ADEI - Final Deliverable

Data Processing, Analysis and Modelization

Joan DOmingo Navarro

January 10, 2022

Contents

1 Introduction

1.1 Data Description

Descripció de les variables del dataset que usarem

- manufacturer Factor: Audi, BMW, Mercedes or Volkswagen
- model: Car model
- year: registration year
- price: price in £
- transmission: type of gearbox
- mileage: distance used
- fuelType: engine fuel
- tax: road tax
- mpg: Consumption in miles per gallon
- engineSize: size in litres

1.2 Load Required Packages

Carreguem els paquets necessaris

1.3 Some useful functions

```
# Some useful functions
calcQ <- function(x) {
  s.x <- summary(x)
  iqr<-s.x[5]-s.x[2]
  list(souti=s.x[2]-3*iqr, mouti=s.x[2]-1.5*iqr, min=s.x[1], q1=s.x[2], q2=s.x[3],
       q3=s.x[5], max=s.x[6], mouts=s.x[5]+1.5*iqr, souts=s.x[5]+3*iqr ) }

countNA <- function(x) {
  mis_x <- NULL
  for (j in 1:ncol(x)) {mis_x[j] <- sum(is.na(x[,j])) }
  mis_x <- as.data.frame(mis_x)
  rownames(mis_x) <- names(x)
  mis_i <- rep(0,nrow(x))
  for (j in 1:ncol(x)) {mis_i <- mis_i + as.numeric(is.na(x[,j])) }
  list(mis_col=mis_x,mis_ind=mis_i) }

countX <- function(x,X) {
  n_x <- NULL
  for (j in 1:ncol(x)) {n_x[j] <- sum(x[,j]==X) }
  n_x <- as.data.frame(n_x)
  rownames(n_x) <- names(x)
  nx_i <- rep(0,nrow(x))
  for (j in 1:ncol(x)) {nx_i <- nx_i + as.numeric(x[,j]==X) }
  list(nx_col=n_x,nx_ind=nx_i) }
```

1.4 Load data

Carreguem el dataset i definim el nostre directori de treball

```
setwd("D:/Drive/UNI/Q7/ADEI/")
filepath<-"D:/Drive/UNI/Q7/ADEI/"
# Data set: MyOldCars-Raw
load(paste0(filepath,"MyOldCars-Raw.RData"))
```

2 Validation of the Data Set: description of the process

2.1 Univariate Descriptive Analysis

Ens informem primer de les variables del data set

```
names(df)
```

```
##  [1] "model"        "year"         "price"        "transmission" "mileage"
##  [6] "fuelType"     "tax"          "mpg"          "engineSize"   "manufacturer"
```

```
summary(df)
```

```
##     model                year          price        transmission
##  Length:5000        Min.   :1999   Min.   :  1975   Length:5000
##  Class :character   1st Qu.:2016   1st Qu.: 13998   Class :character
##  Mode  :character   Median :2017   Median : 19498   Mode  :character
##                     Mean   :2017   Mean   : 21552
##                     3rd Qu.:2019   3rd Qu.: 26350
##                     Max.   :2020   Max.   :139559
##     mileage         fuelType              tax             mpg
##  Min.   :     1   Length:5000        Min.   :  0.0   Min.   : 11.00
##  1st Qu.:  5785   Class :character   1st Qu.:125.0   1st Qu.: 44.80
##  Median : 16741   Mode  :character   Median :145.0   Median : 53.30
```

```
##  Mean   : 23457                    Mean   :125.5   Mean   : 53.81
##  3rd Qu.: 34125                    3rd Qu.:145.0   3rd Qu.: 61.40
##  Max.   :240494                    Max.   :580.0   Max.   :470.80
##    engineSize    manufacturer
##  Min.   :0.000   Length:5000
##  1st Qu.:1.500   Class :character
##  Median :2.000   Mode  :character
##  Mean   :1.927
##  3rd Qu.:2.000
##  Max.   :6.600
```

Inicialitzem els vectors de missings, outliers i errors

```
imis<-rep(0,nrow(df))  # rows - trips
jmis<-rep(0,2*ncol(df))  # columns - variables
mis1<-countNA(df)
iouts<-rep(0,nrow(df))  # rows - trips
jouts<-rep(0,2*ncol(df))  # columns - variables
ierrs<-rep(0,nrow(df))  # rows - trips
jerrs<-rep(0,2*ncol(df))  # columns - variables
```

2.1.1 New variables

Creem la nova variable age a partir de la variable year

```
df$age <-  2021 - df$year
```

2.1.2 Variables qualitatives (factors)

Factoritzem les variables categoriques i mostrem un barplot de cada una d'aquestes

2.1.2.1 Model Car model

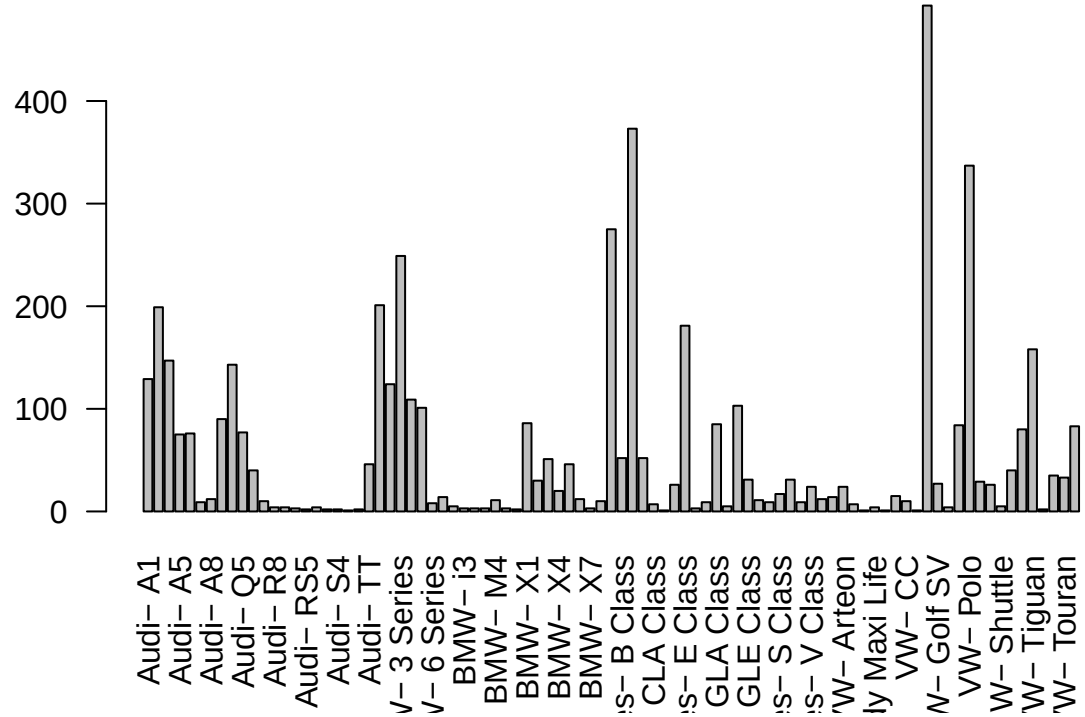
Factoritzem la variable model amb combinació del manufacturer per que quedi una variable categòrica amb forma 'Manufacturer - Model'

```
df$model<-factor(paste0(df$manufacturer,"-",df$model))
levels(df$model)
```

```
##  [1] "Audi- A1"            "Audi- A3"            "Audi- A4"
##  [4] "Audi- A5"            "Audi- A6"            "Audi- A7"
##  [7] "Audi- A8"            "Audi- Q2"            "Audi- Q3"
## [10] "Audi- Q5"            "Audi- Q7"            "Audi- Q8"
## [13] "Audi- R8"            "Audi- RS3"           "Audi- RS4"
## [16] "Audi- RS5"           "Audi- RS6"           "Audi- S3"
## [19] "Audi- S4"            "Audi- S5"            "Audi- SQ5"
## [22] "Audi- TT"            "BMW- 1 Series"       "BMW- 2 Series"
## [25] "BMW- 3 Series"       "BMW- 4 Series"       "BMW- 5 Series"
## [28] "BMW- 6 Series"       "BMW- 7 Series"       "BMW- 8 Series"
## [31] "BMW- i3"             "BMW- M2"             "BMW- M3"
## [34] "BMW- M4"             "BMW- M5"             "BMW- M6"
## [37] "BMW- X1"             "BMW- X2"             "BMW- X3"
## [40] "BMW- X4"             "BMW- X5"             "BMW- X6"
## [43] "BMW- X7"             "BMW- Z4"             "Mercedes- A Class"
## [46] "Mercedes- B Class"   "Mercedes- C Class"   "Mercedes- CL Class"
## [49] "Mercedes- CLA Class" "Mercedes- CLK"       "Mercedes- CLS Class"
## [52] "Mercedes- E Class"   "Mercedes- G Class"   "Mercedes- GL Class"
## [55] "Mercedes- GLA Class" "Mercedes- GLB Class" "Mercedes- GLC Class"
## [58] "Mercedes- GLE Class" "Mercedes- GLS Class" "Mercedes- M Class"
## [61] "Mercedes- S Class"   "Mercedes- SL CLASS"  "Mercedes- SLK"
## [64] "Mercedes- V Class"   "Mercedes- X-CLASS"   "VW- Amarok"
```

```
## [67] "VW- Arteon"           "VW- Beetle"           "VW- Caddy Life"
## [70] "VW- Caddy Maxi Life" "VW- California"       "VW- Caravelle"
## [73] "VW- CC"               "VW- Eos"              "VW- Golf"
## [76] "VW- Golf SV"          "VW- Jetta"            "VW- Passat"
## [79] "VW- Polo"             "VW- Scirocco"         "VW- Sharan"
## [82] "VW- Shuttle"          "VW- T-Cross"          "VW- T-Roc"
## [85] "VW- Tiguan"           "VW- Tiguan Allspace" "VW- Touareg"
## [88] "VW- Touran"           "VW- Up"
```

```
barplot(summary(df$model), las=2)
```

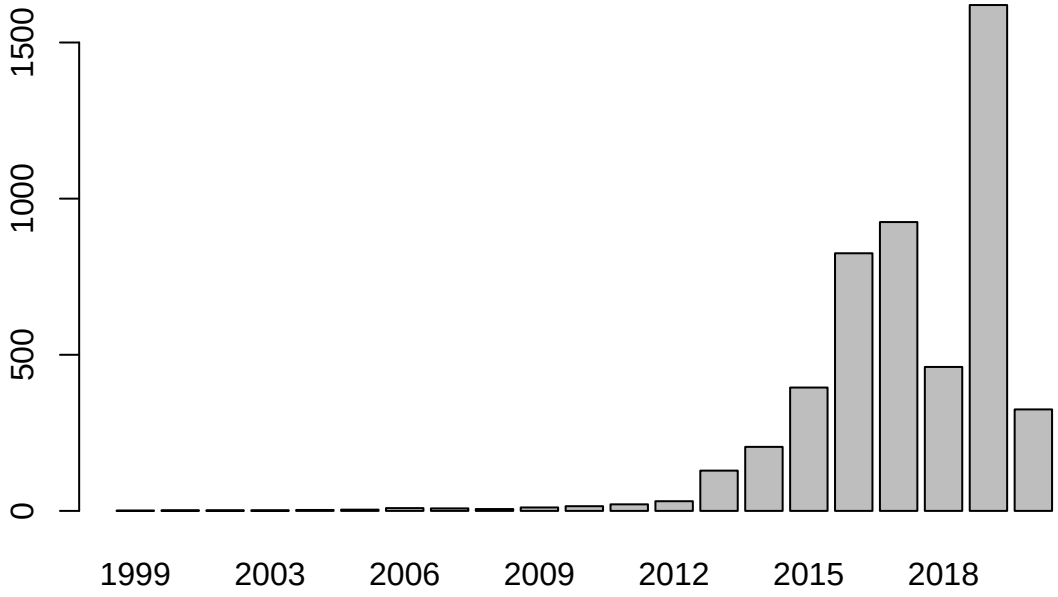


2.1.2.2 Year Registration year

```
df$year<-factor(df$year)
levels(df$year)
```

```
##  [1] "1999" "2001" "2002" "2003" "2004" "2005" "2006" "2007" "2008" "2009"
## [11] "2010" "2011" "2012" "2013" "2014" "2015" "2016" "2017" "2018" "2019"
## [21] "2020"
```

```
barplot(summary(df$year))
```




```
summary(df$year)
```

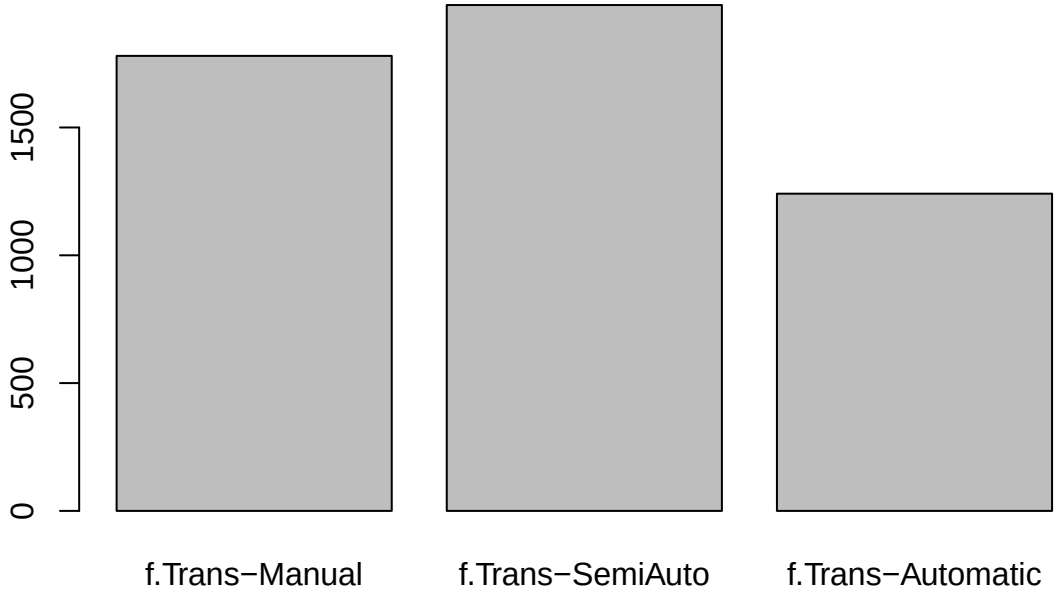
```
## 1999 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
##    1    2    2    2    3    4    9    8    6   11   15   21   31  129  205  395
## 2016 2017 2018 2019 2020
##  825  925  461 1620  325
```

2.1.2.3 Transmission Type of gearbox

```
df$transmission <- factor( df$transmission,
                           levels = c("Manual","Semi-Auto","Automatic"),labels =
                             paste0("f.Trans-",c("Manual","SemiAuto","Automatic")))
levels(df$transmission)
```

```
## [1] "f.Trans-Manual"    "f.Trans-SemiAuto"  "f.Trans-Automatic"
```

```
barplot(summary(df$transmission))
```

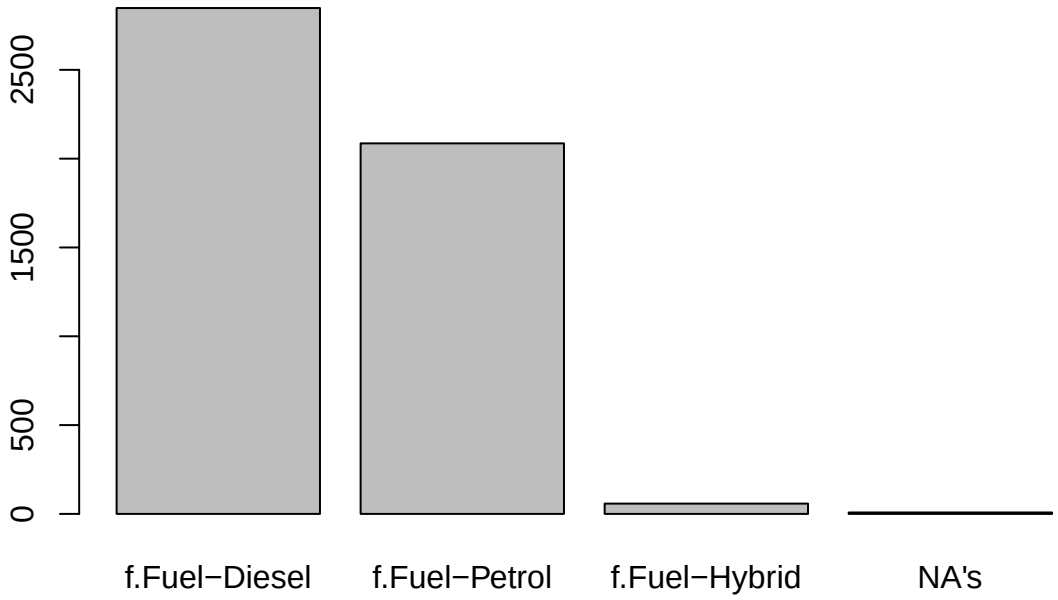


2.1.2.4 **FuelType** Engine fuel

```
df$fuelType <- factor( df$fuelType, levels = c("Diesel","Petrol","Hybrid"),
                       labels = paste0("f.Fuel-",c("Diesel","Petrol","Hybrid")))
levels(df$fuelType)
```

```
## [1] "f.Fuel-Diesel" "f.Fuel-Petrol" "f.Fuel-Hybrid"
```

```
barplot(summary(df$fuelType)) # es pot veure que te NA's, després els contem
```

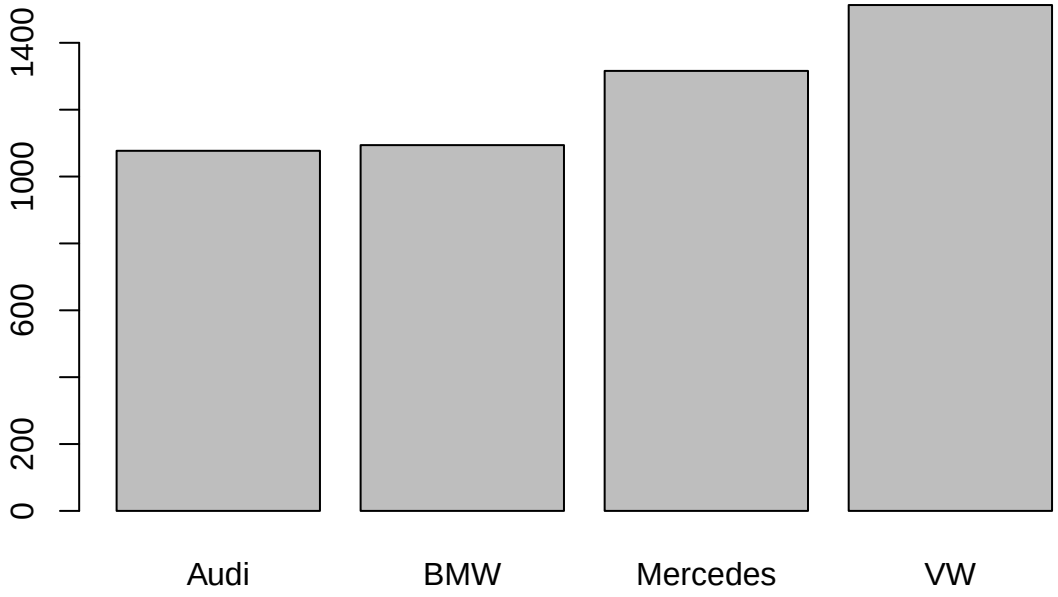


2.1.2.5 Manufacturer Audi, BMW, Mercedes or Volkswagen

```
df$manufacturer <- factor(df$manufacturer)
levels(df$manufacturer)
```

```
## [1] "Audi"     "BMW"      "Mercedes" "VW"
```

```
barplot(summary(df$manufacturer))
```




```
summary(df[c("model", "year", "transmission", "fuelType", "manufacturer")])
```

```
##                model           year                 transmission
##  VW- Golf          : 493   2019   :1620   f.Trans-Manual   :1780
##  Mercedes- C Class: 373   2017   : 925   f.Trans-SemiAuto :1979
##  VW- Polo          : 337   2016   : 825   f.Trans-Automatic:1241
##  Mercedes- A Class: 275   2018   : 461
##  BMW- 3 Series     : 249   2015   : 395
##  BMW- 1 Series     : 201   2020   : 325
##  (Other)           :3072   (Other): 449
##           fuelType      manufacturer
##  f.Fuel-Diesel:2848   Audi    :1077
##  f.Fuel-Petrol:2086   BMW     :1094
##  f.Fuel-Hybrid:  58   Mercedes:1316
##  NA's          :   8   VW      :1513
##
##
##
```

2.1.3 Variables quantitatives (check outliers/errors)

Busquem errors i outliers en les variables numeriques i els convertim en missings per després imputar-los. Considerem errors depenent de cada variable, però en general valors negatius o 0. En els boxplots queden marcats els outliers extrems. També es poden veure a les variables length(sel) (nombre d'errors) i length(llout) (nombre de outliers extrems).

2.1.3.1 **Price** Price in £

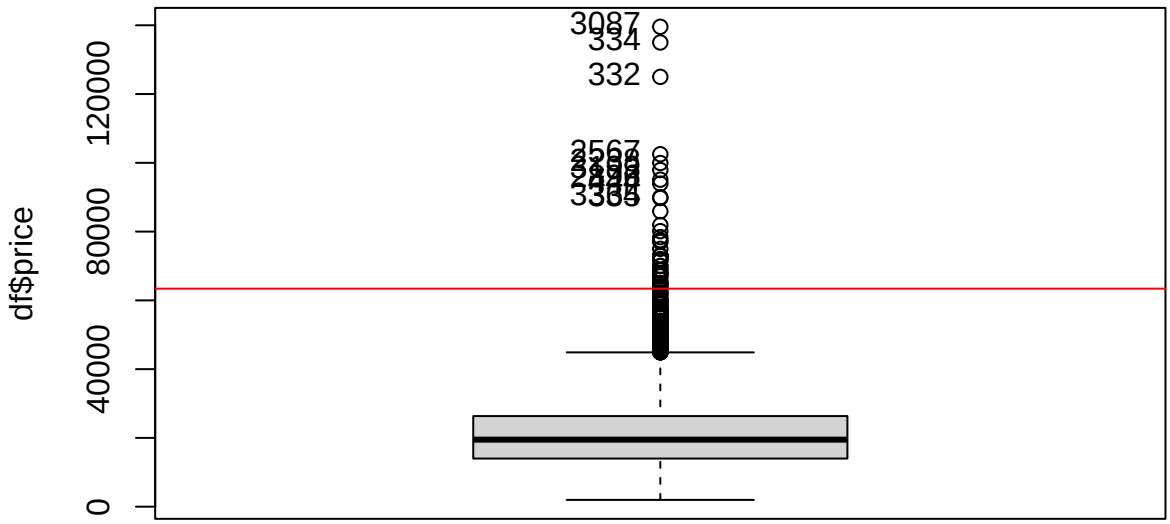
```
summary(df$price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1975   13998   19498   21552   26350  139559
```

```
#error detection

sel<-which(df$price <=0)
ierrs[sel]<-ierrs[sel]+1
jerrs[3]<-length(sel)
df[sel,"price"]<-NA

#outlier detection
Boxplot(df$price)
```

```
##  [1] 3087  334  332 2567 2238 3193 2896  444 3364  335
```

```
var_out<-calcQ(df$price)
abline(h=var_out$souts,col="red")
abline(h=var_out$souti,col="red")
```




```
# Outliers:
llout<-which((df$price>var_out$souts)|(df$price<var_out$souti))
iouts[llout]<-iouts[llout]+1
jouts[3]<-length(llout)
df[llout,"price"]<-NA
```

2.1.3.2 Mileage Distance used

```
summary(df$mileage)
```

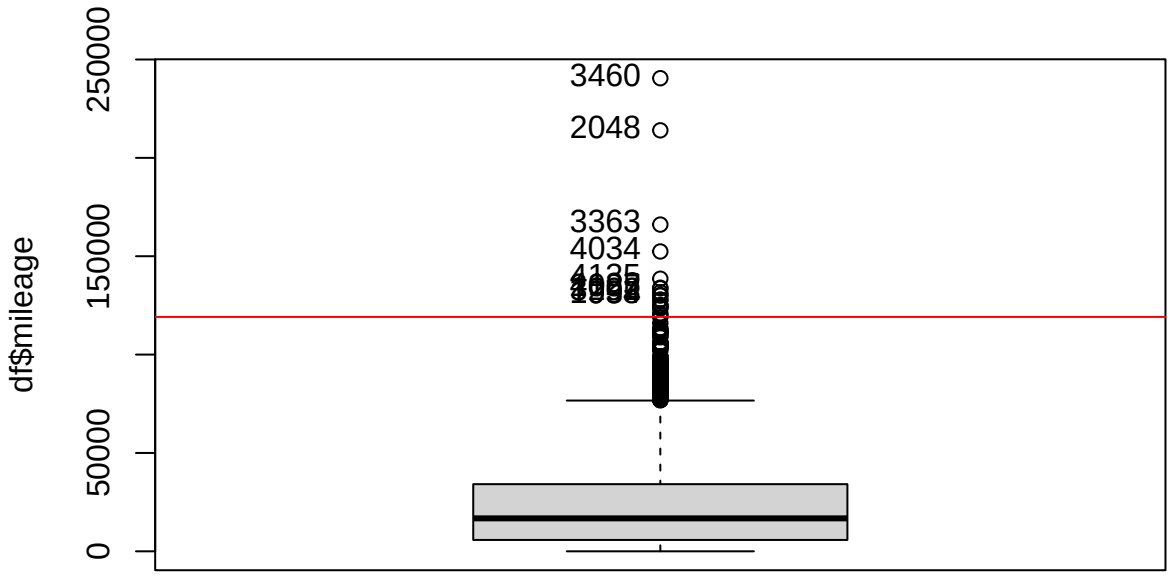
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       1    5785   16741   23457   34125  240494
```

```
#error detection
sel<-which(df$mileage <0)
ierrs[sel]<-ierrs[sel]+1
```

```
jerrs[5]<-length(sel)
df[sel,"mileage"]<-NA

#outlier detection
Boxplot(df$mileage)
```

```
##  [1] 3460 2048 3363 4034 4135 4125 2087 3292 1994 1935
```

```
var_out<-calcQ(df$mileage)
abline(h=var_out$souts,col="red")
abline(h=var_out$souti,col="red")
```




```
# Outliers:
llout<-which((df$mileage>var_out$souts)|(df$mileage<var_out$souti))
iouts[llout]<-iouts[llout]+1
jouts[5]<-length(llout)
df[llout,"mileage"]<-NA
```

2.1.3.3 Tax Road tax

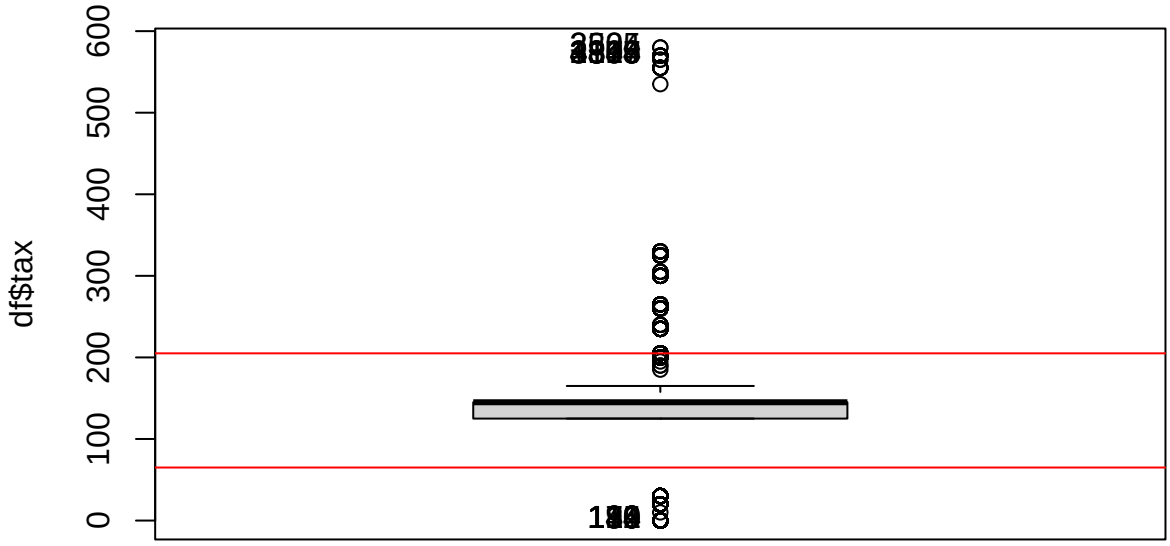
```
summary(df$tax)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   125.0   145.0   125.5   145.0   580.0
```

```
#error detection
sel<-which(df$tax <0)
ierrs[sel]<-ierrs[sel]+1
jerrs[7]<-length(sel)
df[sel,"tax"]<-NA

