

Red Wine Quality Machine Learning Project

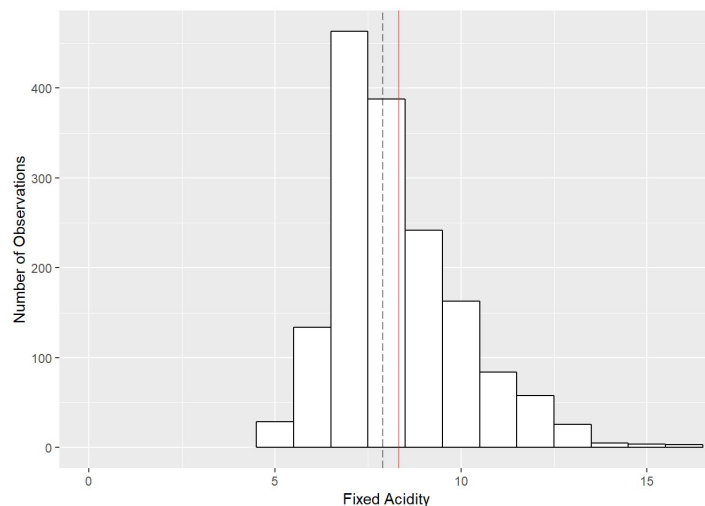
Deliverable 2:

Attribute Review:

I included a link to descriptions of how each attribute can affect one another and the dependent variable of overall wine quality in order to understand the background information and logically understand some of the statistical correlation. This information may also allow me to logically create and understand linear regression algorithms.

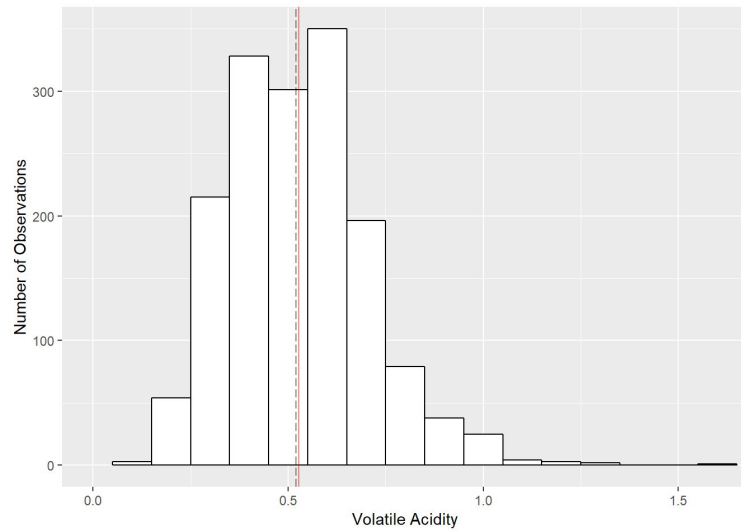
As previously stated in Deliverable 1, the variables included in the dataset are Fixed Acidity, Volatile Acidity, Citric Acid, Residual Sugar, Chlorides, Free Sulfur, Total Sulfur, Density, pH, Sulfates, Alcohol, and Quality.

1. Fixed Acidity: grams/liter



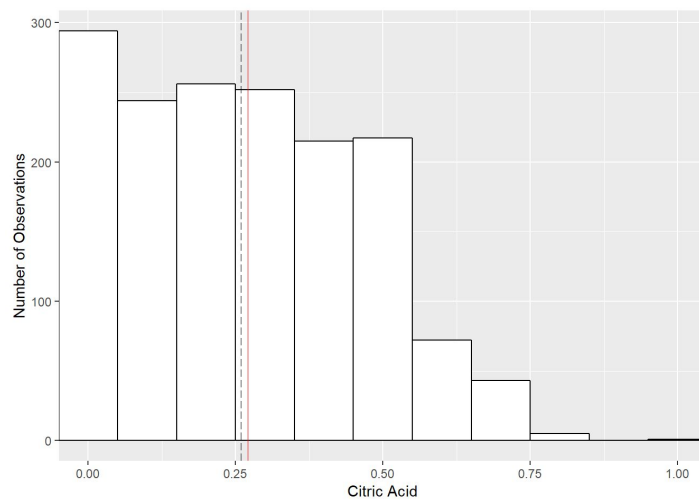
<https://waterhouse.ucdavis.edu/whats-in-wine/fixed-acidity>

2. Volatile Acidity: grams/liter



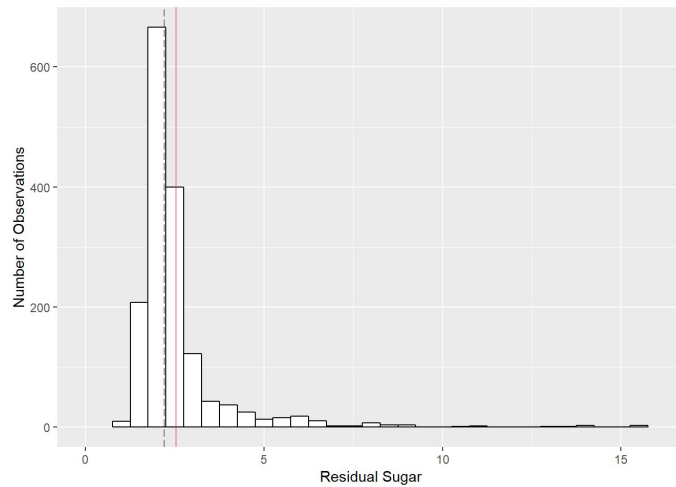
<https://waterhouse.ucdavis.edu/whats-in-wine/volatile-acidity>

