Assignment\_3\_Solutions.R

Ashok Bhowmick

Sat Jul 29 20:43:44 2017

## CMTH 642 Assignment 3   
## Ashok Bhowmick, ID 500861640  
  
setwd("C:/Data Science/Ryerson Courese CMTH 642/Assignment")  
whitewine <- read.table("winequality-white.csv", header = TRUE, sep = ";")  
View(whitewine)  
  
## Question No. 1  
## Data Characteristics  
head(whitewine)

## fixed.acidity volatile.acidity citric.acid residual.sugar chlorides  
## 1 7.0 0.27 0.36 20.7 0.045  
## 2 6.3 0.30 0.34 1.6 0.049  
## 3 8.1 0.28 0.40 6.9 0.050  
## 4 7.2 0.23 0.32 8.5 0.058  
## 5 7.2 0.23 0.32 8.5 0.058  
## 6 8.1 0.28 0.40 6.9 0.050  
## free.sulfur.dioxide total.sulfur.dioxide density pH sulphates alcohol  
## 1 45 170 1.0010 3.00 0.45 8.8  
## 2 14 132 0.9940 3.30 0.49 9.5  
## 3 30 97 0.9951 3.26 0.44 10.1  
## 4 47 186 0.9956 3.19 0.40 9.9  
## 5 47 186 0.9956 3.19 0.40 9.9  
## 6 30 97 0.9951 3.26 0.44 10.1  
## quality  
## 1 6  
## 2 6  
## 3 6  
## 4 6  
## 5 6  
## 6 6

str(whitewine)

## 'data.frame': 4898 obs. of 12 variables:  
## $ fixed.acidity : num 7 6.3 8.1 7.2 7.2 8.1 6.2 7 6.3 8.1 ...  
## $ volatile.acidity : num 0.27 0.3 0.28 0.23 0.23 0.28 0.32 0.27 0.3 0.22 ...  
## $ citric.acid : num 0.36 0.34 0.4 0.32 0.32 0.4 0.16 0.36 0.34 0.43 ...  
## $ residual.sugar : num 20.7 1.6 6.9 8.5 8.5 6.9 7 20.7 1.6 1.5 ...  
## $ chlorides : num 0.045 0.049 0.05 0.058 0.058 0.05 0.045 0.045 0.049 0.044 ...  
## $ free.sulfur.dioxide : num 45 14 30 47 47 30 30 45 14 28 ...  
## $ total.sulfur.dioxide: num 170 132 97 186 186 97 136 170 132 129 ...  
## $ density : num 1.001 0.994 0.995 0.996 0.996 ...  
## $ pH : num 3 3.3 3.26 3.19 3.19 3.26 3.18 3 3.3 3.22 ...  
## $ sulphates : num 0.45 0.49 0.44 0.4 0.4 0.44 0.47 0.45 0.49 0.45 ...  
## $ alcohol : num 8.8 9.5 10.1 9.9 9.9 10.1 9.6 8.8 9.5 11 ...  
## $ quality : int 6 6 6 6 6 6 6 6 6 6 ...

summary(whitewine)

## fixed.acidity volatile.acidity citric.acid residual.sugar   
## Min. : 3.800 Min. :0.0800 Min. :0.0000 Min. : 0.600   
## 1st Qu.: 6.300 1st Qu.:0.2100 1st Qu.:0.2700 1st Qu.: 1.700   
## Median : 6.800 Median :0.2600 Median :0.3200 Median : 5.200   
## Mean : 6.855 Mean :0.2782 Mean :0.3342 Mean : 6.391   
## 3rd Qu.: 7.300 3rd Qu.:0.3200 3rd Qu.:0.3900 3rd Qu.: 9.900   
## Max. :14.200 Max. :1.1000 Max. :1.6600 Max. :65.800   
## chlorides free.sulfur.dioxide total.sulfur.dioxide  
## Min. :0.00900 Min. : 2.00 Min. : 9.0   
## 1st Qu.:0.03600 1st Qu.: 23.00 1st Qu.:108.0   
## Median :0.04300 Median : 34.00 Median :134.0   
## Mean :0.04577 Mean : 35.31 Mean :138.4   
## 3rd Qu.:0.05000 3rd Qu.: 46.00 3rd Qu.:167.0   
## Max. :0.34600 Max. :289.00 Max. :440.0   
## density pH sulphates alcohol   
## Min. :0.9871 Min. :2.720 Min. :0.2200 Min. : 8.00   
## 1st Qu.:0.9917 1st Qu.:3.090 1st Qu.:0.4100 1st Qu.: 9.50   
## Median :0.9937 Median :3.180 Median :0.4700 Median :10.40   
## Mean :0.9940 Mean :3.188 Mean :0.4898 Mean :10.51   
## 3rd Qu.:0.9961 3rd Qu.:3.280 3rd Qu.:0.5500 3rd Qu.:11.40   
## Max. :1.0390 Max. :3.820 Max. :1.0800 Max. :14.20   
## quality   
## Min. :3.000   
## 1st Qu.:5.000   
## Median :6.000   
## Mean :5.878   
## 3rd Qu.:6.000   
## Max. :9.000

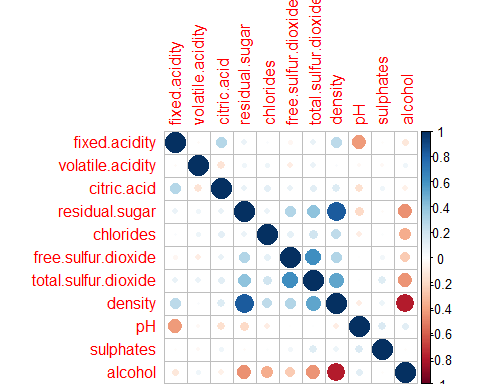
## Check for missing values  
sum(is.na(whitewine == TRUE))