#outlier detection
Boxplot(df$tax)
```

```
##  [1]   10   12   82   94  111  115  134  140  179  180 2504 3325  509 1305 2179
## [16] 3136 3314 3317 4844 4878
```

```
var_out<-calcQ(df$tax)
abline(h=var_out$souts,col="red")
abline(h=var_out$souti,col="red")
```




```
# Outliers:
llout<-which((df$tax>var_out$souts)|(df$tax<var_out$souti))
iouts[llout]<-iouts[llout]+1
jouts[7]<-length(llout)
df[llout,"tax"]<-NA
```

2.1.3.4 Mpg Consumption in miles per gallon

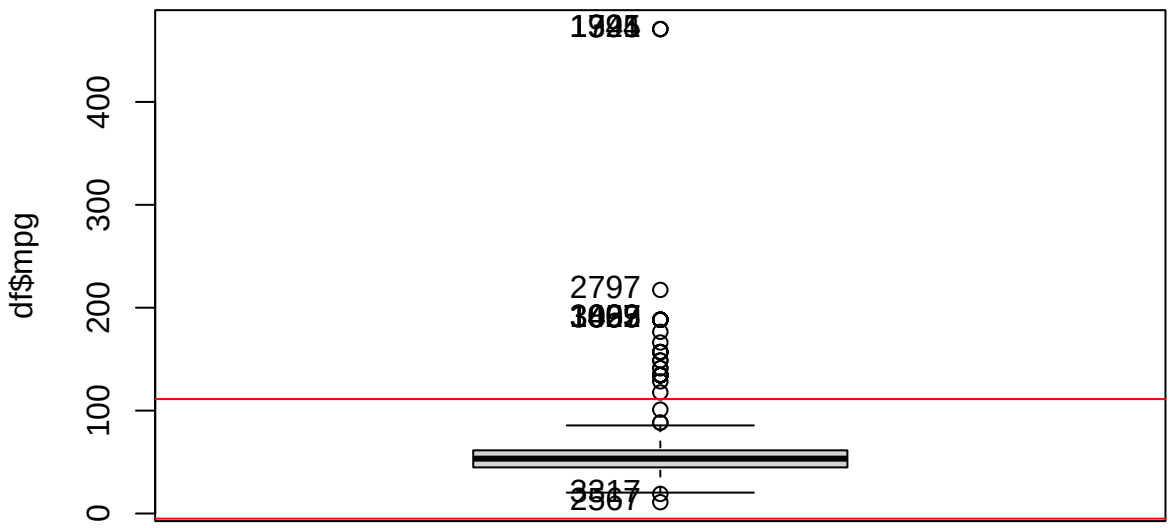
```
summary(df$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.00   44.80   53.30   53.81   61.40  470.80
```

```
#error detection
sel<-which(df$mpg <=0)
ierrs[sel]<-ierrs[sel]+1
jerrs[8]<-length(sel)
df[sel,"mpg"]<-NA

#outlier detection
Boxplot(df$mpg)
```

```
##  [1] 2567 3317 1324 1791 1945 2797  495 1053 1429 1802 3037 3067
```

```
var_out<-calcQ(df$mpg)
abline(h=var_out$souts,col="red")
abline(h=var_out$souti,col="red")
```




```
# Outliers:
llout<-which((df$mpg>var_out$souts)|(df$mpg<var_out$souti))
iouts[llout]<-iouts[llout]+1
jouts[8]<-length(llout)
df[llout,"mpg"]<-NA
```

2.1.3.5 EngineSize Size in litres

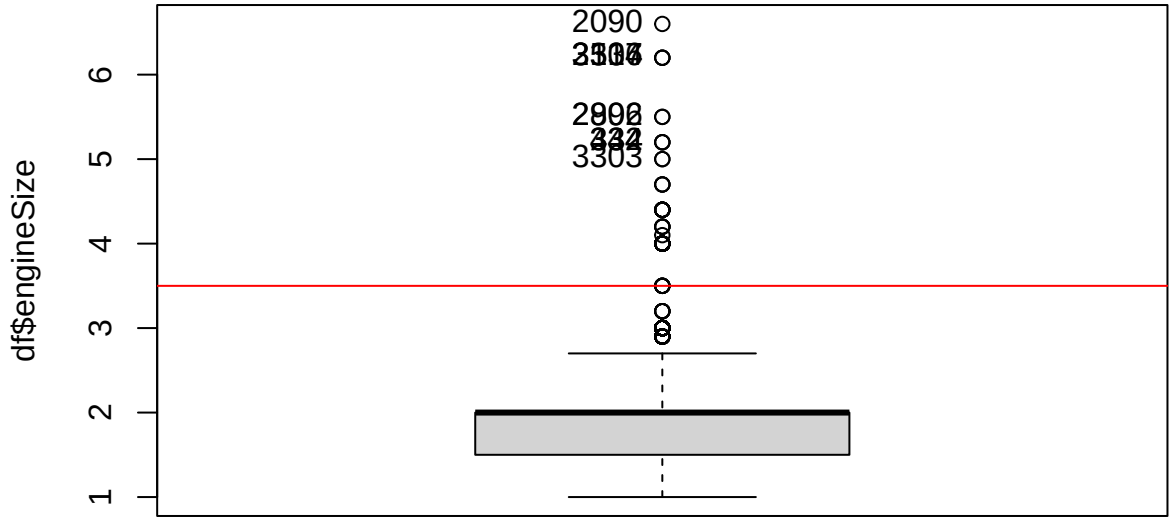
```
summary(df$engineSize)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.500   2.000   1.927   2.000   6.600
```

```
#error detection
sel<-which(df$engineSize <=0)
ierrs[sel]<-ierrs[sel]+1
jerrs[9]<-length(sel)
df[sel,"engineSize"]<-NA

#outlier detection
Boxplot(df$engineSize)
```

```
##  [1] 2090 2504 3136 3317 2802 2996  332  334  444 3303
```

```
var_out<-calcQ(df$engineSize)
abline(h=var_out$souts,col="red")
abline(h=var_out$souti,col="red")
```




```
# Outliers:
llout<-which((df$engineSize>var_out$souts)|(df$engineSize<var_out$souti))
iouts[llout]<-iouts[llout]+1
jouts[9]<-length(llout)
df[llout,"engineSize"]<-NA

df$engineSize <- factor(cut(df$engineSize,breaks=c(0, 1.5, 2, 6),include.lowest = T ), labels = c("Small
summary(df$engineSize)
```

```
##  Small Medium    Big   NA's
##   1400   2526    995     79
```

2.1.3.6 Age Car age

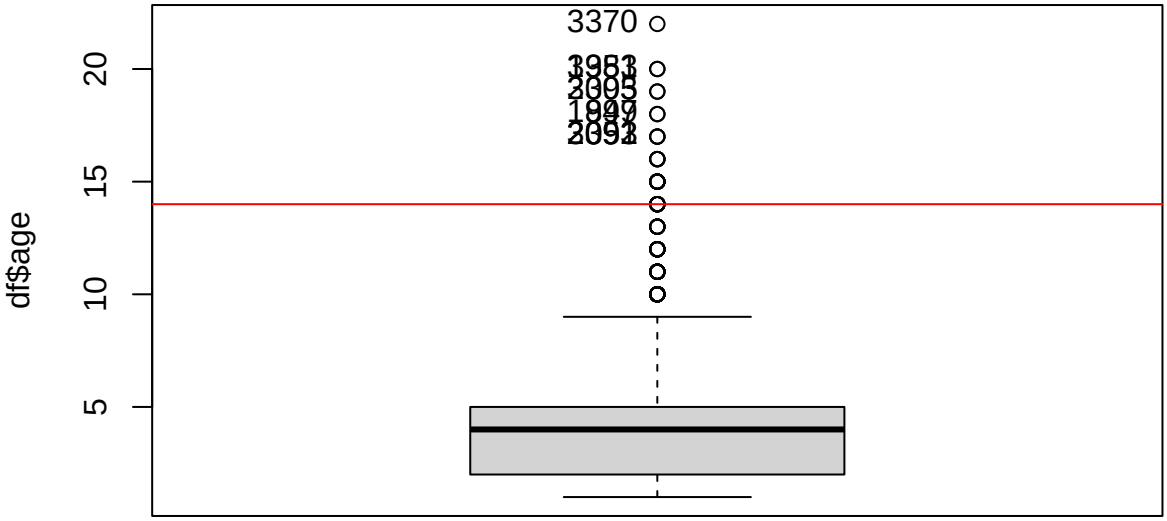
```
summary(df$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   4.000   3.794   5.000  22.000
```

```
#error detection
sel<-which(df$age <0)
ierrs[sel]<-ierrs[sel]+1
jerrs[11]<-length(sel)
df[sel,"age"]<-NA

#outlier detection
Boxplot(df$age)
```

```
##  [1] 3370 1981 3353 2095 3303  997 1849 2092 2093 3351
```

```
var_out<-calcQ(df$age)
abline(h=var_out$souts,col="red")
abline(h=var_out$souti,col="red")
```




```
# Outliers:
llout<-which((df$age>var_out$souts)|(df$age<var_out$souti))
iouts[llout]<-iouts[llout]+1
jouts[11]<-length(llout)
df[llout,"age"]<-NA
```

2.2 Data Quality Report

Per variable, count: Number of missing values, Number of errors (including inconsistencies), Number of outliers, Rank variables according the sum of missing values (and errors).

```
#missings
mis1<-countNA(df)
imis<-mis1$mis_ind
missings <- sort.list(mis1$mis_col$mis_x, decreasing = TRUE)
for (i in missings){
  print(paste(names(df)[i], " : ", mis1$mis_col$mis_x[i]))
}
```

```
## [1] "tax  :  1275"
## [1] "engineSize  :  79"
## [1] "mpg  :  42"
## [1] "price  :  40"
## [1] "age  :  23"
## [1] "mileage  :  20"
## [1] "fuelType  :  8"
## [1] "model  :  0"
## [1] "year  :  0"
## [1] "transmission  :  0"
## [1] "manufacturer  :  0"
```

```
#errors
errors <- sort.list(jerrs, decreasing = TRUE)
for (i in errors){
```

```
  if(!is.na(names(df)[i])) {print(paste(names(df)[i], " : ", jerrs[i]))}
}
```

```
## [1] "engineSize  :  12"
## [1] "model  :  0"
## [1] "year  :  0"
## [1] "price  :  0"
## [1] "transmission  :  0"
## [1] "mileage  :  0"
## [1] "fuelType  :  0"
## [1] "tax  :  0"
## [1] "mpg  :  0"
## [1] "manufacturer  :  0"
## [1] "age  :  0"
```

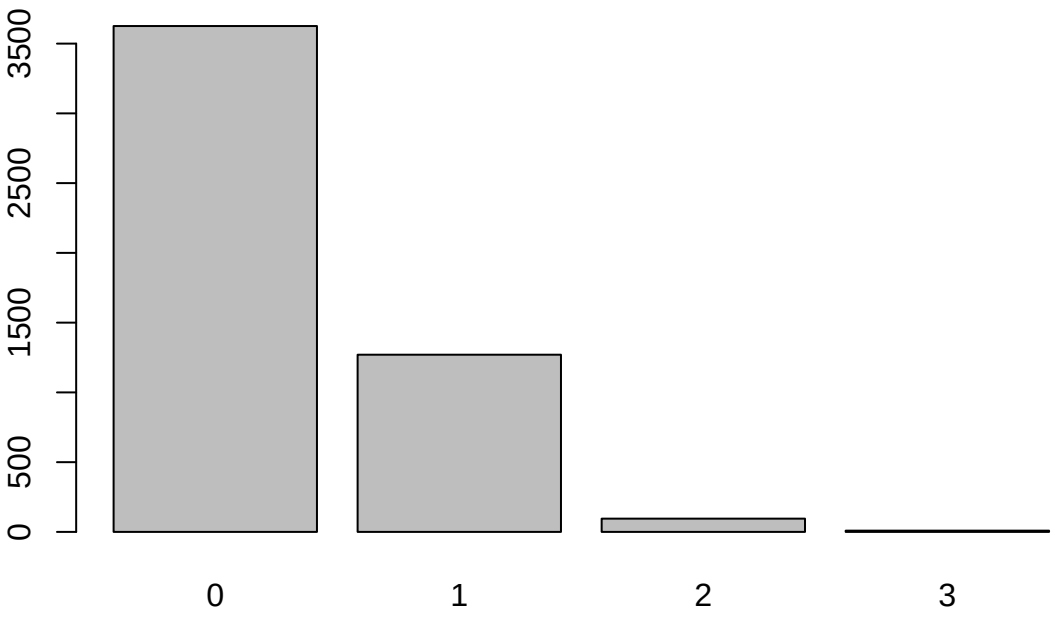
```
#outliers
outliers <- sort.list(jouts, decreasing = TRUE)
for (i in outliers){
  if(!is.na(names(df)[i])) {print(paste(names(df)[i], " : ", jouts[i]))}
}
```

```
## [1] "tax  :  1275"
## [1] "engineSize  :  67"
## [1] "mpg  :  42"
## [1] "price  :  40"
## [1] "age  :  23"
## [1] "mileage  :  20"
## [1] "model  :  0"
## [1] "year  :  0"
## [1] "transmission  :  0"
## [1] "fuelType  :  0"
## [1] "manufacturer  :  0"
```

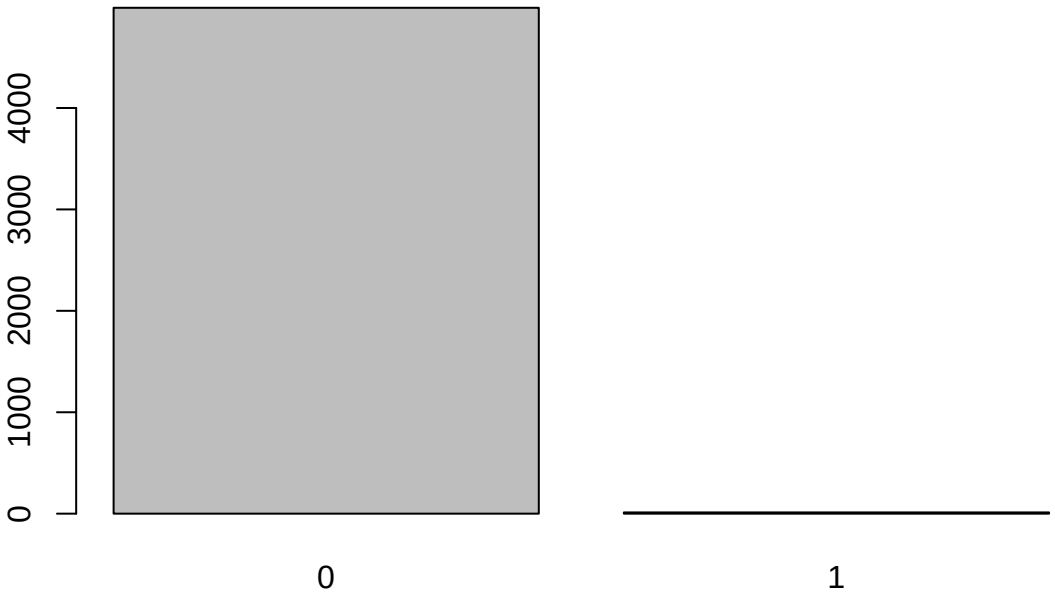
Clarifiquem que el nombre de missings es conta al moment, per tant també es consideren missings els valor imputats com a NA a la secció anterior. El nombre real de missings abans del tractament seria nº de missings - nº d'outliers - nº d'errors.

Per individuals, count: number of missing values, number of errors, number of outliers

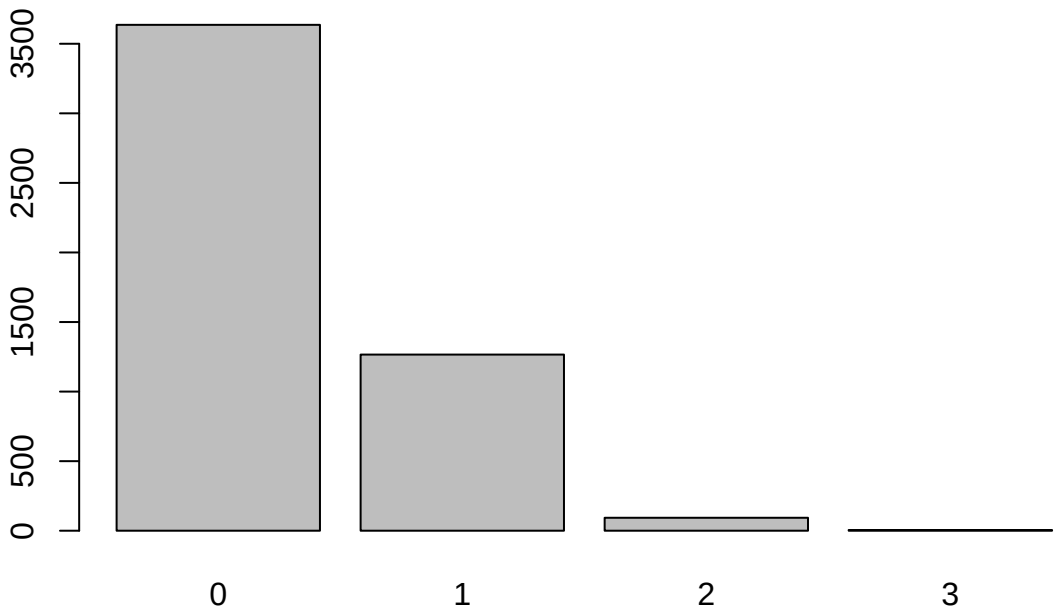
```
barplot(table(imis))
```




```
barplot(table(ierrs))
```



```
barplot(table(iouts))
```



Create variable adding the total number missing values, outliers and errors.

```
total_m<-0
total_e<-0
total_o<-0
for (i in imis) {
  total_m<- total_m + i
}
for (i in ierrs) {
  total_e<- total_e + i
}
for (i in iouts) {
  total_o<- total_o + i
}
total_m
```

```
## [1] 1487
```

```
total_e
```

```
## [1] 12
```

```
total_o
```

```
## [1] 1467
```

```
missing_before <- total_m - total_o - total_e
missing_before #els vists a fuelType
```

```
## [1] 8
```

3 Data Imputation

Per la variable target, borrem les observacions amb NA's. Per la resta de variables fem imputació numèrica o quantitativa

3.1 Target variable

```
sel <- which( is.na( df$price ) )
df <- df[ -sel, ]
```

3.2 Explicative numeric variables

```
library(missMDA)
# Now one by one describe vars and put them on lists
names(df)
```

```
##  [1] "model"        "year"         "price"        "transmission" "mileage"
##  [6] "fuelType"     "tax"          "mpg"          "engineSize"   "manufacturer"
## [11] "age"
```

```
vars_con<-names(df)[c(3,5,7:8, 11)]
vars_dis<-names(df)[c(1:2, 4, 6, 9:10)]
vars_res<-names(df)[c(3)]

summary(df[,vars_con])
```

```
##      price          mileage            tax             mpg
##  Min.   : 1975   Min.   :     1   Min.   :125.0   Min.   : 19.00
##  1st Qu.:13995   1st Qu.:  5905   1st Qu.:145.0   1st Qu.: 44.80
##  Median :19490   Median : 16869   Median :145.0   Median : 53.30
##  Mean   :21080   Mean   : 23138   Mean   :147.1   Mean   : 52.95
##  3rd Qu.:25997   3rd Qu.: 34026   3rd Qu.:145.0   3rd Qu.: 61.40
##  Max.   :62980   Max.   :119000   Max.   :205.0   Max.   :100.90
##                  NA's   :20       NA's   :1274    NA's   :40
##       age
##  Min.   : 1.000
##  1st Qu.: 2.000
##  Median : 4.000
##  Mean   : 3.749
##  3rd Qu.: 5.000
##  Max.   :14.000
##  NA's   :23
```

```
res.impca<-imputePCA(df[,vars_con],ncp=4)
summary(res.impca$completeObs)
```

```
##      price          mileage            tax             mpg
##  Min.   : 1975   Min.   :     1   Min.   :125.0   Min.   : 19.00
##  1st Qu.:13995   1st Qu.:  5936   1st Qu.:145.0   1st Qu.: 45.60
##  Median :19490   Median : 16994   Median :145.0   Median : 53.30
##  Mean   :21080   Mean   : 23275   Mean   :147.3   Mean   : 52.95
##  3rd Qu.:25997   3rd Qu.: 34228   3rd Qu.:148.2   3rd Qu.: 61.40
##  Max.   :62980   Max.   :119000   Max.   :205.0   Max.   :100.90
##       age
##  Min.   : 1.000
##  1st Qu.: 2.000
##  Median : 4.000
```

```
##  Mean   : 3.769
##  3rd Qu.: 5.000
##  Max.   :14.000
```

```
df[ , vars_con ]<-res.impca$completeObs
```

3.3 Explicative categorical variables

```
summary(df[,vars_dis])
```

```
##                model           year                  transmission
##  VW- Golf          : 493   2019   :1595   f.Trans-Manual   :1780
##  Mercedes- C Class: 373   2017   : 925   f.Trans-SemiAuto :1951
##  VW- Polo          : 337   2016   : 824   f.Trans-Automatic:1229
##  Mercedes- A Class: 268   2018   : 457
##  BMW- 3 Series     : 248   2015   : 395
##  BMW- 1 Series     : 201   2020   : 315
##  (Other)           :3040   (Other): 449
##           fuelType     engineSize     manufacturer
##  f.Fuel-Diesel:2830   Small :1400   Audi    :1067
##  f.Fuel-Petrol:2066   Medium:2526   BMW     :1082
##  f.Fuel-Hybrid:  56   Big   : 978   Mercedes:1298
##  NA's          :   8   NA's  :  56   VW      :1513
##
##
##
```

```
res.immca<-imputeMCA(df[,vars_dis],ncp=10)
summary(res.immca$completeObs)
```

```
##                model           year                  transmission
##  VW- Golf          : 493   2019   :1595   f.Trans-Manual   :1780
##  Mercedes- C Class: 373   2017   : 925   f.Trans-SemiAuto :1951
##  VW- Polo          : 337   2016   : 824   f.Trans-Automatic:1229
##  Mercedes- A Class: 268   2018   : 457
##  BMW- 3 Series     : 248   2015   : 395
##  BMW- 1 Series     : 201   2020   : 315
##  (Other)           :3040   (Other): 449
##           fuelType     engineSize     manufacturer
##  f.Fuel-Diesel:2836   Small :1409   Audi    :1067
##  f.Fuel-Petrol:2068   Medium:2553   BMW     :1082
##  f.Fuel-Hybrid:  56   Big   : 998   Mercedes:1298
##                                     VW      :1513
##
##
##
```

```
df[ , vars_dis ]<-res.immca$completeObs
```

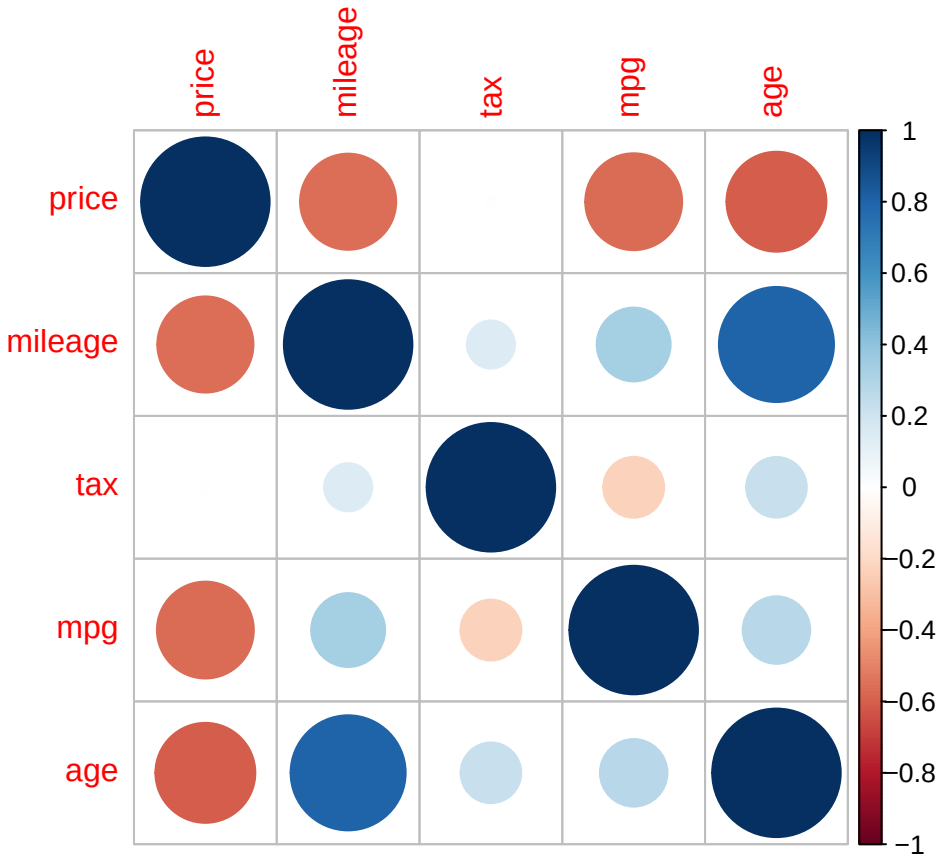
Describe these variables, to which other variables exist higher associations.

- Compute the correlation with all other variables. Rank these variables according the correlation

```
library(mvoutlier)
library(FactoMineR)
res <- cor(df[,vars_con])
round(res, 2)
```

```
##         price mileage   tax   mpg   age
## price    1.00   -0.56 -0.01 -0.57 -0.61
## mileage -0.56    1.00  0.14  0.33  0.80
## tax     -0.01    0.14  1.00 -0.22  0.22
## mpg     -0.57    0.33 -0.22  1.00  0.28
## age     -0.61    0.80  0.22  0.28  1.00
```

```
library(corrplot)
corrplot(res)
```



3.4 Discretitzation

Discretitzem les variables numèriques convertint-les en factors segons els seus quartils. Per fer això, mirem els quartils de cada variable i busquem uns intervals on hi hagi un numero semblant de mostres.

3.4.1 Price

```
summary(df$price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1975   13995   19490   21080   25997   62980
```

```
quantile(df$price,seq(0,1,0.25),na.rm=TRUE)
```

```
##      0%     25%     50%     75%    100%
##  1975.0 13995.0 19490.0 25996.5 62980.0
```

```
df$f.price<-factor(cut(df$price/1000,breaks=c(0,15,20,26, 90),include.lowest = T ))
levels(df$f.price)<-paste("f.price-",levels(df$f.price),sep="")
table(df$f.price,useNA="always")
```

```
##
##  f.price-[0,15] f.price-(15,20] f.price-(20,26] f.price-(26,90]            <NA>
##            1468            1228            1035            1229               0
```

3.4.2 Mileage

```
summary(df$mileage)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       1    5936   16994   23275   34228  119000
```

```
quantile(df$mileage,seq(0,1,0.25),na.rm=TRUE)
```

```
##        0%       25%       50%       75%      100%
##      1.00   5936.25  16993.50  34228.25 119000.00
```

```
df$aux<-factor(cut(df$mileage,breaks=c(0,5750,17800,36000, 195000),include.lowest = T ))
summary(df$aux)
```

```
##        [0,5.75e+03] (5.75e+03,1.78e+04]  (1.78e+04,3.6e+04]  (3.6e+04,1.95e+05]
##                1220                1315                1277                1148
```

```
tapply(df$mileage,df$aux,median)
```

```
##        [0,5.75e+03] (5.75e+03,1.78e+04]  (1.78e+04,3.6e+04]  (3.6e+04,1.95e+05]
##              2512.5             10565.0             26023.0             50359.5
```

```
df$f.miles<-factor(cut(df$mileage/1000,breaks=c(0,6,18,36, 195),include.lowest = T ))
levels(df$f.miles)<-paste("f.miles-",levels(df$f.miles),sep="")
table(df$f.miles,useNA="always")
```

```
##
##    f.miles-[0,6]   f.miles-(6,18]  f.miles-(18,36] f.miles-(36,195]
##             1287             1261             1264             1148
##             <NA>
##                0
```

3.4.3 Tax

```
summary(df$tax)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   125.0   145.0   145.0   147.3   148.2   205.0
```

```
quantile(df$tax,seq(0,1,0.25),na.rm=TRUE)
```

```
##       0%      25%      50%      75%     100%
## 125.0000 145.0000 145.0000 148.1787 205.0000
```

```
df$aux<-factor(cut(df$tax,breaks=c(0, 125, 145, 570),include.lowest = T ))
summary(df$aux)
```

```
##   [0,125] (125,145] (145,570]
##       279      2943      1738
```

```
tapply(df$tax,df$aux,median)
```

```
##   [0,125] (125,145] (145,570]
##       125       145       150
```

```
df$f.tax<-factor(cut(df$tax,breaks=c(0, 125, 145, 570),include.lowest = T ))
levels(df$f.tax)<-paste("f.tax-",levels(df$f.tax),sep="")
table(df$f.tax,useNA="always")
```

