Section 1.2 – Displaying Quantitative Data with Graphs (pp. 27-50)

Review Categorical vs. Quantitative variables

Dotplots A simple graph that shows each data value as a dot above its location on a number line (small data sets)

Example: Number of turnovers for 2009 Oakland Raiders during 16 regular-season NFL games:
3, 0, 3, 3, 3, 2, 4, 1, 2, 3, 1, 0, 1, 2, 3, 2

Draw horizontal axis and label variable name; (2) Scale the axis; (3) Make a dot above location for each individual.
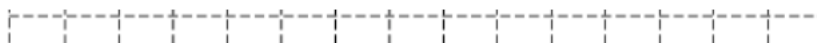
How to Examine the Distribution of a Quantitative Variable (SOCS – AP EXAM!!)

In any graph, look for the **overall pattern** and for striking **departures** from the pattern.

<u>Shape</u>: Concentrate on main features. Look for **clusters** and obvious **gaps**. Look for **potential outliers**. Look for **rough symmetry** or **clear skewness**. Look for number of **modes** or **peaks**. (Discuss **unimodal, bimodal, multimodal** – examples on p. 30)

Unimodal - Describes a graph of quantitative data with a single peak.

Bimodal - Describes a graph of quantitative data with two clear peaks.

Multimodal - Describes a graph of quantitative data with more than two clear peaks.

- A distribution is roughly **symmetric** if the right and left sides of the graph are approximately the same

- A distribution is **skewed to the right** if the right side of the graph is much longer than the left side. It is **skewed to the left** if the left side of the graph is much longer than the right side. (The direction of the long tail gives the direction of skewness.)

<u>Center</u>: Find a value that divides the observations in half. We will use the **mean** and **median** to do this.

<u>Spread</u>: The spread of the distribution tells us how much **variability** there is in the data. Describe smallest and largest values. Also use the range (Max - Min). (Other measures later.)
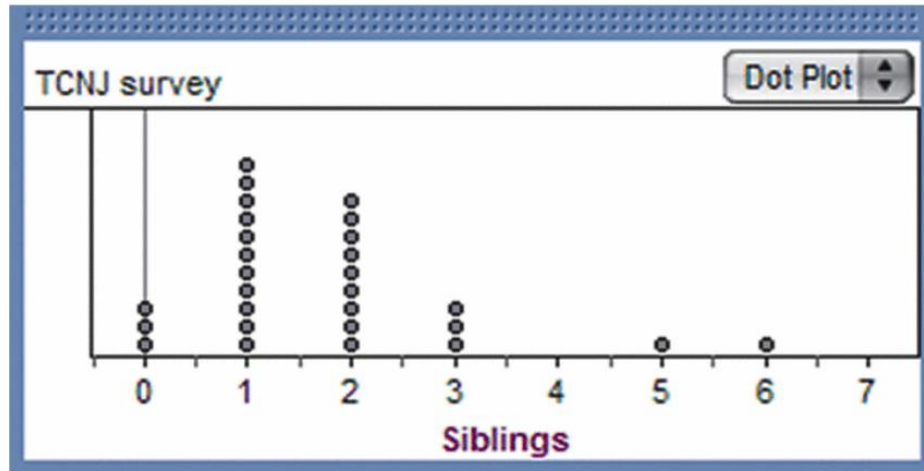
<u>O</u>utliers: What values differ markedly from the bulk of the data? (Rules later)

Teamwork:
1. Describe the distribution of the Raiders' Turnovers.

2. Check Your Understanding pg. 31

The Fathom dotplot displays data on the number of siblings reported by each student in a statistics class.

