WHAT MAKES PORTUGUESE VINHO VERDE WINE FINE?

Siyabonga Mabuza & Chilekwa Lingoma

12/15/2021

Table of Contents

[Abstract 1](#_Toc93102278)

[Data Characteristics 2](#_Toc93102279)

[Scatter Plot matrix and Correlations 18](#_Toc93102280)

[First Order Logistic Regression Model 20](#_Toc93102281)

[Jittered Response vs. Predicted Values Plot 22](#_Toc93102282)

[Residuals vs Fitted Plot 23](#_Toc93102283)

[Variance Inflaction Factors, Cook’s distance, and High Leverage 24](#_Toc93102284)

[Model Selection 25](#_Toc93102285)

[Two Way Interaction effects 28](#_Toc93102286)

[Model Selection on Interaction Model 31](#_Toc93102287)

[Final Model 45](#_Toc93102288)

[Model Diagnostics 49](#_Toc93102289)

[Likelihood ratio test for final model and Goodness of fit test 49](#_Toc93102290)

[Residual Analysis & Diagnostic Analysis 49](#_Toc93102291)

[Scatter Plot for Final Model 50](#_Toc93102292)

[Model Plot 50](#_Toc93102293)

[Influence Analysis 51](#_Toc93102294)

[ROC Curve 52](#_Toc93102295)

[Conclusions 53](#_Toc93102296)

# Abstract

What makes fine wine? The red wine variant of the Portuguese “Vinho Verde” wine refers to Portuguese wine that originated in the historic Minho province in the far north of the country. The main goal of this investigation is to find which features of these kinds of wine provide the most information about its quality.

This investigation aims to answer the following questions:

* What physiochemical quality is the most important to making good quality wine?
* What role does the level of alcohol have in the quality of wine
* Does less residual sugar mean less quality?

This dataset contains one categorical response variable (High or Poor) quality wine and 12 physicochemical predictor variables. The 12 physicochemical predictor variables are: Fixed Acidity, Volatile Acidity, Citric Acid, Residual Sugar, Chlorides, Free Sulfur Dioxide, Total Sulfur Dioxide, Density, pH, Sulphates, and Alcohol.

The response variable is Quality. It was originally ranked on a scale of 1-10 with 1 through 5 being wine of poor quality and 6 through 10 being wine high quality. We changed this to 1 for high quality and 0 for poor quality. There are 1599 wine variants.

Our final model had an AUC of 84% prediction for quality of wine explained by the predictors of our final model. Thus, our model is significant in predicting the quality of wine. Log of Alcohol was the most significant predictor in the final model so interestingly, wines with higher alcohol contents have higher quality. Furthermore, residual sugar was one of the first predictors removed using stepwise regression. It seems that residual sugar has no significance towards the quality of wine.

# Data Characteristics

library(readxl)  
winequality <- read\_excel("C:/Users/Siyabonga Mabuza/Desktop/wine.xlsx")  
wine.data = winequality  
names(wine.data) [1] = "quality"  
attach(wine.data)  
head(wine.data)

## # A tibble: 6 x 12  
## quality `fixed acidity` `volatile acidity` `citric acid` `residual sugar`  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 0 7.4 0.7 0 1.9  
## 2 0 7.8 0.88 0 2.6  
## 3 0 7.8 0.76 0.04 2.3  
## 4 1 11.2 0.28 0.56 1.9  
## 5 0 7.4 0.7 0 1.9  
## 6 0 7.4 0.66 0 1.8  
## # ... with 7 more variables: chlorides <dbl>, free sulfur dioxide <dbl>,  
## # total sulfur dioxide <dbl>, density <dbl>, pH <dbl>, sulphates <dbl>,  
## # alcohol <dbl>

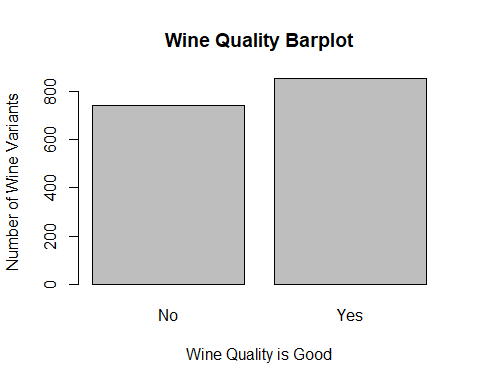
First, we will examine the characteristics of the response variable.

library(vcd)

## Warning: package 'vcd' was built under R version 4.1.2

## Loading required package: grid

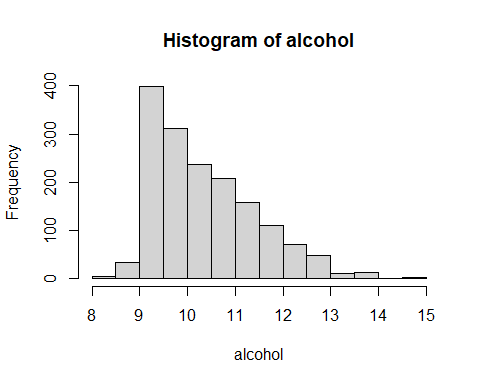
library(MASS)  
barplot (table (ifelse ( quality ==1, "Yes", "No")), xlab="Wine Quality is Good",  
ylab="Number of Wine Variants", main = "Wine Quality Barplot")



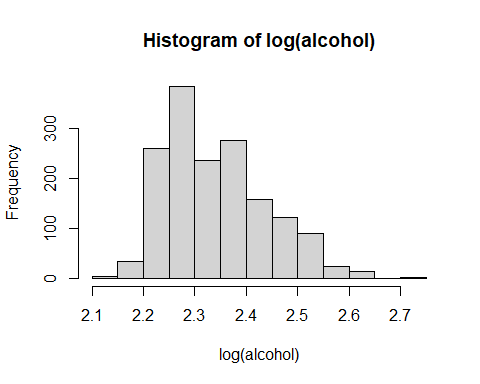
This dataset has a sample size of size of 1599 wine variants. Within this sample, 855 of the wine variants have good quality and 744 do not have good quality. Based on this, no transformation is required for the response variable.

Now we will examine the characteristics of the predictor variables.

library(ggplot2)  
library(tidyr)  
hist(alcohol)

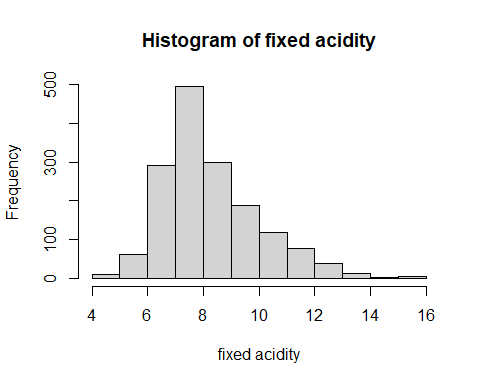


hist(log(alcohol))

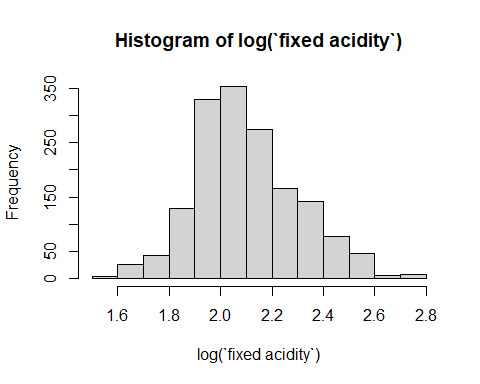


Alcohol is right skewed, a log transformation slightly alleviates this. Thus, we will use a log transformation for alcohol.

hist(`fixed acidity`)

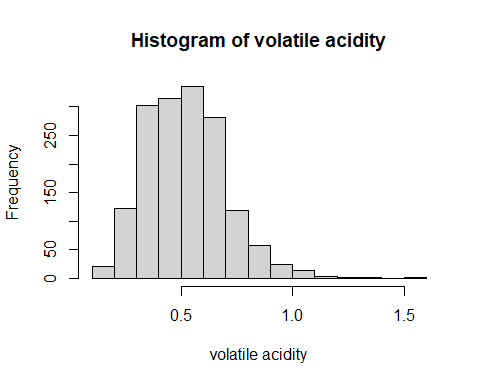


hist(log(`fixed acidity`))

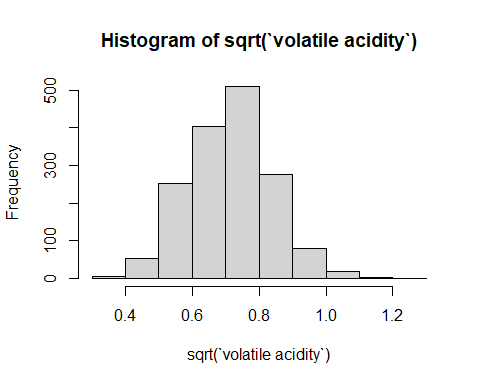


Fixed acidity is right skewed. A log transformation resolves this, thus, we will use a log transformation for fixed acidity.

hist(`volatile acidity`)

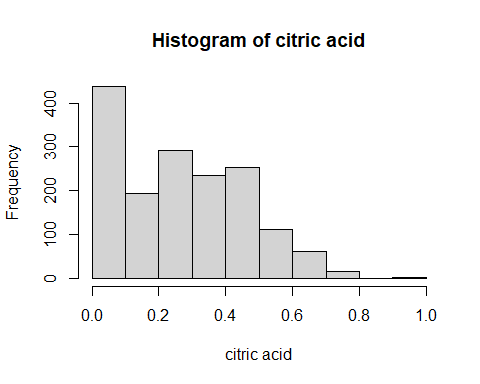


hist(sqrt(`volatile acidity`))

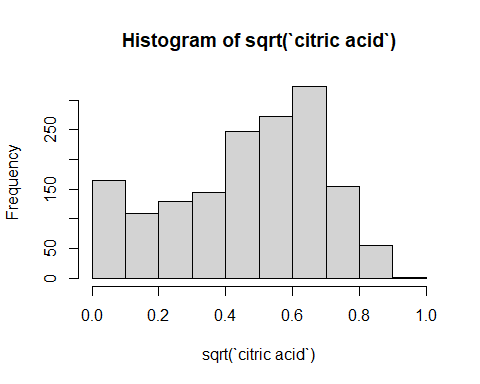


Volatile Acidity is slightly rightly skewed. A square root transformation alleviates this. Thus, we will use a square root transformation for Volatile Acidity.

hist(`citric acid`)

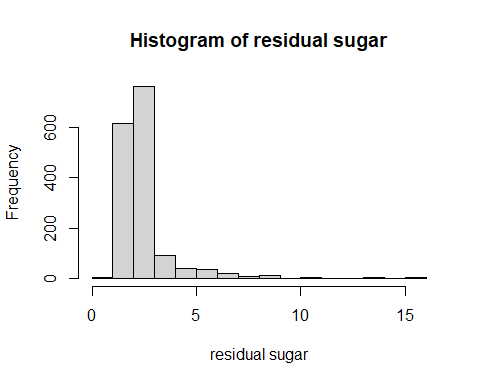


hist(sqrt(`citric acid`))

