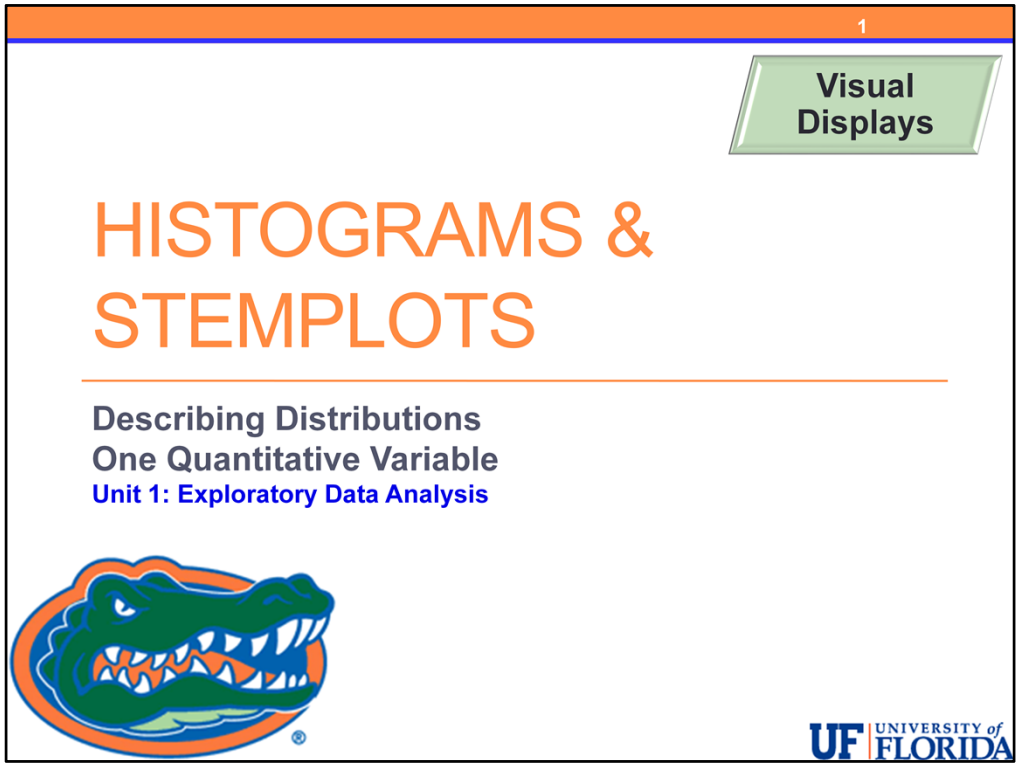
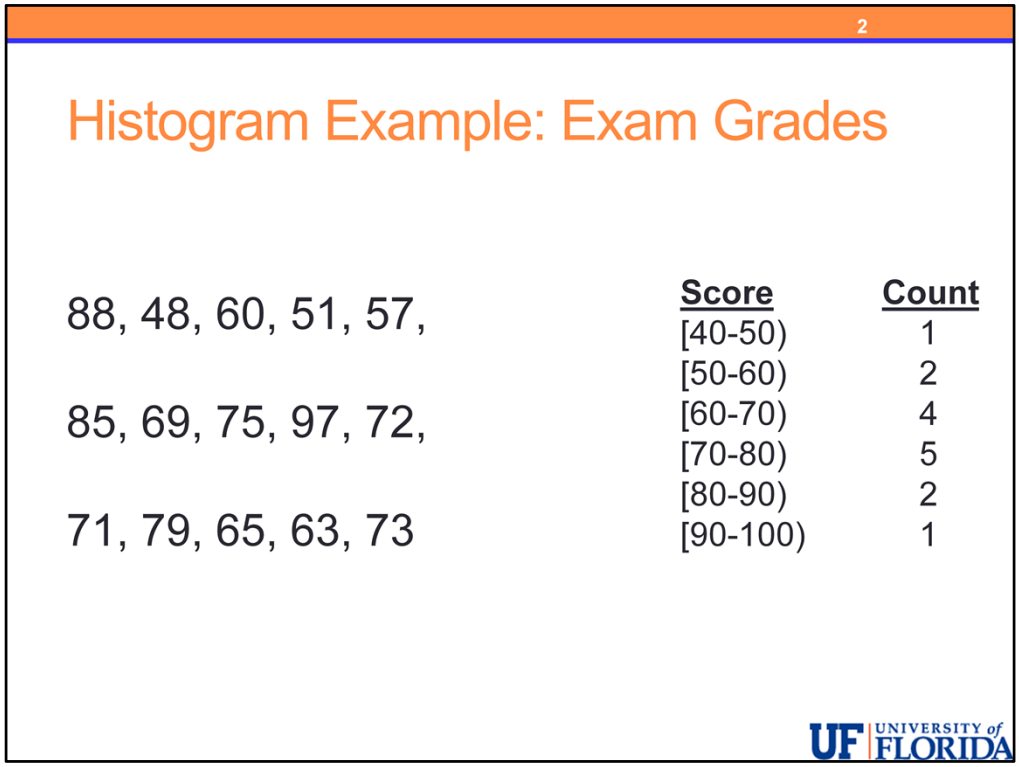


