wine quality prediction

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Import to R the following file:

winequality <- read.csv(file = "http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv", header = TRUE, sep = ";")

Check data characteristics. Is there missing data?

str(winequality)

## '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(winequality)

## 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

sum(is.na(winequality) == TRUE)

## [1] 0

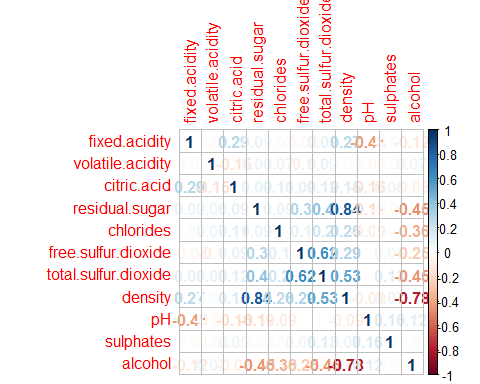
There is no missing data

The correlation between the attributes other than wine quality?

wine <- winequality[,-12]  
wine\_cor <- cor(wine)  
#install.packages("corrplot")  
library(corrplot)

## corrplot 0.84 loaded

corrplot(wine\_cor, method = "number")

 Taking a look at the correlation coefficients r for the predictor variables, we see that density is strongly correlated with residual.sugar (r=0.84) and alcohol (r=???0.78), and moderately correlated with total.sulfur.dioxide (r=0.53). free.sulfur.dioxide and total.sulfur.dioxide are also moderately correlated with each other (r=0.62) although this is trivially known because of course, free sulfur dioxide is incorporated into the total sulfur dioxide. So, we should actually remove the variables residual.sugar and density, as well as total.sulfur.dioxide because of its direct relationship with free.sulfur.dioxide, in order to address problems with multicollinearity. We’re going to withhold removing alcohol, to see the if the initial effect of removing just these three correlated variables is enough to address the issue.

winequality <- subset(winequality, select = -c(4,7,8))  
str(winequality)

## 'data.frame': 4898 obs. of 9 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 ...  
## $ 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 ...  
## $ 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 ...

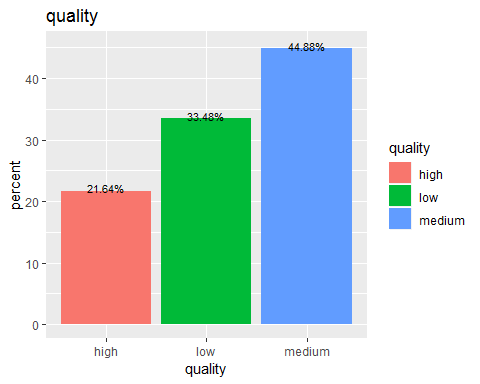
wine <- winequality[,-9]

Reduce the levels of rating for quality to three levels as high, medium and low

winequality$quality[which(winequality$quality %in% c(3,4,5))] = "low"  
winequality$quality[which(winequality$quality %in% c(6))] = "medium"  
winequality$quality[which(winequality$quality %in% c(7,8,9))] = "high"  
winequality$quality <- as.factor(winequality$quality)

Graph the frequency distribution of wine quality.

#install.packages("ggplot2")  
library(ggplot2)  
ggplot(winequality, aes(x = quality, fill = quality)) +  
 geom\_bar(aes(y = prop.table(..count..) \* 100), position='dodge') +  
 geom\_text(aes(y = prop.table(..count..) \* 100 + 0.5,   
 label = paste0(round(prop.table(..count..) \* 100, 2), '%')),   
 stat = 'count',   
 position = position\_dodge(.9),   
 size = 3) +   
 labs(x = 'quality', y = 'percent', fill = 'quality', title = 'quality')

 Normalize the data set.

normalize <- function(x) {  
 return ((x - min(x)) / (max(x) - min(x))) }  
winedata <- as.data.frame(lapply(wine, normalize))

Divide the data to training and testing groups.

split\_data <- sample(nrow(winedata), floor(nrow(winedata)\*0.7))  
winedata\_train <- winedata[split\_data,]  
winedata\_test <- winedata[-split\_data,]  
  
winedata\_train\_labels <- winequality[split\_data, 9]  
winedata\_test\_labels <- winequality[-split\_data, 9]

Use the KNN algorithm to predict the quality of wine using its attributes.

#install.packages("class")  
#install.packages("gmodels")  
library(class)  
library(gmodels)  
  
winedata\_test\_pred <- knn(train = winedata\_train, test = winedata\_test, cl = winedata\_train\_labels, k=10, prob = TRUE)  
winedata\_test\_pred

