

Wine Quality Exploratory Data Analysis

Jeff Daniels

15 October 2017

Introduction

The Wine Quality for White Wines dataset created by Paulo Cortez (Univ. Minho), et al. will be explored. The dataset contains samples of the white variants of the Portuguese “Vinho Verde” wine. This dataset contains quantitative variables such as acidity and density and the qualitative variable of quality as judged by wine experts. Hopefully the quantitative variables will have some effect on the perceived quality of the wine. If so, it may be possible for consumers to distinguish quality wines before tasting them. For producers, manipulating these variables may create higher quality wine.

Summary of the Data

The ‘wineQualityWhites.csv’ file is loaded into a data frame. The structure of the data frame is 4898 observations of 13 variables.

```
str(wt)
```

```
## 'data.frame': 4898 obs. of 13 variables:
## $ X           : int 1 2 3 4 5 6 7 8 9 10 ...
## $ fixed.acidity : num 7 6.3 8.1 7.2 7.2 8.1 6.2 7 6.3 8.1 ...
## $ volatile.acidity : num 0.27 0.3 0.28 0.23 0.23 0.28 0.32 0.27 0.3 0.22 ...
## $ citric.acid   : num 0.36 0.34 0.4 0.32 0.32 0.4 0.16 0.36 0.34 0.43 ...
## $ residual.sugar: num 20.7 1.6 6.9 8.5 8.5 6.9 7 20.7 1.6 1.5 ...
## $ chlorides     : num 0.045 0.049 0.05 0.058 0.058 0.05 0.045 0.045 0.049 0.044 ...
## $ free.sulfur.dioxide: num 45 14 30 47 47 30 30 45 14 28 ...
## $ total.sulfur.dioxide: num 170 132 97 186 186 97 136 170 132 129 ...
## $ density       : num 1.001 0.994 0.995 0.996 0.996 ...
## $ pH            : num 3 3.3 3.26 3.19 3.19 3.26 3.18 3 3.3 3.22 ...
## $ sulphates     : num 0.45 0.49 0.44 0.4 0.4 0.44 0.47 0.45 0.49 0.45 ...
## $ alcohol        : num 8.8 9.5 10.1 9.9 9.9 10.1 9.6 8.8 9.5 11 ...
## $ quality        : int 6 6 6 6 6 6 6 6 6 6 ...
```

After loading the data, a summary is examined.

```
summary(wt)
```

```

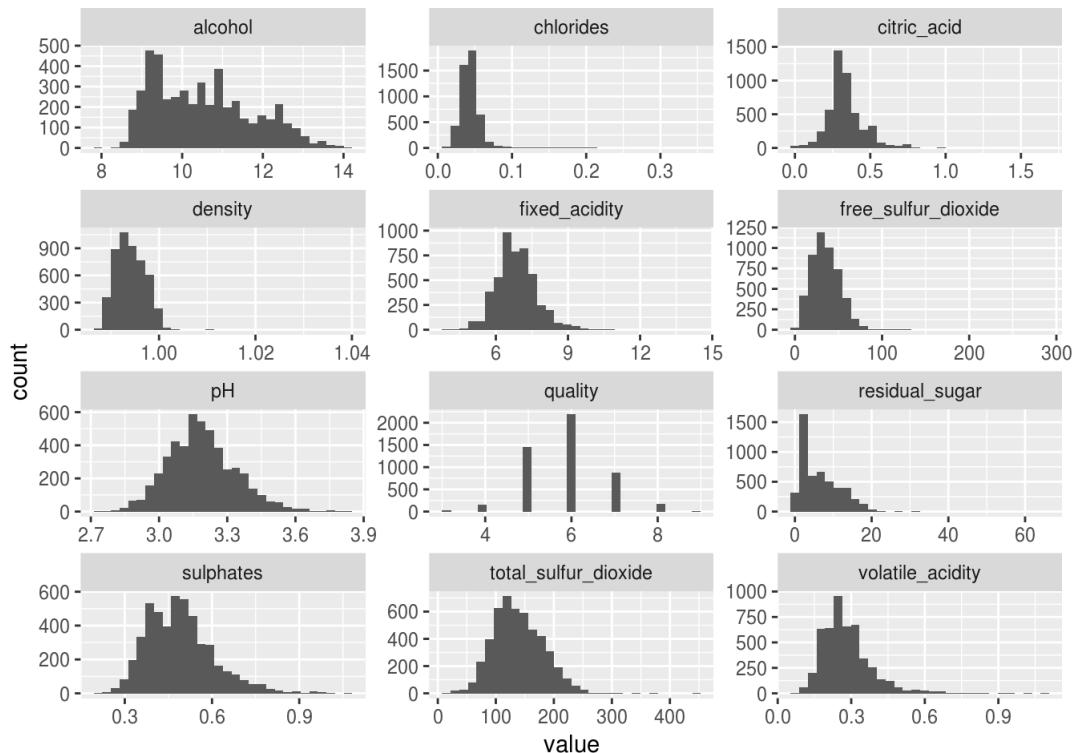
##      X      fixed.acidity volatile.acidity citric.acid
##  Min.   : 1   Min.   : 3.800  Min.   :0.0800  Min.   :0.0000
##  1st Qu.:1225 1st Qu.: 6.300  1st Qu.:0.2100  1st Qu.:0.2700
##  Median :2450  Median : 6.800  Median :0.2600  Median :0.3200
##  Mean   :2450  Mean   : 6.855  Mean   :0.2782  Mean   :0.3342
##  3rd Qu.:3674 3rd Qu.: 7.300  3rd Qu.:0.3200  3rd Qu.:0.3900
##  Max.   :4898  Max.   :14.200  Max.   :1.1000  Max.   :1.6600
##      residual.sugar chlorides     free.sulfur.dioxide
##  Min.   : 0.600  Min.   :0.00900  Min.   : 2.00
##  1st Qu.: 1.700 1st Qu.:0.03600  1st Qu.: 23.00
##  Median : 5.200  Median :0.04300  Median : 34.00
##  Mean   : 6.391  Mean   :0.04577  Mean   : 35.31
##  3rd Qu.: 9.900 3rd Qu.:0.05000  3rd Qu.: 46.00
##  Max.   :65.800  Max.   :0.34600  Max.   :289.00
##      total.sulfur.dioxide density          pH      sulphates
##  Min.   : 9.0      Min.   :0.9871  Min.   :2.720  Min.   :0.2200
##  1st Qu.:108.0    1st Qu.:0.9917  1st Qu.:3.090  1st Qu.:0.4100
##  Median :134.0    Median :0.9937  Median :3.180  Median :0.4700
##  Mean   :138.4    Mean   :0.9940  Mean   :3.188  Mean   :0.4898
##  3rd Qu.:167.0    3rd Qu.:0.9961  3rd Qu.:3.280  3rd Qu.:0.5500
##  Max.   :440.0    Max.   :1.0390  Max.   :3.820  Max.   :1.0800
##      alcohol         quality
##  Min.   : 8.00  Min.   :3.000
##  1st Qu.: 9.50  1st Qu.:5.000
##  Median :10.40  Median :6.000
##  Mean   :10.51  Mean   :5.878
##  3rd Qu.:11.40  3rd Qu.:6.000
##  Max.   :14.20  Max.   :9.000

```

The summary shows that there are no null values in the dataset, it is very clean. Right away we can ignore the X variable because it is just a label for each individual wine. The quantitative variables are all numbers and the quality variable is an integer. The wine is evaluated separately by three experts on a scale of 0-10 and only the median score is presented in the dataset.

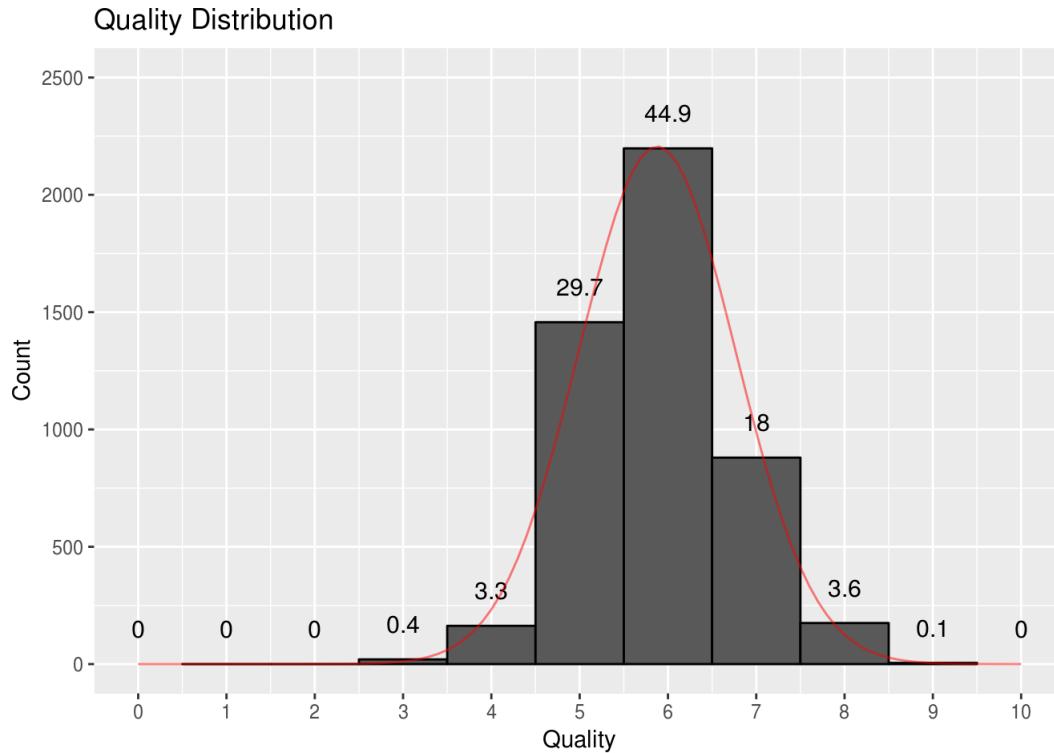
Univariate Plots Section

I shall begin with univariate plots. I'll make a histogram plot for each of the variables before analyzing them further.



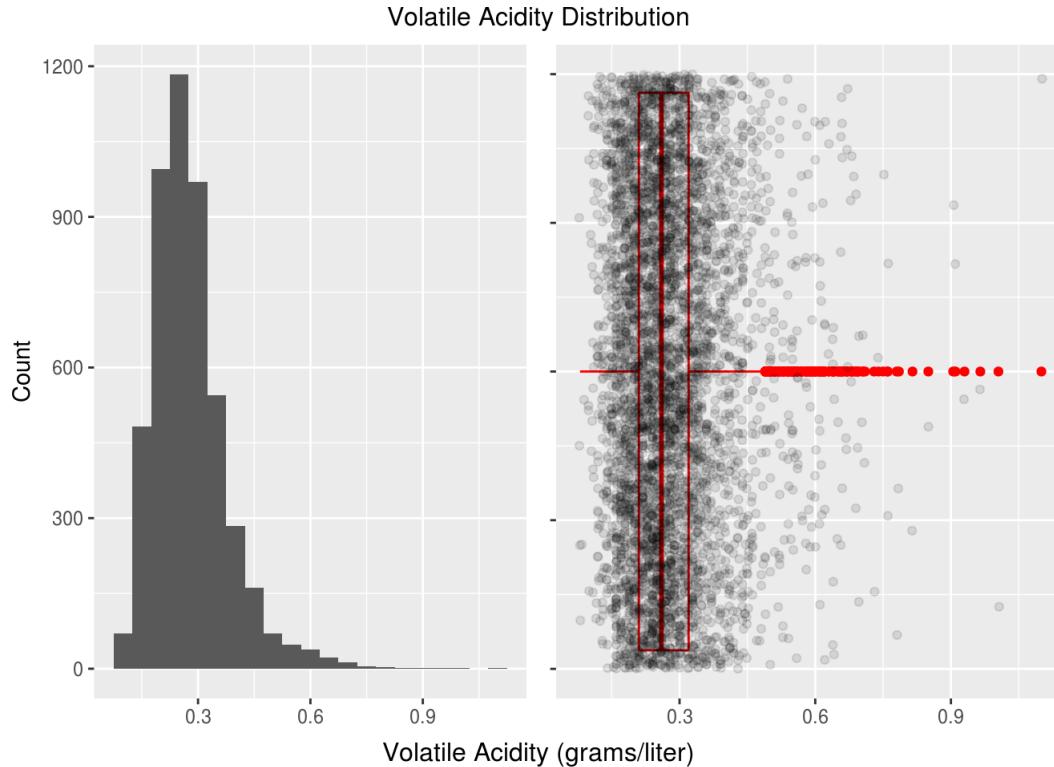
The quality histogram showed a smooth normal distribution of ratings with most wines earning a rating of 6. Very few wines scored a 2 or 9 and none scored a 1 or a 10. There were some exceptional wines, but none that were perfectly good or bad. The average quality for wines is 5.9.