1. Describe the shape of the distribution

2. Describe the center of the distribution

3. Describe the spread of the distribution

4. Identify any potential outliers

Comparing Distributions (Very important – example on p. 32)

Example – Household Size: U.K. versus South Africa
Comparing distributions

How do the numbers of people living in households in the United Kingdom (U.K.) and South Africa compare? To help answer this question, we used CensusAtSchool's "Random Data Selector" to choose 50 students from each country. The figures below is a dotplot of the household sizes reported by the survey respondents.
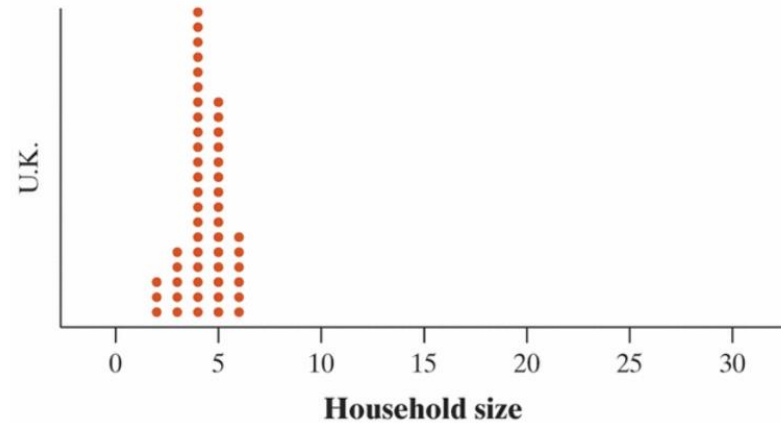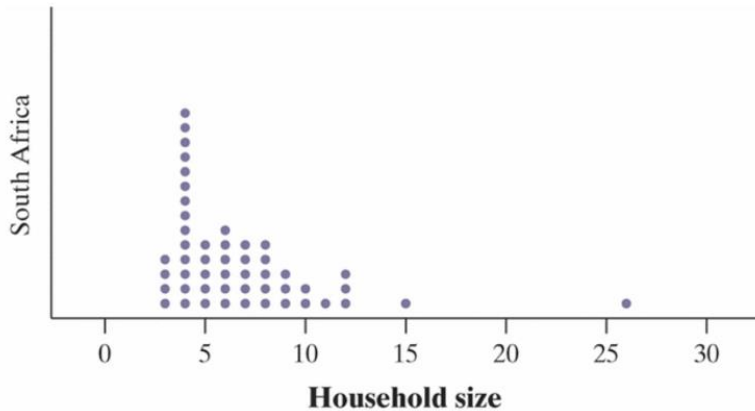
Compare the distributions of household size for these two countries (don't forget your SOCS!).

AP EXAM TIP When comparing distributions of quantitative data, it's not enough just to list values for the center and spread of each distribution. You must explicitly compare these values, using words like "greater than," "less than," or "about the same as." (for example: the center of _____ is greater than the center of _____. A very common mistake on the AP Exam is describing the characteristics of the distributions separately w/o making these explicit comparisons.

Stemplots – Another simple method for displaying fairly small data sets. (p. 33)
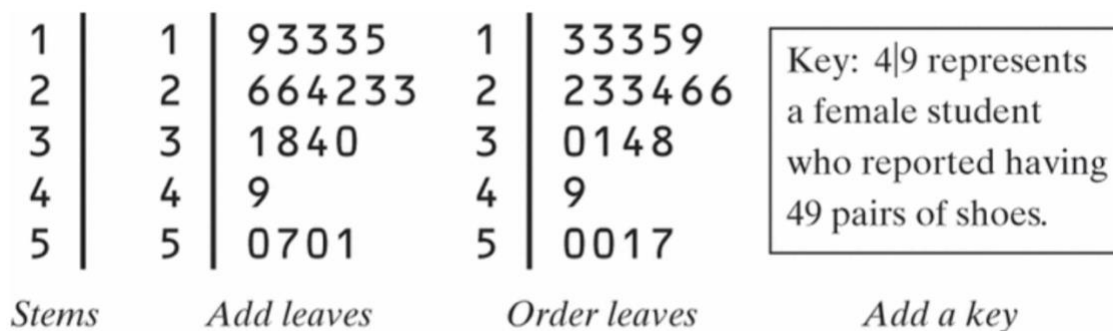
Example – How Many Shoes?
Making a Stemplot

