Red Wine Quality - A study in R

by Mario Bonilla

Introduction

The dataset is related to the red variant of the Portuguese "Vinho Verde" wine. For more details, consult: http://www.vinhoverde.pt/en/ (http://www.vinhoverde.pt/en/) or the reference [Cortez et al., 2009]. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.).

These datasets can be viewed as classification or regression tasks. The classes are ordered and not balanced (e.g. there are much more normal wines than excellent or poor ones).

General Data Information

Number of Instances: red wine - 1599 observations. ***Number of Attributes***: 11 + output attribute (quality).

Input variables (based on physicochemical tests):

1 - ***fixed acidity*** (tartaric acid - g / dm^3): most acids involved with wine or fixed or nonvolatile (do not evaporate readily)

2 - **volatile acidity** (acetic acid - g / dm^3): the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste

3 - **citric acid** (g / dm^3): found in small quantities, citric acid can add 'freshness' and flavor to wines

4 - **residual sugar** (g / dm^3): 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 / dm^3): the amount of salt in the wine

6 - **free sulfur dioxide** (mg / dm^3): 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 / dm^3): 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 / cm^3): 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): a wine additive which can contribute to sulfur dioxide gas (S02) 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): based on sensory data

---

General Data Analysis

Loading the Data and getting just a glimpse on it:

```
wine <- read.csv("wineQualityReds.csv")
```

Variable names:

```
##  [1] "X"                   "fixed.acidity"        "volatile.acidity"
##  [4] "citric.acid"         "residual.sugar"       "chlorides"
##  [7] "free.sulfur.dioxide" "total.sulfur.dioxide" "density"
## [10] "pH"                  "sulphates"            "alcohol"
## [13] "quality"
```

Data Structure:

```
## '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 0.07
3 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 ...
```

First 6 records:

```
##   X fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1 1           7.4             0.70        0.00            1.9     0.076
## 2 2           7.8             0.88        0.00            2.6     0.098
## 3 3           7.8             0.76        0.04            2.3     0.092
## 4 4          11.2             0.28        0.56            1.9     0.075
## 5 5           7.4             0.70        0.00            1.9     0.076
## 6 6           7.4             0.66        0.00            1.8     0.075
##   free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                  11                  34  0.9978 3.51      0.56     9.4
## 2                  25                  67  0.9968 3.20      0.68     9.8
## 3                  15                  54  0.9970 3.26      0.65     9.8
## 4                  17                  60  0.9980 3.16      0.58     9.8
## 5                  11                  34  0.9978 3.51      0.56     9.4
## 6                  13                  40  0.9978 3.51      0.56     9.4
##   quality
## 1       5
## 2       5
## 3       5
## 4       6
## 5       5
## 6       5
```

Some Statistics:

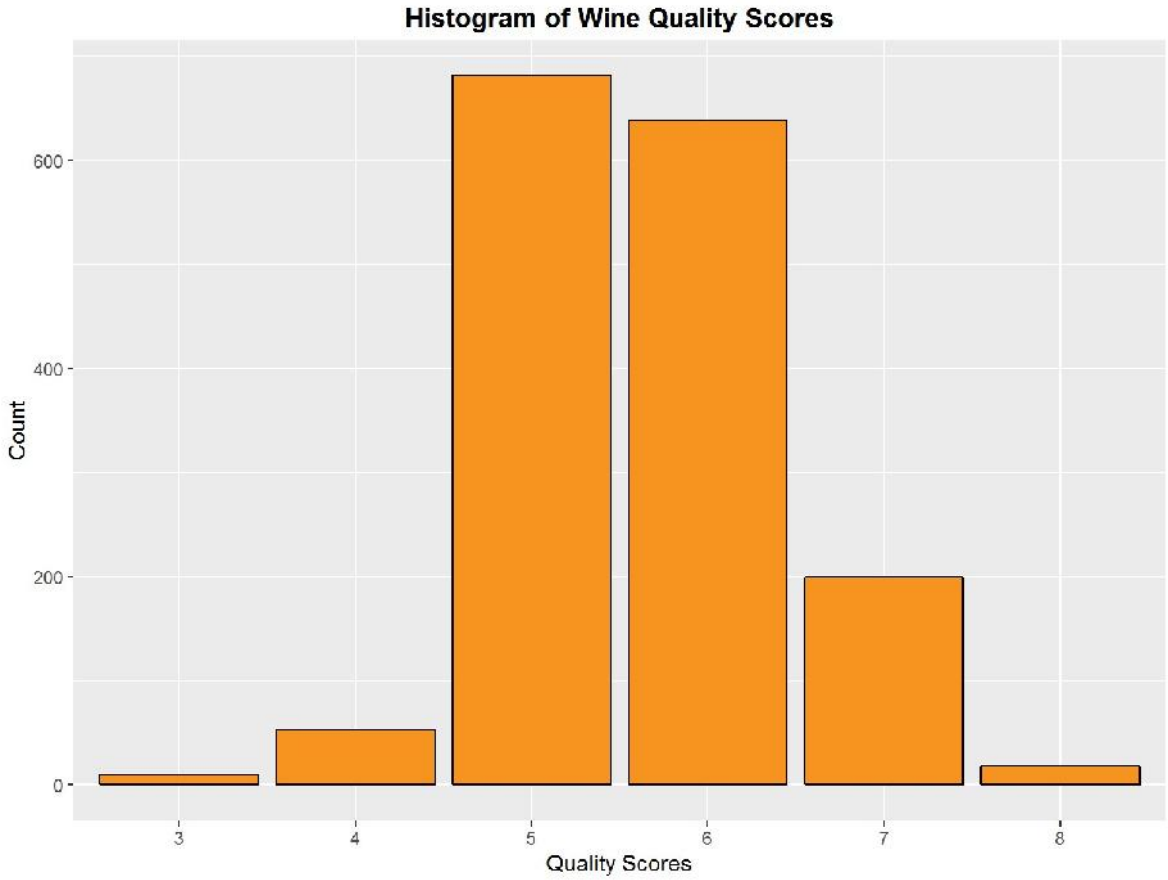
```
##        X          fixed.acidity   volatile.acidity  citric.acid
##  Min.   :   1.0   Min.   : 4.60   Min.   :0.1200   Min.   :0.000
##  1st Qu.: 400.5   1st Qu.: 7.10   1st Qu.:0.3900   1st Qu.:0.090
##  Median : 800.0   Median : 7.90   Median :0.5200   Median :0.260
##  Mean   : 800.0   Mean   : 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
```

Univariate Plots and Analysis Section

In this section we will have a close look at all the variables from the dataset.

1) General Quality

Most of the observations fall into the 5 and 6 quality score, followed by far by 7 quality score. That means than the wines tasted where judged to be of average quality with only a few of them either bad (3, 4 scores) or good/very good (7, 8 scores).




```
##   3   4   5   6   7   8
##  10  53 681 638 199  18
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.000   5.000   6.000   5.636   6.000   8.000
```

We would like to comment on this because the grade was set sujectively by the tasters.

Let us check the Wine Spectator standards:

Wine Spectator tasters review wines on the following 100-point scale:

```
95-100 Classic: a great wine
90-94 Outstanding: a wine of superior character and style
85-89 Very good: a wine with special qualities
80-84 Good: a solid, well-made wine
75-79 Mediocre: a drinkable wine that may have minor flaws
50-74 Not recommended
```

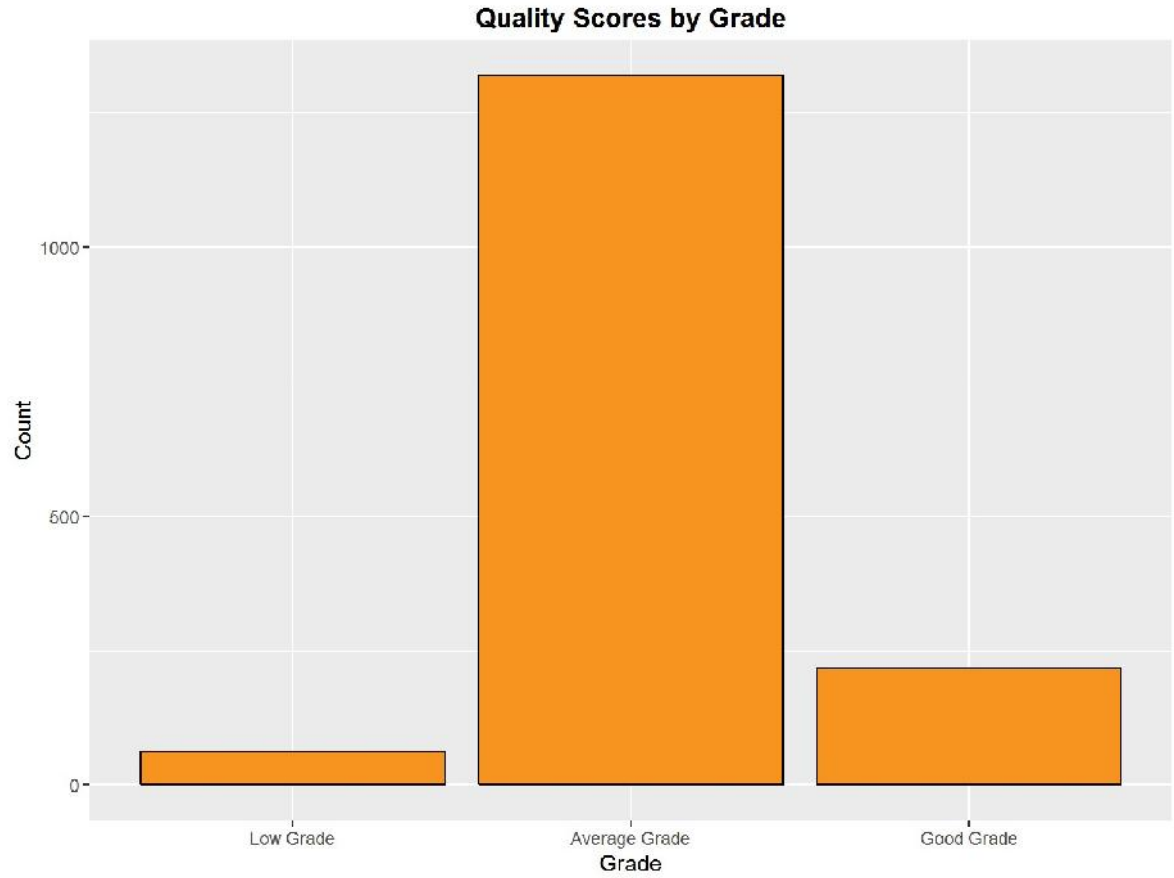
Based on this scale, more than 90% of the wine tasted in the database object of this study would be less than "Mediocre", with both the media and the mean in the "not recommended" bracket.

```
quality
Min.   :3.000
1st Qu.:5.000
Median :6.000
Mean   :5.636
3rd Qu.:6.000
Max.   :8.000
```

We have performed several transformation in the database:

- We have deleted the variable "X", because it was not giving real valuable information. It was just a record numbering value.
- Besides, we transformed the output variable "quality"" from an integer to an ordered factor.
- After that, we created a "Grade" variable setting "quality" variable scores less than five to "Low Grade", from five to seven to "Average Grade" and above seven to "Good Grade".

We plotted the resulting new variable:

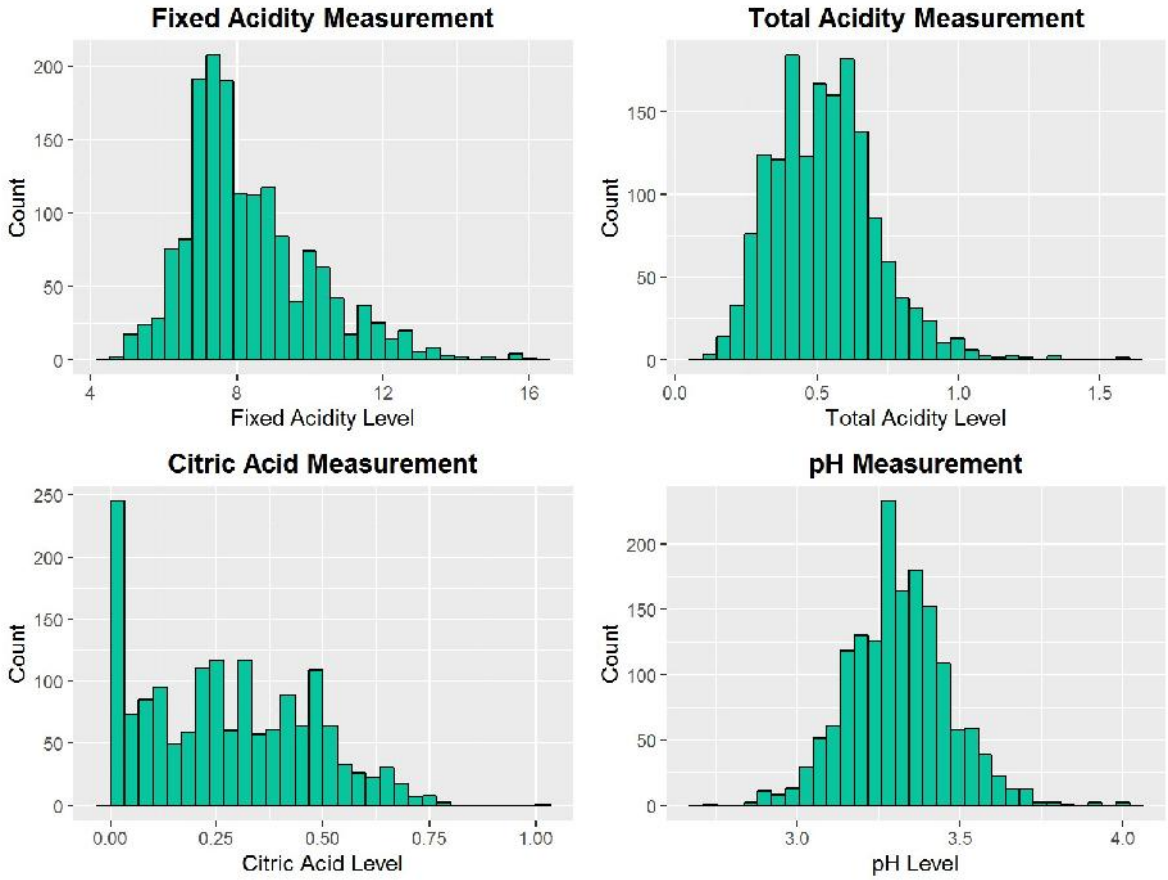



```
##     Low Grade Average Grade    Good Grade
##            63          1319           217
```

2) Acidity Variables

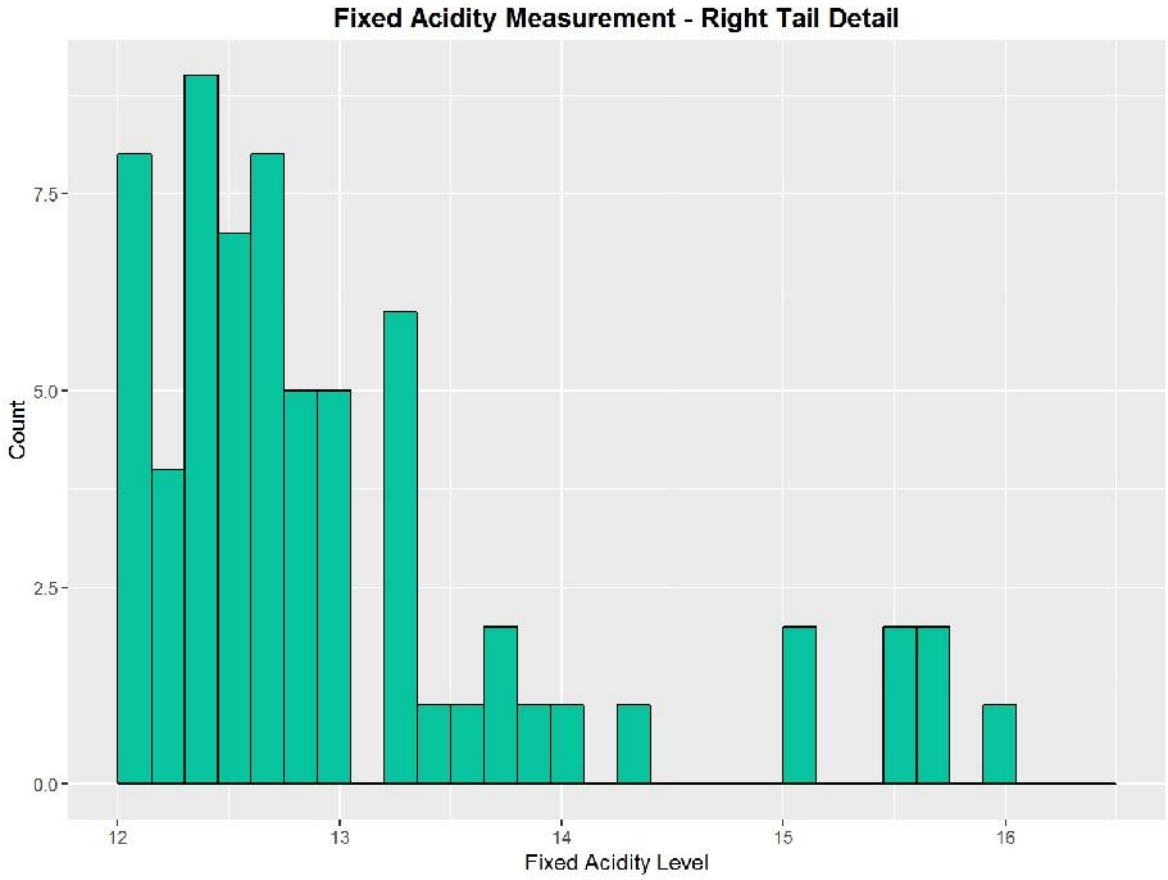
The three variables (fixed and volatile acidity, and pH) show a "Normal" distribution of the observations. The variable "fixed.acidity" is a little long tailed to the right, whereas the pH distribution is the most balanced. We could find very high values in some observations in the three of them.

The citric acid variable is the most skewed; the first quantile is 0.09 and the median is 0.26.



