# Italian Wine in the US Market

## Authors: Andrea Salusso mat. 885032, Niklas Selch 887250, Giulio Cico 887723, David Burca 887725, Mattia Donadei 884911

Italy World Wine, an Italian wine company, needs to improve its marketing methods for selling wine in the United States. In order to increase its performance in the same market, the company aims to acquire a strategic orientation that is more suited to the needs of the US market. To this end, they commissioned a study to determine how several sorts of elements, such as intrinsic and extrinsic product attributes, habits and consumption scenarios, consumer perceptions, and demographic variables, influence the inclination to buy Italian wine on the American market.

knitr::opts\_chunk$set(error = TRUE)

So first of all we load the dataset and exclude all the variables with na as output so that we are left off with just variables with values. We order the rows starting from 1 to the n value of the variables in the original dataset.

wine <- read.table("wine.txt", header = T)  
download <- na.exclude(wine)  
rownames(download) <- 1:nrow(download)  
  
wine<-download  
datawine<-download  
ori<-download  
n<-nrow(wine)

Question 1: Is There a difference in the way consumers perceive Italian wines and other countries’ wines?

We start by creating a binomial variable “div” dividing our into people who prefer to drink Italian Wine and those who drink non-italian Wine. If we compare the answers of the two groups on general Questions. We can understand where there is the difference in perception

s<-rep(0,n)  
  
  
s[ori[,31]!=1]<-"Other Buy"  
s[ori[,31]==1]<-"IT Buy"

Now we need to create factors that we are going to compare

#For you drinking wine signifies  
meaning <- ori[,9]   
meaning[meaning==1]<- "Traveleing"  
meaning[meaning==2]<- "relaxing"  
meaning[meaning==3]<- "romatic"  
meaning[meaning==4]<- "fun"  
meaning[meaning==5]<- "eating"  
meaning[meaning==6]<- "entering cultures"  
meaning[meaning==7]<- "interlectual satisfaction"  
meaning[meaning==8]<- "sensory enjoyment"  
meaning <- factor(meaning)  
wine[,9]<- meaning  
  
# What is your perception of wine as a commercial product?  
  
commer <- ori[,10]  
commer[commer==1]<- "for an elite"  
commer[commer==2]<- "for everybody"  
commer[commer==3]<- "tradional not for young people"  
commer <- factor(commer)  
wine[,10]<- commer  
  
#How much money do you spend on your wine   
money <- ori[,54]  
money[money == 1] <- "7-14"  
money[money == 2] <- "15-21"  
money[money == 3] <- "over"  
money <- factor(money)  
wine[,54] <- money  
  
#which are the important feature while purchasing wine  
  
#place of origin  
origin <- ori[,11]  
origin <- factor(origin)  
wine[,11] <- origin  
  
#quality   
qua <- ori[,15]  
qua <- factor(qua)  
wine[,15] <- qua  
  
#wine brand   
brand <- ori[,12]  
brand <- factor(brand)  
wine[,12] <- brand  
  
#I tend to buy the wines I like and know and very rarely try new ones  
new <- ori[,25]  
new <- factor(new)  
wine[25] <- new  
  
#You judge a wines Italian provenance as?  
jud <- ori[,35]  
jud[jud==1]<- "guarantee of its quality"  
jud[jud ==2]<- "prompts to buy"  
jud[jud==3]<- "irrelevant"  
jud[jud==4]<- "deters from buying"  
jud <- factor(jud)  
wine[,35] <- jud  
  
#compared to american wine italian is?  
comp <- ori[,36]  
comp[comp==1] <-"higher quality"  
comp[comp==2] <-"lower quality"  
comp[comp==3] <-"same quality"  
comp<- factor(comp)  
wine[,36] <- comp  
  
#do you know italian Wine?  
know <- ori[,32]  
know [know==1]<- "yes"  
know [know==2]<- "no"  
know<- factor(know)  
wine[,32] <- know

Now we are able to compare the answers to the different questions. In order to understand the differences in Perception of Wine in the General

Do you know any Italian Wine?

tab1<-table(s,know)  
tab1.prop<-round(prop.table(tab1),3)   
tab1.prop1<-round(prop.table(tab1,1),3)  
tab1.prop2<-round(prop.table(tab1,2),3)  
  
res<-list(addmargins(tab1),  
 addmargins(tab1.prop\*100),  
 addmargins(tab1.prop1\*100,2),  
 addmargins(tab1.prop2\*100,1))  
res

## [[1]]  
## know  
## s no yes Sum  
## IT Buy 17 146 163  
## Other Buy 59 158 217  
## Sum 76 304 380  
##   
## [[2]]  
## know  
## s no yes Sum  
## IT Buy 4.5 38.4 42.9  
## Other Buy 15.5 41.6 57.1  
## Sum 20.0 80.0 100.0  
##   
## [[3]]  
## know  
## s no yes Sum  
## IT Buy 10.4 89.6 100.0  
## Other Buy 27.2 72.8 100.0  
##   
## [[4]]  
## know  
## s no yes  
## IT Buy 22.4 48.0  
## Other Buy 77.6 52.0  
## Sum 100.0 100.0

barchartGC(~know+s, ylim=c(0,320))

## Error in barchartGC(~know + s, ylim = c(0, 320)): could not find function "barchartGC"

The result on the Question is like expected, by comparing the answers in the proportion table, we can see that people who drink Italian have a higher proportion of knowing any Italian wine. Also, there is a small proportion of 17 people who drink italian wine.

Let’s have a look at the way how the different groups choose which wine they buy.

tab1<-table(s,qua) #quality   
tab1.prop<-round(prop.table(tab1),3)   
tab1.prop1<-round(prop.table(tab1,1),3)  
tab1.prop2<-round(prop.table(tab1,2),3)  
  
res<-list(addmargins(tab1),  
 addmargins(tab1.prop\*100),  
 addmargins(tab1.prop1\*100,2),  
 addmargins(tab1.prop2\*100,1))  
res

## [[1]]  
## qua  
## s 1 3 4 5 Sum  
## IT Buy 1 5 51 106 163  
## Other Buy 2 18 58 139 217  
## Sum 3 23 109 245 380  
##   
## [[2]]  
## qua  
## s 1 3 4 5 Sum  
## IT Buy 0.3 1.3 13.4 27.9 42.9  
## Other Buy 0.5 4.7 15.3 36.6 57.1  
## Sum 0.8 6.0 28.7 64.5 100.0  
##   
## [[3]]  
## qua  
## s 1 3 4 5 Sum  
## IT Buy 0.6 3.1 31.3 65.0 100.0  
## Other Buy 0.9 8.3 26.7 64.1 100.0  
##   
## [[4]]  
## qua  
## s 1 3 4 5  
## IT Buy 33.3 21.7 46.8 43.3  
## Other Buy 66.7 78.3 53.2 56.7  
## Sum 100.0 100.0 100.0 100.0

tab1<-table(s,origin)  
tab1.prop<-round(prop.table(tab1),3)   
tab1.prop1<-round(prop.table(tab1,1),3)  
tab1.prop2<-round(prop.table(tab1,2),3)  
  
res<-list(addmargins(tab1),  
 addmargins(tab1.prop\*100),  
 addmargins(tab1.prop1\*100,2),  
 addmargins(tab1.prop2\*100,1))  
res

