



MAT 1272 STATISTICS

LESSON 2

2.1 Organizing and Graphing Qualitative Data



2.1.1 Raw Data

Raw Data

Data recorded in the sequence in which they are collected and before they are processed or ranked are called *raw data*.

Suppose we collect information on the ages (in years) of 50 students selected from a university. The data values, in the order they are collected, are recorded in Table 2.1. For instance, the first student's age is 21, the second student's age is 19 (second number in the first row), and so forth. The data in Table 2.1 **are quantitative** raw data

Table 2.1 Ages of 50 Students

21	19	24	25	29	34	26	27	37	33
18	20	19	22	19	19	25	22	25	23
25	19	31	19	23	18	23	19	23	26
22	28	21	20	22	22	21	20	19	21
25	23	18	37	27	23	21	25	21	24

Qualitative data

Suppose we ask the same 50 students about their student status. The responses of the students are recorded in Table 2.2. In this table, F, SO, J, and SE are the abbreviations for freshman, sophomore, junior, and senior, respectively. This is an example of qualitative (or categorical) raw data.

Table 2.2 Status of 50 Students

J	F	SO	SE	J	J	SE	J	J	J
F	F	J	F	F	F	SE	SO	SE	J
J	F	SE	SO	SO	F	J	F	SE	SE
SO	SE	J	SO	SO	J	J	SO	F	SO
SE	SE	F	SE	J	SO	F	J	SO	SO

2.1.2 Frequency Distributions

The number of adults who belong to a certain category is called the frequency of that category. A frequency distribution exhibits how the frequencies are distributed over various categories. Table 2.3 is called a *frequency distribution table* or simply a *frequency table*.

Table 2.3 Worries About Not Having Enough Money to Pay Normal Monthly Bills		
Variable →	Response	Number of Adults ← Frequency column
	Very worried	162
	Moderately worried	203
Category →	Not too worried	305 ← Frequency
	Not worried at all	325
	Others	20
		Sum = 1015

EXAMPLE 2-1 What Variety of Donuts Is Your Favorite?








© Jack Puccio/Stockphoto

A sample of 30 persons who often consume donuts were asked what variety of donuts is their favorite. The responses from these 30 persons are as follows:

glazed	filled	other	plain	glazed	other
frosted	filled	filled	glazed	other	frosted
glazed	plain	other	glazed	glazed	filled
frosted	plain	other	other	frosted	filled
filled	other	frosted	glazed	glazed	filled

Construct a frequency distribution table for these data.

Table 2.4 Frequency Distribution of Favorite Donut Variety

Donut Variety	Tally	Frequency (f)
Glazed		8
Filled		7
Frosted		5
Plain		3
Other		7
		Sum = 30

2.1.3 Relative Frequency and Percentage Distributions

The relative frequency of a category is obtained by dividing the frequency of that category by the sum of all frequencies. Thus, the relative frequency shows what fractional part or proportion of the total frequency belongs to the corresponding category. A *relative frequency distribution* lists the relative frequencies for all categories.

Calculating Relative Frequency of a Category

$$\text{Relative frequency of a category} = \frac{\text{Frequency of that category}}{\text{Sum of all frequencies}}$$

The percentage for a category is obtained by multiplying the relative frequency of that category by 100. A *percentage distribution* lists the percentages for all categories.

Calculating Percentage

$$\text{Percentage} = (\text{Relative frequency}) \cdot 100\%$$

Constructing relative frequency and percentage distributions.

EXAMPLE 2-2 What Variety of Donuts Is Your Favorite?

Determine the relative frequency and percentage distributions for the data in Table 2.4.

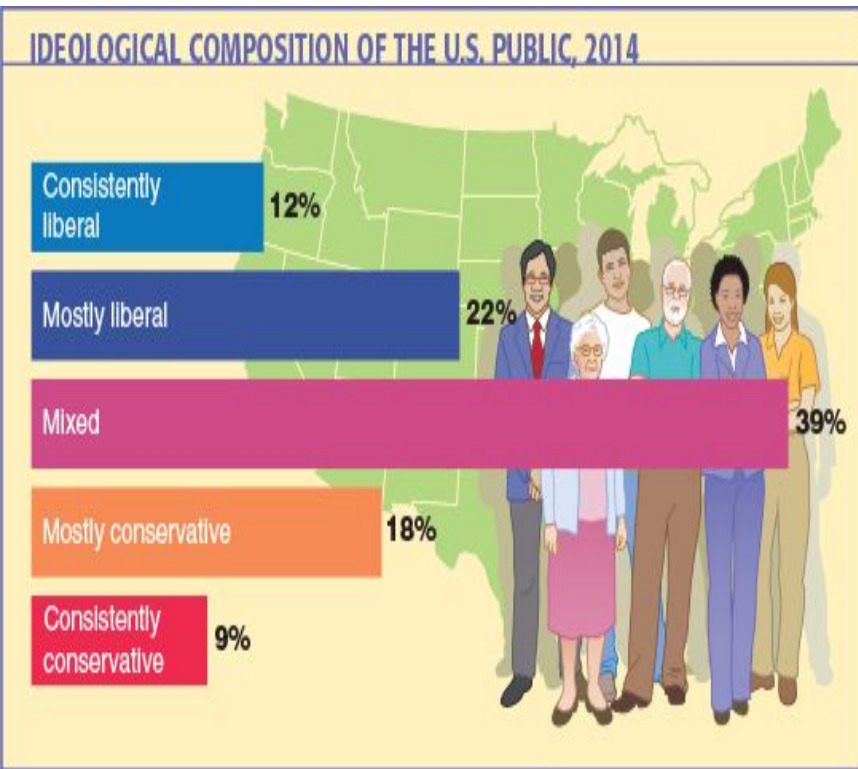
Solution

The relative frequencies and percentages from Table 2.4 are calculated and listed in Table 2.5. Based on this table, we can state that 26.7% of the people in the sample said that glazed donut is their favorite. two categories, we can state that 50% of the persons included in the sample said that glazed or filled donut is their favorite. The other numbers in Table 2.5 can be interpreted in similar ways.

