Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

SKU: 9781337119917-SOLUTIONS

Teaching Tips for Each Chapter

CHAPTER 1: GETTING STARTED

Double-Blind Studies (Section 1.3)

The double-blind method of data collection, mentioned at the end of Section 1.3, is an important part of standard research practice. A typical use is in testing new medications. Because the researcher does not know which patients are receiving the experimental drug and which are receiving the established drug (or a placebo), the researcher is prevented from doing things subconsciously that might skew the results.

If, for instance, the researcher communicates a more optimistic attitude to patients in the experimental group, this could influence how they respond to diagnostic questions or actually might influence the course of their illness. And if the researcher wants the new drug to prove effective, this could subconsciously influence how he or she handles information related to each patient's case. All such factors are eliminated in double-blind testing.

The following appears in the physician's dosing instructions package insert for the prescription drug QUIXIN™:

> In randomized, double-masked, multicenter controlled clinical trials where patients were dosed for 5 days, QUIXIN™ demonstrated clinical cures in 79% of patients treated for bacterial conjunctivitis on the final study visit day (days 6–10).

Note the phrase *double-masked.* Apparently, this is a synonym for *double-blind.* Since *double-blind* is used widely in the medical literature and in clinical trials, why do you suppose that the company chose to use *double-masked* instead?

Perhaps this will provide some insight: QUIXIN™ is a topical antibacterial solution for the treatment of conjunctivitis; i.e., it is an antibacterial eye drop solution used to treat an inflammation of the conjunctiva, the mucous membrane that lines the inner surface of the eyelid and the exposed surface of the eyeball. Perhaps, since QUIXIN™ is a treatment for eye problems, the manufacturer decided the word *blind* should not appear *anywhere* in the discussion.

Source: Package insert. QUIXIN™ is manufactured by Santen Oy, P.O. Box 33, FIN-33721 Tampere, Finland, and marketed by Santen, Inc., Napa, CA 94558, under license from Daiichi Pharmaceutical Co., Ltd., Tokyo, Japan.

CHAPTER 2: ORGANIZING DATA

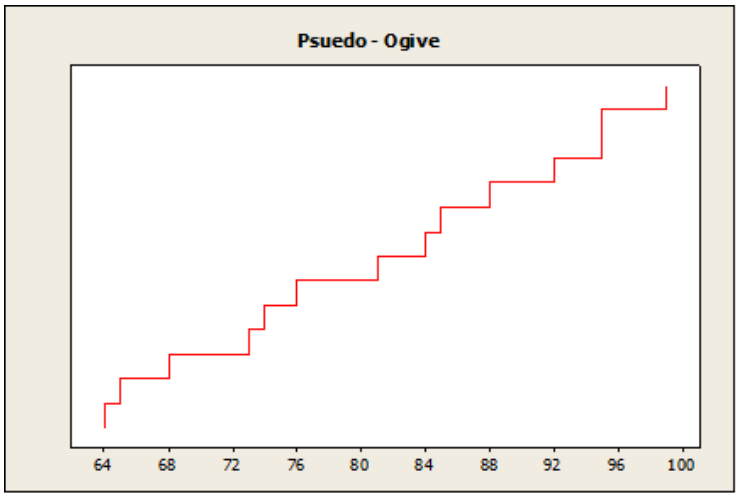
Emphasize when to use the various graphs discussed in this chapter: bar graphs when comparing data sets, circle graphs for displaying how data are dispersed into several categories, time-series graphs to display how data change over time, histograms or frequency polygons to display relative frequencies of grouped data, and stem-and-leaf displays for displaying grouped data in a way that does not lose the detail of the original raw data.

Drawing and Using Ogives (Section 2.1)

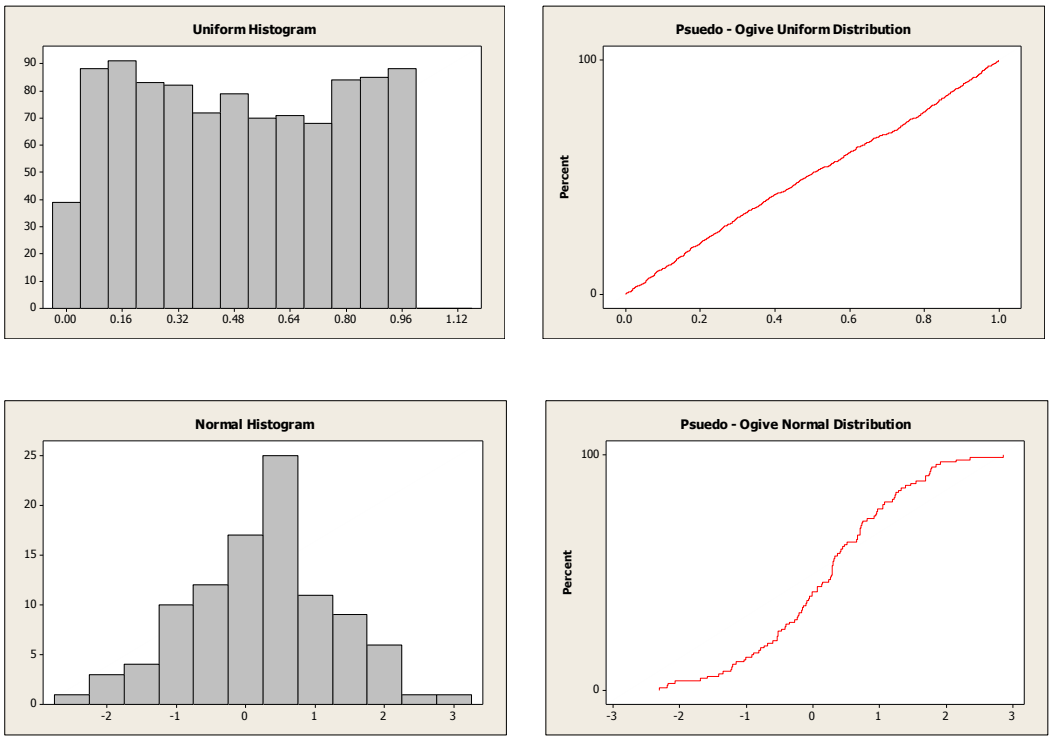
The text describes how an ogive, which is a graph displaying a cumulative-frequency distribution, can be constructed easily using a frequency table. However, a graph of the same basic sort can be constructed even more quickly than that. Simply arrange the data values in ascending order and then plot one point for each data value, where the x coordinate is the data value and the y coordinate starts at 1 for the first point and increases by 1 for each successive point. Finally, connect adjacent points with line segments. In the resulting graph, for any x, the corresponding y value will be (roughly) the number of data values less than or equal to x.