## [[1]]  
## origin  
## s 1 2 3 4 5 Sum  
## IT Buy 13 10 43 51 46 163  
## Other Buy 23 34 49 63 48 217  
## Sum 36 44 92 114 94 380  
##   
## [[2]]  
## origin  
## s 1 2 3 4 5 Sum  
## IT Buy 3.4 2.6 11.3 13.4 12.1 42.8  
## Other Buy 6.1 8.9 12.9 16.6 12.6 57.1  
## Sum 9.5 11.5 24.2 30.0 24.7 99.9  
##   
## [[3]]  
## origin  
## s 1 2 3 4 5 Sum  
## IT Buy 8.0 6.1 26.4 31.3 28.2 100.0  
## Other Buy 10.6 15.7 22.6 29.0 22.1 100.0  
##   
## [[4]]  
## origin  
## s 1 2 3 4 5  
## IT Buy 36.1 22.7 46.7 44.7 48.9  
## Other Buy 63.9 77.3 53.3 55.3 51.1  
## Sum 100.0 100.0 100.0 100.0 100.0

tab1<-table(s,brand)  
tab1.prop<-round(prop.table(tab1),3)   
tab1.prop1<-round(prop.table(tab1,1),3)  
tab1.prop2<-round(prop.table(tab1,2),3)  
  
res<-list(addmargins(tab1),  
 addmargins(tab1.prop\*100),  
 addmargins(tab1.prop1\*100,2),  
 addmargins(tab1.prop2\*100,1))  
res

## [[1]]  
## brand  
## s 1 2 3 4 5 Sum  
## IT Buy 17 25 57 49 15 163  
## Other Buy 33 47 56 60 21 217  
## Sum 50 72 113 109 36 380  
##   
## [[2]]  
## brand  
## s 1 2 3 4 5 Sum  
## IT Buy 4.5 6.6 15.0 12.9 3.9 42.9  
## Other Buy 8.7 12.4 14.7 15.8 5.5 57.1  
## Sum 13.2 19.0 29.7 28.7 9.4 100.0  
##   
## [[3]]  
## brand  
## s 1 2 3 4 5 Sum  
## IT Buy 10.4 15.3 35.0 30.1 9.2 100.0  
## Other Buy 15.2 21.7 25.8 27.6 9.7 100.0  
##   
## [[4]]  
## brand  
## s 1 2 3 4 5  
## IT Buy 34.0 34.7 50.4 45.0 41.7  
## Other Buy 66.0 65.3 49.6 55.0 58.3  
## Sum 100.0 100.0 100.0 100.0 100.0

Looking on these numbers we can clearly see, that the group of Italian wine drinkers. Values origin and quality are higher when choosing a wine. In terms of brand, the numbers are more or less the same.

For you drinking wine signifies?

tab1<-table(s,meaning)  
tab1.prop<-round(prop.table(tab1),3)   
tab1.prop1<-round(prop.table(tab1,1),3)  
tab1.prop2<-round(prop.table(tab1,2),3)  
  
res<-list(addmargins(tab1),  
 addmargins(tab1.prop\*100),  
 addmargins(tab1.prop1\*100,2),  
 addmargins(tab1.prop2\*100,1))  
barchartGC(~meaning+s, ylim=c(0,300))

## Error in barchartGC(~meaning + s, ylim = c(0, 300)): could not find function "barchartGC"

You judge a wine’s Italian provenance as…

tab1<-table(s,jud)  
  
tab1.prop<-round(prop.table(tab1),3)   
tab1.prop1<-round(prop.table(tab1,1),3)  
tab1.prop2<-round(prop.table(tab1,2),3)  
  
res<-list(addmargins(tab1),  
 addmargins(tab1.prop\*100),  
 addmargins(tab1.prop1\*100,2),  
 addmargins(tab1.prop2\*100,1))  
res

## [[1]]  
## jud  
## s guarantee of its quality irrelevant prompts to buy Sum  
## IT Buy 57 33 73 163  
## Other Buy 50 77 90 217  
## Sum 107 110 163 380  
##   
## [[2]]  
## jud  
## s guarantee of its quality irrelevant prompts to buy Sum  
## IT Buy 15.0 8.7 19.2 42.9  
## Other Buy 13.2 20.3 23.7 57.2  
## Sum 28.2 29.0 42.9 100.1  
##   
## [[3]]  
## jud  
## s guarantee of its quality irrelevant prompts to buy Sum  
## IT Buy 35.0 20.2 44.8 100.0  
## Other Buy 23.0 35.5 41.5 100.0  
##   
## [[4]]  
## jud  
## s guarantee of its quality irrelevant prompts to buy  
## IT Buy 53.3 30.0 44.8  
## Other Buy 46.7 70.0 55.2  
## Sum 100.0 100.0 100.0

barchartGC(~jud, ylim=c(0,180))

## Error in barchartGC(~jud, ylim = c(0, 180)): could not find function "barchartGC"

barchartGC(~jud+s, ylim=c(0,180))

## Error in barchartGC(~jud + s, ylim = c(0, 180)): could not find function "barchartGC"

proptestGC(~s+jud, success = "prompts to buy", p=0, graph = T,first = "IT Buy", alternative = "two.sided" )

## Error in proptestGC(~s + jud, success = "prompts to buy", p = 0, graph = T, : could not find function "proptestGC"

By having an look on this we see that the P-value of interference is quite high. So we can say that the proportion of consumers that are prompt to buy italian wine, is equally dived over the two groups

proptestGC(~s+jud, success = "guarantee of its quality", p=0, graph = T, first = "IT Buy", alternative = "two.sided" )

## Error in proptestGC(~s + jud, success = "guarantee of its quality", p = 0, : could not find function "proptestGC"

By performing the same test on the guarantee of its quality section. We can see that the P-value is low, so the proportion of consumers is higher among the usual italian wine drinkers

Last but not least, we can also see that on large italian wine drinkers spend more money on wine

tab1<-table(s,money)  
  
tab1.prop<-round(prop.table(tab1),3)   
tab1.prop1<-round(prop.table(tab1,1),3)  
tab1.prop2<-round(prop.table(tab1,2),3)  
  
res<-list(addmargins(tab1),  
 addmargins(tab1.prop\*100),  
 addmargins(tab1.prop1\*100,2),  
 addmargins(tab1.prop2\*100,1))  
res

## [[1]]  
## money  
## s 15-21 7-14 over Sum  
## IT Buy 41 4 118 163  
## Other Buy 67 9 141 217  
## Sum 108 13 259 380  
##   
## [[2]]  
## money  
## s 15-21 7-14 over Sum  
## IT Buy 10.8 1.1 31.1 43.0  
## Other Buy 17.6 2.4 37.1 57.1  
## Sum 28.4 3.5 68.2 100.1  
##   
## [[3]]  
## money  
## s 15-21 7-14 over Sum  
## IT Buy 25.2 2.5 72.4 100.1  
## Other Buy 30.9 4.1 65.0 100.0  
##   
## [[4]]  
## money  
## s 15-21 7-14 over  
## IT Buy 38.0 30.8 45.6  
## Other Buy 62.0 69.2 54.4  
## Sum 100.0 100.0 100.0

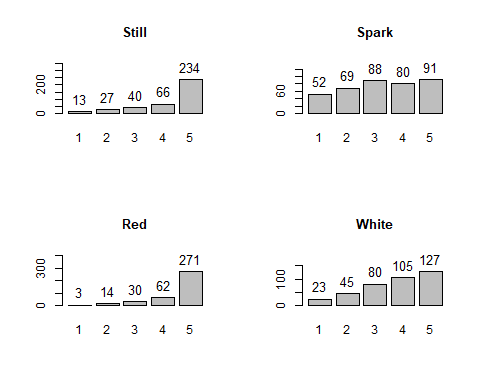
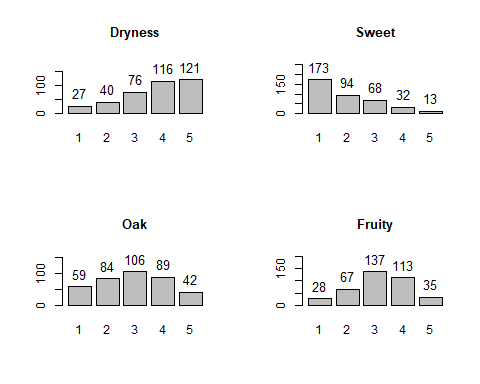
barchartGC(~money, ylim=c(0,180))

## Error in barchartGC(~money, ylim = c(0, 180)): could not find function "barchartGC"

Question 2: What are the key factors in obtaining success on the American market? (developing a strong brand? Red wines? White wines?, wines with particular characteristics?)

