FREQUENCY DISTRIBUTION** A grouping of data into mutually exclusive classes showing the number of observations in each class.

How do we develop a frequency distribution? The first step is to tally the data into a table that shows the classes and the number of observations in each class. The steps in constructing a frequency distribution are best described by an example. Remember, our goal is to construct tables, charts, and graphs that will quickly reveal the concentration and shape of the data.

### Example

We return to the situation where Ms. Kathryn Ball of AutoUSA wants tables, charts, and graphs to show the typical selling price on various dealer lots. Table 2-4 reports only the price of the 80 vehicles sold last month at Whitner Autoplex. What is the *typical* selling price? What is the *highest* selling price? What is the *lowest* selling price? Around what value do the selling prices tend to cluster?

**TABLE 2-4** Prices of Vehicles Sold Last Month at Whitner Autoplex

\$23,197	\$23,372	\$20,454	\$23,591	\$26,651	\$27,453	\$17,266
18,021	28,683	30,872	19,587	23,169	35,851	19,251
20,047	24,285	24,324	24,609	28,670	15,546	15,935
19,873	25,251	25,277	28,034	24,533	27,443	19,889
20,004	17,357	20,155	19,688	23,657	26,613	20,895
20,203	23,765	25,783	26,661	32,277	20,642	21,981
24,052	25,799	15,794	18,263	35,925	17,399	17,968
20,356	21,442	21,722	19,331	22,817	19,766	20,633
20,962	22,845	26,285	27,896	29,076	32,492	18,890
21,740	22,374	24,571	25,449	28,337	20,642	23,613
24,220	30,655	22,442	17,891	20,818	26,237	20,445
21,556	21,639	24,296				

Lowest

Highest

### Solution

We refer to the unorganized information in Table 2-4 as **raw data** or **ungrouped data**. With a little searching, we can find the lowest selling price (\$15,546) and the highest selling price (\$35,925), but that is about all. It is difficult to determine a typical selling price. It is also difficult to visualize where the selling prices tend to cluster. The raw data are more easily interpreted if organized into a frequency distribution.

**Step 1: Decide on the number of classes.** The goal is to use just enough groupings or **classes** to reveal the shape of the distribution. Some judgment is needed here. Too many classes or too few classes might not reveal the basic shape of the data set. In the vehicle selling price example, three classes would not give much insight into the pattern of the data (see Table 2-5).

A useful recipe to determine the number of classes ( $k$ ) is the “2 to the  $k$  rule.” This guide suggests you select the smallest number ( $k$ ) for

The steps for organizing data into a frequency distribution.



### Statistics in Action

In 1788, James Madison, John Jay, and Alexander Hamilton anonymously published a series of essays entitled *The Federalist*. These Federalist papers were an attempt to convince the people of New York that they should ratify the Constitution. In the course of history, the authorship of most of these papers became known, but 12 remained contested. Through the use of statistical analysis, and particularly the study of the frequency of the use of various words, we can now conclude that James Madison is the likely author of the 12 papers. In fact, the statistical evidence that Madison is the author is overwhelming.

TABLE 2-5 An Example of Too Few Classes

Vehicle Selling Price (\$)	Number of Vehicles
15,000 up to 24,000	48
24,000 up to 33,000	30
33,000 up to 42,000	2
Total	80

the number of classes such that  $2^k$  (in words, 2 raised to the power of  $k$ ) is greater than the number of observations ( $n$ ).

In the Whitner Autoplex example, there were 80 vehicles sold. So  $n = 80$ . If we try  $k = 6$ , which means we would use 6 classes, then  $2^6 = 64$ , somewhat less than 80. Hence, 6 is not enough classes. If we let  $k = 7$ , then  $2^7 = 128$ , which is greater than 80. So the recommended number of classes is 7.

**Step 2: Determine the class interval or width.** Generally the **class interval** or **width** should be the same for all classes. The classes all taken together must cover at least the distance from the lowest value in the data up to the highest value. Expressing these words in a formula:

$$i \geq \frac{H - L}{k}$$

where  $i$  is the class interval,  $H$  is the highest observed value,  $L$  is the lowest observed value, and  $k$  is the number of classes.

In the Whitner Autoplex case, the lowest value is \$15,546 and the highest value is \$35,925. If we need 7 classes, the interval should be at least  $(\$35,925 - \$15,546)/7 = \$2,911$ . In practice this interval size is usually rounded up to some convenient number, such as a multiple of 10 or 100. The value of \$3,000 might readily be used in this case.

Unequal class intervals present problems in graphically portraying the distribution and in doing some of the computations, which we will see in later chapters. Unequal class intervals, however, may be necessary in certain situations to avoid a large number of empty, or almost empty, classes. Such is the case in Table 2-6. The Internal Revenue

TABLE 2-6 Adjusted Gross Income for Individuals Filing Income Tax Returns

Adjusted Gross Income		Number of Returns (in thousands)
No adjusted gross income		178.2
\$ 1 up to	\$ 5,000	1,204.6
5,000 up to	10,000	2,595.5
10,000 up to	15,000	3,142.0
15,000 up to	20,000	3,191.7
20,000 up to	25,000	2,501.4
25,000 up to	30,000	1,901.6
30,000 up to	40,000	2,502.3
40,000 up to	50,000	1,426.8
50,000 up to	75,000	1,476.3
75,000 up to	100,000	338.8
100,000 up to	200,000	223.3
200,000 up to	500,000	55.2
500,000 up to	1,000,000	12.0
1,000,000 up to	2,000,000	5.1
2,000,000 up to	10,000,000	3.4
10,000,000 or more		0.6

Service used unequal-sized class intervals to report the adjusted gross income on individual tax returns. Had they used an equal-sized interval of, say, \$1,000, more than 1,000 classes would have been required to describe all the incomes. A frequency distribution with 1,000 classes would be difficult to interpret. In this case the distribution is easier to understand in spite of the unequal classes. Note also that the number of income tax returns or “frequencies” is reported in thousands in this particular table. This also makes the information easier to understand.

**Step 3: Set the individual class limits.** State clear class limits so you can put each observation into only one category. This means you must avoid overlapping or unclear class limits. For example, classes such as “\$1,300–\$1,400” and “\$1,400–\$1,500” should not be used because it is not clear whether the value \$1,400 is in the first or second class. Classes stated as “\$1,300–\$1,400” and “\$1,500–\$1,600” are frequently used, but may also be confusing without the additional common convention of rounding all data at or above \$1,450 up to the second class and data below \$1,450 down to the first class. In this text we will generally use the format \$1,300 up to \$1,400 and \$1,400 up to \$1,500 and so on. With this format it is clear that \$1,399 goes into the first class and \$1,400 in the second.

Because we round the class interval up to get a convenient class size, we cover a larger than necessary range. For example, 7 classes of width \$3,000 in the Whitner Autoplex case result in a range of  $7(\$3,000) = \$21,000$ . The actual range is \$20,379, found by  $\$35,925 - \$15,546$ . Comparing that value to \$21,000 we have an excess of \$621. Because we need to cover only the distance  $(H - L)$ , it is natural to put approximately equal amounts of the excess in each of the two tails. Of course, we should also select convenient class limits. A guideline is to make the lower limit of the first class a multiple of the class interval. Sometimes this is not possible, but the lower limit should at least be rounded. So here are the classes we could use for this data.

\$15,000 up to 18,000
18,000 up to 21,000
21,000 up to 24,000
24,000 up to 27,000
27,000 up to 30,000
30,000 up to 33,000
33,000 up to 36,000

**Step 4: Tally the vehicle selling prices into the classes.** To begin, the selling price of the first vehicle in Table 2–4 is \$23,197. It is tallied in the \$21,000 up to \$24,000 class. The second selling price in the first column of Table 2–4 is \$18,021. It is tallied in the \$18,000 up to \$21,000 class. The other selling prices are tallied in a similar manner. When all the selling prices are tallied, the table would appear as:

Class	Tallies
\$15,000 up to \$18,000	
\$18,000 up to \$21,000	
\$21,000 up to \$24,000	
\$24,000 up to \$27,000	
\$27,000 up to \$30,000	
\$30,000 up to \$33,000	
\$33,000 up to \$36,000	

**Step 5: Count the number of items in each class.** The number of observations in each class is called the **class frequency**. In the \$15,000 up to \$18,000

class there are 8 observations, and in the \$18,000 up to \$21,000 class there are 23 observations. Therefore, the class frequency in the first class is 8 and the class frequency in the second class is 23. There is a total of 80 observations or frequencies in the entire set of data.

Often it is useful to express the data in thousands, or some convenient units, rather than the actual data. Table 2-7, for example, reports the vehicle selling prices in thousands of dollars, rather than dollars.

**TABLE 2-7** Frequency Distribution of Selling Prices at Whitner Autoplex Last Month

Selling Prices (\$ thousands)	Frequency
15 up to 18	8
18 up to 21	23
21 up to 24	17
24 up to 27	18
27 up to 30	8
30 up to 33	4
33 up to 36	2
Total	<u>80</u>

Now that we have organized the data into a frequency distribution, we can summarize the pattern in the selling prices of the vehicles for the AutoUSA lot of Whitner Autoplex in Raytown, Missouri. Observe the following:

1. The selling prices ranged from about \$15,000 up to about \$36,000.
2. The selling prices are concentrated between \$18,000 and \$27,000. A total of 58, or 72.5 percent, of the vehicles sold within this range.
3. The largest concentration, or highest frequency, is in the \$18,000 up to \$21,000 class. The middle of this class is \$19,500. So we say that a typical selling price is \$19,500.

