

An Exploration of Red Wine by Joseph Mapula

This report explores a dataset containing physical attributes for approximately 1599 red wines and their corresponding quality rating as determined by wine experts.

The style is intended to be a stream of consciousness exploration followed by a more put together analysis of prior findings. The structure will begin with analyzing one variable at a time then proceed to incorporate more, all with the guiding question: **which chemical properties influence the quality of red wines?**

This is modeled after the Udacity example project and rubric.

Introduction to the data:

This is a collection of data on 1599 red wine samples with values for 11 objective tests as well as median values for ratings by wine experts (12 variables total).

From the text file provided...

Input variables (based on physicochemical tests):

1. – fixed acidity (tartaric acid - g / dm³)
 - most acids involved with wine or fixed or nonvolatile (do not evaporate readily)
2. – volatile acidity (acetic acid - g / dm³)
 - the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste
3. – citric acid (g / dm³)
 - found in small quantities, citric acid can add ‘freshness’ and flavor to wines
4. – residual sugar (g / dm³)
 - the amount of sugar remaining after fermentation stops, it’s rare to find wines with less than 1 gram/liter and wines with greater than 45 grams/liter are considered sweet
5. – chlorides (sodium chloride - g / dm³)
 - the amount of salt in the wine
6. – free sulfur dioxide (mg / dm³)
 - the free form of SO₂ exists in equilibrium between molecular SO₂ (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine
7. – total sulfur dioxide (mg / dm³)
 - amount of free and bound forms of S₀₂; in low concentrations, SO₂ is mostly undetectable in wine, but at free SO₂ concentrations over 50 ppm, SO₂ becomes evident in the nose and taste of wine
8. – density (g / cm³)
 - the density of water is close to that of water depending on the percent alcohol and sugar content
9. – pH
 - describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale
10. – sulphates (potassium sulphate - g / dm³)
 - wine additive which can contribute to sulfur dioxide gas (SO₂) levels, which acts as an antimicrobial and antioxidant
11. – alcohol (% by volume)
 - the percent alcohol content of the wine

Output variable (based on sensory data):

12. – quality (score between 0 and 10)
 - at least 3 wine experts rated the quality of each wine, providing a rating between 0 (very bad) and 10 (very excellent)

Univariate Plots Section

First, we'll begin with getting a high-level view, then by examining one variable at a time.

What does our dataset look like?

```
## 'data.frame': 1599 obs. of 13 variables:
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...
## $ fixed.acidity : num 7.4 7.8 7.8 11.2 7.4 7.4 7.9 7.3 7.8 7.5 ...
## $ volatile.acidity : num 0.7 0.88 0.76 0.28 0.7 0.66 0.6 0.65 0.58 0.5 ...
## $ citric.acid : num 0 0 0.04 0.56 0 0 0.06 0 0.02 0.36 ...
## $ residual.sugar : num 1.9 2.6 2.3 1.9 1.9 1.8 1.6 1.2 2 6.1 ...
## $ chlorides : num 0.076 0.098 0.092 0.075 0.076 0.075 0.069 0.065 0.073 0.071 ...
## $ free.sulfur.dioxide : num 11 25 15 17 11 13 15 15 9 17 ...
## $ total.sulfur.dioxide : num 34 67 54 60 34 40 59 21 18 102 ...
## $ density : num 0.998 0.997 0.997 0.998 0.998 ...
## $ pH : num 3.51 3.2 3.26 3.16 3.51 3.51 3.3 3.39 3.36 3.35 ...
## $ sulphates : num 0.56 0.68 0.65 0.58 0.56 0.56 0.46 0.47 0.57 0.8 ...
## $ alcohol : num 9.4 9.8 9.8 9.8 9.4 9.4 9.4 10 9.5 10.5 ...
## $ quality : int 5 5 5 6 5 5 5 7 7 5 ...
```

What are the ranges for our variables?

```
##      X      fixed.acidity  volatile.acidity  citric.acid
## Min.   : 1.0    Min.   : 4.60    Min.   :0.1200    Min.   :0.000
## 1st Qu.: 400.5    1st Qu.: 7.10    1st Qu.:0.3900    1st Qu.:0.090
## Median : 800.0    Median : 7.90    Median :0.5200    Median :0.260
## Mean   : 800.0    Mean   : 8.32    Mean   :0.5278    Mean   :0.271
## 3rd Qu.:1199.5    3rd Qu.: 9.20    3rd Qu.:0.6400    3rd Qu.:0.420
## Max.   :1599.0    Max.   :15.90    Max.   :1.5800    Max.   :1.000
## residual.sugar  chlorides      free.sulfur.dioxide
## Min.   : 0.900    Min.   :0.01200    Min.   : 1.00
## 1st Qu.: 1.900    1st Qu.:0.07000    1st Qu.: 7.00
## Median : 2.200    Median :0.07900    Median :14.00
## Mean   : 2.539    Mean   :0.08747    Mean   :15.87
## 3rd Qu.: 2.600    3rd Qu.:0.09000    3rd Qu.:21.00
## Max.   :15.500    Max.   :0.61100    Max.   :72.00
## total.sulfur.dioxide  density      pH      sulphates
## Min.   : 6.00    Min.   :0.9901    Min.   :2.740    Min.   :0.3300
## 1st Qu.: 22.00    1st Qu.:0.9956    1st Qu.:3.210    1st Qu.:0.5500
## Median : 38.00    Median :0.9968    Median :3.310    Median :0.6200
## Mean   : 46.47    Mean   :0.9967    Mean   :3.311    Mean   :0.6581
## 3rd Qu.: 62.00    3rd Qu.:0.9978    3rd Qu.:3.400    3rd Qu.:0.7300
## Max.   :289.00    Max.   :1.0037    Max.   :4.010    Max.   :2.0000
## alcohol      quality
## Min.   : 8.40    Min.   :3.000
## 1st Qu.: 9.50    1st Qu.:5.000
## Median :10.20    Median :6.000
## Mean   :10.42    Mean   :5.636
## 3rd Qu.:11.10    3rd Qu.:6.000
## Max.   :14.90    Max.   :8.000
```

A few observations:

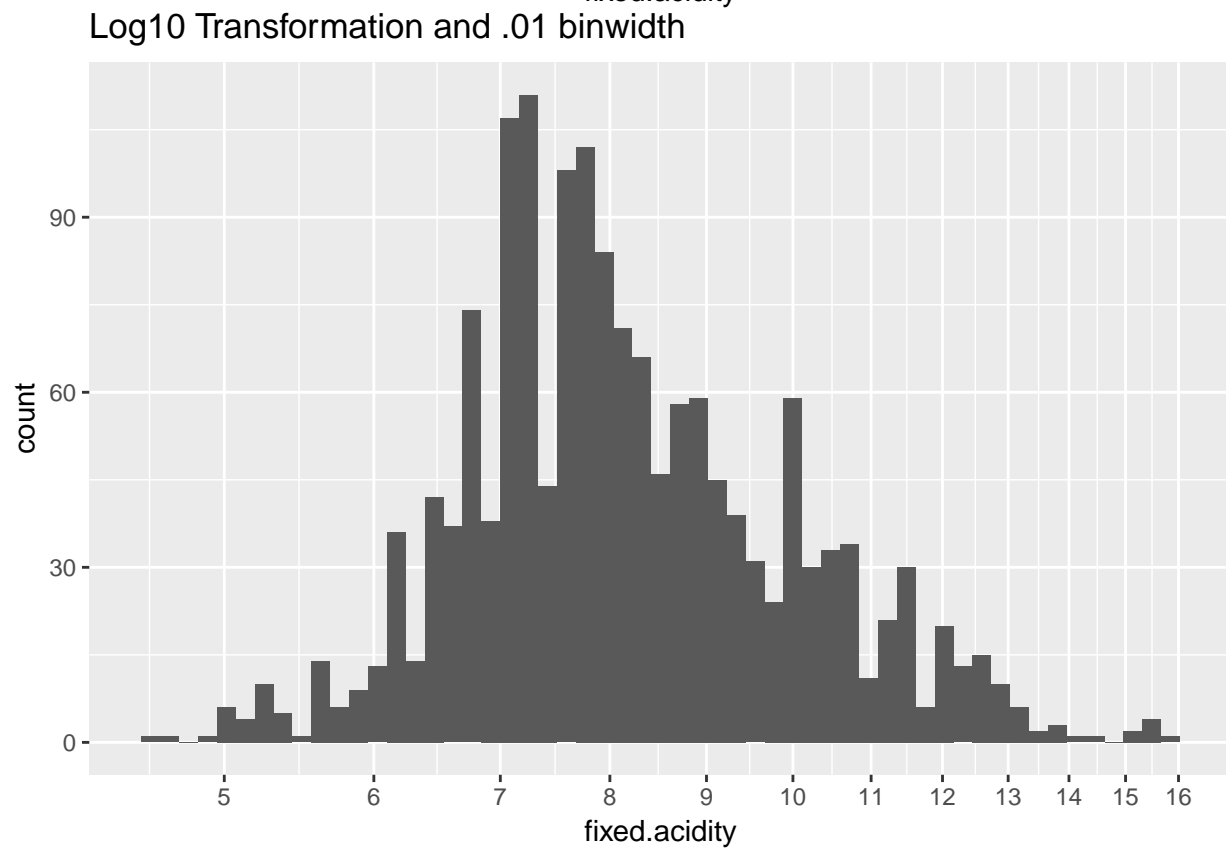
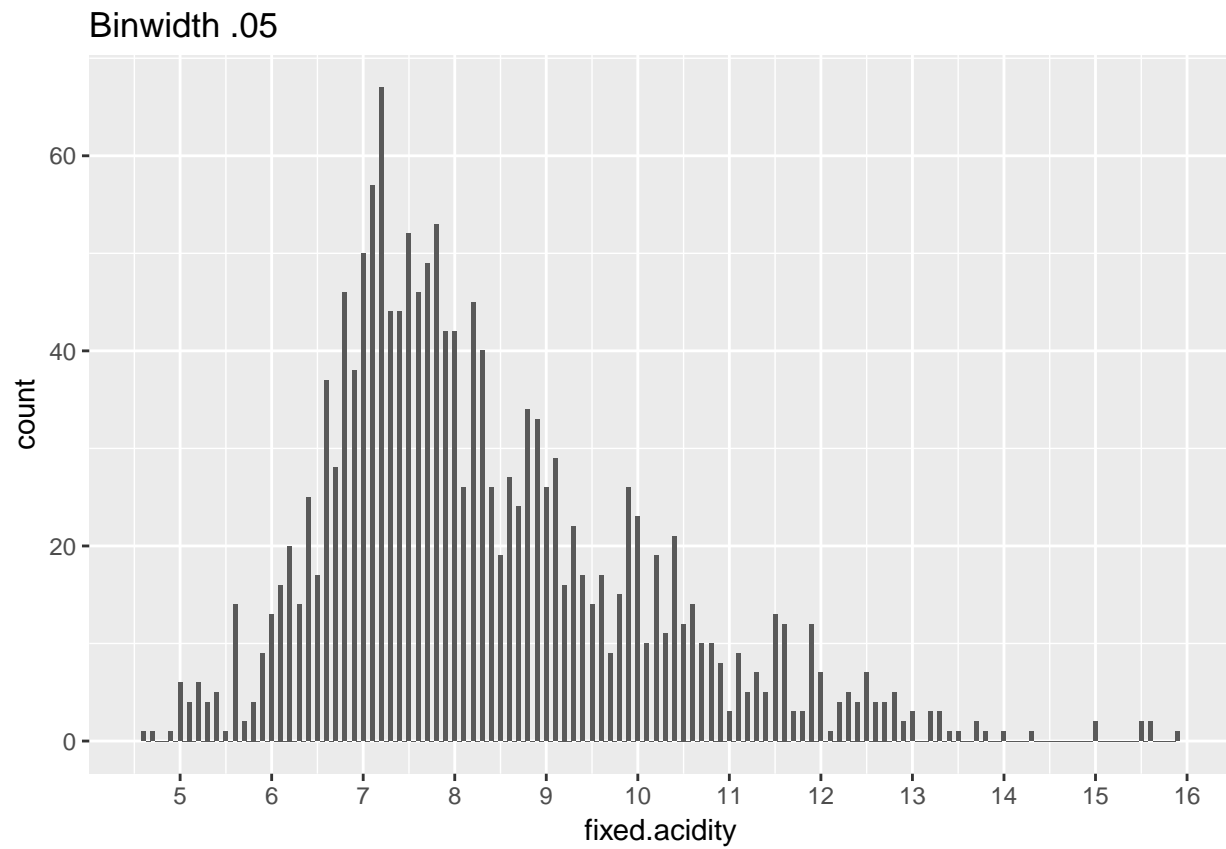
Many of the variables seem to operate on different scales. Most variables seem fine but there are a few that may have outliers to keep an eye out for: each acid, residual.sugar, chlorides, sulfur.dioxide (total and free), and sulphates.

It would be very useful if we knew more about our variables. What does a difference of 10g/dm^3 vs 100g/dm^3 for citric acid even signify taste wise? What are the typical ranges for most wines? What were the judges looking for in standardizing something as subjective as taste?

Some variables may be useful to modify depending on the visualization (changing quality into an ordered factor).

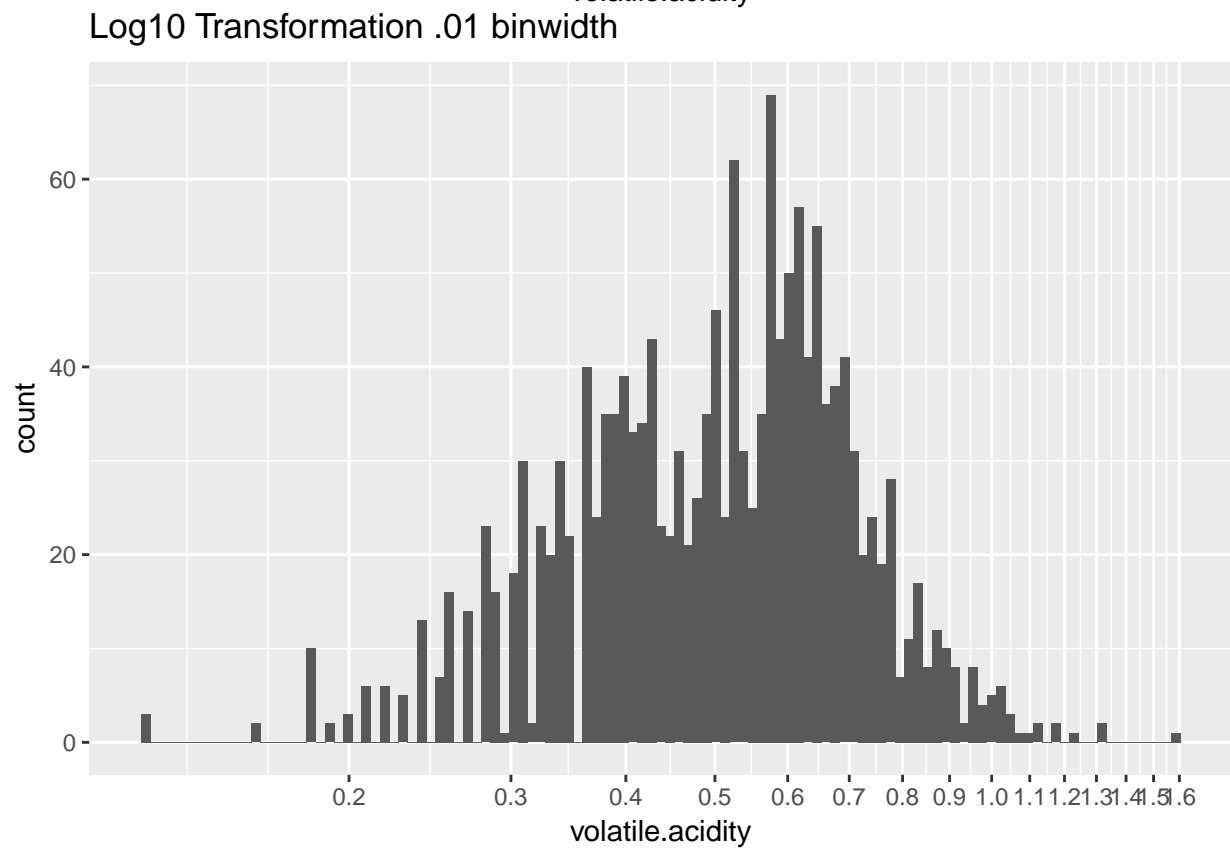
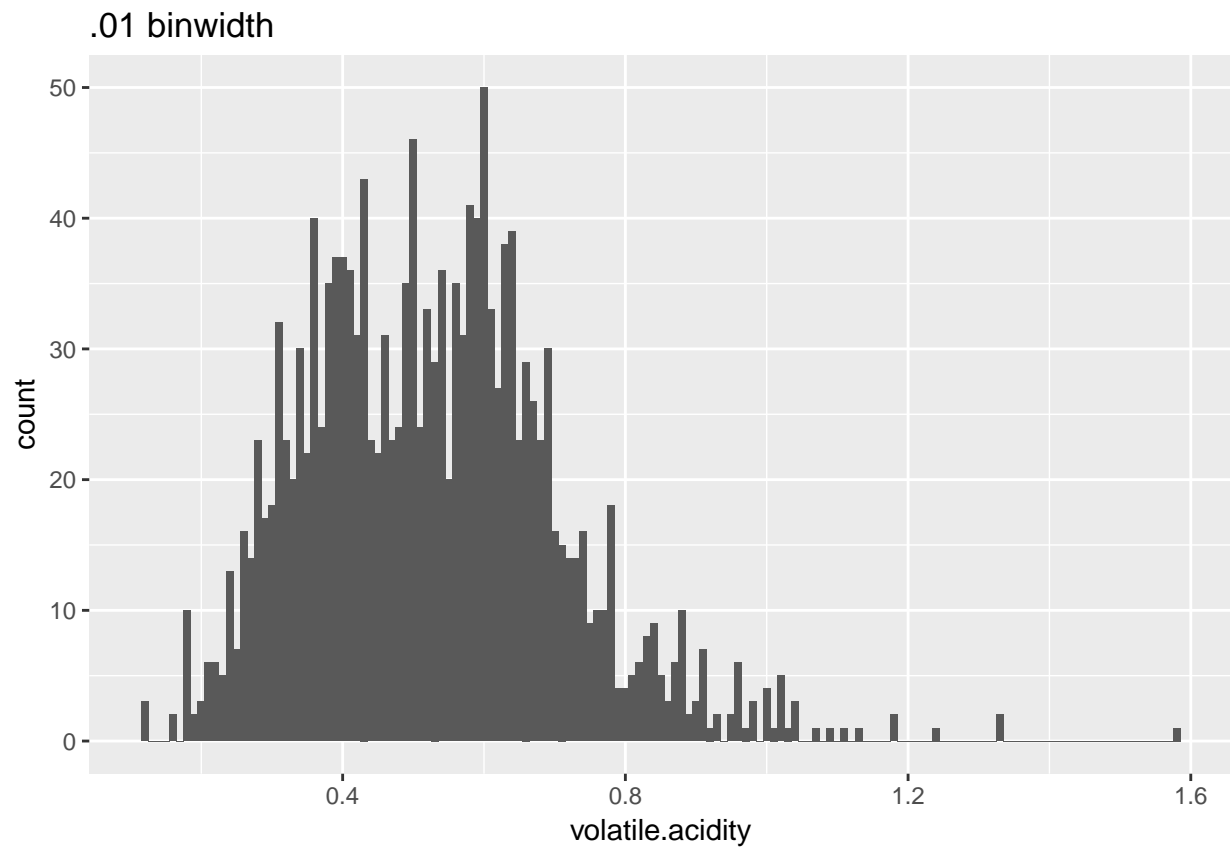
Let's dive in a bit more and take a look at each variable to understand their distributions.

Fixed Acidity



Varies at the .01 level. It seems most fixed acidity values fall between 6.5 and 10 with most under 8.5. They also vary in small increments and are right skewed. A log10 transformation of the x axis normalized the data a bit more.

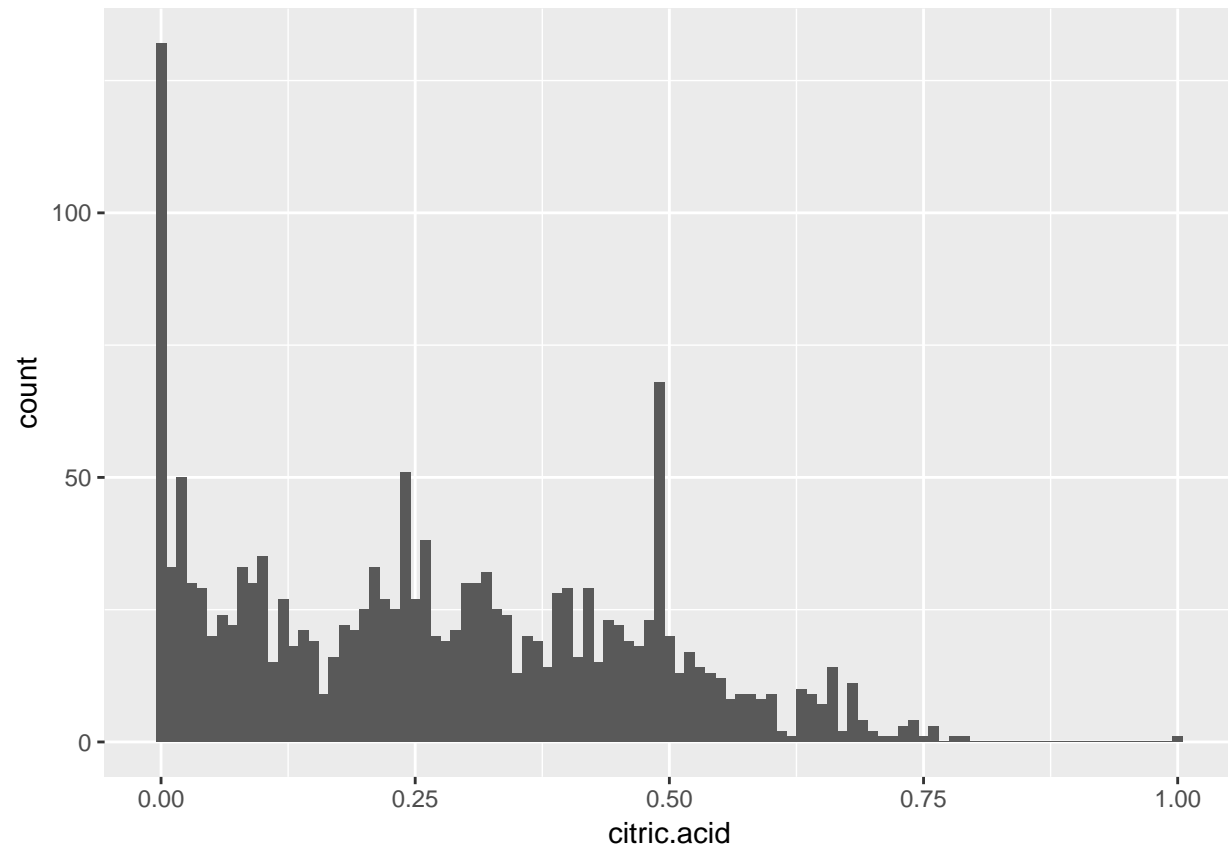
Volatile Acidity



Varies at the .001/.01 level. Volatile acidity looks very similar to the fixed acidity variable.