Although we rarely if ever need to construct graphs by hand, we will discuss the process of creating a histogram and stemplot in order to help you develop a deeper understanding of these visual displays

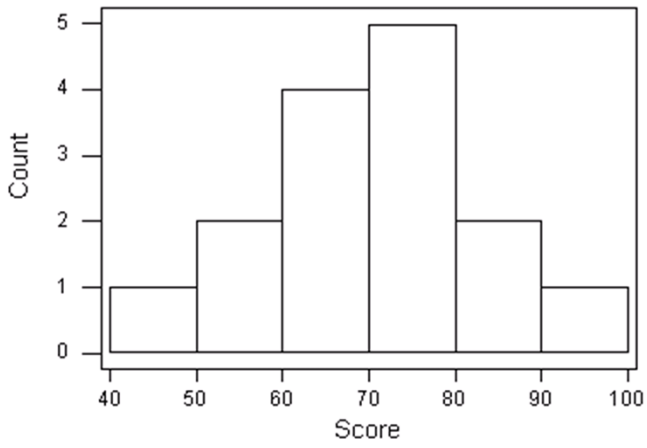


We will use this data on exam grades for 15 students to illustrate creating a histogram.

We break the data into intervals, being sure to define them so that no value is missed or counted in more than one interval. Computers use algorithms to achieve reasonable intervals for a wide range of datasets and many different algorithms are in use. We won't worry about this detail as we will use software to obtain our histograms in practice.

Once we have well-defined intervals, we count how many observations fall into each interval. For this example, we have intervals which begin at 40 and go to 100 by 10. The first interval consists of any value 40 or more but strictly less than 50. Any value of 50 will be placed in the next interval. In this dataset we do have one value, 60, which falls on the "boundary" and you can see that this value is counted in the interval [60, 70). Here we use the algebra notation with a bracket to indicate inclusion of the lower bound (60) and parentheses to indicate the exclusion of the upper bound (70).

Histogram Example: Exam Grades

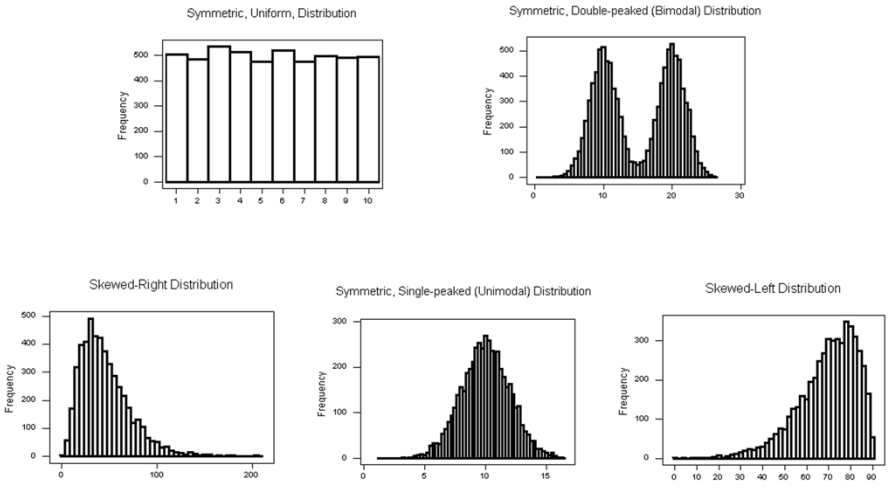


UF UNIVERSITY of FLORIDA

Once we have the frequency in each interval, we display this information in a histogram using the height of each bar to represent the frequency or percentage. Unlike bar charts, histograms have no gaps between the bars.

