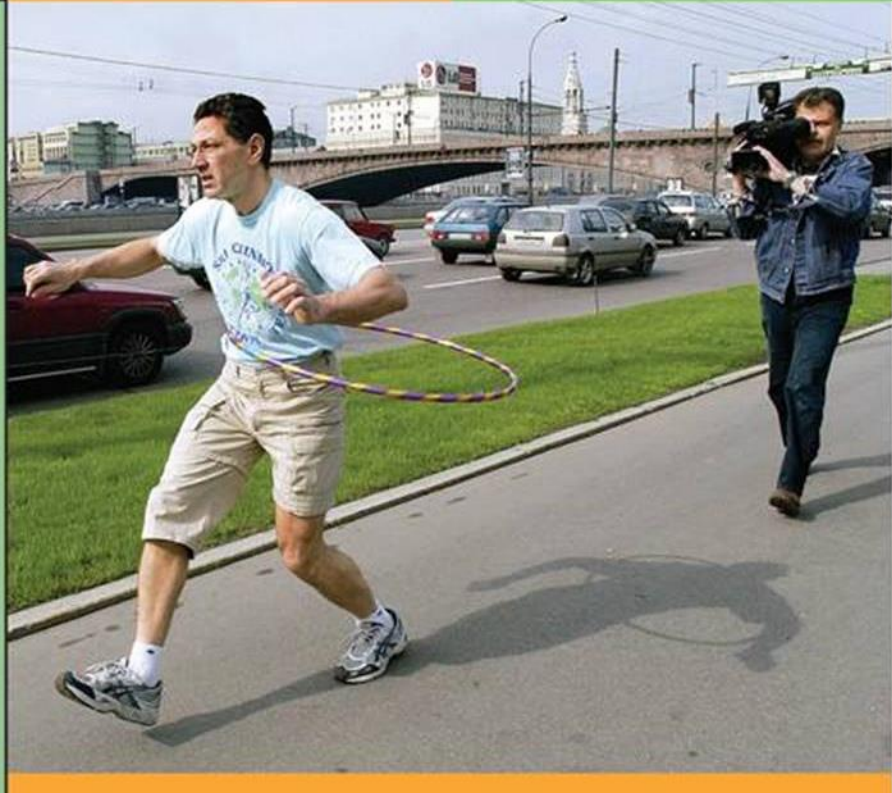


Organizing Data



2



**Section
2.3**

**Stem-and-Leaf
Displays**



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

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*.

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.

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.

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.

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.

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

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 (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

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.

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

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

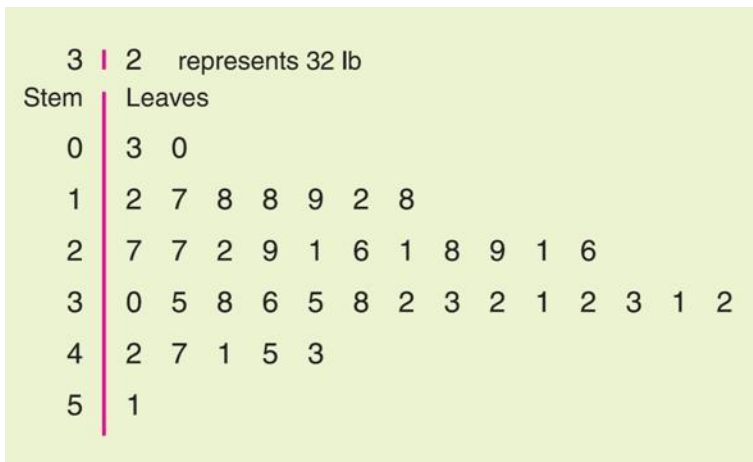
In our case, the tens digits will form the stems, and the units digits will form the leaves.

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

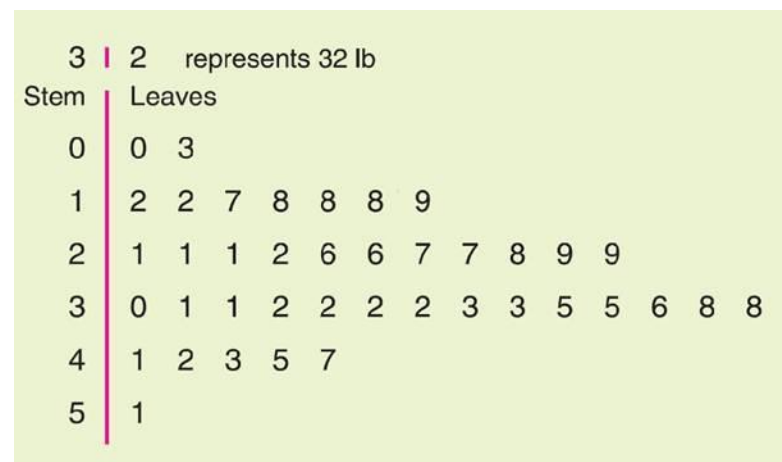
For the weight 18, the stem is again 1, but the leaf is 8.

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). Finally, we order the leaves as shown in Figure 2-15(b).



(a) Leaves Not Ordered



(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

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(b), 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.