553.283 Introduction to R
Homework 1

Note: If a question asks you for a numerical answer, your submission for that question must consist of the R command that produces that answer followed immediately by the output. The easiest way to do this is to write an R script, clear your console, run each line in the command line, and paste it into a document for submission.

1. Use R as you would a calculator to find numeric answers to the following:
   a. $1 + 2(3 + 4)$
   b. $4^3 + 3^{2+1}$
   c. $\sqrt{(4 + 3)(2 + 1)}$
   d. $\left(\frac{1+2}{3+4}\right)^2$
2. Let our small data set be $\{2, 5, 4, 10, 8\}$. Use the vectorization of functions to perform the following:
   a. Enter this data into a data vector **x**.
   b. Find the square of each number.
   c. Subtract 6 from each number.
   d. Subtract 9 from each number and then square the answers.
3. The asking price of used MINI Coopers varies from seller to seller. An online classifieds listing has these values in thousands of dollars:

   15.9 21.4 19.9 21.9 20.0 16.5 17.9 17.5

   Enter in the data and apply one of R's functions to find answers to the following:
   a. What is the smallest amount? The largest?
   b. Find the average amount.
   c. Find the differences of the largest and smallest amounts from the mean.
4. Your cell-phone bill varies from month to month. The monthly amounts in dollars for the last year were

   46 33 39 37 46 30 48 32 49 35 30 48.

   Enter this data into a variable called *bill*.
   a. What is the total amount you spent on your cell phone last year?
   b. What is the smallest amount you spent in a month? What is the largest?
   c. In how many months was the amount greater than $40? What percentage is this?

5. Define x and y with

$$x = c(1, 3, 5, 7, 9)$$
$$y = c(2, 3, 5, 7, 11, 13).$$

Try to guess the results of these R commands:

$x + 1$
$y * 2$
$length(x)$ and $length(y)$
$sum(x > 5)$ and $sum(x[x > 5])$
$y[3]$
$y[-3]$
$y[x]$
$y[y >= 7]$

6. Load in the dataset *train.csv* from the course webpage. Verify that you have done this by giving the number of observations in the data.