1 Overview

For your final project, you will investigate some data, run analysis and disseminate your findings through a report.

The final project is designed (feasible) to be completed alone, but may be completed in a groups.

Deadlines:

| Proposal | Tuesday | March | 18 |
| Report | Friday | April | 04 |

2 Requirements

Your final project requires choosing a data set that allows you to do at least 3 of the following

- Perform a residual analysis

- Apply multiple regression

- Use indicator/dummy variables

- Perform variable selection

- Perform inference (resampling or classical) on a coefficient or some other aspect of the model (non-trivially)

- Apply a non-parametric regression with at least two predictors simultaneously (e.g. kernel, polynomial, knn)

- Apply a GLM (e.g., implement your own)

- Apply a random effect model

- Apply transformations with investigation

- Other skills/exercises relevant to the course with approval (suggest in the proposal)

- Justification that one of the above exercises will be conducted thoroughly enough that it counts as 2 (suggest in proposal)

The 3 methods that you choose must be tied together cohesively to convey a single message or bring the audience to understanding the answer of some investigative question. They cannot be a bunch of separate analyses. That means the order you present your results these will likely be important.

You may choose a data set that has been used in another class provided that none of the analysis you perform on that data is repeated. Similarly, you can't perform/repeat an analysis on the data found elsewhere.

You may perform a simulation for your project, but you will then also be graded on designing a simulation that is appropriate and answers the investigative question.

You may *not* choose exercises that are the focus of another course (e.g., you can't use time series or machine learning topics).

2.1 Proposal

You must submit a proposal (half a page to a page, even point-form is fine) identifying

- Team members

- The data set that is planned for analysis and a brief description of the nature of the data including the variables.

- The three exercises that you'll perform.

- Attestation that no team member has analyzed the data in the 3 ways for credit before. (This means you can reuse a data set from another course if you're analyzing it completely differently).

Students will receive feedback from the proposal and may resubmit if they wish. However, there are no extensions to future deadlines if the proposal is resubmitted.

Major deviations from the proposal require submitting a new proposal e.g., change in your three exercises, change in data set, change in team members.

2.2 Report

The final project will be a report that is a maximum of 10 pages of text, double spaced, using a standard 12 point font and kerning, and at least 1 inch margins. Unlike assignments, Figures, tables, and references come *after* your main document and do not count toward your 10 page limit.

Your document also should end with an appendix containing your code and no code/output should be in your main document. Code in the appendix should be structured and readable, to be easily cross-referenced to your document. Though not required, it is recommended that you include references to your code from your document (e.g., "Implementation of the bootstrapped regression is found in Code Block B of the appendix with additional helper functions from other blocks.")

Other supererogatory material may be placed in the appendix for additional reference. But because it is an appendix, assume that I will not read it.

The entire report must be submitted as a single PDF.

2.2.1 References

Your data must be properly referenced, and at the very least, the topic of extension will require you to cite other sources. Use any standard citation style where a separate bibliography is used (avoid footnote citations). Do not submit your data.

2.2.2 Layout

You may present your report as you see fit but it should be sensible. For example, you'll require sections to introduce your data, your two methods, your analysis, and maybe a discussion.

If applicable, you should explain how you have manipulated the data prior to analysis, what the data's original or common use is, etc.. Similarly, for each analysis, you should (briefly) explain the nature of the approach, its goal, and why and how you plan to use it.

In a report from graduate students taking a non-introductory university course, I expect appropriate introductions, conclusions, quality of writing including flow of text, and so on. This will be considered part of style/presentation when marked.

3   Marking

You will be marked in part with the following considerations:

Analyses	Correctness of method description, analysis, justification, appropriate level of detail in explanation of methods and data analysis
Style/Presentation	Presentation of analysis, flow of text, etc.. Relevant (appropriate) figures and tables are used.
Usefulness	Analyses provided (contextually) relevant, insightful, or interesting results. Note that novel is purposely *not* a criterion.

The marking scheme will likely be a geometric average of these criteria and other criteria will be included if appropriate.