The plot is overlaid with a normal function and the percentage of each quality out of the total sample.

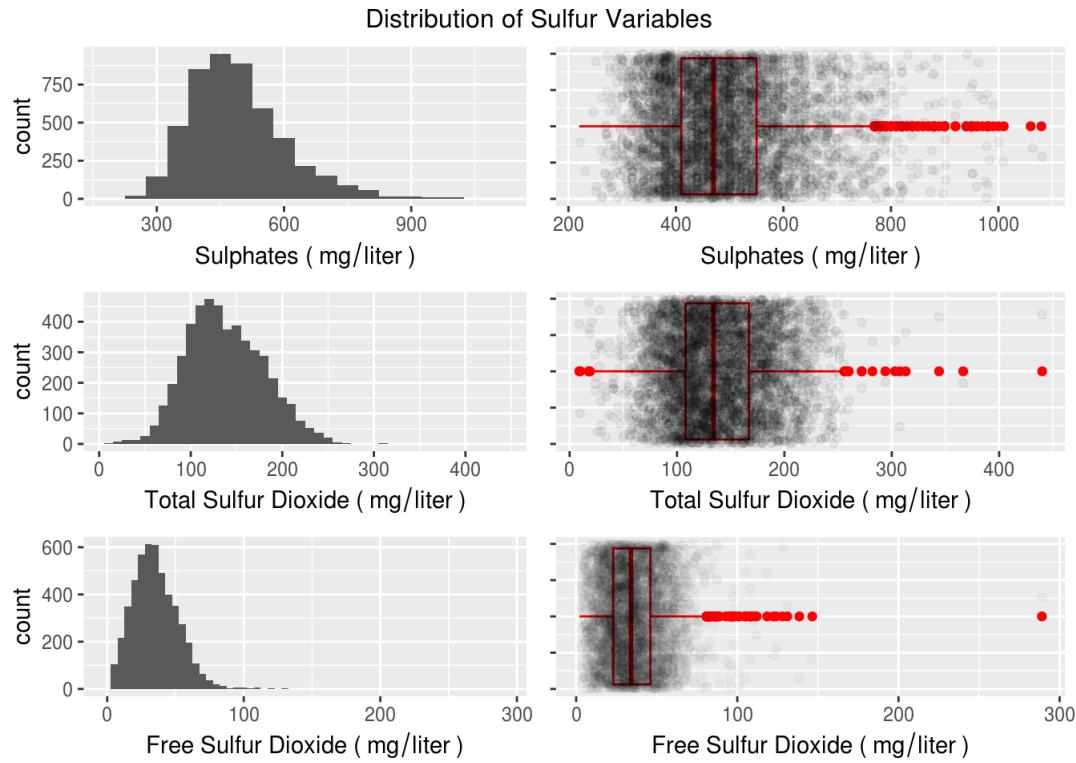


Most of the plots seem to have the normal distributions that the summary implied with similar mean and median values. Some plots are skewed to the right and these values probably represent flawed wines as the result of winemaker error.

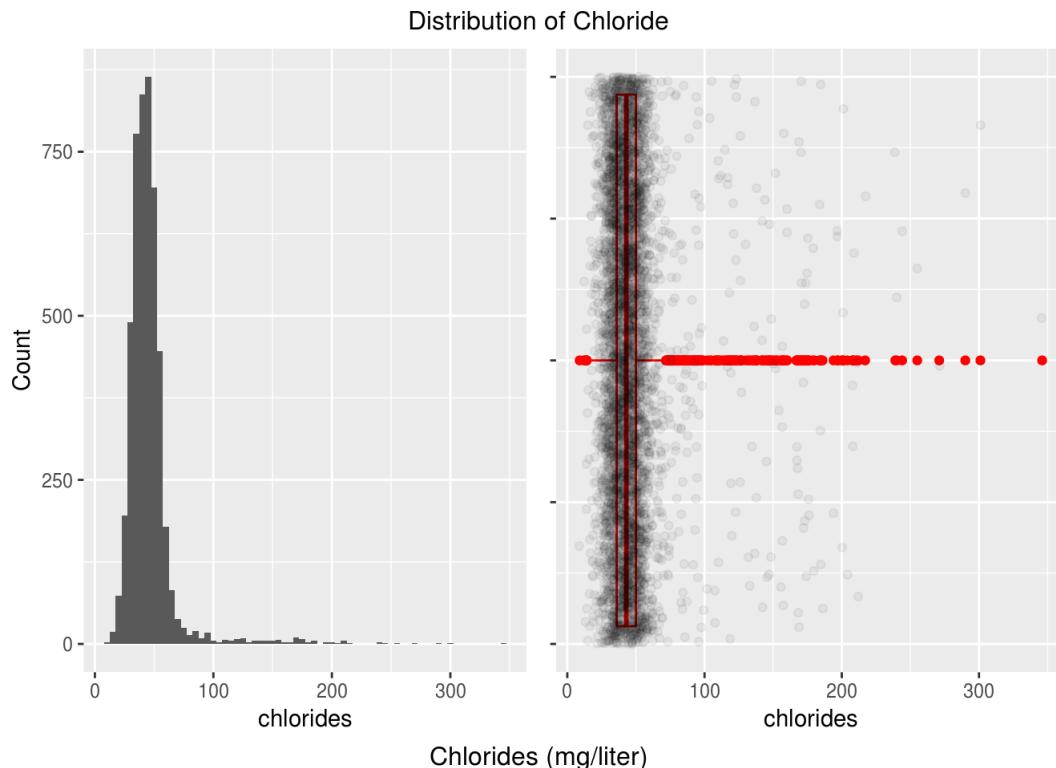
Volatile acidity is flaw so values will try to be reduced. When they are not, the histogram gets skewed. Because a skewed graph like this illustrates flawed and somewhat outlier values for an variable that has negative effects on wine quality, I don't think it would be very useful to scale the x-axis with a log scale to make it appear like there is a normal distribution.



Too many sulfates added could also be a winemaking error if they were accidentally added in too great quantities or are hiding some other flaw. The free sulfur dioxide and total sulfur dioxide should be pretty well correlated with sulfates so it is not surprising that they also have skewed histograms.



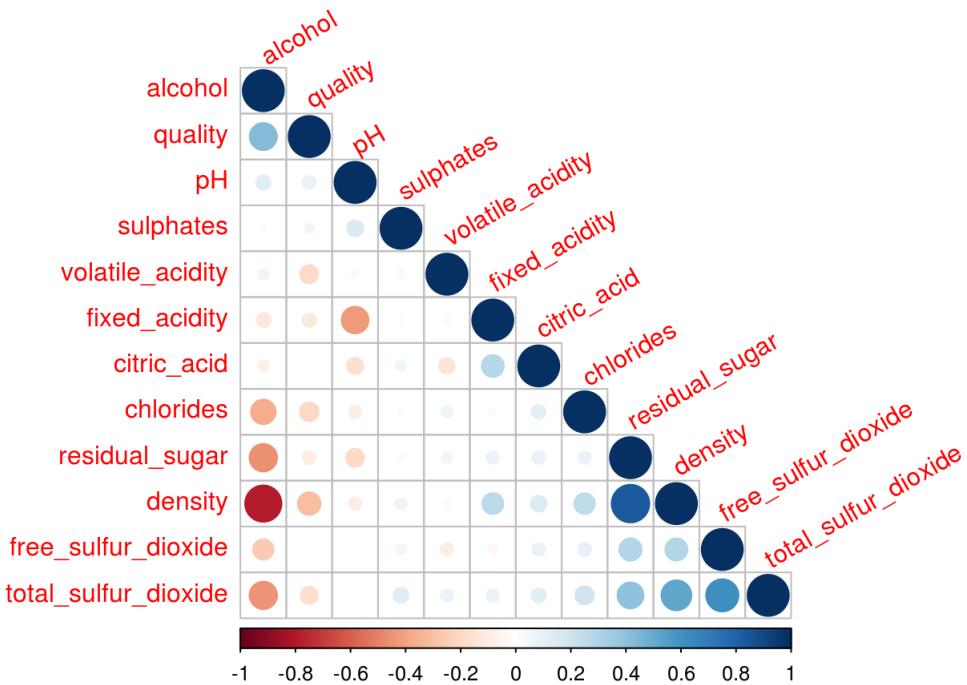
A closer look at the chlorides graph shows a mostly normal distribution with a very long skinny tail. This could mean that chlorides are typically normally distributed but wine makers can intervene and add sodium chlorides if they think it will contribute to flavor. Or perhaps the normal distribution is skewed by wines grown or made near the ocean or on a dried up salt bed. About 97% of the wines fall into a normal distribution.



Correlations

Before plotting bivariate plots the correlations associated with quality are calculated. I'll do this with a correlogram. The correlogram shows positive correlations in blue and negative in red. The size of the circle represents the magnitude of the correlation. The correlogram has been sorted for clarity.

The strongest correlation with quality is with the alcohol variable. It is amusing that wines that had the most potential for intoxication were rated higher. Perhaps raters like the taste of ethanol. I refuse to believe that the amount of alcohol present is solely responsible for how well a wine is received.



The next strongest correlation was a negative and related to density. Statistically this makes sense since there is a strong negative correlation between alcohol and density. From a winemaking perspective this also makes sense because more alcohol in an aqueous solution would make it less dense. Let me briefly give an explanation about how wine is made so that I can introduce new variables that may affect wine quality.

Winemaking Basics

Wine is an alcoholic beverage and is made in roughly the same way all alcohol is made: yeast converts sugar into alcohol.

Sugar \rightarrow Yeast \rightarrow Alcohol

Winemaking converts grape juice, a sugar solution, into a liquid that has water, alcohol and some residual sugar. So we have model of winemaking for our analysis:

Water + Sugar \rightarrow Yeast \rightarrow Water + Alcohol + Residual Sugar

Because the densities of all these components are different and the proportions are changed by the yeast, the density of grape juice will be different from wine.

Fructose 1.69 g/cm³

Water 1.00 g/cm³

Ethanol 0.79 g/cm³

So as the fructose sugars are converted into ethanol, our aqueous solution becomes less dense. The density of grape juice will be referred to as Original Gravity (g/cm³) and the grape juice has a quantity of sugar that I will call Starting Sugar. After conversion by the yeast, the grape juice becomes wine that has a density that corresponds to the Density in our dataset which from here on I will refer to as Final Gravity and a quantity of sugar that corresponds to Residual Sugar in our dataset.