For example, here is the graph for the data set 64, 65, 68, 73, 74, 76, 81, 84, 85, 88, 92, 95, 95, and 99:

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual



This graph is not technically an ogive because the possibility of duplicate data values (such as 95 in this example) means that the graph will not necessarily be a function. But the graph can be used to get a quick fix on the general shape of the cumulative distribution curve. And by implication, the graph can be used to get a quick idea of the shape of the frequency distribution, as illustrated below.



The pseudo-ogive obtained from the example data set suggests a uniform distribution on the interval 63–100 or thereabouts.

CHAPTER 3: AVERAGES AND VARIATION

Students should be instructed in the various ways that sets of numeric data can be represented by a single number. The concepts of this section illustrate for students the need for this kind of representation.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Organizing Data

2

Copyright © Cengage Learning. All rights reserved.

Section 2.1

Frequency Distributions, Histograms, and Related Topics

Copyright © Cengage Learning. All rights reserved.

Focus Points

- Organize raw data using a frequency table.
- Construct histograms, relative-frequency histograms, and ogives.
- Recognize basic distribution shapes: uniform, symmetric, skewed, and bimodal.
- Interpret graphs in the context of the data setting.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Frequency Tables

Frequency Tables

When we have a large set of quantitative data, it's useful to organize it into smaller intervals or *classes* and count how many data values fall into each class. A frequency table does just that.

A **frequency table** partitions data into classes or intervals and shows how many data values are in each class. The classes or intervals are constructed so that each data value falls into exactly one class.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 1 – *Frequency table*

A task force to encourage car pooling did a study of one-way commuting distances of workers in the downtown Dallas area. A random sample of 60 of these workers was taken. The commuting distances of the workers in the sample are given in Table 2-1. Make a frequency table for these data.

13	47	10	3	16	20	17	40	4	2
7	25	8	21	19	15	3	17	14	6
12	45	1	8	4	16	11	18	23	12
6	2	14	13	7	15	46	12	9	18
34	13	41	28	36	17	24	27	29	9
14	26	10	24	37	31	8	16	12	16

One-Way Commuting Distances (in Miles) for 60 Workers in Downtown Dallas

Table 2-1

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 1 – *Solution*

a. First decide how many classes you want. Five to 15 classes are usually used. If you use fewer than five classes, you risk losing too much information. If you use more than 15 classes, the data may not be sufficiently summarized.

Let the spread of the data and the purpose of the frequency table be your guides when selecting the number of classes. In the case of the commuting data, let's use *six* classes.

b. Next, find the *class width* for the six classes.

Example 1 – *Solution*

cont'd

Procedure:

HOW TO FIND THE CLASS WIDTH (INTEGER DATA)

1. Compute $\dfrac{\text{Largest data value} - \text{smallest data value}}{\text{Desired number of classes}}$
2. Increase the computed value to the next highest whole number.

Example 1 – *Solution*

cont'd

To find the class width for the commuting data, we observe that the largest distance commuted is 47 miles and the smallest is 1 mile. Using six classes, the class width is 8, since

Class width = $\dfrac{47 - 1}{6} \approx 7.7$ (increase to 8)

c. Now we determine the data range for each class.

The **lower class limit** is the lowest data value that can fit in a class. The **upper class limit** is the highest data value that can fit in a class. The **class width** is the difference between the *lower* class limit of one class and the *lower* class limit of the next class.

Example 1 – *Solution*

cont'd

The smallest commuting distance in our sample is 1 mile. We use this *smallest* data value as the lower class limit of the *first* class.

Since the class width is 8, we add 8 to 1 to find that the *lower* class limit for the *second* class is 9.

Following this pattern, we establish *all* the *lower class limits*.

Then we fill in the *upper class limits* so that the classes span the entire range of data.

Example 1 – *Solution*

cont'd

Table 2-2, shows the upper and lower class limits for the commuting distance data.

Class Limits Lower–Upper	Class Boundaries Lower–Upper	Tally	Frequency	Class Midpoint
1–8	0.5–8.5	𝍸 𝍸 ||||	14	4.5
9–16	8.5–16.5	𝍸 𝍸 𝍸 𝍸 |	21	12.5
17–24	16.5–24.5	𝍸 𝍸 |	11	20.5
25–32	24.5–32.5	𝍸 |	6	28.5
33–40	32.5–40.5	||||	4	36.5
41–48	40.5–48.5	||||	4	44.5

Frequency Table of One-Way Commuting Distances for 60 Downtown Dallas Workers (Data in Miles)

Table 2-2

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 1 – *Solution*

cont'd

d. Now we are ready to tally the commuting distance data into the six classes and find the frequency for each class.

Procedure:

HOW TO TALLY DATA

Tallying data is a method of counting data values that fall into a particular class or category.

To tally data into classes of a frequency table, examine each data value. Determine which class contains the data value and make a tally mark or vertical stroke (|) beside that class. For ease of counting, each fifth tally mark of a class is placed diagonally across the prior four marks (𝍸).

The *class frequency* for a class is the number of tally marks corresponding to that class.

Example 1 – *Solution*

cont'd

Table 2-2 shows the tally and frequency of each class.

e. The center of each class is called the *midpoint* (or *class mark*). The midpoint is often used as a representative value of the entire class. The midpoint is found by adding the lower and upper class limits of one class and dividing by 2.

$$\text{Midpoint} = \frac{\text{Lower class limit} + \text{upper class limit}}{2}$$

Table 2-2 shows the class midpoints.

Example 1 – *Solution*

cont'd

f. There is a space between the upper limit of one class and the lower limit of the next class. The halfway points of these intervals are called *class boundaries*. These are shown in Table 2-2.

Procedure:

HOW TO FIND CLASS BOUNDARIES (INTEGER DATA)

To find **upper class boundaries,** add 0.5 unit to the upper class limits.

To find **lower class boundaries,** subtract 0.5 unit from the lower class limits.