By presenting this information to Ms. Ball, we give her a clear picture of the distribution of selling prices for last month.

We admit that arranging the information on selling prices into a frequency distribution does result in the loss of some detailed information. That is, by organizing the data into a frequency distribution, we cannot pinpoint the exact selling price, such as \$23,197 or \$26,237. Further, we cannot tell that the actual selling price for the least expensive vehicle was \$15,546 and for the most expensive \$35,925. However, the lower limit of the first class and the upper limit of the largest class convey essentially the same meaning. Likely, Ms. Ball will make the same judgment if she knows the lowest price is about \$15,000 that she will if she knows the exact price is \$15,546. The advantages of condensing the data into a more understandable and organized form more than offset this disadvantage.

### Self-Review 2-2

The commissions earned for the first quarter of last year by the 11 members of the sales staff at Master Chemical Company are:

\$1,650	\$1,475	\$1,510	\$1,670	\$1,595	\$1,760	\$1,540	\$1,495	\$1,590	\$1,625	\$1,510
---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

- (a) What are the values such as \$1,650 and \$1,475 called?
- (b) Using \$1,400 up to \$1,500 as the first class, \$1,500 up to \$1,600 as the second class, and so forth, organize the quarterly commissions into a frequency distribution.
- (c) What are the numbers in the right column of your frequency distribution called?
- (d) Describe the distribution of quarterly commissions, based on the frequency distribution. What is the largest concentration of commissions earned? What is the smallest, and the largest? What is the typical amount earned?



## Class Intervals and Class Midpoints

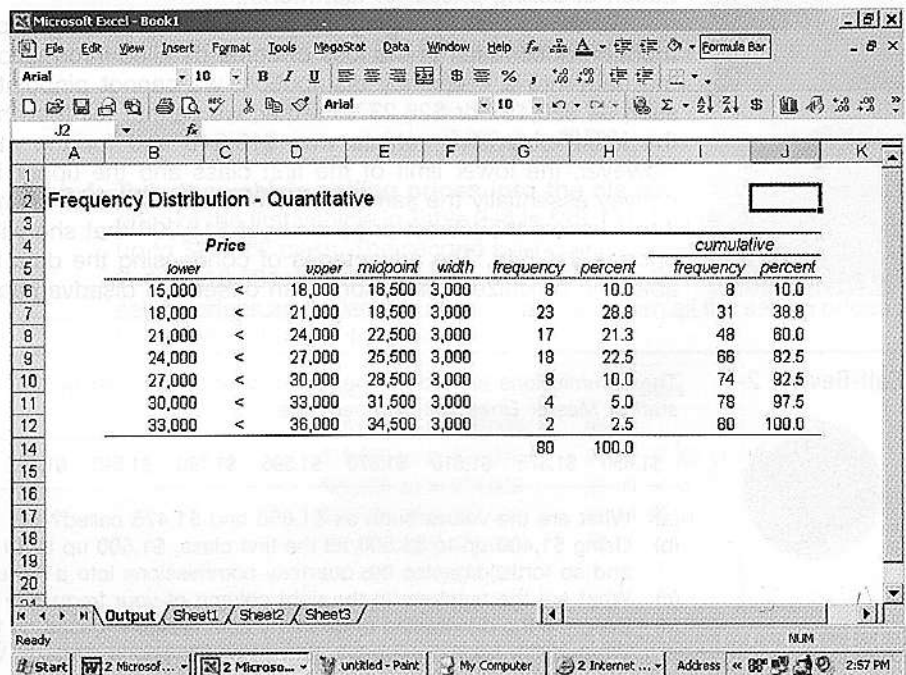
We will use two other terms frequently: **class midpoint** and **class interval**. The midpoint is halfway between the lower limits of two consecutive classes. It is computed by adding the lower limits of consecutive classes and dividing the result by 2. Referring to Table 2-7, for the first class the lower class limit is \$15,000 and the next limit is \$18,000. The class midpoint is \$16,500, found by  $(\$15,000 + \$18,000)/2$ . The midpoint of \$16,500 best represents, or is typical of, the selling price of the vehicles in that class.

To determine the class interval, subtract the lower limit of the class from the lower limit of the next class. The class interval of the vehicle selling price data is \$3,000, which we find by subtracting the lower limit of the first class, \$15,000, from the lower limit of the next class; that is,  $\$18,000 - \$15,000 = \$3,000$ . You can also determine the class interval by finding the difference between consecutive midpoints. The midpoint of the first class is \$16,500 and the midpoint of the second class is \$19,500. The difference is \$3,000.

## A Software Example

As we mentioned in Chapter 1, there are many software packages that perform statistical calculations. Throughout this text we will show the output from Microsoft Excel; from MegaStat, which is an add-in to Microsoft Excel; and from MINITAB. The commands necessary to generate the outputs are given in the **Software Commands** section at the end of each chapter.

The following is a frequency distribution, produced by MegaStat, showing the prices of the 80 vehicles sold last month at the Whitner Autoplex lot in Raytown, Missouri. The form of the output is somewhat different than the frequency distribution of Table 2-7, but the overall conclusions are the same.



Frequency Distribution - Quantitative									
Price		midpoint	width	frequency	percent	cumulative			
lower	upper					frequency	percent		
15,000	<	18,000	3,000	8	10.0	8	10.0		
18,000	<	21,000	3,000	23	28.8	31	38.8		
21,000	<	24,000	3,000	17	21.3	48	60.0		
24,000	<	27,000	3,000	18	22.5	66	82.5		
27,000	<	30,000	3,000	8	10.0	74	92.5		
30,000	<	33,000	3,000	4	5.0	78	97.5		
33,000	<	36,000	3,000	2	2.5	80	100.0		
				80	100.0				



## Self-Review 2-3



Barry Bonds of the San Francisco Giants established a new single season home run record by hitting 73 home runs during the 2001 season. The longest of these home runs traveled 488 feet and the shortest 320 feet. You need to construct a frequency distribution of these home run lengths.

- How many classes would you use?
- What class interval would you suggest?
- What actual classes would you suggest?

## Relative Frequency Distribution

A relative frequency distribution converts the frequency to a percent.

It may be desirable as we did earlier with qualitative data to convert class frequencies to relative class frequencies to show the fraction of the total number of observations in each class. In our vehicle sales example, we may want to know what percent of the vehicle prices are in the \$21,000 up to \$24,000 class. In another study, we may want to know what percent of the employees used 5 up to 10 personal leave days last year.

To convert a frequency distribution to a *relative* frequency distribution, each of the class frequencies is divided by the total number of observations. From the distribution of vehicle selling prices (Table 2-7, where the selling price is reported in thousands of dollars), the relative frequency for the \$15,000 up to \$18,000 class is 0.10, found by dividing 8 by 80. That is, the price of 10 percent of the vehicles sold at Whitner Autoplex is between \$15,000 and \$18,000. The relative frequencies for the remaining classes are shown in Table 2-8.

TABLE 2-8 Relative Frequency Distribution of the Prices of Vehicles Sold Last Month at Whitner Autoplex

Selling Price (\$ thousands)	Frequency	Relative Frequency	Found by
15 up to 18	8	0.1000	8/80
18 up to 21	23	0.2875	23/80
21 up to 24	17	0.2125	17/80
24 up to 27	18	0.2250	18/80
27 up to 30	8	0.1000	8/80
30 up to 33	4	0.0500	4/80
33 up to 36	2	0.0250	2/80
Total	80	1.0000	

## Self-Review 2-4



Refer to Table 2-8, which shows the relative frequency distribution for the vehicles sold last month at Whitner Autoplex.

- How many vehicles sold for \$18,000 up to \$21,000?
- What percent of the vehicles sold for a price between \$18,000 and \$21,000?
- What percent of the vehicles sold for \$30,000 or more?

## Exercises

- A set of data consists of 38 observations. How many classes would you recommend for the frequency distribution?
- A set of data consists of 45 observations between \$0 and \$29. What size would you recommend for the class interval?
- A set of data consists of 230 observations between \$235 and \$567. What class interval would you recommend?

10. A set of data contains 53 observations. The lowest value is 42 and the largest is 129. The data are to be organized into a frequency distribution.
- How many classes would you suggest?
  - What would you suggest as the lower limit of the first class?
11. Wachesaw Manufacturing, Inc., produced the following number of units in the last 16 days.

27	27	27	28	27	25	25	28
26	28	26	28	31	30	26	26

The information is to be organized into a frequency distribution.

- How many classes would you recommend?
  - What class interval would you suggest?
  - What lower limit would you recommend for the first class?
  - Organize the information into a frequency distribution and determine the relative frequency distribution.
  - Comment on the shape of the distribution.
12. The Quick Change Oil Company has a number of outlets in the metropolitan Seattle area. The daily number of oil changes at the Oak Street outlet in the past 20 days are:

65	98	55	62	79	59	51	90	72	56
70	62	66	80	94	79	63	73	71	85

The data are to be organized into a frequency distribution.