## [1] 0

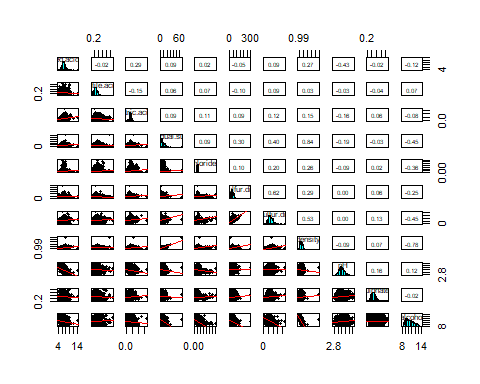
## Clean Data ready for analysis  
wine <- whitewine[, c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11)]  
##  
  
## Question No. 2  
library(corrplot)  
library(psych)  
cor(wine)

## fixed.acidity volatile.acidity citric.acid  
## fixed.acidity 1.00000000 -0.02269729 0.28918070  
## volatile.acidity -0.02269729 1.00000000 -0.14947181  
## citric.acid 0.28918070 -0.14947181 1.00000000  
## residual.sugar 0.08902070 0.06428606 0.09421162  
## chlorides 0.02308564 0.07051157 0.11436445  
## free.sulfur.dioxide -0.04939586 -0.09701194 0.09407722  
## total.sulfur.dioxide 0.09106976 0.08926050 0.12113080  
## density 0.26533101 0.02711385 0.14950257  
## pH -0.42585829 -0.03191537 -0.16374821  
## sulphates -0.01714299 -0.03572815 0.06233094  
## alcohol -0.12088112 0.06771794 -0.07572873  
## residual.sugar chlorides free.sulfur.dioxide  
## fixed.acidity 0.08902070 0.02308564 -0.0493958591  
## volatile.acidity 0.06428606 0.07051157 -0.0970119393  
## citric.acid 0.09421162 0.11436445 0.0940772210  
## residual.sugar 1.00000000 0.08868454 0.2990983537  
## chlorides 0.08868454 1.00000000 0.1013923521  
## free.sulfur.dioxide 0.29909835 0.10139235 1.0000000000  
## total.sulfur.dioxide 0.40143931 0.19891030 0.6155009650  
## density 0.83896645 0.25721132 0.2942104109  
## pH -0.19413345 -0.09043946 -0.0006177961  
## sulphates -0.02666437 0.01676288 0.0592172458  
## alcohol -0.45063122 -0.36018871 -0.2501039415  
## total.sulfur.dioxide density pH  
## fixed.acidity 0.091069756 0.26533101 -0.4258582910  
## volatile.acidity 0.089260504 0.02711385 -0.0319153683  
## citric.acid 0.121130798 0.14950257 -0.1637482114  
## residual.sugar 0.401439311 0.83896645 -0.1941334540  
## chlorides 0.198910300 0.25721132 -0.0904394560  
## free.sulfur.dioxide 0.615500965 0.29421041 -0.0006177961  
## total.sulfur.dioxide 1.000000000 0.52988132 0.0023209718  
## density 0.529881324 1.00000000 -0.0935914935  
## pH 0.002320972 -0.09359149 1.0000000000  
## sulphates 0.134562367 0.07449315 0.1559514973  
## alcohol -0.448892102 -0.78013762 0.1214320987  
## sulphates alcohol  
## fixed.acidity -0.01714299 -0.12088112  
## volatile.acidity -0.03572815 0.06771794  
## citric.acid 0.06233094 -0.07572873  
## residual.sugar -0.02666437 -0.45063122  
## chlorides 0.01676288 -0.36018871  
## free.sulfur.dioxide 0.05921725 -0.25010394  
## total.sulfur.dioxide 0.13456237 -0.44889210  
## density 0.07449315 -0.78013762  
## pH 0.15595150 0.12143210  
## sulphates 1.00000000 -0.01743277  
## alcohol -0.01743277 1.00000000

corrplot(cor(wine))



pairs.panels(wine, ellipses = FALSE, smooth = FALSE, lm = TRUE)

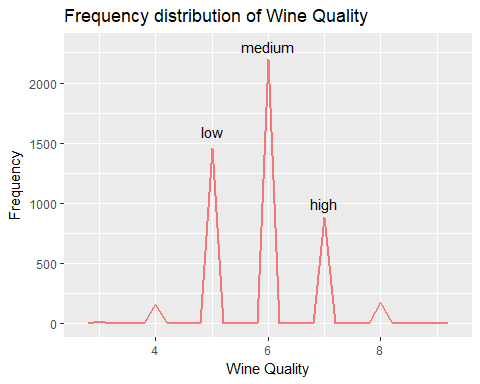


##  
  
## Question No. 3  
library(ggplot2)

##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

w <- ggplot(whitewine, aes(quality))  
w + geom\_freqpoly(size=1, alpha=0.5, color="red", binwidth = 0.2) + labs(x="Wine Quality", y="Frequency", title="Frequency distribution of Wine Quality") + annotate("text", x=6, y=2300, label="medium") + annotate("text", x=7, y=1000, label="high") + annotate("text", x=5, y=1600, label="low")



##  
  
## Question No. 4  
library(dplyr)

##   
## Attaching package: 'dplyr'

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

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

winequality <- mutate(whitewine, grade = quality)  
x <- winequality[, c(12, 13)]  
x[x$grade > 6, ] <- "high"  
x[x$grade < 6, ] <- "low"  
x[x$grade == 6, ] <- "medium"  
View(x)  
winequality <- cbind(whitewine, x)  
winequality <- winequality[, -13]  
View(winequality)  
##  
  
## Question No. 5  
## package clusterSim  
library(clusterSim)

## Loading required package: cluster

## Loading required package: MASS

##   
## Attaching package: 'MASS'

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

##   
## This is package 'modeest' written by P. PONCET.  
## For a complete list of functions, use 'library(help = "modeest")' or 'help.start()'.

winenormal <- data.Normalization(whitewine)  
View(winenormal)  
summary(winenormal)

