Project 5

Mark Ciganovic

August 26, 2017

The Analysis of the Quality of Red Wines Based Upon Their Chemical Constituents

Introduction: This paper will analyze the quality (score between 0 and 10) of red wine based upon certain attributes. These attributes are fixed acidity (tartaric acid - g / dm^3), volatile acidity (acetic acid - g / dm^3), citric acid (g / dm^3), residual sugar (g / dm^3), chlorides (sodium chloride - g / dm^3, free sulfur dioxide (mg / dm^3), total sulfur dioxide (mg / dm^3), density (g / cm^3), pH, sulphates (potassium sulphate - g / dm3), and alcohol (% by volume). The dataset consists of 1599 records for analysis in which no row(instance) has missing information for any column(attribute). The way to understand the dataset is to see quality as the output or dependent variable, whereas the eleven attributes are the input or independent variables that affect quality.

Data Processing

The steps in processing the data in list form

1. Download the file from link provided, along with all the necessary libraries for the analysis.

```
fileURL <- "https://s3.amazonaws.com/udacity-hosted-downloads/ud651/wineQualityReds.csv"
if (!file.exists("wineQualityReds.csv")){
  download.file(fileURL, "wineQualityReds.csv")
}
library(tidyverse)
library(corrplot)
library(gridExtra)
library(moments)
```

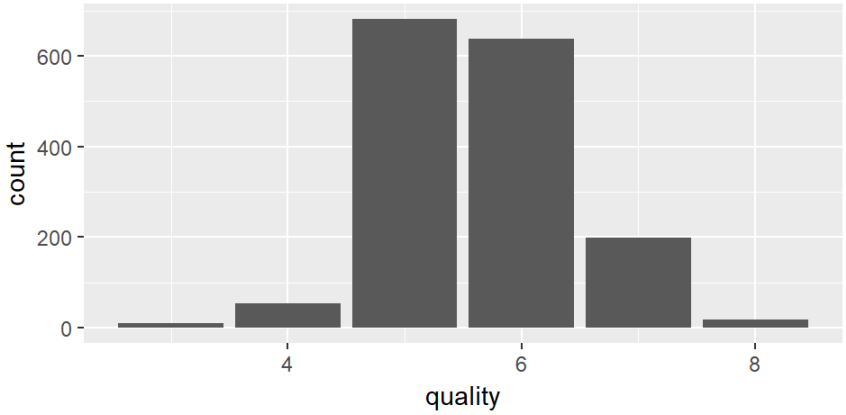
2. Read the file into a dataset called winedata, setting all the attributes to proper data type.

```
winedata <- read_csv("wineQualityReds.csv", col_types = "_ddddddddddi")
options(digits=2, tibble.width = Inf, tibble.print_max = Inf)
```

Data Analysis - Stream of Consciousness Analysis

1. Univariate Analysis: I will analyze the quality variable by both using a bar graph and descriptive statistics.

```
ggplot(aes(x = quality), data = winedata) + geom_bar()
```

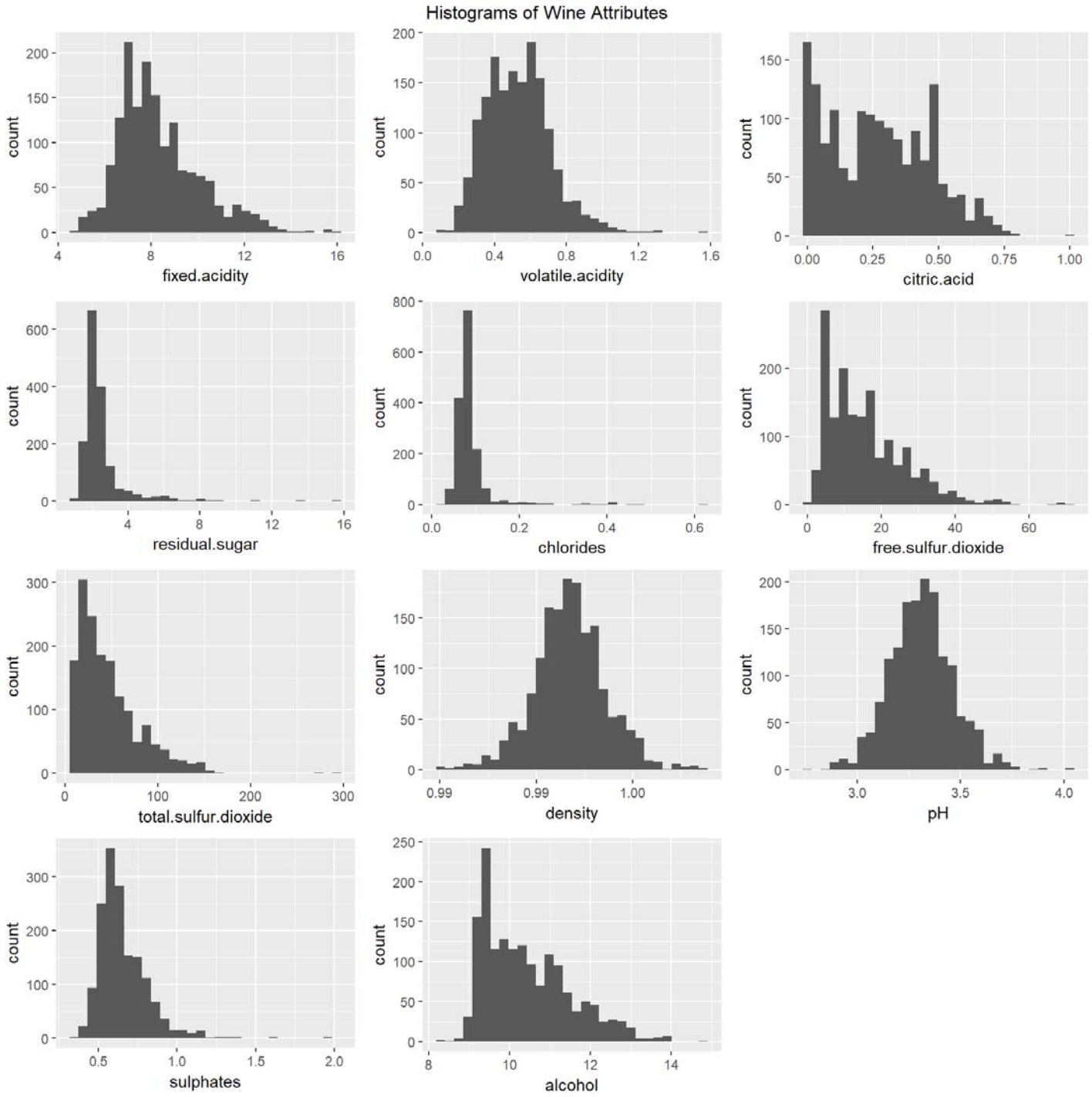



```
winedata %>% group_by(quality) %>% summarise(n = n())
```

```
## # A tibble: 6 x 2
##   quality     n
##     <int> <int>
## 1       3    10
## 2       4    53
## 3       5   681
## 4       6   638
## 5       7   199
## 6       8    18
```

The univariate analysis for the output / dependent variable was based upon the attribute of quality. The boxplot shows the distribution of qualities follows a normal distribution / bell curve, in which the curve is a very steep one. showing that most of the samples are in the middle, with few at the edges, and none at the extreme edges of 0,1,2,9, and 10. The descriptive statistics of the quality attribute also confirm this with greater specificity.

```
h <- Map(function(variable, names){
  return(ggplot(aes(x = variable), data = winedata) + geom_histogram(bins = 30) +
           xlab(names))}, select(winedata, -quality), names(winedata)[1:11])
grid.arrange(grobs = h, top = "Histograms of Wine Attributes")
```



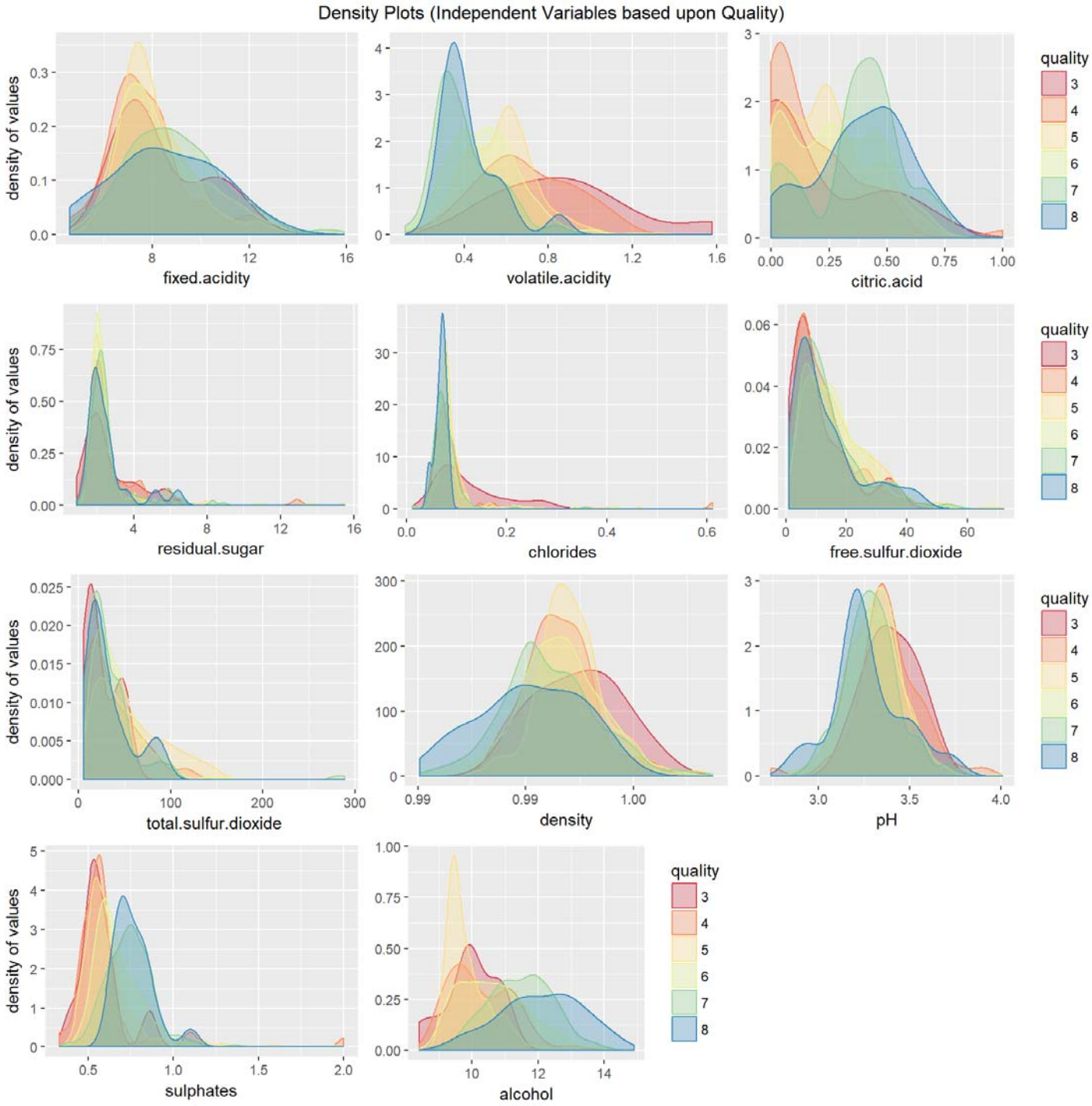

```
winedata %>% gather(attribute, value, 1:11) %>% group_by(attribute) %>%
  summarise(min = min(value),
                        first_quantile = quantile(value, .25),
                        median = median(value),
                        mean = mean(value),
                        third_quantile = quantile(value, .75),
                        max = max(value), skew = skewness(value))
```

```
## # A tibble: 11 x 8
##               attribute   min first_quantile median   mean third_quantile    max  skew
##                   <chr> <dbl>          <dbl>  <dbl>  <dbl>          <dbl>  <dbl> <dbl>
##  1             alcohol 8.400           9.50 10.200 10.423          11.10  14.90 0.860
##  2           chlorides 0.012           0.07  0.079  0.087           0.09   0.61 5.675
##  3         citric.acid 0.000           0.09  0.260  0.271           0.42   1.00 0.318
##  4             density 0.990           1.00  0.997  0.997           1.00   1.00 0.071
##  5       fixed.acidity 4.600           7.10  7.900  8.320           9.20  15.90 0.982
##  6 free.sulfur.dioxide 1.000           7.00 14.000 15.875          21.00  72.00 1.249
##  7                  pH 2.740           3.21  3.310  3.311           3.40   4.01 0.194
##  8      residual.sugar 0.900           1.90  2.200  2.539           2.60  15.50 4.536
##  9           sulphates 0.330           0.55  0.620  0.658           0.73   2.00 2.426
## 10 total.sulfur.dioxide 6.000          22.00 38.000 46.468          62.00 289.00 1.514
## 11    volatile.acidity 0.120           0.39  0.520  0.528           0.64   1.58 0.671
```

The univariate analysis for the input / independent variables was based upon the attributes of fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol. The histograms show the distribution of qualities follow a normal distribution / bell curve for the attributes of density and pH, and a positive skew for all of the other attributes. The descriptive statistics of the input variable attributes also confirm this with greater specificity.

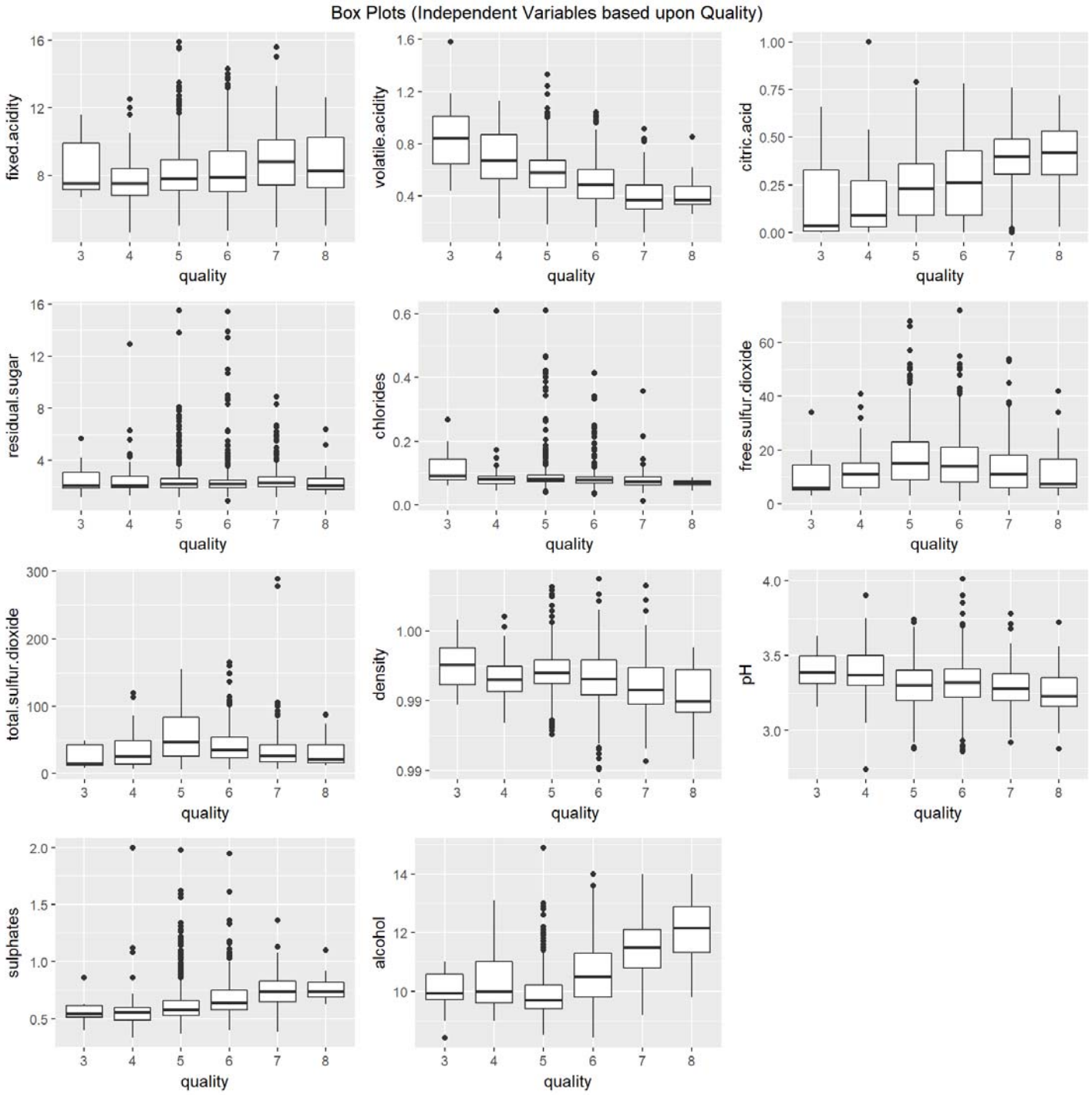
2. Bivariate Analysis: I will analyze each of the eleven attributes by a density plot based upon quality, a boxplot grouped by quality, and then summary statistics grouped by quality.

```
d <- Map(function(variable, names){
  return(ggplot(aes(x = variable), data = winedata) +
           geom_density(aes(group = quality, color = factor(quality),
                            fill = factor(quality)), alpha = 0.3) +
           labs(y  = "density of values", x = names, color = "quality",
                fill = "quality") +
           scale_colour_brewer(palette = "Spectral") +
           scale_fill_brewer(palette = "Spectral"))}, select(winedata, -quality),
  names(winedata)[1:11])
grid.arrange(d[[1]] + theme(legend.position = "none"),
             d[[2]] + theme(legend.position = "none") + ylab(NULL),
             d[[3]] + ylab(NULL), d[[4]] + theme(legend.position = "none"),
             d[[5]] + theme(legend.position = "none") + ylab(NULL),
             d[[6]] + ylab(NULL), d[[7]] + theme(legend.position = "none"),
             d[[8]] + ylab(NULL) + theme(legend.position = "none"),
             d[[9]] + ylab(NULL), d[[10]] + theme(legend.position = "none"),
             d[[11]] + ylab(NULL),
             top = "Density Plots (Independent Variables based upon Quality)")
```



For the first bivariate analysis, I have chosen to take all of the density plots, and place them in a grid. This way we can easily view the different input variables and see how they affect the quality variable. From determining correlations of variables based upon density plots, I woul say that the variables that correlate positively with quality are fixed acidity, citric acid, sulphates, and alcohol. The attributes that negatively correlate with quality are volatile acidity, density, and pH. The attributes that seem to have very little correlation with quality are residual sugar, chlorides, free sulfur dioxide, and total sulfur dioxide. I am taking into consideration that basing off of visualizing does add a strong subjective element, so others might disagree with me. I need to bracket this out of my mind when making a judgement for the next set of plots.

```
b <- Map(function(variable, names){
  return(ggplot(aes(x = factor(quality), y = variable), data = winedata) +
           geom_boxplot() + labs(x = "quality", y = names))},
  select(winedata, -quality), names(winedata)[1:11])
grid.arrange(grobs = b,
             top = "Box Plots (Independent Variables based upon Quality)")
```



For the second bivariate analysis, I have chosen to take all of the box plots, and place them in a grid. This way we can easily view the different input variables and see how they affect the output variable of quality. From determining correlations of variables based upon the box plots, I would say that the variables that correlate positively with quality are citric acid, sulphates, and alcohol. The attributes that negatively correlate with quality are volatile acidity, density, and pH. The attributes that seem to have very little correlation with quality are fixed acidity, residual sugar, chlorides, free sulfur dioxide, and total sulfur dioxide. I

am taking into consideration that basing off of visualizing does add a strong subjective element, so others might disagree with me.

```
winedata %>% gather(attribute, value, 1:11) %>% group_by(attribute, quality) %>%
  summarise(min = min(value),
            first_quantile = quantile(value, .25),
            median = median(value),
            mean = mean(value),
            third_quantile = quantile(value, .75),
            max = max(value))
```

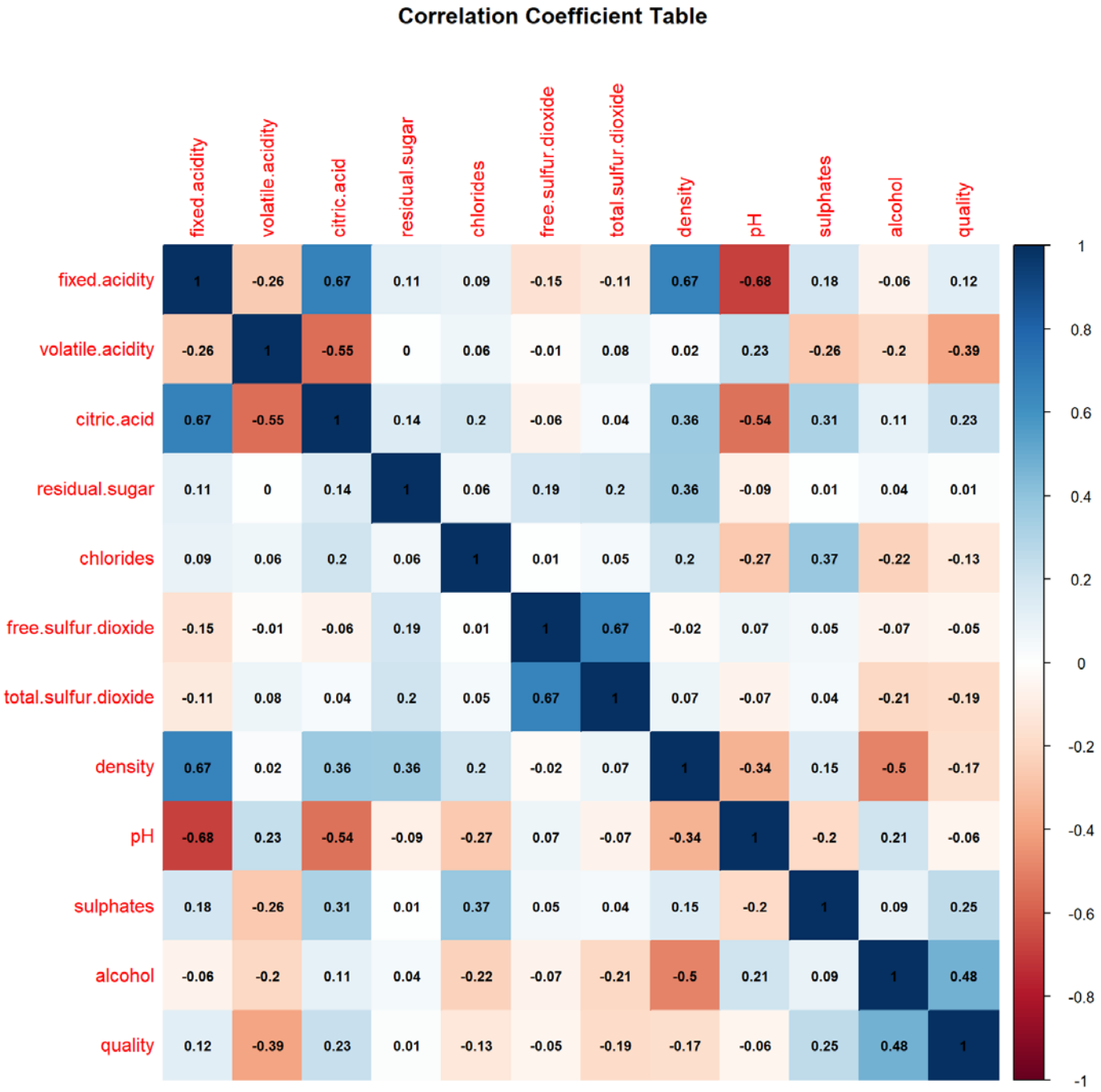
```
## # A tibble: 66 x 8
## # Groups:   attribute [?]
##               attribute quality    min first_quantile median   mean third_quantile     max
##                   <chr>   <int>  <dbl>          <dbl>  <dbl>  <dbl>          <dbl>   <dbl>
##  1             alcohol       3  8.400          9.725  9.925  9.955         10.575  11.000
##  2             alcohol       4  9.000          9.600 10.000 10.265         11.000  13.100
##  3             alcohol       5  8.500          9.400  9.700  9.900         10.200  14.900
##  4             alcohol       6  8.400          9.800 10.500 10.630         11.300  14.000
##  5             alcohol       7  9.200         10.800 11.500 11.466         12.100  14.000
##  6             alcohol       8  9.800         11.325 12.150 12.094         12.875  14.000
##  7           chlorides       3  0.061          0.079  0.090  0.122          0.143   0.267
##  8           chlorides       4  0.045          0.067  0.080  0.091          0.089   0.610
##  9           chlorides       5  0.039          0.074  0.081  0.093          0.094   0.611
## 10           chlorides       6  0.034          0.068  0.078  0.085          0.088   0.415
## 11           chlorides       7  0.012          0.062  0.073  0.077          0.087   0.358
## 12           chlorides       8  0.044          0.062  0.071  0.068          0.075   0.086
## 13         citric.acid       3  0.000          0.005  0.035  0.171          0.328   0.660
## 14         citric.acid       4  0.000          0.030  0.090  0.174          0.270   1.000
## 15         citric.acid       5  0.000          0.090  0.230  0.244          0.360   0.790
## 16         citric.acid       6  0.000          0.090  0.260  0.274          0.430   0.780
## 17         citric.acid       7  0.000          0.305  0.400  0.375          0.490   0.760
## 18         citric.acid       8  0.030          0.302  0.420  0.391          0.530   0.720
## 19             density       3  0.995          0.996  0.998  0.997          0.999   1.001
## 20             density       4  0.993          0.996  0.997  0.997          0.997   1.001
## 21             density       5  0.993          0.996  0.997  0.997          0.998   1.003
## 22             density       6  0.990          0.995  0.997  0.997          0.998   1.004
## 23             density       7  0.991          0.995  0.996  0.996          0.997   1.003
## 24             density       8  0.991          0.994  0.995  0.995          0.997   0.999
## 25       fixed.acidity       3  6.700          7.150  7.500  8.360          9.875  11.600
## 26       fixed.acidity       4  4.600          6.800  7.500  7.779          8.400  12.500
## 27       fixed.acidity       5  5.000          7.100  7.800  8.167          8.900  15.900
## 28       fixed.acidity       6  4.700          7.000  7.900  8.347          9.400  14.300
## 29       fixed.acidity       7  4.900          7.400  8.800  8.872         10.100  15.600
## 30       fixed.acidity       8  5.000          7.250  8.250  8.567         10.225  12.600
## 31  free.sulfur.dioxide       3  3.000          5.000  6.000 11.000         14.500  34.000
## 32  free.sulfur.dioxide       4  3.000          6.000 11.000 12.264         15.000  41.000
## 33  free.sulfur.dioxide       5  3.000          9.000 15.000 16.984         23.000  68.000
## 34  free.sulfur.dioxide       6  1.000          8.000 14.000 15.712         21.000  72.000
## 35  free.sulfur.dioxide       7  3.000          6.000 11.000 14.045         18.000  54.000
## 36  free.sulfur.dioxide       8  3.000          6.000  7.500 13.278         16.500  42.000
## 37                  pH       3  3.160          3.312  3.390  3.398          3.495   3.630
## 38                  pH       4  2.740          3.300  3.370  3.382          3.500   3.900
## 39                  pH       5  2.880          3.200  3.300  3.305          3.400   3.740
## 40                  pH       6  2.860          3.220  3.320  3.318          3.410   4.010
## 41                  pH       7  2.920          3.200  3.280  3.291          3.380   3.780
## 42                  pH       8  2.880          3.162  3.230  3.267          3.350   3.720
## 43      residual.sugar       3  1.200          1.875  2.100  2.635          3.100   5.700
## 44      residual.sugar       4  1.300          1.900  2.100  2.694          2.800  12.900
## 45      residual.sugar       5  1.200          1.900  2.200  2.529          2.600  15.500
## 46      residual.sugar       6  0.900          1.900  2.200  2.477          2.500  15.400
## 47      residual.sugar       7  1.200          2.000  2.300  2.721          2.750   8.900
## 48      residual.sugar       8  1.400          1.800  2.100  2.578          2.600   6.400
## 49           sulphates       3  0.400          0.512  0.545  0.570          0.615   0.860
## 50           sulphates       4  0.330          0.490  0.560  0.596          0.600   2.000
## 51           sulphates       5  0.370          0.530  0.580  0.621          0.660   1.980
## 52           sulphates       6  0.400          0.580  0.640  0.675          0.750   1.950
```

```
## 53            sulphates       7  0.390          0.650  0.740  0.741          0.830   1.360
## 54            sulphates       8  0.630          0.690  0.740  0.768          0.820   1.100
## 55 total.sulfur.dioxide       3  9.000         12.500 15.000 24.900         42.500  49.000
## 56 total.sulfur.dioxide       4  7.000         14.000 26.000 36.245         49.000 119.000
## 57 total.sulfur.dioxide       5  6.000         26.000 47.000 56.514         84.000 155.000
## 58 total.sulfur.dioxide       6  6.000         23.000 35.000 40.870         54.000 165.000
## 59 total.sulfur.dioxide       7  7.000         17.500 27.000 35.020         43.000 289.000
## 60 total.sulfur.dioxide       8 12.000         16.000 21.500 33.444         43.000  88.000
## 61     volatile.acidity       3  0.440          0.647  0.845  0.884          1.010   1.580
## 62     volatile.acidity       4  0.230          0.530  0.670  0.694          0.870   1.130
## 63     volatile.acidity       5  0.180          0.460  0.580  0.577          0.670   1.330
## 64     volatile.acidity       6  0.160          0.380  0.490  0.497          0.600   1.040
## 65     volatile.acidity       7  0.120          0.300  0.370  0.404          0.485   0.915
## 66     volatile.acidity       8  0.260          0.335  0.370  0.423          0.472   0.850
```

For the third bivariate analysis, I have decided to create a table of summary statistics for all of the independent variables grouped by quality. This will help to add specificity to the information obtained from the plots.

3. Corrplot of Variables: I will create a plot that has all of the correlation coefficients of all of the variables related to all of the other variables.

```
corrplot(cor(winedata), method = "color", addCoef.col="black", number.cex = .75,
         title = "Correlation Coefficient Table", mar=c(0,0,1,0))
```



	fixed.acidity	volatile.acidity	citric.acid	residual.sugar	chlorides	free.sulfur.dioxide	total.sulfur.dioxide	density	pH	sulphates	alcohol	quality
fixed.acidity	1	-0.26	0.67	0.11	0.09	-0.15	-0.11	0.67	-0.68	0.18	-0.06	0.12
volatile.acidity	-0.26	1	-0.55	0	0.06	-0.01	0.08	0.02	0.23	-0.26	-0.2	-0.39
citric.acid	0.67	-0.55	1	0.14	0.2	-0.06	0.04	0.36	-0.54	0.31	0.11	0.23
residual.sugar	0.11	0	0.14	1	0.06	0.19	0.2	0.36	-0.09	0.01	0.04	0.01
chlorides	0.09	0.06	0.2	0.06	1	0.01	0.05	0.2	-0.27	0.37	-0.22	-0.13
free.sulfur.dioxide	-0.15	-0.01	-0.06	0.19	0.01	1	0.67	-0.02	0.07	0.05	-0.07	-0.05
total.sulfur.dioxide	-0.11	0.08	0.04	0.2	0.05	0.67	1	0.07	-0.07	0.04	-0.21	-0.19
density	0.67	0.02	0.36	0.36	0.2	-0.02	0.07	1	-0.34	0.15	-0.5	-0.17
pH	-0.68	0.23	-0.54	-0.09	-0.27	0.07	-0.07	-0.34	1	-0.2	0.21	-0.06
sulphates	0.18	-0.26	0.31	0.01	0.37	0.05	0.04	0.15	-0.2	1	0.09	0.25
alcohol	-0.06	-0.2	0.11	0.04	-0.22	-0.07	-0.21	-0.5	0.21	0.09	1	0.48
quality	0.12	-0.39	0.23	0.01	-0.13	-0.05	-0.19	-0.17	-0.06	0.25	0.48	1

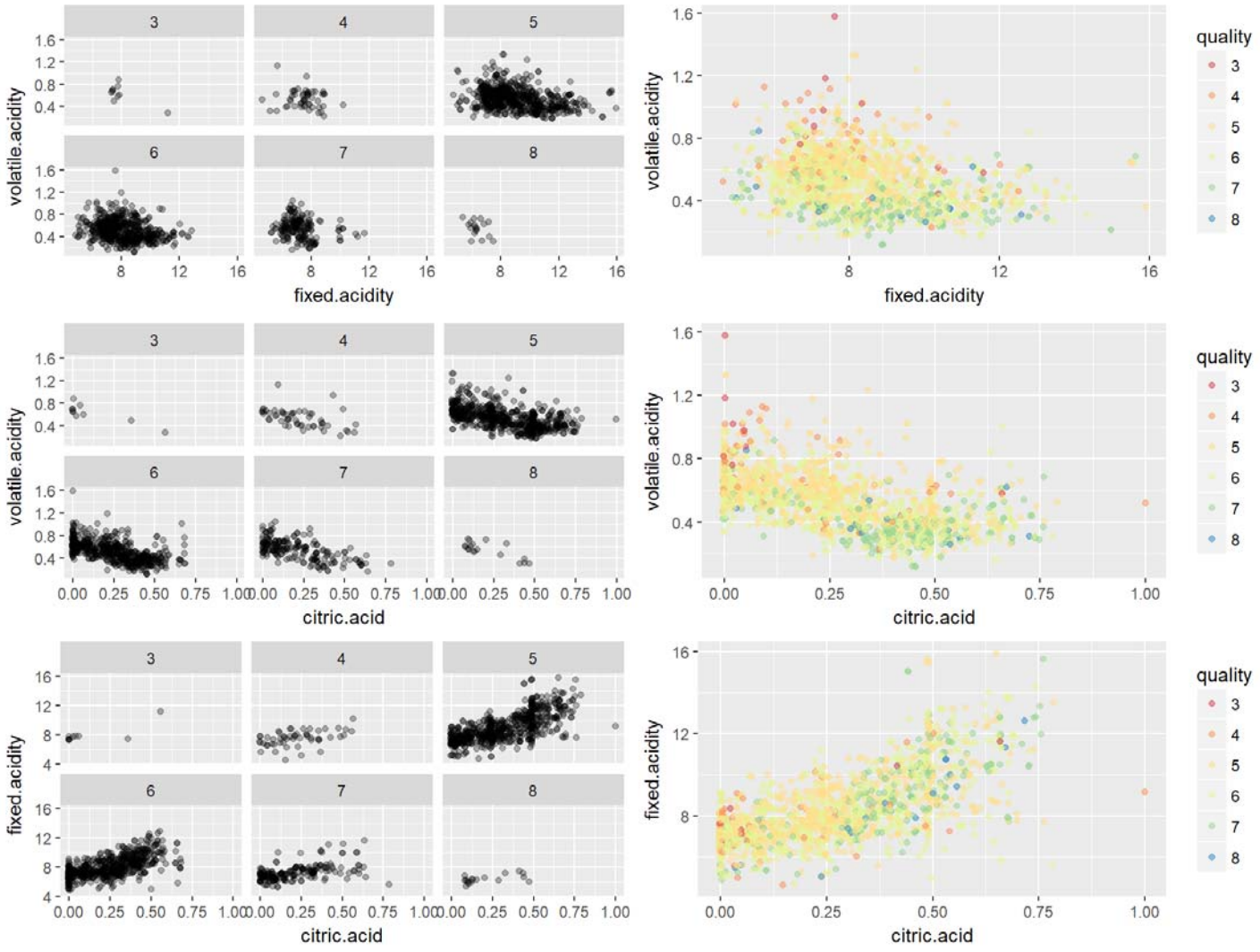
In this section, I used the correlation plot to check my interpretation of the previous two set of graphs, I find that the three most strongly postively correlated variables with quality were those that I noticed a correlation on both set of graphs. These are alcohol at a .48, sulphates at a .25, and citric acid at a .23. The next variable as far as positive correlation would be fixed acidity at .12, which I saw a correlation on the density plots, but not on the boxplots with quality. The next in order from positive to negative would be residual sugar at .01, which I saw no correlation on either set of graphs with quality, then free sulfur dioxides at -.05, which I saw no correlation on either set of graphs with quality, and pH at a -.06, which I saw a negative correlation on both graphs with quality. So far this seems to be the only outlier on my perception of the graphs as far as correlation is concerned. The next attribute of chlorides correlates with quality at -.13, I saw no correlation in either graph. The next attribute of density correlates with quality at -.17, I saw a negative correlation on both graphs, but the next attribute of total sulfur dioxides correlates with quality at -.19, I saw no correlation on either graph. The last variable, which is volatile acidity correlates with quality at -.39, which had a noticeable negative correlation on both graphs. This shows that my visualizing the ranking of correlations were mostly correct, but with a few outliers. This is why I checked my results using a correlation plot.

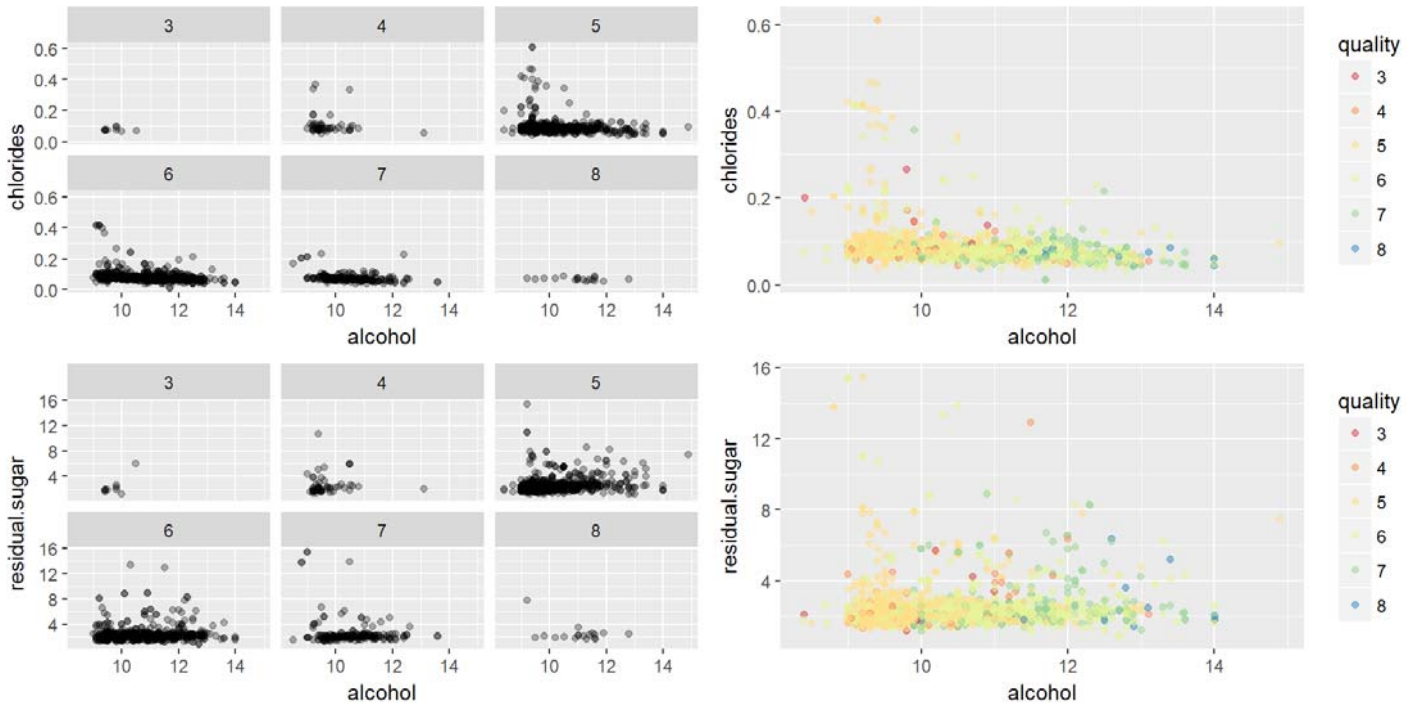
4. Multivariate Analysis: I will select certain attributes to see how they correlate with each other, and how these two correlate with quality. The attributes that I will select are fixed acidity, volatile acidity, citric acid, alcohol, chlorides, and residual sugar.

```
s <- Map(function(x_variable, y_variable, x_names, y_names){
  return(ggplot(aes(x = x_variable, y = y_variable), data = winedata) +
           geom_jitter(aes(color = factor(quality)), alpha = .5) +
           labs(color = "quality") +
           scale_colour_brewer(palette = "Spectral") +
           scale_fill_brewer(palette = "Spectral") +
           labs(x = x_names, y = y_names))},
  winedata[,c(1,3,3,11,11)], winedata[, c(2,2,1,5,4)],
  names(winedata)[c(1,3,3,11,11)], names(winedata)[c(2,2,1,5,4)])

w <-  Map(function(x_variable, y_variable, x_names, y_names){
  return(ggplot(aes(x = x_variable, y = y_variable), data = winedata) +
           geom_jitter(alpha = .3) + facet_wrap(~quality) +
           labs(x = x_names, y = y_names))},
  winedata[,c(1,3,3,11,11)], winedata[, c(2,2,1,5,4)],
  names(winedata)[c(1,3,3,11,11)], names(winedata)[c(2,2,1,5,4)])

Map(function(w, s){
  return(grid.arrange(w, s, ncol = 2))}, w, s)
```





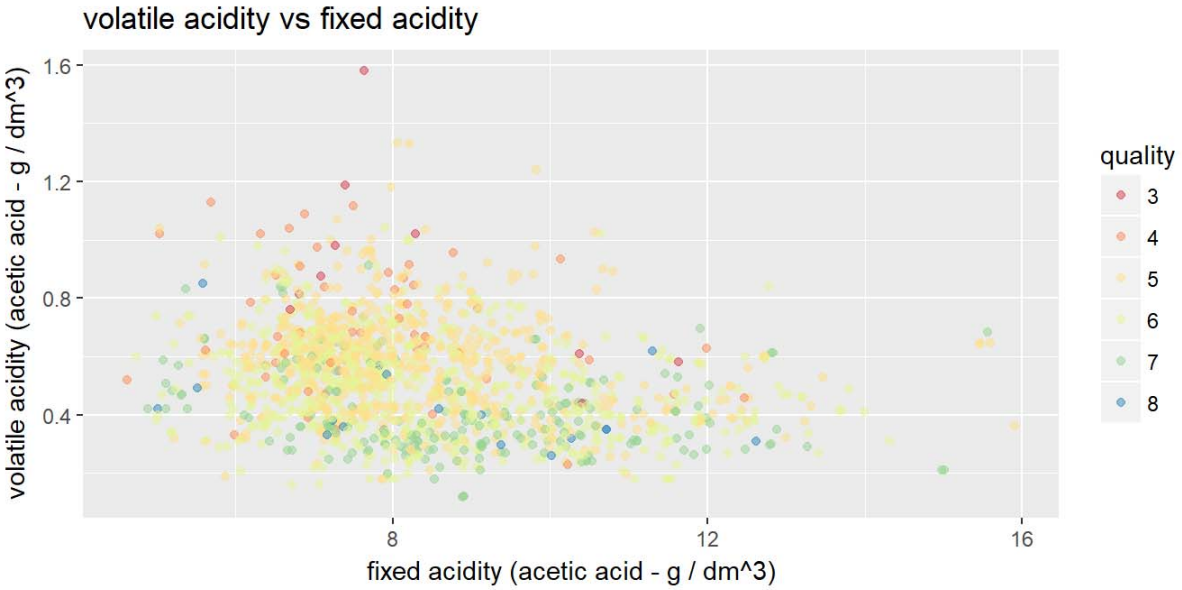
These graphs are rather interesting in that they show a relationship between the different types of acids, and how they impact quality. This is what I will develop further in the section titled, Data Analysis - Final Plots and Summaries. Also, I found it interesting that there is a greater negative correlation between alcohol and chlorides as opposed to alcohol and sugar. I thought the fermentation process would indicate that latter two are more negatively correlated than the former. Why this is not the case, and why there is a greater negative correlation between alcohol and chlorides compared to alcohol and sugar would be a topic for further research, and an initial guess might be due to the quality of the grapes that one uses for wine.

Data Analysis - Final Plots and Summaries

For my final plots and summaries, I have decided to select three graphs that deal with the relationships between the different types of acids and how they impact quality. The information presented in the stream of consciousness part of the paper shows that quality seems to be positively impacted by the tart taste of fixed acidity (tartatic acid), and the citrusy taste of citric acid, while being negatively impacted by volatile acidity (acetic acid). Acetic acid is the component that gives vinegar the vinegar type taste, so I could see how that negatively impacts the taste of wine.

1. Scatter plot of volatile acidity (acetic acid) vs fixed acidity (tartaric acid)

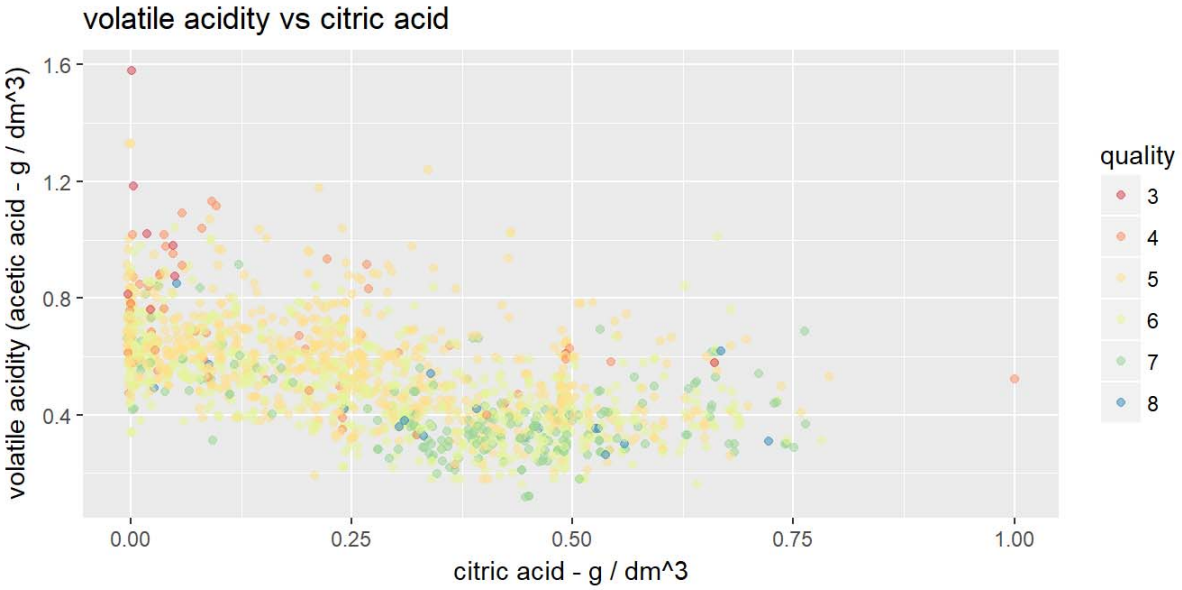
```
s[[1]] + labs(y  = "volatile acidity (acetic acid - g / dm^3)",
          x = "fixed acidity (acetic acid - g / dm^3)",
          title = "volatile acidity vs fixed acidity")
```



For my first final plot, it shows that there is a negative correlation between fixed acidity and acetic acid, with higher concentrations of acetic acid leading to lower quality wines, and higher concentrations of tartaric acid leading to higher quality wines. The issue now would be to isolate the chemical process responsible for the production and breakdown of specific acids.