- How many classes would you recommend?
  - What class interval would you suggest?
  - What lower limit would you recommend for the first class?
  - Organize the number of oil changes into a frequency distribution.
  - Comment on the shape of the frequency distribution. Also determine the relative frequency distribution.
13. The manager of the BiLo Supermarket in Mt. Pleasant, Rhode Island, gathered the following information on the number of times a customer visits the store during a month. The responses of 51 customers were:

5	3	3	1	4	4	5	6	4	2	6	6	6	7	1
1	14	1	2	4	4	4	5	6	3	5	3	4	5	6
8	4	7	6	5	9	11	3	12	4	7	6	5	15	1
1	10	8	9	2	12									

- Starting with 0 as the lower limit of the first class and using a class interval of 3, organize the data into a frequency distribution.
  - Describe the distribution. Where do the data tend to cluster?
  - Convert the distribution to a relative frequency distribution.
14. The food services division of Cedar River Amusement Park, Inc., is studying the amount families who visit the amusement park spend per day on food and drink. A sample of 40 families who visited the park yesterday revealed they spent the following amounts:

\$77	\$18	\$63	\$84	\$38	\$54	\$50	\$59	\$54	\$56	\$36	\$26	\$50	\$34	\$44
41	58	58	53	51	62	43	52	53	63	62	62	65	61	52
60	60	45	66	83	71	63	58	61	71					

- Organize the data into a frequency distribution, using seven classes and 15 as the lower limit of the first class. What class interval did you select?
- Where do the data tend to cluster?
- Describe the distribution.
- Determine the relative frequency distribution.

## Graphic Presentation of a Frequency Distribution

Sales managers, stock analysts, hospital administrators, and other busy executives often need a quick picture of the trends in sales, stock prices, or hospital costs. These trends can often be depicted by the use of charts and graphs. Three charts that will help portray a frequency distribution graphically are the histogram, the frequency polygon, and the cumulative frequency polygon.

### Histogram

A **histogram** for a frequency distribution based on quantitative data is very similar to the bar chart showing the distribution of qualitative data. The classes are marked on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are represented by the heights of the bars. However, there is one important difference based on the nature of the data. Quantitative data is usually measured using scales that are continuous, not discrete. Therefore, the horizontal axis represents all possible values, and the bars are drawn adjacent to each other to show the continuous nature of the data.

**HISTOGRAM** A graph in which the classes are marked on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are represented by the heights of the bars and the bars are drawn adjacent to each other.

We summarized the selling prices—a continuous variable—of the 80 vehicles sold last month at Whitner Autoplex with a frequency distribution. We construct a histogram to illustrate this frequency distribution.

### Example

Below is the frequency distribution.

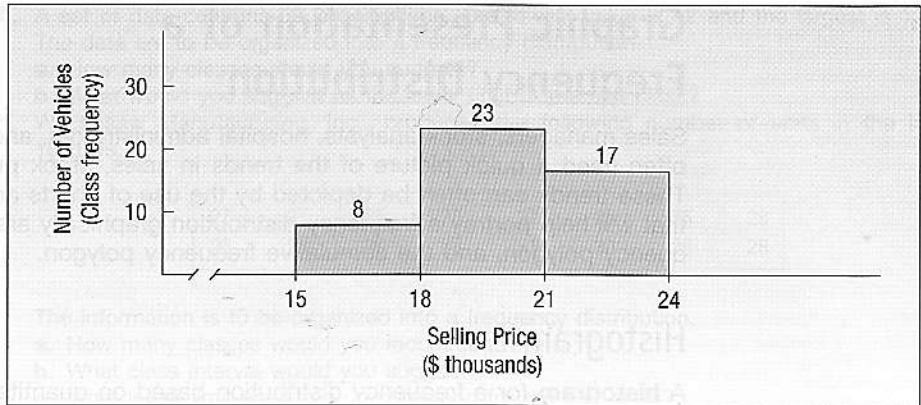
Selling Prices (\$ thousands)	Frequency
15 up to 18	8
18 up to 21	23
21 up to 24	17
24 up to 27	18
27 up to 30	8
30 up to 33	4
33 up to 36	2
Total	80

Construct a histogram. What conclusions can you reach based on the information presented in the histogram?

### Solution

The class frequencies are scaled along the vertical axis (*Y*-axis) and either the class limits or the class midpoints along the horizontal axis. To illustrate the construction of the histogram, the first three classes are shown in Chart 2-3.

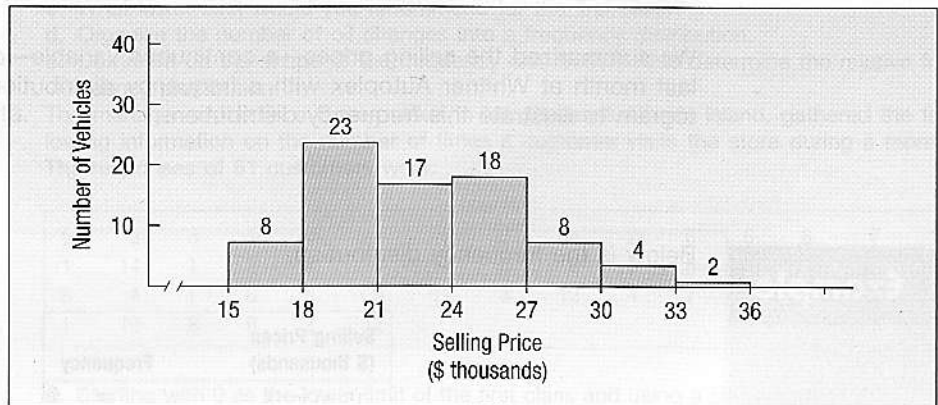




**CHART 2-3** Construction of a Histogram

From Chart 2-3 we note that there are eight vehicles in the \$15,000 up to \$18,000 class. Therefore, the height of the column for that class is 8. There are 23 vehicles in the \$18,000 up to \$21,000 class. So, logically, the height of that column is 23. The height of the bar represents the number of observations in the class.

This procedure is continued for all classes. The complete histogram is shown in Chart 2-4. Note that there is no space between the bars. This is a feature of the histogram. Why is this so? Because the variable plotted on the horizontal axis is quantitative and of the interval, or in this case the ratio, scale of measurement. In bar charts, as described previously, the vertical bars are separated.



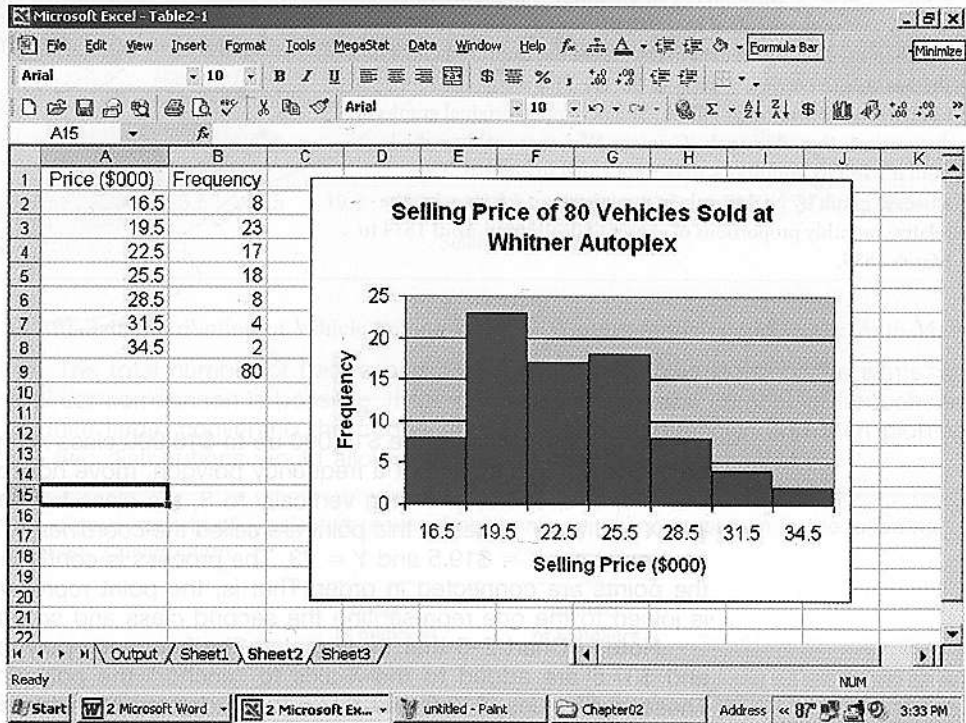
**CHART 2-4** Histogram of the Selling Prices of 80 Vehicles at Whitner Autoplex

From the histogram in Chart 2-4, we conclude:

1. The lowest selling price is about \$15,000, and the highest is about \$36,000.
2. The largest class frequency is the \$18,000 up to \$21,000 class. A total of 23 of the 80 vehicles sold are within this price range.
3. Fifty-eight of the vehicles, or 72.5 percent, had a selling price between \$18,000 and \$27,000.

Thus, the histogram provides an easily interpreted visual representation of a frequency distribution. We should also point out that we would have reached the same conclusions and the shape of the histogram would have been the same had we used a relative frequency distribution instead of the actual frequencies. That is, if we had used the relative frequencies of Table 2-8, found on page 33, we would have had a histogram of the same shape as Chart 2-4. The only difference is that the vertical axis would have been reported in percent of vehicles instead of the number of vehicles.

We use the Microsoft Excel system to produce the histogram for the Whitner Autoplex vehicle sales data (which is shown on page 28). Note that class midpoints are used as the labels for the classes. The software commands to create this output are given in the **Software Commands** section at the end of the chapter.



## Frequency Polygon

In a frequency polygon the class midpoints are connected with a line segment.