## fixed.acidity volatile.acidity citric.acid residual.sugar   
## Min. : 3.800 Min. :0.0800 Min. :0.0000 Min. : 0.600   
## 1st Qu.: 6.300 1st Qu.:0.2100 1st Qu.:0.2700 1st Qu.: 1.700   
## Median : 6.800 Median :0.2600 Median :0.3200 Median : 5.200   
## Mean : 6.855 Mean :0.2782 Mean :0.3342 Mean : 6.391   
## 3rd Qu.: 7.300 3rd Qu.:0.3200 3rd Qu.:0.3900 3rd Qu.: 9.900   
## Max. :14.200 Max. :1.1000 Max. :1.6600 Max. :65.800   
## chlorides free.sulfur.dioxide total.sulfur.dioxide  
## Min. :0.00900 Min. : 2.00 Min. : 9.0   
## 1st Qu.:0.03600 1st Qu.: 23.00 1st Qu.:108.0   
## Median :0.04300 Median : 34.00 Median :134.0   
## Mean :0.04577 Mean : 35.31 Mean :138.4   
## 3rd Qu.:0.05000 3rd Qu.: 46.00 3rd Qu.:167.0   
## Max. :0.34600 Max. :289.00 Max. :440.0   
## density pH sulphates alcohol   
## Min. :0.9871 Min. :2.720 Min. :0.2200 Min. : 8.00   
## 1st Qu.:0.9917 1st Qu.:3.090 1st Qu.:0.4100 1st Qu.: 9.50   
## Median :0.9937 Median :3.180 Median :0.4700 Median :10.40   
## Mean :0.9940 Mean :3.188 Mean :0.4898 Mean :10.51   
## 3rd Qu.:0.9961 3rd Qu.:3.280 3rd Qu.:0.5500 3rd Qu.:11.40   
## Max. :1.0390 Max. :3.820 Max. :1.0800 Max. :14.20   
## quality   
## Min. :3.000   
## 1st Qu.:5.000   
## Median :6.000   
## Mean :5.878   
## 3rd Qu.:6.000   
## Max. :9.000

##  
  
## Question No. 6  
wine\_index <- sample(1:nrow(winenormal), 0.7\*nrow(winenormal))  
wine\_train\_set <- winenormal[wine\_index, ]  
wine\_test\_set <- winenormal[-wine\_index, ]  
str(wine\_train\_set)

## 'data.frame': 3428 obs. of 12 variables:  
## $ fixed.acidity : num 7.1 8 6.9 6.2 7.1 6.1 5.6 6.5 6.7 7.1 ...  
## $ volatile.acidity : num 0.42 0.45 0.28 0.255 0.43 0.3 0.41 0.29 0.16 0.21 ...  
## $ citric.acid : num 0.2 0.36 0.28 0.27 0.61 0.47 0.22 0.52 0.49 0.35 ...  
## $ residual.sugar : num 2.8 8.8 12.2 1.3 11.8 1.4 7.1 7.9 2.4 2.5 ...  
## $ chlorides : num 0.038 0.026 0.042 0.037 0.045 0.049 0.05 0.049 0.046 0.04 ...  
## $ free.sulfur.dioxide : num 28 50 52 30 54 50 44 35 57 41 ...  
## $ total.sulfur.dioxide: num 109 151 139 86 155 187 154 192 187 186 ...  
## $ density : num 0.99 0.993 0.995 0.988 0.997 ...  
## $ pH : num 3.23 3.07 3.03 3.05 3.11 3.19 3.3 3.16 3.62 3.32 ...  
## $ sulphates : num 0.47 0.25 0.56 0.59 0.45 0.45 0.4 0.51 0.81 0.56 ...  
## $ alcohol : num 13.4 12.7 10.4 12.9 8.7 9.5 10.5 9.5 10.4 12.5 ...  
## $ quality : num 6 8 6 7 5 5 5 6 6 6 ...

str(wine\_test\_set)

## 'data.frame': 1470 obs. of 12 variables:  
## $ fixed.acidity : num 8.1 6.2 6.3 6.6 6.2 7.6 8.5 5.8 6.5 6.7 ...  
## $ volatile.acidity : num 0.28 0.32 0.3 0.17 0.66 0.67 0.24 0.27 0.39 0.24 ...  
## $ citric.acid : num 0.4 0.16 0.34 0.38 0.48 0.14 0.39 0.2 0.23 0.39 ...  
## $ residual.sugar : num 6.9 7 1.6 1.5 1.2 ...  
## $ chlorides : num 0.05 0.045 0.049 0.032 0.029 0.074 0.044 0.044 0.051 0.173 ...  
## $ free.sulfur.dioxide : num 30 30 14 28 29 25 20 22 25 63 ...  
## $ total.sulfur.dioxide: num 97 136 132 112 75 168 142 179 149 157 ...  
## $ density : num 0.995 0.995 0.994 0.991 0.989 ...  
## $ pH : num 3.26 3.18 3.3 3.25 3.33 3.05 3.2 3.37 3.24 3.1 ...  
## $ sulphates : num 0.44 0.47 0.49 0.55 0.39 0.51 0.53 0.37 0.35 0.34 ...  
## $ alcohol : num 10.1 9.6 9.5 11.4 12.8 9.3 10 10.2 10 9.4 ...  
## $ quality : num 6 6 6 7 8 5 6 5 5 6 ...