Table 2.5 Relative Frequency and Percentage
Distributions of Favorite Donut Variety

Donut Variety	Relative Frequency	Percentage
Glazed	$8/30 = .267$	$.267(100) = 26.7$
Filled	$7/30 = .233$	$.233(100) = 23.3$
Frosted	$5/30 = .167$	$.167(100) = 16.7$
Plain	$3/30 = .100$	$.100(100) = 10.0$
Other	$7/30 = .233$	$.233(100) = 23.3$
	Sum = 1.000	Sum = 100%

The sum of the relative frequencies is always 1.00 (or approximately 1.00 if the relative frequencies are rounded), and the sum of the percentages is always 100 (or approximately 100 if the percentages are rounded)

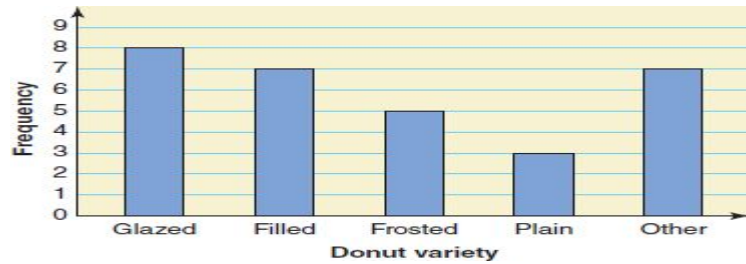


Data source: Pew Research Center

Pew Research Center conducted a national survey of 10,013 adults January 23 to March 16, 2014, to find the political views of adults in the United States. As the above bar chart shows, 12% of the adults polled said that they were consistently liberal, 22% indicated that they were mostly liberal, and so on. In this survey, Pew Research Center also found that, overall, the percentage of Americans who indicated that they were consistently conservative or consistently liberal has increased from 10% to 21% during the past two decades. Note that in this chart, the bars are drawn horizontally.

In order they appear

Bar Graph

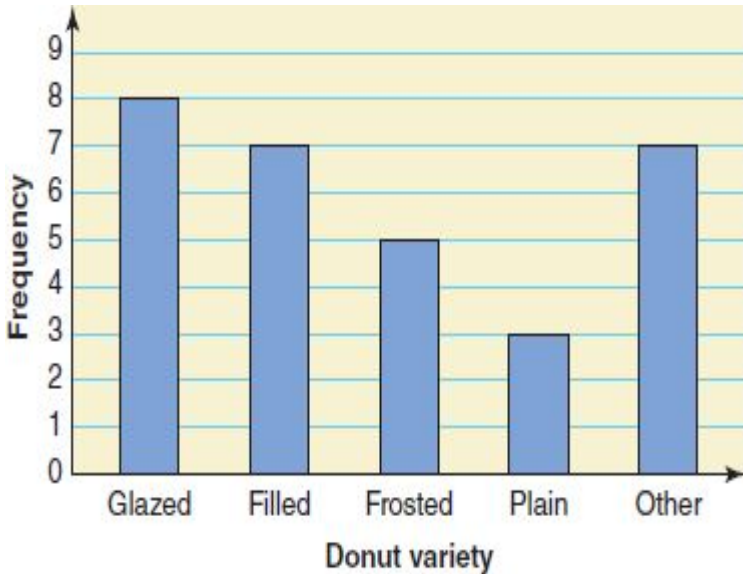


A graph made of bars whose heights represent the frequencies of respective categories is called a *bar graph*.

Pareto Chart

A Pareto chart is a bar graph with bars arranged by their heights in descending order. To make a Pareto chart, arrange the bars according to their heights such that the bar with the largest height appears first on the left side, and then subsequent bars are arranged in descending order with the bar with the smallest height appearing last on the right side.