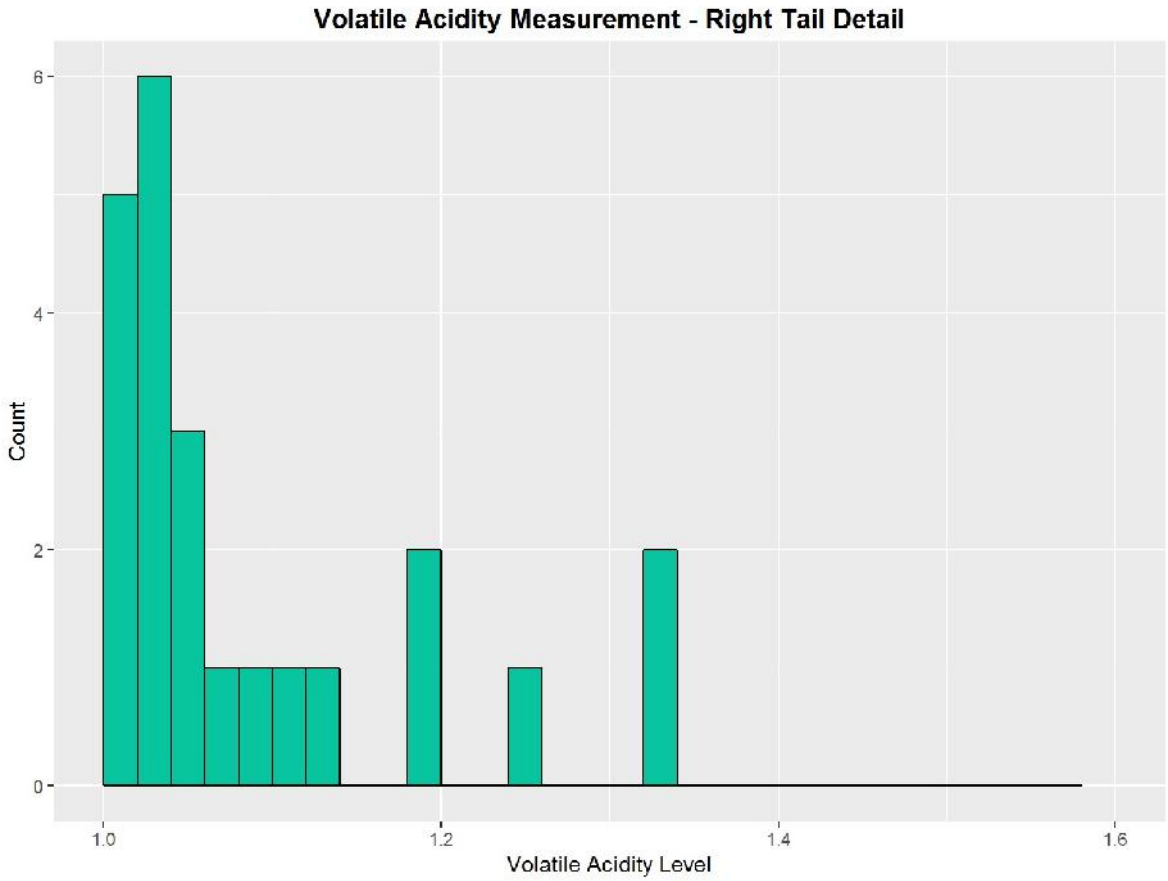
Let us have a look in more detail:

We clearly see now some extreme records, in the 15-16 interval, in the citric acid (at > 1).



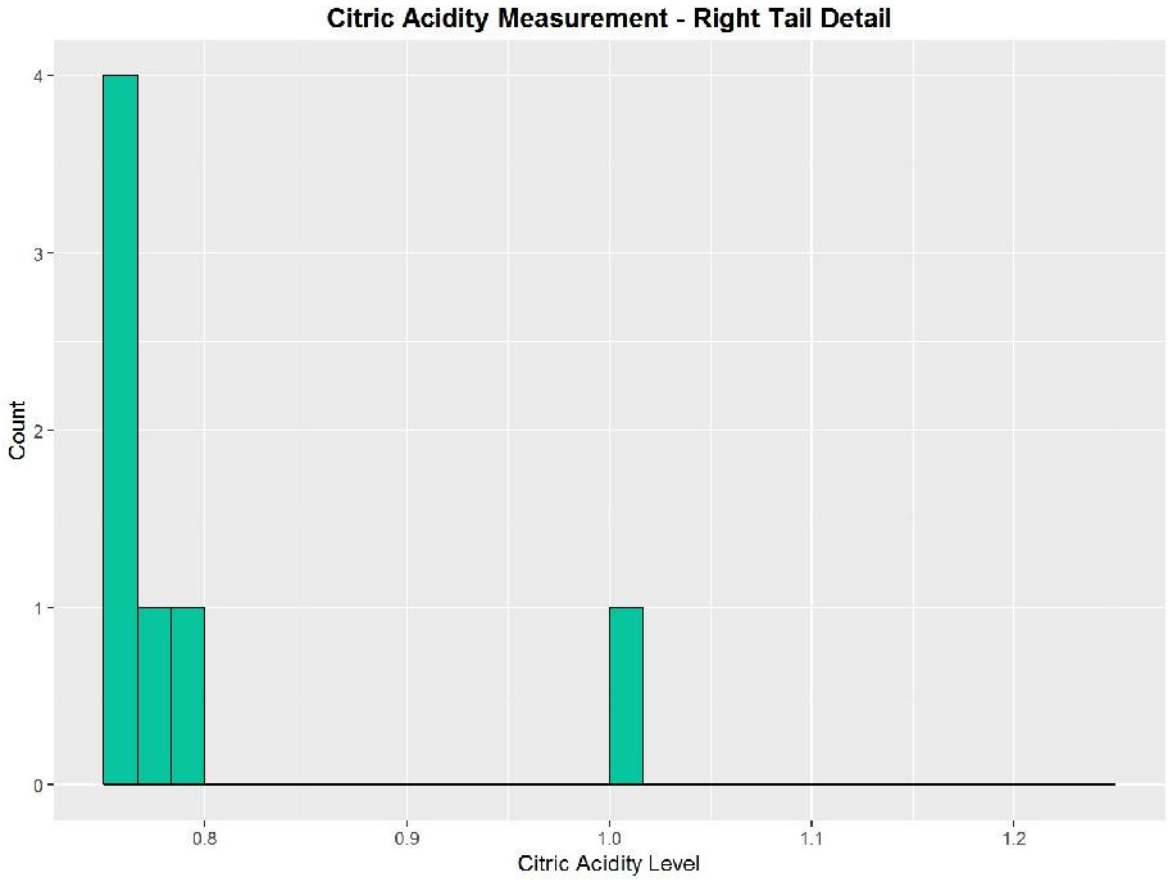
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.60    7.10    7.90    8.32    9.20   15.90
```

We can observe the same pattern in the volatile acidity (from 1.1 to 1.4 approx.)…



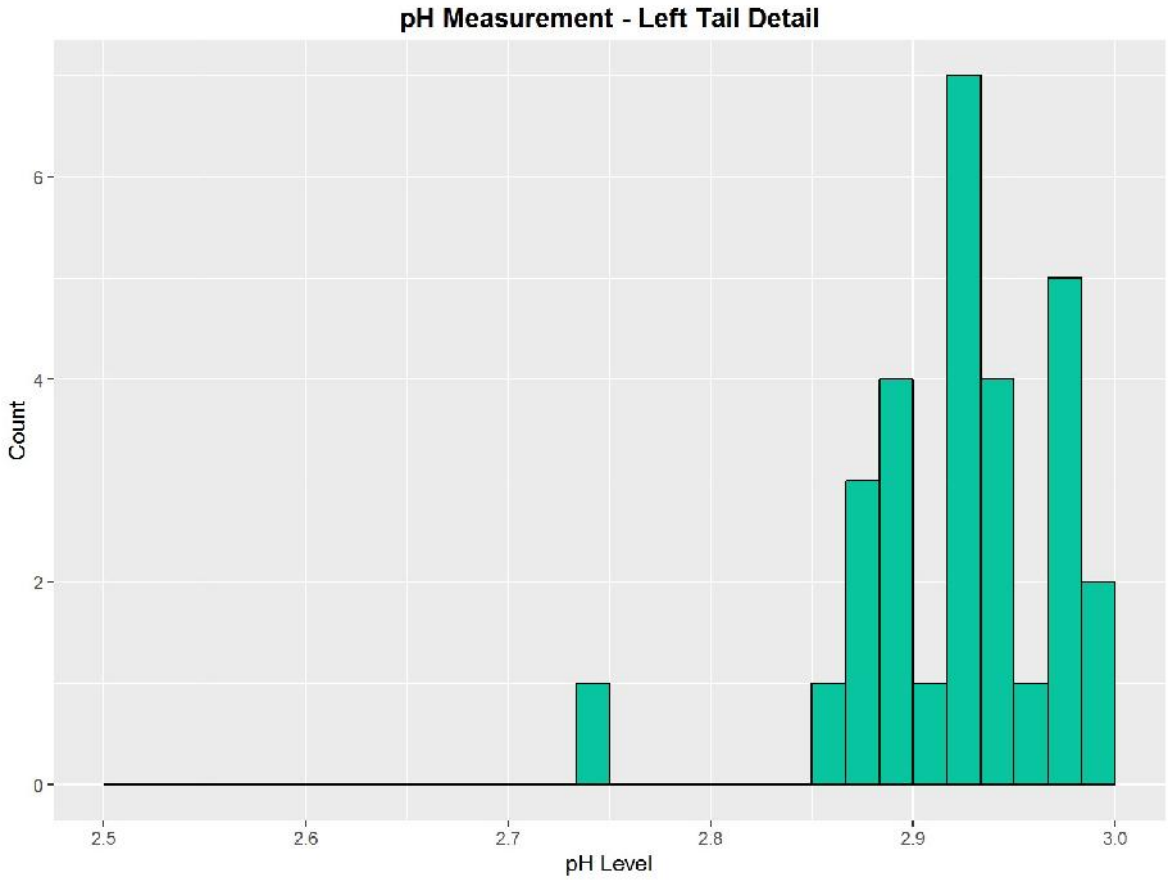

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1200  0.3900  0.5200  0.5278  0.6400  1.5800
```

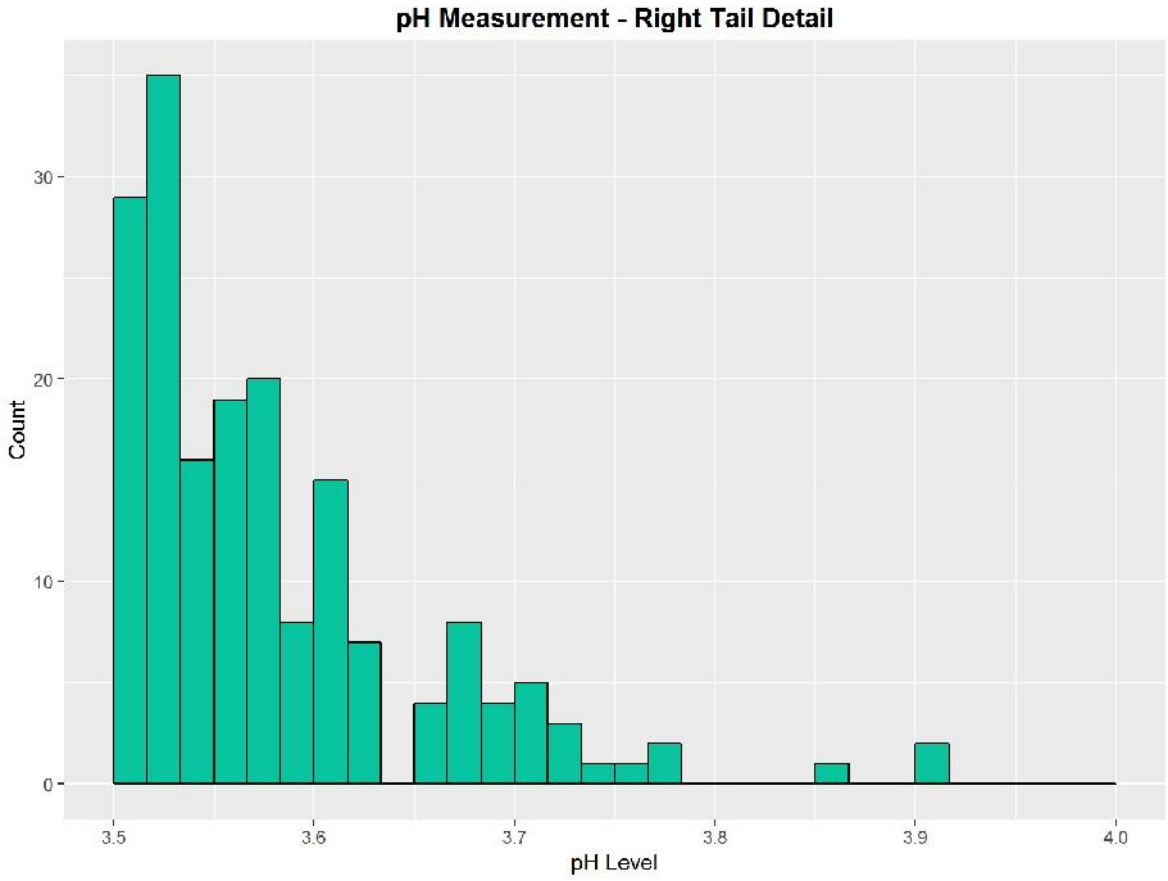
…in the citric acid (at > 1).




