

Mid-term Expectations

You may work with one or two other persons on the midterm and final projects if you wish. Once you decide on working solo or as a group, your decision remains for the rest of the course (i.e., you can't decide to work alone or join someone after submitting the midterm.

Please read the final project page before reading any further.

Throughout the term you will progressively create your final project. Your mid-term project is to submit the work you have completed midway through the course for a progress evaluation, where you have fully completed standards 1.1-4.4 and 7.1-7.4 as shown below. This progress check will allow your peers and me to provide you direction for final completion. This mid-term report will be rendered as an R Markdown HTML or PDF product.

Mid-term expectations, which are based on the final project standards, are listed below:

| Section | Standard | Possible Points |
|-------------------|--|-----------------|
| Introduction | 1.1 Provide an introduction that explains the problem statement you are addressing. Why should I be interested in this? 1.2 Provide a short explanation of how you plan to address this problem statement (the data used and the methodology employed) 1.3 Discuss your current proposed approach/analytic technique you think will address (fully or partially) this problem. 1.4 Explain how your analysis will help the consumer of your analysis. | 5 |
| Packages Required | 2.1 All packages used are loaded upfront so the reader knows which are required to replicate the analysis. 2.2 Messages and warnings resulting from loading the package are suppressed. 2.3 Explanation is provided regarding the purpose of each package (there are over 10,000 packages, don't assume that I know why you loaded each package). | 5 |
| Data Preparation | 3.1 Original source where the data was obtained is cited and, if possible, hyperlinked. 3.2 Source data is thoroughly explained (i.e. what was the original purpose of the data, when was it collected, how many variables did the original have, explain any peculiarities of the source data such as how missing values are recorded, or how data was imputed, etc.). 3.3 Data importing and cleaning steps are explained in the text (tell me why you are doing the data cleaning activities that you perform) and follow a logical process. | 10 |

| Section | Standard | Possible Points |
|------------------------------------|---|-----------------|
| | <p>3.4 Once your data is clean, show what the final data set looks like. However, do not print off a data frame with 200+ rows; show me the data in the most condensed form possible.</p> <p>3.5 Provide summary information about the variables of concern in your cleaned data set. Do not just print off a bunch of code chunks with <code>str()</code>, <code>summary()</code>, etc. Rather, provide me with a consolidated explanation, either with a table that provides summary info for each variable or a nicely written summary paragraph with inline code.</p> | |
| Proposed Exploratory Data Analysis | <p>4.1 Discuss how you plan to uncover new information in the data that is not self-evident. What are different ways you could look at this data to answer the questions you want to answer? Do you plan to slice and dice the data in different ways, create new variables, or join separate data frames to create new summary information? How could you summarize your data to answer key questions?</p> <p>4.2 What types of plots and tables will help you to illustrate the findings to your questions?</p> <p>4.3 What do you <u>not</u> know how to do right now that you need to learn to answer your questions?</p> <p>4.4 Do you plan on incorporating any machine learning techniques (i.e. linear regression, discriminant analysis, cluster analysis) to answer your questions?</p> | 5 |
| Formatting & Other Requirements | <p>7.1 All code is visible, proper coding style is followed, and code is well commented (see section regarding style).</p> <p>7.2 Coding is systematic - complicated problem broken down into sub-problems that are individually much simpler. Code is efficient, correct, and minimal. Code uses appropriate data structure (list, data frame, vector/matrix/array). Code checks for common errors.</p> <p>7.3 Achievement, mastery, cleverness, creativity: Tools and techniques from the course are applied very competently and, perhaps, somewhat creatively. Perhaps student has gone beyond what was expected and required, e.g., extraordinary effort, additional tools not addressed by this course, unusually sophisticated application of tools from course.</p> <p>7.4 .Rmd fully executes without any errors and HTML produced matches the HTML report submitted by student.</p> | 15 |