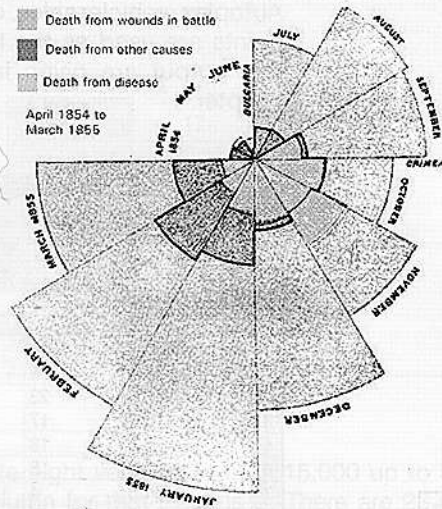
A **frequency polygon** also shows the shape of a distribution and is similar to a histogram. It consists of line segments connecting the points formed by the intersections of the class midpoints and the class frequencies. The construction of a frequency polygon is illustrated in Chart 2-5 (on page 38). We use the vehicle prices for the cars sold last month at Whitner Autoplex. The midpoint of each class is scaled on the X-axis and the class frequencies on the Y-axis. Recall that the class midpoint is the value at the center of a class and represents the typical values in that class. The class frequency is the number of observations in a particular class. The vehicle selling prices at Whitner Autoplex are:

Selling Price (\$ thousands)	Midpoint	Frequency
15 up to 18	16.5	8
18 up to 21	19.5	23
21 up to 24	22.5	17
24 up to 27	25.5	18
27 up to 30	28.5	8
30 up to 33	31.5	4
33 up to 36	34.5	2
Total		80



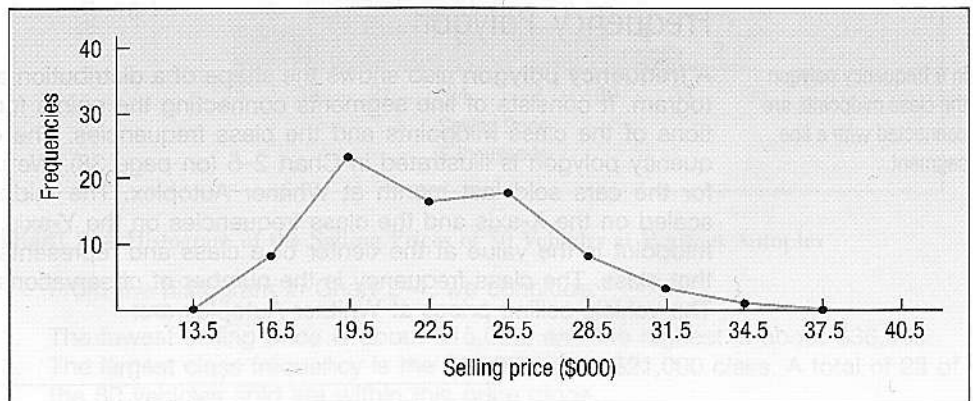
### Statistics in Action

Florence Nightingale is known as the founder of the nursing profession. However, she also saved many lives by using statistical analysis. When she encountered an unsanitary condition or an undersupplied hospital, she improved the conditions and then used statistical data to document the improvement. Thus, she was able to convince others of the need for medical reform, particularly in the area of sanitation. She developed original graphs to demonstrate that, during the Crimean War, more soldiers died from unsanitary conditions than were killed in combat. The adjacent graph by Nightingale is a polar-area graph showing the relative monthly proportions of causes of death from April 1854 to March 1855.



As noted previously, the \$15,000 up to \$18,000 class is represented by the midpoint \$16,500. To construct a frequency polygon, move horizontally on the graph to the midpoint, \$16.5, and then vertically to 8, the class frequency, and place a dot. The  $X$  and the  $Y$  values of this point are called the *coordinates*. The coordinates of the next point are  $X = \$19.5$  and  $Y = 23$ . The process is continued for all classes. Then the points are connected in order. That is, the point representing the lowest class is joined to the one representing the second class and so on.

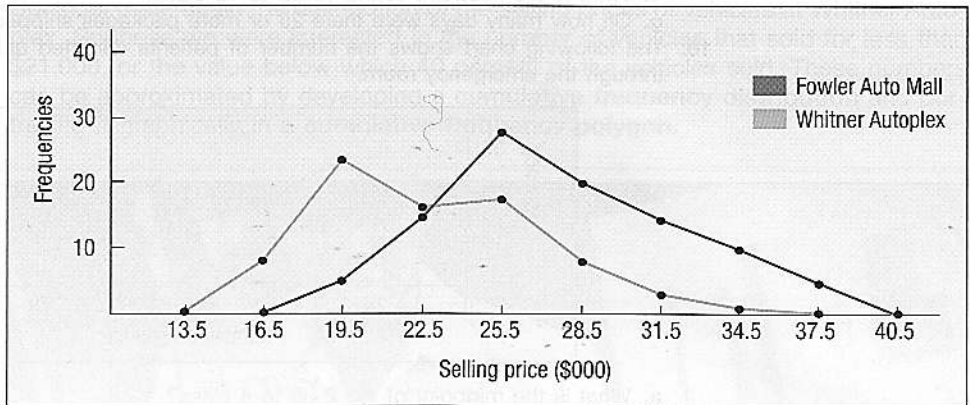
Note in Chart 2-5 that, to complete the frequency polygon, midpoints of \$13.5 and \$37.5 are added to the  $X$ -axis to “anchor” the polygon at zero frequencies. These two values, \$13.5 and \$37.5, were derived by subtracting the class interval of \$3.0 from the lowest midpoint (\$16.5) and by adding \$3.0 to the highest midpoint (\$34.5) in the frequency distribution.



**CHART 2-5** Frequency Polygon of the Selling Prices of 80 Vehicles at Whitner Autoplex

Both the histogram and the frequency polygon allow us to get a quick picture of the main characteristics of the data (highs, lows, points of concentration, etc.). Although the two representations are similar in purpose, the histogram has the advantage of depicting each class as a rectangle, with the height of the rectangular bar representing the number in each class. The frequency polygon, in turn, has an advantage over the histogram. It allows us to compare directly two or more frequency distributions. Suppose Ms. Ball of AutoUSA wants to compare the Whitner Autoplex lot in Raytown, Missouri, with a similar lot, Fowler Auto Mall in Grayling, Michigan. To do this, two

frequency polygons are constructed, one on top of the other, as in Chart 2-6. It is clear from Chart 2-6 that the typical vehicle selling price is higher at the Fowler Auto Mall.

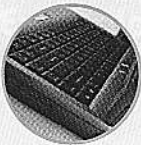


**CHART 2-6** Distribution of Vehicle Selling Prices at Whitner Autoplex and Fowler Auto Mall

The total number of frequencies at the two dealerships is about the same, so a direct comparison is possible. If the difference in the total number of frequencies is quite large, converting the frequencies to relative frequencies and then plotting the two distributions would allow a clearer comparison.

### Self-Review 2-5

The annual imports of a selected group of electronic suppliers are shown in the following frequency distribution.

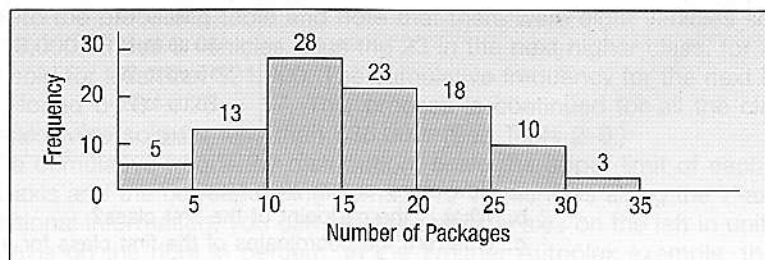


Imports (\$ millions)	Number of Suppliers
2 up to 5	6
5 up to 8	13
8 up to 11	20
11 up to 14	10
14 up to 17	1

- Portray the imports as a histogram.
- Portray the imports as a relative frequency polygon.
- Summarize the important facets of the distribution (such as classes with the highest and lowest frequencies).

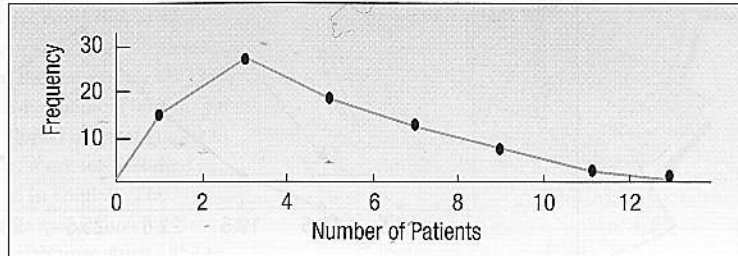
## Exercises

15. Molly's Candle Shop has several retail stores in the coastal areas of North and South Carolina. Many of Molly's customers ask her to ship their purchases. The following chart shows the number of packages shipped per day for the last 100 days.



- What is this chart called?
- What is the total number of frequencies?
- What is the class interval?

- d. What is the class frequency for the 10 up to 15 class?  
 e. What is the relative frequency of the 10 up to 15 class?  
 f. What is the midpoint of the 10 up to 15 class?  
 g. On how many days were there 25 or more packages shipped?
16. The following chart shows the number of patients admitted daily to Memorial Hospital through the emergency room.



- a. What is the midpoint of the 2 up to 4 class?  
 b. How many days were 2 up to 4 patients admitted?  
 c. Approximately how many days were studied?  
 d. What is the class interval?  
 e. What is this chart called?
17. The following frequency distribution reports the number of frequent flier miles, reported in thousands, for employees of Brumley Statistical Consulting, Inc., during the first quarter of 2007.