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.090   0.260   0.271   0.420   1.000
```

In the pH case, the extreme observations are found in both tails (at 2.5 approx. and around 3.9).


pH Measurement - Left Tail Detail
Count
0
2
4
6
2.5
2.6
2.7
2.8
2.9
3.0
pH Level


pH Measurement - Right Tail Detail
Count
0
10
20
30
3.5
3.6
3.7
3.8
3.9
4.0
pH Level

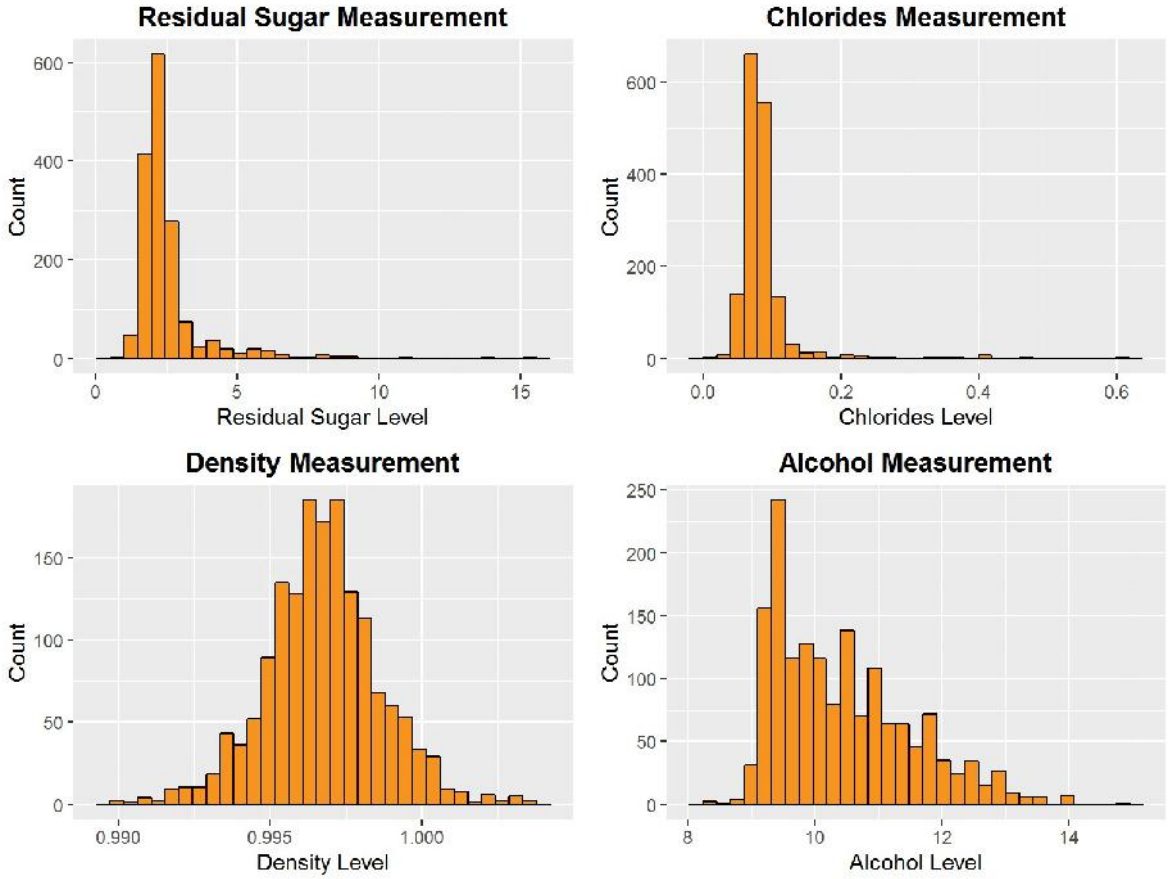
3) Alcohol, Sugar and Salt

About the sugar and salt in the wines (next plots), most of the wines have low values. However, there are some wines with very high values. In the case of sugar, the mean is 2.539 and the 3rd Q. is 2.6 but there are maximum values up to 15.5, almost six times the average.

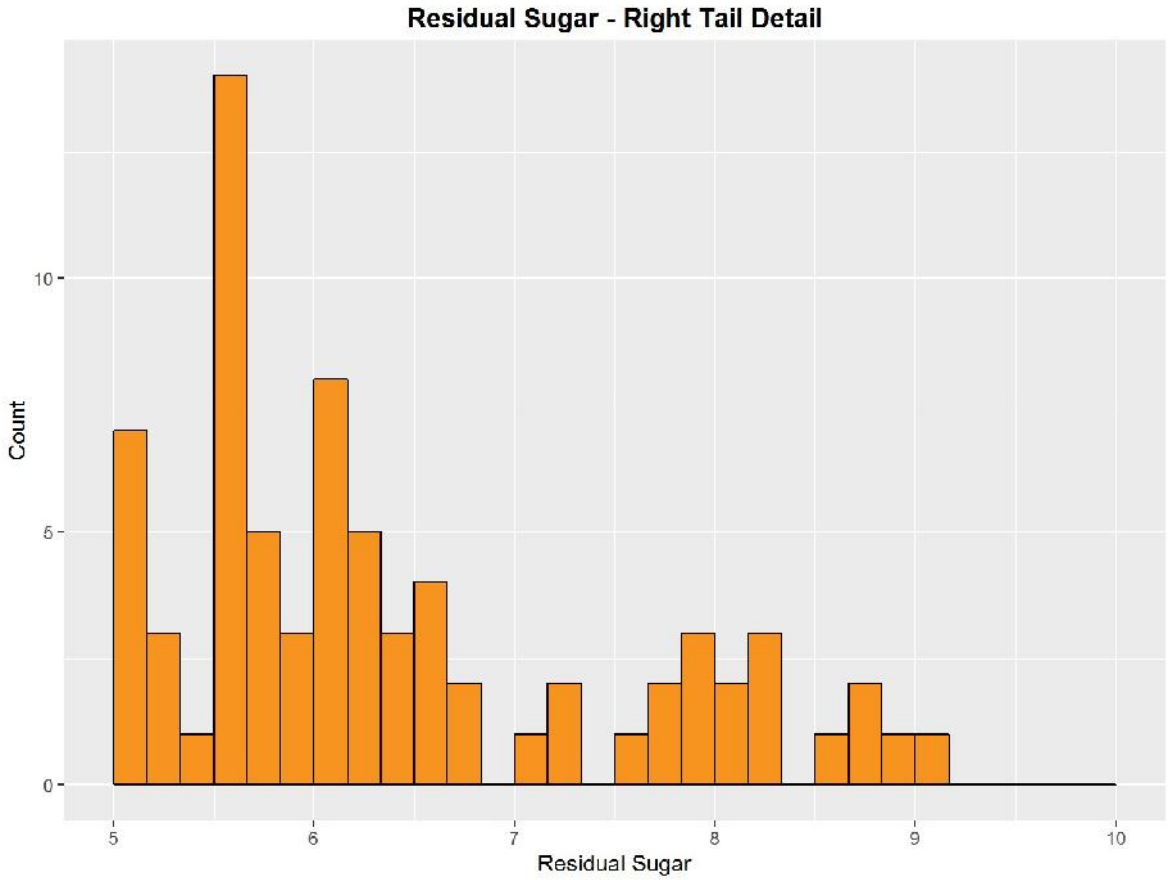
We find the same behaviour in the chloride (salt); there is a big gap between the mean (0.08) - 3rd Q. (0.09) and the maximum value (0.6).

The density follows a nice bell shape distributions, with some variance observed.

The alcohol variable shows a particular shape. Only 25% of the wines have less than 9.5° of alcohol, being the average slightly above of 10.4°.

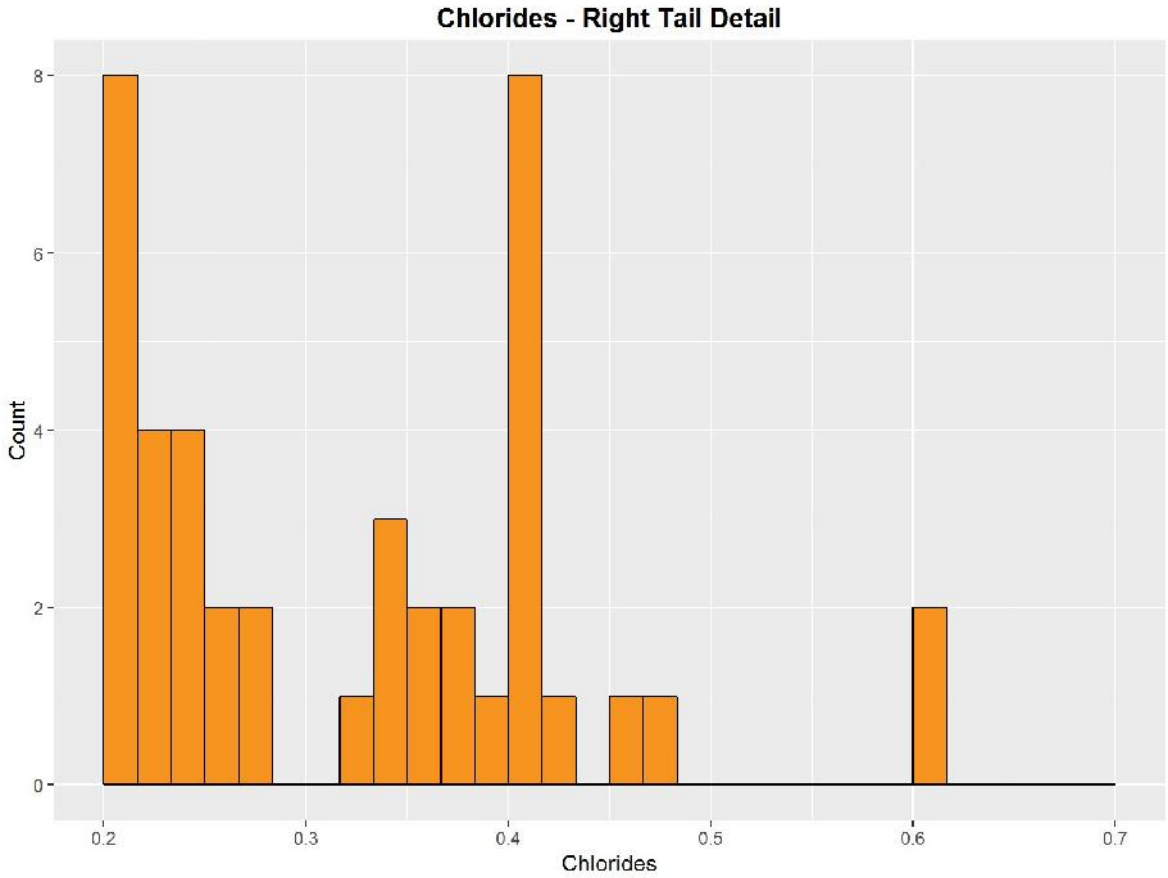


Residual sugar, right tail:



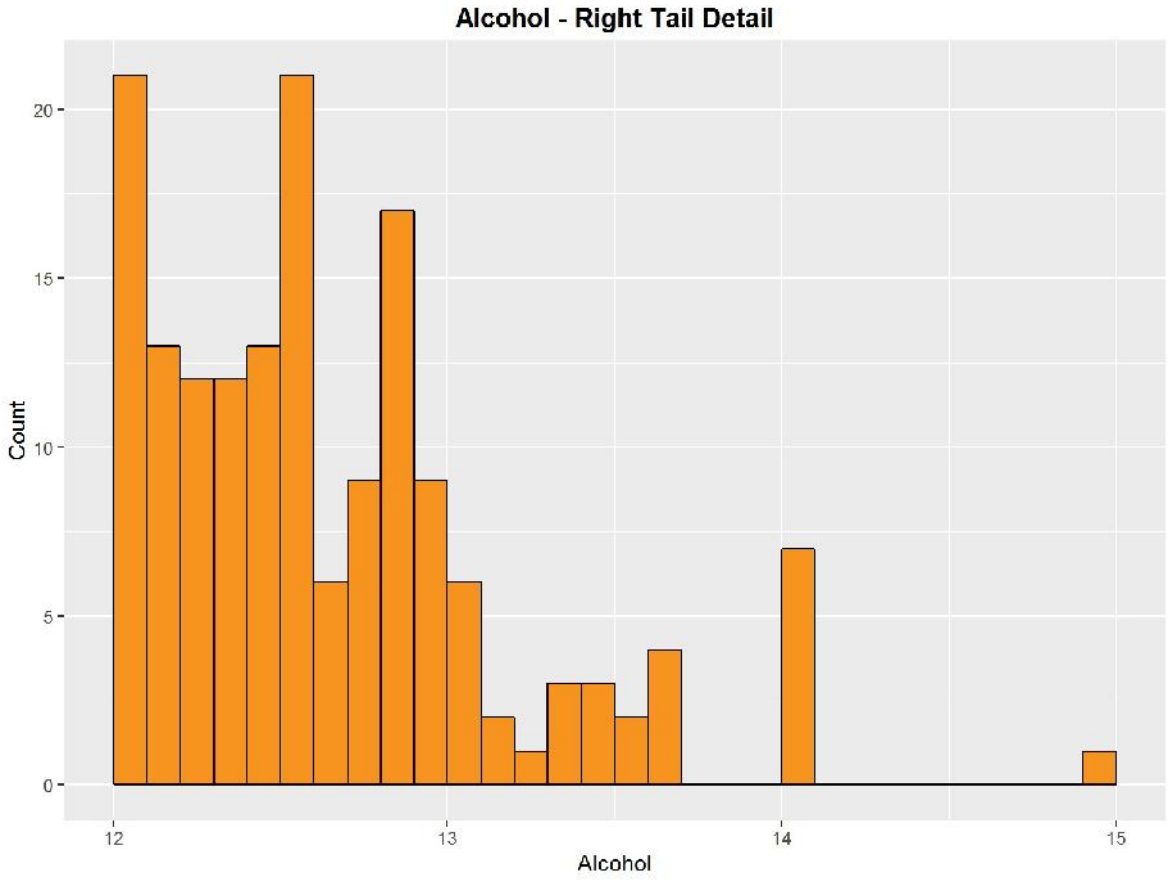

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.900   1.900   2.200   2.539   2.600  15.500
```

Chlorides, right tail:




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01200 0.07000 0.07900 0.08747 0.09000 0.61100
```

Alcohol, right tail:




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.40    9.50   10.20   10.42   11.10   14.90
```

4) Sulphites

All wines contain sulphur dioxide in various forms, collectively known as sulphites. Even in completely unsulphured wine it is present at concentrations of up to 10 milligrams per liter.

In this case we have to be particular vigilant regarding high levels of SO2 because they can become very dangerous for the human health.

There are three reasons you might not want sulphur dioxide added to your wine:

1. **Taste**: sulphur dioxide has an unpleasant smell, like that of a struck match, detectable at very low concentrations.
2. **Health**: sulphur dioxide can cause potentially fatal allergic reactions and has been linked with numerous other health problems, including hangover.
3. **Principle**: adding sulphur dioxide breaks the principle of naturalness in wine.

Red wines contain far less sulfur than white wines. That's because the tannin in red wines acts as a preservative, making sulfur dioxide less necessary.

The World Health Organisation recommends a maximum daily intake of 0.7mg of sulphur dioxide per kilogram of bodyweight.

For a man of average weight this is less than a third of a bottle of a white wine with a concentration of 200 mg/l.

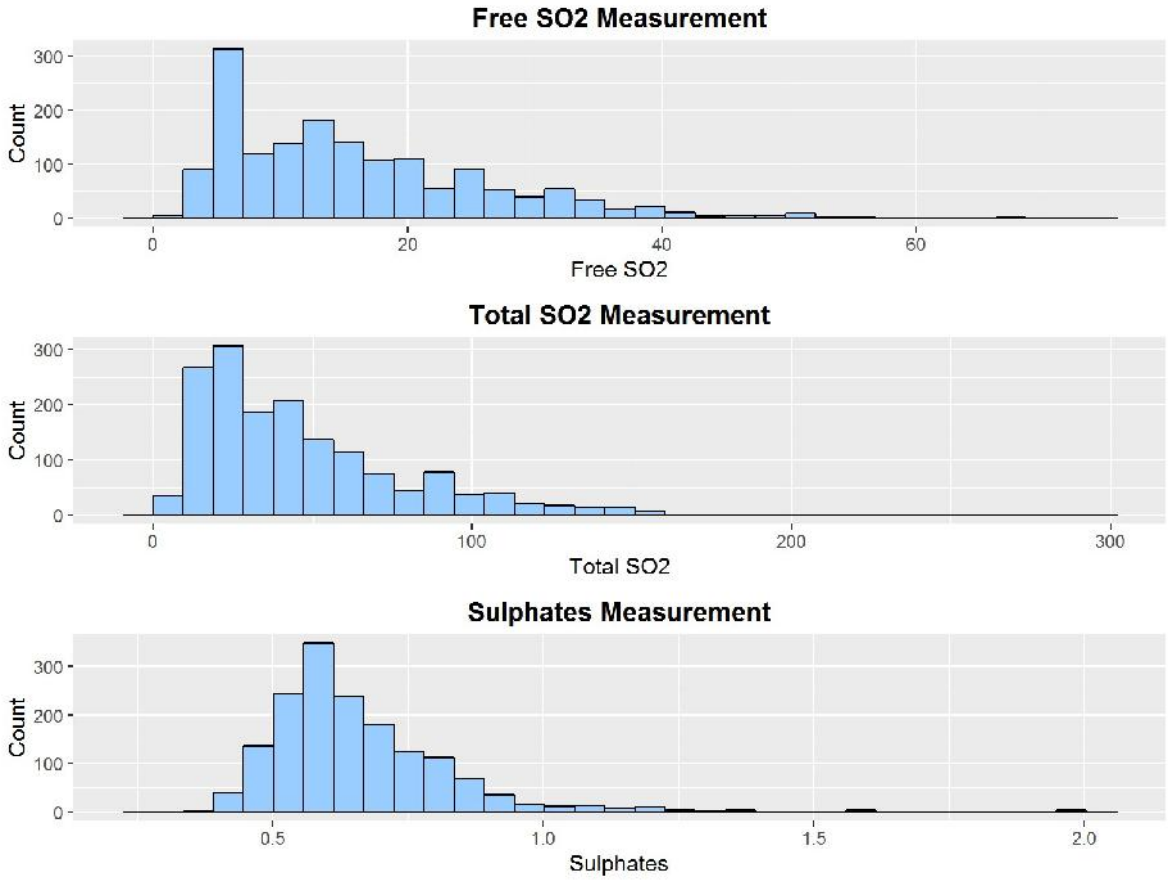
Country/Region	Wine Style	Legal Limit (mg/L)
Europe	Red	160
	White	210
	Rose	210
	Sweet	400
USA		350
Australia		250

Regular consumption of conventional wines means regularly exceeding the RDA of sulphur dioxide by a large margin. Wine is not the only product with high levels of SO2. It is also found in canned tuna, pizza dough, jams, gelatin, trail mix, cheese, deli meat and even prescription pills.

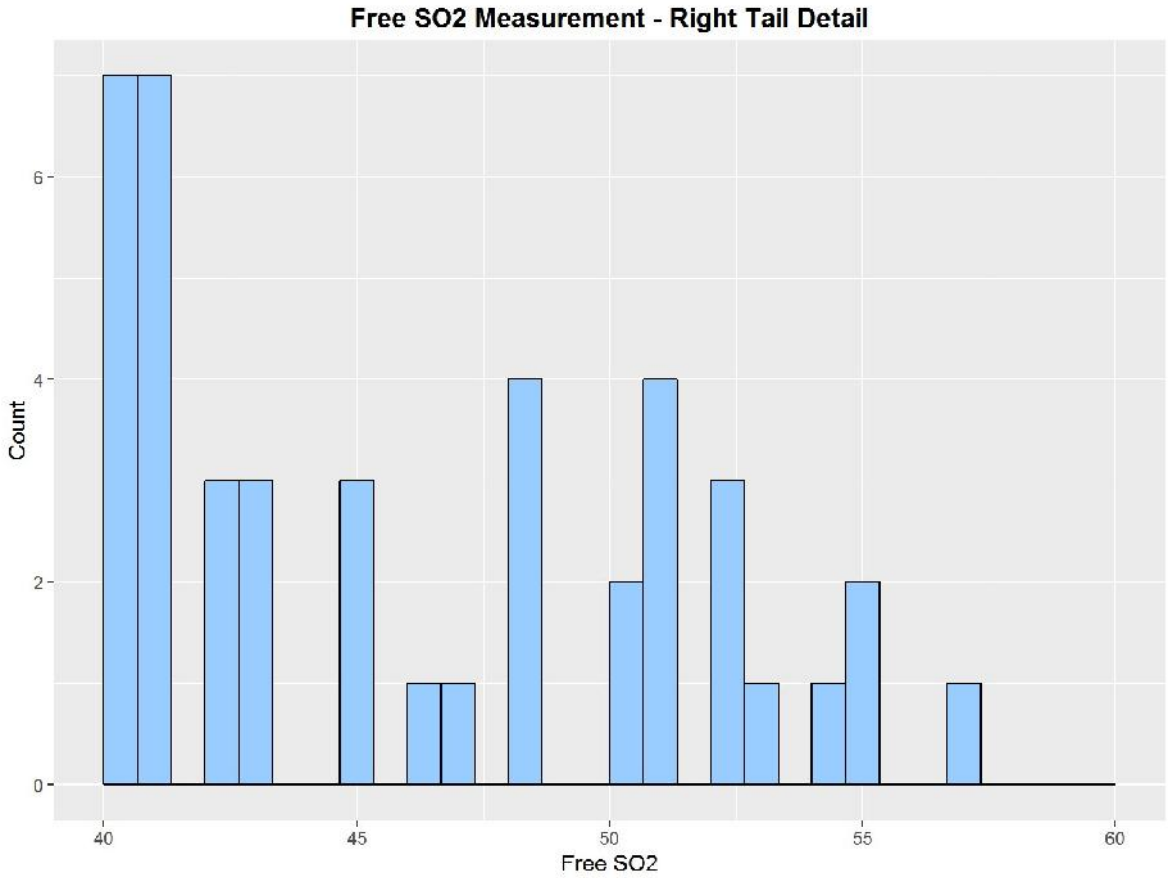
Sulphur dioxide can cause allergic reactions in some people. It is dangerous for asthmatics even at very low levels.

Besides, it contributes significantly to hangovers. Heavy drinkers who also have to get up in the morning would be advised to stick to natural wine.

The sulphur family of variables present a slightly skewed to the left distribution (see next plots). Although most wines are within the legal limits under the U.S. (350) and the Australian (250) legislation, some of them exceed by far the E.U. parameters (max. 160 in the EU, for our case, red wine).

