

BST260 2022 Final Project

In this project, you will identify a question you are interested in, and a dataset you can use to answer or explore that question. The analysis you conduct should be fully reproducible using Rmarkdown and GitHub. **Your final report will be a 2-4 page long knitted .pdf document, not including figures/tables and code.** Two possible routes are:

Route 1. Selecting a dataset that requires extensive scraping or wrangling, and using at least one major analysis.

Route 2. Selecting a dataset that is essentially ready-to-go (i.e. no extensive web scraping or wrangling necessary), and using at least two major analyses.

Major analyses include the following topics we covered in class and in the book chapters denoted in ():

- Association tests (15.10)
- Bayesian statistics (16.4)
- Hierarchical models (16.6)
- Regression (17)
- Linear models (18)
- Smoothing (28)
- Cross-validation (29)
- Machine learning (any topic covered in 31)
- Other (get teaching staff approval)

Format and grading criteria:

- **Introduction (.5-1 page of text, 25 pts):**
 - Introduce (in words) the dataset and question (5 pts)
 - 1-2 key plots or tables illustrating your exploratory data analysis (10 pts)
 - Description and justification of the methodology you will use in your analysis (10 pts)
- **Results (1 - 2.5 pages of text, 50 pts):**
 - The text in the results should guide the reader through your analysis and describe what each plot or table is showing, and how it relates to the central question you are trying to ask. Feel free to reiterate key concepts of the methodology and how they help analyze the question or topic.
 - Route 1
 - Explanation of why your extensive web scraping or wrangling was necessary (5 pts)
 - Two tables showing relevant parts of the dataset 1) before and 2) after (10 pts)
 - 3+ key plots or tables illustrating your major analysis (35 pts)
 - Route 2

- 6+ key plots or tables illustrating your two major analyses (50 pts)
- **Conclusion (.5 pages of text, 15 pts):**
 - Summary of your question, methods and results
 - Additional topics can include:
 - Was your analysis successful? Why or why not?
 - What would you do if you had more time?
- **References (does not count towards page limit; at the minimum, reference the source of data)**
- **Appendix (does not count towards page limit, 10 pts)**
 - All figures/tables and code
 - See <https://bookdown.org/yihui/rmarkdown-cookbook/code-appendix.html>