Data Analytics Programming (COMP 162/COMP 293A) R Project Description

Purpose of the project:

- Apply the data analytics tools covered in class to a dataset of interest to you.
- Practice using R to explore, visualize, and analyze data.
- Gain experience communicating your results in a presentation.
- Learn how to develop your own hypotheses about a dataset and test them.

Due dates:

- Project presentation: Tuesday or Thursday, March 7th or 9th in class
- Project write-up: Wednesday, March 8th at 5 PM

Presentation requirements (approx. 5 minutes per group)

If you have a partner, both of you should speak during the presentation.

- Introduce your dataset, and describe your motivation for analyzing it.
- Description of the variables that you'll analyze in the presentation.
- At least two plots, with an interpretation of each one.
- Explanation of the correlation you performed, and your interpretation of the result.
- Explanation of the linear regression you performed, and your interpretation of the result.
- Explanation of the t-test you performed, and your interpretation of the result.
- Your overarching conclusions from your analysis.

Write-up requirements

- All code required to reproduce your analyses.
- The output of your code.
- *From Homework 2:*
    - A link to your dataset.
    - Description of what each observation of the dataset represents.
    - Description of what each variable in the dataset represents.
    - Description of your motivation for analyzing this dataset.
- *From Homework 3:*
    - The number of rows and columns in your dataset (from the **`str()`** function).
    - The summary of your variables provided by the **`summary()`** function.
- *From Homework 4:*
    - At least one scatterplot showing the relationship between two quantitative variables in your dataset, colored by a categorical variable in the dataset.
    - At least one boxplot showing the relationship between a categorical and quantitative variable in your dataset.
    - **Challenge problem (required for 293A, bonus for 162):** A third plot that shows another relationship in your data, and your interpretation of it.
- *Additional analyses:*
    - At least one correlation between two quantitative variables in your dataset, and your interpretation of the result.

- At least one linear regression between two quantitative variables in your dataset, and your interpretation of the result.
- At least one t-test between a categorical and quantitative variable in your dataset, and your interpretation of the result.
- **Challenge problem (required for 293A, bonus for 162):** A third statistical test (correlation, linear regression, or t-test) and your interpretation of the result.