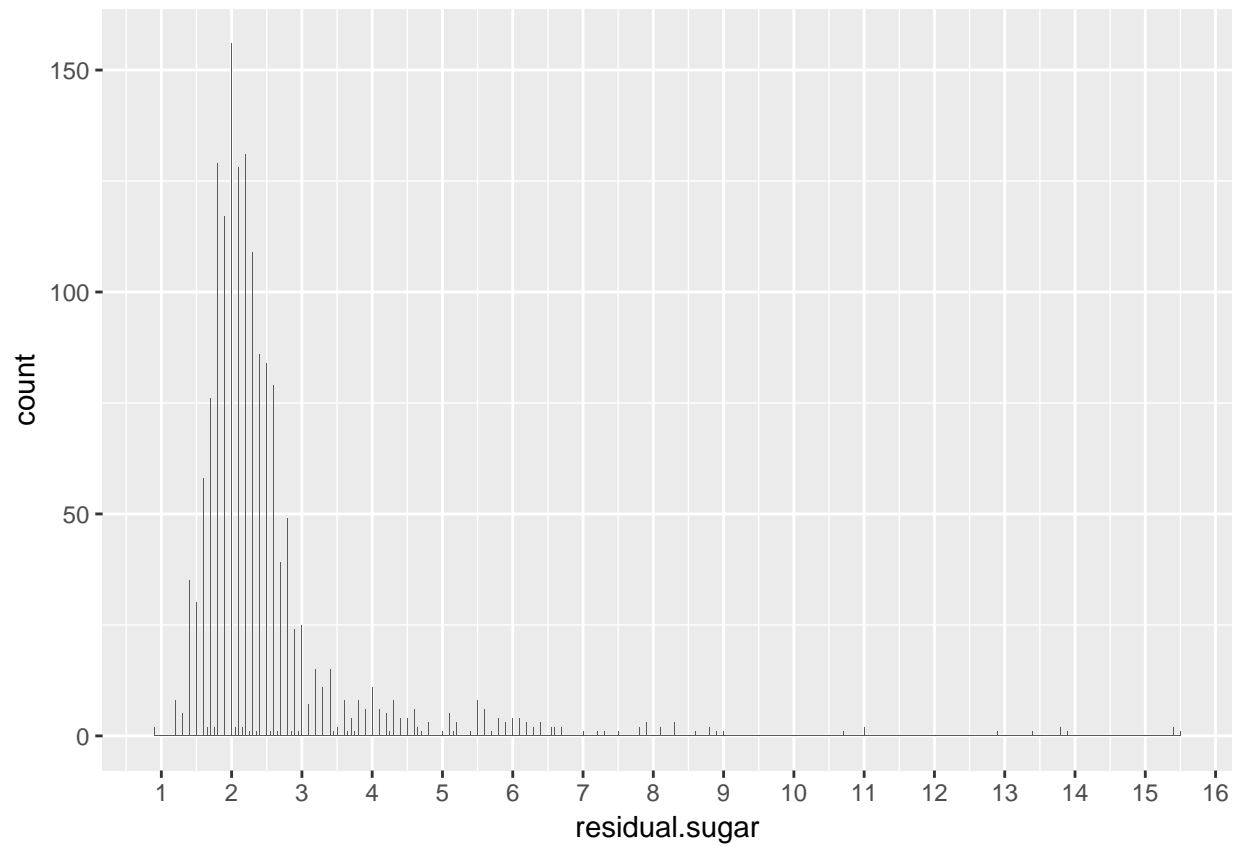
VA ranges from 0-1.6 instead while fixed acidity varied on 0-16 and seems to be on a tenth of the scale of fixed acidity.

Citric acid



Citric acid seems to be more of an optional additive as many wines contain no citric acidity at all. With a range of 0-1, most values are under .5 with spikes at regular intervals such as 0, .25 and .5.

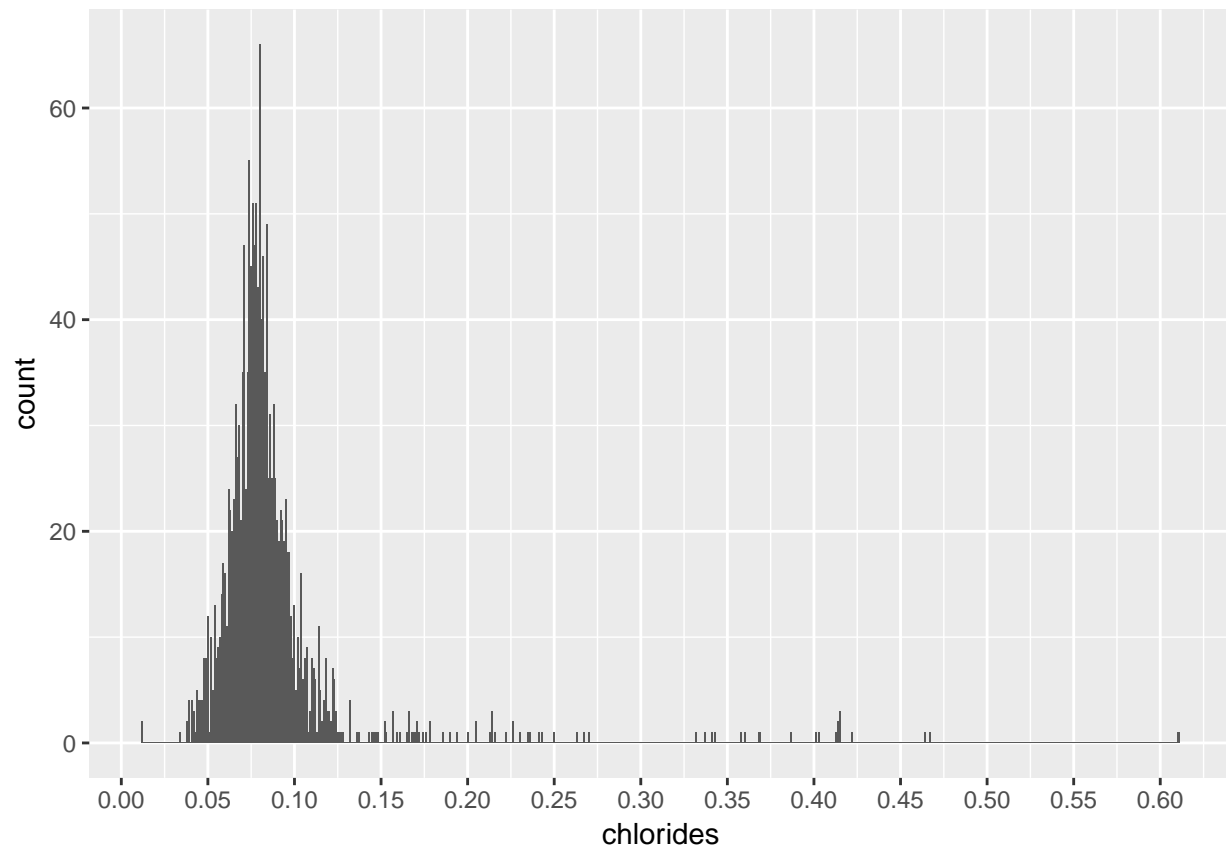
Residual sugar



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.900	1.900	2.200	2.539	2.600	15.500

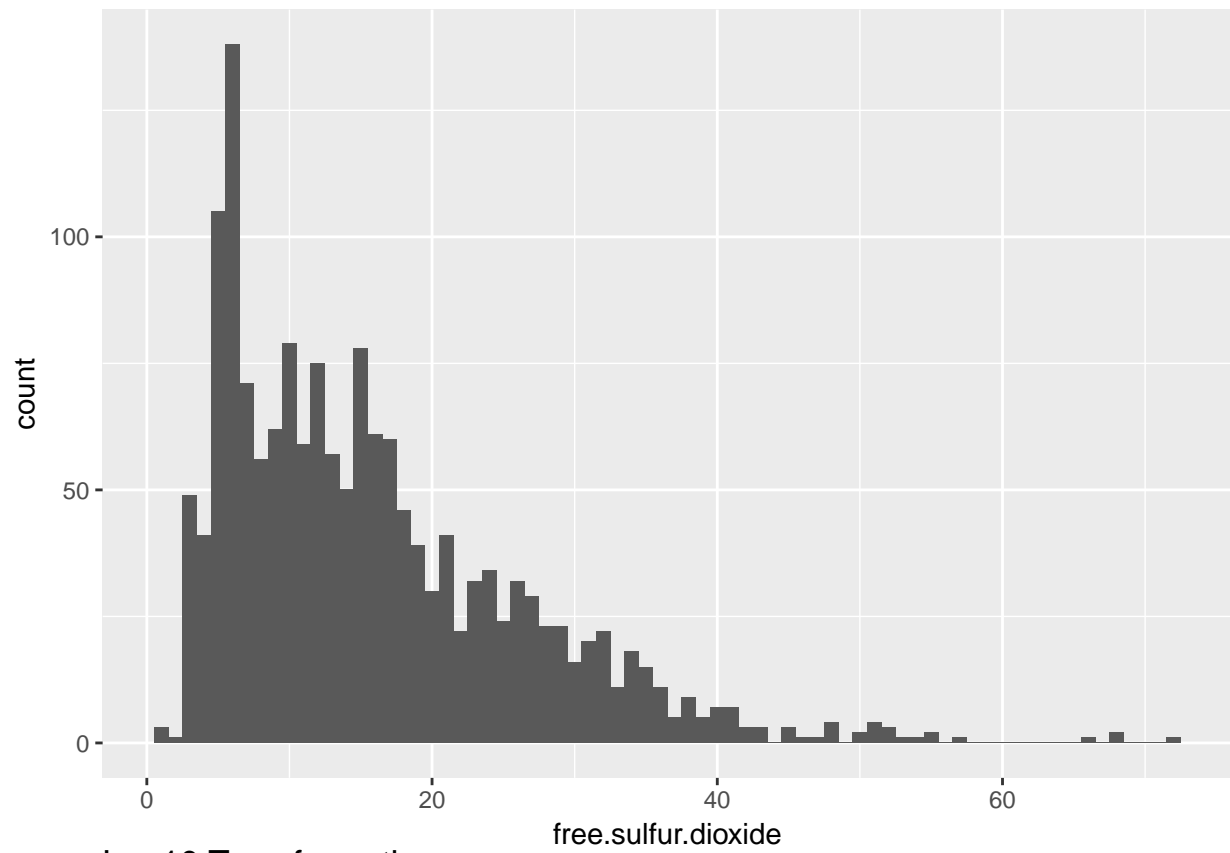
Residual sugar varies on either the tenths or hundredths scale. Most levels fall between 1 and 3

Chlorides

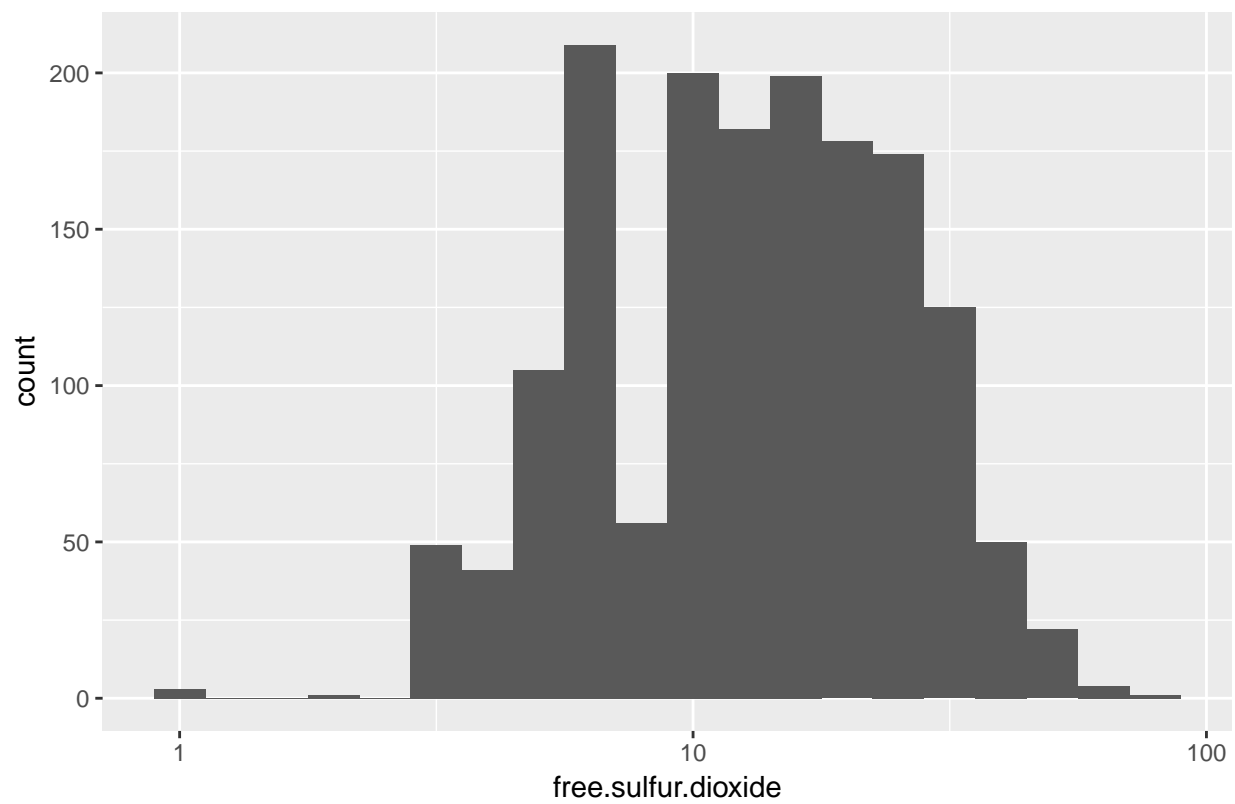


Chlorides vary on the hundredths or thousandths scale and most values lie between .05 and .1 with some ranging to .4 and even above .6. I don't know salt's role in wine specifically, but I know salt helps bring out the natural flavors in food but the amount needed usually depends on the intensity of the original flavors (i.e. you would likely salt your fish less heavily than your eggs). My guess is salt can ruin a wine but won't increase its quality/flavor very much.

Free Sulfur



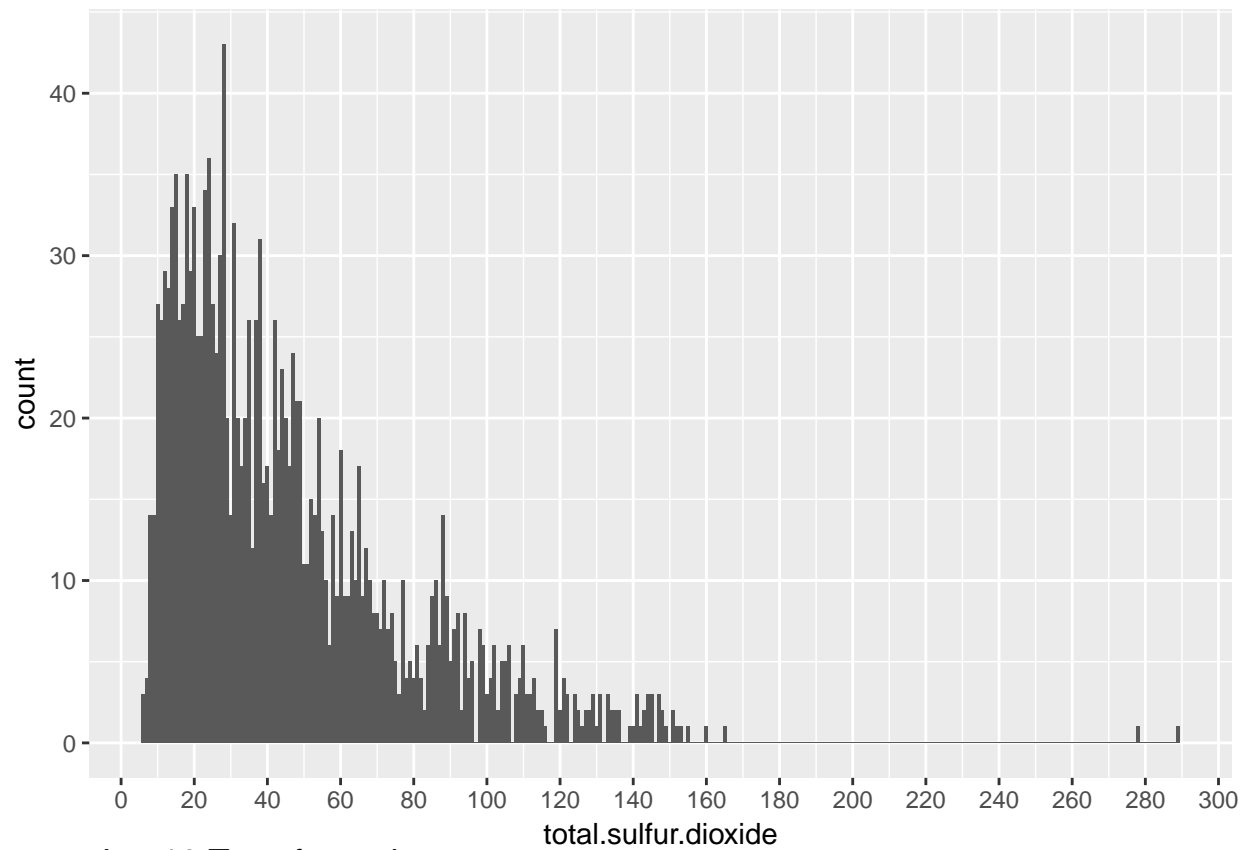
Log10 Transformation



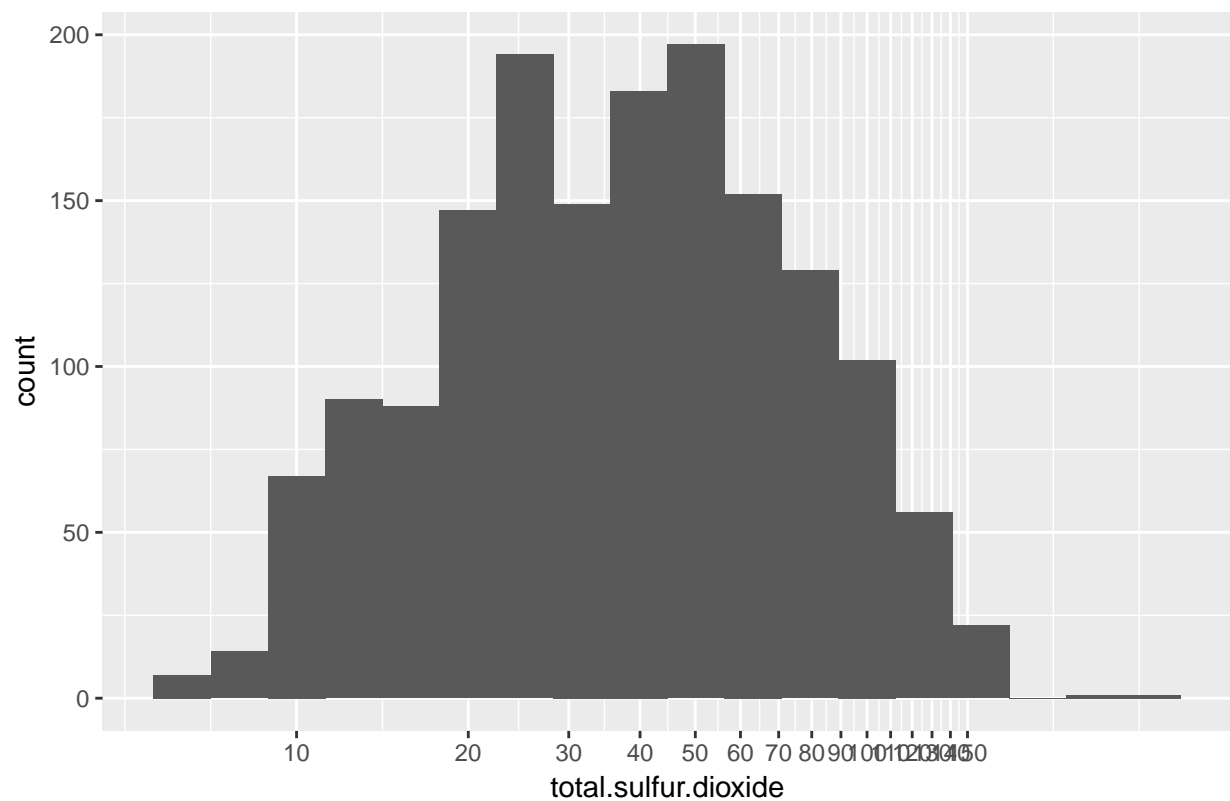
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.00	7.00	14.00	15.87	21.00	72.00

Free Sulfur seems to be an integer with most values between 7-15. The distribution seems a bit long tailed to the right, but appears more normal after applying a log10 transformation.

Total Sulfur Dioxide

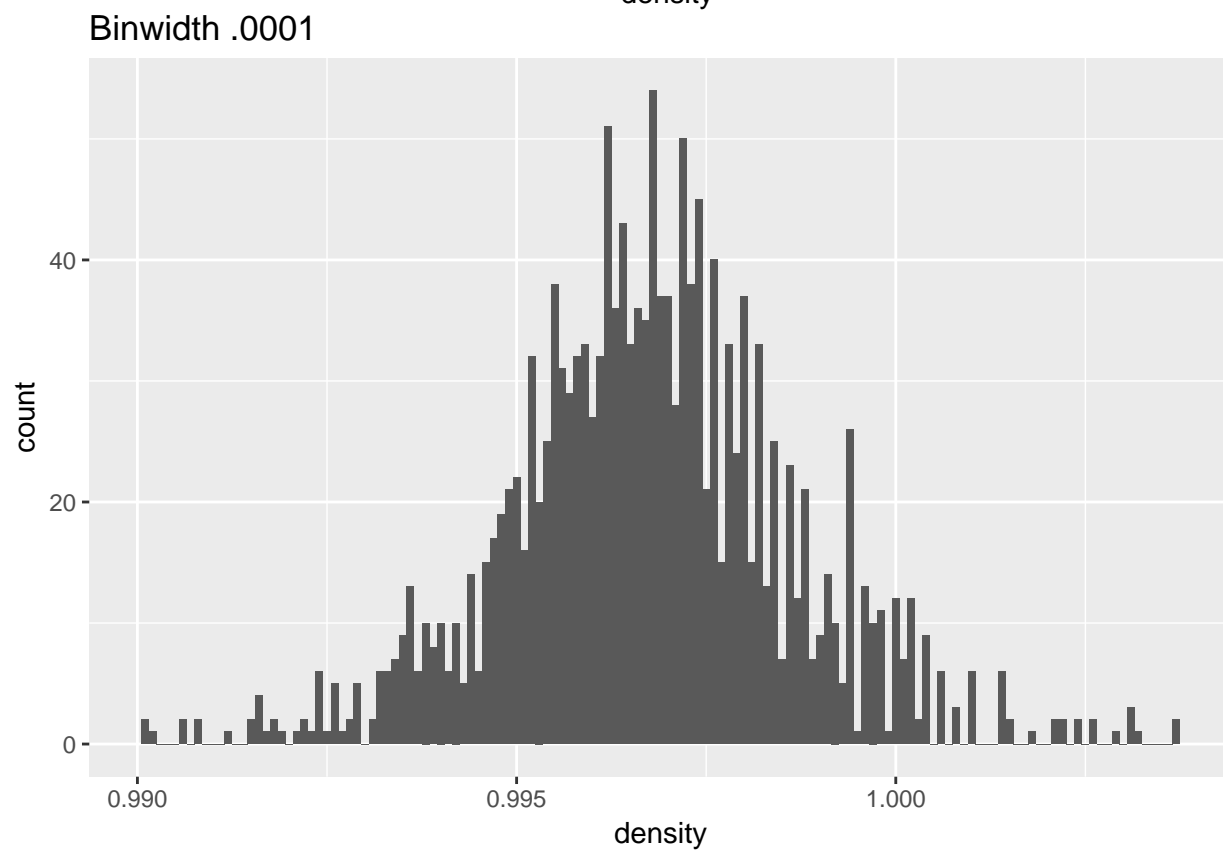
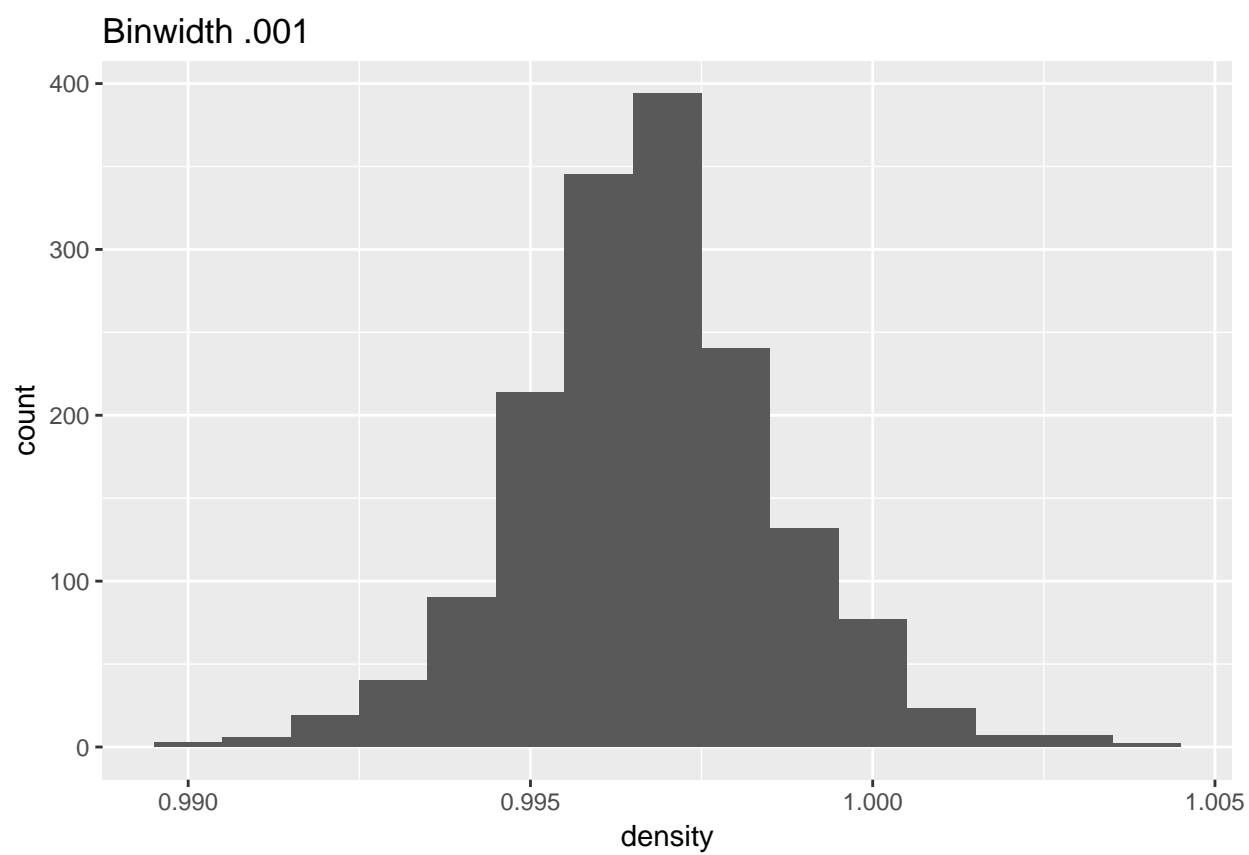