3. Citric Acid: grams/liter



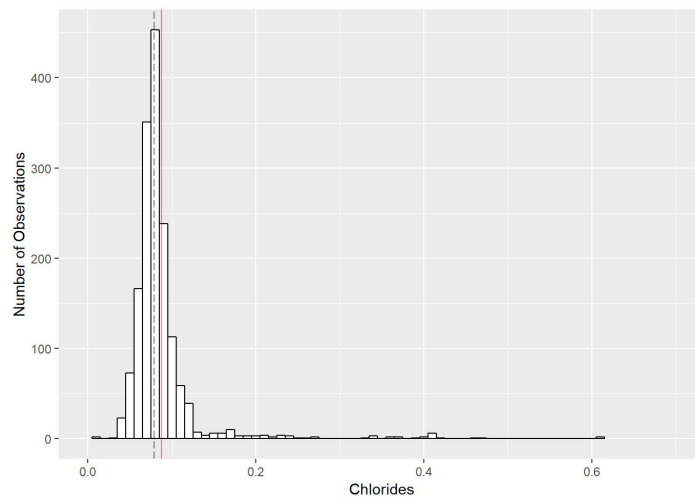
<https://wineserver.ucdavis.edu/industry-info/enology/methods-and-techniques/common-chemical-reagents/citric-acid>

4. Residual Sugar: grams/liter



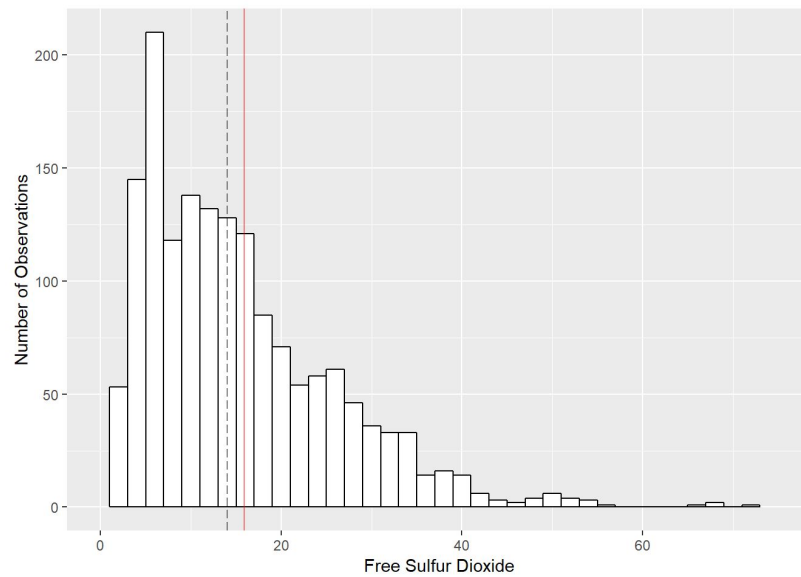
<https://winemakermag.com/technique/501-measuring-residual-sugar-techniques>

5. Chlorides: grams/liter



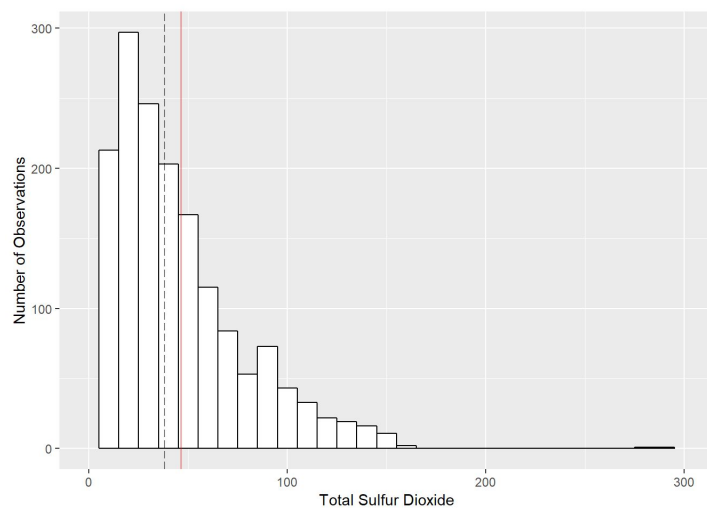
<https://wineserver.ucdavis.edu/industry-info/enology/methods-and-techniques/common-chemical-reagents/chlorine-dioxide>

