

## 2.3: Stem and Leaf Displays

### Learning Objectives

- Create and interpret basic stem and leaf displays
- Create and interpret back-to-back stem and leaf displays
- Judge whether a stem and leaf display is appropriate for a given data set

A stem and leaf display is a graphical method of displaying data. It is particularly useful when your data are not too numerous. In this section, we will explain how to construct and interpret this kind of graph.

As usual, an example will get us started. Consider Table 2.3.1 that shows the number of touchdown passes (TD passes) thrown by each of the 31 teams in the National Football League in the 2000 season.

**Table 2.3.1:** Number of touchdown passes

37	33	33	32	29	28	28	23	22	
22	22	21	21	21	20	20	19	19	
18	18	18	18	16	15	14	14	14	(2.3.1)
12	12	9	6						

A stem and leaf display of the data is shown in Figure 2.3.1. The left portion of Figure 2.3.1 contains the stems. They are the numbers 3, 2, 1, and 0, arranged as a column to the left of the bars. Think of these numbers as 10's digits. A stem of 3, for example, can be used to represent the 10's digit in any of the numbers from 30 to 39. The numbers to the right of the bar are leaves, and they represent the 1's digits. Every leaf in the graph therefore stands for the result of adding the leaf to 10 times its stem.

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

**Figure 2.3.1:** Stem and leaf display of the number of touchdown passes

To make this clear, let us examine Figure 2.3.1 more closely. In the top row, the four leaves to the right of stem 3 are 2, 3, 3, and 7. Combined with the stem, these leaves represent the numbers 32, 33, 33, and 37, which are the numbers of TD passes for the first four teams in Table 2.3.1. The next row has a stem of 2 and 12 leaves. Together, they represent 12 data points, namely, two occurrences of 20 TD passes, three occurrences of 21 TD passes, three occurrences of 22 TD passes, one occurrence of 23 TD passes, two occurrences of 28 TD passes, and one occurrence of 29 TD passes. We leave it to you to figure out what the third row represents. The fourth row has a stem of 0 and two leaves. It stands for the last two entries in Table 2.3.1, namely 9 TD passes and 6 TD passes. (The latter two numbers may be thought of as 09 and 06.)

One purpose of a stem and leaf display is to clarify the shape of the distribution. You can see many facts about TD passes more easily in Figure 2.3.1 than in Table 2.3.1. For example, by looking at the stems and the shape of the plot, you can tell that most of the teams had between 10 and 29 passing TDs, with a few having more and a few having less. The precise numbers of TD passes can be determined by examining the leaves.

We can make our figure even more revealing by splitting each stem into two parts. Figure 2.3.2 shows how to do this. The top row is reserved for numbers from 35 to 39 and holds only the 37 TD passes made by the first team in Table 2.3.1. The second row is reserved for the numbers from 30 to 34 and holds the 32, 33, and 33 TD passes made by the next three teams in the table. You can see for yourself what the other rows represent.

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

**Figure 2.3.2:** Stem and leaf display with the stems split in two

Figure 2.3.2 is more revealing than Figure 1 because the latter figure lumps too many values into a single row. Whether you should split stems in a display depends on the exact form of your data. If rows get too long with single stems, you might try splitting them into two or more parts.

There is a variation of stem and leaf displays that is useful for comparing distributions. The two distributions are placed back to back along a common column of stems. The result is a "back-to-back stem and leaf graph." Figure 2.3.3 shows such a graph. It compares the numbers of TD passes in the 1998 and 2000 seasons. The stems are in the middle, the leaves to the left are for the 1998 data, and the leaves to the right are for the 2000 data. For example, the second-to-last row shows that in 1998 there were teams with 11, 12, and 13 TD passes, and in 2000 there were two teams with 12 and three teams with 14 TD passes.

11		4								
		3	7							
332		3	2	3	3					
8865		2	8	8	9					
44331110		2	0	0	1	1	1	2	2	3
987776665		1	5	6	8	8	8	8	9	9
321		1	2	2	4	4	4	4		
7		0	6	9						

**Figure 2.3.3:** Back-to-back stem and leaf display.

The left side shows the 1998 TD data and the right side shows the 2000 TD data. Figure 2.3.3 helps us see that the two seasons were similar, but that only in 1998 did any teams throw more than 40 TD passes.

There are two things about the football data that make them easy to graph with stems and leaves. First, the data are limited to whole numbers that can be represented with a one-digit stem and a one-digit leaf. Second, all the numbers are positive. If the data include numbers with three or more digits, or contain decimals, they can be rounded to two-digit accuracy.

Negative values are also easily handled. Let us look at another example.

