R TUTORIAL, #5:
MEASURES OF POSITION

The (>) symbol indicates something that you will type in.
A bullet (•) indicates what the R program should output (and other comments).

TEST SCORES DATA SET

> Type: scores <- c(80, 76, 100, 83, 100)
> Type: scores
- Check to make sure you entered the values in correctly.

z SCORES

• Find the z scores of the data values.
> Type: scale(scores)
- You should see the *z* scores in column form.
- The "attributes" listed at the end indicate that the mean (87.8 points) is being used as the measure of center and that the **sample** standard deviation (about 11.4 points) is being used as the measure of spread or "scale."

• We can list the data values and their *z* scores side-by-side:
> Type: data.frame(scores, scale(scores))
- What is true of the data values with positive *z* scores?
- What is true of the data values with negative *z* scores?

• We can also get the *z* scores from the formula for *z* scores:
> Type: (scores - mean(scores)) / sd(scores)
- We are using the formula for sample data: $z = \frac{x - \bar{x}}{s}$.

PRESIDENTS' AGES DATA SET

The following is a list of the ages of the U.S. presidents when they became president. (The ages are in years, rounded down.)

57 61 57 57 58 57 61 54 68 51 49 64 50 48 65 52 56 46 54 49 51 47 55 55
54 42 51 56 55 51 54 51 60 62 43 55 56 61 52 69 64 46 54

(President Obama was 47, but he is not included in the table presented in the class notes. The notes were written before 2009.)

ENTERING THE DATA SET

> Type: ages = scan( )
> Press RETURN or ENTER.
- You should see '1:' in your console window.

> Copy and paste the 43 numbers in the PRESIDENTS' AGES DATA SET above.
- Notice that, unlike for the 'c' command, commas (,) are NOT used with 'scan'.
- Copying can be done by using CTRL-C.
- Pasting can be done by using CTRL-V.

PERCENTILES

• We will find the 92^{nd} percentile (or the 0.92 quantile) of the "ages" data set.
> Type: quantile(ages, 0.92)
- Quantiles can be computed in nine different ways by R. Often, quantiles are not actual data values; they may be weighted averages or interpolations, for example.

DECILES

• We will find the nine deciles of "ages."
> Type: quantile(ages, seq(0.1, 0.9, by=0.1))

QUARTILES

• We will find the quartiles of "ages."
> Type: quantile(ages)

• The 'seq' argument is not necessary if we want quartiles; quartiles are assumed.

• The output includes the min (labeled with "0%") and the max (labeled with "100%"); these are not technically quartiles, though.

> Type: median(ages)

• Does this look familiar?

IQR (INTER-QUARTILE RANGE)

• The IQR (inter-quartile range) is the distance between the first and third quartiles. That is, $IQR = Q_3 - Q_1$. It is a measure of spread that excludes outliers.
> Type: IQR(ages)

FIVE-NUMBER SUMMARY

• We will find the five-number summary of "ages."
> Type: fivenum(ages)

• We will find the five-number summary of "ages," as well as the mean.
> Type: summary(ages)

BOXPLOTS

• We will do a boxplot of "ages"; we will visualize the five-number summary.
> Type: boxplot(ages)

• Note: "Hinges" are used instead of Q_1 and Q_3, but they should be close. The "lower hinge" is the median of the lower half of the data set; the "upper hinge" is the median of the upper half.

• The IQR corresponds to the height of the box.

• The "whiskers" extend to up to 1.5 times the IQR above the top of the box and below the bottom of the box. Outliers lie outside those whiskers.

• Let's do a red, horizontal boxplot.
> Type: boxplot(ages, horizontal=T, col=rgb(1,0,0))