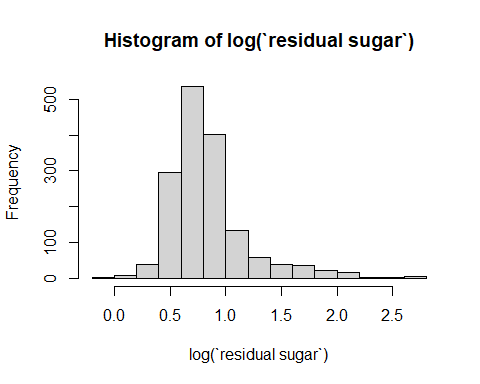


Citric Acid is rightly skewed, a square root transformation slightly alleviates this. Thus, we will make a square root transformation for Citric Acid.

hist(`residual sugar`)

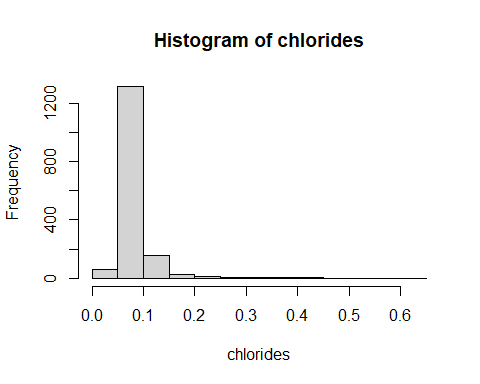


hist(log(`residual sugar`))

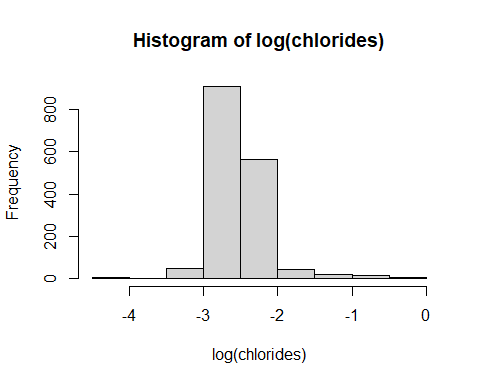


Residual Sugar is rightly skewed, a log transformation alleviates this. Thus, we will use a log transformation for residual sugar.

hist(chlorides)

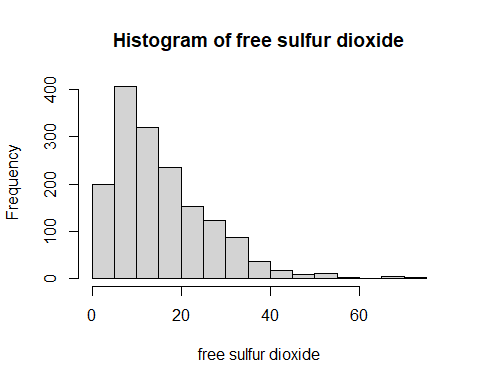


hist(log(chlorides))

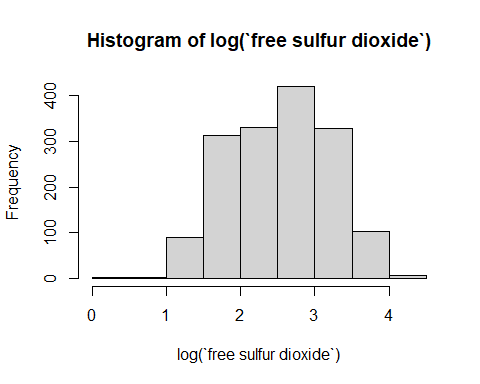


Chlorides is rightly skewed, a log transformation resolves this. Thus, we will use a log transformation for Chlorides.

hist(`free sulfur dioxide`)

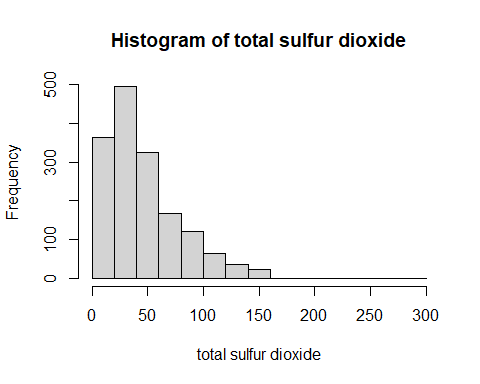


hist(log(`free sulfur dioxide`))

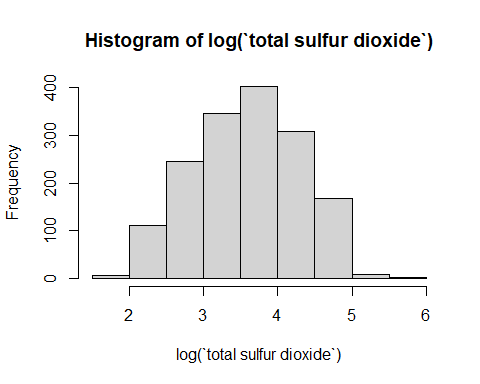


Free Sulfur Dioxide is rightly skewed, a log transformation resolves this. Thus we will use a log transformation for Free Sulfur Dioxide.

hist(`total sulfur dioxide`)

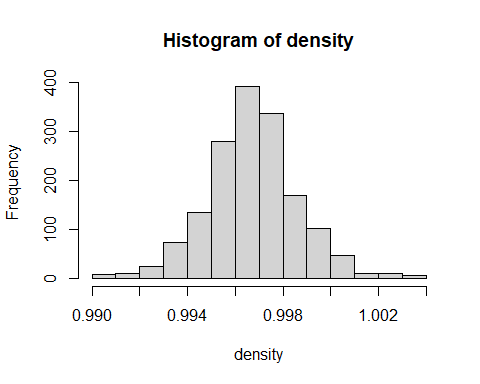


hist(log(`total sulfur dioxide`))



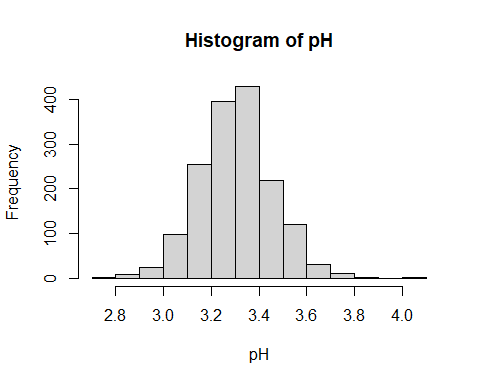
Total Sulfur Dioxide is rightly skewed, a log transformation resolves this. Thus, Total Sulfur Dioxide will be log transformed.

hist(density)



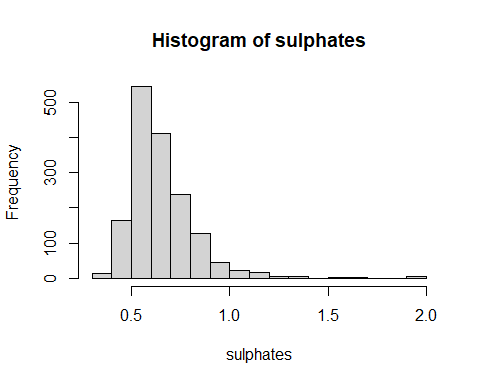
No transformation is necessary for density.

hist(pH)

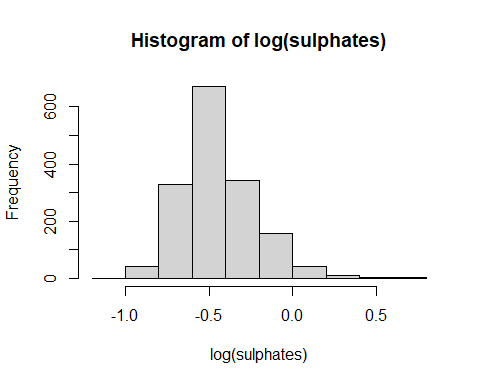


No transformation is necessary for pH.

hist(sulphates)



hist(log(sulphates))

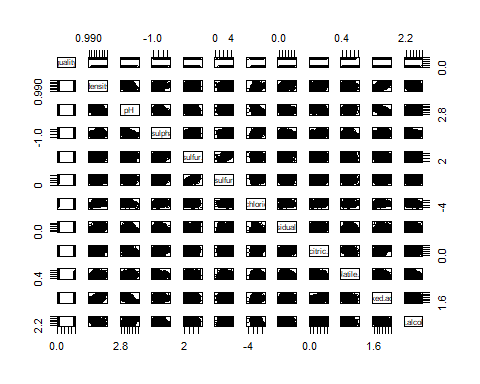


Sulphates is right skewed, a log transformation alleviates this. Thus, a log transformation will be used for sulphates.

# Scatter Plot matrix and Correlations

We will now create a scatterplot and a correlation plot with coefficients in order to check for any linear association between the variables.

wine.data$log.sulphates = log(sulphates)  
wine.data$log.totalsulfur.dioxide = log(`total sulfur dioxide`)  
wine.data$log.freesulfur.dioxide = log(`free sulfur dioxide`)  
wine.data$log.chlorides = log(chlorides)  
wine.data$log.residual.sugar = log(`residual sugar`)  
wine.data$sqrt.citric.acid = sqrt(`citric acid`)  
wine.data$sqrt.volatile.acidity = sqrt(`volatile acidity`)  
wine.data$log.fixed.acidity = log(`fixed acidity`)  
wine.data$log.alcohol = log(alcohol)  
  
plot(wine.data [, c(1, 9:10, 13:21)])

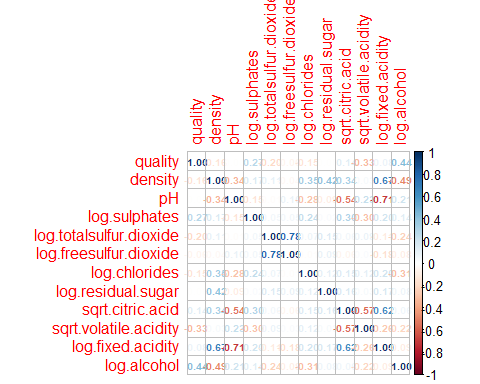


transcorr = cor(wine.data [, c(1, 9:10, 13:21)], use= "complete.obs")  
library("corrplot")

## Warning: package 'corrplot' was built under R version 4.1.2

## corrplot 0.92 loaded

## corrplot 0.84 loaded  
corrplot (transcorr, method = "number", number.cex=0.6)



Log Free Sulfur Dioxide and Log Total Sulfur Dioxide have a relatively high positive correlation (0.78). This is expected because the Total sulfur dioxide measures free and bound forms of Sulfur Dioxide. Density and log Fixed Acidity are reasonably positively correlated(0.67). pH and log Fixed Acidity have a relatively high negative correlation (0.67).

There are no variables that have extremely high positive or negative correlations. However, there are numerous variables with low positive and negative correlations.