Frequency Tables

Basic frequency tables show how many data values fall into each class. It's also useful to know the *relative frequency* of a class. The relative frequency of a class is the proportion of all data values that fall into that class. To find the relative frequency of a particular class, divide the class frequency *f* by the total of all frequencies *n* (sample size).

Class	Frequency *f*	Relative Frequency *f/n*
1–8	14	14/60 ≈ 0.23
9–16	21	21/60 ≈ 0.35
17–24	11	11/60 ≈ 0.18
25–32	6	6/60 ≈ 0.10
33–40	4	4/60 ≈ 0.07
41–48	4	4/60 ≈ 0.07

Relative Frequencies of One-Way Commuting Distances

Table 2-3

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Frequency Tables

$$\text{Relative frequency} = \frac{f}{n} = \frac{\text{Class frequency}}{\text{Total of all frequencies}}$$

Table 2-3 shows the relative frequencies for the commuter data of Table 2-1.

13	47	10	3	16	20	17	40	4	2
7	25	8	21	19	15	3	17	14	6
12	45	1	8	4	16	11	18	23	12
6	2	14	13	7	15	46	12	9	18
34	13	41	28	36	17	24	27	29	9
14	26	10	24	37	31	8	16	12	16

One-Way Commuting Distances (in Miles) for 60 Workers in Downtown Dallas

Table 2-1

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Frequency Tables

Since we already have the frequency table (Table 2-2), the relative-frequency table is obtained easily.

Class Limits Lower–Upper	Class Boundaries Lower–Upper	Tally	Frequency	Class Midpoint
1–8	0.5–8.5	卌 卌 ||||	14	4.5
9–16	8.5–16.5	卌 卌 卌 卌 |	21	12.5
17–24	16.5–24.5	卌 卌 |	11	20.5
25–32	24.5–32.5	卌 |	6	28.5
33–40	32.5–40.5	||||	4	36.5
41–48	40.5–48.5	||||	4	44.5

Frequency Table of One-Way Commuting Distances for 60 Downtown Dallas Workers (Data in Miles)

Table 2-2

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Frequency Tables

The sample size is $n = 60$. Notice that the sample size is the total of all the frequencies. Therefore, the relative frequency for the first class (the class from 1 to 8) is

$$\text{Relative frequency} = \frac{f}{n} = \frac{14}{60} \approx 0.23$$

The symbol ≈ means "approximately equal to." We use the symbol because we rounded the relative frequency. Relative frequencies for the other classes are computed in a similar way.

Frequency Tables

The total of the relative frequencies should be 1.

However, rounded results may make the total slightly higher or lower than 1.

Frequency Tables

Procedure:

HOW TO MAKE A FREQUENCY TABLE

1. Determine the number of classes and the corresponding class width.
2. Create the distinct classes. We use the convention that the *lower class limit* of the first class is the smallest data value. Add the class width to this number to get the *lower class limit* of the next class.
3. Fill in *upper class limits* to create distinct classes that accommodate all possible data values from the data set.
4. Tally the data into classes. Each data value should fall into exactly one class. Total the tallies to obtain each *class frequency.*
5. Compute the *midpoint* (class mark) for each class.
6. Determine the *class boundaries*.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Frequency Tables

Procedure:

HOW TO MAKE A RELATIVE-FREQUENCY TABLE

First make a frequency table. Then, for each class, compute the *relative frequency*, f/n, where f is the class frequency and n is the total sample size.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Histograms and Relative-Frequency Histograms

Histograms and Relative-Frequency Histograms

Histograms and *relative-frequency histograms* provide effective visual displays of data organized into frequency tables. In these graphs, we use bars to represent each class, where the width of the bar is the class width.

For histograms, the height of the bar is the class frequency, whereas for relative-frequency histograms, the height of the bar is the relative frequency of that class.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Histograms and Relative-Frequency Histograms

Procedure:

HOW TO MAKE A HISTOGRAM OR A RELATIVE-FREQUENCY HISTOGRAM

1. Make a frequency table (including relative frequencies) with the designated number of classes.
2. Place class boundaries on the horizontal axis and frequencies or relative frequencies on the vertical axis.
3. For each class of the frequency table, draw a bar whose width extends between corresponding class boundaries. For histograms, the height of each bar is the corresponding class frequency. For relative-frequency histograms, the height of each bar is the corresponding class relative frequency.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 2 – *Histogram and Relative-Frequency Histogram*

Make a histogram and a relative-frequency histogram with six bars for the data in Table 2-1 showing one-way commuting distances.

13	47	10	3	16	20	17	40	4	2
7	25	8	21	19	15	3	17	14	6
12	45	1	8	4	16	11	18	23	12
6	2	14	13	7	15	46	12	9	18
34	13	41	28	36	17	24	27	29	9
14	26	10	24	37	31	8	16	12	16

One-Way Commuting Distances (in Miles) for 60 Workers in Downtown Dallas

Table 2-1

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 2 – *Solution*

The first step is to make a frequency table and a relative-frequency table with six classes. We'll use Table 2-2 and Table 2-3.

Class Limits Lower–Upper	Class Boundaries Lower–Upper	Tally	Frequency	Class Midpoint
1–8	0.5–8.5	𝍸 𝍸 ||||	14	4.5
9–16	8.5–16.5	𝍸 𝍸 𝍸 𝍸 |	21	12.5
17–24	16.5–24.5	𝍸 𝍸 |	11	20.5
25–32	24.5–32.5	𝍸 |	6	28.5
33–40	32.5–40.5	||||	4	36.5
41–48	40.5–48.5	||||	4	44.5

Frequency Table of One-Way Commuting Distances for 60 Downtown Dallas Workers (Data in Miles)

Table 2-2

Example 2 – *Solution*

cont'd

Class	Frequency f	Relative Frequency f/n
1–8	14	$14/60 \approx 0.23$
9–16	21	$21/60 \approx 0.35$
17–24	11	$11/60 \approx 0.18$
25–32	6	$6/60 \approx 0.10$
33–40	4	$4/60 \approx 0.07$
41–48	4	$4/60 \approx 0.07$

