36-613 Rubric for Final Projects

IDMRAD report, due Friday, Oct 13th at 5 PM EDT (155 points)

As your final project, you will write an **IDMRAD** (Introduction-Data-Methods-Results-Discussion) report to aid in the development of your project for 36-611. For this project, you will be using multiple sources of data:

- LinkedIn Job Postings Dataset,
- Daily stock prices (see example script available on Canvas),
- additional data that you find relevant in your research for 36-611 project.

Your IDMRAD report should follow the template covered in 36-617 and contain the following sections in this order of appearance:

- *Abstract*: Summarize I, D, M, R and D of paper
- *(I)ntroduction*: Why would anyone want to read this paper? What questions will be addressed?
- *(D)ata*: What dataset was used for this study? Typically: Variable definitions, sample size, quick summaries and initial descriptive EDA. **Your report for this class will contain more extensive EDA in this section than expected for your reports in 36-617**.
- *(M)ethods*: What did you do, to address these questions? This should explain any kind of statistical analyses used to complement your graphs and why they are relevant in your report.
- *(R)esults*: What did you find?
- (A)nd *(D)iscussion*: What does it all mean? Typically: answer questions, discuss generalizations, and limitations.

And then include appropriate references in the end.

You'll submit reports as a group assignment on Gradescope, which asks you to list all the group members when you submit; this will make it easy for us to update the gradebook. **You do NOT have to use RMarkdown to generate your report**, feel free to use whatever you are most comfortable with for document formatting (but you MUST submit the report as a PDF regardless if it's Word, RMarkdown, or LaTeX generated).

Research questions of interest

Your report should tell a somewhat cohesive story. To help structure your report, **you are required to come up with at least three interesting, overarching questions for your dataset that are relevant for your 36-611 project.** You'll use EDA with a variety of data visualizations to walk the reader through a comprehensive analysis of those questions, as well as demonstrate your findings and conclusions. You should state these research questions clearly in the *Introduction* of your report.

Graph requirements

In order to demonstrate your ability to create high-quality graphics, you will have requirements on the graphs you must use:

1. You are required to make **at least six** graphs with additional requirements:

- **at least three** graphs from Lectures 2 through 5
- **at least three** graphs from Lectures 6 through the end of the semester.

In other words, you can NOT just turn in 8 classical EDA graphs. You are required to make at least one graph that we will be learning about in the second half of the class.

2. Each group can have no more than two graphs that show a single variable (e.g., one-variable bar charts, histograms, density estimates).
3. You can **NOT** have more than 3 of the same type of graph. For example, don't just use 5 smoothed density plots. (Note: a facetted graph counts as one graph, and NOT as the number of facets.)

Interactive graphic requirement

Additionally, you are required to make **an interactive Shiny app or Tableau Public dashboard** that enables exploration of your dataset in a manner that is relevant to your research questions. This app/dashboard must include a data visualization that is reactive to user input, and is NOT simply an interactive presentation of a table. You must include a public link to your app/dashboard in the *Data* section of your report. The visualizations in your app/dashboard do NOT count towards the counts in the graph requirements above. **You may include static versions of visualization(s) in your app/dashboard to count towards your required graphs.**

Grading rubric

Abstract (10 pts)

- **Basic item (5pts)**: Does the report include an *Abstract*?
- **Detailed item (rated 0 to 5pts)**: What is the quality of this *Abstract*? Does it clearly and concisely the I, D, M, R, and D sections of your report?

Introduction (25 pts)

- **Basic item (5pts)**: Does the report include an *Introduction*?
- **Basic item (15pts)**: Does the *Introduction* section include at least three clearly stated research questions? (minus 5pts for each question that is missing or not clearly stated)
- **Detailed item (rated 0 to 5)**: What are the quality of the research questions? Are they well-motivated by your RFI request with your assigned dataset? Or are they shallow? For example, a shallow question would be in the form, *What does the distribution of this variable look like?* A more interesting question would motivate why we would like to inspect particular distributions with relevant context for your RFI.

Data (50 pts)

- **Basic item (5pts)**: Does the report include a *Data* section?
- **Basic item (5pts)**: Does the *Data* section include a description of the dataset?
- **Detailed item (rated 0 to 5)**: What is the quality of this description? Does it clearly communicate what the rows and columns (i.e., observations and variables) are in the dataset and include basic summaries such as the number of observations?
- **Basic item (15 pts)**: Does the *Data* section have 6 to 10 graphs that follow the graphs requirements above? (each criterion is worth 5pts)
- **Basic item (5 pts)**: What is the quality of these graphs? Are these graphs easily readable, interpretable, and properly labeled?
- **Detailed item (0 to 5)**: Are the graphs well-motivated, given the research questions of the report? In other words, do the graphs address/answer the research questions, or do they only provide tangential (or even irrelevant) details?
- **Basic item (5pts)**: Does the *Data* section include a link to an interactive Shiny app or Tableau Public dashboard?
- **Detailed item (0 to 5)**: What is the quality of the app/dashboard? Is the interactivity well-motivated and relevant to the research questions of interest?

Methods (15 pts)

- **Basic item (5pts)**: Does the report include a *Methods* section?
- **Basic item (5 pts)**: Does the report include any kind of statistical analyses to complement their graphs?
- **Detailed item (0 to 5)**: Are the statistical analyses appropriate given the type of data and research questions of interest?

Results (25 pts)

- **Basic item (5pts)**: Does the report include a *Results* section?
- **Basic item (10 pts)**: Does the report provide further descriptions/interpretations for each graph? (minus 2.5pts for each description that is missing, up to 10pts)
- **Detailed item (0 to 5)**: What is the quality of these descriptions? After reading these descriptions, is it crystal clear what is being displayed in the graph and what the main takeaways are?
- **Detailed item (0 to 5)**: Are the statistical analyses interpreted correctly?

Discussion (30 pts)

- **Basic item (5pts)**: Does the report include a *Discussion* section?
- **Basic item (10 pts)**: Does the report provide some clear conclusions that can be made from their graphs and analyses?
- **Detailed item (0 to 5)**: What is the accuracy of the claims being made? Are the claims well-supported by the graphs and analyses presented? Are the conclusions well-aligned with the research questions of the report?

- **Basic item (5 pts)**: Does the report discuss questions that have not been answered by the report, but could be answered with future work?
- **Detailed item (0 to 5)**: Did the team provide adequate reasons as to why these questions were left as future work (e.g., they need more data, need more nuanced statistical techniques they haven't learned, etc.)? Are these future-work questions well-motivated given what the team has completed for this project?