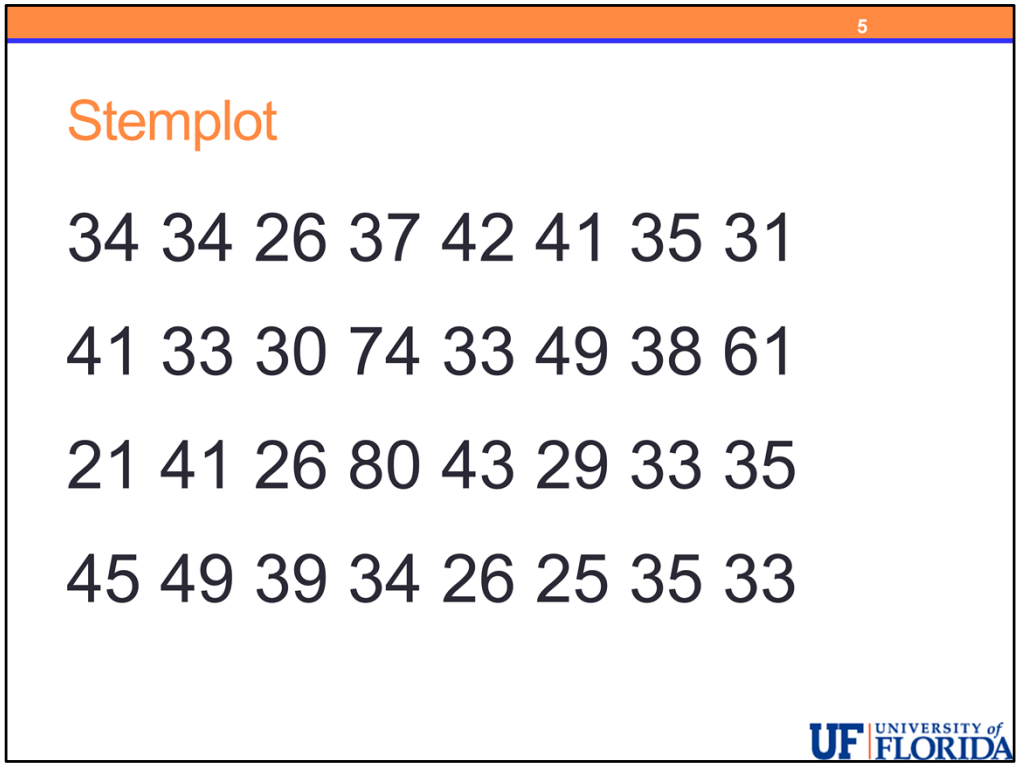
Although some information is lost - in that we can not recover the exact data - we can use these graphs to answer some questions about the data and to help us visualize the distribution.

Histograms

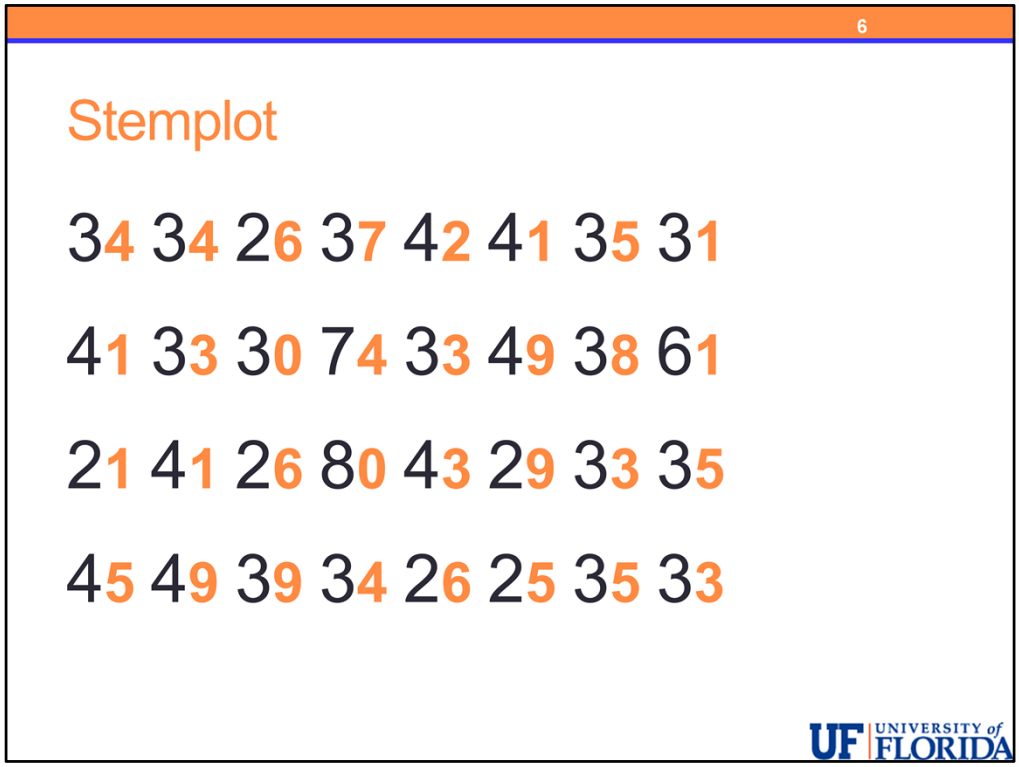


UF UNIVERSITY of FLORIDA

Here are a few other histograms to get you thinking about the possible variety we might see in distributions.