Frequent Flier Miles (000)	Number of Employees
0 up to 3	5
3 up to 6	12
6 up to 9	23
9 up to 12	8
12 up to 15	2
Total	50

- a. How many employees were studied?  
 b. What is the midpoint of the first class?  
 c. Construct a histogram.  
 d. A frequency polygon is to be drawn. What are the coordinates of the plot for the first class?  
 e. Construct a frequency polygon.  
 f. Interpret the frequent flier miles accumulated using the two charts.
18. Ecommerce.com, a large Internet retailer, is studying the lead time (elapsed time between when an order is placed and when it is filled) for a sample of recent orders. The lead times are reported in days.

Lead Time (days)	Frequency
0 up to 5	6
5 up to 10	7
10 up to 15	12
15 up to 20	8
20 up to 25	7
Total	40

- a. How many orders were studied?  
 b. What is the midpoint of the first class?  
 c. What are the coordinates of the first class for a frequency polygon?  
 d. Draw a histogram.  
 e. Draw a frequency polygon.  
 f. Interpret the lead times using the two charts.

## Cumulative Frequency Distributions

Consider once again the distribution of the selling prices of vehicles at Whitner Autoplex. Suppose we were interested in the number of vehicles that sold for less than \$21,000, or the value below which 40 percent of the vehicles sold. These numbers can be approximated by developing a **cumulative frequency distribution** and portraying it graphically in a **cumulative frequency polygon**.



### Example

The frequency distribution of the vehicle selling prices at Whitner Autoplex is repeated from Table 2-7.

Selling Price (\$ thousands)	Frequency
15 up to 18	8
18 up to 21	23
21 up to 24	17
24 up to 27	18
27 up to 30	8
30 up to 33	4
33 up to 36	2
Total	80

Construct a cumulative frequency polygon. Fifty percent of the vehicles were sold for less than what amount? Twenty-five of the vehicles were sold for less than what amount?

### Solution

As the name implies, a cumulative frequency distribution and a cumulative frequency polygon require *cumulative frequencies*. To construct a cumulative frequency distribution, refer to the preceding table and note that there were eight vehicles sold for less than \$18,000. Those 8 vehicles, plus the 23 in the next higher class, for a total of 31, were sold for less than \$21,000. The cumulative frequency for the next higher class is 48, found by  $8 + 23 + 17$ . This process is continued for all the classes. All the vehicles were sold for less than \$36,000. (See Table 2-9.)

To plot a cumulative frequency distribution, scale the upper limit of each class along the X-axis and the corresponding cumulative frequencies along the Y-axis. To provide additional information, you can label the vertical axis on the left in units and the vertical axis on the right in percent. In the Whitner Autoplex example, the vertical axis on the left is labeled from 0 to 80 and on the right from 0 to 100 percent. The value of 50 percent corresponds to 40 vehicles sold.

TABLE 2-9 Cumulative Frequency Distribution for Vehicle Selling Price

Selling Price (\$ thousands)	Frequency	Cumulative Frequency	Found by
15 up to 18	8	8	
18 up to 21	23	31	← 8 + 23
21 up to 24	17	48	8 + 23 + 17
24 up to 27	18	66	8 + 23 + 17 + 18
27 up to 30	8	74	⋮
30 up to 33	4	78	
33 up to 36	2	80	
Total	<u>80</u>		

To begin the plotting, 8 vehicles sold for less than \$18,000, so the first plot is at  $X = 18$  and  $Y = 8$ . The coordinates for the next plot are  $X = 21$  and  $Y = 31$ . The rest of the points are plotted and then the dots connected to form the chart below.

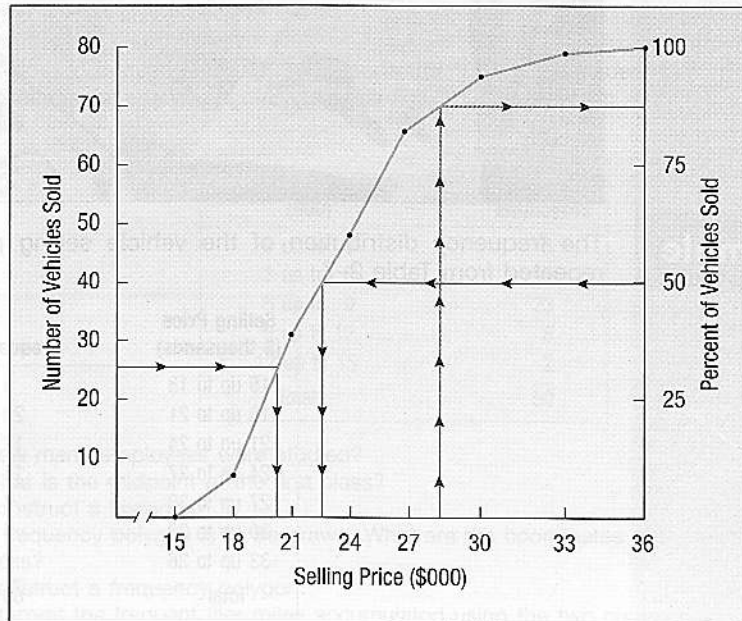


CHART 2-7 Cumulative Frequency Distribution for Vehicle Selling Price

To find the selling price below which half the cars sold, we draw a horizontal line from the 50 percent mark on the right-hand vertical axis over to the polygon, then drop down to the  $X$ -axis and read the selling price. The value on the  $X$ -axis is about 22.5, so we estimate that 50 percent of the vehicles sold for less than \$22,500.

To find the price below which 25 of the vehicles sold, we locate the value of 25 on the left-hand vertical axis. Next, we draw a horizontal line from the value of 25 to the polygon, and then drop down to the  $X$ -axis and read the price. It is about 20.5, so we estimate that 25 of the vehicles sold for less than \$20,500. We can also make estimates of the percent of vehicles that sold for less than a particular amount. To explain, suppose we want to estimate the percent of vehicles that sold for less than \$28,500. We begin by locating the value of 28.5 on the  $X$ -axis, move vertically to the polygon, and then horizontally to the vertical axis on the right. The value is about 87 percent, so we conclude that 87 percent of the vehicles sold for less than \$28,500.

## Self-Review 2-6

A sample of the hourly wages of 15 employees at Home Depot in Brunswick, Georgia, was organized into the following table.

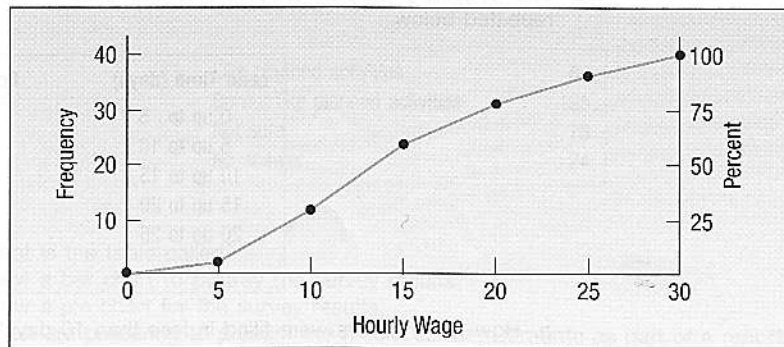


Hourly Wages	Number of Employees
\$ 8 up to \$10	3
10 up to 12	7
12 up to 14	4
14 up to 16	1

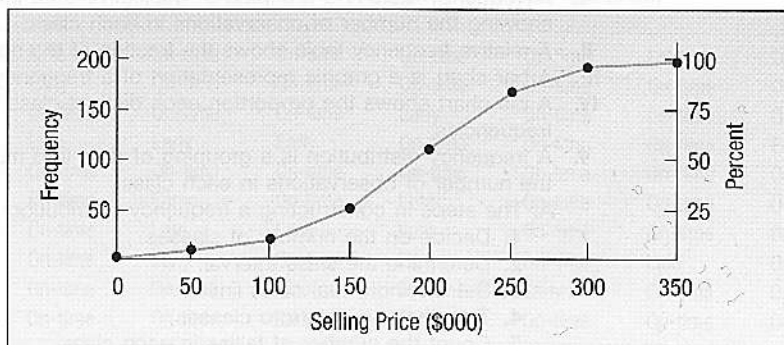
- What is the table called?
- Develop a cumulative frequency distribution and portray the distribution in a cumulative frequency polygon.
- On the basis of the cumulative frequency polygon, how many employees earn \$11 an hour or less? Half of the employees earn an hourly wage of how much or more? Four employees earn how much or less?

## Exercises

19. The following chart shows the hourly wages of a sample of certified welders in the Atlanta, Georgia area.



- How many welders were studied?
  - What is the class interval?
  - About how many welders earn less than \$10.00 per hour?
  - About 75 percent of the welders make less than what amount?
  - Ten of the welders studied made less than what amount?
  - What percent of the welders make less than \$20.00 per hour?
20. The following chart shows the selling price (\$000) of houses sold in the Billings, Montana area.





- a. How many homes were studied?
  - b. What is the class interval?
  - c. One hundred homes sold for less than what amount?
  - d. About 75 percent of the homes sold for less than what amount?
  - e. Estimate the number of homes in the \$150,000 up to \$200,000 class.
  - f. About how many homes sold for less than \$225,000?
21. The frequency distribution representing the number of frequent flier miles accumulated by employees at Brumley Statistical Consulting Company is repeated from Exercise 17.

Frequent Flier Miles (000)	Frequency
0 up to 3	5
3 up to 6	12
6 up to 9	23
9 up to 12	8
12 up to 15	2
Total	<u>50</u>

- a. How many employees accumulated less than 3,000 miles?
  - b. Convert the frequency distribution to a cumulative frequency distribution.
  - c. Portray the cumulative distribution in the form of a cumulative frequency polygon.
  - d. Based on the cumulative frequency polygon, about 75 percent of the employees accumulated how many miles or less?
22. The frequency distribution of order lead time at Ecommerce.com from Exercise 18 is repeated below.