##  
  
## Question No. 7  
wine\_train\_set\_new <- wine\_train\_set[-12]  
wine\_test\_set\_new <- wine\_test\_set[-12]  
nrow(wine\_train\_set\_new)

## [1] 3428

nrow(wine\_test\_set\_new)

## [1] 1470

##  
wine\_train\_set\_label <- wine\_train\_set$quality  
wine\_test\_set\_label <- wine\_test\_set$quality  
##  
library(caret)

## Loading required package: lattice

library(class)  
library(gmodels)  
## knn test for k = 3  
wine\_knn\_prediction <- knn(train = wine\_train\_set\_new, test = wine\_test\_set\_new, cl = wine\_train\_set\_label, k = 3)  
## knn test for k=11  
wine\_knn\_prediction2 <- knn(train = wine\_train\_set\_new, test = wine\_test\_set\_new, cl = wine\_train\_set\_label, k = 11)  
##  
  
## Question No. 8  
## knn performance for k=3  
CrossTable(x=wine\_test\_set\_label, y=wine\_knn\_prediction, prop.chisq = FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1470   
##   
##   
## | wine\_knn\_prediction   
## wine\_test\_set\_label | 3 | 4 | 5 | 6 | 7 | 8 | Row Total |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 3 | 0 | 0 | 4 | 3 | 2 | 0 | 9 |   
## | 0.000 | 0.000 | 0.444 | 0.333 | 0.222 | 0.000 | 0.006 |   
## | 0.000 | 0.000 | 0.010 | 0.004 | 0.007 | 0.000 | |   
## | 0.000 | 0.000 | 0.003 | 0.002 | 0.001 | 0.000 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 4 | 0 | 7 | 16 | 14 | 3 | 0 | 40 |   
## | 0.000 | 0.175 | 0.400 | 0.350 | 0.075 | 0.000 | 0.027 |   
## | 0.000 | 0.179 | 0.039 | 0.020 | 0.011 | 0.000 | |   
## | 0.000 | 0.005 | 0.011 | 0.010 | 0.002 | 0.000 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 5 | 1 | 16 | 195 | 196 | 50 | 5 | 463 |   
## | 0.002 | 0.035 | 0.421 | 0.423 | 0.108 | 0.011 | 0.315 |   
## | 0.500 | 0.410 | 0.474 | 0.278 | 0.177 | 0.167 | |   
## | 0.001 | 0.011 | 0.133 | 0.133 | 0.034 | 0.003 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 6 | 1 | 13 | 153 | 383 | 111 | 7 | 668 |   
## | 0.001 | 0.019 | 0.229 | 0.573 | 0.166 | 0.010 | 0.454 |   
## | 0.500 | 0.333 | 0.372 | 0.543 | 0.392 | 0.233 | |   
## | 0.001 | 0.009 | 0.104 | 0.261 | 0.076 | 0.005 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 7 | 0 | 3 | 39 | 79 | 101 | 10 | 232 |   
## | 0.000 | 0.013 | 0.168 | 0.341 | 0.435 | 0.043 | 0.158 |   
## | 0.000 | 0.077 | 0.095 | 0.112 | 0.357 | 0.333 | |   
## | 0.000 | 0.002 | 0.027 | 0.054 | 0.069 | 0.007 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 8 | 0 | 0 | 4 | 28 | 16 | 8 | 56 |   
## | 0.000 | 0.000 | 0.071 | 0.500 | 0.286 | 0.143 | 0.038 |   
## | 0.000 | 0.000 | 0.010 | 0.040 | 0.057 | 0.267 | |   
## | 0.000 | 0.000 | 0.003 | 0.019 | 0.011 | 0.005 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 9 | 0 | 0 | 0 | 2 | 0 | 0 | 2 |   
## | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.001 |   
## | 0.000 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 | |   
## | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|  
## Column Total | 2 | 39 | 411 | 705 | 283 | 30 | 1470 |   
## | 0.001 | 0.027 | 0.280 | 0.480 | 0.193 | 0.020 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|  
##   
##

## knn performance for k = 11  
CrossTable(x=wine\_test\_set\_label, y=wine\_knn\_prediction2, prop.chisq = FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1470   
##   
##   
## | wine\_knn\_prediction2   
## wine\_test\_set\_label | 4 | 5 | 6 | 7 | 8 | Row Total |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 3 | 0 | 4 | 5 | 0 | 0 | 9 |   
## | 0.000 | 0.444 | 0.556 | 0.000 | 0.000 | 0.006 |   
## | 0.000 | 0.010 | 0.006 | 0.000 | 0.000 | |   
## | 0.000 | 0.003 | 0.003 | 0.000 | 0.000 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 4 | 2 | 22 | 13 | 3 | 0 | 40 |   
## | 0.050 | 0.550 | 0.325 | 0.075 | 0.000 | 0.027 |   
## | 0.143 | 0.053 | 0.016 | 0.013 | 0.000 | |   
## | 0.001 | 0.015 | 0.009 | 0.002 | 0.000 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 5 | 8 | 183 | 228 | 44 | 0 | 463 |   
## | 0.017 | 0.395 | 0.492 | 0.095 | 0.000 | 0.315 |   
## | 0.571 | 0.437 | 0.282 | 0.196 | 0.000 | |   
## | 0.005 | 0.124 | 0.155 | 0.030 | 0.000 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 6 | 3 | 169 | 400 | 93 | 3 | 668 |   
## | 0.004 | 0.253 | 0.599 | 0.139 | 0.004 | 0.454 |   
## | 0.214 | 0.403 | 0.495 | 0.415 | 0.600 | |   
## | 0.002 | 0.115 | 0.272 | 0.063 | 0.002 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 7 | 1 | 33 | 128 | 70 | 0 | 232 |   
## | 0.004 | 0.142 | 0.552 | 0.302 | 0.000 | 0.158 |   
## | 0.071 | 0.079 | 0.158 | 0.312 | 0.000 | |   
## | 0.001 | 0.022 | 0.087 | 0.048 | 0.000 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 8 | 0 | 8 | 32 | 14 | 2 | 56 |   
## | 0.000 | 0.143 | 0.571 | 0.250 | 0.036 | 0.038 |   
## | 0.000 | 0.019 | 0.040 | 0.062 | 0.400 | |   
## | 0.000 | 0.005 | 0.022 | 0.010 | 0.001 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## 9 | 0 | 0 | 2 | 0 | 0 | 2 |   
## | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.001 |   
## | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | |   
## | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## Column Total | 14 | 419 | 808 | 224 | 5 | 1470 |   
## | 0.010 | 0.285 | 0.550 | 0.152 | 0.003 | |   
## --------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
##   
##

## Question 8: Exploring other options  
##  
## By decision tree approach  
library(party)

## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

library(partykit)

