Preparing Your Mid-term Project Eval

Start applying what you have learned to the data set that you've identified for your final project. Your mid-term project eval is due by the end of next week. This mid-term report will be rendered as an R Markdown HTML product and you need to show **all** your code. You will need to upload your HTML report to RPubs and then send me the url via Canvas. The following is what I expect in this report. It would be wise to review some examples of previous students' final reports to get a taste of what I expect.[1]

[1] The grading rubric for what I expect is outlined at http://zzz1990771.github. io/data_wrangling/mid-term and you can view some example reports at the bottom of the following webpage: http: // zzz1990771.github.io/data_wrangling/final-project

Section 1: 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.

Section 2: Packages Reauired

- 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).

Section 3: 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.
- 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.
- 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 str(), summary(), 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.

Section 4: Proposed Exploratory Data Analysis

- 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?
- 4.2 What types of plots and tables will help you to illustrate the findings to your questions?
- 4.3 What do you not know how to do right now that you need to learn to answer your questions?
- 4.4 Do you plan on incorporating any machine learning techniques (i.e. linear regression, discriminant analysis, cluster analysis) to answer your questions?

Formatting & Other Requirements

- All code is visible, proper coding style is followed, and code is well commented (see section regarding syle).
- 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.
- 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.

- .Rmd fully executes without any errors and HTML produced matches the HTML report submitted by student.