The following equations were used for calculating Original Gravity and Starting Sugar:

$$FinalGravity = Density$$

$$OriginalGravity = \frac{7.36}{1000} * Alcohol + FinalGravity$$

$$Brix = (((182.4601 * OriginalGravity - 775.6821) * OriginalGravity + 1262.7794) * OriginalGravity - 669.5622)$$

$$StartingSugar = Brix * 10$$

The efficiency of sugar conversion by the yeast is called attenuation and is calculated with this equation:

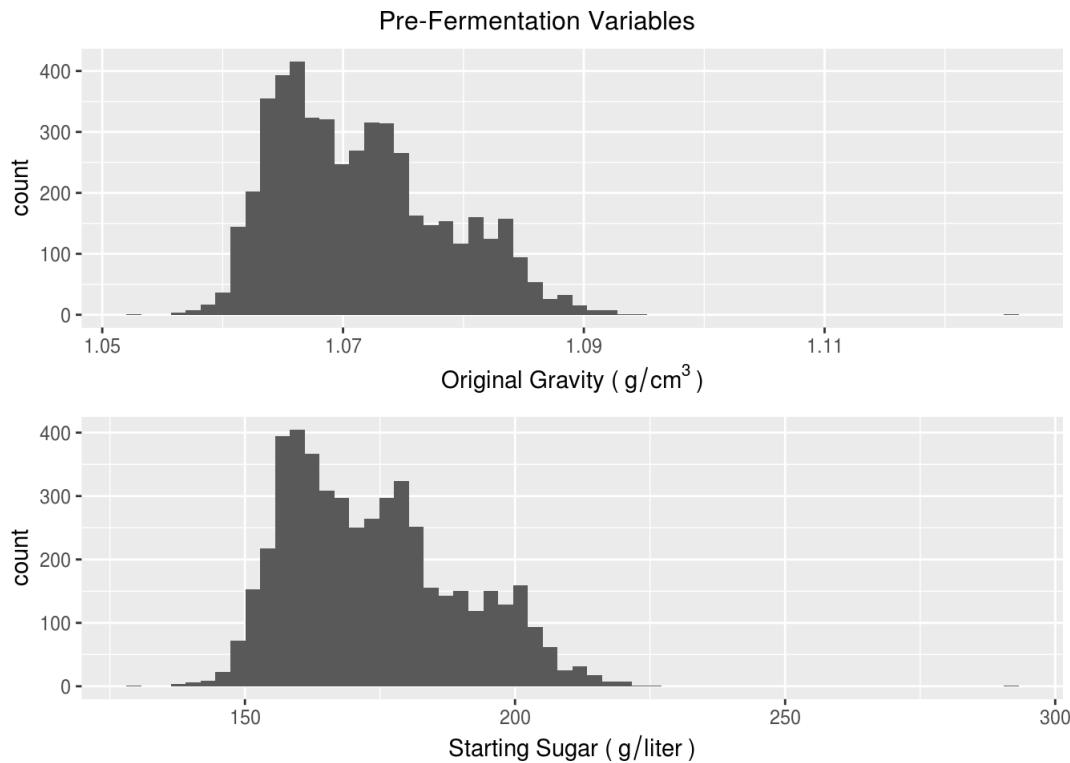
$$Attenuation = \frac{StartingSugar - ResidualSugar}{StartingSugar}$$

It must be noted that most of these equations are approximations of values that must be tested in a lab to be accurate. They should work well enough as statistical transformations for our analysis. Most importantly, they will give us an estimate of what the wine was like before fermentation. Because the dataset only contains variables for finished wines, I think these new variables will yield interesting insight.

Let us take a look at our new variables. The original gravity and starting sugar should look pretty similar if we think of grape juice as mostly water and sugar. These variables can give us clues about what the grape juice looked like before fermentation and comparing their corresponding variables, final gravity and residual sugar, we can get a sense of what happened during fermentation.

The original gravity of the wines doesn't cleanly fit a normal distribution which is what I was expecting. I assumed that because original gravity for a single wine is from the contributions of many many grapes, the large sample of grapes would have characteristics that fit a normal distribution. The histograms suggest a few peaks and unpredictable variation in original gravities. Perhaps this reflects some variation caused by growing regions, grape varieties or even human intervention.

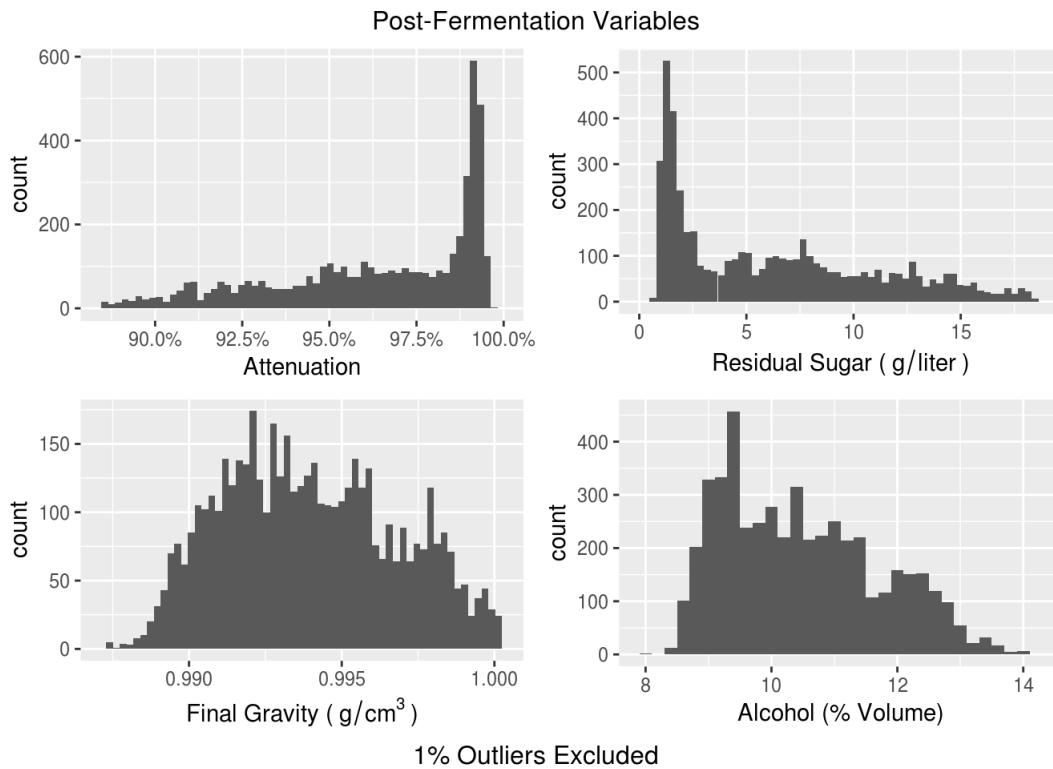
The outlier in these plots represent one observation of a wine that has the highest maximum value of residual sugar in the sample. It is moderately high in fixed and volatile acidity and alcohol, and sulfates. Though it seems an aberration, it did receive a rating of 6 so it must be an acceptable example of Vinho Verde. Recall that our starting sugar variable is a function of residual sugar and alcohol. An artificially high and unnatural starting sugar could be erroneously estimated if there is sugar or alcohol added post fermentation. In this example it seems that sugar was added.



Plotting attenuation reveals that wines have a tendency towards full attenuation. Most wines seem to be well attenuated with many values close to 100% attenuation. This variable may not take into account any added sugar. This plot should almost be a mirror image of residual sugar. Both plots are shown below with 1% outliers removed.

The residual sugar histogram is very skewed to the right because there are a great number of wines that have almost no residual sugar. As the residual sugar increases, the counts decrease. There seem to be some outlier values, but 99% of wines finished with residual sugar below 18.8 g/liter.

The skewed values of alcohol might be because of a natural limit to how much alcohol a wine can have. As wine becomes more alcoholic the yeast activity is inhibited and a ceiling for alcohol values is approached. The amount of alcohol in a wine seems to have a greater effect on the final gravity than the residual sugar as shown by their similarly shaped histograms. There is simply more alcohol than sugar.



Univariate Analysis

Most of the plots displayed normal distribution with outliers displaying large values which probably correspond to undesirable wine qualities. Mostly, these histograms give us a nice representation of what a Vinho Verde wine is supposed to look like in terms of measurable variables. This isn't very useful if we don't have other styles of wine to compare the character of Vinho Verde to.

Quality is definitely the most interesting variable in the dataset because it is considered an output variable and all the others are inputs. Before making bivariate plots I looked at the correlations with quality and it was overwhelmingly related to the level of alcohol. This incited me to look at what could affect alcohol levels the most and create new variables related to alcohol production.

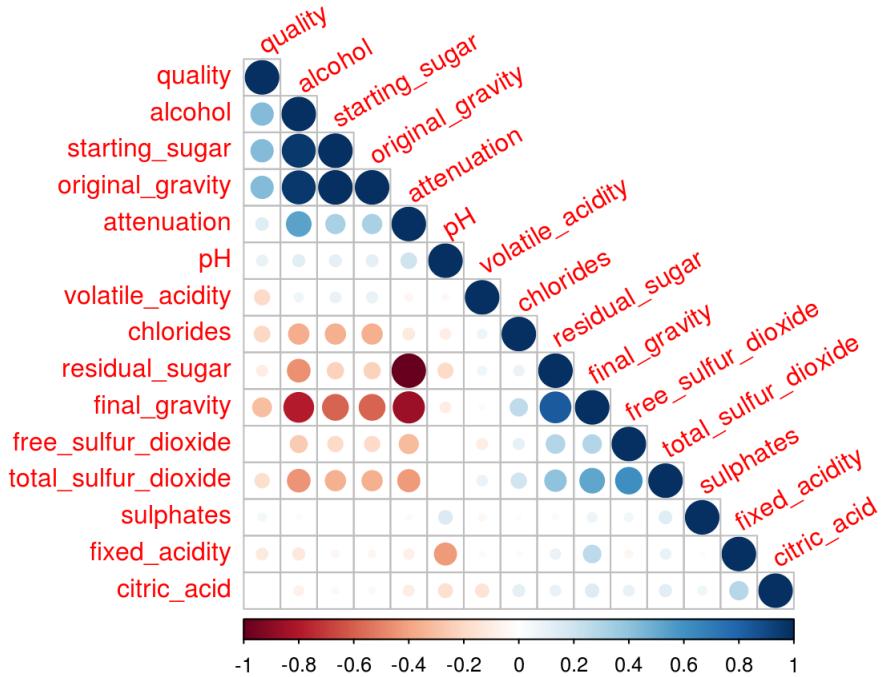
An estimate for the amount of Starting Sugar and Original Gravity was calculated from the alcohol and density variables. The variable density was renamed to Final Gravity for clarity. Attenuation was calculated from Starting Sugar and Residual Sugar values.

In addition to adding new variables, I deleted the id variable because each row already has an id in the data frame and this variable isn't useful if I don't know which commercial wines correspond to each id. I also converted the units of the sulphates variable so that they were the same as free and total sulfur dioxide.

I would like to know if quality wines have variables that ascend, descend, or converge on a value. For example, as alcohol levels go up, does quality also go up so that an alcoholic wine is considered ideal. I would assume that as volatile acidity descends, quality scores go up. Maybe there is a fixed amount of residual sugar or acidity that is expected so that the highest quality wines have values close to the median for certain variables. Since I am interested in what variables affect quality, I must proceed with a bivariate analysis.

Correlations with Quality Revisited

Let us revisit the correlogram plot with our new variables.