6. Free Sulfur: milligrams/liter



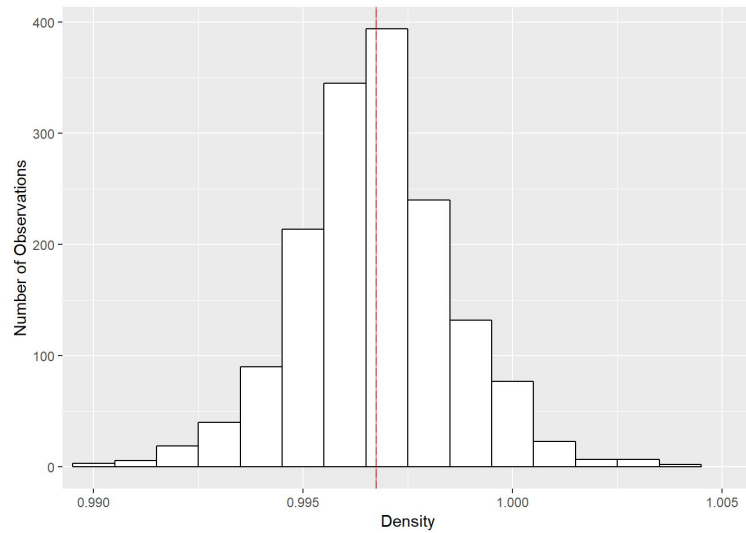
https://morewinemaking.com/articles/SO2_management