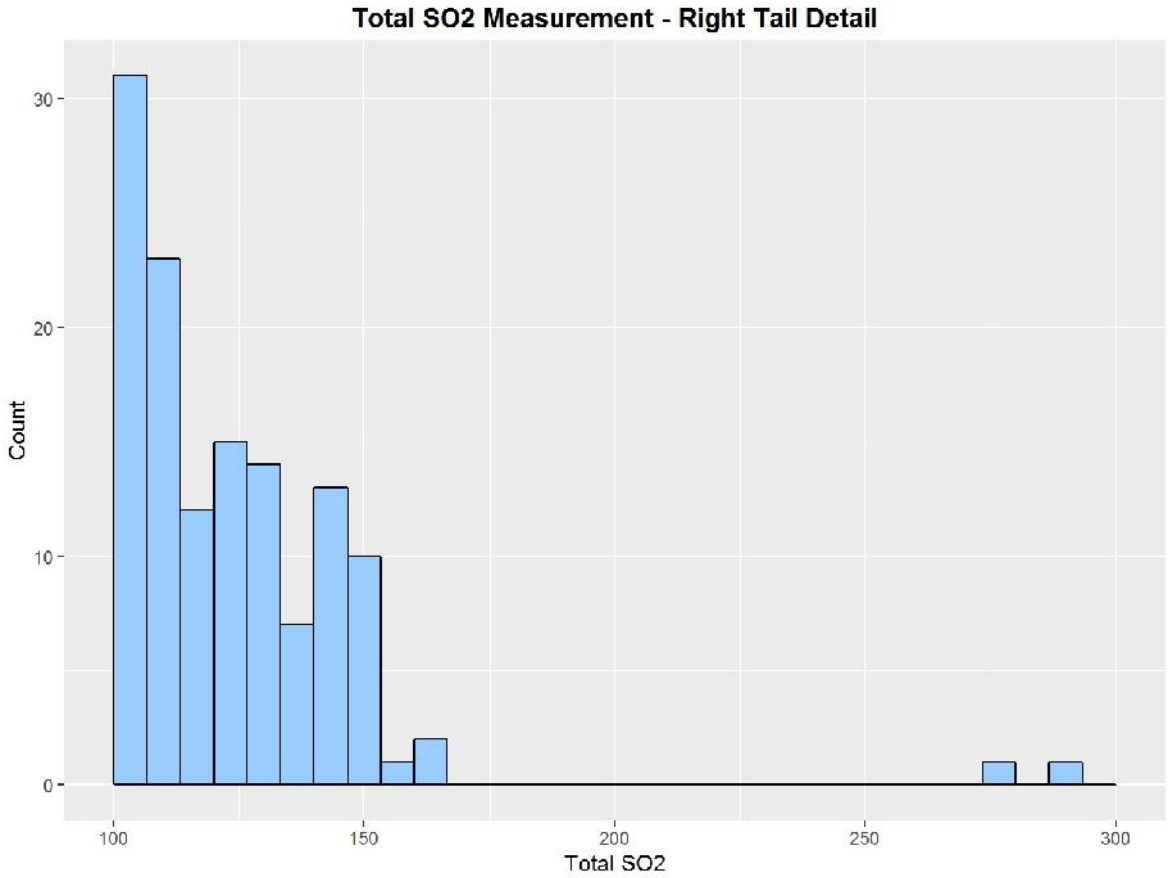


For instance, we find observations in the free sulphur dioxide variable up to 4.5 times the average value:



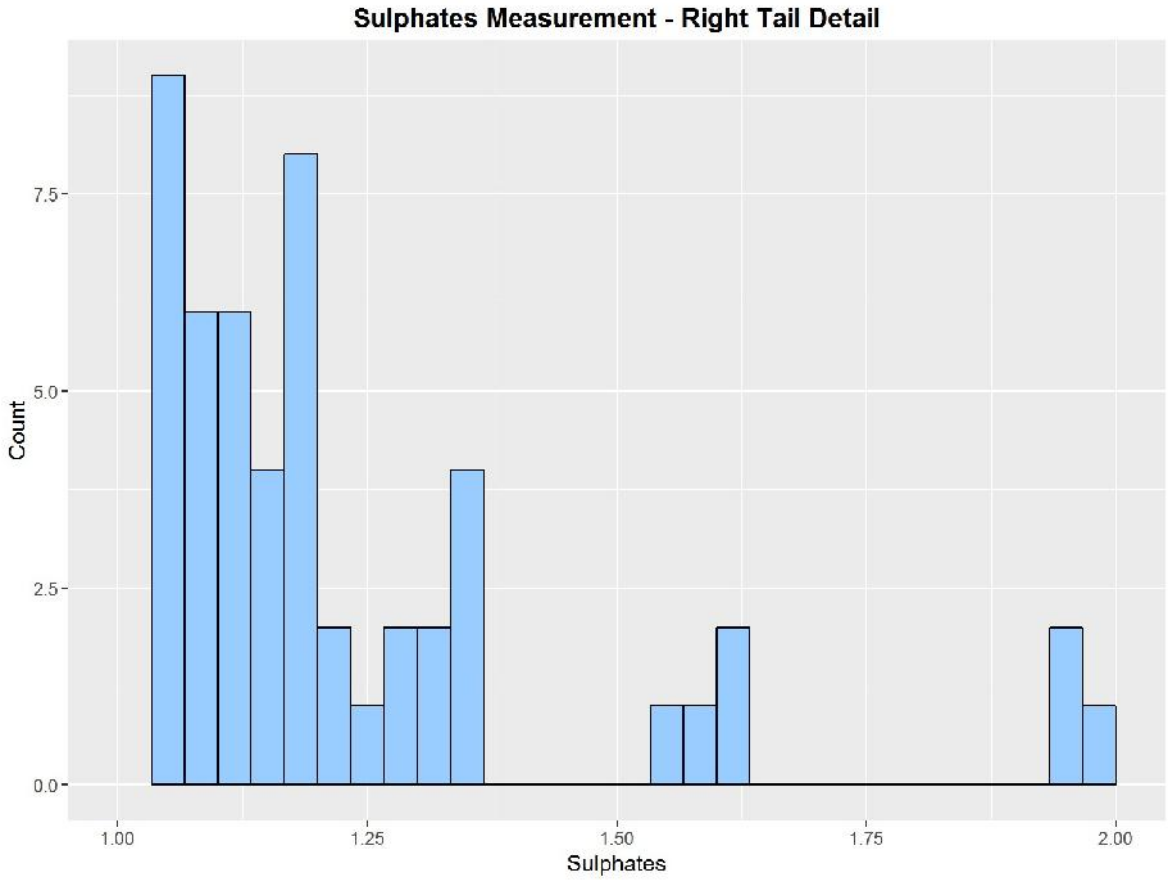

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    7.00   14.00   15.87   21.00   72.00
```

That is true also for the total sulphur dioxide concentration; in this case there are observations up to 6 times the average values. The maximum values would suffice in the U.S. and Australian but not in Europe. This is particularly important because the red "Vinho Verde" in produced in Portugal (EU).




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.00   22.00   38.00   46.47   62.00  289.00
```

With the sulphates, there are observations up to 3 times the average.