How many pairs of shoes does a typical teenager have? To find out, a group of AP Statistics students conducted a survey. They selected a random sample of 20 female students from their school. Then they recorded the number of pairs of shoes that each respondent reported having. Here are the data:

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

Below are the steps for making a stemplot.
1. Separate each observation into a stem, consisting of all but the final digit, and a leaf, the final digit. Write the stems in a vertical column with the smallest at the top, and draw a vertical line at the right of this column. Do not skip any stems, even if there is no data value for a particular stem. For the data above, the tens digits are the stems, and the ones digits are the leaves. The stems run from 1 to 5.

2. Write each leaf in the row to the right of its stem. For example, the female student with 50 pairs of shoes would have stem 5 and leaf 0, while the student with 31 pairs of shoes would have stem 3 and leaf 1.

3. Arrange the leaves in increasing order out from the stem.

4. Provide a key that explains in context what the stems and leaves represent.

```
1 |        1 | 93335      1 | 33359       Key: 4|9 represents
2 |        2 | 664233     2 | 233466      a female student
3 |        3 | 1840       3 | 0148        who reported having
4 |        4 | 9          4 | 9           49 pairs of shoes.
5 |        5 | 0701       5 | 0017

Stems     Add leaves      Order leaves        Add a key
```
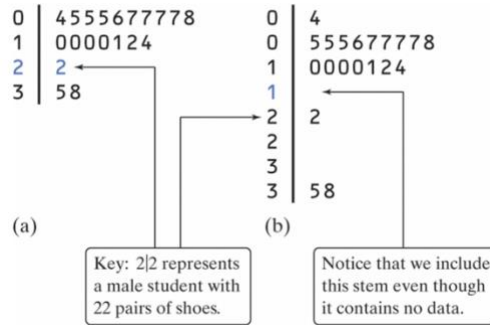
(Common mistake on AP Exam: forget the KEY and the Labels.)

The AP Statistics students in the previous example also collected data from a random sample of 20 male students at their school. Below are the numbers of pairs of shoes reported by each male in the sample:
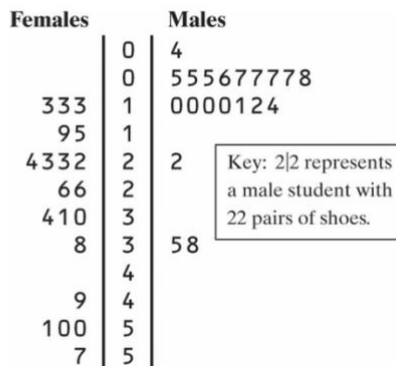
14   7   6   5   12   38   8   7   10   10
10   11   4   5   22   7   5   10   35   7

What would happen if we tried the same approach as before: using the first digits as stems and the last digits as leaves? The completed stemplot is shown in the figure below. What shape does this distribution have? It is difficult to tell with so few stems. We can get a better picture of male shoe ownership by **splitting stems.**

```
0 | 4555677778        0 | 4
1 | 0000124           0 | 555677778
2 | 2 ←               1 | 0000124
3 | 58                1 |
                      2 | 2
                      2 |
                      3 |
                      3 | 58
(a)                   (b)
```

Key: 2|2 represents a male student with 22 pairs of shoes.

Notice that we include this stem even though it contains no data.

Splitting stems - A method for spreading out a stemplot that has too few stems.