Log10 Transformation



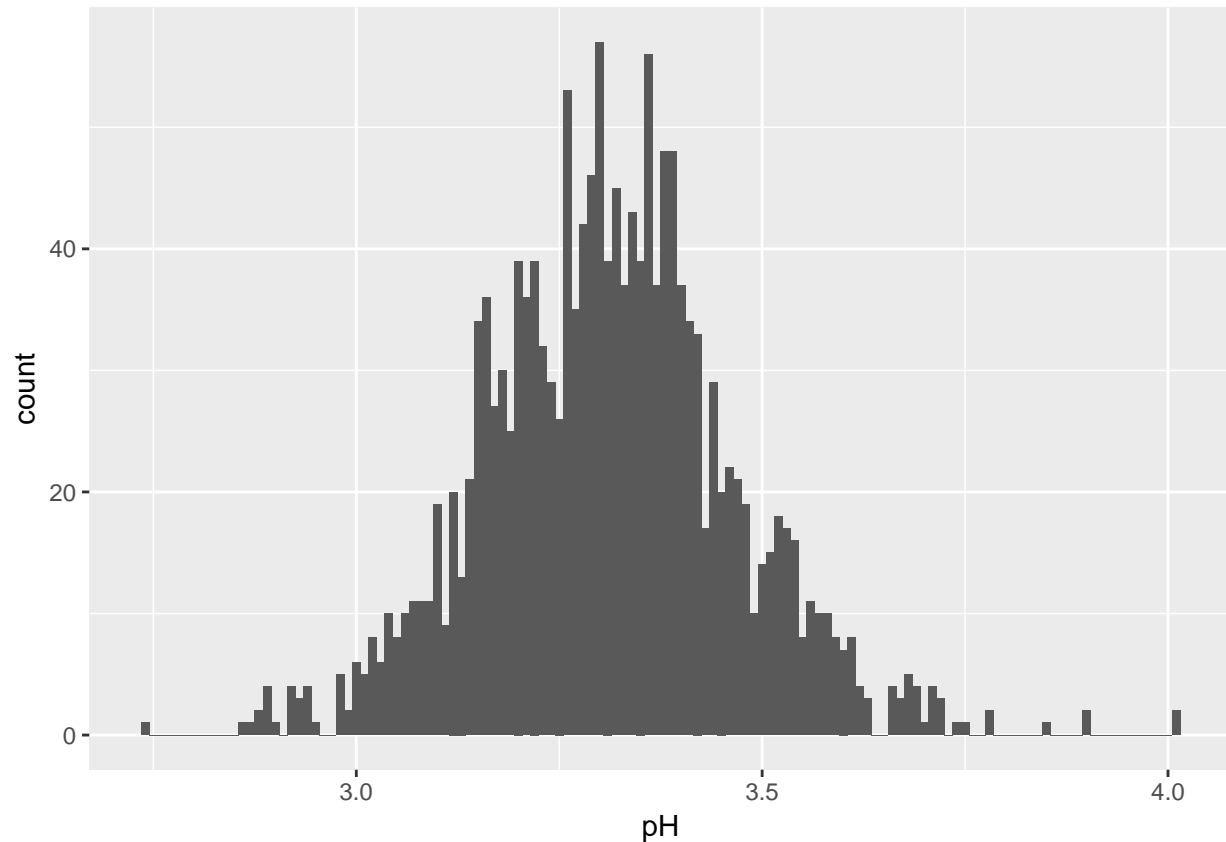
Total Sulfur seems fairly similar to free sulfur. Most values are a bit higher and lie between 20-70 but range up to 280. The distribution is also long tailed and appears more normal after log10 transformation.

Density



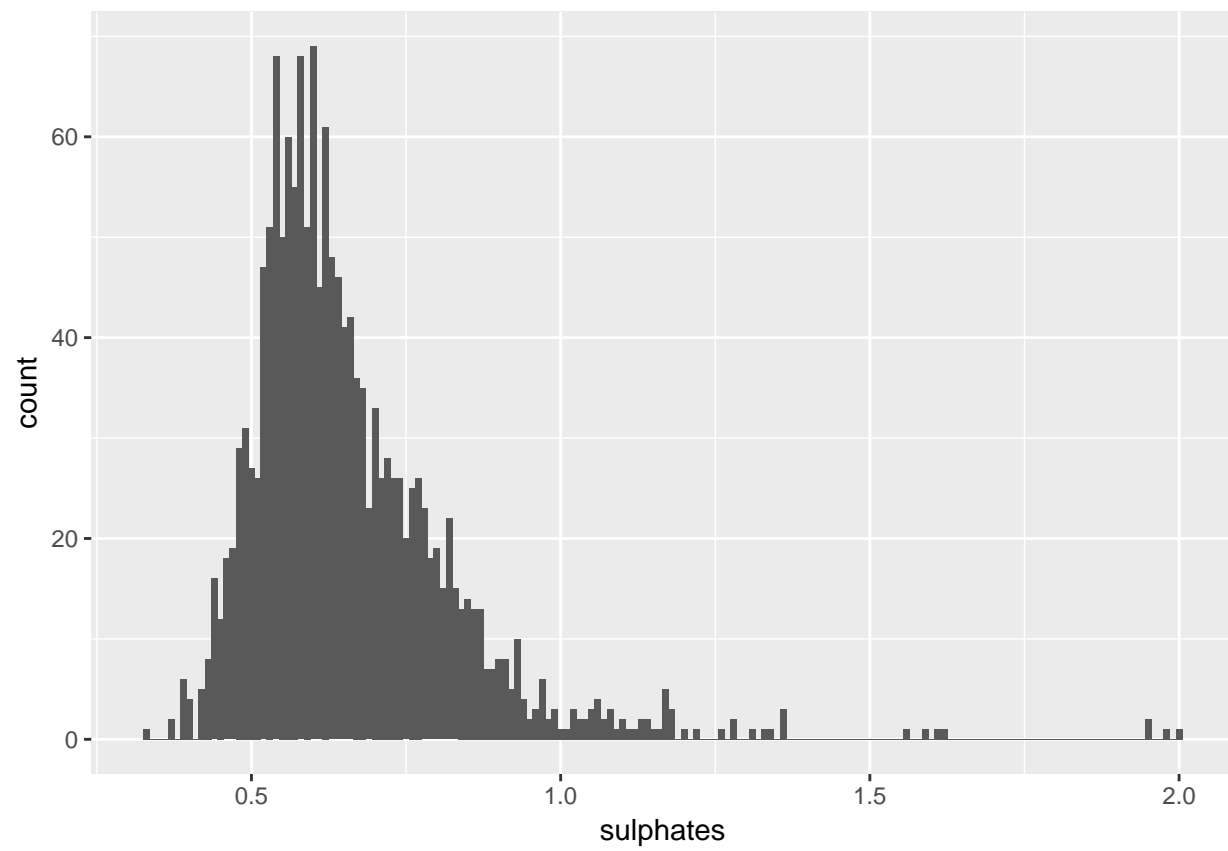
Density varies on orders of magnitude smaller than other variables and only ranges from .99 to not even 1.004. This distribution seems normal. I doubt the density of wine significantly changes its flavor rather than its texture. However, if density depends on other factors that significantly change flavor, then it may be a good summary variable for other variables that need to be in balance (like maybe salt and citric acid).

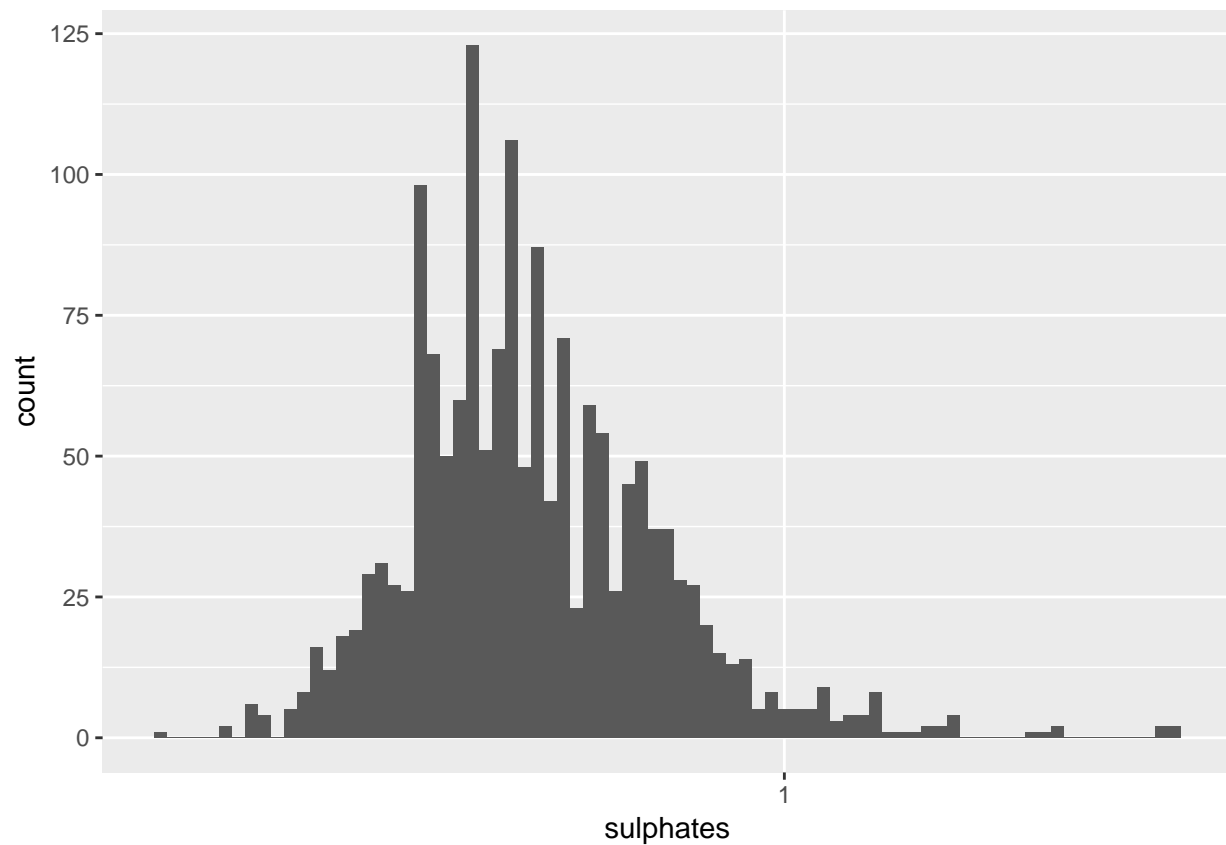
PH



PH is on a logarithmic scale which may be useful knowledge to note for later. Varies on the hundredths level and appears to have a normal distribution. Intuitively I imagine pH would serve as a summary variable in the same way that density would. pH may be a good indicator of the amount of acids/bases that affect flavor being in balance. In addition, I would also think pH may work a similar way to salt in that a certain range would be ideal and outside of that an otherwise delicious wine would be too acidic/basic to really enjoy by itself. This would work in the same way flavored vinaigrettes work.

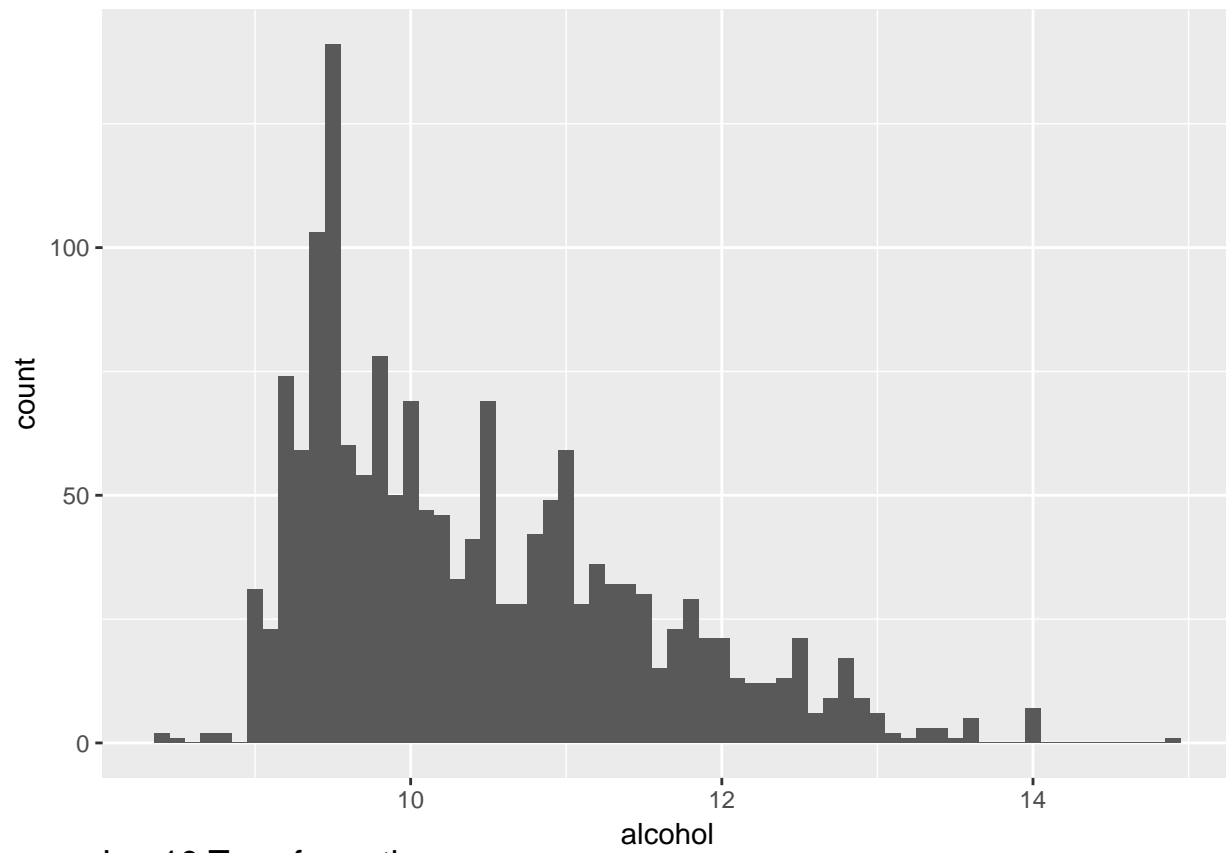
Sulphates



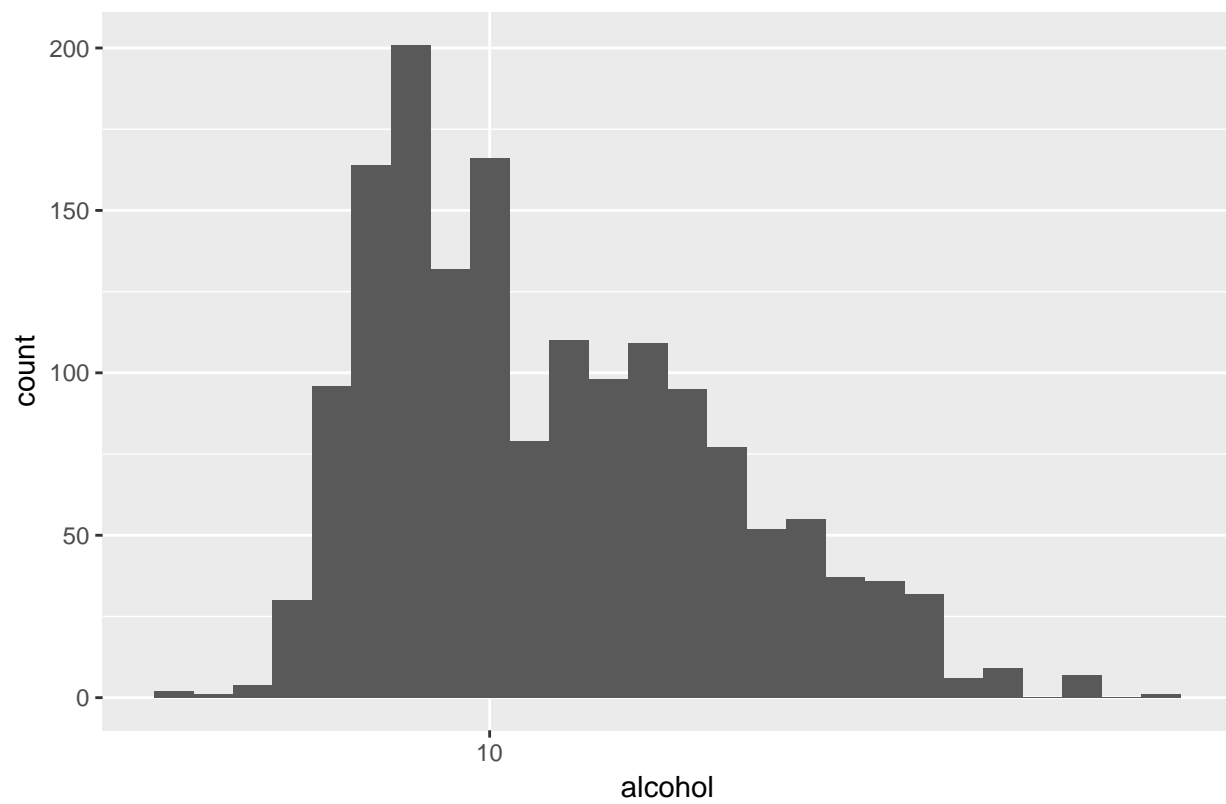


Sulphates vary on the hundredths level with most data falling between .5 and .75. The data is long tailed to the right which is slightly fixed by a log10 transformation. I imagine given main use of sulphates is as an antimicrobial/antioxidant yet it also contributes to total sulfur dioxide which ‘can be evident in the nose and taste of wine’, then the ideal level of sulphates will likely be a moderate range rather than any extreme.

Alcohol

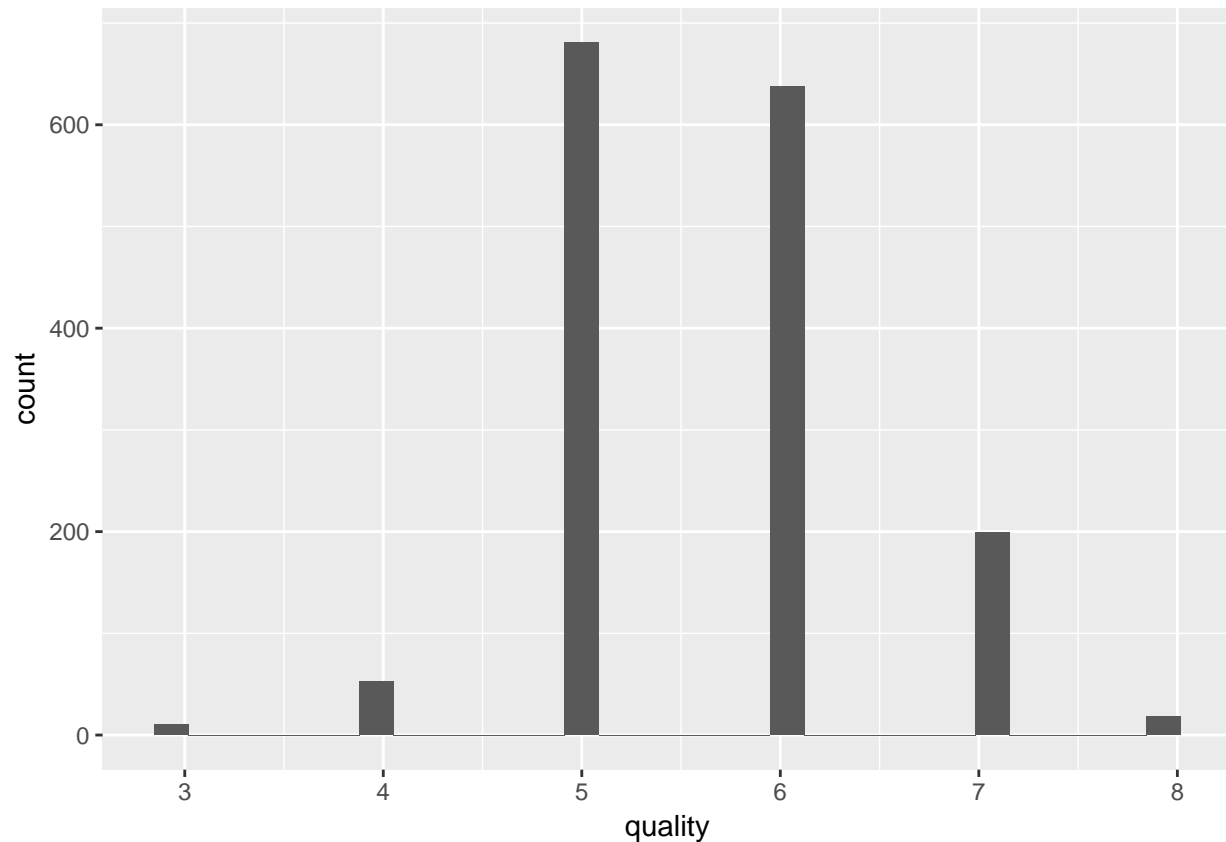


Log10 Transformation



Alcohol varies between 8.4 and 14.9 with most values between 7-10 or 12. Skewed to the right. I don't know if alcohol levels will have a significant impact on quality but I imagine higher levels of alcohol will mean a harsher taste given that more alcoholic drinks tend to be more harsh (to most people and from my own experience). This likely depends on what the judges criteria are for a high quality wine.

Quality



Most wines were either rated at a 5, 6, or 7. Very few wines received a 3, 4, or 8. None really made it to the extremes of the scale. The data actually looks like a normal distribution so the criteria may actually be structured so that most wines fall within the 5-6 region or maybe most wines are just average.

