**Student: Seif Kungulio**

**Date: 03/16/2025**

**Subject: Project 1**

**Class: DSCI 512**

**Section: 01W**

**Instructor: Dr. Nengbing Tao**

**File Name: Project1\_Kungulio\_Seif.docx**

1. Read the dataset in Boston.csv into R. Call the loaded data Boston. Make sure that you have the directory set to the correct location for the data.
2. How many rows are in the data frame? How many columns? What do the rows and columns represent?
3. Select the 1st, 100th, and 500th rows with columns tax and medv.
4. Look at the data using cor function. Are any of the predictors associated with per capita crime rate? If so, explain the relationship based on correlation coefficents.
5. Make some pairwise scatterplots of the predictors, crim, rad, tax, indus, and lstat in this data set. Describe your findings.
6. Do any of the suburbs of Boston appear to have particularly high crime rates by looking at the histogram of crim? What is the range of crim by using range() function in R?
7. How many of the suburbs in this dataset bound the Charles River?
8. What is the median pupil-teach ratio among the towns in this dataset? What’s the mean?
9. In this dataset, how many of the suburbs average more than seven rooms per dwelling? More than eight rooms per dwelling? Comment on the suburbs that average more than eight rooms per dwelling.
10. Convert chas to a factor. Boxplot the medv against chas. Are houses around the Charles River more expensive?