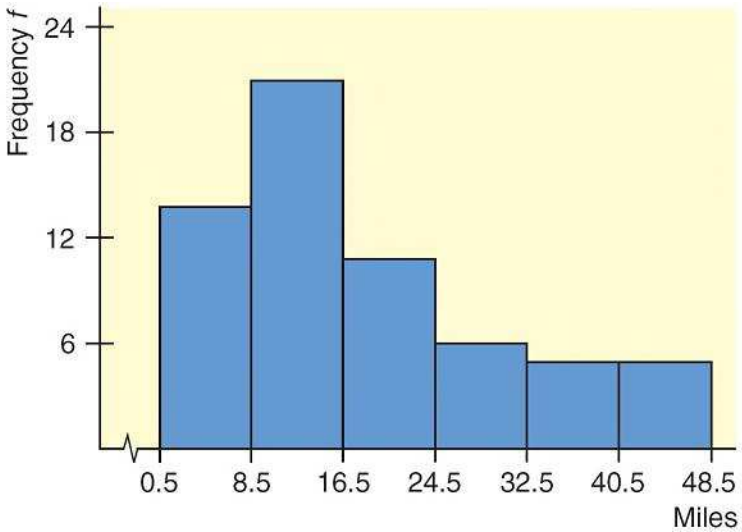
Relative Frequencies of One-Way Commuting Distances

Table 2-3

Example 2 – *Solution*

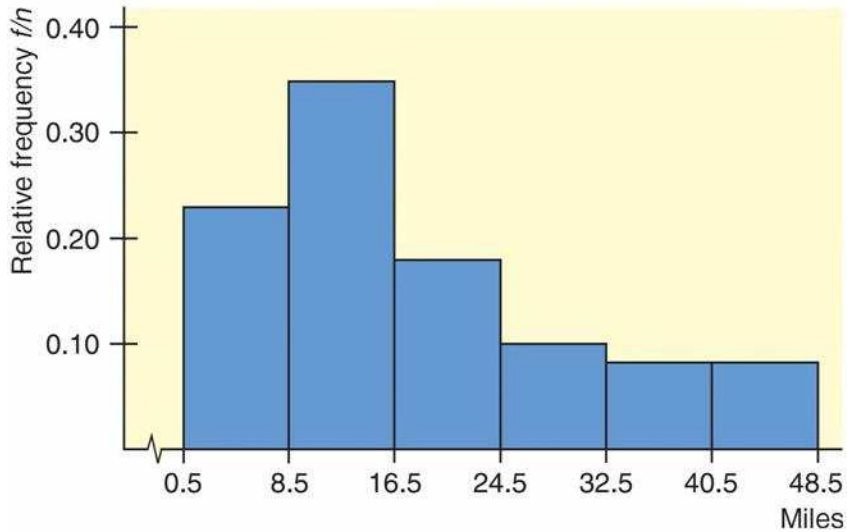
cont'd

Figures 2-2 and 2-3 show the histogram and relative-frequency histogram. In both graphs, class boundaries are marked on the horizontal axis.



Histogram for Dallas Commuters: One-Way Commuting Distances

Figure 2-2



Relative-Frequency Histogram for Dallas Commuters: One-Way Commuting Distances

Figure 2-3

Example 2 – *Solution*

cont'd

For each class of the frequency table, make a corresponding bar with horizontal width extending from the lower boundary to the upper boundary of the respective class.

For a histogram, the height of each bar is the corresponding class frequency.

For a relative-frequency histogram, the height of each bar is the corresponding relative frequency.

Example 2 – *Solution*

cont'd

Notice that the basic shapes of the graphs are the same. The only difference involves the vertical axis.

The vertical axis of the histogram shows frequencies, whereas that of the relative-frequency histogram shows relative frequencies.

Distribution Shapes

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Distribution Shapes

Histograms are valuable and useful tools. If the raw data came from a random sample of population values, the histogram constructed from the sample values should have a distribution shape that is reasonably similar to that of the population.

Several terms are commonly used to describe histograms and their associated population distributions.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Distribution Shapes

(a) ***Mound-shaped symmetrical:*** This term refers to a histogram in which both sides are (more or less) the same when the graph is folded vertically down the middle. Figure 2-8(a) shows a typical mound-shaped symmetrical histogram.

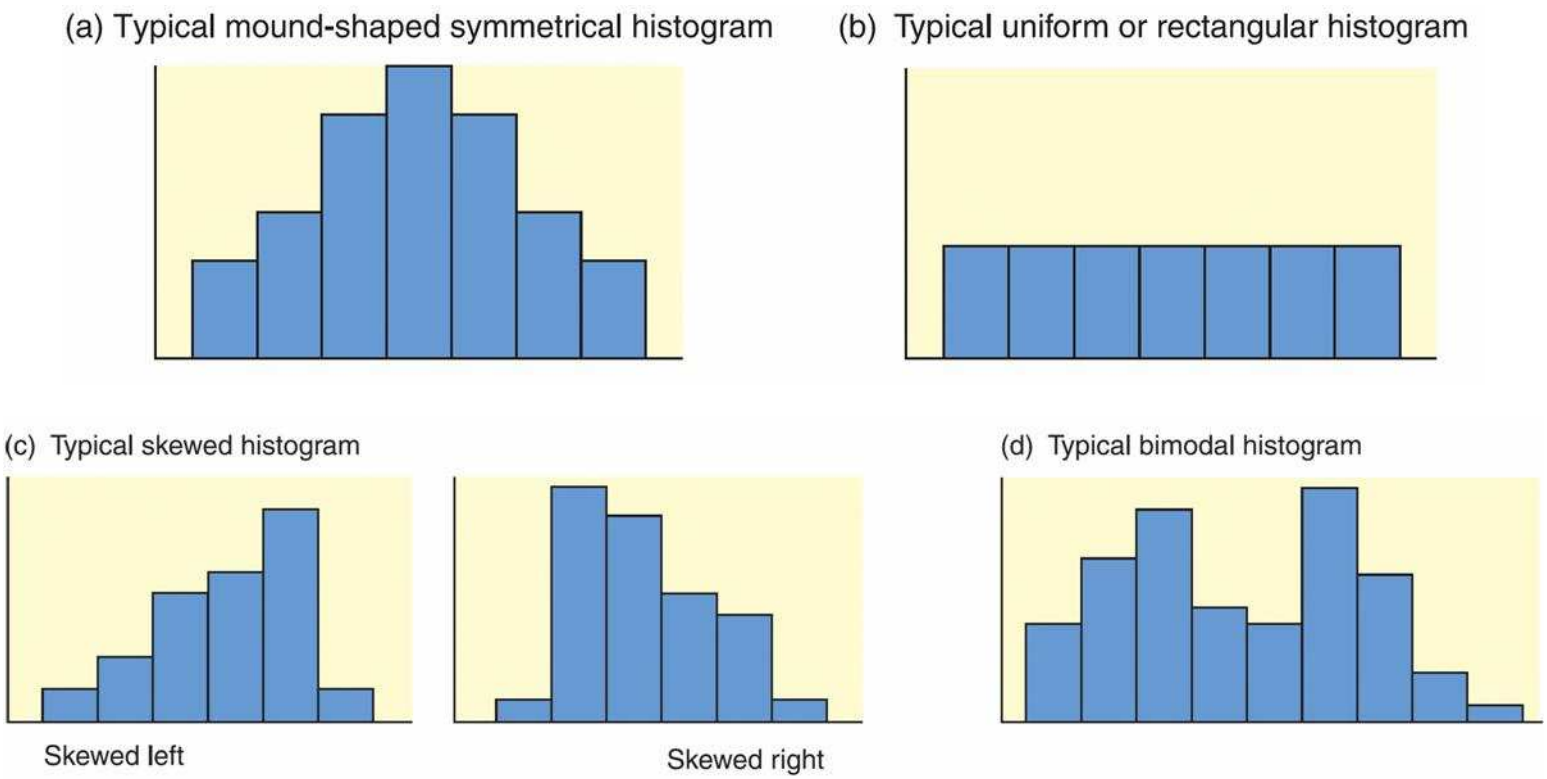
(b) ***Uniform or rectangular:*** These terms refer to a histogram in which every class has equal frequency. From one point of view, a uniform distribution is symmetrical with the added property that the bars are of the same height. Figure 2-8(b) illustrates a typical histogram with a uniform shape.