## [1] low medium high medium medium low medium high medium  
## [10] low medium medium low medium medium low medium low   
## [19] medium low high low low medium low low medium  
## [28] medium medium low low low medium medium low medium  
## [37] low high high medium medium medium low low medium  
## [46] low medium high low medium low low medium medium  
## [55] low medium medium low low low medium high medium  
## [64] medium medium low low medium high medium low low   
## [73] low medium medium low high medium low high low   
## [82] low medium medium medium high medium medium medium medium  
## [91] low medium medium high medium high low medium medium  
## [100] low low medium low medium medium low low high   
## [109] low low medium high medium medium medium low high   
## [118] medium high medium low medium low high medium low   
## [127] low medium medium medium low medium medium low low   
## [136] low medium low low medium low low low medium  
## [145] medium high medium medium low low medium low high   
## [154] low medium medium high low medium medium high low   
## [163] low medium low low low low medium low low   
## [172] high high medium high medium low low medium medium  
## [181] low low low medium medium medium high medium high   
## [190] medium low medium medium low low high low high   
## [199] medium low low medium medium low low medium low   
## [208] low medium low medium low medium low medium medium  
## [217] high medium medium high medium low medium medium low   
## [226] medium medium medium low medium high low low medium  
## [235] high medium medium high low low medium low low   
## [244] medium medium low medium medium high medium high low   
## [253] medium medium low medium high low low low high   
## [262] medium medium medium high high medium medium medium low   
## [271] medium low low low low medium medium low medium  
## [280] high low medium low low low high low low   
## [289] medium high low medium high medium medium medium medium  
## [298] high medium low medium low low high medium low   
## [307] medium medium medium medium low low medium low medium  
## [316] medium medium low low low low low medium low   
## [325] medium low high low high medium low medium low   
## [334] low high low medium low high low low high   
## [343] low medium low low low low low medium medium  
## [352] low low low medium medium low high medium low   
## [361] low high low low medium low high low medium  
## [370] low low low low high medium high high high   
## [379] high low medium high low medium medium medium low   
## [388] low low low low low low medium medium medium  
## [397] medium medium medium medium high low low medium low   
## [406] low low low low medium medium medium medium medium  
## [415] high medium high low medium high medium medium high   
## [424] medium medium medium medium medium low high medium medium  
## [433] medium high low low low medium low medium medium  
## [442] medium high low medium medium low medium medium low   
## [451] medium medium medium high low low low medium medium  
## [460] medium low high low medium medium medium medium low   
## [469] high low low medium medium high medium medium medium  
## [478] medium low medium medium high medium low low medium  
## [487] low medium low high medium medium low low low   
## [496] medium medium high medium low high medium medium medium  
## [505] medium low low low medium medium medium high medium  
## [514] medium medium medium medium low low medium medium medium  
## [523] low medium low low medium medium low low medium  
## [532] medium medium medium high high low high high medium  
## [541] low medium medium high low low medium high high   
## [550] medium low medium medium medium high low medium low   
## [559] medium medium low low medium low low medium medium  
## [568] low medium medium medium medium low low high medium  
## [577] high medium high medium medium medium high high medium  
## [586] low high high high low medium medium medium medium  
## [595] high high medium high low low medium medium low   
## [604] medium medium medium medium high low low medium low   
## [613] low low medium medium low medium high medium medium  
## [622] low low medium medium medium low low low low   
## [631] low medium low low medium high medium low high   
## [640] low low medium medium high low high low low   
## [649] low low medium low medium low low medium high   
## [658] medium medium low low medium medium medium medium low   
## [667] low low low medium low medium medium low low   
## [676] medium high low medium medium medium medium low medium  
## [685] medium low high low high high medium low low   
## [694] high low low medium low medium medium low low   
## [703] medium low high high low low medium medium low   
## [712] medium medium low low low medium low medium medium  
## [721] high medium low low low high low medium low   
## [730] high low low medium high low low low medium  
## [739] low medium high medium low low low low low   
## [748] medium medium medium medium medium medium medium medium medium  
## [757] medium medium medium low low low high medium low   
## [766] low high medium medium medium medium low medium medium  
## [775] medium low low low medium medium medium low low   
## [784] high low low low low medium medium low low   
## [793] medium medium medium low low low low high low   
## [802] low low low medium low low high low medium  
## [811] low low medium high low low high medium low   