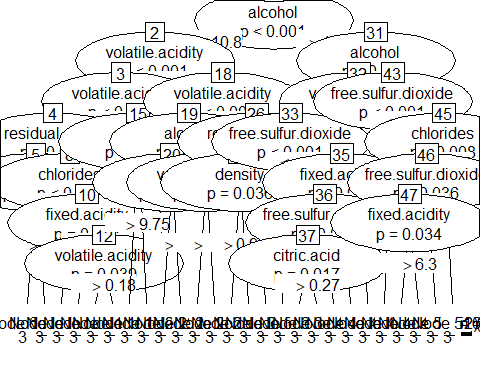
##   
## Attaching package: 'partykit'

## The following objects are masked from 'package:party':  
##   
## cforest, ctree, ctree\_control, edge\_simple, mob, mob\_control,  
## node\_barplot, node\_bivplot, node\_boxplot, node\_inner,  
## node\_surv, node\_terminal

wine\_tree\_model <- ctree(quality ~ fixed.acidity+volatile.acidity+citric.acid+residual.sugar+chlorides+free.sulfur.dioxide+total.sulfur.dioxide+density+alcohol, data = wine\_train\_set)  
print(wine\_tree\_model)

##   
## Model formula:  
## quality ~ fixed.acidity + volatile.acidity + citric.acid + residual.sugar +   
## chlorides + free.sulfur.dioxide + total.sulfur.dioxide +   
## density + alcohol  
##   
## Fitted party:  
## [1] root  
## | [2] alcohol <= 10.8  
## | | [3] volatile.acidity <= 0.25  
## | | | [4] volatile.acidity <= 0.205  
## | | | | [5] residual.sugar <= 12.55  
## | | | | | [6] fixed.acidity <= 8.7: 5.956 (n = 390, err = 196.3)  
## | | | | | [7] fixed.acidity > 8.7: 5.067 (n = 15, err = 14.9)  
## | | | | [8] residual.sugar > 12.55  
## | | | | | [9] chlorides <= 0.055: 6.208 (n = 77, err = 28.7)  
## | | | | | [10] chlorides > 0.055  
## | | | | | | [11] fixed.acidity <= 6.8: 6.500 (n = 8, err = 4.0)  
## | | | | | | [12] fixed.acidity > 6.8  
## | | | | | | | [13] volatile.acidity <= 0.18: 6.778 (n = 9, err = 1.6)  
## | | | | | | | [14] volatile.acidity > 0.18: 7.867 (n = 15, err = 1.7)  
## | | | [15] volatile.acidity > 0.205  
## | | | | [16] alcohol <= 9.75: 5.615 (n = 286, err = 123.7)  
## | | | | [17] alcohol > 9.75: 5.904 (n = 239, err = 116.8)  
## | | [18] volatile.acidity > 0.25  
## | | | [19] volatile.acidity <= 0.445  
## | | | | [20] alcohol <= 10.1  
## | | | | | [21] alcohol <= 8.6: 4.769 (n = 13, err = 4.3)  
## | | | | | [22] alcohol > 8.6: 5.331 (n = 740, err = 277.9)  
## | | | | [23] alcohol > 10.1  
## | | | | | [24] volatile.acidity <= 0.33: 5.781 (n = 196, err = 151.6)  
## | | | | | [25] volatile.acidity > 0.33: 5.333 (n = 51, err = 19.3)  
## | | | [26] volatile.acidity > 0.445  
## | | | | [27] residual.sugar <= 2.1  
## | | | | | [28] density <= 0.994: 4.852 (n = 27, err = 13.4)  
## | | | | | [29] density > 0.994: 4.000 (n = 8, err = 2.0)  
## | | | | [30] residual.sugar > 2.1: 5.059 (n = 85, err = 28.7)  
## | [31] alcohol > 10.8  
## | | [32] alcohol <= 11.85  
## | | | [33] volatile.acidity <= 0.45  
## | | | | [34] free.sulfur.dioxide <= 11: 5.108 (n = 37, err = 37.6)  
## | | | | [35] free.sulfur.dioxide > 11  
## | | | | | [36] fixed.acidity <= 6.8  
## | | | | | | [37] free.sulfur.dioxide <= 42  
## | | | | | | | [38] citric.acid <= 0.27: 6.041 (n = 73, err = 30.9)  
## | | | | | | | [39] citric.acid > 0.27: 6.388 (n = 183, err = 77.5)  
## | | | | | | [40] free.sulfur.dioxide > 42: 6.571 (n = 84, err = 42.6)  
## | | | | | [41] fixed.acidity > 6.8: 6.079 (n = 265, err = 183.3)  
## | | | [42] volatile.acidity > 0.45: 4.538 (n = 13, err = 7.2)  
## | | [43] alcohol > 11.85  
## | | | [44] free.sulfur.dioxide <= 10: 5.606 (n = 33, err = 37.9)  
## | | | [45] free.sulfur.dioxide > 10  
## | | | | [46] chlorides <= 0.06  
## | | | | | [47] free.sulfur.dioxide <= 21  
## | | | | | | [48] fixed.acidity <= 6.3: 6.692 (n = 52, err = 29.1)  
## | | | | | | [49] fixed.acidity > 6.3: 6.250 (n = 68, err = 34.8)  
## | | | | | [50] free.sulfur.dioxide > 21: 6.707 (n = 450, err = 247.3)  
## | | | | [51] chlorides > 0.06: 5.727 (n = 11, err = 2.2)  
##   
## Number of inner nodes: 25  
## Number of terminal nodes: 26

plot(wine\_tree\_model, type = "extended")



wine\_tree\_prediction <- predict(wine\_tree\_model, wine\_test\_set)  
head(wine\_tree\_prediction)

## 6 7 9 16 21 24   
## 5.331081 5.331081 5.331081 6.387978 6.706667 4.851852

table(wine\_tree\_prediction, wine\_test\_set$quality)