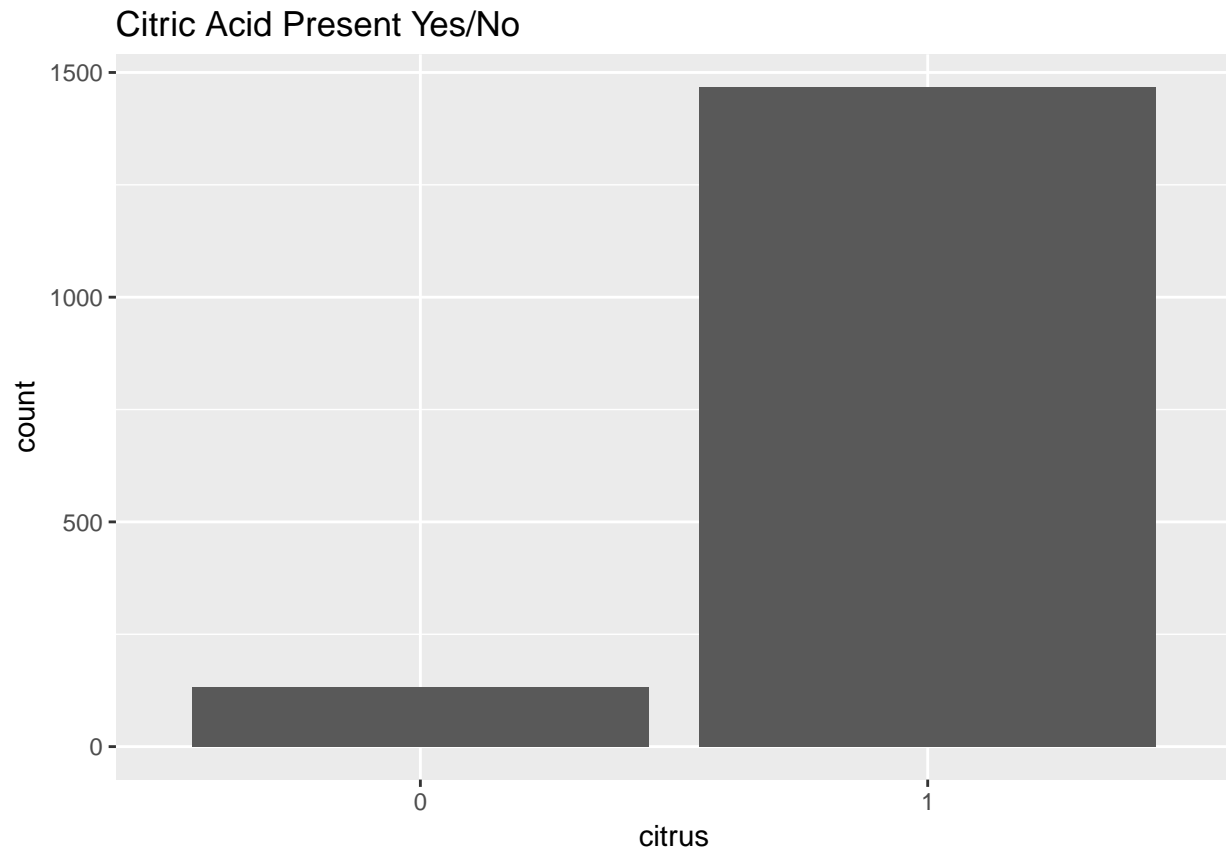
General Comments:

Potential changes to variables:

- Citrus: creating a binary variable of whether or not any citric acid is present
- Putting variables on the same scale. Some variables are in g while others in mg
- Creating ratios: combining fixed and volatile acidity for total acidity or getting a ratio of volatile to fixed
- Cutting variables and creating categories for levels of different variables: highly citrus, salty, highly alcoholic, moderate, highly rated etc.

Citric Acid Presence

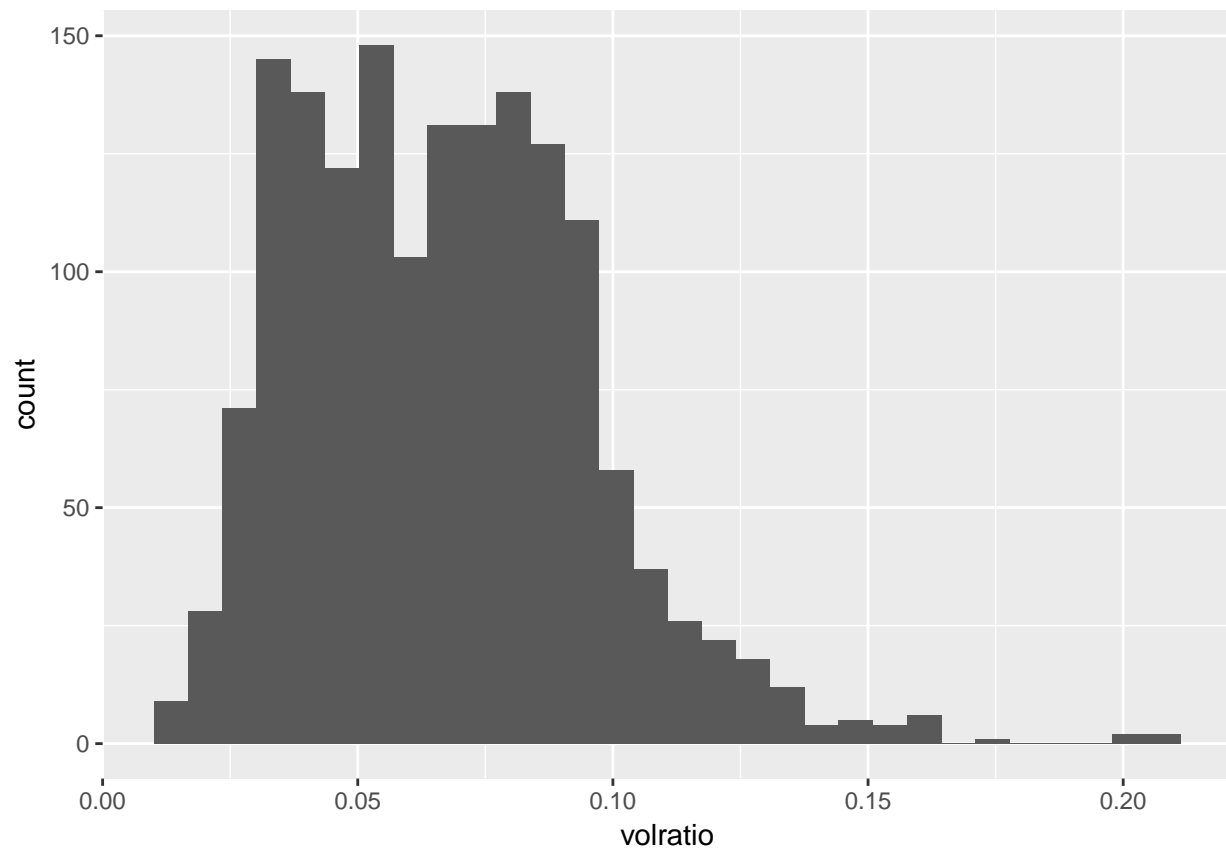
```
##      0      1
## 132 1467
```



I created a binary variable for citric acid presence (yes/no). Most wines have at least some levels of citric acid which is interesting since most wines aren't necessarily advertised as having citrus fruits added. Wine makers may just be adding citric acid for the purpose of 'freshness' rather than any actual citrus flavor. It may be useful to later compare levels of citric acid in groups or separate wines with no citric acid added at all.

% Ratio of Volatile acid to Fixed Acid

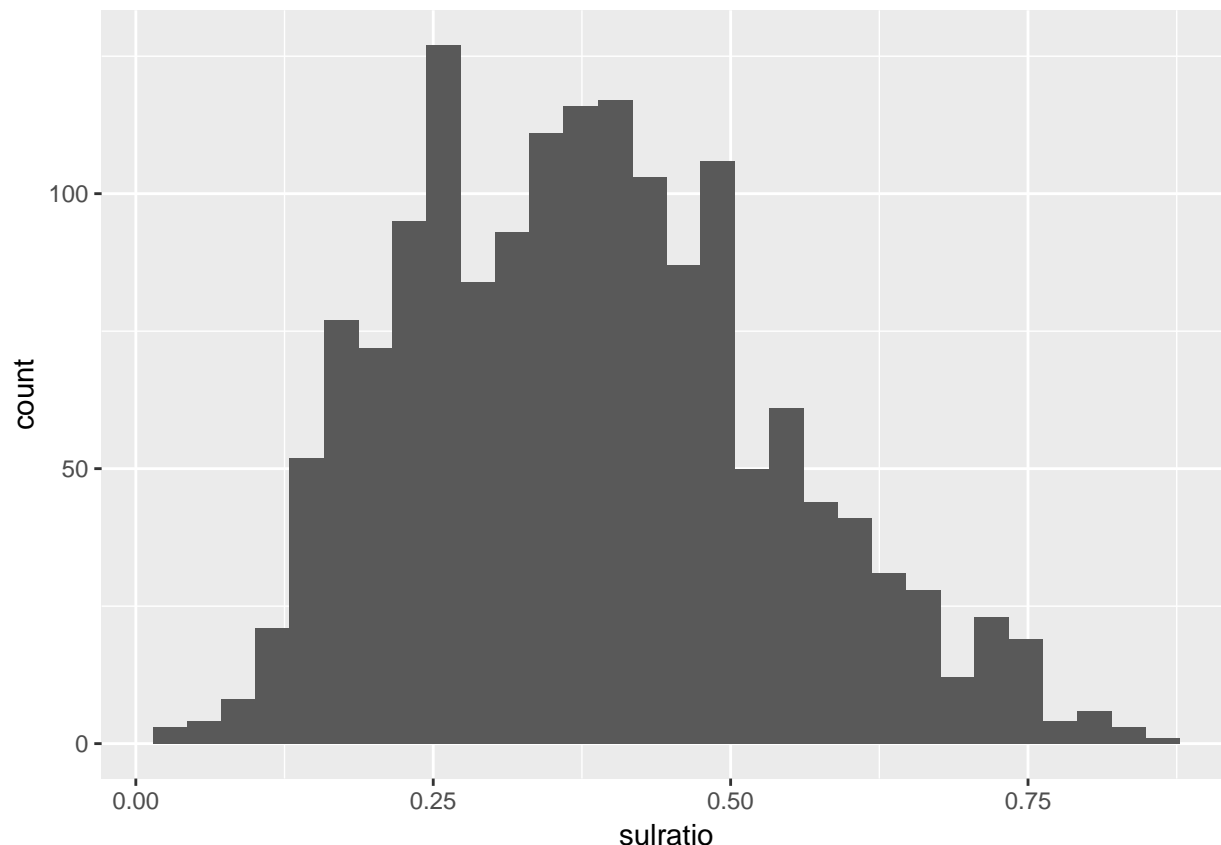
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01348 0.04405 0.06569 0.06706 0.08581 0.20800
```



Volratio is the ratio of volatile to fixed acidity. Most wines fall around the .03-.1 range

% Ratio of Free Sulfur

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.02273 0.25926 0.37500 0.38231 0.48485 0.85714
```



Sulratio is the ratio of free to total sulfur dioxide.

Univariate Analysis

What is the structure of your dataset?

The data-set consists of 12 variables with 1599 observations. Most variables are objective tests and measurements of the wine's chemical/physical makeup but also includes the wine's median quality rating by experts.

- Fixed acidity values fall between 6.5 and 10 with a right skew.
- Volatile acidity seems to have a similar range but on a tenth of the scale and most values lying between .4 and .7.
- Citric acid seems to be an additive as a large number of observations have none at all. Most values are under .5 with a spike at .25. Transformed: Log10
- Residual sugar has a median of 2.2 and most levels between 1 and 3 increasing in small amounts and ranging up to almost 16.
- Chlorides vary on the hundredths with most values between .05 and .1
- Free Sulfur varies widely but most values fall between 7-15
- Total Sulfur varies widely as well with most values between 20-70
- Density is fairly consistent between .99 and 1 but has very precise readings
- PH is measured on a logarithmic scale with most wines falling between 3 and 3.5

- Sulphates have most values between .5 and .75 and is slightly skewed to the right
- Alcohol is more so skewed to the right with most wines between 7-10
- Quality: Most wines are rated a 5, 6, or 7 and none go to the extremes of the 1-10 scale.

What is/are the main feature(s) of interest in your dataset?

The main features of interest are quality, volatile acidity, and total sulfur dioxide.

What other features in the dataset do you think will help support your investigation into your feature(s) of interest?

Other features will likely be alcohol, citric acid, and chlorides.

Did you create any new variables from existing variables in the dataset?

Yes, I chose to create a few new variables:

- Quality2 is the same as Quality but an ordered factor
- Citrus is a binary ordered factor of whether or not a wine has any citrus in it
- Volratio is the ratio of volatile acidity to fixed acidity
- Sulratio is the ratio of free sulfur dioxide to total sulfur dioxide

Of the features you investigated, were there any unusual distributions? Did you perform any operations on the data to tidy, adjust, or change the form of the data? If so, why did you do this?

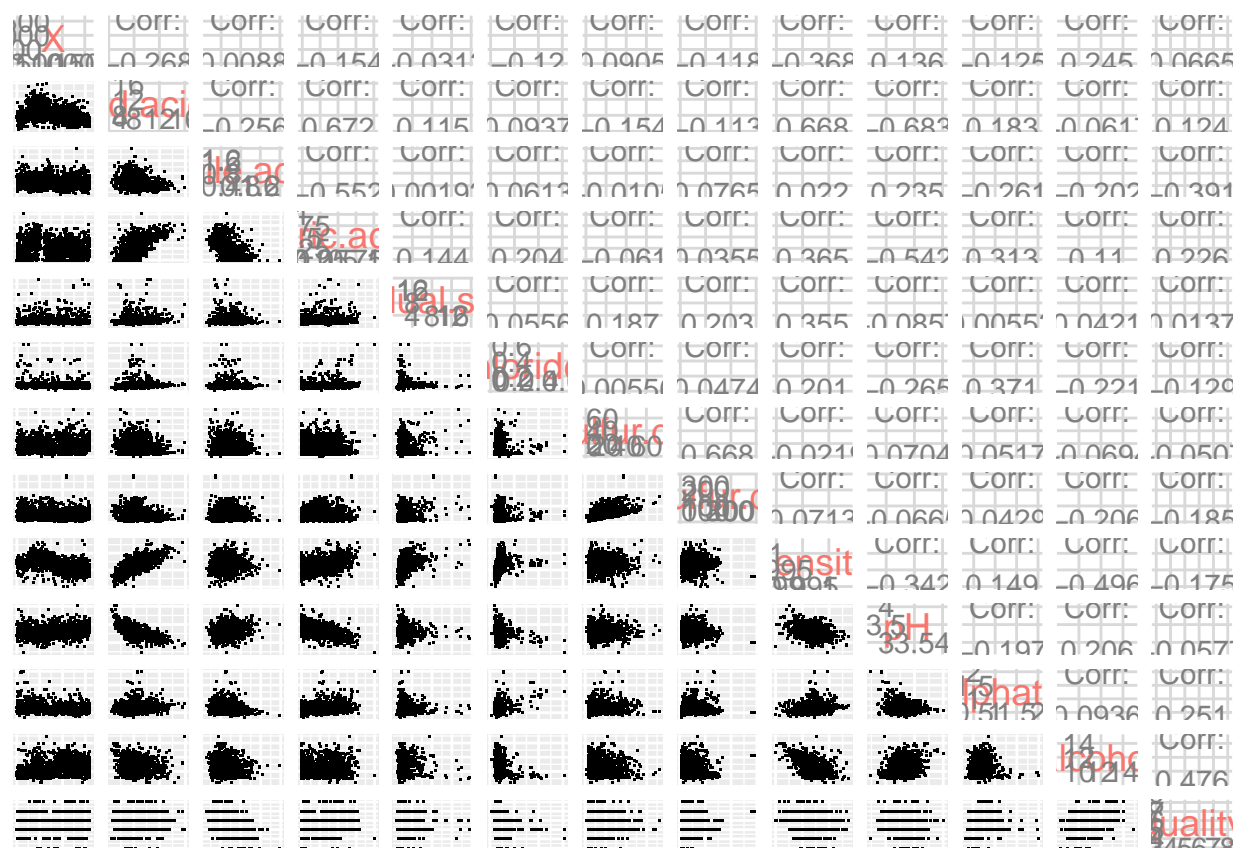
From the univariate analysis, I would think that the variables that vary marginally but have a high specificity may actually be of interest in determining quality of a wine. I see no other reason why they would be measured and reported at such minute levels. I also imagine density and pH may be summary variables for other factors that influence quality.

Bivariate Plots Section

Now that we understand the individual variables a bit better

My hope here is to explore the features of interest from earlier along with most pairs that have at least a moderate correlation.

Corelation Matrices



The above focuses on the original variable set. Now it's quite a bit easier to see a few pairs worth exploring such as:

- -0.552 volatile acidity and citric acid
- 0.672 fixed acid and citric acid
- -0.391 volatile acid and quality
- 0.476 alcohol and quality

Time to explore a bit more...

Quality by Citric Acid Presence

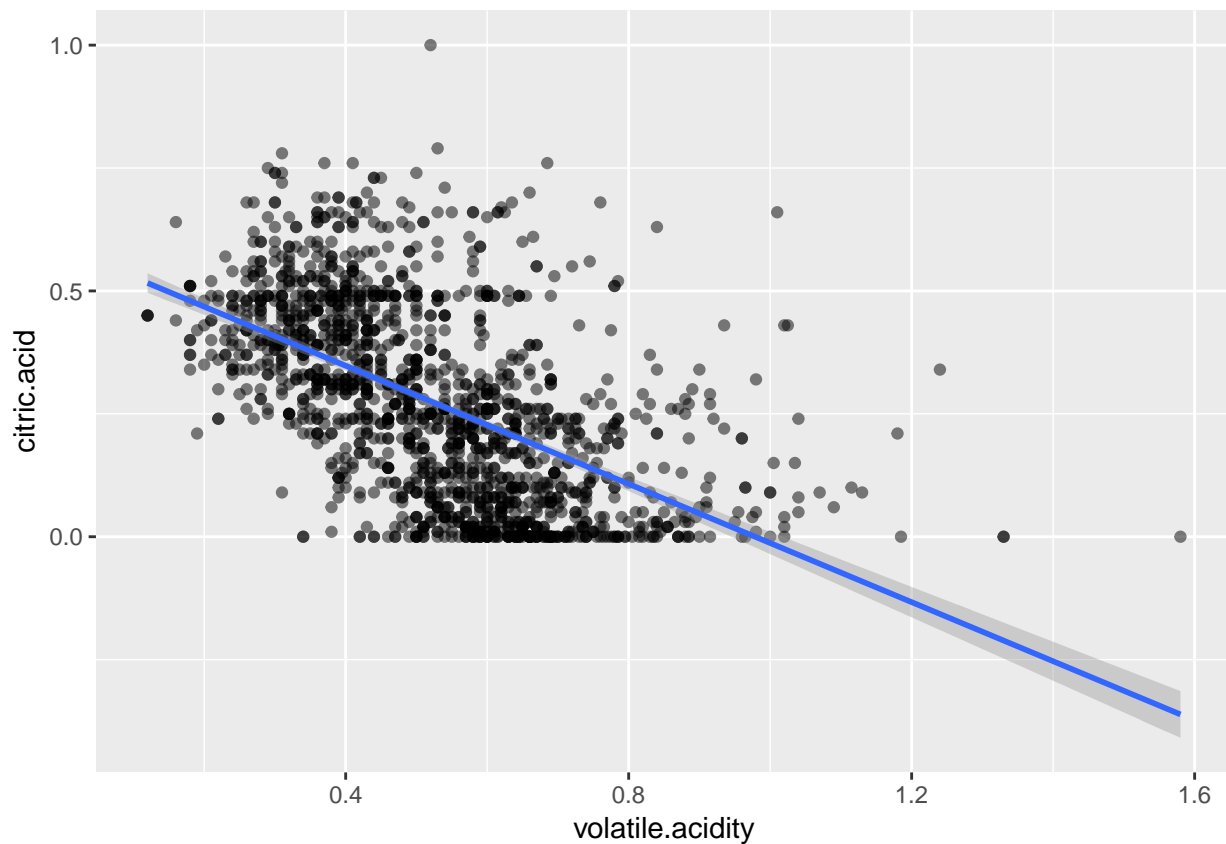
```
## rw$quality: 3
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0050  0.0350  0.1710  0.3275  0.6600
## -----
## rw$quality: 4
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0300  0.0900  0.1742  0.2700  1.0000
## -----
## rw$quality: 5
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0900  0.2300  0.2437  0.3600  0.7900
## -----
```



```
## rw$quality: 6
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0900  0.2600  0.2738  0.4300  0.7800
## -----
## rw$quality: 7
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.3050  0.4000  0.3752  0.4900  0.7600
## -----
## rw$quality: 8
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0300  0.3025  0.4200  0.3911  0.5300  0.7200
```

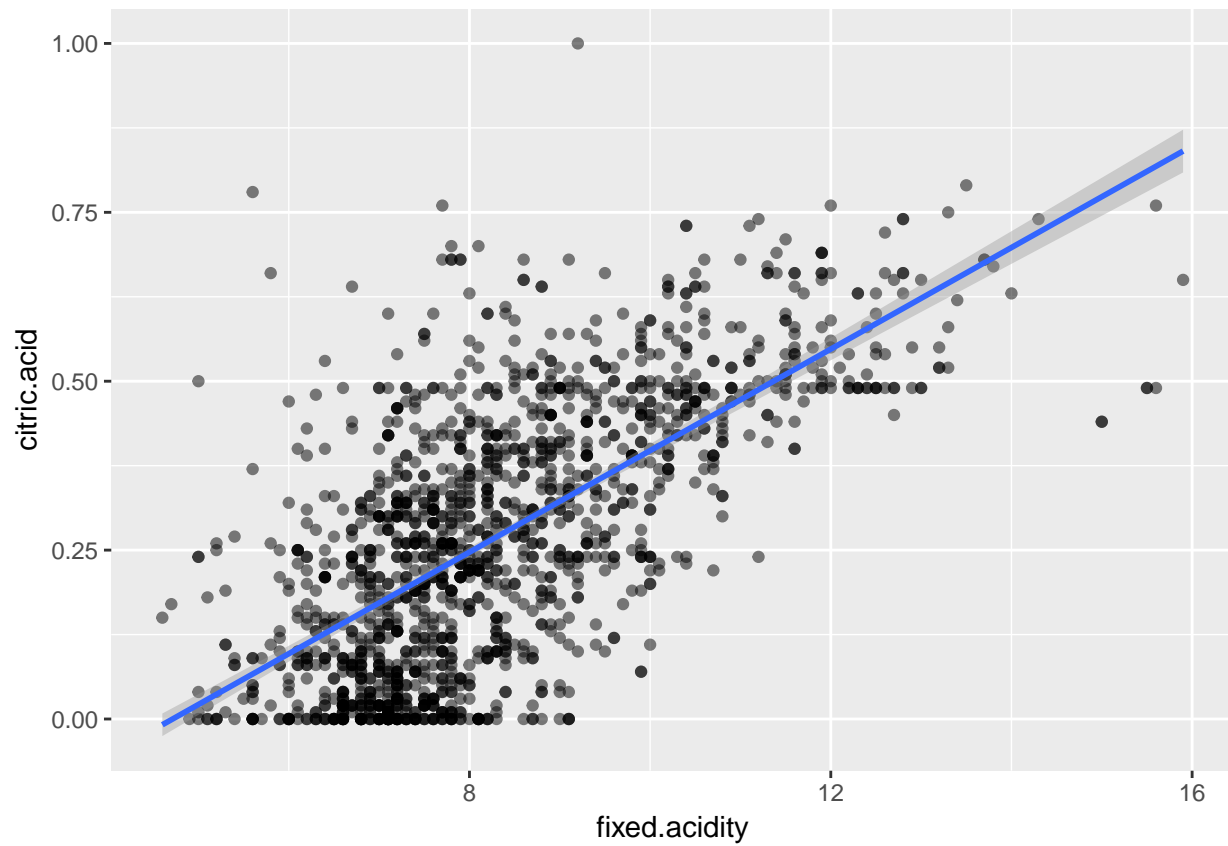
I found it surprising that citrus didn't have a higher correlation with quality, but it still seems of interest as higher quality wines have higher amounts of citric acid on average. In fact, the average value of citric acidity for the highest quality wines is over double that of the lowest quality wines.