Lead Time (days)	Frequency
0 up to 5	6
5 up to 10	7
10 up to 15	12
15 up to 20	8
20 up to 25	7
Total	<u>40</u>

- a. How many orders were filled in less than 10 days? In less than 15 days?
- b. Convert the frequency distribution to a cumulative frequency distribution.
- c. Develop a cumulative frequency polygon.
- d. About 60 percent of the orders were filled in less than how many days?

## Chapter Summary

- I. A frequency table is a grouping of qualitative data into mutually exclusive classes showing the number of observations in each class.
- II. A relative frequency table shows the fraction of the number of frequencies in each class.
- III. A bar chart is a graphic representation of a frequency table.
- IV. A pie chart shows the proportion each distinct class represents of the total number of frequencies.
- V. A frequency distribution is a grouping of data into mutually exclusive classes showing the number of observations in each class.
  - A. The steps in constructing a frequency distribution are:
    1. Decide on the number of classes.
    2. Determine the class interval.
    3. Set the individual class limits.
    4. Tally the raw data into classes.
    5. Count the number of tallies in each class.

- B. The class frequency is the number of observations in each class.
  - C. The class interval is the difference between the limits of two consecutive classes.
  - D. The class midpoint is halfway between the limits of consecutive classes.
- VI. A relative frequency distribution shows the percent of observations in each class.
- VII. There are three methods for graphically portraying a frequency distribution.
- A. A histogram portrays the number of frequencies in each class in the form of a rectangle.
  - B. A frequency polygon consists of line segments connecting the points formed by the intersection of the class midpoint and the class frequency.
  - C. A cumulative frequency distribution shows the number or percent of observations below given values.

## Chapter Exercises

23. Describe the similarities and differences of qualitative and quantitative variables. Be sure to include:
- a. What level of measurement is required for each variable type?
  - b. Can both types be used to describe both samples and populations?
24. Describe the similarities and differences of a frequency table and a frequency distribution. Be sure to include which requires qualitative data and which requires quantitative data.
25. Alexandra Damonte will be building a new resort in Myrtle Beach, South Carolina. She must decide how to design the resort based on the type of activities that the resort will offer to its customers. A recent poll of 300 potential customers showed the following results about customers' preferences for planned resort activities:

Like planned activities	63
Do not like planned activities	135
Not sure	78
No answer	24

- a. What is the table called?
  - b. Draw a bar chart to portray the survey results.
  - c. Draw a pie chart for the survey results.
  - d. If you are preparing to present the results to Ms. Damonte as part of a report, which graph would you prefer to show? Why?
26. Speedy Swift is a package delivery service that serves the greater Atlanta, Georgia metropolitan area. To maintain customer loyalty, one of Speedy Swift's performance objectives is on-time delivery. To monitor its performance, each delivery is measured on the following scale: early (package delivered before the promised time), on-time (package delivered within 5 minutes of the promised time), late (package delivered more than 5 minutes past the promised time), lost (package never delivered). Speedy Swift's objective is to deliver 99 percent of all packages either early or on-time. Another objective is to never lose a package.

Speedy collected the following data for last month's performance:

On-time	On-time	Early	Late	On-time	On-time	On-time	On-time	Late	On-time
Early	On-time	On-time	Early	On-time	On-time	On-time	On-time	On-time	On-time
Early	On-time	Early	On-time	On-time	On-time	Early	On-time	On-time	On-time
Early	On-time	On-time	Late	Early	Early	On-time	On-time	On-time	Early
On-time	Late	Late	On-time	On-time	On-time	On-time	On-time	On-time	On-time
On-time	Late	Early	On-time	Early	On-time	Lost	On-time	On-time	On-time
Early	Early	On-time	On-time	Late	Early	Lost	On-time	On-time	On-time
On-time	On-time	Early	On-time	Early	On-time	Early	On-time	Late	On-time
On-time	Early	On-time	On-time	On-time	Late	On-time	Early	On-time	On-time
On-time	On-time	On-time	On-time	On-time	Early	Early	On-time	On-time	On-time

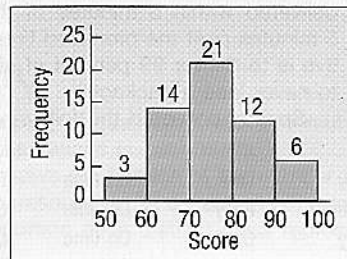
- a. What scale is used to measure delivery performance? What kind of variable is delivery performance?
  - b. Construct a frequency table for delivery performance for last month.
  - c. Construct a relative frequency table for delivery performance last month.
  - d. Construct a bar chart of the frequency table for delivery performance for last month.
  - e. Construct a pie chart of on-time delivery performance for last month.
  - f. Analyze the data summaries and write an evaluation of last month's delivery performance as it relates to Speedy Swift's performance objectives. Write a general recommendation for further analysis.
27. A data set consists of 83 observations. How many classes would you recommend for a frequency distribution?
  28. A data set consists of 145 observations that range from 56 to 490. What size class interval would you recommend?
  29. The following is the number of minutes to commute from home to work for a group of automobile executives.

28	25	48	37	41	19	32	26	16	23	23	29	36
31	26	21	32	25	31	43	35	42	38	33	28	

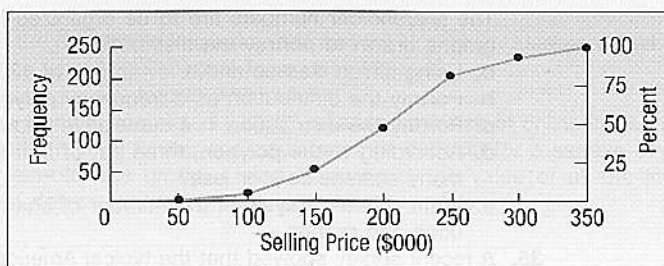
- a. How many classes would you recommend?
  - b. What class interval would you suggest?
  - c. What would you recommend as the lower limit of the first class?
  - d. Organize the data into a frequency distribution.
  - e. Comment on the shape of the frequency distribution.
30. The following data give the weekly amounts spent on groceries for a sample of households.

\$271	\$363	\$159	\$ 76	\$227	\$337	\$295	\$319	\$250
279	205	279	266	199	177	162	232	303
192	181	321	309	246	278	50	41	335
116	100	151	240	474	297	170	188	320
429	294	570	342	279	235	434	123	325

- a. How many classes would you recommend?
  - b. What class interval would you suggest?
  - c. What would you recommend as the lower limit of the first class?
  - d. Organize the data into a frequency distribution.
31. The following histogram shows the scores on the first exam for a statistics class.



- a. How many students took the exam?
  - b. What is the class interval?
  - c. What is the class midpoint for the first class?
  - d. How many students earned a score of less than 70?
32. The following chart summarizes the selling price of homes sold last month in the Sarasota, Florida, area.



- What is the chart called?
  - How many homes were sold during the last month?
  - What is the class interval?
  - About 75 percent of the houses sold for less than what amount?
  - One hundred seventy-five of the homes sold for less than what amount?
33. A chain of sport shops catering to beginning skiers, headquartered in Aspen, Colorado, plans to conduct a study of how much a beginning skier spends on his or her initial purchase of equipment and supplies. Based on these figures, it wants to explore the possibility of offering combinations, such as a pair of boots and a pair of skis, to induce customers to buy more. A sample of cash register receipts revealed these initial purchases:

\$140	\$ 82	\$265	\$168	\$ 90	\$114	\$172	\$230	\$142
86	125	235	212	171	149	156	162	118
139	149	132	105	162	126	216	195	127
161	135	172	220	229	129	87	128	126
175	127	149	126	121	118	172	126	

- Arrive at a suggested class interval. Use six classes, and let the lower limit of the first class be \$70.
  - What would be a better class interval?
  - Organize the data into a frequency distribution using a lower limit of \$80.
  - Interpret your findings.
34. Following is the number of shareholders for a selected group of large companies (in thousands):

Company	Number of Shareholders (thousands)	Company	Number of Shareholders (thousands)
Southwest Airlines	144	Standard Oil (Indiana)	173
General Public Utilities	177	Home Depot	195
Occidental Petroleum	266	Detroit Edison	220
Middle South Utilities	133	Eastman Kodak	251
DaimlerChrysler	209	Dow Chemical	137
Standard Oil of California	264	Pennsylvania Power	150
Bethlehem Steel	160	American Electric Power	262
Long Island Lighting	143	Ohio Edison	158
RCA	246	Transamerica Corporation	162
Greyhound Corporation	151	Columbia Gas System	165
Pacific Gas & Electric	239	International Telephone & Telegraph	223
Niagara Mohawk Power	204	Union Electric	158
E. I. du Pont de Nemours	204	Virginia Electric and Power	162
Westinghouse Electric	195	Public Service Electric & Gas	225
Union Carbide	176	Consumers Power	161
BankAmerica	175		
Northeast Utilities	200		

The shareholder numbers are to be organized into a frequency distribution and several graphs drawn to portray the distribution.

- Using seven classes and a lower limit of 130, construct a frequency distribution.
  - Portray the distribution as a frequency polygon.
  - Portray the distribution in a cumulative frequency polygon.
  - According to the polygon, three out of four (75 percent) of the companies have how many shareholders or less?
  - Write a brief analysis of the number of shareholders based on the frequency distribution and graphs.
35. A recent survey showed that the typical American car owner spends \$2,950 per year on operating expenses. Below is a breakdown of the various expenditure items. Draw an appropriate chart to portray the data and summarize your findings in a brief report.