## [820] high medium medium medium medium high low medium medium  
## [829] medium medium low high high medium medium high low   
## [838] low medium medium medium medium high medium medium medium  
## [847] low medium medium medium medium medium medium low medium  
## [856] medium medium low medium medium high medium medium medium  
## [865] low low high low low low medium low medium  
## [874] high low low low low low medium medium medium  
## [883] medium medium medium low low medium medium medium low   
## [892] medium high high medium high medium high medium medium  
## [901] high high medium medium medium low low low medium  
## [910] medium low medium high high high high medium high   
## [919] high high medium medium medium low medium low low   
## [928] high medium medium low high high high medium high   
## [937] high high high medium medium low medium medium medium  
## [946] low high low low medium medium medium medium medium  
## [955] medium medium high medium high high low high medium  
## [964] medium medium medium medium high low high medium medium  
## [973] medium medium low medium low low medium high high   
## [982] high medium high medium medium medium medium medium medium  
## [991] high high high medium medium medium medium high medium  
## [1000] medium medium low low high medium high high high   
## [1009] high medium high high medium high medium medium medium  
## [1018] low medium high medium medium medium medium low low   
## [1027] high medium high low high low medium medium low   
## [1036] low medium high medium medium medium medium medium medium  
## [1045] medium low medium high medium medium medium low medium  
## [1054] medium low medium medium low medium high medium medium  
## [1063] low medium low high low high medium high medium  
## [1072] medium medium medium high high high medium medium high   
## [1081] high high high high medium high medium high medium  
## [1090] high medium medium low low low high high medium  
## [1099] medium medium low medium medium medium high high medium  
## [1108] low medium medium low low medium medium high low   
## [1117] low medium medium medium low low medium low high   
## [1126] medium low low medium medium medium high high medium  
## [1135] low high medium medium low low low medium low   
## [1144] high medium high high high medium high low medium  
## [1153] medium low medium low medium medium high low low   
## [1162] low medium low low low medium medium medium medium  
## [1171] high low high high low low medium high medium  
## [1180] medium medium medium high medium medium high low medium  
## [1189] high low medium high medium high high high high   
## [1198] low low medium high low medium medium medium high   
## [1207] low medium medium medium low medium low medium medium  
## [1216] low low medium medium medium medium low medium low   
## [1225] medium low low low low high low low low   
## [1234] low high high medium medium medium medium medium medium  
## [1243] medium medium medium high low medium low high low   
## [1252] low medium low high medium medium low medium high   
## [1261] medium low low low medium high high medium medium  
## [1270] high high medium medium medium high medium low medium  
## [1279] high medium medium high medium medium medium low medium  
## [1288] low medium low medium low low medium medium low   
## [1297] high high medium medium medium low high low low   
## [1306] medium medium medium medium medium medium high high low   
## [1315] high high medium high low low high low medium  
## [1324] medium medium low medium low medium medium low medium  
## [1333] high low high medium medium medium medium low low   
## [1342] low low medium medium low medium low medium low   
## [1351] medium high low medium low high medium medium medium  
## [1360] medium high medium medium high high medium high medium  
## [1369] medium medium high medium medium medium medium high high   
## [1378] low medium medium low high low medium medium medium  
## [1387] medium low low high high medium medium medium medium  
## [1396] medium medium medium high medium low low medium medium  
## [1405] high high high high low low high low low   
## [1414] medium medium low medium high medium medium medium medium  
## [1423] medium medium low high low medium high low medium  
## [1432] high low medium high medium medium low medium medium  
## [1441] low low high medium medium medium medium low high   
## [1450] medium medium medium medium medium low medium low high   
## [1459] medium high high low medium low medium medium medium  
## [1468] medium medium high   
## attr(,"prob")  
## [1] 0.6000000 0.7000000 0.5000000 0.7000000 0.7272727 0.6000000  
## [7] 0.6000000 0.5000000 0.5000000 0.6000000 0.6000000 0.6363636  
## [13] 0.5000000 0.6363636 0.7000000 0.9000000 0.7000000 0.9090909  
## [19] 0.5000000 0.8000000 0.6000000 0.6000000 0.7000000 0.8000000  
## [25] 0.9000000 0.7000000 0.9000000 0.7000000 0.7000000 0.7000000  
## [31] 0.4000000 0.6363636 0.6000000 0.5000000 0.5000000 0.6000000  
## [37] 0.7000000 0.8000000 0.4000000 0.5454545 0.7000000 0.5000000  
## [43] 0.7500000 0.8000000 0.7000000 0.8000000 0.8000000 0.6000000  
## [49] 0.6000000 0.6000000 0.9000000 0.8000000 0.7000000 0.5000000  
## [55] 0.7000000 0.5000000 0.7000000 0.6000000 0.6000000 0.5454545  
## [61] 0.5000000 0.7000000 0.6000000 0.5000000 0.5000000 0.6000000  
## [67] 0.4000000 0.5000000 0.4166667 0.6000000 0.7000000 0.6000000  
## [73] 0.8000000 0.6000000 0.6000000 1.0000000 