7. Total Sulfur: milligrams/liter



<https://www.extension.iastate.edu/wine/total-sulfur-dioxide-why-it-matters-too>

8. Density: grams/cubic centimeter

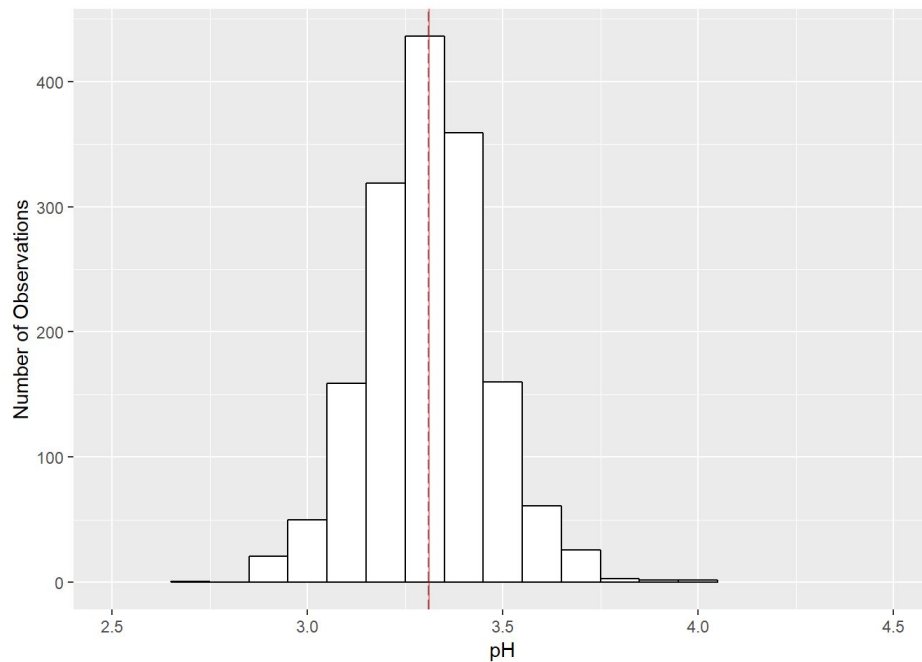


9. pH: 1-13

1-6=acidic

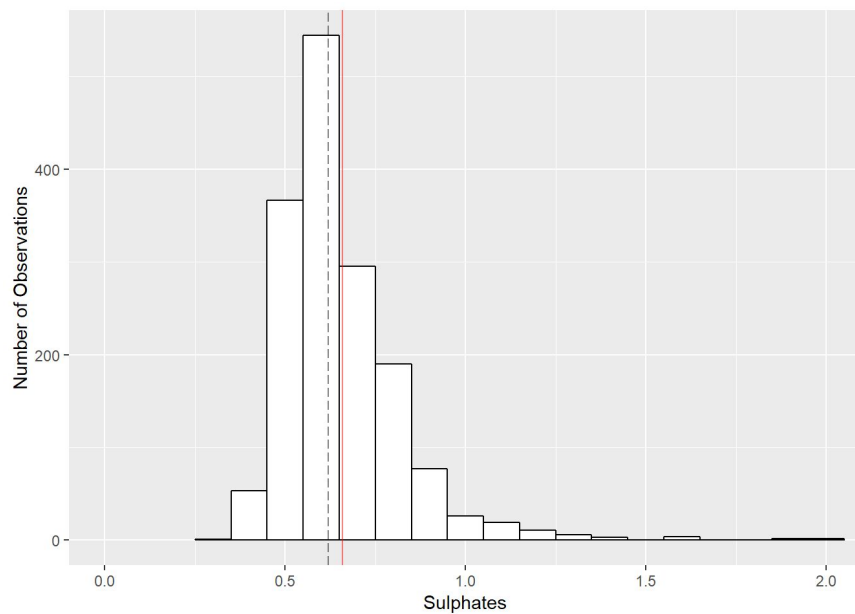
7=neutral

8-13=basic



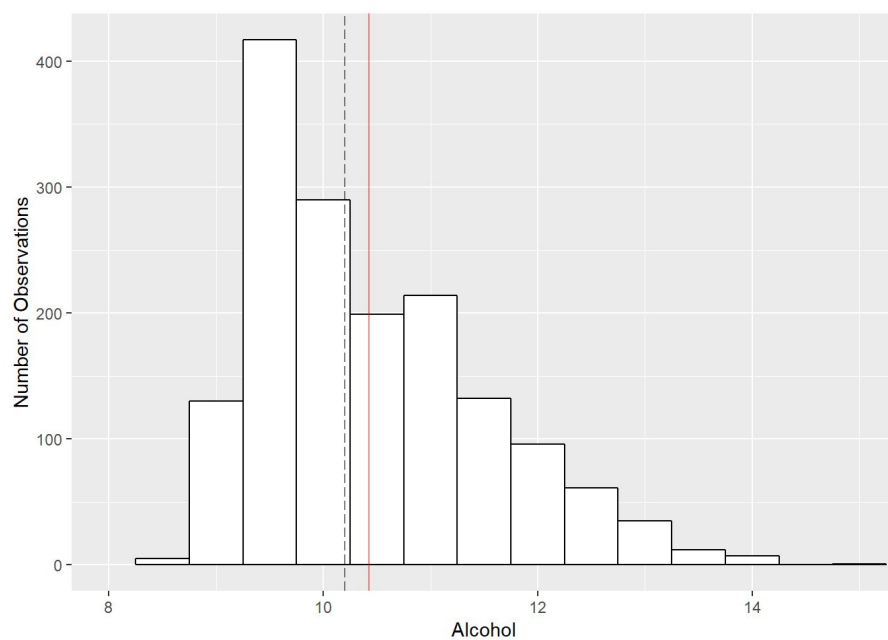
<https://www.winespectator.com/drvinny/show/id/5035>

10. Sulfates: grams/liter



<https://www.winespectator.com/drvinny/show/id/Difference-Between-Sulfites-Sulfates-Wine-54706>