Expenditure Item	Amount
Fuel	\$ 603
Interest on car loan	279
Repairs	930
Insurance and license	646
Depreciation	492
Total	<u>\$2,950</u>

36. Midland National Bank selected a sample of 40 student checking accounts. Below are their end-of-the-month balances.

\$404	\$ 74	\$234	\$149	\$279	\$215	\$123	\$ 55	\$ 43	\$321
87	234	68	489	57	185	141	758	72	863
703	125	350	440	37	252	27	521	302	127
968	712	503	489	327	608	358	425	303	203

- Tally the data into a frequency distribution using \$100 as a class interval and \$0 as the starting point.
  - Draw a cumulative frequency polygon.
  - The bank considers any student with an ending balance of \$400 or more a "preferred customer." Estimate the percentage of preferred customers.
  - The bank is also considering a service charge to the lowest 10 percent of the ending balances. What would you recommend as the cutoff point between those who have to pay a service charge and those who do not?
37. Residents of the state of South Carolina earned a total of \$69.5 billion in 2005 in adjusted gross income. Seventy-three percent of the total was in wages and salaries; 11 percent in dividends, interest, and capital gains; 8 percent in IRAs and taxable pensions; 3 percent in business income pensions; 2 percent in Social Security, and the remaining 3 percent from other sources. Develop a pie chart depicting the breakdown of adjusted gross income. Write a paragraph summarizing the information.
38. A recent study of home technologies reported the number of hours of personal computer usage per week for a sample of 60 persons. Excluded from the study were people who worked out of their home and used the computer as a part of their work.

9.3	5.3	6.3	8.8	6.5	0.6	5.2	6.6	9.3	4.3
6.3	2.1	2.7	0.4	3.7	3.3	1.1	2.7	6.7	6.5
4.3	9.7	7.7	5.2	1.7	8.5	4.2	5.5	5.1	5.6
5.4	4.8	2.1	10.1	1.3	5.6	2.4	2.4	4.7	1.7
2.0	6.7	1.1	6.7	2.2	2.6	9.8	6.4	4.9	5.2
4.5	9.3	7.9	4.6	4.3	4.5	9.2	8.5	6.0	8.1

- a. Organize the data into a frequency distribution. How many classes would you suggest? What value would you suggest for a class interval?
- b. Draw a histogram. Interpret your result.
39. Merrill Lynch recently completed a study regarding the size of online investment portfolios (stocks, bonds, mutual funds, and certificates of deposit) for a sample of clients in the 40- to 50-year-old age group. Listed following is the value of all the investments in thousands of dollars for the 70 participants in the study.

\$669.9	\$ 7.5	\$ 77.2	\$ 7.5	\$125.7	\$516.9	\$ 219.9	\$645.2
301.9	235.4	716.4	145.3	26.6	187.2	315.5	89.2
136.4	616.9	440.6	408.2	34.4	296.1	185.4	526.3
380.7	3.3	363.2	51.9	52.2	107.5	82.9	63.0
228.6	308.7	126.7	430.3	82.0	227.0	321.1	403.4
39.5	124.3	118.1	23.9	352.8	156.7	276.3	23.5
31.3	301.2	35.7	154.9	174.3	100.6	236.7	171.9
221.1	43.4	212.3	243.3	315.4	5.9	1,002.2	171.7
295.7	437.0	87.8	302.1	268.1	899.5		

- a. Organize the data into a frequency distribution. How many classes would you suggest? What value would you suggest for a class interval?
- b. Draw a histogram. Interpret your result.
40. In spring 2005, a total of 5.9 percent of the prime time viewing audience watched shows on ABC, 7.6 percent watched shows on CBS, 5.5 percent on Fox, 6.0 percent on NBC, 2.0 percent on Warner Brothers, and 2.2 percent on UPN. A total of 70.8 percent of the audience watched shows on other cable networks, such as CNN and ESPN. You can find the latest information on TV viewing from the following website: <http://tv.zap2it.com/news/ratings>. Develop a pie chart or a bar chart to depict this information. Write a paragraph summarizing your findings.
41. The American Heart Association reported the following percentage breakdown of expenses. Draw a pie chart depicting the information. Interpret.

Category	Percent
Research	32.3
Public Health Education	23.5
Community Service	12.6
Fund Raising	12.1
Professional and Educational Training	10.9
Management and General	8.6

42. Annual revenues, by type of tax, for the state of Georgia are as follows. Develop an appropriate chart or graph and write a brief report summarizing the information.

Type of Tax	Amount (\$000)
Sales	\$2,812,473
Income (Individual)	2,732,045
License	185,198
Corporate	525,015
Property	22,647
Death and Gift	37,326
Total	\$6,314,704

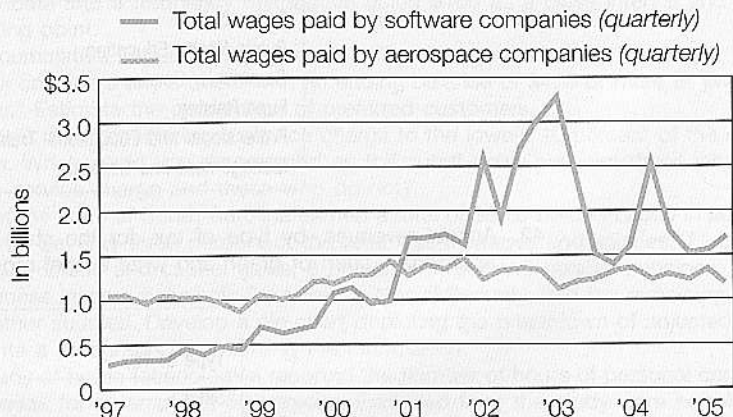
43. Annual imports from selected Canadian trading partners are listed below for the year 2005. Develop an appropriate chart or graph and write a brief report summarizing the information.

Partner	Annual Imports (\$ millions)
Japan	\$9,550
United Kingdom	4,556
South Korea	2,441
China	1,182
Australia	618

44. Farming has changed from the early 1900s. In the early 20th century, machinery gradually replaced animal power. For example, in 1910 U.S. farms used 24.2 million horses and mules and only about 1,000 tractors. By 1960, 4.6 million tractors were used and only 3.2 million horses and mules. In 1920 there were over 6 million farms in the United States. Today there are less than 2 million. Listed below is the number of farms, in thousands, for each of the 50 states. Write a paragraph summarizing your findings.

47	1	8	46	76	26	4	3	39	45
4	21	80	63	100	65	91	29	7	15
7	52	87	39	106	25	55	2	3	8
14	38	59	33	76	71	37	51	1	24
35	86	185	13	7	43	36	20	79	9

45. One of the most popular candies in the United States is M&M's, which are produced by the Mars Company. In the beginning M&M's were all brown; more recently they were produced in red, green, blue, orange, brown, and yellow. You can read about the history of the product, find ideas for baking, purchase the candies in the colors of your school or favorite team, and learn the percent of each color in the standard bags at <http://global.mms.com/us/about/products/milkchocolate/>. Recently the purchase of a 14-ounce bag of M&M's Plain had 444 candies with the following breakdown by color: 130 brown, 98 yellow, 96 red, 35 orange, 52 blue, and 33 green. Develop a chart depicting this information and write a paragraph summarizing the results.
46. The following graph shows the total wages paid by software and aircraft companies in the state of Washington from 1997 until 2005. Write a brief report summarizing this information.



47. A pie chart shows the market shares of cola products. The "slice" for Pepsi-Cola has a central angle of 90 degrees. What is its market share?

## exercises.com



48. Monthly and year-to-date truck sales are available at the website: <http://www.pickuptruck.com>. Go to this site and search under **News** to obtain the most recent information on sales. Make a pie chart or a bar chart showing the most recent information. What is the best selling truck? What are the four or five best selling trucks? What is their market share? You may wish to group some of the trucks into a category called "Other" to get a better picture of market share. Comment on your findings.