(c) ***Skewed left or skewed right:*** These terms refer to a histogram in which one tail is stretched out longer than the other. The direction of skewness is on the side of the *longer* tail. So, if the longer tail is on the left, we say the histogram is skewed to the left. Figure 2-8(c) shows a typical histogram skewed to the left and another skewed to the right.

(d) ***Bimodal:*** This term refers to a histogram in which the two classes with the largest frequencies are separated by at least one class. The top two frequencies of these classes may have slightly different values. This type of situation sometimes indicates that we are sampling from two different populations. Figure 2-8(d) illustrates a typical histogram with a bimodal shape.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Distribution Shapes



Types of Histograms

Figure 2-8

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Cumulative-Frequency Tables and Ogives

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Cumulative-Frequency Tables and Ogives

Sometimes we want to study cumulative totals instead of frequencies. Cumulative frequencies tell us how many data values are smaller than an upper class boundary.

Once we have a frequency table, it is a fairly straightforward matter to add a column of cumulative frequencies.

The **cumulative frequency** for a class is the sum of the frequencies for *that class* and *all previous classes*.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Cumulative-Frequency Tables and Ogives

An *ogive (*pronounced "oh-jī ve") is a graph that displays cumulative frequencies.

Procedure:

HOW TO MAKE AN OGIVE

1. Make a frequency table showing class boundaries and cumulative frequencies.
2. For each class, make a dot over the *upper class boundary* at the height of the cumulative class frequency. The coordinates of the dots are (upper class boundary, cumulative class frequency). Connect these dots with line segments.
3. By convention, an ogive begins on the horizontal axis at the lower class boundary of the first class.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 3 – *Cumulative-Frequency Table and Ogive*

Aspen, Colorado, is a world-famous ski area. If the daily high temperature is above 40°F, the surface of the snow tends to melt. It then freezes again at night.

This can result in a snow crust that is icy. It also can increase avalanche danger.

Example 3 – *Cumulative-Frequency Table and Ogive*

cont'd

Table 2-11 gives a summary of daily high temperatures (°F) in Aspen during the 151-day ski season.

Class Boundaries			
Lower	Upper	Frequency	Cumulative Frequency
10.5	20.5	23	23
20.5	30.5	43	66 (sum 23 + 43)
30.5	40.5	51	117 (sum 66 + 51)
40.5	50.5	27	144 (sum 117 + 27)
50.5	60.5	7	151 (sum 144 + 7)

High Temperatures During the Aspen Ski Season (°F)

Table 2-11

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 3 – *Cumulative-Frequency Table and Ogive*

cont'd

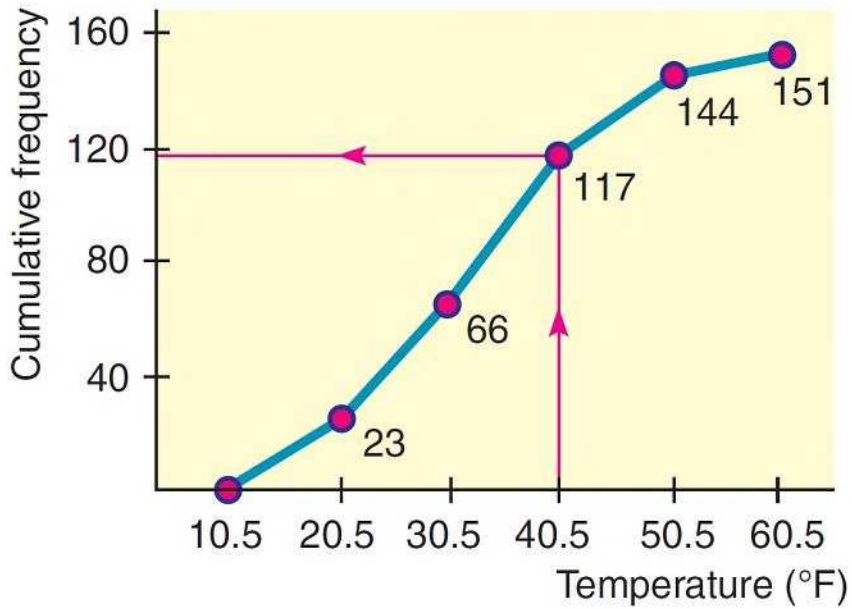
a. The cumulative frequency for a class is computed by adding the frequency of that class to the frequencies of previous classes. Table 2-11 shows the cumulative frequencies.

b. To draw the corresponding ogive, we place a dot at cumulative frequency 0 on the lower class boundary of the first class. Then we place dots over the *upper class boundaries* at the height of the cumulative class frequency for the corresponding class.