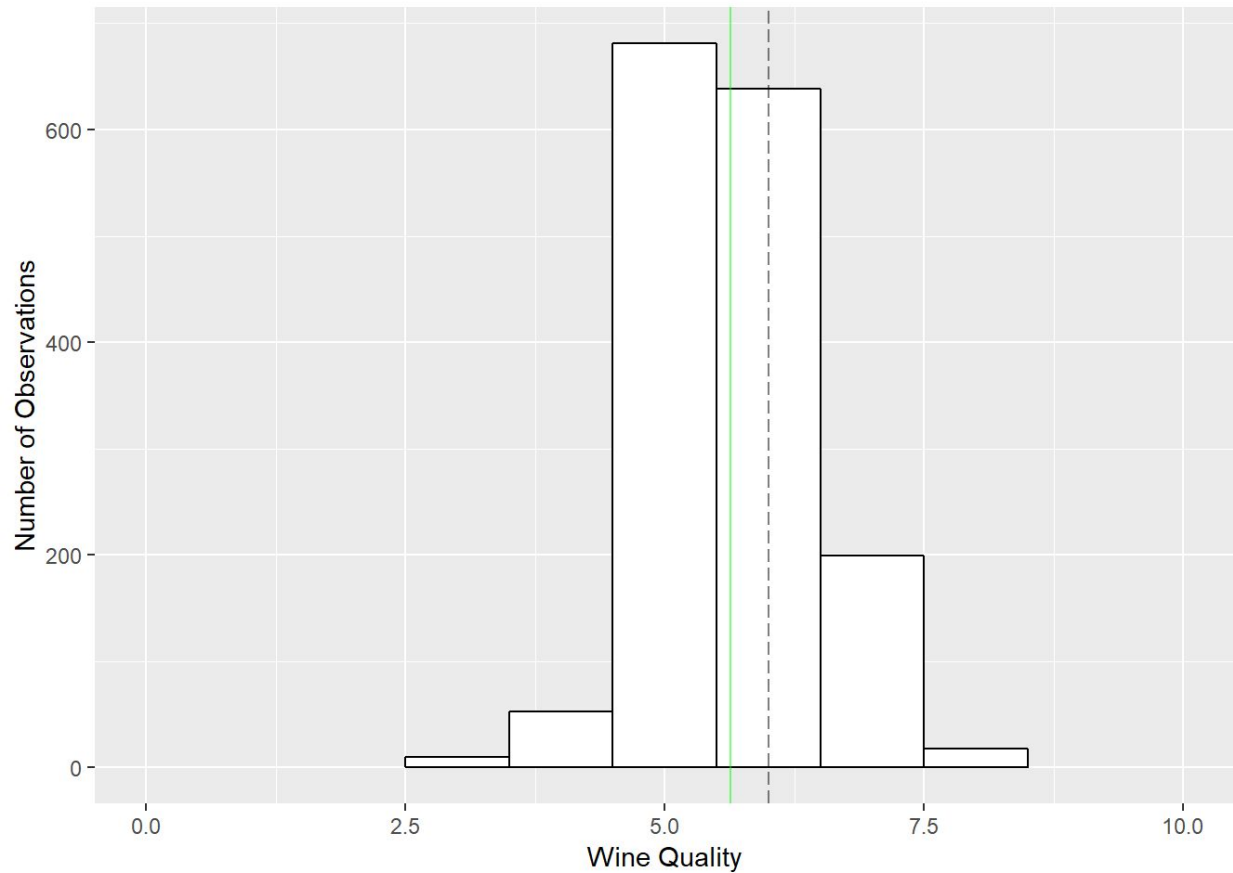
11. Alcohol: % volume



12. Quality: 1-10 where 10 is best.

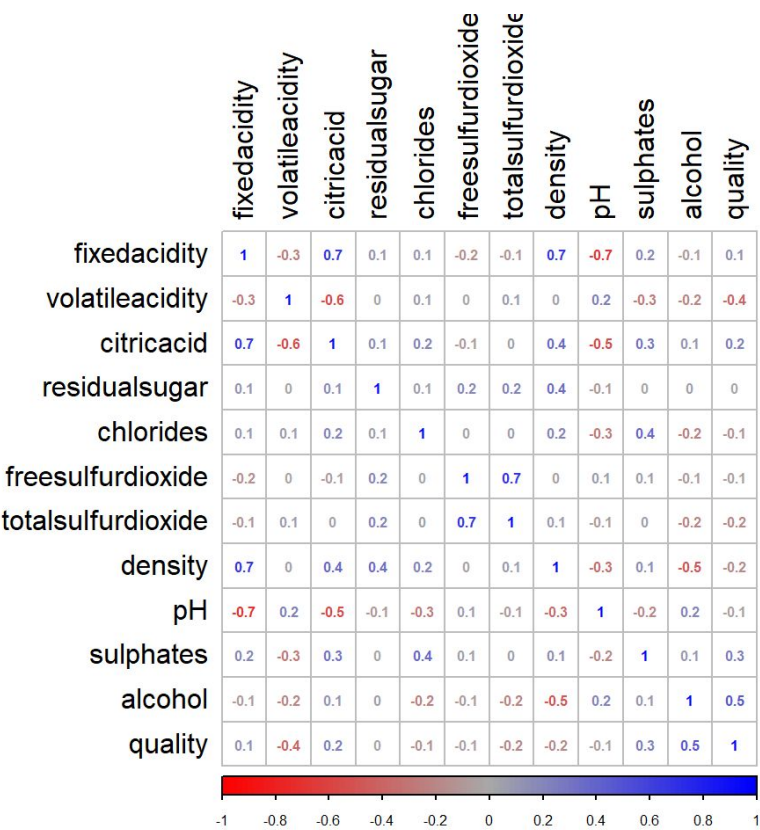
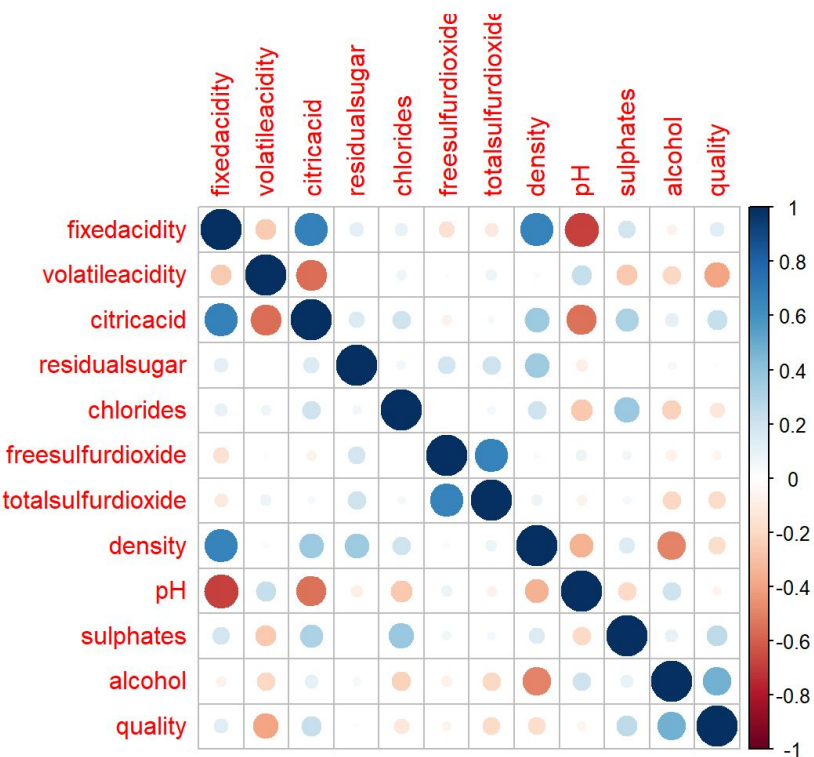
Wine Quality:

Wine Quality Distribution



The dataset contains mostly of average wines. There are not a lot of terrible wines or spectacular wines, so determining an algorithm that will be able to predict wine quality may be difficult, especially for the more likely to occur 5-6 ranking wines.