0.6000000 0.6000000  
## [79] 0.6000000 0.8000000 0.8000000 0.8000000 0.9000000 0.5000000  
## [85] 0.5000000 0.6000000 0.7000000 0.6000000 0.7000000 0.4000000  
## [91] 0.7000000 0.7272727 0.5000000 0.5000000 0.6000000 0.5000000  
## [97] 0.7000000 0.5454545 0.6000000 0.8000000 0.5000000 0.7000000  
## [103] 0.5000000 0.7857143 0.4000000 0.6363636 0.5000000 0.6000000  
## [109] 0.5000000 0.5000000 0.5000000 0.6000000 0.5000000 0.5000000  
## [115] 0.6000000 0.6000000 0.4545455 0.6000000 0.6000000 0.5000000  
## [121] 0.7000000 0.9000000 0.8000000 0.8000000 0.5000000 0.4000000  
## [127] 0.8000000 0.6000000 0.7000000 0.5000000 0.4000000 0.4545455  
## [133] 0.5000000 0.6000000 0.9090909 0.7272727 0.7272727 0.9090909  
## [139] 0.6363636 0.7000000 0.7000000 0.9000000 0.9000000 0.7000000  
## [145] 0.6000000 0.7000000 0.8000000 0.6363636 0.8000000 0.6000000  
## [151] 0.4615385 0.5000000 0.5000000 1.0000000 0.6000000 0.5000000  
## [157] 0.5000000 0.5454545 0.4000000 0.5454545 0.8000000 0.7000000  
## [163] 1.0000000 0.6000000 0.7272727 0.6000000 0.6363636 0.6000000  
## [169] 0.6000000 0.4000000 0.4000000 0.4000000 0.9000000 0.9090909  
## [175] 0.5000000 0.5000000 0.6000000 0.5000000 0.4000000 0.5000000  
## [181] 0.4000000 0.9000000 0.6363636 0.7000000 0.6000000 0.6000000  
## [187] 0.5454545 0.8000000 0.5000000 0.5000000 0.9090909 0.6363636  
## [193] 0.6363636 0.5000000 0.7000000 0.8000000 0.8181818 0.4000000  
## [199] 0.5000000 0.7000000 0.9000000 0.7000000 0.5454545 0.8000000  
## [205] 0.9000000 0.7000000 1.0000000 0.6000000 0.5000000 0.5454545  
## [211] 0.5000000 0.8000000 0.4545455 0.9000000 0.8000000 0.8000000  
## [217] 0.7000000 0.4545455 0.8000000 0.7000000 0.6000000 0.5000000  
## [223] 0.7272727 0.7272727 0.9000000 0.6666667 0.5000000 0.4000000  
## [229] 0.7000000 0.6000000 0.6000000 0.4000000 0.6000000 0.4545455  
## [235] 0.6000000 0.5000000 0.5000000 0.5000000 0.5454545 1.0000000  
## [241] 0.7000000 0.6000000 0.9000000 0.5000000 0.6000000 0.5000000  
## [247] 0.4000000 0.8000000 0.9000000 0.5000000 0.6000000 0.6000000  
## [253] 0.8000000 0.6000000 0.8000000 0.7000000 0.5454545 0.7272727  
## [259] 0.9000000 0.6000000 0.7000000 0.9230769 0.4000000 0.5000000  
## [265] 0.9000000 0.9000000 0.7000000 0.6000000 0.5000000 0.5000000  
## [271] 0.5454545 0.5454545 0.7000000 0.5000000 0.7000000 0.9000000  
## [277] 0.6000000 0.5000000 0.5384615 0.4000000 0.7000000 0.5000000  
## [283] 0.7000000 0.9000000 0.4000000 0.7000000 0.9000000 0.4000000  
## [289] 0.4545455 0.8000000 0.6000000 0.4000000 0.7000000 0.8000000  
## [295] 0.6000000 0.8181818 0.5000000 0.4000000 0.7000000 0.7000000  
## [301] 0.6000000 0.8181818 0.6000000 0.5000000 0.6000000 0.5000000  
## [307] 0.5000000 0.4000000 0.8000000 0.7000000 0.6000000 1.0000000  
## [313] 0.8181818 0.6000000 0.5000000 0.6000000 0.4545455 0.6000000  
## [319] 0.5000000 0.5000000 0.9000000 0.5000000 0.6000000 0.5000000  
## [325] 0.7272727 0.6000000 0.8000000 0.4000000 0.5000000 0.4000000  
## [331] 0.4000000 0.8000000 0.6000000 0.8000000 0.9090909 0.4000000  
## [337] 0.5000000 0.5000000 0.4000000 0.5000000 0.6000000 0.6363636  
## [343] 0.8000000 0.5000000 0.7000000 0.5000000 0.6000000 0.8000000  
## [349] 0.8000000 0.6000000 0.6000000 0.6000000 0.6000000 0.7000000  
## [355] 0.8000000 0.5000000 0.5000000 0.6000000 0.7000000 1.0000000  
## [361] 0.6363636 0.6000000 0.7000000 0.9090909 0.5000000 0.7000000  
## [367] 0.5000000 0.9000000 0.6000000 0.7000000 0.3636364 0.8000000  
## [373] 0.4545455 0.5000000 0.5000000 0.4000000 0.9000000 0.8000000  
## [379] 0.5000000 0.7000000 0.5000000 0.9000000 0.6000000 0.9000000  
## [385] 0.7000000 0.5000000 0.7000000 0.7000000 0.9000000 0.7000000  
## [391] 0.5000000 0.7000000 0.5000000 0.8000000 0.6000000 0.8000000  
## [397] 0.8000000 0.7000000 0.8000000 0.7000000 0.8000000 0.5000000  
## [403] 0.7000000 0.8000000 0.6000000 0.5000000 0.5000000 0.7000000  
## [409] 1.0000000 0.7000000 0.7000000 0.6000000 0.6000000 0.5000000  
## [415] 0.5000000 0.7000000 0.4000000 0.5454545 0.5000000 0.5000000  
## [421] 0.5454545 0.5454545 0.6000000 0.5000000 0.7000000 0.8000000  
## [427] 0.6000000 0.6000000 1.0000000 0.5000000 0.5000000 0.6000000  
## [433] 0.6000000 0.5000000 0.8000000 0.6000000 0.4000000 0.8000000  
## [439] 0.6000000 0.4545455 0.4000000 0.4000000 0.5454545 0.7272727  
## [445] 0.5000000 0.7000000 0.4000000 0.5000000 0.6000000 0.7272727  
## [451] 0.7000000 0.4000000 0.8000000 0.5000000 0.6363636 0.5000000  
## [457] 0.8000000 0.5000000 0.6000000 0.7000000 0.8181818 0.5000000  
## [463] 0.7000000 0.8000000 0.5000000 0.7000000 0.7000000 0.6000000  
## [469] 0.5000000 0.5000000 0.7000000 0.6000000 0.6000000 0.5000000  
## [475] 0.6000000 0.6000000 0.5454545 0.6000000 0.7000000 0.5000000  
## [481] 0.5000000 0.5000000 0.6363636 0.6000000 0.6000000 0.7000000  
## [487] 0.8000000 0.6000000 1.0000000 0.5000000 0.5000000 0.7000000  
## [493] 0.7000000 0.8000000 0.8181818 0.6000000 0.7000000 0.7000000  
## [499] 0.5000000 0.5000000 0.5833333 0.6000000 0.6363636 0.6000000  
## [505] 0.6000000 0.6000000 0.9000000 0.5000000 0.6000000 0.5000000  
## [511] 0.5000000 0.4000000 0.8000000 0.5000000 0.5000000 0.5000000  
## [517] 0.5000000 0.6000000 0.7000000 0.4000000 0.6000000 0.4000000  
## [523] 0.5000000 0.5000000 0.6000000 0.8000000 0.4000000 0.5000000  
## [529] 0.9000000 0.9000000 0.5454545 0.5000000 0.4545455 0.5000000  
## [535] 0.6000000 0.4000000 0.5000000 0.6000000 0.4000000 0.6000000  
## [541] 0.9090909 0.6000000 0.8000000 0.6000000 