# First Order Logistic Regression Model

We now fit a logistic regression model that includes all of the variables and their transformations.

fit1 = glm (data=wine.data, quality ~ log(alcohol) + log(`fixed acidity`) + sqrt(`volatile acidity`) + sqrt(`citric acid`) + log(`residual sugar`) + log(chlorides) + log(`free sulfur dioxide`) + log(`total sulfur dioxide`) + log(sulphates) + pH + density, family = "binomial")  
summary(fit1)

##   
## Call:  
## glm(formula = quality ~ log(alcohol) + log(`fixed acidity`) +   
## sqrt(`volatile acidity`) + sqrt(`citric acid`) + log(`residual sugar`) +   
## log(chlorides) + log(`free sulfur dioxide`) + log(`total sulfur dioxide`) +   
## log(sulphates) + pH + density, family = "binomial", data = wine.data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1893 -0.8328 0.3130 0.8199 2.3241   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 53.68411 86.63539 0.620 0.53548   
## log(alcohol) 8.65912 1.12686 7.684 1.54e-14 \*\*\*  
## log(`fixed acidity`) 1.71561 0.86090 1.993 0.04628 \*   
## sqrt(`volatile acidity`) -4.78121 0.70380 -6.793 1.09e-11 \*\*\*  
## sqrt(`citric acid`) -1.25723 0.42668 -2.947 0.00321 \*\*   
## log(`residual sugar`) 0.09247 0.23720 0.390 0.69667   
## log(chlorides) -0.48109 0.21919 -2.195 0.02817 \*   
## log(`free sulfur dioxide`) 0.41729 0.15067 2.770 0.00561 \*\*   
## log(`total sulfur dioxide`) -0.63177 0.15605 -4.049 5.15e-05 \*\*\*  
## log(sulphates) 2.35108 0.32795 7.169 7.55e-13 \*\*\*  
## pH -0.01264 0.73752 -0.017 0.98633   
## density -72.49657 88.04194 -0.823 0.41026   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2209.0 on 1598 degrees of freedom  
## Residual deviance: 1655.6 on 1587 degrees of freedom  
## AIC: 1679.6  
##   
## Number of Fisher Scoring iterations: 4

exp (fit1$coefficients)

## (Intercept) log(alcohol)   
## 2.064024e+23 5.762434e+03   
## log(`fixed acidity`) sqrt(`volatile acidity`)   
## 5.560081e+00 8.385836e-03   
## sqrt(`citric acid`) log(`residual sugar`)   
## 2.844395e-01 1.096876e+00   
## log(chlorides) log(`free sulfur dioxide`)   
## 6.181067e-01 1.517844e+00   
## log(`total sulfur dioxide`) log(sulphates)   
## 5.316505e-01 1.049685e+01   
## pH density   
## 9.874420e-01 3.274462e-32

exp (confint (fit1))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 2.969629e-51 1.265343e+97  
## log(alcohol) 6.479989e+02 5.387842e+04  
## log(`fixed acidity`) 1.033575e+00 3.028360e+01  
## sqrt(`volatile acidity`) 2.068596e-03 3.271604e-02  
## sqrt(`citric acid`) 1.225687e-01 6.535636e-01  
## log(`residual sugar`) 6.888600e-01 1.747374e+00  
## log(chlorides) 4.006964e-01 9.471070e-01  
## log(`free sulfur dioxide`) 1.129945e+00 2.040923e+00  
## log(`total sulfur dioxide`) 3.910032e-01 7.212038e-01  
## log(sulphates) 5.559977e+00 2.013029e+01  
## pH 2.325267e-01 4.197641e+00  
## density 3.180629e-107 3.368764e+43

Upon fitting the first order model, there are eight variables that are significant. These variables are:

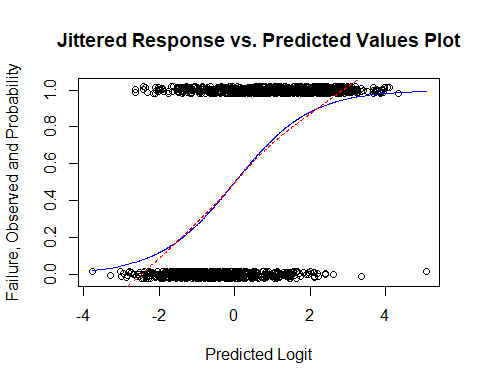
* log of alcohol
* log of fixed acidity
* square root of volatile.acidity
* square root of citric.acid
* log of chlorides
* log of free sulfur dioxide
* log of total sulfur dioxide
* log of sulphates

Log of free sulfur dioxide and log of total sulfur dioxide which also happen to have the highest correlation between the variables are the two most significant variables.

# Jittered Response vs. Predicted Values Plot

Next we will plot the jittered response vs. predicted values.

pred.logit = predict (fit1)  
plot(jitter(quality,0.1) ~ pred.logit, data = wine.data, main = 'Jittered Response vs. Predicted Values Plot', xlab="Predicted Logit", ylab="Failure, Observed and Probability")  
pred.r = predict(fit1,type='response')  
pred.order =order(pred.logit)  
lines(pred.logit[pred.order],pred.r[pred.order],col='blue')  
lines(lowess(pred.logit[pred.order], pred.r[pred.order]),col='red',lty=2)

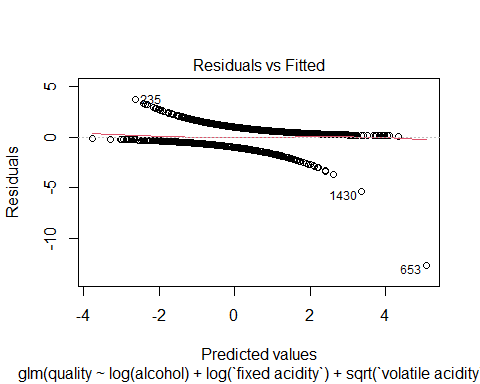


Initially, there seems to be a higher density of good quality and not good quality as the predicted logit value of the model increases. However, once the predicted logit value surpasses two, there seems to be a higher density of good quality.

# Residuals vs Fitted Plot

We will now examine the residuals vs fitted plot.

plot(fit1, which=1)



Overall, the residual distribution from our model appears to have little deviation from zero, so the distribution of the residuals is not concerning. There are some obvious outliers, particularly point 653.

# Variance Inflaction Factors, Cook’s distance, and High Leverage

First, we will check for multicollinearity.

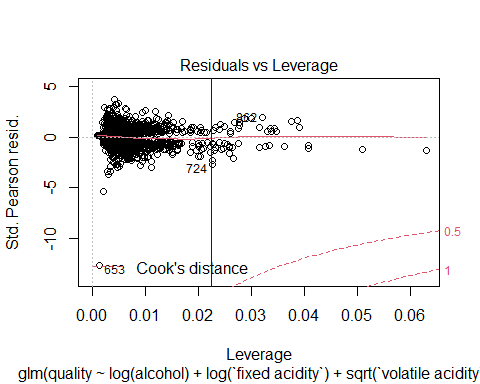
car:: vif (fit1)

## log(alcohol) log(`fixed acidity`)   
## 2.472997 7.883668   
## sqrt(`volatile acidity`) sqrt(`citric acid`)   
## 1.725502 2.704049   
## log(`residual sugar`) log(chlorides)   
## 1.964659 1.406245   
## log(`free sulfur dioxide`) log(`total sulfur dioxide`)   
## 2.848886 3.116081   
## log(sulphates) pH   
## 1.385253 3.650965   
## density   
## 6.814337

Log of fixed acidity and density have relatively high has VIF values (They are above the cautionary value of 5), 7.88 and 6.81, respectfully. However, because this is not extremely high we will keep these variables.

Next, we will examine leverage and Cook’s distance.

plot(fit1, which=5)  
abline(v=0.0225, col="blue")



There do not seem to be any points with high cook’s distance. The high leverage cutoff point is 3(11+1)/1599=0.0225. There are numerous points with high leverage. This is concerning.

# Model Selection

We will now fit the initial model using AIC stepwise regression.

fit2= step(fit1, direction = "both")

## Start: AIC=1679.63  
## quality ~ log(alcohol) + log(`fixed acidity`) + sqrt(`volatile acidity`) +   
## sqrt(`citric acid`) + log(`residual sugar`) + log(chlorides) +   
## log(`free sulfur dioxide`) + log(`total sulfur dioxide`) +   
## log(sulphates) + pH + density  
##   
## Df Deviance AIC  
## - pH 1 1655.6 1677.6  
## - log(`residual sugar`) 1 1655.8 1677.8  
## - density 1 1656.3 1678.3  
## <none> 1655.6 1679.6  
## - log(`fixed acidity`) 1 1659.6 1681.6  
## - log(chlorides) 1 1660.5 1682.5  
## - log(`free sulfur dioxide`) 1 1663.3 1685.3  
## - sqrt(`citric acid`) 1 1664.4 1686.4  
## - log(`total sulfur dioxide`) 1 1672.2 1694.2  
## - sqrt(`volatile acidity`) 1 1705.6 1727.6  
## - log(sulphates) 1 1710.6 1732.6  
## - log(alcohol) 1 1718.8 1740.8  
##   
## Step: AIC=1677.63  
## quality ~ log(alcohol) + log(`fixed acidity`) + sqrt(`volatile acidity`) +   
## sqrt(`citric acid`) + log(`residual sugar`) + log(chlorides) +   
## log(`free sulfur dioxide`) + log(`total sulfur dioxide`) +   
## log(sulphates) + density  
##   
## Df Deviance AIC  
## - log(`residual sugar`) 1 1655.8 1675.8  
## - density 1 1656.7 1676.7  
## <none> 1655.6 1677.6  
## + pH 1 1655.6 1679.6  
## - log(chlorides) 1 1660.8 1680.8  
## - log(`free sulfur dioxide`) 1 1663.4 1683.4  
## - sqrt(`citric acid`) 1 1664.5 1684.5  
## - log(`fixed acidity`) 1 1664.9 1684.9  
## - log(`total sulfur dioxide`) 1 1672.4 1692.4  
## - sqrt(`volatile acidity`) 1 1705.6 1725.6  
## - log(sulphates) 1 1712.0 1732.0  
## - log(alcohol) 1 1742.4 1762.4  
##   
## Step: AIC=1675.81  
## quality ~ log(alcohol) + log(`fixed acidity`) + sqrt(`volatile acidity`) +   
## sqrt(`citric acid`) + log(chlorides) + log(`free sulfur dioxide`) +   
## log(`total sulfur dioxide`) + log(sulphates) + density  
##   
## Df Deviance AIC  
## - density 1 1656.8 1674.8  
## <none> 1655.8 1675.8  
## + log(`residual sugar`) 1 1655.6 1677.6  
## + pH 1 1655.8 1677.8  
## - log(chlorides) 1 1660.9 1678.9  
## - log(`free sulfur dioxide`) 1 1663.6 1681.6  
## - sqrt(`citric acid`) 1 1664.5 1682.5  
## - log(`fixed acidity`) 1 1665.3 1683.3  
## - log(`total sulfur dioxide`) 1 1672.4 1690.4  
## - sqrt(`volatile acidity`) 1 1706.0 1724.0  
## - log(sulphates) 1 1713.4 1731.4  
## - log(alcohol) 1 1774.9 1792.9  
##   
## Step: AIC=1674.77  
## quality ~ log(alcohol) + log(`fixed acidity`) + sqrt(`volatile acidity`) +   
## sqrt(`citric acid`) + log(chlorides) + log(`free sulfur dioxide`) +   
## log(`total sulfur dioxide`) + log(sulphates)  
##   
## Df Deviance AIC  
## <none> 1656.8 1674.8  
## + density 1 1655.8 1675.8  
## + pH 1 1656.3 1676.3  
## + log(`residual sugar`) 1 1656.7 1676.7  
## - log(chlorides) 1 1662.1 1678.1  
## - log(`free sulfur dioxide`) 1 1664.8 1680.8  
## - sqrt(`citric acid`) 1 1665.5 1681.5  
## - log(`fixed acidity`) 1 1667.2 1683.2  
## - log(`total sulfur dioxide`) 1 1674.4 1690.4  
## - sqrt(`volatile acidity`) 1 1710.8 1726.8  
## - log(sulphates) 1 1713.5 1729.5  
## - log(alcohol) 1 1823.5 1839.5