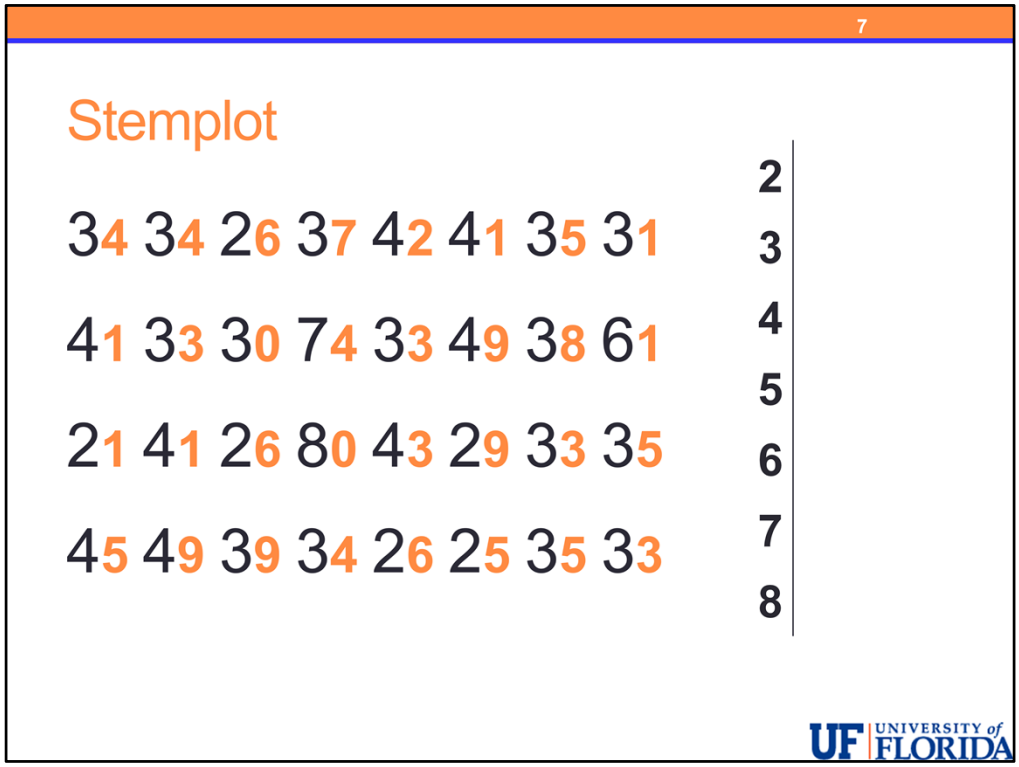
To create a stemplot, we first need to have data in which all values are rounded the same, we have that here with the data containing the ages of best actress Oscar winners.



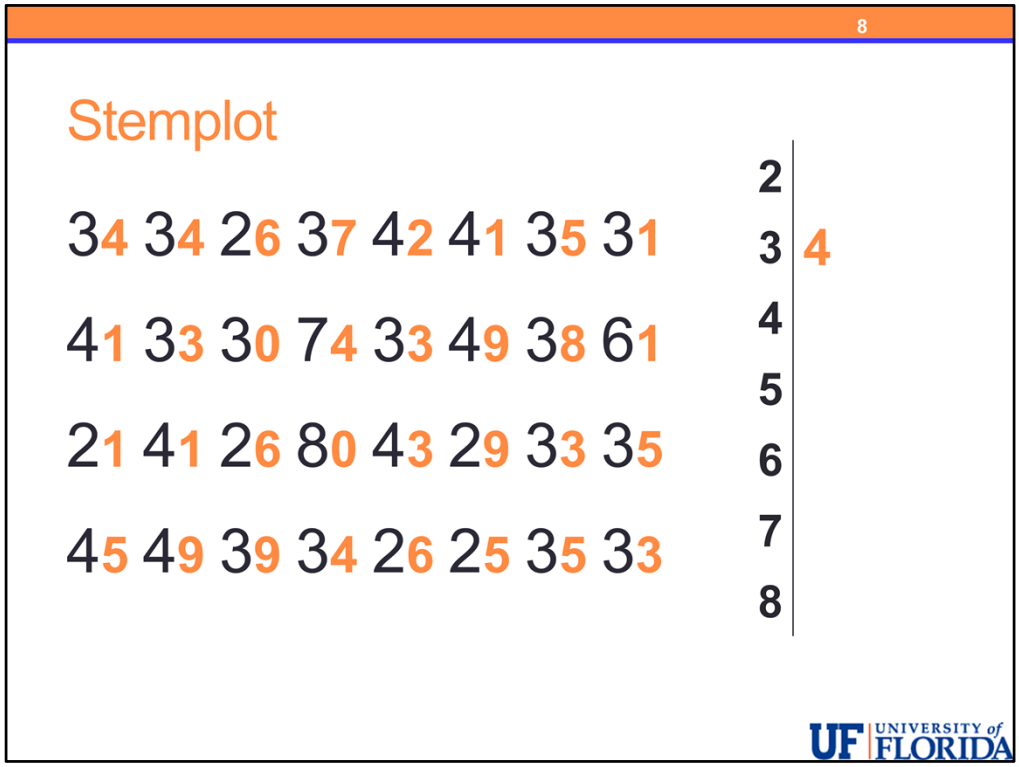
The right most digit becomes the leaf – presented here in bold and orange

All remaining digits become the stems – seen here in black with a larger font.