First we load the and remove all the NA data

par(mfrow=c(2,2),cex=0.8,cex.main=1,cex.axis=0.9)  
for (i in 36:43){  
freq<-table(datawine[,i+1])  
bp<-barplot(freq,ylim=c(0,max(freq)\*(1+0.5)),  
main=colnames(datawine)[i+1])  
text(bp, freq, labels=freq,pos=3)  
}



par(mfrow=c(2,2),cex=0.6,cex.main=1, cex.axis=0.9)

dataw<-na.exclude(datawine)  
rownames(dataw)<-1:nrow(dataw)  
data<-dataw  
original<-dataw  
n<-nrow(data)

We use the generalized linear model to create a linear regression using a binomial variable of success ##Success on Quality and willingness to buy We create a binomial variable success based on v35 that report 1 if the respondent chose “Superior quality” or “prompt to buy”, and 0 for the others

win<-rep(0,n)  
win[original[,35] == 1 || original[,35] ==2]<-1  
  
win[original[,35] ==3 ]<-0  
win<-factor(win)

I used the generalized linear model, which is a more flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution.

lmodel<-glm(win~data[,11] + data[,12]+ data[,13] + data[, 14] + data[,15] + data[, 16] + data[, 17] + data[, 18] + data[,19] + data[,20]+ data[,21] + data[,22] + data[, 23] + data[, 24] + data[,37]+ data[,38] + data[, 39] + data[, 40] + data[, 41] + data[, 42] + data[, 43] + data[, 44],family=binomial(link="logit"))  
summary(lmodel)

##   
## Call:  
## glm(formula = win ~ data[, 11] + data[, 12] + data[, 13] + data[,   
## 14] + data[, 15] + data[, 16] + data[, 17] + data[, 18] +   
## data[, 19] + data[, 20] + data[, 21] + data[, 22] + data[,   
## 23] + data[, 24] + data[, 37] + data[, 38] + data[, 39] +   
## data[, 40] + data[, 41] + data[, 42] + data[, 43] + data[,   
## 44], family = binomial(link = "logit"))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.6450 -0.6441 0.4814 0.7157 2.2620   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.301597 1.499978 -0.201 0.8406   
## data[, 11] 0.159547 0.144307 1.106 0.2689   
## data[, 12] -0.027831 0.158778 -0.175 0.8609   
## data[, 13] -0.357359 0.150224 -2.379 0.0174 \*   
## data[, 14] 0.035658 0.139253 0.256 0.7979   
## data[, 15] -0.183716 0.215659 -0.852 0.3943   
## data[, 16] 0.277273 0.139562 1.987 0.0470 \*   
## data[, 17] -0.069773 0.133733 -0.522 0.6019   
## data[, 18] 0.037280 0.154050 0.242 0.8088   
## data[, 19] 0.663837 0.152416 4.355 1.33e-05 \*\*\*  
## data[, 20] -0.160403 0.127609 -1.257 0.2088   
## data[, 21] 0.315256 0.143555 2.196 0.0281 \*   
## data[, 22] -0.009064 0.135753 -0.067 0.9468   
## data[, 23] 0.019813 0.167423 0.118 0.9058   
## data[, 24] 0.330299 0.138896 2.378 0.0174 \*   
## data[, 37] 0.112923 0.131928 0.856 0.3920   
## data[, 38] 0.055869 0.137348 0.407 0.6842   
## data[, 39] 0.013680 0.111297 0.123 0.9022   
## data[, 40] -0.106117 0.142289 -0.746 0.4558   
## data[, 41] -0.098554 0.124477 -0.792 0.4285   
## data[, 42] 0.097432 0.112781 0.864 0.3876   
## data[, 43] -0.174761 0.176199 -0.992 0.3213   
## data[, 44] -0.286606 0.129044 -2.221 0.0264 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 457.28 on 379 degrees of freedom  
## Residual deviance: 356.61 on 357 degrees of freedom  
## AIC: 402.61  
##   
## Number of Fisher Scoring iterations: 5

We observe that the relevant variables are Price, Information provided on the label, Provenance on the label, Description of taste and aroma on the label, Wines history and White wine because they have the lowest p-value associated with the value in the z value column. We make a logistic regression using only the relevant factor to check if we are able to predict the success of Italian wine knowing the relevant factor.

print(paste("likelihood ratio test, pvalue", round(with(lmodel,  
pchisq(null.deviance - deviance, df.null - df.residual, lower.tail = FALSE)), 5)))

## [1] "likelihood ratio test, pvalue 0"

nullmodel<-glm(win ~ 1, family = binomial(link = "logit"))  
print(paste("pseudo R-squared:", round(1 - logLik(lmodel) / logLik(nullmodel),  
3)))

## [1] "pseudo R-squared: 0.22"

model<-glm(win~ data[, 14] + data[, 16] + data[,19] + data[,21] + data[, 24] + data[,44] ,family=binomial(link="logit"))  
summary(model)

##   
## Call:  
## glm(formula = win ~ data[, 14] + data[, 16] + data[, 19] + data[,   
## 21] + data[, 24] + data[, 44], family = binomial(link = "logit"))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4008 -0.7484 0.4967 0.7387 2.1970   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.85593 0.72640 -3.932 8.44e-05 \*\*\*  
## data[, 14] 0.06399 0.12465 0.513 0.6077   
## data[, 16] 0.24499 0.11984 2.044 0.0409 \*   
## data[, 19] 0.67775 0.10979 6.173 6.70e-10 \*\*\*  
## data[, 21] 0.24294 0.12271 1.980 0.0477 \*   
## data[, 24] 0.22900 0.11991 1.910 0.0562 .   
## data[, 44] -0.23061 0.10925 -2.111 0.0348 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 457.28 on 379 degrees of freedom  
## Residual deviance: 370.42 on 373 degrees of freedom  
## AIC: 384.42  
##   
## Number of Fisher Scoring iterations: 4

we compare lmodel which is our original model with the null model, which is the intercept model. We do reject the null model if the values are smaller than 0.05.