Example 3 – *Cumulative-Frequency Table and Ogive*

cont'd

Finally, we connect the dots. Figure 2-9 shows the corresponding ogive.



Ogive for Daily High Temperatures (°F) During Aspen Ski Season

Figure 2.9

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 3 – *Cumulative-Frequency Table and Ogive*

cont'd

c. Looking at the ogive, estimate the total number of days with a high temperature lower than or equal to 40°F.

Solution:

The red lines on the ogive in Figure 2-9, we see that 117 days have had high temperatures of no more than 40°F.

Organizing Data

2

Copyright © Cengage Learning. All rights reserved.

Section 2.2

Bar Graphs, Circle Graphs, and Time-Series Graphs

Copyright © Cengage Learning. All rights reserved.

Focus Points

- Determine types of graphs appropriate for specific data.
- Construct bar graphs, Pareto charts, circle graphs, and time-series graphs.
- Interpret information displayed in graphs.

Bar Graphs, Circle Graphs, and Time-Series Graphs

Histograms provide a useful visual display of the distribution of data.

However, the data must be quantitative. In this section, we examine other types of graphs, some of which are suitable for qualitative or category data as well.

Let's start with *bar graphs*. These are graphs that can be used to display quantitative or qualitative data.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Bar Graphs, Circle Graphs, and Time-Series Graphs

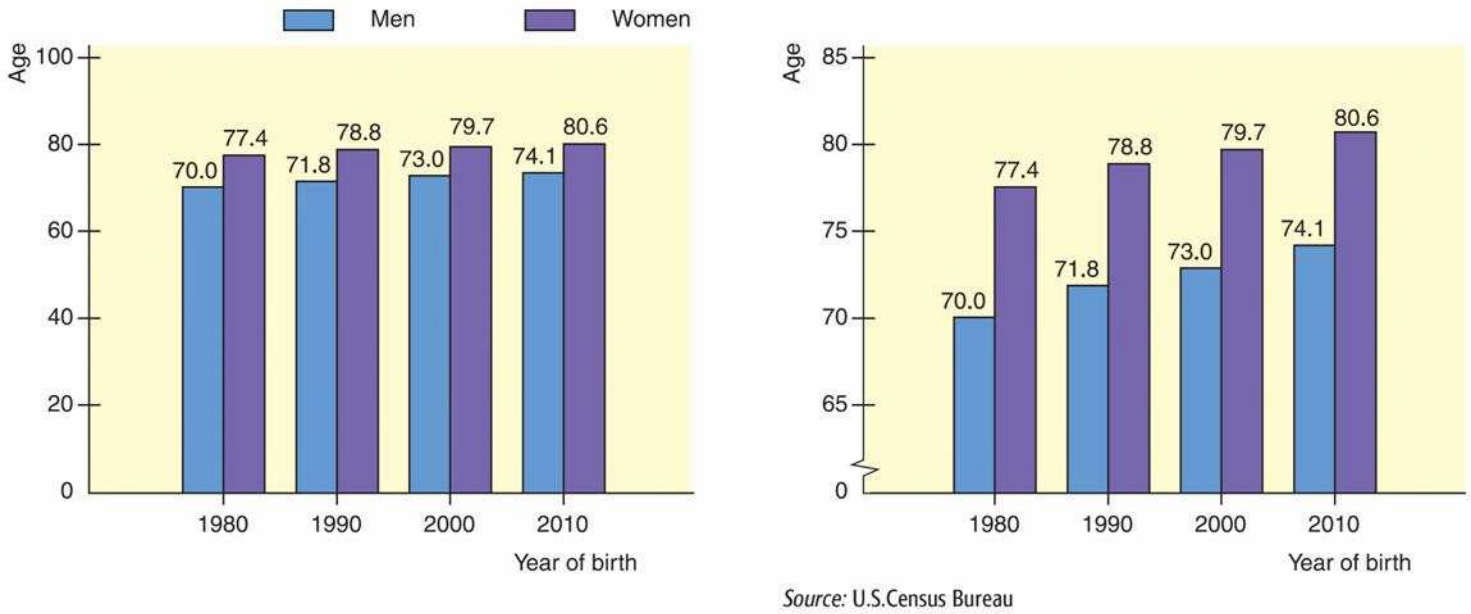
Features of a Bar Graph

1. Bars can be vertical or horizontal.
2. Bars are of uniform width and uniformly spaced.
3. The lengths of the bars represent values of the variable being displayed, the frequency of occurrence, or the percentage of occurrence. The same measurement scale is used for the length of each bar.
4. The graph is well annotated with title, labels for each bar, and vertical scale or actual value for the length of each bar.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 4 – *Bar Graph*

Figure 2-11 shows two bar graphs depicting the life expectancies for men and women born in the designated year. Let's analyze the features of these graphs.



Life Expectancy

Figure 2-11

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 4 – *Solution*

The graphs are called *clustered bar graphs* because there are two bars for each year of birth.

One bar represents the life expectancy for men, and the other represents the life expectancy for women.

The height of each bar represents the life expectancy (in years).

Bar Graphs, Circle Graphs, and Time-Series Graphs

An important feature illustrated in Figure 2-11(b) is that of a *changing scale*. Notice that the scale between 0 and 65 is compressed.

The changing scale amplifies the apparent difference between life spans for men and women, as well as the increase in life spans from those born in 1980 to the projected span of those born in 2010.

Changing Scale

Whenever you use a change in scale in a graphic, warn the viewer by using a squiggle ~ on the changed axis. Sometimes, if a single bar is unusually long, the bar length is compressed with a squiggle in the bar itself.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Bar Graphs, Circle Graphs, and Time-Series Graphs

A **Pareto chart** is a bar graph in which the bar height represents frequency of an event. In addition, the bars are arranged from left to right according to decreasing height.

Another popular pictorial representation of data is the *circle graph* or *pie chart*. It is relatively safe from misinterpretation and is especially useful for showing the division of a total quantity into its component parts.

