## Instructions for R Homework 1*

This assignment will explore a sample of homes in Boston suburbs. It comes from the paper `Hedonic housing prices and the demand for clean air` which tries to estimate how much people are willing to pay to live in homes with cleaner air.

To use this dataset, use the function `read.csv` with the url `https://mattbutner.github.io/data/housing_df.csv` as done in the introduction_to_R.pdf document. Remember that you need to include the code that loads the dataset into your R Markdown file.

This data set has the following variables

| VARIABLE | INFO |
|---|---|
| CRIM | Per capita crime rate by town |
| ZN | Proportion of residential land zoned for lots over 25,000 sq. ft. |
| INDUS | Proportion of non-retail business acres per town |
| CHAS | Charles River dummy variable (=1 if census tract touches river; 0 otherwise) |
| NOX | Nitric oxides concentration (parts per 10 million) |
| RM | Average number of rooms per dwelling |
| AGE | Proportion of owner-occupied units built priort to 1940 |
| DIS | weighted distances to five Boston employment centres |
| RAD | index of accessibility to radial highways |
| TAX | Full-value property-tax rate per $10,000 |
| PTRATIO | Pupil-teacher ratio by town |
| B | Formula involving % Black |
| LSTAT | % lower status of the population |
| MEDV | Median value of owner-occupied homes in $1000's |

* For all HW assignments, I need to see all the code used

## Question 1

Look in the `Environment` panel of R studio, how many variables and how many observations are in this data set? Which of the variables are quantitative, which of the variables are categorical?

## Question 2

Each observation is a town in the suburbs or boston. First, we want to get a sense of the distribution of the *median* value of owner-occupied homes of towns in the Boston suburb at this time.

Use the function `mean()` and `sd()` to find the average median value of owner-occupied homes in $1000's. Report the R code and number to two digits (see the option chunk option `echo = TRUE/FALSE` from the rmarkdown website).

## Question 3

To practice with R coding. Use the functions `sqrt()`, `sum()`, `^2`, `length()`, and `mean()`, but not `var()` or `sd()`, calculate the sample standard deviation median value of owner-occupied homes in $1000's. Report the R code and number to two digits.

## Question 4

Use the `hist()` function, create a histogram of NOX pollution. With this distribution is the mean or median a better measure of central tendency?

## Question 5

Now with a sense of the distribution of the two variables of interest. Let's make a scatter plot of `MEDV` on the x-axis and `NOX` on the y-axis. You can use the `plot(x = ____, y = ____)` function for this. Include the plot in the output. To improve the plot, you can use additional parameters `xlab = ""`, and `ylab = ""` to improve the axis labels.

Use the chunk option `fig.cap = ""` to describe the relationship between neighborhood NOX levels and home prices.