UDACITY

PROJECT

Explore and Summarize Data

A part of the Data Analyst Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

Excellent work on passing all the specifications on your first try. That is not an easy feat for this project, I commend you for all your hard work here, and I wish you all the best for your next project.

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.

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.

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.

Well done for ensuring all code lines are less than 80 characters long, and consistent formatting used throughout the code.

As a reference, here is a style guide to R code to follow: Hadley Wickham's R style guide.

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

The report is well-presented. Well done on applying Markdown syntax properly.

(Optional) Dynamic table of contents

This is not critical, but since the report is quite long, I think a table of content would be a useful feature to incorporate here. The following Markdown syntax can be added on top of the document to create a dynamic table of contents that can also act as a navigator:

```
---
title: "Analysis Title"
author: Your Name
date: [Month] [date], [year]
output:
  html_document:
    toc: true
    toc_depth: 3
    toc_float: true
---
```

Quality of Analysis

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

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

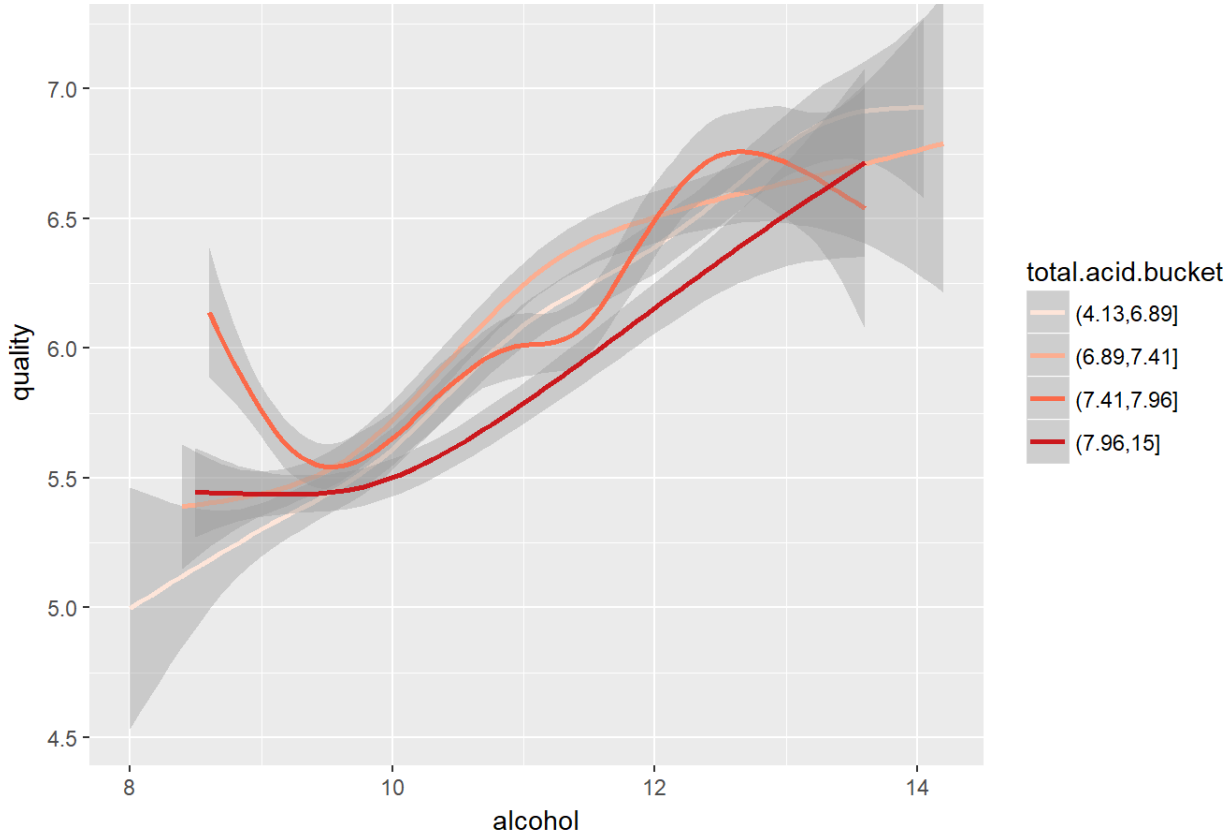
The project contains at least 20 visualizations. The visualizations are varied and show multiple comparisons and trends. Relevant statistics (e.g. mean, median, confidence intervals, correlations) are computed throughout the analysis when an inference is made about the data.

More than 20 visualizations are provided, and statistics have been calculated and presented. Currently, all summary statistics are grouped on top of the report which is enough to pass this specification, but it is a good idea to present them right after relevant plots and discussions to make them more accessible to readers.

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.

All plots look polished, good work!

(Optional) Improper use of line plot



Line plot is most effective when the goal is to visualize changes in the x-axis data. In other words, one point in the x-axis must have a continuity or dependant on a point on its left and predates the point on its right i.e. periodical. That does not seem to be the case in the above visualization.

Read more here: https://infogram.com/blog/the-line-chart-how-and-when-to-use-it/

I would say a boxplot or violin plot might be a better visualization type to use here, especially since readers can find other useful statistics from it, like quartiles and outliers for each category.

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 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)

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.

I do not mark this off since this is a minor issue; in final plot three, I suggest including unit of measurement for residual sugar as well.

Adding a unit of measurement even though it is obvious is *super important*. NASA can tell you more about it:

> NASA lost its $125-million Mars Climate Orbiter because spacecraft engineers failed to convert from English to metric measurements when exchanging vital data before the craft was launched, space agency officials said Thursday.
>
> A navigation team at the Jet Propulsion Laboratory used the metric system of millimeters and meters in its calculations, while Lockheed Martin Astronautics in Denver, which designed and built the spacecraft, provided crucial acceleration data in the English system of inches, feet and pounds.
>
> As a result, JPL engineers mistook acceleration readings measured in English units of pound-seconds for a metric measure of force called newton-seconds.
>
> In a sense, the spacecraft was lost in translation.

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.

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.

You have written a proper Reflection section. A report on struggles and successes throughout analysis have been provided, including reasons for doing certain analysis, and provided a couple of ideas for future work. You have done an excellent job here, well done!

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

Student FAQ