What if we want to compare the number of pairs of shoes that males and females have? That calls for a **back-to-back stemplot** with common stems. The leaves on each side are ordered out from the common stem. The figure below is a back-to-back stemplot for the male and female shoe data. Note that we have used the split stems from the above figure as the common stems. The values on the right are the male data. The values on the left are the female data, ordered out from the stem from right to left

```
Females        Males
         0 | 4
         0 | 555677778
    333  1 | 0000124
     95  1 |
   4332  2 | 2      Key: 2|2 represents
     66  2 |        a male student with
    410  3 |        22 pairs of shoes.
      8  3 | 58
         4 |
      9  4 |
    100  5 |
      7  5 |
```
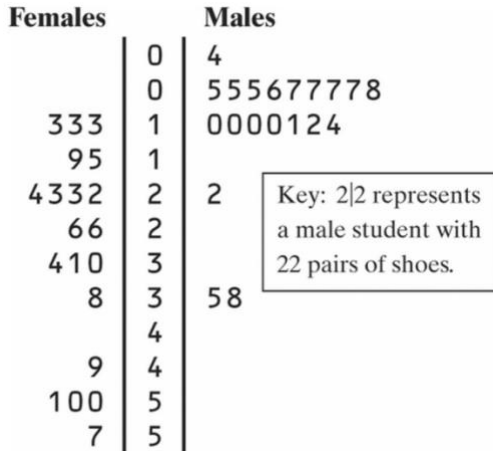
Below are a few tips to consider when making stemplots:

- Stemplots do not work well for large data sets, where each stem must hold a large number of leaves.

- There is no magic number of stems to use, but five is a good minimum. Too few or too many stems will make it difficult to see the distribution's shape.

- If you split stems, be sure that each stem is assigned an equal number of possible leaf digits (two stems, each with five possible leaves; or five stems, each with two possible leaves).

- You can get more flexibility by rounding the data so that the final digit after rounding is suitable as a leaf. Do this when the data have too many digits. For example, in reporting teachers' salaries, using all five digits (for example, $42,549) would be unreasonable. It would be better to round to the nearest thousand and use 4 as a stem and 3 as a leaf.

- Instead of rounding, you can also truncate (remove one or more digits) when data have too many digits. The teacher's salary of $42,549 would truncate to $42,000.
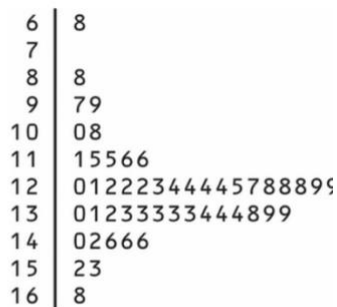
Check Your Understanding pg. 34

1. Use the back-to-back stemplot below to write a few sentences comparing the number of pairs of shoes owned by males and females. Be sure to address shape, center, spread, and outliers.

```
    Females        Males
              0 | 4
              0 | 555677778
        333   1 | 0000124
         95   1 |
       4332   2 | 2     Key: 2|2 represents
         66   2 |       a male student with
        410   3 |       22 pairs of shoes.
          8   3 | 58
              4 |
          9   4 |
        100   5 |
          7   5 |
```

Multiple Choice: Select the best answer for Questions 2 through 4.
Here is a stemplot of the percents of residents aged 65 and older in the 50 states and the District of Columbia. The stems are whole percents and the leaves are tenths of a percent.

```
 6 | 8
 7 |
 8 | 8
 9 | 79
10 | 08
11 | 15566
12 | 0122234444578889S
13 | 01233333444899
14 | 02666
15 | 23
16 | 8
```

Key: 8|8 represents a state
in which 8.8% of residents
are 65 and older.

2. The low outlier is Alaska. What percent of Alaska residents are 65 or older?

(a) 0.68   (b) 6.8   (c) 8.8   (d) 16.8   (e) 68

3. Ignoring the outlier, the shape of the distribution is

(a) skewed to the right   (c) skewed to the left.   (e) skewed to the middle. (b) roughly symmetric   (d) bimodal.

4. The center of the distribution is close to
(a) 13.3%.   (b) 12.8%.   (c) 12.0%.   (d) 11.6%.   (e) 6.8% to 16.8%.