STAT 350
Fall 2020
Final Project

Due Date: December 8th, 2020 @23:59

Motivation:
Upon successful completion of this project, students will possess a working knowledge of Github and R Markdown, and mastery in the practice of regression analysis. These skills are highly valued globally by employers in search of data scientists.

Project Description:

Projects can be done in groups or alone – your choice. If you wish to be in a group, but do not know your classmates, I can construct groups for you. Groups can be no larger than four people (do not ask for more, the answer is no). Please make sure to be organized!

If you wish to form your own group, you MUST email me by 11:59pm on October 30. If you wish to be placed into a group, you must inform me by 11:59pm on October 30. Otherwise, you are working solo and will not be placed into a group may not hand in a group project.

Each group will be assigned a dataset. Collectively, group members are to perform a complete regression analysis of their data, details of which must be presented on **Github** (https://github.com) using**Markdown** (https://rmarkdown.rstudio.com/articles_intro.html).

The following sections must be included:
Abstract (150 words or less)
Introduction (must contain a thorough description of the questions of interest)
Data Description (must contain data visualizations that are properly labelled and explained)
Methods (must contain a complete description of all analysis tools used)
Results (all figures should be properly labelled and discussed)
Conclusion (must contain a concise discussion of what has been learned about the application from the analysis)
Appendix (must include all data and R Markdown files for reproducibility)

Data:
Datasets are found at https://lionbridge.ai/datasets/10-open-datasets-for-linear-regression/. I will assign your group a dataset. So, do NOT start until the assignment is made

Note: BEFORE commencing your analysis, **you must introduce one new additional data point** into your assigned dataset. A description of this unique data point must be included in your *Data Description* section along with some rationale for the values chosen.

Grading Scheme:

8 Overall presentation and organization of materials (includes quality of writing and quality of the Github presentation, …)

5 Quality of data visualizations

8 Correctness of analysis

6 Quality and selection of relevant figures

8 Interpretation of results

--

35