Volatile Acidity and Citric Acid



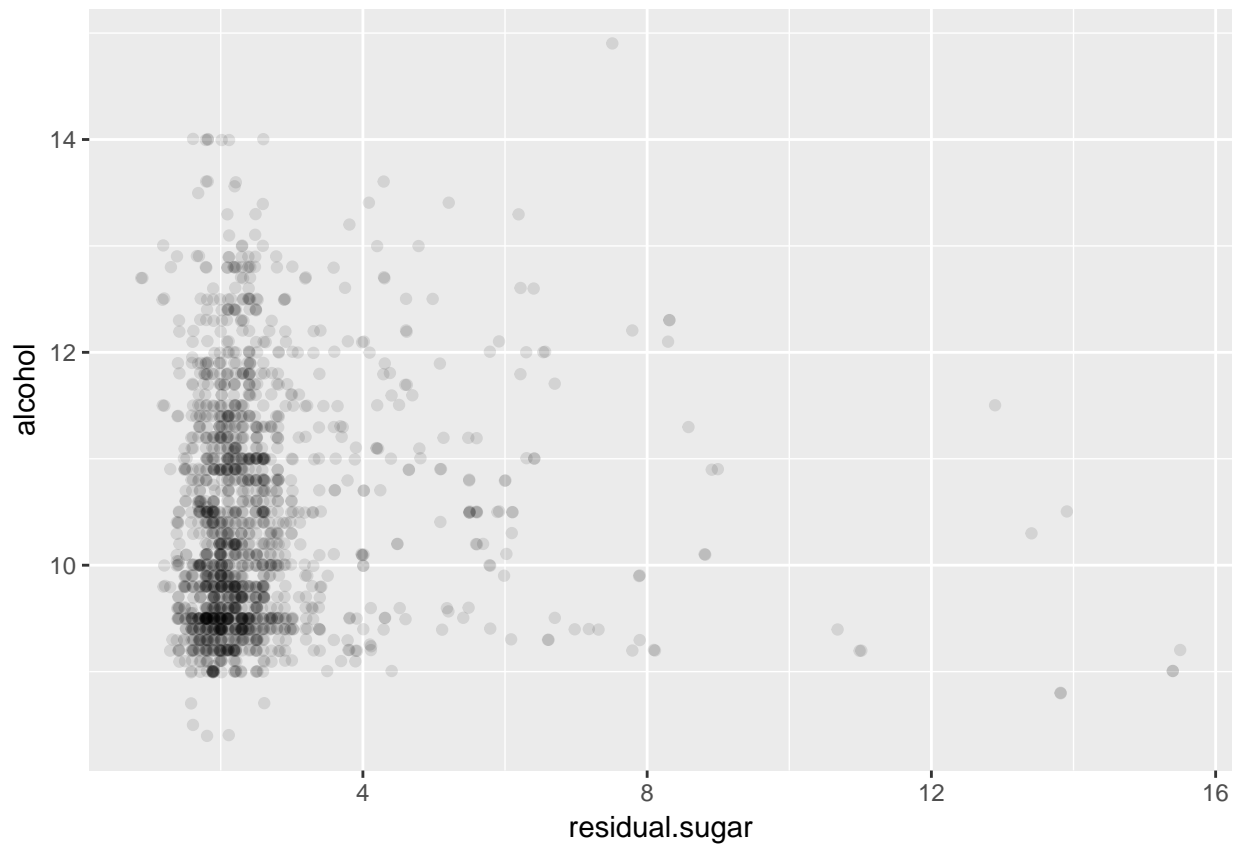
Volatile acidity and citric acid have a moderate, negative correlation (-.552).

Fixed acidity and Citric Acid

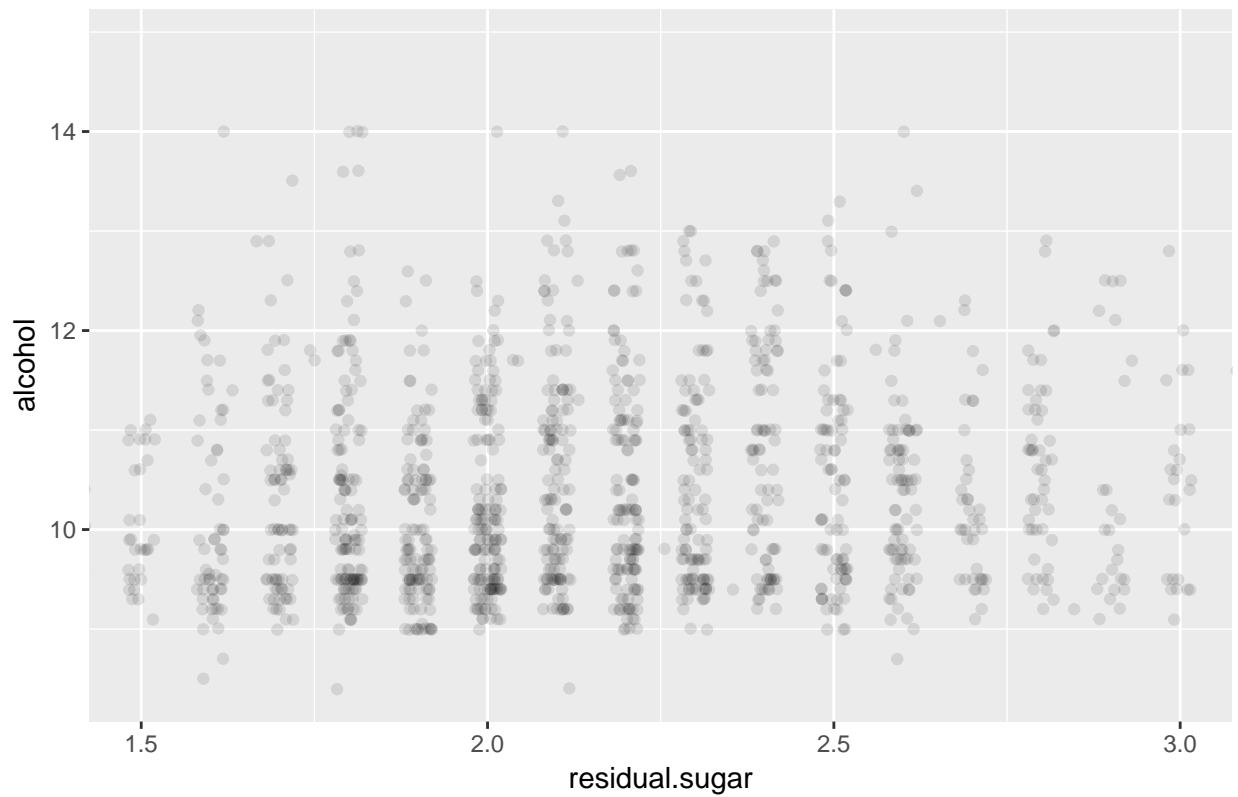


Fixed acidity and citric acid are positively correlated (.672) which is somewhat surprising since I would think Fixed acidity and Volatile acidity would correlate well with each other (and at least correlate in the same direction with Citric Acid)

Residual Sugar and Alcohol

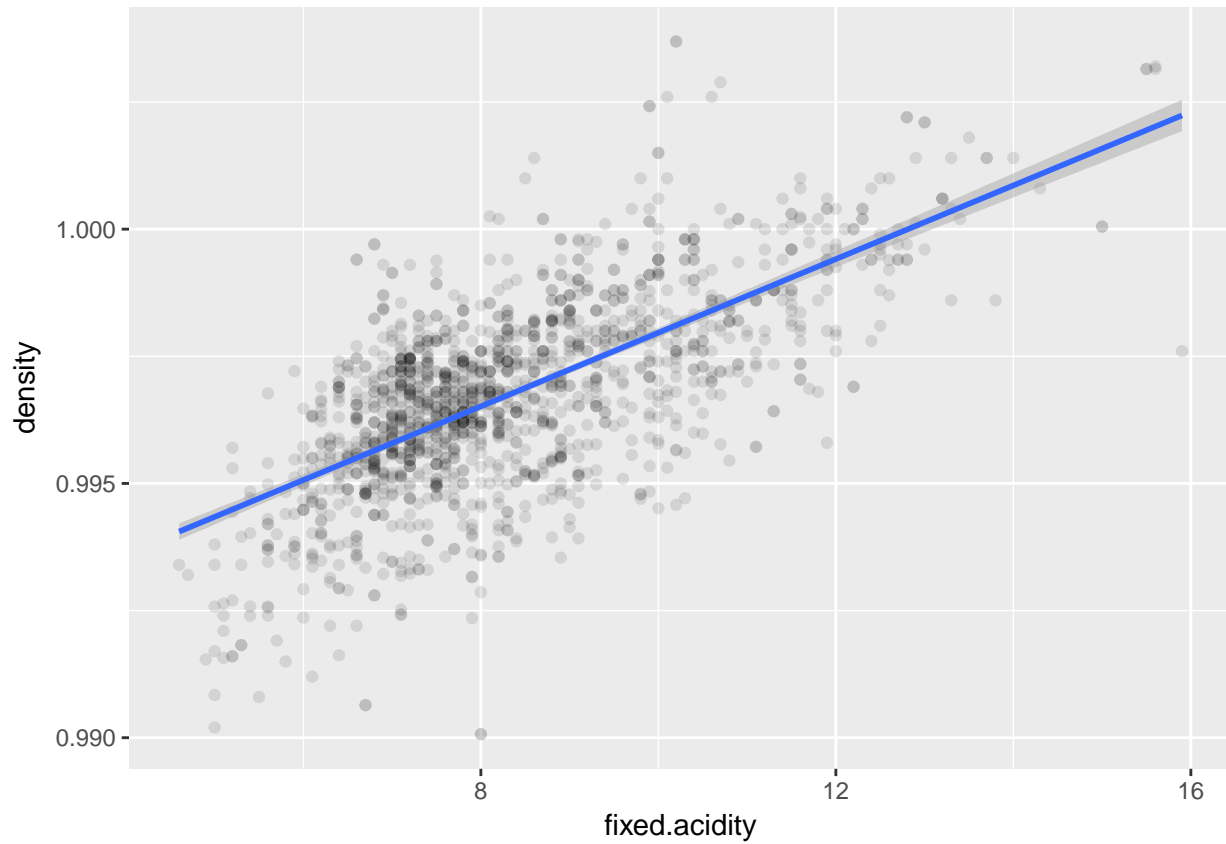


Zooming in

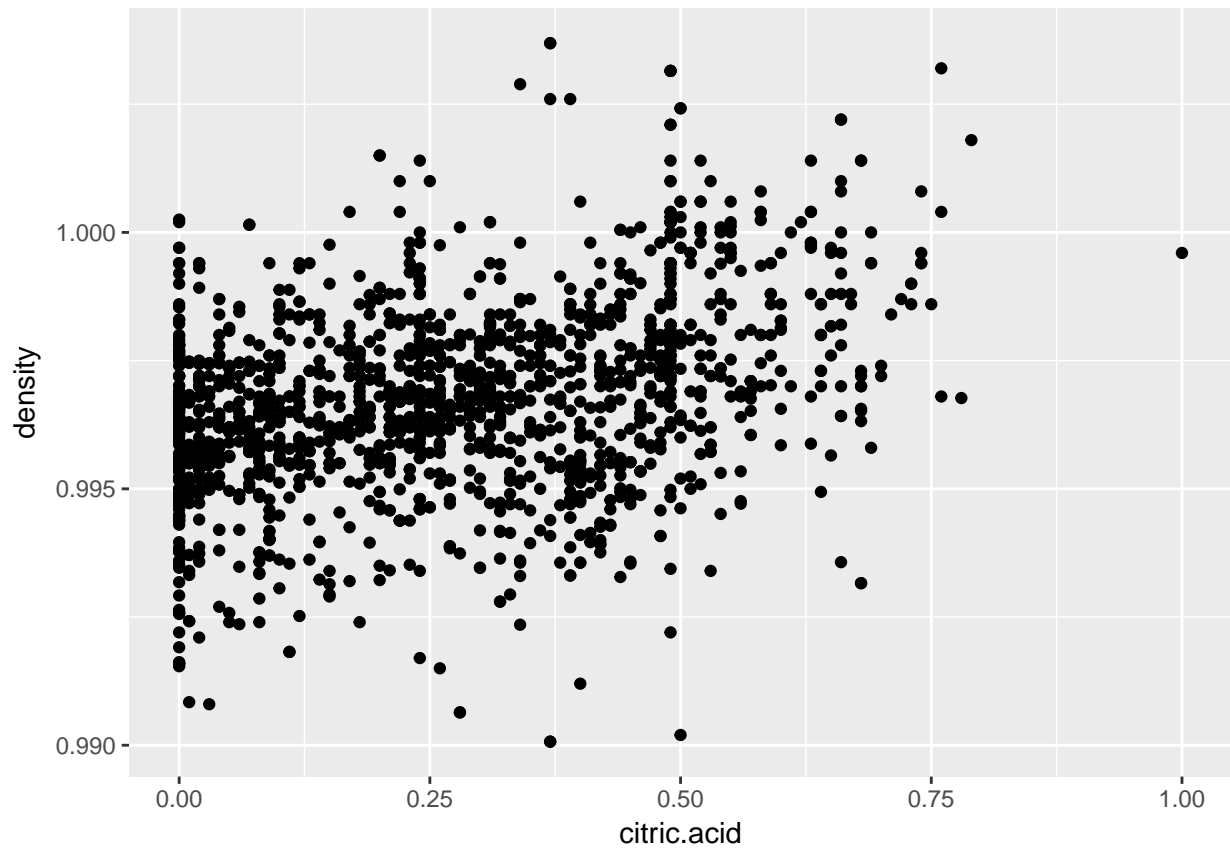


Alcohol and residual sugar have a correlation of .0421 which is unexpectedly low since sugar is a precursor to alcohol. After graphing and limiting the axes there are still vertical bands across the same x value with no clear relationship.

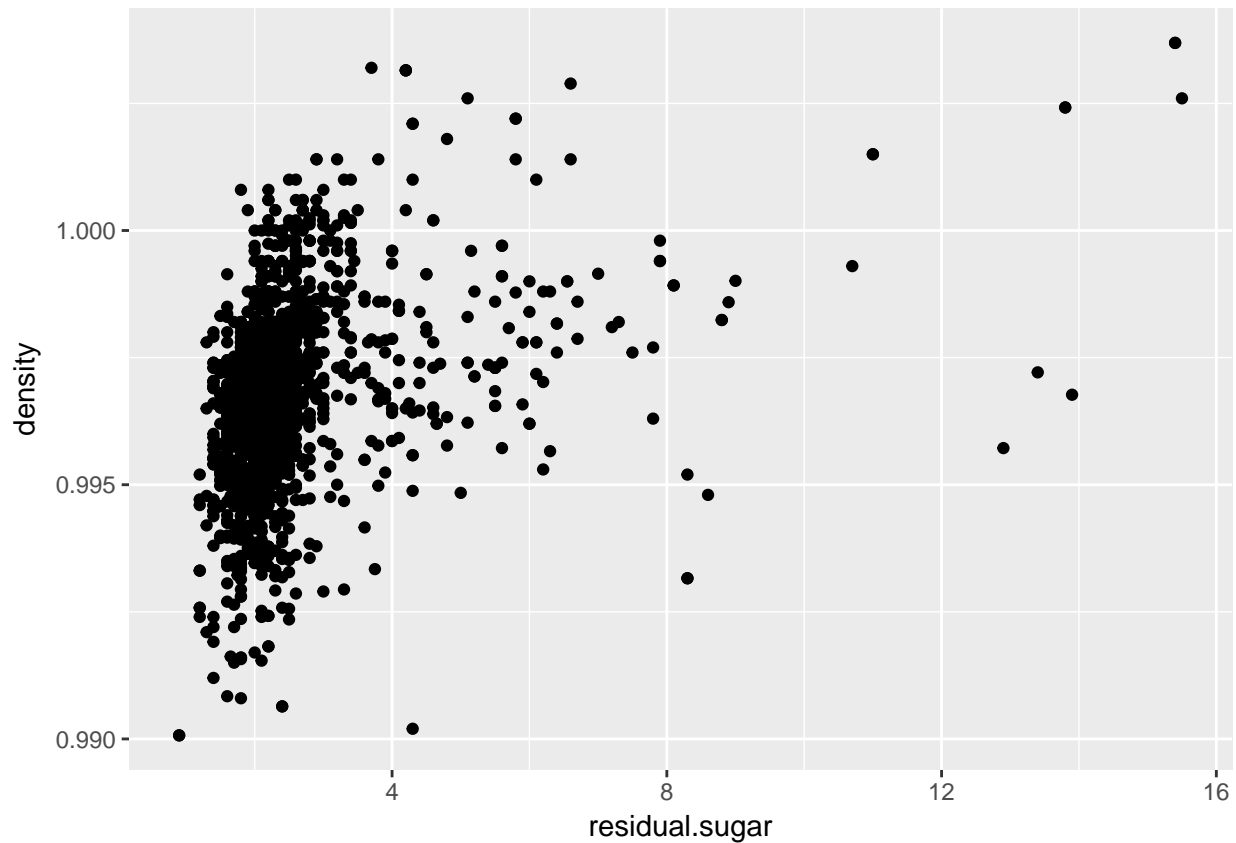
Density correlations



Fixed acidity and density are strongly correlated (.668)



Citric acid and density are lightly correlated .365

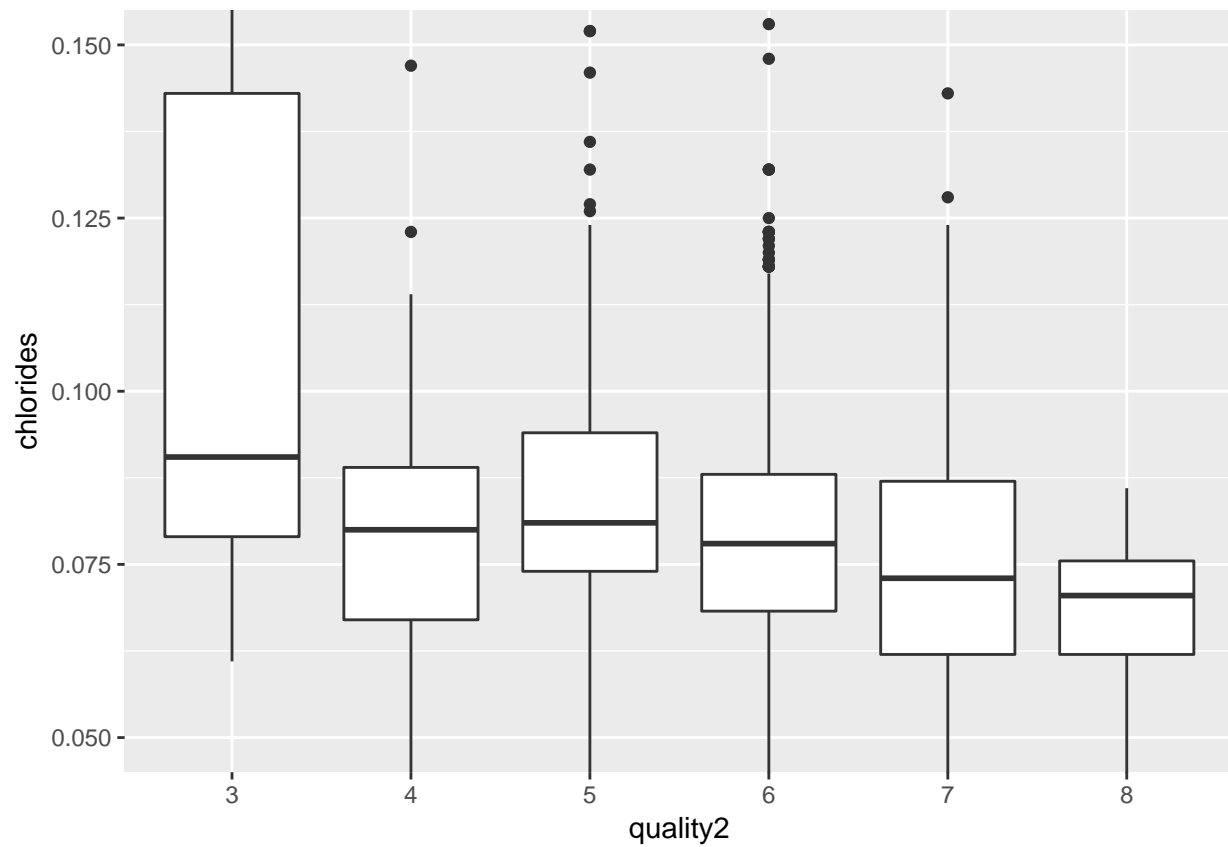


Residual sugar and density are lightly correlated .355

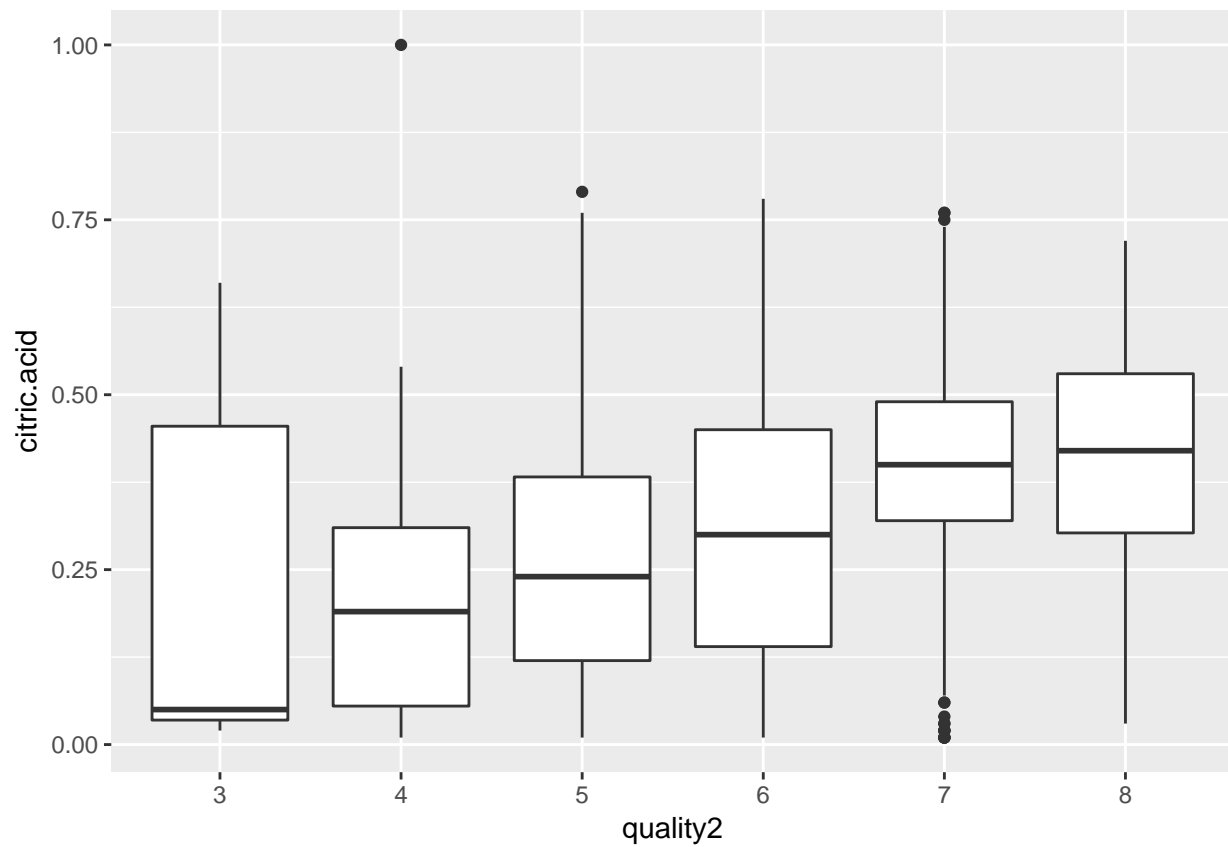
Density is positively correlated with fixed acidity, citric acidity, and residual sugar. Density may prove a useful variable in analysis for summarizing the above variables.

Quality plots

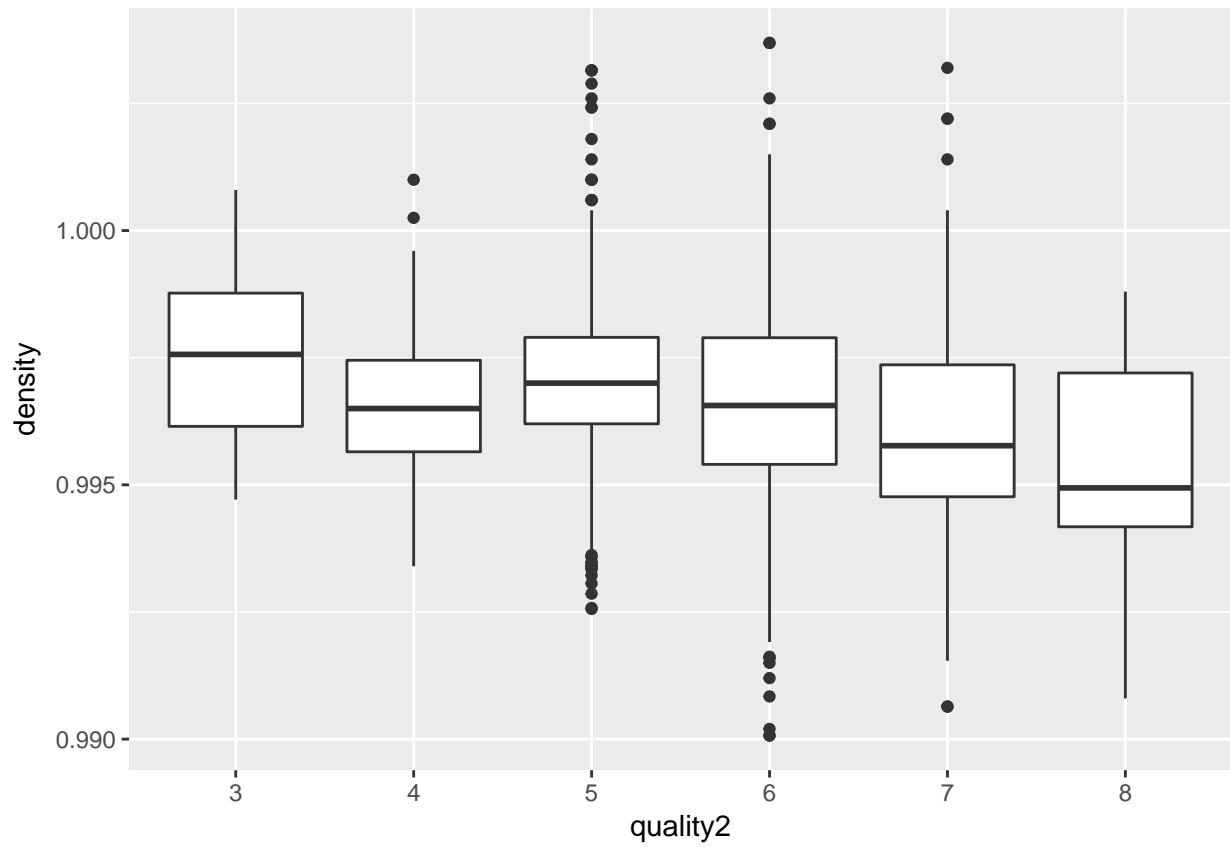
Now, let's begin exploring the relationship of a few variables with our key outcome variable: Quality. Please note, Quality2 is the same information as our original quality variable but is an ordered factor rather than an integer.