```
##
##   f.tax-[0,125] f.tax-(125,145] f.tax-(145,570]            <NA>
##             279            2943            1738               0
```

3.4.4 Mpg

```
summary(df$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   19.00   45.60   53.30   52.95   61.40  100.90
```

```
quantile(df$mpg,seq(0,1,0.25),na.rm=TRUE)
```

```
##    0%   25%   50%   75%  100%
##  19.0  45.6  53.3  61.4 100.9
```

```
df$aux<-factor(cut(df$mpg,breaks=c(0, 45, 54, 62, 101),include.lowest = T ))
summary(df$aux)
```

```
##   [0,45]  (45,54]  (54,62] (62,101]
##     1239     1430     1163     1128
```

```
tapply(df$mpg,df$aux,median)
```

```
##   [0,45]  (45,54]  (54,62] (62,101]
##     39.2     49.6     58.9     67.3
```

```
df$f.mpg<-factor(cut(df$mpg,breaks=c(0, 45, 54, 62, 101),include.lowest = T ))
levels(df$f.mpg)<-paste("f.mpg-",levels(df$f.mpg),sep="")
table(df$f.mpg,useNA="always")
```

```
##
##   f.mpg-[0,45]  f.mpg-(45,54]  f.mpg-(54,62] f.mpg-(62,101]           <NA>
##           1239            1430            1163            1128               0
```

3.4.5 Age

```
summary(df$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   4.000   3.769   5.000  14.000
```

```
quantile(df$age,seq(0,1,0.25),na.rm=TRUE)
```

```
##   0%  25%  50%  75% 100%
##    1    2    4    5   14
```

```
df$aux<-factor(cut(df$age,breaks=c(0, 2, 4.1, 5.1, 15),include.lowest = T ))
summary(df$aux)
```

```
##     [0,2]   (2,4.1] (4.1,5.1]  (5.1,15]
##      1910      1383       825       842
```

```
tapply(df$age,df$aux,median)
```

```
##     [0,2]   (2,4.1] (4.1,5.1]  (5.1,15]
##         2         4         5         7
```

```
df$f.age<-factor(cut(df$age,breaks=c(0, 2, 4.1, 5.1, 15),include.lowest = T ))
levels(df$f.age)<-paste("f.age-",levels(df$f.age),sep="")
table(df$f.age,useNA="always")
```

```
##
##     f.age-[0,2]   f.age-(2,4.1] f.age-(4.1,5.1]  f.age-(5.1,15]            <NA>
##            1910            1383             825             842               0
```

```
summary(df)
```

```
##                 model           year          price
##  VW- Golf          : 493   2019   :1595   Min.   : 1975
##  Mercedes- C Class: 373   2017   : 925   1st Qu.:13995
##  VW- Polo          : 337   2016   : 824   Median :19490
##  Mercedes- A Class: 268   2018   : 457   Mean   :21080
##  BMW- 3 Series     : 248   2015   : 395   3rd Qu.:25997
##  BMW- 1 Series     : 201   2020   : 315   Max.   :62980
##  (Other)           :3040   (Other): 449
##            transmission    mileage                fuelType         tax
##  f.Trans-Manual   :1780   Min.   :     1   f.Fuel-Diesel:2836   Min.   :125.0
##  f.Trans-SemiAuto :1951   1st Qu.:  5936   f.Fuel-Petrol:2068   1st Qu.:145.0
##  f.Trans-Automatic:1229   Median : 16994   f.Fuel-Hybrid:  56   Median :145.0
##                           Mean   : 23275                        Mean   :147.3
##                           3rd Qu.: 34228                        3rd Qu.:148.2
##                           Max.   :119000                        Max.   :205.0
##
##       mpg          engineSize     manufacturer       age
##  Min.   : 19.00   Small :1409   Audi    :1067   Min.   : 1.000
##  1st Qu.: 45.60   Medium:2553   BMW     :1082   1st Qu.: 2.000
##  Median : 53.30   Big   : 998   Mercedes:1298   Median : 4.000
##  Mean   : 52.95                 VW      :1513   Mean   : 3.769
##  3rd Qu.: 61.40                                 3rd Qu.: 5.000
##  Max.   :100.90                                 Max.   :14.000
##
##            f.price            aux                  f.miles
##  f.price-[0,15] :1468   [0,2]    :1910   f.miles-[0,6]   :1287
##  f.price-(15,20]:1228   (2,4.1]  :1383   f.miles-(6,18]  :1261
##  f.price-(20,26]:1035   (4.1,5.1]: 825   f.miles-(18,36] :1264
##  f.price-(26,90]:1229   (5.1,15] : 842   f.miles-(36,195]:1148
##
##
##
##              f.tax                f.mpg                  f.age
```

```
##  f.tax-[0,125]   : 279   f.mpg-[0,45]  :1239   f.age-[0,2]    :1910
##  f.tax-(125,145]:2943   f.mpg-(45,54] :1430   f.age-(2,4.1]  :1383
##  f.tax-(145,570]:1738   f.mpg-(54,62] :1163   f.age-(4.1,5.1]: 825
##                         f.mpg-(62,101]:1128   f.age-(5.1,15] : 842
##
##
##
```

4 Feature Selection for Numeric Target and Binary Target

4.1 Numeric Profiling

Fem profiling de la variable target numérica (price) per trobar la relació d'aquesta amb les altres variables

```
library(FactoMineR)
summary(df$price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1975   13995   19490   21080   25997   62980
```

```
# The "variable to describe cannot have NA
res.condes<-condes(df[,c(vars_res,vars_con,vars_dis)],3)

res.condes$quanti  # Global association to numeric variables
```

```
##         correlation       p.value
## age       0.8016308  0.000000e+00
## mpg       0.3304369 1.130881e-126
## tax       0.1405900  2.564861e-23
## price.1  -0.5597160  0.000000e+00
## price    -0.5597160  0.000000e+00
```

```
res.condes$quali # Global association to factors
```

```
##                       R2      p.value
## year          0.654968100 0.000000e+00
## model         0.096690680 3.220007e-59
## fuelType      0.048648606 2.080017e-54
## transmission 0.048508064 2.999643e-54
## engineSize    0.019672549 4.106100e-22
## manufacturer 0.005974259 1.595680e-06
```

```
res.condes$category  # Partial association to significative levels in factors
```

```
##                                  Estimate      p.value
## year=2013                      2130.1101 2.984028e-70
## fuelType=f.Fuel-Diesel         3049.0855 6.735687e-53
## transmission=f.Trans-Manual    4875.6871 8.020897e-37
## year=2012                     13424.3199 3.077209e-30
## year=2011                     20894.3223 5.335843e-28
## year=2010                     22633.3156 1.152127e-21
## year=2009                     28633.8664 9.257206e-20
## year=2006                     32252.4969 2.694915e-18
## year=2008                     35403.6080 8.155160e-14
## year=2007                     19990.8241 3.626864e-11
## engineSize=Big                 3791.2230 2.381582e-10
## model=VW- CC                  33889.4607 1.570778e-07
## model=Audi- A6                10132.2628 6.077243e-07
## year=2005                     22986.9413 7.487475e-07
```

```
## model=VW- Passat                 8306.4616 7.655008e-06
## year=2002                       33894.9413 2.624900e-05
## model=BMW- 3 Series              2769.8591 1.976121e-04
## model=Audi- A3                   3178.4035 3.356072e-04
## model=Mercedes- CLK             72464.4607 6.215127e-04
## manufacturer=Audi                1783.6216 8.954549e-04
## model=Audi- S5                  68284.4607 1.236100e-03
## engineSize=Medium                 855.4346 1.242162e-03
## model=VW- Jetta                 32416.2107 1.489320e-03
## model=VW- Eos                   67104.4607 1.491533e-03
## model=Mercedes- SLK             19986.9051 2.223914e-03
## manufacturer=BMW                 1467.6842 4.914435e-03
## transmission=f.Trans-Automatic   1173.3074 5.136314e-03
## model=Mercedes- CL Class         6147.0761 5.266096e-03
## model=VW- Golf                    309.1178 5.917118e-03
## model=Mercedes- M Class         17609.5377 6.305173e-03
## model=Mercedes- GL Class        16811.1274 8.750760e-03
## model=BMW- 1 Series              1681.5154 9.003003e-03
## model=BMW- 5 Series              3261.0053 1.027223e-02
## model=VW- Scirocco               6964.3573 2.256754e-02
## year=2001                        2553.4413 2.987534e-02
## year=2003                        2236.9413 3.145751e-02
## model=Audi- TT                   4605.4824 3.223203e-02
## model=Audi- RS4                 23540.1274 4.074341e-02
## model=Audi- SQ5                 28470.4607 4.662521e-02
## model=Audi- A4                   1237.1682 4.878500e-02
## model=Audi- Q8                 -21513.5393 4.865973e-02
## model=Mercedes- GLE Class      -10170.6726 4.668128e-02
## model=Mercedes- B Class         -8431.5778 4.058885e-02
## model=VW- Golf SV              -11529.8356 2.705093e-02
## model=Mercedes- GLB Class      -24416.5393 2.326959e-02
## model=Mercedes- X-CLASS        -19092.2893 7.527791e-03
## model=Mercedes- SL CLASS       -14563.9726 1.969203e-03
## model=Mercedes- GLC Class       -9001.7993 1.819621e-03
## model=BMW- 2 Series             -8430.0924 1.443793e-03
## model=Mercedes- V Class        -16487.8310 1.378544e-03
## model=VW- Arteon               -17221.4143 7.675468e-04
## manufacturer=Mercedes           -2403.0402 2.022722e-05
## model=BMW- X2                  -20035.2060 7.697543e-06
## year=2017                      -27638.7452 2.274745e-07
## model=Audi- Q2                 -14938.4393 2.644523e-08
## model=VW- T-Cross              -22779.7643 2.317464e-09
## model=VW- T-Roc                -19467.1393 1.089217e-12
## year=2018                      -38027.9121 4.245809e-13
## engineSize=Small                -4646.6576 2.557317e-20
## transmission=f.Trans-SemiAuto   -6048.9945 8.885520e-51
## fuelType=f.Fuel-Petrol          -6710.8284 1.062097e-55
## year=2020                      -52260.2904 1.247191e-73
## year=2014                       -4065.2928 5.440908e-75
## year=2016                      -18177.2991 1.693021e-78
## year=2015                      -10400.0899 3.321170e-88
## year=2019                      -48223.2794 0.000000e+00
```

S'utilitza per fer totes les combinacions possibles de variables numèriques i factorials. Tindrem les variables que tenen un pvalor a partir d'un llindar del pvalor acceptat. Només surten les variables que tenen una certa relació.

La variable age, dins de els quantitatives, és la que te més relació amb price. Això indica que com més nou es un cotxe més preu té. Dins les qualitatives, la que te més relació és year, que implica la mateixa explicació que age.

4.2 Factorial Profiling

Fem profiling pel factor Audi si/no

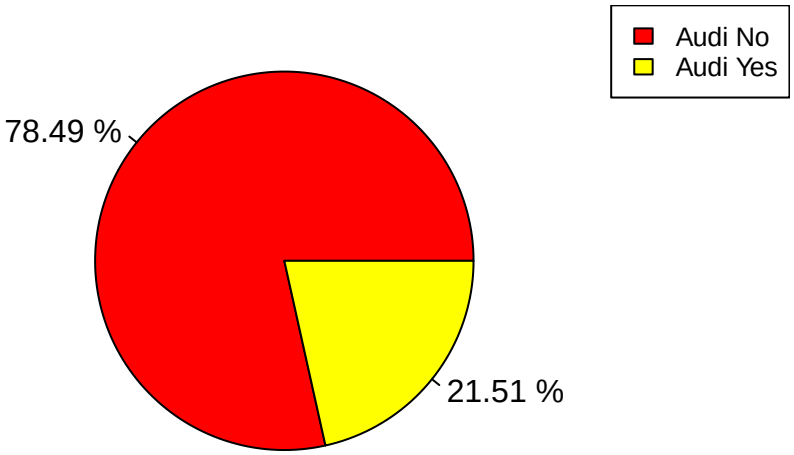
```r
# Binary Target: Audi?

df$Audi<-ifelse(df$manufacturer == "Audi",1,0)
df$Audi<-factor(df$Audi,labels=paste("Audi",c("No","Yes")))
summary(df$Audi)
```

```
##  Audi No Audi Yes
##     3893     1067
```

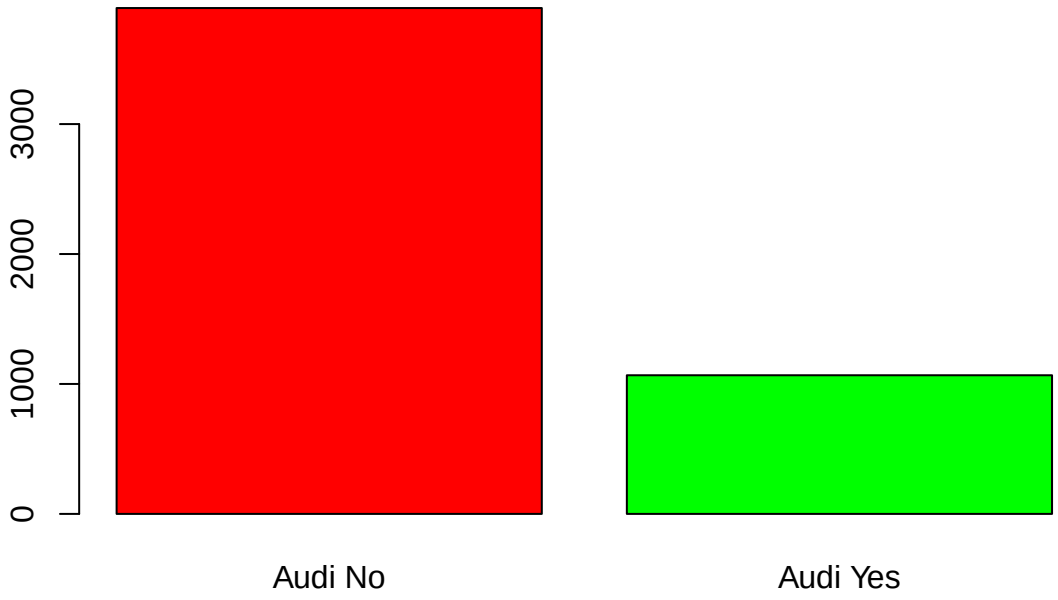
```r
# Pie
piepercent<-round(100*(table(df$Audi)/nrow(df)),dig=2); piepercent
```

```
##
##  Audi No Audi Yes
##    78.49    21.51
```

```r
pie(table(df$Audi),col=heat.colors(2),labels=paste(piepercent,"%"))
legend("topright", levels(df$Audi), cex = 0.8, fill = heat.colors(2))
```




```r
# Bar Chart
barplot(table(df$Audi),main="Barplot Binary Outcome - Factor",col=c("red","green"))
```




```
library(FactoMineR)
# The "variable to describe cannot have NA
res.catdes<-catdes(df,18)

res.catdes$quanti.var  # Global association to numeric variables
```

```
##                Eta2      P-value
## mpg     0.010872811 1.813506e-13
## tax     0.004710517 1.310622e-06
## price   0.003507012 3.001979e-05
## mileage 0.002222870 8.954549e-04
## age     0.001690095 3.781779e-03
```

```
res.catdes$quanti # Partial association of numeric variables to levels of outcome factor
```

```
## $`Audi No`
##            v.test Mean in category Overall mean sd in category   Overall sd
## mpg      7.342906        53.572357    52.950636      11.268286    11.388988
## age     -2.895027         3.725345     3.769041       2.011869     2.030232
## mileage -3.320122     22735.873207 23275.094475   21352.529529 21845.901298
## price   -4.170284     20776.133830 21080.082258    9789.536765  9803.729018
## tax     -4.833162       146.912383   147.330971      10.888491    11.649622
##              p.value
## mpg     2.090052e-13
## age     3.791265e-03
## mileage 8.997799e-04
## price   3.042201e-05
## tax     1.343814e-06
##
## $`Audi Yes`
##            v.test Mean in category Overall mean sd in category   Overall sd
## tax      4.833162       148.858207   147.330971      13.976080    11.649622
## price    4.170284     22189.052484 21080.082258    9775.519370  9803.729018
## mileage  3.320122     25242.468792 23275.094475   23453.636518 21845.901298
```

```
## age      2.895027         3.928467     3.769041       2.088126     2.030232
## mpg     -7.342906        50.682254    52.950636      11.538249    11.388988
##              p.value
## tax     1.343814e-06
## price   3.042201e-05
## mileage 8.997799e-04
## age     3.791265e-03
## mpg     2.090052e-13
```

```
#Multivariate Outliers
library(mvoutlier)

summary(df[,vars_con])
```

```
##      price          mileage            tax             mpg
##  Min.   : 1975   Min.   :     1   Min.   :125.0   Min.   : 19.00
##  1st Qu.:13995   1st Qu.:  5936   1st Qu.:145.0   1st Qu.: 45.60
##  Median :19490   Median : 16994   Median :145.0   Median : 53.30
##  Mean   :21080   Mean   : 23275   Mean   :147.3   Mean   : 52.95
##  3rd Qu.:25997   3rd Qu.: 34228   3rd Qu.:148.2   3rd Qu.: 61.40
##  Max.   :62980   Max.   :119000   Max.   :205.0   Max.   :100.90
##       age
##  Min.   : 1.000
##  1st Qu.: 2.000
##  Median : 4.000
##  Mean   : 3.769
##  3rd Qu.: 5.000
##  Max.   :14.000
```

```
names(df)
```

```
##  [1] "model"        "year"         "price"        "transmission" "mileage"
##  [6] "fuelType"     "tax"          "mpg"          "engineSize"   "manufacturer"
## [11] "age"          "f.price"      "aux"          "f.miles"      "f.tax"
## [16] "f.mpg"        "f.age"        "Audi"
```

```
vars_con
```

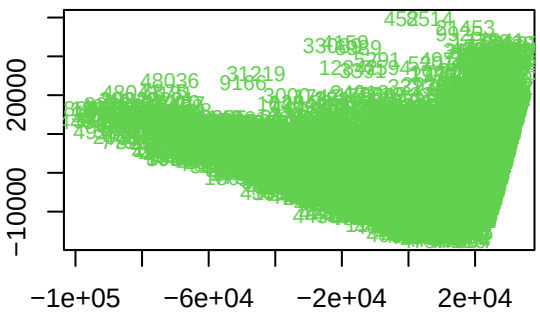
```
## [1] "price"   "mileage" "tax"     "mpg"     "age"
```

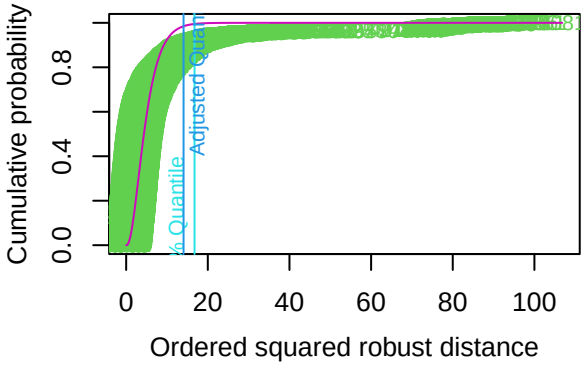
```
names(df)
```

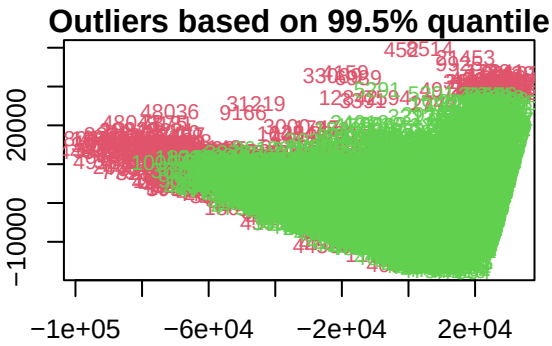
```
##  [1] "model"        "year"         "price"        "transmission" "mileage"
##  [6] "fuelType"     "tax"          "mpg"          "engineSize"   "manufacturer"
## [11] "age"          "f.price"      "aux"          "f.miles"      "f.tax"
## [16] "f.mpg"        "f.age"        "Audi"
```

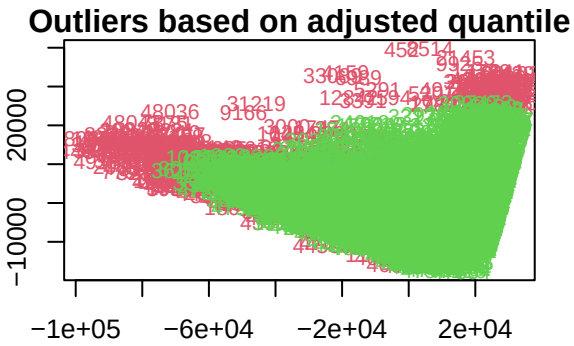
```
mout<-aq.plot(df[,c(3,5,7,8,11)],delta=qchisq(0.995,5),quan=0.995)
```

```
## Projection to the first and second robust principal components.
## Proportion of total variation (explained variance): 0.9994102
```





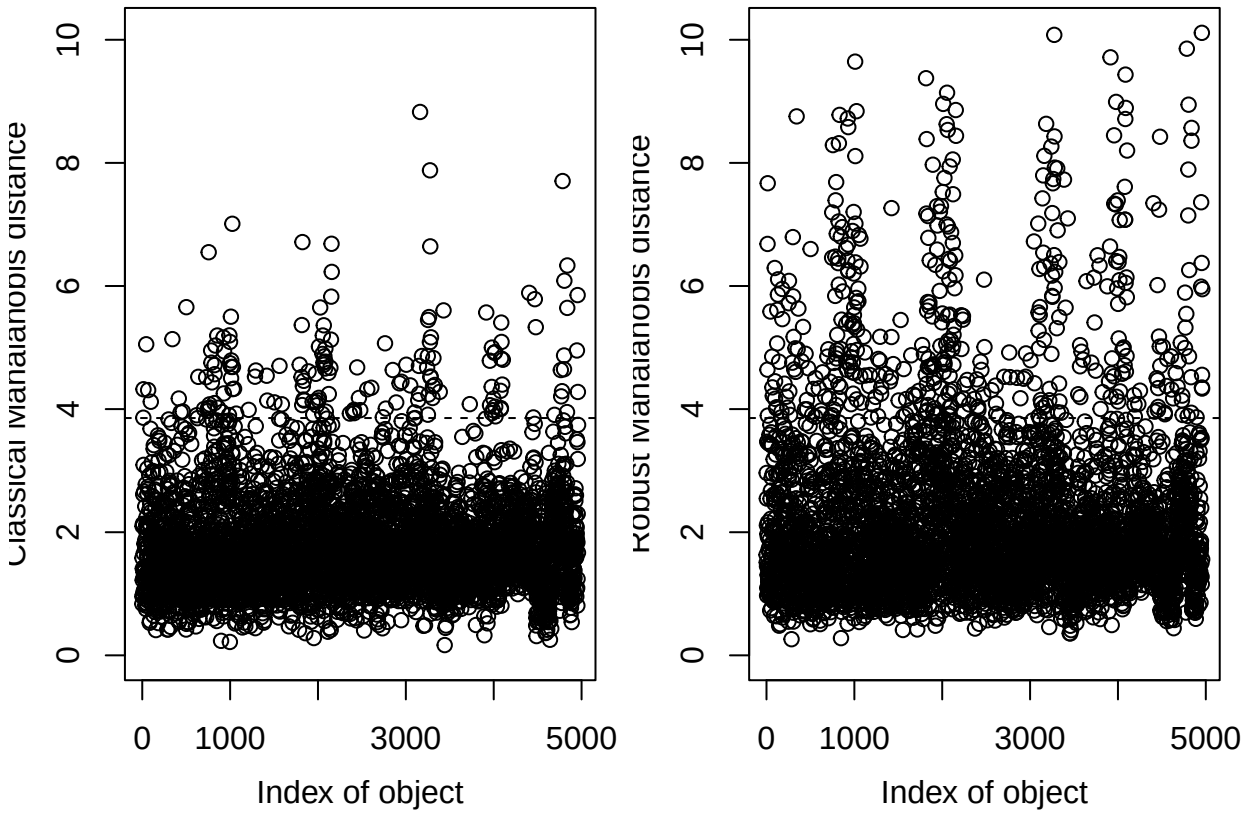





```
library(chemometrics)
summary(df[,vars_con])
```

```
##      price          mileage            tax             mpg
##  Min.   : 1975   Min.   :     1   Min.   :125.0   Min.   : 19.00
##  1st Qu.:13995   1st Qu.:  5936   1st Qu.:145.0   1st Qu.: 45.60
##  Median :19490   Median : 16994   Median :145.0   Median : 53.30
##  Mean   :21080   Mean   : 23275   Mean   :147.3   Mean   : 52.95
##  3rd Qu.:25997   3rd Qu.: 34228   3rd Qu.:148.2   3rd Qu.: 61.40
##  Max.   :62980   Max.   :119000   Max.   :205.0   Max.   :100.90
##       age
##  Min.   : 1.000
##  1st Qu.: 2.000
##  Median : 4.000
##  Mean   : 3.769
##  3rd Qu.: 5.000
##  Max.   :14.000
```

```
mout<-Moutlier(df[,c(3,5,8,11)],quantile = 0.995, plot = TRUE)
```




```
ll<-which(mout$rd>5)
Boxplot(mout$rd)
```

```
##  [1] 4955 3275 4784 3915 1009 4086 1816 2055 3979 2011
```

```
df[ll,c(vars_res,vars_con)]
```

```
##       price price.1   mileage      tax       mpg       age
## 152    8300    8300  98940.00 147.8410  74.30000  8.000000
## 158   16200   16200  89334.00 146.0412  62.80000  5.000000
## 452   57995   57995  34926.00 155.2474  29.40000  4.000000
## 1068  12498   12498  78269.00 142.4866  74.30000  5.000000
## 1321  12998   12998  68248.00 143.1516  72.40000  5.000000
## 1324  12995   12995  82500.00 145.0000  54.30000  7.000000
## 1401  13298   13298  68468.00 143.4178  65.70000  4.000000
## 1410  13498   13498  77061.00 144.6967  67.30000  5.000000
## 1822  11995   11995  68643.00 140.7108  74.30000  4.000000
## 1889  14995   14995  77352.00 147.5120  64.20000  6.000000
## 2592  10262   10262  74000.00 144.1992  68.90000  5.000000
## 2609  12691   12691  84000.00 205.0000  44.80000  7.000000
## 3067  13680   13680  81386.00 145.0000  52.30000  5.000000
## 3113  10500   10500  74924.00 144.6965  67.30000  5.000000
## 3146  13541   13541  79965.00 149.3248  64.20000  7.000000
## 3457  16499   16499  90778.00 150.0000  52.30000  4.000000
## 3616   6998    6998  89831.00 147.8949  74.30000  8.000000
## 4159  47000   47000  49106.00 145.0000  39.20000  4.000000
## 4977  51990   51990  19800.00 162.9012  22.80000  7.000000
## 7427   5495    5495 106000.00 156.4624  64.20000 11.000000
## 7441   8500    8500  96000.00 150.0000  53.30000  7.000000
## 7494   8488    8488  73000.00 172.6270  31.00000 14.000000
## 7770   9690    9690  80950.00 147.4629  64.20000  6.000000
## 7773   5995    5995  77000.00 167.2825  36.20000 12.000000
## 7796   1975    1975 104000.00 163.3429  38.10000 10.287472
## 7875  17775   17775  96753.00 156.3599  41.50000  7.000000
```