##   
## wine\_tree\_prediction 3 4 5 6 7 8 9  
## 4 0 0 0 1 0 0 0  
## 4.53846153846154 1 1 2 2 3 0 0  
## 4.76923076923077 0 1 7 1 0 0 0  
## 4.85185185185185 1 3 8 0 0 0 0  
## 5.05882352941176 1 4 24 13 1 0 0  
## 5.06666666666667 0 0 1 2 1 0 0  
## 5.10810810810811 0 6 7 5 2 0 0  
## 5.33108108108108 0 12 188 122 7 0 0  
## 5.33333333333333 0 0 12 13 1 0 0  
## 5.60606060606061 0 0 1 9 2 0 0  
## 5.61538461538461 3 1 58 59 4 1 0  
## 5.72727272727273 0 0 0 4 0 0 0  
## 5.78061224489796 1 4 33 26 6 0 0  
## 5.90376569037657 0 2 25 56 8 2 0  
## 5.95641025641026 0 3 47 92 25 5 0  
## 6.04109589041096 0 0 1 17 6 1 0  
## 6.07924528301887 0 0 23 67 28 7 0  
## 6.20779220779221 0 0 3 18 13 2 0  
## 6.25 1 2 2 15 7 0 0  
## 6.3879781420765 1 0 9 36 26 3 0  
## 6.5 0 0 0 0 2 0 0  
## 6.57142857142857 0 1 3 17 9 5 0  
## 6.69230769230769 0 0 0 13 5 3 0  
## 6.70666666666667 0 0 9 79 75 26 2  
## 6.77777777777778 0 0 0 1 1 0 0  
## 7.86666666666667 0 0 0 0 0 1 0

##  
  
## Stepwise regression approach  
start <- lm(quality ~ 1, data = winenormal)  
end <- lm(quality ~ fixed.acidity+volatile.acidity+citric.acid+residual.sugar+chlorides+free.sulfur.dioxide+total.sulfur.dioxide+density+alcohol, data = winenormal)  
##  
res.both <- step(start,scope=list(upper=end),direction="both",test="F")