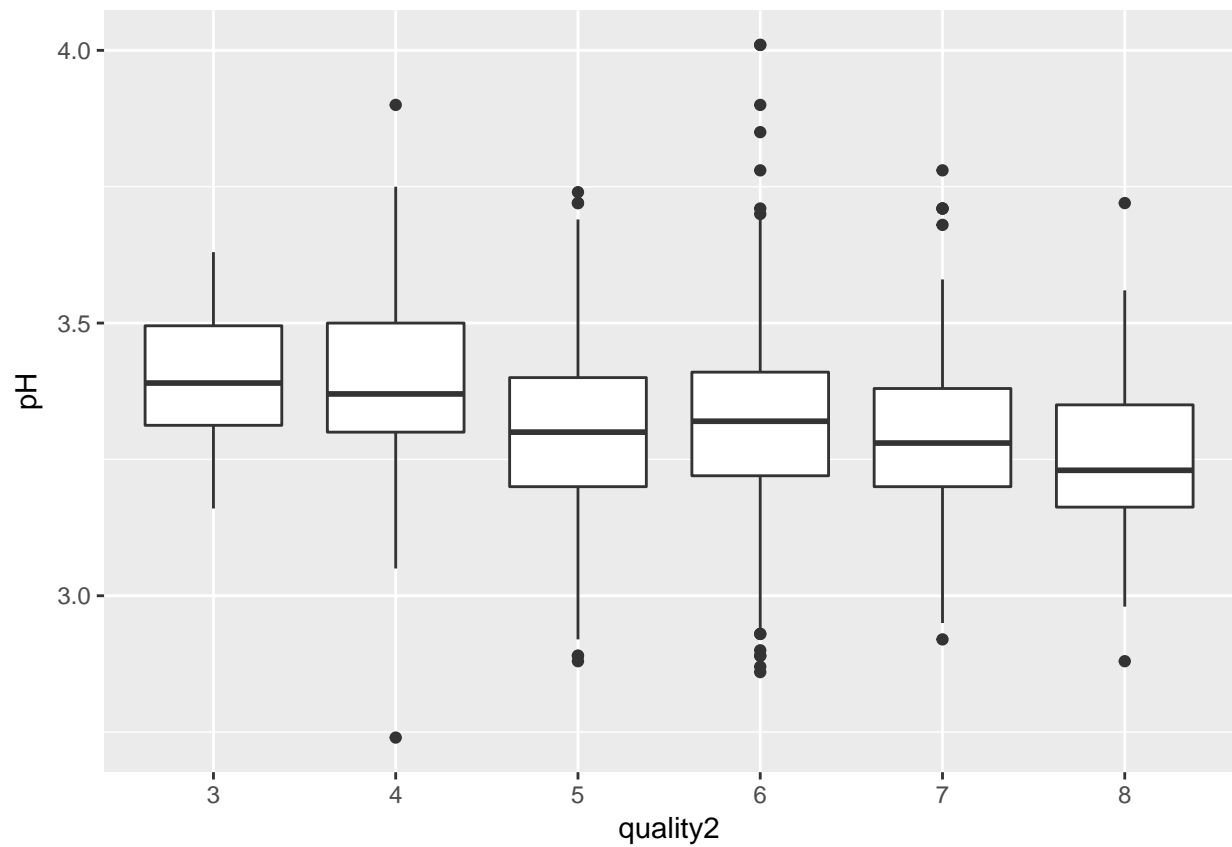
Higher quality wines tend to have a marginally lower median chloride level. Lower quality wines have a much larger range for chloride levels than higher quality wines.



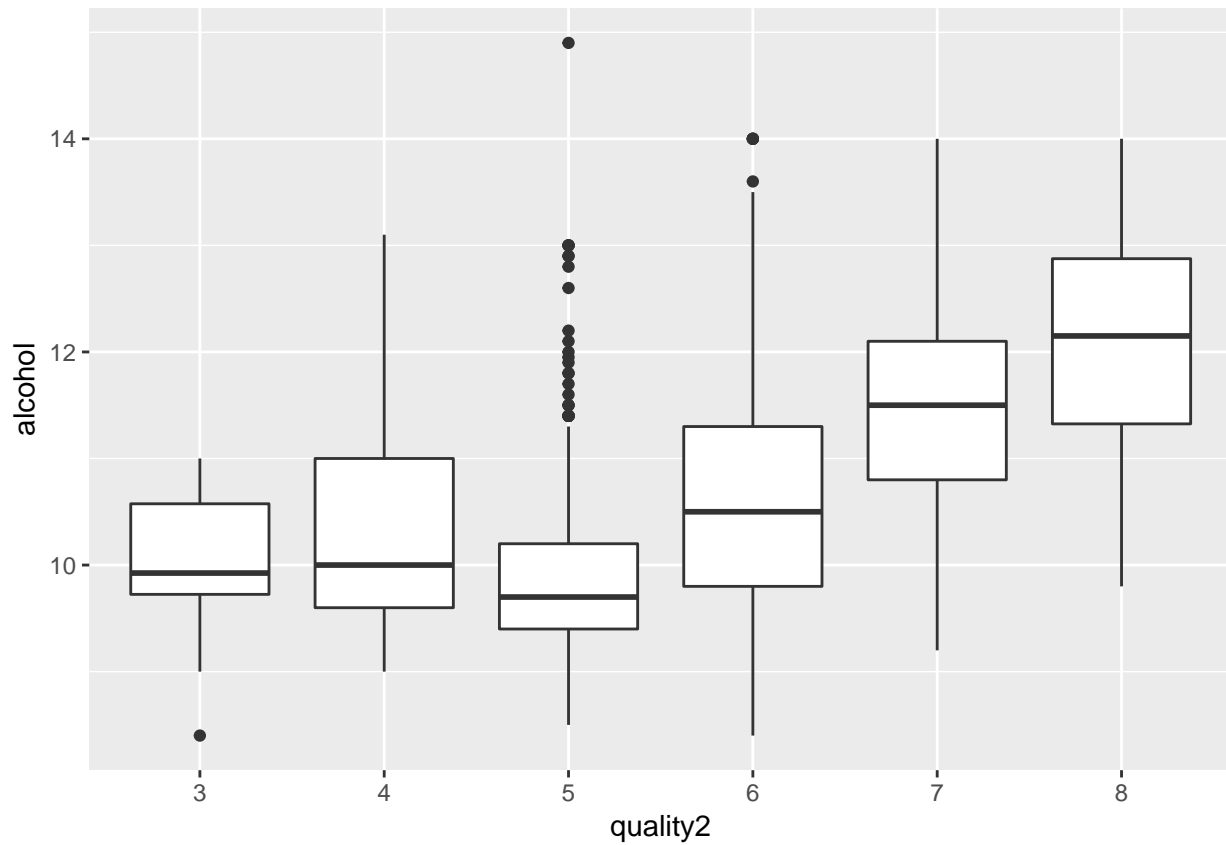
Of the wines with citric acid, wines of higher quality have a higher median citric acid value. It seems the lowest quality wines have a very large range comparatively.



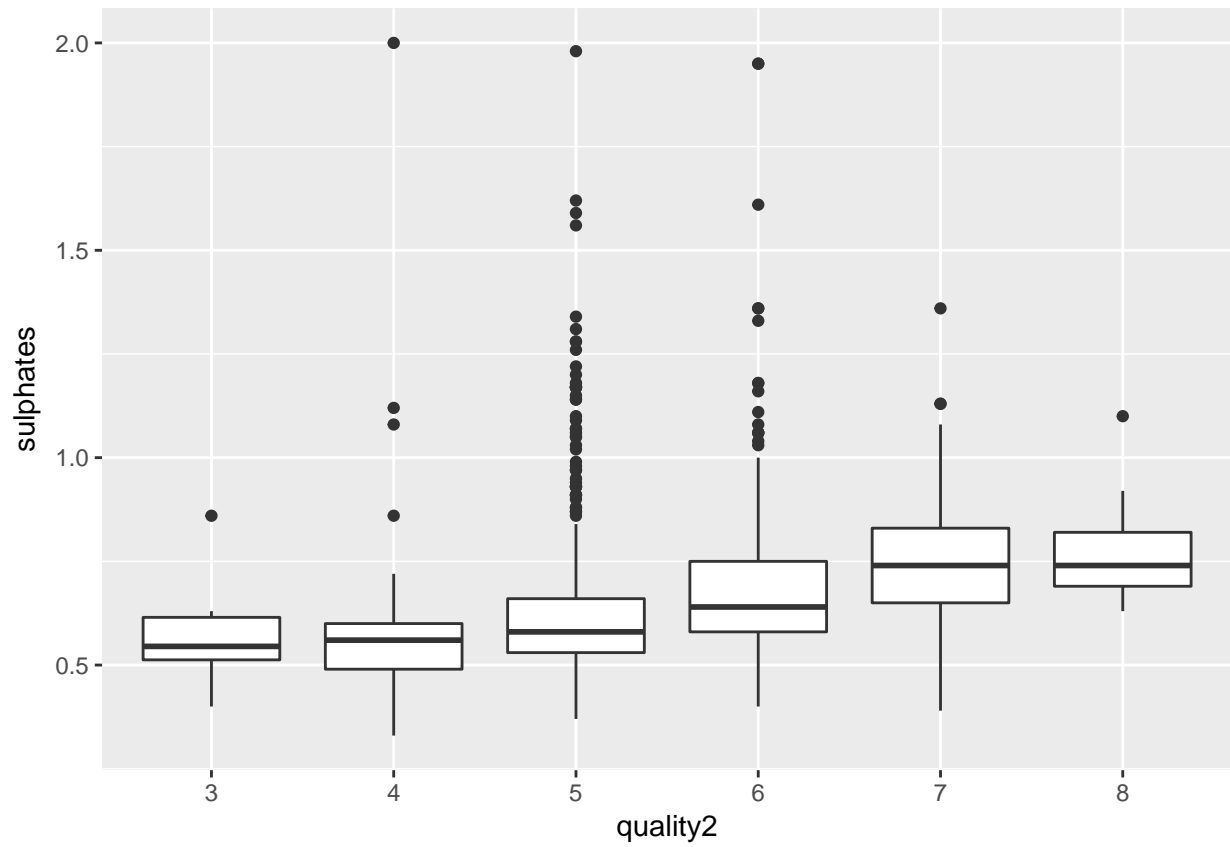
Higher quality wines had the lowest median values for density and oddly enough, the highest range which is different from our previous graphs.



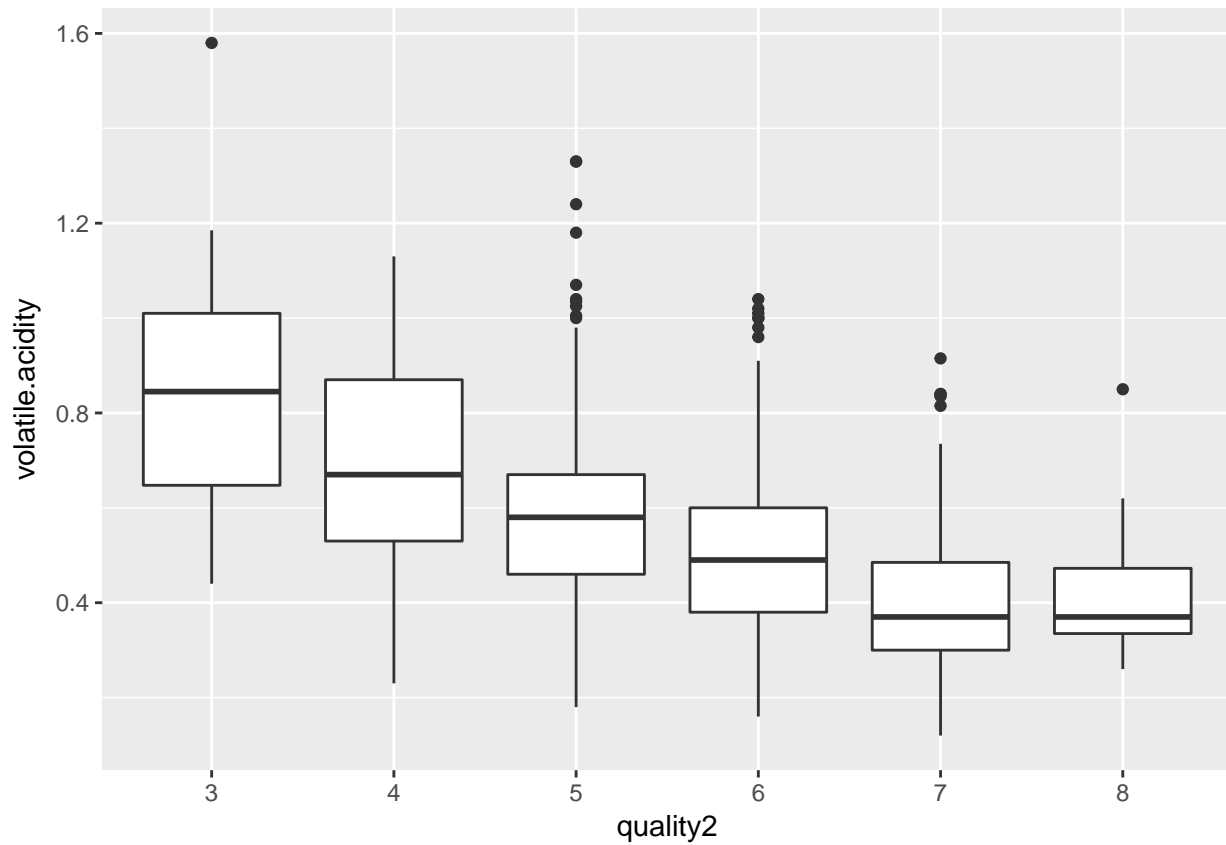
Higher quality wines have a marginally lower pH. Overall, most wines seem to have a fairly tight range for pH.



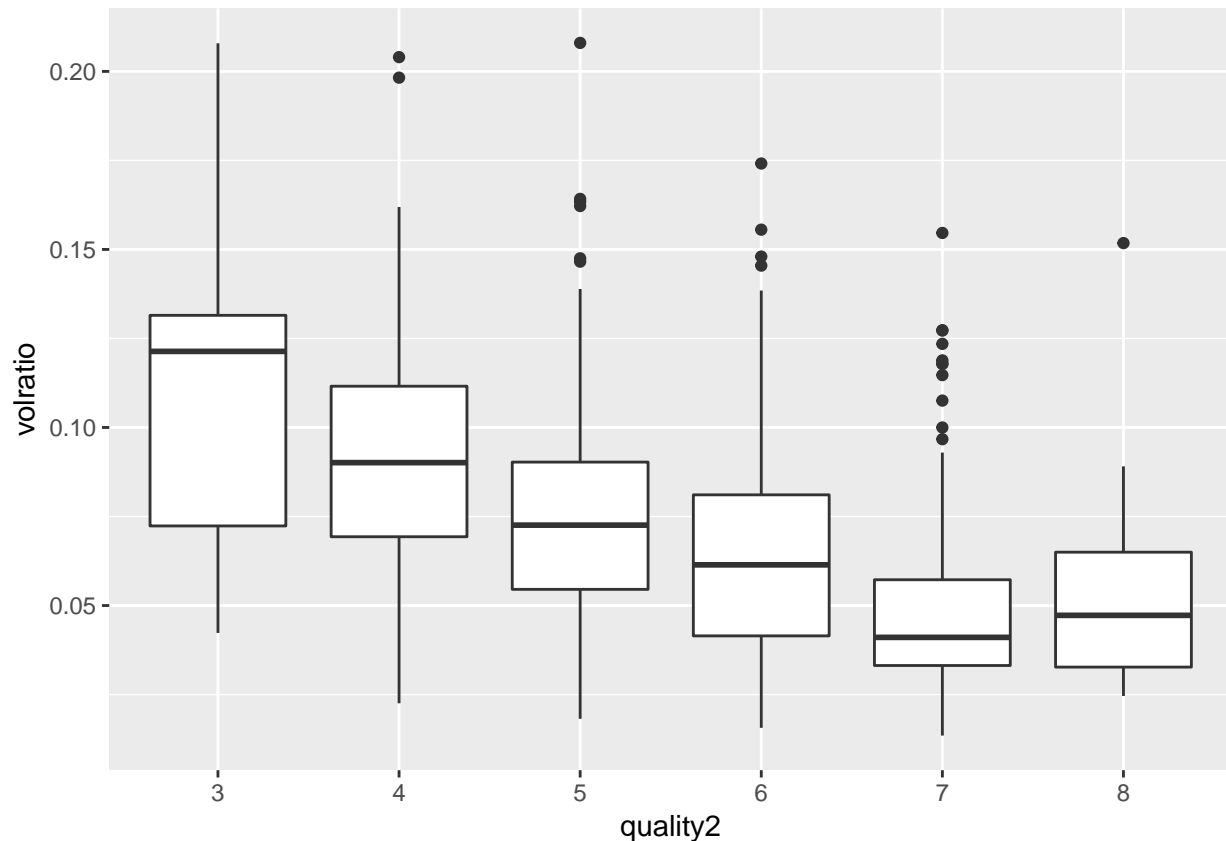
From earlier, we know that alcohol and quality are positively correlated. Here, we can clearly see this relationship. Also of note, lower quality wines have a smaller range than higher quality wines.



Sulphates and quality are also slightly positively correlated (.251). This is easier to see when breaking up the median value of sulphates across quality scores.



Volatile acidity and quality are negatively correlated (-.391). Wines with a higher quality score have lower median values of volatile acidity. No surprises here as this 'at too high of levels can lead to an unpleasant, vinegar taste'.



As expected, higher quality wines have a lower median value for their volatility ratio (volatile acidity/fixed acidity).

Bivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. How did the feature(s) of interest vary with other features in the dataset?

Wines with citric acid on average are rated higher. Citric acid correlates negatively with volatile acid, but positive with fixed acid. Citric acid may contribute to fixed acid levels and not volatile acids.

On average, higher rated wines had a higher median and mean alcohol and sulphate levels. Wines at the outer ends of the quality ratings had higher amounts of total sulfur dioxide than those with moderate quality ratings.

As expected, on average, higher quality wines had lower levels of volatile acidity than lower quality wines.

Did you observe any interesting relationships between the other features (not the main feature(s) of interest)?

- Residual sugar did not have a strong correlation with alcohol.
- Density is positively correlated with fixed acid, citric acid, and residual sugar. Density may prove a useful variable in analysis for summarizing the above variables.

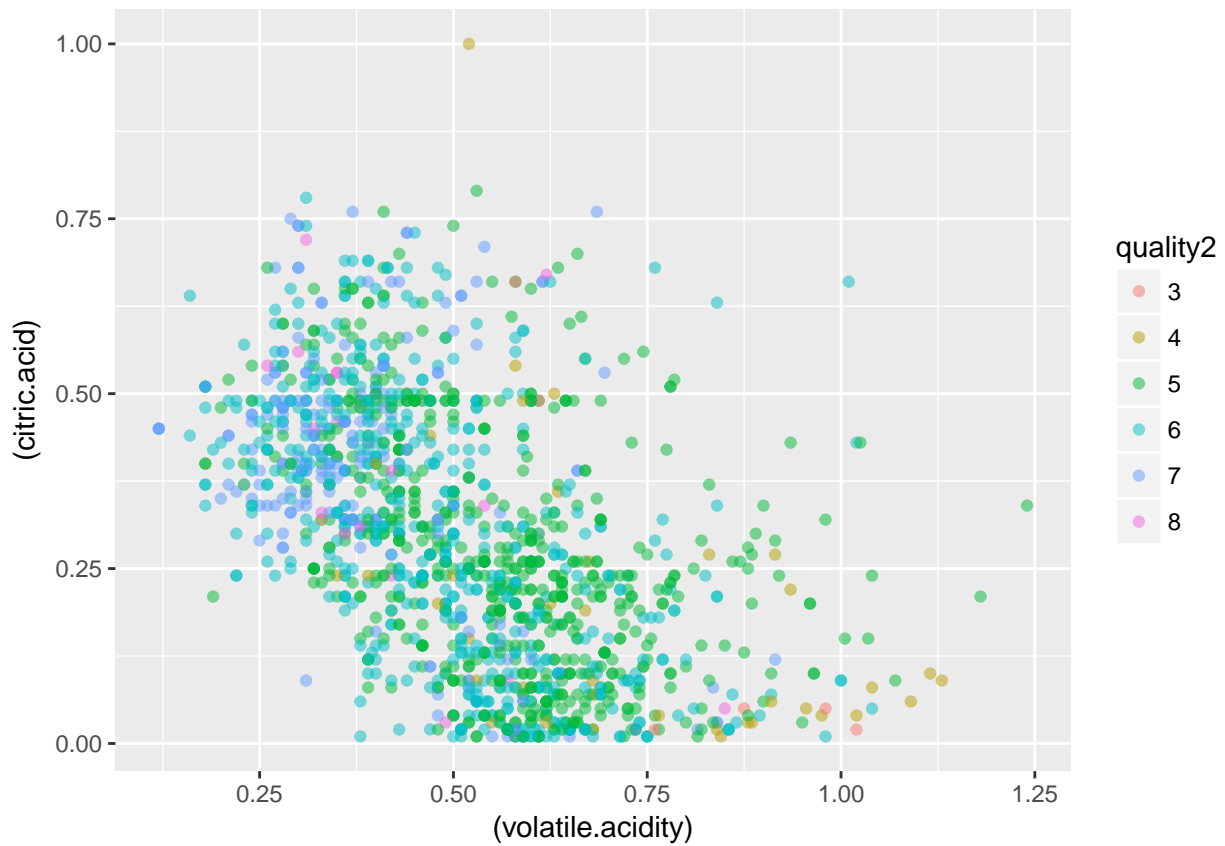
What was the strongest relationship you found?

Fixed acidity and pH had the strongest correlation at $-.683$. This makes sense as the more acidic, the lower in pH a substance would be.