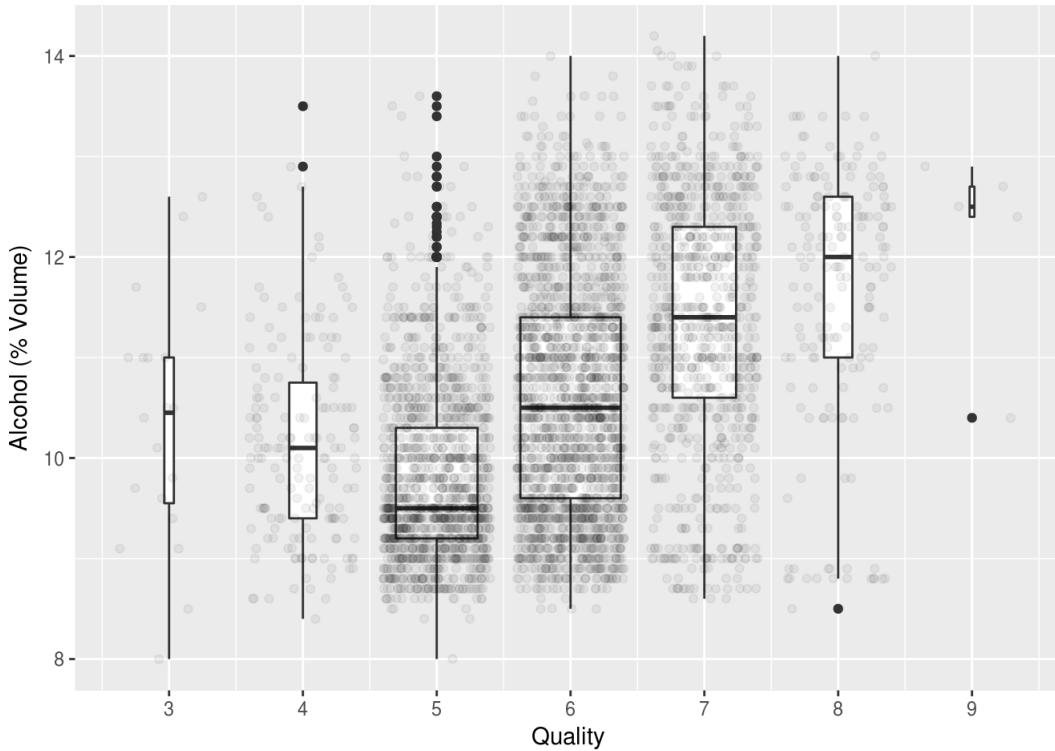
Starting sugar and original gravity both have a similar positive correlation with quality that alcohol does, around 0.44. This makes sense because the more sugar you start out with, the more alcohol you are likely to end up with. We can also observe a negative correlation between final gravity and quality of -.31. This is interesting because it is a much stronger correlation than the one residual sugar has with quality, -0.10. One would think that because they are somewhat related that their quality correlations would be closer.

The -0.21 chloride and quality correlation is interesting because it is about as strong as the -0.19 correlation between volatile acid and quality. Volatile acid is explicitly a negative characteristic of wine. Before I did not know if chlorides negatively affected quality. I just thought if you added salt to something, it would taste better. Clearly too much salt is bad for quality. Chlorides seem to positively correlate with final gravity and total sulfur dioxide.

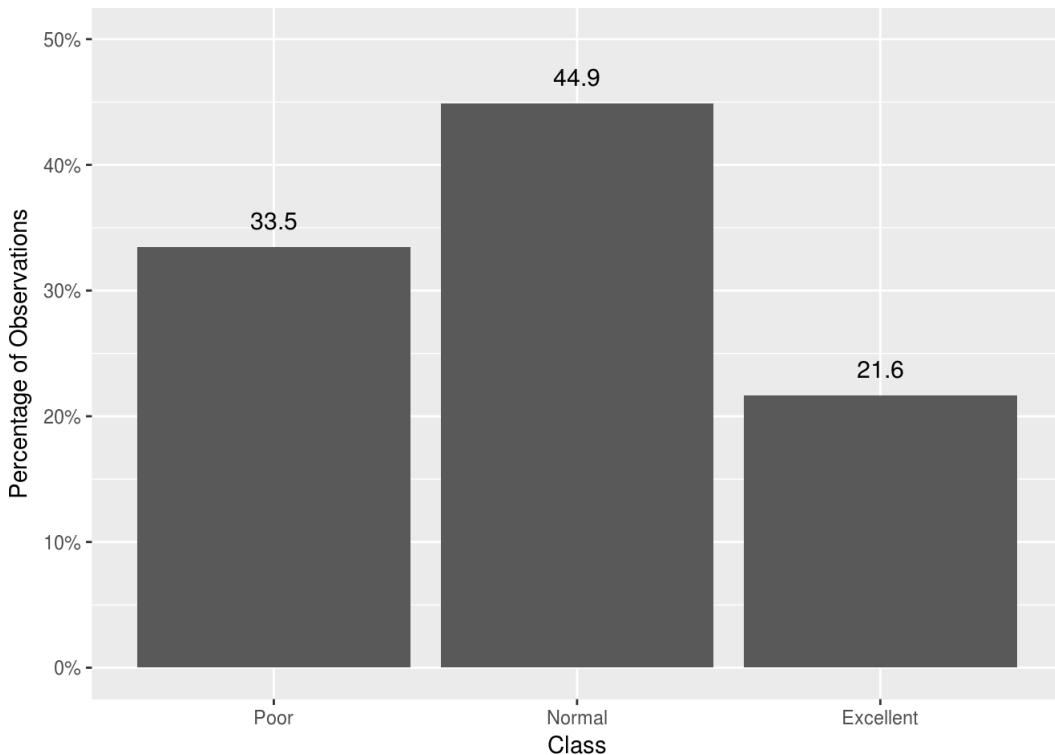
Total sulfur dioxide has a -0.17 correlation with quality which also seems significant. Total sulfur dioxide has a 0.53 correlation with final gravity which leads me to believe that more sulfur is added to wines if they end up with too little alcohol and more residual sugar which could increase the likelihood of spoilage. Predictably, total sulfur dioxide has a strong positive correlation with free sulfur dioxide but surprisingly not so much with sulfates. Other variables that I thought would prevent spoilage, fixed acidity and pH, don't seem to have strong correlations with total sulfur dioxide. There is a 0.20 correlation with chlorides. I don't know how sulfur dioxides and chlorides are related, but they are probably both common in bad wines.

Bivariate Plots

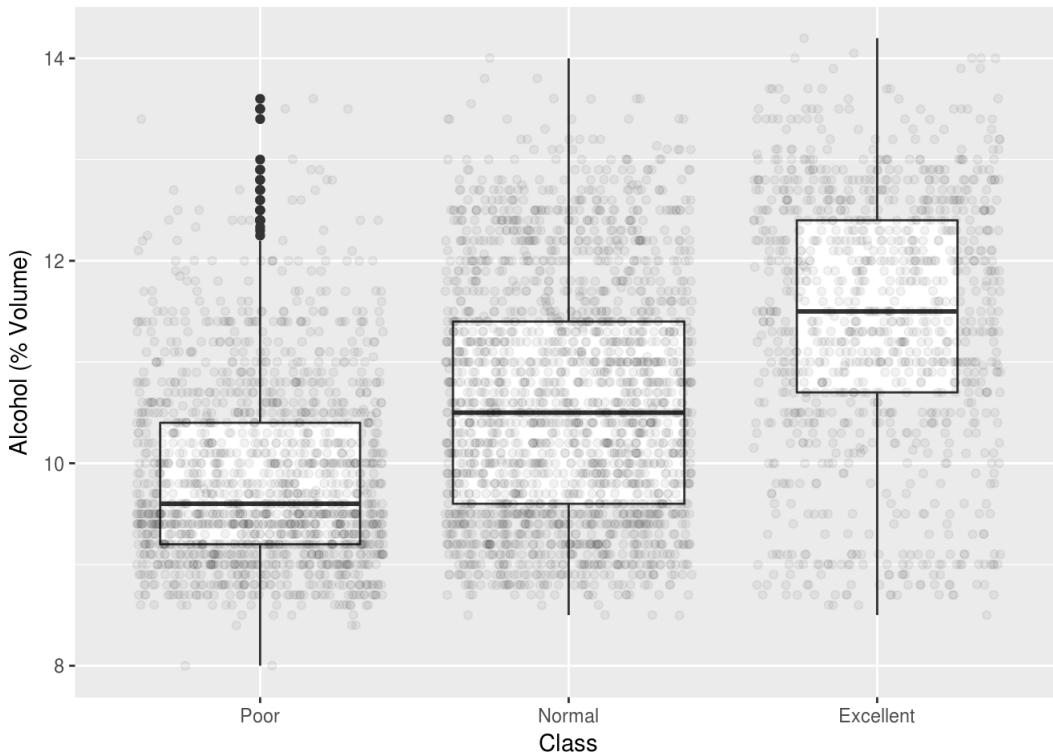
The first bivariate plot I will examine is Quality vs. Alcohol. Given a level of alcohol, what kind of quality score should we expect. A box plot gives a clear picture of the relationship between alcohol and quality. As quality increases, the alcohol content increases. If the box plot width is not scaled with the number of observations in each group, then it might appear as if there was less variation in alcohol values for the highest quality wines. Now might be a good time to group the observations with ranges of quality values so that I don't over generalize about quality based on a small number of observations in the best or worst quality wines.



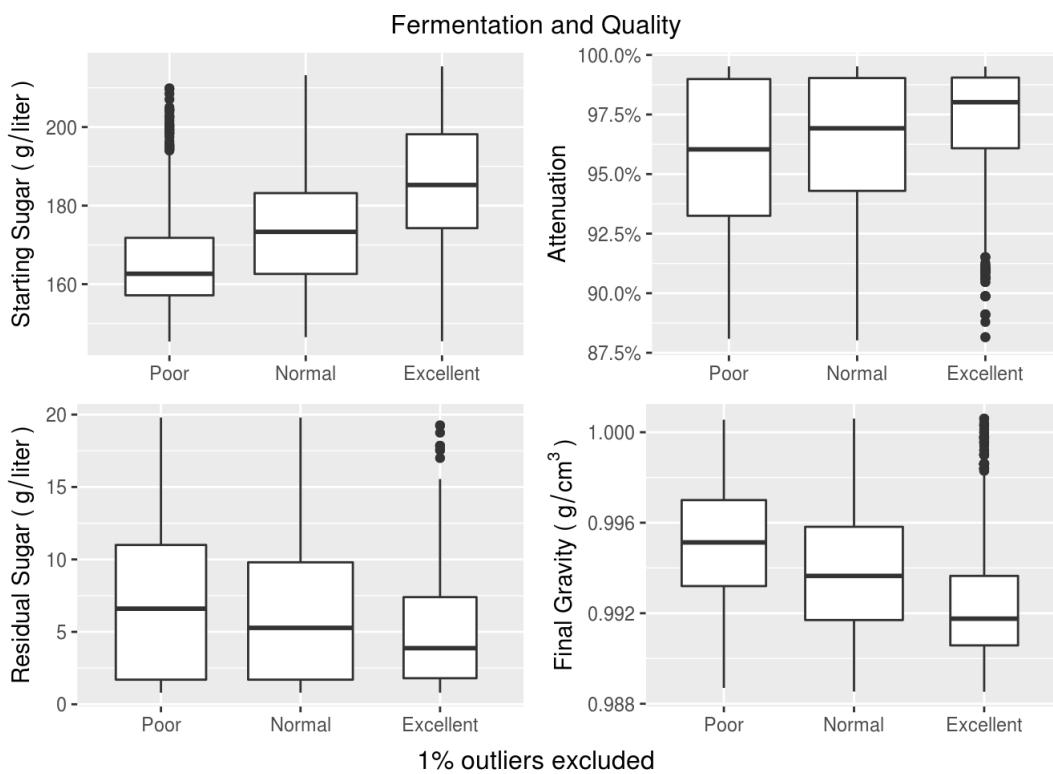
Three classes shall be made from the quality variable: Poor, Normal, Excellent. Wines with quality scores less than 6 are poor, 6 are normal, and above 6 are excellent. Hopefully this will make trends clearer.



The relationship between alcohol and quality is clearer now that the wines are grouped in classes. Alcohol increases as quality increases. The variations between classes is not that different because the sample groups are larger.