Our stems range from 2 through 8

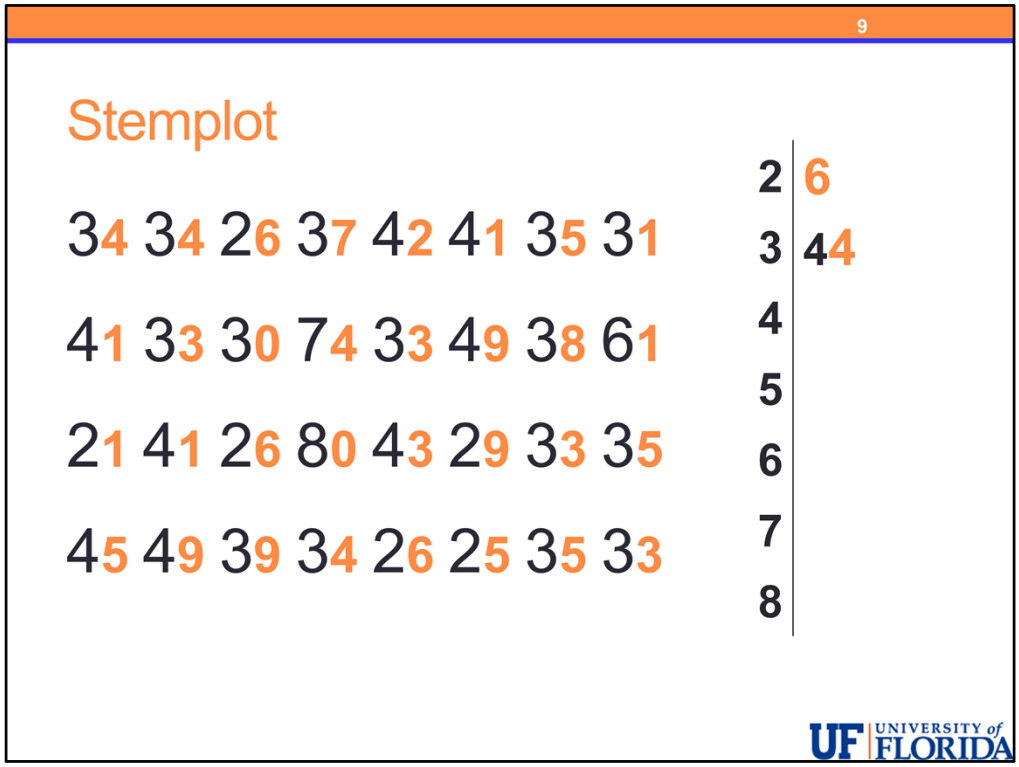


We write the stems to the left of a vertical line, starting at the top with the smallest stem, we put the stems in order. Don't skip any stems in the range of your data – even if there is no data for that stem.