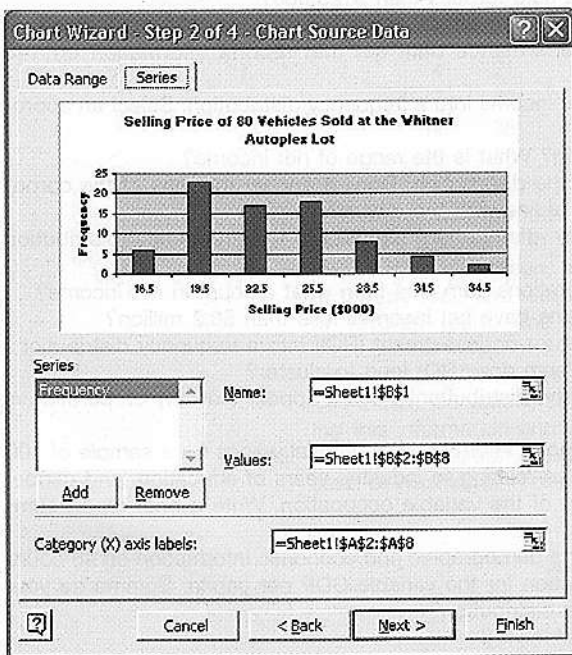
## Data Set Exercises

49. Refer to the Real Estate data at the end of the book, which reports information on homes sold in the Denver, Colorado, area during the last year.
- Organize the data on number of bedrooms into a frequency distribution.
    - What is the typical number of bedrooms?
    - What are the fewest and the most number of bedrooms offered in the market?
  - Select an appropriate class interval and organize the selling prices into a frequency distribution.
    - Around what values do the data tend to cluster?
    - What is the largest selling price? What is the smallest selling price?
  - Draw a cumulative frequency distribution based on the frequency distribution developed in part (b).
    - How many homes sold for less than \$200,000?
    - Estimate the percent of the homes that sold for more than \$220,000.
    - What percent of the homes sold for less than \$125,000?
  - Write a report summarizing the selling prices of the homes.
50. Refer to the Global Financial Performance data set that reports information on 148 corporations.
- Organize the information on net income into a frequency distribution. Select an appropriate class interval.
    - What is a typical net income? What is the range of net income?
    - Comment on the shape of the distribution. Does it appear that any of the corporation net incomes are out of line with the others?
  - Draw a cumulative frequency distribution based on the frequency distribution developed in part (a).
    - Eighty percent of the corporations earn less than what amount in net income?
    - About how many corporations have net incomes less than \$8.2 million?
  - Organize the information on return-on-investment (ROI) into a frequency distribution.
    - What is the typical ROI? Where does ROI tend to cluster?
    - Comment on the shape of the distribution. Does it appear that any corporation is out of line with the others?
51. Refer to the Wage data, which reports information on annual wages for a sample of 100 workers. Also included are variables relating to industry, years of education, and gender for each worker. Draw a bar chart of the variable occupation. Write a brief report summarizing your findings.
52. Refer to the CIA data, which reports demographic and economic information on 46 countries. Develop a frequency distribution for the variable GDP per capita. Summarize your findings. What is the shape of the distribution?



## Software Commands

- The Excel commands for the pie chart on page 25 are:
  - Set cell A1 as the active cell and type the words *Use of Sales*. In cells A2 through A5 type *Prizes, Education, Bonuses, and Expense*.
  - Set B1 as the active cell and type *Amount (\$ Millions)* and in cells B2 through B5 enter the data.
  - From the Toolbar select **Chart Wizard**. Select **Pie** as the type of chart, select the chart type in the upper left corner, and then click on **Next**.
  - For the **Data Range** type A1:B5, indicate that the data are in **Columns**, and then click on **Next**.
  - Click on the chart title area and type *Ohio Lottery Expenses 2004*. Then click **Finish**.



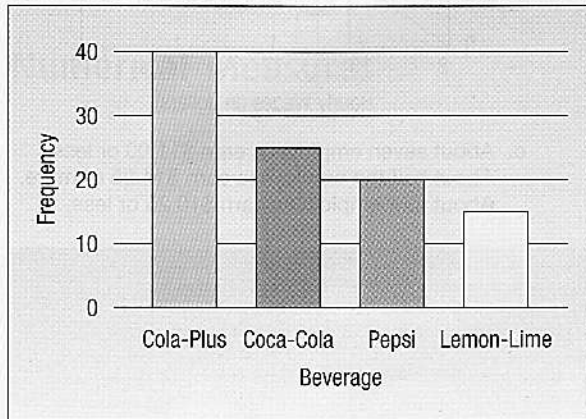
- The MegaStat commands for the frequency distribution on page 32 are:
  - Open Excel and from the CD provided, select **Data Sets**, and select the Excel format; go to Chapter 2, and select **Whitner-2005**. Click on **MegaStat, Frequency Distribution**, and select **Quantitative**.
  - In the dialog box, input the range from A1:A81, select **Equal width intervals**, use 3,000 as the interval width, 15,000 as the lower boundary of the first interval, select **Histogram**, and then click **OK**.

- The Excel commands for the histogram on page 37 are:
  - In cell A1 indicate that the column of data is the selling price and in B1 that it is the frequency. In cells A2 to A8 insert the midpoints of the selling prices in \$000. In B2 to B8 record the class frequencies.
  - With your mouse arrow on A1, click and drag to highlight the cells A1:B8.
  - From the Toolbar select **Chart Wizard**, under **Chart type** select **Column**, under **Chart sub-type** select the vertical bars in the upper left corner, and finally click on **Next** in the lower right corner.
  - At the top select the **Series** tab. Under the Series list box, **Price** is highlighted. Select **Remove**. (We do not want Price to be a part of the values.) At the bottom, in the **Category (X) axis labels** text box, click the icon at the far right. Put your cursor on cell A2, click and drag to cell A8. There will be a running box around cells A2 to A8. Touch the **Enter** key. This identifies the column of **Prices** as the X-axis labels. Click on **Next**.
  - At the top of the dialog box click on **Titles**. Click in the **Chart title** box and key in *Selling Price of 80 Vehicles Sold at the Whitner Autoplex Lot*. Tab to the **Category (X) axis** box and key in the label *Selling Price in (\$000)*. Tab to the **Category (Y) axis** box and key in *Frequency*. At the top select **Legend** and remove the check from the **Show legend** box. Click **Finish**.
  - To make the chart larger, click on the middle handle of the top line and drag the line to row 1. Make sure the handles show on the chart box. With your right mouse button, click on one of the columns. Select **Format Data Series**. At the top select the **Options** tab. In the **Gap width** text box, click the down arrow until the gap width reads 0, and click **OK**.

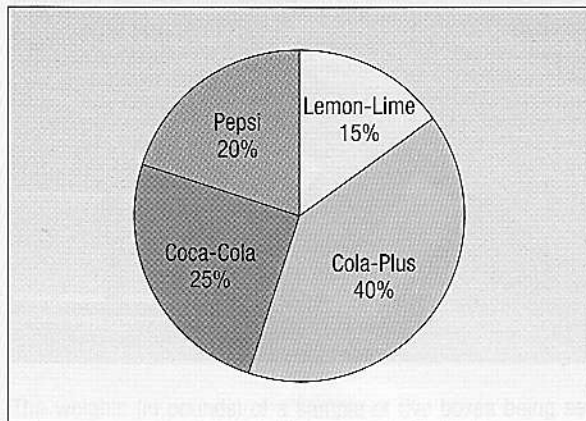


## Chapter 2 Answers to Self-Review

- 2-1 a. Qualitative data, because the customers' response to the taste test is the name of a beverage.  
 b. Frequency table. It shows the number of people who prefer each beverage.  
 c.



d.



- 2-2 a. The raw data or ungrouped data.

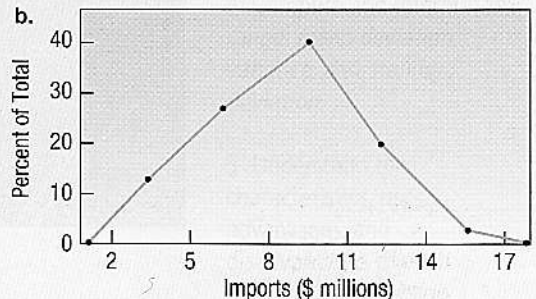
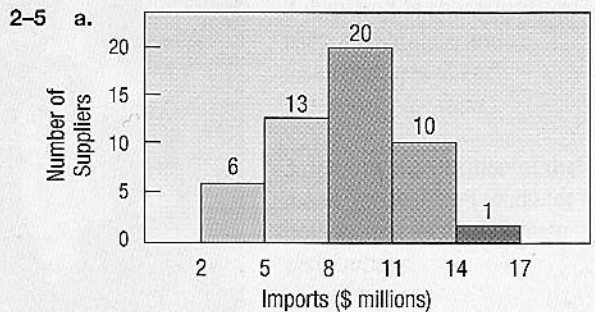
b.

Commission	Number of Salespeople
\$1,400 up to \$1,500	2
1,500 up to 1,600	5
1,600 up to 1,700	3
1,700 up to 1,800	1
Total	11

- c. Class frequencies.  
 d. The largest concentration of commissions is \$1,500 up to \$1,600. The smallest commission is about \$1,400 and the largest is about \$1,800. The typical amount earned is \$15,500.

- 2-3 a.  $2^6 = 64 < 73 < 128 = 2^7$ . So 7 classes are recommended.  
 b. The interval width should be at least  $(488 - 320)/7 = 24$ . Class intervals of 25 or 30 feet are both reasonable.  
 c. If we use a class interval of 25 feet and begin with a lower limit of 300 feet, eight classes would be necessary. A class interval of 30 feet beginning with 300 feet is also reasonable. This alternative requires only seven classes.

- 2-4 a. 23  
 b. 28.75%, found by  $(23/80) \times 100$   
 c. 7.5%, found by  $(6/80) \times 100$



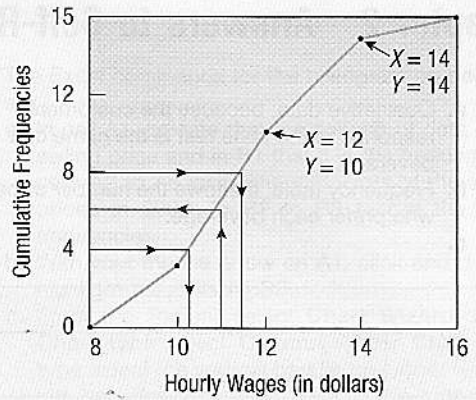
The plots are: (3.5, 12), (6.5, 26), (9.5, 40), (12.5, 20), and (15.5, 2).

- c. The smallest annual volume of imports by a supplier is about \$2 million, the largest about \$17 million. The highest frequency is between \$8 million and \$11 million.

2-6 a. A frequency distribution.

b.

Hourly Wages	Cumulative Number
Less than \$8	0
Less than \$10	3
Less than \$12	10
Less than \$14	14
Less than \$16	15



c. About seven employees earn \$11.00 or less.  
 About half the employees earn \$11.25 or more.  
 About four employees earn \$10.25 or less.