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3300  0.5500  0.6200  0.6581  0.7300  2.0000
```

A possible solution to the sulphites problem could be the implementation of an Organic way of producing wine.

Bivariate Plots Sections and Analysis

Correlation

In this section we will study the relationships between two given variables, giving special attention to the variable Quality (or Grade).

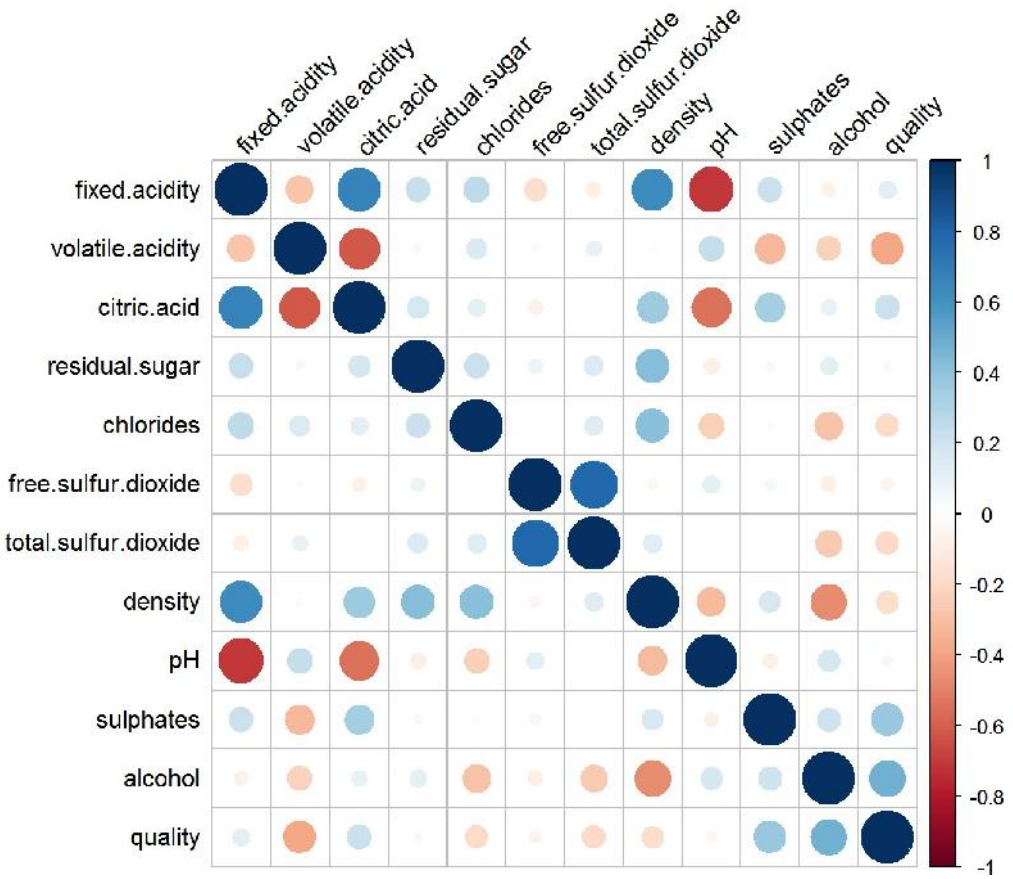
As a summary, we have obtained the correlation between the quality and the rest of variables:

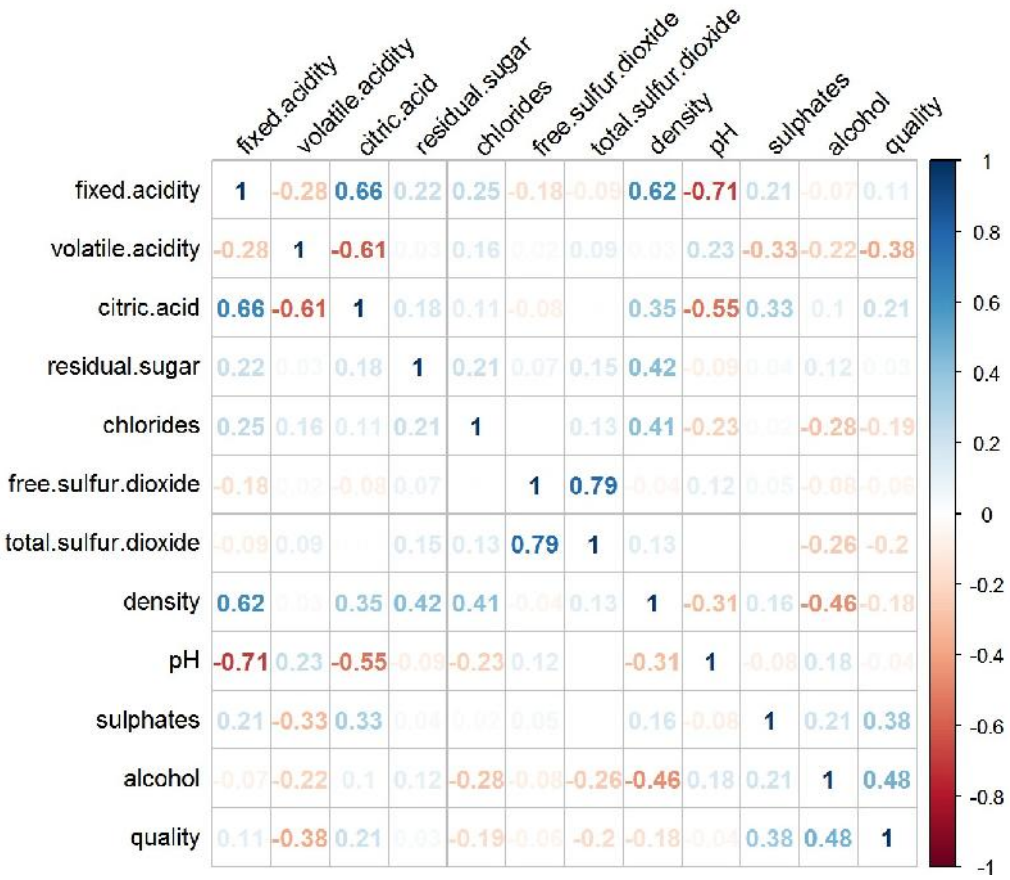
```
##                              [,1]
## volatile.acidity     -0.39055778
## citric.acid           0.22637251
## residual.sugar        0.01373164
## chlorides            -0.12890656
## free.sulfur.dioxide  -0.05065606
## total.sulfur.dioxide -0.18510029
## density              -0.17491923
## pH                   -0.05773139
## sulphates             0.25139708
## alcohol               0.47616632
```

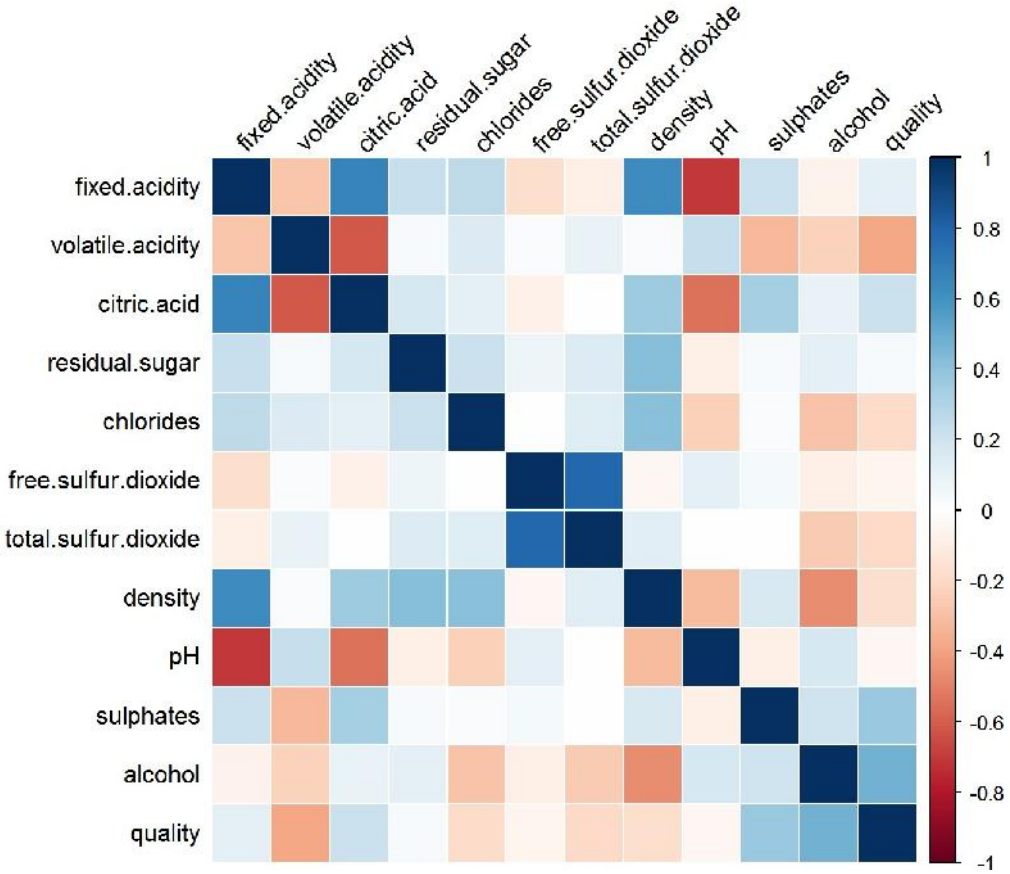
We can easily notice that the quality is most positive correlated with the alcohol (+0.48), followed at distance by sulphates (+0.25) and citric acid (+0.23).

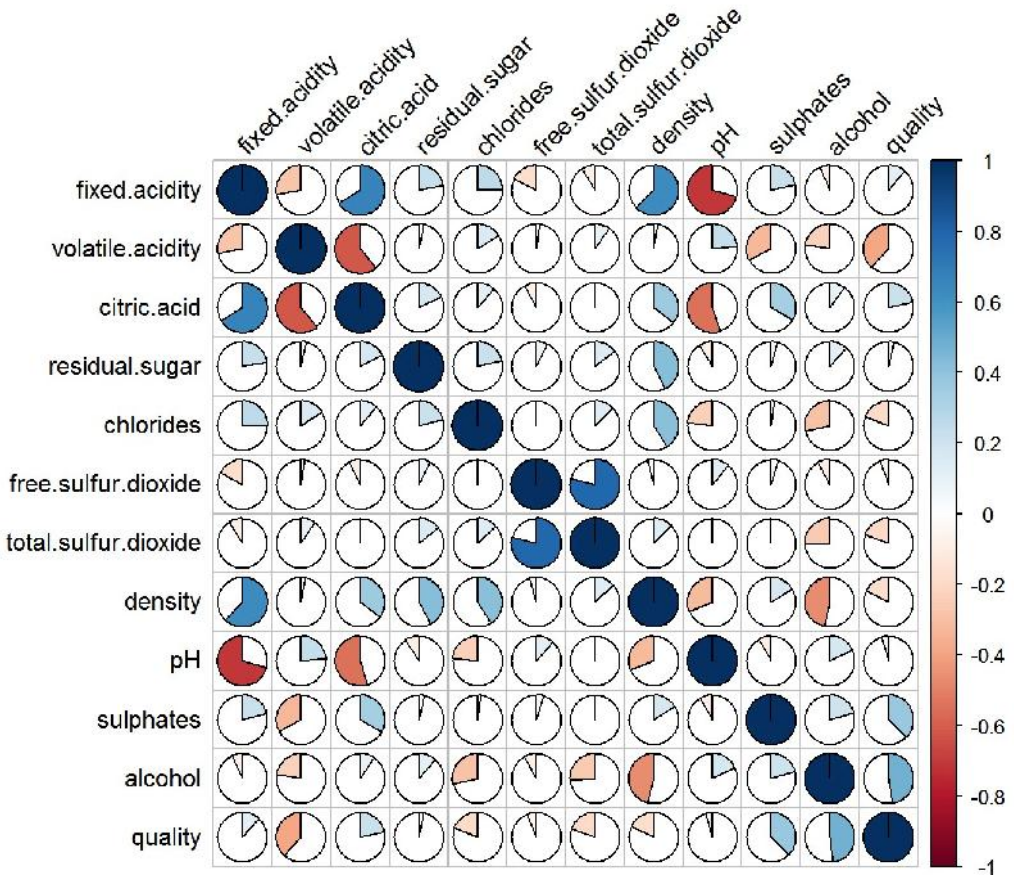
On the negative correlation side, we can find the volatile acidity (-0.39), the total sulfur dioxide (-0.18) and the chlorides (-0.13) but all of them with relatively weak correlation figures.

Let us see now some graphical representations of cross correlation among all variables. We have provided several graphical possibilities; please use the one more suitable for your taste.




fixed.acidity
volatile.acidity
citric.acid
residual.sugar
chlorides
free.sulfur.dioxide
total.sulfur.dioxide
density
pH
sulphates
alcohol
quality
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


fixed.acidity
volatile.acidity
citric.acid
residual.sugar
chlorides
free.sulfur.dioxide
total.sulfur.dioxide
density
pH
sulphates
alcohol
quality
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

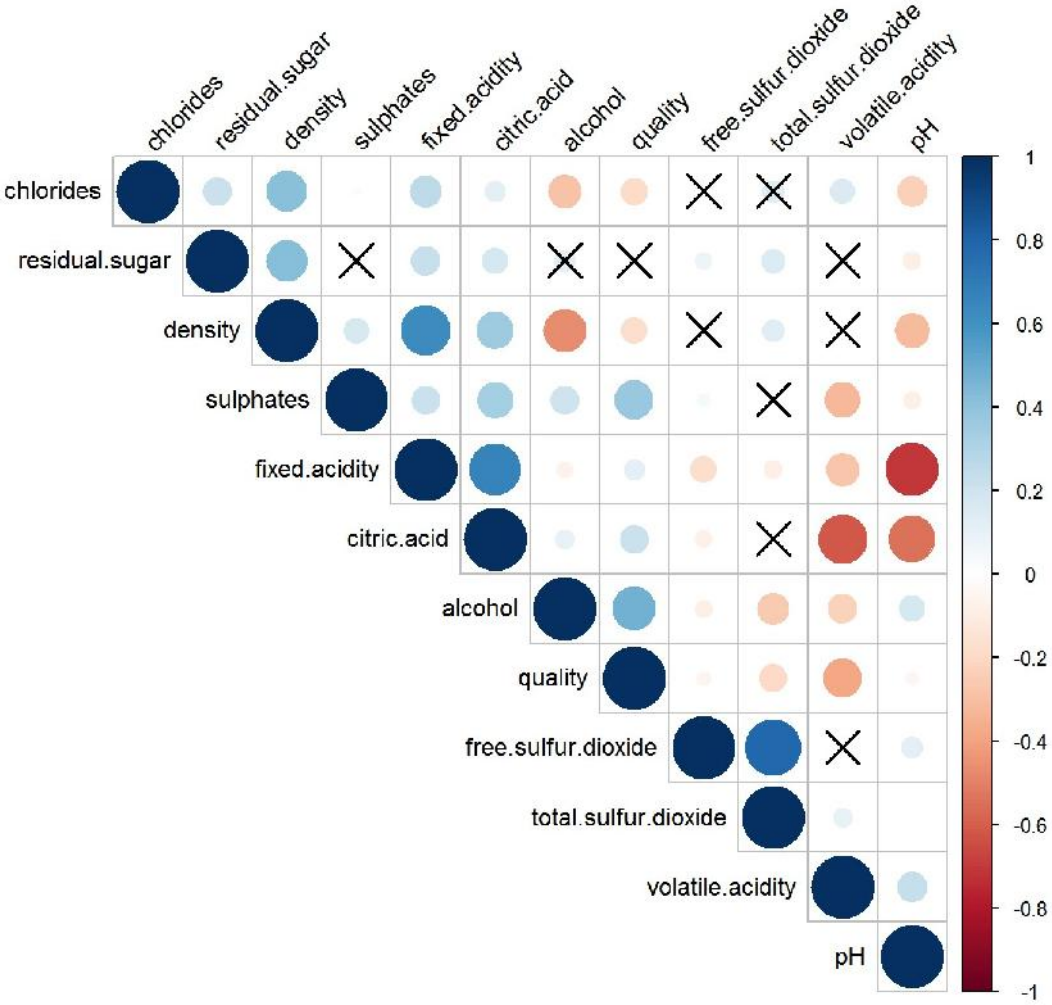


We found the expected strong correlations between the variables of a similar or different nature:

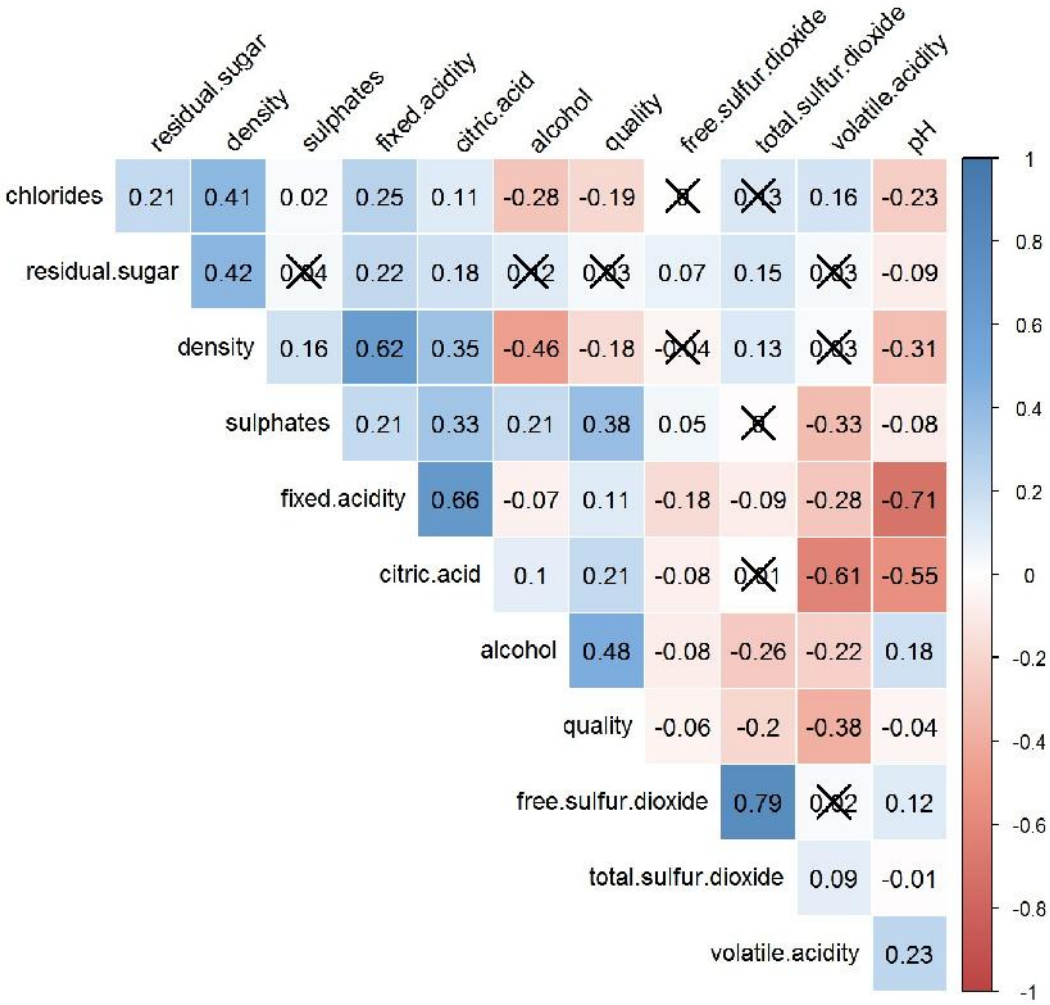
```
+ SO2 = total snd free sulfur + 0.79
+ Acidity = fixed acidity and citric acid + 0.66
- pH = fixed acidity and pH -0.71
- Acidity = citric acid and volatile acidity -0.61
- pH = citric acid and pH -0.55
```

The next two plots show basically the same information as the above ones. Just in this case we have added some statistical information: the correlation values marked with "X" are not statistically significant @ 95%.

First graphic, with visual information:

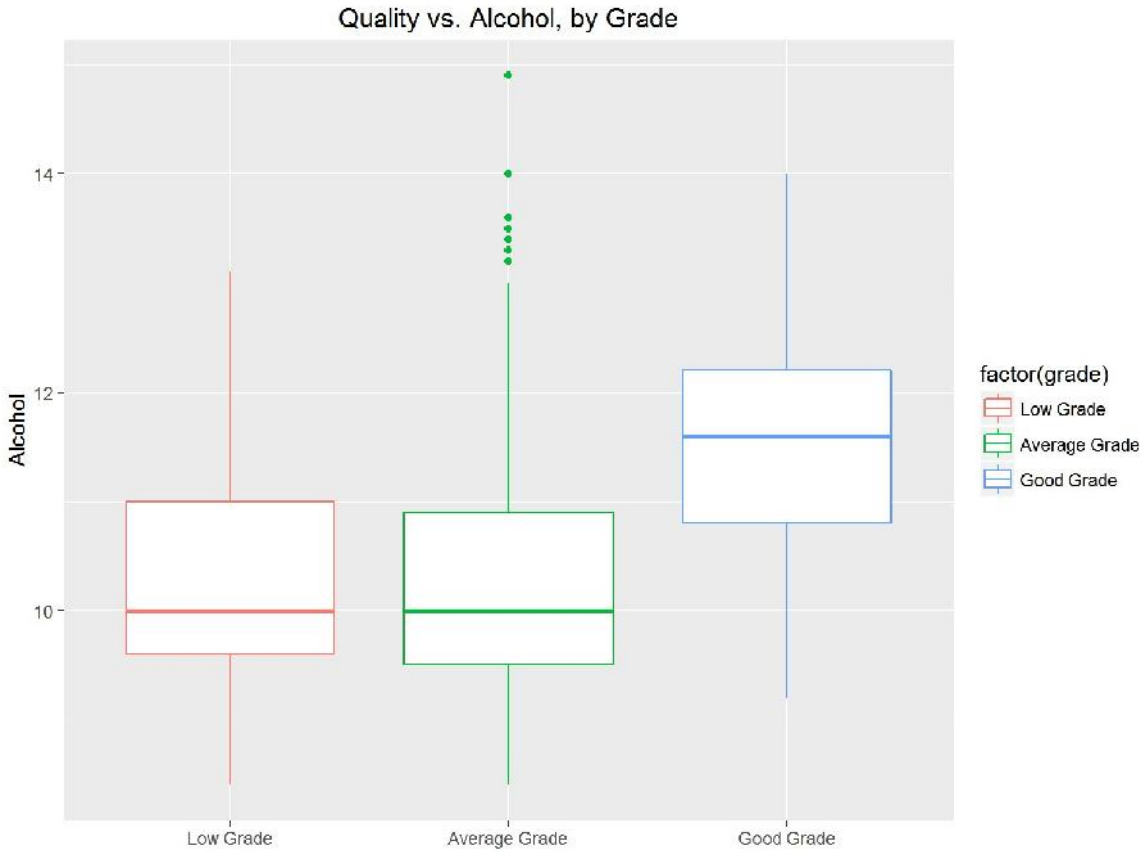


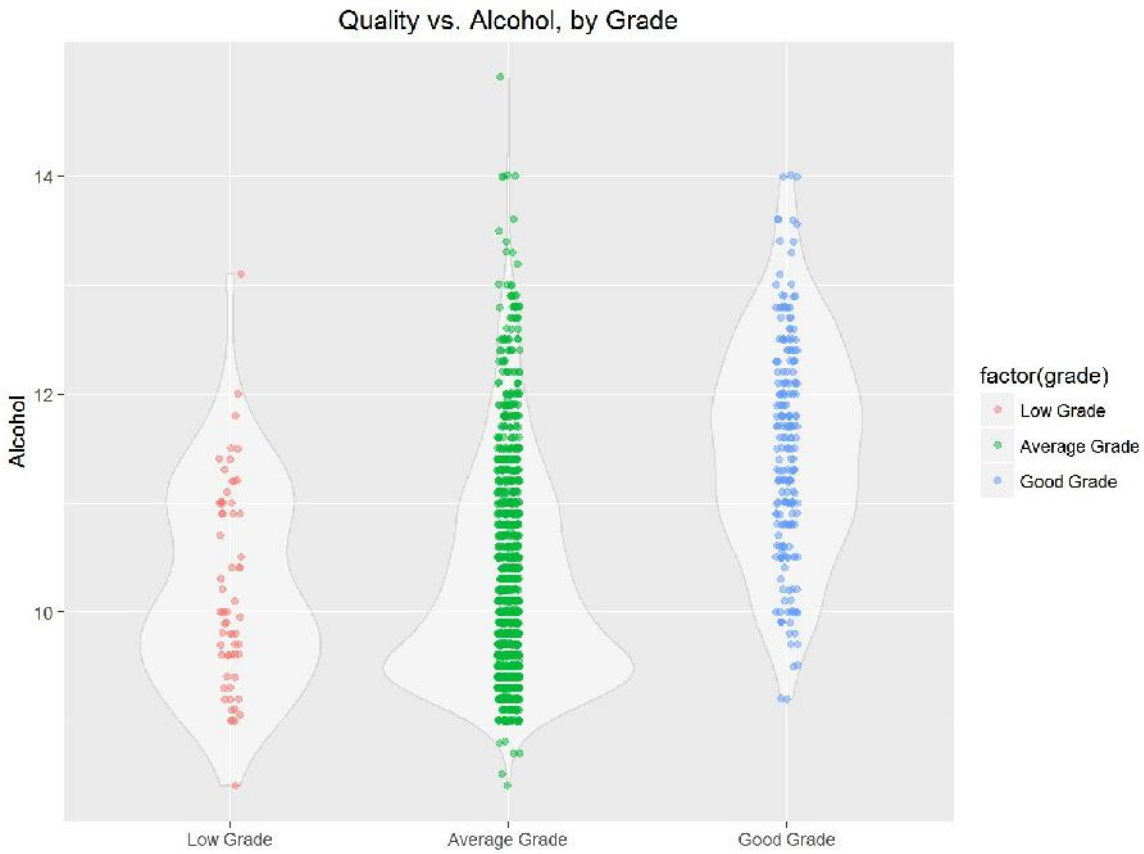
Second graphic, with numerical information (real correlation figures):



Finally, we plotted the variable most positively correlated with quality, that is the alcohol volume in %.

We have used the new variable that we added to the dataset before, "Grade", and set it to an ordered factor for better comprehension of the results.

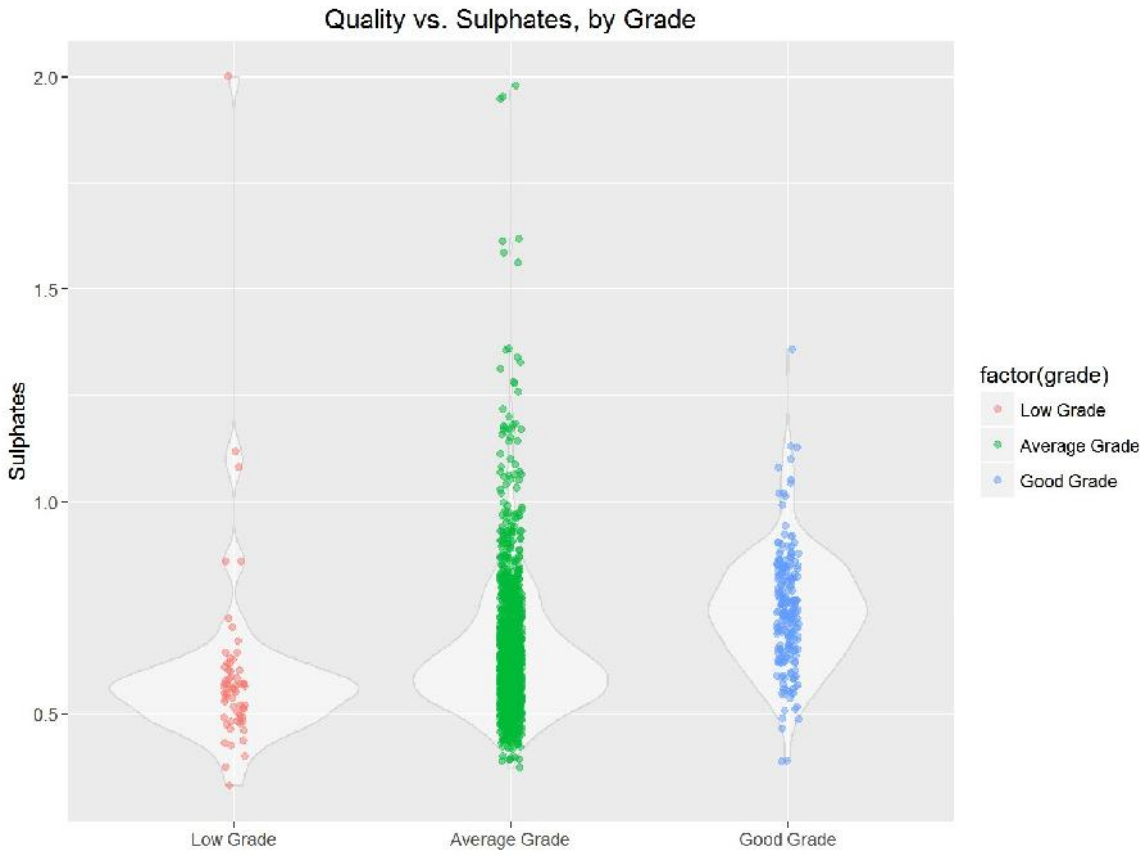




The more alcohol a given wine has, the more chances to get a good grade.

The average value of alcohol for the low and average graded wines is very similar.

We could find some outliers in the plots above, especially in the average grade, but we found even more outliers when we plotted the second most positive correlated variable with quality, sulphates:



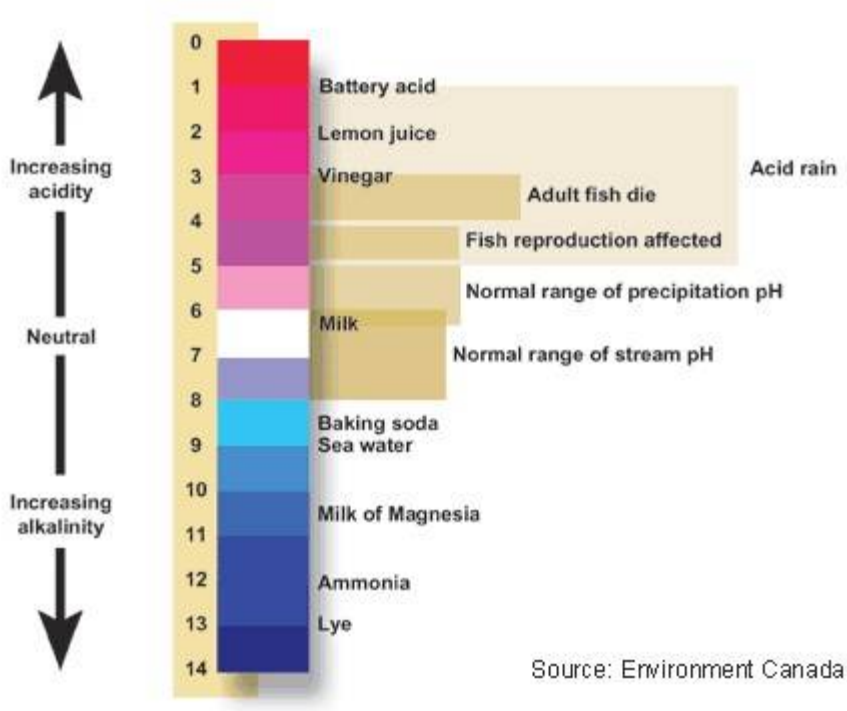
For both the low and normal grade, the sulphates concentrate a little bit above the 0.5 mark, while the good graded wine at the 0.75 concentration mark.

We would like to point out some not logical relationship between the pH and the acidity of the wine.

The general pH levels that the wines from our database have vary from 2.74 to 4.01.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.740   3.210   3.310   3.311   3.400   4.010
```