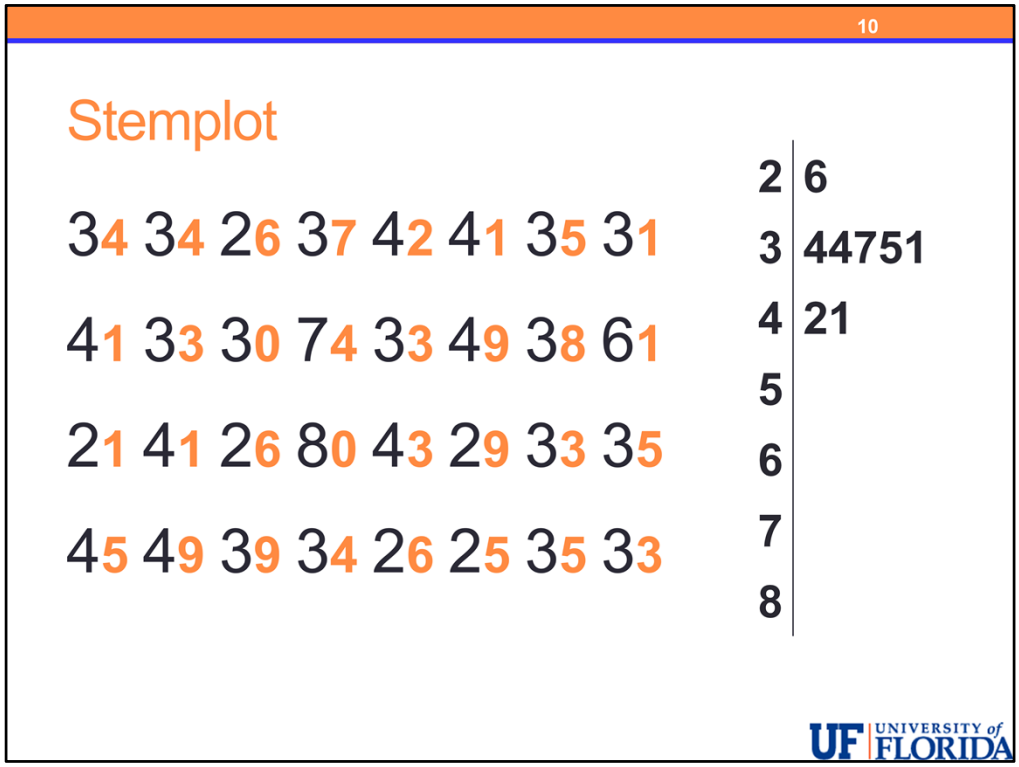


Then we add the leaves for each number to the right of the stem.

For 34, we place a leaf of 4 next to the stem for 3.

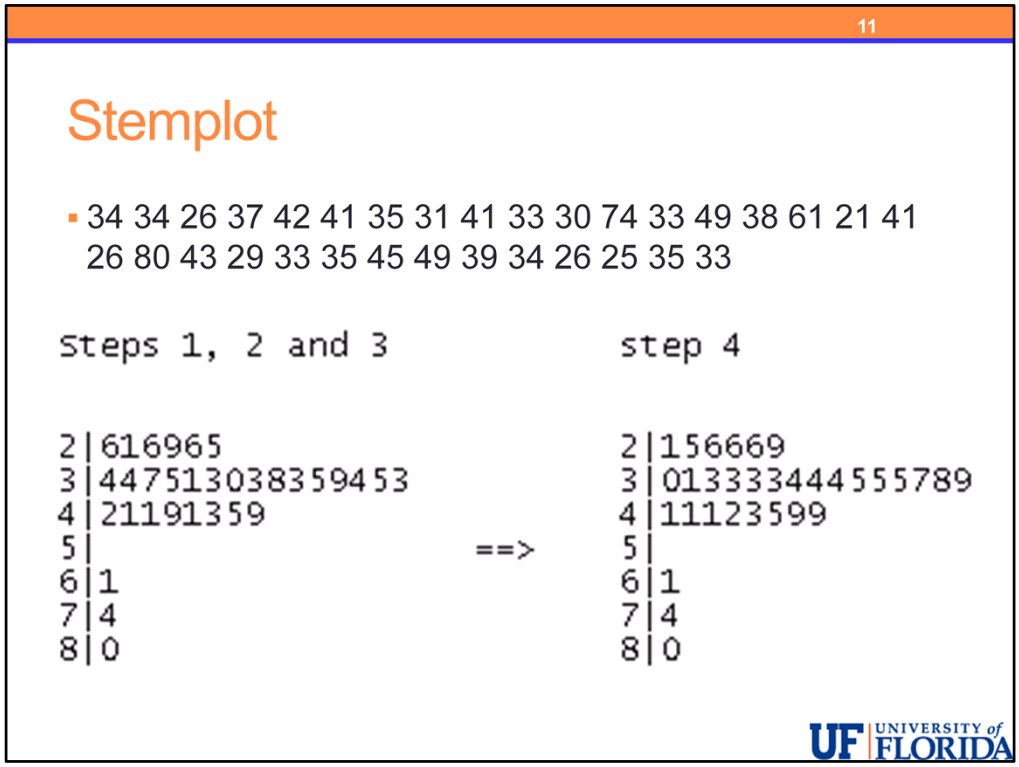


There was another 34 followed by a 26.



Continuing in this way through the first row we would get what you see here on the right

You can pause the video and finish the rest for yourself. Next we will show the final steps.



