Data Analysis with R Project Rubric [Nanodegree]

Data Analysis with R Project Rubric

Overview

This rubric is here to help you understand the expectations for the analysis that you create. It is the same rubric that the person evaluating your project will use. We will refer to this person as the "project evaluator" in this document. You should look at the rubric **before you begin working** on your analysis **and before you submit it**.

How to Use: before you begin

1. Look at the bold headings under the criteria column to understand what the project evaluator will be looking for in your project.
2. Go through each criteria item in more detail.
3. Familiarize yourself with what is required for your project to "meet specifications" or to be "completely Udacious" (“exceeds specifications”). In order to gain a certificate, you need to "meet specifications", however, to gain the most benefit and learn most from the experience, we encourage you to aim for "completely Udacious".

How to Use: before you submit

1. Once your analysis is complete, go through each criteria item and do your best to honestly evaluate where you think your project falls.
2. If you think your project "does not meet specifications" for **any**

criteria item, then you should make some changes to your analysis.

3. Once you're confident that your project "meets specifications" or is "completely Udacious," go ahead and follow the Project Submission Instructions to submit!

How Grading Works

1. Your project evaluator will use this rubric to evaluate your analysis.
2. Your grade will simply be "pass" or "doesn't pass."
    a. You pass if your project meets or exceeds specifications in each criteria.
    b. If any criteria item "does not meet specifications," you will not pass. You will be able to make changes and re-submit the project.

The Rubric

Criteria	Does not meet specifications	Meets specifications	Exceeds specifications (Completely Udacious)
Code Functionality			
Does the code work?	Some code is not functional, producing an error or preventing the RMD document from being knit.	All code is functional.	*Not Applicable*
Does project utilize good coding practices?	The project sometimes uses repetitive code where a function would be more appropriate. The code uses constants or column numbers to access variables or subsets of data.	The project almost never uses repetitive code where a function would be more appropriate. The code references variables by name instead of using constants or column numbers.	The code is never repetitive and makes use of functions where appropriate and uses sound practices to access variables, subset data, or perform complex operations.

Project Readability			
Is the R code in the student's RMD file commented in a way that is useful and not superfluous?	Code is not commented or complex code is not adequately explained with comments. It is not always clear what the code is doing.	All complex code is adequately explained with comments. It is always clear what the code is doing.	All complex code is well explained with comments, and comments are not overused to explain obvious code. It is always clear what the code is is doing and how and why any unusual coding decisions were made.
Does student's code use formatting techniques (indents, spaces, line breaks, etc…) to improve readability? (Refer to Hadley Wickham's R Style Guide)	The code does not use formatting techniques or formatting techniques do not improve readability. Some lines are longer than 80 characters.	The code uses formatting techniques to improve code readability. All lines are shorter than 80 characters.	The code uses formatting techniques in a consistent and effective manner to improve code readability. All lines are shorter than 80 characters.
Is Markdown used to improve the presentation of the knitted HTML file? (e.g. section headers, text and paragraph spacing, document styles)	There are significant areas in the knitted HTML file where use of Markdown in the RMD would greatly improve readability.	Markdown syntax is used in the code to improve readability of the knitted file.	***Not Applicable***
Quality of Analysis			
Is the data set explored in many ways?	The project does not appropriately use univariate, bivariate, and multivariate plots to explore most of the expected relationships in the data set.	The project appropriately uses univariate, bivariate, and multivariate plots to explore most of the expected relationships in the data set.	The project uses many plot types to explore expected and unexpected relationships in the data. A variety of leading questions, dead-ends, and alternate approaches are presented.
Are questions and observations included as text throughout the analysis? (i.e. Every plot or set of related plots is followed by text interpreting the plot(s).)	Questions or findings are missing in multiple places which make it unclear what the student was thinking or what the student found.	Questions and findings are placed between blocks of R code regularly so it is clear what the student was thinking throughout the analysis.	***Not Applicable***
Is the flow of the analysis easy to follow?	Reasons for making each plot or set of plots are not always clear from the text.	Reasoning is provided for the plots made throughout the analysis. Plots made follow a logical flow.	***Not Applicable***