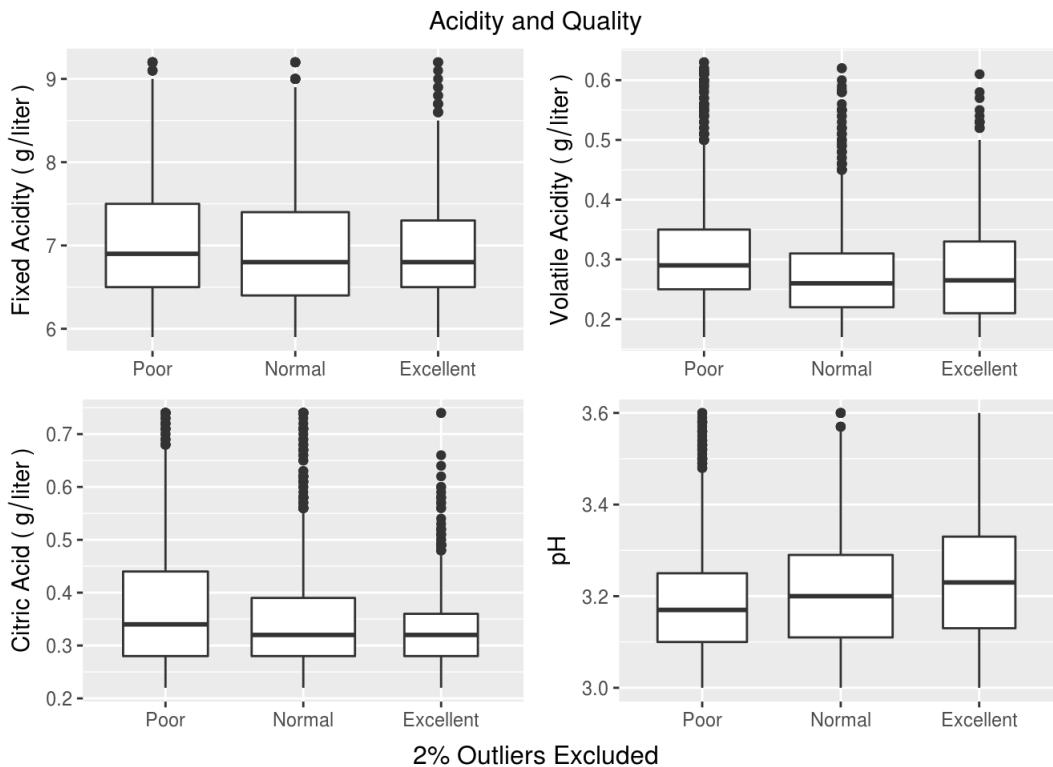


Alcohol is largely influenced by how much sugar there was initially and the degree of attenuation. These plots show how alcohol level increases as quality increases. Alcohol levels in excellent wines are the result of more starting sugar and higher attenuation.

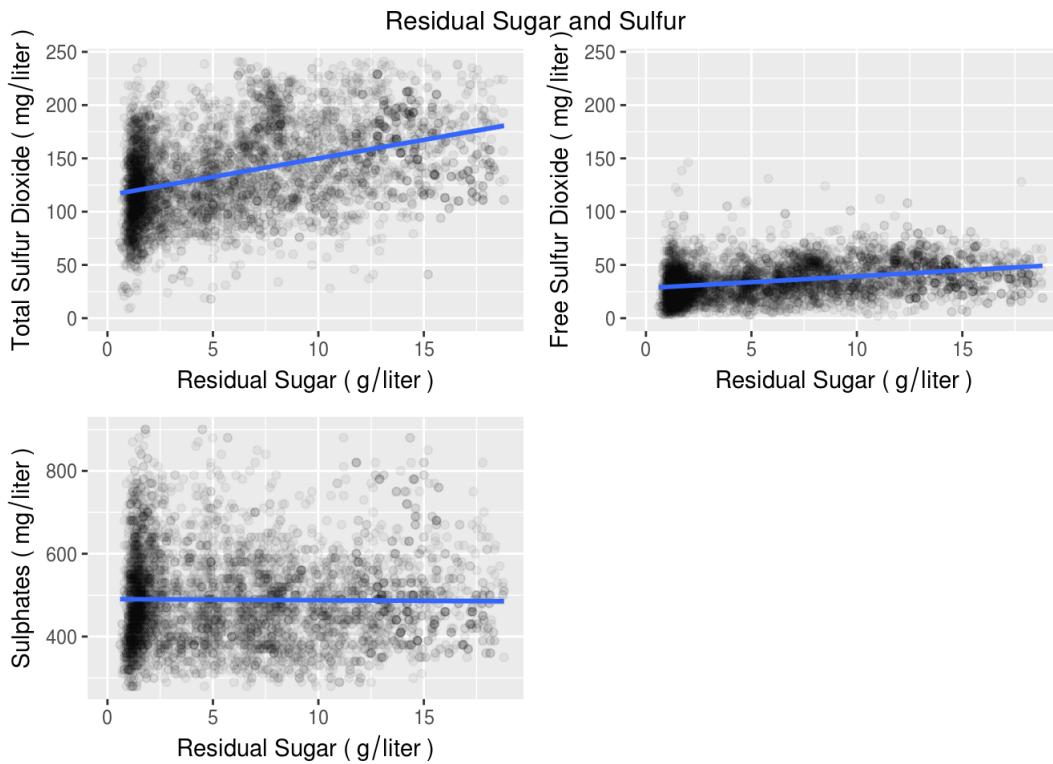


We know that one fermentation input, starting sugar can be related to quality. The other starting variable I would like to examine is acidity. I know that pH is supposed to decrease during fermentation, but I think acid levels stay relatively constant. Is there a relationship between acidity of the grapes and the quality of the finished wine?

The three types of acidity and pH are compared below. The differences are not that great but subtle trends can be made out if we zoom in and ignore the outliers. Poor wines have more volatile acidity. As quality increases, the amount of citric acid decreases slightly. There is a corresponding increase of pH when quality increases that describes excellent wines as less acidic.

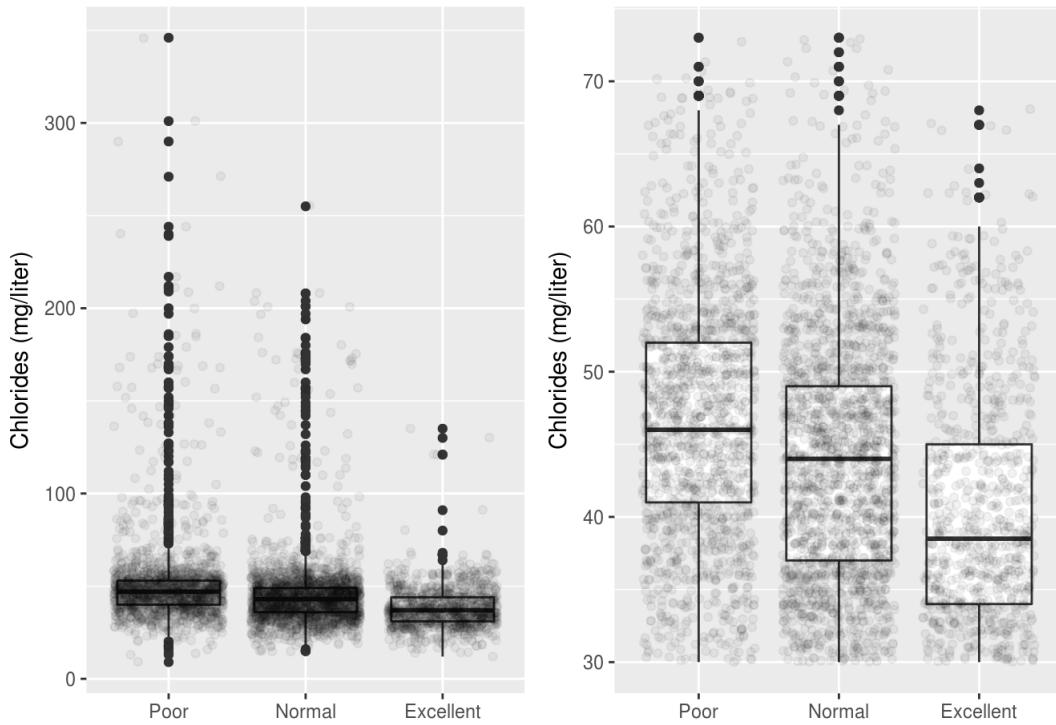


There seems to be a significant correlation between total sulfur dioxide and residual sugar. I assume this is because you add more sulfates to wine to prevent spoilage from having too much residual sugar. It is interesting to me that the free sulfur dioxide doesn't seem to increase as much for a given residual sugar. Also, the curve for sulphates is flat. I should probably research winemaking to understand its relationship to sulfur dioxide.



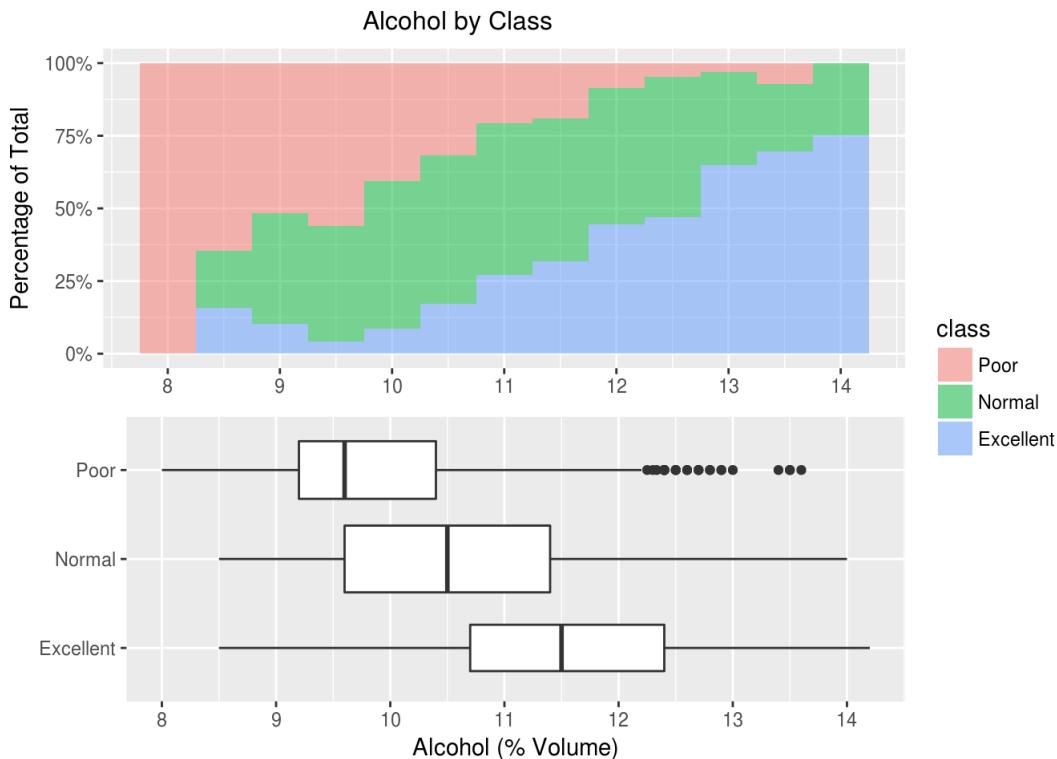
There is a negative correlation between chlorides and quality that is hard to see through all the outliers. Outliers are ignored by filtering out 5% of the sample so that chlorides and quality can be examined. There does seem to be negative correlation with quality but the quantities are actually very small and might be below the taste threshold. The fact that there are many more outliers in poor wines with both large and small values is worth mentioning.

Chlorides and Quality



A histogram plot with all the bin heights equal to 1 reveals, for a given level of alcohol, what portion are in each class. What we see is that more alcohol a wine has, the more likely it is to be judged excellent. If a wine is low in alcohol it is more likely to be poor. Normal wines seem to be uniformly present at most alcohol levels. In a sense, if we know the level of alcohol, the probability of belonging to a certain class can be estimated.