## Start: AIC=-1188.69  
## quality ~ 1  
##   
## Df Sum of Sq RSS AIC F value Pr(>F)   
## + alcohol 1 728.73 3112.3 -2217.1 1146.3955 < 2.2e-16 \*\*\*  
## + density 1 362.30 3478.7 -1672.0 509.9111 < 2.2e-16 \*\*\*  
## + chlorides 1 169.28 3671.7 -1407.5 225.7271 < 2.2e-16 \*\*\*  
## + volatile.acidity 1 145.64 3695.4 -1376.0 192.9582 < 2.2e-16 \*\*\*  
## + total.sulfur.dioxide 1 117.28 3723.7 -1338.6 154.1982 < 2.2e-16 \*\*\*  
## + fixed.acidity 1 49.62 3791.4 -1250.4 64.0805 1.480e-15 \*\*\*  
## + residual.sugar 1 36.57 3804.4 -1233.5 47.0641 7.724e-12 \*\*\*  
## <none> 3841.0 -1188.7   
## + citric.acid 1 0.33 3840.7 -1187.1 0.4153 0.5193   
## + free.sulfur.dioxide 1 0.26 3840.7 -1187.0 0.3259 0.5681   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=-2217.14  
## quality ~ alcohol  
##   
## Df Sum of Sq RSS AIC F value Pr(>F)   
## + volatile.acidity 1 193.99 2918.3 -2530.4 325.3965 < 2.2e-16 \*\*\*  
## + free.sulfur.dioxide 1 56.18 3056.1 -2304.4 89.9863 < 2.2e-16 \*\*\*  
## + residual.sugar 1 46.96 3065.3 -2289.6 74.9887 < 2.2e-16 \*\*\*  
## + fixed.acidity 1 14.51 3097.7 -2238.0 22.9269 1.732e-06 \*\*\*  
## + chlorides 1 12.42 3099.8 -2234.7 19.6110 9.699e-06 \*\*\*  
## + density 1 10.48 3101.8 -2231.7 16.5454 4.825e-05 \*\*\*  
## + citric.acid 1 2.18 3110.1 -2218.6 3.4373 0.06380 .   
## + total.sulfur.dioxide 1 2.08 3110.2 -2218.4 3.2719 0.07054 .   
## <none> 3112.3 -2217.1   
## - alcohol 1 728.73 3841.0 -1188.7 1146.3955 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=-2530.37  
## quality ~ alcohol + volatile.acidity  
##   
## Df Sum of Sq RSS AIC F value Pr(>F)   
## + residual.sugar 1 70.27 2848.0 -2647.8 120.7536 < 2.2e-16 \*\*\*  
## + free.sulfur.dioxide 1 40.48 2877.8 -2596.8 68.8482 < 2.2e-16 \*\*\*  
## + density 1 25.64 2892.6 -2571.6 43.3788 4.985e-11 \*\*\*  
## + fixed.acidity 1 16.11 2902.2 -2555.5 27.1649 1.945e-07 \*\*\*  
## + total.sulfur.dioxide 1 11.17 2907.1 -2547.2 18.7964 1.484e-05 \*\*\*  
## + chlorides 1 4.47 2913.8 -2535.9 7.5125 0.006149 \*\*   
## <none> 2918.3 -2530.4   
## + citric.acid 1 0.30 2918.0 -2528.9 0.5053 0.477230   
## - volatile.acidity 1 193.99 3112.3 -2217.1 325.3965 < 2.2e-16 \*\*\*  
## - alcohol 1 777.09 3695.4 -1376.0 1303.4593 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=-2647.76  
## quality ~ alcohol + volatile.acidity + residual.sugar  
##   
## Df Sum of Sq RSS AIC F value Pr(>F)   
## + free.sulfur.dioxide 1 21.00 2827.0 -2682.0 36.3519 1.768e-09 \*\*\*  
## + density 1 20.81 2827.2 -2681.7 36.0104 2.104e-09 \*\*\*  
## + fixed.acidity 1 19.00 2829.0 -2678.5 32.8649 1.047e-08 \*\*\*  
## + total.sulfur.dioxide 1 1.90 2846.1 -2649.0 3.2689 0.07067 .   
## + chlorides 1 1.64 2846.4 -2648.6 2.8170 0.09334 .   
## + citric.acid 1 1.59 2846.4 -2648.5 2.7296 0.09857 .   
## <none> 2848.0 -2647.8   
## - residual.sugar 1 70.27 2918.3 -2530.4 120.7536 < 2.2e-16 \*\*\*  
## - volatile.acidity 1 217.30 3065.3 -2289.6 373.4168 < 2.2e-16 \*\*\*  
## - alcohol 1 819.45 3667.4 -1411.2 1408.1494 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=-2682.01  
## quality ~ alcohol + volatile.acidity + residual.sugar + free.sulfur.dioxide  
##   
## Df Sum of Sq RSS AIC F value Pr(>F)   
## + density 1 19.04 2807.9 -2713.1 33.1733 8.945e-09 \*\*\*  
## + fixed.acidity 1 15.50 2811.5 -2706.9 26.9636 2.157e-07 \*\*\*  
## + total.sulfur.dioxide 1 2.41 2824.6 -2684.2 4.1733 0.04112 \*   
## + chlorides 1 2.22 2824.8 -2683.8 3.8406 0.05008 .   
## + citric.acid 1 2.21 2824.8 -2683.8 3.8285 0.05045 .   
## <none> 2827.0 -2682.0   
## - free.sulfur.dioxide 1 21.00 2848.0 -2647.8 36.3519 1.768e-09 \*\*\*  
## - residual.sugar 1 50.79 2877.8 -2596.8 87.9070 < 2.2e-16 \*\*\*  
## - volatile.acidity 1 200.33 3027.3 -2348.7 346.7334 < 2.2e-16 \*\*\*  
## - alcohol 1 839.45 3666.4 -1410.5 1452.9275 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=-2713.11  
## quality ~ alcohol + volatile.acidity + residual.sugar + free.sulfur.dioxide +   
## density  
##   
## Df Sum of Sq RSS AIC F value Pr(>F)   
## + fixed.acidity 1 4.164 2803.8 -2718.4 7.2634 0.007061 \*\*   
## + chlorides 1 1.414 2806.5 -2713.6 2.4637 0.116565   
## <none> 2807.9 -2713.1   
## + citric.acid 1 0.431 2807.5 -2711.9 0.7509 0.386249   
## + total.sulfur.dioxide 1 0.084 2807.9 -2711.3 0.1468 0.701631   
## - density 1 19.041 2827.0 -2682.0 33.1733 8.945e-09 \*\*\*  
## - free.sulfur.dioxide 1 19.237 2827.2 -2681.7 33.5146 7.514e-09 \*\*\*  
## - residual.sugar 1 53.793 2861.7 -2622.2 93.7179 < 2.2e-16 \*\*\*  
## - alcohol 1 160.305 2968.2 -2443.2 279.2835 < 2.2e-16 \*\*\*  
## - volatile.acidity 1 191.207 2999.2 -2392.4 333.1207 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=-2718.38  
## quality ~ alcohol + volatile.acidity + residual.sugar + free.sulfur.dioxide +   
## density + fixed.acidity  
##   
## Df Sum of Sq RSS AIC F value Pr(>F)   
## + chlorides 1 1.694 2802.1 -2719.3 2.9554 0.085654 .   
## <none> 2803.8 -2718.4   
## + total.sulfur.dioxide 1 0.139 2803.7 -2716.6 0.2419 0.622836   
## + citric.acid 1 0.041 2803.8 -2716.4 0.0707 0.790353   
## - fixed.acidity 1 4.164 2807.9 -2713.1 7.2634 0.007061 \*\*   
## - density 1 7.709 2811.5 -2706.9 13.4470 0.000248 \*\*\*  
## - free.sulfur.dioxide 1 17.599 2821.4 -2689.7 30.6994 3.170e-08 \*\*\*  
## - residual.sugar 1 33.765 2837.6 -2661.8 58.9011 1.989e-14 \*\*\*  
## - alcohol 1 156.104 2959.9 -2455.0 272.3120 < 2.2e-16 \*\*\*  
## - volatile.acidity 1 194.294 2998.1 -2392.2 338.9311 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=-2719.34  
## quality ~ alcohol + volatile.acidity + residual.sugar + free.sulfur.dioxide +   
## density + fixed.acidity + chlorides  
##   
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 2802.1 -2719.3   
## - chlorides 1 1.694 2803.8 -2718.4 2.9554 0.0856539 .   
## + total.sulfur.dioxide 1 0.123 2802.0 -2717.6 0.2147 0.6431194   
## + citric.acid 1 0.002 2802.1 -2717.3 0.0032 0.9549224   
## - fixed.acidity 1 4.444 2806.5 -2713.6 7.7546 0.0053783 \*\*   
## - density 1 7.036 2809.1 -2709.1 12.2786 0.0004623 \*\*\*  
## - free.sulfur.dioxide 1 18.045 2820.1 -2689.9 31.4902 2.115e-08 \*\*\*  
## - residual.sugar 1 31.368 2833.5 -2666.8 54.7404 1.610e-13 \*\*\*  
## - alcohol 1 149.393 2951.5 -2466.9 260.7094 < 2.2e-16 \*\*\*  
## - volatile.acidity 1 188.136 2990.2 -2403.1 328.3209 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

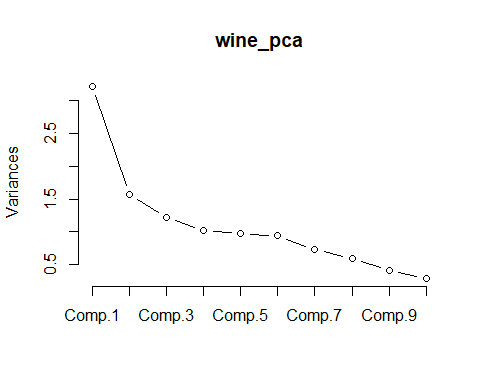
summary(res.both)

