PROJECT

Explore and Summarize Data

A part of the Data Analyst Nanodegree Program

PROJECT REVIEW | CODE REVIEW | NOTES

Meets Specifications

SHARE YOUR ACCOMPLISHMENT

Code Functionality

✓ **All code is functional (e.g. No Error is produced and RMD document is not prevented from being knit.)**

✓ **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.**

Well Done for demonstrating the use of functions that reduce repetitions and simplify the code.

Project Readability

✓ **All complex code is adequately explained with comments. It is always clear what the code is doing and how and why any unusual coding decisions were made.**

✓ **The code uses formatting techniques in a consistent and effective manner to improve code readability. All lines are shorter than 80 characters.**

There are some places where you exceed the maximum line length. This seems picky but the limit is widespread convention that ensure that future programmers can read your code easily no matter what their text editor and window size preferences are. One way to hem things in is by breaking up lists with line breaks. RStudio does the indentation automatically when you add a line break in the middle of a parameter list. RStudio also has a built in feature for finding overly long lines. in the Code Editing section of the preferences there's an option called "Show margin" that puts a line length indicator in the code editor.

✓ **Markdown syntax is used in the RMD file to improve readability of the knitted file.**

Quality of Analysis

✓ **The project appropriately uses univariate, bivariate, and multivariate plots to explore most of the expected relationships in the data set.**

The analysis make use of different chart type that explore many aspects about the data set.

✓ **Questions and findings are placed between blocks of R code regularly so it is clear what the student was thinking throughout the analysis.**

The discussion between code block include relevant questions and interesting findings.

✓ **Reasoning is provided for the plots made throughout the analysis. Plots made follow a logical flow. Comments following plots accurately reflect the plots' contents.**

The analysis follow a logical flow where the results of one analysis lead to another.

For the univariate section, please consider to expand the discussion about the outliers for each feature. You can even remove outliers if you find it appropriate, that will make the following analysis more robust.
http://www.public.iastate.edu/~maitra/stat501/lectures/Outliers.pdf

For the bivariate section , it is awesome that you are using the correlation values to focus and guide the following analysis. Please consider to expand the discussion under each chart to include some of your insights or findings . that will make it easier for you readers to follow the line of yoru thoughts.

✓ **The project contains at least 20 visualizations. The visualizations are varied and show multiple comparisons and trends. Relevant statistics (e.g. mean, median, confidence intervals, correlations) are computed throughout the analysis when an inference is made about the data.**

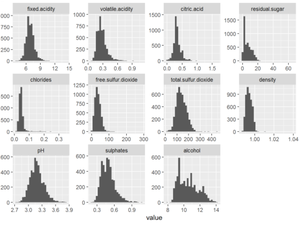
The analysis include many figures that depict comparison, trends and relations between features.

Please consider to include the relevant statistics that you calculated next to each figure. That will be usefull for your readers but also for the analysis. For example the mean median and quartiles can quantify the distribution depicted with the histograms, and the correlation values can be useful to quantify the relations depicted in the bivariate and multivariate sections.

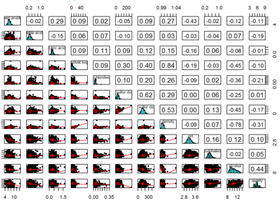
✓ **Visualizations made in the project depict the data in an appropriate manner that allows plots to be readily interpreted. Choice of plot type, variables, and aesthetic parameters (e.g. bin width, color, axis breaks) is appropriate.**

Most of the charts are well done, so I only have few comments here.

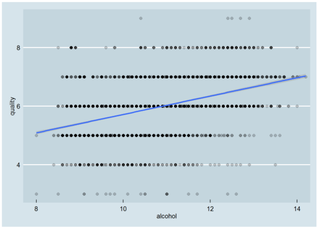
Since the report include a seperate figures that depict the distribution of each feature, perhaps the multi figure histogram is redundant.



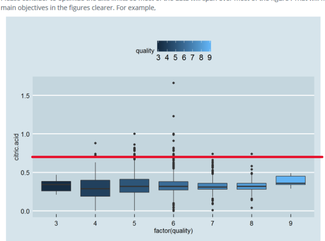
Please consider to increase the size of the correlation figure so the name of the features will be clearer.
https://rstudio.github.io/dygraphs/r-markdown.html



For the figure here bellow , please note that the overplotting make it difficult to appreciate the distribution of alcohol for each quality rating, please consider to use box plot (like you did below) or jitter plot .



Please consider to optimize the axis limits so most of the data will span over most of the figure . That will make the main objectives in the figures clearer. For example,



Final Plots and Summary

✓ **The project includes a Final Plots and Summary section containing three plots and commentary. All plots in this section reflect what has been explored in the main body of the analysis.**

The final plot section include 3 figures that represents the analysis done in the exploratory sections.

✓ **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.**

✓ **All plots have appropriately selected variables and are plotted in a way that accurately conveys the data/information (i.e findings in Final Plot 1 do not depend on the findings of Final Plot 2).**

✓ **All plots are labeled appropriately (axis labels, plot titles, axis units) and can be read and interpreted easily. Plots are scaled appropriately.**

Please include title for each figure in the final plot section.

Since you are using blue as one of the color coding, perhaps white background will make the figure clearer.

✓ **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.**

Please consider to include more statistics to support your findings in the final plot section.

Reflection

✓ **The project includes a Reflection section discussing the analysis performed.**

✓ **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 explains any important decisions in the analysis and how those decisions affected the analysis.**

Please consider to expand the discussion in the reflection section to describe also some of the struggles and successes throughout the analysis.

DOWNLOAD PROJECT

Have a question about your review? Email us at review-support@udacity.com and include the link to this review.

RETURN TO PATH