2.1.4 Graphical Presentation of Qualitative Data



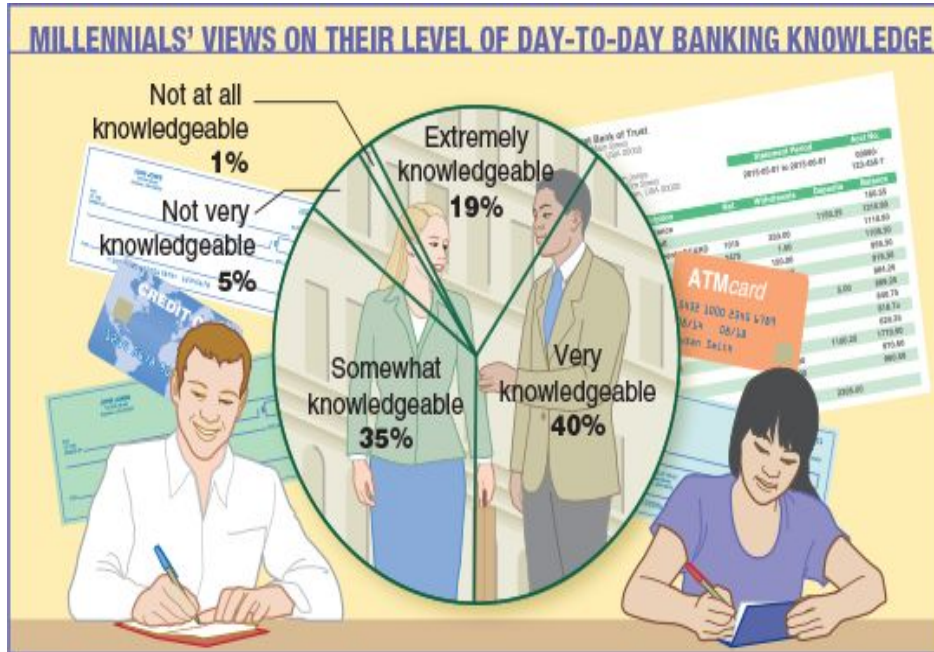
Bar Graphs

To construct a bar graph (also called a *bar chart*), we mark the various categories on the horizontal axis as in Figure 2.1. Note that all categories are represented by intervals of the same width. We mark the frequencies on the vertical axis. Then we draw one bar for each category such that the height of the bar represents the frequency of the corresponding category. We leave a small gap between adjacent bars. Figure 2.1 gives the bar graph for the frequency distribution of Table 2.4.

Figure 2.2 Pareto chart for the frequency distribution of Table 2.4.

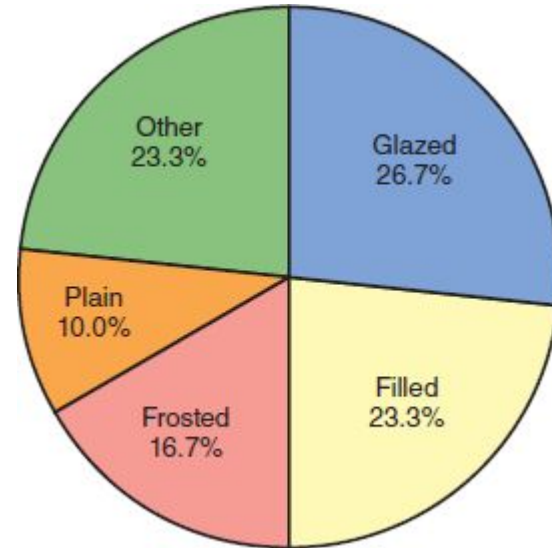
Pie Chart

A circle divided into portions that represent the relative frequencies or percentages of a population or a sample belonging to different categories is called a *pie chart*.



Data source: TD Bank: The Millennial Financial Behaviors & Needs Survey

Figure 2.3 Pie chart for the percentage distribution of Table 2.5.



Sara polled students in the cafeteria to determine what they would like added to the menu. Her results are shown below. Compute the relative frequency (percentage) of each choice to the nearest percent. [Note: Due to rounding, relative frequencies may not add to exactly 100%.]

Response	Frequency	Relative Frequency
Pizza	25	<input type="text" value="44"/> %
Waffles	17	<input type="text" value="30"/> %
Tiramisu	9	<input type="text" value="16"/> %
Jerk Chicken	6	<input type="text" value="11"/> %
Total	57	100%

What is the difference between frequency and relative frequency?

Refer to the table in this slide to do the interpretation and explain the similarities and differences.

2.3 The following data give the results of a sample survey. The letters Y, N, and D represent the three categories.

D N N Y Y Y N Y D Y
Y Y Y Y N Y Y N N Y
N Y Y N D N Y Y Y Y
Y Y N N Y Y N N D Y

(a) Prepare a frequency distribution table.

ANSWER ⊕

WORKED SOLUTION ⊕

(b) Calculate the relative frequencies and percentages for all categories.

ANSWER ⊕

WORKED SOLUTION ⊕

(c) What percentage of the elements in this sample belong to category Y?

ANSWER ⊕

WORKED SOLUTION ⊕

(d) What percentage of the elements in this sample belong to category N or D?

ANSWER ⊕

WORKED SOLUTION ⊕

(e) Draw a pie chart for the percentage distribution.

WORKED SOLUTION ⊕

(f) Make a Pareto chart for the percentage distribution.

WORKED SOLUTION ⊕

The letters Y, N, and D represent the three categories.

D N N Y Y Y N Y D Y
Y Y Y Y N Y Y N N Y
N Y Y N D N Y Y Y Y
Y Y N N Y Y N N D Y

a

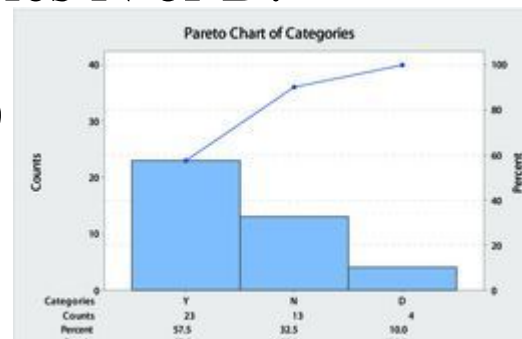
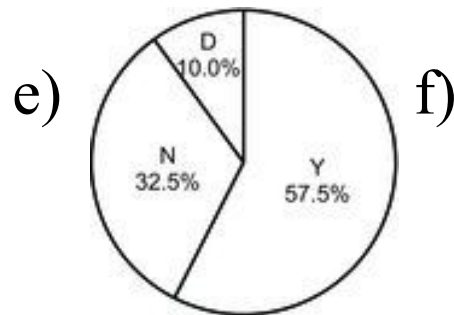
Category	Frequency	Relative Frequency	Percentage
Y	23	$23/40 = 0.575$	57.5
N	13	$13/40 = 0.325$	32.5
D	4	$4/40 = 0.100$	10.0

c) 57.5% of the elements belong to category Y.

b

Category	Frequency	Relative Frequency	Percentage
Y	23	$23/40 = 0.575$	57.5
N	13	$13/40 = 0.325$	32.5
D	4	$4/40 = 0.100$	10.0

d) $17/40 = 42.5\%$ of the elements belong to categories N or D.



