

Project Review 2

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

requires changes

Same as mentioned in last review: In univariate plots session, plots should be detailed enough to give a complete picture of the data, and that coordinates are chosen and displayed appropriately. **Multivariate plots should not be the first time a variable appears in the report, ideally they should have been explored in both the univariate and bivariate sections.**

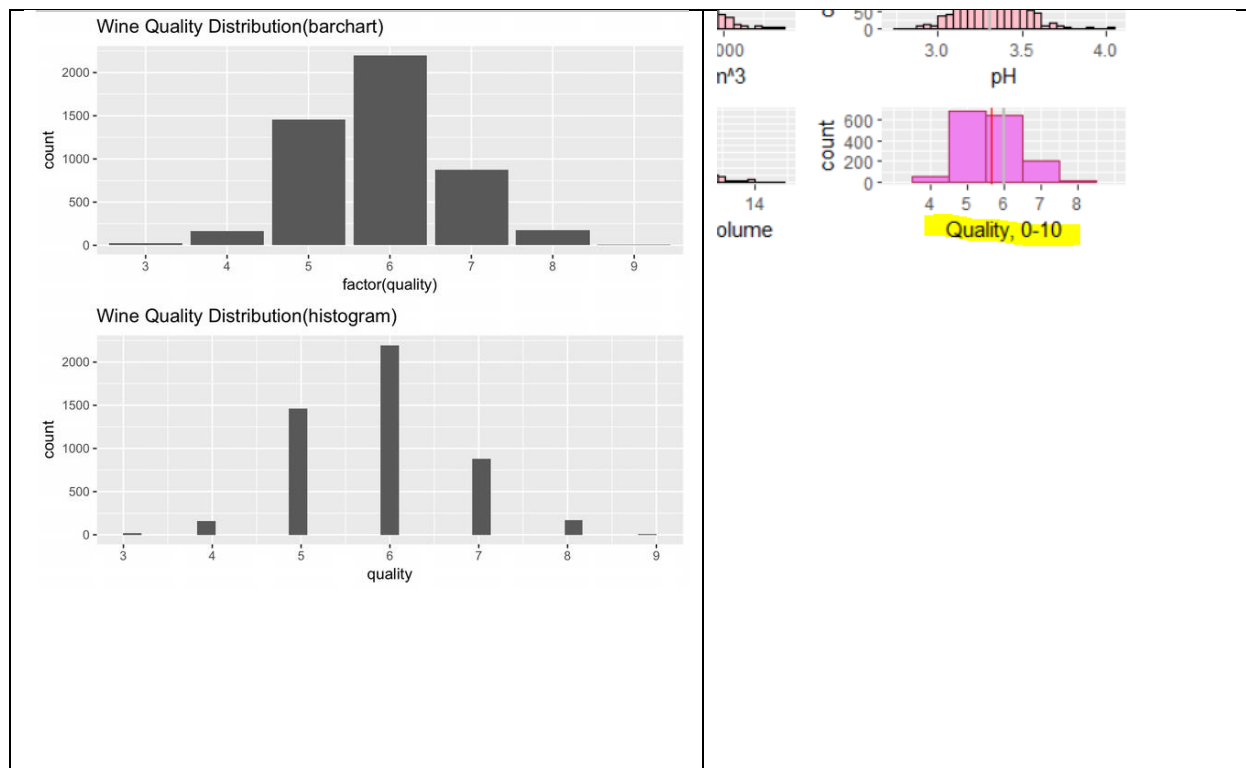
`ggpairs` is bivariate plot, please move it to bivariate plot section.

I've moved the *ggpairs* to the Bivariate Section. This is despite Reviewer 1 understanding my approach's rationale:

1. "If you would still prefer to present them before other univariate plots, I suggest adding another section above the Univariate Plots section for these plots."
 - a. I went with this option with a subsection named *Data Overview*.

The summary() was already there in my data overview section, but I moved it to the univariate section. Regarding the bin discussion, I did have the larger bin version in my report.

Reviewer 2's Suggestion	My Histogram
-------------------------	--------------



Reviewer 2's other suggestions on histogram, bar charts, and subsetting continuous data were very insightful.

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

- This unit I marked not pass, as mentioned above, I would suggested try improve plots' visualizations and choose better final plots as a summary of this EDA.
- If you choose ggpairs here, please polish the plot by fix the labels and values overlapping problem in ggpairs.

This is a great suggestion. I've incorporated "fig.height= , fig.width= " in three plot matrix. I will also took the reviewer's *ggcorr* by including one in my report. I actually used a simple correlation for some other analyses not shown in this final report. E.g.

```
cor(red1[c("fixed.acidity", "volatile.acidity", "citric.acid",
          "residual.sugar", "chlorides", "quality")]))
```

I see the value of *ggcorr*, especially when you have many variables and color coding would help to quickly identify highly correlated variables.

I respect the reviewer 2's other comments such as "*ggpairs is more like a tool rather than a plot that shows the significant findings. I would suggest not put them as final plots,*". This EDA course is not a data visualization course, which I have completed with Tableau course. If I have to redo the course, I'd

certainly consider having plots to show significant findings. My final significant findings were not based so much on plots,. Rather, they were based on the multivariate statistical analyses.

Personally, I cannot tell if two or more lines based on a variable stratification are different unless they are very obvious. These plot plus the uni-, bi-, and multivariate plot really helped in my variable selection for the statistical analysis. Again, in answering the research question – which is the overarching goal of this project – I utilized the regression analyses.