Plotting and simulation exercises, Assignment 1
EEB 201: R bootcamp
September 18, 2019
Kirk Lohmueller

Imagine that you wish to conduct a study of height of people living in Los Angeles. One strategy is to take a sample of 100 people and compute the average height. Another study design would involve taking a sample of 1000 people and computing the average height. You wish to test the extent to which the different sample sizes affect your estimates of the average height.

To test this, conduct a simple simulation study in R. Assume that the heights of people from LA are normally distributed with a mean of 69 inches and a standard deviation of 10 inches.

In order to conduct the simulation study, you should do the following:

1. Write a function (called "get_heights") in R to draw a sample of individuals (either 100 or 1000) from the population. Hint: You will want to use "rnorm" within your function. Store the random heights that you've generated in a variable called "heights".

2. Within your function, compute the average height from your "heights" vector.

3. Make your function return the average height.

4. Use a "for" loop to call your "get_heights" function 1000 times, with taking a sample of size 100 from the population. Save the mean height from each replicate in a vector called "mean_heights_100".

5. Use a "for" loop to call your "get_heights" function 1000 times, with taking a sample of size 1000 from the population. Save the mean height from each replicate in a vector called "mean_heights_1000".

6. Plot a histogram of the distribution of the average heights for your sample size of 100 and 1000 individuals. The two sets of data should be plotted on the same axes. Add a legend. Label the axes. Plot the data from the 100 samples in red and the data from the 1000 samples in blue. Your plot should look something like the one shown on the next page.