Upon fitting the model using stepwise regression, eight variables are worth keeping. Namely:

* log of alcohol
* log of fixed acidity
* square root of volatile.acidity
* square root of citric.acid
* log of chlorides
* log of free sulfur dioxide
* log of total sulfur dioxide
* log of sulphates

We will now use fit this model using stepwise regression with the BIC criterion.

n = dim (wine.data)[1]  
fit3=step(fit2, direction='both', k=log(n))

## Start: AIC=1723.17  
## quality ~ log(alcohol) + log(`fixed acidity`) + sqrt(`volatile acidity`) +   
## sqrt(`citric acid`) + log(chlorides) + log(`free sulfur dioxide`) +   
## log(`total sulfur dioxide`) + log(sulphates)  
##   
## Df Deviance AIC  
## - log(chlorides) 1 1662.1 1721.1  
## <none> 1656.8 1723.2  
## - log(`free sulfur dioxide`) 1 1664.8 1723.8  
## - sqrt(`citric acid`) 1 1665.5 1724.5  
## - log(`fixed acidity`) 1 1667.2 1726.2  
## - log(`total sulfur dioxide`) 1 1674.4 1733.4  
## - sqrt(`volatile acidity`) 1 1710.8 1769.8  
## - log(sulphates) 1 1713.5 1772.5  
## - log(alcohol) 1 1823.5 1882.5  
##   
## Step: AIC=1721.11  
## quality ~ log(alcohol) + log(`fixed acidity`) + sqrt(`volatile acidity`) +   
## sqrt(`citric acid`) + log(`free sulfur dioxide`) + log(`total sulfur dioxide`) +   
## log(sulphates)  
##   
## Df Deviance AIC  
## <none> 1662.1 1721.1  
## - log(`free sulfur dioxide`) 1 1670.5 1722.1  
## + log(chlorides) 1 1656.8 1723.2  
## - log(`fixed acidity`) 1 1672.1 1723.7  
## - sqrt(`citric acid`) 1 1673.0 1724.7  
## - log(`total sulfur dioxide`) 1 1679.7 1731.4  
## - log(sulphates) 1 1713.5 1765.1  
## - sqrt(`volatile acidity`) 1 1727.4 1779.0  
## - log(alcohol) 1 1866.6 1918.3

Log of chlorides was removed from the model using the BIC criterion. Thus, the best model that can be derived from this first stepwise model is quality ~ log of alcohol + log of fixed acidity + square root of volatile.acidity + square root of citric.acid + log of free sulfur dioxide + log of total sulfur dioxide + log of sulphates.

# Two Way Interaction effects

Now, we will check for interaction effects. We will be using the seven predictors kept in the last model.

my.center = function (x) (x - mean (x))  
logfreesulfurdioxide=my.center(log(`free sulfur dioxide`))  
logfixedacidity=my.center(log(`fixed acidity`))  
sqrtcitricacid=my.center(sqrt(`citric acid`))  
logtotalsulfurdioxide=my.center(log(`total sulfur dioxide`))  
logsulphates=my.center(log(sulphates))  
sqrtvolatileacidity=my.center(sqrt(`volatile acidity`))  
logalcohol=my.center(log(alcohol))  
fit4 = glm (quality ~ ( logfreesulfurdioxide + logfixedacidity + sqrtcitricacid + logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity + logalcohol)^2, family = "binomial")  
summary(fit4)

##   
## Call:  
## glm(formula = quality ~ (logfreesulfurdioxide + logfixedacidity +   
## sqrtcitricacid + logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol)^2, family = "binomial")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.6645 -0.7853 0.2122 0.7706 2.4062   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.6906 0.1109 6.230 4.67e-10  
## logfreesulfurdioxide 0.5656 0.1625 3.480 0.000501  
## logfixedacidity 1.9001 0.4953 3.836 0.000125  
## sqrtcitricacid -2.0005 0.5141 -3.891 9.96e-05  
## logtotalsulfurdioxide -0.9708 0.1699 -5.714 1.10e-08  
## logsulphates 2.7119 0.3710 7.310 2.68e-13  
## sqrtvolatileacidity -5.2403 0.7537 -6.953 3.57e-12  
## logalcohol 10.7133 0.8456 12.670 < 2e-16  
## logfreesulfurdioxide:logfixedacidity -1.6007 0.9856 -1.624 0.104355  
## logfreesulfurdioxide:sqrtcitricacid 0.6536 1.0301 0.635 0.525728  
## logfreesulfurdioxide:logtotalsulfurdioxide -0.5672 0.1464 -3.874 0.000107  
## logfreesulfurdioxide:logsulphates -0.1585 0.6876 -0.230 0.817730  
## logfreesulfurdioxide:sqrtvolatileacidity -1.5097 1.5695 -0.962 0.336099  
## logfreesulfurdioxide:logalcohol 4.1376 1.8837 2.196 0.028057  
## logfixedacidity:sqrtcitricacid -5.8835 1.9596 -3.002 0.002679  
## logfixedacidity:logtotalsulfurdioxide 0.2555 0.9850 0.259 0.795350  
## logfixedacidity:logsulphates 4.0804 2.2152 1.842 0.065475  
## logfixedacidity:sqrtvolatileacidity -3.0171 4.1305 -0.730 0.465118  
## logfixedacidity:logalcohol -7.2480 4.8690 -1.489 0.136586  
## sqrtcitricacid:logtotalsulfurdioxide -0.1998 0.9904 -0.202 0.840090  
## sqrtcitricacid:logsulphates -4.2996 2.0179 -2.131 0.033110  
## sqrtcitricacid:sqrtvolatileacidity 5.4777 3.6442 1.503 0.132802  
## sqrtcitricacid:logalcohol 0.7904 5.3913 0.147 0.883444  
## logtotalsulfurdioxide:logsulphates -1.4536 0.6469 -2.247 0.024634  
## logtotalsulfurdioxide:sqrtvolatileacidity 2.0802 1.5937 1.305 0.191791  
## logtotalsulfurdioxide:logalcohol -2.5375 1.8750 -1.353 0.175945  
## logsulphates:sqrtvolatileacidity -2.7315 3.2768 -0.834 0.404507  
## logsulphates:logalcohol 3.6119 3.9712 0.910 0.363081  
## sqrtvolatileacidity:logalcohol -4.4091 8.5391 -0.516 0.605612  
##   
## (Intercept) \*\*\*  
## logfreesulfurdioxide \*\*\*  
## logfixedacidity \*\*\*  
## sqrtcitricacid \*\*\*  
## logtotalsulfurdioxide \*\*\*  
## logsulphates \*\*\*  
## sqrtvolatileacidity \*\*\*  
## logalcohol \*\*\*  
## logfreesulfurdioxide:logfixedacidity   
## logfreesulfurdioxide:sqrtcitricacid   
## logfreesulfurdioxide:logtotalsulfurdioxide \*\*\*  
## logfreesulfurdioxide:logsulphates   
## logfreesulfurdioxide:sqrtvolatileacidity   
## logfreesulfurdioxide:logalcohol \*   
## logfixedacidity:sqrtcitricacid \*\*   
## logfixedacidity:logtotalsulfurdioxide   
## logfixedacidity:logsulphates .   
## logfixedacidity:sqrtvolatileacidity   
## logfixedacidity:logalcohol   
## sqrtcitricacid:logtotalsulfurdioxide   
## sqrtcitricacid:logsulphates \*   
## sqrtcitricacid:sqrtvolatileacidity   
## sqrtcitricacid:logalcohol   
## logtotalsulfurdioxide:logsulphates \*   
## logtotalsulfurdioxide:sqrtvolatileacidity   
## logtotalsulfurdioxide:logalcohol   
## logsulphates:sqrtvolatileacidity   
## logsulphates:logalcohol   
## sqrtvolatileacidity:logalcohol   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2209.0 on 1598 degrees of freedom  
## Residual deviance: 1572.7 on 1570 degrees of freedom  
## AIC: 1630.7  
##   
## Number of Fisher Scoring iterations: 5

There are five significant interactions. Namely:

* log of free sulfur dioxide: log of total sulfur dioxide
* log of free sulfur dioxide: log of alcohol
* log of fixed acidity: square root of citric acid
* square root of citric acid: log of sulphates
* log of total sulfur dioxide: log of sulphates

# Model Selection on Interaction Model

We now use stepwise regression on the interaction model, fit4, to see if there is a model with interaction terms that fits better than fit3.

fit5=step(fit4, direction="both")