2.7 In a 2013 survey of employees conducted by Financial Finesse Inc., employees were asked about their overall financial stress levels. The following table shows the results of this survey (www.financialfinesse.com)

Financial Stress Level	Percentage of Responses
No financial stress	14
Some financial stress	63
High financial stress	18
Overwhelming financial stress	5

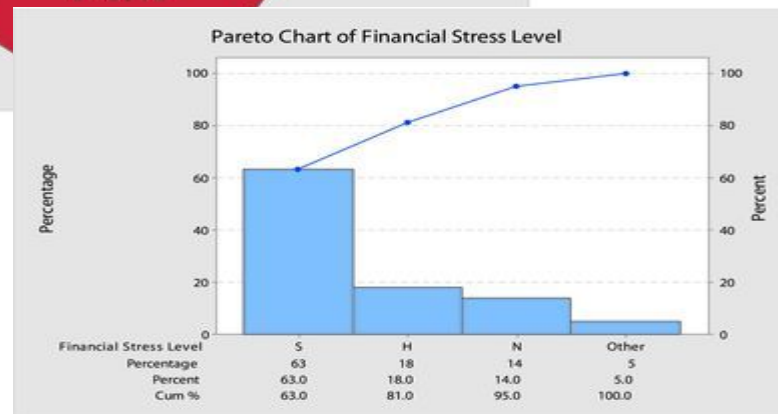
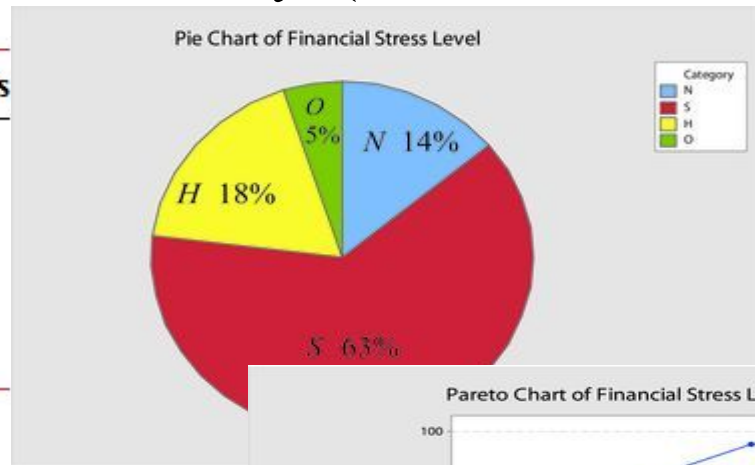
- (a) Draw a pie chart for this percentage distribution.
- (b) Make a Pareto chart for this percentage distribution.

2.7 In a 2013 survey of employees conducted by Financial Finesse Inc., employees were asked about their overall financial stress levels. The following table shows the results of this survey (www.financialfinesse.com)

Financial Stress Level	Percentage of Responses
No financial stress	14
Some financial stress	63
High financial stress	18
Overwhelming financial stress	5

(a) Draw a pie chart for this percentage distribution.

(b) Make a Pareto chart for this percentage distribution.



A staff member at a local grocery store was assigned the job of inspecting all containers of yogurt in the store to determine the number of days to expiry date for each container. Containers that had already expired but were still on the shelves were given a value of 0 for number of days to expiry. The following table gives the frequency distribution of the number of days to expiry date.

a. How many containers of yogurt were inspected?

Number of Days	Number of Containers
0 to 5	33
6 to 11	65
12 to 17	44
18 to 23	22
24 to 29	11

175

 containers of yogurt

Number of Days (Class Limits)	Class Width	Class Midpoint
0 to 5	6	2.5
6 to 11	6	8.5
12 to 17	6	14.5
18 to 23	6	20.5
24 to 29	6	26.5

Table of frequency, relative frequency, %

Number of Days (Class Limits)	Frequency	Relative Frequency	Percentage
0 to 5	33	0.189	18.9
6 to 11	65	0.371	37.1
12 to 17	44	0.251	25.1
18 to 23	22	0.126	12.6
24 to 29	11	0.063	6.3

e. Explain why you cannot determine exactly how many containers have already expired.

- ☐ It is necessary to know the average number of purchases per customer.
- ☐ There are too many containers not included in the class 0 to 5.
- ☐ There are too few containers included in the class 0 to 5.
- ☐ The containers that have not already expired skews the data.
- ☒ 0 is included in the class 0 to 5.

f. What is the largest number of containers that may have already expired?

The largest number of containers that could already have expired is .

Frequency distribution

Variable	Weekly Earnings (dollars)	Number of Employees f	Frequency column
	801 to 1000	4	
	1001 to 1200	11	
Third class	1201 to 1400	39	{ Frequency of the third class
	1401 to 1600	24	
	1601 to 1800	16	
	1801 to 2000	6	
Lower limit of the sixth class			Upper limit of the sixth class

Frequency Distribution for Quantitative Data

A *frequency distribution* for quantitative data lists all the classes and the number of values that belong to each class. Data presented in the form of a frequency distribution are called *grouped data*.

2.2.3 Relative Frequency and Percentage Distributions

$$\text{Relative frequency of a class} = \frac{\text{Frequency of that class}}{\text{Sum of all frequencies}} = \frac{f}{\sum f}$$

$$\text{Percentage} = (\text{Relative frequency}) \cdot 100\%$$

Finding Class Width

To find the width of a class, subtract its lower limit from the lower limit of the next class. Thus:

$$\text{Width of a class} = \text{Lower limit of the next class} - \text{Lower limit of the current class}$$

$$\text{Width of the first class} = 1001 - 801 = 200$$

Calculating Class Midpoint or Mark

$$\text{Class midpoint or mark} = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

Thus, the midpoint of the first class in Table 2.6 or Table 2.7 is calculated as follows:

$$\text{Midpoint of the first class} = \frac{801 + 1000}{2} = 900.5$$

The class midpoints for the frequency distribution of Table 2.6 are listed in the third column of Table 2.7.

