DUE: October 22nd, 2018 at 11:59 PM

QMM 1001 - Statistics for Data Analytics
Lab 4 /26

Create and submit an R script which, when run, will print the answers to the following questions. Your R script must include a title with your name and student number and comments for each question number and letter.

1. **(10 marks)** The data set "NHL.Data.csv" shows information about hockey players. Read this dataset into R and use it to answer the following questions.
    a. Create a scatterplot of player weight vs. player height. Add appropriate axis titles, a main title, and change the marker symbol to ▲. (5 marks)
    b. Create a subset of the data that only consists of the quantitative variables in the data set (weight, height, and age). (2 marks)
    c. Find the correlations for all pairs of variables using a correlation table. Which two variables have the strongest correlation? Which two variables have the weakest correlation? (3 marks)

2. **(16 marks)** The data set "House.Price.csv" shows house price in dollars and the living space of the house in square feet. Read this data set into R and use it to answer the following questions.
    a. Which variable is the independent/explanatory variable? (1 mark)
    b. Which variable is the dependent/response variable? (1 mark)
    c. Create an appropriately labelled scatterplot. (2 marks)
    d. Perform a regression to predict the house price and print the results. (2 marks)
    e. What is the equation of the line of best fit? (1 mark)
    f. Plot the line of best fit on your scatterplot. Change the colour to any colour other than black. (1 mark)
    g. Using your regression equation, what is the predicted price of a 2000 square foot house? Show the calculation using R. (2 marks)

h. Using your regression equation, what is the predicted square footage of a house that costs $235 000? Show the calculation using R. (2 marks)

i. Create a residual plot for your regression. Describe your residual plot, and list any noteworthy features if there are any. If your residual plot possesses any noteworthy features, what does this imply about your regression? (4 marks)

Save your R Script as: Last Name, First Name LAB 4

Upload your R Script to the "R Assignment – Lab 4" drop box on Moodle before October 22nd at 11:59 PM.