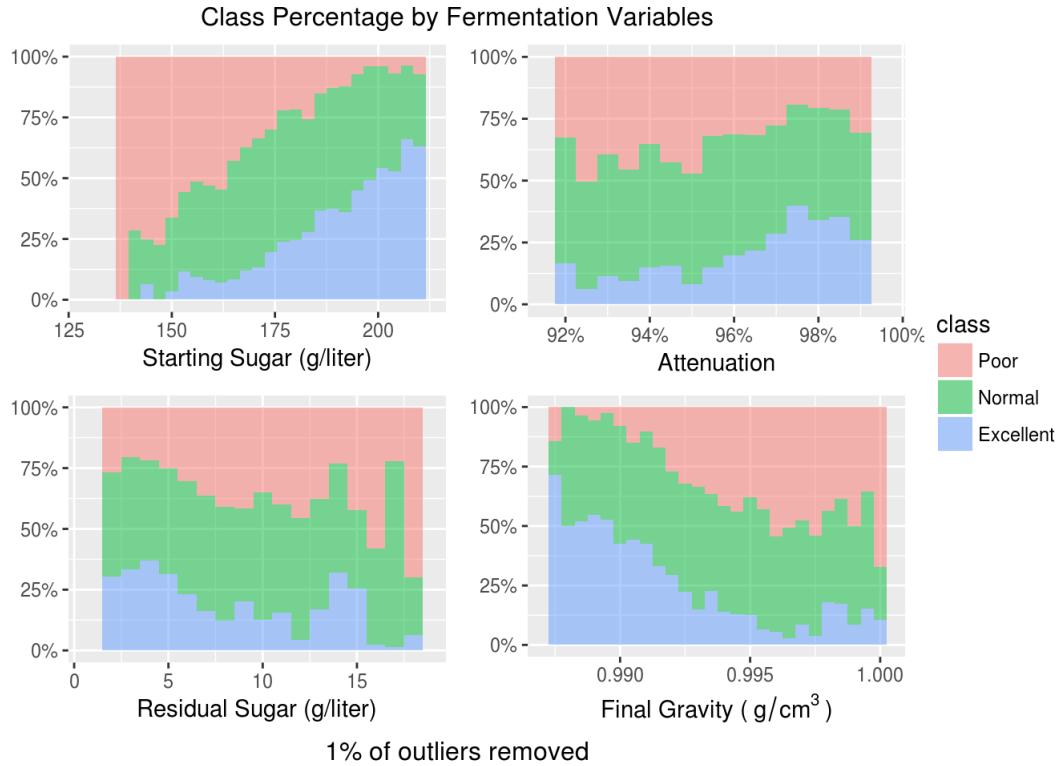
A boxplot can show the probability of a certain class having a given alcohol content, but this histogram shows the probability of a class for a given value. Even if we encounter a wine with an alcohol content near the median value of an excellent wine, it is still only has about a 30% chance to be an excellent wine.



Class percentages for a given variable value seem to illustrate the data well. Let us compare the class percentages for other variables related to fermentation.

Like the in alcohol plot, normal wines seem to be uniformly present at all variable values. The poor and excellent wines occur in greater proportions at opposite ends of the variable ranges. Wines with higher starting sugars are more likely to be excellent. A higher residual sugar makes a wine more likely to be poor. Wines with low attenuation were unlikely to be excellent. Wines with lower final gravities are more likely to be excellent.

All these observations illustrate what we already discovered. Excellent wines start out with more sugar and ferment more completely to yield wines with more alcohol.

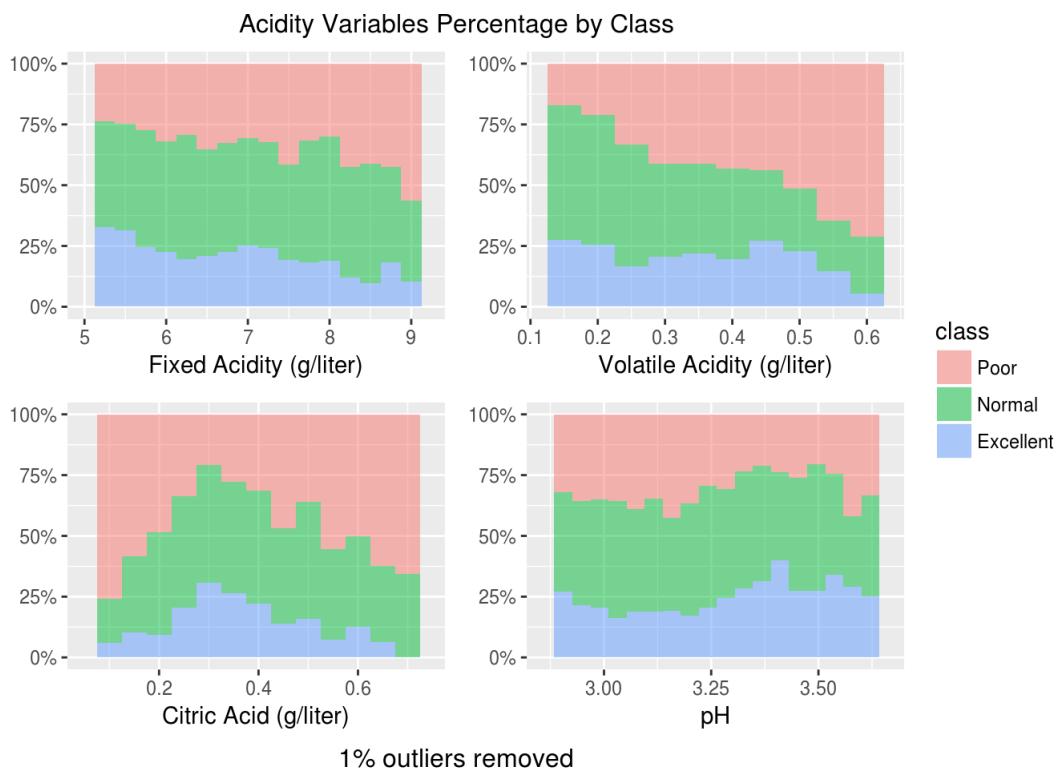


1% of outliers removed

Is there a relationship between acidity and the likelihood of a wine being excellent and poor? As volatile acidity increases, the portion of poor wine increases. Because volatile acidity is flaw, this makes sense. Interestingly, as fixed acidity increases, the probability of a wine being poor also increases. Low fixed acidity has a small increase in proportion of excellent wines.

There is an interesting peak at 0.3 g/liter of citric acid where the probability of a wine being excellent peaks. This is the first such curve I have found in the data set. Quality is usually correlated linearly with ideal for a variable being in the upper or lower range of values. For citric acid, the upper and lower ranges of values make wine more likely to be poor.

There may be a similar local maximum for excellent wines in the pH curve at about 3.4. This relationship doesn't seem to be as strong as the others, but it could suggest that more acidic wines score lower on quality.



1% outliers removed

Bivariate Analysis

A new variable, class, was created to make analysis of quality easier. This illustrated a strong relationship between alcohol levels and quality. By examining original gravity and attenuation, both factors in the alcohol level, we found that quality wines generally had both high original gravities and attenuation.

Acidity levels don't seem to affect quality much. This surprised me since I thought acidity was a major flavor component in wine. I will study interactions of acidity with other variables in the multivariate section. Are there other characteristics that balance out acidity, like residual sugar or alcohol that can make a wine score higher in quality?

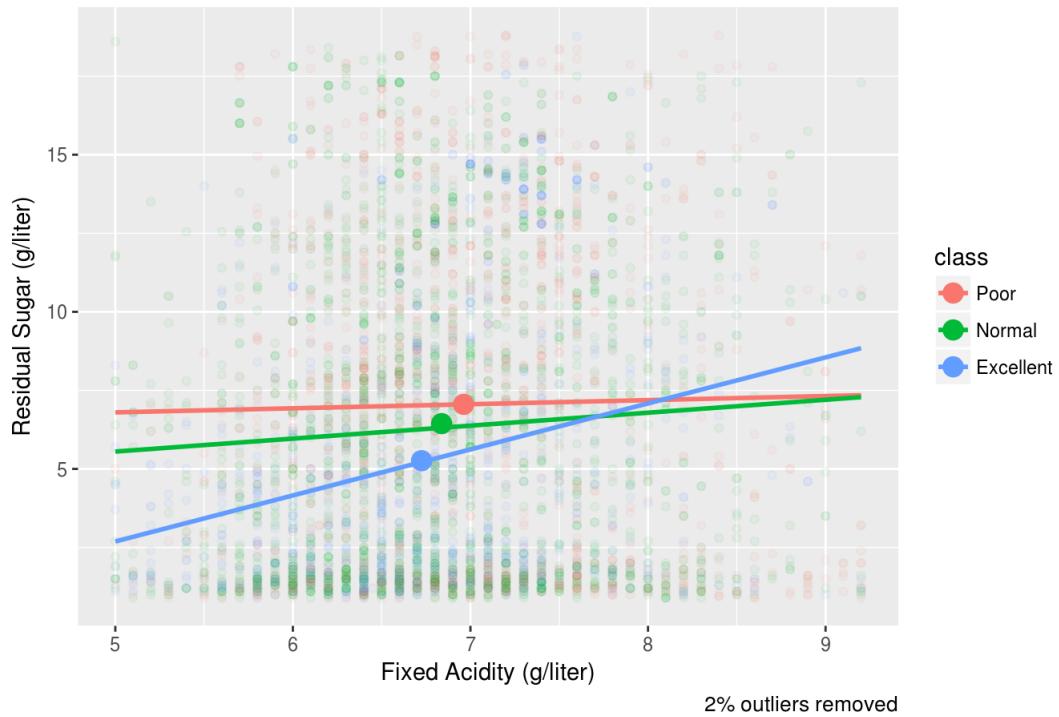
I don't really understand how sulfurs interact with wine and how they are used. This was shown when I tried to find a relationship between sulphates and residual sugar and there was none. I don't think multivariate analysis will illuminate this for me. I must do research outside of this dataset.

Histogram plots with bins adding up to 1 opened up a possibility of predicting the likelihood of a wine being in a certain quality class given a variable's value. Alcohol content still seemed to correlate with quality. At very high alcohol levels, wines were usually excellent. But I was also able to show that excellent wines could share the same alcohol level with poor wines.

Multivariate Plots

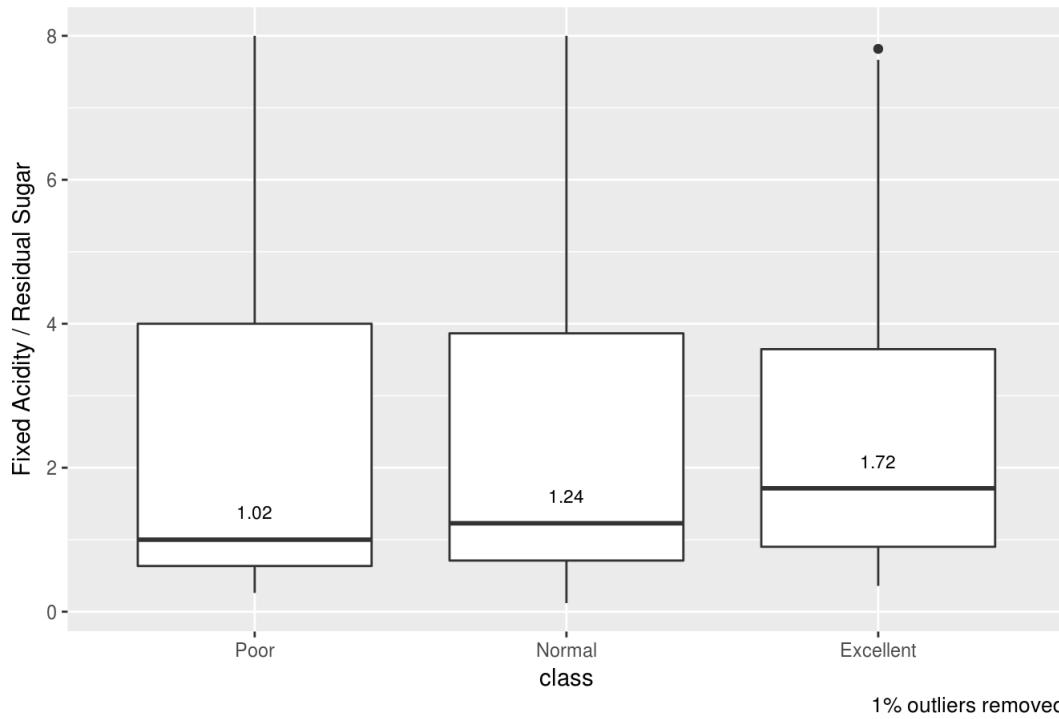
Still trying to relate acidity to quality somehow, I plotted residual sugar versus fixed acidity. The mean values for each class are shown with the larger circles. Excellent wines have less residual sugar and poor wines have more. We know that less residual sugar is from higher attenuation and results in higher alcohol content which is positively correlated to quality. This plot may not be telling us much. The mean values of fixed acidity are lower for excellent wines which may represent something useful.