0.6000000 0.6000000  
## [547] 0.5000000 0.8000000 0.6000000 0.5000000 0.6000000 0.4000000  
## [553] 0.8000000 0.4000000 0.7000000 0.7272727 0.6000000 0.5000000  
## [559] 0.5000000 0.4545455 0.8000000 0.6000000 0.5000000 0.6000000  
## [565] 0.7000000 0.6000000 0.5000000 0.8000000 0.6000000 0.7000000  
## [571] 0.7000000 0.8000000 1.0000000 0.7000000 0.8000000 0.5000000  
## [577] 0.6000000 0.7272727 0.6000000 0.5000000 0.5000000 0.5454545  
## [583] 0.5454545 0.5454545 0.5000000 0.8000000 0.6363636 0.4000000  
## [589] 0.6000000 1.0000000 0.7000000 0.6000000 0.6000000 0.6000000  
## [595] 0.6000000 0.7000000 0.8000000 0.5000000 0.6000000 0.6000000  
## [601] 0.9000000 0.6000000 0.6000000 0.6000000 0.6000000 0.5000000  
## [607] 0.6000000 0.6363636 0.8000000 0.8000000 0.7000000 0.8000000  
## [613] 0.6000000 0.7000000 0.5000000 0.5000000 0.7000000 0.6000000  
## [619] 0.8000000 0.6000000 0.6000000 0.7000000 0.5454545 0.5000000  
## [625] 0.7000000 0.6000000 0.6000000 0.5000000 0.5000000 0.8000000  
## [631] 0.8000000 0.9000000 0.6000000 0.7000000 0.5000000 0.8000000  
## [637] 0.4000000 0.7000000 1.0000000 0.6000000 0.6000000 0.7000000  
## [643] 0.7000000 0.4000000 0.5000000 0.5454545 0.4000000 0.5000000  
## [649] 0.6000000 0.6000000 0.5000000 0.6000000 0.4000000 0.4000000  
## [655] 0.8000000 0.5000000 0.4545455 0.7000000 0.7000000 0.5000000  
## [661] 0.5454545 0.7000000 0.6000000 0.6000000 0.6363636 0.4545455  
## [667] 0.9090909 0.5000000 0.7000000 0.8000000 0.9090909 0.6363636  
## [673] 0.4000000 0.9000000 0.6000000 0.6363636 0.6000000 0.5000000  
## [679] 0.4000000 0.6000000 0.4545455 0.6000000 0.7000000 0.6000000  
## [685] 0.4000000 0.9000000 0.5000000 0.8000000 0.3636364 0.5000000  
## [691] 0.5000000 0.6000000 0.6000000 1.0000000 0.9000000 0.6000000  
## [697] 0.8000000 0.8000000 0.5000000 0.6363636 0.6000000 0.4000000  
## [703] 0.7000000 0.7272727 0.5000000 0.5000000 0.5833333 0.6000000  
## [709] 0.7000000 0.7000000 0.6363636 0.6000000 0.6000000 0.4000000  
## [715] 0.7272727 0.7000000 0.6000000 1.0000000 0.4000000 0.6000000  
## [721] 0.4000000 0.7000000 0.8000000 0.8000000 0.7000000 0.4000000  
## [727] 0.8000000 0.6000000 0.6363636 0.8000000 0.8000000 0.7000000  
## [733] 0.5000000 0.6000000 1.0000000 0.6000000 0.5000000 0.4000000  
## [739] 0.5000000 0.5000000 0.5000000 0.5833333 0.9000000 0.8181818  
## [745] 0.9000000 0.8000000 0.6000000 0.6000000 0.6000000 0.8000000  
## [751] 0.7000000 0.7000000 0.5454545 0.3636364 0.6000000 0.5454545  
## [757] 0.5000000 0.5000000 0.7000000 0.8181818 0.5000000 0.8181818  
## [763] 0.7000000 0.5000000 0.5000000 0.8000000 1.0000000 0.4000000  
## [769] 0.5000000 0.6000000 0.6000000 0.8181818 0.7000000 0.6000000  
## [775] 0.5833333 0.7000000 0.7000000 0.8000000 0.5000000 0.4000000  
## [781] 0.8181818 0.5000000 0.6363636 0.4000000 0.6000000 0.6000000  
## [787] 0.6000000 0.6000000 0.4545455 0.4545455 0.7000000 0.7000000  
## [793] 0.6000000 0.7692308 0.6000000 0.9000000 0.4000000 0.9000000  
## [799] 0.8000000 0.6000000 0.5000000 0.6000000 0.4000000 0.6000000  
## [805] 0.5000000 0.6000000 0.9000000 0.7000000 0.7272727 0.8000000  
## [811] 0.5000000 0.9000000 0.6000000 0.5000000 0.5000000 0.5000000  
## [817] 0.8000000 0.7000000 0.8000000 0.6000000 0.5000000 0.6000000  
## [823] 0.5000000 0.5000000 0.8000000 0.4000000 0.7000000 0.7000000  
## [829] 0.5000000 0.3636364 0.3636364 0.5000000 0.5000000 0.3636364  
## [835] 0.7000000 0.7000000 0.8000000 0.7000000 0.4000000 0.7000000  
## [841] 0.5000000 0.5000000 0.7000000 0.5454545 0.5454545 0.6000000  
## [847] 0.9000000 0.5000000 0.9000000 0.6000000 0.6000000 0.6000000  
## [853] 0.5000000 0.5454545 0.6000000 0.6000000 0.5454545 0.6000000  
## [859] 0.5000000 0.5000000 0.7272727 0.6000000 0.7500000 0.5454545  
## [865] 0.7000000 0.7000000 0.5454545 0.4000000 0.5000000 0.6000000  
## [871] 0.6000000 0.6363636 0.5000000 0.5000000 0.6000000 0.6363636  
## [877] 1.0000000 0.5000000 0.7272727 0.4000000 0.8000000 0.6000000  
## [883] 0.6000000 0.6000000 0.7000000 0.8000000 0.6000000 0.6000000  
## [889] 0.6000000 0.5000000 0.6000000 0.5000000 0.4000000 0.6000000  
## [895] 0.5000000 0.6000000 0.8000000 0.7000000 0.5000000 0.6000000  
## [901] 0.9090909 0.5000000 0.7000000 0.4000000 0.5000000 0.6000000  
## [907] 0.6000000 0.7000000 0.8000000 0.8000000 0.7000000 0.7000000  
## [913] 0.6000000 0.5000000 0.5000000 0.6000000 0.7000000 0.6000000  
## [919] 0.7000000 0.8000000 0.8000000 0.8000000 0.5000000 0.5000000  
## [925] 0.5000000 0.9000000 1.0000000 0.9000000 0.7000000 0.5000000  
## [931] 0.8000000 0.5000000 0.8000000 0.5000000 0.6363636 0.9000000  
## [937] 1.0000000 0.5454545 0.7000000 0.5000000 0.5000000 0.7000000  
## [943] 0.5000000 0.7000000 0.5000000 0.7000000 0.7000000 0.6000000  
## [949] 0.5000000 0.5000000 0.6363636 0.9090909 0.4000000 1.0000000  
## [955] 0.7000000 0.6000000 0.8000000 0.5000000 0.7000000 0.7000000  
## [961] 0.7000000 0.4000000 0.7272727 0.4000000 0.7272727 0.5000000  
## [967] 0.5000000 1.0000000 1.0000000 0.6000000 0.6000000 0.5000000  
## [973] 0.5000000 1.0000000 0.7272727 0.5000000 0.5000000 0.5000000  
## [979] 0.8181818 0.8000000 0.6000000 0.6000000 0.6000000 0.8000000  
## [985] 0.7000000 0.9000000 0.7000000 0.7000000 0.5000000 0.7272727  
## [991] 0.6000000 0.6000000 0.6000000 0.7000000 0.6000000 0.6000000  
## [997] 0.8181818 0.6363636 0.5000000 0.6000000 0.6000000 0.4000000  
## [1003] 0.6000000 0.5000000 0.5000000 0.7000000 0.5000000 0.4000000  
## [1009] 0.6000000 0.7000000 0.6000000 