Table 2.7 Class Widths and Class Midpoints for Table 2.6

Class Limits	Class Width	Class Midpoint
801 to 1000	200	900.5
1001 to 1200	200	1100.5
1201 to 1400	200	1300.5
1401 to 1600	200	1500.5
1601 to 1800	200	1700.5
1801 to 2000	200	1900.5

2.2.2 Constructing Frequency Distribution Tables

When constructing a frequency distribution table, we need to make the following three major decisions.

(1) Number of Classes

Usually the number of classes for a frequency distribution table varies from 5 to 20, depending mainly on the number of observations in the data set.¹One rule to help decide on the number of classes is Sturge's formula:

$$c = 1 + 3.3 \log n$$

where c is the number of classes and n is the number of observations in the data set. The value of $\log n$ can be obtained by using a calculator. It is preferable to have more classes as the size of a data set increases. The decision about the number of classes is arbitrarily made by the data organizer.

(2) Class Width

Although it is not uncommon to have classes of different sizes, most of the time it is preferable to have the same width for all classes. To determine the class width when all classes are the same size, first find the difference between the largest and the smallest values in the data. Then, the approximate width of a class is obtained by dividing this difference by the number of desired classes.

Calculation of Class Width

$$\text{Approximate class width} = \frac{\text{Largest value} - \text{Smallest value}}{\text{Number of classes}}$$

Usually this approximate class width is rounded to a convenient number, which is then used as the class width. Note that rounding this number may slightly change the number of classes initially intended.

(3) Lower Limit of the First Class or the Starting Point

Any convenient number that is equal to or less than the smallest value in the data set can be used as the lower limit of the first class.

Example 2-3 illustrates the procedure for constructing a frequency distribution table for quantitative data.

Constructing a frequency distribution table for quantitative data.

EXAMPLE 2-3 Values of Baseball Teams, 2015

The following table gives the value (in million dollars) of each of the 30 baseball teams as estimated by *Forbes* magazine (*source: Forbes Magazine*, April 13, 2015). Construct a frequency distribution table.

Values of Baseball Teams, 2015			
Team	Value (millions of dollars)	Team	Value (millions of dollars)
Arizona Diamondbacks	840	Milwaukee Brewers	875
Atlanta Braves	1150	Minnesota Twins	895
Baltimore Orioles	1000	New York Mets	1350
Boston Red Sox	2100	New York Yankees	3200
Chicago Cubs	1800	Oakland Athletics	725
Chicago White Sox	975	Philadelphia Phillies	1250
Cincinnati Reds	885	Pittsburgh Pirates	900
Cleveland Indians	825	San Diego Padres	890
Colorado Rockies	855	San Francisco Giants	2000
Detroit Tigers	1125	Seattle Mariners	1100
Houston Astros	800	St. Louis Cardinals	1400
Kansas City Royals	700	Tampa Bay Rays	605
Los Angeles Angels of Anaheim	1300	Texas Rangers	1220
Los Angeles Dodgers	2400	Toronto Blue Jays	870
Miami Marlins	650	Washington Nationals	1280

Solution

In these data, the minimum value is 605, and the maximum value is 3200. Suppose we decide to group these data using six classes of equal width. Then,

$$\text{Approximate width of each class} = \frac{3200 - 605}{6} = 432.5$$

Now we round this approximate width to a convenient number, say 450. The lower limit of the first class can be taken as 605 or any number less than 605. Suppose we take 601 as the lower limit of the first class. Then our classes will be

601-1050, 1051-1500, 1501-1950, 1951-2400, 2401-2850, and 2851-3300

We record these five classes in the first column of Table 2.8.

Table 2.8 Frequency Distribution of the Values of Baseball Teams, 2015

Value of a Team (in million \$)	Tally	Number of Teams (f)
601-1050		16
1051-1500		9
1501-1950		1
1951-2400		3
2401-2850		0
2851-3300		1
		$\Sigma f = 30$

Now we read each value from the given data and mark a tally in the second column of Table 2.8 next to the corresponding class. The first value in our original data set is 840, which belongs to the 601-1050 class. To record it, we mark a tally in the second column next to the 601-1050 class. We continue this process until all the data values have been read and entered in the tally column. Note that tallies are marked in blocks of five for counting convenience. After the tally column is completed, we count the tally marks for each class and write those numbers in the third column. This gives the column of frequencies. These frequencies represent the number of baseball teams with values in the corresponding classes. For example, 16 of the teams have values in the interval \$601-\$1050 million.

Using the Σ notation (see Section 1.7 of Chapter 1), we can denote the sum of frequencies of all classes by Σf . Hence,

$$\Sigma f = 16 + 9 + 1 + 3 + 0 + 1 = 30$$

The number of observations in a sample is usually denoted by n . Thus, for the sample data, Σf is equal to n . The number of observations in a population is denoted by N . Consequently, Σf is equal to N for population data. Because the data set on the values of baseball teams in Table 2.8 is for all 30 teams, it represents a population. Therefore, in Table 2.8 we can denote the sum of frequencies by N instead of Σf .

Note that when we present the data in the form of a frequency distribution table, as in Table 2.8, we lose the information on individual observations. We cannot know the exact value of any team from Table 2.8. All we know is that 16 teams have values in the interval \$601-\$1050 million, and so forth.