## Start: AIC=1630.71  
## quality ~ (logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol)^2  
##   
## Df Deviance AIC  
## - sqrtcitricacid:logalcohol 1 1572.7 1628.7  
## - sqrtcitricacid:logtotalsulfurdioxide 1 1572.8 1628.8  
## - logfreesulfurdioxide:logsulphates 1 1572.8 1628.8  
## - logfixedacidity:logtotalsulfurdioxide 1 1572.8 1628.8  
## - sqrtvolatileacidity:logalcohol 1 1573.0 1629.0  
## - logfreesulfurdioxide:sqrtcitricacid 1 1573.1 1629.1  
## - logfixedacidity:sqrtvolatileacidity 1 1573.2 1629.2  
## - logsulphates:sqrtvolatileacidity 1 1573.4 1629.4  
## - logsulphates:logalcohol 1 1573.5 1629.5  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1573.6 1629.6  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1574.4 1630.4  
## - logtotalsulfurdioxide:logalcohol 1 1574.5 1630.5  
## <none> 1572.7 1630.7  
## - logfixedacidity:logalcohol 1 1574.9 1630.9  
## - sqrtcitricacid:sqrtvolatileacidity 1 1575.0 1631.0  
## - logfreesulfurdioxide:logfixedacidity 1 1575.4 1631.4  
## - logfixedacidity:logsulphates 1 1576.2 1632.2  
## - logfreesulfurdioxide:logalcohol 1 1577.5 1633.5  
## - sqrtcitricacid:logsulphates 1 1577.6 1633.6  
## - logtotalsulfurdioxide:logsulphates 1 1577.8 1633.8  
## - logfixedacidity:sqrtcitricacid 1 1581.8 1637.8  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1588.0 1644.0  
##   
## Step: AIC=1628.73  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:sqrtcitricacid +   
## logfreesulfurdioxide:logtotalsulfurdioxide + logfreesulfurdioxide:logsulphates +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logtotalsulfurdioxide +   
## logfixedacidity:logsulphates + logfixedacidity:sqrtvolatileacidity +   
## logfixedacidity:logalcohol + sqrtcitricacid:logtotalsulfurdioxide +   
## sqrtcitricacid:logsulphates + sqrtcitricacid:sqrtvolatileacidity +   
## logtotalsulfurdioxide:logsulphates + logtotalsulfurdioxide:sqrtvolatileacidity +   
## logtotalsulfurdioxide:logalcohol + logsulphates:sqrtvolatileacidity +   
## logsulphates:logalcohol + sqrtvolatileacidity:logalcohol  
##   
## Df Deviance AIC  
## - sqrtcitricacid:logtotalsulfurdioxide 1 1572.8 1626.8  
## - logfreesulfurdioxide:logsulphates 1 1572.8 1626.8  
## - logfixedacidity:logtotalsulfurdioxide 1 1572.8 1626.8  
## - logfreesulfurdioxide:sqrtcitricacid 1 1573.1 1627.1  
## - sqrtvolatileacidity:logalcohol 1 1573.2 1627.2  
## - logfixedacidity:sqrtvolatileacidity 1 1573.3 1627.3  
## - logsulphates:sqrtvolatileacidity 1 1573.4 1627.4  
## - logsulphates:logalcohol 1 1573.6 1627.6  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1573.6 1627.6  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1574.4 1628.4  
## - logtotalsulfurdioxide:logalcohol 1 1574.7 1628.7  
## <none> 1572.7 1628.7  
## - sqrtcitricacid:sqrtvolatileacidity 1 1575.0 1629.0  
## - logfreesulfurdioxide:logfixedacidity 1 1575.4 1629.4  
## - logfixedacidity:logsulphates 1 1576.2 1630.2  
## - logfixedacidity:logalcohol 1 1576.3 1630.3  
## + sqrtcitricacid:logalcohol 1 1572.7 1630.7  
## - sqrtcitricacid:logsulphates 1 1577.7 1631.7  
## - logfreesulfurdioxide:logalcohol 1 1577.7 1631.7  
## - logtotalsulfurdioxide:logsulphates 1 1577.8 1631.8  
## - logfixedacidity:sqrtcitricacid 1 1582.3 1636.3  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1588.0 1642.0  
##   
## Step: AIC=1626.78  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:sqrtcitricacid +   
## logfreesulfurdioxide:logtotalsulfurdioxide + logfreesulfurdioxide:logsulphates +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logtotalsulfurdioxide +   
## logfixedacidity:logsulphates + logfixedacidity:sqrtvolatileacidity +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## sqrtcitricacid:sqrtvolatileacidity + logtotalsulfurdioxide:logsulphates +   
## logtotalsulfurdioxide:sqrtvolatileacidity + logtotalsulfurdioxide:logalcohol +   
## logsulphates:sqrtvolatileacidity + logsulphates:logalcohol +   
## sqrtvolatileacidity:logalcohol  
##   
## Df Deviance AIC  
## - logfixedacidity:logtotalsulfurdioxide 1 1572.8 1624.8  
## - logfreesulfurdioxide:logsulphates 1 1572.8 1624.8  
## - sqrtvolatileacidity:logalcohol 1 1573.3 1625.3  
## - logfreesulfurdioxide:sqrtcitricacid 1 1573.3 1625.3  
## - logfixedacidity:sqrtvolatileacidity 1 1573.3 1625.3  
## - logsulphates:sqrtvolatileacidity 1 1573.5 1625.5  
## - logsulphates:logalcohol 1 1573.7 1625.7  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1573.9 1625.9  
## - logtotalsulfurdioxide:logalcohol 1 1574.8 1626.8  
## <none> 1572.8 1626.8  
## - sqrtcitricacid:sqrtvolatileacidity 1 1575.1 1627.1  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1575.2 1627.2  
## - logfreesulfurdioxide:logfixedacidity 1 1575.6 1627.6  
## - logfixedacidity:logsulphates 1 1576.3 1628.3  
## - logfixedacidity:logalcohol 1 1576.3 1628.3  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1572.7 1628.7  
## + sqrtcitricacid:logalcohol 1 1572.8 1628.8  
## - logfreesulfurdioxide:logalcohol 1 1577.7 1629.7  
## - sqrtcitricacid:logsulphates 1 1577.8 1629.8  
## - logtotalsulfurdioxide:logsulphates 1 1578.0 1630.0  
## - logfixedacidity:sqrtcitricacid 1 1582.5 1634.5  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1588.1 1640.1  
##   
## Step: AIC=1624.81  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:sqrtcitricacid +   
## logfreesulfurdioxide:logtotalsulfurdioxide + logfreesulfurdioxide:logsulphates +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logsulphates +   
## logfixedacidity:sqrtvolatileacidity + logfixedacidity:logalcohol +   
## sqrtcitricacid:logsulphates + sqrtcitricacid:sqrtvolatileacidity +   
## logtotalsulfurdioxide:logsulphates + logtotalsulfurdioxide:sqrtvolatileacidity +   
## logtotalsulfurdioxide:logalcohol + logsulphates:sqrtvolatileacidity +   
## logsulphates:logalcohol + sqrtvolatileacidity:logalcohol  
##   
## Df Deviance AIC  
## - logfreesulfurdioxide:logsulphates 1 1572.9 1622.9  
## - sqrtvolatileacidity:logalcohol 1 1573.3 1623.3  
## - logfixedacidity:sqrtvolatileacidity 1 1573.3 1623.3  
## - logfreesulfurdioxide:sqrtcitricacid 1 1573.4 1623.4  
## - logsulphates:sqrtvolatileacidity 1 1573.5 1623.5  
## - logsulphates:logalcohol 1 1573.7 1623.7  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1573.9 1623.9  
## <none> 1572.8 1624.8  
## - logtotalsulfurdioxide:logalcohol 1 1575.1 1625.1  
## - sqrtcitricacid:sqrtvolatileacidity 1 1575.1 1625.1  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1575.2 1625.2  
## - logfixedacidity:logalcohol 1 1576.3 1626.3  
## - logfixedacidity:logsulphates 1 1576.4 1626.4  
## + logfixedacidity:logtotalsulfurdioxide 1 1572.8 1626.8  
## + sqrtcitricacid:logalcohol 1 1572.8 1626.8  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1572.8 1626.8  
## - logfreesulfurdioxide:logfixedacidity 1 1577.3 1627.3  
## - sqrtcitricacid:logsulphates 1 1577.8 1627.8  
## - logtotalsulfurdioxide:logsulphates 1 1578.0 1628.0  
## - logfreesulfurdioxide:logalcohol 1 1578.1 1628.1  
## - logfixedacidity:sqrtcitricacid 1 1582.6 1632.6  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1588.5 1638.5  
##   
## Step: AIC=1622.87  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:sqrtcitricacid +   
## logfreesulfurdioxide:logtotalsulfurdioxide + logfreesulfurdioxide:sqrtvolatileacidity +   
## logfreesulfurdioxide:logalcohol + logfixedacidity:sqrtcitricacid +   
## logfixedacidity:logsulphates + logfixedacidity:sqrtvolatileacidity +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## sqrtcitricacid:sqrtvolatileacidity + logtotalsulfurdioxide:logsulphates +   
## logtotalsulfurdioxide:sqrtvolatileacidity + logtotalsulfurdioxide:logalcohol +   
## logsulphates:sqrtvolatileacidity + logsulphates:logalcohol +   
## sqrtvolatileacidity:logalcohol  
##   
## Df Deviance AIC  
## - sqrtvolatileacidity:logalcohol 1 1573.3 1621.3  
## - logfixedacidity:sqrtvolatileacidity 1 1573.4 1621.4  
## - logfreesulfurdioxide:sqrtcitricacid 1 1573.4 1621.4  
## - logsulphates:sqrtvolatileacidity 1 1573.6 1621.6  
## - logsulphates:logalcohol 1 1573.8 1621.8  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1573.9 1621.9  
## <none> 1572.9 1622.9  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1575.2 1623.2  
## - logtotalsulfurdioxide:logalcohol 1 1575.2 1623.2  
## - sqrtcitricacid:sqrtvolatileacidity 1 1575.2 1623.2  
## - logfixedacidity:logalcohol 1 1576.4 1624.4  
## - logfixedacidity:logsulphates 1 1576.5 1624.5  
## + logfreesulfurdioxide:logsulphates 1 1572.8 1624.8  
## + logfixedacidity:logtotalsulfurdioxide 1 1572.8 1624.8  
## + sqrtcitricacid:logalcohol 1 1572.8 1624.8  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1572.9 1624.9  
## - logfreesulfurdioxide:logfixedacidity 1 1577.6 1625.6  
## - sqrtcitricacid:logsulphates 1 1577.9 1625.9  
## - logfreesulfurdioxide:logalcohol 1 1578.2 1626.2  
## - logfixedacidity:sqrtcitricacid 1 1582.7 1630.7  
## - logtotalsulfurdioxide:logsulphates 1 1585.6 1633.6  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1588.7 1636.7  
##   
## Step: AIC=1621.34  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:sqrtcitricacid +   
## logfreesulfurdioxide:logtotalsulfurdioxide + logfreesulfurdioxide:sqrtvolatileacidity +   
## logfreesulfurdioxide:logalcohol + logfixedacidity:sqrtcitricacid +   
## logfixedacidity:logsulphates + logfixedacidity:sqrtvolatileacidity +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## sqrtcitricacid:sqrtvolatileacidity + logtotalsulfurdioxide:logsulphates +   
## logtotalsulfurdioxide:sqrtvolatileacidity + logtotalsulfurdioxide:logalcohol +   
## logsulphates:sqrtvolatileacidity + logsulphates:logalcohol  
##   
## Df Deviance AIC  
## - logfixedacidity:sqrtvolatileacidity 1 1573.7 1619.7  
## - logfreesulfurdioxide:sqrtcitricacid 1 1573.9 1619.9  
## - logsulphates:sqrtvolatileacidity 1 1574.0 1620.0  
## - logsulphates:logalcohol 1 1574.3 1620.3  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1574.5 1620.5  
## <none> 1573.3 1621.3  
## - sqrtcitricacid:sqrtvolatileacidity 1 1575.5 1621.5  
## - logtotalsulfurdioxide:logalcohol 1 1575.6 1621.6  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1576.0 1622.0  
## - logfixedacidity:logalcohol 1 1576.5 1622.5  
## + sqrtvolatileacidity:logalcohol 1 1572.9 1622.9  
## - logfixedacidity:logsulphates 1 1577.0 1623.0  
## + sqrtcitricacid:logalcohol 1 1573.1 1623.1  
## + logfreesulfurdioxide:logsulphates 1 1573.3 1623.3  
## + logfixedacidity:logtotalsulfurdioxide 1 1573.3 1623.3  