

Project Review

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.

Old

```
```{r Data Overview1}
ggpairs(red1[c(7:13)],
 lower = list(continuous = wrap("points", shape = I('.'))),
 upper = list(combo = wrap("box", outlier.shape = I('.'))))
```
```

New

```
```{r Data Overview1}

ggpairs(red1[c("free.sulfur.dioxide", "total.sulfur.dioxide", "density",
 "pH", "sulphates", "alcohol", "quality")],
 lower = list(continuous = wrap("points", shape = I('.'))),
 upper = list(combo = wrap("box", outlier.shape = I('.'))))
```
```

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

I've made all coding lines to be less than 80 characters.

I have the margin showed using the reviewer's suggested resource: <https://support.rstudio.com/hc/en-us/articles/200549016-Customizing-RStudio#editing>

Old

```
```{r Plotting for Normality Check 2, results='hide', message=FALSE, warning=FALSE}
Examples
2. Histogram
h01 <- ggplot(red1, aes(x=total.sulfur.dioxide)) +
 geom_histogram(color='red', fill='cyan') +
 ggtitle('Histogram: Total Sulfur Dioxide') +
 geom_vline(aes(xintercept = mean(total.sulfur.dioxide)), col='red', size=0.5) +
 geom_vline(aes(xintercept = median(total.sulfur.dioxide)), col = 'grey', size=1) +
 annotate("text", x = mean(red1$total.sulfur.dioxide) * 1.5, y = 150,
 label = paste0("Avg: ", round(mean(red1$total.sulfur.dioxide),1))) +
 annotate("text", x = median(red1$total.sulfur.dioxide) * 1.1, y = 200,
 label = paste0("Med: ", round(median(red1$total.sulfur.dioxide),1)))
```
```

New

```

```{r Plotting for Normality Check 2, results='hide', message=FALSE}
Examples
2. Histogram
h01 <- ggplot(red1, aes(x=total.sulfur.dioxide)) +
 geom_histogram(color='red', fill='cyan') +
 ggtitle('Histogram: Total Sulfur Dioxide') +
 geom_vline(aes(xintercept = mean(total.sulfur.dioxide)),
 col='red',size=0.5) +
 geom_vline(aes(xintercept = median(total.sulfur.dioxide)),
 col = 'grey', size=1) +
 annotate("text", x = mean(red1$total.sulfur.dioxide) * 1.5, y = 150,
 label = paste0("Avg: ", round(mean(red1$total.sulfur.dioxide),1))) +
 annotate("text", x = median(red1$total.sulfur.dioxide) * 1.1, y = 200,
 label = paste0("Med: ", round(median(red1$total.sulfur.dioxide),1)))

```

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

#### Incorrect plot placements + add some bivariate plots

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*.
2. "Is a multivariate plot, it should not be placed in the Bivariate Analysis section. Notice that each plot in this ggpairs plot has three variables (the third being quality)."
  - a. I moved the plot matrix stratified by quality category from the bivariate to multivariate.
3. "Please add a couple of proper bivariate plots in your report. Scatterplots or boxplots with two variables would do well here"
  - a. I added a boxplot matrix.
  - b. I added an correlation plot with smoothing line to look for possible non-linear relationship.

#### Explore all variables before using them in more complex plots

"quality.3 variable was used in some multivariate plots but it has not been fully explored in previous sections. Please include a histogram of this variable in the Univariate Plots section. This is important so readers may get a complete perspective on the variables before seeing them used in further analysis."

I added a histogram on the categorized quality variable in the 'Quality Categorized' section.

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

"Units of measurements need to be added in final plots e.g. "Alcohol (%)". Please also update other variables."

I've added unit of measurement in the final plots using *labs()*.