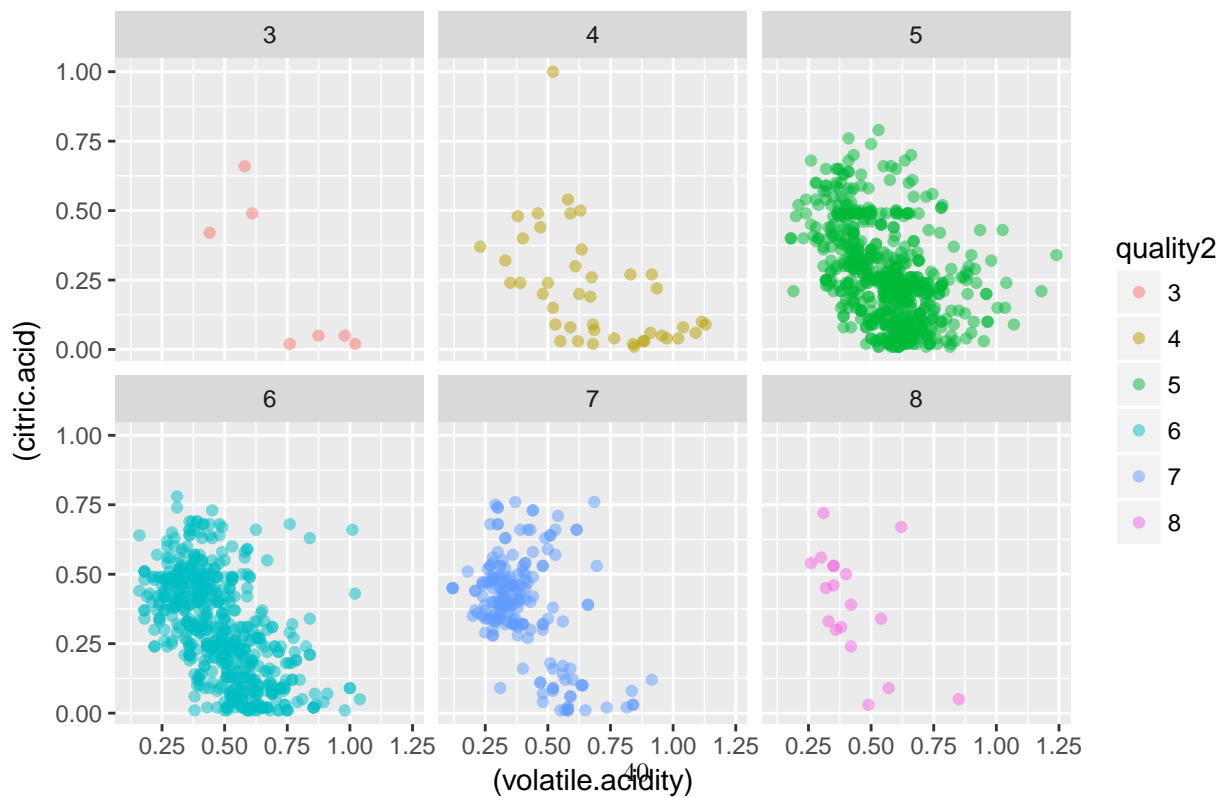
Aside from free and total sulfur dioxide having a $.668$ correlation, fixed acidity and density also had a $.668$ correlation with fixed acidity and citric acid having a $.672$ correlation.

Multivariate Plots Section

Volatile Acidity and Citric Acid

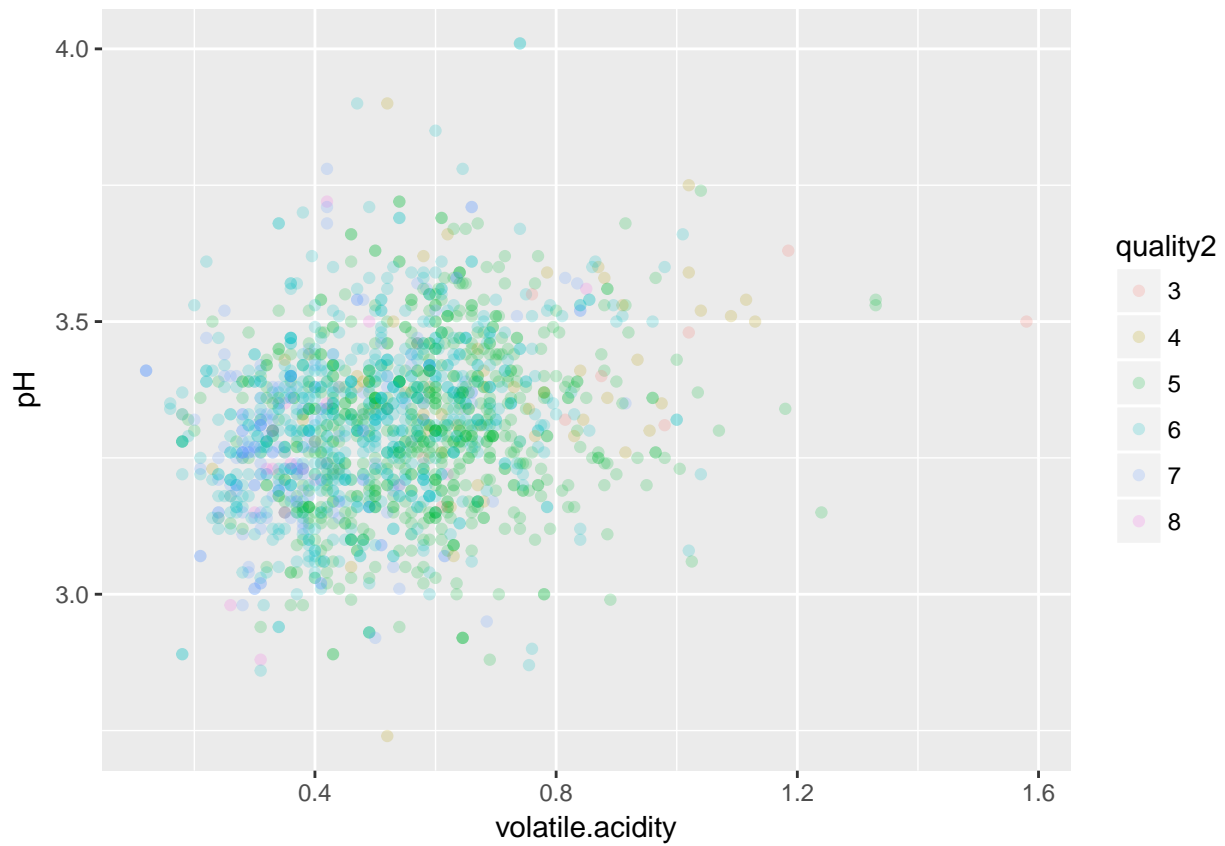


Volatile and Citric Acid by Quality

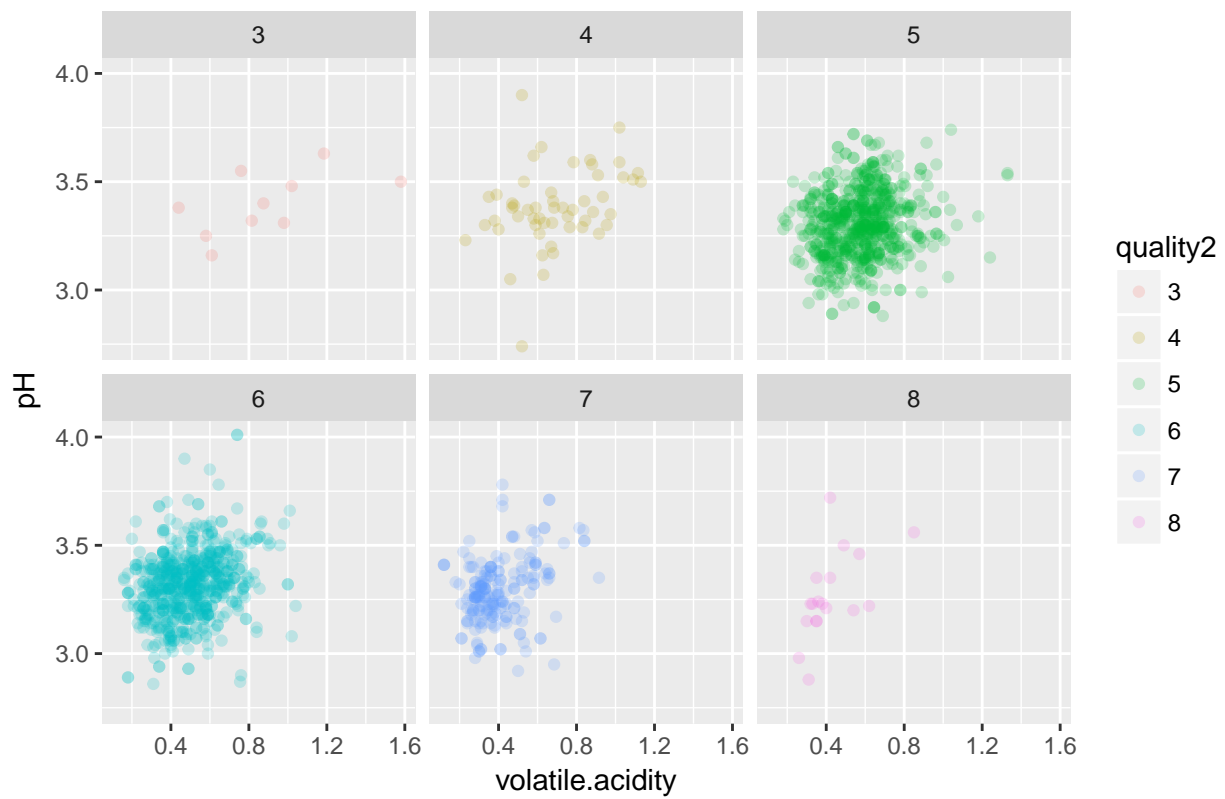


As expected, higher quality wines tend to have low volatile acidity and higher amounts of citric acid.

Volatile Acidity and pH

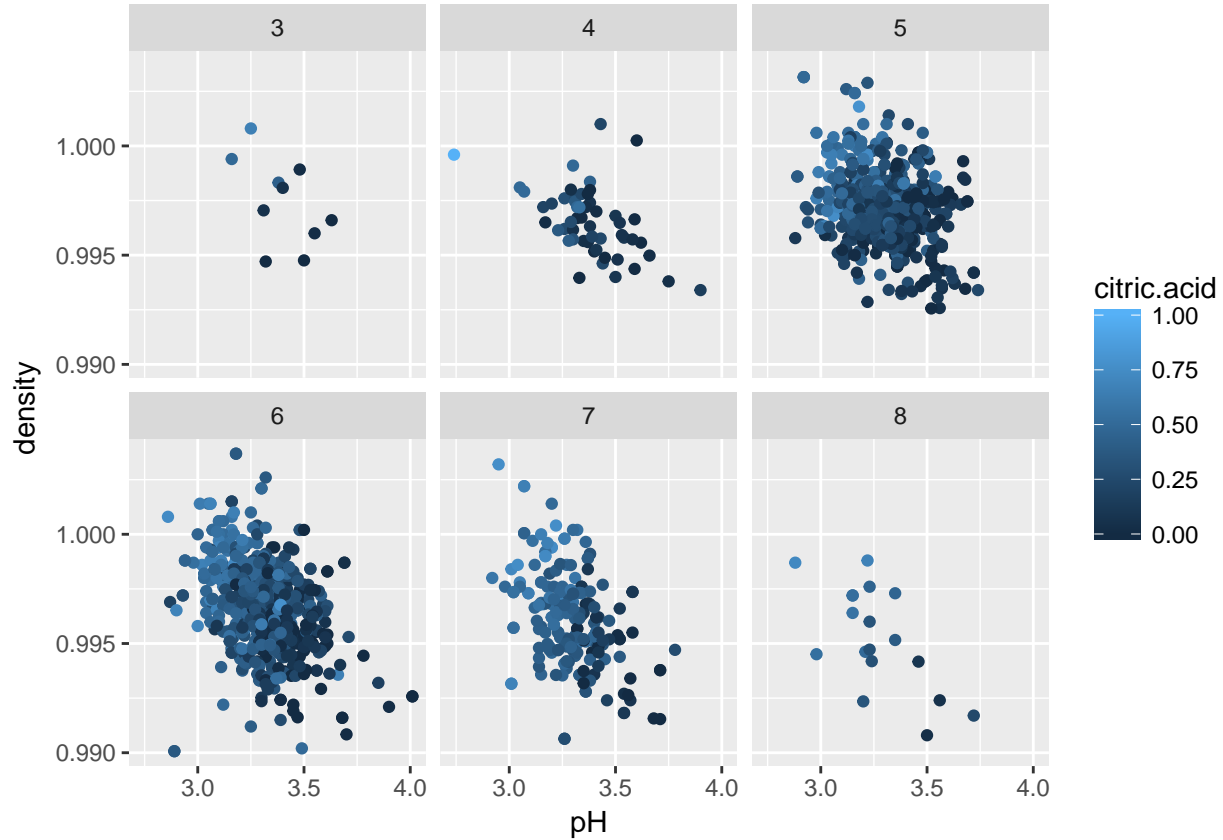


Volatile Acidity and pH by Quality



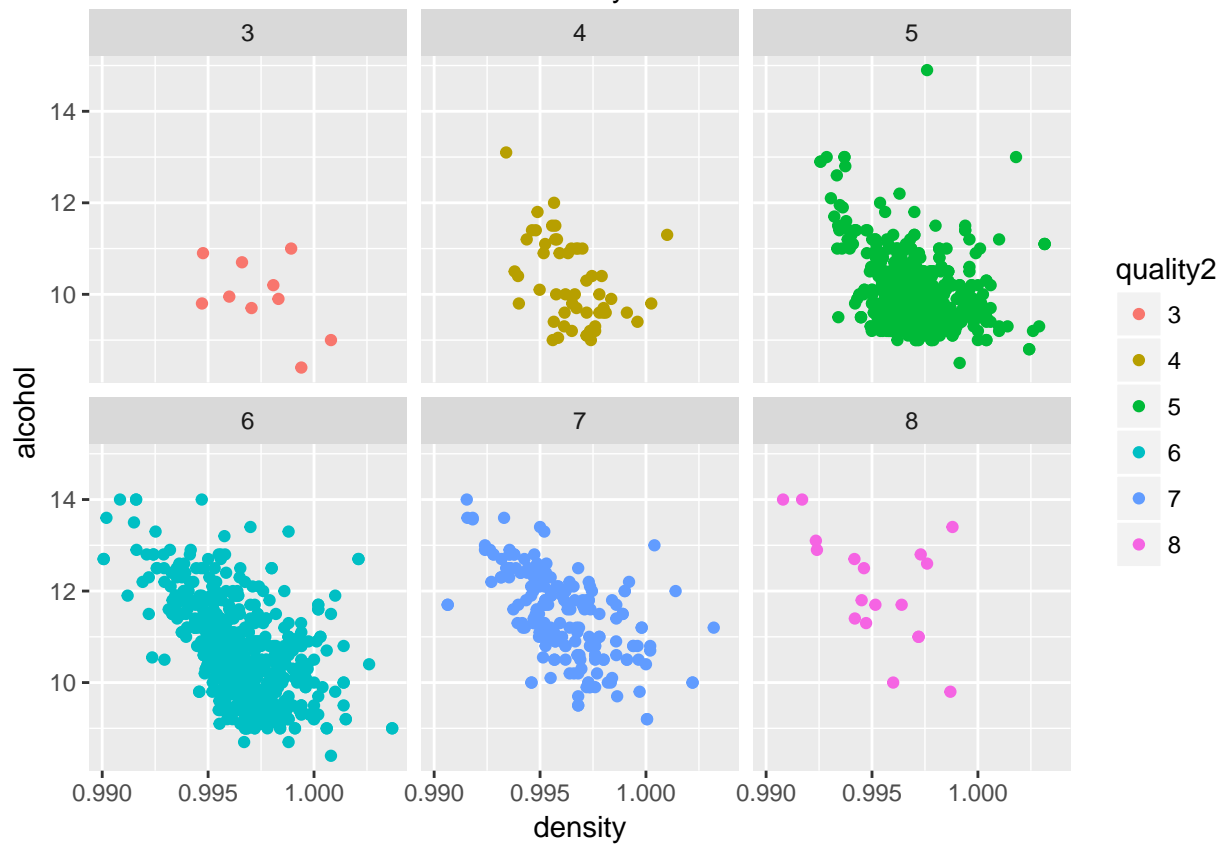
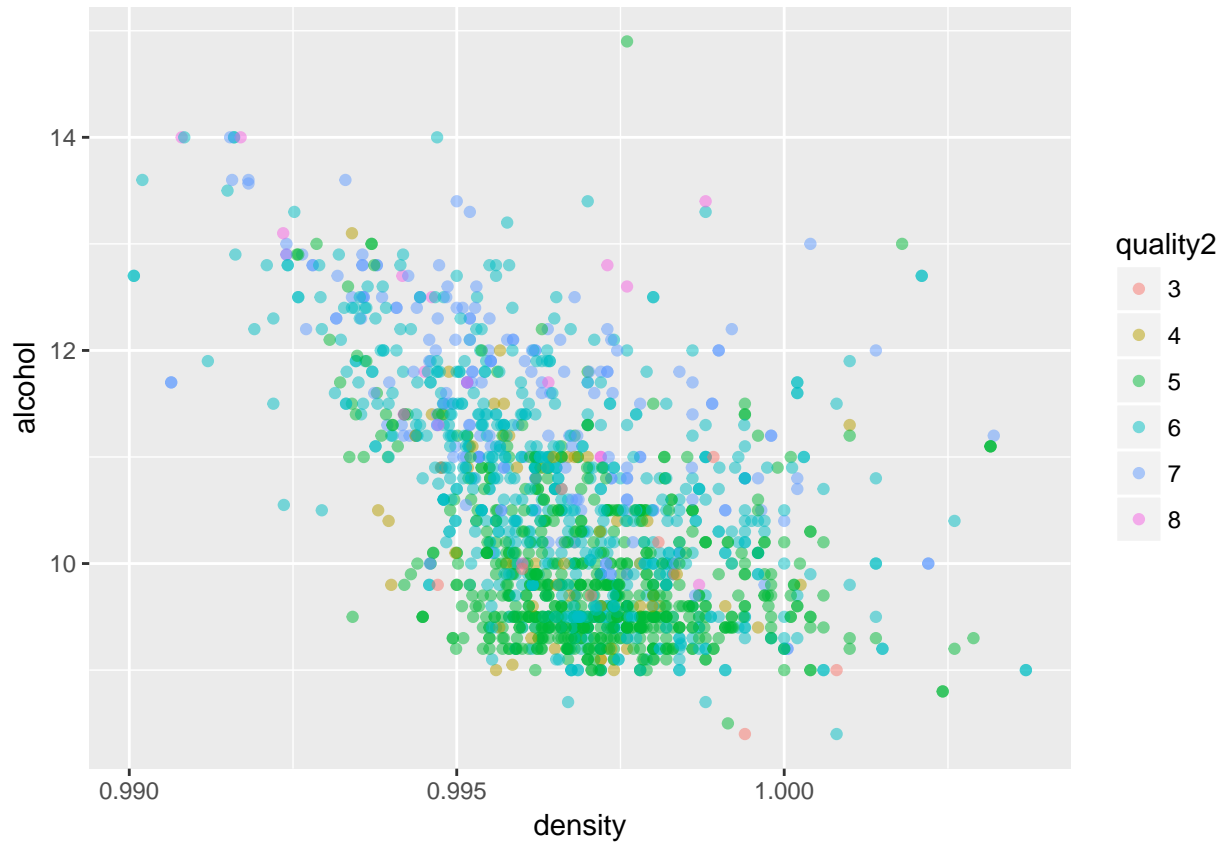
Higher rated wines tend to have lower volatile acidity and pH. This relationship isn't too strong but was worth exploring as summary variables.

Density and pH

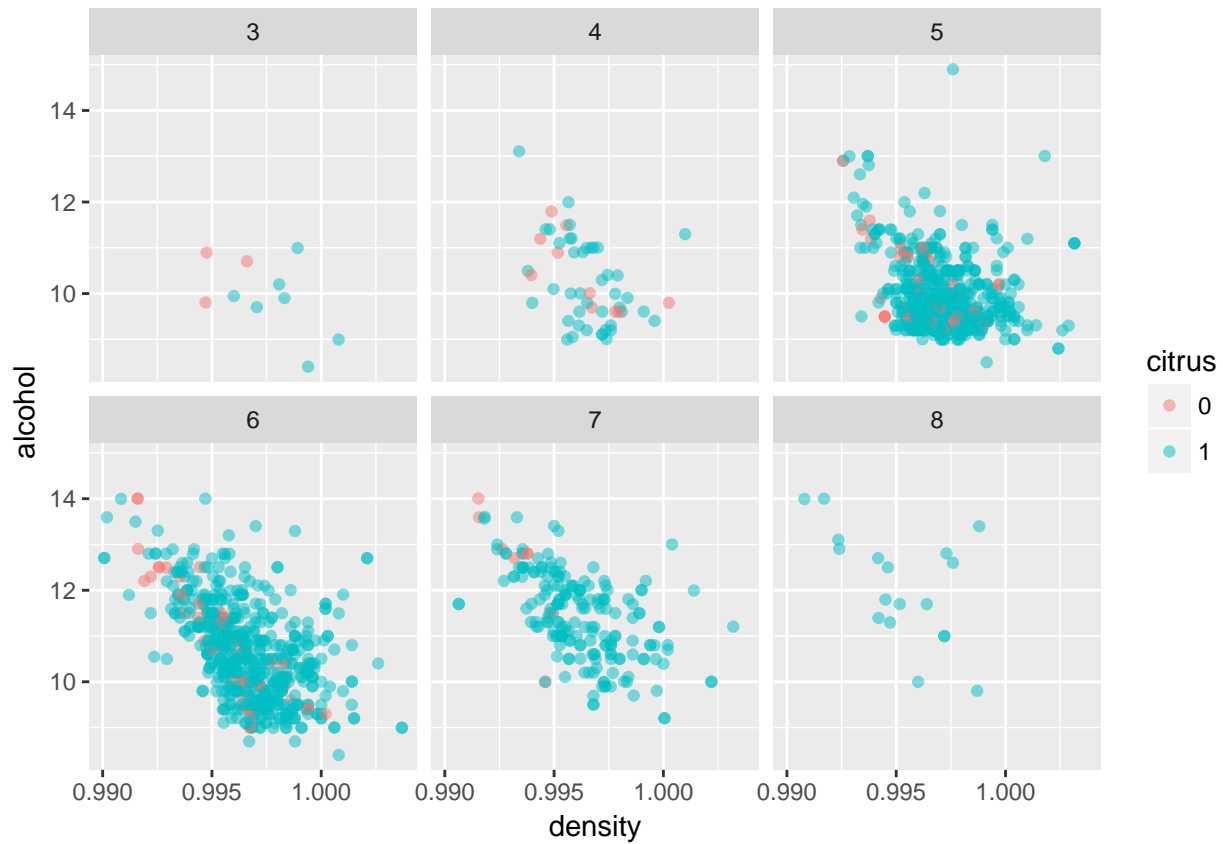


Density and pH are negatively correlated. Most lower rated wines have a higher pH. Wines with higher amounts of citric acid seem to have lower pH and higher density.

