# CONTENT OVERVIEW

This unit on stemplots is the first in a series of units to focus on graphical representation of quantitative data. The emphasis in this unit is not simply to construct a stemplot but to use the plot to interpret the data's story.

The clearest picture of the distribution of values of a variable is just that – a picture. A *stemplot* (or stem-and-leaf plot) is a simple kind of graph that is constructed using the numbers themselves. Here's an example of head sizes in inches of 30 male soldiers. The head size was measured by putting a tape measure around each soldier's forehead.

| | | | | | |
|---|---|---|---|---|---|
| 23.0 | 22.2 | 21.7 | 22.0 | 22.3 | 22.6 |
| 22.7 | 21.5 | 22.7 | 24.9 | 20.8 | 23.3 |
| 24.2 | 23.5 | 23.9 | 23.4 | 20.8 | 21.5 |
| 23.0 | 24.0 | 22.7 | 22.6 | 23.9 | 21.8 |
| 23.1 | 21.9 | 21.0 | 22.4 | 23.5 | 22.5 |

To make a stemplot of these measurements, we first separate each observation into a stem, which is the first digit or digits, and a leaf, the final digit. The stems can consist of any number of digits, but the leaves generally have only a single digit. For the head circumference data, the measurements range from about 21 inches to about 25 inches and are measured to tenths of an inch. We'll take the whole inches as stems and the tenths as leaves.

First arrange the stems in order with a vertical line to their right as shown in Figure 2.6.

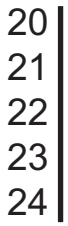


*Figure 2.6. Setting up the stem of the stemplot.*

Next, go through the list of observations, putting each leaf on the proper stem. The first soldier's head size was 23.0 inches, so we put leaf 0 on stem 23. The second head size is 22.2, so we put leaf 2 on stem 22. When we are finished, we have the display in Figure 2.7.

```
20 | 88
21 | 755890
22 | 2036777645
23 | 035940915
24 | 920
```

*Figure 2.7. Stemplot with unordered leaves.*

As a final step, arrange the leaves in order from smallest to largest. Figure 2.8 shows the completed stemplot. (This final step is unnecessary if technology is used to order the data from smallest to largest.)

```
20 | 88
21 | 055789
22 | 0234566777
23 | 001345599
24 | 029
```

*Figure 2.8. The completed stemplot.*

If there are too many stems with no leaves or only one leaf, it often helps to truncate the numbers and then to make a stemplot of the truncated numbers. (Truncation is faster than rounding.) If the leaves are crowded onto too few stems, expand the stem. For example, each stem can be split into two, one for leaf digits 0, 1, 2, 3, 4 and the other for leaf digits 5, 6, 7, 8, 9. (Or split each stem into five, using leaf digits 0 and 1, 2 and 3, 4 and 5, 6 and 7, and 8 and 9 for the five stems.)

Splitting stems can often reveal new information, as was the case of the fuel economy of Toyota's 2012 vehicles that was shown in Figures 2.3 and 2.4. Hence, don't be afraid to experiment with different stems or truncation to see what additional information might be learned. Finally, placing stemplots back-to-back is a good way to compare two datasets. (See Figure 2.5.)

Making the stemplot isn't the end in itself. It is a tool to help unlock the data's story. For example, the completed stemplot in Figure 2.8 gives a picture of the distribution of soldiers' head sizes. From the stemplot, we learn that the smallest head size was 20.8 and the largest was 24.9. The shape of the distribution is mound shaped (one peak). Although the two sides of the stemplot would not line up exactly if we folded the plot along the 22 stem, they come pretty close. So, we can say that this distribution is roughly symmetric. A middle value is somewhere in the 22-inch range (in other words, somewhere between 22.0 and 22.7).

The art of looking at stemplots intelligently is as important as the skill of making them. In looking at any distribution, always look first for the overall pattern of the distribution and then for any striking deviations from that pattern. In sizing up the overall pattern, look for and try to describe the following:

- center and spread
- one peak or several
- a regular shape, such as symmetric

For now, identify a center by looking at the stemplot and selecting a number that appears to best measure the middle of the distribution. (In later units, we will cover specific measures of center such as the mean and median).