Correlation Matrices:



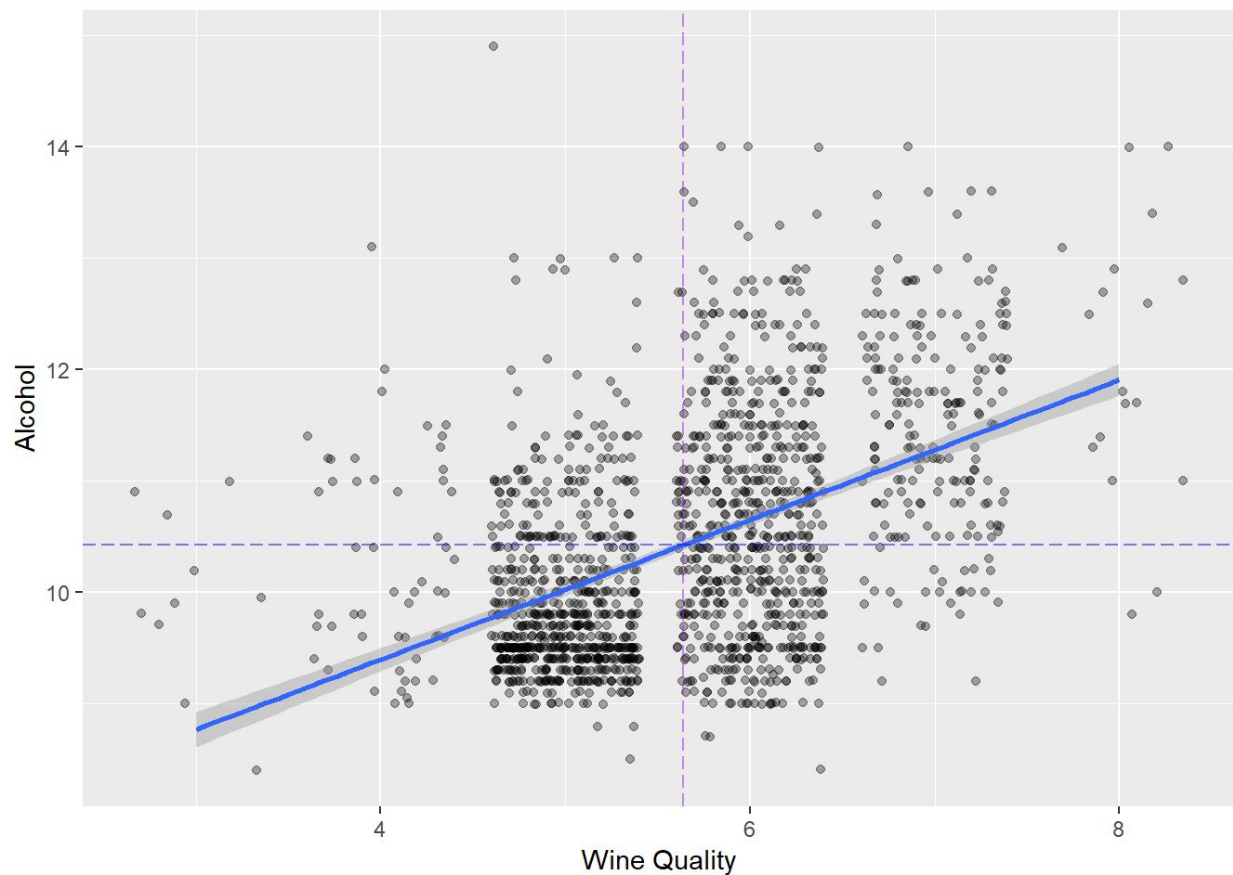
The correlation matrices I created show that at first glance, the variables that correlate most with quality are alcohol, volatile acidity, sulphates, and citric acid. Alcohol, sulphates, and citric acid have a positive correlation while volatile acidity has a negative correlation.

Residual sugar appears to have little effect on the quality of wine, but it does correlate with density which has a negative correlation on wine quality.