2.2.3 Relative Frequency and Percentage Distributions

Using Table 2.8, we can compute the relative frequency and percentage distributions in the same way as we did for qualitative data in Section 2.1.3. The relative frequencies and percentages for a quantitative data set are that relative frequency is the same as proportion.

Calculating Relative Frequency and Percentage

$$\begin{aligned}\text{Relative frequency of a class} &= \frac{\text{Frequency of that class}}{\text{Sum of all frequencies}} = \frac{f}{\sum f} \\ \text{Percentage} &= (\text{Relative frequency}) \cdot 100\%\end{aligned}$$

Example 2-4 illustrates how to construct relative frequency and percentage distributions.

Constructing relative frequency and percentage distributions.

EXAMPLE 2-4 Values of Baseball Teams, 2015

Calculate the relative frequencies and percentages for Table 2.8.

Solution

The relative frequencies and percentages for the data in Table 2.8 are calculated and listed in the second and third columns, respectively, of Table 2.9.

Table 2.9 Relative Frequency and Percentage Distributions of the Values of Baseball Teams

Value of a Team (in million \$)	Relative Frequency	Percentage
601-1050	16/30 = .533	53.3
1051-1500	9/30 = .300	30.0
1501-1950	1/30 = .033	3.3
1951-2400	3/30 = .100	10.0
2401-2850	0/30 = .000	0.0
2851-3300	1/30 = .033	3.3
	Sum = 1.000	Sum = 100%

Histogram

A *histogram* is a graph in which classes are marked on the horizontal axis and the frequencies, relative frequencies, or percentages are marked on the vertical axis. The frequencies, relative frequencies, or percentages are represented by the heights of the bars. In a histogram, the bars are drawn adjacent to each other.

Figures 2.4 and 2.5 show the frequency and the percentage histograms, respectively, for the data of Tables 2.8 and 2.9 of Sections 2.2.2 and 2.2.3. The two histograms look alike because they represent the same data. A relative frequency histogram can be drawn for the relative frequency distribution of Table 2.9 by marking the relative frequencies on the vertical axis.

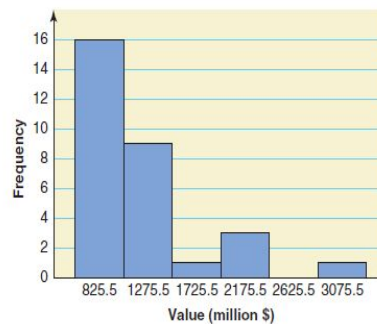


Figure 2.4 Frequency histogram for Table 2.8.

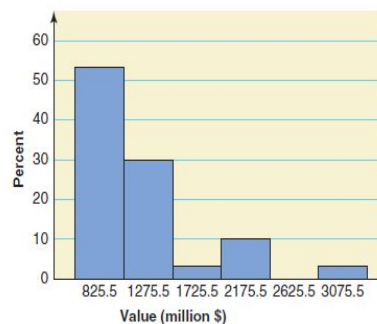
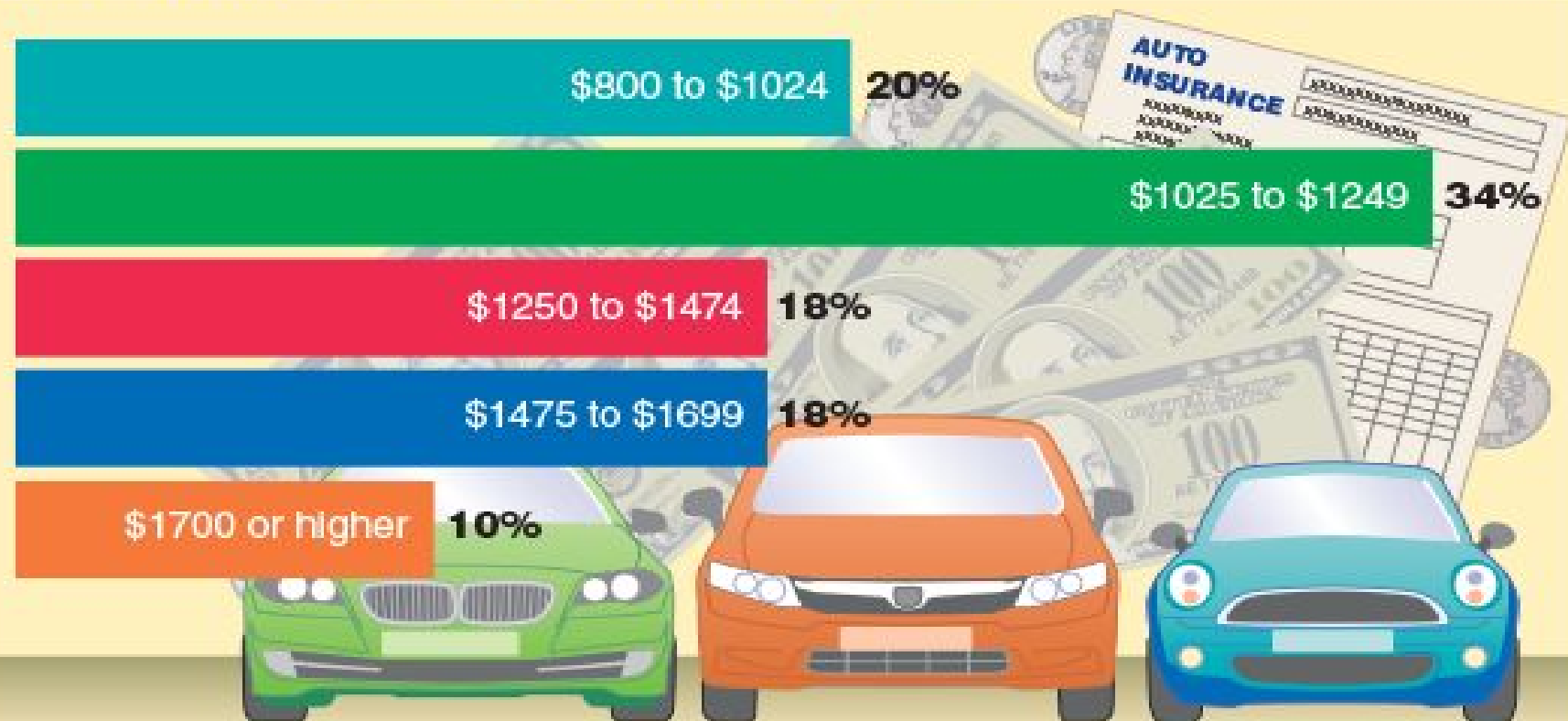