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1573.3 1623.3  
## - logfreesulfurdioxide:logfixedacidity 1 1577.9 1623.9  
## - sqrtcitricacid:logsulphates 1 1578.4 1624.4  
## - logfreesulfurdioxide:logalcohol 1 1578.6 1624.6  
## - logfixedacidity:sqrtcitricacid 1 1583.6 1629.6  
## - logtotalsulfurdioxide:logsulphates 1 1586.2 1632.2  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1588.9 1634.9  
##   
## Step: AIC=1619.72  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:sqrtcitricacid +   
## logfreesulfurdioxide:logtotalsulfurdioxide + logfreesulfurdioxide:sqrtvolatileacidity +   
## logfreesulfurdioxide:logalcohol + logfixedacidity:sqrtcitricacid +   
## logfixedacidity:logsulphates + logfixedacidity:logalcohol +   
## sqrtcitricacid:logsulphates + sqrtcitricacid:sqrtvolatileacidity +   
## logtotalsulfurdioxide:logsulphates + logtotalsulfurdioxide:sqrtvolatileacidity +   
## logtotalsulfurdioxide:logalcohol + logsulphates:sqrtvolatileacidity +   
## logsulphates:logalcohol  
##   
## Df Deviance AIC  
## - logfreesulfurdioxide:sqrtcitricacid 1 1574.4 1618.4  
## - logsulphates:sqrtvolatileacidity 1 1574.5 1618.5  
## - logsulphates:logalcohol 1 1574.7 1618.7  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1574.8 1618.8  
## - sqrtcitricacid:sqrtvolatileacidity 1 1575.5 1619.5  
## <none> 1573.7 1619.7  
## - logtotalsulfurdioxide:logalcohol 1 1576.0 1620.0  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1576.6 1620.6  
## - logfixedacidity:logalcohol 1 1577.1 1621.1  
## + logfixedacidity:sqrtvolatileacidity 1 1573.3 1621.3  
## + sqrtvolatileacidity:logalcohol 1 1573.4 1621.4  
## + sqrtcitricacid:logalcohol 1 1573.5 1621.5  
## - logfixedacidity:logsulphates 1 1577.5 1621.5  
## + logfreesulfurdioxide:logsulphates 1 1573.7 1621.7  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1573.7 1621.7  
## + logfixedacidity:logtotalsulfurdioxide 1 1573.7 1621.7  
## - logfreesulfurdioxide:logfixedacidity 1 1578.4 1622.4  
## - logfreesulfurdioxide:logalcohol 1 1579.1 1623.1  
## - sqrtcitricacid:logsulphates 1 1579.1 1623.1  
## - logfixedacidity:sqrtcitricacid 1 1584.1 1628.1  
## - logtotalsulfurdioxide:logsulphates 1 1586.3 1630.3  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1589.9 1633.9  
##   
## Step: AIC=1618.4  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:logtotalsulfurdioxide +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logsulphates +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## sqrtcitricacid:sqrtvolatileacidity + logtotalsulfurdioxide:logsulphates +   
## logtotalsulfurdioxide:sqrtvolatileacidity + logtotalsulfurdioxide:logalcohol +   
## logsulphates:sqrtvolatileacidity + logsulphates:logalcohol  
##   
## Df Deviance AIC  
## - logsulphates:sqrtvolatileacidity 1 1575.3 1617.3  
## - logsulphates:logalcohol 1 1575.4 1617.4  
## - sqrtcitricacid:sqrtvolatileacidity 1 1576.2 1618.2  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1576.3 1618.3  
## <none> 1574.4 1618.4  
## - logtotalsulfurdioxide:logalcohol 1 1576.5 1618.5  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1577.1 1619.1  
## + logfreesulfurdioxide:sqrtcitricacid 1 1573.7 1619.7  
## + logfixedacidity:sqrtvolatileacidity 1 1573.9 1619.9  
## - logfixedacidity:logalcohol 1 1577.9 1619.9  
## - logfixedacidity:logsulphates 1 1578.1 1620.1  
## + sqrtvolatileacidity:logalcohol 1 1574.1 1620.1  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1574.1 1620.1  
## + sqrtcitricacid:logalcohol 1 1574.2 1620.2  
## + logfreesulfurdioxide:logsulphates 1 1574.4 1620.4  
## + logfixedacidity:logtotalsulfurdioxide 1 1574.4 1620.4  
## - logfreesulfurdioxide:logfixedacidity 1 1578.9 1620.9  
## - sqrtcitricacid:logsulphates 1 1579.5 1621.5  
## - logfreesulfurdioxide:logalcohol 1 1579.8 1621.8  
## - logfixedacidity:sqrtcitricacid 1 1584.9 1626.9  
## - logtotalsulfurdioxide:logsulphates 1 1587.0 1629.0  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1589.9 1631.9  
##   
## Step: AIC=1617.28  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:logtotalsulfurdioxide +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logsulphates +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## sqrtcitricacid:sqrtvolatileacidity + logtotalsulfurdioxide:logsulphates +   
## logtotalsulfurdioxide:sqrtvolatileacidity + logtotalsulfurdioxide:logalcohol +   
## logsulphates:logalcohol  
##   
## Df Deviance AIC  
## - logsulphates:logalcohol 1 1576.5 1616.5  
## - sqrtcitricacid:sqrtvolatileacidity 1 1576.7 1616.7  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1577.0 1617.0  
## <none> 1575.3 1617.3  
## - logtotalsulfurdioxide:logalcohol 1 1577.6 1617.6  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1577.8 1617.8  
## + logsulphates:sqrtvolatileacidity 1 1574.4 1618.4  
## + logfreesulfurdioxide:sqrtcitricacid 1 1574.5 1618.5  
## + logfixedacidity:sqrtvolatileacidity 1 1574.6 1618.6  
## - logfixedacidity:logsulphates 1 1578.6 1618.6  
## - logfixedacidity:logalcohol 1 1579.0 1619.0  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1575.0 1619.0  
## + sqrtvolatileacidity:logalcohol 1 1575.0 1619.0  
## + sqrtcitricacid:logalcohol 1 1575.1 1619.1  
## + logfixedacidity:logtotalsulfurdioxide 1 1575.2 1619.2  
## + logfreesulfurdioxide:logsulphates 1 1575.3 1619.3  
## - sqrtcitricacid:logsulphates 1 1579.5 1619.5  
## - logfreesulfurdioxide:logfixedacidity 1 1579.7 1619.7  
## - logfreesulfurdioxide:logalcohol 1 1581.0 1621.0  
## - logfixedacidity:sqrtcitricacid 1 1586.3 1626.3  
## - logtotalsulfurdioxide:logsulphates 1 1588.7 1628.7  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1590.7 1630.7  
##   
## Step: AIC=1616.53  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:logtotalsulfurdioxide +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logsulphates +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## sqrtcitricacid:sqrtvolatileacidity + logtotalsulfurdioxide:logsulphates +   
## logtotalsulfurdioxide:sqrtvolatileacidity + logtotalsulfurdioxide:logalcohol  
##   
## Df Deviance AIC  
## - sqrtcitricacid:sqrtvolatileacidity 1 1578.1 1616.1  
## - logtotalsulfurdioxide:logalcohol 1 1578.5 1616.5  
## <none> 1576.5 1616.5  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1578.6 1616.6  
## + logsulphates:logalcohol 1 1575.3 1617.3  
## + logsulphates:sqrtvolatileacidity 1 1575.4 1617.4  
## - logfixedacidity:logsulphates 1 1579.5 1617.5  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1579.5 1617.5  
## - logfixedacidity:logalcohol 1 1579.7 1617.7  
## + logfreesulfurdioxide:sqrtcitricacid 1 1575.8 1617.8  
## + logfixedacidity:sqrtvolatileacidity 1 1575.8 1617.8  
## + sqrtcitricacid:logalcohol 1 1576.2 1618.2  
## + sqrtvolatileacidity:logalcohol 1 1576.2 1618.2  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1576.3 1618.3  
## + logfixedacidity:logtotalsulfurdioxide 1 1576.5 1618.5  
## + logfreesulfurdioxide:logsulphates 1 1576.5 1618.5  
## - logfreesulfurdioxide:logfixedacidity 1 1580.8 1618.8  
## - sqrtcitricacid:logsulphates 1 1581.1 1619.1  
## - logfreesulfurdioxide:logalcohol 1 1582.1 1620.1  
## - logfixedacidity:sqrtcitricacid 1 1587.0 1625.0  
## - logtotalsulfurdioxide:logsulphates 1 1591.6 1629.6  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1592.0 1630.0  
##   
## Step: AIC=1616.1  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:logtotalsulfurdioxide +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logsulphates +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## logtotalsulfurdioxide:logsulphates + logtotalsulfurdioxide:sqrtvolatileacidity +   
## logtotalsulfurdioxide:logalcohol  
##   
## Df Deviance AIC  
## - logtotalsulfurdioxide:logalcohol 1 1580.0 1616.0  
## <none> 1578.1 1616.1  
## + sqrtcitricacid:sqrtvolatileacidity 1 1576.5 1616.5  
## + logsulphates:logalcohol 1 1576.7 1616.7  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1580.8 1616.8  
## - logfixedacidity:logalcohol 1 1581.0 1617.0  
## - logfixedacidity:logsulphates 1 1581.2 1617.2  
## + logfreesulfurdioxide:sqrtcitricacid 1 1577.3 1617.3  
## + logsulphates:sqrtvolatileacidity 1 1577.5 1617.5  
## + sqrtvolatileacidity:logalcohol 1 1577.8 1617.8  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1577.9 1617.9  
## + sqrtcitricacid:logalcohol 1 1577.9 1617.9  
## + logfixedacidity:logtotalsulfurdioxide 1 1578.0 1618.0  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1582.1 1618.1  
## + logfreesulfurdioxide:logsulphates 1 1578.1 1618.1  
## + logfixedacidity:sqrtvolatileacidity 1 1578.1 1618.1  
## - logfreesulfurdioxide:logfixedacidity 1 1582.3 1618.3  
## - sqrtcitricacid:logsulphates 1 1583.4 1619.4  
## - logfreesulfurdioxide:logalcohol 1 1583.6 1619.6  
## - logfixedacidity:sqrtcitricacid 1 1589.6 1625.6  
## - logtotalsulfurdioxide:logsulphates 1 1592.7 1628.7  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1594.7 1630.7  
##   
## Step: AIC=1616  
## quality ~ logfreesulfurdioxide + logfixedacidity + sqrtcitricacid +   
## logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:logtotalsulfurdioxide +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logsulphates +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## logtotalsulfurdioxide:logsulphates + logtotalsulfurdioxide:sqrtvolatileacidity  
##   
## Df Deviance AIC  
## <none> 1580.0 1616.0  
## + logtotalsulfurdioxide:logalcohol 1 1578.1 1616.1  
## + sqrtcitricacid:sqrtvolatileacidity 1 1578.5 1616.5  
## - logfixedacidity:logalcohol 1 1582.6 1616.6  
## - logfreesulfurdioxide:sqrtvolatileacidity 1 1582.8 1616.8  
## + logsulphates:logalcohol 1 1579.0 1617.0  
## - logfixedacidity:logsulphates 1 1583.0 1617.0  
## + logsulphates:sqrtvolatileacidity 1 1579.3 1617.3  
## + logfreesulfurdioxide:sqrtcitricacid 1 1579.3 1617.3  
## + logfixedacidity:logtotalsulfurdioxide 1 1579.6 1617.6  
## + sqrtcitricacid:logtotalsulfurdioxide 1 1579.8 1617.8  
## + sqrtvolatileacidity:logalcohol 1 1579.8 1617.8  
## - logfreesulfurdioxide:logfixedacidity 1 1583.8 1617.8  
## + logfreesulfurdioxide:logsulphates 1 1580.0 1618.0  
## + logfixedacidity:sqrtvolatileacidity 1 1580.0 1618.0  
## + sqrtcitricacid:logalcohol 1 1580.0 1618.0  
## - logfreesulfurdioxide:logalcohol 1 1584.1 1618.1  
## - logtotalsulfurdioxide:sqrtvolatileacidity 1 1584.5 1618.5  
## - sqrtcitricacid:logsulphates 1 1585.0 1619.0  
## - logfixedacidity:sqrtcitricacid 1 1592.0 1626.0  
## - logtotalsulfurdioxide:logsulphates 1 1594.0 1628.0  
## - logfreesulfurdioxide:logtotalsulfurdioxide 1 1596.6 1630.6