```
## 7977  11990   11990  83000.00 146.0629  62.80000  5.000000
## 8052   7990    7990  88000.00 153.3586  62.80000  9.000000
## 8137  12999   12999  77019.00 145.0000  47.90000  5.000000
## 8143  13063   13063  83855.00 160.0000  50.40000  6.000000
## 8181  17940   17940  95012.00 158.2092  41.50000  8.000000
## 8201  10599   10599  35577.00 163.1191  38.70000 10.000000
## 8232   9470    9470 111873.00 151.8107  61.40000  8.000000
## 8310  10199   10199  98791.00 143.5017  70.60000  5.000000
## 8454  11513   11513  82785.00 144.6495  67.30000  5.000000
## 8514  60750   60750  26885.00 150.0000  37.20000  3.000000
## 8575  14795   14795  77240.00 145.0000  67.30000  4.000000
## 8619  19950   19950  68454.00 150.0000  58.90000  4.000000
## 8970  13000   13000  81190.00 144.1648  68.90000  5.000000
## 8987  17999   17999  65675.00 164.3208  28.50000  9.000000
## 9166  27490   27490  74896.00 152.6569  42.20000  5.000000
## 9202   7295    7295 104460.00 141.3635  83.10000  6.000000
## 9221  16494   16494  94000.00 162.9187  38.20000 10.000000
## 9225  17000   17000  77700.00 125.0000  58.90000  5.000000
## 9270  11000   11000  83990.00 148.3281  61.40000  6.000000
## 9282  10695   10695  97150.00 142.9499  72.40000  5.000000
## 9349  14990   14990  81957.00 205.0000  47.10000  8.000000
## 9730   8395    8395  85000.00 200.0000  48.70000 10.000000
## 9756  10999   10999  89330.00 148.2928  61.40000  6.000000
## 9877  11999   11999  81459.00 145.0000  55.40000  8.000000
## 9881   7995    7995  93000.00 168.9130  30.70000 12.000000
## 9882  12495   12495  93000.00 200.0000  47.90000  9.000000
## 9884   7495    7495  82041.61 165.0000  49.60000 11.000000
## 9888  10795   10795  78120.00 148.3658  61.40000  6.000000
## 9916   5895    5895  62000.00 163.8626  22.40000  7.726460
## 9925  59900   59900  17283.00 150.0000  29.40000  4.000000
## 10009  7000    7000 112380.00 144.0722  74.30000  6.000000
## 10030  7695    7695 113500.00 145.0000  54.30000  9.000000
## 10034 11495   11495  93820.00 165.4387  30.10000 10.000000
## 10038  3450    3450  92500.00 150.0000  53.30000 13.000000
## 10094  7500    7500  94366.00 152.8748  64.20000  9.000000
## 10126  7850    7850  63000.00 173.8828  27.20000 14.000000
## 10172 17999   17999  66000.00 163.7549  20.30000  7.291866
## 10215  8288    8288  25000.00 150.1658  74.30000  9.000000
## 10426 15888   15888  74000.00 165.8102  29.40000 10.000000
## 10429 11995   11995  85000.00 150.1631  61.40000  7.000000
## 10502 12999   12999  92764.00 145.0000  53.30000  7.000000
## 10679 15500   15500  74907.00 145.0000  52.30000  4.000000
## 10681 15500   15500  77823.00 125.0000  54.30000  5.000000
## 11058 10495   10495  83380.00 125.0000  57.60000  8.000000
## 12844 27852   27852  57501.00 139.1498  85.60000  5.000000
## 14235 10995   10995  75500.00 160.0000  44.80000  7.000000
## 14240 11590   11590  86500.00 143.5894  70.60000  5.000000
## 15297 11740   11740  81838.00 125.0000  57.60000  7.000000
## 17543 15990   15990  68000.00 147.3535  59.06484  5.000000
## 17979 10995   10995 103500.00 144.5103  67.30000  5.000000
## 18006  8995    8995  90000.00 160.0000  51.40000  6.000000
## 18037  7995    7995  59000.00 185.0000  42.20000 14.000000
## 18074  3495    3495  70000.00 162.4273  29.70000  8.227539
## 18138  4150    4150  92659.74 125.0000  60.10000 12.000000
## 18179  9250    9250 105000.00 162.9848  43.50000 11.000000
## 18210 15900   15900  94000.00 161.2359  37.70000  9.000000
## 18340  8990    8990  90000.00 160.0000  50.40000 10.000000
## 18347 10990   10990  88000.00 145.0000  55.40000  8.000000
## 18383 10490   10490  95000.00 125.0000  58.90000  8.000000
## 18671 13699   13699  72111.00 146.1429  62.80000  5.000000
## 18810 12789   12789  85366.00 145.0000  62.80000  4.000000
## 19033 15499   15499  78792.00 145.1971  65.70000  5.000000
## 19186 10000   10000  86800.00 144.6158  67.30000  5.000000
## 19199 13795   13795  80950.00 145.1747  65.70000  5.000000
```

```
## 19251 13295   13295  84140.00 145.1513  65.70000  5.000000
## 19564  4995    4995  81000.00 161.8179  34.40000  8.734569
## 19580  5695    5695 110931.00 160.0000  51.40000 12.000000
## 19618 11999   11999  74014.00 147.9592  62.80000  6.000000
## 19663 10699   10699  93056.00 145.9125  68.90000  6.000000
## 19745 10991   10991  81978.00 125.0000  60.10000  7.000000
## 19801 12999   12999  68949.00 200.0000  43.50000  5.000000
## 19838 11295   11295  94590.00 147.8201  62.80000  6.000000
## 19917  2990    2990  88000.00 161.6345  37.70000  9.229031
## 19929 12695   12695 105730.00 125.0000  58.90000  6.000000
## 20087  2995    2995 105031.06 169.2817  40.90000 14.000000
## 20272 10990   10990  76000.00 160.1384  41.50000  9.000000
## 20301 14299   14299  83354.00 145.0000  54.30000  5.000000
## 20305 11799   11799  97290.00 145.9677  62.80000  5.000000
## 20342 12500   12500 106170.00 160.0000  51.40000  6.000000
## 20416 10295   10295 102900.00 147.3210  64.20000  6.000000
## 20419  4995    4995  93800.00 156.9800  62.80000 11.000000
## 20439 12995   12995  26000.00 164.9202  33.20000 10.000000
## 20468 57000   57000  11297.00 145.0000  22.10000  4.000000
## 20525  6990    6990  78316.00 165.0667  23.70000  8.659869
## 20638  5295    5295  77000.00 205.0000  42.80000 12.000000
## 20645  8999    8999  97800.00 141.4156  83.10000  6.000000
## 20665 11490   11490  89977.00 125.0000  60.10000  9.000000
## 20709  7740    7740  85000.00 141.4940  83.10000  6.000000
## 20868 16500   16500  62368.00 150.0000  58.90000  4.000000
## 20874 13300   13300  83259.00 145.1571  65.70000  5.000000
## 21099 14000   14000  91707.00 150.0000  53.30000  5.000000
## 21143 16000   16000  80895.00 145.0000  54.30000  5.000000
## 21152 10650   10650  88378.00 144.1062  68.90000  5.000000
## 21185  8900    8900  79809.00 147.8283  68.90000  7.000000
## 21208 12400   12400  69736.00 146.1524  62.80000  5.000000
## 21261 18000   18000  63197.00 160.0000  49.60000  4.000000
## 21373  9000    9000  59777.00 165.6991  35.80000 11.000000
## 21421  7995    7995  41500.00 200.0000  41.50000 11.000000
## 21453 61948   61948  16000.00 156.9142  30.40000  5.000000
## 21503 16990   16990  18000.00 168.1161  29.10000 11.000000
## 21524  7340    7340 104000.00 173.1076  28.80000 14.000000
## 22174 12691   12691  78000.00 125.0000  61.40000  6.000000
## 22302 14270   14270  65000.00 142.4408  68.90000  4.000000
## 22316 15561   15561  71610.00 125.0000  56.60000  5.000000
## 24689  8751    8751  91046.00 150.0000  53.30000  8.000000
## 24730 23000   23000  43442.00 164.3809  23.00000  8.000000
## 30490  9489    9489  96351.00 145.0000  52.30000  8.000000
## 31118 14850   14850  90000.00 148.3068  61.40000  6.000000
## 31126 12400   12400  72000.00 141.8683  76.40000  5.000000
## 31132 16450   16450  70000.00 145.7305  64.20000  5.000000
## 31219 30990   30990  72000.00 153.6345  39.20000  5.000000
## 31283 14490   14490  78000.00 147.5053  64.20000  6.000000
## 31301 13990   13990  72000.00 125.0000  51.40000  5.000000
## 31615 11591   11591  94000.00 145.9105  68.90000  6.000000
## 31693 16499   16499  76640.00 145.0000  70.60000  3.000000
## 31703  4990    4990  97717.00 200.0000  43.50000 11.000000
## 31762 16489   16489  62480.00 145.2304  60.10000  4.000000
## 31765 13940   13940  82732.00 125.0000  50.40000  6.000000
## 31812 56900   56900  12000.00 140.0000 100.90000  2.000000
## 31964 11695   11695 103820.00 145.8462  68.90000  6.000000
## 32524 12980   12980  85000.00 147.8914  62.80000  6.000000
## 32610 12195   12195  94440.00 145.5489  64.20000  5.000000
## 32680  9995    9995  51238.00 165.3207  37.20000 11.000000
## 32687  5975    5975  90000.00 205.0000  47.90000 11.000000
## 32696  3995    3995  86638.43 169.8776  27.70000 12.000000
## 32732 16879   16879  74252.00 147.3341  59.00881  5.000000
## 32773 12437   12437  95008.00 125.0000  56.50000  6.000000
## 32870  3790    3790  98000.00 167.2316  24.60000 10.071211
```

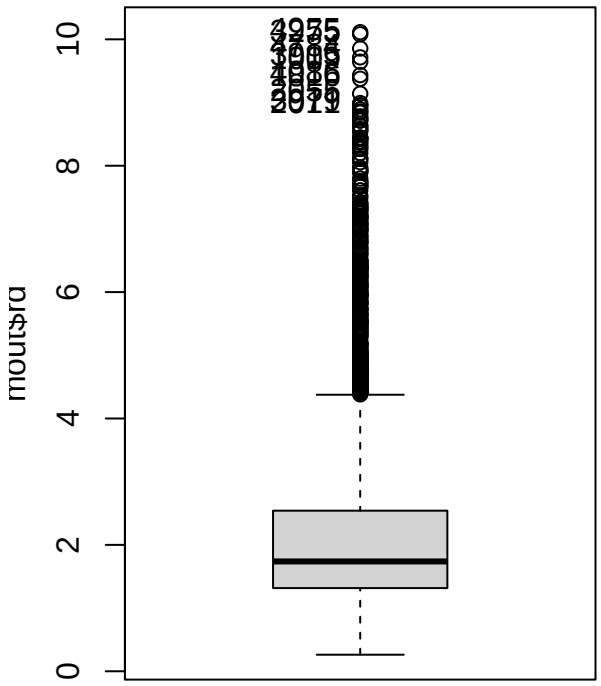
```
## 32994 15995   15995  56098.00 167.8600  29.10000 11.000000
## 33000 12995   12995  45000.00 175.2504  23.30000 14.000000
## 33030 18975   18975  69000.00 174.6334  19.00000 13.000000
## 33048  4995    4995  75000.00 163.4090  28.00000  8.485506
## 33066  8000    8000  91890.00 143.5367  70.60000  5.000000
## 33089 45000   45000  52438.00 163.7360  25.20000  8.000000
## 33197  6850    6850  65000.00 162.1246  28.00000  7.745573
## 33272  7000    7000 109447.00 149.4606  68.90000  8.000000
## 33321  3950    3950  91000.00 164.5594  30.00000  9.513761
## 33500  7490    7490  89000.00 160.0000  52.30000 10.000000
## 33520 13950   13950  85000.00 125.0000  61.40000  6.000000
## 34043 13995   13995  96000.00 147.8237  62.80000  6.000000
## 34193 12800   12800  79011.00 148.3695  61.40000  6.000000
## 34409  9990    9990  59000.00 169.2410  30.40000 12.000000
## 35767 12990   12990  74166.00 146.5503  67.30000  6.000000
## 36500  7999    7999  75151.00 149.0778  53.30000  5.000000
## 37263  5991    5991  87803.00 141.6718  88.30000  7.000000
## 37320  8798    8798  75577.00 150.9157  53.30000  6.000000
## 37684 13499   13499  80227.00 144.6758  67.30000  5.000000
## 37924  7480    7480  85240.00 144.2536  74.30000  6.000000
## 38753 13500   13500  76077.00 146.1158  62.80000  5.000000
## 39109  8940    8940  81351.00 142.4493  74.30000  5.000000
## 39150 10291   10291 106000.00 145.9030  62.80000  5.000000
## 39542 12500   12500  95890.00 145.5668  64.11777  5.000000
## 39550 11000   11000  93310.00 146.4145  67.30000  6.000000
## 39721  4995    4995 111913.00 145.0000  52.30000 11.000000
## 39749  6395    6395 113000.00 147.5973  68.90000  7.000000
## 39750  4295    4295 105000.00 151.4558  74.30000 10.000000
## 39881  2695    2695  88000.00 200.0000  41.50000  9.946728
## 39903 12500   12500  87490.00 144.6231  67.30000  5.000000
## 40007 17495   17495  67000.00 145.0000  57.70000  4.000000
## 40048  8995    8995  73000.00 166.6954  38.20000 12.000000
## 40054  4490    4490  83000.00 145.0000  50.40000  9.000000
## 40120  4495    4495  88000.00 160.0000  49.60000 13.000000
## 40140  3195    3195  89000.00 190.0000  40.40000  9.797236
## 40164  6251    6251  97181.00 147.7010  68.90000  7.000000
## 40640  7520    7520  74178.00 145.4882  70.60000  6.000000
## 40766  9500    9500  89100.00 146.0108  62.80000  5.000000
## 40791  8490    8490  85000.00 144.6204  67.30000  5.000000
## 40793  7990    7990  85000.00 144.1158  68.90000  5.000000
## 40849  5991    5991 110000.00 149.9695  61.40000  7.000000
## 40871  9292    9292 104000.00 144.4989  67.30000  5.000000
## 40880  8049    8049  82934.00 145.9666  68.90000  6.000000
## 40898  9295    9295  99160.00 146.7910  60.10000  5.000000
## 41006  2995    2995  90000.00 200.0000  47.10000 12.000000
## 41052  9495    9495  74784.00 144.1904  68.90000  5.000000
## 41066  9200    9200  87241.00 145.0000  68.90000  4.000000
## 43917  3195    3195  58500.00 160.0000  43.50000 13.000000
## 44342  3695    3695 102000.00 155.5071  67.30000 11.000000
## 44558  4250    4250  79000.00 125.0000  51.40000 11.000000
## 44565  2495    2495  41637.00 145.0000  48.70000 12.000000
## 44615  5740    5740  80000.00 150.0000  47.90000  9.000000
## 44642  3295    3295 116000.00 165.3158  35.80000 11.008121
## 44982 10500   10500  72414.00 150.0000  53.30000  6.000000
## 47233  4995    4995  78977.00 150.5775  60.10000  7.000000
## 47744  7428    7428  59361.00 165.2549  37.20000 11.000000
## 47780  8699    8699  51206.00 163.4778  37.20000 10.000000
## 47830 10999   10999  16969.00 160.0408  37.20000  8.000000
## 47890  3850    3850  40000.00 171.1903  36.20000 14.000000
## 48036 20995   20995  96000.00 164.1199  34.40000 10.000000
## 48043 14495   14495 106000.00 163.5345  34.40000  9.734175
## 48062  5950    5950 119000.00 171.2592  28.50000 13.000000
## 48117 14495   14495  83451.00 157.1540  39.20000  7.000000
## 48417  8995    8995 111000.00 169.4896  28.50000 12.000000
```

```
## 48418  6000    6000 105000.00 173.6910  26.90000 14.000000
## 49573  8852    8852  69222.00 145.0000  67.30000  4.000000
## 49583  2295    2295 111000.00 155.0000  48.70000 13.000000
## 49623  8998    8998  84932.00 148.7204  60.10000  6.000000
## 49652  8888    8888 108808.00 125.0000  58.90000  5.000000
## 49720  2995    2995  92640.00 200.0000  48.00000 10.125615
```

```
df$mout <- 0
df$mout[ ll ]<-1
df$mout <- factor( df$mout, labels=c( "NoMOut","YesMOut"))

summary(df$mout)
```

```
##  NoMOut YesMOut
##    4736     224
```

```
df<-df[-13] #treiem la variable aux
#save(df, vars_con, vars_dis, vars_res, file = "MyOldCars-5000Clean.RData")
```



Podem veure que el factor Audi si/no te relació amb les milles per galons (si no és Audi) i amb la taxa de circulació i el preu (si ho és).

```
names(df)
```

```
##  [1] "model"        "year"         "price"        "transmission" "mileage"
##  [6] "fuelType"     "tax"          "mpg"          "engineSize"   "manufacturer"
## [11] "age"          "f.price"      "f.miles"      "f.tax"        "f.mpg"
## [16] "f.age"        "Audi"         "mout"
```

```
vars_con<-names(df)[c(5,7:8, 11)]
vars_dis<-names(df)[c(1:2, 4, 6, 9:10, 12:16)]
vars_res<-names(df)[c(3,17)]
vars_dis
```

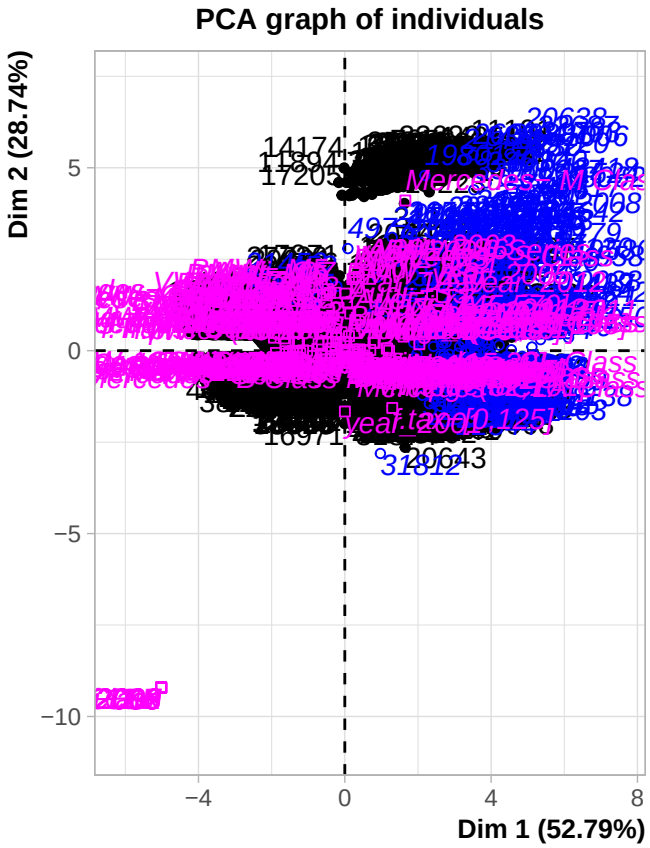
```
##  [1] "model"        "year"         "transmission" "fuelType"     "engineSize"
##  [6] "manufacturer" "f.price"      "f.miles"      "f.tax"        "f.mpg"
## [11] "f.age"
```

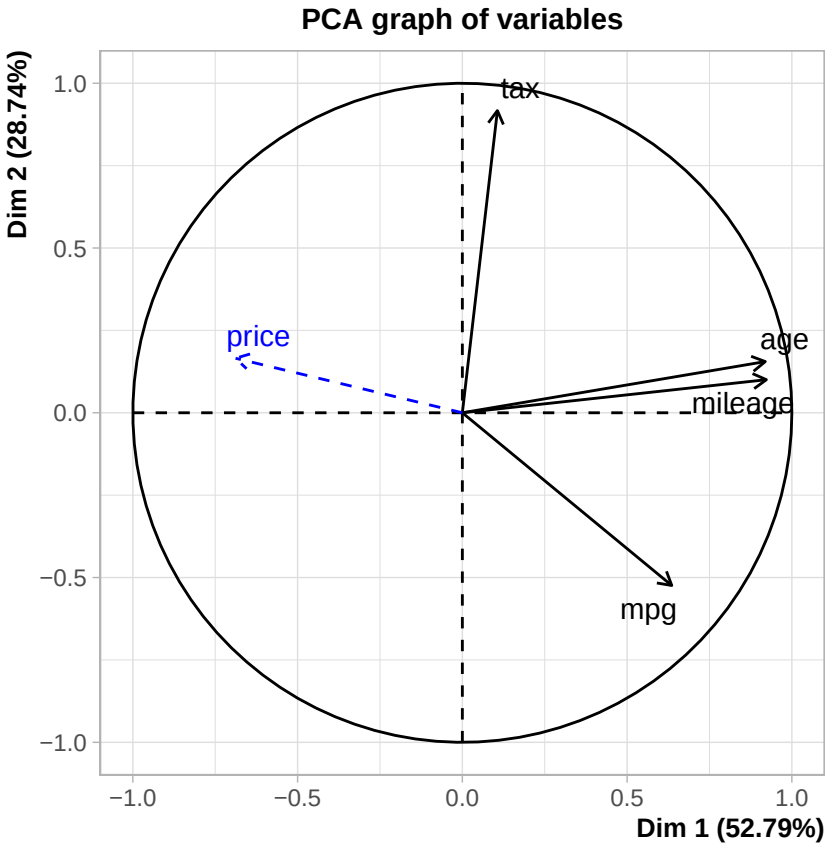
5 PCA analysis

```
c(vars_res, vars_dis,vars_con, "mout")
```

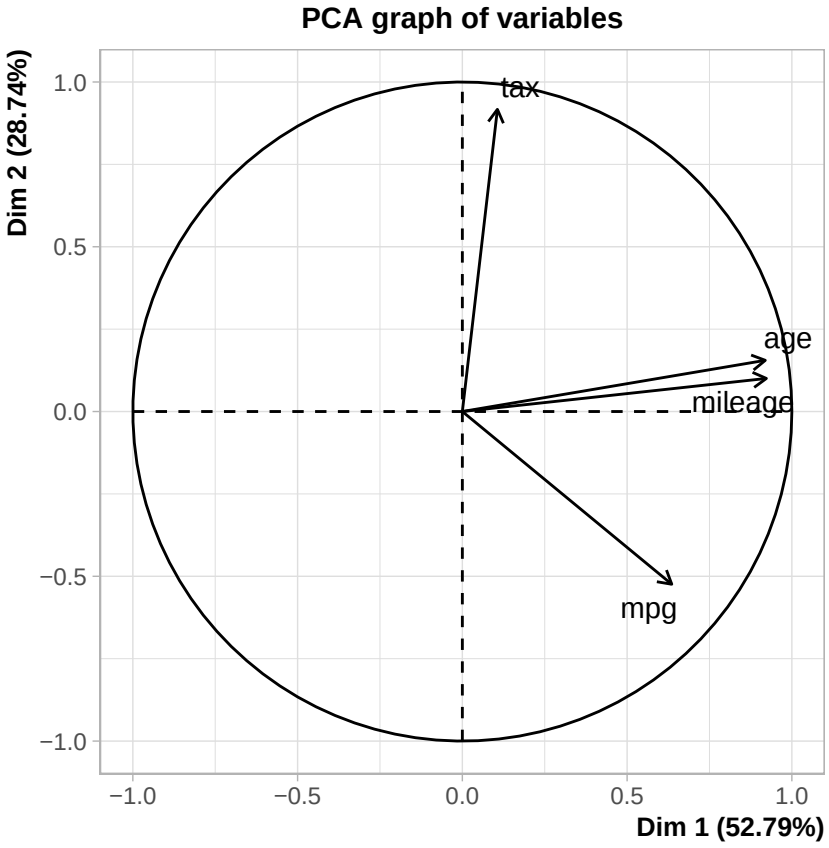
```
##  [1] "price"        "Audi"         "model"        "year"         "transmission"
##  [6] "fuelType"     "engineSize"   "manufacturer" "f.price"      "f.miles"
## [11] "f.tax"        "f.mpg"        "f.age"        "mileage"      "tax"
## [16] "mpg"          "age"          "mout"
```

```
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 )
```

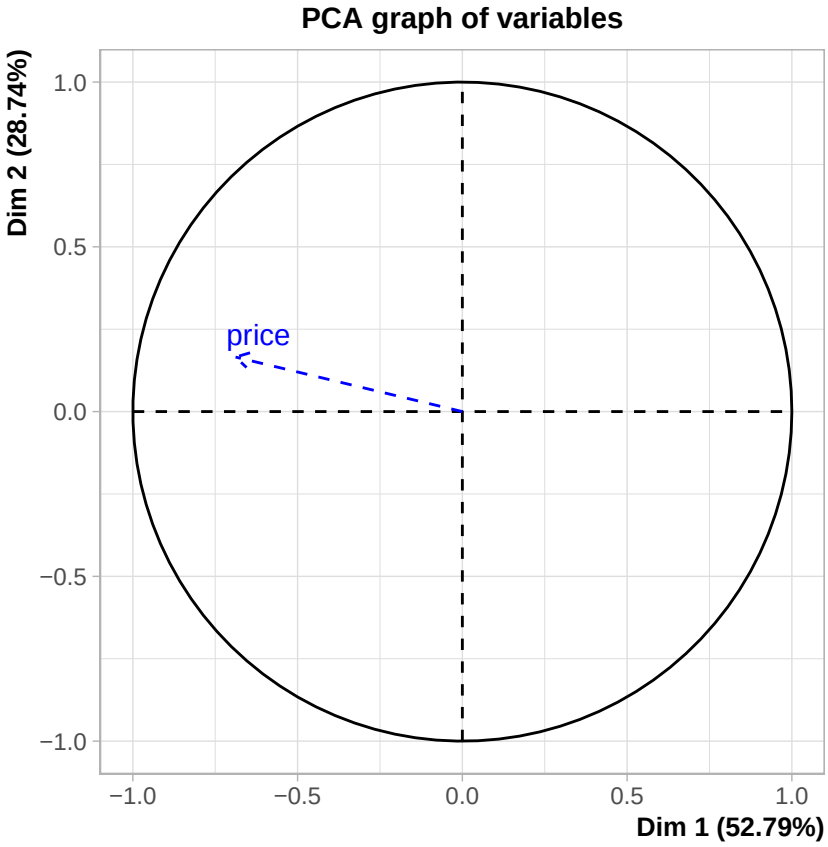




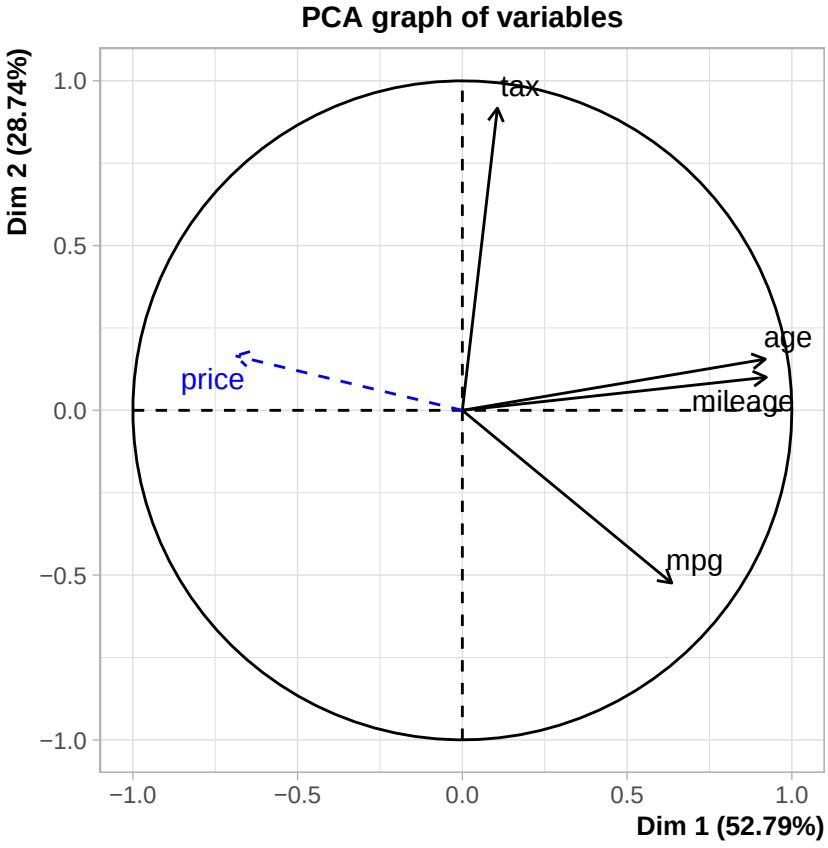

```
plot.PCA(res.pca,choix=c("var"),invisible=c("quanti.sup"))
```



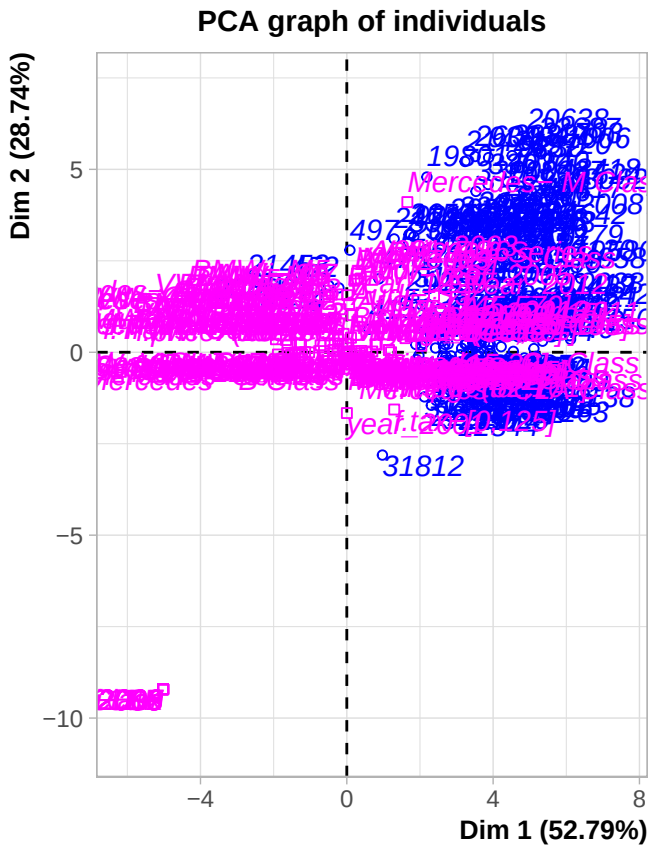
```
plot.PCA(res.pca,choix=c("var"),invisible=c("var"))
```




```
plot.PCA(res.pca,choix=c("var"),invisible=c("quali"))
```



```
plot.PCA(res.pca,choix=c("ind"),invisible=c("ind"))
```



5.1 Eigenvalues and dominant axes analysis

Segons el criteri de Kaiser hem d'agafar 2 dimensions, que són les que tenen un eigenvalue major a 1. Segons Elbow hem d'agafar 2 o 3.

