Deliverable 2

Lab 2 - PCA, CA and Clustering

Othman Benmoussa & Eloi Cruz

May 1, 2022

Contents

1 Load Required Packages: to be increased over the course

```
# Load Required Packages: to be increased over the course
options(contrasts=c("contr.treatment","contr.treatment"))

requiredPackages <- c("effects","FactoMineR","car","missMDA","mvoutlier","chemometrics", "factoextra","F

#use this function to check if each package is on the local machine
#if a package is installed, it will be loaded
#if any are not, the missing package(s) will be installed and loaded
package.check <- lapply(requiredPackages, FUN = function(x) {
  if (!require(x, character.only = TRUE)) {
    install.packages(x, dependencies = TRUE)
    library(x, character.only = TRUE)
  }
})
#verify they are loaded
search()
```

1.1 Load Processed data

```
# Clear plots
if(!is.null(dev.list())) dev.off()

# Clean workspace
rm(list=ls())
setwd("/Users/othmanbenmoussa/Downloads/ADEI-master-79098fa31eef4ee27f4e3f58437a95dcf6a573c1/Lab3PCA")
#filepath("/Users/othmanbenmoussa/Downloads/ADEI-master-79098fa31eef4ee27f4e3f58437a95dcf6a573c1/Lab3PCA
#filepath<-"/"Users/othmanbenmoussa/Desktop/FIB/ADEI/LABO
#setwd("C:/Users/Eloi/Documents/ADEI/ADEI/Lab3PCA/")
#filepath<-"C:/Users/Eloi/Documents/ADEI/ADEI/Lab3PCA/"
# green_tripdata_2016-01)

load("MyOldCars-5000Clean.RData")
options(contrasts=c("contr.treatment","contr.treatment"))
```

We assume that NA are not present in the variables. Our working dataframe is already clean.

2 Principal Component Analysis

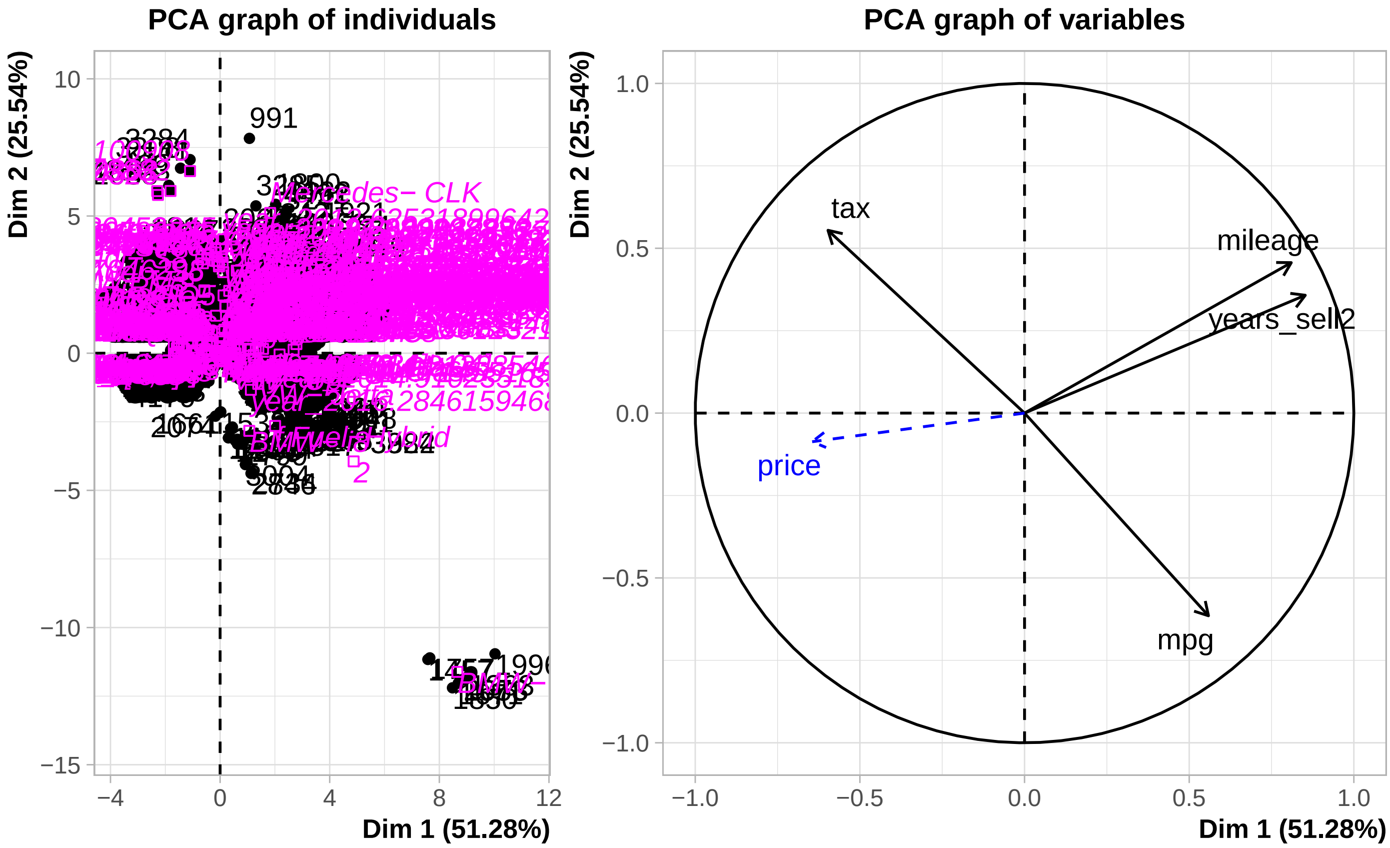
In this section we do the principal component analysis to reduce the number of variables that we are using to analize the data.

```
#Reasignation of variables because there were some errors in the first lab
vars_con<-names(df)[c(5,7,8,14)]
vars_dis<-names(df)[c(1,2,4,6,9,10, 12,13,15,16,17)]
vars_res<-names(df)[c(3,11)]



#Remove remaining NA's
df = df[complete.cases(df),]

#Calculate the PCA
res.pca<-PCA(df[,c(vars_res, vars_dis,vars_con)],quali.sup=c(2:13),quanti.sup= c(1))
```




```r
#library(FactoMineR)
#plot.PCA(res.pca,choix=c("var"),axes=c(1,2))
```

First of all we can interpret the result of the execution of the PCA function. We can see that the first dimension created contains a variance of 51% of the observations and that the second dimension created contains a variance of the 25% of the observations. We know that variables that conform a 90 degree angle are not related. Dimensions in the same direction are related. As we can see mileage is very positively strong related to the age of the car. The consumption of the car is inverse related to the tax. There is no relation between the tax or mpg and the mileage and the age. What is more cars with a lot of mileage or that are very old are cheaper than cars that are new.

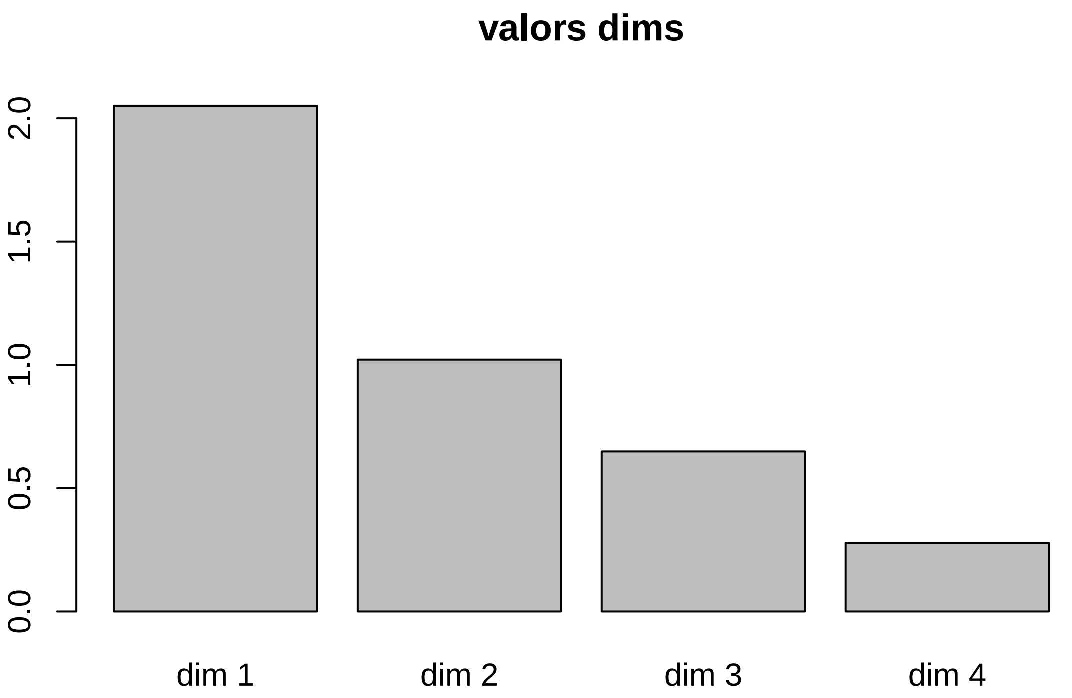
```r
# Multivariant outliers should be included as supplementary observations
#ll <- which( df$mout == "YesMOut")
#res.pca<-PCA(df[,c(vars_res, vars_dis, vars_con)],quali.sup=c(2:13),quanti.sup= c(1), ind.sup = ll )
```

2.1 Eigenvalues and dominant axes analysis. How many axes we have to interpret according to Kaiser and Elbow's rule?

```r
round(res.pca$eig,2)
```

```
##        eigenvalue percentage of variance cumulative percentage of variance
## comp 1       2.05                  51.28                             51.28
## comp 2       1.02                  25.54                             76.81
## comp 3       0.65                  16.23                             93.04
## comp 4       0.28                   6.96                            100.00
```

```r
barplot(res.pca$eig[,1],main="valors dims",names.arg=paste("dim",1:nrow(res.pca$eig)))
```



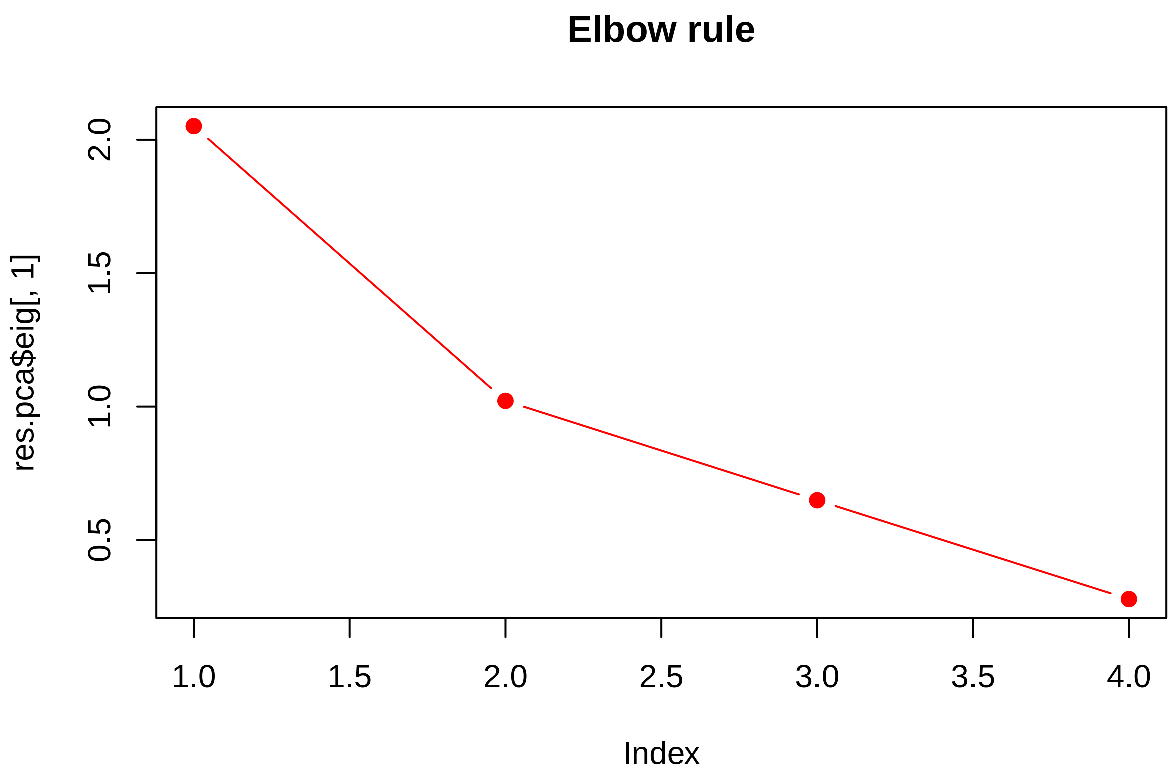

```r
sum(res.pca$eig[,1])
```

```
## [1] 4
```

In one hand, According to Kaiser criteria the PCs with eignvalue greater than 1 have more variance that the original values and for this reason the axes that we have to leave are the ones with a value less than 1. Using this criteria we will have to interpret only two axes.

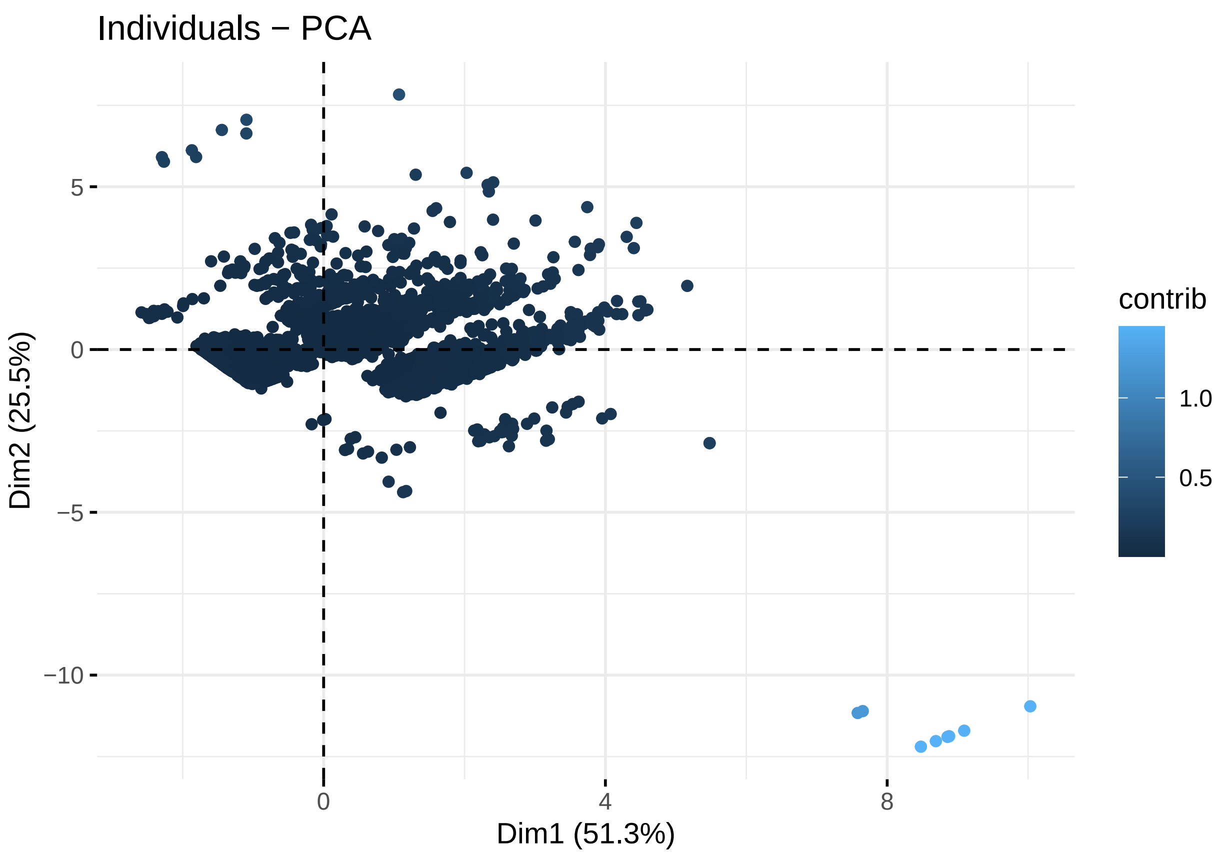
In the other hand, according to the elbow rule criteria we have to choose before the line flaterns out, so we will choose two dimensions too.

```r
plot(res.pca$eig[,1],main="Elbow rule", type="b", pch = 19, col = "red")
```



2.2 Individuals point of view. Are they any individuals "too contributive"?

```
fviz_pca_ind(res.pca, col.ind="contrib", geom = "point") #+scale_color_gradient2(low="darkslateblue", m
```

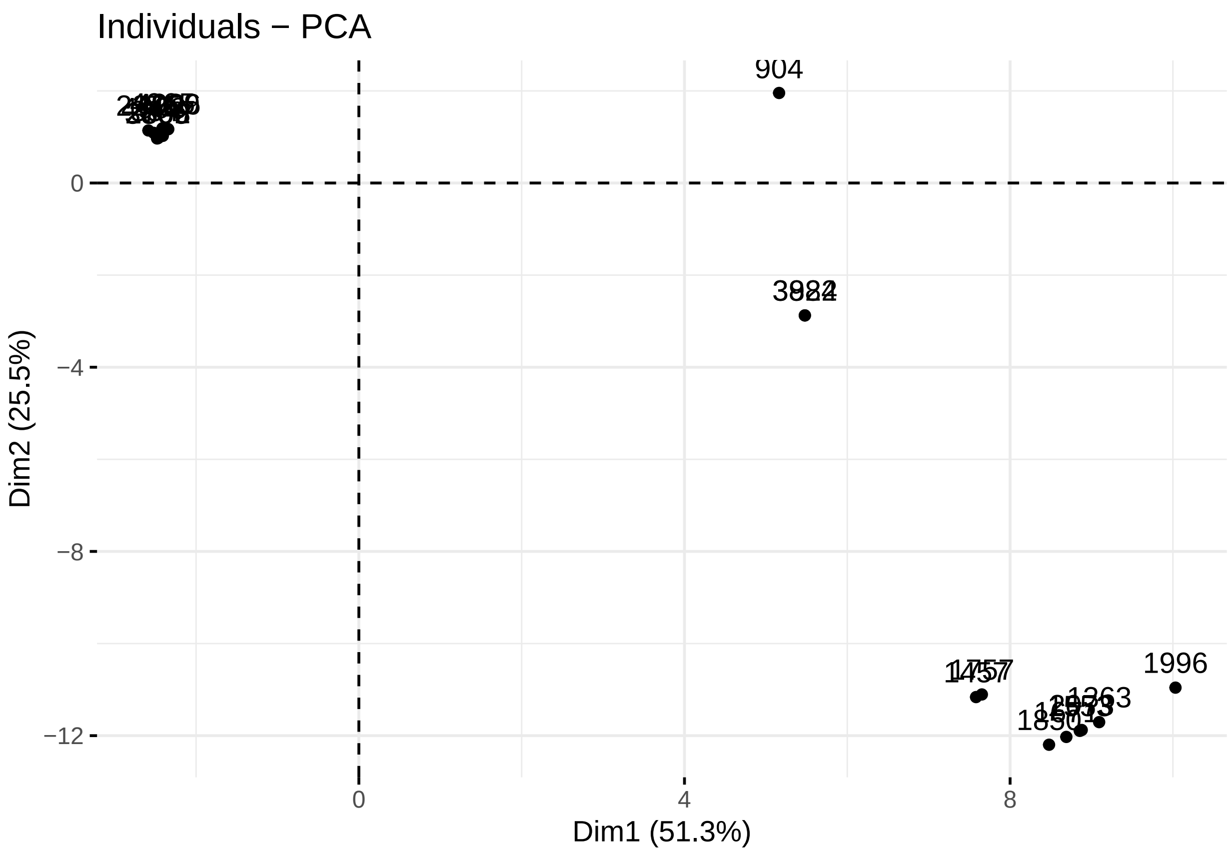


We can see that there are individuals that are too contributive, especially at the extremes parts of the two dimensions. Let's understand them better.

2.2.1 Extreme Individuals

We will leave just the extreme values in the following plots according to each of the two dimensions

```
rang<-order(res.pca$ind$coord[,1])
contrib.extremes<-c(row.names(df)[rang[1]], row.names(df)[rang[length(rang)]])
contrib.extremes<-c(row.names(df)[rang[1:10]], row.names(df) [rang[(length(rang)-10):length(rang)]])
fviz_pca_ind(res.pca, select.ind = list(names=contrib.extremes))
```




```
df[which(row.names(df) %in% row.names(df)[rang[length(rang)]]), 1:19]
```

```
##        model year price      transmission mileage      fuelType tax   mpg
## 1996 BMW- i3 2015 12500 f.Trans-Automatic   79830 f.Fuel-Hybrid  20 470.8
##      engineSize manufacturer Audi total years_sell years_sell2
## 1996      Petit    f.Man-BMW   No     2   Semi nou           2
##                      aux       f.price          f.miles          f.tax
## 1996 (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323] f.tax-(1,125]
##                  mpg_d
## 1996 mpg_d-(61.4,471]
```

We notice that the extreme models are BMW i3. We can think that they don't follow the same characteristics as others because of their engine which is hybrid, which makes them extreme.

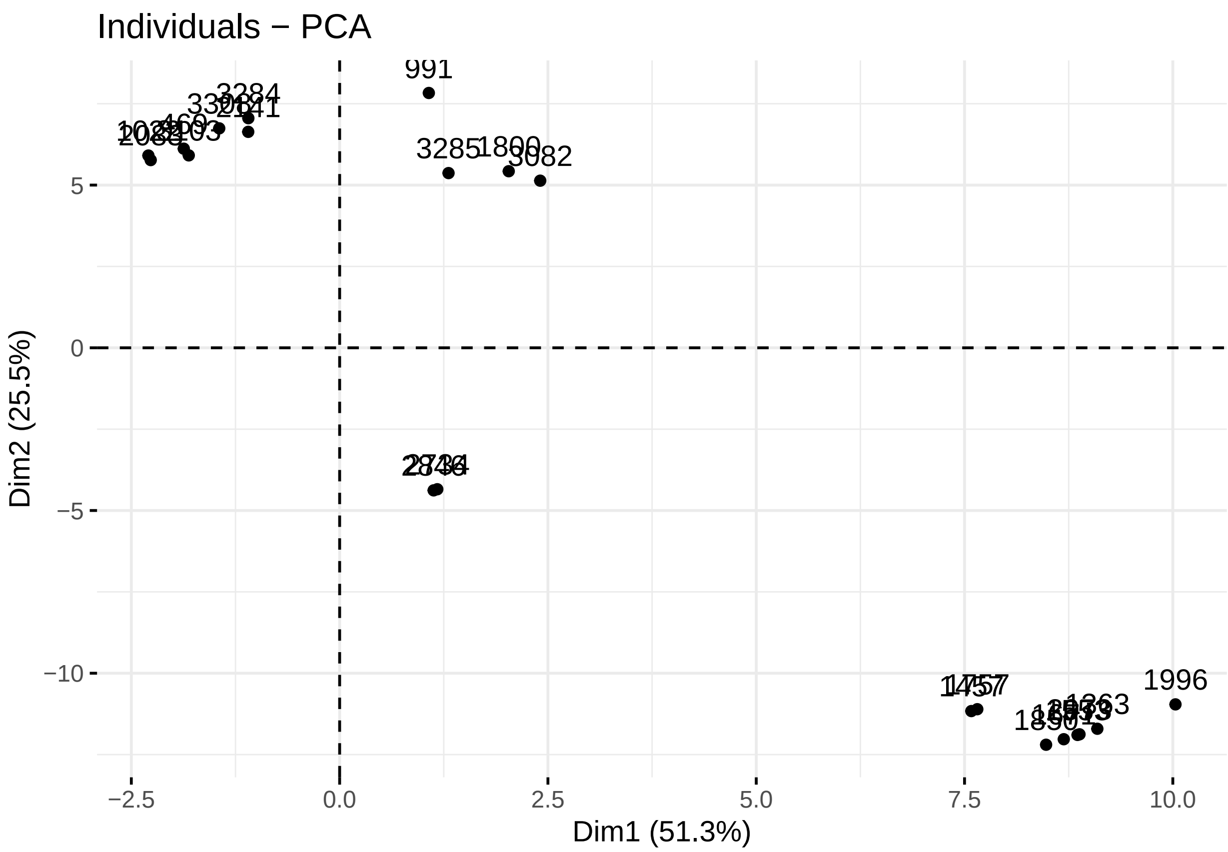
```
df[which(row.names(df) %in% row.names(df)[rang[1]]),1:19]
```

```
##                  model year price      transmission mileage      fuelType tax
## 2275 Mercedes- X-CLASS 2019 37891 f.Trans-Automatic     303 f.Fuel-Diesel 265
##       mpg engineSize   manufacturer Audi total years_sell years_sell2
## 2275 31.4       Gran f.Man-Mercedes   No     0    Molt nou           1
##                 aux       f.price       f.miles           f.tax          mpg_d
## 2275 [0,5.89e+03] Segmento -  A f.miles-[0,6] f.tax-(150,570] mpg_d-[0,44.8]
```

Mercedes X class is a big car with a big tax value, a very high mileage and a high price, this makes it a unusual model that doesn't follow the usual distribution.

In dimension 2 We will

```
rang<-order(res.pca$ind$coord[,2])
contrib.extremes<-c(row.names(df)[rang[1]], row.names(df)[rang[length(rang)]])
contrib.extremes<-c(row.names(df)[rang[1:10]], row.names(df) [rang[(length(rang)-10):length(rang)]])
fviz_pca_ind(res.pca, select.ind = list(names=contrib.extremes))
```




```
df[which(row.names(df) %in% row.names(df)[rang[length(rang)]]), 1:19]
```

```
##        model year price      transmission mileage      fuelType tax mpg
## 991 Audi- Q7 2013  8995 f.Trans-Automatic  136000 f.Fuel-Diesel 540  31
##     engineSize manufacturer Audi total years_sell years_sell2
## 991       Gran   f.Man-Audi  Yes     1       Vell           3
##                         aux       f.price          f.miles           f.tax
## 991 (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323] f.tax-(150,570]
##               mpg_d
## 991 mpg_d-[0,44.8]
```

```
df[which(row.names(df) %in% row.names(df)[rang[1]]),1:19]
```

```
##        model year price      transmission mileage      fuelType tax   mpg
## 1850 BMW- i3 2016 19850 f.Trans-Automatic   19995 f.Fuel-Hybrid  20 470.8
##      engineSize manufacturer Audi total years_sell years_sell2
## 1850      Petit    f.Man-BMW   No     2   Semi nou           2
##                          aux       f.price         f.miles         f.tax
## 1850 (1.69e+04,3.4e+04] Segmento -  C f.miles-(17,34] f.tax-(1,125]
##                  mpg_d
## 1850 mpg_d-(61.4,471]
```

Detection of multivariant outliers and influent data.

Since we've commented before that we don't consider multivariate outliers, no action should be taken here.