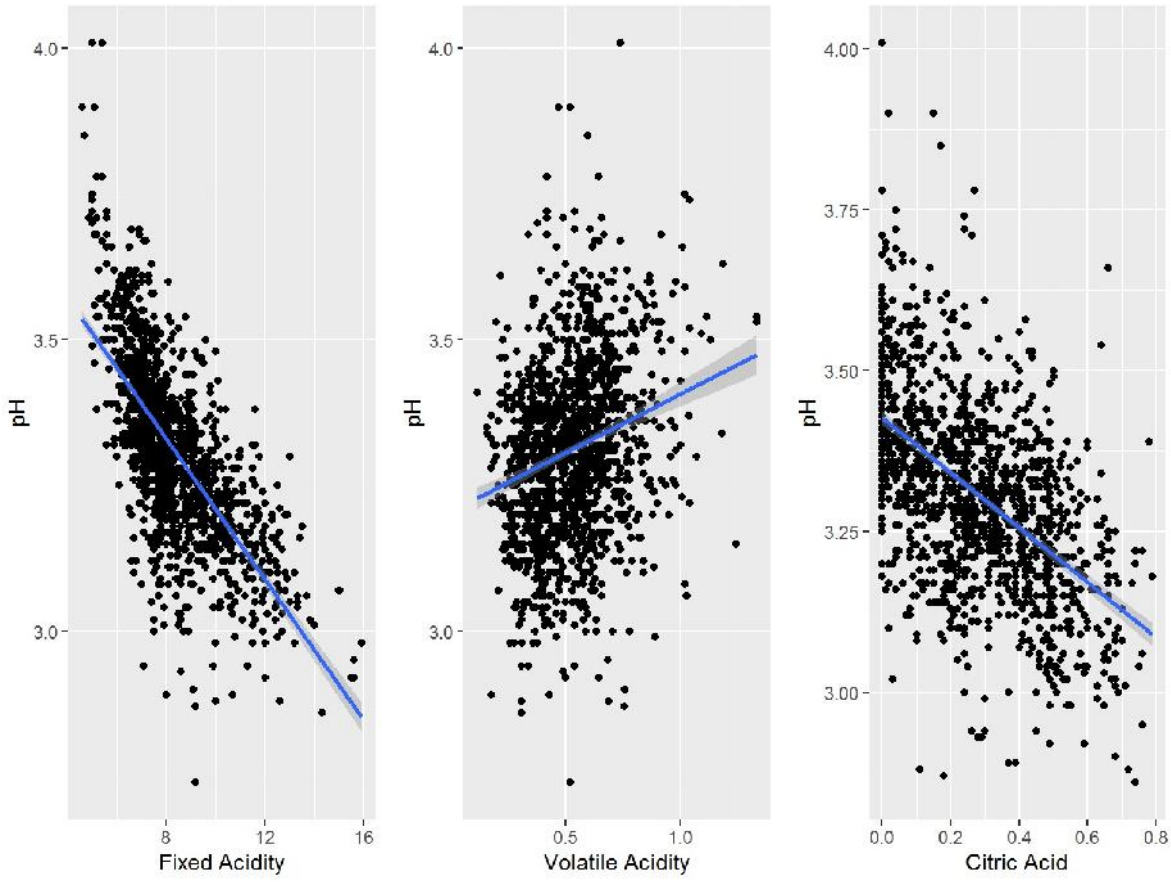
Therefore, we are in the "acid" part of the pH scale:



We could say that the lower the pH, the higher the acidity. However, that is not true in our dataset values.

We have plotted the pH against the "acidity" variables, and while the fixed and the citric acid variables behave as expected (less pH, more acidity), the volatile acidity is going the other way round (less pH, less acidity).

Perhaps the data and the procedures of measurement should be revised; right now we do not have a clear explanation to this phenomenon.



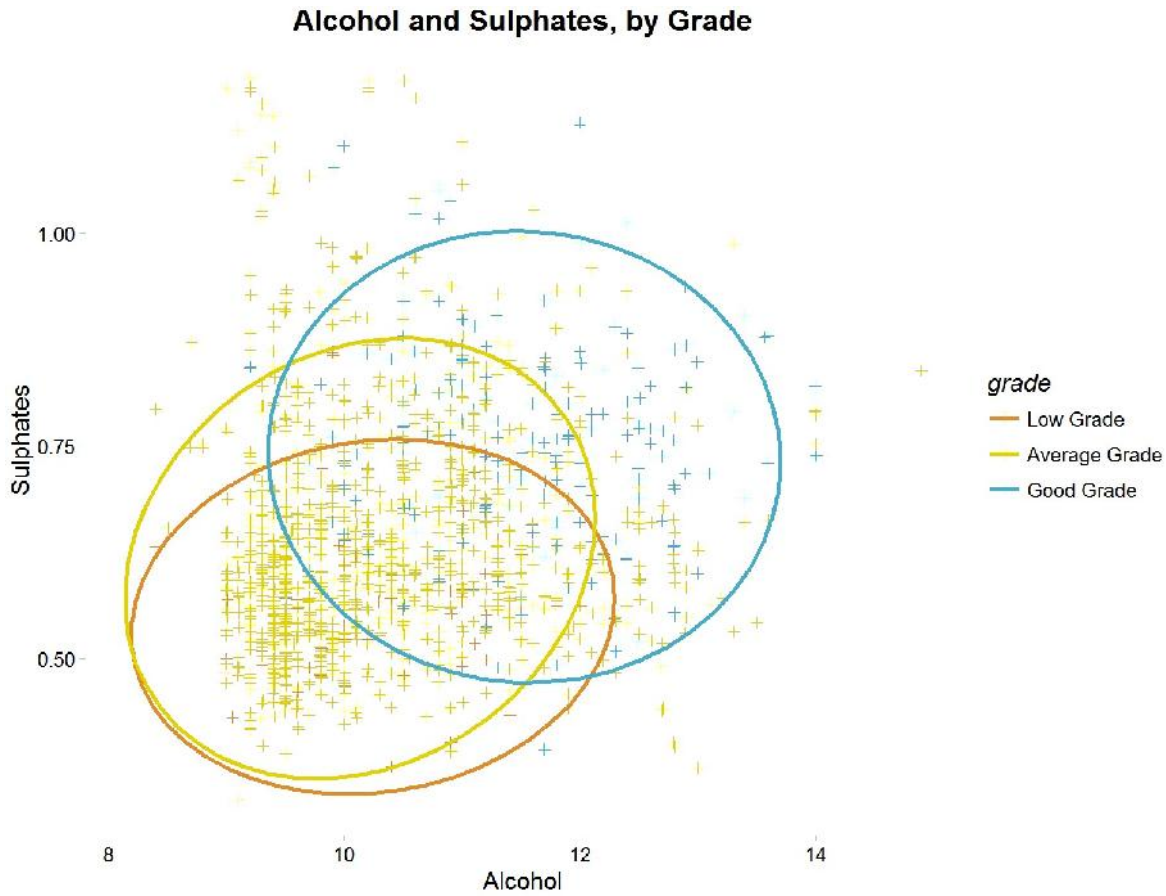
Multivariate Plots and Analysis Section

In this section we will study the interaction among several variables.

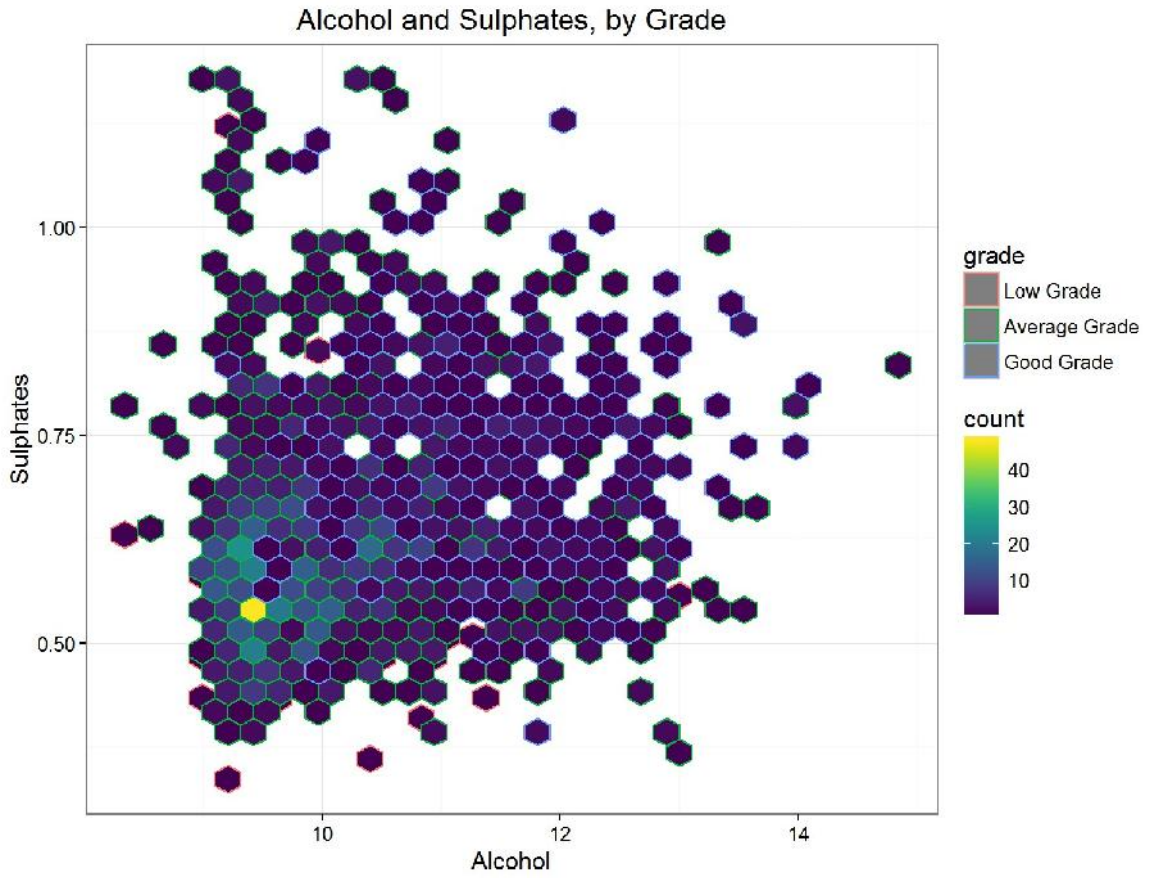
We can start with the two variables most positively correlated with quality, the alcohol and the sulphates. We will use the new variable "Grade" for getting a clearer understanding.

This graphic clearly shows that good graded wines (blue circle, blue dots) have in general more alcohol and sulphates than low and average graded ones.

While the amount of alcohol is very similar in the low and average categories, the average graded wines contain on average more sulphates than the low graded ones.

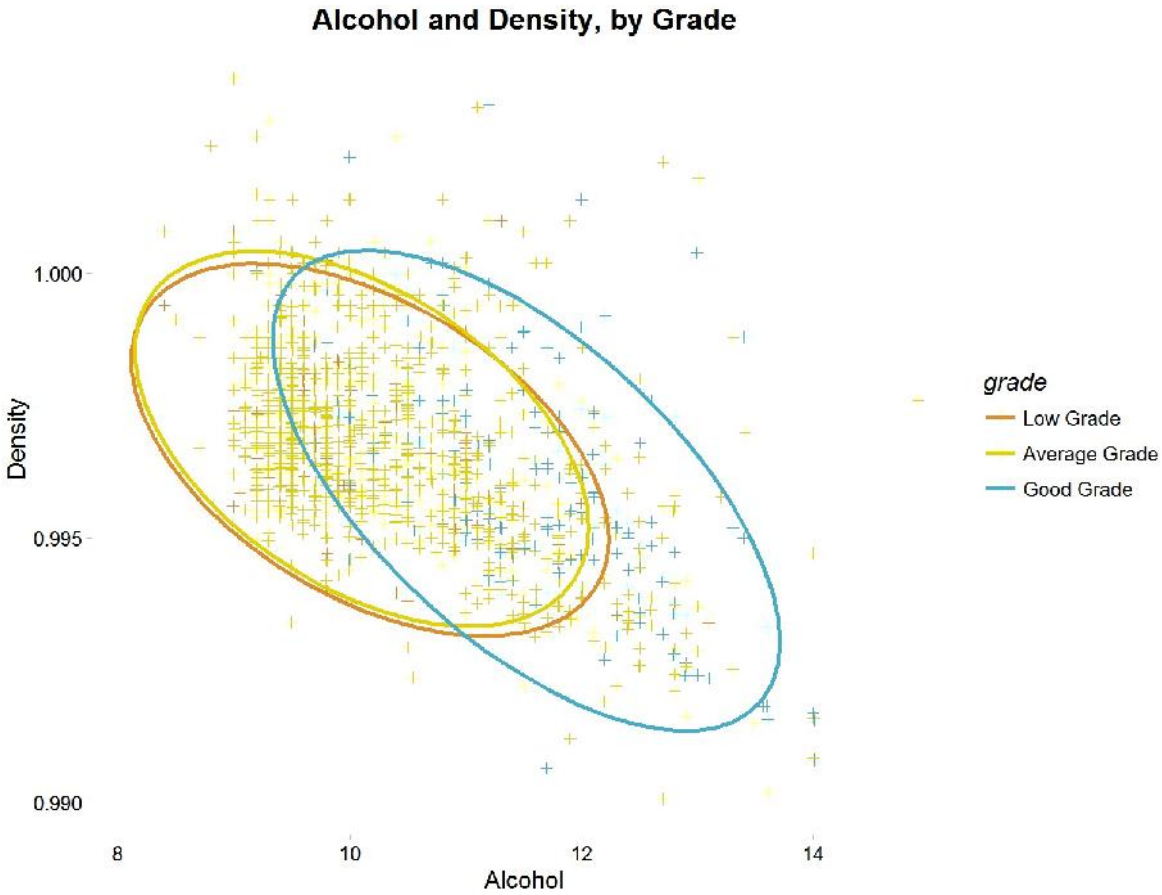


Another way to look at it:

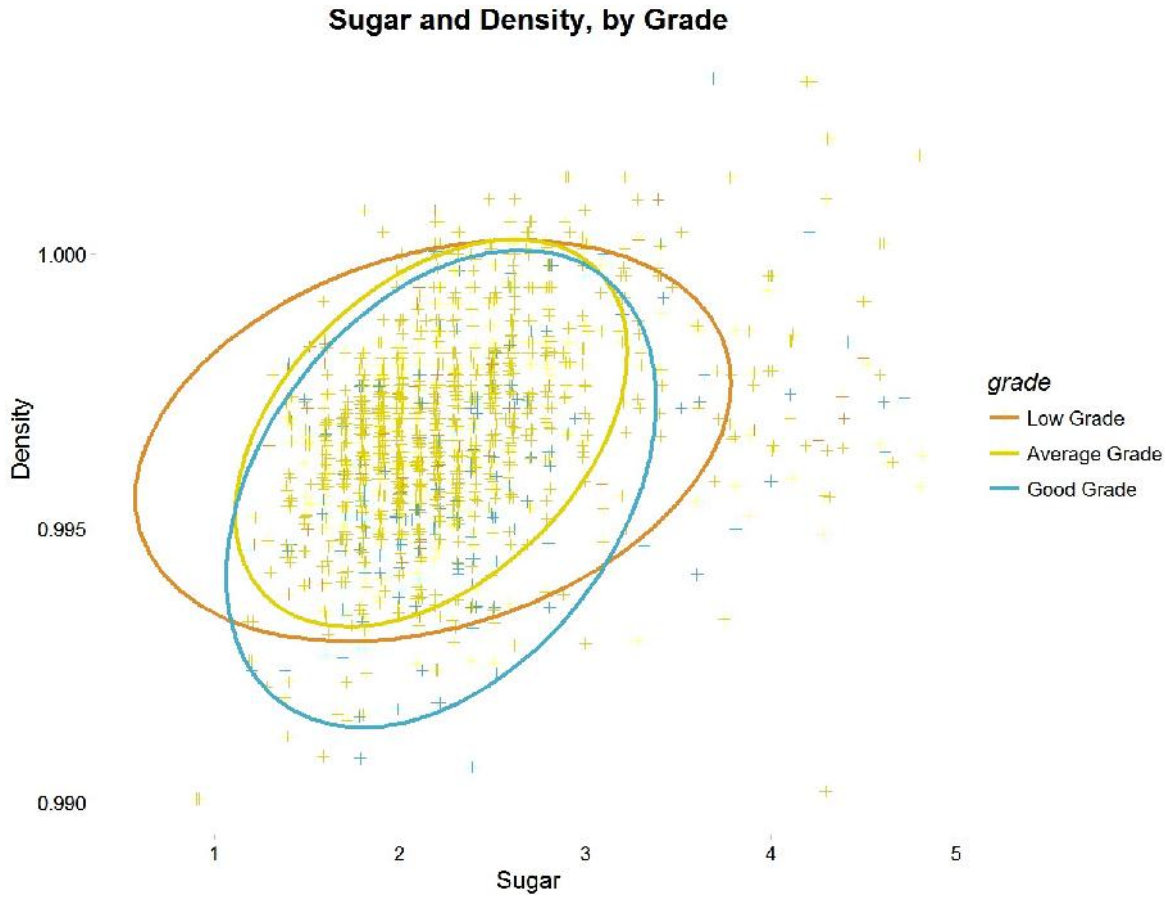


One might think of other variables that could have a great deal of influence on what grade a wine obtains, like if the wine is too sweet or not, or if it is very dense or not, for example:

- the combination of alcohol and density: good wines tend to be less dense and low and average wines have practically the same alcohol and density parameters.