```
summary(res.pca,nb.dec=2,nbind=1,nbelements=1,ncp=2)
```

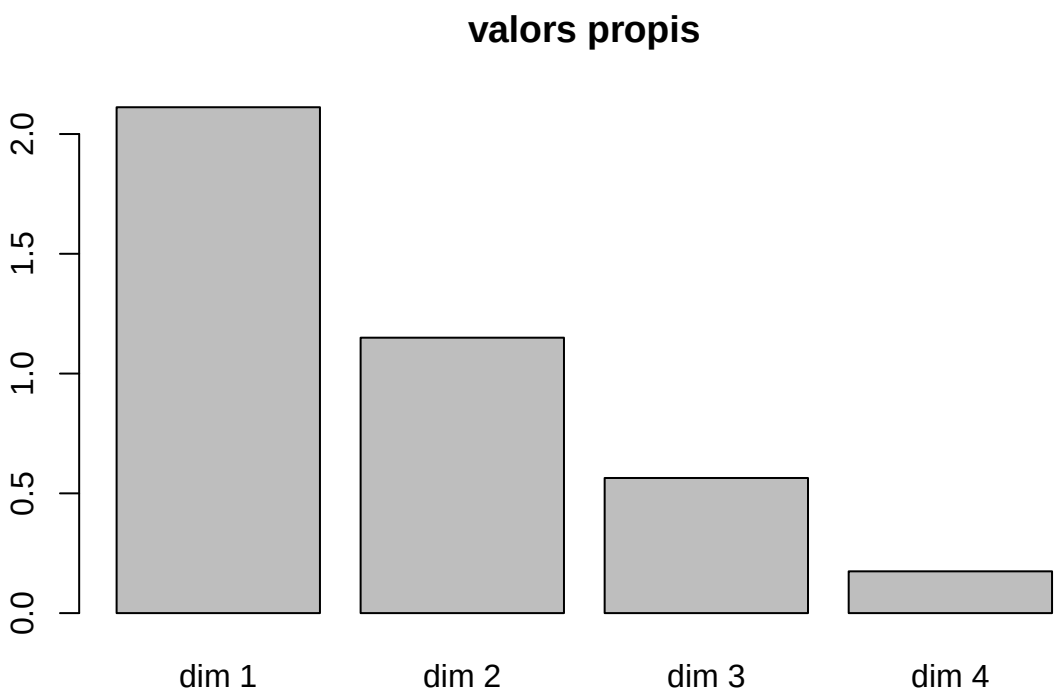
```
##
## Call:
## PCA(X = df[, c(vars_res, vars_dis, vars_con)], ind.sup = ll,
##      quanti.sup = c(1), quali.sup = c(2:13))
##
##
## Eigenvalues
##                        Dim.1  Dim.2  Dim.3  Dim.4
## Variance                2.11   1.15   0.56   0.17
## % of var.              52.79  28.74  14.10   4.36
## Cumulative % of var.   52.79  81.54  95.64 100.00
##
## Individuals (the 1 first)
##            Dist   Dim.1   ctr  cos2   Dim.2   ctr  cos2
## 54      |  1.45 |  1.44  0.02  0.98 | -0.09  0.00  0.00 |
##
## Supplementary individuals (the 1 first)
##            Dist   Dim.1  cos2   Dim.2  cos2
## 152     |  5.38 |  5.18  0.93 | -0.11  0.00 |
##
## Variables (the 1 first)
##           Dim.1   ctr  cos2   Dim.2   ctr  cos2
## mileage |  0.92 40.31  0.85 |  0.10  0.88  0.01 |
##
```

```
## Supplementary continuous variable
##           Dim.1  cos2   Dim.2  cos2
## price   | -0.69  0.47 |  0.17  0.03 |
##
## Supplementary categories (the 1 first)
##            Dist   Dim.1  cos2 v.test   Dim.2  cos2 v.test
## Audi No |  0.07 |  0.00  0.00   0.16 | -0.06  0.74  -7.61 |
```

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

```
##        eigenvalue percentage of variance cumulative percentage of variance
## comp 1       2.11                 52.79                             52.79
## comp 2       1.15                 28.74                             81.54
## comp 3       0.56                 14.10                             95.64
## comp 4       0.17                  4.36                            100.00
```

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



5.2 Individuals point of view

Fem servir els extreme individuals per entendre millor els axes. Cap es Multivariate Outlier i la mitat son Audi, sent els 2 més influents Audi. Tots comparteixen molts atributs, com ser Manual, Diesel, amb anys entre 8 i 10, preu menor a 15.000 lliures, etc.

```
inds <- res.pca$ind$coord
inds <- as.data.frame(inds)
rang<-inds[order(inds$Dim.1, decreasing = TRUE),]

rang[1,]
```

```
##         Dim.1     Dim.2      Dim.3     Dim.4
## 9642 5.100293 0.9513074 -0.9979927 0.1390043
```

```
res.pca$ind$coord[row.names(rang)[1:10],1]
```

```
##     9642     8253    39252     9608    20084     7892     9633    35755
## 5.100293 4.556650 4.502366 4.495247 4.448843 4.359922 4.359520 4.338991
##    49676    20408
## 4.178526 4.156100
```

```
df[which(row.names(df) %in% row.names(res.pca$ind$coord[row.names(rang)[1:10],])),1:18]
```

```
##               model year price      transmission mileage      fuelType      tax
## 7892       Audi- TT 2011 10250    f.Trans-Manual   80000 f.Fuel-Diesel 145.0000
## 8253       Audi- A1 2012  8299    f.Trans-Manual   76975 f.Fuel-Diesel 151.5180
## 9608       Audi- A3 2011  6875    f.Trans-Manual   70000 f.Fuel-Diesel 154.8696
## 9633       Audi- A1 2012  6790    f.Trans-Manual   65794 f.Fuel-Diesel 149.8895
## 9642       Audi- A3 2011  5290    f.Trans-Manual   87000 f.Fuel-Diesel 154.7498
## 20084 BMW- 1 Series 2011  6000    f.Trans-Manual   80183 f.Fuel-Diesel 145.0000
## 20408 BMW- 3 Series 2013  9195    f.Trans-Manual   82260 f.Fuel-Diesel 151.5653
## 35755      VW- Golf 2013  6498    f.Trans-Manual   81025 f.Fuel-Diesel 149.6457
## 39252      VW- Golf 2013  6500    f.Trans-Manual   79971 f.Fuel-Diesel 147.9576
## 49676        VW- CC 2011  7440 f.Trans-Automatic   74000 f.Fuel-Diesel 150.0000
##        mpg engineSize manufacturer age        f.price          f.miles
## 7892  53.3     Medium         Audi  10 f.price-[0,15] f.miles-(36,195]
## 8253  68.9     Medium         Audi   9 f.price-[0,15] f.miles-(36,195]
## 9608  64.2     Medium         Audi  10 f.price-[0,15] f.miles-(36,195]
## 9633  74.3     Medium         Audi   9 f.price-[0,15] f.miles-(36,195]
## 9642  64.2     Medium         Audi  10 f.price-[0,15] f.miles-(36,195]
## 20084 55.4     Medium          BMW  10 f.price-[0,15] f.miles-(36,195]
## 20408 62.8     Medium          BMW   8 f.price-[0,15] f.miles-(36,195]
## 35755 68.9     Medium           VW   8 f.price-[0,15] f.miles-(36,195]
## 39252 74.3     Medium           VW   8 f.price-[0,15] f.miles-(36,195]
## 49676 53.3     Medium           VW  10 f.price-[0,15] f.miles-(36,195]
##                  f.tax          f.mpg          f.age     Audi   mout
## 7892  f.tax-(125,145]  f.mpg-(45,54] f.age-(5.1,15] Audi Yes NoMOut
## 8253  f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi Yes NoMOut
## 9608  f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi Yes NoMOut
## 9633  f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi Yes NoMOut
## 9642  f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi Yes NoMOut
## 20084 f.tax-(125,145]  f.mpg-(54,62] f.age-(5.1,15]  Audi No NoMOut
## 20408 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15]  Audi No NoMOut
## 35755 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15]  Audi No NoMOut
## 39252 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15]  Audi No NoMOut
## 49676 f.tax-(145,570]  f.mpg-(45,54] f.age-(5.1,15]  Audi No NoMOut
```

5.3 Interpreting the axes

Les variables que més contribueixen en la primera dimensió son mileage i age, que estan molt relacionades. En la segona dimensió la més representativa es tax. Les variables qualitatives amb més contribució a la dimensió 1 són year, el factor de mileage i el factor de age (com les quantitatives mileage i age).

```
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  0.92  0.10  0.85  0.01 40.31  0.88
## tax      0.11  0.92  0.01  0.84  0.54 73.13
## mpg      0.64 -0.52  0.40  0.27 19.13 23.89
## age      0.92  0.16  0.85  0.02 40.02  2.10
```

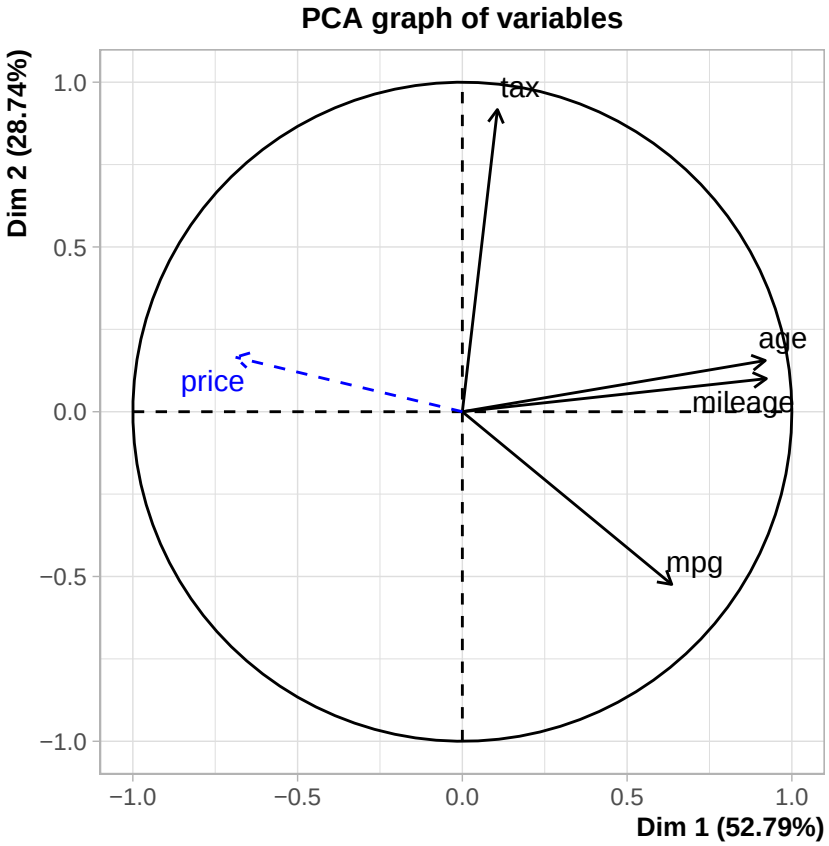
```
round(cbind(res.pca$var$cos2[,1:2],res.pca$var$contrib[,1:2]),2)
```

```
##         Dim.1 Dim.2 Dim.1 Dim.2
## mileage  0.85  0.01 40.31  0.88
## tax      0.01  0.84  0.54 73.13
## mpg      0.40  0.27 19.13 23.89
## age      0.85  0.02 40.02  2.10
```

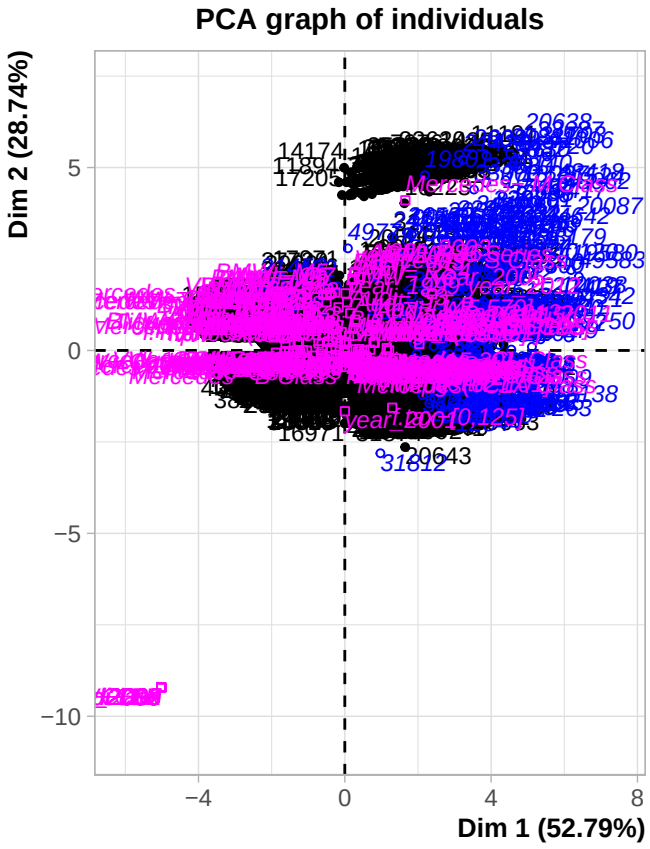
```
res.des<-dimdesc(res.pca)
res.des$Dim.1$quali
```

```
##                       R2       p.value
## year          0.857422800  0.000000e+00
## f.price       0.506990767  0.000000e+00
## f.miles       0.801836418  0.000000e+00
## f.mpg         0.365971770  0.000000e+00
## f.age         0.805337957  0.000000e+00
## f.tax         0.221017477 1.996805e-257
## model         0.120573810  2.492417e-80
## transmission  0.071762554  2.918926e-77
## fuelType      0.071331141  8.765539e-77
## engineSize    0.001712024  1.733587e-02
```

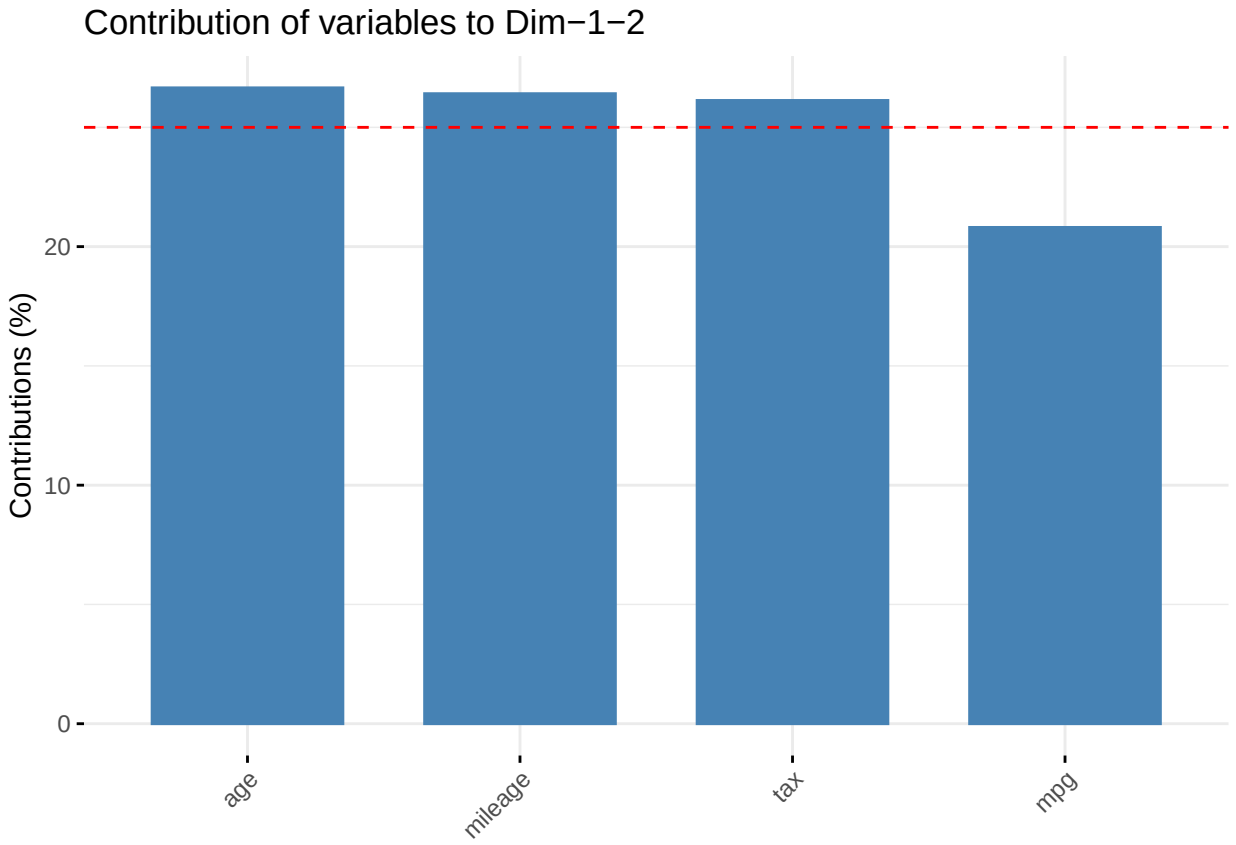
```
plot.PCA(res.pca,choix=c("var"),axes=c(1,2))
```




```
plot.PCA(res.pca,choix=c("ind"),cex=0.8)
```




```
fviz_contrib(res.pca, choice="var", axes = 1:2)
```



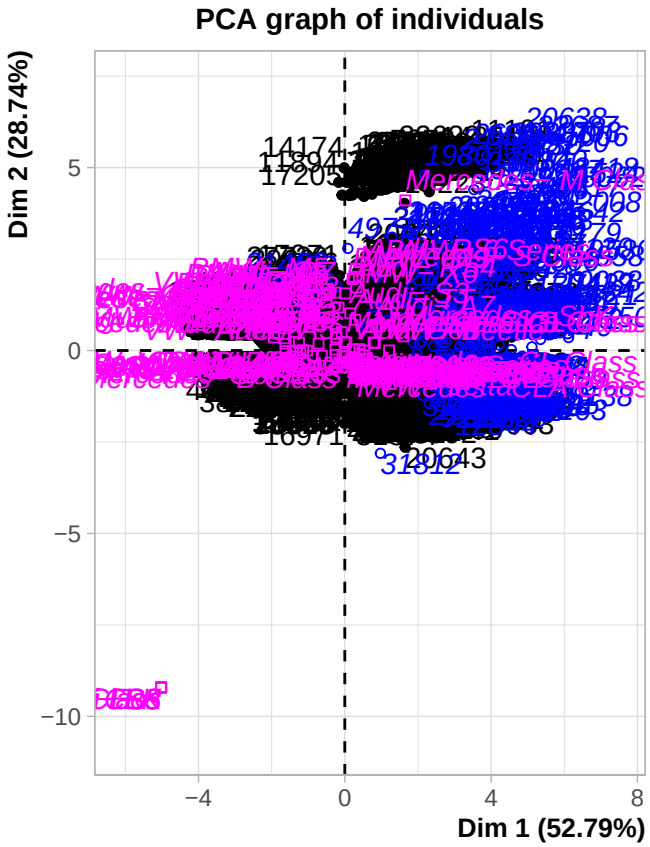
5.4 PCA with supplementary variables

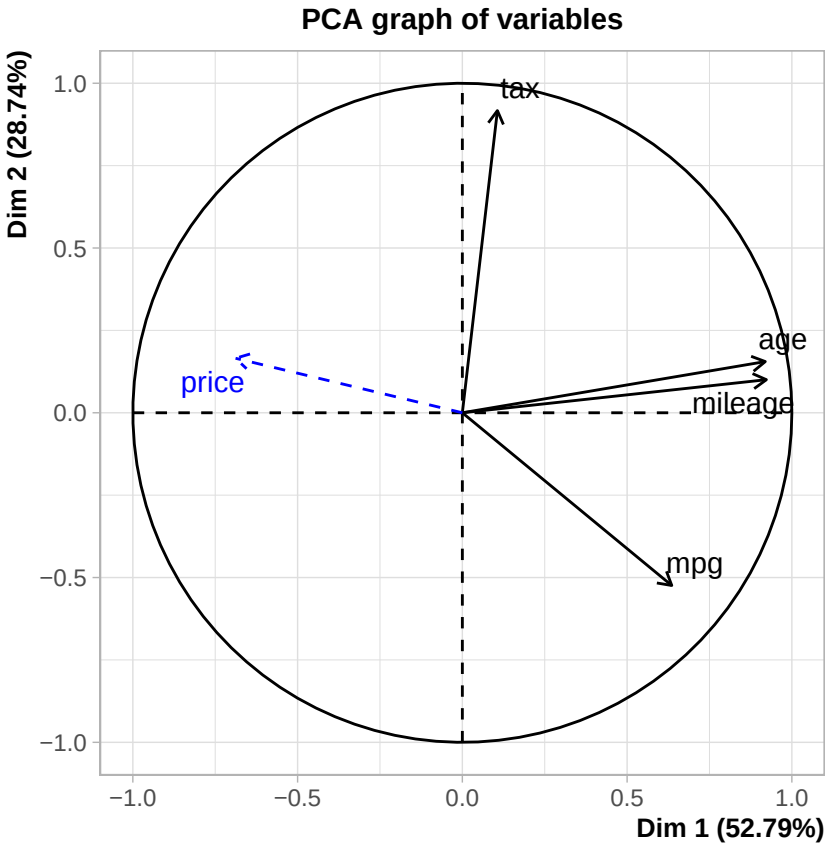
Fent servir com a variables suplementaries el factor Audi SI/NO, el model, el price i els MOut podem tornar a veure uns resultats mols semblants als de l'apartat anterior. La variable Audi SI/No, sobretot Audi SI, .

```
c(vars_res, vars_dis,vars_con)
```

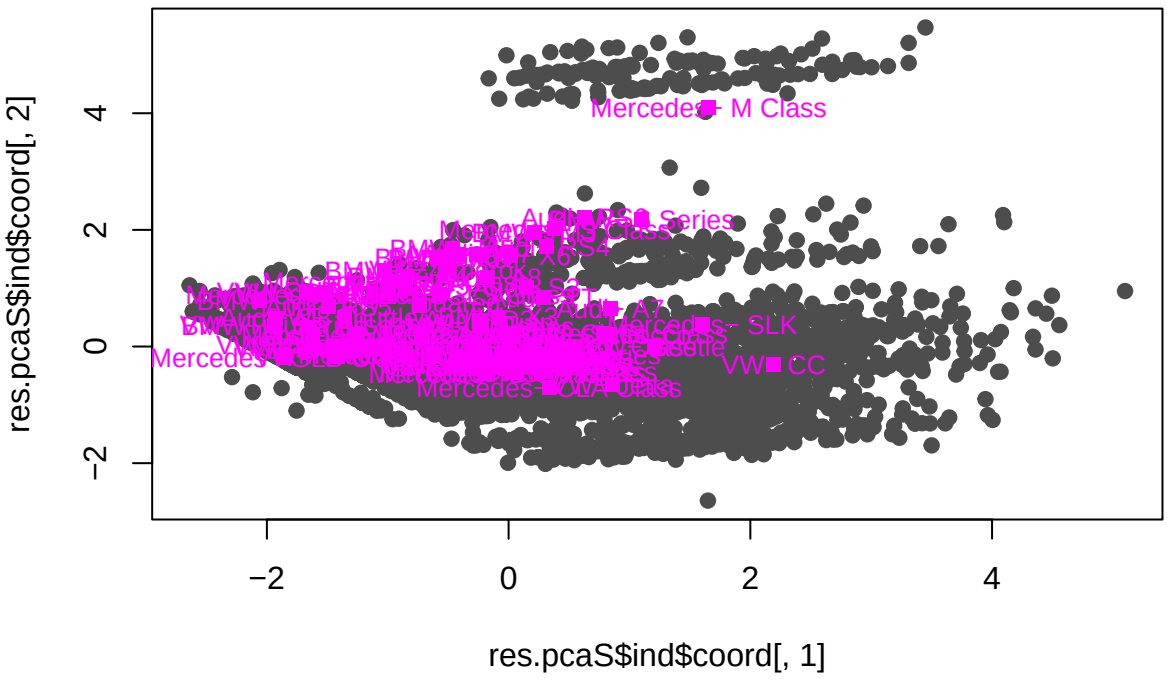
```
##  [1] "price"        "Audi"         "model"        "year"         "transmission"
##  [6] "fuelType"     "engineSize"   "manufacturer" "f.price"      "f.miles"
## [11] "f.tax"        "f.mpg"        "f.age"        "mileage"      "tax"
## [16] "mpg"          "age"
```

```
ll <- which( df$mout == "YesMOut")

res.pcaS<-PCA(df[,c(vars_res, vars_con, "model")],quali.sup=c(2,7),quanti.sup= c(1), ind.sup = ll)
```






```
plot(res.pcaS$ind$coord[,1],res.pcaS$ind$coord[,2],pch=19,col="grey30")
points(res.pcaS$quali.sup$coord[,1],res.pcaS$quali.sup$coord[,2],pch=15,col="magenta")
text(res.pcaS$quali.sup$coord[,1],res.pcaS$quali.sup$coord[,2],labels=names(res.pcaS$quali.sup$coord[,1]
```



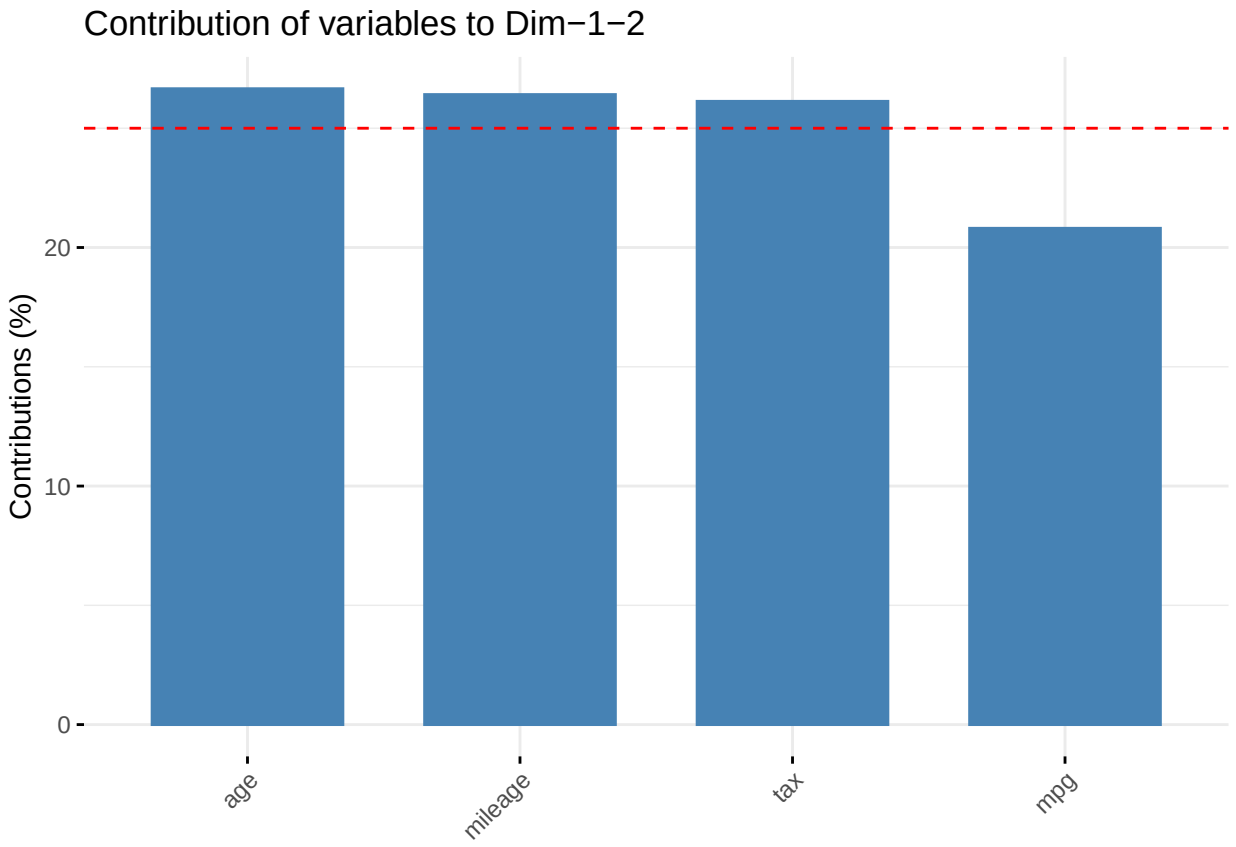