The total quantity, or 100%, is represented by the entire circle. Each wedge of the circle represents a component part of the total.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Bar Graphs, Circle Graphs, and Time-Series Graphs

These proportional segments are usually labeled with corresponding percentages of the total.

In a **circle graph** or **pie chart,** wedges of a circle visually display proportional parts of the total population that share a common characteristic.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Bar Graphs, Circle Graphs, and Time-Series Graphs

We will use a *time-series graph*. A time-series graph is a graph showing data measurements in chronological order.

To make a time-series graph, we put time on the horizontal scale and the variable being measured on the vertical scale. In a basic time-series graph, we connect the data points by line segments.

In a **time-series graph,** data are plotted in order of occurrence at regular intervals over a period of time.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 5 – *Time-Series Graph*

Suppose you have been in the walking/jogging exercise program for 20 weeks, and for each week you have recorded the distance you covered in 30 minutes. Your data log is shown in Table 2-14.

Week	1	2	3	4	5	6	7	8	9	10
Distance	1.5	1.4	1.7	1.6	1.9	2.0	1.8	2.0	1.9	2.0
Week	11	12	13	14	15	16	17	18	19	20
Distance	2.1	2.1	2.3	2.3	2.2	2.4	2.5	2.6	2.4	2.7

Distance (in Miles) Walked/Jogged in 30 Minutes

Table 2-14

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 5(a) – *Time-Series Graph* cont'd

Make a time-series graph.

Solution:

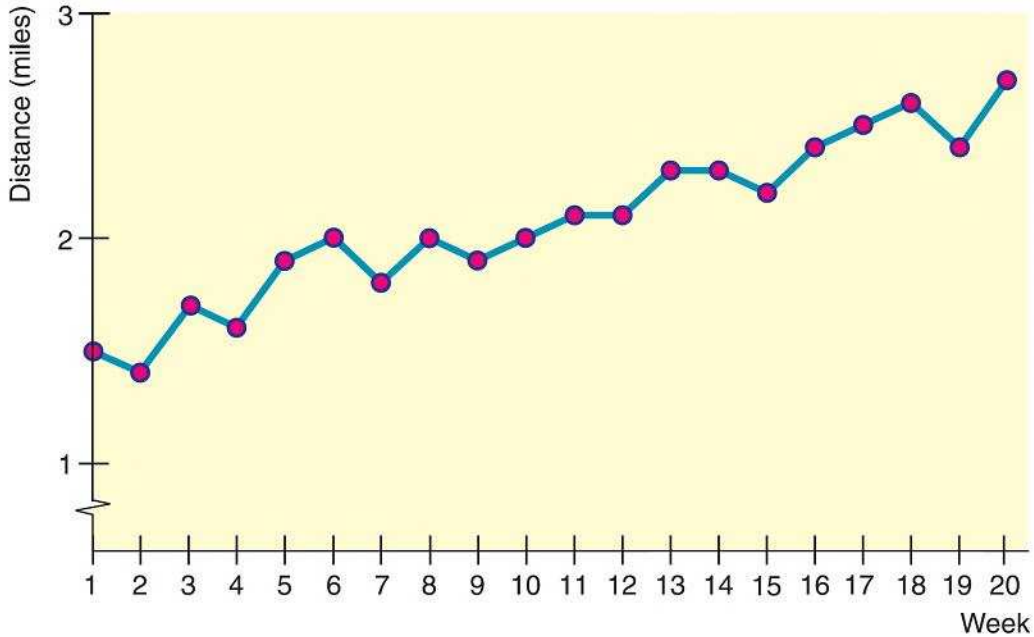
The data are appropriate for a time-series graph because they represent the same measurement (distance covered in a 30-minute period) taken at different times.

The measurements are also recorded at equal time intervals (every week). To make our time-series graph, we list the weeks in order on the horizontal scale. Above each week, plot the distance covered that week on the vertical scale.

Example 5(a) – *Solution*

cont'd

Then connect the dots. Figure 2-14 shows the time-series graph. Be sure the scales are labeled.



Time-Series Graph of Distance (in miles) Jogged in 30 Minutes

Figure 2-14

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 5(b) – *Time-Series Graph* cont'd

From looking at Figure 2-14, can you detect any patterns?

Solution:

There seems to be an upward trend in distance covered. The distances covered in the last few weeks are about a mile farther than those for the first few weeks.

However, we cannot conclude that this trend will continue. Perhaps you have reached your goal for this training activity and now wish to maintain a distance of about 2.5 miles in 30 minutes.

Bar Graphs, Circle Graphs, and Time-Series Graphs

Data sets composed of similar measurements taken at regular intervals over time are called *time series*.

Time series are often used in economics, finance, sociology, medicine, and any other situation in which we want to study or monitor a similar measure over a period of time. A time-series graph can reveal some of the main features of a time series.

Time-series data consist of measurements of the same variable for the same subject taken at regular intervals over a period of time.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Bar Graphs, Circle Graphs, and Time-Series Graphs

Procedure:

HOW TO DECIDE WHICH TYPE OF GRAPH TO USE

Bar graphs are useful for quantitative or qualitative data. With qualitative data, the frequency or percentage of occurrence can be displayed. With quantitative data, the measurement itself can be displayed, as was done in the bar graph showing life expectancy. Watch that the measurement scale is consistent or that a jump scale squiggle is used.

Pareto charts identify the frequency of events or categories in decreasing order of frequency of occurrence.

Circle graphs display how a *total* is dispersed into several categories. The circle graph is very appropriate for qualitative data, or any data for which percentage of occurrence makes sense. Circle graphs are most effective when the number of categories or wedges is 10 or fewer.

Time-series graphs display how data change over time. It is best if the units of time are consistent in a given graph. For instance, measurements taken every day should not be mixed on the same graph with data taken every week.

For any graph: Provide a title, label the axes, and identify units of measure. As Edward Tufte suggests in his book *The Visual Display of Quantitative Information*, don't let artwork or skewed perspective cloud the clarity of the information displayed.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Copyright © Cengage Learning. All rights reserved.

Section 2.3

Stem-and-Leaf Displays

Copyright © Cengage Learning. All rights reserved.

Focus Points

- Construct a stem-and-leaf display from raw data.
- Use a stem-and-leaf display to visualize data distribution.
- Compare a stem-and-leaf display to a histogram.