2.3 Interpreting the axes: Variables point of view coordinates, quality of representation, contribution of the variables

```
round(cbind(res.pca$var$coord[,1:2],res.pca$var$cos2[,1:2],res.pca$var$contrib[,1:2]),2)
```

```
##              Dim.1 Dim.2 Dim.1 Dim.2 Dim.1 Dim.2
```

```
## mileage       0.81  0.46  0.66  0.21 31.99 20.39
## tax          -0.60  0.55  0.36  0.31 17.37 30.10
## mpg           0.56 -0.61  0.31  0.38 15.21 37.00
## years_sell2  0.85  0.36  0.73  0.13 35.43 12.51
```

```
round(cbind(res.pca$var$cos2[,1:2],res.pca$var$contrib[,1:2]),2)
```

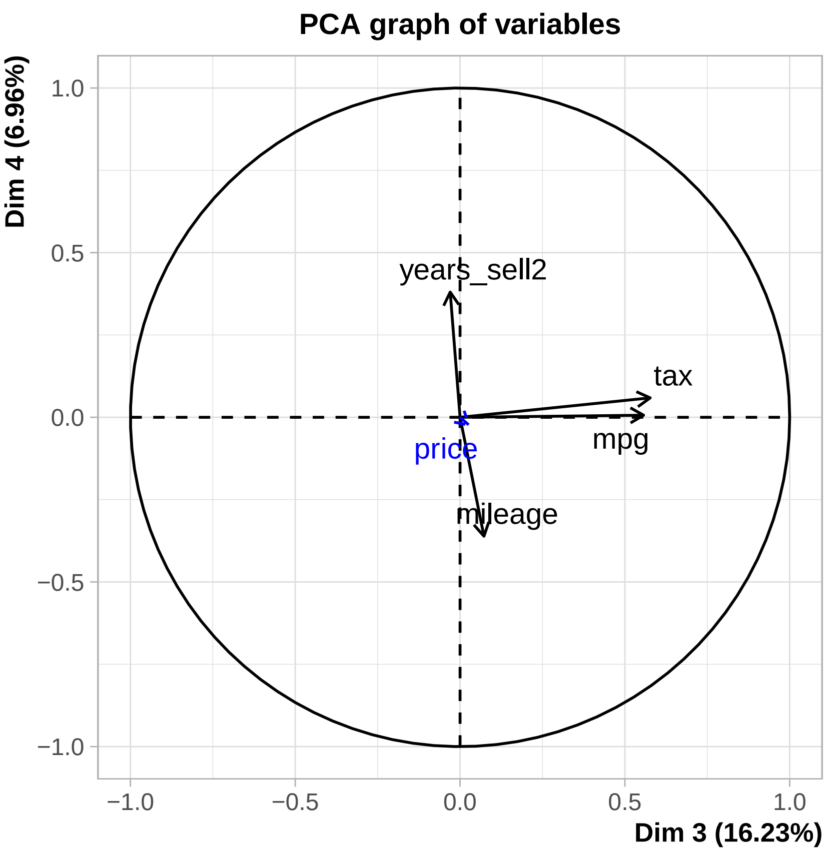
```
##              Dim.1 Dim.2 Dim.1 Dim.2
## mileage       0.66  0.21 31.99 20.39
## tax           0.36  0.31 17.37 30.10
## mpg           0.31  0.38 15.21 37.00
## years_sell2  0.73  0.13 35.43 12.51
```

```
# dimdes easies this description from the variables
res.des<-dimdesc(res.pca)
## #

res.des$Dim.1$quanti
```

```
##              correlation      p.value
## years_sell2   0.8524794 0.000000e+00
## mileage       0.8099881 0.000000e+00
## mpg           0.5585470 0.000000e+00
## total         0.2578467 3.487248e-76
## tax          -0.5968714 0.000000e+00
## price        -0.6450006 0.000000e+00
```

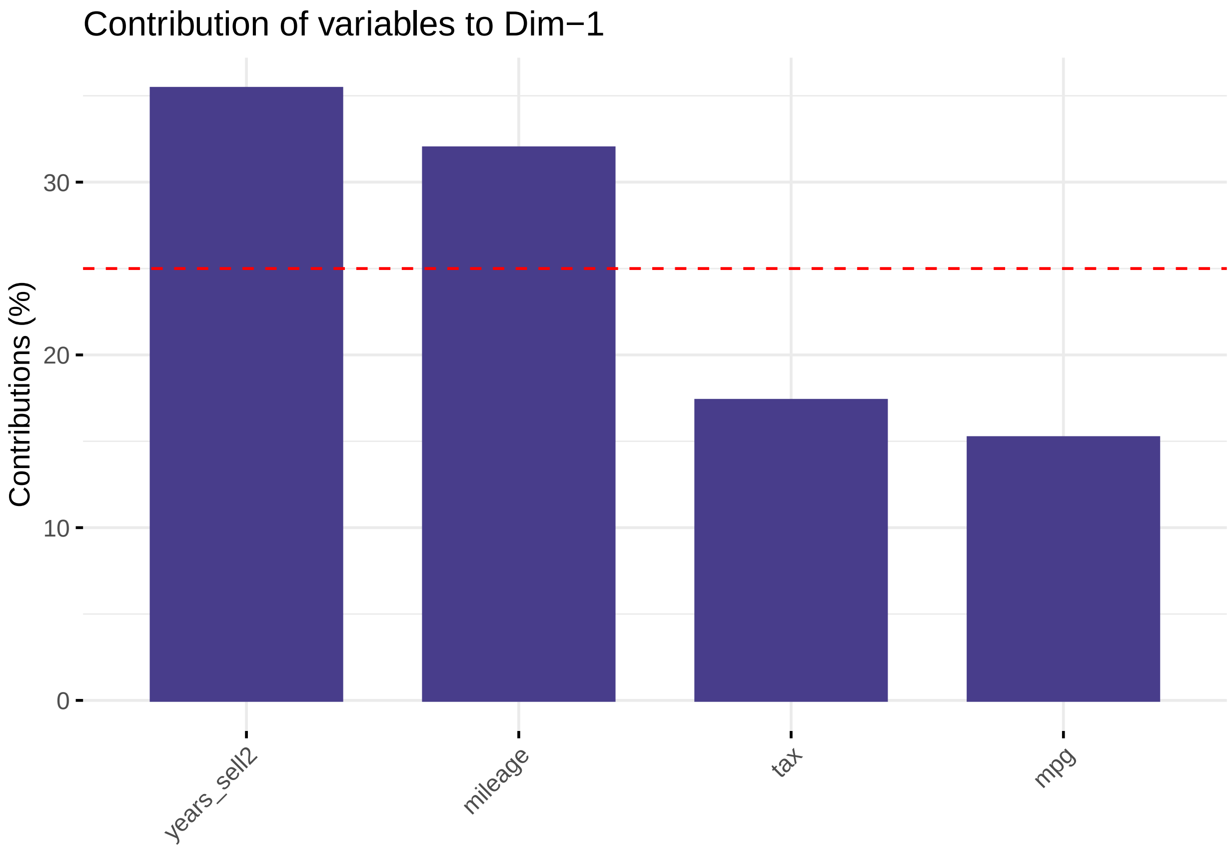
```
## # we can need more than 2 axes to have a good representation of the clouds
#plot.PCA(res.pca,choix=c("ind"),cex=0.8)
#plot.PCA(res.pca,choix=c("ind"),invisible=c("quali"),axes=c(3,4))
plot.PCA(res.pca,choix=c("var"),axes=c(3,4))
```



We cansee that this dimensions reprensent a new type of individuals because the relation betweeen distance and years sell is contrary to what we waw in the first PCA with dimensions 1 and 2. Tax and mpg are correlated in a positive way too despite what we saw in the first graphic.

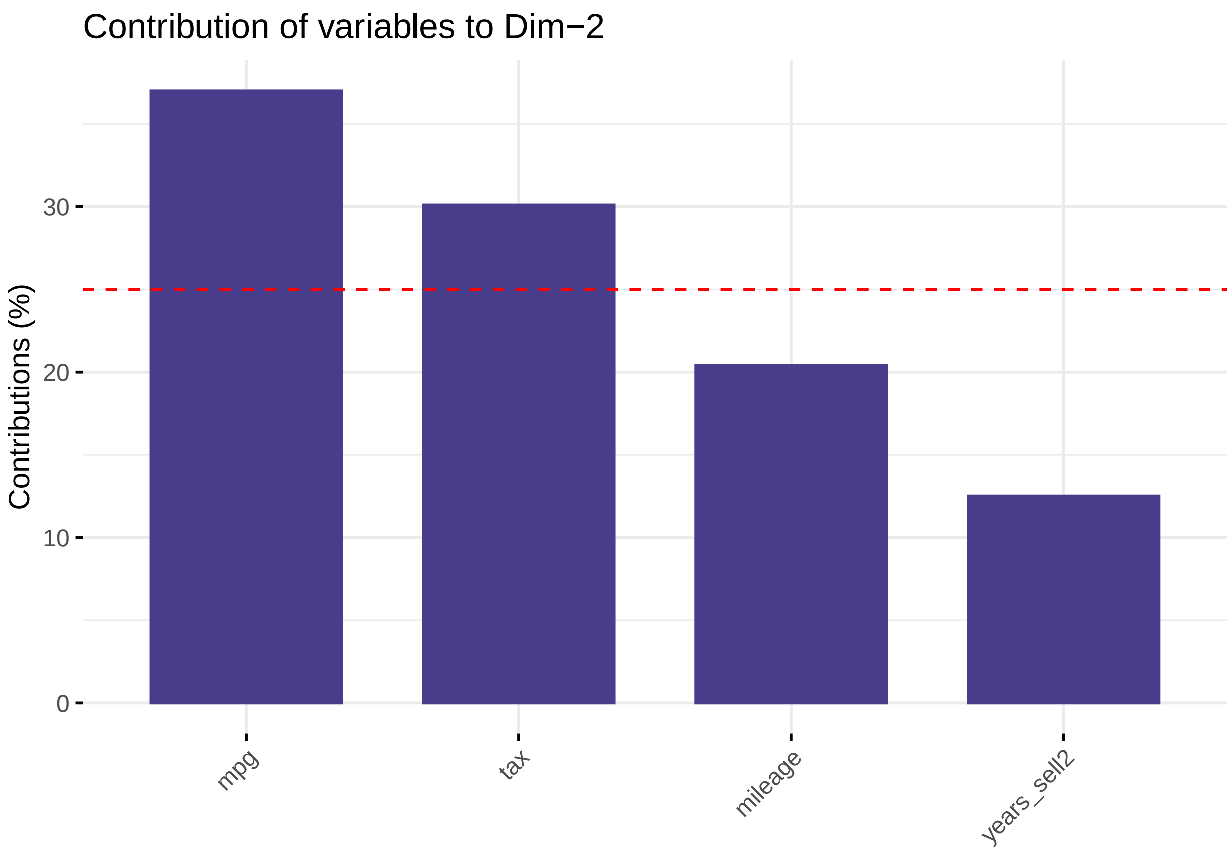
```r
res.des <- dimdesc(res.pca)
```

```r
fviz_contrib(res.pca, fill = "darkslateblue" , color = "darkslateblue", choice = "var", axes = 1, top =
```



We see that price and mileage are the most contributieve to the first dimension.

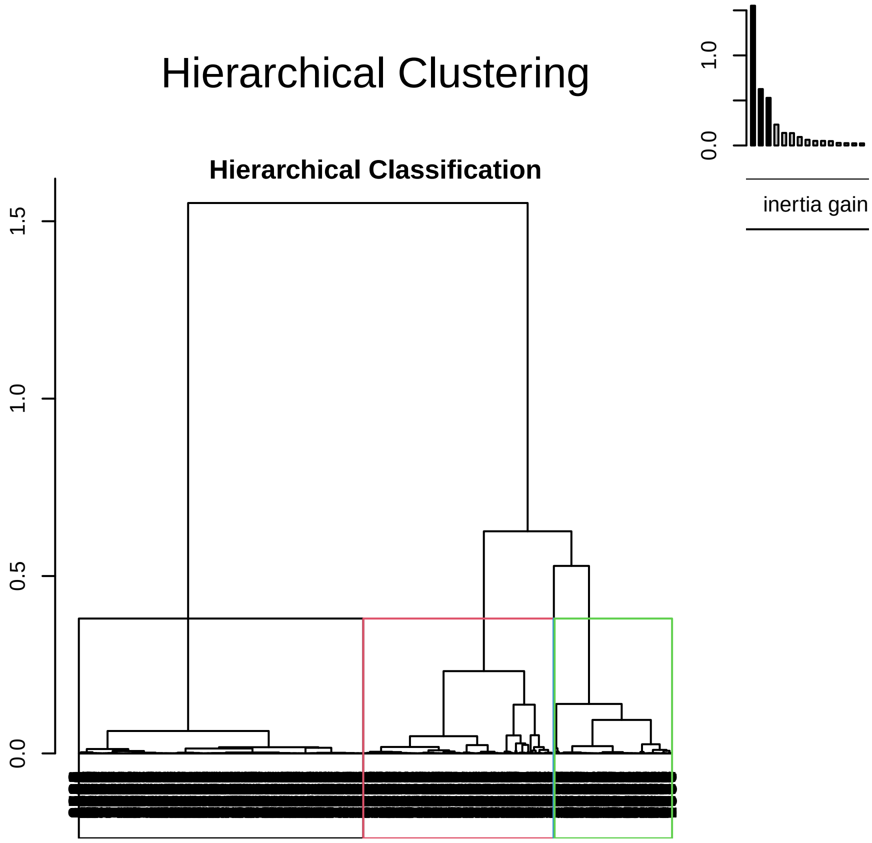
```r
fviz_contrib(res.pca, fill = "darkslateblue", color = "darkslateblue", choice = "var", axes = 2, top = 5
```

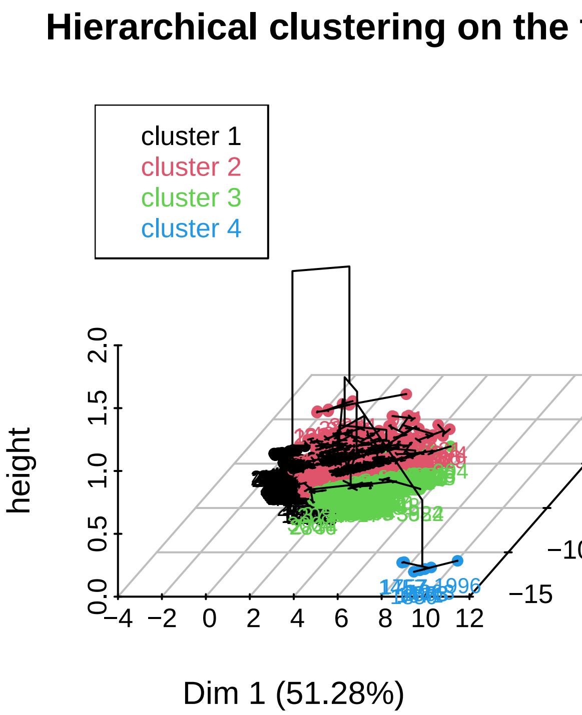


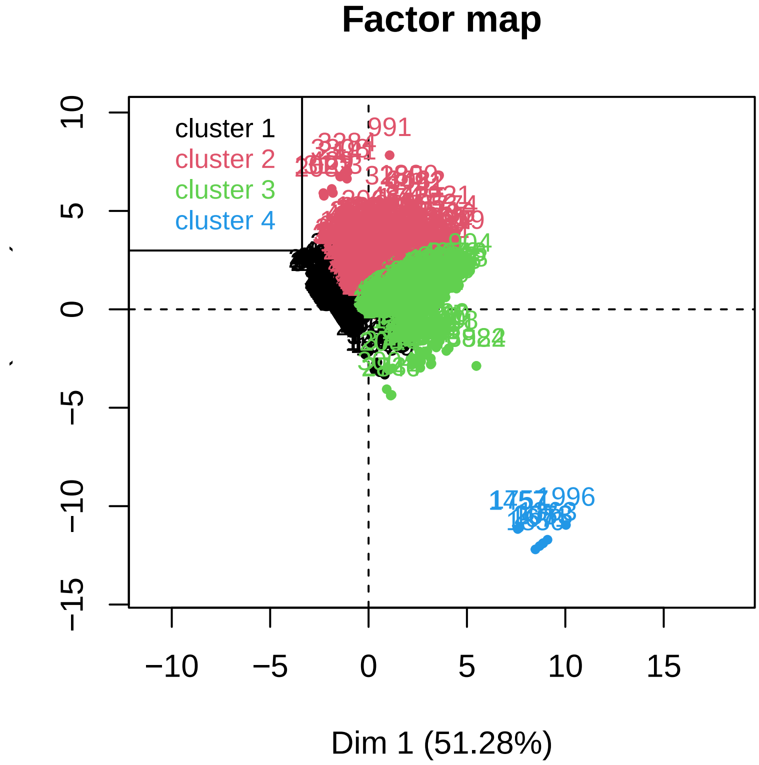
We see that mpg and mileage are the most contributive to the second dimension.

3 Hierarchial Clustering

```r
res.hcpc <- HCPC(res.pca,nb.clust = 4, order = TRUE)
```








```r
#res.hcpc <- HCPC(res.pca,nb.clust = 2, order = TRUE)
```

Note: If we chose the default number of cluster it would be 3, as we can guess from the inertia reduction plot. In our case, due to the amount of data we have and when we reduce the clusters to 3, it gives us two big clusters and a small one (the black one above)which doesn't contain much cars and informations. Choosing four clusters keeps that small cluster but makes the two initial big clusters divide into three big clusters which is much more interesting than only dealing with two big clusters.

3.1 Description of clusters

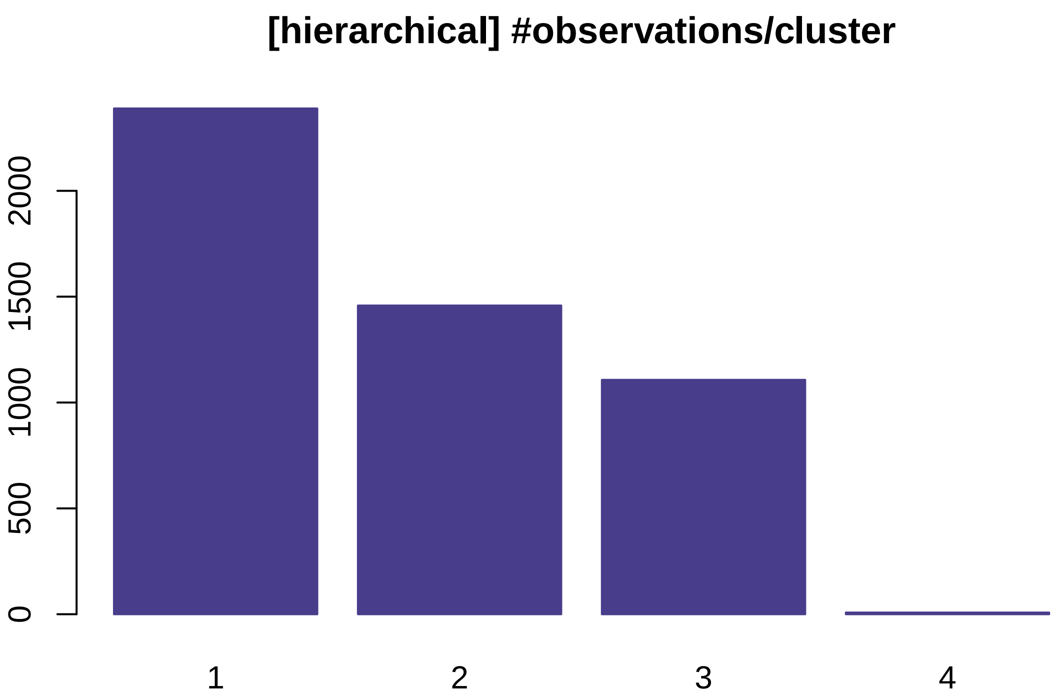
3.1.1 Number of observations in each cluster:

```
table(res.hcpc$data.clust$clust)
```

```
##
##    1    2    3    4
## 2389 1458 1107    8
```

We see that the first cluster doesn't represent many vehicles as we explained before, the three other cluser are well represented which is interesting.

```
barplot(table(res.hcpc$data.clust$clust), col="darkslateblue", border="darkslateblue", main="[hierarchic
```




```
res.hcpc$desc.var$test.chi2
```

```
##                    p.value  df
## model         0.000000e+00 258
## year          0.000000e+00 237
## years_sell    0.000000e+00   6
## aux           0.000000e+00   9
## f.price       0.000000e+00   9
## f.miles       0.000000e+00   9
## fuelType     1.506740e-125   6
## transmission 2.730693e-109   6
## engineSize    3.172321e-81   6
## manufacturer  1.836385e-31   9
## Audi          4.711223e-02   3
```

for model and year: df>90. The distributions follows a normal low We can see that price,age and tax constitute the main characterizes of the clusters.

Next, we want to see for each cluster which are the categories that characterize them.The clusters that contain more individuals are the first, the second and the fourth one. Cluster number 4 has less individuals. We proceed to analyze them.

3.1.2 Description of the clusters with the qualitative variables:

```
res.hcpc$desc.var$category
quali_var_decription_1<-res.hcpc$desc.var$category
```

Cluster 1: -> The cluster 1 only contains the BMW i3 model Cluster 2: -> This cluster almost only contains the smallest *tax_pay* (98%), 76% of its cars are in the cheapest category *Segement D* and 92% are *Semi Nou* Cluster 3: -> This cluster contains *mid priced* vehicules (Segment B and Segment C) and 86% of the vehicules the pay the *most taxes* Cluster 4: ->This cluster contains almost all the most expensive cars (*price*: Segment A), they represent 75% of the cars of this cluster. These cars are in the middle category for *tax_pay* (145-150) and are in the *age* category: Molt Nou

3.1.3 Description of the clusters with the quantitative variables:

```
res.hcpc$desc.var$quanti
quanti_var_decription_1<-res.hcpc$desc.var$quanti
```

Cluster 1: -> The first cluster contains cars with a huge *mpg* and an above average *tax* pay, we can thus think of the Hybrid cars such as the BMW i3. Cluster 2: -> The second cluster represents cars that are cheap, have a low *tax_pay*, an above average *mpg* and an the biggest *mileage*. These are the second most common cars as they used to be effective but are turning *old*. Cluster 3: -> The third cluster represents the most cars as the *mileage* and *mpg* are approximately near the average, the car seem to be of a diesel nature as the *tax* is high Cluster 4: ->The fourth cluster represents ¨*new* cars that are pretty *expensive*.

3.1.4 The description of the clusters by the individuals

```
res.hcpc$desc.ind$para
```

```
## Cluster: 1
##       4101       1479        243       1342       3716
## 0.02985832 0.03005169 0.03147738 0.03254836 0.03677123
## ------------------------------------------------------------
## Cluster: 2
##      3269      1231      4598       675      1253
## 0.1468722 0.1768802 0.1852476 0.1886373 0.1900470
## ------------------------------------------------------------
## Cluster: 3
##      1819       907      1391       590      1037
## 0.1271236 0.1305122 0.1659142 0.1768643 0.1788078
## ------------------------------------------------------------
## Cluster: 4
##      1553      2073      1671      1363      1850
## 0.4969180 0.4973243 0.5845558 0.6403553 0.8447913
```

What we obtain are the more representative individuals,paragons, for each cluster. We get the rownames of each paragon in every single cluster.

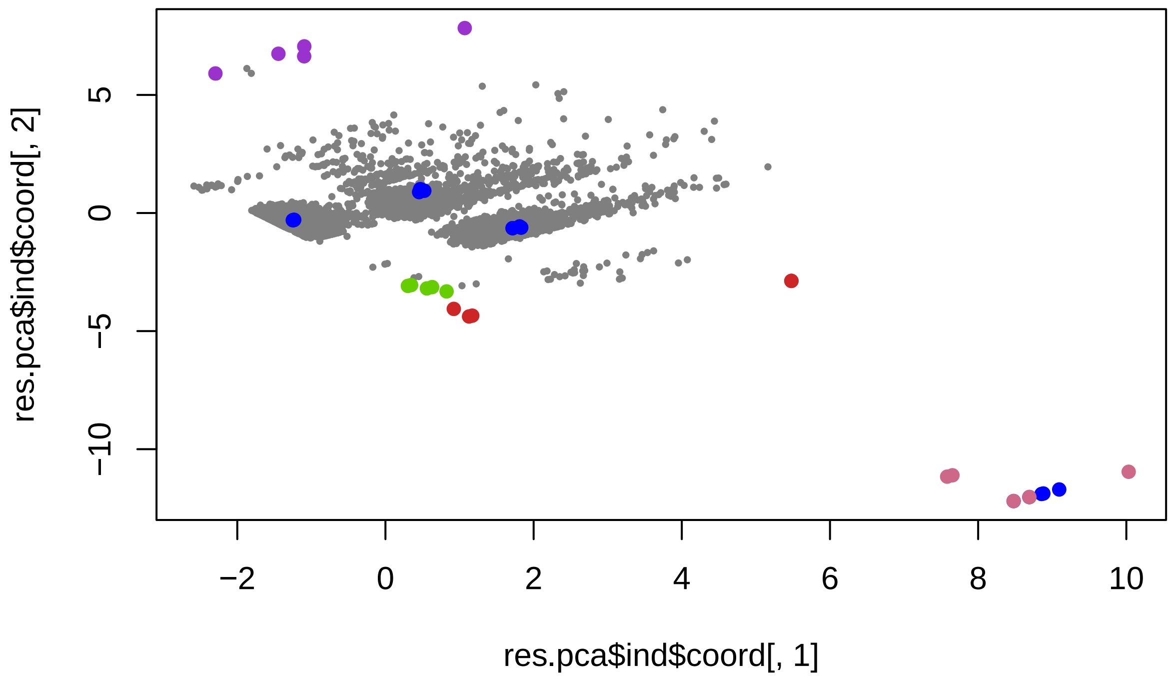
```
res.hcpc$desc.ind$dist
```

```
## Cluster: 1
##     1799     1240     1803     1244     1436
## 4.920916 4.689066 4.650195 4.511287 4.482129
## ------------------------------------------------------------
## Cluster: 2
##      991     3284     3308     2141     1022
## 9.701481 8.707059 8.515805 8.440471 7.907765
## ------------------------------------------------------------
```

```
## Cluster: 3
##     3922     3884     2846     2734     3004
## 7.123103 7.122217 6.333806 6.330527 5.807111
## ------------------------------------------------------------
## Cluster: 4
##     1457     1757     1996     1850     1671
## 18.27829 18.27193 18.26846 18.19178 18.17783
```

```
para1<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[1]]))
dist1<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[1]]))
para2<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[2]]))
dist2<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[2]]))
para3<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[3]]))
dist3<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[3]]))
para4<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[4]]))
dist4<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[4]]))
```

```
plot(res.pca$ind$coord[,1],res.pca$ind$coord[,2],col="grey50",cex=0.5,pch=16)
points(res.pca$ind$coord[para1,1],res.pca$ind$coord[para1,2],col="blue",cex=1,pch=16)
points(res.pca$ind$coord[dist1,1],res.pca$ind$coord[dist1,2],col="chartreuse3",cex=1,pch=16)
points(res.pca$ind$coord[para2,1],res.pca$ind$coord[para2,2],col="blue",cex=1,pch=16)
points(res.pca$ind$coord[dist2,1],res.pca$ind$coord[dist2,2],col="darkorchid3",cex=1,pch=16)
points(res.pca$ind$coord[para3,1],res.pca$ind$coord[para3,2],col="blue",cex=1,pch=16)
points(res.pca$ind$coord[dist3,1],res.pca$ind$coord[dist3,2],col="firebrick3",cex=1,pch=16)
points(res.pca$ind$coord[para4,1],res.pca$ind$coord[para4,2],col="blue",cex=1,pch=16)
points(res.pca$ind$coord[dist4,1],res.pca$ind$coord[dist4,2],col="palevioletred3",cex=1,pch=16)
```



3.2 Partition quality

We are going to evaluate the partition quality.

3.2.1 Gain in inertia (in %)

```
# ( between sum of squares / total sum of squares ) * 100
((res.hcpc$call$t$within[1]-res.hcpc$call$t$within[5])/res.hcpc$call$t$within[1])*100
```

```
## [1] 73.45002
```

The quality of this reduction if of 73.45%.

