**Explore and Summarize Data**

* [Review](https://review.udacity.com/#!/reviews/1573809)
* [Code Review](https://review.udacity.com/#!/reviews/1573809)
* [History](https://review.udacity.com/#!/reviews/1573809)

### Meets Specifications

Dear student,  
Great work on this resubmission! It now meets all the required criteria!   
I hope you've enjoyed this project, as it gives us a taste on how to analyze a dataset from scratch and discover some interesting relationships between variables.   
These are skills we will continue to develop every day on as they are the core of the job of a data analyst.   
Before going ahead, please take a moment to see some suggestions I left below. I hope you find them useful!  
Keep up the great work and good luck on the 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.

#### SUGGESTION

Even though most code blocks are reasonably short and easy to read, it would still be a good idea to add a comment in each code block where data cleansing or more than a single visualization are being created.

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

Awesome code formatting! 👍

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.

#### SUGGESTION

I really appreciate the use of box-plots here. These are important plots as they allow the reader to understand how the data is distributed.  
My suggestion here would be to increment these box-plots with overlays that would show the actual data points. This is nice because it will give the reader an idea of how big the data sample is and how trustworthy we can be of the insights we extract. To create this type of plot, please take a look at [this page](https://stackoverflow.com/questions/23675735/how-to-add-boxplots-to-scatterplot-with-jitter).

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.

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.

### 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).

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

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.

#### SUGGESTIONS

In this section, I would advise on thinking about the following:

* Do you have any idea on how statistical tests could be used to compare wines anyhow? What kind of test would you like to apply for each analysis we have conducted here thus far?
* The trends we observed here are for white wines. Do you believe the behavior of the variables would be maintained for red wines? What do you expect would change?
* Also, we are only analyzing one specific type of wine from one specific country. What do you think would change for other wines from other types of the globe?
* Finally, it would also be nice to think about why we have so many ‘average’ wines compared to poor and good ones. In practice, the ratings we have are averages that we obtained from different experts. Does that bias our analysis anyhow? Do you think ‘averaging’ the ratings provided by different tasters relate anyhow with having too many “average” final ratings?