Figure 2.5 Percentage distribution histogram for Table 2.9.

CAR INSURANCE PREMIUMS PER YEAR IN 50 STATES



Data source: www.insure.com

Polygon

A graph formed by joining the midpoints of the tops of successive bars in a histogram with straight lines is called a *polygon*.

Figure 2.6 shows the frequency polygon for the frequency distribution of Table 2.8.

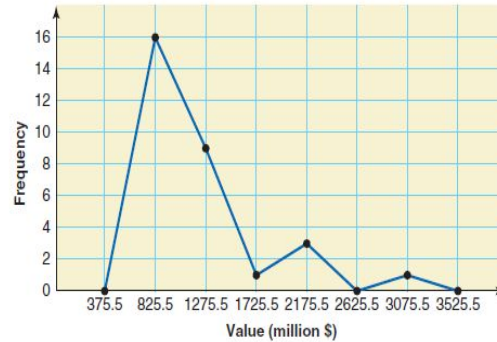


Figure 2.6 Frequency polygon for Table 2.8.

For a very large data set, as the number of classes is increased (and the width of classes is decreased), the frequency polygon eventually becomes a smooth curve. Such a curve is called a *frequency distribution curve* or simply a *frequency curve*. Figure 2.7 shows the frequency curve for a large data set with a large number of classes.

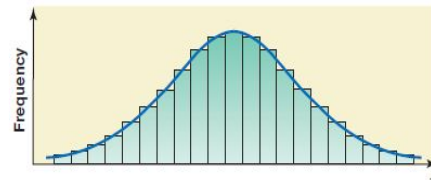


Figure 2.7 Frequency distribution curve.

EXAMPLE 2-6 Number of Vehicles Owned by Households



© Jorge Salcedo/Stockphoto

The administration in a large city wanted to know the distribution of the number of vehicles owned by households in that city. A sample of 40 randomly selected households from this city produced the following data on the number of vehicles owned.

5 1 1 2 0 1 1 2 1 1
1 3 3 0 2 5 1 2 3 4
2 1 2 2 1 2 2 1 1 1
4 2 1 1 2 1 1 4 1 3

Construct a frequency distribution table for these data using single-valued classes.

Solution

The observations in this data set assume only six distinct values: 0, 1, 2, 3, 4, and 5. Each of these six values is used as a class in the frequency distribution in Table 2.12, and these six classes are listed in the first column of that table. To obtain the frequencies of these classes, the observations in the data that belong to each class are counted, and the results are recorded in the second column of Table 2.12. Thus, in these data, 2 households own no vehicle, 18 own one vehicle each, 11 own two vehicles each, and so on.

The data of Table 2.12 can also be displayed in a bar graph, as shown in Figure 2.8. To construct a bar graph, we mark the classes, as intervals, on the horizontal axis with a little gap between consecutive intervals. The bars represent the frequencies of respective classes.

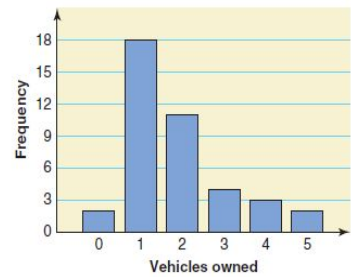


Figure 2.8 Bar graph for Table 2.12.

The frequencies of Table 2.12 can be converted to relative frequencies and percentages the same way as in Table 2.9. Then, a bar graph can be constructed to display the relative frequency or percentage distribution by marking the relative frequencies or percentages, respectively, on the vertical axis.

Table 2.12 Frequency Distribution of the Number of Vehicles Owned	
Vehicles Owned	Number of Households (<i>f</i>)
0	2
1	18
2	11
3	4
4	3
5	2
$\sum f = 40$	

2.2.6 Cumulative Frequency Distributions

Consider again Example 2-3 of Section 2.2.2 about the values of baseball teams. Suppose we want to know how many baseball teams had values of \$1500 million or less in 2015. Such a question can be answered by using a cumulative frequency distribution. Each class in a cumulative frequency distribution table gives the total number of values that fall below a certain value. A cumulative frequency distribution is constructed for quantitative data only.

Cumulative Frequency Distribution

A cumulative frequency distribution gives the total number of values that fall below the upper boundary of each class.

In a cumulative frequency distribution table, each class has the same lower limit but a different upper limit. Example 2-7 illustrates the procedure for preparing a cumulative frequency distribution.

Example 2-7

EXAMPLE 2-7 Values of Baseball Teams, 2015

Using the frequency distribution of Table 2.8, reproduced here, prepare a cumulative frequency distribution for the values of the baseball teams.