In case we wanted to achieve an 80% of the clustering representativity we would need 10 clusters.

```
((res.hcpc$call$t$within[1]-res.hcpc$call$t$within[10])/res.hcpc$call$t$within[1])*100
```

```
## [1] 85.60135
```

3.2.2 Save the results into dataframe

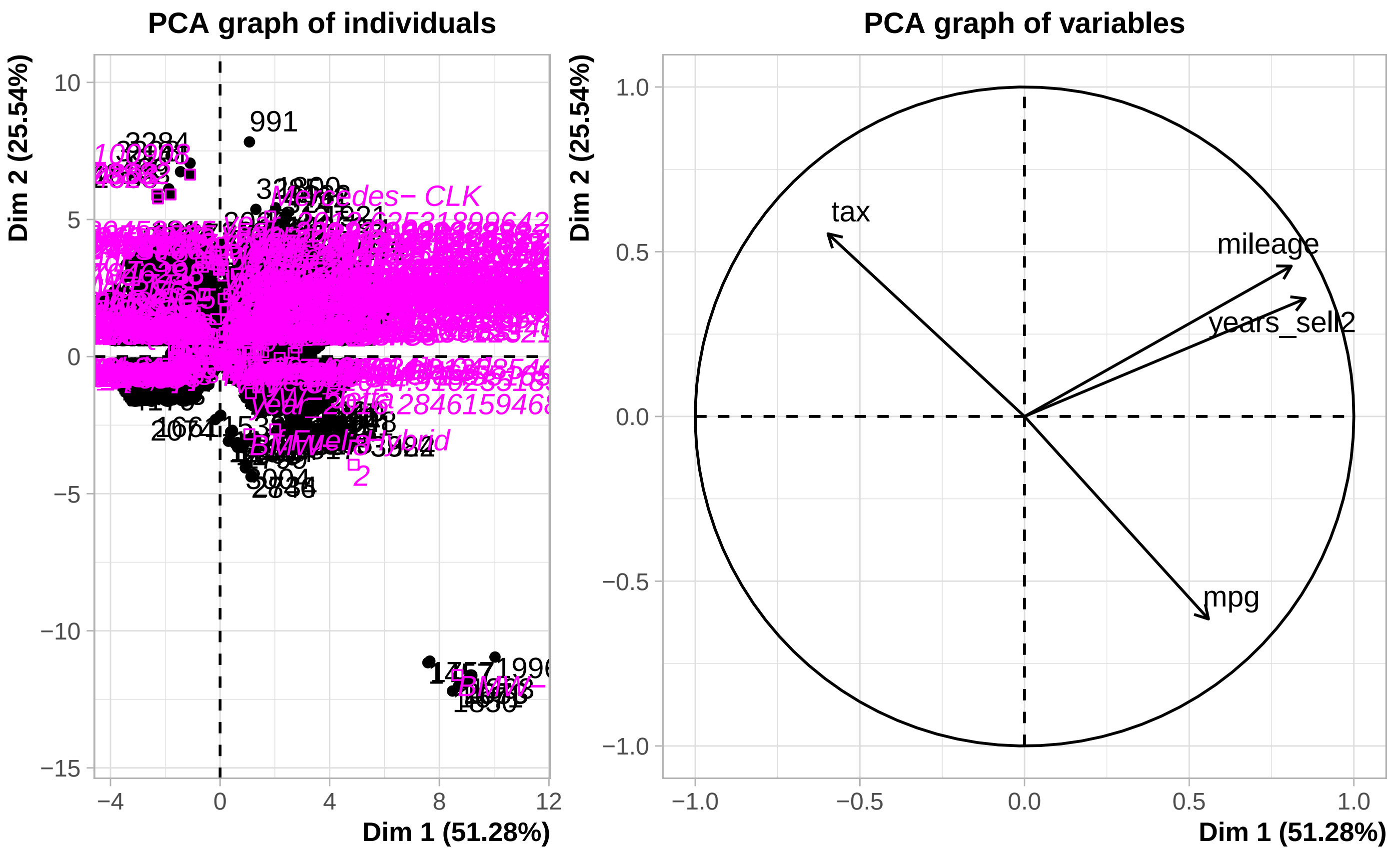
```
res.hcpc$call$t$inert.gain[1:5]
```

```
## [1] 1.5513914 0.6260955 0.5285420 0.2319721 0.1395193
```

```
df$hcpck<-res.hcpc$data.clust$clust
```

4 K-means Classification

```
res.pca <- PCA(df[,c(vars_con,vars_dis)], scale. = T,quali.sup=c(5:15))
```




```
ppcc<-res.pca$ind$coord[,1:2] # We choose the two first  principal components as per kaiser
summary(ppcc)
```

```
##      Dim.1             Dim.2
##  Min.   :-2.5854   Min.   :-12.1972
##  1st Qu.:-1.2838   1st Qu.: -0.4931
##  Median :-0.1962   Median : -0.1717
##  Mean   : 0.0000   Mean   :  0.0000
##  3rd Qu.: 1.1477   3rd Qu.:  0.3382
##  Max.   :10.0318   Max.   :  7.8304
```

```
dim(ppcc)
```

```
## [1] 4962    2
```

We will estimate the optimal number of clusters

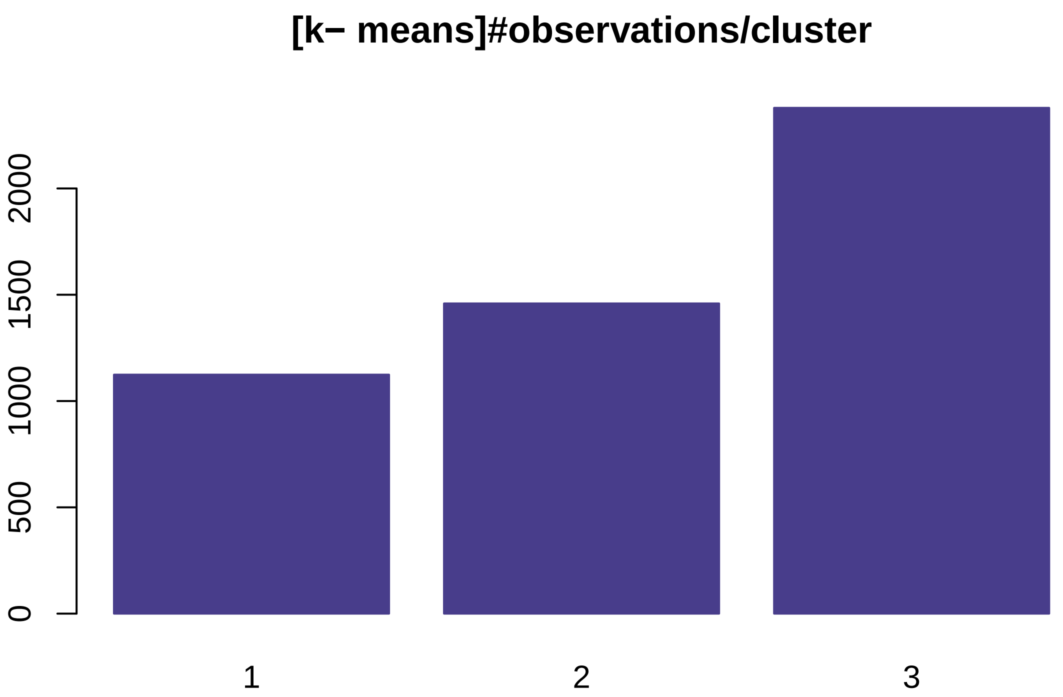
```
#library("factoextra")
#fviz_nbclust(ppcc, kmeans, method = "gap_stat")
```

4.1 Classification

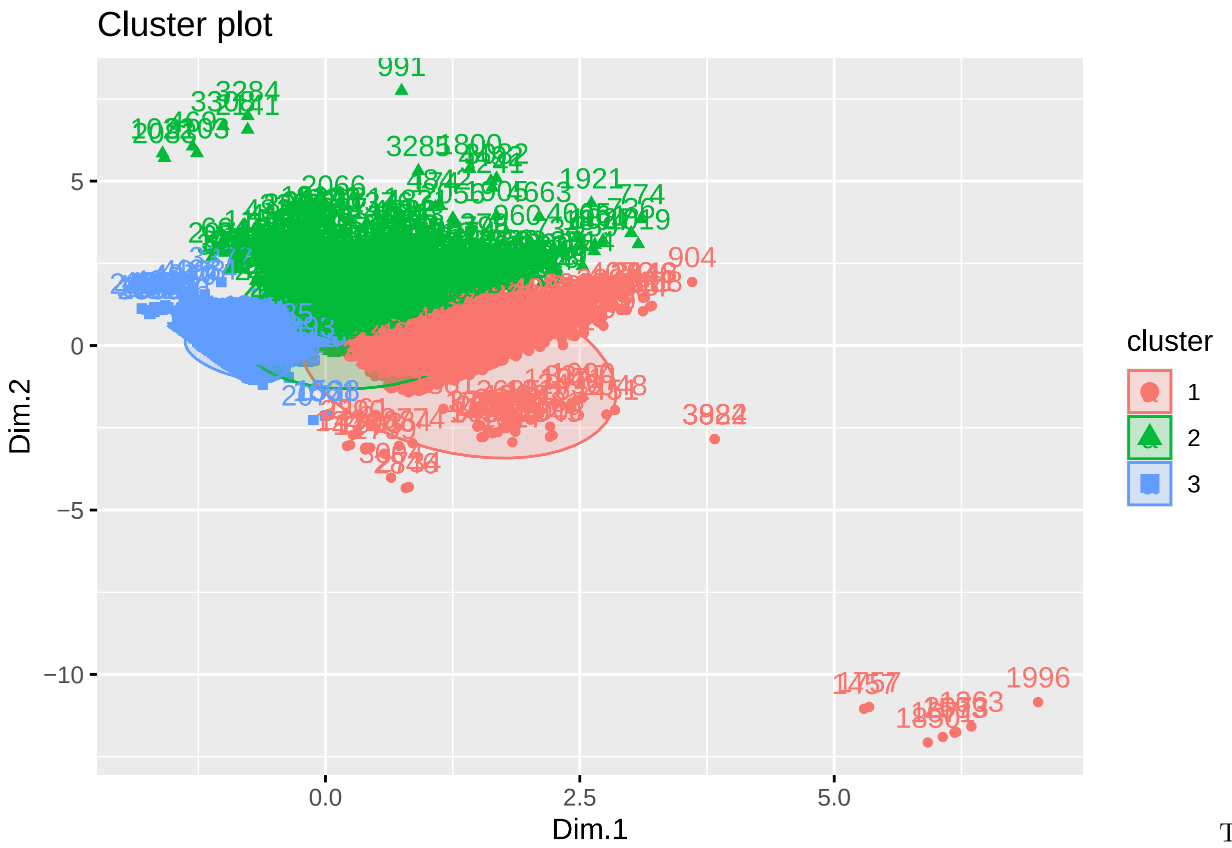
We will compute, dist, a matrix which shows the distances to each one of the clusters

```
#d<-dist(ppcc) # coordenates are real - Euclidean metric
#kc<-kmeans(d,3,iter.max=30,trace=TRUE) #calclulate the distances, into a matrix
#kc
#kc<-kmeans(dist,3,iter.max=30,trace=TRUE)
set.seed(123)
kc <- kmeans(ppcc, 3, nstart = 25)
```

```
df$claKM<-0
df$claKM<-kc$cluster
df$claKM<-factor(df$claKM)
barplot(table(df$claKM),col="darkslateblue",border="darkslateblue",main="[k- means]#observations/cluster
```

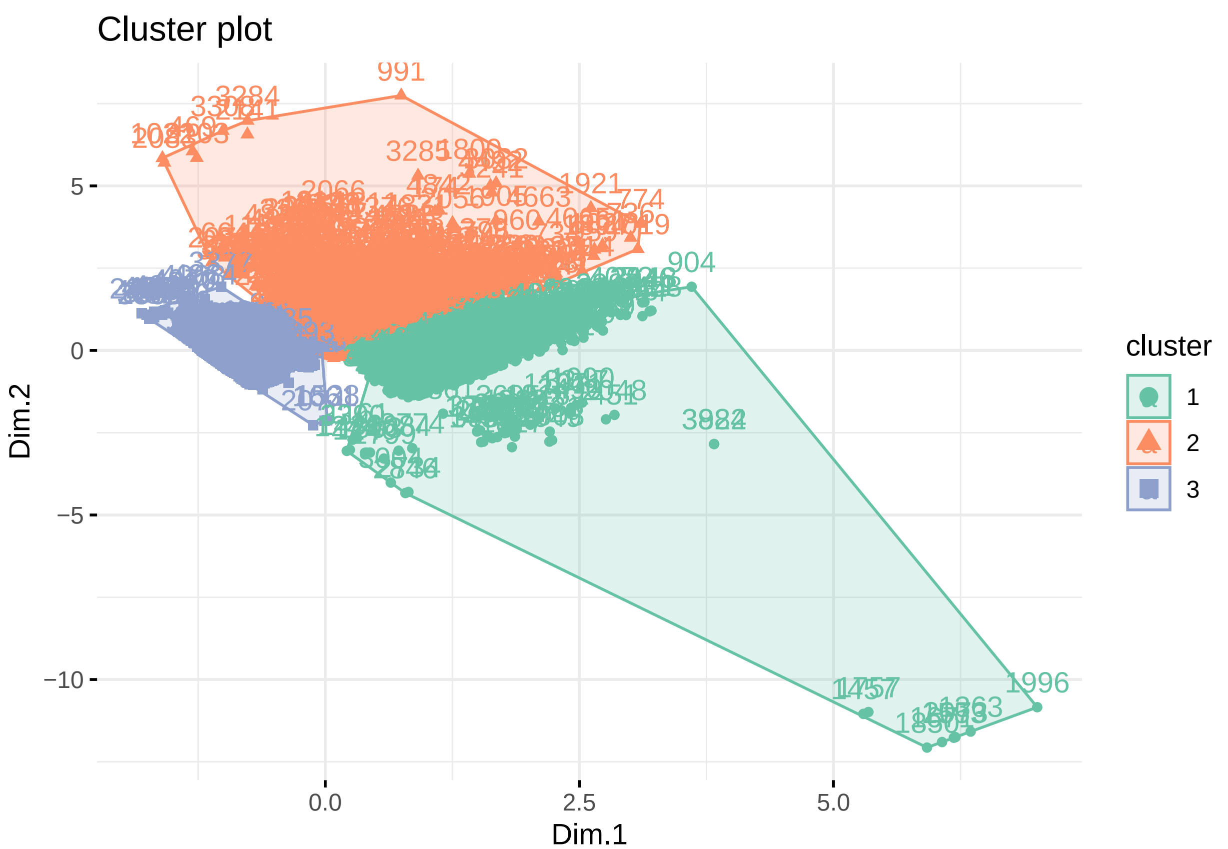


```r
fviz_cluster(kc,ppcc, ellipse.type = "norm")
```



To have a better perception of the clusters:

```r
# Change the color palette and theme
fviz_cluster(kc, ppcc,
   palette = "Set2", ggtheme = theme_minimal())
```



```
100*(kc$betweenss/kc$totss)
```

```
## [1] 67.35681
```

4.2 K-means clusters characteristics

If we want to know the characteristics of each cluster, we need to execute a catdes to obtain these characteristics. In the following output we get them:

```
res.cat <- catdes(df,20) #the 20th variable of df is claKM
res.cat
```

We start wit the description of the categorical variables that characterize the clusters, so in this output we do not have dimensions because it is the total association. We can see the intensity of the variables, in our case the variables that affect more to the clustering are *Tax_pay*, *Age* and *price*.

- Cluster 1: -Every car in this cluster is in the *lowest tax segment*.76% of the cars that pays the less taxes are in this cluster. 94% of the total *Semi nou cars* are in this category, while it contains approximately (44%) half of the *semi nou cars*.73% of the *cheapest cars* are also in this cluster. This explains why this cluster contains a large amount of cars, cars that are majoritarely *manual* (62%). 77% of these cars have been bought *between 2013 and 2016*.

- Cluster 2: -Almost 95% of all the cars in this cluster are in the second *highest tax segment*. Almost all the cars (96%) with the *lowest miles*(0,6) are in this cluster, which explains that the cars in this cluster are almost all *new* (80% of the cars have been bought after 2018). Almost all the *segment A* cars are in this cluster (97%) in addition to the *B category* (77%). This cluster thus contains the *newly bought cars* that are expensive

- Cluster 3: -This car contains the cars that pay the most in *taxes* (75% of them), 85% of the cars in this cluster blong to *cheapest cars* (Segments C and D) which explains why their tax pay is big (as we will see later on in the CA analysis). 42% of the cars in this cluster are in the category of cars with the *most mileage*. This cluster thus contains *old, cheap cars that pay the most taxes*

We can notice that the cluster have been chosen in the basis of the tax segments in addition to the age of these cars. We will later develop this point in the CA analysis to dress a comparaison between age and tax in addition to price and tax

We now proceed to see the quantitative variables that characterizes the clusters. • Cluster 1: -The *Tax_pay* is below average by about 90 euros which consolidates our analysis of the categorical variable. The *Years_sell* Mean is slightly above average by half a year, the *mileage* is consequently also higher than the average. The *price* is below average by about 8000 euros with a smaller sd than in other categories.This confirms what we saw in the categorical variables as cars in this category are *getting old, unexpensive and tax economical.* • Cluster 2: -The *tax*variable is slightly above average by 24 dollars, *price* is by about 7000 dollars with a big sd which is normal (as the price grows, the sd naturally grows), we thus expect cars with a really huge price in this category. *Mileage* is below average as well as the *age* variable which confirms what we saw in the categorical variables, cars in this category are *new and expensive but not too tax consuming.* • Cluster 3: -the *tax* variable is above average by about 30 dollars, *mileage* is also way above average as well as *years*. *price*is below average by about 6000 dollars. This confirms what we saw with the categorical variables as the cars in this category are *old, cheap and tax consuming*

This accordance between the categorical and continuous variables makes us confirm that these clusters have been assigned based principally on *tax_pay*, *age* and *price*

4.3 Comparaison of clusters (confusion table)

We want to compare the hierarchical clustering, previously done, and the k-means clustering, so proceed to do the following.

```
df$hcpck<-res.hcpc$data.clust$clust
tt<-table(df$hcpck,df$claKM)
tt
```

```
##
##         1    2    3
```

```
##   1    7    3 2379
##   2    3 1455    0
##   3 1106    1    0
##   4    8    0    0
```

In order to have a better visualization of the table we add names to the columns and the rows:

```
df$hcpck<-factor(df$hcpck,labels=c("kHP-1","kHP-2","kHP-3","kHP-4"))
df$claKM<- factor(df$claKM,levels=c(1,2,3),labels=c("kKM-1","kKM-2","kKM-3"))
tt<-table(df$hcpck,df$claKM); tt
```

```
##
##          kKM-1 kKM-2 kKM-3
##   kHP-1     7     3  2379
##   kHP-2     3  1455     0
##   kHP-3  1106     1     0
##   kHP-4     8     0     0
```

```
100*sum(diag(tt)/sum(tt))
```

```
## [1] 29.46393
```

We have a concordance of 27% between the two ways of clustering which is not really good

5 CA analysis

CA analysis for your data should contain your factor version of the numeric target (previous) in K=5 (variable aux_price created before) levels and 2 factors:

We set the numeric variable as the price of the car

With the price factor, we proceed to create a variable that associates the price with different factors such as tax price (f.tax), engineSize and Years_Sell. For each of these variables, we create a contingency table and look up for correlations and links between the different categories of these variables.

5.1 Price vs f.Tax

```
tt<-table(df[,c("f.price","f.tax")]);tt
```

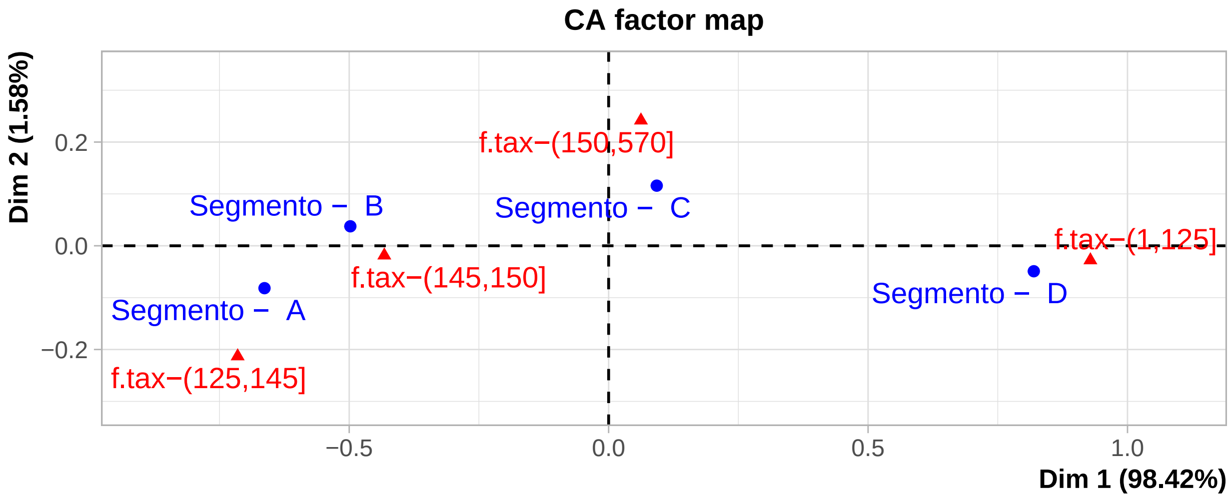
```
##                  f.tax
## f.price           f.tax-(1,125] f.tax-(125,145] f.tax-(145,150] f.tax-(150,570]
##   Segmento -  D             983               2             397             122
##   Segmento -  C             372               5             658             141
##   Segmento -  B              67              12             864              97
##   Segmento -  A               5              19            1145              73
```

```
chisq.test(tt, simulate.p.value = TRUE)
```

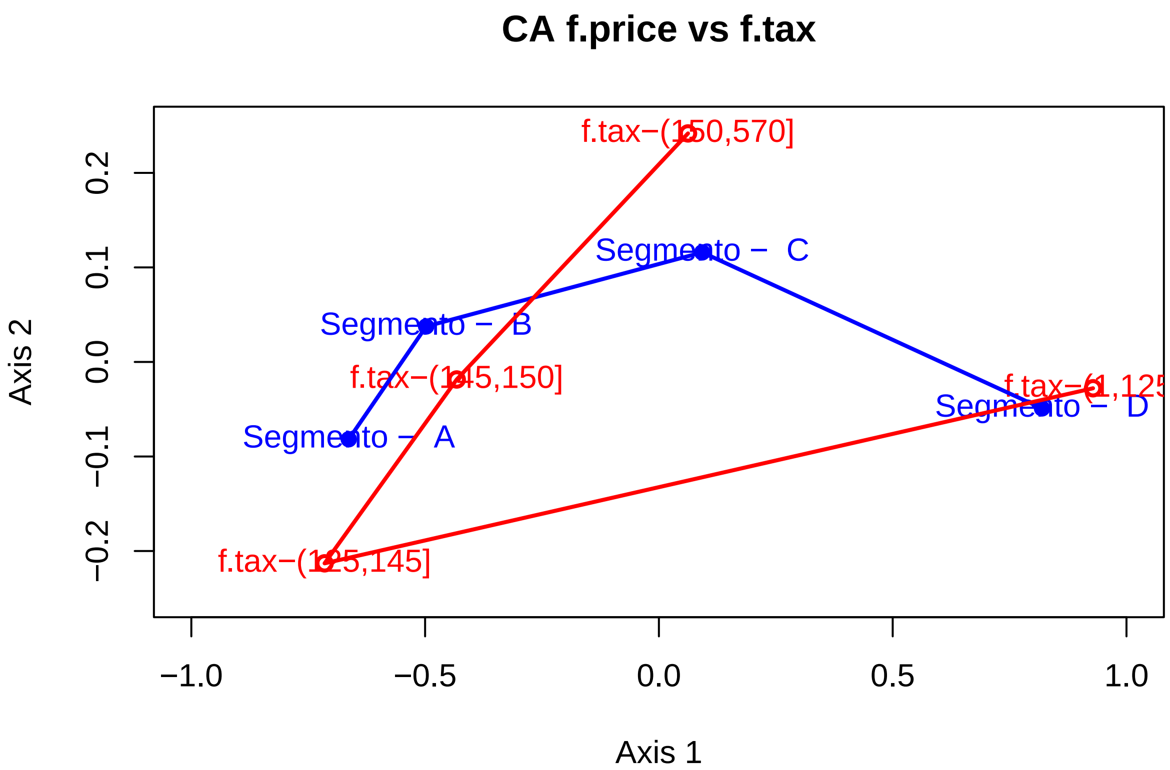
```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  tt
## X-squared = 1853.2, df = NA, p-value = 0.0004998
```

We get a p-value smaller than 0.05 so we can deny the H0 hypothesis. There is thus a link between the columns and the rows We are now going to take a look to the simple correspondences.

```
res.ca <- CA(tt)
```




```
plot(res.ca$row$coord[,1],res.ca$row$coord[,2],pch=19,col="blue",xlim=c(-1,1),ylim=c(-0.25,0.25),xlab="
points(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red")
text(res.ca$row$coord[,1],res.ca$row$coord[,2],lwd=2,col="blue",labels=levels(df$f.price))
text(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red",labels=levels(df$f.tax))
lines(res.ca$row$coord[,1],res.ca$row$coord[,2],lwd=2,col="blue")
lines(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red")
```



The majority of the expensive cars (Segment A and B) are new and more environment friendly thanks to new technologies, they consequently have a less expensive tax price . Cheapest cars(Segment D), have a small mpg and have thus the least tax price (<145). We can't give additional informations about the Segment C as there is no tax category that is really near it. .

```
summary_price_tax<-summary(res.ca)$call
```

```
##
## Call:
## CA(X = tt)
##
## The chi square of independence between the two variables is equal to 1853.25 (p-value =  0 ).
```

```
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3
## Variance               0.368   0.006   0.000
## % of var.             98.424   1.575   0.000
## Cumulative % of var.  98.424 100.000 100.000
##
## Rows
##                   Iner*1000     Dim.1     ctr    cos2     Dim.2     ctr    cos2
## Segmento -  D   |   204.254 |   0.819  55.365   0.996 |  -0.049  12.426   0.004
## Segmento -  C   |     5.222 |   0.093   0.554   0.390 |   0.116  54.129   0.610
## Segmento -  B   |    52.234 |  -0.498  14.129   0.994 |   0.038   5.024   0.006
## Segmento -  A   |   111.779 |  -0.663  29.953   0.985 |  -0.082  28.421   0.015
##                     Dim.3     ctr    cos2
## Segmento -  D   |   0.000   1.899   0.000 |
## Segmento -  C   |  -0.001  21.617   0.000 |
## Segmento -  B   |   0.002  59.888   0.000 |
## Segmento -  A   |  -0.001  16.596   0.000 |
##
## Columns
##                   Iner*1000     Dim.1     ctr    cos2     Dim.2     ctr    cos2
## f.tax-(1,125]   |   248.146 |   0.928  67.443   0.999 |  -0.028   3.789   0.001
## f.tax-(125,145] |     4.264 |  -0.715   1.065   0.918 |  -0.213   5.908   0.082
## f.tax-(145,150] |   115.640 |  -0.432  31.400   0.998 |  -0.019   3.614   0.002
## f.tax-(150,570] |     5.438 |   0.062   0.092   0.062 |   0.242  86.690   0.938
##                     Dim.3     ctr    cos2
## f.tax-(1,125]   |   0.000   0.010   0.000 |
## f.tax-(125,145] |   0.014  92.262   0.000 |
## f.tax-(145,150] |   0.000   3.236   0.000 |
## f.tax-(150,570] |   0.001   4.492   0.000 |
```

We can see from the summary is that we have a chi square statistic of 2659.613, great enough to reject the H0, which means the intensity of the relation between tax and price is high. If we take a look at the variances from the different dimensions, we can see that all together sum more than 1.

