Project

The purpose of the data project is for you to conduct a reproducible analysis with a data set of your choosing. There are two components to the project, the proposal, which will be graded on a pass/fail basis, and the final report. The outline for each of these are provided in the templates. When submitting the assignments, include the R Markdown file (change the name to include your last name, for example `Bryer-Proposal.Rmd` and `Bryer-Project.Rmd`) along with any supplementary files necessary to run the R Markdown file (e.g. data files, screenshots, etc.). Suggestions for possible data sources are included below, however you are free to use data not listed below. The only requirement is that you are allowed to share the data. Projects will be shared with others on this website so should be presented in a way that other students can reproduce your analysis.

Project Proposal

The proposal can be more informal using bullet points where necessary and include R code and output. You must address the following areas:

- Research question

- What are the cases, and how many are there?

- Describe the method of data collection.

- What type of study is this (observational/experiment)?

- Data Source: If you collected the data, state self-collected. If not, provide a citation/link.

- Response: What is the response variable, and what type is it (numerical/categorical)?

- Explanatory: What is the explanatory variable(s), and what type is it (numerical/categorival)?

- Relevant summary statistics

- Download project proposal template

- Download project template

Example data project proposal (Source Rmarkdown file)

Final Project

- You are required to attend ONLY ONE of those time slots. You will do your presentation, watch the other presentations, and provide peer feedback (will be shared anonymously afterward).

Checklist / Suggested Outline

- Abstract (no more than 300 words)
- Overview slide
  - Context on the data collection
  - Description of the dependent variable (what is being measured)
  - Description of the independent variable (what is being measured; include at least 2 variables)
  - Research question
- Summary statistics
- Include appropriate data visualizations.
- Statistical output

- Include the appropriate statistics for your method used.
- For null hypothesis tests (e.g. t-test, chi-squared, ANOVA, etc.), state the null and alternative hypotheses along with relevant statistic and p-value (and confidence interval if appropriate).
- For regression models, include the regression output and interpret the R-squared value.
  - Conclusion
    - Why is this analysis important?
    - Limitations of the analysis?

Rubric

Domain	Accomplished	Proficient	Needs Improvement
Abstract	Abstract is less than 300 words, free of grammatical errors, summarizes the analysis conducted, has a conclusion and implicaitons	NA	NA
Introduction	The research question is clearly stated, can be answered by the data, and the context of the problem clearly explained.	The research question is unclear and/or not supported by the data.	Research question is ambiguous, unclear, or not stated.
Data Display	Includes appropriate, well-labeled, accurate displays (graphs and tables) of the data.	Includes appropriate, accurate displays of the data.	Includes appropriate but no accurate displays of the data.
Data Analysis	The appropriate statistical test(s) was used for the data and interpretation was clear.	The appropriate statistical test(s) was used but interpretation was not fully clear or well articulated.	The incorrect statistical test was used an/or not justified for the data as presented.
Conclusion	Conclusion includes a clear answer to the statistical question that is consistent with the data analysis and the method of data collection.	Conclusion includes an answer to the statistical question that is consistent with the data but not with the data collection method.	Conclusion does not include an answer to the statistical question that is consistent with the data analysis.
Overall Presentation	Attractive, well-organized, well-written presentation	Presentation has two of the three qualities: attractive, well-organized, well-written.	Presentation is not attractive, organized, or written. There are numerous errors throughout.

Example Data Sources

You are not to use data sources used in class or the textbooks. Possible data sources include, but are not limited to: