Rohan George

#Ruid 191004212

Wine Quality Analysis

Introduction:

The data is of the Red & White wine variants of the Portuguese "Vinho Verde" wine taken from UCI repository and it consist of more than 6000 samples of both Red & White Wines.
The Data consist of various Parameters of Wine.

https://archive.ics.uci.edu/ml/datasets/Wine+Quality

About the data:

The Data consist of various Parameters of Wine of both Red and White wine along with its Quality

```
dim(winedata)
[1] 6497   14
```

The data has 6497 samples with 14 columns

The various Parameters in Data are:

- fixed acidity
- volatile acidity
- citric acid
- residual sugar
- Chlorides
- free sulfur dioxide
- total sulfur dioxide
- density
- pH
- sulphates
- alcohol
- Quality
- Type
- high-quality

```
names(winedata)
 [1] "fixed.acidity"        "volatile.acidity"     "citric.acid"          "residual.sugar
"       "chlorides"
 [6] "free.sulfur.dioxide"  "total.sulfur.dioxide" "density"              "pH"
"sulphates"
[11] "alcohol"              "quality"              "type"                 "highquality"
```

Purpose of Analysis:

The Purpose of this analysis is to study different parameters of wine and also analyze the impact of each parameter on the quality of red and white wine using data analysis techniques.
The analysis will more focus on the ph value, Alcohol content and both SO2 (free & total) because of following reasons:

1. pH value is the most important parameter to determine the Quality of the wine.
2. Alcohol content is the parameter that is considered while buying a wine as many prefer more alcohol in a wine.
3. SO2 tends to cause some allergic reaction therefore even this is important parameter in determining the Wine Quality.

Loading the data:

```
library(ggplot2)

library (gridExtra)

## Warning: package 'gridExtra' was built under R version 3.1.1

## Loading required package: grid

library(GGally)

## Warning: package 'GGally' was built under R version 3.1.1

winedata<-read.csv('winequality-redWhite.csv')
```

Summary of Data:

```
> summary(winedata)

 fixed.acidity    volatile.acidity  citric.acid     residual.sugar     chlorides
free.sulfur.dioxide
 Min.   : 3.800   Min.   :0.0800   Min.   :0.0000   Min.   : 0.600   Min.   :0.00900
Min.   :  1.00
 1st Qu.: 6.400   1st Qu.:0.2300   1st Qu.:0.2500   1st Qu.: 1.800   1st Qu.:0.03800
1st Qu.: 17.00
 Median : 7.000   Median :0.2900   Median :0.3100   Median : 3.000   Median :0.04700
Median : 29.00
 Mean   : 7.215   Mean   :0.3397   Mean   :0.3186   Mean   : 5.443   Mean   :0.05603
Mean   : 30.53
 3rd Qu.: 7.700   3rd Qu.:0.4000   3rd Qu.:0.3900   3rd Qu.: 8.100   3rd Qu.:0.06500
3rd Qu.: 41.00
 Max.   :15.900   Max.   :1.5800   Max.   :1.6600   Max.   :65.800   Max.   :0.61100
Max.   :289.00

 total.sulfur.dioxide    density             pH          sulphates         alcohol
quality       type
 Min.   :  6.0        Min.   :0.9871   Min.   :2.720   Min.   :0.2200   Min.   : 8.00
Min.   :3.000   R:1599
 1st Qu.: 77.0        1st Qu.:0.9923   1st Qu.:3.110   1st Qu.:0.4300   1st Qu.: 9.50
1st Qu.:5.000   W:4898
 Median :118.0        Median :0.9949   Median :3.210   Median :0.5100   Median :10.30
Median :6.000
 Mean   :115.7        Mean   :0.9947   Mean   :3.219   Mean   :0.5313   Mean   :10.49
Mean   :5.818
 3rd Qu.:156.0        3rd Qu.:0.9970   3rd Qu.:3.320   3rd Qu.:0.6000   3rd Qu.:11.30
3rd Qu.:6.000
 Max.   :440.0        Max.   :1.0390   Max.   :4.010   Max.   :2.0000   Max.   :14.90
Max.   :9.000

 highquality
 Min.   :0.0000
 1st Qu.:0.0000
 Median :1.0000
 Mean   :0.6331
 3rd Qu.:1.0000
 Max.   :1.0000
```

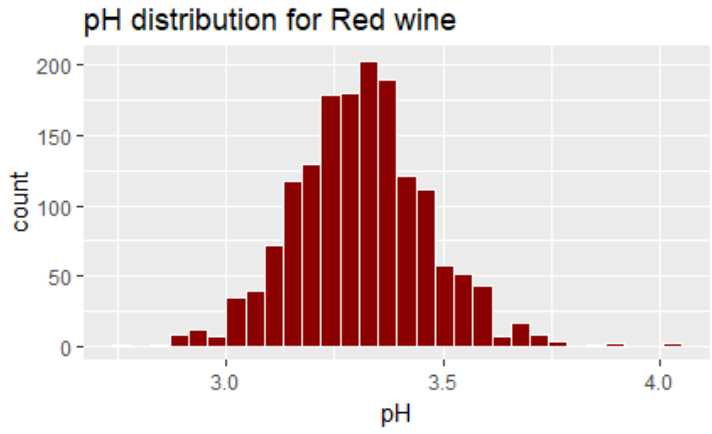
Observations from the Summary

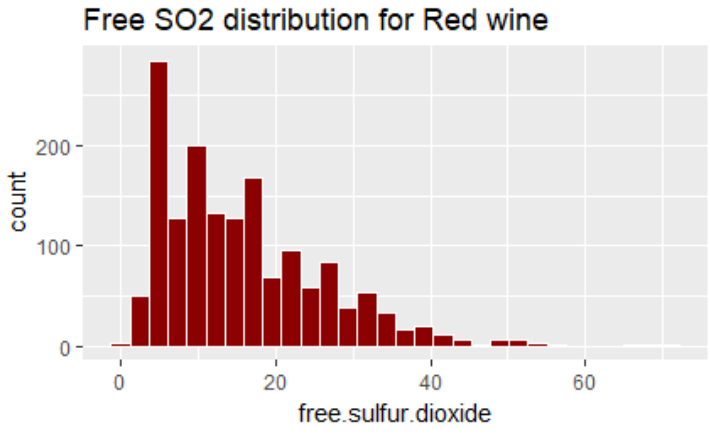
1. Big range for both Free Sulfur. Dioxide and Total Sulfur.dioxide across the samples with median 29.00 and 118.0 respectively
2. The sample consists of 1599 Red Wine and 4898 White wine.
3. Alcohol content ranges from 8.00 to 14.90 for the samples in dataset with median of 10.30
4. Quality of the samples range from 3 to 9 with 6 being the median.
5. Range for fixed acidity is quite high with minimum being 3.8 and maximum being 15.9.
6. pH value varies from 2.720 to 4.010 with a median being 3.210.
7. Volatile acidity ranges from 0.08 to 1.58, 0.29 being the median.
8. The data consist of wide range of residual sugar from 0.600 to 65.800
9. Density is present in smaller range of 0.9871 to 1.0390 compare to other parameters.

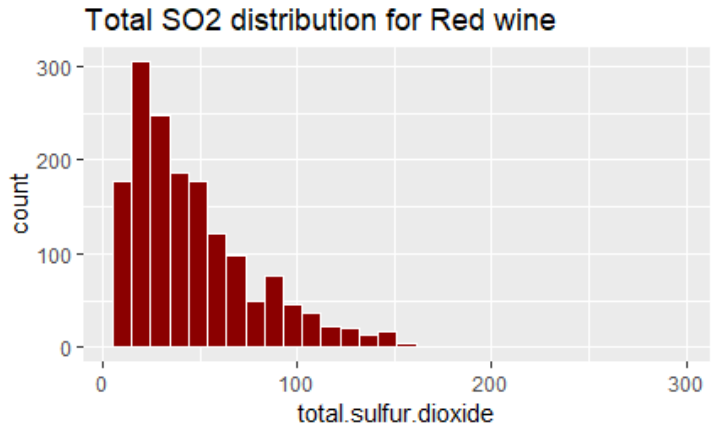
Data Exploration:

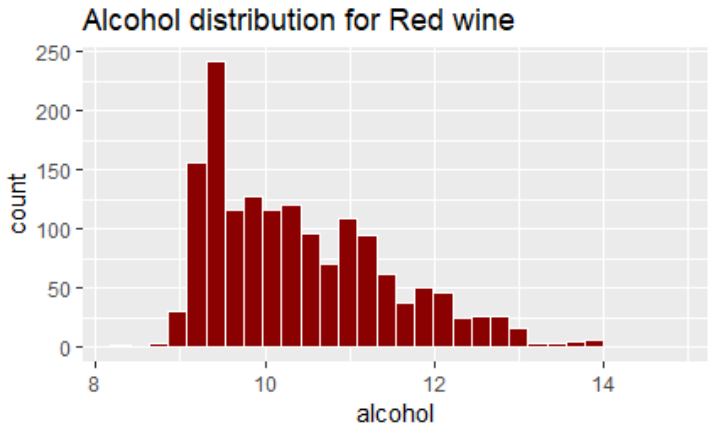
Understanding the RedWine distribution

```
R1<-ggplot(aes(x=pH),data =  subset(wine,type %in% c("R")))+
  geom_histogram(color =I('white'),fill = I('#8B0000'))+
  ggtitle('pH distribution for Red wine')
plot(R1)
R2<-ggplot(aes(x=free.sulfur.dioxide),
         data =  subset(wine,type %in% c("R")))+
  geom_histogram(color =I('white'),fill = I('#8B0000'))+
  ggtitle('Free SO2 distribution for Red wine')
plot(R2)
R3<-ggplot(aes(x=total.sulfur.dioxide),
         data =  subset(wine,type %in% c("R")))+
  geom_histogram(color =I('white'),fill = I('#8B0000'))+
  ggtitle('Total SO2 distribution for Red wine')
plot(R3)
R4<-ggplot(aes(x=alcohol),
         data =  subset(wine,type %in% c("R")))+
  geom_histogram(color =I('white'),fill = I('#8B0000'))+
  ggtitle('Alcohol distribution for Red wine')
plot(R4)
#Link for chart colors in R: http://research.stowers.org/mcm/efg/R/Color/Cha
grid.arrange(R1,R2,R3,R4,ncol=2) #ncol= 2 signifies each column with 2 graphs
```







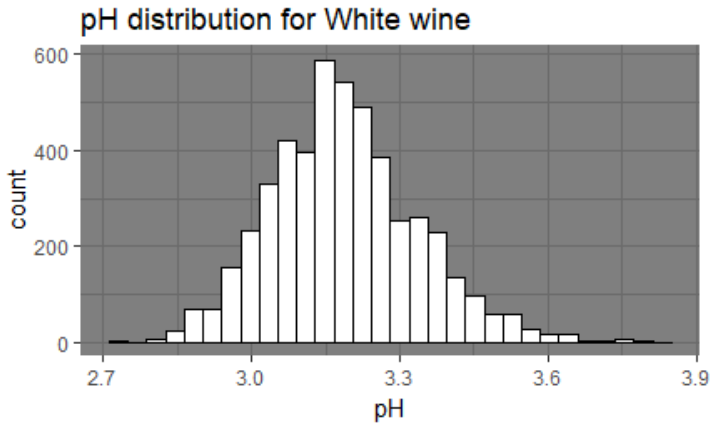


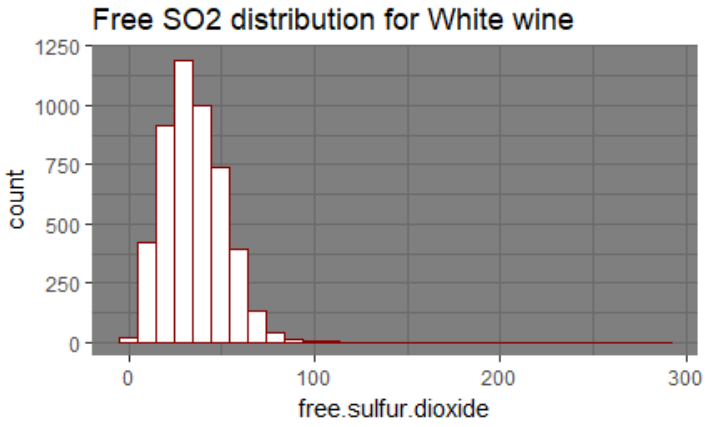
Observations:

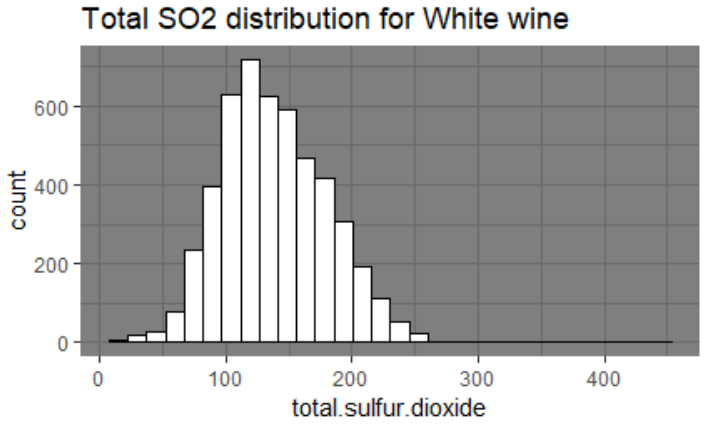
- The pH value shows a normal distribution with most of the samples exhibiting values between 3.0 and 3.5
- The free sulfur dioxide is between the 1-100 count in most of the samples with peaking around 50 mark and highest being 250
- The total sulfur dioxide ranges from 0 and 300 and exhibiting peak around 150.
- The alcohol content varies from 8 to 14 with many peaks around 10 and lower peak between 13 and 14.

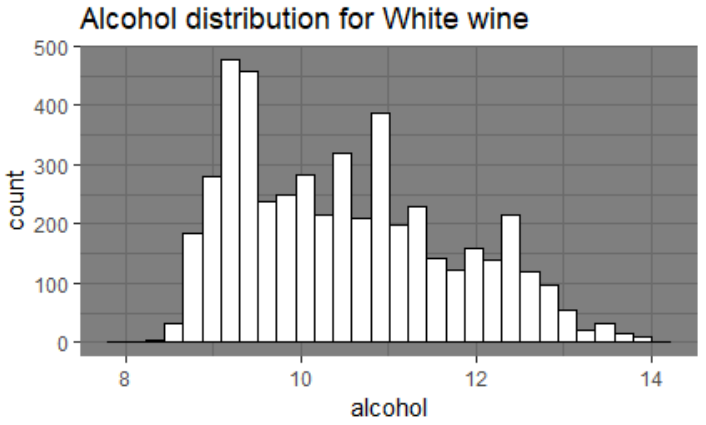
Understanding the White Wine Distributions

```
W1<-ggplot(aes(x=pH),
       data =  subset(wine,type %in% c("W")))+
  geom_histogram(color =I('black'),fill = I('#FFFFFF'))+
  ggtitle('pH distribution for White wine') + theme_dark(base_size = 11)
plot(W1)

#Link to theme types in R : http://www.sthda.com/english/wiki/ggplot2-themes-and-background-colors-the-3-elements

W2<-ggplot(aes(x=free.sulfur.dioxide),
       data =  subset(wine,type %in% c("W")))+
  geom_histogram(color =I('darkred'),fill = I('#FFFFFF'))+
  ggtitle('Free SO2 distribution for White wine') + theme_dark(base_size = 11)
plot(W2)

W3<-ggplot(aes(x=total.sulfur.dioxide),
       data =  subset(wine,type %in% c("W")))+
  geom_histogram(color =I('black'),fill = I('#FFFFFF'))+
  ggtitle('Total SO2 distribution for White wine') + theme_dark(base_size = 11)
plot(W3)

W4<-ggplot(aes(x=alcohol),
       data =  subset(wine,type %in% c("W")))+
  geom_histogram(color =I('black'),fill = I('#FFFFFF'))+
  ggtitle('Alcohol distribution for White wine') + theme_dark(base_size = 11)
plot(W4)

grid.arrange(W1,W2,W3,W4,ncol=2)
```







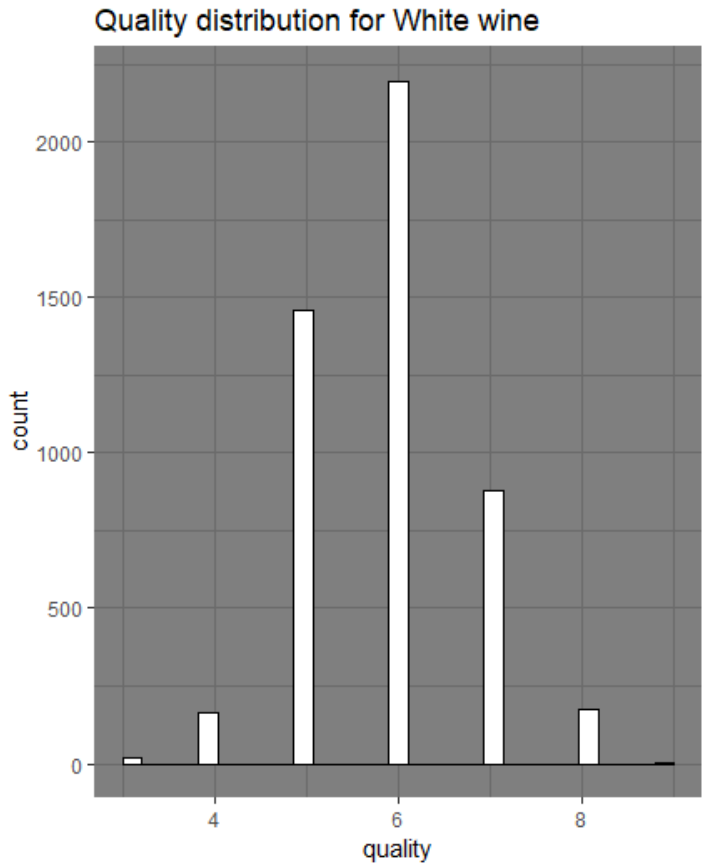


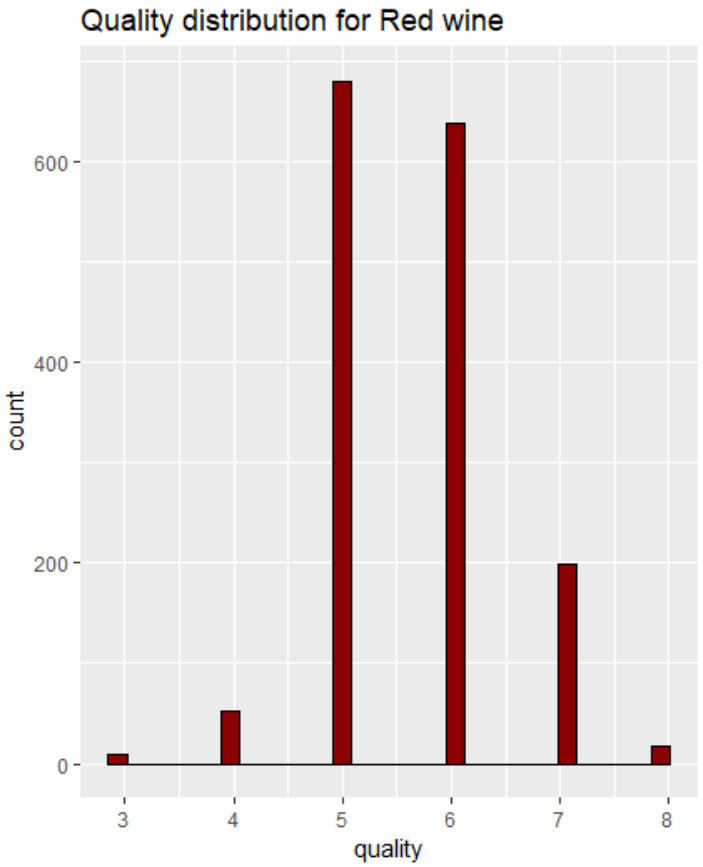
Observations:

- pH value exhibits quite normal distribution but some sample which exhibits pH values beyond 3.5 and 4
- The free sulfur dioxide spreads between 0 to 120 with peak around 50.
- The total sulfur dioxide spreads between 0 and 250 and exhibiting peak around 150.and also lower peak exhibited around the 20 mark.
- The alcohol content varies from 7 to 14 with major peaks around 10 with a large sample

Quality Distributions for Red and White wine:

```
Q1<-ggplot(aes(x=quality),
        data =  subset(wine,type %in% c("W")))+
  geom_histogram(color =I('black'),fill = I('#FFFFFF'))+
  ggtitle('Quality distribution for White wine')+ theme_dark(base_size = 11)
Q2<-ggplot(aes(x=quality),
        data =  subset(wine,type %in% c("R")))+
  geom_histogram(color =I('black'),fill = I('#8B0000'))+
  ggtitle('Quality distribution for Red wine')
grid.arrange(Q1,Q2,ncol=2)
```





Observations on Quality distribution on Red and white wine:

- Both Red & White Wine show Normal Distribution.
- White wine Quality spreads from 1 to 9, and it shows peak at 6 and some samples peak at 5.
- Red Wine spreads from 3 to 8 and it shows peak at 5 and many samples peak at 6.
- Among both Red & White wines only White wine exhibits 9 as the Quality.

Data Analyzation:

Correlation using Scatterplot:

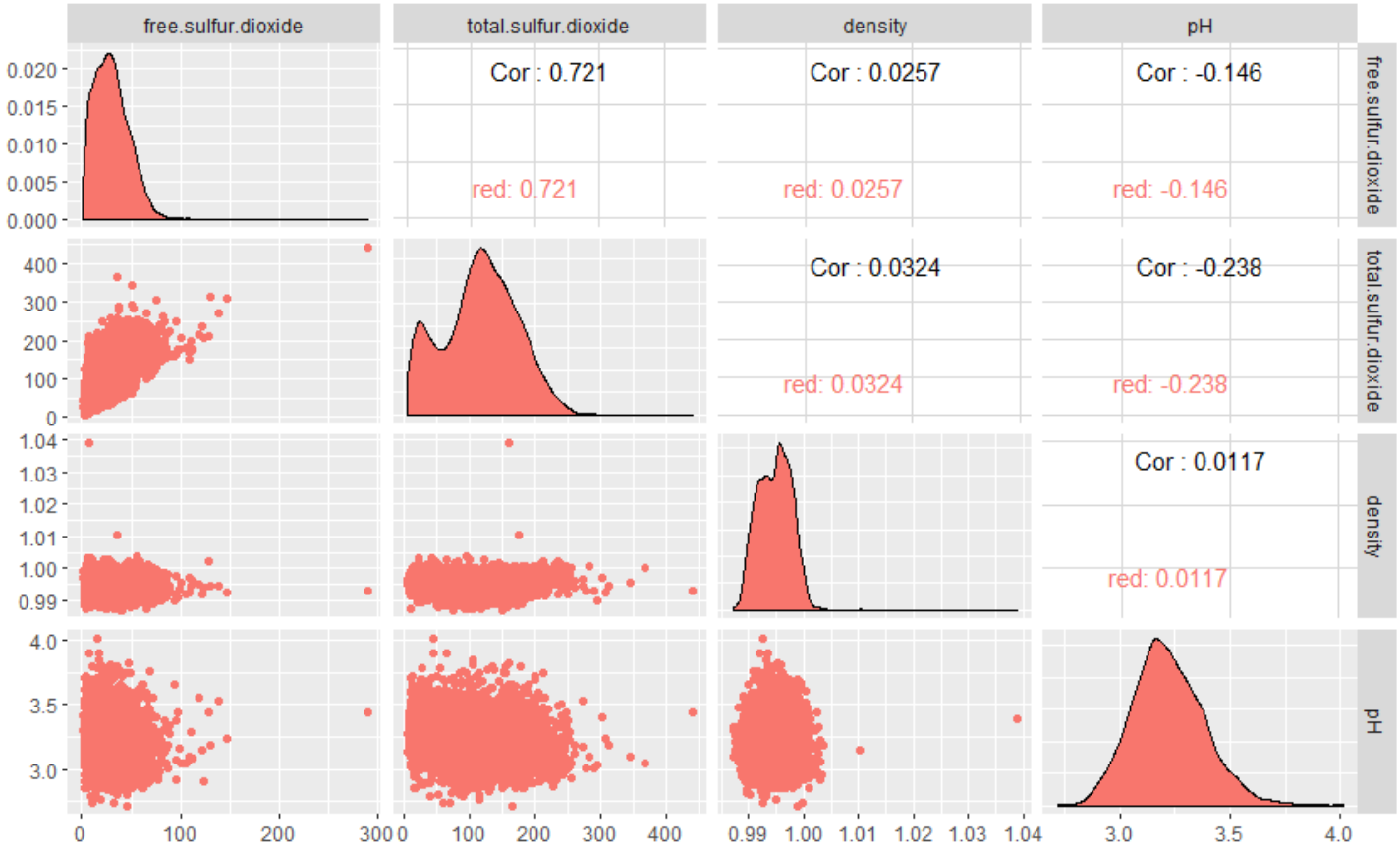
Correlation:

It is a technique of finding the relationship between two quantitative variables.

Scatterplot:

A scatterplot matrix is derived to understand the overall variable behaviour and correlations.

```
# library(GGally)

ggpairs(wine,columns = 6:9, ggplot2::aes(colour='red'))
```

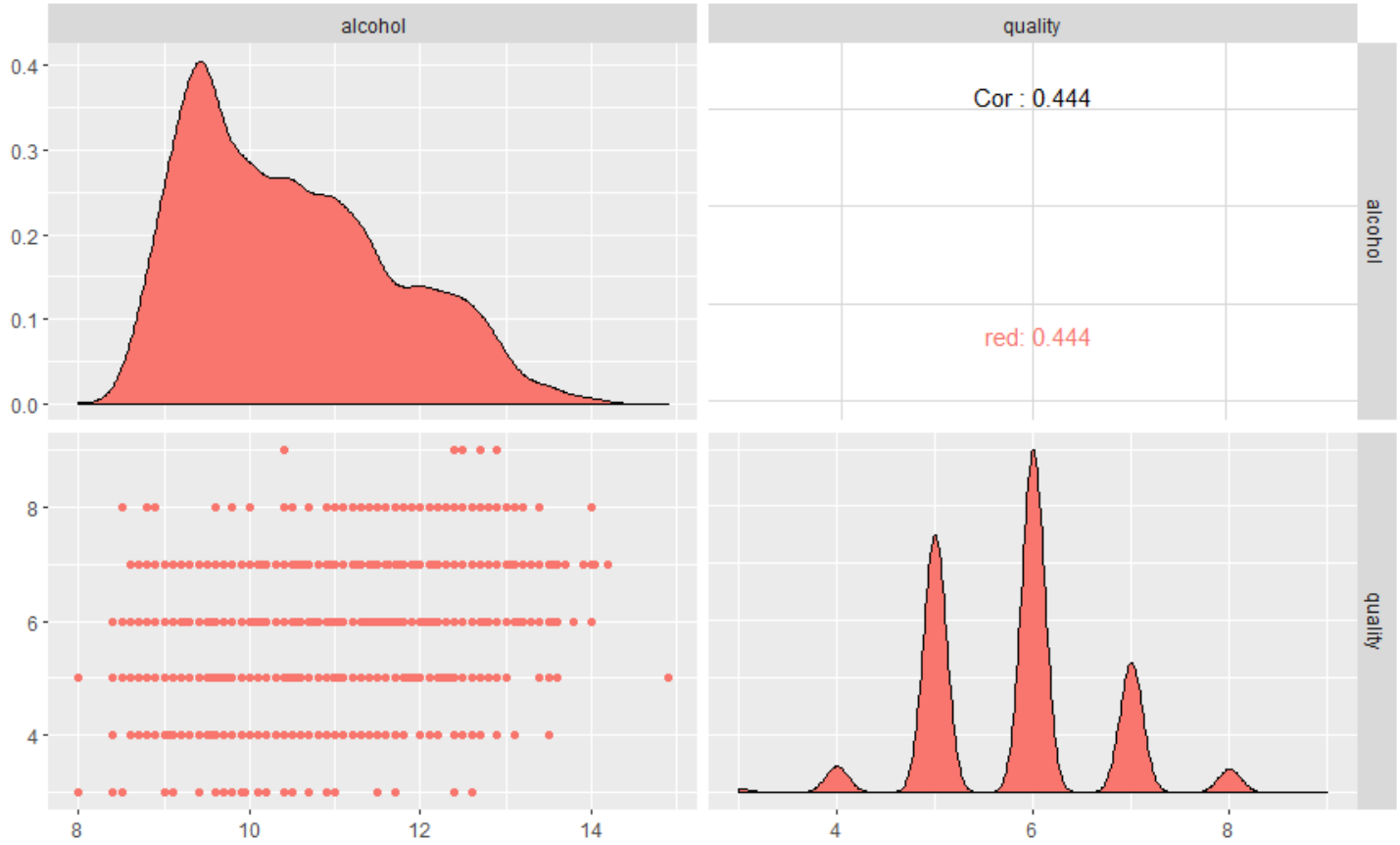


Observations from Scatterplot:

- The above scatterplot shows the correlation between 4 columns i.e. FreeSO2, Total SO2, Density, PH.
- FreeSO2
  FreeSO2 exhibits positive correlation with Total SO2, Negative correlation with pH.
- TotalSO2
  It shows Positive correlation between free so2, Negative correlation with pH.
- Density
  There is no correlation of density with FreeSO2, TotalSO2 and PH.
- PH
  PH shows no correlation with density and Negative correlation with TotalSO2 and Free SO2

Correlation using scatterplot between Alcohol and Quality.

```
# AlcoholQuality

ggpairs(winedata,columns = 11:12, ggplot2::aes(colour='red'))
```

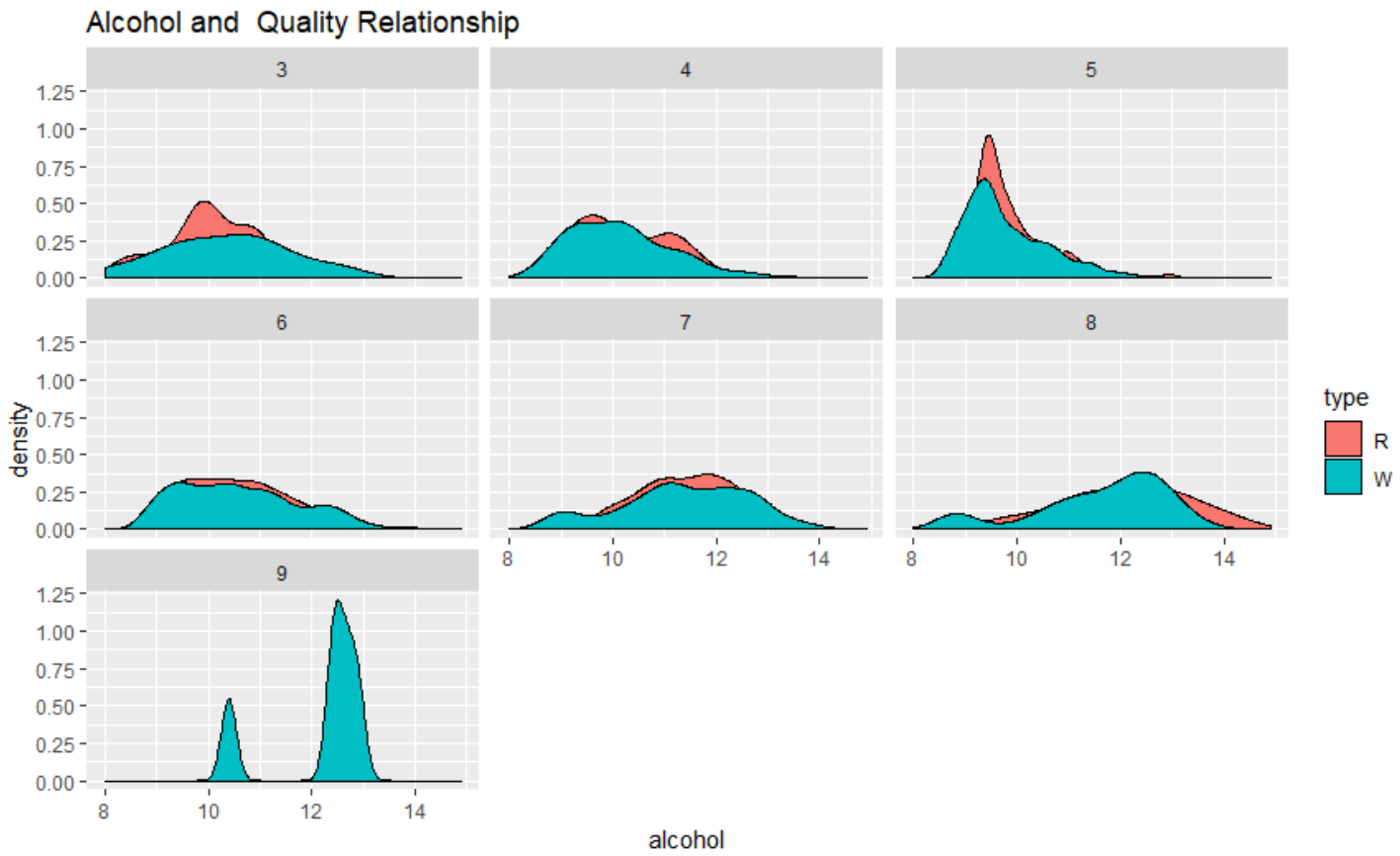


Observations:

Scatterplot shows that Alcohol and Quality as a Positive Correlation.

Alcohol and Quality Relationship:

```
ggplot(aes(x=alcohol),data =wine) +
  geom_density(aes(fill=type))+
  facet_wrap(~quality)+
  ggtitle('Alcohol and  Quality Relationship')
```

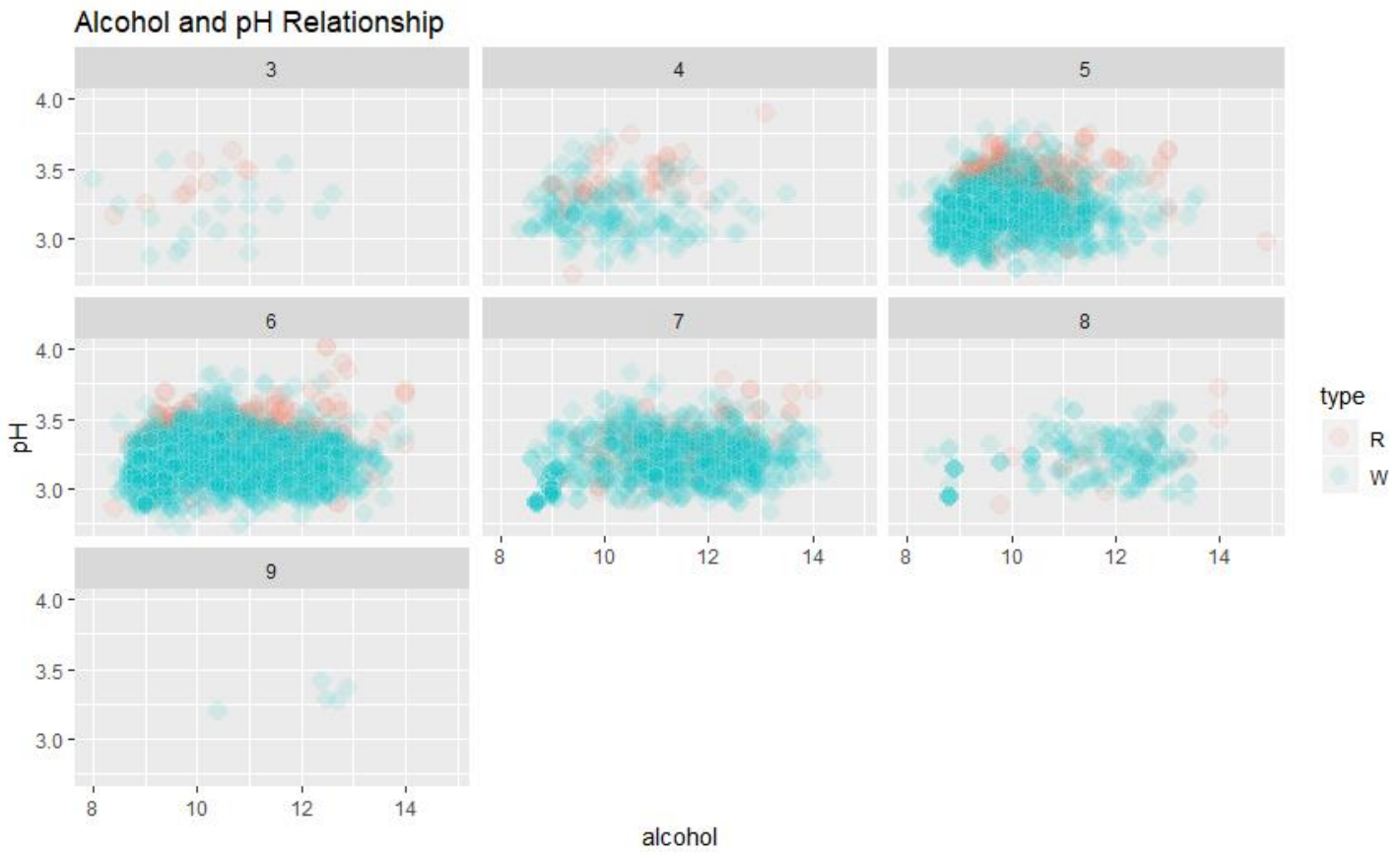


Observations:

The overall alcohol content impact for both Red & White wine remains same even though some samples have higher Alcohol content for Red wine for quality levels 3 and 5 when compared with White Wine.

Alcohol & PH relationship:

```
ggplot(aes(x=alcohol,y=pH),data = winedata) +
  geom_jitter(aes(color = type,bg = type),alpha=1/10,,pch=21,cex=4)+
  facet_wrap(~quality)+
  scale_color_brewer(type = 'div')+
  ggtitle('Alcohol and pH Relationship')
```

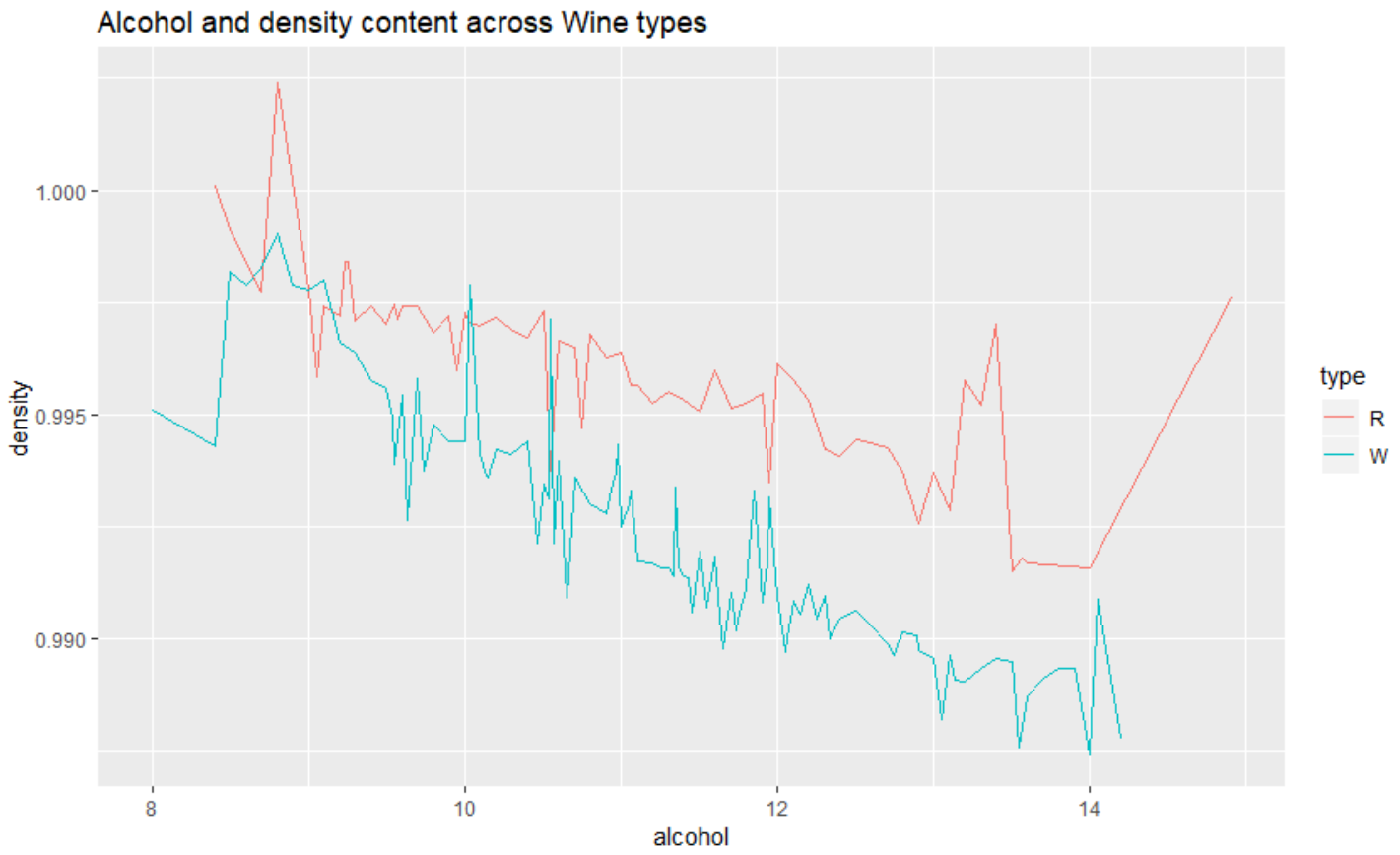


Observations:

This shows that as quality rating increases, the White wine has instances of more PH value then Red wine for similar Alcohol content.
Therefore we can even conclude that more PH leads to better Quality.

Alcohol and density content across wine types:

```
ggplot(aes(x=alcohol,y=density),
      data = winedata)+
  geom_line(aes(color=type),stat='summary',fun.y=median)+
  ggtitle('Alcohol And density content across Wine types')
```

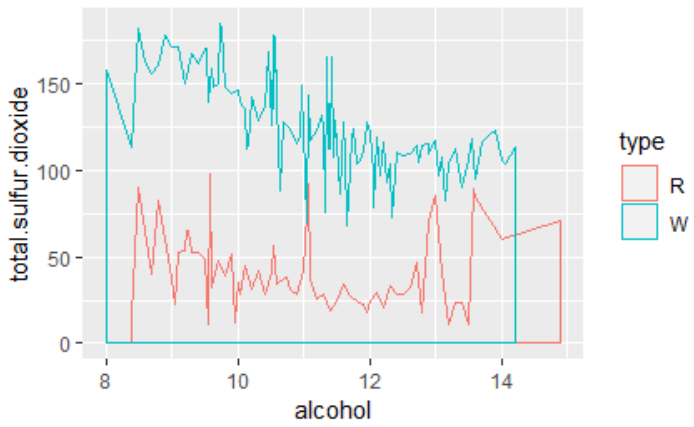


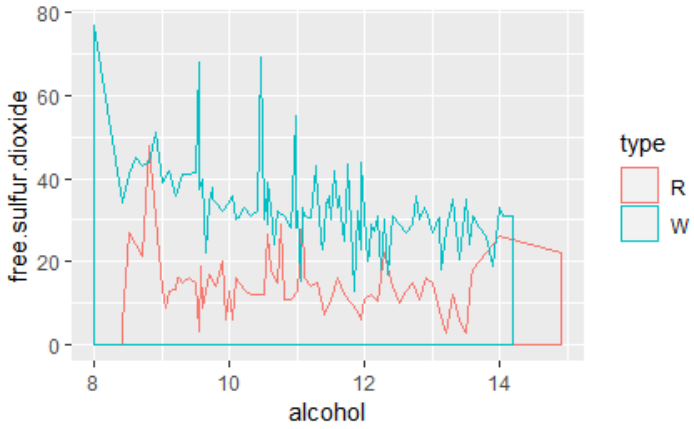
Observations:

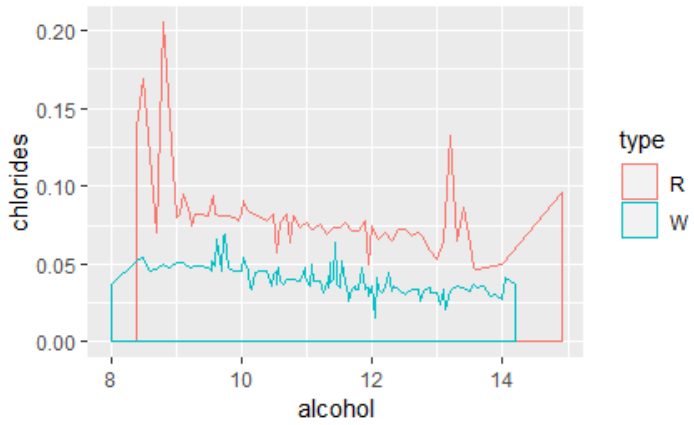
As analyzed before we know that Alcohol exhibits strong negative correlation with density but further from above analysis we can say that as the density decrease the alcohol content also decreases along both Red and white wine

Analysing Alcohol with SO2 (free & Total) and chlorides:

```
alcohol1<-ggplot(aes(x=alcohol,y=total.sulfur.dioxide),
        data = winedata)+
  geom_density(aes(color=type),stat='summary',fun.y=median)
alcohol2<-ggplot(aes(x=alcohol,y=free.sulfur.dioxide),
        data = winedata)+
  geom_density(aes(color=type),stat='summary',fun.y=median)
alcohol3<-ggplot(aes(x=alcohol,y=chlorides),
        data = winedata)+
  geom_density(aes(color=type),stat='summary',fun.y=median)


grid.arrange(alcohol1,alcohol2,alcohol3,ncol=2)
```





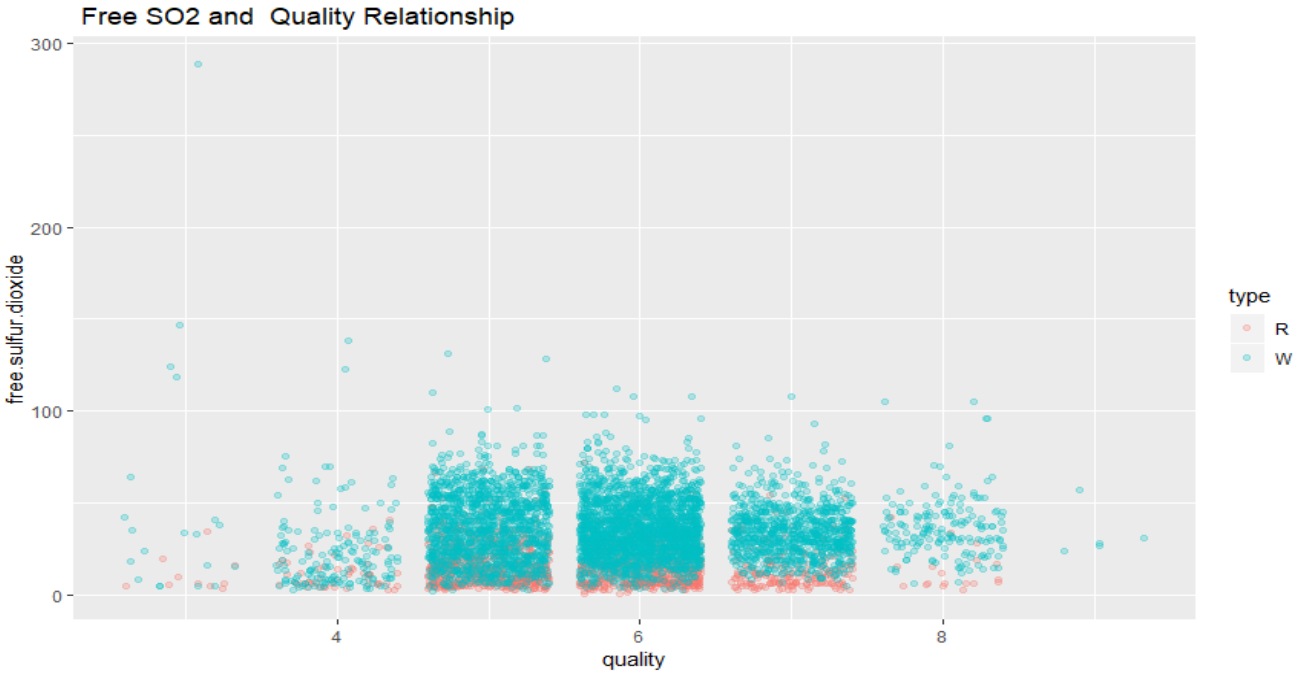


Observations:

- Total SO2
  As the Alcohol content increases it seems like the Total SO2 decreases across white wine.
  White wine contains more SO2 as compared with Red wine across all alcohol content.
- FreeSO2
  It decreases for the major of the samples as the alcohol content increases for White wine.
  And White wine consist of more FreeSO2 content than Red wine for major samplesfor all alcohol contents.
- Chloride
  White wine shows lower level of chloride content across all alcohol contents
  The chloride content in Red wine is greater than White wine in all alcohol contents.

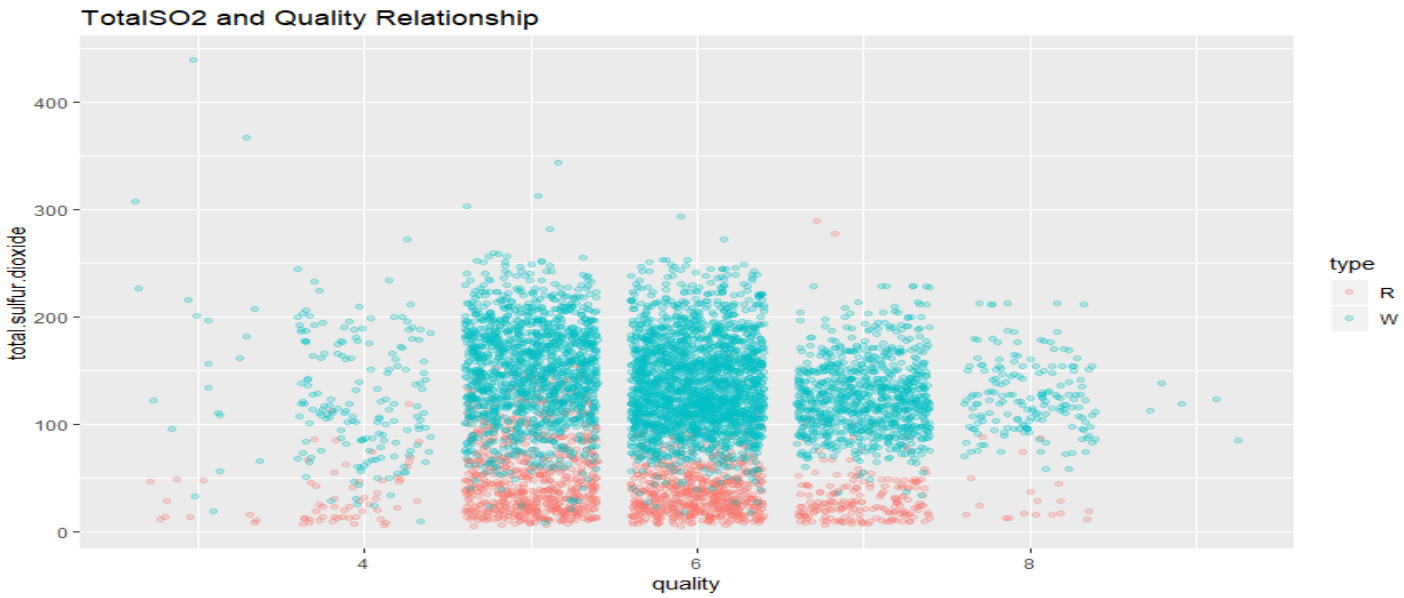
FreeSo2 and Quality Relationship:

```
ggplot(aes(x = quality, y = free.sulfur.dioxide), data = wine) +
  geom_point(aes(color=type),alpha=1/4, position = 'jitter')+
  ggtitle(' FreeSO2 and  Quality Relationship')
```



Total SO2 And Quality

```
ggplot(aes(x = quality, y = total.sulfur.dioxide), data = wine) +
  geom_point(aes(color=type),alpha=1/4, position = 'jitter')+
  ggtitle('TotalSO2 and Quality Relationship')
```

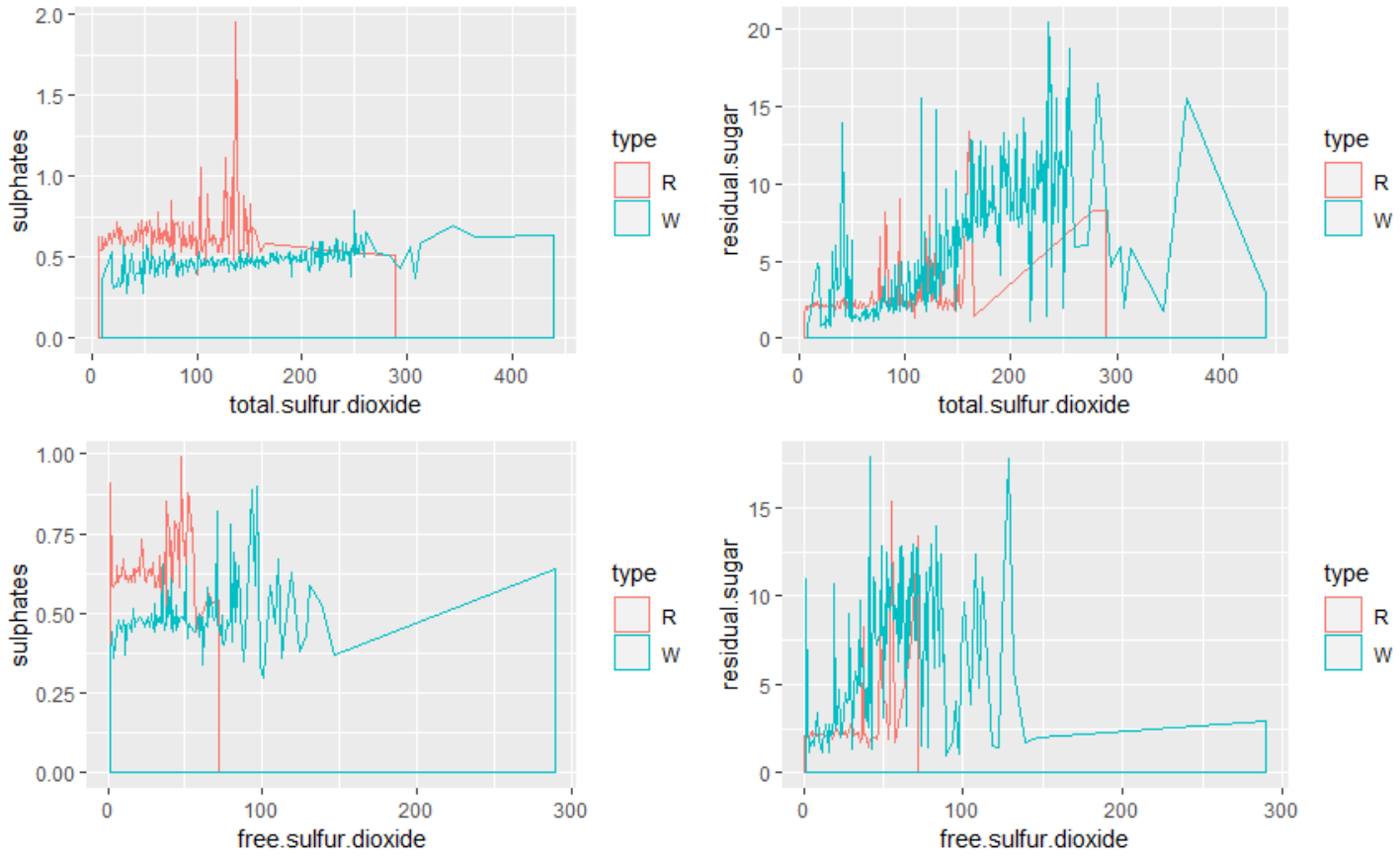


Observations:

Both the above plot of FreeSO2 and Quality, TotalSO2 and Quality Shows that White wine consist of higher content of freeSO2 as compared to Red wine for same Quality levels.

Analyzing TotalSO2 with sulphates and Residual Sugar

```
X1<-ggplot(aes(x=total.sulfur.dioxide,y=sulphates),
        data = winedata)+
 geom_density(aes(color=type),stat='summary',fun.y=median)
X2<-ggplot(aes(x=total.sulfur.dioxide,y=residual.sugar),
        data = winedata)+
 geom_density(aes(color=type),stat='summary',fun.y=median)
x3<-ggplot(aes(x=free.sulfur.dioxide,y=sulphates),
        data = winedata)+
 geom_density(aes(color=type),stat='summary',fun.y=median)
x4<-ggplot(aes(x=free.sulfur.dioxide,y=residual.sugar),
        data = winedata)+
 geom_density(aes(color=type),stat='summary',fun.y=median)


grid.arrange(X1,X2,x3,x4, ncol=2)
```

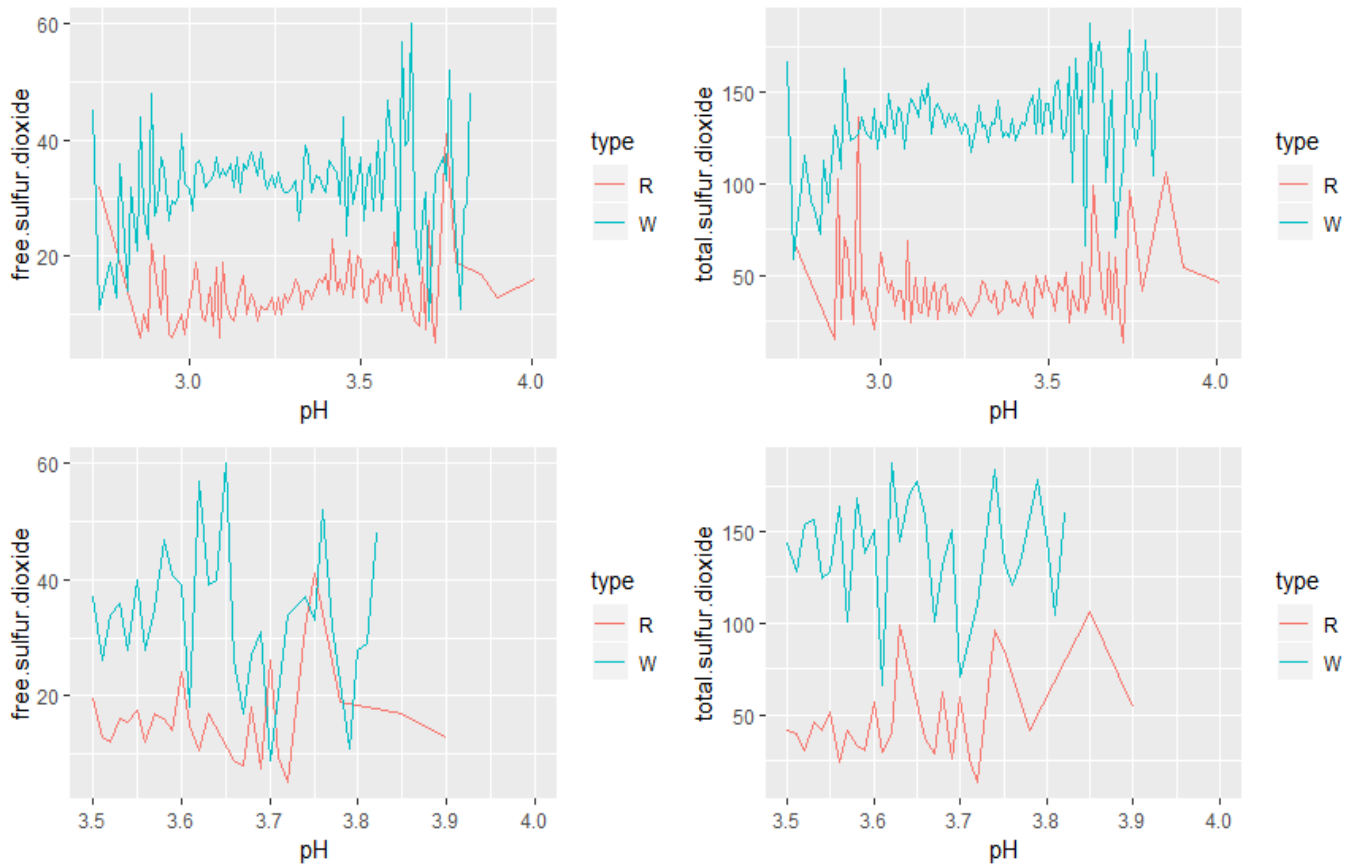


Observations:

- Sulphate and TotalSO2
  The Red Wine exhibits more content of sulphate as compared to White wine upto 250 mark and then White wine exhibits more content of sulphate.
  Only White wine exhibits sulphates as well as TotalSO2 beyond the 280 mark.
- Residual Sugar and TotalSO2
  White wine Shows high Residual sugar content as compare to White Wine.
- Sulphate and FreeSO2
  Sulphate content is very high for Red wine as compared to White wine and Red wine does not contain sulphate beyond 70 mark.
- Residual Sugar and FreeSO2
  White wine contains high level of sugar as compare to Red Wine.

PH and Sulphur dioxide Relationship:

```
Y1<-ggplot(aes(x=pH,y=free.sulfur.dioxide),
       data = winedata)+
 geom_line(aes(color=type),stat='summary',fun.y=median)
Y2<-ggplot(aes(x=pH,y=total.sulfur.dioxide),
        data = winedata)+
 geom_line(aes(color=type),stat='summary',fun.y=median)
Y3<-ggplot(aes(x=pH,y=free.sulfur.dioxide),
        data = winedata)+
 geom_line(aes(color=type),stat='summary',fun.y=median)+
 xlim(3.5,4.0)
Y4<-ggplot(aes(x=pH,y=total.sulfur.dioxide),
        data = winedata)+
 geom_line(aes(color=type),stat='summary',fun.y=median)+
 xlim(3.5,4.0)


grid.arrange(Y1,Y2,Y3,Y4,ncol=2)
```



Observations:

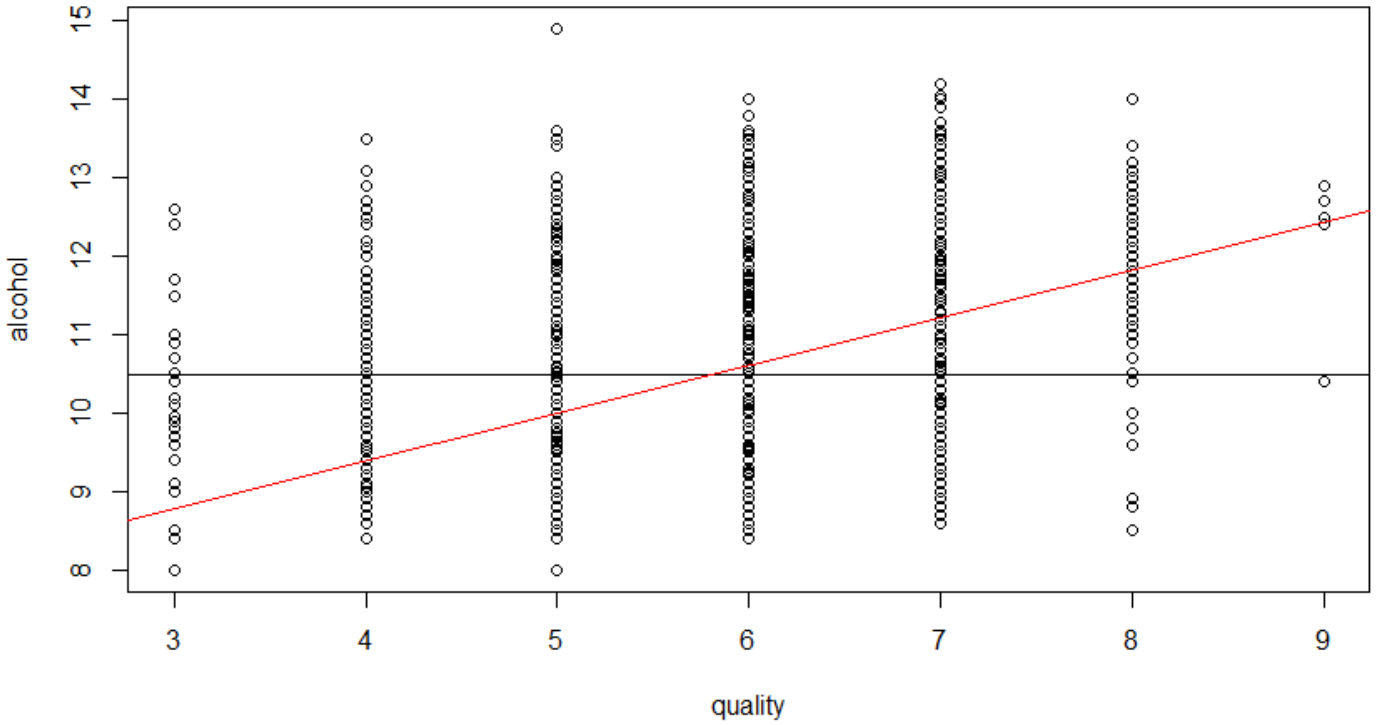
- PH and FreeSO2
  White Wine exhibits more content of FreeSO2 as compared to Red wine for major samples, Red Wine shows more content of FreeSO2 compare to White wine for PH 1-2 marks.
  White Wine does not exhibit FreeSO2 beyond PH value 3.8
- PH and TotalSO2
  Again, White Wine exhibits more content of TotalSO2 as compared to Red wine.
  There seems to be a huge difference between White and Red Wine in terms of TotalSO2 content for PH value 3-3.5.

Linear Regression on Alcohol and Quality:

Linear Regression:

linear regression is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).

```
plot(alcohol~quality,data=winedata)

mean.alcohol=mean(winedata$alcohol)

abline(h=mean.alcohol)

model2=lm(alcohol~quality,data=winedata)

abline(model2,col="red")
```



Observations:

We can exhibit from the plot that Quality increases with increase in alcohol content.

Summary of Linear Regression on Alcohol and Quality

```
> summary(model2)

Call:
lm(formula = alcohol ~ quality, data = winedata)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3157 -0.7952 -0.1952  0.7048  4.9048

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.96086    0.08933   77.92   <2e-16 ***
quality      0.60686    0.01518   39.97   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.069 on 6495 degrees of freedom
Multiple R-squared:  0.1974,   Adjusted R-squared:  0.1973
F-statistic:  1598 on 1 and 6495 DF,  p-value: < 2.2e-16
```

```
SSE = sum(model2$residuals^2)
> SSE
[1] 7416.618
```

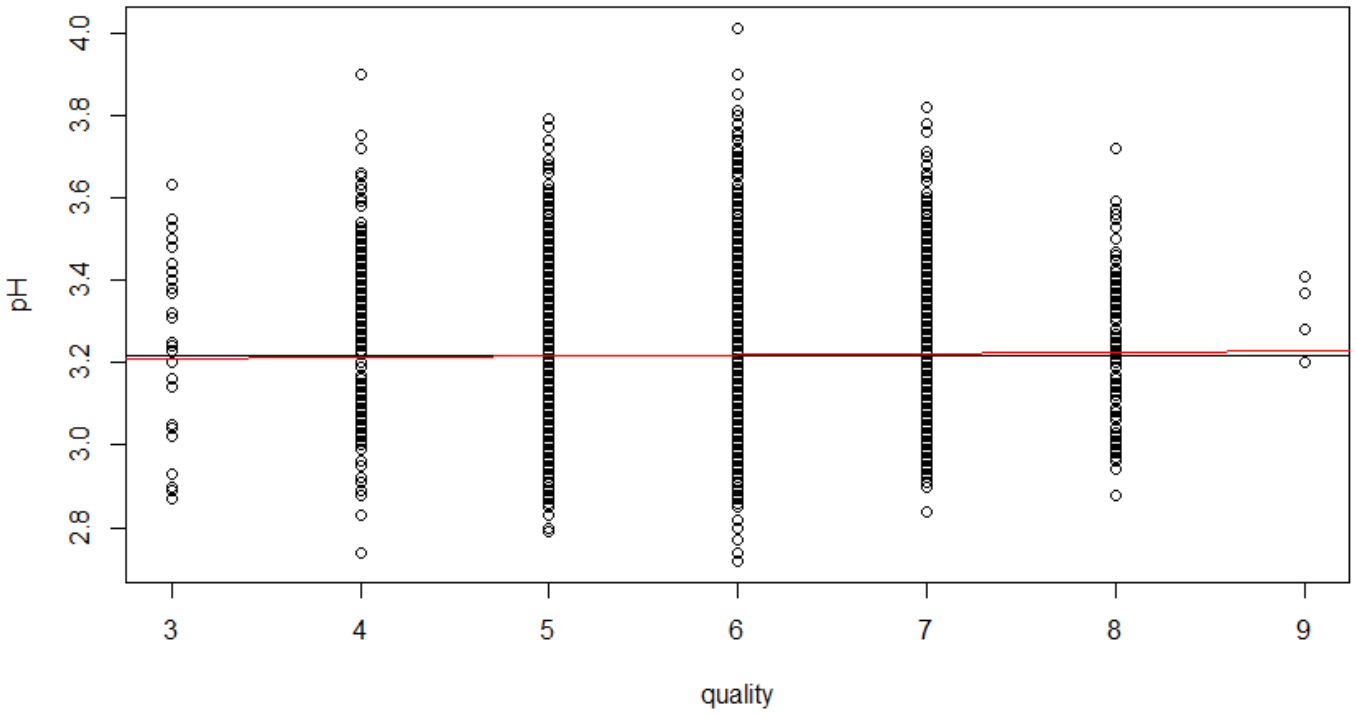
Multiple R-squared value is 0.1974 which is the R-squared value.

Whereas, the adjusted R-squared value is 0.1973

Sums of squared errors is 7416.618

Linear regression on PH and Quality:

```
plot(pH~quality,data=winedata)
mean.pH=mean(winedata$pH)
abline(h=mean.pH)
model1=lm(pH~quality,data=winedata)
abline(model1,col="red")
```



```
> summary(model1)

Call:
lm(formula = pH ~ quality, data = winedata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.49915 -0.10915 -0.01197  0.10085  0.79085

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.197604   0.013439 237.930   <2e-16 ***
quality     0.003591   0.002284   1.572    0.116
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1608 on 6495 degrees of freedom
Multiple R-squared:  0.0003805,       Adjusted R-squared:  0.0002266
F-statistic: 2.472 on 1 and 6495 DF,  p-value: 0.1159
```

```
SSE = sum(model1$residuals^2)
> SSE
[1] 167.8741
```

0bservation:

PH value is important but does not have a significant impact in deciding the Quality of the Wine.

Multiple R-squared value is 0.0003 which is the R-squared value.

Whereas, the adjusted R-squared value is 0.

Sum of squared errors is 167.8741

Multiple Regression on Quality and Alcohol-PH content:

Multiple Regression:

Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables.

```
regression <- lm(quality ~ (alcohol+pH),data=winedata)

summary(regression)
```

```
> summary(regression)

Call:
lm(formula = quality ~ (alcohol + pH), data = winedata)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4934 -0.4930 -0.0374  0.5123  3.2083

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.982487   0.204145  14.610  < 2e-16 ***
alcohol      0.328408   0.008194  40.080  < 2e-16 ***
pH          -0.189438   0.060781  -3.117  0.00184 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7819 on 6494 degrees of freedom
Multiple R-squared:  0.1986,   Adjusted R-squared:  0.1984
F-statistic: 804.7 on 2 and 6494 DF,  p-value: < 2.2e-16
```

```
> SSE = sum(regression$residuals^2)
> SSE
[1] 3969.796
```

Multiple R-squared value is 0.1986 which is the R-squared value.

Whereas, the adjusted R-squared value is 0.1984

Which is more than the previous models which consisted of alcohol and PH independently and indicates that this model with both Alcohol and PH is better than previous models.

Sum of Squared error is 3969.796 better than previous models.

Multiple Regression on all parameter contents

```
multipleregression <- lm(quality ~
(volatile.acidity+citric.acid+fixed.acidity+alcohol+pH+total.sulfur.dioxide+free.sulfur.dioxide+sulphates+density+residual.sugar),data=winedata)

summary(multipleregression)
```

```
 > summary(multipleregression)

Call:
lm(formula = quality ~ (volatile.acidity + citric.acid + fixed.acidity +
    alcohol + pH + total.sulfur.dioxide + free.sulfur.dioxide +
    sulphates + density + residual.sugar), data = winedata)

Residuals:
    Min      1Q  Median      3Q     Max
-3.7554 -0.4577 -0.0407  0.4664  2.9926

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           5.922e+01  1.165e+01   5.081 3.86e-07 ***
volatile.acidity     -1.349e+00  7.596e-02 -17.764  < 2e-16 ***
citric.acid          -1.277e-01  7.866e-02  -1.623    0.105
fixed.acidity         7.149e-02  1.535e-02   4.657 3.28e-06 ***
alcohol               2.672e-01  1.673e-02  15.971  < 2e-16 ***
pH                    4.634e-01  8.884e-02   5.216 1.88e-07 ***
total.sulfur.dioxide -2.434e-03  2.748e-04  -8.859  < 2e-16 ***
free.sulfur.dioxide   5.910e-03  7.501e-04   7.880 3.82e-15 ***
sulphates             7.470e-01  7.471e-02   9.998  < 2e-16 ***
density              -5.856e+01  1.188e+01  -4.928 8.50e-07 ***
residual.sugar        4.524e-02  5.025e-03   9.003  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7354 on 6486 degrees of freedom
Multiple R-squared:  0.2919,   Adjusted R-squared:  0.2908
F-statistic: 267.4 on 10 and 6486 DF,  p-value: < 2.2e-16
```

```
> SSE = sum(multipleregression$residuals^2)
> SSE
[1] 3507.674
```

Multiple R-squared value is 0.2919 which is the R-squared value.

Whereas, the adjusted R-squared value is 0.2908

Which is more than all the previous models.

Sum of Squared error is 3507.674 better than all previous models.

Conclusion:

Analysis performed on the dataset can be concluded as:

- pH value is considered an important parameter when determining the quality of the Wine. As defined by the linear regression PH and Quality relation had the least sum of squared errors.
- Alcohol is also the important factor to determine the quality of the wine concluded from the linear regression performed on the alcohol and Quality where it was observed that the Quality of wine increases with increase in alcohol content and it exhibited strong positive correlation with quality.
- The analysis over the samples also indicate that there are no specific values of pH which provides bias for quality ratings and a higher density of Red Wine samples did indicate a higher PH values as compared to White wine samples for the same quality ratings.
- White Wine exhibits more content of FreeSO2 as compared to Red wine for major samples, Red Wine shows more content of FreeSO2 compare to White wine for PH 1-2 marks.
- White Wine does not exhibit FreeSO2 beyond PH value 3.8
- White Wine exhibits more content of TotalSO2 as compared to Red wine.
- There seems to be a huge difference between White and Red Wine in terms of TotalSO2 content for PH value 3-3.5.
- The wines with higher Alcohol content exhibited lower SO2 content as compared with samples with lower Alcohol content.
- White wine Shows high Residual sugar content as compare to White Wine
- Therefore, customer conscious of sugar is more likely to buy a Red wine than White wine.
- White wine shows lower level of chloride content across all alcohol contents
- The chloride content in Red wine is greater than White wine in all alcohol contents.

References:

https://archive.ics.uci.edu/ml/datasets/Wine+Quality

Multiple Linear Regression in R | R Tutorial 5.3 | MarinStatsLectures (https://www.youtube.com/watch?v=q1RD5ECsSB0)

http://learntech.uwe.ac.uk/da/default.aspx?pageid=1442

statisticsfun (https://www.youtube.com/watch?v=WRp_MpYQFbg)

https://rstudio-pubs-static.s3.amazonaws.com/326428_e7b3416aa30140e79f57b8ed70ca64e4.html