



---

## PROJECT

### Explore and Summarize Data

A part of the Data Analyst Nanodegree Program

---

## PROJECT REVIEW

CODE REVIEW
NOTES

SHARE YOUR ACCOMPLISHMENT!  

## Requires Changes

### 6 SPECIFICATIONS REQUIRE CHANGES

You are on a great start!

This is definitely one of the better projects for a first try!

I look forward to your next submission :)

## 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 showing the use of custom functions to reduce repetitive codes :)

## 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.

Great job commenting where necessary :)

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 lines over 80 characters.

To help meet this specification, go to Tools - Global Options - Code - Display and check "show margin" to make a vertical line appear at the 80 character mark.

Using line breaks where possible will help keep your codes clean and consistent.

For example, line 237 should be:

```
chart <- createhistogram(wine,
                          "alcohol",
                          exclude_outlier = TRUE,
                          color = "gray",
                          bins = 28,
                          title = "log 10",
                          unit="% / vol")
```

*Note the spacing after commas and spacing before&after arithmetic signs.*

Please note that you should use consistent formatting(consistent line breaks, etc) throughout your code even if the lines are not going over 80 characters.

This guide will be helpful:

<https://google.github.io/styleguide/Rguide.xml>

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

Amazing use of tabs to clean up the format of the project file :)

Just make sure to check for simple typos and left over # tags as in the below:

We can identify 3 different quality groups. The low one with score up to 4. The medium ones with score of 5, 6 or 7. The good ones with scores of 8 or 9.

Most of the wines are considered having a medium quality. We can see we have very few wines considered as bad (quality = 3) and even fewer rated very good (quality = 9). We have no "crappy" wines (quality = 0 or 1) nor outstanding wines (quality = 10). ####

## Quality of Analysis

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

Great job exploring the univariate distributions in depth before moving onto more complex relationships :)

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

Well done providing detailed insights and findings for the plots created :)

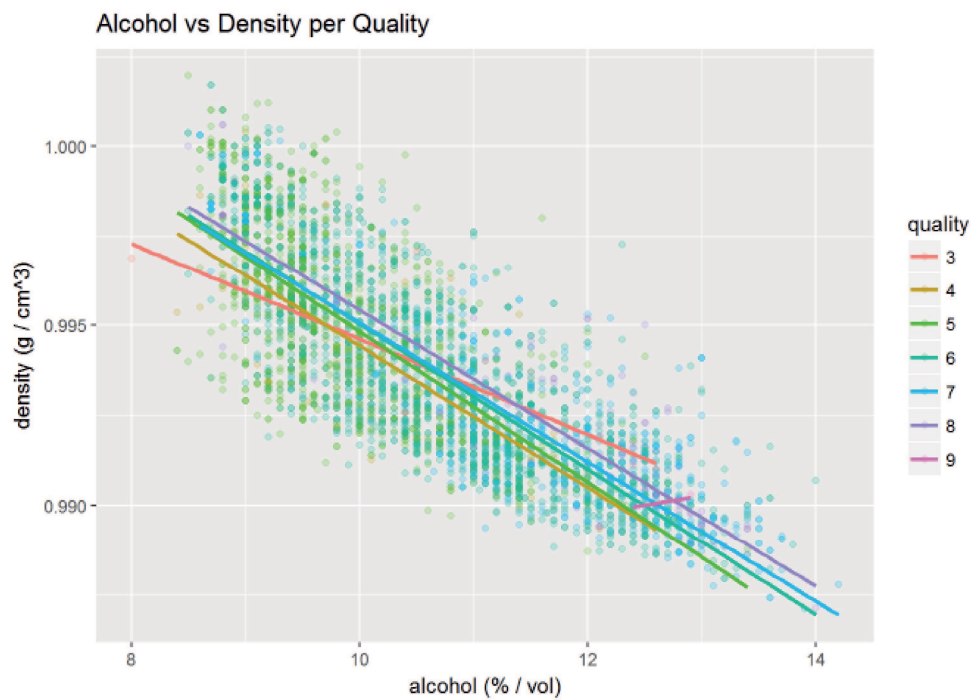
Reasoning is provided for the plots made throughout the analysis. Plots made follow a logical flow. Comments following plots accurately reflect the plots' contents.

The project has a nice flow and is easy to follow :)

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.

Great job supporting the plots with relevant stats such as correlation factors :)

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.

