IST772 – Week 1 – Breakout 2

Instructions: In your group or with your partner, one person should share their R-Studio window and do the typing, pointing, and clicking. The other person(s) should dictate the code and comments. Share your code and comments for the rest of the class to see in the shared document provided by your instructor.

1. Here is R code to create a small data set: **smallData <- c(71,71,72,74)**
    Without using any functions except sum( ), length( ) and sqrt( ), calculate the mean, the sum of squared deviations from the mean, the variance, and the standard deviation. Use the code window in order to save all of your code and comments. Why do people like to discuss the standard deviation of a variable as opposed to the variance?

2. The mean of a *very large data set* of temperature readings is 72. The median is also 72. What, if anything, can you say about the shape of the distribution? What, if anything, can you say about the mode? Add a comment with your answers.

3. Daily temperature readings collected from Alpha City for one year showed a mean of 72 (F) and a standard deviation of 6 degrees. Annual temperature readings collected from Omega City showed a mean of 72 (F) and a standard deviation of 18 degrees. In which city would you prefer to live and why? Add a comment with your answers.

4. Run this line of code in R-Studio: **myDist <- rbinom(1000,144,0.5)**
    Calculate the mean, median, and standard deviation of myDist (you can use functions this time!). How many data points have values higher than the median? Create a histogram of myDist and note the shape. If you drew a vertical line at the 99th percentile of the histogram, how many data points would have values higher than that line? Add a comment with your answers.

5. In your new role as a traffic analyst, you set up your measuring station alongside a busy road where the speed limit is 30 MPH. Over the course of 24 hours, you collect measurements of the *speed* of each vehicle that passes, the number of new vehicles that pass your location each minute, and, through a cool new algorithm, the approximate *weight* of each vehicle. Use random number generators in R-Studio to create plausible distributions of observations for each of these three variables. Calculate the mean, median, and standard deviation of each distribution and create a histogram for each one, noting the shape in a comment.

6. Paste all of your code and comments into the code share window here: https://codeshare.io/2jg7kR