We proceeded then to calculate the pseudo r squared, which cannot be tested, ranges between 0 and 1 and is used especially when comparing two models. In this case we get 0.19

print(paste("likelihood ratio test, pvalue:",round(with(model,  
pchisq(null.deviance - deviance, df.null - df.residual,lower.tail = FALSE)), 5)))

## [1] "likelihood ratio test, pvalue: 0"

nullmodel <- glm(win ~ 1, family = binomial(link = "logit"))  
print(paste("pseudo R-squared:", round(1 - logLik(model) / logLik(nullmodel),  
3)))

## [1] "pseudo R-squared: 0.19"

We use the set seed function to be sure that the results are as reliable as possible. we choose as argument the whole set of values contained in these variables, selecting a sample of 150.

set.seed(n)  
  
index.tr<-sample(c(1:n),150)  
  
trmodel1<-glm( win[index.tr] ~ data[index.tr, 14] + data[index.tr, 16] + data[index.tr, 19] + data[index.tr, 21] + data[index.tr, 24]+ data[index.tr, 44],family=binomial(link='logit'))  
train.index<-index.tr  
test.index<-seq(1:n)[-train.index]  
test<-data.frame(cbind(data[test.index, 13], data[test.index, 16], data[test.index, 19], data[test.index, 21], data[test.index, 24], data[test.index, 44]))  
predicted1<-exp(trmodel1$coefficients[1]+ trmodel1$coefficients[2:7]%\*%t(test))/(1+exp(trmodel1$coefficients[1]+ trmodel1$coefficients[2:7]%\*%t(test)))  
  
predicted<-ifelse(predicted1>0.5,1,0)  
  
table(win[test.index], predicted)

## predicted  
## 0 1  
## 0 30 41  
## 1 7 152

We organize the test dataset into a dataframe so that we can move on to calculate the predicted probabilities.

We assign the value 1 for the “superior quality” or “prompt to buy” like we did at the beginning, to predict probabilities larger than 0.5.

Finally we calculate the accuracy of this model, in this case 0.21: the misclassification error is given by the proportion of cases where we fail to predict the correct response, calculated through the mean obtained by the number of responses predicted not corresponding to the actual na.rm refers to the logical parameter that tells the function whether or not to remove NA values from the calculation, in this case yes.

The accuracy is the percentage of cases where we correctly classify: equal to 1-misclassification error

round(prop.table(table(win[test.index],predicted),1),2)

## predicted  
## 0 1  
## 0 0.42 0.58  
## 1 0.04 0.96

misClassificError <- mean(predicted!= win[test.index], na.rm=T)  
print(paste("misClassification Error:", round(misClassificError,2)))

## [1] "misClassification Error: 0.21"

print(paste("Accuracy:",round(1-misClassificError,2)))

## [1] "Accuracy: 0.79"

The predictive model has a quite good accuracy of 79%, but looking at the confusion matrix we can observe that the model is not good to predict the unsuccess of Italian wine.

##Conclusion Looking at the calculated relapse we notice that shoppers of Italian wine are not price sensitive, since the evaluated std of the variable cost is negative, and pay consideration to the information given on the label. Among the information on the two that are related to the success of Italian wine are the provenance of the wine and the portrayal of taste and aroma. The negative sign of the variable white wine implies that the more the buyer like white wines the less they may have a positive judge on Italian wine. These comes about affirm that too habitual Italian wine customers are not cost touchy and lean toward ruddy wine, whereas there’s no correlation between routine Italian wine utilization and the bundling of wine.

Question 3

data[, 37:44][is.na(data[, 37:44])] <- 3  
correlation<-round(cor(data[,37:44]),3)  
#change the names of the variables to have them in a more compact format in the output  
colnames(correlation)<-c("Dryness","Sweet","Oak","Fruity","Still","Spark","Red","White")  
rownames(correlation)<-c("Dryness","Sweet","Oak","Fruity","Still","Spark","Red","White")  
correlation

## Dryness Sweet Oak Fruity Still Spark Red White  
## Dryness 1.000 -0.425 0.025 -0.192 0.180 0.073 0.255 0.060  
## Sweet -0.425 1.000 -0.032 0.259 -0.160 -0.087 -0.224 0.051  
## Oak 0.025 -0.032 1.000 -0.110 -0.041 -0.218 0.067 -0.091  
## Fruity -0.192 0.259 -0.110 1.000 -0.038 0.198 0.014 0.164  
## Still 0.180 -0.160 -0.041 -0.038 1.000 0.070 0.213 0.066  
## Spark 0.073 -0.087 -0.218 0.198 0.070 1.000 -0.057 0.361  
## Red 0.255 -0.224 0.067 0.014 0.213 -0.057 1.000 -0.196  
## White 0.060 0.051 -0.091 0.164 0.066 0.361 -0.196 1.000

library(psych)

## Warning: package 'psych' was built under R version 4.0.5

KMO(data[,37:44])

## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = data[, 37:44])  
## Overall MSA = 0.59  
## MSA for each item =   
## Dryness Sweet Oak Fruity Still Spark Red White   
## 0.62 0.62 0.63 0.57 0.67 0.56 0.56 0.53

cortest.bartlett(data[,37:44])

## R was not square, finding R from data

## $chisq  
## [1] 299.1622  
##   
## $p.value  
## [1] 3.597914e-47  
##   
## $df  
## [1] 28

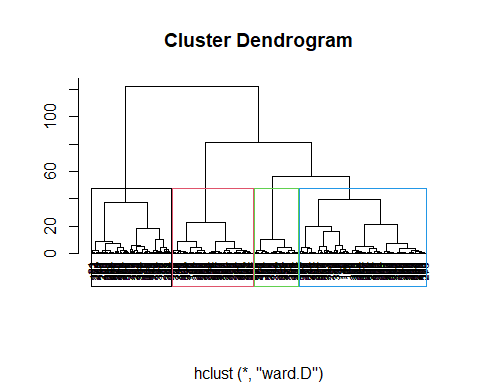
Fanalysis<-factanal(data[,37:44],3)  
Fanalysis

##   
## Call:  
## factanal(x = data[, 37:44], factors = 3)  
##   
## Uniquenesses:  
## Dryness Sweet Oak Fruity Still Spark Red White   
## 0.622 0.513 0.919 0.713 0.906 0.490 0.005 0.720   
##   
## Loadings:  
## Factor1 Factor2 Factor3  
## Dryness 0.594 0.156   
## Sweet -0.691   
## Oak -0.282   
## Fruity -0.386 0.151 0.340   
## Still 0.223 0.194   
## Spark 0.112 0.703   
## Red 0.185 0.964 -0.178   
## White -0.110 0.518   
##   
## Factor1 Factor2 Factor3  
## SS loadings 1.077 1.038 0.998  
## Proportion Var 0.135 0.130 0.125  
## Cumulative Var 0.135 0.264 0.389  
##   
## Test of the hypothesis that 3 factors are sufficient.  
## The chi square statistic is 9.51 on 7 degrees of freedom.  
## The p-value is 0.218

Fanalysis3<-factanal(data[,37:44],3, rotation="varimax",scores="regression")  
Fanalysis3

##   
## Call:  
## factanal(x = data[, 37:44], factors = 3, scores = "regression", rotation = "varimax")  
##   
## Uniquenesses:  
## Dryness Sweet Oak Fruity Still Spark Red White   
## 0.622 0.513 0.919 0.713 0.906 0.490 0.005 0.720   
##   
## Loadings:  
## Factor1 Factor2 Factor3  
## Dryness 0.594 0.156   
## Sweet -0.691   
## Oak -0.282   
## Fruity -0.386 0.151 0.340   
## Still 0.223 0.194   
## Spark 0.112 0.703   
## Red 0.185 0.964 -0.178   
## White -0.110 0.518   
##   
## Factor1 Factor2 Factor3  
## SS loadings 1.077 1.038 0.998  
## Proportion Var 0.135 0.130 0.125  
## Cumulative Var 0.135 0.264 0.389  
##   
## Test of the hypothesis that 3 factors are sufficient.  
## The chi square statistic is 9.51 on 7 degrees of freedom.  
## The p-value is 0.218

#We will analyse the mouth feeling, the color and the taste in order to understand if the customer's perception in multidimensional  
f.scores<-Fanalysis3$scores  
  
mouth\_feel<-f.scores[,1]  
colour<- f.scores[,2]  
Taste<- f.scores[,3]  
  
f.load <-Fanalysis3$loadings  
  
cluster\_data <-f.scores  
  
distance\_matrix<- dist(cluster\_data, method = "euclidean")  
  
clu<-hclust(distance\_matrix, method = "ward.D")  
plot (clu, xlab = "", ylab= "", hang =-1, cex=0.6)  
m<-rect.hclust(clu, k=4,which = c(1,2,3,4), border = 1:4)