Residual Sugar vs. Fixed Acidity

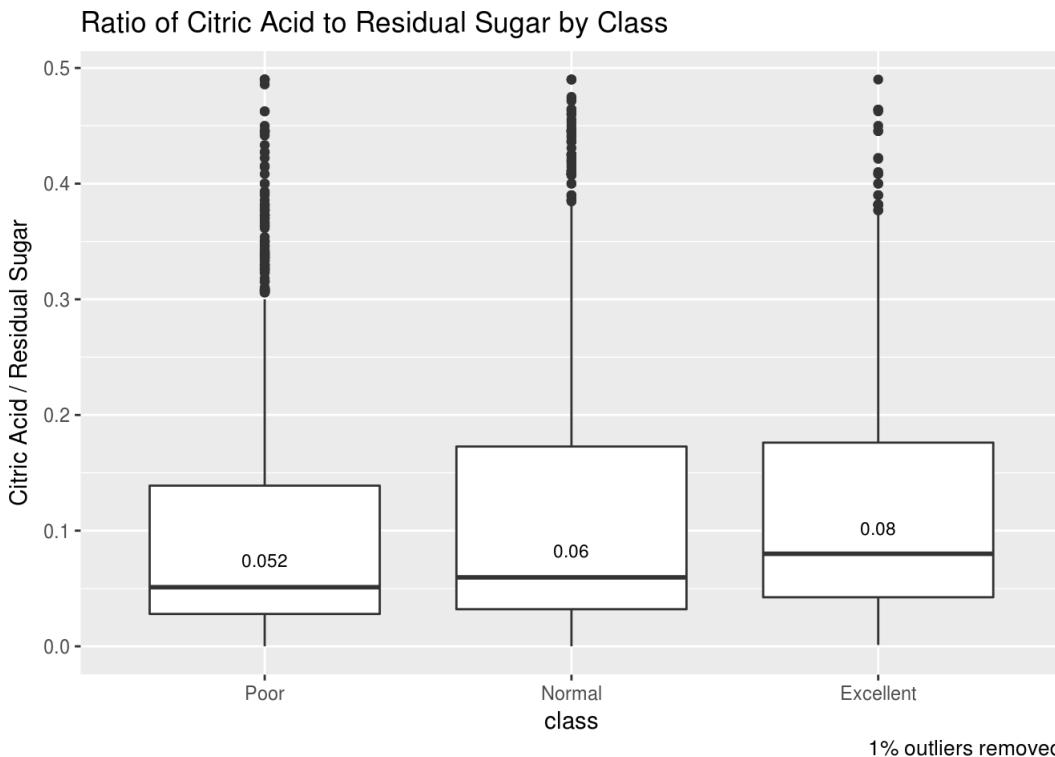


Plotting the ratio of fixed acidity and residual sugar tells us a little more about the balance of sugars and acids. The amount of acid in wine typically varies between 1 to 4 times as much as the residual sugar. This seems like a level of balance that could be perceptible. The median ratio among classes is shown below. The tendency for excellent wines to have a higher ratio of fixed acidity to residual sugar should not be surprising since there is not much variation in fixed acidity by excellent wines have less residual sugar. Or one way of looking at this is that ratio in excellent wines is about 70% higher than in poor wines. Perhaps the takeaway should be that acid should be slightly greater than sugar.

Ratio of Fixed Acidity to Residual Sugar by Class

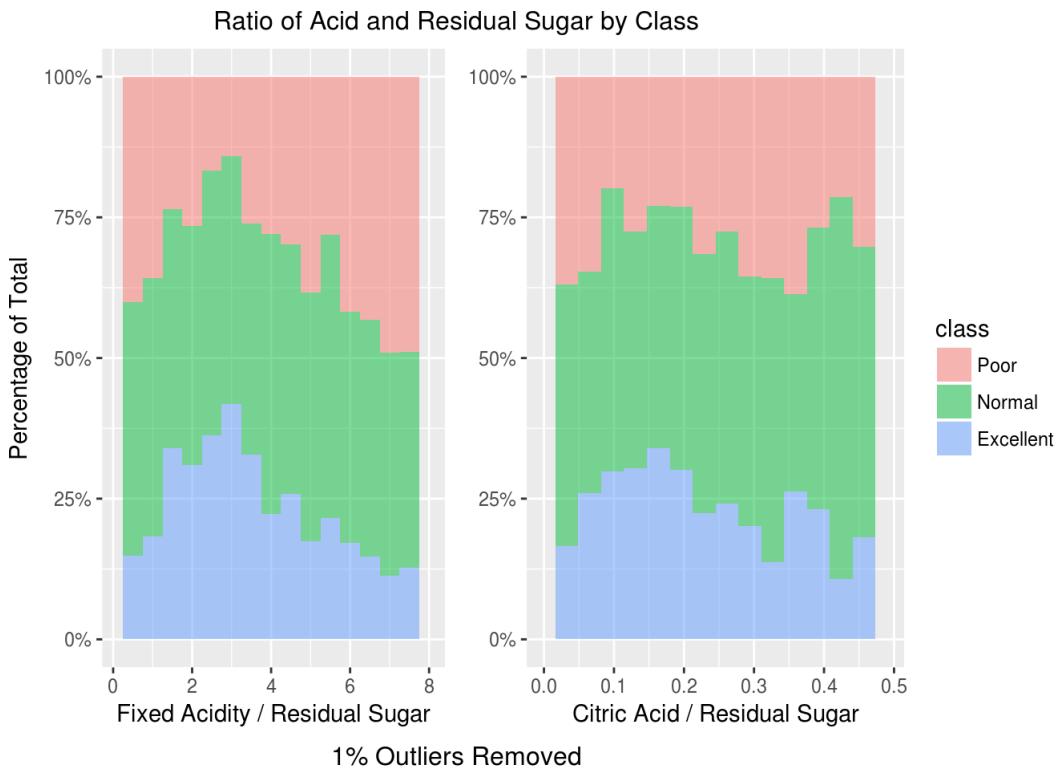


A similar relationship is shown with regards to citric acid. For a given amount of residual sugar, excellent wines tend to have more citric acid. The levels of citric acid are much lower than tartaric acid so this relationship may be harder for the palette to perceive.

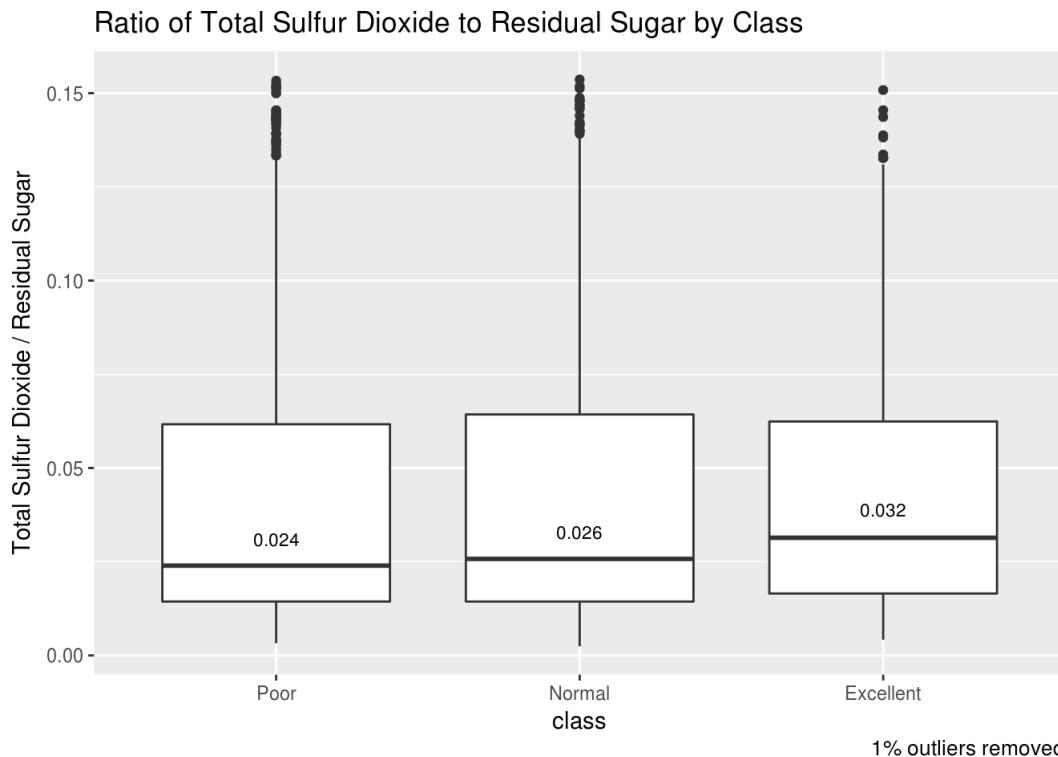


Now I would like to examine the likelihood of a wine belonging to one of the classes given an acid to sugar ratio. I'll use the histogram plot where each bin scales up to 1. We may have finally found the relationship of acid to quality that I have been looking for! There is definitely a ratio between fixed acidity and residual sugar where the likelihood of a poor wine is minimized and excellent wines maximized at around 3. If the ratio is less or more than this value, the likelihood of a wine being poor increases. Normal wines seem to exist across the entire range of ratios. A wine with a ratio of 1 is about as likely to be normal as one with a ratio of 8 which seems a wide disparity to me. Still, wines at the extreme ratios can also be excellent.

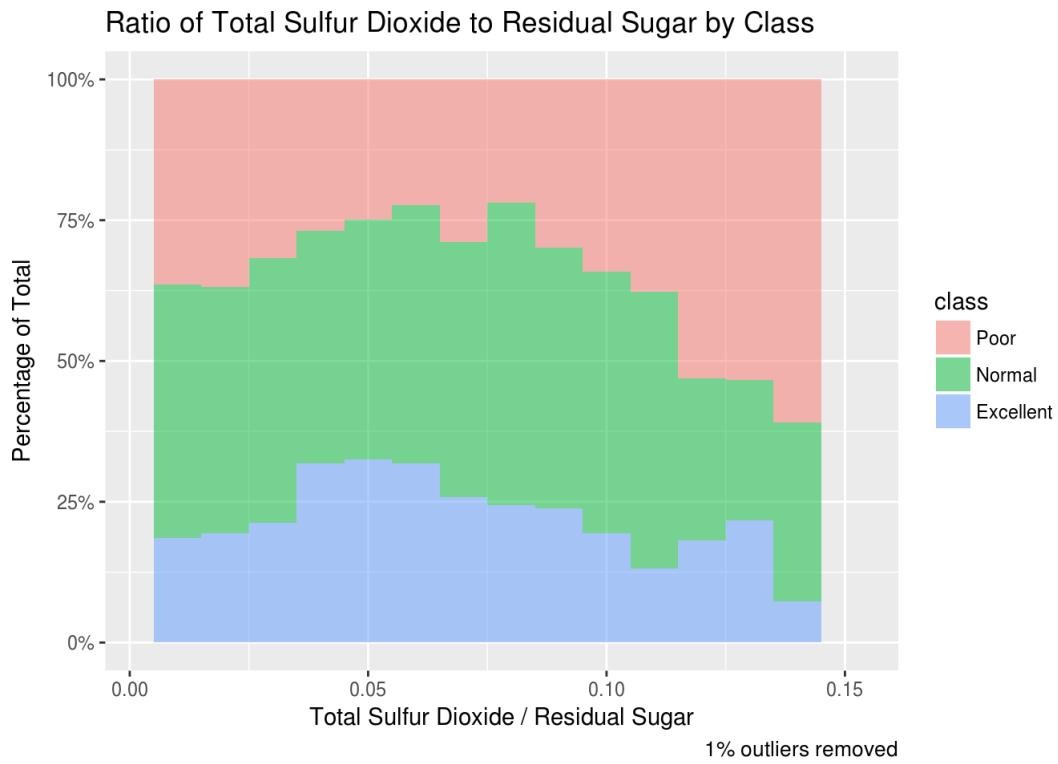
There is a similar peak likelihood for excellent wines that have a citric acid to residual sugar ratio of about 0.15. The effect doesn't look as strong as the ratio of fixed acidity.