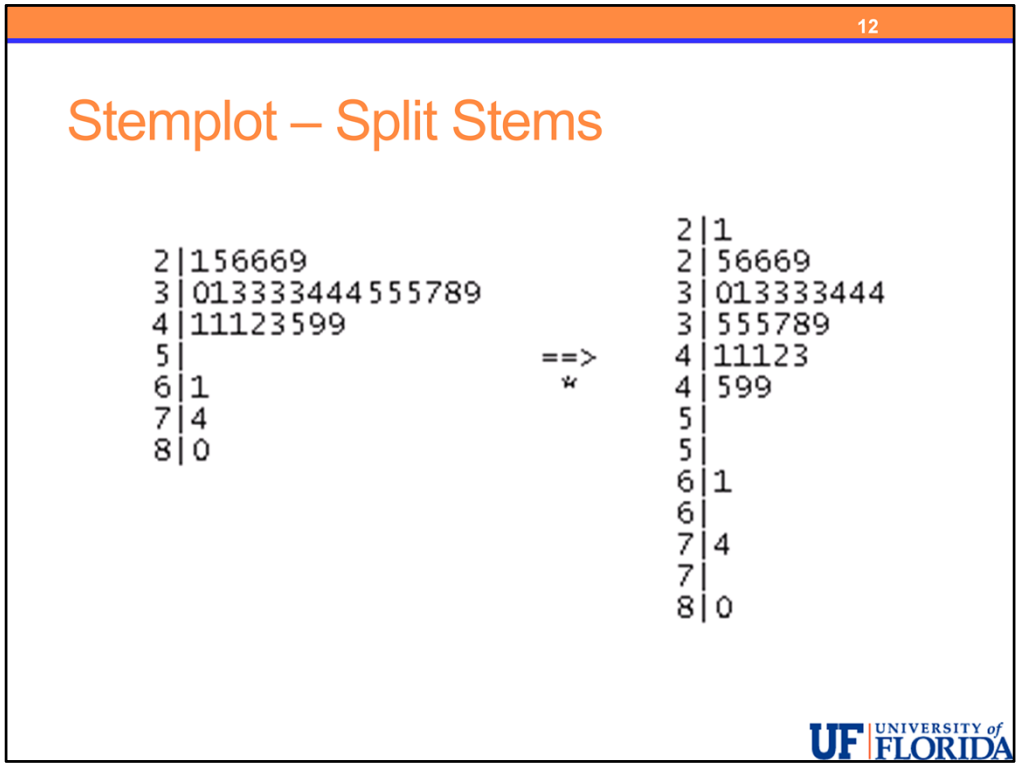
When we put all leaves in the display for our data we will have the stemplot on the lower left. The leaves are not in order (unless our original data was in order).

Next we sort the leaves to provide the final stemplot.

Note that it preserves the original data AND sorts the data.

These are two nice features of using a stemplot

(definitely helpful if you are ever stuck on a deserted island needing to quickly visualize the distribution of one quantitative variable!)

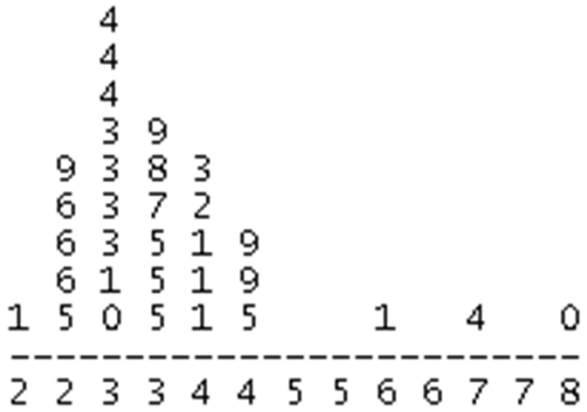


One additional option is to split the stems. This is useful whenever there are a large number of values in only a few stems.

There are a few easy ways to split stems. The easiest is to split the leaves for each stem equally into two groups of 5 digits; 0-4, and 5-9 which we have done here in the stemplot on the right. This gives two sets of leaves for each stem. You can see that we repeat the stem on the left. If a lower or upper stem is not needed for the data it does not need to be displayed, as for the stem of 8 in our display – we only needed the first 8 as the largest observation was 80.