QUESTION 3

#It's crucial to visualize the scores of different attributes through frequency plots  
Pinto<-cleandata

## Error in eval(expr, envir, enclos): object 'cleandata' not found

#setting the plot  
par(mfrow=c(2,4), cex=0.3, cex.main=2, cex.axis= 1)  
for(i in 37:44){  
   
 fre<-table(Pinto[,i])   
   
 fre.p <- round(fre/dim(data)[1]\*100,1)   
   
 bp <- barplot(fre.p, ylim=c(0, max(fre.p)\*(1+0.5)),   
 main=colnames(data[i]))  
   
 text(bp, fre.p, label=paste(fre.p, "%"), pos=3, cex=1.5)  
}

## Error in table(Pinto[, i]): object 'Pinto' not found

test\_results<-matrix(0,3,8)  
  
for (i in 1:8){  
 #mean of the attribute scores greater than the indifference level  
 if (mean(Pinto[,36+i])>3){  
   
 t\_test<-t.test(Pinto[,36+i], data=Pinto, mu=3, alternative = "greater")  
 test\_results[1,i]<- t\_test$estimate  
 test\_results[2,i]<- t\_test$statistic  
 test\_results[3,i] <- t\_test$p.value  
   
 }  
 #mean of the attribute scores equal or lower than the indifference level  
 else{  
 t\_test<-t.test(Pinto[,36+i], data=Pinto, mu=3, alternative = "less")  
 test\_results[1,i]<- t\_test$estimate  
 test\_results[2,i]<- t\_test$statistic  
 test\_results[3,i] <- t\_test$p.value  
 }  
   
   
}

## Error in mean(Pinto[, 36 + i]): object 'Pinto' not found

colnames(test\_results)<-c(colnames(Pinto[37:44]))

## Error in is.data.frame(x): object 'Pinto' not found

rownames(test\_results) <- c("mean", "t-test", "p-value")  
test\_results

## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]  
## mean 0 0 0 0 0 0 0 0  
## t-test 0 0 0 0 0 0 0 0  
## p-value 0 0 0 0 0 0 0 0

Dryness: has a significant positive impact on participants’ perception. Since the p-value is next to 0, we reject the hypothesis that the mean of the scores for this attribute is next to the level of indifference 3.

Sweetness: has a significant negative impact on participants’ perceptions. Since the p-value is next to 0, we reject the hypothesis that the mean of the scores for this attribute is next to the level of indifference 3.

Oaktaste: does not have a significant impact on participants’ perception. Since the p-value is 11.38, far above the 5% level of significance, we cant reject the H0.

Fruitytaste: has a positive impact on the participants’ perception. The p-value is lower than 5% but, if we compare its t-test with the ones of attributes like dryness and sweetness, we can state that respondents attitude toward this characteristic is less significant.

Still: has a significant positive impact on the participants’ perception. Since the p-value is next to 0, we reject the hypothesis that the mean of the scores for this attribute is next to the level of indifference 3.

Spark: has a positive impact on the participants’ perception. Since the p-value is next to 0, we reject the hypothesis that the mean of the scores for this attribute is next to the level of indifference 3. Despite this, the p-value of this attribute is significantly greater than the ones of attributes like dryness and sweetness and its t-test is lower. Therefore we can accept the H1 but we have to underline that it has a smaller impact on respondents perception.

Red: has a significant positive impact on the participants’ perception. Since the p-value is next to 0, we reject the hypothesis that the mean of the scores for this attribute is next to the level of indifference 3.

White: has a significant positive impact on the participants’ perception. Since the p-value is next to 0, we reject the hypothesis that the mean of the scores for this attribute is next to the level of indifference 3.

#Put the attributes on a scale from the most impacting attribute to the least impacting one  
most\_imp<- order(test\_results[3,])  
  
test\_results<-test\_results[, most\_imp]  
  
colnames(test\_results)

## NULL

#Build a correlation table to see if there is any correlation  
data[, 37:44][is.na(data[, 37:44])] <- 3  
correlation<-round(cor(data[,37:44]),3)  
#change the names of the variables to have them in a more compact format in the output  
colnames(correlation)<-c("Dryness","Sweet","Oak","Fruity","Still","Spark","Red","White")  
rownames(correlation)<-c("Dryness","Sweet","Oak","Fruity","Still","Spark","Red","White")  
correlation

## Dryness Sweet Oak Fruity Still Spark Red White  
## Dryness 1.000 -0.425 0.025 -0.192 0.180 0.073 0.255 0.060  
## Sweet -0.425 1.000 -0.032 0.259 -0.160 -0.087 -0.224 0.051  
## Oak 0.025 -0.032 1.000 -0.110 -0.041 -0.218 0.067 -0.091  
## Fruity -0.192 0.259 -0.110 1.000 -0.038 0.198 0.014 0.164  
## Still 0.180 -0.160 -0.041 -0.038 1.000 0.070 0.213 0.066  
## Spark 0.073 -0.087 -0.218 0.198 0.070 1.000 -0.057 0.361  
## Red 0.255 -0.224 0.067 0.014 0.213 -0.057 1.000 -0.196  
## White 0.060 0.051 -0.091 0.164 0.066 0.361 -0.196 1.000

#Perform a KMO test to see if we can exploit the correlation  
library(psych)  
KMO(data[,37:44])

## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = data[, 37:44])  
## Overall MSA = 0.59  
## MSA for each item =   
## Dryness Sweet Oak Fruity Still Spark Red White   
## 0.62 0.62 0.63 0.57 0.67 0.56 0.56 0.53

#The overall MSA is quite low (lower than 0.7 as it is suggested) but usually marketers accept a overall MSA>0.5  
#Perform a Barlett test to get to know the chi-square test for independence  
cortest.bartlett(data[,37:44])

## R was not square, finding R from data

## $chisq  
## [1] 299.1622  
##   
## $p.value  
## [1] 3.597914e-47  
##   
## $df  
## [1] 28

#The chi-square test for independence, p-value is very low therefore we reject the hypothesis that the variables are independent.

Now we can proceed with the factorial analysis:

Fanalysis<-factanal(data[,37:44],3)  
Fanalysis

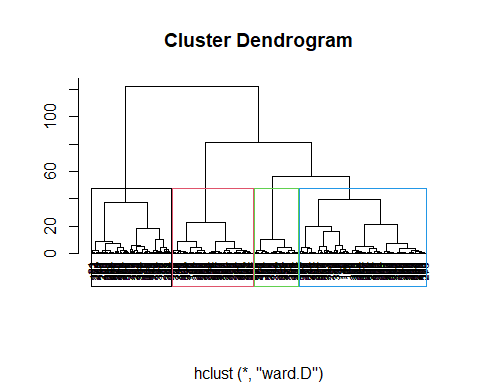
##   
## Call:  
## factanal(x = data[, 37:44], factors = 3)  
##   
## Uniquenesses:  
## Dryness Sweet Oak Fruity Still Spark Red White   
## 0.622 0.513 0.919 0.713 0.906 0.490 0.005 0.720   
##   
## Loadings:  
## Factor1 Factor2 Factor3  
## Dryness 0.594 0.156   
## Sweet -0.691   
## Oak -0.282   
## Fruity -0.386 0.151 0.340   
## Still 0.223 0.194   
## Spark 0.112 0.703   
## Red 0.185 0.964 -0.178   
## White -0.110 0.518   
##   
## Factor1 Factor2 Factor3  
## SS loadings 1.077 1.038 0.998  
## Proportion Var 0.135 0.130 0.125  
## Cumulative Var 0.135 0.264 0.389  
##   
## Test of the hypothesis that 3 factors are sufficient.  
## The chi square statistic is 9.51 on 7 degrees of freedom.  
## The p-value is 0.218