0.6000000 0.5714286 0.7000000  
## [1015] 0.6363636 0.4000000 0.6000000 0.9000000 0.8000000 0.6000000  
## [1021] 0.8181818 0.9000000 0.8181818 0.6000000 0.7000000 0.7000000  
## [1027] 0.6000000 0.7000000 0.9000000 0.5000000 0.6000000 0.5000000  
## [1033] 0.6000000 0.8000000 0.5000000 0.5000000 0.6000000 0.6000000  
## [1039] 0.9000000 0.6000000 0.6000000 0.7000000 0.5000000 0.7000000  
## [1045] 0.5000000 0.8000000 1.0000000 0.5000000 0.5000000 0.6000000  
## [1051] 0.5000000 0.8000000 0.7272727 0.6000000 0.5000000 1.0000000  
## [1057] 0.6000000 0.6000000 0.8000000 0.9000000 0.5454545 0.6000000  
## [1063] 0.7000000 1.0000000 0.9000000 0.4000000 0.6363636 0.7000000  
## [1069] 0.5000000 0.7000000 0.9000000 0.7000000 0.5454545 0.8000000  
## [1075] 0.7000000 0.8000000 0.7000000 0.7000000 0.7000000 0.5454545  
## [1081] 0.7000000 1.0000000 0.7000000 0.6000000 0.7000000 0.6363636  
## [1087] 0.6000000 0.7000000 0.6000000 0.8000000 0.7000000 0.7000000  
## [1093] 0.6363636 0.6000000 1.0000000 0.6000000 0.6000000 0.5000000  
## [1099] 0.5454545 0.6000000 0.4000000 0.6000000 0.5454545 0.6000000  
## [1105] 0.8181818 0.9000000 0.5000000 0.5000000 0.6000000 0.6000000  
## [1111] 0.8000000 0.8000000 0.7000000 0.6666667 0.5000000 0.8000000  
## [1117] 0.4000000 0.6000000 0.7000000 0.6000000 0.8000000 0.8000000  
## [1123] 0.5000000 0.6000000 0.7000000 0.9000000 0.9000000 0.6000000  
## [1129] 0.7000000 0.7000000 0.5000000 0.6000000 0.9000000 0.4545455  
## [1135] 0.6000000 0.7000000 0.5000000 0.8181818 0.5000000 0.6000000  
## [1141] 0.9000000 0.8000000 0.6000000 0.7000000 0.6000000 0.6000000  
## [1147] 0.5384615 1.0000000 0.5000000 0.8000000 0.5454545 0.5000000  
## [1153] 0.7000000 1.0000000 0.5000000 1.0000000 0.8181818 0.7000000  
## [1159] 0.7000000 0.5000000 0.9000000 0.9000000 0.7000000 0.6000000  
## [1165] 0.5000000 0.6000000 0.7500000 0.8000000 0.6000000 0.5000000  
## [1171] 0.6000000 0.6000000 0.8000000 0.7000000 0.7272727 0.7272727  
## [1177] 0.6000000 0.9000000 0.7000000 0.6000000 0.8000000 0.5384615  
## [1183] 0.4000000 0.5384615 0.6000000 0.4000000 0.8181818 0.5000000  
## [1189] 0.5000000 0.5000000 0.6000000 1.0000000 0.6000000 0.6000000  
## [1195] 0.8181818 0.7000000 0.5000000 0.5000000 0.6000000 0.9000000  
## [1201] 0.6000000 0.7000000 0.4545455 0.6000000 0.6000000 0.7000000  
## [1207] 0.5000000 0.6000000 0.7000000 0.7000000 0.5000000 0.5000000  
## [1213] 0.8000000 0.6666667 0.6000000 0.8000000 0.5000000 0.5454545  
## [1219] 0.8000000 0.5000000 0.5000000 0.6000000 0.5000000 0.7000000  
## [1225] 1.0000000 0.9000000 1.0000000 0.9090909 1.0000000 0.6000000  
## [1231] 0.8000000 0.6000000 0.6000000 0.6666667 1.0000000 1.0000000  
## [1237] 0.5000000 0.5000000 0.6000000 0.5000000 0.6000000 0.6000000  
## [1243] 0.5000000 0.6000000 0.5000000 0.4666667 0.6000000 0.4375000  
## [1249] 0.6000000 0.9000000 0.6000000 0.7000000 0.8181818 0.8000000  
## [1255] 0.6000000 0.7000000 0.6000000 0.5000000 0.6000000 0.4000000  
## [1261] 0.5000000 0.7272727 0.5454545 0.7272727 0.6000000 0.4000000  
## [1267] 0.6000000 0.6000000 0.7272727 1.0000000 1.0000000 0.7000000  
## [1273] 0.7000000 0.6000000 0.8000000 0.6000000 0.7000000 0.5000000  
## [1279] 0.8000000 0.5000000 0.7272727 0.5000000 0.9090909 0.4000000  
## [1285] 0.6000000 0.4000000 0.7000000 0.9000000 0.5000000 0.8000000  
## [1291] 0.7000000 0.7000000 0.6000000 0.6000000 0.6000000 0.7000000  
## [1297] 0.5000000 0.5000000 0.6000000 0.6000000 0.6000000 0.6000000  
## [1303] 0.6153846 0.7000000 0.6000000 0.5454545 0.4000000 0.6000000  
## [1309] 0.7000000 0.7000000 0.6000000 0.6000000 0.6000000 0.8000000  
## [1315] 1.0000000 0.6000000 0.5000000 0.5000000 0.6666667 0.8000000  
## [1321] 0.6363636 0.8000000 0.7000000 0.8000000 0.6000000 0.8000000  
## [1327] 0.6363636 1.0000000 0.6000000 0.7000000 0.5000000 0.5000000  
## [1333] 0.5000000 0.8181818 0.4000000 0.7272727 0.6000000 0.6000000  
## [1339] 0.5000000 0.4000000 0.4000000 0.9090909 0.8181818 1.0000000  
## [1345] 0.7000000 0.6000000 0.7000000 0.9000000 0.7000000 0.8000000  
## [1351] 0.4000000 0.7000000 0.7000000 0.6000000 0.7000000 0.7000000  
## [1357] 0.6000000 0.6000000 0.6000000 0.5000000 0.7000000 0.8000000  
## [1363] 0.8000000 0.4000000 0.6363636 0.5000000 0.7000000 0.9000000  
## [1369] 0.9000000 0.6000000 0.6363636 0.8000000 0.6000000 0.7000000  
## [1375] 0.8000000 0.5833333 0.5000000 0.5000000 0.5000000 0.7000000  
## [1381] 0.5000000 0.6363636 0.6363636 0.7000000 0.5000000 0.6666667  
## [1387] 0.6000000 0.7000000 0.7272727 0.4000000 0.8000000 0.5000000  
## [1393] 0.5000000 0.6000000 0.5454545 0.4166667 0.6000000 0.5000000  
## [1399] 0.5000000 0.5000000 0.8000000 0.7000000 0.6363636 0.6363636  
## [1405] 0.8000000 0.6000000 0.5000000 0.7000000 0.8000000 0.6000000  
## [1411] 0.4000000 0.6000000 0.8000000 0.5000000 0.6000000 0.8000000  
## [1417] 0.6000000 0.5000000 0.6363636 0.6363636 0.4000000 0.6000000  
## [1423] 0.7000000 0.6000000 0.6000000 0.8181818 1.0000000 0.8000000  
## [1429] 0.6000000 0.4000000 0.7000000 0.9000000 0.5000000 0.5000000  
## [1435] 0.7000000 0.9000000 0.6000000 0.6000000 0.6000000 0.8000000  
## [1441] 0.9000000 0.8000000 0.9000000 0.9000000 0.5000000 0.8181818  
## [1447] 0.8000000 0.4000000 0.7000000 0.6000000 0.5000000 0.8000000  
## [1453] 0.8000000 0.6000000 0.8000000 0.7000000 0.8000000 0.5000000  
## [1459] 0.9000000 0.5000000 0.5000000 0.6000000 0.5000000 0.6000000  
## [1465] 0.6363636 0.6000000 0.9000000 0.5000000 0.6000000 0.7000000  
## Levels: high low medium