- the combination of sugar and density: like before, good wines tend to be less dense, and about the sugar, we can find low graded wines with both little or much sugar. Lots of outliers with a great amount of sugar (more than 3.7), and high density (more than 1).

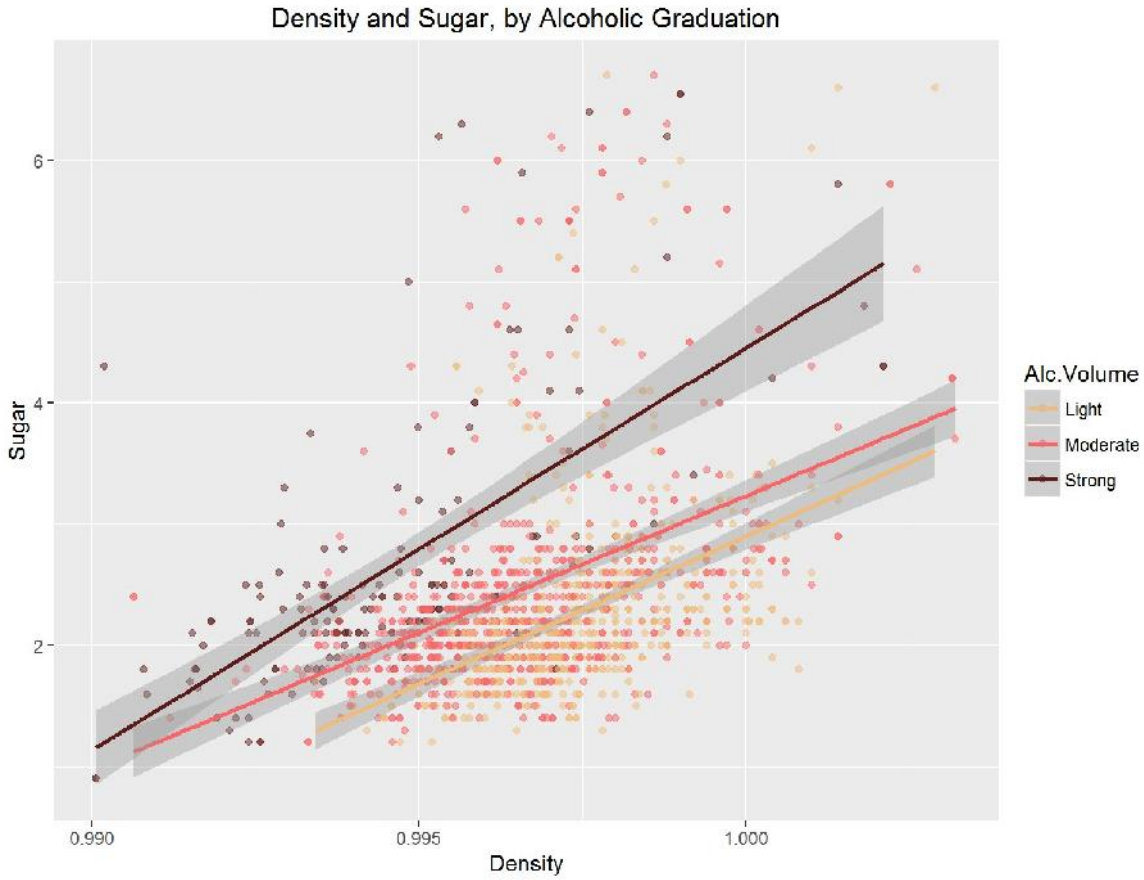


We have also created a new variable to categorize the amount of alcohol present in the wines, called "Alc.Volume":

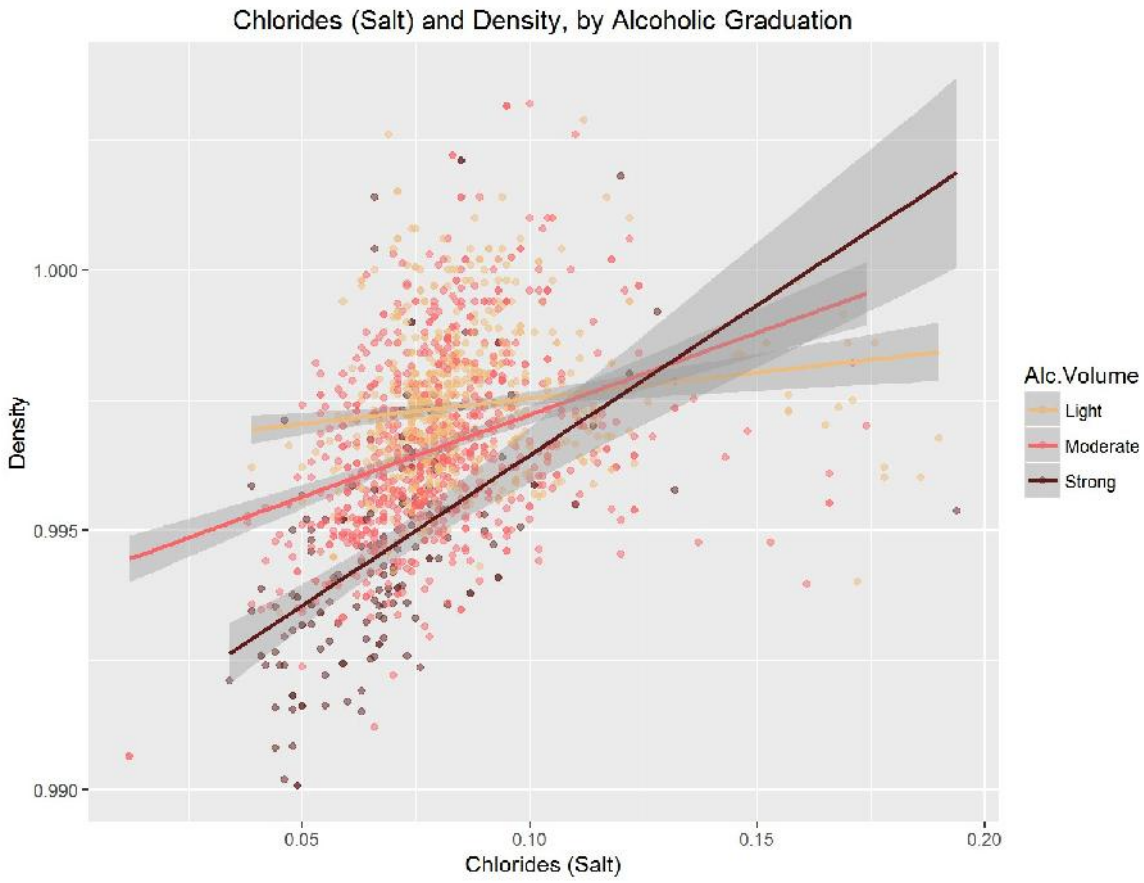
```
Light: Alcohol (% by volume) below 10%.
Moderate: Alcohol (% by volume) higher than 10% and below 12%.
Strong: Alcohol (% by volume) higher than 12%.
```

This new variable will be used in the next plots:

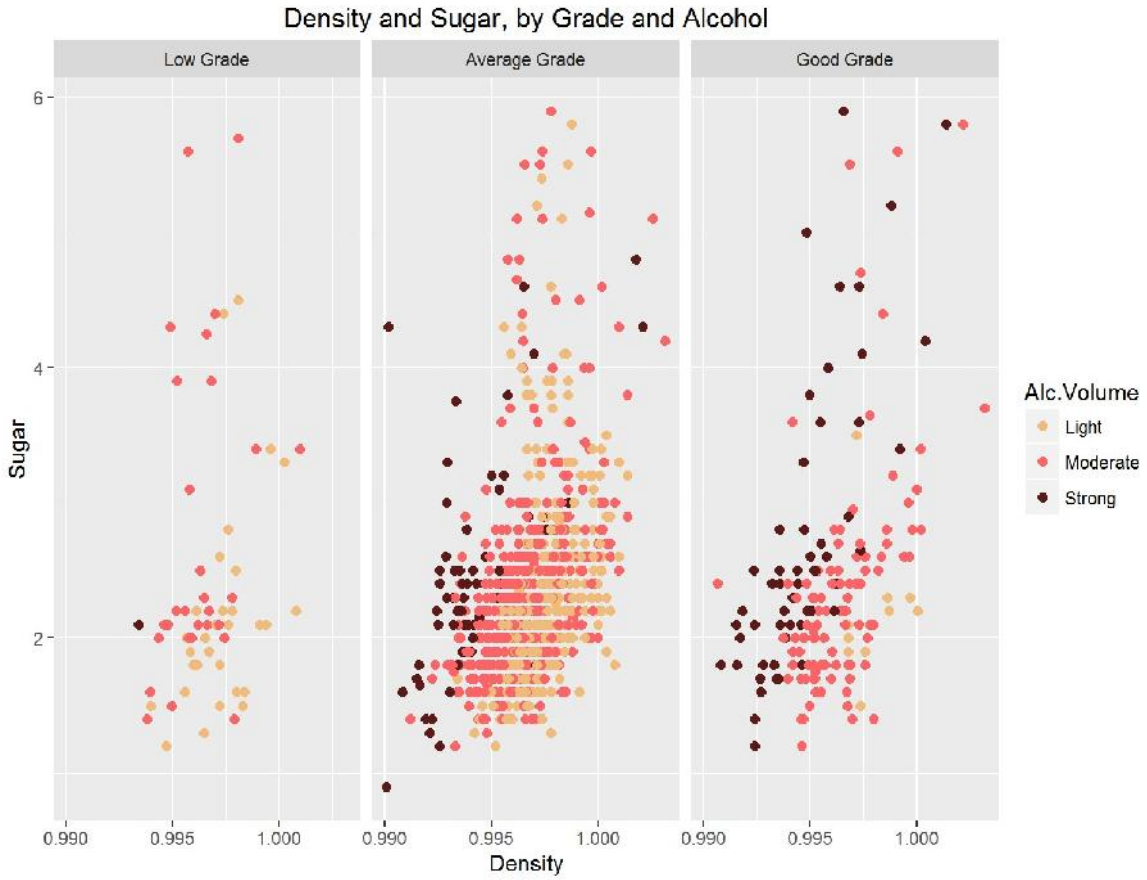
- Density vs. Sugar: the strong wines tend to have more sugar and more density than then moderate and light ones.



- the more salt, the more density, for every alcohol category.

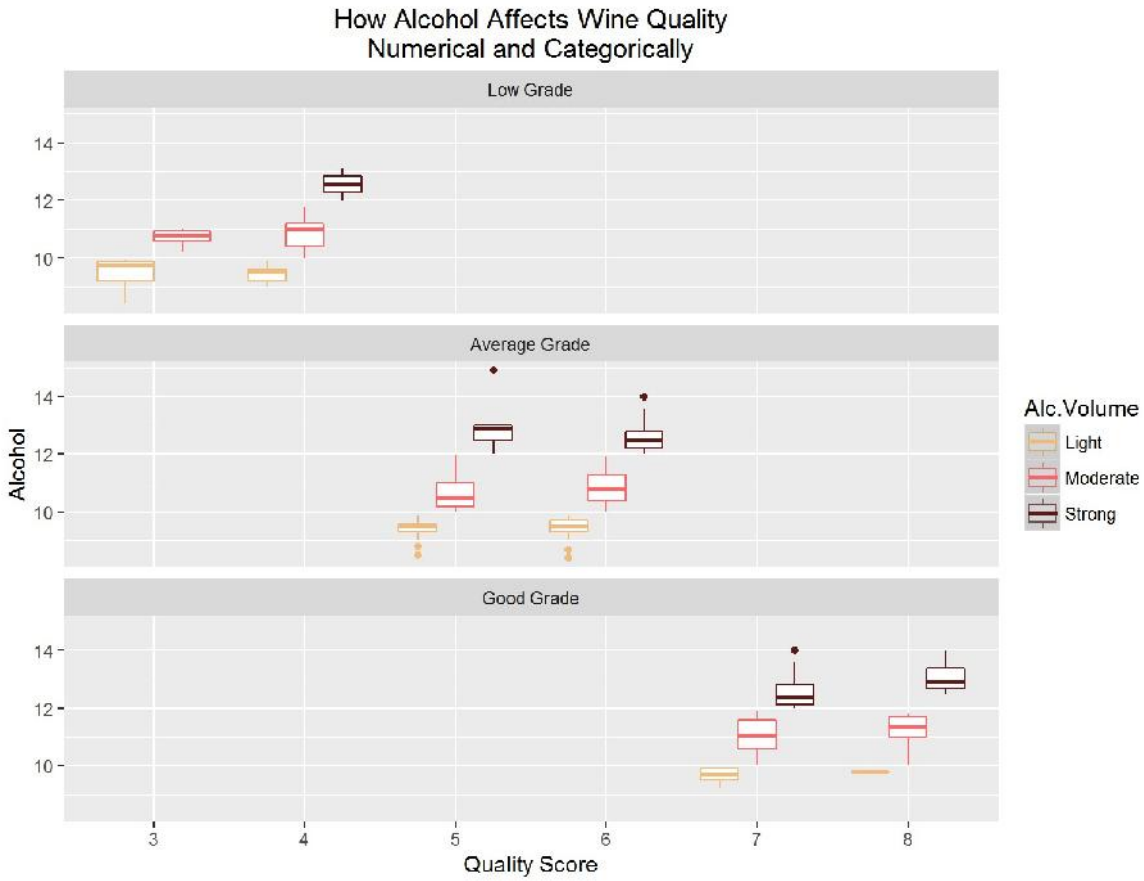


Mixing the Grade and the Alcohol categories:

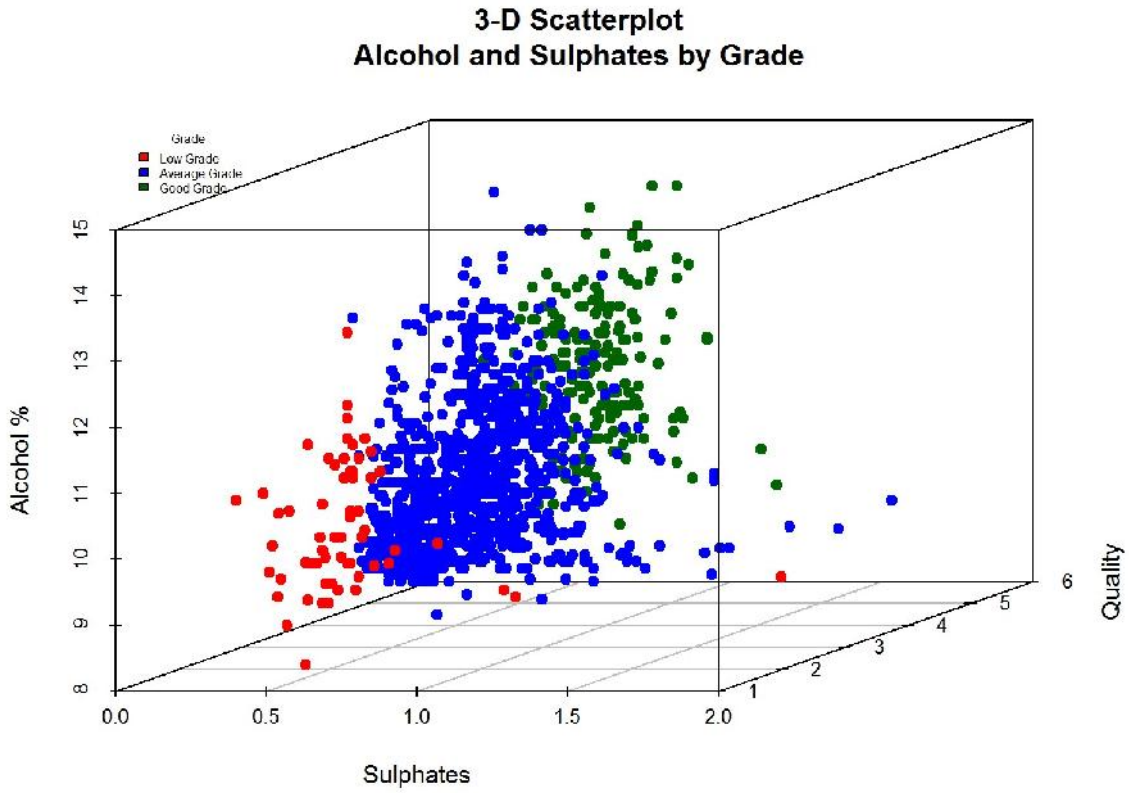


In the nest plot we can see how alcohol affects the quality of the wine, both in numerical values as well as categorized by light, moderate and strong alcohol concentration and low, average and good grade.