Fanalysis3<-factanal(data[,37:44],3, rotation="varimax",scores="regression")  
Fanalysis3

##   
## Call:  
## factanal(x = data[, 37:44], factors = 3, scores = "regression", rotation = "varimax")  
##   
## Uniquenesses:  
## Dryness Sweet Oak Fruity Still Spark Red White   
## 0.622 0.513 0.919 0.713 0.906 0.490 0.005 0.720   
##   
## Loadings:  
## Factor1 Factor2 Factor3  
## Dryness 0.594 0.156   
## Sweet -0.691   
## Oak -0.282   
## Fruity -0.386 0.151 0.340   
## Still 0.223 0.194   
## Spark 0.112 0.703   
## Red 0.185 0.964 -0.178   
## White -0.110 0.518   
##   
## Factor1 Factor2 Factor3  
## SS loadings 1.077 1.038 0.998  
## Proportion Var 0.135 0.130 0.125  
## Cumulative Var 0.135 0.264 0.389  
##   
## Test of the hypothesis that 3 factors are sufficient.  
## The chi square statistic is 9.51 on 7 degrees of freedom.  
## The p-value is 0.218

Despite the cumulative variance is a bit low, the p-value allow us to validate the correlation structure. The factors are going to be: F1: mouth feel –> is described as dryness and roughness rather than a taste and it is subject to the presence/absence of sugar. F2: color –> this factor is correlated with the attributes Red and White which are the possible colors of the wine. F3: Taste –> this factor is strongly correlated with the fruity taste and sparkling, attributes correlated with the aroma of the wine. Now we can tell that the respondents’ perception is a multidimensional trait.

#We will analyse the mouth feeling, the color and the taste in order to understand if the customer's perception in multidimensional  
fact.scores<-Fanalysis3$scores  
  
mouth\_feel<-fact.scores[,1]  
colour<- fact.scores[,2]  
Taste<- fact.scores[,3]  
  
fact.load <-Fanalysis3$loadings  
  
cluster\_data <-fact.scores  
  
distance\_matrix<- dist(cluster\_data, method = "euclidean")  
  
clu<-hclust(distance\_matrix, method = "ward.D")  
plot (clu, xlab = "", ylab= "", hang =-1, cex=0.6)  
membership<-rect.hclust(clu, k=4,which = c(1,2,3,4), border = 1:4)



memb<-cutree(clu, k=4)  
clu.scores<-matrix(0,4,3)  
  
rownames(clu.scores)<-c("clu1", "clu2", "clu3", "clu4")  
  
for (i in 1:4){  
 for (j in 1:3){  
 clu.scores[i,j]<-round(mean(fact.scores[memb==i,j]),3)  
 }  
}  
  
colnames(clu.scores)<-c("mouth feel", "colour", "Taste")  
clu.scores

## mouth feel colour Taste  
## clu1 -0.510 0.521 0.071  
## clu2 -0.081 -1.478 0.272  
## clu3 0.503 0.305 -0.872  
## clu4 0.677 0.599 0.903

From here we can recognize 4 different types customers: Cluster 1: coloursensitivecustomer/tasteindifferentcustomers –> they select their wines giving particular attention to wine’s color. Moreover they are really indifferent to its taste. Cluster 2: tastesensitivecustomer –> they have a really taste score and they don’t care about mouth feel and color Cluster 3: colourindifferentcustomers/mouthfeelsensitivecustomer –> they are almost indifferent to wine’s color and select their wine based on the mouth feel therefore the quantity of sugar present in the wine. Cluster 4: tasters –> they make their choices without caring particularly on the external aspect of the wine, they value the peculiar aspects like the taste.

This analysis will be answering question 4 of our research proposal:What are the key words or images to make our wine desirable for US consumers?. In order to achieve this, we are going to analyses the differences of frequent Italian wine consumers and those who aren’t.

QUESTION 4:

In order to complete this question, first we are going to create a binomial variable which will divide the sample in consumers who usually buy Italian wine and consumers who do not. This division will be based on V31 from the questionnaire. Our assumption is that we assume that habitual Italian wine consumers answered the questionnaire with Italian wine in mind, while all other consumers based their answers on wine in general. We are going to analyze the difference of means of the answers from V11 to v18 from the two sub-samples, to understand which attributes are valued more by habitual Italian wine consumers compared to other consumers. We are going to make inference at a 95% confidence interval in order to make a conclusion on the american market and not only on the sample.

data<-download#source the data matrix  
data <- na.exclude(data)#removes all non answers from data  
   
  
#question 4  
n <- nrow(data)  
s<-rep(0,n)  
  
#stater code  
s[data[,31]!=1]<-"buy other"  
s[data[,31]==1]<-"buy ita"  
s<-factor(s)  
data[,31]<-s

In the code seen above I divide the data into people who buy Italian wines and those who buy other wines. Then I name the variable for ease of code interpretation. This will allow us to analyse the differences between the two sub-samples and their preferences. ‘S’ is converted to a factor. This is done because factors are easier to work with thus simplifying the analysis process.

#code 1  
j<-1  
v<-j+10  
vv<-data[,v]  
  
print(paste("we analyse the variable: ",data[j]))

## [1] "we analyse the variable: c(1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, \n127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 171, 173, 174, 175, 176, 177, 178, 179, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 200, 201, 202, 203, 204, 206, 207, 208, 209, 210, 211, 212, 214, 215, 217, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, \n237, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 252, 253, 255, 257, 258, 260, 261, 262, 263, 267, 268, 269, 270, 271, 272, 273, 275, 276, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 291, 292, 293, 294, 296, 297, 299, 300, 302, 304, 305, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 320, 321, 322, 324, 325, 326, 328, 329, 332, 333, 335, 336, 338, 339, 340, 341, 342, 343, 344, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 357, 358, 359, 360, 361, 362, 363, 364, 365, \n366, 369, 371, 372, 373, 376, 377, 378, 379, 380, 381, 382, 383, 384, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 420, 421, 424, 425, 426, 427, 428, 430, 431, 432, 433)"

#[1] "we analyse the variable: Pl\_Origin"  
s<-factor(s)  
data[,'s'] <- as.factor(data[,'s'])

## Error in `[.data.frame`(data, , "s"): undefined columns selected

ttestGC(vv ~ s, mu=0, graph = T, first = "buy ita")

## Error in ttestGC(vv ~ s, mu = 0, graph = T, first = "buy ita"): could not find function "ttestGC"

#This code format will be used to analyse all the attributes

With a P-value of 0.22 we can determine that frequent wine drinkers are influenced by the place of origin of the wine more than other consumers. This is because, due to the 95% confidence interval, we reject the null hypothesis.