Are there a variety of relevant visualizations and statistical summaries?	The project contains fewer than 20 visualizations. Relevant statistics, such as means, medians, quartiles, or confidence intervals, are not reported to support inferences regarding the data.	The project contains at least 20 visualizations. The visualizations are varied and show multiple comparisons and trends. Relevant statistics are computed throughout the analysis when an inference is made about the data.	The project contains a variety of visualizations that show multiple comparisons and trends. Relevant statistics are calculated throughout the analysis where they would benefit inferences and are included in visualizations.
Is the data visualized using appropriate plots and parameter choices?	There are visualizations in the project that would benefit interpretability through choice of different plot type, variables plotted, or parameter choice (e.g. bin width, color, axis breaks).	Visualizations made in the project depict the data in an appropriate manner that allows plots to be readily interpreted.	Visualizations made in the project depict the data in a way that demonstrates a deep understanding of the data's structure.
## Final Plots and Summary			
Has a Final Plots and Summary section been included in the project?	The project does not include a dedicated Final Plots and Summary section containing three plots and commentary, or at least one plot does not follow from what was explored in the main analysis.	The project includes a Final Plots and Summary section containing three plots and commentary. Plots reflect what has been explored in the main body of the analysis.	***Not Applicable***
Are the final three plots varied and do they meet some of the following criteria: ● Draw comparisons. ● Identify trends. ● Engage a wide audience. ● Explain a complicated finding. ● Clarify a gap between perception and reality. ● Enable the reader to digest large amounts of information.	The plots chosen for the section seem to have been selected arbitrarily, do not fulfill at least 2 of the criteria, or are all of the same plot type.	The plots are well chosen and the plots fulfill at least 2 of the criteria. The plots are varied and reveal interesting trends and relationships.	Each plot reveals an important and different comparison or trend in the data. The plots incorporate many of the variables from the data set in a way that allows the plots to convey a lot of information while still being interpreted easily. The plots fulfill 4 or more of the criteria.
Are the plots appropriate?	One or more plots can be significantly improved by selection of a different variable, different plot type, or different sequence of plots.	All plots have appropriately selected variables and are plotted in a way that accurately conveys the data/information.	***Not Applicable***

Are the plots polished? ● are axes labeled? ● are units labeled on each axis? ● are plots titled? ● are all labels, titles, and units readable?	One or more plots are missing axes labels, plot titles, axes units, or are scaled inappropriately.	All plots are labeled appropriately and can be read and interpreted easily.	***Not Applicable***
Are the plots explained?	The reasoning and findings are not explained for each plot, the text about one plot is not descriptive enough to stand alone, or comments do not reflect the contents of an associated plot.	The reasoning and findings from each plot are explained and the text about each plot is descriptive enough to stand alone. Comments reflect the contents of the plots that they are associated with.	The reasoning and findings from each plot are explained concisely with appropriate variable transformations, other plot decisions, and/or statistics. The text about each figure is descriptive and adds information that the graphic itself would not easily explain.
Reflection			
Has a Reflection section been included in the project?	The project does not include a dedicated Reflection section to reflect upon the analysis performed.	The project includes a Reflection section discussing the analysis performed.	***Not Applicable***
Does the section provide a written reflection of the analysis? Consider the following in your reflections: ● Where did I run into difficulties in the analysis? ● Where did I find successes? ● How could the analysis be enriched in future work (e.g. additional data and analyses)?	The section does not communicate struggles, successes, and ideas for improvement.	The section reflects on how the analysis was conducted and reports on the struggles and successes throughout the analysis. The section provides at least one idea or question for future work.	The section provides a rich and well-written reflection of struggles, successes, and lessons learned. The section poses ideas or questions for future work. The section explains any important decisions in the analysis and how those decisions affected the analysis.