Value of a Team (in million \$)	Number of Teams (<i>f</i>)
601-1050	16
1051-1500	9
1501-1950	1
1951-2400	3
2401-2850	0
2851-3300	1

Solution

Table 2.13 gives the cumulative frequency distribution for the values of the baseball teams. As we can observe, 601 (which is the lower limit of the first class in Table 2.8) is taken as the lower limit of each class in Table 2.13. The upper limits of all classes in Table 2.13 are the same as those in Table 2.8. To obtain the cumulative frequency of a class, we add the frequency of that class in Table 2.8 to the frequencies of all preceding classes. The cumulative frequencies are recorded in the second column of Table 2.13.

Table 2.13 Cumulative Frequency Distribution of Values of Baseball Teams, 2015	
Class Limits	Cumulative Frequency
601-1050	16
601-1500	$16 + 9 = 25$
601-1950	$16 + 9 + 1 = 26$
601-2400	$16 + 9 + 1 + 3 = 29$
601-2850	$16 + 9 + 1 + 3 + 0 = 29$
601-3300	$16 + 9 + 1 + 3 + 0 + 1 = 30$

From Table 2.13, we can determine the number of observations that fall below the upper limit of each class. For example, 26 teams were valued between \$601 and \$1950 million.

The cumulative relative frequencies are obtained by dividing the cumulative frequencies by the total number of observations in the data set. The cumulative percentages are obtained by multiplying the cumulative relative frequencies by 100.

Calculating Cumulative Relative Frequency and Cumulative Percentage

Cumulative relative frequency = $\frac{\text{Cumulative frequency of a class}}{\text{Total observations in the data set}}$

Cumulative percentage = $(\text{Cumulative relative frequency}) \cdot 100\%$

Table 2.14 contains both the cumulative relative frequencies and the cumulative percentages for Table 2.13. We can observe, for example, that 90% of the teams were valued between \$601 and \$1800 million.

Table 2.14 Cumulative Relative Frequency and Cumulative Percentage Distributions for Values of Baseball Teams, 2015		
Class Limits	Cumulative Relative Frequency	Cumulative Percentage
601-1050	$16/30 = .5333$	53.33
601-1500	$25/30 = .8333$	83.33
601-1950	$26/30 = .8667$	86.67
601-2400	$29/30 = .9667$	96.67
601-2850	$29/30 = .9667$	96.67
601-3300	$30/30 = 1.000$	100.00

2.2.7 Shapes of Histograms

A histogram can assume any one of a large number of shapes. The most common of these shapes are

1. Symmetric
2. Skewed
3. Uniform or rectangular

A symmetric histogram is identical on both sides of its central point. The histograms shown in Figure 2.9 are symmetric around the dashed lines that represent their central points.

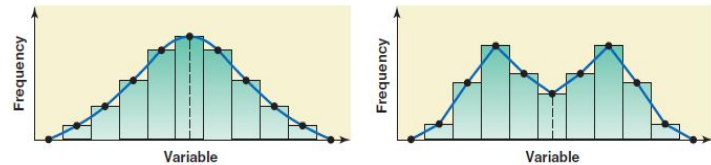


Figure 2.9 Symmetric histograms.

A **skewed histogram** is nonsymmetric. For a skewed histogram, the tail on one side is longer than the tail on the other side. A skewed-to-the-right histogram has a longer tail on the right side (see Figure 2.10a). A skewed-to-the-left histogram has a longer tail on the left side (see Figure 2.10b).

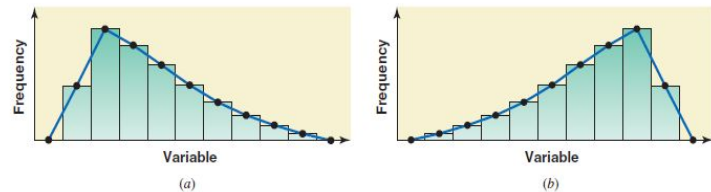


Figure 2.10 (a) A histogram skewed to the right. (b) A histogram skewed to the left.

A uniform or rectangular histogram has the same frequency for each class. Figure 2.11 is an illustration of such a case.

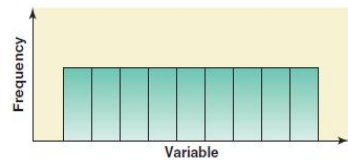


Figure 2.11 A histogram with uniform distribution.

2.2.8 Truncating Axes

Describing data using graphs gives us insights into the main characteristics of the data. But graphs, unfortunately, can also be used, intentionally or unintentionally, to distort the facts and deceive the reader. The following are two ways to manipulate graphs to convey a particular opinion or impression.

1. *Changing the scale* either on one or on both axes—that is, shortening or stretching one or both of the axes.
2. *Truncating the frequency axis*—that is, starting the frequency axis at a number greater than zero.

Suppose 400 randomly selected adults were asked whether or not they are happy with their jobs. Of them, 156 said that they are happy, 136 said that they are not happy, and 108 had no opinion. Converting these numbers to percentages, 39% of these adults said that they are happy, 34% said that they are not happy, and 27% had no opinion. Let us denote the three opinions by A, B, and C, respectively. The following table shows the results of this survey.

Opinion	Percentage
A	39
B	34
C	27
Sum = 100	

Now let us make two bar graphs—one showing the complete vertical axis and the second using a truncated vertical axis. Figure 2.13 shows a bar graph with the complete vertical axis. By looking at this bar graph, we can observe that the opinions represented by three categories in fact differ by small percentages. But now look at Figure 2.14 in which the vertical axis has been truncated to start at 25%. By looking at this bar chart, if we do not pay attention to the vertical axis, we may erroneously conclude that the opinions represented by three categories vary by large percentages.

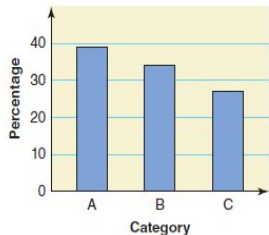


Figure 2.13 Bar graph without truncation of the vertical axis.