Evaluate the model performance.

CrossTable(x=winedata\_test\_labels, y=winedata\_test\_pred, prop.chisq=FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1470   
##   
##   
## | winedata\_test\_pred   
## winedata\_test\_labels | high | low | medium | Row Total |   
## ---------------------|-----------|-----------|-----------|-----------|  
## high | 143 | 28 | 126 | 297 |   
## | 0.481 | 0.094 | 0.424 | 0.202 |   
## | 0.513 | 0.056 | 0.182 | |   
## | 0.097 | 0.019 | 0.086 | |   
## ---------------------|-----------|-----------|-----------|-----------|  
## low | 22 | 308 | 173 | 503 |   
## | 0.044 | 0.612 | 0.344 | 0.342 |   
## | 0.079 | 0.618 | 0.250 | |   
## | 0.015 | 0.210 | 0.118 | |   
## ---------------------|-----------|-----------|-----------|-----------|  
## medium | 114 | 162 | 394 | 670 |   
## | 0.170 | 0.242 | 0.588 | 0.456 |   
## | 0.409 | 0.325 | 0.569 | |   
## | 0.078 | 0.110 | 0.268 | |   
## ---------------------|-----------|-----------|-----------|-----------|  
## Column Total | 279 | 498 | 693 | 1470 |   
## | 0.190 | 0.339 | 0.471 | |   
## ---------------------|-----------|-----------|-----------|-----------|  
##   
##