Exploratory Data Analysis

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Exploratory Data Analysis

Together with histograms and other graphics techniques, the stem-and-leaf display is one of many useful ways of studying data in a field called *exploratory data analysis* (often abbreviated as *EDA*).

John W. Tukey wrote one of the definitive books on the subject, *Exploratory Data Analysis* (Addison-Wesley).

Another very useful reference for EDA techniques is the book *Applications, Basics, and Computing of Exploratory Data Analysis* by Paul F. Velleman and David C. Hoaglin (Duxbury Press).

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Exploratory Data Analysis

Exploratory data analysis techniques are particularly useful for detecting patterns and extreme data values.

They are designed to help us explore a data set, to ask questions we had not thought of before, or to pursue leads in many directions.

EDA techniques are similar to those of an explorer. An explorer has a general idea of destination but is always alert for the unexpected.

Exploratory Data Analysis

An explorer needs to assess situations quickly and often simplify and clarify them. An explorer makes pictures—that is, maps showing the relationships of landscape features.

The aspects of rapid implementation, visual displays such as graphs and charts, data simplification, and robustness (that is, analysis that is not influenced much by extreme data values) are key ingredients of EDA techniques.

Exploratory Data Analysis

In addition, these techniques are good for exploration because they require very few prior assumptions about the data.

EDA methods are especially useful when our data have been gathered for general interest and observation of subjects.

For example, we may have data regarding the ages of applicants to graduate programs. We don't have a specific question in mind.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Exploratory Data Analysis

We want to see what the data reveal. Are the ages fairly uniform or spread out?

Are there exceptionally young or old applicants? If there are, we might look at other characteristics of these applicants, such as field of study.

EDA methods help us quickly absorb some aspects of the data and then may lead us to ask specific questions to which we might apply methods of traditional statistics.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Exploratory Data Analysis

In contrast, when we design an experiment to produce data to answer a specific question, we focus on particular aspects of the data that are useful to us.

If we want to determine the average highway gas mileage of a specific sports car, we use that model car in well-designed tests.

We don’t need to worry about unexpected road conditions, poorly trained drivers, different fuel grades, sudden stops and starts, etc. Our experiment is designed to control outside factors.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Exploratory Data Analysis

Consequently, we do not need to “explore” our data as much. We can often make valid assumptions about the data.

Methods of traditional statistics will be very useful to analyze such data and answer our specific questions.

Stem-and-Leaf Display

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Stem-and-Leaf Display

In this text, we will introduce the EDA techniques: stem-and-leaf displays.

A **stem-and-leaf display** is a method of exploratory data analysis that is used to rank-order and arrange data into groups.

We know that frequency distributions and histograms provide a useful organization and summary of data. However, in a histogram, we lose most of the specific data values.

Stem-and-Leaf Display

A stem-and-leaf display is a device that organizes and groups data but allows us to recover the original data if desired.

In the next example, we will make a stem-and-leaf display.

Example 6 – *Stem-and-Leaf Display*

Many airline passengers seem weighted down by their carry-on luggage. Just how much weight are they carrying?

The carry-on luggage weights in pounds for a random sample of 40 passengers returning from a vacation to Hawaii were recorded (see Table 2-15).

30	27	12	42	35	47	38	36	27	35
22	17	29	3	21	0	38	32	41	33
26	45	18	43	18	32	31	32	19	21
33	31	28	29	51	12	32	18	21	26

Weights of Carry-On Luggage in Pounds

Table 2-15

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 6 – *Stem-and-Leaf Display* cont'd

To make a stem-and-leaf display, we break the digits of each data value into *two* parts.

The left group of digits is called a *stem*, and the remaining group of digits on the right is called a *leaf*.

We are free to choose the number of digits to be included in the stem.

The weights in our example consist of two-digit numbers.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual

Example 6 – *Stem-and-Leaf Display* cont'd

For a two-digit number, the stem selection is obviously the left digit.

In our case, the tens digits will form the stems, and the units digits will form the leaves.

For example, for the weight 12, the stem is 1 and the leaf is 2.

For the weight 18, the stem is again 1, but the leaf is 8.

Example 6 – *Stem-and-Leaf Display* cont'd

In the stem-and-leaf display, we list each possible stem once on the left and all its leaves in the same row on the right, as in Figure 2-15(a).

3 | 2 represents 32 lb

Stem	Leaves
0	3 0
1	2 7 8 8 9 2 8
2	7 7 2 9 1 6 1 8 9 1 6
3	0 5 8 6 5 8 2 3 2 1 2 3 1 2
4	2 7 1 5 3
5	1

(a) Leaves Not Ordered

3 | 2 represents 32 lb

Stem	Leaves
0	0 3
1	2 2 7 8 8 8 9
2	1 1 1 2 6 6 7 7 8 9 9
3	0 1 1 2 2 2 2 3 3 5 5 6 8 8
4	1 2 3 5 7
5	1

(b) Final Display with Leaves Ordered

Stem-and-Leaf Displays of Airline Carry-On Luggage Weights

Figure 2-15

Example 6 – *Stem-and-Leaf Display* cont'd

Finally, we order the leaves as shown in Figure 2-15(b). Figure 2-15 shows a stem-and-leaf display for the weights of carry-on luggage.

From the stem-and-leaf display in Figure 2-15, we see that two bags weighed 27 lb, one weighed 3 lb, one weighed 51 lb, and so on.

We see that most of the weights were in the 30-lb range, only two were less than 10 lb, and six were over 40 lb.

Example 6 – *Stem-and-Leaf Display* cont'd

Note that the lengths of the lines containing the leaves give the visual impression that a sideways histogram would present.

As a final step, we need to indicate the scale. This is usually done by indicating the value represented by the stem and one leaf.

Stem-and-Leaf Display

Procedure:

HOW TO MAKE A STEM-AND-LEAF DISPLAY

1. Divide the digits of each data value into two parts. The leftmost part is called the *stem* and the rightmost part is called the *leaf*.
2. Align all the stems in a vertical column from smallest to largest. Draw a vertical line to the right of all the stems.
3. Place all the leaves with the same stem in the same row as the stem, and arrange the leaves in increasing order.
4. Use a label to indicate the magnitude of the numbers in the display. We include the decimal position in the label rather than with the stems or leaves.

Understandable Statistics Concepts and Methods 12th edition Brase Solutions Manual