#code 2  
v<-j+10  
vv<-data[,v]  
print(paste("we analyse the variable: ",data[j]))

## [1] "we analyse the variable: c(1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, \n127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 171, 173, 174, 175, 176, 177, 178, 179, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 200, 201, 202, 203, 204, 206, 207, 208, 209, 210, 211, 212, 214, 215, 217, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, \n237, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 252, 253, 255, 257, 258, 260, 261, 262, 263, 267, 268, 269, 270, 271, 272, 273, 275, 276, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 291, 292, 293, 294, 296, 297, 299, 300, 302, 304, 305, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 320, 321, 322, 324, 325, 326, 328, 329, 332, 333, 335, 336, 338, 339, 340, 341, 342, 343, 344, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 357, 358, 359, 360, 361, 362, 363, 364, 365, \n366, 369, 371, 372, 373, 376, 377, 378, 379, 380, 381, 382, 383, 384, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 420, 421, 424, 425, 426, 427, 428, 430, 431, 432, 433)"

## [1] "we analyse the variable: W\_Brand"  
print(ttestGC(vv ~ s, mu=0, graph = T, first = "buy ita"))

## Error in ttestGC(vv ~ s, mu = 0, graph = T, first = "buy ita"): could not find function "ttestGC"

j<-j+1

#code 3  
v<-j+10  
vv<-data[,v]  
print(paste("we analyse the variable: ",data[j]))

## [1] "we analyse the variable: c(3, 2, 2, 6, 3, 3, 6, 6, 3, 3, 6, 3, 6, 3, 3, 3, 6, 6, 3, 3, 6, 6, 6, 6, 6, 3, 3, 6, 3, 6, 3, 6, 6, 6, 6, 6, 3, 3, 6, 3, 3, 2, 6, 6, 2, 6, 6, 6, 6, 6, 6, 6, 6, 2, 6, 6, 6, 3, 6, 3, 6, 2, 2, 6, 6, 6, 1, 2, 3, 6, 6, 3, 6, 6, 3, 3, 6, 6, 6, 3, 6, 3, 3, 6, 6, 3, 3, 3, 3, 3, 2, 3, 3, 6, 6, 6, 6, 3, 6, 3, 3, 6, 3, 6, 6, 6, 3, 6, 3, 6, 6, 3, 6, 6, 3, 6, 3, 3, 3, 6, 6, 6, 3, 3, 6, 3, 3, 6, 6, 6, 6, 3, 6, 3, 6, 6, 3, 3, 6, 2, 6, 3, 6, 2, 1, 2, 3, 1, 1, 2, 1, 2, 6, 6, 6, 2, 3, 6, 3, 4, 5, 4, 5, 4, 4, 2, 3, \n5, 4, 3, 3, 3, 4, 2, 3, 4, 3, 6, 6, 6, 6, 6, 3, 3, 6, 6, 2, 4, 5, 4, 4, 4, 1, 2, 4, 4, 6, 3, 4, 4, 6, 3, 4, 4, 6, 6, 4, 5, 6, 4, 6, 4, 4, 3, 3, 6, 6, 6, 3, 6, 6, 2, 4, 3, 3, 3, 4, 4, 5, 3, 4, 3, 6, 4, 3, 6, 6, 3, 5, 4, 6, 3, 1, 2, 4, 5, 4, 6, 4, 3, 6, 3, 5, 4, 4, 6, 4, 4, 4, 4, 4, 4, 4, 6, 5, 6, 4, 4, 4, 6, 5, 4, 6, 5, 3, 4, 3, 6, 3, 4, 6, 4, 3, 6, 3, 6, 1, 2, 6, 3, 4, 6, 5, 6, 6, 6, 4, 6, 5, 5, 6, 6, 4, 4, 5, 5, 4, 3, 4, 4, 6, 6, 3, 3, 6, 6, 4, 5, 3, 3, 4, 4, 6, 4, 6, 4, 4, 6, 6, 4, 4, 5, 5, 4, \n5, 4, 6, 3, 6, 6, 4, 6, 6, 5, 4, 6, 4, 4, 6, 6, 6, 3, 5, 4, 6, 5, 6, 6, 4, 5, 4, 6, 6, 4, 6, 5, 4, 4, 4, 5, 4, 4, 3, 6, 6, 4, 6, 6, 6, 3)"

## [1] "we analyse the variable: Price"  
print(ttestGC(vv ~ s, mu=0, graph = T, first = "buy ita"))

## Error in ttestGC(vv ~ s, mu = 0, graph = T, first = "buy ita"): could not find function "ttestGC"

j<-j+1  
  
#this test shows us that frequent consumers arent as influenced by price when buying compared to normal consumers

# code 4  
v<-j+10  
vv<-data[,v]  
print(paste("we analyse the variable: ",data[j]))

## [1] "we analyse the variable: c(2, 2, 3, 4, 1, 5, 5, 5, 1, 3, 5, 1, 5, 1, 1, 5, 5, 5, 5, 4, 5, 1, 5, 1, 2, 5, 1, 3, 5, 4, 2, 1, 1, 5, 5, 5, 3, 5, 3, 1, 2, 4, 1, 2, 5, 3, 1, 3, 5, 2, 5, 1, 5, 2, 3, 5, 3, 5, 5, 3, 5, 2, 3, 5, 5, 5, 5, 4, 5, 1, 5, 5, 5, 2, 2, 1, 5, 5, 1, 5, 5, 5, 5, 3, 5, 5, 5, 3, 4, 5, 5, 2, 1, 5, 5, 5, 5, 2, 2, 5, 3, 5, 3, 4, 3, 5, 2, 1, 5, 5, 5, 5, 5, 3, 5, 2, 2, 2, 1, 3, 3, 5, 3, 1, 5, 5, 4, 5, 5, 5, 3, 5, 5, 2, 2, 3, 2, 2, 1, 1, 5, 2, 5, 4, 4, 2, 5, 1, 3, 1, 3, 1, 1, 5, 5, 2, 1, 5, 5, 5, 5, 2, 5, 5, 1, 1, 3, \n5, 4, 3, 3, 2, 5, 1, 5, 3, 2, 5, 3, 3, 2, 1, 1, 5, 5, 5, 3, 2, 5, 5, 5, 5, 2, 5, 5, 4, 4, 3, 2, 5, 5, 5, 5, 4, 5, 2, 5, 5, 5, 4, 5, 3, 5, 4, 5, 5, 2, 5, 1, 1, 5, 4, 1, 1, 1, 1, 5, 5, 4, 2, 5, 5, 5, 2, 3, 5, 5, 5, 5, 5, 5, 5, 2, 1, 3, 5, 4, 3, 5, 2, 5, 3, 5, 5, 5, 5, 2, 1, 3, 5, 5, 5, 5, 4, 4, 5, 5, 5, 3, 5, 5, 2, 2, 3, 5, 5, 1, 5, 5, 5, 1, 5, 3, 3, 2, 4, 1, 4, 4, 5, 3, 4, 3, 1, 4, 5, 1, 5, 5, 3, 4, 4, 5, 5, 4, 5, 5, 1, 5, 5, 5, 5, 3, 4, 5, 5, 1, 5, 5, 3, 2, 5, 3, 2, 2, 2, 3, 2, 3, 4, 5, 5, 4, 1, \n5, 5, 2, 4, 5, 4, 2, 2, 2, 5, 3, 3, 5, 2, 3, 5, 4, 1, 5, 1, 2, 5, 2, 3, 3, 5, 3, 1, 4, 5, 2, 1, 3, 4, 5, 3, 3, 2, 1, 2, 2, 5, 4, 1, 3, 5)"

## [1] "we analyse the variable: Pack"  
print(ttestGC(vv ~ s, mu=0, graph = T, first = "buy ita"))

## Error in ttestGC(vv ~ s, mu = 0, graph = T, first = "buy ita"): could not find function "ttestGC"

j<-j+1  
  
#as there is a small pvalue that, however, is bigger than 0, packaging plays an important role in the buying decisions of a frequent wine consumer

#code 5  
v<-j+10  
vv<-data[,v]  
print(paste("we analyse the variable: ",data[j]))

