**Lab 2**

(a) Read in Home Depot Data (all csv files)

<https://www.kaggle.com/c/home-depot-product-search-relevance/data>

How many columns and rows in each file?

(b) Show the right down corner element of each file in R (namely, last row, last column).

(c) Output the odd numbers of columns and even numbers of rows of train.csv

(d) Save into R objects and load them, using dput, dget, save, load, save.image

(e) Install the readr Package, and use it to read in the data then.

(f) Any difference in terms of speed in loading the data? Write a simple code to print out the time cost of reading the test.csv data using either read.csv or read\_csv from readr Package.

2.

(a) Create a new vector called “test” containing five numbers of your choice.

(b) Create a second vector called “students” containing five common names of your choice.

(c) Determine the class of “students” and “test”.

(d) Create a data frame containing two columns, “students” and “tests” as defined above.

(e) Convert “test” to character class, and confirm that you were successful.

3.

(a) Select just the Sepal.Length and Species columns from the *iris* data set and save the result to a new data.frame named iris.2

(b) Calculate the mean of the Sepal.length column in iris.2

(c) Calculate the mean of sepal.length, but only for the setosa species

(d) Calculate the number of sepal lengths that are more than one standard deviation below the average sepal length

4. Write R commands for the following questions

1. 1000, 1000, 998, 998, 996, 996, …… , 4, 4, 2, 2
2. Generate a sequence of 10 “a” and 5 “b”
3. Print reverse the order of (b)
4. Find the row numbers in the *iris* data set, where the Petal.Length is larger than 5 and Petal.Width is less than 1.7. And print out this part of the *iris* data set.
5. Guess what the following matrix would look like and the results of the following commands and compare with the real results.

x <- matrix(c(rep(6,3), seq(10,2,-3),c(NA,3,4), 6,1,10), 4, 3)

print(x[,x[2,] > 4])

print(x[,2] < 4)

print(x[x[,2] < 4,])

sum(x[x > 6])

sum(x[x > 6],na.rm=T)

order(x[,3])

x[order(x[,3]),]