summary(fit5)

##   
## Call:  
## glm(formula = quality ~ logfreesulfurdioxide + logfixedacidity +   
## sqrtcitricacid + logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:logtotalsulfurdioxide +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logsulphates +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## logtotalsulfurdioxide:logsulphates + logtotalsulfurdioxide:sqrtvolatileacidity,   
## family = "binomial")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5643 -0.8000 0.2575 0.7704 2.5149   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.6546 0.0977 6.701 2.07e-11  
## logfreesulfurdioxide 0.5121 0.1521 3.368 0.000757  
## logfixedacidity 1.8466 0.4704 3.925 8.67e-05  
## sqrtcitricacid -1.7743 0.4627 -3.835 0.000126  
## logtotalsulfurdioxide -0.9154 0.1590 -5.757 8.54e-09  
## logsulphates 2.5583 0.3363 7.608 2.78e-14  
## sqrtvolatileacidity -4.9766 0.7175 -6.936 4.04e-12  
## logalcohol 10.4915 0.8160 12.857 < 2e-16  
## logfreesulfurdioxide:logfixedacidity -0.9829 0.5006 -1.964 0.049565  
## logfreesulfurdioxide:logtotalsulfurdioxide -0.5702 0.1406 -4.056 5.00e-05  
## logfreesulfurdioxide:sqrtvolatileacidity -2.1510 1.2626 -1.704 0.088445  
## logfreesulfurdioxide:logalcohol 2.2392 1.1059 2.025 0.042888  
## logfixedacidity:sqrtcitricacid -5.8500 1.6811 -3.480 0.000502  
## logfixedacidity:logsulphates 3.7196 2.1485 1.731 0.083408  
## logfixedacidity:logalcohol -5.3576 3.3417 -1.603 0.108876  
## sqrtcitricacid:logsulphates -3.7213 1.7090 -2.177 0.029449  
## logtotalsulfurdioxide:logsulphates -1.6151 0.4474 -3.610 0.000306  
## logtotalsulfurdioxide:sqrtvolatileacidity 2.7828 1.3028 2.136 0.032676  
##   
## (Intercept) \*\*\*  
## logfreesulfurdioxide \*\*\*  
## logfixedacidity \*\*\*  
## sqrtcitricacid \*\*\*  
## logtotalsulfurdioxide \*\*\*  
## logsulphates \*\*\*  
## sqrtvolatileacidity \*\*\*  
## logalcohol \*\*\*  
## logfreesulfurdioxide:logfixedacidity \*   
## logfreesulfurdioxide:logtotalsulfurdioxide \*\*\*  
## logfreesulfurdioxide:sqrtvolatileacidity .   
## logfreesulfurdioxide:logalcohol \*   
## logfixedacidity:sqrtcitricacid \*\*\*  
## logfixedacidity:logsulphates .   
## logfixedacidity:logalcohol   
## sqrtcitricacid:logsulphates \*   
## logtotalsulfurdioxide:logsulphates \*\*\*  
## logtotalsulfurdioxide:sqrtvolatileacidity \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2209 on 1598 degrees of freedom  
## Residual deviance: 1580 on 1581 degrees of freedom  
## AIC: 1616  
##   
## Number of Fisher Scoring iterations: 5

Using stepwise regression, 10 interaction effects were significant. These are:

* log of fixed acidity:log of alcohol
* log of free sulfur dioxide:square root of volatile acidity
* log of fixed acidity:log of sulphates
* log of free sulfur dioxide:log of fixed acidity
* log of free sulfur dioxide:log of alcohol
* log of total sulfur dioxide:square root of volatile acidity
* square root of citric acid:log of sulphates
* log of fixed acidity:square root of citric acid
* log of total sulfur dioxide:log of sulphates
* log of free sulfur dioxide:log of total sulfur dioxide

We will therefore use these interactions for our final model as well as the significant predictors from the previous model in our final model.

Now, we will plot two of the interaction effects and what these interactions mean.

categorize = function (x, quantiles=(1:3)/4) {  
 cutoffs = quantile (x, quantiles)  
 n.cutoffs = length (cutoffs)  
 result = rep ("C1", length (x))  
 for (j in 1:n.cutoffs) {  
 result [x > cutoffs [j]] = paste ("C", j+1, sep="")  
 }  
 return (result)  
}  
library (ggplot2)  
library (dplyr)

##   
## Attaching package: 'dplyr'

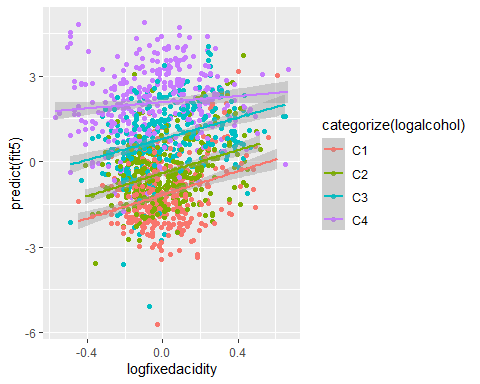
## The following object is masked from 'package:MASS':  
##   
## select

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

with (wine.data, qplot (x=logfixedacidity, y=predict(fit5), color=categorize (logalcohol)) +  
geom\_smooth (method="lm"))

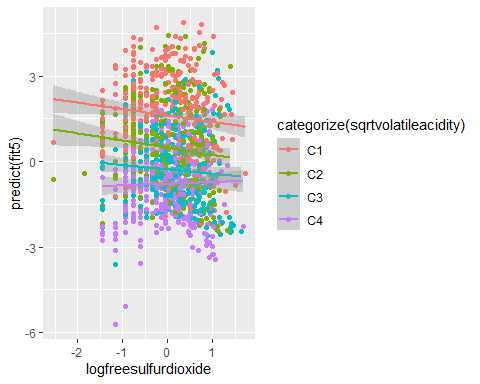
## `geom\_smooth()` using formula 'y ~ x'



From the interaction plot of log fixed acidity and log of alcohol we can see a negative linear relationship between the two. The interaction between acidity and alcohol is negative – as alcohol increases, the slope between quality and acidity decreases.

with (wine.data, qplot (x=logfreesulfurdioxide, y=predict(fit5), color=categorize (sqrtvolatileacidity)) +  
geom\_smooth (method="lm"))

## `geom\_smooth()` using formula 'y ~ x'



Similar comment as above: As volatile acidity increases (C1 to C4), the slope between quality and free sulfur dioxide increases (The C4 slope is positive, while the others are negative and about the same).

# Final Model

summary(fit5)

##   
## Call:  
## glm(formula = quality ~ logfreesulfurdioxide + logfixedacidity +   
## sqrtcitricacid + logtotalsulfurdioxide + logsulphates + sqrtvolatileacidity +   
## logalcohol + logfreesulfurdioxide:logfixedacidity + logfreesulfurdioxide:logtotalsulfurdioxide +   
## logfreesulfurdioxide:sqrtvolatileacidity + logfreesulfurdioxide:logalcohol +   
## logfixedacidity:sqrtcitricacid + logfixedacidity:logsulphates +   
## logfixedacidity:logalcohol + sqrtcitricacid:logsulphates +   
## logtotalsulfurdioxide:logsulphates + logtotalsulfurdioxide:sqrtvolatileacidity,   
## family = "binomial")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5643 -0.8000 0.2575 0.7704 2.5149   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.6546 0.0977 6.701 2.07e-11  
## logfreesulfurdioxide 0.5121 0.1521 3.368 0.000757  
## logfixedacidity 1.8466 0.4704 3.925 8.67e-05  
## sqrtcitricacid -1.7743 0.4627 -3.835 0.000126  
## logtotalsulfurdioxide -0.9154 0.1590 -5.757 8.54e-09  
## logsulphates 2.5583 0.3363 7.608 2.78e-14  
## sqrtvolatileacidity -4.9766 0.7175 -6.936 4.04e-12  
## logalcohol 10.4915 0.8160 12.857 < 2e-16  
## logfreesulfurdioxide:logfixedacidity -0.9829 0.5006 -1.964 0.049565  
## logfreesulfurdioxide:logtotalsulfurdioxide -0.5702 0.1406 -4.056 5.00e-05  
## logfreesulfurdioxide:sqrtvolatileacidity -2.1510 1.2626 -1.704 0.088445  
## logfreesulfurdioxide:logalcohol 2.2392 1.1059 2.025 0.042888  
## logfixedacidity:sqrtcitricacid -5.8500 1.6811 -3.480 0.000502  
## logfixedacidity:logsulphates 3.7196 2.1485 1.731 0.083408  
## logfixedacidity:logalcohol -5.3576 3.3417 -1.603 0.108876  
## sqrtcitricacid:logsulphates -3.7213 1.7090 -2.177 0.029449  
## logtotalsulfurdioxide:logsulphates -1.6151 0.4474 -3.610 0.000306  
## logtotalsulfurdioxide:sqrtvolatileacidity 2.7828 1.3028 2.136 0.032676  
##   
## (Intercept) \*\*\*  
## logfreesulfurdioxide \*\*\*  
## logfixedacidity \*\*\*  
## sqrtcitricacid \*\*\*  
## logtotalsulfurdioxide \*\*\*  
## logsulphates \*\*\*  
## sqrtvolatileacidity \*\*\*  
## logalcohol \*\*\*  
## logfreesulfurdioxide:logfixedacidity \*   
## logfreesulfurdioxide:logtotalsulfurdioxide \*\*\*  
## logfreesulfurdioxide:sqrtvolatileacidity .   
## logfreesulfurdioxide:logalcohol \*   
## logfixedacidity:sqrtcitricacid \*\*\*  
## logfixedacidity:logsulphates .   
## logfixedacidity:logalcohol   
## sqrtcitricacid:logsulphates \*   
## logtotalsulfurdioxide:logsulphates \*\*\*  
## logtotalsulfurdioxide:sqrtvolatileacidity \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2209 on 1598 degrees of freedom  
## Residual deviance: 1580 on 1581 degrees of freedom  
## AIC: 1616  
##   
## Number of Fisher Scoring iterations: 5

# Odds ratios  
#logfreesulfurdioxide  
exp (0.5121)

## [1] 1.668792

#logfixedacidity   
exp (1.8466)

## [1] 6.338233

#sqrtcitricacid   
exp (-1.7743)

## [1] 0.1696021

#logtotalsulfurdioxide  
exp (-0.9154)

## [1] 0.4003565

#logsulphates   
exp (2.5583)

## [1] 12.91385

#sqrtvolatileacidity  
exp (-4.9766)