## [1] "we analyse the variable: c(2, 2, 3, 5, 5, 5, 5, 5, 4, 3, 5, 5, 5, 3, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 2, 3, 2, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 2, 5, 5, 5, 4, 5, 4, 5, 1, 5, 5, 5, 5, 2, 4, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 1, 5, 5, 5, 5, 5, 5, 3, 4, 5, 5, 4, 5, 5, 4, 5, 5, 5, 4, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 3, 2, 5, 5, 5, 4, 3, 3, 5, 5, 5, 2, 5, 3, 5, 5, 5, 3, 5, 5, 3, 4, 2, 2, 5, 3, 5, 5, 3, 5, 5, 2, 2, 3, 2, 2, 5, 5, 3, 4, 2, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, \n5, 5, 4, 3, 2, 5, 2, 5, 3, 5, 5, 3, 4, 4, 1, 2, 3, 5, 5, 3, 4, 5, 5, 5, 5, 2, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 1, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 1, 5, 5, 5, 5, 5, 2, 5, 4, 3, 4, 4, 5, 4, 5, 4, 5, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 5, 5, 5, 4, 3, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 4, 1, 3, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, \n5, 5, 5, 3, 5, 5, 4, 5, 5, 5, 5, 4, 4, 4, 5, 5, 5, 4, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 4, 4, 4, 5, 5, 5, 5, 4, 5, 5, 5)"

## [1] "we analyse the variable: Quality"  
print(ttestGC(vv ~ s, mu=0, graph = T, first = "buy ita"))

## Error in ttestGC(vv ~ s, mu = 0, graph = T, first = "buy ita"): could not find function "ttestGC"

j<-j+1  
#quality is obviously an important attribute for frequent consumers, this is demonstrated by this ttest.

#code 6  
v<-j+10  
vv<-data[,v]  
print(paste("we analyse the variable: ",data[j]))

## [1] "we analyse the variable: c(5, 2, 5, 5, 5, 5, 5, 5, 1, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 4, 5, 2, 5, 5, 5, 5, 5, 4, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 4, 5, 5, 4, 5, 5, 5, 5, 5, 5, 1, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 3, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, \n5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 4, 5, 5, 5, 3, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 2, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 3, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 1, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, \n5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 4, 5, 5, 3, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5)"

## [1] "we analyse the variable: Info\_lab"  
print(ttestGC(vv ~ s, mu=0, graph = T, first = "buy ita"))

## Error in ttestGC(vv ~ s, mu = 0, graph = T, first = "buy ita"): could not find function "ttestGC"

j<-j+1  
#

#code 7  
v<-j+10  
vv<-data[,v]  
print(paste("we analyse the variable: ",data[j]))

## [1] "we analyse the variable: c(5, 2, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 1, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 2, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, \n5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 4, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, \n5, 5, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 2, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5)"

## [1] "we analyse the variable: Sal\_rec"  
print(ttestGC(vv ~ s, mu=0, graph = T, first = "buy ita"))

## Error in ttestGC(vv ~ s, mu = 0, graph = T, first = "buy ita"): could not find function "ttestGC"

j<-j+1

#Code 8  
v<-j+10  
vv<-data[,v]  
print(paste("we analyse the variable: ",data[j]))

## [1] "we analyse the variable: c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 4, 1, 1, 2, 1, 1, 1, 2, 1, 1, 1, 2, 2, 1, 2, 2, 1, 4, 1, 4, 4, 1, 1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 3, 2, 1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 1, 1, 1, 2, 1, 1, 2, 4, 1, 1, 1, 1, 1, 2, 2, 1, 2, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1, 1, 2, 1, 4, 2, 2, 1, 1, 4, 4, 1, 1, 1, 4, 2, 1, 1, 1, 1, 4, 2, 1, 1, 1, 1, 1, 1, 1, 1, 4, 1, \n1, 2, 1, 2, 4, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 2, 1, 1, 1, 1, 1, 4, 2, 1, 1, 1, 2, 4, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 2, 2, 4, 1, 1, 2, 1, 1, 1, 1, 1, 1, 3, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 2, 1, 1, 1, 1, 4, 4, 2, 1, 2, 1, 1, 2, 1, 2, 1, 1, 2, 1, 1, 2, 1, 1, 1, 2, 1, 1, 1, 1, 1, 4, 1, 2, 1, 1, 1, 1, 1, 1, 4, 1, 1, 1, 2, 1, 4, 4, 2, 1, 1, 2, 1, 4, 1, 1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 4, 4, 1, 1, 1, 2, 1, 4, 1, 1, 1, 2, 1, 2, 1, 1, \n4, 1, 1, 4, 4, 1, 1, 1, 1, 1, 1, 1, 4, 4, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 2, 4, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1)"

## [1] "we analyse the variable: F\_rec"  
print(ttestGC(vv ~ s, mu=0, graph = T, first = "buy ita"))

## Error in ttestGC(vv ~ s, mu = 0, graph = T, first = "buy ita"): could not find function "ttestGC"

j<-j+1

As we can see from this part of the analysis, the two attributes which are most important for frequent italian wine consumers are: place of origin and the information on the label.

#code 9 barplot  
#In this barplot we are going to use the results of the previous analysis in order to understand what information frequent consumers look for on the label of a wine bottle. We are going to look at variables 19-24 taken from the questionnaire.  
  
frequent <- data[data[, 31] == 1,]  
  
m<-rep(0,6)  
  
i<-1  
for (v in 19:24){  
 vv<-frequent[,v]  
 m[i]<-mean(vv)  
 i<-i+1 #always increases by one   
}  
m<-sort(m, decreasing = T) #sort m list  
  
bp<-barplot(m, name = c("Prov","Tech\_info", "d\_tase", "food ", "brand", "history"), ylim=c(0,5))

## Warning in min(w.l): no non-missing arguments to min; returning Inf

## Warning in max(w.r): no non-missing arguments to max; returning -Inf

## Error in plot.window(xlim, ylim, log = log, ...): need finite 'xlim' values

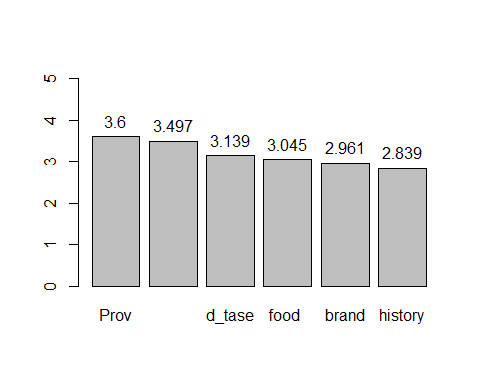


text(bp, m, labels = round(m,3), pos = 3)

## Error in text.default(bp, m, labels = round(m, 3), pos = 3): zero-length 'labels' specified

# Looking at the barplot, we can see that the provenance of the wine, taste and aroma and technical info are the top 3 attributes that frequent italian wine consumers seek on the label.

# with this barplot we will compare the previous barplot with the results with the whole data sample. This is done to understand whether the sought after attributes differ based on the consumer.  
m<-rep(0,6)  
  
i<-1  
for (v in 19:24){  
 vv<-data[,v]  
 m[i]<-mean(vv)  
 i<-i+1  
}  
m<-sort(m, decreasing = T)  
  
bp<-barplot(m, name = c("Prov","Tech\_info", "d\_tase", "food ", "brand", "history"), ylim=c(0,5))  
  
 text(bp, m, labels = round(m,3), pos = 3)



# as we can see the results are the same. Although the means are slightly lower than before, the top 3 sought after attributes remain the same none the less. This analysis tells us that including this information on wine labels will attract the most customers in general. Regardless if they frequently consume italian wine or not.