k <- 1:10  
for(x in k){  
 winedata\_test\_pred <- knn(winedata\_train, winedata\_test,  
 winedata\_train\_labels, k = x)  
 accuracy <- mean(winedata\_test\_pred == winedata\_test\_labels)  
 cat('\n','When k =', x, "accuracy is ", accuracy)  
}

##   
## When k = 1 accuracy is 0.6605442  
## When k = 2 accuracy is 0.5721088  
## When k = 3 accuracy is 0.5857143  
## When k = 4 accuracy is 0.5884354  
## When k = 5 accuracy is 0.5836735  
## When k = 6 accuracy is 0.5795918  
## When k = 7 accuracy is 0.5857143  
## When k = 8 accuracy is 0.570068  
## When k = 9 accuracy is 0.577551  
## When k = 10 accuracy is 0.5748299

As we can see, the model has the highest accuracy of 66% when k=1.Then the next highest accuracy is 60% when k=3 and k=5. When we repeat the process for k from 1 to 50 we see some decrease and then increase.

accuracy <- rep(0, 50)  
k <- 1:50  
for(x in k){  
 winedata\_test\_pred <- knn(winedata\_train, winedata\_test,  
 winedata\_train\_labels, k = x)  
 accuracy[x] <- mean(winedata\_test\_pred == winedata\_test\_labels)  
 }  
plot(k, accuracy, type = 'b')

