This homework assignment is due on Wednesday 2/15.

In class on 2/8 we discussed the following topics:

1. Three key characteristics describing a distribution of numbers: the center, the spread, and the shape.
2. To measure center we have the mean or median
3. To measure spread with have the IQR or the standard deviation.
4. We discussed how boxplots can show the center, spread, and shape of a data set in a manner that allows many comparisons.
5. The standard deviation is defined by

$$s = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2},$$

   where $\sum$ is short for add up all the terms of the type $(x_i - \bar{x})2$. (It may take some time to get comfortable with this notation.)

6. Why use the standard deviation? We discussed two reasons why. First, if the data is "bell-shaped," then roughly 68% of the data is with 1 standard deviation of the sample mean, 95% is within 2 standard deviations, and 99.8% is within 3 standard deviations of the mean.

   The above is true for a certain shape of data. In general, for any data set atleast $(1-1/k^2)\cdot 100\%$ of the data is in the interval $(\bar{x}-ks, \bar{x}+ks)$. This is known as Chebyshev's lemma.

   Because of this, we define how big a value is within a data set by its z-score:

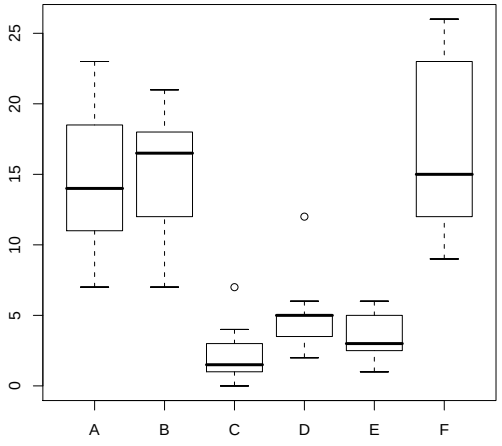
$$z = \frac{x_i - \bar{x}}{s}.$$

7. We discussed using the following words to describe shape: unimodal, bimodal, multimodal; symmetric versus skewed; long-tailed or not.

Some problems:

Exercise 1. From the diagram, answer the following

```
> boxplot(count ~ spray, InsectSprays)
```

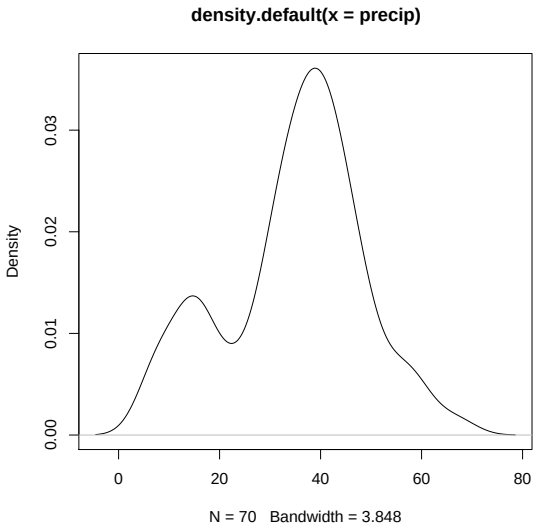


The data is on the effectiveness of 6 insect spays. What is plotted is the distribution of bites for different sprays.

1. Which sprays (if any) have a similar center?
2. Which sprays (if any) have a similar spread?
3. Which sprays (if any) have outliers that are marked?
4. Which spray contains the largest measurement? The smallest?

Exercise 2. For the density plot estimate the median and IQR

```
> plot(density(precip))
```



Exercise 3. Find the standard deviation of 10, 12, 19, and 15 by filling in the following chart, and then using the formula.

x	(x-xbar)	(x-xbar)^2
10	-4	16
12	*	4
19	5	*
15	*	1
56	*	*

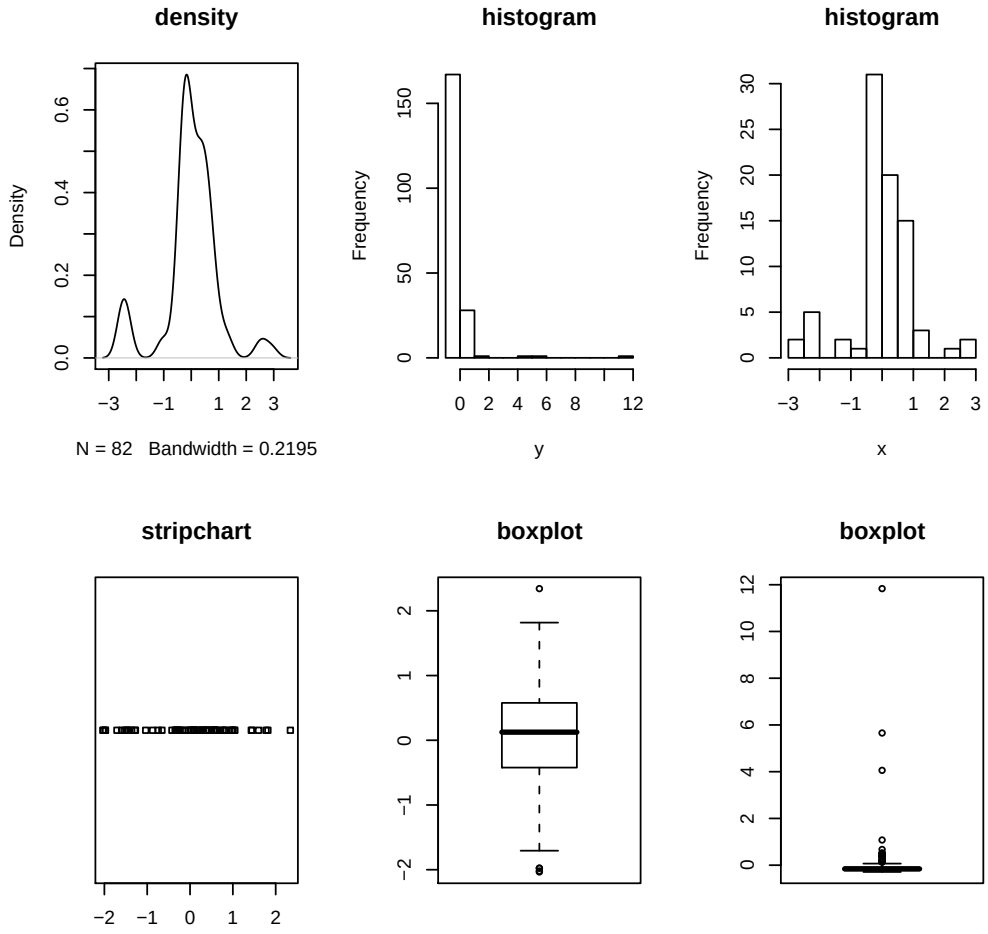
Exercise 4. For the data set

```
4,1,3,9,18
```

1. Compute the standard deviation
2. For each x_i, compute the z-score.
3. What percent of z scores are less than 1.5 in absolute value? Chebyshev's theorem says that atleast 55% are.

Exercise 5. For bell shaped data, the interval $[Q_1, Q_3]$ spans 50% of the data, whereas the interval $[\bar{x}-s, \bar{x}+s]$ spans approximately 68% of the data. If the median and mean are basically the same, which is larger the IQR or s? The IQR or $2s$?

Exercise 6. The six graphs show only 3 different data sets. Can you match up the pairs?



1. Which distribution(s), if any, are symmetric?
2. Which distribution(s), if any, are multimodal?
3. Which distribution(s), if any, are unimodal?