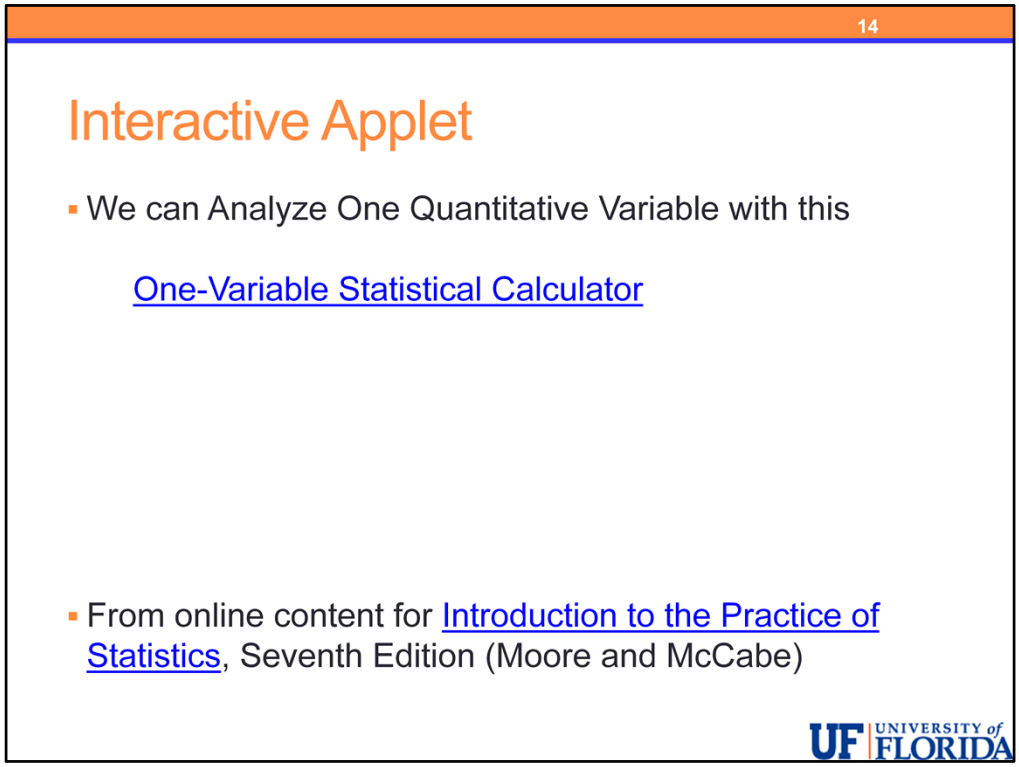
Leaves for each stem can also be split into 5 groups of 2 digits if there is a large amount of data (0-1, 2-3, 4-5, 6-7, 8-9) and some software packages split stems in almost any way they choose!

Stemplots



UF UNIVERSITY of FLORIDA

When we rotate the stemplot 90 degrees counterclockwise, the plot resembles a histogram.



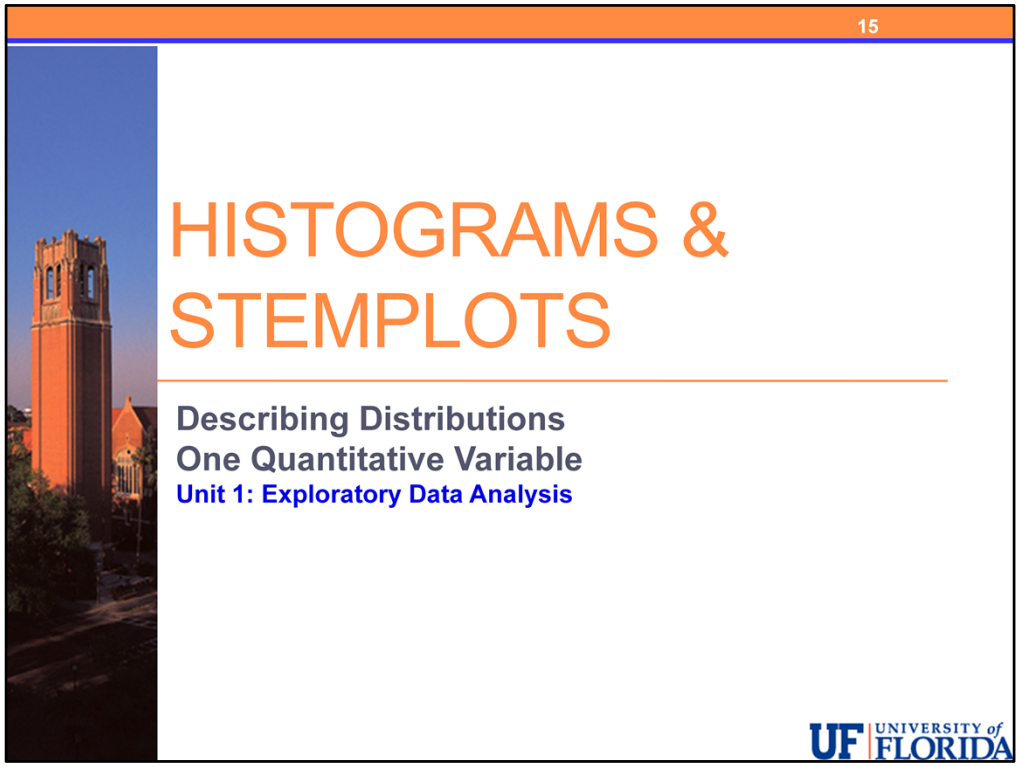
Remember the interactive applet for analyzing one variable, it might help you pull together the ideas in this section on describing the distribution of one quantitative variable.

{web address of One-Variable Statistical Calculator applet:

http://content.bfwpub.com/webroot_pubcontent/Content/BCS_4/IPS7e/Student/Statistical%20Applets/onevar.html

}

{web address of textbook: http://bcs.whfreeman.com/ips7e }



Histograms and to a lesser degree stemplots are helpful displays for visualizing the distribution of one quantitative variable.