Previously I found that total sulfur dioxide increased with residual sugar which I thought was related to an effort to prevent spoilage. Is there a difference in application of sulphates among the wine classes? For a given amount of residual sugar, the amount of total sulfur dioxide seems to be constant between the classes.



A scaled histogram plot also shows how a high level of total sulfur dioxide makes it likely that the wine will be poor. For the most part, there doesn't seem to be a big difference in class proportions for a given ratio of Sulfur Dioxide to Residual sugar. Perhaps there is just a standard level of sulfur added to wines for a given amount of residual sugar that is more or less followed by all wine producers. If there isn't a difference in quality, there is no reason to vary sulphate levels away from standard practice.



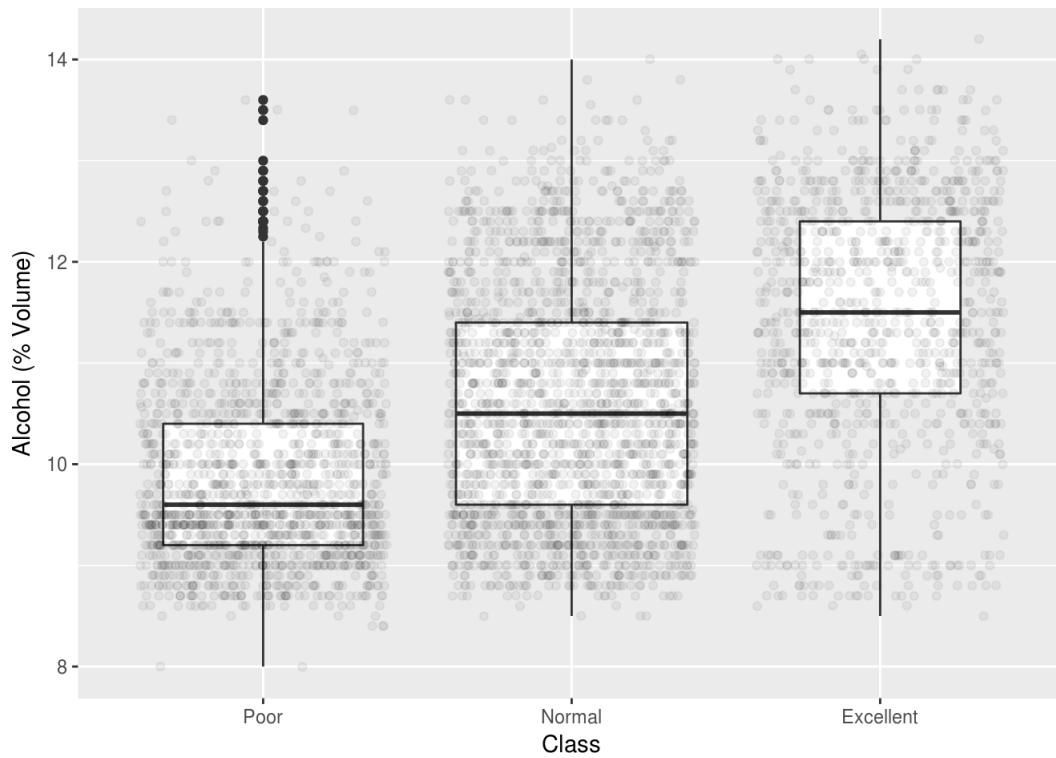
Multivariate Analysis

Wine quality is the sum of many variables and I was able to examine a few interactions between variables. Although acidity didn't seem to affect quality, its ratio in proportion with residual sugar seemed to peak at a specific value. Wines that didn't have a good balance of acidity to sugar were generally poor.

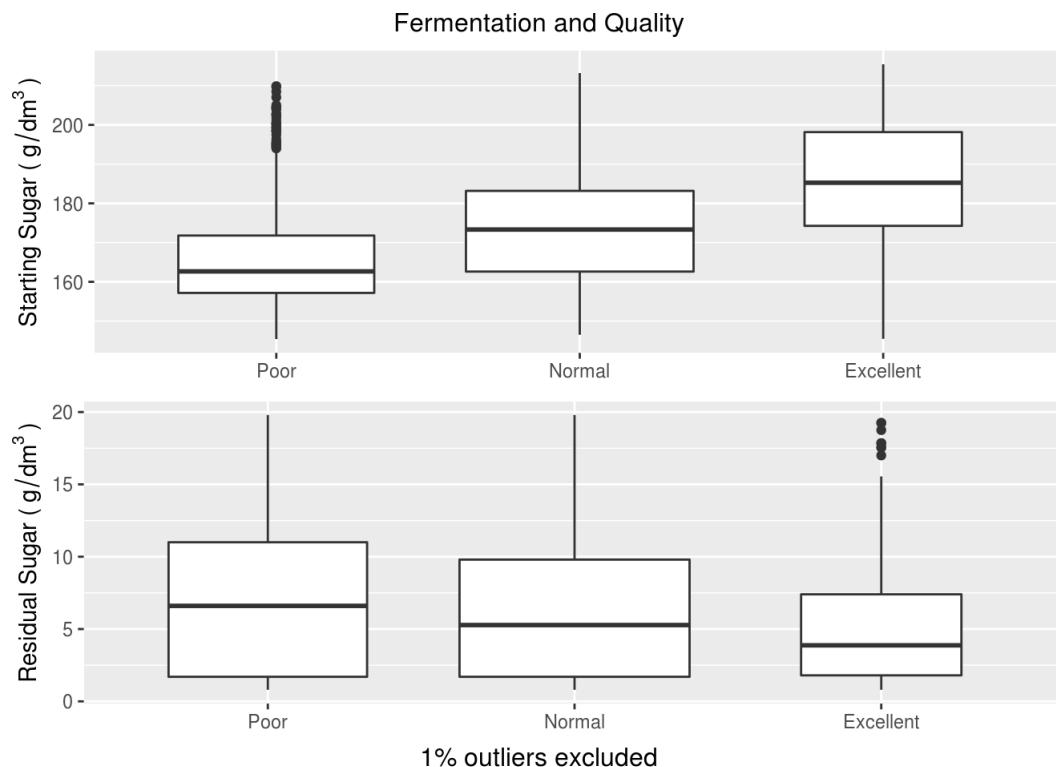
The only relationship I found with sulphates and quality was that if a wine had a high level of total sulfur dioxide versus residual sugar, it was probably poor.

Summary

The simple box plot of alcohol grouped by class I think summarizes the dataset pretty well. More alcohol means better quality. I wish there was more to say about wine quality but this plot seems to sum things up better than any other plot I made.

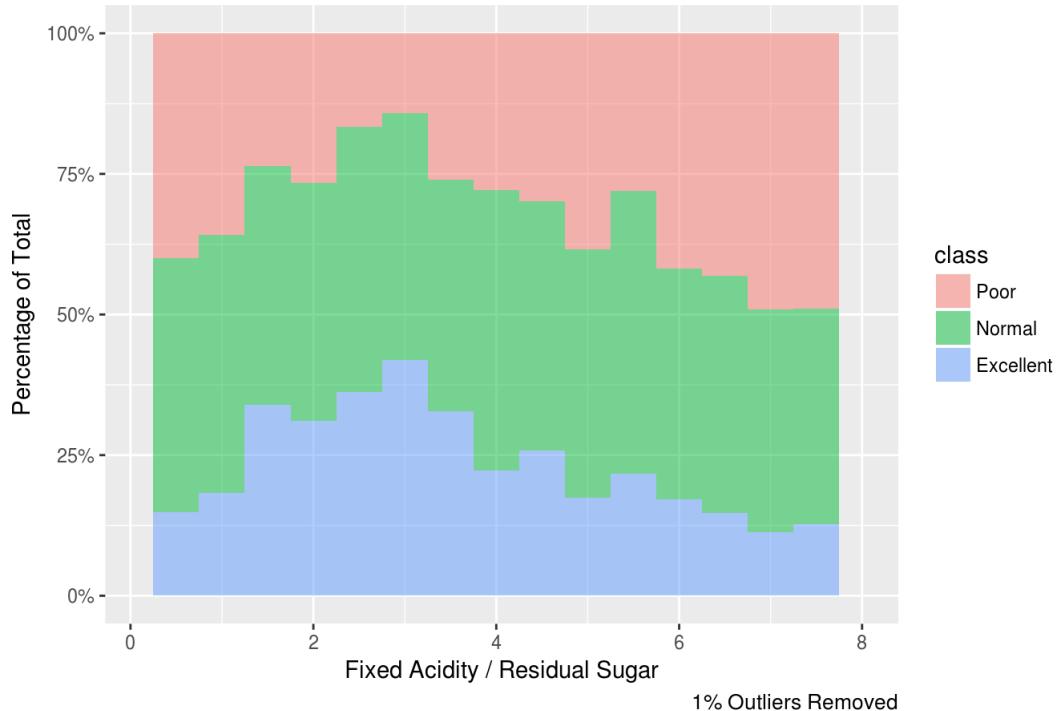


Since alcohol is important, I wanted to figure out how it was made. I was able to determine that wines that are high in alcohol started out as grape juice that had more starting sugar and were fermented more completely resulting in less residual sugar. These are the secondary variables that contribute to the high alcohol levels in excellent wines.



Finally, because I tried so hard to find a variable that wasn't related to fermentation I would like to present the ratio of fixed acidity to residual sugar plot. It was also the only plot I was able to make that showed a variable related to quality that didn't have a linear relationship.

Ratio of Acid and Residual Sugar by Class



Wine starts out as grape juice with a given amount of sugar and acidity. If you have a lot of sugar, you have the potential to make a very alcoholic wine but you must make sure your yeast works efficiently, converting most of the sugars to alcohol. But you want some sugars leftover to balance the acidity. More alcohol is generally better, but an acidity balance seems to move toward a fixed ratio. Make an alcoholic wine with a good acidity to sugar balance, and you increase the likelihood of the wine being judged favorably.

Reflection

R is a powerful numeric tool with many useful packages for analyzing and visualizing data. Being able to type functions into an IDE seems more efficient and faster than working in a spreadsheet. Using R Markdown makes analyzing and writing about data seem seamless compared to using an office suite.

Data exploration is like jumping down a rabbit hole for me. I had lots of ideas about how to look at the data that didn't work out. Trying to find the useful way of interpreting data and expressing it visually can send you in many different directions that don't pay off. Exploratory data analysis requires creativity. For me, it also requires some restraint so that I don't spend too much time trying to figure out a particular relationship that may or may not exist.

This dataset was fun to examine. I know a lot about alcohol from making and drinking beer. Learning about wine was fun and I was able to bring my knowledge and experience to the analysis. I am glad I am through with this dataset though. I got too obsessed with trying to show certain relationships with quality.

This dataset needs to be compared with other wine quality assessment datasets. Comparing the white wine dataset to the red wine dataset would have been an interesting addition to this analysis. I am glad I didn't try to model this dataset. In principle, I would rather people drink wine to assess its quality rather than train a machine how to rate wine.