

DATA101 Final Project Presentation and Paper

Prof. Joshua Davidson

Due Dates

Presentation: December 3, 5, and 8 in class. Dates will be assigned to groups randomly. Please let me know if you have an outstanding commitment that makes one date not possible. Please post slides to Blackboard in the Final Project Presentation assignment page in the Week 13/14 folder by 8:00 AM the day you are presenting so I can download them to my computer.

Paper: 9:00 AM section: December 16 by 4:00 PM on Blackboard. 10:00 AM section: December 18 by 4:00 PM on Blackboard. Please submit two PDF files for the final paper: 1 file with all the code included, and 1 file with all the code suppressed. Please post both PDFs to the Final Project Paper assignment page in the Week 13/14 folder.

Final Project Structure

- **Presentation:** 5-6 minutes, followed by 2-3 minutes of Q&A from the class. The presentation should feature the sections defined below. Please evenly divide the giving of the presentation among your group members.
- **Paper:** Approximately 1000 words (not counting figures/tables or citations), written as a clear, analytic narrative, featuring the sections defined below. **Unlike our labs, please suppress all code in one version of your final paper.** This pdf file should read as a manuscript, with all tables, plots, etc. embedded/included in the text, in the order they should appear given the text, but the code that generated the tables and figures should be hidden. In the second pdf file please include all your code. I will be primarily grading the file without the code, but will use the code file for reference.

Final Project Sections

Section 1 - Introduction:

- The introduction should a) introduce your overall research question, b) your motivation for asking this research question(ie, why is this question important to you), and c) your data (where it came from and what it represents).
- **Presentation:** This section should be 1-3 slides and approximately 1 minute of your presentation.
- **Paper:** This section should be approximately 100-150 words and feature at most 1 table/visualization.

Section 2 - Data:

- Describe explicitly how the data was collected, what is the unit of analysis (ie, what does each row represent?), what are the variables (ie, what does each column represent), and what are some issues with the data.
- Provide the dimensions of the data and what each variable represents. If you are only using a subset of variables from a larger dataset please just define the variables you'll be using in the project.
- **Presentation:** This section should be 2-4 slides and approximately 2 minutes of your presentation.
- **Paper:** This section should be approximately 100-150 words and feature at most 1 table/visualization.

Section 3 - Analytic Framework:

- Define the outcome (response, Y) and predictor (explanatory, X) variables you are using to investigate your research question. What are the characteristics of these variables (categorical/binary/continuous), and how do these variables relate to one another? Are you primarily taking a descriptive approach to these data or are you building a model, and why? If you are building a model(s), justify the model specification given your data and variables.
- **Presentation:** This section should be 2-4 slides and approximately 2 minutes of your presentation.
- **Paper:** This section should be approximately 200-250 words and feature at most 1 table/visualization.

Section 4 - Results:

- Discuss the results of your exploratory data analysis and models (if you are building them) in the context of the data and your research questions. Please be sure to provide basic univariate descriptions of your key variables, as well as bivariate relationships between your dependent and independent variables. What trends do you see in the data (plots, tabular output, regression coefficients, etc.)? How strong are these trends? Are there any outliers that go against the predominant trends in the data? What might these outliers represent? If you are building a model, how closely does your model specification match the key assumptions of the regression framework (shape and distribution of residuals)?
- **Presentation:** This section should be 4-6 slides and approximately 2 minutes of your presentation.
- **Paper:** This section should be approximately 300-500 words and feature at most 10 tables/visualizations (pick the “best” ones that add the most to your analysis).

Section 5 - Implications

- Summarize your project and discuss the implications of your findings for the topic area you are studying. Why is what you found important? What are some areas in your study that remain unresolved? How could you (or someone else) improve on your study going forward?
- **Presentation:** This section should be 1-3 slides and approximately 1 minute of your presentation.
- **Paper:** This section should be approximately 100-150 words and feature at most 1 table/visualization.

Citations

Please cite all data sources, articles referenced, and/or images that are copyrighted (not in the public domain). Failure to cite all these sources is an honor code violation.

Rubric

The final project presentation is worth 18 points (or 40% of your final project grade).

The final project paper is worth 18 points (or 40% of your final project grade).

16-18 points: The presentation/paper will have a clear research question, a meaningful motivation that explains why this question is important to you personally, explicit description of the data and its characteristics, defined response/explanatory variables, explicit and justified analytic framework, clear and focused discussion of results including both univariate and bivariate trends, thoughtfully chosen visualizations that add meaning to the analysis, and clear connections between the results and the topic area studied. The presentation will be clearly rehearsed in advance and keep to the 10 minute time limit. The paper will read as linear narrative, with plots embedded in the text, and edited for grammar and clarity.

14-16 points: The presentation/paper will have a research question but it may be vague or hard to understand, the motivation generated only some interest and featured limited personal connections, data and its characteristics were described thought not in full detail, response/explanatory variables were suggested but not clearly defined, the analytic framework lacked some major rationales, the discussion of results missed major points and/or did not discuss either univariate or bivariate relationships in enough detail, chosen visualizations did not add enough meaning to the analysis (or too many or too few visualizations were present), and cursory connections between the results and the topic area were made. The presentation will

not be well rehearsed leading to limited engagement and/or will go over the 10 minute time limit. The paper will not read clearly as narrative, code may be present, plots may not relate to the text, and the paper will feature grammatical issues.

0-12 points: The presentation/paper will not be a research question or it will be nearly impossible to understand, the motivation will be lacking, data and its characteristics are not described in any kind of detail, response/explanatory variables are not defined, the analytic framework lacks rationale (or is entirely missing), discussion of results lacks univariate or bivariate relationships and does not address the findings in a meaningful way, no visualizations are present or they are wholly unclear, and no connections between the results and the topic area were made. The presentation will not be rehearsed leading to very little engagement and/or will go well over the 10 minute time limit. The paper will not be a narrative document, code is present, plots do not relate to the text, and major grammatical issues are present throughout.