##   
## Call:  
## lm(formula = quality ~ alcohol + volatile.acidity + residual.sugar +   
## free.sulfur.dioxide + density + fixed.acidity + chlorides,   
## data = winenormal)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.8099 -0.5034 -0.0309 0.4707 3.2250   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.094e+01 1.374e+01 3.707 0.000212 \*\*\*  
## alcohol 3.125e-01 1.936e-02 16.146 < 2e-16 \*\*\*  
## volatile.acidity -1.993e+00 1.100e-01 -18.120 < 2e-16 \*\*\*  
## residual.sugar 4.064e-02 5.493e-03 7.399 1.61e-13 \*\*\*  
## free.sulfur.dioxide 3.819e-03 6.806e-04 5.612 2.12e-08 \*\*\*  
## density -4.815e+01 1.374e+01 -3.504 0.000462 \*\*\*  
## fixed.acidity -4.171e-02 1.498e-02 -2.785 0.005378 \*\*   
## chlorides -9.264e-01 5.389e-01 -1.719 0.085654 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.757 on 4890 degrees of freedom  
## Multiple R-squared: 0.2705, Adjusted R-squared: 0.2694   
## F-statistic: 259 on 7 and 4890 DF, p-value: < 2.2e-16

##  
## shows chlorides, fixed.acidity, total.sulfur.dioxide, citric.acid are low significant  
## Regression   
summary(lm(quality ~ volatile.acidity+residual.sugar+free.sulfur.dioxide+density+alcohol, data = winenormal))

##   
## Call:  
## lm(formula = quality ~ volatile.acidity + residual.sugar + free.sulfur.dioxide +   
## density + alcohol, data = winenormal)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.7991 -0.5089 -0.0349 0.4652 3.1701   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 70.962196 11.950699 5.938 3.09e-09 \*\*\*  
## volatile.acidity -1.992534 0.109170 -18.252 < 2e-16 \*\*\*  
## residual.sugar 0.047968 0.004955 9.681 < 2e-16 \*\*\*  
## free.sulfur.dioxide 0.003925 0.000678 5.789 7.51e-09 \*\*\*  
## density -68.509916 11.894845 -5.760 8.95e-09 \*\*\*  
## alcohol 0.297280 0.017789 16.712 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7576 on 4892 degrees of freedom  
## Multiple R-squared: 0.269, Adjusted R-squared: 0.2682   
## F-statistic: 360 on 5 and 4892 DF, p-value: < 2.2e-16

## PCA investigation for best regression  
library(lattice)  
winenormal.new <- winenormal[-12]  
wine\_pca <- princomp(winenormal.new, cor = TRUE)  
screeplot(wine\_pca, type = "line")



wine\_pca$loadings

##   
## Loadings:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7  
## fixed.acidity 0.157 0.588 -0.121 0.251 -0.104 0.198  
## volatile.acidity 0.591 -0.274 0.643 0.122 -0.269  
## citric.acid 0.144 0.345 -0.504 -0.149 0.132 -0.705  
## residual.sugar 0.427 0.214 0.274 -0.289 -0.213  
## chlorides 0.212 0.102 -0.711 -0.329 0.396   
## free.sulfur.dioxide 0.300 -0.290 -0.279 0.306 0.177 0.494 0.167  
## total.sulfur.dioxide 0.407 -0.244 -0.124 0.293 0.276   
## density 0.512 0.129 -0.328 -0.110  
## pH -0.129 -0.581 -0.127 -0.120 -0.193 -0.427  
## sulphates -0.223 -0.433 -0.442 0.401 -0.481 0.309  
## alcohol -0.437 -0.106 0.141 0.337 0.139 -0.129  
## Comp.8 Comp.9 Comp.10 Comp.11  
## fixed.acidity 0.588 -0.331 0.132 -0.171   
## volatile.acidity 0.146 0.224   
## citric.acid -0.152 0.202   
## residual.sugar -0.388 -0.409 -0.490   
## chlorides -0.100 -0.394   
## free.sulfur.dioxide -0.144 0.567   
## total.sulfur.dioxide 0.247 0.155 -0.709   
## density 0.760   
## pH 0.534 -0.261 0.111 -0.141   
## sulphates -0.270   
## alcohol -0.196 -0.621 -0.273 0.358   
##   
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8  
## SS loadings 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000  
## Proportion Var 0.091 0.091 0.091 0.091 0.091 0.091 0.091 0.091  
## Cumulative Var 0.091 0.182 0.273 0.364 0.455 0.545 0.636 0.727  
## Comp.9 Comp.10 Comp.11  
## SS loadings 1.000 1.000 1.000  
## Proportion Var 0.091 0.091 0.091  
## Cumulative Var 0.818 0.909 1.000

## from this analysis it shows that the following should provide the best regression for quality  
summary(glm(quality ~ citric.acid+residual.sugar+chlorides+free.sulfur.dioxide+density+pH, data = winenormal))

##   
## Call:  
## glm(formula = quality ~ citric.acid + residual.sugar + chlorides +   
## free.sulfur.dioxide + density + pH, data = winenormal)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.8054 -0.5623 0.0122 0.4722 4.5421   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.335e+02 7.380e+00 31.636 < 2e-16 \*\*\*  
## citric.acid 5.769e-01 9.574e-02 6.025 1.81e-09 \*\*\*  
## residual.sugar 9.891e-02 4.367e-03 22.650 < 2e-16 \*\*\*  
## chlorides -2.482e+00 5.545e-01 -4.476 7.78e-06 \*\*\*  
## free.sulfur.dioxide 3.573e-03 6.985e-04 5.116 3.25e-07 \*\*\*  
## density -2.325e+02 7.502e+00 -30.992 < 2e-16 \*\*\*  
## pH 8.406e-01 7.834e-02 10.731 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.6181935)  
##   
## Null deviance: 3841.0 on 4897 degrees of freedom  
## Residual deviance: 3023.6 on 4891 degrees of freedom  
## AIC: 11553  
##   
## Number of Fisher Scoring iterations: 2

##  
## End of Assignment 3