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 here! There are only few Explore and Summarize Data projects that passed all specifications on first try, good work on making your project one of them.

### 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.**

Variables are referenced by their names throughout the project, good work.

As a suggestion, you can remove repetition by using custom functions. For example, in Univariate section, you can create a user defined function that wraps a `ggplot` function, then create the plots in that block using that function. For example, line 76 and 77 can be replaced as follows:

```
create_plot <- function(varname, binwidth = 0.1) {
  return(qplot(x = aes_string(x = varname), data = df, binwidth = binwidth))
}
```

You can then use it to draw ggplots with far less code, e.g.

```
create_plot('residual.sugar', 0.1)
create_plot('residual.sugar', 0.05) + scale_x_log10(breaks=c(1,1.2,2,5,8,10,20,25,30))
```

### 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.**

Some lines are much longer than 80 characters that they are not easily readable. Try reviewing all project's code again and separate lines of code with newlines to adjust their length. This is a good practice to ensure your code are readable throughout a wide array of screen sizes your colleagues might be using.

RStudio can draw a vertical line at 80 characters for you so that it's easy to check. This webpage shows how to do that with 'show margin'.

As another motivation for doing this, know that this habit is a standard for Google engineers.

However, seeing that this is the only specification you missed, and I think it is quite minor, I can mark this as passing the specification.

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

Well done for properly using Markdown syntax for better report's readability.

## Quality of Analysis

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

Well done for using the right plot types within exploration sections. "Alcohol by Quality" can be argued as part of Bivariate Plots section, but as I see this plot is only used to conclude your findings from other univariate plots, I think it is alright to leave it here (although I would place it under Univariate Analysis section, if it were me).

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

Great job accompanying all R code blocks (plots and statistics) with comments.

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

You have done a really good job ensuring the comments surrounding plots and statistics follow a logical flow. What I like the most for this section is how you included which direction your analysis is going to have early above in the introduction section.

**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.**

Throughout the project you accompanied your visualizations with heaps of statistical calculations and insightful discussions. You have also approached the dataset with multiple visualizations showing different point of views of the information. Excellent job here.

**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.**

Superb job here! I found it really rarely a project meets this specification on its first try. Variables were scaled appropriately, and the diverging color theme used to encode quantity helped readers quickly find which areas contained worse, medium, or good wines separated by different color spectrum.

I have only one minor suggestion here: Variable names are somewhat hard to read in `pairs.panels` plot. I suggest to enlarge figure size by adding knitr chunk options `fig.width` and `fig.height`.

## 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.**

Good job including three final plots you have explored in your previous exploration 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 of your plots have all the needed visual cues and statistical information that fulfill all the points in this rubric. Well done for putting so much thoughts in your analysis and design of your visualizations.

**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).**

Good job making sure final plots are ordered properly and using the right variables.

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

Great job labeling and scaling all the final plots appropriately. I see that unit of measurements are included in axis labels, well done.

**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.**

You have provided very clear reasoning and findings for each of your final plots. Well done. I find it really great that you included your insights in your descriptions to the final plots, as well as explaining the plots in detail.

## 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.**

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

Student FAQ