```
res.pcaS$quali.sup$coord
```

```
##                      Dim.1      Dim.2      Dim.3      Dim.4
```

```
## Audi No              0.001760855 -0.0613609737  0.036116723 -0.0004894634
## Audi Yes            -0.006578554  0.2292445977 -0.134932076  0.0018286353
## Audi- A1             0.532806538 -0.2279545726  0.189448304 -0.1246478776
## Audi- A3             0.403455637 -0.0905462276  0.054512274 -0.0117170444
## Audi- A4            -0.057928670 -0.1538371317 -0.176078491  0.0745701625
## Audi- A5            -0.253541750  0.0258927824 -0.287087576  0.0588910623
## Audi- A6             0.401136357  0.0361371599 -0.087499107  0.1085267912
## Audi- A7             0.843894610  0.6518415518 -0.733886329 -0.2699756555
## Audi- A8            -0.171693998  1.1768555956  0.253942129 -0.3266879414
## Audi- Q2            -0.874509724 -0.1439696265 -0.042031760 -0.0570347590
## Audi- Q3            -0.243690778  0.4479434731 -0.396183390  0.0295657024
## Audi- Q5            -0.410757056  1.5757463352 -0.143868855  0.0809960668
## Audi- Q7            -0.262214433  1.5821781838 -0.330608817  0.1781752448
## Audi- Q8            -1.925697330  0.4847466335 -0.886442833 -0.0296240256
## Audi- R8            -5.021106174 -9.2107840334 -9.345191194  0.0777486508
## Audi- RS3           -0.999311615  0.9561443254 -1.152138552  0.0218348779
## Audi- RS4            0.316585668  1.7219583256 -1.881031174  0.2713345884
## Audi- RS5           -1.488950127  0.6938656313 -1.283067855 -0.2227733885
## Audi- RS6            0.628760535  2.2101891883 -1.730019573 -0.1890354710
## Audi- S3             0.147459089  1.0150838235 -0.768468170 -0.0670732366
## Audi- S4            -1.621314822  0.1841337088 -0.430041014  0.0029257099
## Audi- S5            -5.021106174 -9.2107840334 -9.345191194  0.0777486508
## Audi- SQ5           -0.851664109  1.0463014491 -1.080376679 -0.4408984135
## Audi- TT             0.286811795  0.8405953429 -0.332754596 -0.1176613836
## BMW- 1 Series        0.491405581 -0.1329646118  0.041855888 -0.0398499149
## BMW- 2 Series       -0.251467868 -0.1188387758  0.078085765 -0.0553150420
## BMW- 3 Series        0.256781367 -0.1810564355  0.015943863  0.1129213660
## BMW- 4 Series        0.035713563 -0.1642917371  0.133258032  0.0329883652
## BMW- 5 Series        0.277022108 -0.2810634989  0.244136506  0.0847020254
## BMW- 6 Series        1.100191780  2.1821887231  0.489165442  0.0448045360
## BMW- 7 Series       -0.227080573  0.3483387289 -0.338330996 -0.1022826427
## BMW- 8 Series       -1.937903066  0.3546201839 -0.678520590  0.1119629315
## BMW- i3              0.101413172 -0.3327053844 -0.172405407 -0.6156726229
## BMW- M2             -2.060972078  0.7953342886 -1.121464603  0.2646170360
## BMW- M3              0.210570280  1.9542336728 -1.530445478 -0.1422808830
## BMW- M4             -0.997294996  1.0680050084 -1.029624050 -0.2316266079
## BMW- M5             -0.467398877  1.6817488845 -1.637202221  0.2486147644
## BMW- M6             -0.590183749  1.5252856186 -1.574822713 -0.3072141910
## BMW- X1              0.298415905 -0.4057615813 -0.132231735 -0.0253888702
## BMW- X2             -1.415064116 -0.0073660552 -0.084624044  0.0394037504
## BMW- X3             -0.086435759  0.5038649724 -0.357675432  0.0193869844
## BMW- X4             -0.833476981  1.2078496151 -0.100035991  0.1238398797
## BMW- X5             -0.500070001  1.5119132188 -0.190167909  0.0972149132
## BMW- X6              0.018989669  1.5600970660 -0.590004195  0.0942551315
## BMW- Z4             -1.028884295  1.2969622635 -0.499851140  0.0123645876
## Mercedes- A Class    0.273574709 -0.4313611413  0.410472884 -0.0052911790
## Mercedes- B Class   -0.199470580 -0.4576597324  0.420796176  0.0294202097
## Mercedes- C Class   -0.072002653 -0.3077967539  0.134299923  0.0064199972
## Mercedes- CL Class   0.533518924 -0.0016143084 -0.065477385  0.0884990197
## Mercedes- CLA Class  0.336080932 -0.6971789049 -0.191618702  0.1395519006
## Mercedes- CLK       -5.021106174 -9.2107840334 -9.345191194  0.0777486508
## Mercedes- CLS Class  0.041816070  0.2181018772 -0.181331641 -0.1327771437
## Mercedes- E Class    0.132697370 -0.3351952973  0.206894330  0.0151481595
## Mercedes- G Class   -5.021106174 -9.2107840334 -9.345191194  0.0777486508
## Mercedes- GL Class   0.784471832  0.2194195331 -0.513425850 -0.2619668013
## Mercedes- GLA Class  0.036825271 -0.4520716634 -0.024370359 -0.0646055731
## Mercedes- GLB Class -1.861059918 -0.1895334706  0.209664724  0.2436354887
## Mercedes- GLC Class -0.628644689 -0.1434966342 -0.361030699  0.1085223011
## Mercedes- GLE Class -0.929543407  1.1085915402 -0.504545993  0.0582411361
## Mercedes- GLS Class -0.571344463  0.8965918236 -0.958147593 -0.1714246724
## Mercedes- M Class    1.655129820  4.1047021682  0.386120376 -0.2297942371
## Mercedes- S Class    0.382964581  2.0156435601  0.480620823 -0.2058310572
## Mercedes- SL CLASS  -1.031172849 -0.0039435938 -0.130626020 -0.1454166225
## Mercedes- SLK        1.602749501  0.3857431358  0.323835428 -0.6221507215
```

```
## Mercedes- V Class   -1.252282129  0.2083734563 -0.412208015  0.0052577532
## Mercedes- X-CLASS   -1.593374407  0.8851004446 -0.626614924 -0.0773318193
## VW- Amarok          -0.501381164  1.3173010587 -1.087306477  0.1526386764
## VW- Arteon          -1.139612900  0.0062439129 -0.059056294  0.0245426983
## VW- Beetle           1.209552540  0.0005605478 -0.764485571  0.1257493240
## VW- Caddy Life      -1.621160640 -0.0165864450  0.748606397  0.2833369203
## VW- Caddy Maxi Life -1.217773226 -0.0699471516  0.146659395  0.0866595551
## VW- California      -1.675869752  0.9479698460 -0.758543821  0.2382434809
## VW- Caravelle       -1.525249933  0.9214831593 -0.710218653  0.0484335217
## VW- CC               2.194408615 -0.3133857074 -0.718919957  0.1036955423
## VW- Eos             -5.021106174 -9.2107840334 -9.345191194  0.0777486508
## VW- Golf             0.214704273  0.0300846483  0.068263607  0.0215352962
## VW- Golf SV         -0.394122369 -0.3292591717  0.030691726 -0.1106601699
## VW- Jetta            0.859228942 -0.6481666290  0.720053886  0.0387709040
## VW- Passat           0.212641254 -0.2974972680  0.091418951  0.2802949491
## VW- Polo             0.331720223 -0.2217803251  0.184915059 -0.1530301903
## VW- Scirocco         0.819143470  0.2068973522 -0.418453887 -0.1302302303
## VW- Sharan          -0.717552308  0.1908426189 -0.275651969  0.0607110777
## VW- Shuttle         -1.361538509  0.5050827538 -0.497947785  0.1707856755
## VW- T-Cross         -1.415074359 -0.1632640077  0.252384106 -0.0131289051
## VW- T-Roc           -1.324428219 -0.0047366264 -0.007923737  0.0357237389
## VW- Tiguan          -0.232371873  0.3601498412 -0.260195270  0.0546941874
## VW- Tiguan Allspace -1.660882211  0.3346642737 -0.683619453  0.0970745845
## VW- Touareg         -0.734655925  0.7121587341 -0.914273480  0.0086221187
## VW- Touran           0.156808651 -0.2372681710 -0.029730566  0.0950145262
## VW- Up               0.469710588 -0.3462173145  0.471703134 -0.2836128783
```

```
fviz_contrib(res.pcaS, choice="var", axes = 1:2)
```



6 K-Means Classification

6.1 Description of clusters

Un cop calculat el K-Means amb k=14, podem veure una inèrcia explicada del 92%. Les variables categoriques que estan més relacionades amb els cluster són el model i el manufacturer. Com es obvi, quan Audi = No, les variables que estan directament relacionades són el models de cotxe no Audi, encara que també ho són els cotxes amb engineSize = Big, i inversament relacionades les que si son audi. Quan Audi = Yes, apart dels modesl Audi, estan relacionats els cotxes amb engineSize Medium i amb mpg menor a 45. Les variables numeriques més relacionades amb els cluster són mpg, tax i price, ja sigui Audi o no.

```
ppcc<-res.pca$ind$coord[,1:4]
dim(ppcc)
```

```
## [1] 4736    4
```

```
kc<-kmeans(dist(ppcc),12, iter.max = 30, trace=F)
kc$size
```

```
##  [1] 330 356 193 185 455 820 137 648 657 309 142 504
```

```
llvout<-which(df$mout=="YesMOut")
df[-llvout,"claKMMC"]<-kc$cluster
kc$betweenss/kc$totss
```

```
## [1] 0.9125898
```

```
catdes(df,17)
```

```
##
## Link between the cluster variable and the categorical variables (chi-square test)
## =================================================================================
##                    p.value df
## model         0.000000e+00 87
## manufacturer  0.000000e+00  3
## engineSize    2.963844e-20  2
## f.mpg         5.001165e-17  3
## f.miles       2.467586e-05  3
## fuelType      9.706971e-05  2
## f.price       5.015511e-04  3
## mout          1.744646e-03  1
## transmission  2.730973e-03  2
## f.age         3.887650e-02  3
##
## Description of each cluster by the categories
## =============================================
## $`Audi No`
##                                  Cla/Mod    Mod/Cla      Global       p.value
## manufacturer=VW                100.00000 38.8646288 30.50403226 5.282569e-196
## manufacturer=Mercedes          100.00000 33.3418957 26.16935484 1.149117e-162
## manufacturer=BMW               100.00000 27.7934755 21.81451613 1.899203e-131
## model=VW- Golf                 100.00000 12.6637555  9.93951613  9.720293e-56
## model=Mercedes- C Class        100.00000  9.5812998  7.52016129  9.888214e-42
## model=VW- Polo                 100.00000  8.6565631  6.79435484  1.305024e-37
## model=Mercedes- A Class        100.00000  6.8841510  5.40322581  8.160316e-30
## model=BMW- 3 Series            100.00000  6.3704084  5.00000000  1.402828e-27
## model=BMW- 1 Series            100.00000  5.1631133  4.05241935  2.278596e-22
## model=Mercedes- E Class        100.00000  4.6493707  3.64919355  3.607674e-20
## model=VW- Tiguan               100.00000  4.0585667  3.18548387  1.185258e-17
## engineSize=Big                  87.47495 22.4248651 20.12096774  5.661366e-16
```

```
## model=BMW- 2 Series            100.00000  3.1852042  2.50000000  5.874847e-14
## model=BMW- 4 Series            100.00000  2.7998973  2.19758065  2.453892e-12
## model=BMW- 5 Series            100.00000  2.5944002  2.03629032  1.786443e-11
## model=Mercedes- GLC Class      100.00000  2.5687131  2.01612903  2.288983e-11
## model=BMW- X1                  100.00000  2.2090932  1.73387097  7.314034e-10
## model=Mercedes- GLA Class      100.00000  2.1834061  1.71370968  9.363423e-10
## model=VW- Passat               100.00000  2.1577190  1.69354839  1.198636e-09
## model=VW- Up                   100.00000  2.1320319  1.67338710  1.534317e-09
## model=VW- T-Roc                100.00000  2.0549705  1.61290323  3.216984e-09
## f.mpg=f.mpg-(62,101]            84.48582 24.4798356 22.74193548  1.033457e-08
## f.miles=f.miles-(6,18]          83.10864 26.9201130 25.42338710  2.515764e-06
## model=Mercedes- CL Class       100.00000  1.3357308  1.04838710  3.145337e-06
## model=Mercedes- B Class        100.00000  1.3357308  1.04838710  3.145337e-06
## model=BMW- X3                  100.00000  1.3100437  1.02822581  4.018860e-06
## model=BMW- X5                  100.00000  1.1045466  0.86693548  2.849082e-05
## f.mpg=f.mpg-(54,62]             82.71711 24.7110198 23.44758065  4.538207e-05
## model=VW- T-Cross              100.00000  1.0274852  0.80645161  5.933030e-05
## model=VW- Touareg              100.00000  0.8990496  0.70564516  2.012522e-04
## model=VW- Touran               100.00000  0.8476753  0.66532258  3.279107e-04
## f.price=f.price-[0,15]          81.53951 30.7474955 29.59677419  6.200007e-04
## model=Mercedes- SL CLASS       100.00000  0.7706139  0.60483871  6.817030e-04
## model=Mercedes- GLE Class      100.00000  0.7706139  0.60483871  6.817030e-04
## model=BMW- X2                  100.00000  0.7706139  0.60483871  6.817030e-04
## model=VW- Scirocco             100.00000  0.7449268  0.58467742  8.699476e-04
## fuelType=f.Fuel-Hybrid          94.64286  1.3614179  1.12903226  1.019501e-03
## model=VW- Golf SV              100.00000  0.6935525  0.54435484  1.416498e-03
## model=VW- Sharan               100.00000  0.6678654  0.52419355  1.807344e-03
## model=Mercedes- CLS Class      100.00000  0.6678654  0.52419355  1.807344e-03
## mout=NoMOut                     78.88514 95.9671205 95.48387097  2.560076e-03
## model=VW- Arteon               100.00000  0.6164911  0.48387097  2.941833e-03
## model=Mercedes- V Class        100.00000  0.6164911  0.48387097  2.941833e-03
## engineSize=Small                81.05039 29.3347033 28.40725806  5.317514e-03
## fuelType=f.Fuel-Diesel          79.86601 58.1813511 57.17741935  6.479754e-03
## model=BMW- X4                  100.00000  0.5137426  0.40322581  7.788991e-03
## model=VW- Caravelle            100.00000  0.3853070  0.30241935  2.627335e-02
## model=Mercedes- S Class        100.00000  0.3853070  0.30241935  2.627335e-02
## model=VW- Amarok               100.00000  0.3596198  0.28225806  3.350039e-02
## model=BMW- 7 Series            100.00000  0.3596198  0.28225806  3.350039e-02
## transmission=f.Trans-Automatic  80.47193 25.4045723 24.77822581  4.981989e-02
## model=Audi- SQ5                  0.00000  0.0000000  0.04032258  4.624298e-02
## model=Audi- S4                   0.00000  0.0000000  0.04032258  4.624298e-02
## model=Audi- S3                   0.00000  0.0000000  0.04032258  4.624298e-02
## model=Audi- RS5                  0.00000  0.0000000  0.04032258  4.624298e-02
## year=2012                       61.29032  0.4880555  0.62500000  3.007444e-02
## year=2011                       57.14286  0.3082456  0.42338710  2.957037e-02
## f.age=f.age-(5.1,15]            75.41568 16.3113280 16.97580645  1.858717e-02
## model=Audi- RS4                  0.00000  0.0000000  0.06048387  9.933194e-03
## f.miles=f.miles-(36,195]        75.52265 22.2707424 23.14516129  5.768199e-03
## mout=YesMOut                    70.08929  4.0328795  4.51612903  2.560076e-03
## model=Audi- RS6                  0.00000  0.0000000  0.08064516  2.132120e-03
## model=Audi- RS3                  0.00000  0.0000000  0.08064516  2.132120e-03
## f.price=f.price-(26,90]         75.18308 23.7349088 24.77822581  1.304424e-03
## fuelType=f.Fuel-Petrol          76.16054 40.4572309 41.69354839  7.751953e-04
## transmission=f.Trans-Manual     75.84270 34.6776265 35.88709677  7.468804e-04
## model=Audi- Q8                   0.00000  0.0000000  0.10080645  4.573131e-04
## model=Audi- A7                   0.00000  0.0000000  0.18145161  9.607418e-07
## model=Audi- A8                   0.00000  0.0000000  0.24193548  9.354395e-09
## f.mpg=f.mpg-[0,45]              71.10573 22.6303622 24.97983871  9.834443e-13
## engineSize=Medium               73.56052 48.2404315 51.47177419  2.332934e-18
## model=Audi- Q7                   0.00000  0.0000000  0.76612903  2.594801e-26
## model=Audi- TT                   0.00000  0.0000000  0.92741935  9.266585e-32
## model=Audi- A5                   0.00000  0.0000000  1.51209677  1.093130e-51
## model=Audi- A6                   0.00000  0.0000000  1.53225806  2.219825e-52
## model=Audi- Q5                   0.00000  0.0000000  1.55241935  4.504191e-53
```

```
## model=Audi- Q2                   0.00000  0.0000000  1.81451613  4.135500e-62
## model=Audi- A1                   0.00000  0.0000000  2.60080645  1.384331e-89
## model=Audi- Q3                   0.00000  0.0000000  2.88306452  1.384882e-99
## model=Audi- A4                   0.00000  0.0000000  2.96370968 1.865140e-102
## model=Audi- A3                   0.00000  0.0000000  4.01209677 2.421942e-140
## manufacturer=Audi                0.00000  0.0000000 21.51209677  0.000000e+00
##                                    v.test
## manufacturer=VW                 29.866992
## manufacturer=Mercedes           27.179044
## manufacturer=BMW                24.395710
## model=VW- Golf                  15.728020
## model=Mercedes- C Class         13.533726
## model=VW- Polo                  12.817707
## model=Mercedes- A Class         11.341637
## model=BMW- 3 Series             10.882107
## model=BMW- 1 Series              9.728531
## model=Mercedes- E Class          9.199157
## model=VW- Tiguan                 8.554359
## engineSize=Big                   8.096391
## model=BMW- 2 Series              7.510842
## model=BMW- 4 Series              7.005908
## model=BMW- 5 Series              6.722491
## model=Mercedes- GLC Class        6.686291
## model=BMW- X1                    6.159142
## model=Mercedes- GLA Class        6.119900
## model=VW- Passat                 6.080426
## model=VW- Up                     6.040716
## model=VW- T-Roc                  5.920124
## f.mpg=f.mpg-(62,101]             5.725145
## f.miles=f.miles-(6,18]           4.706848
## model=Mercedes- CL Class         4.661094
## model=Mercedes- B Class          4.661094
## model=BMW- X3                    4.610405
## model=BMW- X5                    4.185206
## f.mpg=f.mpg-(54,62]              4.078219
## model=VW- T-Cross                4.015459
## model=VW- Touareg                3.717439
## model=VW- Touran                 3.592190
## f.price=f.price-[0,15]           3.422711
## model=Mercedes- SL CLASS         3.396834
## model=Mercedes- GLE Class        3.396834
## model=BMW- X2                    3.396834
## model=VW- Scirocco               3.329523
## fuelType=f.Fuel-Hybrid           3.285089
## model=VW- Golf SV                3.191268
## model=VW- Sharan                 3.120190
## model=Mercedes- CLS Class        3.120190
## mout=NoMOut                      3.016149
## model=VW- Arteon                 2.973751
## model=Mercedes- V Class          2.973751
## engineSize=Small                 2.787142
## fuelType=f.Fuel-Diesel           2.722462
## model=BMW- X4                    2.661082
## model=VW- Caravelle              2.222147
## model=Mercedes- S Class          2.222147
## model=VW- Amarok                 2.126035
## model=BMW- 7 Series              2.126035
## transmission=f.Trans-Automatic   1.961507
## model=Audi- SQ5                 -1.993169
## model=Audi- S4                  -1.993169
## model=Audi- S3                  -1.993169
## model=Audi- RS5                 -1.993169
## year=2012                       -2.169109
## year=2011                       -2.175798
```

```
## f.age=f.age-(5.1,15]            -2.353709
## model=Audi- RS4                 -2.578146
## f.miles=f.miles-(36,195]        -2.760675
## mout=YesMOut                    -3.016149
## model=Audi- RS6                 -3.071184
## model=Audi- RS3                 -3.071184
## f.price=f.price-(26,90]         -3.215005
## fuelType=f.Fuel-Petrol          -3.361503
## transmission=f.Trans-Manual     -3.371765
## model=Audi- Q8                  -3.504589
## model=Audi- A7                  -4.899514
## model=Audi- A8                  -5.742038
## f.mpg=f.mpg-[0,45]              -7.132804
## engineSize=Medium               -8.739911
## model=Audi- Q7                 -10.612931
## model=Audi- TT                 -11.727018
## model=Audi- A5                 -15.125890
## model=Audi- A6                 -15.230472
## model=Audi- Q5                 -15.334399
## model=Audi- Q2                 -16.631296
## model=Audi- A1                 -20.068753
## model=Audi- Q3                 -21.182499
## model=Audi- A4                 -21.491614
## model=Audi- A3                 -25.219927
## manufacturer=Audi                    -Inf
##
## $`Audi Yes`
##                                   Cla/Mod     Mod/Cla      Global       p.value
## manufacturer=Audi              100.000000 100.0000000 21.51209677  0.000000e+00
## model=Audi- A3                 100.000000  18.6504217  4.01209677 2.421942e-140
## model=Audi- A4                 100.000000  13.7769447  2.96370968 1.865140e-102
## model=Audi- Q3                 100.000000  13.4020619  2.88306452  1.384882e-99
## model=Audi- A1                 100.000000  12.0899719  2.60080645  1.384331e-89
## model=Audi- Q2                 100.000000   8.4348641  1.81451613  4.135500e-62
## model=Audi- Q5                 100.000000   7.2164948  1.55241935  4.504191e-53
## model=Audi- A6                 100.000000   7.1227741  1.53225806  2.219825e-52
## model=Audi- A5                 100.000000   7.0290534  1.51209677  1.093130e-51
## model=Audi- TT                 100.000000   4.3111528  0.92741935  9.266585e-32
## model=Audi- Q7                 100.000000   3.5613871  0.76612903  2.594801e-26
## engineSize=Medium               26.439483  63.2614808 51.47177419  2.332934e-18
## f.mpg=f.mpg-[0,45]              28.894270  33.5520150 24.97983871  9.834443e-13
## model=Audi- A8                 100.000000   1.1246485  0.24193548  9.354395e-09
## model=Audi- A7                 100.000000   0.8434864  0.18145161  9.607418e-07
## model=Audi- Q8                 100.000000   0.4686036  0.10080645  4.573131e-04
## transmission=f.Trans-Manual     24.157303  40.2999063 35.88709677  7.468804e-04
## fuelType=f.Fuel-Petrol          23.839458  46.2043112 41.69354839  7.751953e-04
## f.price=f.price-(26,90]         24.816924  28.5848172 24.77822581  1.304424e-03
## model=Audi- RS6                100.000000   0.3748828  0.08064516  2.132120e-03
## model=Audi- RS3                100.000000   0.3748828  0.08064516  2.132120e-03
## mout=YesMOut                    29.910714   6.2792877  4.51612903  2.560076e-03
## f.miles=f.miles-(36,195]        24.477352  26.3355201 23.14516129  5.768199e-03
## model=Audi- RS4                100.000000   0.2811621  0.06048387  9.933194e-03
## f.age=f.age-(5.1,15]            24.584323  19.4001874 16.97580645  1.858717e-02
## year=2011                       42.857143   0.8434864  0.42338710  2.957037e-02
## year=2012                       38.709677   1.1246485  0.62500000  3.007444e-02
## model=Audi- SQ5                100.000000   0.1874414  0.04032258  4.624298e-02
## model=Audi- S4                 100.000000   0.1874414  0.04032258  4.624298e-02
## model=Audi- S3                 100.000000   0.1874414  0.04032258  4.624298e-02
## model=Audi- RS5                100.000000   0.1874414  0.04032258  4.624298e-02
## transmission=f.Trans-Automatic  19.528072  22.4929709 24.77822581  4.981989e-02
## model=VW- Amarok                 0.000000   0.0000000  0.28225806  3.350039e-02
## model=BMW- 7 Series              0.000000   0.0000000  0.28225806  3.350039e-02
## model=VW- Caravelle              0.000000   0.0000000  0.30241935  2.627335e-02
## model=Mercedes- S Class          0.000000   0.0000000  0.30241935  2.627335e-02
```