5.2 Price vs EngineSize

```r
tt<-table(df[,c("f.price","engineSize")]);tt
```

```
##                 engineSize
## f.price           Petit Mitjà Gran
##    Segmento -  D    956   518   30
##    Segmento -  C    451   649   76
##    Segmento -  B    368   580   92
##    Segmento -  A     94   807  341
```

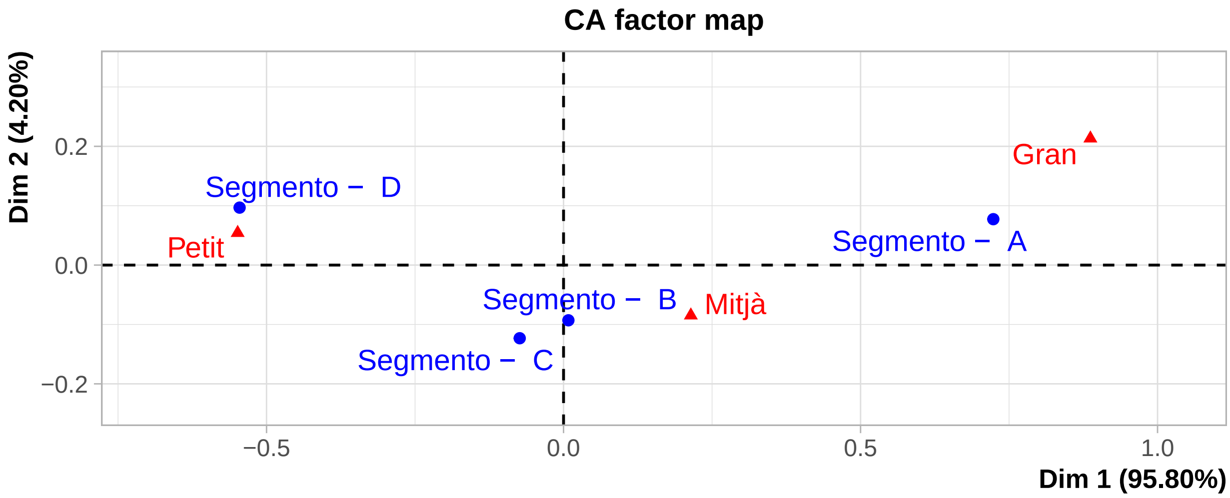
We want to see if the rows and columns are independents, we will do a p-value test H0: Rows and columns are independent

```r
chisq.test(tt, simulate.p.value = TRUE)
```

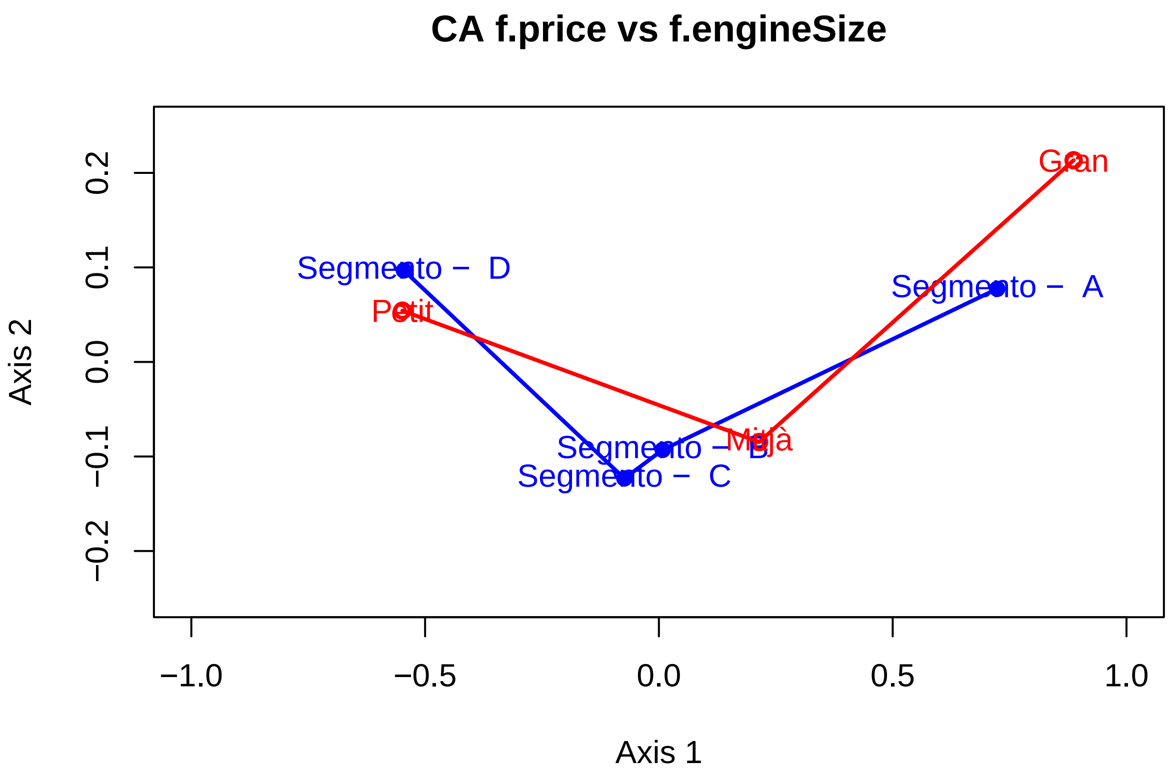
```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  tt
## X-squared = 1152.3, df = NA, p-value = 0.0004998
```

We get a p-value smaller than 0.05 so we can deny the H0 hypothesis. There is thus a link between the columns and the rows We are now going to take a look to the simple correspondences.

```
res.ca <- CA(tt)
```




```
plot(res.ca$row$coord[,1],res.ca$row$coord[,2],pch=19,col="blue",xlim=c(-1,1),ylim=c(-0.25,0.25),xlab="A
points(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red")
text(res.ca$row$coord[,1],res.ca$row$coord[,2],lwd=2,col="blue",labels=levels(df$f.price))
text(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red",labels=levels(df$engineSize))
lines(res.ca$row$coord[,1],res.ca$row$coord[,2],lwd=2,col="blue")
lines(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red")
```



We can see in the plot, that the category "Gran" corresponding to big engines belongs to the summum of highest price category (Segment A), while the smaller engines "MITJÀ" are cheaper (Segment A and Segment B). The smallest engines belong to the bottom of the cheapest category (Segment D). All these results seems logical and follow the cars' distribution of prices we know

```
summary_price_enginesize<-summary(res.ca)$eigenvalues
```

```
##
## Call:
## CA(X = tt)
##
## The chi square of independence between the two variables is equal to 1152.268 (p-value =  1.022839e-2
```

```
##
## Eigenvalues
##                        Dim.1   Dim.2
## Variance               0.222   0.010
## % of var.             95.803   4.197
## Cumulative % of var.  95.803 100.000
##
## Rows
##                   Iner*1000     Dim.1     ctr    cos2     Dim.2     ctr    cos2
## Segmento -  D |     92.992 |  -0.545  40.523   0.969 |   0.097  29.133   0.031 |
## Segmento -  C |      4.889 |  -0.074   0.580   0.264 |  -0.123  36.915   0.736 |
## Segmento -  B |      1.826 |   0.008   0.006   0.008 |  -0.093  18.596   0.992 |
## Segmento -  A |    132.511 |   0.723  58.890   0.989 |   0.077  15.356   0.011 |
##
## Columns
##                   Iner*1000     Dim.1     ctr    cos2     Dim.2     ctr    cos2
## Petit         |    114.488 |  -0.549  50.964   0.990 |   0.054  11.370   0.010 |
## Mitjà         |     27.335 |   0.214  10.626   0.865 |  -0.085  37.902   0.135 |
## Gran          |     90.396 |   0.887  38.410   0.945 |   0.213  50.727   0.055 |
```

We can see from the summary is that we have a chi square statistic of 1155.745, great enough to reject the H0 hypothesis, which means the intensity of the relation is high. If we take a look at the variances from the different dimensions, we can see that all together sum is 1.

```
mean(res.ca$eig[,1])
```

```
## [1] 0.1161092
```

Following the kaiser criteria and the value got in the output, we should retain dimensions with a variance greater than 0.116062. In this case, the first dimension fulfills this because its variance is 0.419, but it is not enough to work with data so, we would choose 2 dimensions for this case.

5.3 Price vs Years-sell

```
tt<-table(df[,c("f.price","years_sell")]);tt
```

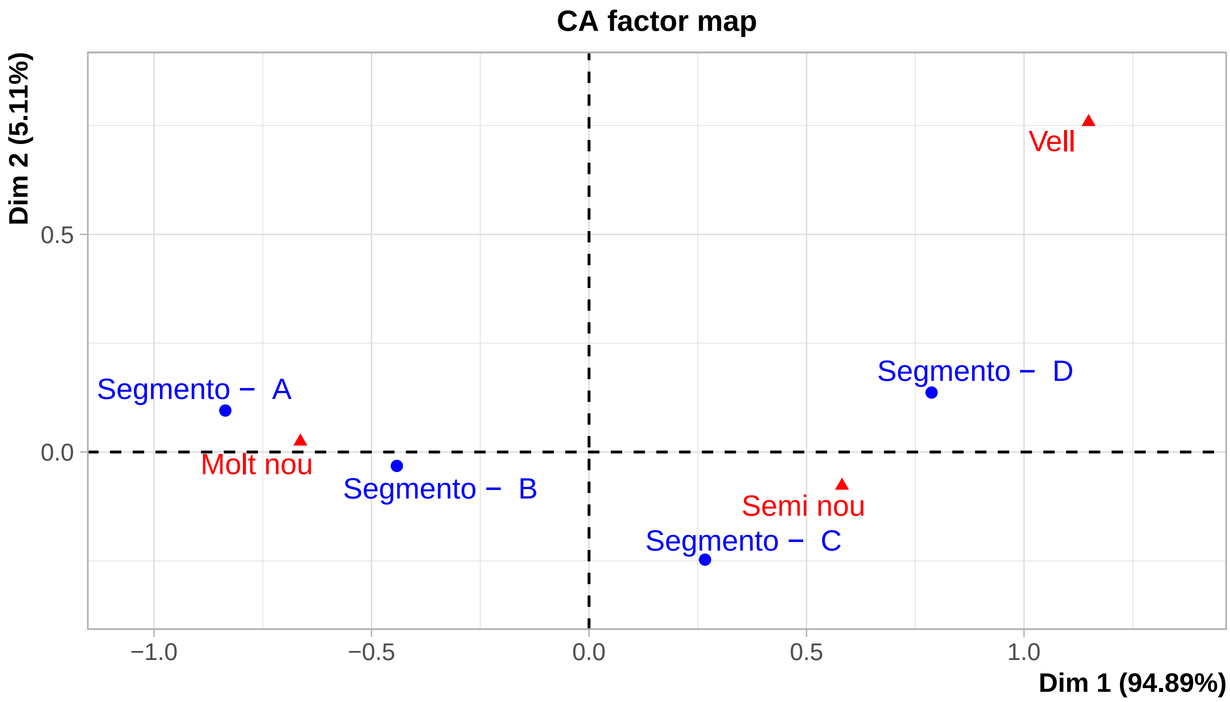
```
##                  years_sell
## f.price           Molt nou Semi nou Vell
##   Segmento -  D        157     1190  157
##   Segmento -  C        390      777    9
##   Segmento -  B        726      312    2
##   Segmento -  A       1121      121    0
```

```
chisq.test(tt, simulate.p.value = TRUE)
```

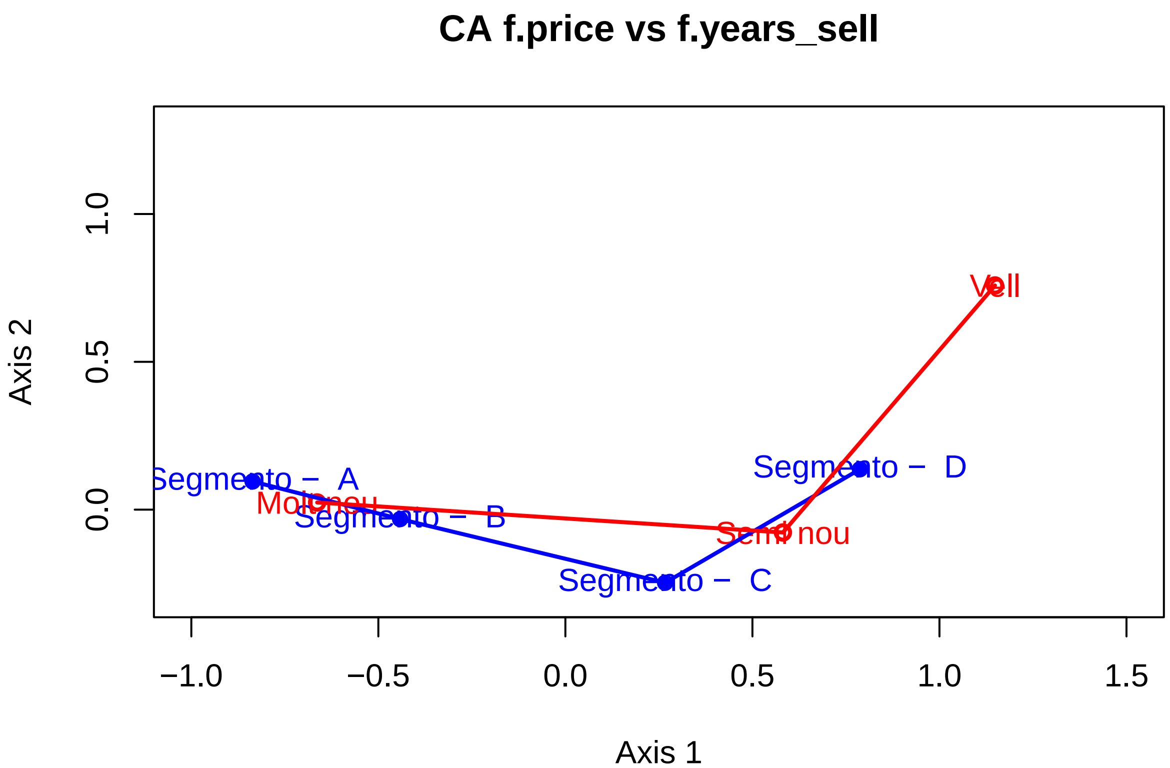
```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  tt
## X-squared = 2199.8, df = NA, p-value = 0.0004998
```

We get a p-value smaller than 0.05 so we can deny the H0 hypothesis. There is thus a link between the columns and the rows We are now going to take a look to the simple correspondences.

```
res.ca <- CA(tt)
```




```
plot(res.ca$row$coord[,1],res.ca$row$coord[,2],pch=19,col="blue",xlim=c(-1,1.5),ylim=c(-0.3,1.3),xlab="A
points(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red")
text(res.ca$row$coord[,1],res.ca$row$coord[,2],lwd=2,col="blue",labels=levels(df$f.price))
text(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red",labels=levels(df$years_sell))
lines(res.ca$row$coord[,1],res.ca$row$coord[,2],lwd=2,col="blue")
lines(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red")
```



We can see in the plot, that the category “MOLT NOU” belongs to the highest prices categories (Segment A and Segment B), while the older cars “SEMI NOU” belong to the cheaper categories (Segment C and Segment D). The oldest cars belong to the cheapest category (Segment D).

```
summary_price_years_sell<-summary(res.ca)$eigenvalues
```

```
##
## Call:
## CA(X = tt)
```

```
##
## The chi square of independence between the two variables is equal to 2199.768 (p-value =  0 ).
##
## Eigenvalues
##                        Dim.1   Dim.2
## Variance               0.421   0.023
## % of var.             94.893   5.107
## Cumulative % of var.  94.893 100.000
##
## Rows
##                   Iner*1000    Dim.1     ctr    cos2     Dim.2     ctr    cos2
## Segmento -  D |    193.611 |   0.787  44.677   0.971 |   0.137  25.009   0.029 |
## Segmento -  C |     31.340 |   0.267   4.006   0.538 |  -0.247  63.999   0.462 |
## Segmento -  B |     41.128 |  -0.442   9.726   0.995 |  -0.032   0.943   0.005 |
## Segmento -  A |    177.244 |  -0.836  41.592   0.987 |   0.095  10.049   0.013 |
##
## Columns
##                   Iner*1000    Dim.1     ctr    cos2     Dim.2     ctr    cos2
## Molt nou      |    212.737 |  -0.664  50.502   0.999 |   0.024   1.251   0.001 |
## Semi nou      |    166.457 |   0.582  38.882   0.983 |  -0.077  12.750   0.017 |
## Vell          |     64.129 |   1.148  10.616   0.696 |   0.758  85.998   0.304 |
```

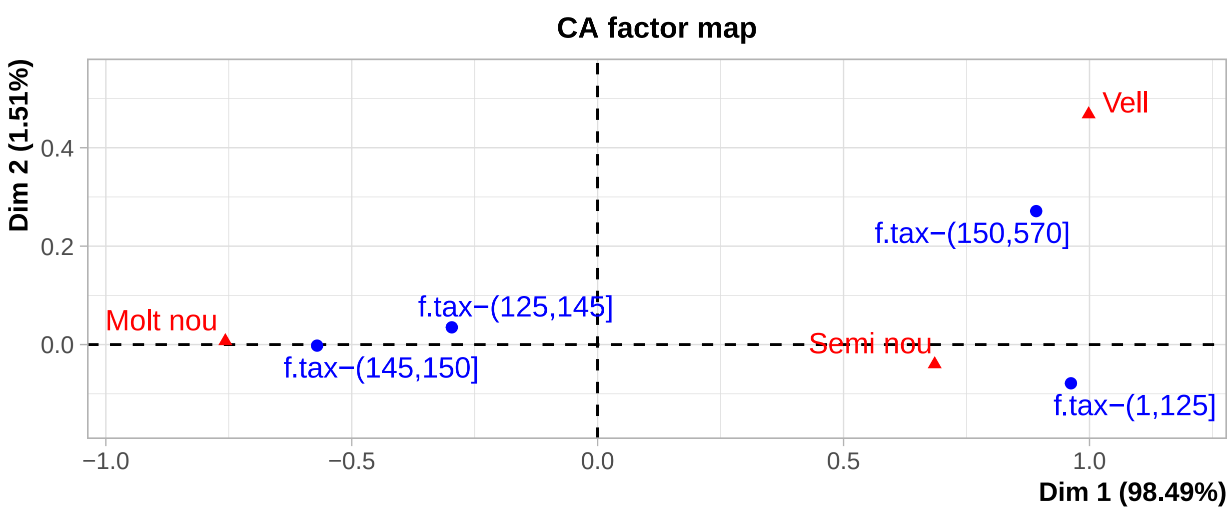
We can see from the summary is that we have a chi square statistic of 2200.099, great enough to reject the H0 hypothesis, which means the intensity of the relation is high. If we take a look at the variances from the different dimensions, we can see that all together sum more than 1.

We also think that it would be interesting to see the link between age and tax price. This will show us if the manufacturers are doing efforts to respect environment (which would be shown by a diminution of tax price)

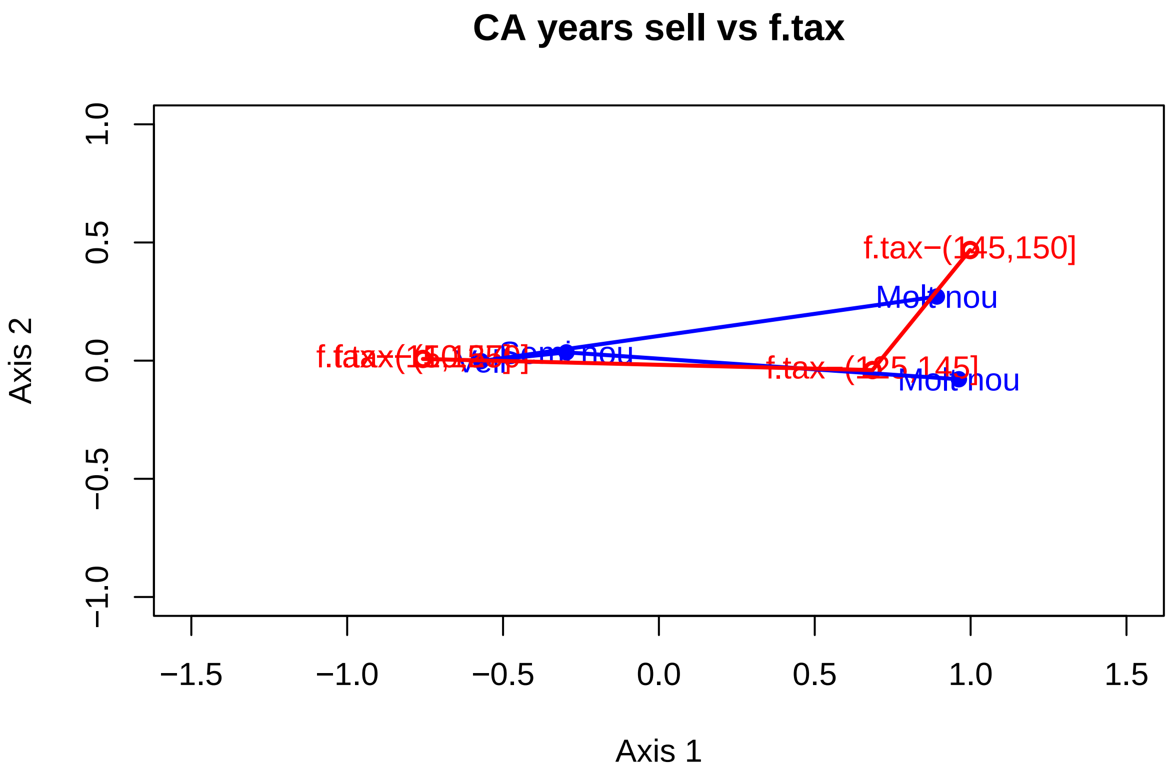
```
tt<-table(df[,c("f.tax","years_sell")]);tt
```

```
##                   years_sell
## f.tax               Molt nou Semi nou Vell
##    f.tax-(1,125]           0     1335   92
##    f.tax-(125,145]        24       13    1
##    f.tax-(145,150]      2349      693   22
##    f.tax-(150,570]        21      359   53
```

```
res.ca_1 <- CA(tt)
```




```
plot(res.ca_1$row$coord[,1],res.ca_1$row$coord[,2],pch=19,col="blue",xlim=c(-1.5,1.5),ylim=c(-1,1),xlab=
points(res.ca_1$col$coord[,1],res.ca_1$col$coord[,2],lwd=2,col="red")
text(res.ca_1$row$coord[,1],res.ca_1$row$coord[,2],lwd=2,col="blue",labels=levels(df$years_sell))
text(res.ca_1$col$coord[,1],res.ca_1$col$coord[,2],lwd=2,col="red",labels=levels(df$f.tax))
lines(res.ca_1$row$coord[,1],res.ca_1$row$coord[,2],lwd=2,col="blue")
lines(res.ca_1$col$coord[,1],res.ca_1$col$coord[,2],lwd=2,col="red")
```



We can consequently confirm our hypotesis, new cars are more respectful to the environment (tax price<150) than the old cars

6 MCA analysis

Now we will proceed with the multiple correspondence analysis to analyse all the categorical variables.

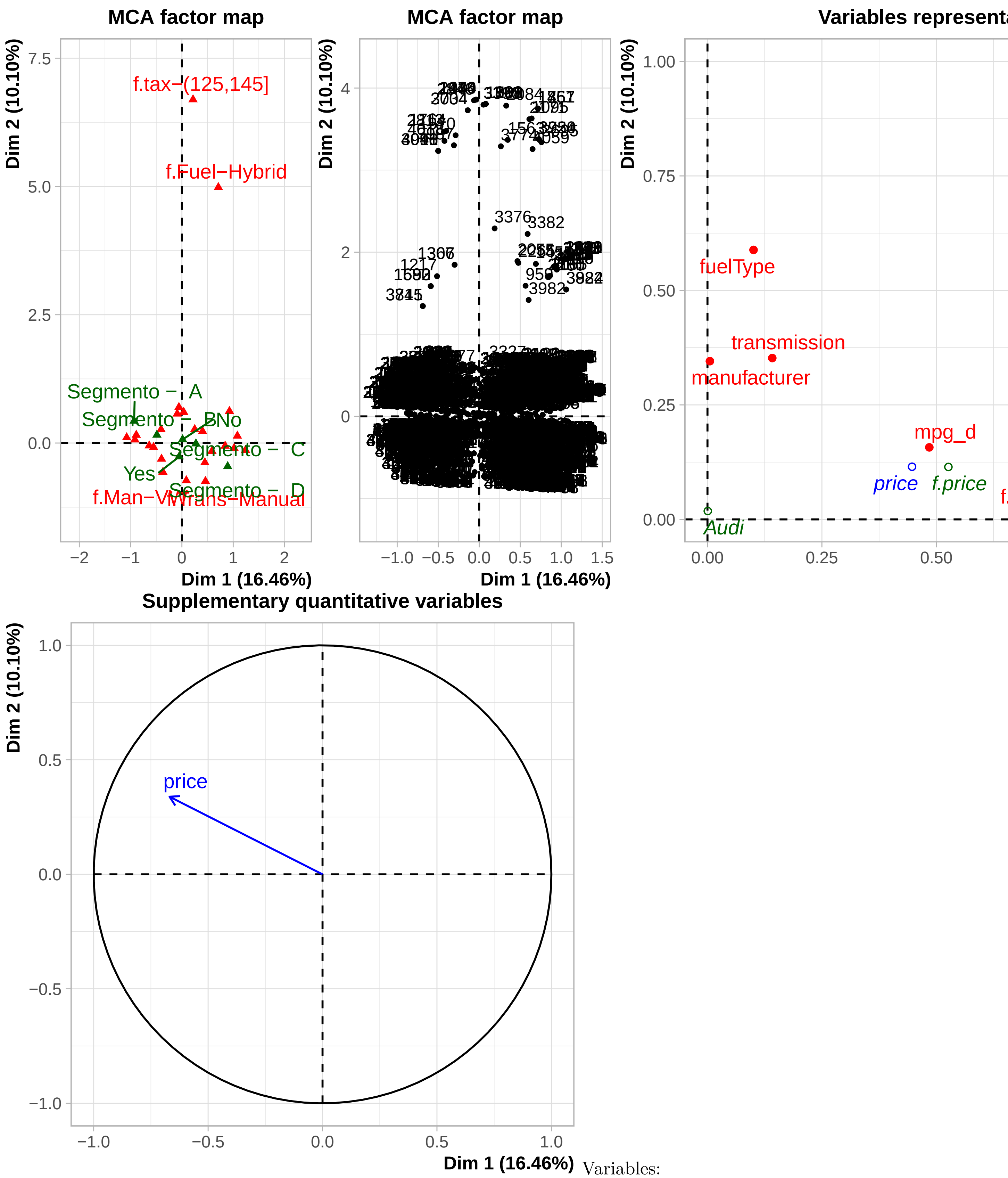
To the analysis of the MCA we will use the variables transmission, fuelTYpe, manufacturer, Audi, years_sell (nou, vell, molt vell), f.price, f.miles and f.tax. The quantitative supplementary variable will be the price one. Teh qualitative variables that will not be used for the computation of MCA will be binary target Audi and factor price.

```
names(df[,c(3,4,6,10,11,13,16,17,18,19)])
```

```
##  [1] "price"        "transmission" "fuelType"     "manufacturer" "Audi"
##  [6] "years_sell"   "f.price"      "f.miles"      "f.tax"        "mpg_d"
```

```
res.mca<-MCA(df[,c(3,4,6,10,11,13,16,17,18,19) ],quali.sup=c(5,7), quanti.sup=1 )
```

```
## Warning: ggrepel: 21 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



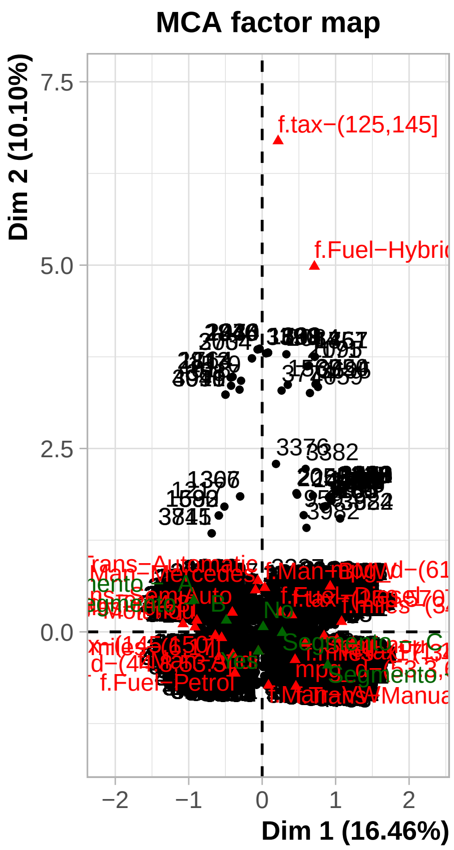
Variables:

The graphic created by the execution of the function MCA shows us that the Dimension 1 gets 16,5% of the variability and the dimension2 gets 10% of the variability. The supplementary quantitative variable price has more correlation to the dimenison 1 than to the dimension t2. As the Audi variable has been used as a supplementary to the analsis it cas no correlation with the dimensions. We will enter in more tdetail in the next sections.

Individuals and categories:

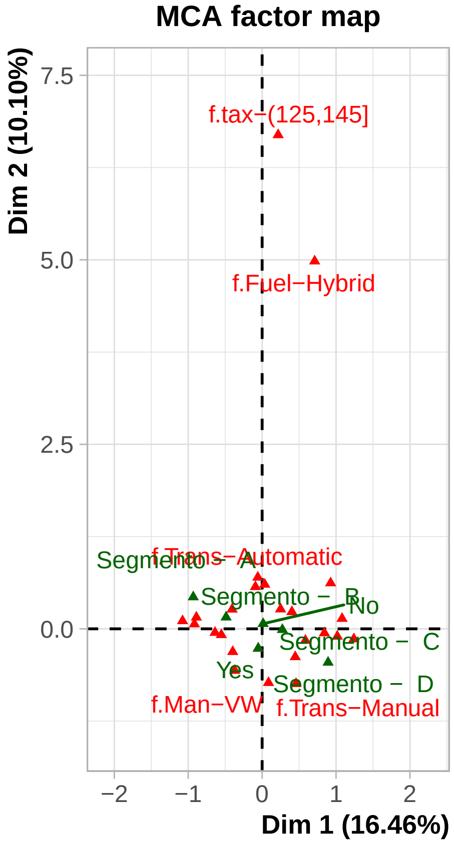
We will enter in more detail in the analysisi of individuals and categories in the next sections but we can see clearly that there are some varibales and individuals that has a strong correlation with the dimension 2.

