**UCS 548 Foundations of Data Science**

**Assignment 4**

Use R to answer the following questions:

1. Vector creation Write R code to generate the following vectors, explore the functions seq() and rep() using the help on commands:
   - 1.3 1.6 1.9 2.2 2.5 2.8 3.1 3.4 3.7 4.0 4.3 4.6 4.9
   - 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
   - 14 12 10 8 6 4 2 0
   - 5 5 12 12 13 13 20 20

2. Loading and exploring data structure Load the iris data that R provides internally by typing data(iris)
   A. What sort of data type is iris?
   B. How many rows (observations) and columns (variables) does the iris dataset have?
   C. Which variable of the data frame iris is a factor and how many levels does it have?
   Select one:
   (a) The variable Species is a factor and it has 5 levels.
   (b) The variable Species is a factor and it has 3 levels.
   (c) The variable 'data.frame' is a factor and it has 150 levels.
   (d) The variable 'data.frame' is a factor and it has 5 levels.

3. Use the "iris" dataset to find
   a) The mean and standard deviation of the sepal width and sepal length for each type of species.
   b) Create a new dataset called iris.class from the iris dataset. Use a loop and ifelse statement to create a vector in the iris.class dataset called Calyx.Width, which is "short" if Sepal.Length is less than 5, and otherwise is "long." (The sepals of a flower are collectively known as the calyx.)

4. Explore dataset- mtcars in R.
   You can get the structure and column names of data by typing the command str(mtcars) and names(mtcars) respectively. Write your code to subset the dataset- mtcars according to the following requirements (NOTE: each requirement is independent.)
   A. Select cars whose cyl (a column in the dataset) value is no smaller than 5.
   B. Show all the fields (columns) of the first 10 cars.
   C. Find all cars matching "Honda".