```
## model=BMW- X4                    0.000000   0.0000000  0.40322581  7.788991e-03
## fuelType=f.Fuel-Diesel          20.133992  53.5145267 57.17741935  6.479754e-03
## engineSize=Small                18.949610  25.0234302 28.40725806  5.317514e-03
## model=VW- Arteon                 0.000000   0.0000000  0.48387097  2.941833e-03
## model=Mercedes- V Class          0.000000   0.0000000  0.48387097  2.941833e-03
## mout=NoMOut                     21.114865  93.7207123 95.48387097  2.560076e-03
## model=VW- Sharan                 0.000000   0.0000000  0.52419355  1.807344e-03
## model=Mercedes- CLS Class        0.000000   0.0000000  0.52419355  1.807344e-03
## model=VW- Golf SV                0.000000   0.0000000  0.54435484  1.416498e-03
## fuelType=f.Fuel-Hybrid           5.357143   0.2811621  1.12903226  1.019501e-03
## model=VW- Scirocco               0.000000   0.0000000  0.58467742  8.699476e-04
## model=Mercedes- SL CLASS         0.000000   0.0000000  0.60483871  6.817030e-04
## model=Mercedes- GLE Class        0.000000   0.0000000  0.60483871  6.817030e-04
## model=BMW- X2                    0.000000   0.0000000  0.60483871  6.817030e-04
## f.price=f.price-[0,15]          18.460490  25.3983130 29.59677419  6.200007e-04
## model=VW- Touran                 0.000000   0.0000000  0.66532258  3.279107e-04
## model=VW- Touareg                0.000000   0.0000000  0.70564516  2.012522e-04
## model=VW- T-Cross                0.000000   0.0000000  0.80645161  5.933030e-05
## f.mpg=f.mpg-(54,62]             17.282889  18.8378632 23.44758065  4.538207e-05
## model=BMW- X5                    0.000000   0.0000000  0.86693548  2.849082e-05
## model=BMW- X3                    0.000000   0.0000000  1.02822581  4.018860e-06
## model=Mercedes- CL Class         0.000000   0.0000000  1.04838710  3.145337e-06
## model=Mercedes- B Class          0.000000   0.0000000  1.04838710  3.145337e-06
## f.miles=f.miles-(6,18]          16.891356  19.9625117 25.42338710  2.515764e-06
## f.mpg=f.mpg-(62,101]            15.514184  16.4011246 22.74193548  1.033457e-08
## model=VW- T-Roc                  0.000000   0.0000000  1.61290323  3.216984e-09
## model=VW- Up                     0.000000   0.0000000  1.67338710  1.534317e-09
## model=VW- Passat                 0.000000   0.0000000  1.69354839  1.198636e-09
## model=Mercedes- GLA Class        0.000000   0.0000000  1.71370968  9.363423e-10
## model=BMW- X1                    0.000000   0.0000000  1.73387097  7.314034e-10
## model=Mercedes- GLC Class        0.000000   0.0000000  2.01612903  2.288983e-11
## model=BMW- 5 Series              0.000000   0.0000000  2.03629032  1.786443e-11
## model=BMW- 4 Series              0.000000   0.0000000  2.19758065  2.453892e-12
## model=BMW- 2 Series              0.000000   0.0000000  2.50000000  5.874847e-14
## engineSize=Big                  12.525050  11.7150890 20.12096774  5.661366e-16
## model=VW- Tiguan                 0.000000   0.0000000  3.18548387  1.185258e-17
## model=Mercedes- E Class          0.000000   0.0000000  3.64919355  3.607674e-20
## model=BMW- 1 Series              0.000000   0.0000000  4.05241935  2.278596e-22
## model=BMW- 3 Series              0.000000   0.0000000  5.00000000  1.402828e-27
## model=Mercedes- A Class          0.000000   0.0000000  5.40322581  8.160316e-30
## model=VW- Polo                   0.000000   0.0000000  6.79435484  1.305024e-37
## model=Mercedes- C Class          0.000000   0.0000000  7.52016129  9.888214e-42
## model=VW- Golf                   0.000000   0.0000000  9.93951613  9.720293e-56
## manufacturer=BMW                 0.000000   0.0000000 21.81451613 1.899203e-131
## manufacturer=Mercedes            0.000000   0.0000000 26.16935484 1.149117e-162
## manufacturer=VW                  0.000000   0.0000000 30.50403226 5.282569e-196
##                                    v.test
## manufacturer=Audi                     Inf
## model=Audi- A3                 25.219927
## model=Audi- A4                 21.491614
## model=Audi- Q3                 21.182499
## model=Audi- A1                 20.068753
## model=Audi- Q2                 16.631296
## model=Audi- Q5                 15.334399
## model=Audi- A6                 15.230472
## model=Audi- A5                 15.125890
## model=Audi- TT                 11.727018
## model=Audi- Q7                 10.612931
## engineSize=Medium               8.739911
## f.mpg=f.mpg-[0,45]              7.132804
## model=Audi- A8                  5.742038
## model=Audi- A7                  4.899514
## model=Audi- Q8                  3.504589
## transmission=f.Trans-Manual     3.371765
```

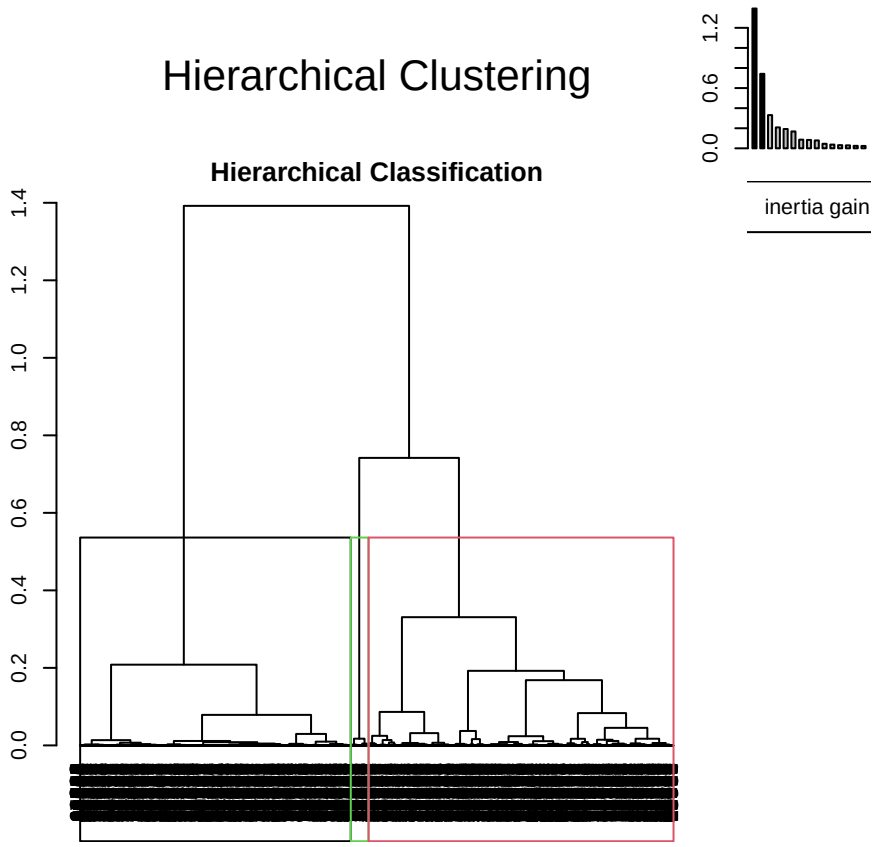
```
## fuelType=f.Fuel-Petrol           3.361503
## f.price=f.price-(26,90]          3.215005
## model=Audi- RS6                  3.071184
## model=Audi- RS3                  3.071184
## mout=YesMOut                     3.016149
## f.miles=f.miles-(36,195]         2.760675
## model=Audi- RS4                  2.578146
## f.age=f.age-(5.1,15]             2.353709
## year=2011                        2.175798
## year=2012                        2.169109
## model=Audi- SQ5                  1.993169
## model=Audi- S4                   1.993169
## model=Audi- S3                   1.993169
## model=Audi- RS5                  1.993169
## transmission=f.Trans-Automatic  -1.961507
## model=VW- Amarok                -2.126035
## model=BMW- 7 Series             -2.126035
## model=VW- Caravelle             -2.222147
## model=Mercedes- S Class         -2.222147
## model=BMW- X4                   -2.661082
## fuelType=f.Fuel-Diesel          -2.722462
## engineSize=Small                -2.787142
## model=VW- Arteon                -2.973751
## model=Mercedes- V Class         -2.973751
## mout=NoMOut                     -3.016149
## model=VW- Sharan                -3.120190
## model=Mercedes- CLS Class       -3.120190
## model=VW- Golf SV               -3.191268
## fuelType=f.Fuel-Hybrid          -3.285089
## model=VW- Scirocco              -3.329523
## model=Mercedes- SL CLASS        -3.396834
## model=Mercedes- GLE Class       -3.396834
## model=BMW- X2                   -3.396834
## f.price=f.price-[0,15]          -3.422711
## model=VW- Touran                -3.592190
## model=VW- Touareg               -3.717439
## model=VW- T-Cross               -4.015459
## f.mpg=f.mpg-(54,62]             -4.078219
## model=BMW- X5                   -4.185206
## model=BMW- X3                   -4.610405
## model=Mercedes- CL Class        -4.661094
## model=Mercedes- B Class         -4.661094
## f.miles=f.miles-(6,18]          -4.706848
## f.mpg=f.mpg-(62,101]            -5.725145
## model=VW- T-Roc                 -5.920124
## model=VW- Up                    -6.040716
## model=VW- Passat                -6.080426
## model=Mercedes- GLA Class       -6.119900
## model=BMW- X1                   -6.159142
## model=Mercedes- GLC Class       -6.686291
## model=BMW- 5 Series             -6.722491
## model=BMW- 4 Series             -7.005908
## model=BMW- 2 Series             -7.510842
## engineSize=Big                  -8.096391
## model=VW- Tiguan                -8.554359
## model=Mercedes- E Class         -9.199157
## model=BMW- 1 Series             -9.728531
## model=BMW- 3 Series            -10.882107
## model=Mercedes- A Class        -11.341637
## model=VW- Polo                 -12.817707
## model=Mercedes- C Class        -13.533726
## model=VW- Golf                 -15.728020
## manufacturer=BMW               -24.395710
## manufacturer=Mercedes          -27.179044
```

```
## manufacturer=VW                -29.866992
##
##
## Link between the cluster variable and the quantitative variables
## ================================================================
##                Eta2      P-value
## mpg     0.010872811 1.813506e-13
## tax     0.004710517 1.310622e-06
## price   0.003507012 3.001979e-05
## mileage 0.002222870 8.954549e-04
## age     0.001690095 3.781779e-03
## claKMMC 0.001318375 1.245694e-02
##
## Description of each cluster by quantitative variables
## ======================================================
## $`Audi No`
##            v.test Mean in category Overall mean sd in category   Overall sd
## mpg      7.342906        53.572357    52.950636      11.268286    11.388988
## claKMMC  2.526813         6.883030     6.822635       3.210074     3.215021
## age     -2.895027         3.725345     3.769041       2.011869     2.030232
## mileage -3.320122     22735.873207 23275.094475   21352.529529 21845.901298
## price   -4.170284     20776.133830 21080.082258    9789.536765  9803.729018
## tax     -4.833162       146.912383   147.330971      10.888491    11.649622
##              p.value
## mpg     2.090052e-13
## claKMMC 1.151029e-02
## age     3.791265e-03
## mileage 8.997799e-04
## price   3.042201e-05
## tax     1.343814e-06
##
## $`Audi Yes`
##            v.test Mean in category Overall mean sd in category   Overall sd
## tax      4.833162       148.858207   147.330971      13.976080    11.649622
## price    4.170284     22189.052484 21080.082258    9775.519370  9803.729018
## mileage  3.320122     25242.468792 23275.094475   23453.636518 21845.901298
## age      2.895027         3.928467     3.769041       2.088126     2.030232
## claKMMC -2.587378         6.597000     6.822635       3.223444     3.215021
## mpg     -7.342906        50.682254    52.950636      11.538249    11.388988
##              p.value
## tax     1.343814e-06
## price   3.042201e-05
## mileage 8.997799e-04
## age     3.791265e-03
## claKMMC 9.670937e-03
## mpg     2.090052e-13
```

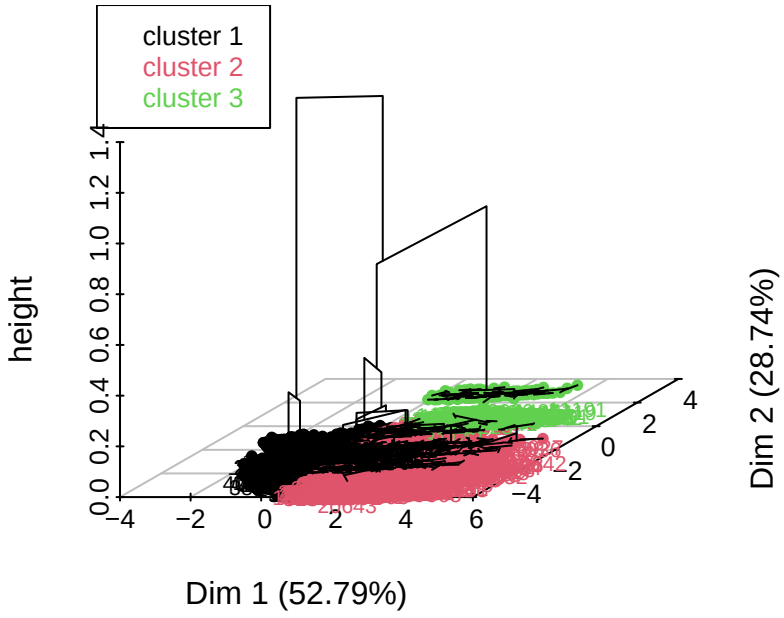
7 Hierarchical Clustering

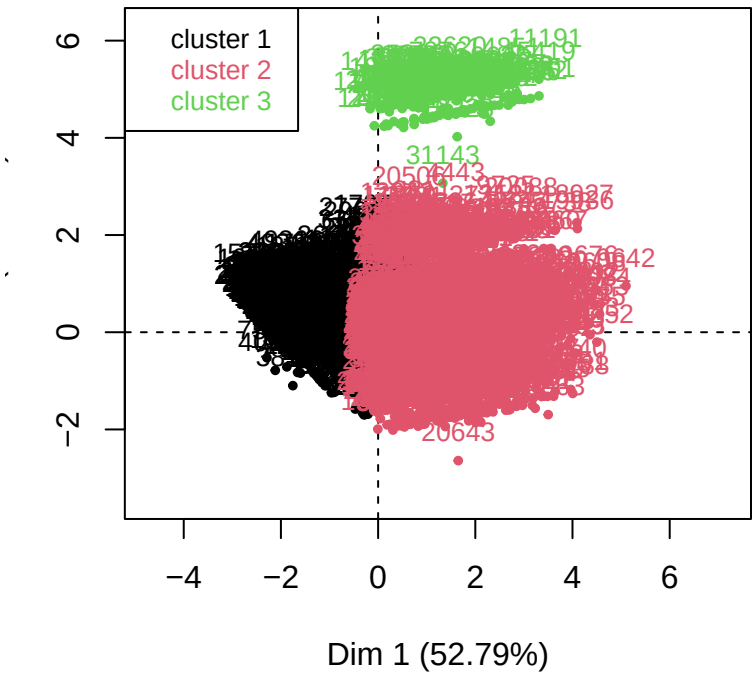
Per no parar l'execució, fiquem nb.clust = -1 per que ens esculli un tall a l'arbre automàticament a un nivell pròpiament suggerit. El resultat son 3 clusters caracteritzats per les variables categòriques year, f.price, f.miles, f.mpg i f.age. Concretament, el primer cluster esta descrit pels cotxes amb edat menor a 2 anys, amb menys de 6.000 milles i de l'any 2019. El segon per cotxes amb preus menors de 15.000 lliures i amb mileage entre 18.000 i 196.000. Per últim, el tercer cluster es caracteritza per cotxes amb una taxa de circulació entre 145 i 570, d'edat entre 5 i 15 anys i de l'any 2015. Les variables numèriques més característiques són el preu, la taxa, el mileage i l'edat. Pel cluster 1 en concret, ho és el preu, pel segon ho és la edat i la mileage i pel tercer la taxa. Podem veure també la dependencia dels cluster amb les dimensions del PCA, sent la Dim.1 la més important, i els individus qie millor descriuen els clausters.

```
res.hcpc<-HCPC(res.pca,order=TRUE, nb.clust = -1)
```



Hierarchical clustering on the factor map






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

```
##                     p.value  df
## year           0.000000e+00  28
## f.price        0.000000e+00   6
## f.miles        0.000000e+00   6
## f.mpg          0.000000e+00   6
## f.age          0.000000e+00   6
## f.tax         2.682761e-260   4
## model         3.295287e-142 164
## transmission   4.842692e-74   4
## fuelType       4.677492e-37   4
## engineSize     2.396488e-23   4
## manufacturer   5.259079e-08   6
## Audi           9.376726e-06   2
```

```
res.hcpc$desc.var$category
```

```
## $`1`
##                                      Cla/Mod     Mod/Cla     Global       p.value
## f.age=f.age-[0,2]                 99.9476166 76.68810289 40.3082770  0.000000e+00
## f.miles=f.miles-[0,6]             99.7668998 51.60771704 27.1748311  0.000000e+00
## year=year_2019                    99.9372647 64.02733119 33.6570946  0.000000e+00
## f.price=f.price-(26,90]           92.6047658 45.29742765 25.6967905 8.076314e-266
## f.tax=f.tax-(125,145]             70.6371191 81.99356913 60.9797297 2.268804e-221
## f.mpg=f.mpg-[0,45]                85.4700855 40.19292605 24.7043919 2.728150e-162
## f.miles=f.miles-(6,18]            81.1155378 40.91639871 26.4991554 1.959506e-131
## year=year_2020                   100.0000000 12.66077170  6.6511824  3.595948e-93
## f.price=f.price-(20,26]           75.5082285 31.35048232 21.8116554  2.220366e-65
## transmission=f.Trans-SemiAuto     64.0041494 49.59807074 40.7094595  1.579665e-39
## fuelType=f.Fuel-Petrol            63.4920635 51.44694534 42.5675676  6.194956e-39
## f.mpg=f.mpg-(45,54]               66.4992826 37.25884244 29.4341216  5.596219e-36
## year=year_2018                    76.4835165 13.98713826  9.6072635  2.475207e-28
## model=VW- T-Roc                   98.7500000  3.17524116  1.6891892  1.792539e-21
```

```
## model=VW- T-Cross              100.0000000  1.60771704  0.8445946  5.654495e-12
## model=Audi- Q2                  83.1460674  2.97427653  1.8792230  1.234127e-09
## model=BMW- X2                  100.0000000  1.20578778  0.6334459  3.774850e-09
## model=VW- Arteon                95.8333333  0.92443730  0.5067568  4.422986e-06
## model=Mercedes- V Class         95.8333333  0.92443730  0.5067568  4.422986e-06
## engineSize=Medium               55.4541684 53.73794212 50.9079392  4.171957e-05
## model=Mercedes- X-CLASS        100.0000000  0.48231511  0.2533784  4.362976e-04
## transmission=f.Trans-Automatic  56.5371025 25.72347267 23.9020270  1.974809e-03
## model=VW- Caravelle             92.3076923  0.48231511  0.2744932  3.156416e-03
## model=Mercedes- GLC Class       65.0000000  2.61254019  2.1114865  1.140256e-02
## model=Mercedes- C Class         58.9385475  8.48070740  7.5591216  1.149714e-02
## manufacturer=Mercedes           55.5023923 27.97427653 26.4780405  1.406415e-02
## model=Mercedes- GLE Class       73.3333333  0.88424437  0.6334459  2.201663e-02
## model=Mercedes- SL CLASS        74.0740741  0.80385852  0.5701014  2.464115e-02
## model=VW- Sharan                73.0769231  0.76366559  0.5489865  3.601886e-02
## model=VW- Touareg               70.9677419  0.88424437  0.6545608  3.977786e-02
## model=VW- Shuttle              100.0000000  0.20096463  0.1055743  3.993596e-02
## model=Mercedes- GLB Class      100.0000000  0.20096463  0.1055743  3.993596e-02
## model=Audi- Q8                 100.0000000  0.20096463  0.1055743  3.993596e-02
## model=BMW- X4                   75.0000000  0.60289389  0.4222973  4.526396e-02
## model=Audi- A3                  45.1612903  3.37620579  3.9273649  4.060539e-02
## model=Mercedes- CL Class        37.5000000  0.72347267  1.0135135  3.766786e-02
## model=Audi- TT                  35.7142857  0.60289389  0.8868243  2.979326e-02
## model=Mercedes- M Class         12.5000000  0.04019293  0.1689189  2.785667e-02
## model=BMW- 6 Series             12.5000000  0.04019293  0.1689189  2.785667e-02
## model=VW- Beetle                 0.0000000  0.00000000  0.1055743  2.403849e-02
## model=BMW- 1 Series             44.3298969  3.45659164  4.0962838  1.983672e-02
## manufacturer=VW                 49.9656829 29.26045016 30.7643581  1.840236e-02
## model=Mercedes- SLK              0.0000000  0.00000000  0.1266892  1.139681e-02
## model=BMW- X1                   38.8235294  1.32636656  1.7947635  1.100446e-02
## model=VW- Scirocco              25.9259259  0.28135048  0.5701014  5.766461e-03
## model=VW- CC                     0.0000000  0.00000000  0.1689189  2.559944e-03
## year=year_2011                   0.0000000  0.00000000  0.2111486  5.744729e-04
## model=VW- Golf                  44.7983015  8.48070740  9.9451014  4.049985e-04
## model=Audi- A1                  34.3750000  1.76848875  2.7027027  3.002904e-05
## model=VW- Up                    29.2682927  0.96463023  1.7314189  1.940469e-05
## engineSize=Big                  45.5913978 17.04180064 19.6368243  2.310225e-06
## year=year_2012                   0.0000000  0.00000000  0.4222973  3.223490e-07
## model=VW- Polo                  33.5347432  4.46141479  6.9890203  5.833965e-13
## f.age=f.age-(2,4.1]             40.0586941 21.94533762 28.7795608  7.495706e-28
## fuelType=f.Fuel-Diesel          44.5110528 47.74919614 56.3555743  1.944426e-36
## year=year_2013                   0.0000000  0.00000000  2.4070946  2.737991e-38
## f.price=f.price-(15,20]         36.0532889 17.40353698 25.3589527  3.162626e-40
## f.mpg=f.mpg-(54,62]             33.8365897 15.31350482 23.7753378  2.360178e-47
## transmission=f.Trans-Manual     36.6348449 24.67845659 35.3885135  1.527081e-59
## year=year_2014                   0.0000000  0.00000000  3.9484797  4.398273e-63
## f.tax=f.tax-[0,125]              3.0534351  0.32154341  5.5320946  1.037470e-73
## year=year_2017                  21.8302095  7.95819936 19.1511824  6.929497e-98
## year=year_2015                   1.1142061  0.16077170  7.5802365 1.127162e-114
## f.tax=f.tax-(145,570]           27.7427491 17.68488746 33.4881757 8.516885e-133
## f.mpg=f.mpg-(62,101]            17.2084130  7.23472669 22.0861486 1.100984e-156
## f.age=f.age-(4.1,5.1]            3.7711313  1.16559486 16.2373311 5.544864e-227
## year=year_2016                   3.6458333  1.12540193 16.2162162 3.367229e-228
## f.miles=f.miles-(18,36]         14.7853736  7.47588424 26.5625000 8.149600e-230
## f.age=f.age-(5.1,15]             0.7194245  0.20096463 14.6748311 1.630094e-241
## f.price=f.price-[0,15]          11.5175097  5.94855305 27.1326014 4.945122e-284
## f.miles=f.miles-(36,195]         0.0000000  0.00000000 19.7635135  0.000000e+00
##                                    v.test
## f.age=f.age-[0,2]                     Inf
## f.miles=f.miles-[0,6]                 Inf
## year=year_2019                        Inf
## f.price=f.price-(26,90]          34.831640
## f.tax=f.tax-(125,145]            31.760575
## f.mpg=f.mpg-[0,45]               27.147256
```

```
## f.miles=f.miles-(6,18]          24.394431
## year=year_2020                  20.475021
## f.price=f.price-(20,26]         17.076502
## transmission=f.Trans-SemiAuto   13.155628
## fuelType=f.Fuel-Petrol          13.051940
## f.mpg=f.mpg-(45,54]             12.522887
## year=year_2018                  11.039092
## model=VW- T-Roc                  9.516417
## model=VW- T-Cross                6.888094
## model=Audi- Q2                   6.075746
## model=BMW- X2                    5.893769
## model=VW- Arteon                 4.590447
## model=Mercedes- V Class          4.590447
## engineSize=Medium                4.097744
## model=Mercedes- X-CLASS          3.517095
## transmission=f.Trans-Automatic   3.093995
## model=VW- Caravelle              2.952080
## model=Mercedes- GLC Class        2.530114
## model=Mercedes- C Class          2.527214
## manufacturer=Mercedes            2.455621
## model=Mercedes- GLE Class        2.290081
## model=Mercedes- SL CLASS         2.246982
## model=VW- Sharan                 2.096714
## model=VW- Touareg                2.056048
## model=VW- Shuttle                2.054411
## model=Mercedes- GLB Class        2.054411
## model=Audi- Q8                   2.054411
## model=BMW- X4                    2.002193
## model=Audi- A3                  -2.047537
## model=Mercedes- CL Class        -2.078451
## model=Audi- TT                  -2.172828
## model=Mercedes- M Class         -2.199299
## model=BMW- 6 Series             -2.199299
## model=VW- Beetle                -2.256514
## model=BMW- 1 Series             -2.329422
## manufacturer=VW                 -2.357421
## model=Mercedes- SLK             -2.530290
## model=BMW- X1                   -2.542557
## model=VW- Scirocco              -2.760774
## model=VW- CC                    -3.016164
## year=year_2011                  -3.443389
## model=VW- Golf                  -3.536805
## model=Audi- A1                  -4.173246
## model=VW- Up                    -4.271633
## engineSize=Big                  -4.724200
## year=year_2012                  -5.109888
## model=VW- Polo                  -7.204307
## f.age=f.age-(2,4.1]            -10.939080
## fuelType=f.Fuel-Diesel         -12.606498
## year=year_2013                 -12.938249
## f.price=f.price-(15,20]        -13.276642
## f.mpg=f.mpg-(54,62]            -14.454152
## transmission=f.Trans-Manual    -16.273297
## year=year_2014                 -16.765025
## f.tax=f.tax-[0,125]            -18.161708
## year=year_2017                 -20.997386
## year=year_2015                 -22.760591
## f.tax=f.tax-(145,570]          -24.522428
## f.mpg=f.mpg-(62,101]           -26.668199
## f.age=f.age-(4.1,5.1]          -32.164463
## year=year_2016                 -32.251357
## f.miles=f.miles-(18,36]        -32.366425
## f.age=f.age-(5.1,15]           -33.187513
## f.price=f.price-[0,15]         -36.014588
```

```
## f.miles=f.miles-(36,195]            -Inf
##
## $`2`
##                                     Cla/Mod     Mod/Cla     Global
## f.price=f.price-[0,15]           86.92607004 53.06413302 27.1326014
## f.miles=f.miles-(36,195]         93.16239316 41.42517815 19.7635135
## f.miles=f.miles-(18,36]          79.80922099 47.69596200 26.5625000
## f.mpg=f.mpg-(62,101]             82.79158700 41.14014252 22.0861486
## year=year_2016                   88.93229167 32.44655582 16.2162162
## f.age=f.age-(4.1,5.1]            88.81664499 32.44655582 16.2373311
## f.age=f.age-(5.1,15]             89.78417266 29.64370546 14.6748311
## year=year_2017                   75.96471885 32.73159145 19.1511824
## f.tax=f.tax-[0,125]              96.94656489 12.06650831  5.5320946
## f.tax=f.tax-(145,570]            63.24085750 47.64845606 33.4881757
## transmission=f.Trans-Manual      61.99284010 49.35866983 35.3885135
## year=year_2015                   87.74373259 14.96437055  7.5802365
## f.mpg=f.mpg-(54,62]              66.16341030 35.39192399 23.7753378
## year=year_2014                   94.65240642  8.40855107  3.9484797
## fuelType=f.Fuel-Diesel           52.34170101 66.36579572 56.3555743
## f.age=f.age-(2,4.1]              58.47395451 37.86223278 28.7795608
## f.price=f.price-(15,20]          58.78434638 33.53919240 25.3589527
## year=year_2013                   88.59649123  4.79809976  2.4070946
## model=VW- Polo                   66.46525680 10.45130641  6.9890203
## model=VW- Up                     70.73170732  2.75534442  1.7314189
## model=Audi- A1                   63.28125000  3.84798100  2.7027027
## manufacturer=VW                  48.45573095 33.53919240 30.7643581
## year=year_2012                   85.00000000  0.80760095  0.4222973
## year=year_2011                  100.00000000  0.47505938  0.2111486
## model=VW- Golf                   52.01698514 11.63895487  9.9451014
## model=VW- CC                    100.00000000  0.38004751  0.1689189
## model=BMW- X1                    61.17647059  2.47030879  1.7947635
## engineSize=Small                 47.88530466 31.73396675 29.4552365
## model=VW- Scirocco               74.07407407  0.95011876  0.5701014
## model=VW- Beetle                100.00000000  0.23752969  0.1055743
## fuelType=f.Fuel-Hybrid           60.78431373  1.47268409  1.0768581
## model=BMW- 1 Series              52.06185567  4.79809976  4.0962838
## model=Audi- A6                   57.81250000  1.75771971  1.3513514
## model=Mercedes- A Class          50.75187970  6.41330166  5.6165541
## engineSize=Big                   47.52688172 20.99762470 19.6368243
## model=Mercedes- GL Class         85.71428571  0.28503563  0.1478041
## model=Audi- A3                   51.61290323  4.56057007  3.9273649
## model=Mercedes- S Class          15.38461538  0.09501188  0.2744932
## model=Audi- Q7                   26.47058824  0.42755344  0.7179054
## model=Mercedes- GLC Class        34.00000000  1.61520190  2.1114865
## model=Mercedes- GLE Class        16.66666667  0.23752969  0.6334459
## model=Mercedes- X-CLASS           0.00000000  0.00000000  0.2533784
## model=Audi- A8                    0.00000000  0.00000000  0.2533784
## model=VW- Caravelle               0.00000000  0.00000000  0.2744932
## model=BMW- X4                     5.00000000  0.04750594  0.4222973
## model=BMW- X5                    14.63414634  0.28503563  0.8657095
## model=VW- Arteon                  4.16666667  0.04750594  0.5067568
## model=Mercedes- V Class           4.16666667  0.04750594  0.5067568
## engineSize=Medium                41.26918291 47.26840855 50.9079392
## transmission=f.Trans-Automatic   38.42756184 20.66508314 23.9020270
## model=Audi- Q5                   15.06849315  0.52256532  1.5413851
## model=Audi- Q2                   16.85393258  0.71258907  1.8792230
## model=BMW- X2                     0.00000000  0.00000000  0.6334459
## model=VW- T-Cross                 0.00000000  0.00000000  0.8445946
## model=VW- T-Roc                   1.25000000  0.04750594  1.6891892
## year=year_2018                   23.51648352  5.08313539  9.6072635
## fuelType=f.Fuel-Petrol           33.58134921 32.16152019 42.5675676
## transmission=f.Trans-SemiAuto    32.72821577 29.97624703 40.7094595
## f.mpg=f.mpg-(45,54]              27.90530846 18.47980998 29.4341216
## f.price=f.price-(20,26]          21.10358180 10.35629454 21.8116554
```