Table 2.3.2 shows data from the case study Weapons and Aggression. Each value is the mean difference over a series of trials between the times it took an experimental subject to name aggressive words (like “punch”) under two conditions. In one condition, the words were preceded by a non-weapon word such as “bug.” In the second condition, the same words were preceded by a weapon word such as “gun” or “knife.” The issue addressed by the experiment was whether a preceding weapon word would speed up (or prime) pronunciation of the aggressive word compared to a non-weapon priming word. A positive difference implies greater priming of the aggressive word by the weapon word. Negative differences imply that the priming by the weapon word was less than for a neutral word.

**Table 2.3.2:** The effects of priming (thousandths of a second)

$$\begin{array}{cccccc} 43.2 & 42.9 & 35.6 & 25.6 & 25.4 & 23.6 \\ 20.5 & 19.9 & 14.4 & 12.7 & 11.3 & 10.2 \\ 10.0 & 9.1 & 7.5 & 5.4 & 4.7 & 3.8 & 2.1 & 1.2 \\ -0.2 & -6.3 & -6.7 & -8.8 & -10.4 & -10.5 & & \\ -14.9 & -14.9 & -15.0 & -18.5 & -27.4 & & & \end{array} \quad (2.3.5)$$

You see that the numbers range from 43.2 to -27.4. The first value indicates that one subject was 43.2 milliseconds faster pronouncing aggressive words when they were preceded by weapon words than when preceded by neutral words. The value -27.4 indicates that another subject was 27.4 milliseconds slower pronouncing aggressive words when they were preceded by weapon words.

The data are displayed with stems and leaves in Figure 2.3.4. Since stem and leaf displays can only portray two whole digits (one for the stem and one for the leaf), the numbers are first rounded. Thus, the value 43.2 is rounded to 43 and represented with a stem of 4 and a leaf of 3. Similarly, 42.9 is rounded to 43. To represent negative numbers, we simply use negative stems. For example, the bottom row of the figure represents the number  $-27$ . The second-to-last row represents the numbers  $-10, -10, -15$ , etc. Once again, we have rounded the original values from Table 2.3.2

$$\begin{array}{c|cccccccc} 4 & 3 & 3 & & & & & & \\ 3 & 6 & & & & & & & \\ 2 & 0 & 0 & 4 & 5 & 6 & & & \\ 1 & 0 & 0 & 1 & 3 & 4 & & & \\ 0 & 1 & 2 & 4 & 5 & 5 & 8 & 9 & \\ -0 & 0 & 6 & 7 & 9 & & & & \\ -1 & 0 & 0 & 5 & 5 & 5 & 9 & & \\ -2 & 7 & & & & & & & \end{array} \quad (2.3.6)$$

**Figure 2.3.4:** Stem and leaf display with negative numbers and rounding

Observe that the figure contains a row headed by "0" and another headed by "-0". The stem of 0 is for numbers between 0 and 9, whereas the stem of -0 is for numbers between 0 and -9. For example, the fifth row of the table holds the numbers 1, 2, 4, 5, 5, 8, 9 and the sixth row holds 0, -6, -7, and -9. Values that are exactly 0 before rounding should be split as evenly as possible between the "0" and "-0" rows. In Table 2.3.2, none of the values are 0 before rounding. The "0" that appears in the "-0" row comes from the original value of -0.2 in the table.

Although stem and leaf displays are unwieldy for large data sets, they are often useful for data sets with up to 200 observations. Figure 2.3.5 portrays the distribution of populations of 185 US cities in 1998. To be included, a city had to have between 100,000 and 500,000 residents.

[illegible]

**Figure 2.3.5:** Stem and leaf display of populations of 185 US cities with populations between 100,000 and 500,000 in 1998.

Since a stem and leaf plot shows only two-place accuracy, we had to round the numbers to the nearest 10,000. For example, the largest number (493,559) was rounded to 490,000 and then plotted with a stem of 4 and a leaf of 9. The fourth highest number (463,201) was rounded to 460,000 and plotted with a stem of 4 and a leaf of 6. Thus, the stems represent units of 100,000 and the leaves represent units of 10,000. Notice that each stem value is split into five parts: 0-1, 2-3, 4-5, 6-7, and 8-9.

Whether your data can be suitably represented by a stem and leaf graph depends on whether they can be rounded without loss of important information. Also, their extreme values must fit into two successive digits, as the data in Figure 2.3.5 fit into the 10,000 and 100,000 places (for leaves and stems, respectively). Deciding what kind of graph is best suited to displaying your data thus requires good judgment. Statistics is not just recipes!