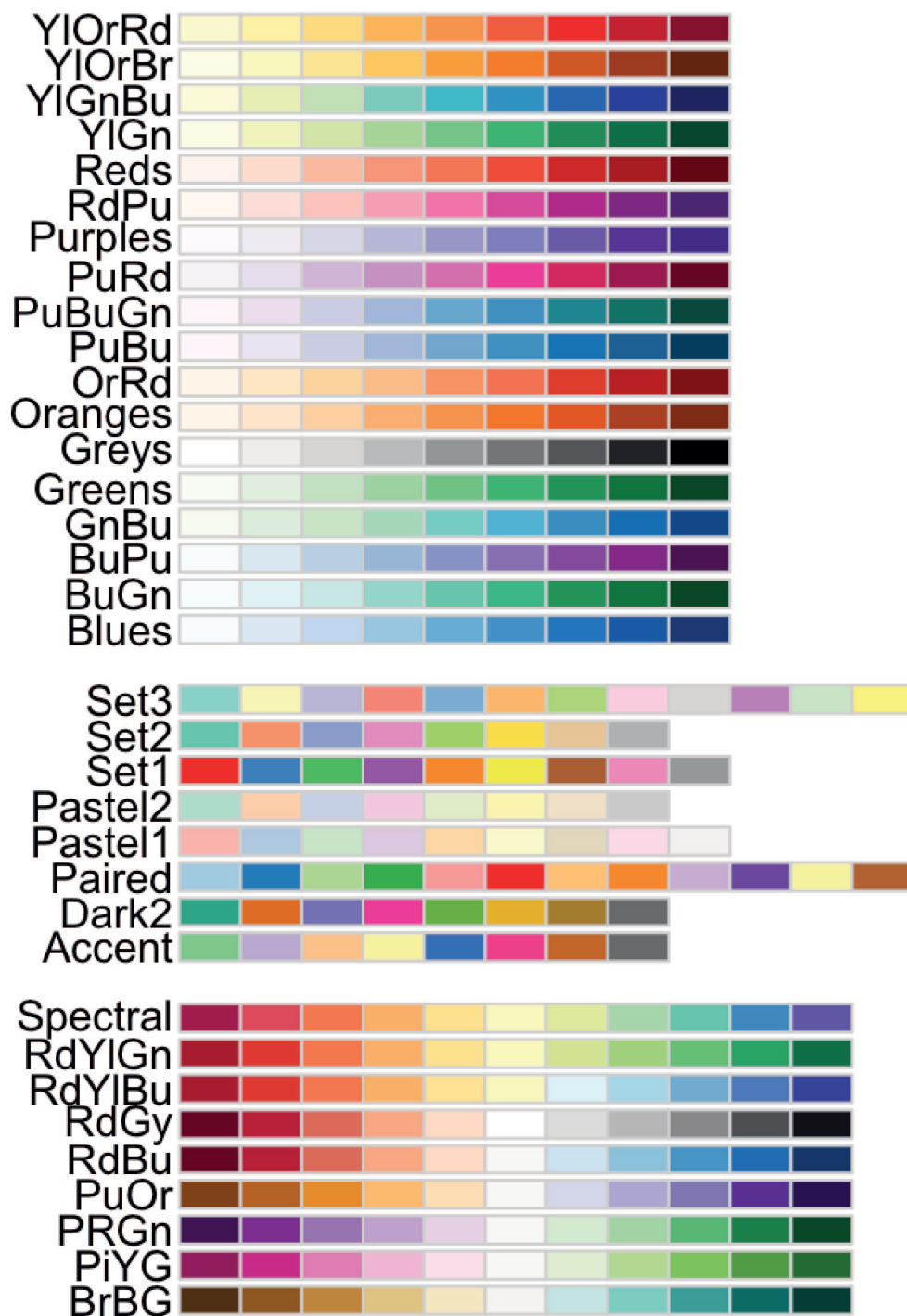


This is a nice plot but the current colors are not effective in showing the differences in quality. You will notice that the 7 colors are pretty much random and the readers have to constantly look back and forth between the points and the color legend.

When you are using colors to encode quantifiable categories such as quality (where there are "bad" and "good" values), make sure to use sequential/diverging color scheme.

Take a look at how `scale_color_brewer` is used in the following link (such as the diamond and green palette or the YlGnBu palette for clearer variation.):

[http://docs.ggplot2.org/current/scale\\_brewer.html](http://docs.ggplot2.org/current/scale_brewer.html)



<http://www.sthda.com/english/wiki/colors-in-r>

## 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.

Great choice of plots!

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).

Once again, please use an appropriate color palette for the last plot.

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

Let's polish the final plots as much as possible.  
Therefore, plot #1 could use a better title ;)

If possible, fix the variable names such as free.sulfur.dioxide to "Free Sulfur Dioxide."

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 be more descriptive with the explanations and include more stats.  
Note that the descriptions should be able to stand alone. In other words, readers should be able to get a pretty good idea just by reading the descriptions.

For example,

In this chart, we can see that some wine grower do not fully master the wine making process.

What specific part/pattern show such finding? Are you referring to the median values? Can you support it with stats? etc.

## 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.

Note that the reflection should be more of a personal reflection rather than a summary. Please elaborate on your struggles and successes during the analysis. Where did you get stuck and how did you overcome the issue? Where were you successful? etc

 RESUBMIT

 [DOWNLOAD PROJECT](#)



## Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

 [Watch Video](#) (3:01)

Have a question about your review? Email us at [review-support@udacity.com](mailto:review-support@udacity.com) and include the link to this review.

RETURN TO PATH



---

[Student FAQ](#)