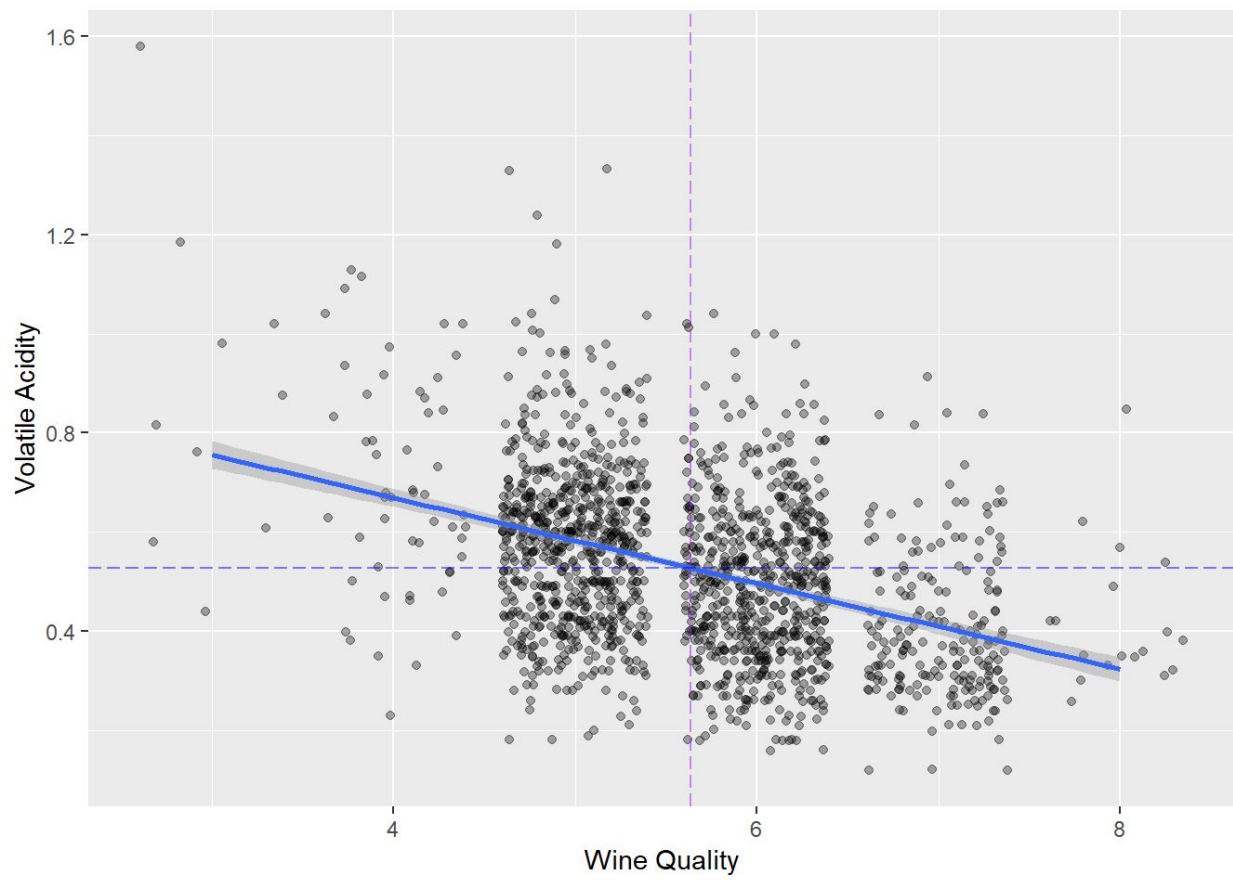
Next, I am going to look more closely at the linear regressions of alcohol, volatile acidity, citric acid, total sulfur, and density in relation to the wine quality.

Simple Linear Regressions:

