In-Class Exercise Set 1: Data Analysis with R

In this exercise set you will work with an updated version of the Boston housing data. It is a cross-section of 506 observations containing information used for hedonic pricing of housing in boston districts.

For a complete description consider:
`https://nowosad.github.io/spData/reference/boston.html`

Sources:

Harrison, David, and Daniel L. Rubinfeld, Hedonic Housing Prices and the Demand for Clean Air, *Journal of Environmental Economics and Management*, Volume 5, (1978), 81-102.

Gilley, O.W., and R. Kelley Pace, On the Harrison and Rubinfeld Data, *Journal of Environmental Economics and Management*, 31 (1996), 403-405.

Exercise 1: Loading and Transforming Data

1. Load the *.txt* data from source:
   `"http://lib.stat.cmu.edu/datasets/boston_corrected.txt"`
   Note that there are some description lines to be skipped. The final data should consist of 506 observations and 21 variables. Use **head()** to compare it to the source file.
2. Check the your data frame for its dimension and potential missing observations.
3. Remove the columns "OBS.", "TOWN", "TOWN.", "TRACT", "LON", "LAT", and "MEDV" from your data frame.
4. Transform all variable names to lower cases using **tolower()**.
5. Change the name of the variable "cmedv" to "medv"
6. Save your new data file as "bostonBI.csv" (comma-separated file).

Exercise 2: Exploring, Manipulating and Visualizing Data

1. Provide summary statistics for the property tax and median value of housing in the boston districts. Are the two variables correlated?
2. Provide a density plot of the median housing value. Change the x-Axis to US$-values.
3. Provide a histogram (binwidth = 5) of the property taxes.
4. Create an ordered factor variable "tax2" with three possible entries

"low"	if property taxes are below 300
"medium"	if property taxes are between 300 and 600
"high"	if property taxes are above 600

   and add it as a new column to your data frame.

5. Provide summary statistics of the median housing value by the tax2 categories "low", "medium", and "high".

6. Provide boxplots of the median housing value by the tax categories "low", "medium", and "high".

7. Provide a scatterplot of the original tax value against the median housing value. Limit the y-axis between 0 and 50 dollars. Add a linear regression line.

8. Briefly comment on your findings.