The pattern repits itself across the mentioned categories.



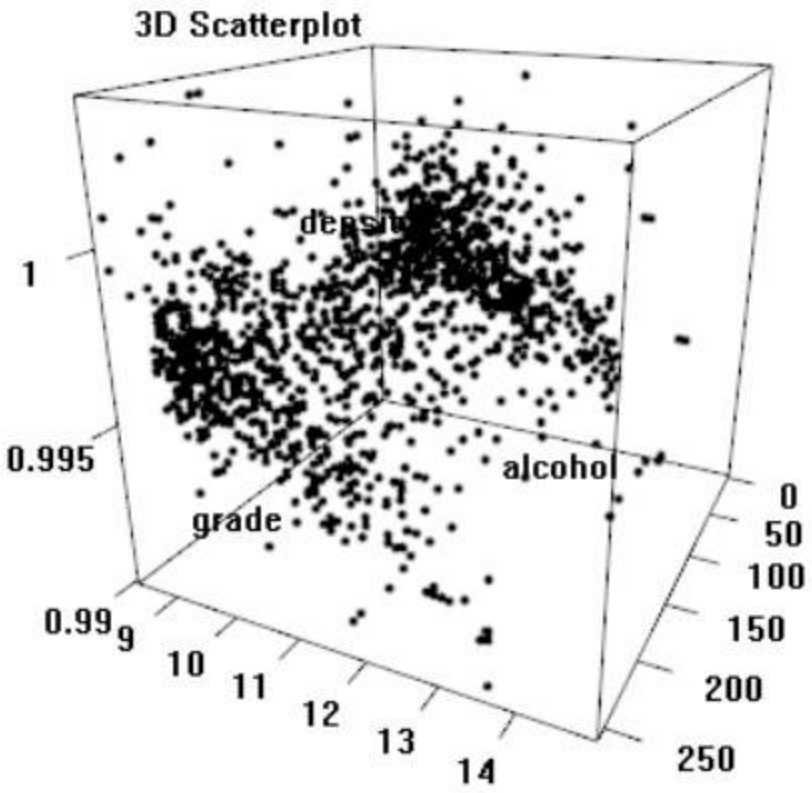
As a bonus, we add a couple of 3D visualizations just to have a taste of the possiblities:



3D mouse-spinning plots: rotate them with your mouse !

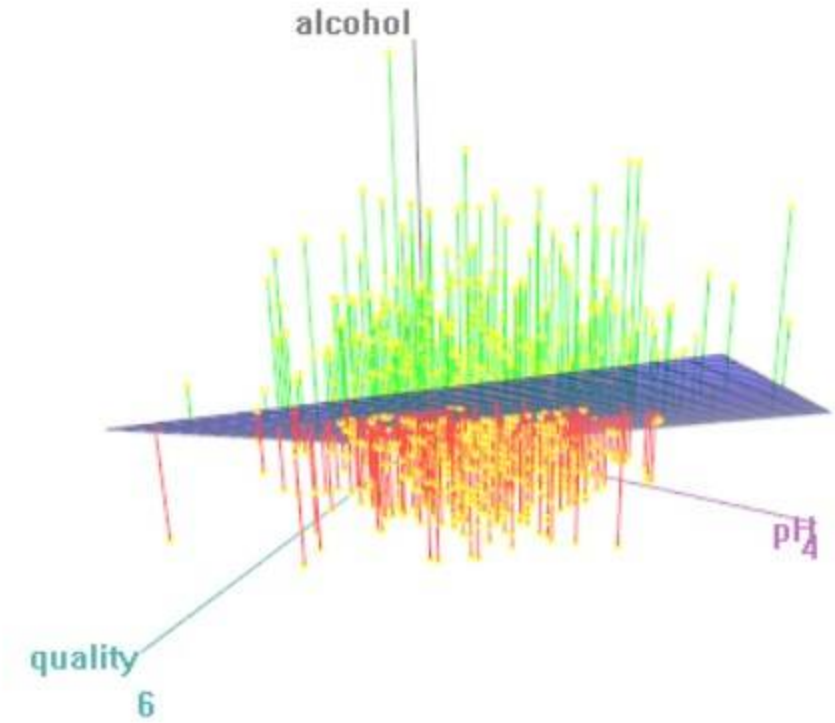
Please, try those on your R interpreter:

```
library(Rcmdr)
attach(wine)
plot3d(alcohol,density,grade, main="3D Scatterplot", size = 4)
```



or:

```
library(Rcmdr)
attach(wine)
scatter3d(alcohol,density,chlorides, main="3D Scatterplot", size = 4)
```

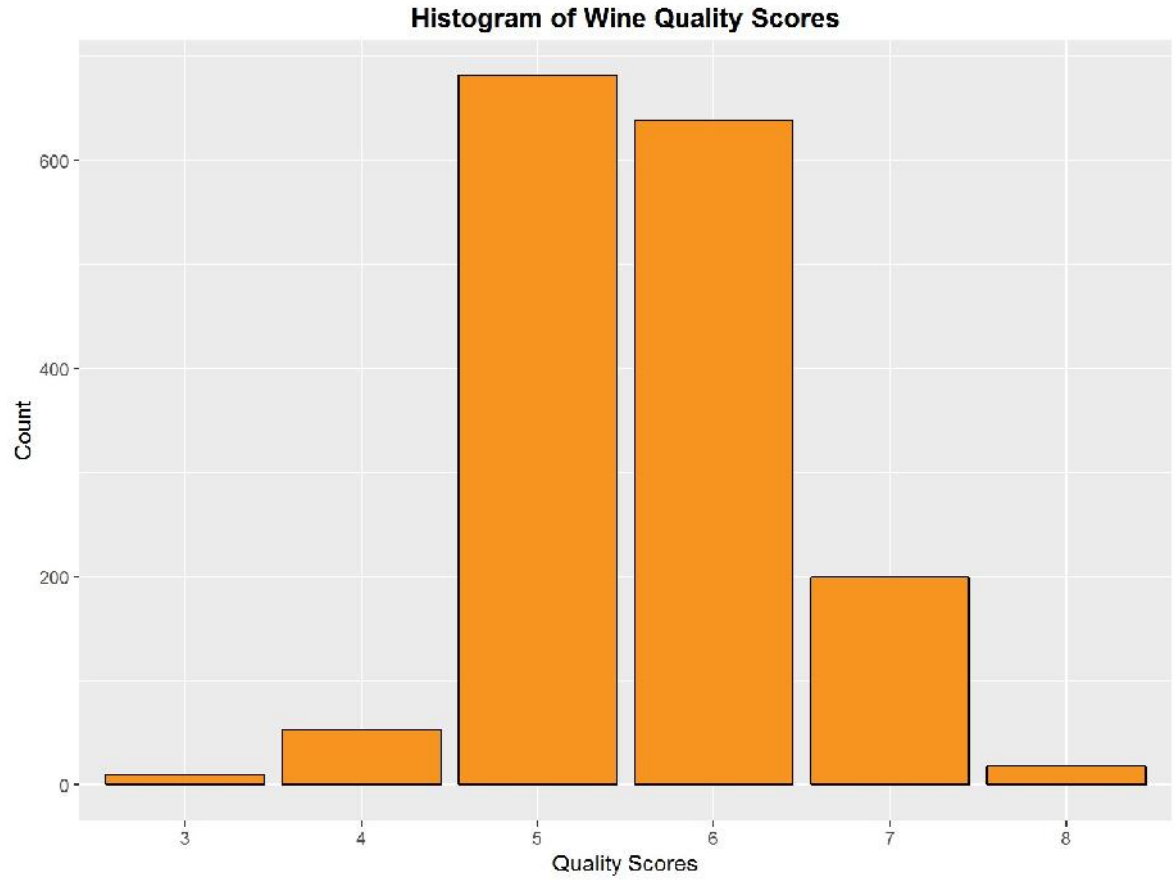


Final Plots and Summary

Plot One

This is the simplest plot but nevertheless it gives us a good view of the results of the output variable, the quality score of the wines.

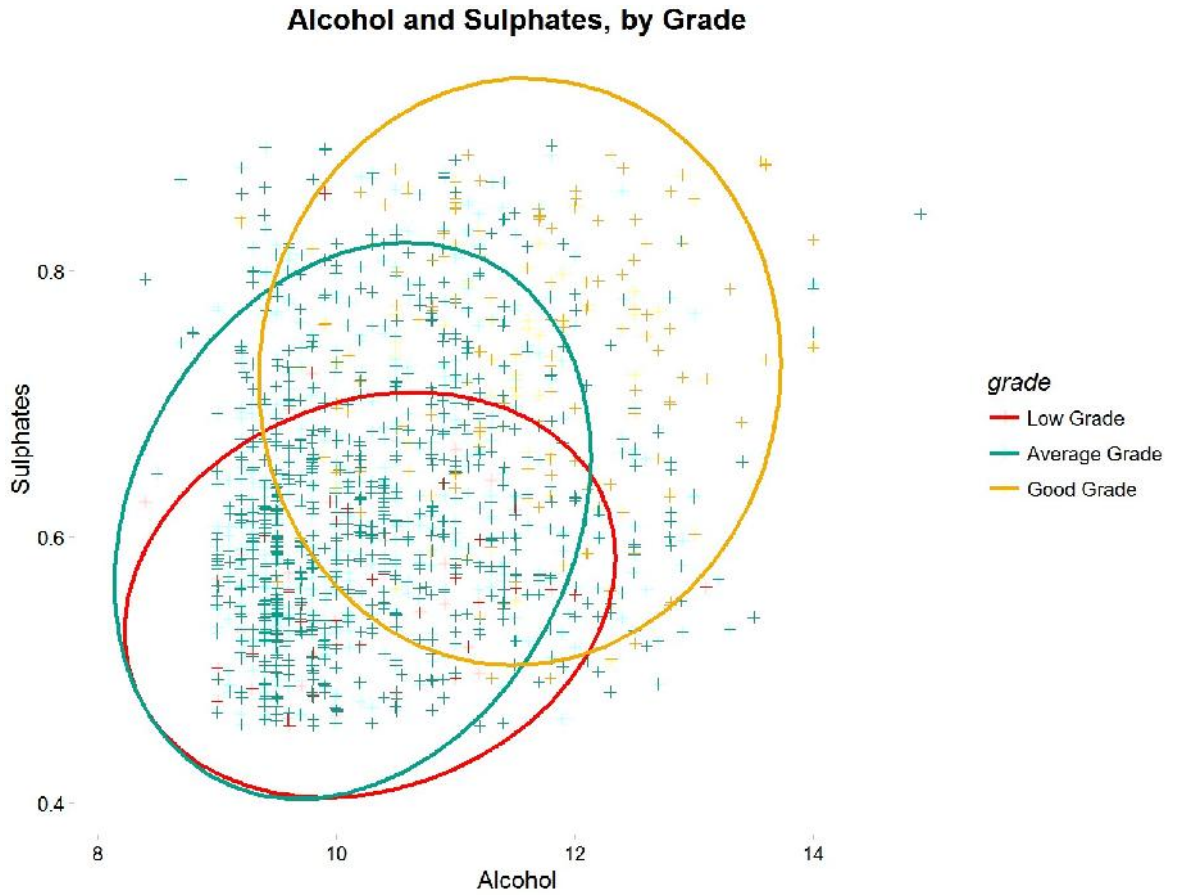
Most of the observations fall into the 5 and 6 quality score, followed by far by 7 quality score. That means than the wines tasted where judged to be of average quality with only a few of them either bad (3, 4 scores) or good/very good (7, 8 scores).



Plot Two

This plot shows the relationship of the two most positively correlated input variables (alcohol and sulphates) with the output variable quality. The quality scores have been categorized from numerical values (0-10) to Low, Average and Good grades for better understanding.

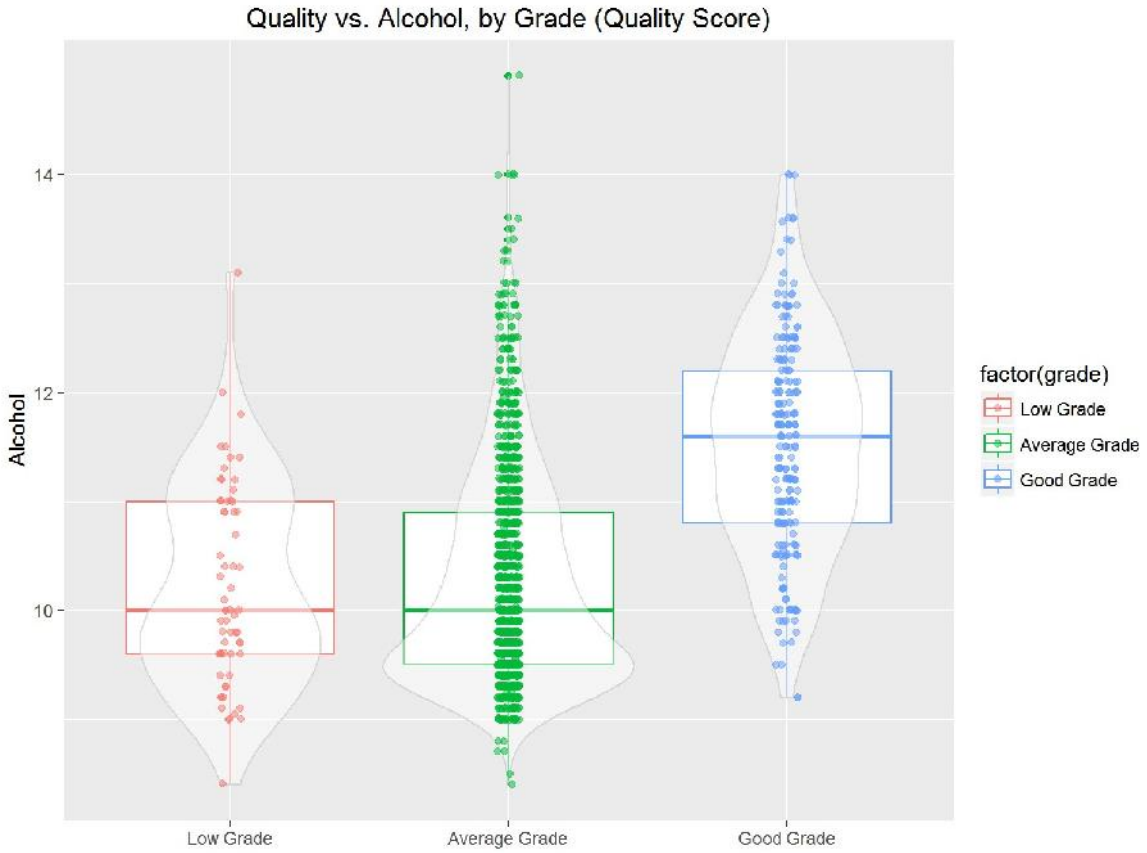
It is clear that the good graded wines contain more alcohol and sulphates than the average and low graded ones. The main difference between the average and the low graded wines appears to be the concentration of sulphates, being the alcohol level very similar between them.



Plot Three

The third plot is a mix between a "boxplot" and a "violin" that shows the relationship between the most positively correlated input variable (alcohol) and the output variable (quality). The quality variable has been categorized into "Low", "Average" and "Good" grade because it is better to understand for anyone. The "violin" shapes show the distribution of the data for each category, where the widest zone corresponds to where more data accumulates, while the boxplots show statistical data about the 1st, 2nd, 3rd and 4th Quantiles.

For instance, for the average grade, the data concentrates below the 10% alcohol volume and the mean is almost exactly 10°. Nevertheless, we can find wines with an alcoholic volume even higher than 14°. The low grade wines, although having the mean similar to the average grade (about 10°), show a different distribution, this time more of a real "violin" shape. Now, it is pretty clear that the good graded wines have in average more alcohol, about 11,5° based on the graphic.



Reflection

Based on this analysis, we have extracted some insights:

- high concentration of wines in the middle ranges of the ranking.
- the more alcohol and sulphates a wine has the better chance to score high in quality.
- other parameters suitable to be directly tasted by the experts like sweetness, acidity and density play a small role in the scoring process.
- pH and acidity variables: while the majority behave as expected, the citric acid defies the general rule (lower pH, higher acidity).

The chemical variables covered in this study can explain a a great deal of the quality, although there are other numerous variables which come into play in order to grade a wine.

The output variable “general quality” -a subjective general scored given by the tasters- is fine but some other indicators are missed. We could think of a quality breakdown into “nose” (smell), taste and color, all of them adding point to the final score. Also, a “blind” vs. “normal” tasting to check if a well known brand can influence the taster decision - bias avoidance.

We might also think of another variables, like barrel wine vs. bottled wine - being the same wine-, professional tasters vs. connoisseurs vs. general wine consumers, and then compare que results. Different grape mixes could also affect the results; the possibilities are endless.

One might think of "price" as a factor of wine quality but we do not agree with that because it could well be the case that the most expensive wines might not be labelled as the best ones, being the price set from other variables like prestige, scarcity of grape in a given year, bad harvest…etc.

Resources

- **Stackoverflow**: http://stackoverflow.com/questions/tagged/r (http://stackoverflow.com/questions/tagged/r)
- **Github**: https://github.com/ (https://github.com/)
- **ggplot2**: http://docs.ggplot2.org/current/ (http://docs.ggplot2.org/current/)
- **The Comprehensive R Archive Network**: https://cran.r-project.org/ (https://cran.r-project.org/)
- **Statistical tools for high-throughput data analysis**: http://www.sthda.com/english/ (http://www.sthda.com/english/)
- **Sulphites in Wine**: http://www.morethanorganic.com/sulphur-in-the-bottle (http://www.morethanorganic.com/sulphur-in-the-bottle)
- **Wikipedia, pH**: https://en.wikipedia.org/wiki/PH (https://en.wikipedia.org/wiki/PH)
- **CSS for Knitr**: http://jessachandler.com/2015/06/customize-r-knitr-html-output-easily/ (http://jessachandler.com/2015/06/customize-r-knitr-html-output-easily/)