## [1] 0.006897474

#logalcohol  
exp (10.4915)

## [1] 36008.13

exp(confint(fit5))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 1.591835e+00 2.335164e+00  
## logfreesulfurdioxide 1.239181e+00 2.250999e+00  
## logfixedacidity 2.532729e+00 1.603692e+01  
## sqrtcitricacid 6.800629e-02 4.177278e-01  
## logtotalsulfurdioxide 2.925968e-01 5.460812e-01  
## logsulphates 6.739987e+00 2.521968e+01  
## sqrtvolatileacidity 1.659551e-03 2.770139e-02  
## logalcohol 7.475516e+03 1.835434e+05  
## logfreesulfurdioxide:logfixedacidity 1.400157e-01 9.983727e-01  
## logfreesulfurdioxide:logtotalsulfurdioxide 4.284881e-01 7.440325e-01  
## logfreesulfurdioxide:sqrtvolatileacidity 9.927055e-03 1.443794e+00  
## logfreesulfurdioxide:logalcohol 1.081704e+00 8.297215e+01  
## logfixedacidity:sqrtcitricacid 1.062165e-04 7.783129e-02  
## logfixedacidity:logsulphates 6.349520e-01 2.914841e+03  
## logfixedacidity:logalcohol 6.895990e-06 3.344696e+00  
## sqrtcitricacid:logsulphates 7.808907e-04 6.395000e-01  
## logtotalsulfurdioxide:logsulphates 8.042046e-02 4.682441e-01  
## logtotalsulfurdioxide:sqrtvolatileacidity 1.232912e+00 2.070798e+02

All the predictors in our final model are significant, the AIC is 1616 while the previous model (fit4) had an AIC of 1679.6. The parameter estimates, standard errors, and p-values are as follows:

Log free sulfur dioxide has an estimate of 0.5121, standard error of 0.1521 and a p-value of 0.000757

Log fixed acidity has an estimate of 1.8466, standard error of 0.4704, and a p-value of 8.67e-05

Square root of citric acid has an estimate of -1.7743, a standard error of 0.4627, and a p-value of 0.000126

Log total sulfur dioxide has an estimate of -0.9154, a standard error of 0.1590, and a p-value of 8.54e-09

Log sulphates has an estimate of 2.5583, an estimate of 0.3363, and a p-value of 2.78e-14

Square root of volatile acidity has an estimate of -4.9766, a standard error of 0.7175, and a p-value of 4.04e-12 log of alcohol has an estimate of 10.4915, a standard error of 0.8160, and a p-value of < 2e-16

Next we will investigate some model diagnostics.

## Model Diagnostics

## Likelihood ratio test for final model and Goodness of fit test

# Likelihood ratio test for final model  
pchisq(fit5$null.deviance - fit5$deviance, fit5$df.null - fit5$df.residual, lower=F)

## [1] 1.03385e-122

# Goodness of fit test  
pchisq (deviance(fit5), df.residual(fit5), lower=F)

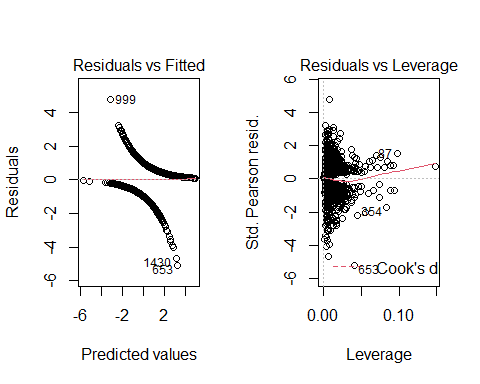
## [1] 0.5024012

The likelihood ratio test has a p-value of 1.034e-122, this means that we have significant predictors.

The goodness of fit test has a p-value of 0.5024012 which is very high and greater that 0.05 thus we can conclude that the model likely fits.

## Residual Analysis & Diagnostic Analysis

par(mfrow = c(1,2))   
plot(fit5, which=c(1,5))

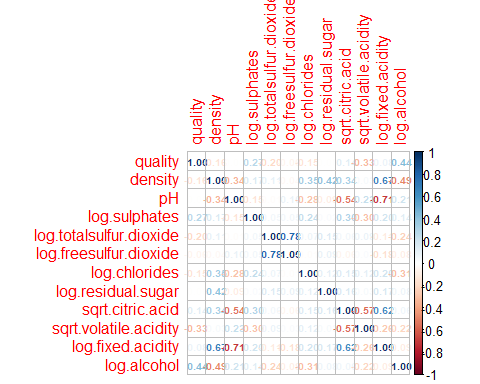


In the residual analysis vs fitted analysis has the smoothing line at 0 thus showing good linearity, points 1430 and 653 deviate a bit far from the smoothing line but still follow the curve.

The residuals vs leverage shows point 653 has large negative residual, otherwise the residuals are generally clustered to the left with one point at the right making the line go upwards a bit.

## Scatter Plot for Final Model

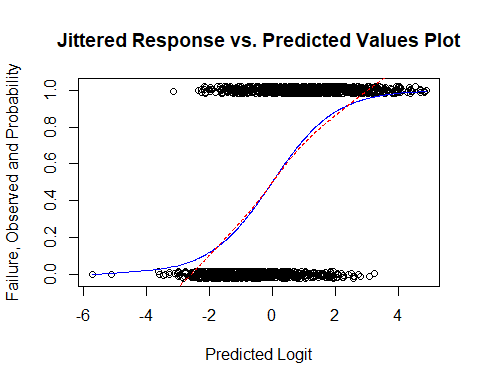
transcorr = cor(wine.data [, c(1, 9:10, 13:21)], use= "complete.obs")  
library("corrplot")  
## corrplot 0.84 loaded  
corrplot (transcorr, method = "number", number.cex=0.6)



The high leverage values do not seem too influential.

## Model Plot

pred.logit = predict (fit5)  
plot(jitter(quality,0.1) ~ pred.logit, data = wine.data, main = 'Jittered Response vs. Predicted Values Plot', xlab="Predicted Logit", ylab="Failure, Observed and Probability")  
pred.r = predict(fit5,type='response')  
pred.order =order(pred.logit)  
lines(pred.logit[pred.order],pred.r[pred.order],col='blue')  
lines(lowess(pred.logit[pred.order], pred.r[pred.order]),col='red',lty=2)



The final model plot shows that there is a concentration of “low quality” responses when the predicted logit value of the model is low and a concentration of “high quality” responses when the logit value is high. This shows the model has an ability to predict better than the first one.

## Influence Analysis

levcut = 2\*(sqrt(4/1599))   
levcut

## [1] 0.1000313

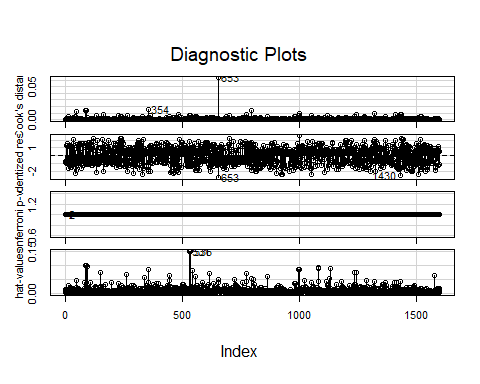
library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

influenceIndexPlot(fit5)



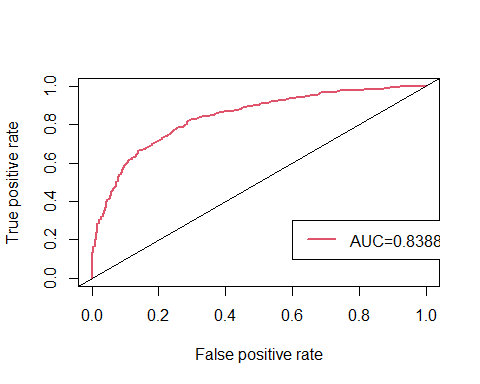
Our hat matrix leverage cutoff (2\*sqrt(p/n)) is 0.1000313 The diagnostic plot shows that only one case exceeds to this cutoff value, as case 3 has the highest leverage at roughly 0.15, this was seen in the cooks distance with one value indicating to have high cooks distance.

## ROC Curve

par (mfrow=c(1,1))   
library(ROCR)  
pred1 <- prediction(fit5$fitted.values, fit5$y)  
perf1 <- performance(pred1,"tpr","fpr")  
auc1 <- performance(pred1,"auc")@y.values[[1]]  
auc1

## [1] 0.8388292

plot(perf1, lwd=2, col=2)  
abline(0,1)  
legend(0.6, 0.3, c(paste ("AUC=", round (auc1, 4), sep="")), lwd=2, col=2)



roc.x = slot (perf1, "x.values") [[1]]  
roc.y = slot (perf1, "y.values") [[1]]  
cutoffs = slot (perf1, "alpha.values") [[1]]  
auc.table = cbind.data.frame(cutoff=pred1@cutoffs,  
tp=pred1@tp, fp=pred1@fp, tn=pred1@tn,  
fn=pred1@fn)  
names (auc.table) = c("Cutoff", "TP", "FP", "TN", "FN")  
auc.table$sensitivity = auc.table$TP / (auc.table$TP + auc.table$FN)  
auc.table$specificity = auc.table$TN / (auc.table$TN + auc.table$FP)  
auc.table$FalsePosRate = 1 - auc.table$specificity  
auc.table$sens\_spec = auc.table$sensitivity + auc.table$specificity  
auc.best = auc.table [auc.table$sens\_spec == max (auc.table$sens\_spec),]  
auc.best

## Cutoff TP FP TN FN sensitivity specificity FalsePosRate sens\_spec  
## 173 0.4408939 703 215 529 152 0.8222222 0.7110215 0.2889785 1.533244

We got an AUC of 0.84 thus the model correctly predicted 84% of the data with a specificity of 0.71 and sensitivity of 0.82. The cutoff is 0.44 with a false positive rate of 0.29. This makes us confident this model is our best model to use.

## Conclusions

Conclusively, having gone through all the process and the necessary steps to find our best model that is reliably predictable, our final model is “fit5”. All seven variables in the model are significant, thus, based on all the processes, these predictors, build our model to predict quality of wine.

Additionally, using the stepwise regression on the centered interaction effect between multiple variables, the model shows that the following had significant interactions:

* log of free sulfur dioxide: log of total sulfur dioxide
* log of free sulfur dioxide: log of alcohol
* log of fixed acidity: square root of citric acid
* square root of citric acid: log of sulphates
* log of total sulfur dioxide: log of sulphates

We also got an AUC of 84% prediction for quality of wine explained by the predictors of our final model.

With all these results we found that our model would be significant in predicting the quality of wine. Log of Alcohol was the most significant predictor in the final model so interestingly, wines with higher alcohol contents have higher quality. Furthermore, residual sugar was one of the first predictors removed using stepwise regression. It seems that residual sugar has no significance towards the quality of wine.

The model from a physiochemical lens seems to give conclusive evidence to what makes a good quality wine. However, if we had other factors considered such as types of grapes used, how long the wine was aged for, location (considering the effects of weather), etc (Which were not provided in the dataset). A better model could have been produced and we would have been able to examine the impact of these aforementioned factors to the quality of Vinho Verde wine.