```r
plot.MCA(res.mca,choix=c("ind"),cex=0.8)
```




```r
plot.MCA(res.mca,choix=c("ind"),invisible=c("ind"),cex=0.8)
```

```
## Warning: ggrepel: 20 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



6.1 Eigenvalues and dominant axes analysis. How many axes we have to consider for next Hierarchical Classification stage?

We will use the Kiser criteria to choose the number of axes to be considered. Wee will choose all the dimensions that have a greater eigenvalue than the mean. As the mean is 1428571, we will use the first 7 dimensions to analyze the data. As we can see in the graphic this 7 dimensions accumulate approximately the 60% of the variability.

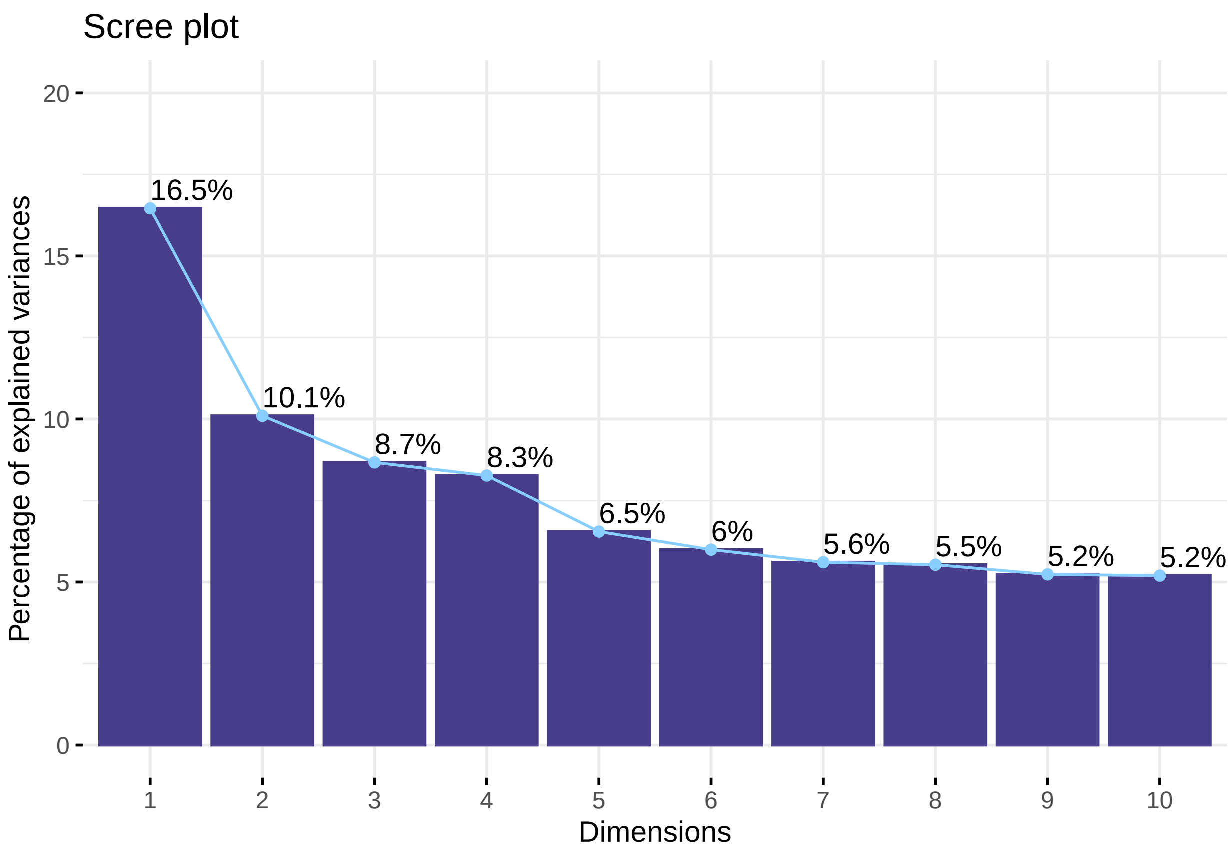
```
mean(res.mca$eig[,1])
```

```
## [1] 0.1428571
```

```
head(get_eigenvalue(res.mca), 10)
```

```
##          eigenvalue variance.percent cumulative.variance.percent
## Dim.1     0.4232739        16.460652                    16.46065
## Dim.2     0.2596686        10.098225                    26.55888
## Dim.3     0.2229031         8.668455                    35.22733
## Dim.4     0.2125692         8.266580                    43.49391
## Dim.5     0.1683757         6.547946                    50.04186
## Dim.6     0.1541116         5.993228                    56.03509
## Dim.7     0.1441866         5.607258                    61.64234
## Dim.8     0.1422142         5.530552                    67.17290
## Dim.9     0.1345968         5.234321                    72.40722
## Dim.10    0.1336127         5.196048                    77.60327
```

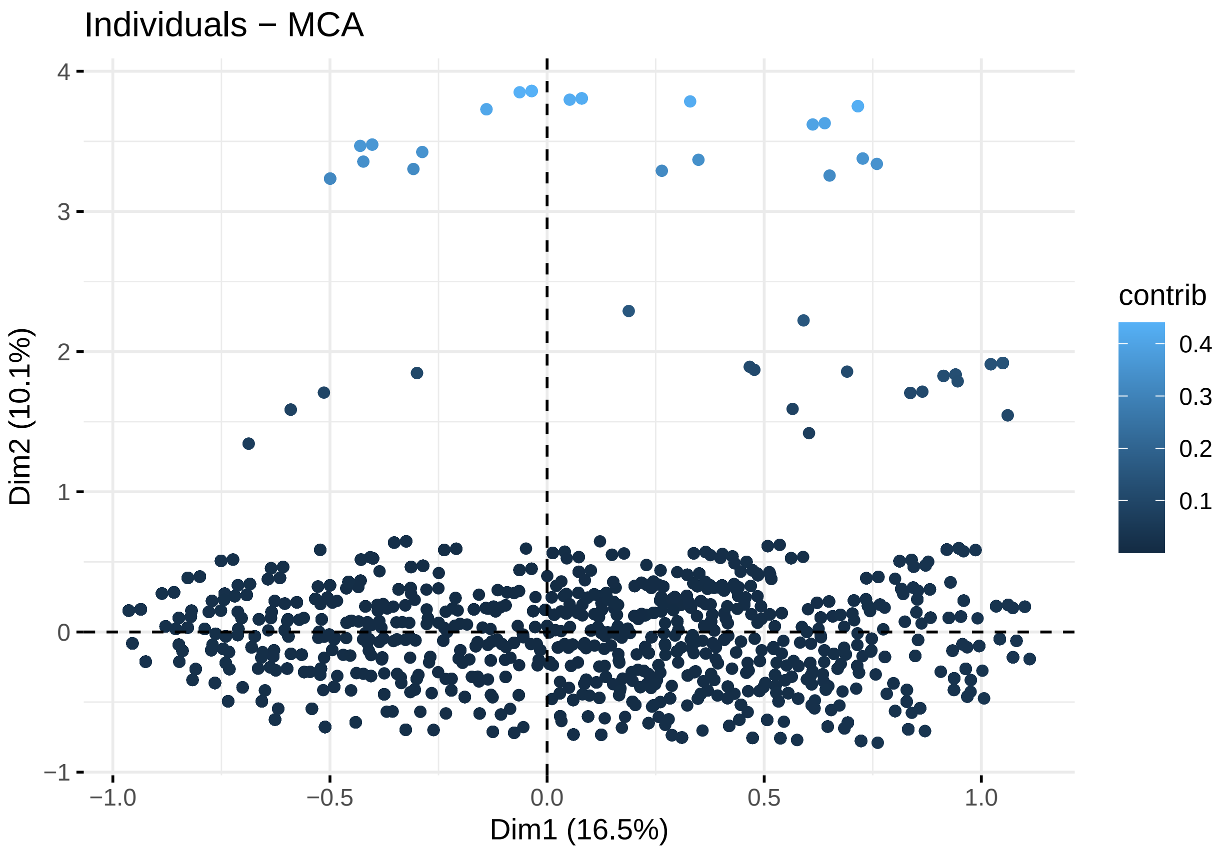
```
fviz_screeplot(
  res.mca,
  addlabels=TRUE,
  ylim=c(0,20),
  barfill="darkslateblue",
  barcolor="darkslateblue",
  linecolor="skyblue1"
)
```



6.2 Individuals point of view.

6.2.1 Are they any individuals “too contributive”?

```
fviz_mca_ind(res.mca, col.ind="contrib", geom = "point")
```



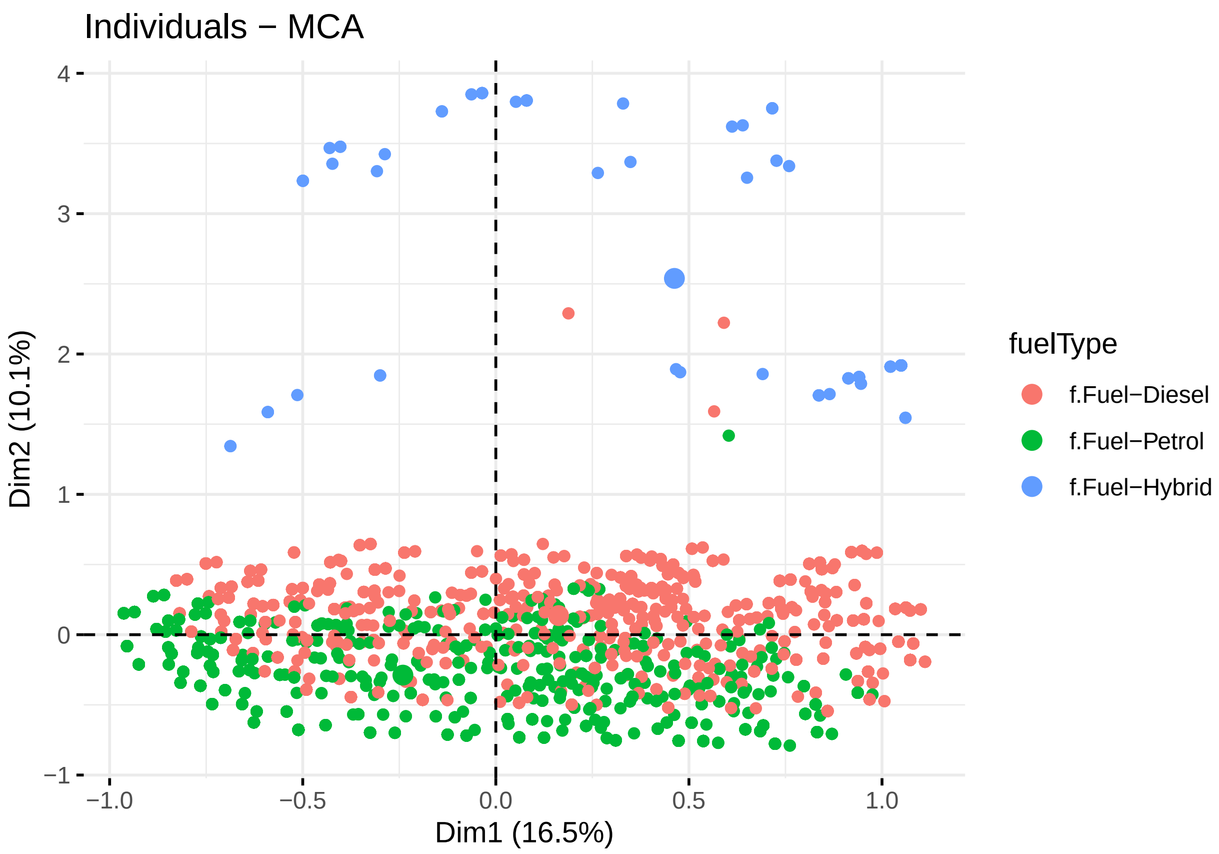
In the dimension 1 we can’t identify any observations that are too contributive. Otherwise, in the dimension 2 the are several individuals that have much weight in the creation of the second dimension.

6.2.2 Are there any groups?

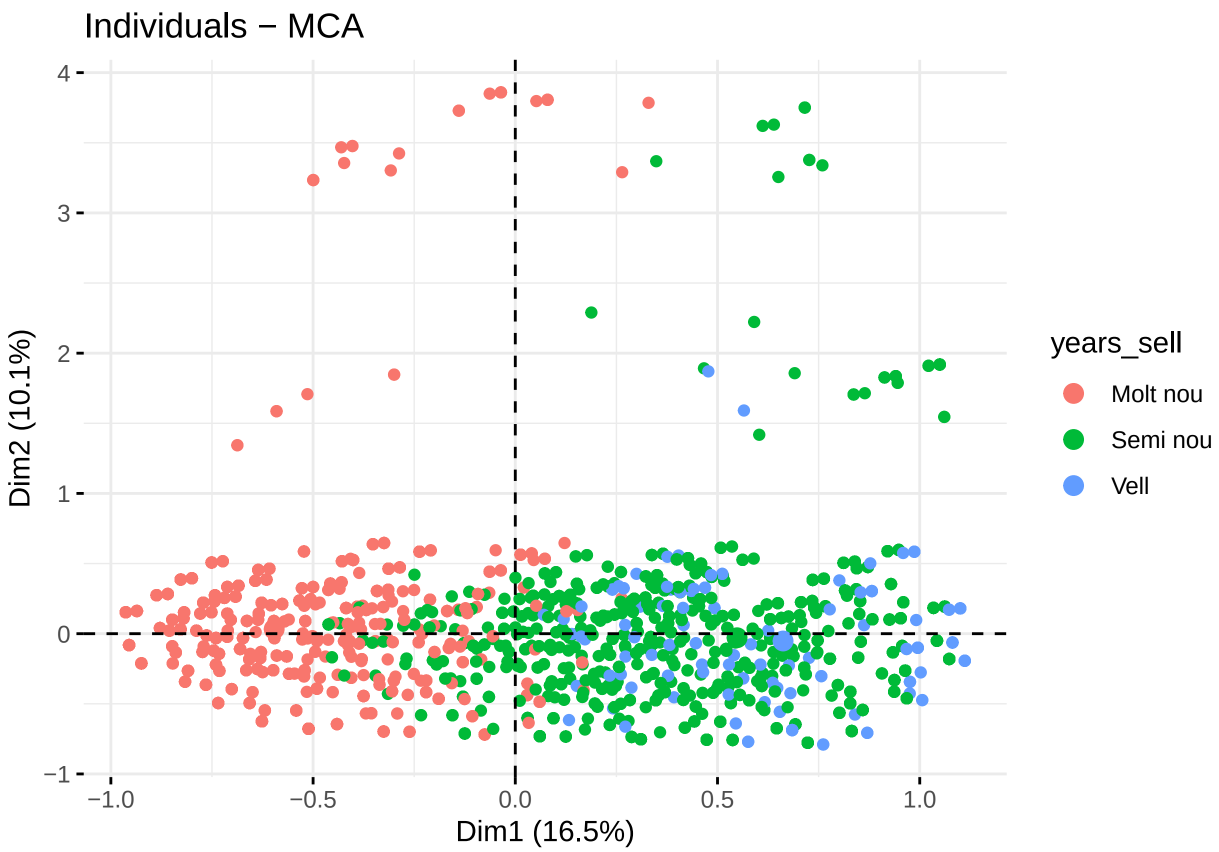
Depending on the qualitative variable used to classify the individuals we can see different types of groups. After proving all of them we have chosen fyelType, years_sell and f.price because are the ones that show more clearly differentiated groups. The first one, fuel Type is strongly related to the dim2. Values higher than 0 are represented by Petrol vehicles, values between 0 and -1 are represented by Diesel vehicles and finally the extreme observations, the ones that are more contributive to the creation of the Dim2 axis are the ones created by Hybrid vehicles.

As we can see the variable years_sell is strongly related to the dimension 1. The newest vehicle sobtain values lower than 0 and the oldest vehicle obtain values higher than 0.

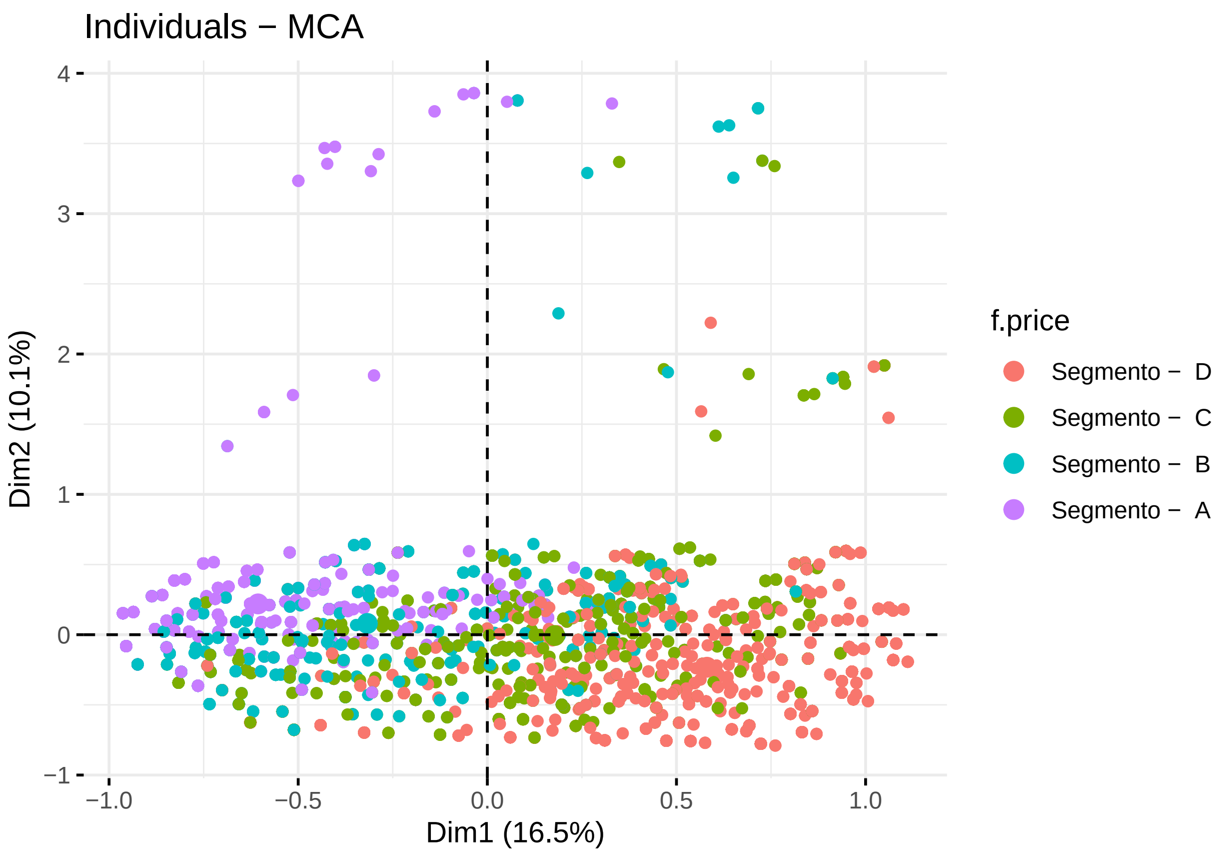
```
fviz_mca_ind(res.mca, label="none", habillage="fuelType")
```




```
fviz_mca_ind(res.mca, label="none", habillage="years_sell")
```



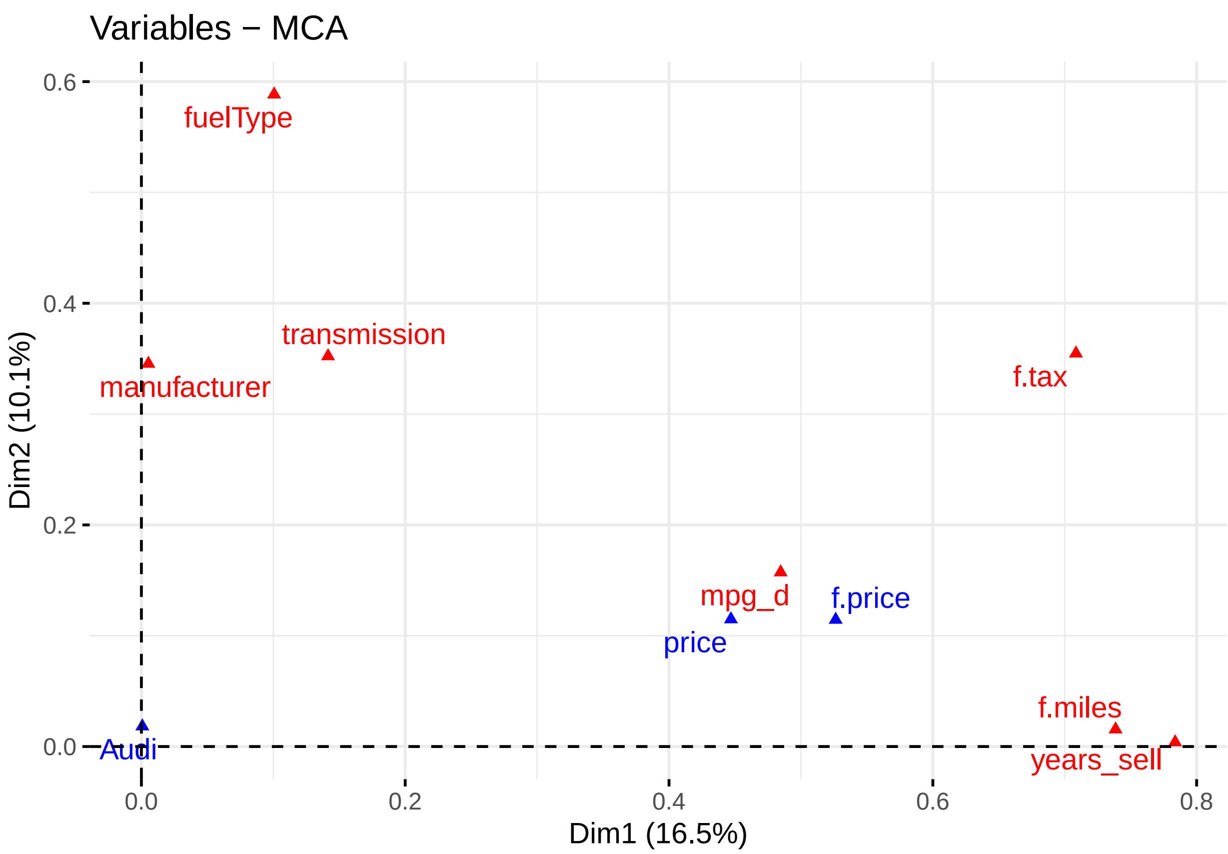

```
fviz_mca_ind(res.mca, label="none", habillage="f.price")
```



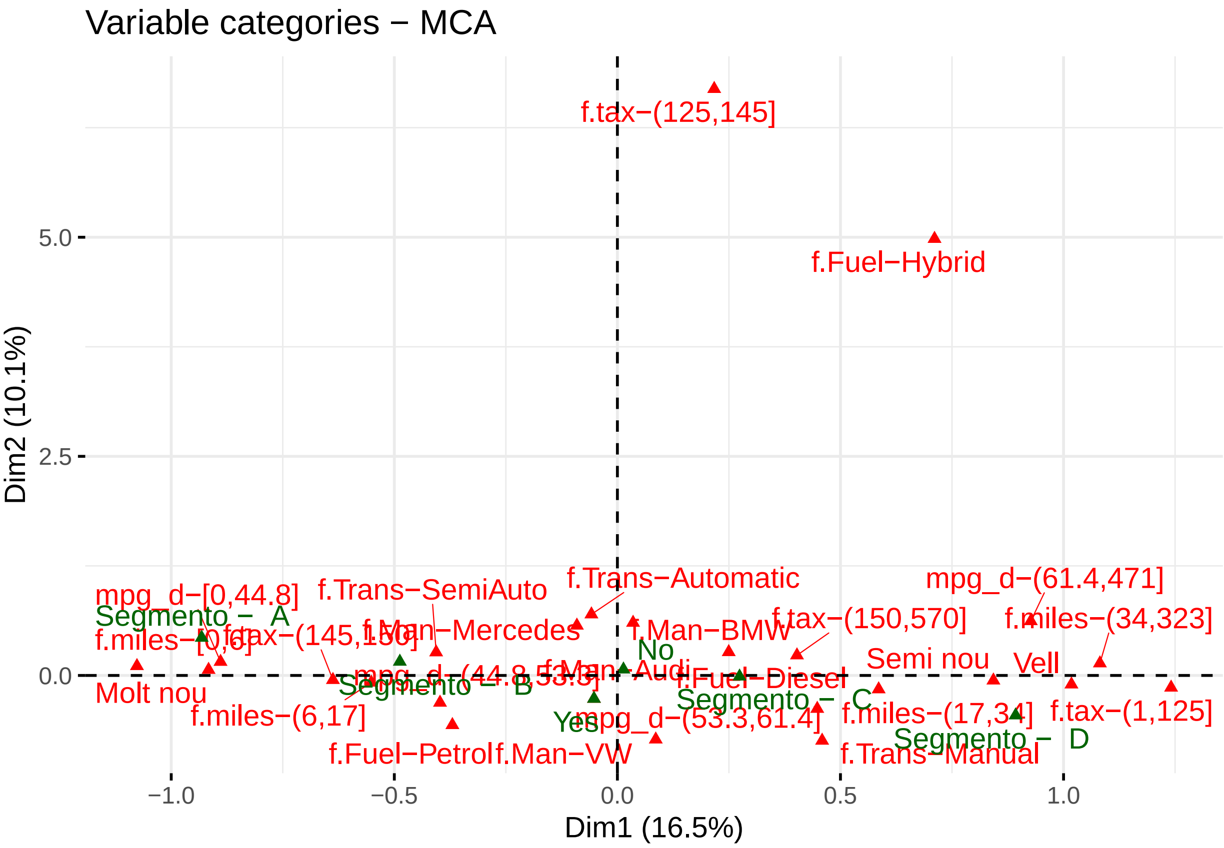
6.3 Interpreting map of categories: average profile versus extreme profiles (rare categories)

TO analyse the categories in out dataset we will use the following two plots.

```
fviz_mca_var(res.mca, choice="mca.cor", repel=TRUE)
```



```r
fviz_mca_var(res.mca, repel=TRUE)
```



The first plot shows us the correlation between variables and the two axes defined.