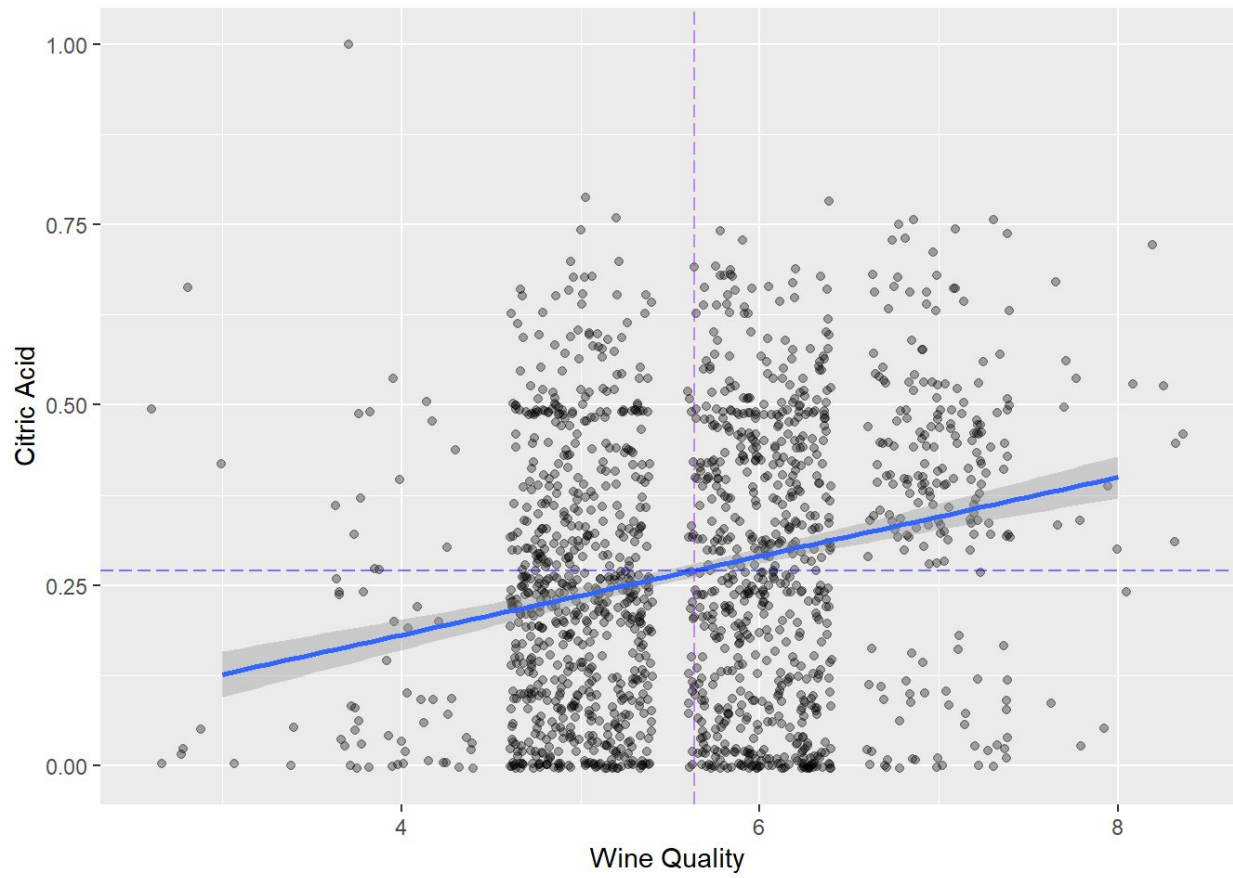
Alcohol Vs. Quality



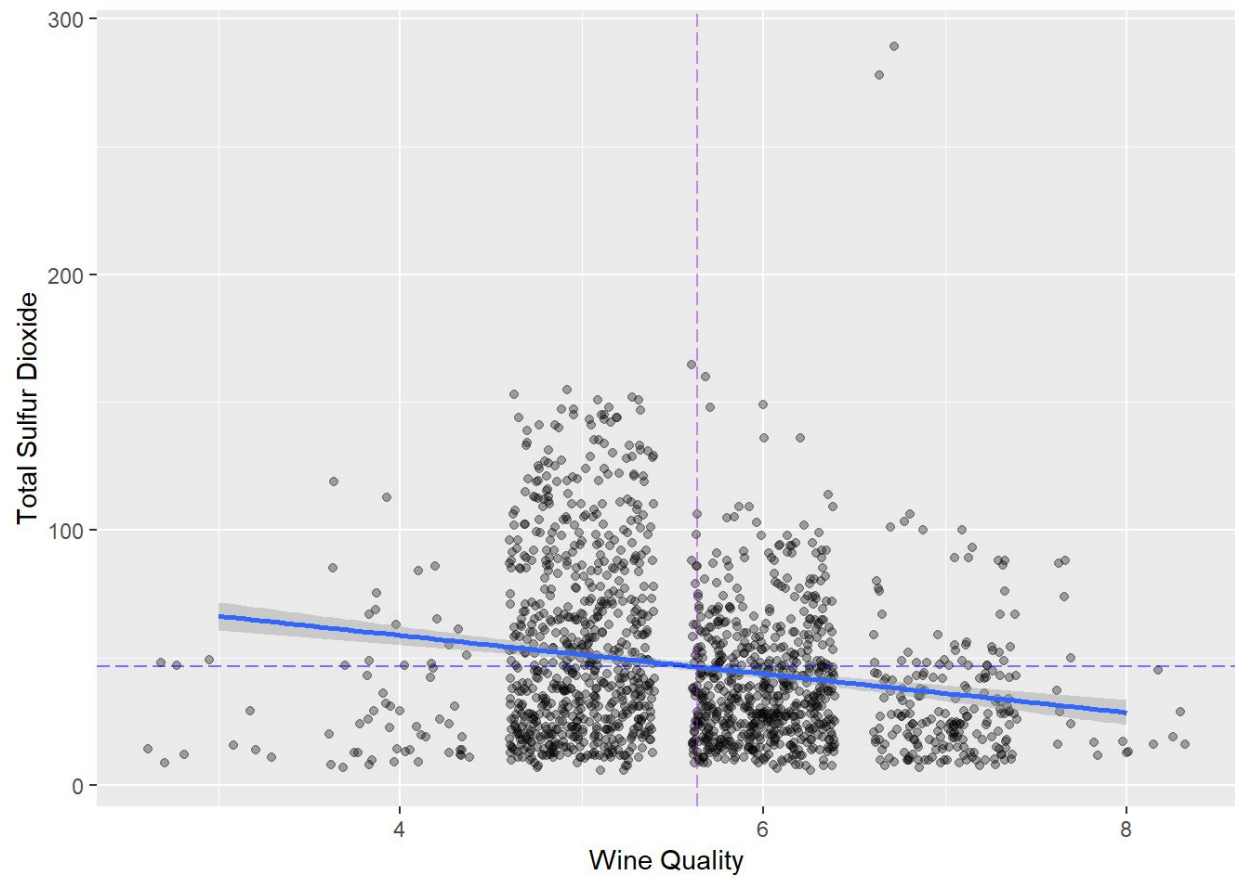
Volatile Acidity Vs. Quality:



Citric Acid Vs. Quality:



Total Sulfur Vs. Quality:



Density Vs. Quality:



