# Instructions

1. The dataset is BostonHousing.csv. We have used it in class. You should have a “BostonHousing.R” file if you followed me in class (You may have different files names. You can continue to use your own file name).
2. You may write your codes and answers in the file “BostonHousing.R” from class. If you would like to start a new R file for the assignment, that is fine too.
3. To turn in:

* Submit the “BostonHousing.R” file (or your own file name) through Canvas via “Assignment 3”.
* No late submission: Deadline is Wednesday, 11:59pm, March 3rd

# Questions

The Boston Housing data contains information collected by the US Bureau of the Census concerning housing in Boston.

Please use RStudio to finish the following (refer to DMBA Book Chapter 6 and class lecture for codes). Write your answers to the questions using comment “#” near your codes.

1. Open the R script “BostonHousing.R” from our class practice. Or you can create a new R script.
2. Change the working directory use *Session -> Set Working Directory -> To Source File Location* from the Menu Bar.
3. Rerun the read.csv command to import the dataset.
4. What is the data type of CHAS? Make sure it is a factor variable. (5 points)
5. In our data exploration, we found that CRIM (crime rate) is skewed. Create a new variable lgCRIM by taking log of CRIM. (5 points)
6. Fit a multiple linear regression model to the median house price (MEDV) as a function of all the variables except the original CRIM variable, the coded CAT..MEDV and RAD. (8 points)
7. What is the coefficient of NOX? Standard error? t-value? p-value? Is it significant? How would you interpret the impact of that variable on median housing price? (8 points)
8. What is the coefficient of CHAS? Standard error? t-value? p-value? Is it significant? How would you interpret the impact of bounding Charles River on median housing price? (8 points)
9. Using the estimated regression model to make predictions and get the accuracy measures. What is RMSE? (8 points)
10. Perform a backward selection of variables on the model in Step 6. Which variables are excluded from the model? (8 points)