The first significant observation that we can see is that the variable fuelType has a huge impact on the creation of the second dimension and that it gets 60% of the variability. This matches the results of the previus section where we saw that the cars where distributed in the plot according to their consumption type. The variables f.tax, transmission and manufacturer have a significant impact too. This might be because the type of fuel of a car conditions the type of transmission, the manufacturer and the tax.

The dimension 1 is significantly created by the variables miles, years_sell and f.tax. That makes sense because this variables are related too as we saw in the previus chapters.

The second plot shows the categories for each of the variables that we have described. Cars with a tax between 125 and 145 and hybrid cars have a very big positive correlation with the dimension 2. The dimension 1 otherwise shows us other relation. For example newest cars are negative correlated to the dimension 1 but positive correlated with cars with very few miles.

6.4 Interpreting the axes association to factor map.

In this part we rank the variables and categories seen in the previus part due to ther correlaiton to the 2 dimensions of the factor map.

```r
res.desc_1 <- dimdesc(res.mca, axes = c(1,2))#Output in Annex
```

6.4.1 Dimension 1

6.4.1.1 Quantitative

- Price (-0,6): The only quantitative variable that we have included in our analysis is the price. As we can see it has a strong negative relation with the dimension 1. That means that it will have a positive strong correlation with alll variables that have hight negative values.

6.4.1.2 Qalitative We can see that there are 3 variables that have the biggest values. This three are highly positive correlated with the dimension1 but they are very correlated between them too. This means that, for example, how much older is a cad, it has muche more miles and has to pay more taxes.

- years_sell (0,78)
- f.miles (0,73)
- f.tax (0,70)

6.4.1.3 Category

The most correlated categories are the ones that are part of the yeats_sell, miles and tax variables. This is shown in the newxt lists where we tank the variables according to their correlation.

Positive correlated: as we can see old cars have a lot of milages and are cheap (segment-D)

- years_sell=Vell (0.46)
- f.tax=f.tax-(1,125] (0.60)
- f.miles=f.miles-(34,323] (0,70)

- f.price= Segmento-D (0,62)

Negative correlated: as we can see new cars have less miles and are more expensive than the ones of the previus list.

- years_sell=Molt nou (-0,80)
- f.tax=f.tax-(145,150] (-0.61)

- f.miles=f.miles-[0,6] (-0,70)
- f.price=Segmento - A (-0,57)

```
res.desc_1[[1]]#Output can be found in the Annex
```

6.4.2 Dimension 2

6.4.2.1 Quantitative

- Price (0,33): The only quantitative variable that we have included in our analysis is the price. As we can see the correlation with the dimension 2 is less important than the correlation with the dimension 1 but in this case is positive.

6.4.2.2 Qalitative

As we have seen in the previus analysis the variable that has more weight in the second dimension is the variable fuel Type with a value of 0,58. Transmision manufacturer and tax are related too buyt in a less significant way.

- fuelType (0,58)

6.4.2.3 Category

The most correlated categories are the ones that build the fuelType variable.

Positive correlated: Hybrid cars are positive correlated with de dimension and with the category f.tax-(125,145].

- f.tax=f.tax-(125,145] (2.55)
- fuelType=f.Fuel-Hybrid (1.74)

Negative correlated: diesel and petrol cars are positive related between them but negative related to transmision, manufacturer and tax.

- fuelType=f.Fuel-Diesel (-0.66)
- fuelType=f.Fuel-Petrol (-1.08)

```
res.desc_1[[2]]#Output in Annex
```

6.5 Perform a MCA taking into account also supplementary variables (use all numeric variables) quantitative and/or categorical. How supplementary variables enhance the axis interpretation?

Now we have added to the suplementaru quantitativ e list the 4 quantitative variables (price, mileage, mpg, tax) and we have added to the computation of the MCA the variables AUdi and engineSize.

```
res.mca<-MCA(df[,c(3,4,5,6,7,8,9,10,11,13,16,17,18,19) ], quanti.sup=c(1,3,5,6), graph = FALSE )
```

6.6 Interpreting the axes association to factor map.

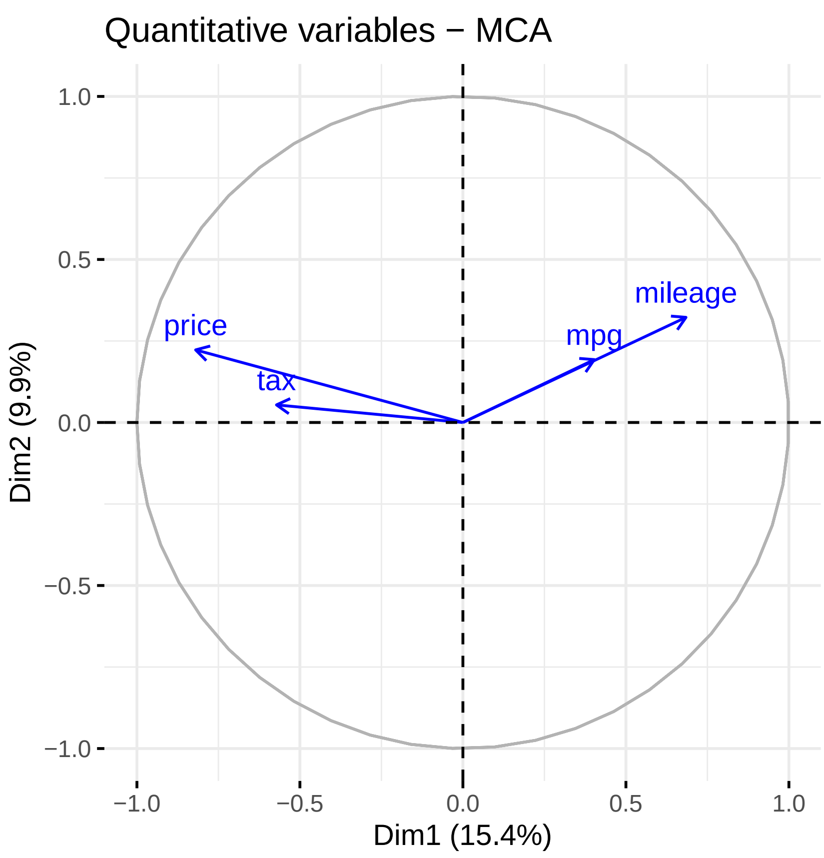
In this part we rank the variables and categories seen in the previus part due to ther correlation to the 2 dimensions of the factor map.

We can see that supplementary quantitativa variables are much more related to the first dimension that to the second dimension. Milage and mpg are veri positively related and negative related to price and tax.

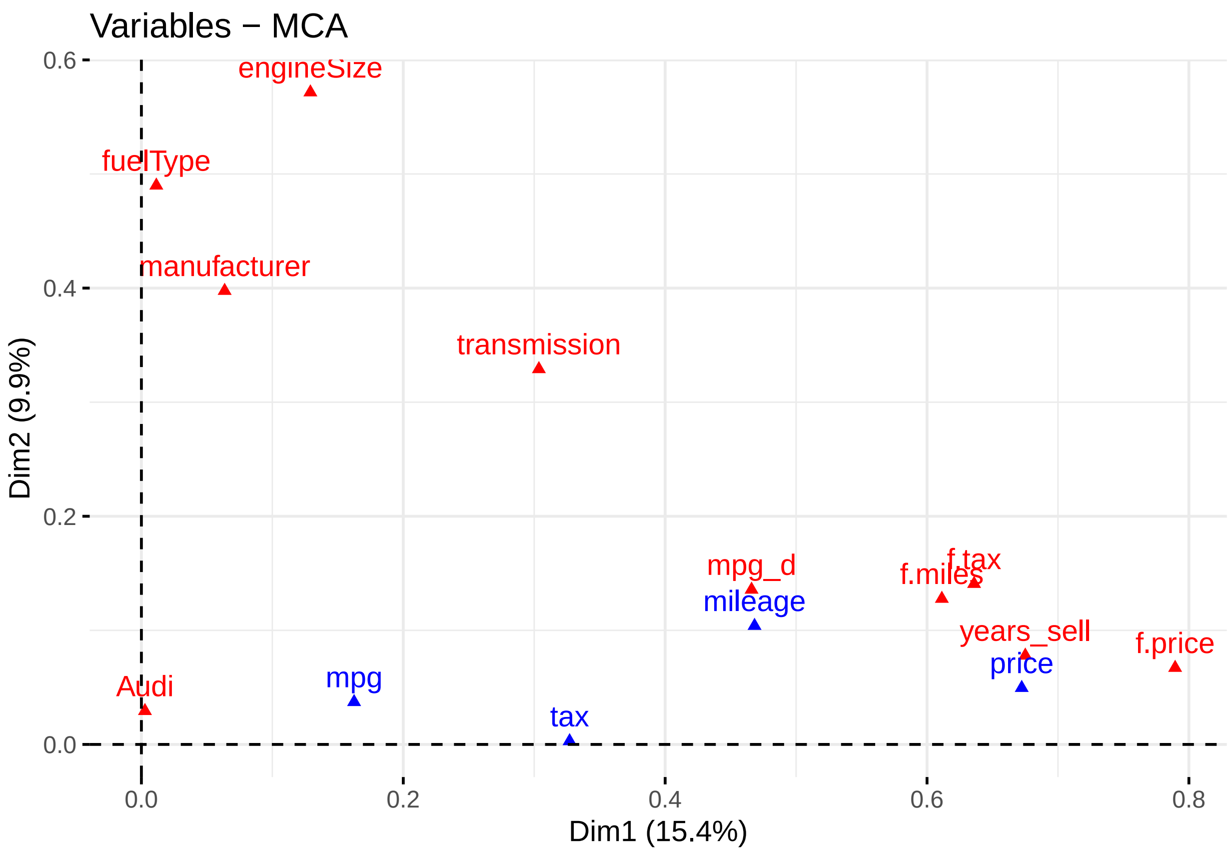
The dimension 2 is more correlated to qualitative variables. As we can see engineSize is the variable more related with the dimension 2 but fuel type remains in the top2.

```
res.desc <- dimdesc(res.mca, axes = c(1,2))

fviz_mca_var(res.mca, choice="quanti.sup")
```



```r
fviz_mca_var(res.mca, choice="mca.cor")
```



6.6.1 Dimension 1

Now we will proceed to analyse variables and categories for dimension 1 with the result of the MCA with all the variables. As we will see adding variables have not changed significantly the creation of this dimension. The amount of variance collected by this dimension is of about 15%.

6.6.1.1 Quantitative Quantitative variables have high correlation to the dimension 1. Mileage and miles per gallon has a strong positive correlation. Tax and price have a negative correlation with the dimension 1.

- mileage (0.68)
- mpg (0.40)
- tax (-0.57)
- price (-0.81)

6.6.1.2 Qalitative We can see that there are 3 variables that have the biggest values. This three are highly positive correlated with the dimension1 but they are very correlated between them too. This means that, for example, how much older is a cad, it has muche more miles and has to pay more taxes. This hasn't changed in relation with the first MCA analysis.

- years_sell (0.67)
- f.miles (0.61)
- f.tax (0.64)

6.6.1.3 Category The most correlated categories are the ones that are part of the price, years, miles and tax variables. This is shown in the next lists where we tank the variables according to their correlation.

Positive correlated

- f.tax=f.tax-(1,125] (0.66)
- f.miles=f.miles-(34,323] (0.59)
- f.price=Segmento-D (0.72)

Negative correlated

- mpg_d=mpg_d-[0,44.8] (-0.61)
- f.miles=f.miles-[0,6] (-0.62)
- f.price=Segmento-A (-0.66)
- years_sell=Molt nou (-0.72)

```
res.desc<-res.desc[[1]]# Output Can be found in the annex
```

6.6.2 Dimension 2

Now we will proceed to analyse variables and categories for dimension 2 with the result of the MCA with all the variables. As we will see this dimension has absorved themajority of the variance generated by the engineSize variable. The amount of variance collected by this dimension is of about 10%.

6.6.2.1 Quantitative The quantitative variables have much more correlation to the dimension 1 than to the dimension 2.

- Price (0,33): The only quantitative variable that we have included in our analysis is the price. As we can see the correlation with the dimension 2 is less important than the correlation with the dimension 1 but in this case is positive.

6.6.2.2 Qualitative The variable guelType remains as the second with more correlation to the second dimension but the engineSize one now is the variable with more correlation. This last one has added some correlation with the manufacturer variable.

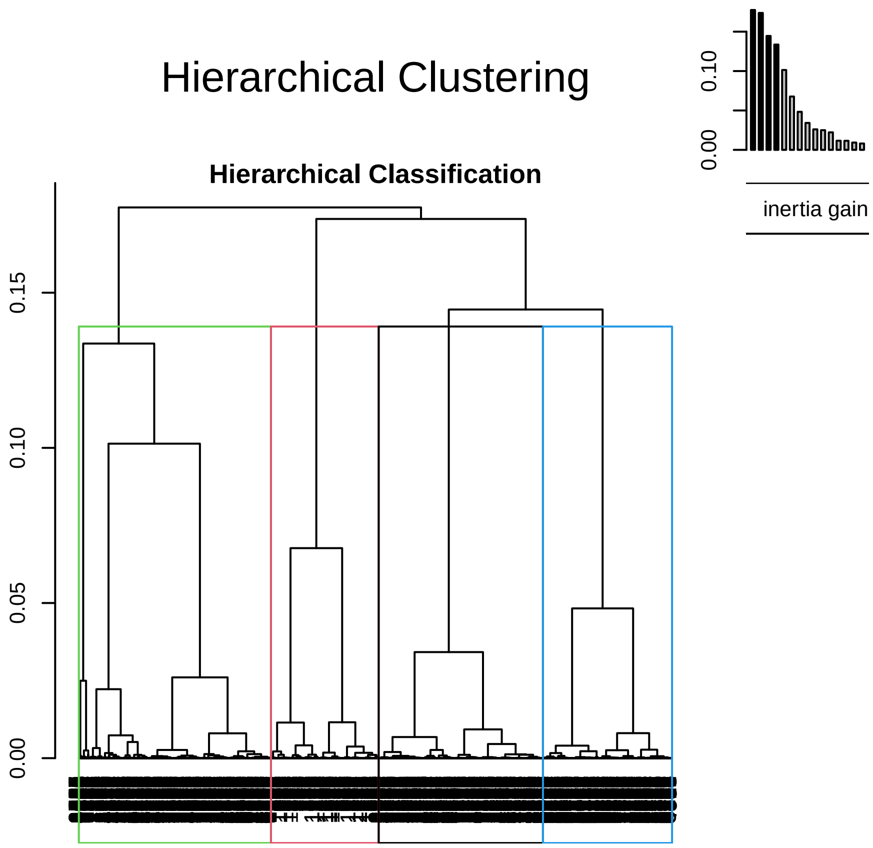
- engineSize (0.57)
- fuelType (0,48)
- manufacturer (0.40)

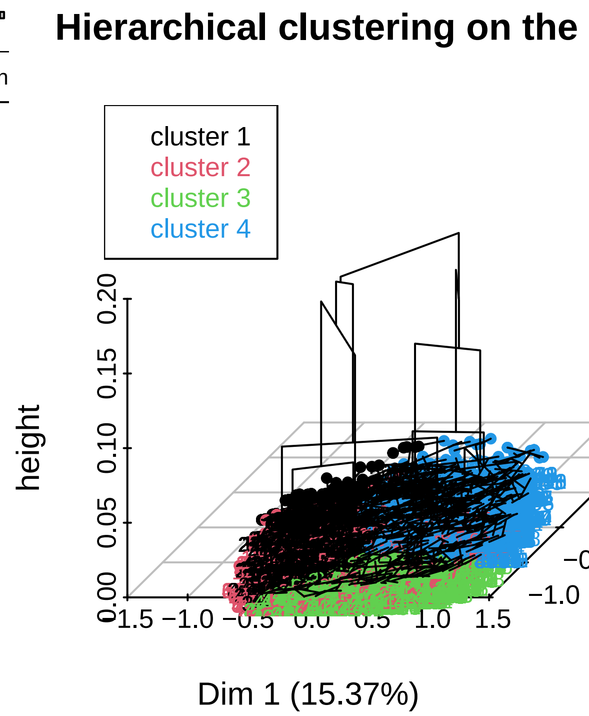
```
res.desc[[2]]
```

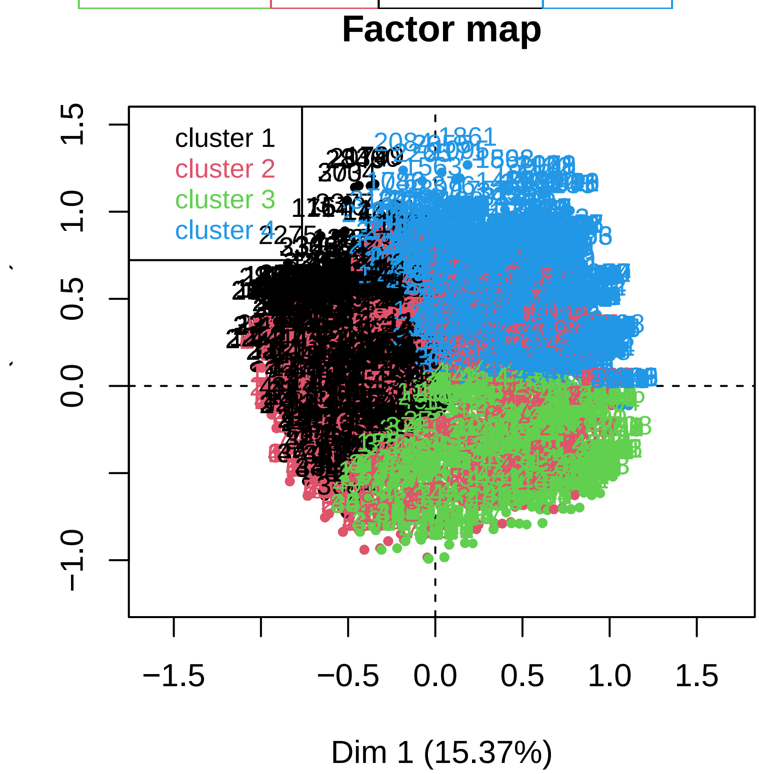
7 Hierarchical Clustering (from MCA)

In the first section of MCA analysis we said that we would use Kaiser criteria to choose the clusters and this mean that we have to choose the 9 clusters that have greater value than the mean. Otherwise, to reduce the complexity of the problem we have executed the function sevveral times and we have found that 4 clusters is a number that groups observations in significant diferent groups.

```
res.hcpcMCA <- HCPC(res.mca,nb.clust = 4, order = TRUE)
```







7.1 Description of clusters

We have four different clusters that are represented in the previus image.

- Cluster1: represented in color black is more correlated to the dim1 that to the dim2. It is correlated i a negative way. Contains 1653 observations.
- Cluster2: represented in color pink is correlated with both dimensions in a approximately equal way and contains 1000 observations.
- Cluster3: represented in color green is strong positive correlated to dim1 and negative correlated to dim2. COntains 943 observations.
- Cluster4: represented in color blue is positive correlated to dim1 and positive correlated to dim2.

Although the number of observations of the cluster 1 is higher than the other clusters, the number of observations is distributed equally between them.

```
table(res.hcpcMCA$data.clust$clust) #Output int the res.hcpcMCA$data.clust$clust annex section
```

7.1.1 Correlation with categories

When we say that a cluster is correlated with a dimension what we are saying is that this cluster is correlated with the variables correlated with this dimension too. Now we will analyze the most significant correlations with the different categories.

Note: to help interpret the result of the output Cla/Mod: % of the individuals who belong to the category and also belong to class Mod/Cla: % of individuals of class that belong to the category Global: % of the observations that are part of the category

- **Cluster 1:**
  - Variable target Audi: The first clear observation that we can make is related to our binary target Audi. All the individuals of Cluter 1 are in the category **Audi=No**. This means that this cluster does not contain any Audi car. The representation of the non audi cars is noticable (41%).
  - Variable target price: The 73% of the most expensive cars **(f.price=Segmento - A)** belong to this group. Of all the observations of the cluster a 55% are very expensive.
  - Variable tax: 96% of the individuals of the cluster 1 are of the category **f.tax=f.tax-(145,150**]. What is more 50% of the individuals that are of this category belong to this cluster.
  - Variable old: 92% of the observations in this cluer are **very young (less than two years old)**. This cluster contains 62% of the newest cars.
  - Manufacturer: 56% of the **Mercedes cars** belong to this cluster and they represent a 44% of all the cluster observations.
  - Transmission: 62% of the cars in this group are **SemiAuto** and 54% of the SemiAuto cars
  - EngineSize: 64% of the observations belong to the category **engineSize=Mitjà**.
- **Cluster 2:**
  - Variable target Audi: This cluster contains all the **Audi=Yes**. What is more all the Audi cars belong to this category. This is useful data because this varieable is one of our target variables.
  - Variable target price: From the point of view of the price of the cars in this cluster we don't get such relevant information. We can see that 25% belong to the cheapest category **(f.price=Segmento - D)** and a 30% belong to the most expensive **(f.price=Segmento - A)**
  - Variable fuel: more or les 50% of the cars in this group are of the type **fuelType=f.Fuel-Petrol** and the other 50% **fuelType=f.Fuel-Diesel**
  - Variable old: 45% of the observations in this cluer are **years_sell=Molt nou** . This cluster contains 20% of the newest cars.
  - EngineSize: 50% of the observations belong to the category **engineSize=Mitjà**.
- **Cluster 3:**
  - Variable target Audi: The first clear observation that we can make is related to our binary target Audi. All the individuals of Cluter 1 are in the category **Audi=No**. This means that this cluster does not contain any Audi car.
  - Variable target price: The 43% of the cheapest cars **(f.price=Segmento - D)** belong to this group. Of all the observations of the cluster a 68% are very expensive.
  - Variable fuel: 85% of the cars in this group are of the type **fuelType=f.Fuel-Petrol**.
  - Manufacturer: 54% of the **VW cars** belong to this cluster and they represent a 90% of all the cluster observations.
  - Transmission: 86% of the cars in this group are **transmission=f.Trans-Manual**.
  - EngineSize: 95% of the observations belong to the category **engineSize=Mitjà**.
- **Cluster 4:**
  - Variable target Audi: The first clear observation that we can make is related to our binary target Audi. All the individuals of Cluter 1 are in the category **Audi=No**. This means that this cluster does not contain any Audi car.
  - Variable target price: 40% of the observations are from the category **f.price=Segmento - C** and another 40% are from the category 40% of the observations are from the category **f.price=Segmento - D**.

- Manufacturer: 50% of the **BMW cars** belong to this cluster and they represent a 40% of all the cluster observations. 40% of the **Mercedes cars** belong to this cluster and they represent a 40% of all the cluster observations.
- Variable fuel: 85% of the cars in this group are of the type **fuelType=f.Fuel-Diesel**.
- Variable old: 91% of the observations in this cluer are not too old**years_sell=Semi nou** (between 3 and 5 years old).
- EngineSize: 71% of the observations belong to the category **engineSize=Mitjà**.

```
res.hcpcMCA$desc.var$category #Output int the res.hcpcMCA$desc.var$category annex section
```