2. Scatter plot of volatile acidity (acetic acid) vs citric acid

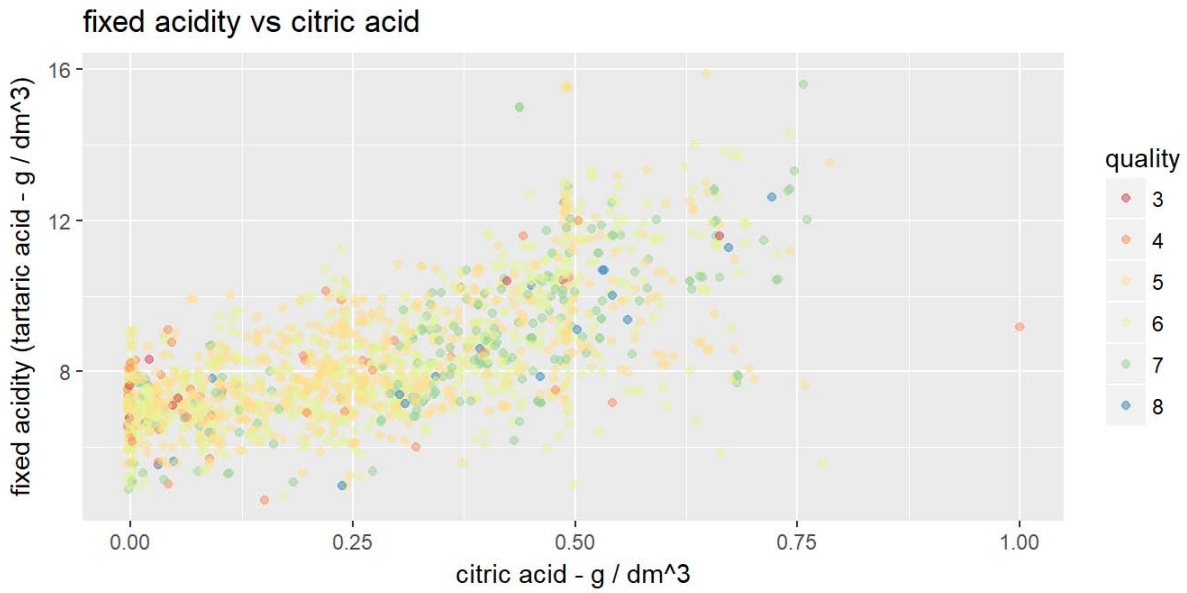
```
s[[2]] + labs(y  = "volatile acidity (acetic acid - g / dm^3)",
         x = "citric acid - g / dm^3",
         title = "volatile acidity vs citric acid")
```



For my second final plot, it shows that there is a negative correlation between fixed acidity and citric acid, with higher concentrations of acetic acid leading to lower quality wines, and higher concentrations of citric acid leading to higher quality wines. The issue now would be to isolate the chemical process responsible for the production and breakdown of specific acids.

3. Scatter plot of fixed acidity (tartaric acid) vs citric acid

```
s[[3]] + labs(y  = "fixed acidity (tartaric acid - g / dm^3)",
          x = "citric acid - g / dm^3",
          title = "fixed acidity vs citric acid")
```



For my third final plot, it shows that there is a positive correlation between tartaric acid and citric acid, with higher concentrations of tartaric acid leading to higher quality wines, and higher concentrations of citric acid leading to higher quality wines. The issue now would be to isolate the chemical process responsible for the production and breakdown of specific acids.

There is definitely an interesting relationship between the three types of acids,and how they impact quality. It makes sense that fixed acidity and citric acid would positively affect quality, since the tart taste of fixed acidity, which is tartaric acid would probably give a slight tart taste to wine, and citric acid would give that slight citrusy sour taste to the wine, which could be seen as adding a desired flavor. It also makes sense that volatile acidity, which is acetic acid would negatively impact the quality of wine, since it has that vinegar quality taste which is not something that someone would want in their wine.

4. Linear Model of Quality based upon other variables: Here I will create a linear model of the eleven attributes and how they relate to quality, which will be expressed as a function of how quality relates to all of the variables.

```
model <- lm((quality ~ .), data = winedata)
formula <- formula(model)
as.formula(
  paste0("quality ~ ", round(coefficients(model)[1],2), "",
    paste(sprintf(" %+.2f*%s ",
                  coefficients(model)[-1],
                  names(coefficients(model)[-1])),
          collapse="")))
```

```
## quality ~ 21.97 + 0.02 * fixed.acidity - 1.08 * volatile.acidity -
##     0.18 * citric.acid + 0.02 * residual.sugar - 1.87 * chlorides +
##     0 * free.sulfur.dioxide - 0 * total.sulfur.dioxide - 17.88 *
##     density - 0.41 * pH + 0.92 * sulphates + 0.28 * alcohol
```

I decided to finish by creating a linear model of the function by which the dependent variable, which is quality depends on specific values of all the independent variables. There is something that strikes me as odd, and that is why a variable like total

sulfur dioxide, which correlates at a -.19 on the correlation graph has a 0 coefficient for the linear model, whereas residual sugar which correlates with a .01 on the correlation plot has a .02 coefficient on the regression model. This is interesting, and something for me to think about when it comes to more detailed analyses.

Reflection

As I reflect upon the process of developing my analyses, I first would like to explain some of the learning process. First, when I visualized the variables in histograms, I found the uneven distribution a problem for getting insights from the data. This is where I discovered the density plot, and how it overcomes this problem. The decision to make boxplots for the variables was an easy one. I remember the ggally library from the coursework, but it always crashed, so I looked for something like that, but I found something better, and that was corrplot. I wanted to add the linear model, because this is where further work should be done. The visualizations should be used more for general insights, and you can get descriptive statistics by using the dplyr library. It is in the linear model that you can get predictive power. I realize that to do that I will need to split data into a test set, and training set, and then train a linear model based upon a training set, and test it on the test set. I see that this is done in another course in the data analysis curriculum, so I will hold off on that type of analysis for that course. As far as the results are concerned, it was interesting to see the attributes that raise the quality of wine, as opposed to those that have no effect, and those that have a negative effect. It seems that alcohol has the strongest positive affect on quality, this could be understood as more of a smooth taste that comes from alcohol as opposed to grape juice. As far as the tastes are concerned, quality is increased from the smooth taste of alcohol, the tart taste of tartaric acid, the citrusy sour taste of citric acid. The tastes that decrease quality would be the vinegar type taste of acetic acid, and the salt taste of chlorides. The taste that seem to have little impact on quality is that of sugar. This is good in that it seems there is no bias to either sweet or dry wines. Given what I specifically focused on, it would be interesting to go into the chemical processes dealing with the three types of acids. The next analysis would be the relation of sulphates to fixed acidity, volatile acidity, and citric acid. Sulphates are used as a preserver of some sort to stop certain chemical processes from taking place. I would analyze the relation between sulphates and these three acids. After that I would analyze sulphates in relation to free and total sulfur dioxide.