Density and Alcohol

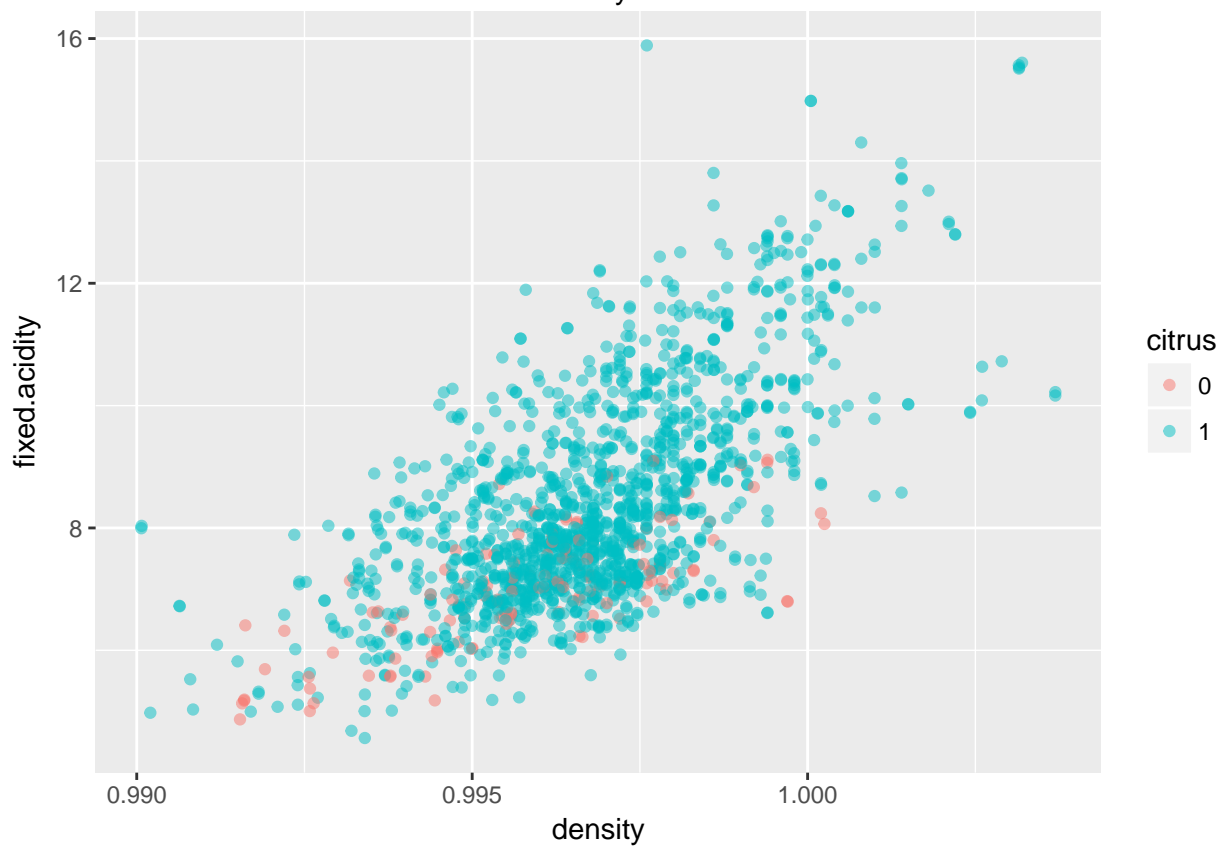
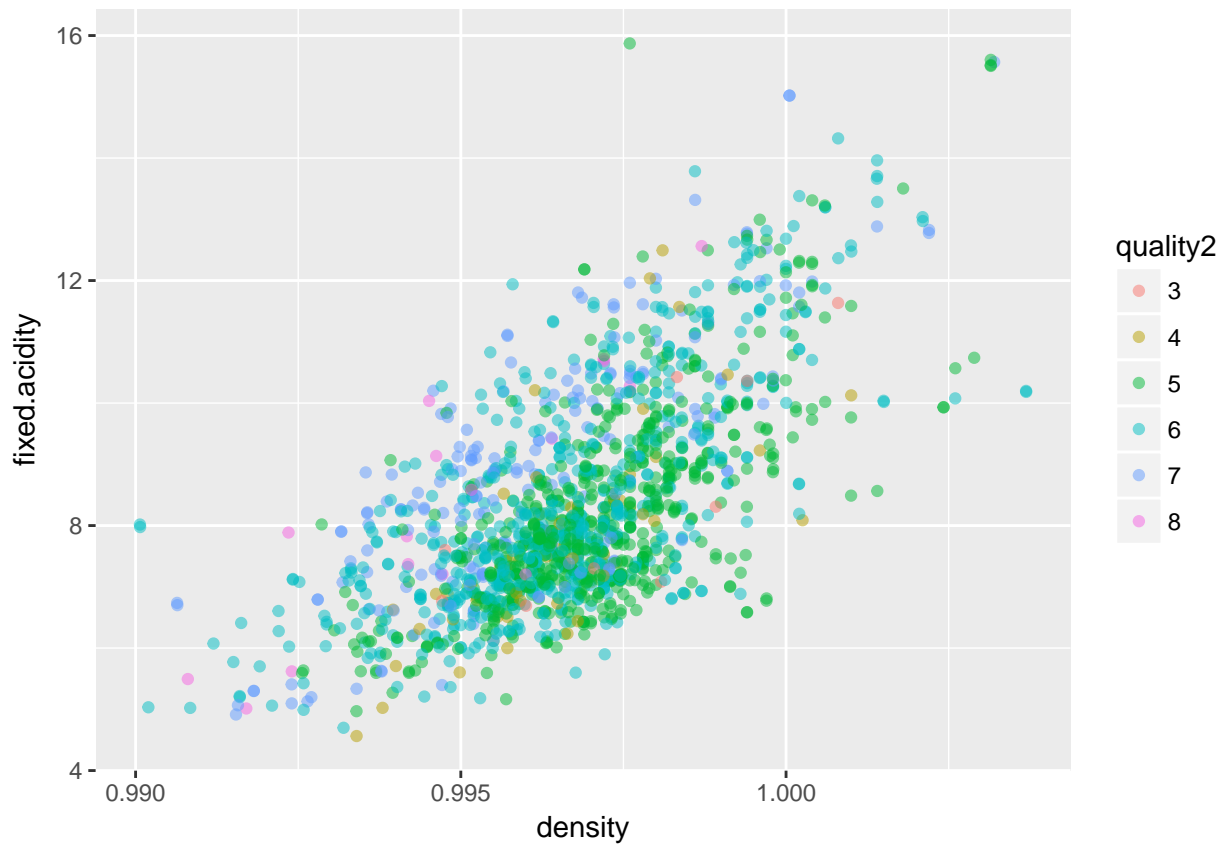


Alcohol and density are negatively correlated. Most of the highest quality wines have a higher abv. Density and quality are slightly negatively correlated (-.175).

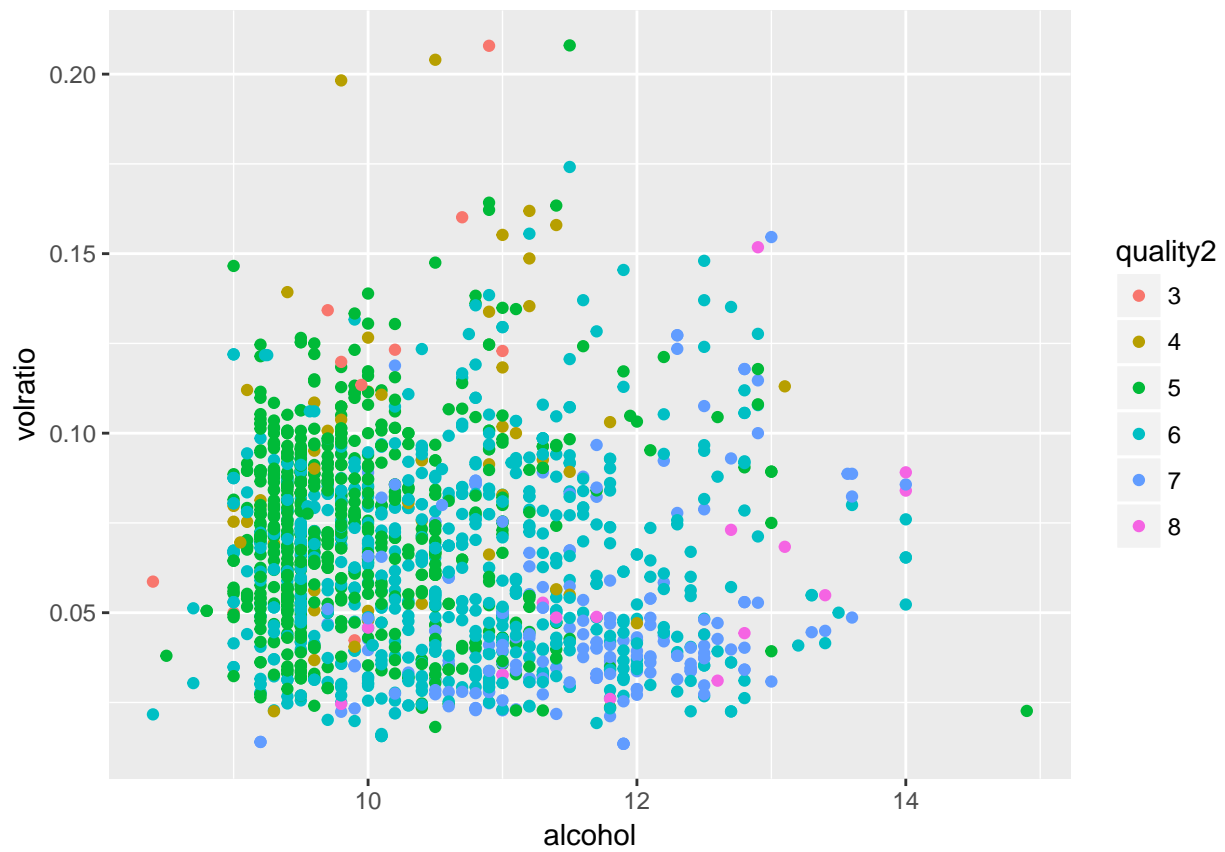


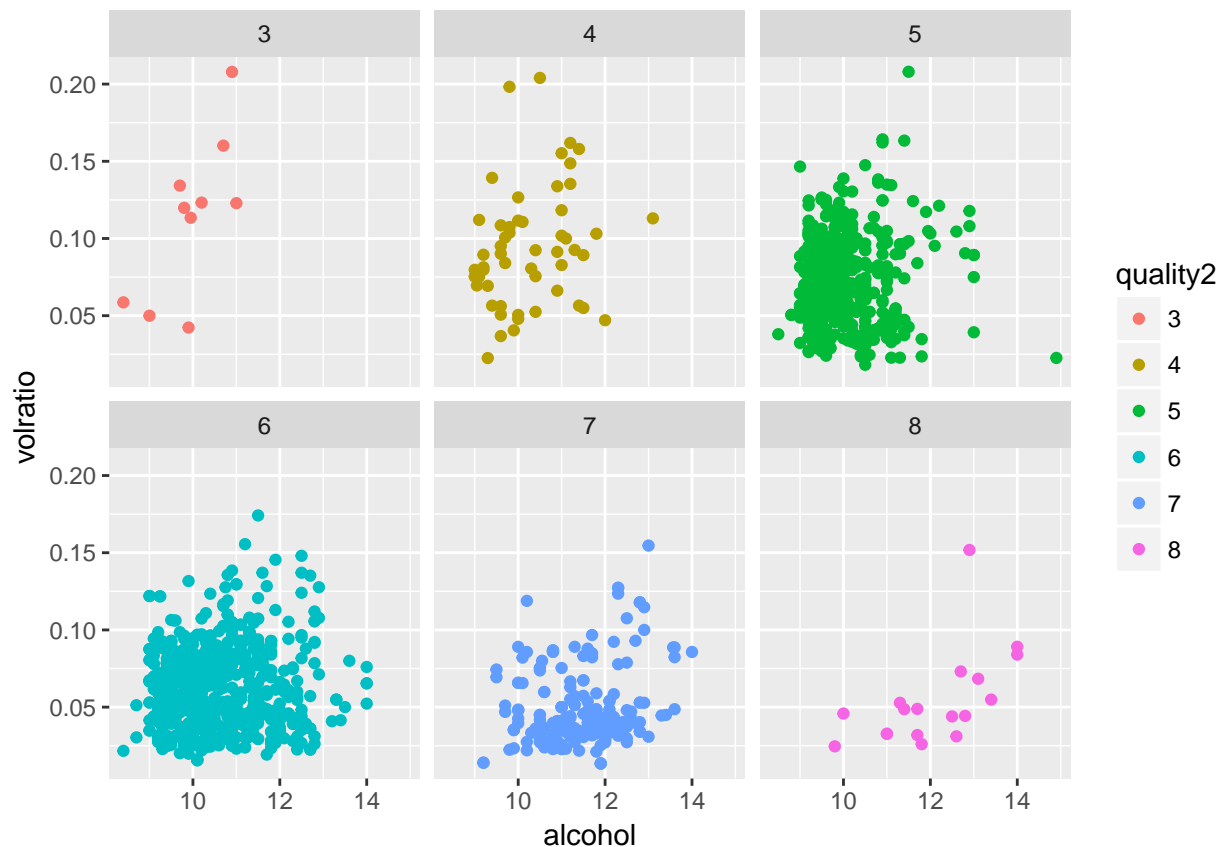
We can see a negative relationship between alcohol and density but it doesn't seem that this changes across wines with citrus added or without. However, density and citric acid are positively correlated.

Density and Fixed Acidity



Density and fixed acidity are positively correlated. While neither have a significant correlation with quality, citric acid is positively correlated with fixed acidity and density.





Higher quality wines tend to have a higher abv and a lower volratio while lower quality wines have to have the opposite. Higher quality wines are also positively correlated with alcohol.

Multivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. Were there features that strengthened each other in terms of looking at your feature(s) of interest?

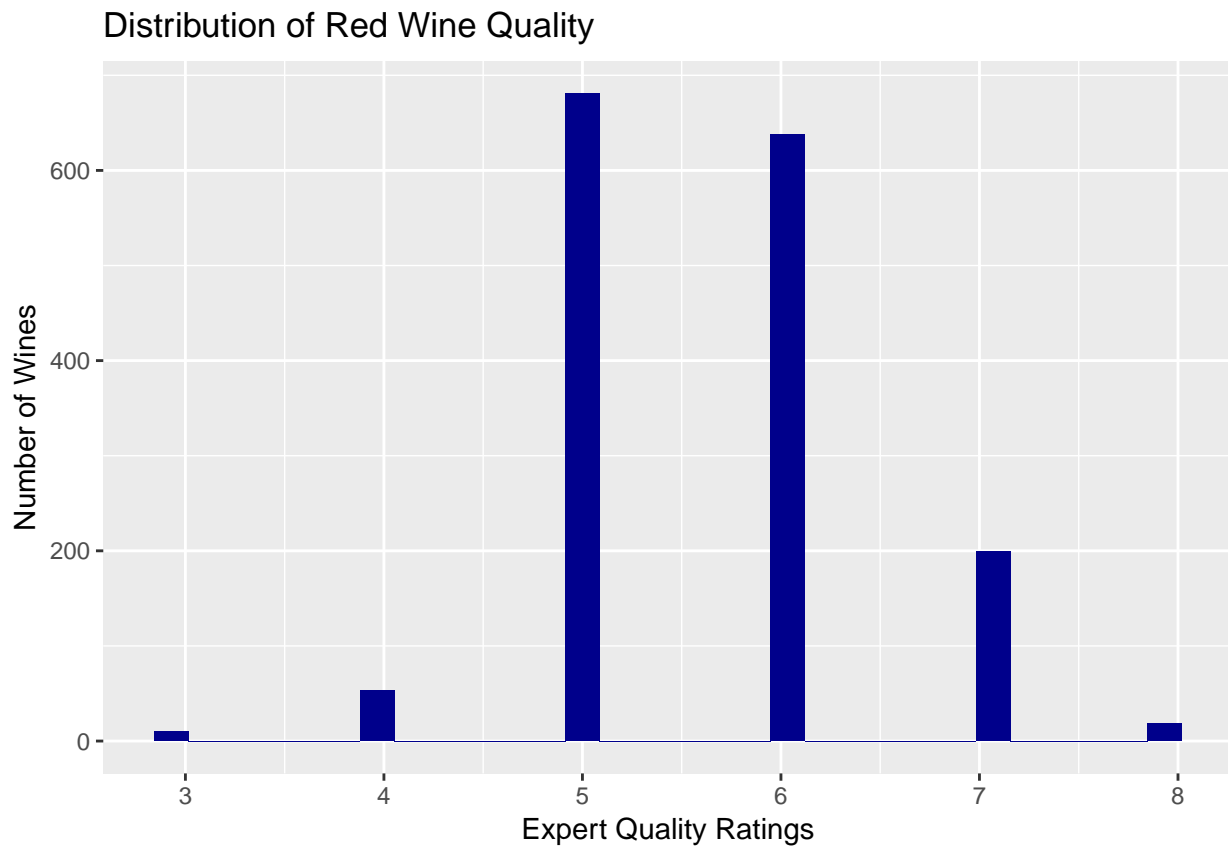
Yes, wines with low amounts of volatile acidity and higher amounts of citric acid, and alcohol tended to be the highest quality wines. Density and pH are negatively correlated, however, neither are significantly correlated with wine quality. Density and pH were negatively correlated with citric acid which may have a slight positive correlation with quality. Both also correlate with alcohol which has a significant positive correlation with quality.

Were there any interesting or surprising interactions between features?

Yes, density and citric acid were positively correlated, which after looking at summary statistics across quality levels, may indicate higher quality wine. However, density was also significantly negatively correlated with alcohol which correlates positively with quality. This may be why density did not have a stronger relationship with quality.

Final Plots and Summary

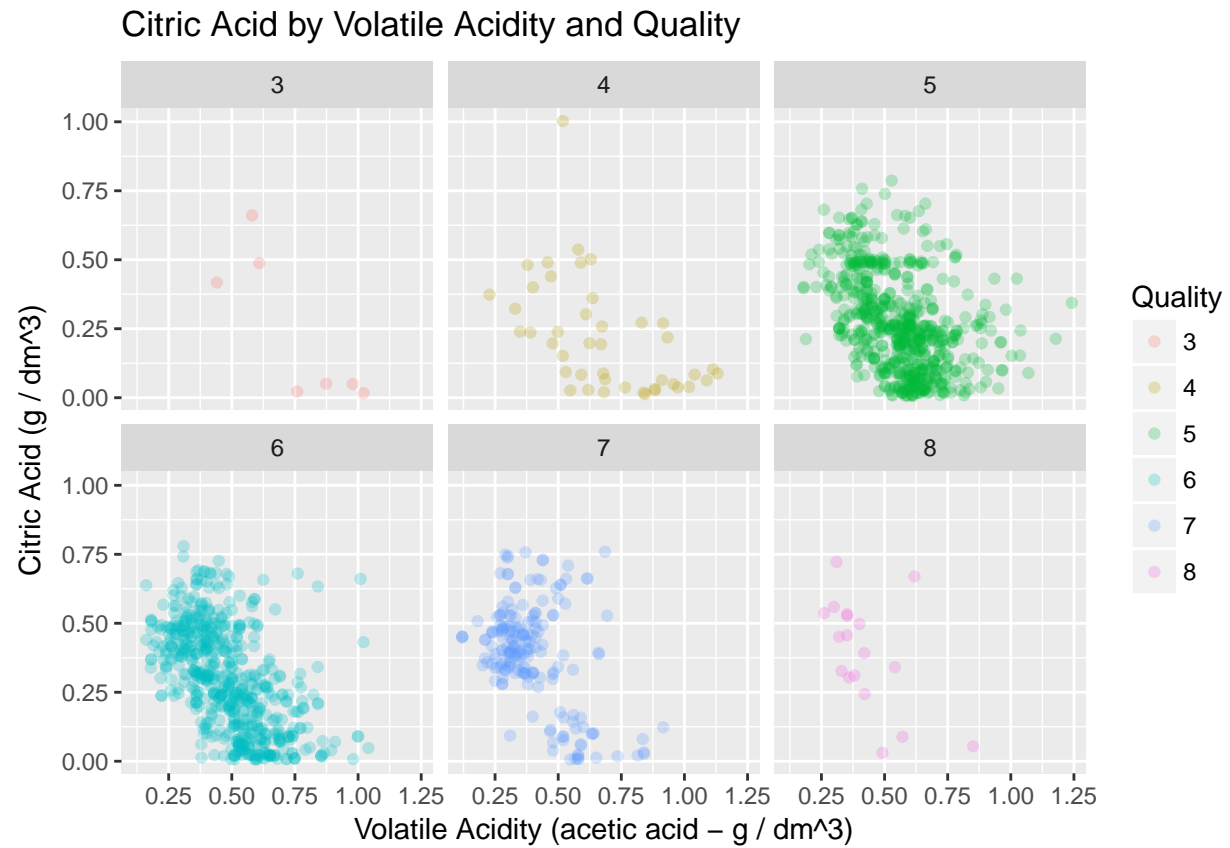
Plot One



Description One

The distribution of wine quality is fairly normal. Most wines fall within the 5-6 range with very few wines on the edges of our data and none on the edges of the 1-10 rating scale.

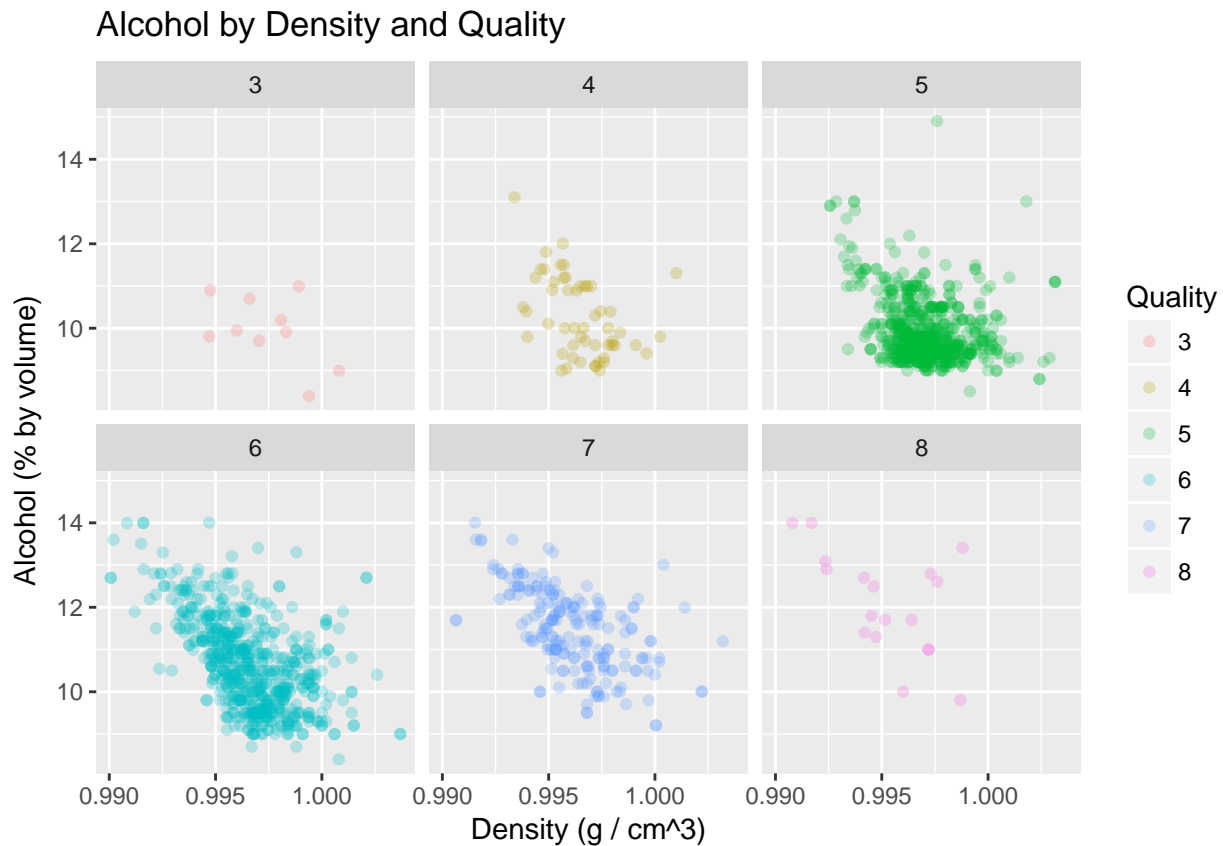
Plot Two



Description Two

Citric acid and volatile acidity have a moderate negative correlation (-.552). The highest quality wines also tend to have low levels of volatile acidity and contain higher levels of citric acid. This makes sense as volatile acidity (acetic acid) can give wine an 'unpleasant, vinegar taste' while citric acid actually 'can add 'freshness' and flavor to wines'.

Plot Three



Description Three

Alcohol and density have a correlation of $-.496$. Alcohol also had the strongest positive relationship with wine quality, meaning higher quality wines tended to have a higher percentage of alcohol. This was fairly surprising and unintuitive but might make sense in that higher alcohol percentage might mean the grapes used were more sweet than bitter and had more sugar to convert into alcohol. Density and quality are slightly negatively correlated ($-.175$) but as many factors influence a wine's density (citric acid, residual sugar, etc.) it may be a useful indicator of when certain ingredients are at extremes rather than keeping track of each of those variables individually.

Reflection

Overall, the red wine data-set was an interesting one to work with. Going in with little knowledge of the variables and their impact on wine taste, the EDA process was a bit more difficult than predicted. Most notably, it was easy to get lost in looking for cross interactions without a clear understanding of what variables influenced each other. I frequently had to take a step back in order to prioritize which pieces to explore. However, I did find that R was very pleasant to work with for EDA (likely a bit more so since the data had been cleaned previously). Visualizations were much easier to construct and manipulate once a direction was targeted.

From there, and from reading the text file on the data, it was easier to identify volatile acidity as a major factor in wine quality; surprisingly much more so than total sulfur dioxide. In addition, alcohol content also seemed to be a larger factor than expected. Higher quality wines tended to have a higher alcohol content which was unintuitive to me as in general, alcohol with a higher ABV tend to be less widely drinkable. Along with this thinking, I expected wines with higher amounts of citric acid to be more drinkable and thus to have a higher positive correlation with quality than was seen.

Other factors like density and pH were harder to understand as they tended to have more cross interaction with variables. From the limited data points and information provided on the quality ratings, it's hard to conceptualize what factor the experts were attuning to. For further exploration, it would be great to have additional data on wines with ratings across the spectrum as well as background knowledge on the experts' rubric. With additional time, it would also be of interest to group the data by quality ratings for exploration as well as cut a few of the variables a bit differently in order to understand the difference between high, moderate, and low levels of key factors.