8 Annex

8.1 Hierarchial clustering

8.1.1 Description of cluster by qualitative variables

```
quali_var_decription_1
```

```
## $`1`
##                                  Cla/Mod     Mod/Cla     Global       p.value
## f.miles=f.miles-[0,6]          99.039231  51.77898702 25.1713019  0.000000e+00
## aux=[0,5.89e+03]               99.074853  49.30933445 23.9621121  0.000000e+00
## years_sell=Molt nou            99.791145 100.00000000 48.2466747  0.000000e+00
## year=year_2019                 99.873976  66.34575136 31.9830713  0.000000e+00
## f.price=Segmento -  A          89.935588  46.75596484 25.0302297 1.965841e-280
## year=year_2018                 99.564270  19.12934282  9.2503023 3.998185e-152
## aux=(5.89e+03,1.69e+04]        76.484194  41.52365006 26.1386538 3.610289e-129
## year=year_2020                 99.712644  14.52490582  7.0133011 1.087181e-114
## f.miles=f.miles-(6,17]         75.000000  39.30514860 25.2317614 2.084043e-110
## f.price=Segmento -  B          69.711538  30.34742570 20.9592906  3.754461e-56
## transmission=f.Trans-SemiAuto  60.940803  48.26287149 38.1297864  9.553153e-46
## model=VW- T-Roc               100.000000   2.80452072  1.3502620  3.312432e-22
## fuelType=f.Fuel-Petrol         55.652596  48.01172039 41.5356711  4.460211e-19
## model=VW- T-Cross             100.000000   1.46504814  0.7053607  6.809066e-12
## model=Audi- Q2                 80.487805   2.76266220  1.6525595  1.485380e-09
## model=BMW- X2                 100.000000   1.08832147  0.5239823  5.196913e-09
## model=Mercedes- B Class        77.192982   1.84177480  1.1487304  7.855256e-06
## model=Mercedes- V Class        88.888889   1.00460444  0.5441354  1.159227e-05
## model=VW- Arteon               90.000000   0.75345333  0.4030633  1.156803e-04
## model=VW- Tiguan Allspace     100.000000   0.41858518  0.2015316  6.627407e-04
## model=BMW- M4                  91.666667   0.46044370  0.2418380  2.291420e-03
## manufacturer=f.Man-Mercedes    51.655119  28.08706572 26.1789601  3.240347e-03
## model=VW- Sharan               78.260870   0.75345333  0.4635228  3.805897e-03
## model=VW- Amarok               90.909091   0.41858518  0.2216848  4.424650e-03
## model=BMW- 2 Series            59.712230   3.47425701  2.8012898  5.749111e-03
## model=VW- Shuttle             100.000000   0.25115111  0.1209190  1.241483e-02
## model=Mercedes- GLC Class      58.620690   2.84637924  2.3377670  2.280874e-02
## model=Audi- Q7                 66.666667   1.00460444  0.7255139  2.689068e-02
## model=Audi- Q5                 59.183673   2.42779406  1.9750101  2.783399e-02
## model=Mercedes- X-CLASS        81.818182   0.37672666  0.2216848  2.867977e-02
## model=Mercedes- GLE Class      64.444444   1.21389703  0.9068924  2.911599e-02
## model=Mercedes- GLS Class      87.500000   0.29300963  0.1612253  3.053750e-02
## model=VW- Touareg              66.666667   0.83717036  0.6045949  4.412939e-02
## model=Mercedes- C Class        52.987013   8.53913771  7.7589682  4.814322e-02
## Audi=No                        48.858447  80.61950607 79.4437727  4.831062e-02
## manufacturer=f.Man-Audi        45.392157  19.38049393 20.5562273  4.831062e-02
## Audi=Yes                       45.392157  19.38049393 20.5562273  4.831062e-02
## model=Mercedes- CL Class       26.470588   0.37672666  0.6852076  1.089010e-02
## model=Mercedes- SLK             0.000000   0.00000000  0.1410721  1.004090e-02
## model=Mercedes- GL Class        9.090909   0.04185852  0.2216848  8.838624e-03
```

```
## model=Audi- A4                  36.134454  1.79991628  2.3982265  7.803700e-03
## model=Mercedes- M Class          0.000000  0.00000000  0.1612253  5.199785e-03
## model=BMW- i3                    0.000000  0.00000000  0.1612253  5.199785e-03
## model=Audi- A3                  38.144330  3.09753035  3.9097138  4.363824e-03
## fuelType=f.Fuel-Hybrid          32.530120  1.13017999  1.6727126  3.925250e-03
## model=VW- CC                     0.000000  0.00000000  0.1813785  2.692259e-03
## model=BMW- X1                   32.608696  1.25575555  1.8540911  2.479182e-03
## model=VW- Beetle                 0.000000  0.00000000  0.2015316  1.393691e-03
## model=Audi- TT                  14.814815  0.16743407  0.5441354  3.542425e-04
## model=Audi- A1                  32.450331  2.05106739  3.0431278  7.779459e-05
## model=BMW- 1 Series             32.275132  2.55336961  3.8089480  7.056242e-06
## model=VW- Polo                  34.972678  5.35789033  7.3760580  1.316745e-07
## model=VW- Scirocco               0.000000  0.00000000  0.5642886  9.616993e-09
## fuelType=f.Fuel-Diesel          43.115685 50.85809962 56.7916163  4.175114e-16
## f.price=Segmento -  C           33.163265 16.32482210 23.7001209  1.723693e-32
## years_sell=Vell                  0.000000  0.00000000  3.3857316  8.025917e-50
## year=year_2013                   0.000000  0.00000000  3.3857316  8.025917e-50
## transmission=f.Trans-Manual     33.351528 25.57555463 36.9205965  4.370353e-58
## year=year_2014                   0.000000  0.00000000  3.9500202  3.033614e-58
## year=year_2015                   0.000000  0.00000000  7.7388150 1.014189e-116
## aux=(1.69e+04,3.4e+04]          15.710919  8.37170364 25.6549778 1.736648e-171
## f.miles=f.miles-(17,34]         15.451664  8.16241105 25.4332930 1.742334e-172
## year=year_2016                   0.000000  0.00000000 17.4324869 2.778950e-284
## year=year_2017                   0.000000  0.00000000 17.7952439 3.814868e-291
## f.price=Segmento -  D           10.438830  6.57178736 30.3103587 3.155589e-299
## f.miles=f.miles-(34,323]         1.501251  0.75345333 24.1636437  0.000000e+00
## aux=(3.4e+04,3.23e+05]           1.579385  0.79531185 24.2442563  0.000000e+00
## years_sell=Semi nou              0.000000  0.00000000 48.3675937  0.000000e+00
##                                    v.test
## f.miles=f.miles-[0,6]                 Inf
## aux=[0,5.89e+03]                      Inf
## years_sell=Molt nou                   Inf
## year=year_2019                        Inf
## f.price=Segmento -  A           35.783902
## year=year_2018                  26.272090
## aux=(5.89e+03,1.69e+04]         24.180019
## year=year_2020                  22.762175
## f.miles=f.miles-(6,17]          22.325614
## f.price=Segmento -  B           15.788147
## transmission=f.Trans-SemiAuto   14.197064
## model=VW- T-Roc                  9.690396
## fuelType=f.Fuel-Petrol           8.924920
## model=VW- T-Cross                6.861606
## model=Audi- Q2                   6.045944
## model=BMW- X2                    5.840741
## model=Mercedes- B Class          4.469091
## model=Mercedes- V Class          4.385118
## model=VW- Arteon                 3.855104
## model=VW- Tiguan Allspace        3.404546
## model=BMW- M4                    3.049606
## manufacturer=f.Man-Mercedes      2.943967
## model=VW- Sharan                 2.893817
## model=VW- Amarok                 2.846185
## model=BMW- 2 Series              2.761758
## model=VW- Shuttle                2.500128
## model=Mercedes- GLC Class        2.276623
## model=Audi- Q7                   2.213101
## model=Audi- Q5                   2.199618
## model=Mercedes- X-CLASS          2.187860
## model=Mercedes- GLE Class        2.181912
## model=Mercedes- GLS Class        2.163048
## model=VW- Touareg                2.012860
## model=Mercedes- C Class          1.976102
## Audi=No                          1.974626
```

```
## manufacturer=f.Man-Audi          -1.974626
## Audi=Yes                         -1.974626
## model=Mercedes- CL Class         -2.546206
## model=Mercedes- SLK              -2.574418
## model=Mercedes- GL Class         -2.618234
## model=Audi- A4                   -2.660447
## model=Mercedes- M Class          -2.794389
## model=BMW- i3                    -2.794389
## model=Audi- A3                   -2.850590
## fuelType=f.Fuel-Hybrid           -2.884107
## model=VW- CC                     -3.000851
## model=BMW- X1                    -3.025871
## model=VW- Beetle                 -3.195954
## model=Audi- TT                   -3.572017
## model=Audi- A1                   -3.951095
## model=BMW- 1 Series              -4.491985
## model=VW- Polo                   -5.276500
## model=VW- Scirocco               -5.737349
## fuelType=f.Fuel-Diesel           -8.133372
## f.price=Segmento -  C           -11.868580
## years_sell=Vell                 -14.840414
## year=year_2013                  -14.840414
## transmission=f.Trans-Manual     -16.066656
## year=year_2014                  -16.089277
## year=year_2015                  -22.966239
## aux=(1.69e+04,3.4e+04]          -27.915367
## f.miles=f.miles-(17,34]         -27.997508
## year=year_2016                  -36.030575
## year=year_2017                  -36.466160
## f.price=Segmento -  D           -36.972613
## f.miles=f.miles-(34,323]              -Inf
## aux=(3.4e+04,3.23e+05]                -Inf
## years_sell=Semi nou                   -Inf
##
## $`2`
##                                 Cla/Mod    Mod/Cla     Global       p.value
## years_sell=Semi nou            55.8750000 91.9753086 48.3675937  0.000000e+00
## year=year_2017                 76.2174405 46.1591221 17.7952439 4.876758e-229
## f.miles=f.miles-(34,323]       51.4595496 42.3182442 24.1636437  5.520092e-78
## aux=(1.69e+04,3.4e+04]         50.5106049 44.1015089 25.6549778  9.716054e-78
## aux=(3.4e+04,3.23e+05]         51.3715711 42.3868313 24.2442563  9.804240e-78
## f.miles=f.miles-(17,34]        50.5546751 43.7585734 25.4332930  3.977753e-77
## f.price=Segmento -  C          46.4285714 37.4485597 23.7001209  2.191110e-46
## years_sell=Vell                69.6428571  8.0246914  3.3857316  6.713849e-28
## year=year_2013                 69.6428571  8.0246914  3.3857316  6.713849e-28
## engineSize=Gran                49.7217069 18.3813443 10.8625554  5.922071e-26
## year=year_2016                 39.7687861 23.5939643 17.4324869  5.350400e-13
## model=BMW- X1                  60.8695652  3.8408779  1.8540911  3.129843e-10
## year=year_2014                 49.4897959  6.6529492  3.9500202  1.721730e-09
## engineSize=Mitjà               32.8895850 57.6131687 51.4711810  2.265518e-08
## year=year_2015                 41.9270833 11.0425240  7.7388150  5.093919e-08
## model=VW- Tiguan               48.0662983  5.9670782  3.6477227  7.476313e-08
## f.price=Segmento -  D          34.5079787 35.5967078 30.3103587  2.197443e-07
## manufacturer=f.Man-BMW         35.5515041 26.7489712 22.1080210  5.266916e-07
## fuelType=f.Fuel-Diesel         32.1149752 62.0713306 56.7916163  1.189542e-06
## model=Audi- Q3                 49.5934959  4.1838134  2.4788392  2.024746e-06
## model=Mercedes- M Class       100.0000000  0.5486968  0.1612253  5.481549e-05
## model=VW- Beetle               90.0000000  0.6172840  0.2015316  1.228914e-04
## model=Audi- TT                 62.9629630  1.1659808  0.5441354  3.563158e-04
## model=Mercedes- GL Class       81.8181818  0.6172840  0.2216848  5.174163e-04
## model=Mercedes- CLS Class      73.3333333  0.7544582  0.3022975  6.123165e-04
## model=BMW- 3 Series            38.7755102  6.5157750  4.9375252  1.242195e-03
## model=Audi- A5                 45.7831325  2.6063100  1.6727126  1.518473e-03
## model=Mercedes- GLA Class      45.5882353  2.1262003  1.3704152  4.636100e-03
```

```
## model=Mercedes- GLC Class       41.3793103  3.2921811  2.3377670  5.436470e-03
## transmission=f.Trans-Automatic  32.3909532 27.5034294 24.9496171  7.681939e-03
## model=BMW- 5 Series             40.8695652  3.2235940  2.3176139  7.969883e-03
## model=Audi- A7                  70.0000000  0.4801097  0.2015316  1.064598e-02
## model=Audi- Q5                  40.8163265  2.7434842  1.9750101  1.494323e-02
## model=Mercedes- SL CLASS        60.0000000  0.6172840  0.3022975  1.613273e-02
## manufacturer=f.Man-Audi         32.3529412 22.6337449 20.5562273  2.026990e-02
## Audi=Yes                        32.3529412 22.6337449 20.5562273  2.026990e-02
## model=Mercedes- E Class         36.8159204  5.0754458  4.0507860  2.057154e-02
## model=Mercedes- SLK             71.4285714  0.3429355  0.1410721  2.956882e-02
## model=VW- Arteon                10.0000000  0.1371742  0.4030633  4.862451e-02
## model=Audi- Q2                  19.5121951  1.0973937  1.6525595  4.324241e-02
## model=Audi- A3                  22.6804124  3.0178326  3.9097138  3.371786e-02
## model=VW- Tiguan Allspace        0.0000000  0.0000000  0.2015316  3.072076e-02
## model=Mercedes- V Class         11.1111111  0.2057613  0.5441354  2.932579e-02
## Audi=No                         28.6149163 77.3662551 79.4437727  2.026990e-02
## f.price=Segmento -  B           26.4423077 18.8614540 20.9592906  1.852663e-02
## model=Mercedes- B Class         12.2807018  0.4801097  1.1487304  2.511434e-03
## model=Mercedes- C Class         21.5584416  5.6927298  7.7589682  3.257214e-04
## fuelType=f.Fuel-Petrol          26.5405143 37.5171468 41.5356711  2.035723e-04
## model=BMW- X2                    0.0000000  0.0000000  0.5239823  1.147376e-04
## model=VW- Polo                  20.4918033  5.1440329  7.3760580  6.547762e-05
## model=VW- Passat                11.5384615  0.8230453  2.0959291  1.398217e-05
## model=VW- Up                     9.8901099  0.6172840  1.8339379  7.427693e-06
## model=Audi- A1                  13.9072848  1.4403292  3.0431278  6.655887e-06
## model=VW- T-Cross                0.0000000  0.0000000  0.7053607  4.897402e-06
## fuelType=f.Fuel-Hybrid           7.2289157  0.4115226  1.6727126  6.831630e-07
## model=VW- Golf                  18.0084746  5.8299040  9.5122934  2.924320e-09
## model=VW- T-Roc                  0.0000000  0.0000000  1.3502620  6.242961e-11
## manufacturer=f.Man-VW           21.6688228 22.9766804 31.1567916  3.354076e-16
## f.miles=f.miles-(6,17]          15.8945687 13.6488340 25.2317614  1.609540e-36
## engineSize=Petit                18.7265918 24.0054870 37.6662636  4.609589e-39
## aux=(5.89e+03,1.69e+04]         14.9575944 13.3058985 26.1386538  9.809586e-44
## year=year_2020                   0.0000000  0.0000000  7.0133011  1.225532e-55
## year=year_2018                   0.0000000  0.0000000  9.2503023  3.229334e-74
## f.price=Segmento -  A            9.5008052  8.0932785 25.0302297  1.130528e-81
## aux=[0,5.89e+03]                 0.2523129  0.2057613 23.9621121 1.211779e-204
## f.miles=f.miles-[0,6]            0.3202562  0.2743484 25.1713019 4.234225e-215
## year=year_2019                   0.0000000  0.0000000 31.9830713 3.079192e-303
## years_sell=Molt nou              0.0000000  0.0000000 48.2466747  0.000000e+00
##                                   v.test
## years_sell=Semi nou                  Inf
## year=year_2017                 32.311154
## f.miles=f.miles-(34,323]       18.694191
## aux=(1.69e+04,3.4e+04]         18.664009
## aux=(3.4e+04,3.23e+05]         18.663526
## f.miles=f.miles-(17,34]        18.588552
## f.price=Segmento -  C          14.299904
## years_sell=Vell                10.949063
## year=year_2013                 10.949063
## engineSize=Gran                10.535574
## year=year_2016                  7.216089
## model=BMW- X1                   6.292207
## year=year_2014                  6.022096
## engineSize=Mitjà                5.590396
## year=year_2015                  5.448001
## model=VW- Tiguan                5.379331
## f.price=Segmento -  D           5.181809
## manufacturer=f.Man-BMW          5.016325
## fuelType=f.Fuel-Diesel          4.857372
## model=Audi- Q3                  4.750939
## model=Mercedes- M Class         4.034085
## model=VW- Beetle                3.840285
## model=Audi- TT                  3.570489
```

```
## model=Mercedes- GL Class          3.471574
## model=Mercedes- CLS Class         3.426099
## model=BMW- 3 Series               3.229010
## model=Audi- A5                    3.171130
## model=Mercedes- GLA Class         2.831288
## model=Mercedes- GLC Class         2.779965
## transmission=f.Trans-Automatic    2.665739
## model=BMW- 5 Series               2.653343
## model=Audi- A7                    2.554109
## model=Audi- Q5                    2.433752
## model=Mercedes- SL CLASS          2.405899
## manufacturer=f.Man-Audi           2.321314
## Audi=Yes                          2.321314
## model=Mercedes- E Class           2.315757
## model=Mercedes- SLK               2.175819
## model=VW- Arteon                 -1.971869
## model=Audi- Q2                   -2.021361
## model=Audi- A3                   -2.123430
## model=VW- Tiguan Allspace        -2.160671
## model=Mercedes- V Class          -2.179079
## Audi=No                          -2.321314
## f.price=Segmento -  B            -2.354921
## model=Mercedes- B Class          -3.021960
## model=Mercedes- C Class          -3.593935
## fuelType=f.Fuel-Petrol           -3.714542
## model=BMW- X2                    -3.857106
## model=VW- Polo                   -3.992148
## model=VW- Passat                 -4.344141
## model=VW- Up                     -4.481049
## model=Audi- A1                   -4.504406
## model=VW- T-Cross                -4.569135
## fuelType=f.Fuel-Hybrid           -4.966090
## model=VW- Golf                   -5.935791
## model=VW- T-Roc                  -6.537802
## manufacturer=f.Man-VW            -8.159862
## f.miles=f.miles-(6,17]          -12.621390
## engineSize=Petit                -13.074438
## aux=(5.89e+03,1.69e+04]         -13.868657
## year=year_2020                  -15.713338
## year=year_2018                  -18.225664
## f.price=Segmento -  A           -19.141908
## aux=[0,5.89e+03]                -30.525067
## f.miles=f.miles-[0,6]           -31.303101
## year=year_2019                  -37.221371
## years_sell=Molt nou                    -Inf
##
## $`3`
##                                    Cla/Mod     Mod/Cla      Global
## years_sell=Semi nou              43.7916667 94.94128275 48.36759371
## f.price=Segmento -  D            54.9202128 74.61607949 30.31035873
## year=year_2016                   59.8843931 46.79313460 17.43248690
## aux=(3.4e+04,3.23e+05]           46.7165420 50.76784101 24.24425635
## f.miles=f.miles-(34,323]         46.7055880 50.58717254 24.16364369
## transmission=f.Trans-Manual      38.2641921 63.32429991 36.92059653
## year=year_2015                   57.8125000 20.05420054  7.73881499
## engineSize=Petit                 32.5842697 55.01355014 37.66626360
## f.miles=f.miles-(17,34]          33.6767036 38.39205059 25.43329303
## aux=(1.69e+04,3.4e+04]           33.4642577 38.48238482 25.65497783
## model=VW- Polo                   44.5355191 14.72448058  7.37605804
## manufacturer=f.Man-VW            30.9184994 43.17976513 31.15679162
## year=year_2014                   50.0000000  8.85275519  3.95002015
## model=Audi- A1                   53.6423841  7.31707317  3.04312777
## model=VW- Golf                   37.5000000 15.98915989  9.51229343
## model=VW- Up                     50.5494505  4.15537489  1.83393793
```

```
## fuelType=f.Fuel-Hybrid          50.6024096  3.79403794  1.67271262
## model=Audi- A3                  39.1752577  6.86540199  3.90971383
## model=VW- Passat                43.2692308  4.06504065  2.09592906
## fuelType=f.Fuel-Diesel          24.7693400 63.05329720 56.79161628
## model=BMW- 1 Series             35.4497354  6.05239386  3.80894800
## model=VW- Scirocco              57.1428571  1.44534779  0.56428859
## model=VW- Golf SV               50.0000000  1.17434508  0.52398227
## years_sell=Vell                 30.3571429  4.60704607  3.38573156
## year=year_2013                  30.3571429  4.60704607  3.38573156
## model=VW- CC                    55.5555556  0.45167118  0.18137848
## model=Audi- A4                  30.2521008  3.25203252  2.39822652
## model=VW- Jetta                100.0000000  0.18066847  0.04030633
## model=BMW- Z4                    0.0000000  0.00000000  0.24183797
## model=BMW- M4                    0.0000000  0.00000000  0.24183797
## model=BMW- 7 Series              0.0000000  0.00000000  0.24183797
## model=BMW- 6 Series              0.0000000  0.00000000  0.26199113
## f.price=Segmento -  C           20.0680272 21.31887986 23.70012092
## model=Mercedes- S Class          0.0000000  0.00000000  0.28214430
## model=Mercedes- B Class         10.5263158  0.54200542  1.14873035
## model=BMW- X5                    7.8947368  0.27100271  0.76582023
## model=Mercedes- CLS Class        0.0000000  0.00000000  0.30229746
## model=BMW- 4 Series             12.0000000  1.08401084  2.01531640
## model=BMW- X4                    0.0000000  0.00000000  0.38291012
## model=VW- Arteon                 0.0000000  0.00000000  0.40306328
## model=Mercedes- GLA Class        8.8235294  0.54200542  1.37041516
## model=VW- Sharan                 0.0000000  0.00000000  0.46352277
## model=BMW- X2                    0.0000000  0.00000000  0.52398227
## model=Mercedes- E Class         13.4328358  2.43902439  4.05078597
## model=Mercedes- V Class          0.0000000  0.00000000  0.54413543
## model=VW- Touareg                0.0000000  0.00000000  0.60459492
## model=VW- T-Cross                0.0000000  0.00000000  0.70536074
## model=Audi- Q7                   0.0000000  0.00000000  0.72551391
## model=BMW- X1                    6.5217391  0.54200542  1.85409109
## model=Mercedes- GLE Class        0.0000000  0.00000000  0.90689238
## model=BMW- X3                    0.0000000  0.00000000  1.02781137
## engineSize=Mitjà                19.4988254 44.98644986 51.47118098
## model=Audi- Q3                   5.6910569  0.63233966  2.47883918
## manufacturer=f.Man-Mercedes     17.3210162 20.32520325 26.17896010
## model=Audi- A5                   2.4096386  0.18066847  1.67271262
## model=VW- T-Roc                  0.0000000  0.00000000  1.35026199
## manufacturer=f.Man-BMW          16.1349134 15.98915989 22.10802096
## transmission=f.Trans-Automatic  16.4781906 18.42818428 24.94961709
## model=Audi- Q2                   0.0000000  0.00000000  1.65255945
## fuelType=f.Fuel-Petrol          17.8068899 33.15266486 41.53567110
## model=Audi- Q5                   0.0000000  0.00000000  1.97501008
## model=Mercedes- GLC Class        0.0000000  0.00000000  2.33776703
## model=VW- Tiguan                 0.0000000  0.00000000  3.64772269
## year=year_2020                   0.2873563  0.09033424  7.01330109
## f.miles=f.miles-(6,17]           9.1054313 10.29810298 25.23176139
## year=year_2018                   0.4357298  0.18066847  9.25030230
## aux=(5.89e+03,1.69e+04]          8.5582113 10.02710027 26.13865377
## transmission=f.Trans-SemiAuto   10.6765328 18.24751581 38.12978638
## engineSize=Gran                  0.0000000  0.00000000 10.86255542
## f.price=Segmento -  B            3.6538462  3.43270099 20.95929061
## aux=[0,5.89e+03]                 0.6728343  0.72267389 23.96211205
## f.miles=f.miles-[0,6]            0.6405124  0.72267389 25.17130189
## f.price=Segmento -  A            0.5636071  0.63233966 25.03022975
## year=year_2019                   0.1260239  0.18066847 31.98307134
## years_sell=Molt nou              0.2088555  0.45167118 48.24667473
##                                     p.value     v.test
## years_sell=Semi nou            2.722668e-314  37.898515
## f.price=Segmento -  D          4.259098e-273  35.309091
## year=year_2016                 2.571091e-160  26.979512
## aux=(3.4e+04,3.23e+05]         3.458544e-109  22.199686
```

```
## f.miles=f.miles-(34,323]       1.393502e-108  22.136951
## transmission=f.Trans-Manual     4.871508e-92  20.347642
## year=year_2015                  5.043800e-56  15.769512
## engineSize=Petit                1.372981e-40  13.338995
## f.miles=f.miles-(17,34]         1.237240e-27  10.893548
## aux=(1.69e+04,3.4e+04]          5.343760e-27  10.759538
## model=VW- Polo                  7.471417e-23   9.841333
## manufacturer=f.Man-VW           7.313715e-22   9.609169
## year=year_2014                  4.294643e-18   8.670702
## model=Audi- A1                  1.884819e-17   8.500681
## model=VW- Golf                  3.195043e-15   7.883043
## model=VW- Up                    3.020970e-09   5.930455
## fuelType=f.Fuel-Hybrid          1.456297e-08   5.666634
## model=Audi- A3                  6.937206e-08   5.392790
## model=VW- Passat                1.622901e-06   4.795475
## fuelType=f.Fuel-Diesel          1.644914e-06   4.792773
## model=BMW- 1 Series             2.742818e-05   4.193831
## model=VW- Scirocco              7.810464e-05   3.950143
## model=VW- Golf SV               2.166326e-03   3.066428
## years_sell=Vell                 1.371614e-02   2.464613
## year=year_2013                  1.371614e-02   2.464613
## model=VW- CC                    3.619636e-02   2.094715
## model=Audi- A4                  4.142961e-02   2.039205
## model=VW- Jetta                 4.973668e-02   1.962222
## model=BMW- Z4                   4.816701e-02  -1.975892
## model=BMW- M4                   4.816701e-02  -1.975892
## model=BMW- 7 Series             4.816701e-02  -1.975892
## model=BMW- 6 Series             3.739512e-02  -2.081424
## f.price=Segmento -  C           3.346964e-02  -2.126404
## model=Mercedes- S Class         2.903052e-02  -2.183072
## model=Mercedes- B Class         2.393764e-02  -2.258129
## model=BMW- X5                   2.272629e-02  -2.278005
## model=Mercedes- CLS Class       2.253562e-02  -2.281217
## model=BMW- 4 Series             8.598760e-03  -2.627608
## model=BMW- X4                   8.178521e-03  -2.644610
## model=VW- Arteon                6.346916e-03  -2.729299
## model=Mercedes- GLA Class       3.857106e-03  -2.889618
## model=VW- Sharan                2.965340e-03  -2.971308
## model=BMW- X2                   1.384706e-03  -3.197820
## model=Mercedes- E Class         1.239135e-03  -3.229715
## model=Mercedes- V Class         1.074157e-03  -3.270347
## model=VW- Touareg               5.012401e-04  -3.480093
## model=VW- T-Cross               1.405506e-04  -3.807197
## model=Audi- Q7                  1.089716e-04  -3.869695
## model=BMW- X1                   4.592045e-05  -4.075475
## model=Mercedes- GLE Class       1.100245e-05  -4.396472
## model=BMW- X3                   2.379320e-06  -4.718207
## engineSize=Mitjà                9.784321e-07  -4.895927
## model=Audi- Q3                  4.373835e-07  -5.051923
## manufacturer=f.Man-Mercedes     3.082765e-07  -5.118316
## model=Audi- A5                  2.236479e-07  -5.178524
## model=VW- T-Roc                 3.967221e-08  -5.492305
## manufacturer=f.Man-BMW          1.119746e-08  -5.711515
## transmission=f.Trans-Automatic  5.901446e-09  -5.819527
## model=Audi- Q2                  8.425911e-10  -6.136689
## fuelType=f.Fuel-Petrol          9.467170e-11  -6.475223
## model=Audi- Q5                  1.363790e-11  -6.761703
## model=Mercedes- GLC Class       1.294102e-13  -7.406772
## model=VW- Tiguan                5.437434e-21  -9.400365
## year=year_2020                  1.961920e-38 -12.963834
## f.miles=f.miles-(6,17]          6.718460e-44 -13.895783
## year=year_2018                  7.824017e-50 -14.842123
## aux=(5.89e+03,1.69e+04]         5.377115e-50 -14.867258
## transmission=f.Trans-SemiAuto   3.507023e-58 -16.080295
```

```
## engineSize=Gran                  8.534026e-64 -16.862208
## f.price=Segmento -  B            4.692767e-77 -18.579682
## aux=[0,5.89e+03]                5.263746e-136 -24.821427
## f.miles=f.miles-[0,6]           6.847079e-145 -25.631227
## f.price=Segmento -  A           1.170564e-145 -25.699944
## year=year_2019                  1.289273e-211 -31.046065
## years_sell=Molt nou              0.000000e+00       -Inf
##
## $‘4‘
##                                   Cla/Mod Mod/Cla     Global      p.value
## model=BMW- i3                   100.0000000     100  0.1612253 1.103363e-25
## fuelType=f.Fuel-Hybrid            9.6385542     100  1.6727126 4.352019e-15
## manufacturer=f.Man-BMW            0.7292616     100 22.1080210 5.594232e-06
## transmission=f.Trans-Automatic    0.6462036     100 24.9496171 1.476116e-05
## engineSize=Petit                  0.4280364     100 37.6662636 4.013814e-04
## years_sell=Semi nou               0.3333333     100 48.3675937 2.977253e-03
## year=year_2019                    0.0000000       0 31.9830713 4.568592e-02
## transmission=f.Trans-Manual       0.0000000       0 36.9205965 2.498414e-02
## transmission=f.Trans-SemiAuto     0.0000000       0 38.1297864 2.139640e-02
## fuelType=f.Fuel-Petrol            0.0000000       0 41.5356711 1.359516e-02
## years_sell=Molt nou               0.0000000       0 48.2466747 5.119388e-03
## engineSize=Mitjà                  0.0000000       0 51.4711810 3.057697e-03
## fuelType=f.Fuel-Diesel            0.0000000       0 56.7916163 1.205918e-03
##                                   v.test
## model=BMW- i3                   10.476869
## fuelType=f.Fuel-Hybrid           7.844350
## manufacturer=f.Man-BMW           4.541172
## transmission=f.Trans-Automatic   4.332223
## engineSize=Petit                 3.539174
## years_sell=Semi nou              2.970077
## year=year_2019                  -1.998284
## transmission=f.Trans-Manual     -2.241648
## transmission=f.Trans-SemiAuto   -2.300916
## fuelType=f.Fuel-Petrol          -2.467786
## years_sell=Molt nou             -2.799424
## engineSize=Mitjà                -2.961877
## fuelType=f.Fuel-Diesel          -3.237477
```

8.1.2 Description of clusters by quantitative variables

```
quanti_var_decription_1
```

```
## $‘1‘
##                v.test Mean in category Overall mean sd in category   Overall sd
## price        41.17280     2.732510e+04 2.116983e+04   1.001034e+04 1.014634e+04
## tax          24.61101     1.466471e+02 1.256561e+02   1.084623e+01 5.788649e+01
## total       -10.81327     3.850984e-02 8.424023e-02   1.924236e-01 2.870255e-01
## mpg         -20.03724     4.809773e+01 5.464395e+01   1.148130e+01 2.217305e+01
## mileage     -47.53185     7.719697e+03 2.298127e+04   6.891752e+03 2.179152e+04
## years_sell2 -66.66933     1.000000e+00 1.551391e+00   0.000000e+00 5.613142e-01
##                   p.value
## price        0.000000e+00
## tax         9.630974e-134
## total        2.978453e-27
## mpg          2.608228e-89
## mileage      0.000000e+00
## years_sell2  0.000000e+00
##
## $‘2‘
##                 v.test Mean in category Overall mean sd in category
## years_sell2  42.806824         2.080247     1.551391   2.716751e-01
## tax          27.468303       160.652851   125.656103   5.002183e+01
```

```
## mileage      26.931671     35898.507382 22981.274838   2.125782e+04
## mpg          -3.371934        52.998354    54.643950   9.193697e+00
## price       -16.674934     17445.978052 21169.829101   6.599298e+03
##               Overall sd       p.value
## years_sell2 5.613142e-01  0.000000e+00
## tax         5.788649e+01 4.200272e-166
## mileage     2.179152e+04 9.353032e-160
## mpg         2.217305e+01  7.464227e-04
## price       1.014634e+04  1.994309e-62
##
## $`3`
##                v.test Mean in category Overall mean sd in category   Overall sd
## years_sell2  32.95952     2.041554e+00 1.551391e+00   2.210440e-01 5.613142e-01
## mileage      27.43003     3.881808e+04 2.298127e+04   2.130771e+04 2.179152e+04
## mpg          22.61803     6.793117e+01 5.464395e+01   1.684542e+01 2.217305e+01
## total        12.79910     1.815718e-01 8.424023e-02   3.901498e-01 2.870255e-01
## price       -31.08281     1.281412e+04 2.116983e+04   4.223167e+03 1.014634e+04
## tax         -59.22925     3.481818e+01 1.256561e+02   2.437484e+01 5.788649e+01
##                   p.value
## years_sell2 3.090313e-238
## mileage     1.202678e-165
## mpg         2.880378e-113
## total        1.658529e-37
## price       4.112469e-212
## tax          0.000000e+00
##
## $`4`
##                v.test Mean in category Overall mean sd in category Overall sd
## mpg         53.122981           470.80  54.64394965   5.684342e-14 22.1730472
## total       16.426442             1.75   0.08424023   4.330127e-01  0.2870255
## years_sell2  2.262111             2.00   1.55139057   0.000000e+00  0.5613142
## tax         -3.760410            48.75 125.65610302   4.979646e+01 57.8864858
##                  p.value
## mpg         0.000000e+00
## total       1.237075e-60
## years_sell2 2.369053e-02
## tax         1.696349e-04
```

8.2 MCA

8.2.1 Interpreting the axes association to factor map.

```
res.desc_1[[1]]
```

8.2.1.1 Dimension 1

```
## $quanti
##       correlation p.value
## price  -0.6685324       0
##
## $quali
##                       R2       p.value
## years_sell   0.783708357  0.000000e+00
## f.price      0.526348699  0.000000e+00
## f.miles      0.738558553  0.000000e+00
## f.tax        0.708527406  0.000000e+00
## mpg_d        0.484668713  0.000000e+00
## transmission 0.141519789 4.833372e-165
## fuelType     0.100633265 6.114288e-115
## manufacturer 0.005301294  8.031455e-06
```

```
##
## $category
##                                   Estimate       p.value
## mpg_d=mpg_d-(61.4,471]          0.58860292 2.035550e-321
## f.tax=f.tax-(1,125]             0.60854447  0.000000e+00
## f.miles=f.miles-(34,323]        0.69736358  0.000000e+00
## f.price=Segmento -  D           0.62182529  0.000000e+00
## years_sell=Semi nou             0.34355373  0.000000e+00
## transmission=f.Trans-Manual     0.29976608 6.787033e-144
## f.miles=f.miles-(17,34]         0.37472853 3.458023e-136
## fuelType=f.Fuel-Diesel          0.03429396  4.133402e-94
## mpg_d=mpg_d-(53.3,61.4]         0.27721154  1.305290e-77
## years_sell=Vell                 0.45735298  8.974331e-42
## f.price=Segmento -  C           0.21929175  3.086396e-27
## f.tax=f.tax-(150,570]           0.06294199  1.382364e-18
## fuelType=f.Fuel-Hybrid          0.33438754  6.024471e-11
## manufacturer=f.Man-VW           0.05955263  4.551059e-05
## transmission=f.Trans-Automatic -0.03668316  1.799889e-02
## manufacturer=f.Man-Mercedes    -0.05546428  1.387306e-04
## mpg_d=mpg_d-(44.8,53.3]        -0.27287507  4.691158e-68
## f.price=Segmento -  B          -0.27609277  3.210704e-72
## fuelType=f.Fuel-Petrol         -0.36868150 2.851125e-112
## transmission=f.Trans-SemiAuto  -0.26308292 1.075159e-117
## f.miles=f.miles-(6,17]         -0.36523262 6.186170e-119
## mpg_d=mpg_d-[0,44.8]           -0.59293939 6.936682e-321
## f.tax=f.tax-(145,150]          -0.61374728  0.000000e+00
## f.miles=f.miles-[0,6]          -0.70685949  0.000000e+00
## f.price=Segmento -  A          -0.56502427  0.000000e+00
## years_sell=Molt nou            -0.80090671  0.000000e+00
##
## attr(,"class")
## [1] "condes" "list"
```

```r
res.desc_1[[2]]
```

8.2.1.2 Dimension 2

```
## $quanti
##       correlation     p.value
## price   0.3389601 1.1795e-133
##
## $quali
##                     R2       p.value
## transmission 0.352265910  0.000000e+00
## fuelType     0.588464464  0.000000e+00
## manufacturer 0.345487645  0.000000e+00
## f.tax        0.354749866  0.000000e+00
## mpg_d        0.157261328 1.384190e-183
## f.price      0.114470234 2.490716e-130
## Audi         0.018251044  1.202426e-21
## f.miles      0.015529383  1.000744e-16
## years_sell   0.003921919  5.867376e-05
##
## $category
##                                   Estimate       p.value
## f.tax=f.tax-(125,145]           2.55269179  0.000000e+00
## fuelType=f.Fuel-Hybrid          1.74321676  0.000000e+00
## transmission=f.Trans-Automatic  0.31838374 3.715684e-192
## manufacturer=f.Man-Mercedes     0.26781187 9.437891e-134
## mpg_d=mpg_d-(61.4,471]          0.30525092 1.009807e-133
## manufacturer=f.Man-BMW          0.28424901 6.072686e-119
```

```
## f.price=Segmento -  A           0.20403226  6.499369e-71
## transmission=f.Trans-SemiAuto   0.09766548  2.512642e-49
## Audi=No                         0.08517693  1.202426e-21
## mpg_d=mpg_d-[0,44.8]            0.06849633  3.987135e-10
## f.price=Segmento -  B           0.06478201  1.085534e-08
## f.miles=f.miles-(34,323]        0.06969395  3.415322e-08
## f.miles=f.miles-[0,6]           0.05383054  1.110592e-05
## years_sell=Molt nou             0.04890632  1.198226e-05
## f.miles=f.miles-(6,17]         -0.04225571  9.042527e-04
## years_sell=Semi nou            -0.01291821  1.009466e-04
## f.tax=f.tax-(145,150]          -0.88381370  4.714238e-06
## f.tax=f.tax-(150,570]          -0.74065897  5.417575e-07
## f.tax=f.tax-(1,125]            -0.92821912  5.737612e-10
## f.miles=f.miles-(17,34]        -0.08126878  8.180896e-11
## Audi=Yes                       -0.08517693  1.202426e-21
## manufacturer=f.Man-Audi        -0.15767351  1.202426e-21
## mpg_d=mpg_d-(44.8,53.3]        -0.16906656  1.081456e-41
## mpg_d=mpg_d-(53.3,61.4]        -0.20468068  6.711637e-56
## f.price=Segmento -  D          -0.24741321 6.613554e-104
## fuelType=f.Fuel-Diesel         -0.65952373 5.141477e-108
## fuelType=f.Fuel-Petrol         -1.08369303 5.795040e-280
## manufacturer=f.Man-VW          -0.39438737 1.006336e-299
## transmission=f.Trans-Manual    -0.41604922  0.000000e+00
##
## attr(,"class")
## [1] "condes" "list"
```

8.2.2 MCA with all suppementary variables

8.2.2.1 Interpreting the axes association to factor map.

```
res.desc[[1]]
```

8.2.2.1.1 Dimension 1

```
##         correlation       p.value
## mileage   0.6842993  0.000000e+00
## mpg       0.4030493 3.286102e-193
## tax      -0.5718568  0.000000e+00
## price    -0.8199518  0.000000e+00
```

```
res.desc[[2]]
```

8.2.2.2 Dimension 2

```
##                       R2       p.value
## transmission 0.303571863  0.000000e+00
## years_sell   0.675052384  0.000000e+00
## f.price      0.789493956  0.000000e+00
## f.miles      0.611316081  0.000000e+00
## f.tax        0.635957502  0.000000e+00
## mpg_d        0.465994394  0.000000e+00
## engineSize   0.129059957 1.588135e-149
## manufacturer 0.063577186  2.701756e-70
## fuelType     0.011400973  4.492969e-13
## Audi         0.002710043  2.438722e-04
```

8.3 Hierarchical Clustering MCA

8.3.1 Description of clusters

```
res.hcpcMCA$desc.var$category
```

8.3.1.1 Correlation with categories

```
## $`1`
##                                   Cla/Mod     Mod/Cla     Global       p.value
## f.tax=f.tax-(145,150]          51.8276762  96.0677556 61.7492946  0.000000e+00
## years_sell=Molt nou            63.5338346  92.0145191 48.2466747  0.000000e+00
## f.price=Segmento -  A          73.3494364  55.1119177 25.0302297 1.588211e-253
## Audi=No                        41.9330289 100.0000000 79.4437727 8.780284e-208
## f.miles=f.miles-[0,6]          63.1705364  47.7313975 25.1713019 1.079805e-141
## transmission=f.Trans-SemiAuto  54.3868922  62.2504537 38.1297864 6.254828e-134
## manufacturer=f.Man-Mercedes    56.9668976  44.7670901 26.1789601  1.290037e-94
## mpg_d=mpg_d-[0,44.8]           56.4715581  41.4398064 24.4457880  1.024526e-82
## f.miles=f.miles-(6,17]         53.9137380  40.8348457 25.2317614  8.624378e-69
## f.price=Segmento -  B          56.0576923  35.2692075 20.9592906  2.112916e-65
## engineSize=Mitjà               41.3860611  63.9443436 51.4711810  9.037761e-36
## manufacturer=f.Man-BMW         48.1312671  31.9419238 22.1080210  5.518332e-31
## engineSize=Gran                47.6808905  15.5474894 10.8625554  2.522215e-13
## transmission=f.Trans-Automatic 39.4184168  29.5220811 24.9496171  1.852427e-07
## f.tax=f.tax-(125,145]          60.5263158   1.3914096  0.7658202  6.561038e-04
## fuelType=f.Fuel-Diesel         34.9183818  59.5281307 56.7916163  5.912707e-03
## fuelType=f.Fuel-Petrol         30.9558467  38.5964912 41.5356711  2.946276e-03
## mpg_d=mpg_d-(53.3,61.4]        23.8019169  18.0278282 25.2317614  3.873884e-17
## manufacturer=f.Man-VW          24.9029754  23.2909861 31.1567916  9.940966e-18
## years_sell=Vell                 0.0000000   0.0000000  3.3857316  6.420378e-31
## f.tax=f.tax-(150,570]           9.6997691   2.5408348  8.7263200  6.198979e-33
## mpg_d=mpg_d-(61.4,471]         17.1779141  11.8572293 22.9947602  6.416750e-43
## f.price=Segmento -  C          13.5204082   9.6188748 23.7001209  4.470515e-68
## f.miles=f.miles-(17,34]        14.1045959  10.7683001 25.4332930  9.258414e-70
## engineSize=Petit               18.1380417  20.5081670 37.6662636  3.401623e-73
## Audi=Yes                        0.0000000   0.0000000 20.5562273 8.780284e-208
## manufacturer=f.Man-Audi         0.0000000   0.0000000 20.5562273 8.780284e-208
## transmission=f.Trans-Manual     7.4235808   8.2274652 36.9205965 2.130331e-221
## f.miles=f.miles-(34,323]        0.9174312   0.6654567 24.1636437 2.671626e-226
## f.tax=f.tax-(1,125]             0.0000000   0.0000000 28.7585651 3.973098e-311
## f.price=Segmento -  D           0.0000000   0.0000000 30.3103587  0.000000e+00
## years_sell=Semi nou             5.5000000   7.9854809 48.3675937  0.000000e+00
##                                    v.test
## f.tax=f.tax-(145,150]                 Inf
## years_sell=Molt nou                   Inf
## f.price=Segmento -  A           34.009943
## Audi=No                         30.760758
## f.miles=f.miles-[0,6]           25.342766
## transmission=f.Trans-SemiAuto   24.628508
## manufacturer=f.Man-Mercedes     20.636528
## mpg_d=mpg_d-[0,44.8]            19.266598
## f.miles=f.miles-(6,17]          17.528885
## f.price=Segmento -  B           17.079397
## engineSize=Mitjà                12.484793
## manufacturer=f.Man-BMW          11.574981
## engineSize=Gran                  7.317719
## transmission=f.Trans-Automatic   5.213568
## f.tax=f.tax-(125,145]            3.407294
## fuelType=f.Fuel-Diesel           2.752583
## fuelType=f.Fuel-Petrol          -2.973288
## mpg_d=mpg_d-(53.3,61.4]         -8.416654
```

```
## manufacturer=f.Man-VW            -8.574626
## years_sell=Vell                 -11.561990
## f.tax=f.tax-(150,570]           -11.953845
## mpg_d=mpg_d-(61.4,471]          -13.733274
## f.price=Segmento -  C           -17.435065
## f.miles=f.miles-(17,34]         -17.655334
## engineSize=Petit                -18.096404
## Audi=Yes                        -30.760758
## manufacturer=f.Man-Audi         -30.760758
## transmission=f.Trans-Manual     -31.762555
## f.miles=f.miles-(34,323]        -32.115586
## f.tax=f.tax-(1,125]             -37.705918
## f.price=Segmento -  D                 -Inf
## years_sell=Semi nou                   -Inf
##
## $`2`
##                                Cla/Mod      Mod/Cla     Global       p.value
## Audi=Yes                     100.000000 100.00000000 20.5562273  0.000000e+00
## manufacturer=f.Man-Audi      100.000000 100.00000000 20.5562273  0.000000e+00
## mpg_d=mpg_d-[0,44.8]          28.276999  33.62745098 24.4457880  8.810469e-14
## fuelType=f.Fuel-Petrol        23.289665  47.05882353 41.5356711  6.359473e-05
## transmission=f.Trans-Manual   23.253275  41.76470588 36.9205965  3.504723e-04
## f.price=Segmento -  A         23.993559  29.21568627 25.0302297  6.266267e-04
## f.miles=f.miles-(34,323]      23.769808  27.94117647 24.1636437  1.776496e-03
## f.tax=f.tax-(150,570]         24.480370  10.39215686  8.7263200  3.756852e-02
## engineSize=Mitjà              19.459671  48.72549020 51.4711810  4.918089e-02
## years_sell=Molt nou           19.340017  45.39215686 48.2466747  4.068097e-02
## transmission=f.Trans-SemiAuto 18.604651  34.50980392 38.1297864  7.357583e-03
## fuelType=f.Fuel-Diesel        19.162527  52.94117647 56.7916163  5.467719e-03
## f.tax=f.tax-(125,145]          2.631579   0.09803922  0.7658202  1.831251e-03
## f.miles=f.miles-(6,17]        17.332268  21.27450980 25.2317614  9.541507e-04
## f.price=Segmento -  D         16.821809  24.80392157 30.3103587  1.371748e-05
## mpg_d=mpg_d-(53.3,61.4]       16.214058  19.90196078 25.2317614  7.625071e-06
## mpg_d=mpg_d-(61.4,471]        15.337423  17.15686275 22.9947602  3.600055e-07
## fuelType=f.Fuel-Hybrid         0.000000   0.00000000  1.6727126  4.235668e-09
## manufacturer=f.Man-BMW         0.000000   0.00000000 22.1080210 8.940990e-127
## manufacturer=f.Man-Mercedes    0.000000   0.00000000 26.1789601 1.655015e-154
## manufacturer=f.Man-VW          0.000000   0.00000000 31.1567916 5.491953e-191
## Audi=No                        0.000000   0.00000000 79.4437727  0.000000e+00
##                                  v.test
## Audi=Yes                            Inf
## manufacturer=f.Man-Audi             Inf
## mpg_d=mpg_d-[0,44.8]           7.457612
## fuelType=f.Fuel-Petrol         3.999059
## transmission=f.Trans-Manual    3.574817
## f.price=Segmento -  A          3.419820
## f.miles=f.miles-(34,323]       3.125257
## f.tax=f.tax-(150,570]          2.079532
## engineSize=Mitjà              -1.967020
## years_sell=Molt nou           -2.046767
## transmission=f.Trans-SemiAuto -2.680211
## fuelType=f.Fuel-Diesel        -2.778104
## f.tax=f.tax-(125,145]         -3.116317
## f.miles=f.miles-(6,17]        -3.303708
## f.price=Segmento -  D         -4.348335
## mpg_d=mpg_d-(53.3,61.4]       -4.475450
## mpg_d=mpg_d-(61.4,471]        -5.088974
## fuelType=f.Fuel-Hybrid        -5.874717
## manufacturer=f.Man-BMW       -23.951372
## manufacturer=f.Man-Mercedes  -26.479830
## manufacturer=f.Man-VW        -29.478123
## Audi=No                            -Inf
##
## $`3`
```

```
##                                Cla/Mod     Mod/Cla     Global       p.value
## manufacturer=f.Man-VW        54.786546  89.8197243 31.1567916  0.000000e+00
## engineSize=Petit             48.368111  95.8642630 37.6662636  0.000000e+00
## transmission=f.Trans-Manual  44.541485  86.5323436 36.9205965 1.955737e-272
## fuelType=f.Fuel-Petrol       38.864629  84.9416755 41.5356711 9.964455e-206
## f.price=Segmento -  D        43.018617  68.6108165 30.3103587 8.969567e-165
## Audi=No                      23.921867 100.0000000 79.4437727 9.495517e-107
## mpg_d=mpg_d-(44.8,53.3]      30.530973  43.9024390 27.3276904  1.645745e-34
## mpg_d=mpg_d-(53.3,61.4]      30.351438  40.2969247 25.2317614  4.422999e-30
## f.miles=f.miles-(17,34]      25.832013  34.5705196 25.4332930  2.950765e-12
## f.tax=f.tax-(1,125]          25.017519  37.8579003 28.7585651  1.796900e-11
## years_sell=Semi nou          21.666667  55.1431601 48.3675937  3.767272e-06
## f.miles=f.miles-(6,17]       22.523962  29.9045599 25.2317614  2.945123e-04
## f.tax=f.tax-(125,145]         2.631579   0.1060445  0.7658202  3.534774e-03
## f.miles=f.miles-[0,6]        14.971978  19.8303287 25.1713019  1.899766e-05
## years_sell=Molt nou          16.290727  41.3573701 48.2466747  2.444245e-06
## fuelType=f.Fuel-Hybrid        0.000000   0.0000000  1.6727126  2.145462e-08
## f.miles=f.miles-(34,323]     12.343620  15.6945917 24.1636437  2.657075e-12
## mpg_d=mpg_d-(61.4,471]        9.903593  11.9830329 22.9947602  3.897659e-21
## f.tax=f.tax-(150,570]         2.078522   0.9544008  8.7263200  1.852979e-29
## f.price=Segmento -  B         6.826923   7.5291622 20.9592906  1.064644e-34
## manufacturer=f.Man-BMW        5.287147   6.1505832 22.1080210  4.707161e-48
## engineSize=Gran               0.000000   0.0000000 10.8625554  2.618044e-53
## transmission=f.Trans-Automatic  3.554120   4.6659597 24.9496171  2.473312e-73
## mpg_d=mpg_d-[0,44.8]          2.967848   3.8176034 24.4457880  1.786768e-78
## manufacturer=f.Man-Mercedes   2.925327   4.0296925 26.1789601  8.716364e-86
## Audi=Yes                      0.000000   0.0000000 20.5562273 9.495517e-107
## manufacturer=f.Man-Audi       0.000000   0.0000000 20.5562273 9.495517e-107
## transmission=f.Trans-SemiAuto   4.386892   8.8016967 38.1297864 3.539147e-111
## f.price=Segmento -  A         0.000000   0.0000000 25.0302297 5.353711e-134
## fuelType=f.Fuel-Diesel        5.039035  15.0583245 56.7916163 2.859813e-190
## engineSize=Mitjà              1.527016   4.1357370 51.4711810 4.689251e-270
##                                 v.test
## manufacturer=f.Man-VW              Inf
## engineSize=Petit                   Inf
## transmission=f.Trans-Manual  35.265929
## fuelType=f.Fuel-Petrol       30.606713
## f.price=Segmento -  D        27.356777
## Audi=No                      21.945813
## mpg_d=mpg_d-(44.8,53.3]      12.251669
## mpg_d=mpg_d-(53.3,61.4]      11.395104
## f.miles=f.miles-(17,34]       6.980049
## f.tax=f.tax-(1,125]           6.721640
## years_sell=Semi nou           4.623825
## f.miles=f.miles-(6,17]        3.620080
## f.tax=f.tax-(125,145]        -2.916946
## f.miles=f.miles-[0,6]        -4.276357
## years_sell=Molt nou          -4.712726
## fuelType=f.Fuel-Hybrid       -5.599843
## f.miles=f.miles-(34,323]     -6.994763
## mpg_d=mpg_d-(61.4,471]       -9.435330
## f.tax=f.tax-(150,570]       -11.269653
## f.price=Segmento -  B       -12.286937
## manufacturer=f.Man-BMW      -14.564748
## engineSize=Gran             -15.369594
## transmission=f.Trans-Automatic -18.113953
## mpg_d=mpg_d-[0,44.8]        -18.754263
## manufacturer=f.Man-Mercedes -19.629148
## Audi=Yes                    -21.945813
## manufacturer=f.Man-Audi     -21.945813
## transmission=f.Trans-SemiAuto -22.404732
## f.price=Segmento -  A       -24.634814
## fuelType=f.Fuel-Diesel      -29.422159
## engineSize=Mitjà            -35.110329
```

```
##
## $`4`
##                                 Cla/Mod    Mod/Cla    Global       p.value
## years_sell=Semi nou          51.4583333  91.753343 48.367594  0.000000e+00
## f.miles=f.miles-(34,323]     62.9691410  56.092125 24.163644 5.422286e-208
## f.tax=f.tax-(1,125]          55.6412053  58.989599 28.758565 5.625916e-171
## Audi=No                      34.1451040 100.000000 79.443773 3.214146e-161
## fuelType=f.Fuel-Diesel       40.8800568  85.586924 56.791616 4.536537e-151
## mpg_d=mpg_d-(61.4,471]       57.5810692  48.811293 22.994760 7.054835e-141
## engineSize=Mitjà             37.6272514  71.396731 51.471181  1.766862e-67
## f.tax=f.tax-(150,570]        63.7413395  20.505201  8.726320  5.120744e-63
## manufacturer=f.Man-BMW       46.5815861  37.964339 22.108021  1.667917e-56
## f.price=Segmento -  C        45.3231293  39.598811 23.700121  1.626386e-54
## f.price=Segmento -  D        40.1595745  44.873700 30.310359  1.224249e-40
## f.miles=f.miles-(17,34]      40.6497623  38.112927 25.433293  3.192446e-34
## manufacturer=f.Man-Mercedes  40.1077752  38.707281 26.178960  6.057153e-33
## transmission=f.Trans-Automatic 37.4798061  34.472511 24.949617  2.087845e-20
## years_sell=Vell              54.1666667   6.760773  3.385732  5.818086e-14
## fuelType=f.Fuel-Hybrid       62.6506024   3.863299  1.672713  1.106755e-11
## mpg_d=mpg_d-(53.3,61.4]      29.6325879  27.563150 25.231761  2.178582e-02
## transmission=f.Trans-Manual  24.7816594  33.729569 36.920597  4.365791e-03
## transmission=f.Trans-SemiAuto 22.6215645  31.797920 38.129786  1.667446e-08
## manufacturer=f.Man-VW        20.3104787  23.328380 31.156792  1.511375e-13
## f.price=Segmento -  B        16.9230769  13.075780 20.959291  6.432574e-18
## mpg_d=mpg_d-[0,44.8]         12.2835944  11.069837 24.445788  1.869901e-45
## mpg_d=mpg_d-(44.8,53.3]      12.4631268  12.555721 27.327690  6.440639e-51
## engineSize=Petit             12.1455324  16.864785 37.666264  2.518102e-82
## f.miles=f.miles-(6,17]        6.2300319   5.794948 25.231761 6.144875e-100
## f.price=Segmento -  A         2.6570048   2.451709 25.030230 2.402219e-146
## Audi=Yes                      0.0000000   0.000000 20.556227 3.214146e-161
## manufacturer=f.Man-Audi       0.0000000   0.000000 20.556227 3.214146e-161
## fuelType=f.Fuel-Petrol        6.8898593  10.549777 41.535671 3.783709e-182
## f.miles=f.miles-[0,6]         0.0000000   0.000000 25.171302 2.486776e-204
## f.tax=f.tax-(145,150]         8.5835509  19.539376 61.749295 5.754164e-309
## years_sell=Molt nou           0.8354219   1.485884 48.246675  0.000000e+00
##                                  v.test
## years_sell=Semi nou                 Inf
## f.miles=f.miles-(34,323]      30.776406
## f.tax=f.tax-(1,125]           27.873282
## Audi=No                       27.056369
## fuelType=f.Fuel-Diesel        26.179609
## mpg_d=mpg_d-(61.4,471]        25.268711
## engineSize=Mitjà              17.356322
## f.tax=f.tax-(150,570]         16.755983
## manufacturer=f.Man-BMW        15.839253
## f.price=Segmento -  C         15.548593
## f.price=Segmento -  D         13.347540
## f.miles=f.miles-(17,34]       12.197823
## manufacturer=f.Man-Mercedes   11.955767
## transmission=f.Trans-Automatic  9.257750
## years_sell=Vell                7.512112
## fuelType=f.Fuel-Hybrid         6.791888
## mpg_d=mpg_d-(53.3,61.4]        2.294081
## transmission=f.Trans-Manual   -2.850446
## transmission=f.Trans-SemiAuto -5.643379
## manufacturer=f.Man-VW         -7.386153
## f.price=Segmento -  B         -8.624585
## mpg_d=mpg_d-[0,44.8]         -14.149912
## mpg_d=mpg_d-(44.8,53.3]      -15.008691
## engineSize=Petit             -19.219992
## f.miles=f.miles-(6,17]       -21.220740
## f.price=Segmento -  A        -25.761400
## Audi=Yes                     -27.056369
## manufacturer=f.Man-Audi      -27.056369
```

```
## fuelType=f.Fuel-Petrol         -28.780309
## f.miles=f.miles-[0,6]          -30.501532
## f.tax=f.tax-(145,150]          -37.573822
## years_sell=Molt nou                  -Inf
```

9 Finally, save the data

```
save.image("EloiOthman_del2.RData")
```