```
## year=year_2020                   0.00000000  0.00000000  6.6511824
## f.miles=f.miles-(6,18]          18.00796813 10.73634204 26.4991554
## f.tax=f.tax-(125,145]           29.36288089 40.28503563 60.9797297
## f.mpg=f.mpg-[0,45]               8.97435897  4.98812352 24.7043919
## f.price=f.price-(26,90]          5.25883320  3.04038005 25.6967905
## f.age=f.age-[0,2]                0.05238345  0.04750594 40.3082770
## f.miles=f.miles-[0,6]            0.23310023  0.14251781 27.1748311
## year=year_2019                   0.06273526  0.04750594 33.6570946
##                                      p.value     v.test
## f.price=f.price-[0,15]         3.793238e-301  37.091912
## f.miles=f.miles-(36,195]       1.832297e-273  35.332952
## f.miles=f.miles-(18,36]        2.653409e-197  29.966863
## f.mpg=f.mpg-(62,101]           1.090547e-183  28.903130
## year=year_2016                 1.779808e-174  28.160551
## f.age=f.age-(4.1,5.1]          1.022876e-173  28.098464
## f.age=f.age-(5.1,15]           3.287816e-162  27.140390
## year=year_2017                 3.954209e-102  21.456696
## f.tax=f.tax-[0,125]             3.161390e-81  19.088258
## f.tax=f.tax-(145,570]           2.086493e-76  18.499435
## transmission=f.Trans-Manual     1.560797e-72  18.012275
## year=year_2015                  7.304928e-71  17.798145
## f.mpg=f.mpg-(54,62]             1.999455e-63  16.811818
## year=year_2014                  3.193301e-51  15.055159
## fuelType=f.Fuel-Diesel          9.678191e-36  12.479343
## f.age=f.age-(2,4.1]             6.353800e-35  12.328604
## f.price=f.price-(15,20]         7.519671e-31  11.548412
## year=year_2013                  2.279143e-23   9.960065
## model=VW- Polo                  7.039050e-17   8.346374
## model=VW- Up                    1.435301e-06   4.820039
## model=Audi- A1                  1.547942e-05   4.321753
## manufacturer=VW                 2.199914e-04   3.694879
## year=year_2012                  2.529084e-04   3.659297
## year=year_2011                  2.973276e-04   3.617617
## model=VW- Golf                  5.213714e-04   3.469529
## model=VW- CC                    1.511849e-03   3.172399
## model=BMW- X1                   1.901273e-03   3.105236
## engineSize=Small                2.124023e-03   3.072319
## model=VW- Scirocco              2.135645e-03   3.070690
## model=VW- Beetle                1.730034e-02   2.380262
## fuelType=f.Fuel-Hybrid          1.972475e-02   2.331543
## model=BMW- 1 Series             3.018208e-02   2.167693
## model=Audi- A6                  3.211588e-02   2.142965
## model=Mercedes- A Class         3.394800e-02   2.120689
## engineSize=Big                  3.530796e-02   2.104809
## model=Mercedes- GL Class        3.671610e-02   2.088907
## model=Audi- A3                  4.595219e-02   1.995832
## model=Mercedes- S Class         3.469314e-02  -2.111922
## model=Audi- Q7                  3.352944e-02  -2.125686
## model=Mercedes- GLC Class       3.294029e-02  -2.132810
## model=Mercedes- GLE Class       1.573666e-03  -3.160744
## model=Mercedes- X-CLASS         8.543909e-04  -3.334545
## model=Audi- A8                  8.543909e-04  -3.334545
## model=VW- Caravelle             4.736769e-04  -3.495217
## model=BMW- X4                   1.374984e-04  -3.812625
## model=BMW- X5                   5.646438e-05  -4.027119
## model=VW- Arteon                1.522314e-05  -4.325435
## model=Mercedes- V Class         1.522314e-05  -4.325435
## engineSize=Medium               7.438403e-06  -4.480742
## transmission=f.Trans-Automatic  2.763907e-06  -4.687627
## model=Audi- Q5                  8.900526e-08  -5.347850
## model=Audi- Q2                  3.226319e-08  -5.528689
## model=BMW- X2                   2.037853e-08  -5.608757
## model=VW- T-Cross               5.378857e-11  -6.560053
## model=VW- T-Roc                 1.491490e-19  -9.045377
```

```
## year=year_2018                   2.405393e-22  -9.723020
## fuelType=f.Fuel-Petrol           9.182236e-39 -13.021928
## transmission=f.Trans-SemiAuto    1.002522e-41 -13.532715
## f.mpg=f.mpg-(45,54]              5.852338e-51 -15.015044
## f.price=f.price-(20,26]          2.526340e-69 -17.598567
## year=year_2020                   5.159611e-85 -19.538582
## f.miles=f.miles-(6,18]          6.578795e-115 -22.784190
## f.tax=f.tax-(125,145]           8.344995e-153 -26.331573
## f.mpg=f.mpg-[0,45]              1.490814e-199 -30.139091
## f.price=f.price-(26,90]         2.754511e-266 -34.862483
## f.age=f.age-[0,2]                0.000000e+00       -Inf
## f.miles=f.miles-[0,6]            0.000000e+00       -Inf
## year=year_2019                   0.000000e+00       -Inf
##
## $`3`
##                                    Cla/Mod     Mod/Cla     Global      p.value
## f.tax=f.tax-(145,570]            9.0163934 100.0000000 33.4881757 1.357012e-70
## f.age=f.age-(5.1,15]             9.4964029  46.1538462 14.6748311 5.936005e-20
## year=year_2015                  11.1420613  27.9720280  7.5802365 8.669210e-14
## model=Audi- Q5                  26.0273973  13.2867133  1.5413851 1.691773e-13
## engineSize=Big                   6.8817204  44.7552448 19.6368243 4.560477e-12
## f.miles=f.miles-(36,195]         6.8376068  44.7552448 19.7635135 6.188781e-12
## year=year_2016                   7.4218750  39.8601399 16.2162162 6.368278e-12
## f.age=f.age-(4.1,5.1]            7.4122237  39.8601399 16.2373311 6.735425e-12
## f.mpg=f.mpg-(45,54]              5.5954089  54.5454545 29.4341216 2.256687e-10
## model=Mercedes- M Class         75.0000000   4.1958042  0.1689189 1.835548e-08
## f.mpg=f.mpg-[0,45]               5.5555556  45.4545455 24.7043919 4.519104e-08
## f.miles=f.miles-(18,36]          5.4054054  47.5524476 26.5625000 5.282307e-08
## model=BMW- X5                   24.3902439   6.9930070  0.8657095 2.502807e-07
## f.price=f.price-(15,20]          5.1623647  43.3566434 25.3589527 2.089372e-06
## model=Audi- Q7                  23.5294118   5.5944056  0.7179054 5.752168e-06
## manufacturer=Audi                5.3000000  37.0629371 21.1148649 9.344095e-06
## Audi=Audi Yes                    5.3000000  37.0629371 21.1148649 9.344095e-06
## transmission=f.Trans-Automatic   5.0353357  39.8601399 23.9020270 1.777940e-05
## model=Mercedes- S Class         38.4615385   3.4965035  0.2744932 2.576324e-05
## year=year_2013                  11.4035088   9.0909091  2.4070946 3.814711e-05
## model=BMW- 6 Series             37.5000000   2.0979021  0.1689189 1.401266e-03
## model=BMW- X4                   20.0000000   2.7972028  0.4222973 2.900047e-03
## model=Audi- A8                  25.0000000   2.0979021  0.2533784 5.180693e-03
## year=year_2012                  15.0000000   2.0979021  0.4222973 2.373554e-02
## model=BMW- X6                   20.0000000   1.3986014  0.2111486 3.752259e-02
## f.price=f.price-(26,90]          2.1364010  18.1818182 25.6967905 3.261538e-02
## manufacturer=Mercedes            2.1531100  18.8811189 26.4780405 3.258995e-02
## f.tax=f.tax-[0,125]              0.0000000   0.0000000  5.5320946 2.570268e-04
## model=Mercedes- C Class          0.2793296   0.6993007  7.5591216 1.547287e-04
## f.price=f.price-[0,15]           1.5564202  13.9860140 27.1326014 1.490966e-04
## manufacturer=VW                  1.5785861  16.0839161 30.7643581 4.928310e-05
## year=year_2020                   0.0000000   0.0000000  6.6511824 4.547923e-05
## model=VW- Polo                   0.0000000   0.0000000  6.9890203 2.684988e-05
## f.age=f.age-(2,4.1]              1.4673514  13.9860140 28.7795608 2.591763e-05
## Audi=Audi No                     2.4089936  62.9370629 78.8851351 9.344095e-06
## year=year_2018                   0.0000000   0.0000000  9.6072635 4.226686e-07
## transmission=f.Trans-Manual      1.3723150  16.0839161 35.3885135 2.260859e-07
## f.miles=f.miles-(6,18]           0.8764940   7.6923077 26.4991554 9.172428e-09
## f.mpg=f.mpg-(62,101]             0.0000000   0.0000000 22.0861486 1.701268e-16
## f.mpg=f.mpg-(54,62]              0.0000000   0.0000000 23.7753378 6.955704e-18
## f.miles=f.miles-[0,6]            0.0000000   0.0000000 27.1748311 8.919438e-21
## engineSize=Small                 0.0000000   0.0000000 29.4552365 8.551168e-23
## year=year_2019                   0.0000000   0.0000000 33.6570946 1.076585e-26
## f.age=f.age-[0,2]                0.0000000   0.0000000 40.3082770 2.039885e-33
## f.tax=f.tax-(125,145]            0.0000000   0.0000000 60.9797297 1.109643e-60
##                                    v.test
## f.tax=f.tax-(145,570]            17.763423
## f.age=f.age-(5.1,15]              9.145490
```

```
## year=year_2015                    7.459742
## model=Audi- Q5                    7.371138
## engineSize=Big                    6.918625
## f.miles=f.miles-(36,195]          6.875235
## year=year_2016                    6.871158
## f.age=f.age-(4.1,5.1]             6.863159
## f.mpg=f.mpg-(45,54]               6.342771
## model=Mercedes- M Class           5.626827
## f.mpg=f.mpg-[0,45]                5.469263
## f.miles=f.miles-(18,36]           5.441536
## model=BMW- X5                     5.157491
## f.price=f.price-(15,20]           4.744582
## model=Audi- Q7                    4.535300
## manufacturer=Audi                 4.431820
## Audi=Audi Yes                     4.431820
## transmission=f.Trans-Automatic    4.291095
## model=Mercedes- S Class           4.208007
## year=year_2013                    4.118425
## model=BMW- 6 Series               3.194390
## model=BMW- X4                     2.978139
## model=Audi- A8                    2.795578
## year=year_2012                    2.261383
## model=BMW- X6                     2.080032
## f.price=f.price-(26,90]          -2.136786
## manufacturer=Mercedes            -2.137099
## f.tax=f.tax-[0,125]              -3.655155
## model=Mercedes- C Class          -3.783355
## f.price=f.price-[0,15]           -3.792569
## manufacturer=VW                  -4.059001
## year=year_2020                   -4.077721
## model=VW- Polo                   -4.198660
## f.age=f.age-(2,4.1]              -4.206657
## Audi=Audi No                     -4.431820
## year=year_2018                   -5.058454
## transmission=f.Trans-Manual      -5.176501
## f.miles=f.miles-(6,18]           -5.745362
## f.mpg=f.mpg-(62,101]             -8.241456
## f.mpg=f.mpg-(54,62]              -8.615632
## f.miles=f.miles-[0,6]            -9.348150
## engineSize=Small                 -9.827745
## year=year_2019                  -10.694794
## f.age=f.age-[0,2]               -12.045840
## f.tax=f.tax-(125,145]           -16.433035
```

```
res.hcpc$desc.var$quanti.var
```

```
##              Eta2 P-value
## price   0.3675013       0
## mileage 0.6025290       0
## tax     0.7078747       0
## mpg     0.3420681       0
## age     0.6478671       0
```

```
res.hcpc$desc.var$quanti
```

```
## $`1`
##              v.test Mean in category Overall mean sd in category   Overall sd
## price    41.122088     26943.987942 21486.968961   9245.5118709  9606.547493
## tax      -7.395706       145.891474   147.045162      2.8064999    11.292649
## mpg     -36.808808        47.256917    52.924126      9.4600903    11.145627
## mileage -53.396398      7425.900198 20567.651340   6268.5175085 17816.742468
## age     -55.275648         2.215016     3.595831      0.8204008     1.808375
```

```
##               p.value
## price    0.000000e+00
## tax      1.406588e-13
## mpg     1.334749e-296
## mileage  0.000000e+00
## age      0.000000e+00
##
## $`2`
##            v.test Mean in category Overall mean sd in category   Overall sd
## age      50.83022         5.089262     3.595831       1.302673     1.808375
## mileage  49.79245     34981.042779 20567.651340   14941.488164 17816.742468
## mpg      40.14386        60.193516    52.924126       8.780995    11.145627
## tax     -12.47095       144.757090   147.045162       8.572586    11.292649
## price   -41.06711     15077.295962 21486.968961    5308.345441  9606.547493
##               p.value
## age     0.000000e+00
## mileage 0.000000e+00
## mpg     0.000000e+00
## tax     1.075422e-35
## price   0.000000e+00
##
## $`3`
##            v.test Mean in category Overall mean sd in category   Overall sd
## tax      57.795181       200.798776   147.045162       3.867212    11.292649
## age      13.700446         5.636364     3.595831       1.203407     1.808375
## mileage  11.230023     37046.587413 20567.651340   14447.377233 17816.742468
## mpg      -9.157181        44.518182    52.924126       3.844543    11.145627
##               p.value
## tax     0.000000e+00
## age     1.009029e-42
## mileage 2.904022e-29
## mpg     5.327086e-20
```

```
res.hcpc$desc.axes$quanti.var
```

```
##               Eta2       P-value
## Dim.1 0.731289430  0.000000e+00
## Dim.2 0.611712476  0.000000e+00
## Dim.3 0.092437457 2.064844e-100
## Dim.4 0.003369622  3.396222e-04
```

```
res.hcpc$desc.axes$quanti
```

```
## $`1`
##           v.test Mean in category  Overall mean sd in category Overall sd
## Dim.4   3.393242       0.01957642  1.020639e-13      0.2555618  0.4176432
## Dim.1 -58.844236      -1.18124799 -4.853850e-14      0.6039866  1.4531959
##            p.value
## Dim.4 0.0006907067
## Dim.1 0.0000000000
##
## $`2`
##           v.test Mean in category  Overall mean sd in category Overall sd
## Dim.1  55.336813       1.30650991 -4.853850e-14      0.8933917  1.4531959
## Dim.4  -3.906177      -0.02650524  1.020639e-13      0.5412479  0.4176432
## Dim.3  -6.973596      -0.08509562  2.135454e-13      0.7907791  0.7510624
## Dim.2 -17.027757      -0.29663628  2.418939e-13      0.8382962  1.0722411
##            p.value
## Dim.1 0.000000e+00
## Dim.4 9.376775e-05
## Dim.3 3.089407e-12
## Dim.2 5.112660e-65
```

```
##
## $`3`
##         v.test Mean in category  Overall mean sd in category Overall sd
## Dim.2 53.10161         4.689433  2.418939e-13      0.2861708  1.0722411
## Dim.3 20.57803         1.272917  2.135454e-13      0.4984770  0.7510624
## Dim.1 11.02776         1.319872 -4.853850e-14      0.8352055  1.4531959
##            p.value
## Dim.2 0.000000e+00
## Dim.3 4.319105e-94
## Dim.1 2.807716e-28
```

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

```
## Cluster: 1
##     28347     37614     37757     36515      8526
## 0.1434348 0.1434468 0.1444324 0.1444727 0.1489091
## ------------------------------------------------------------
## Cluster: 2
##     48854     44075      5925     40884     41563
## 0.2198185 0.2231591 0.2259935 0.2260200 0.2315994
## ------------------------------------------------------------
## Cluster: 3
##     48053     18079     23196      3484     20348
## 0.2417919 0.3708227 0.4035927 0.4127184 0.4298707
```

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

```
## Cluster: 1
##    15290    26643    26743    23733    16173
## 4.252626 4.163025 4.129496 4.129310 4.118647
## ------------------------------------------------------------
## Cluster: 2
##    20084     7892    44340    20091    18188
## 5.979029 5.951906 5.948536 5.831193 5.800116
## ------------------------------------------------------------
## Cluster: 3
##    11191    45419     9625    14851    35336
## 6.201451 5.934944 5.862110 5.809624 5.707131
```

8 CA analysis

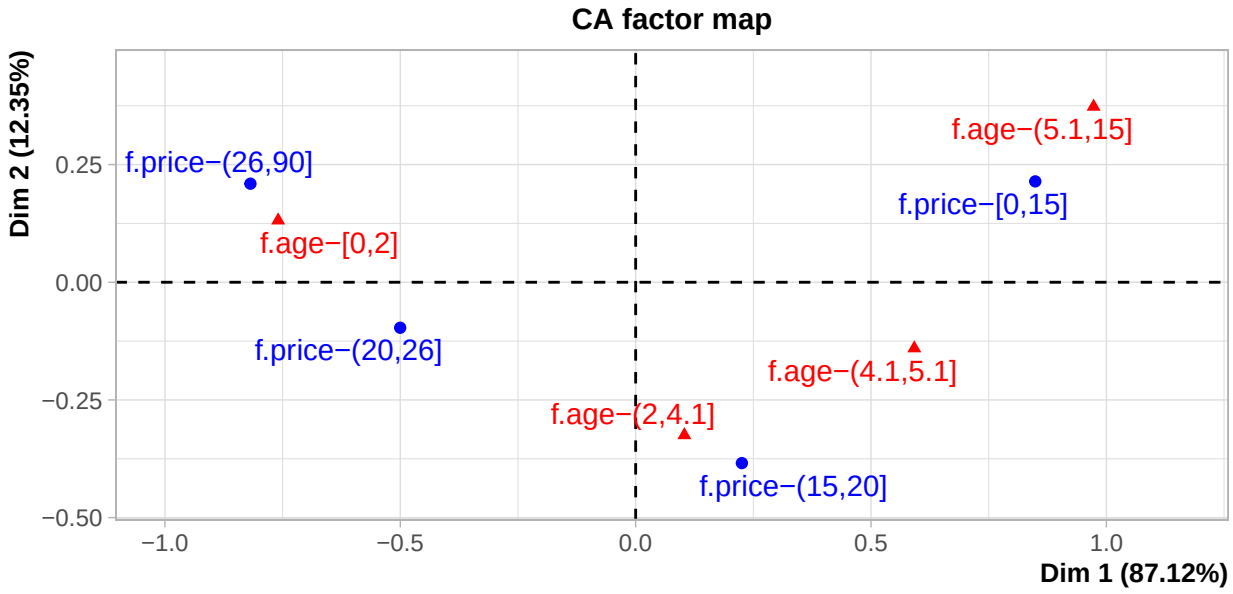
8.1 Eigenvalues and dominant axes analysis

Farem CA sobre les variables f.price - f.age i sobre f.price - fuelType. Podem veure la taula de contingència i fer el test de Chi quadrat. Per f.price-f.age demostrem que les variables no són independents. Per f.price - fuelType podem asegurar el mateix. En els dos casos ens quedem amb tots els 3 eixos.

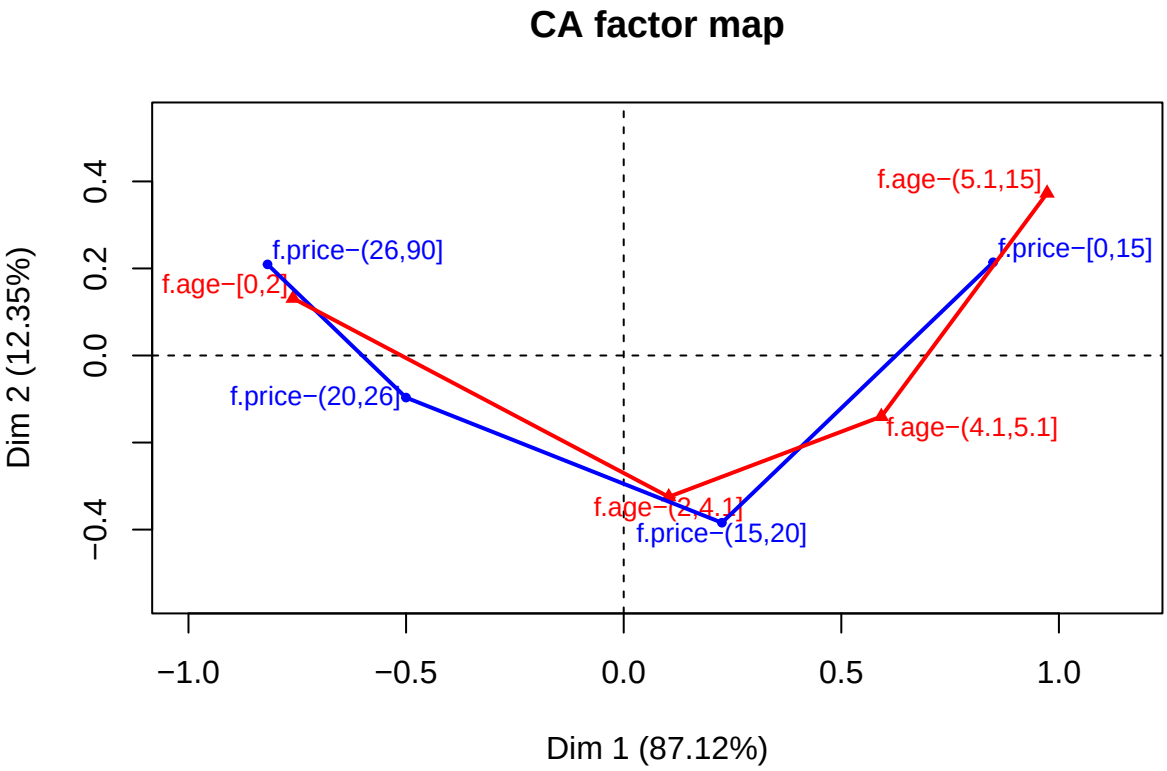
```
par(mfrow=c(1,1))

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

```
##                  f.age
## f.price           f.age-[0,2] f.age-(2,4.1] f.age-(4.1,5.1] f.age-(5.1,15]
##   f.price-[0,15]           81           353             395            639
##   f.price-(15,20]         259           512             305            152
##   f.price-(20,26]         600           327              74             34
##   f.price-(26,90]         970           191              51             17
```

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




```
plot( res.ca, cex=0.8, graph.type = "classic" )
lines( res.ca$row$coord[,1], res.ca$row$coord[,2], col="blue", lwd = 2 )
lines( res.ca$col$coord[,1], res.ca$col$coord[,2], col="red", lwd = 2 )
```



```
summary(res.ca,dig=2)
```

```
##
## Call:
## CA(X = tt)
##
## The chi square of independence between the two variables is equal to 2528.084 (p-value =  0 ).
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3
## Variance               0.444   0.063   0.003
## % of var.             87.124  12.349   0.527
## Cumulative % of var.  87.124  99.473 100.000
##
## Rows
##                    Iner*1000     Dim.1     ctr    cos2     Dim.2     ctr    cos2
## f.price-[0,15]  |   226.920 |   0.849  48.036   0.940 |   0.214  21.585   0.060
## f.price-(15,20] |    49.541 |   0.226   2.840   0.255 |  -0.384  58.060   0.738
## f.price-(20,26] |    55.880 |  -0.500  11.756   0.934 |  -0.097   3.098   0.035
## f.price-(26,90] |   177.354 |  -0.818  37.368   0.936 |   0.209  17.257   0.061
##                      Dim.3     ctr    cos2
## f.price-[0,15]  |  -0.008   0.782   0.000 |
## f.price-(15,20] |   0.039  14.342   0.008 |
## f.price-(20,26] |  -0.091  64.279   0.031 |
## f.price-(26,90] |   0.047  20.597   0.003 |
##
## Columns
##                    Iner*1000     Dim.1     ctr    cos2     Dim.2     ctr    cos2
## f.age-[0,2]     |   228.795 |  -0.759  50.021   0.971 |   0.131  10.555   0.029
## f.age-(2,4.1]   |    33.065 |   0.103   0.673   0.090 |  -0.325  46.735   0.890
## f.age-(4.1,5.1] |    63.299 |   0.592  13.124   0.921 |  -0.140   5.197   0.052
## f.age-(5.1,15]  |   184.534 |   0.973  36.182   0.871 |   0.373  37.514   0.128
##                      Dim.3     ctr    cos2
## f.age-[0,2]     |   0.008   0.916   0.000 |
## f.age-(2,4.1]   |  -0.049  24.710   0.020 |
## f.age-(4.1,5.1] |   0.102  65.046   0.028 |
## f.age-(5.1,15]  |  -0.038   9.328   0.001 |
```

```
chisq.test(tt)
```

```
##
##  Pearson's Chi-squared test
##
## data:  tt
## X-squared = 2528.1, df = 9, p-value < 2.2e-16
```

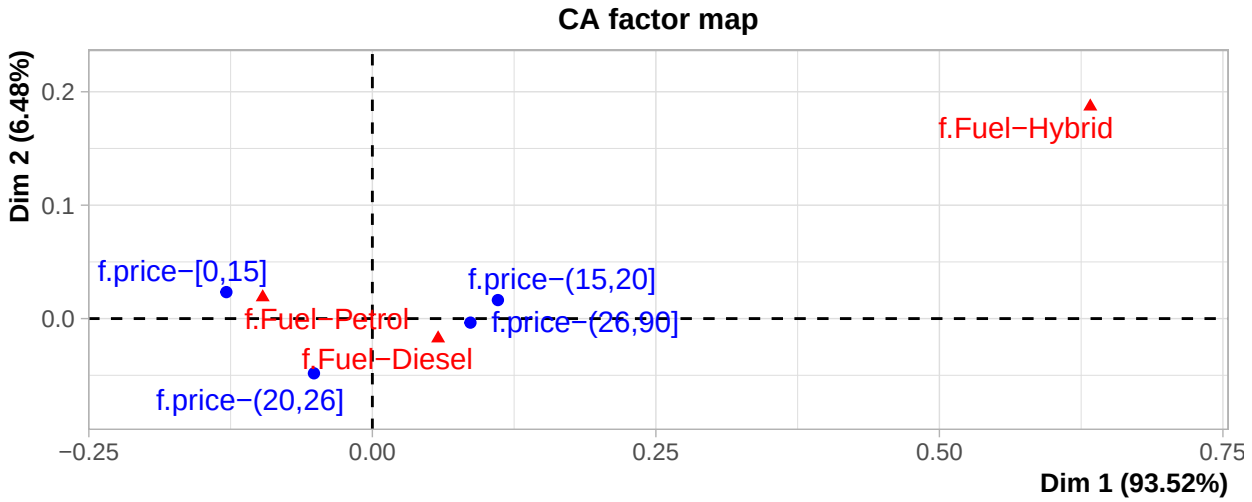
```
res.ca$eig
```

```
##        eigenvalue percentage of variance cumulative percentage of variance
## dim 1 0.444067396             87.1242487                           87.12425
## dim 2 0.062940972             12.3487672                           99.47302
## dim 3 0.002686008              0.5269841                          100.00000
```

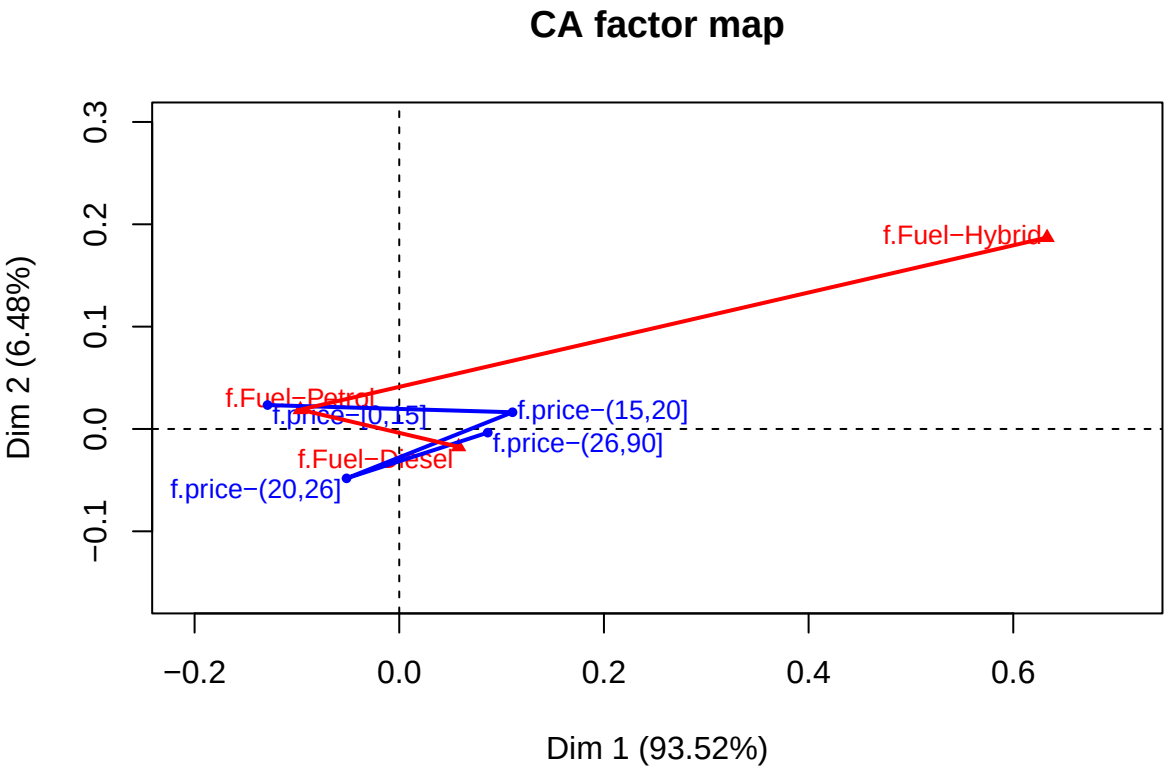
```
tt<-table(df[,c("f.price","fuelType")])
tt
```

```
##                  fuelType
## f.price           f.Fuel-Diesel f.Fuel-Petrol f.Fuel-Hybrid
##   f.price-[0,15]            765           697             6
##   f.price-(15,20]           739           464            25
##   f.price-(20,26]           593           438             4
##   f.price-(26,90]           739           469            21
```

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




```
plot( res.ca, cex=0.8, graph.type = "classic" )
lines( res.ca$row$coord[,1], res.ca$row$coord[,2], col="blue", lwd = 2 )
lines( res.ca$col$coord[,1], res.ca$col$coord[,2], col="red", lwd = 2 )
```



```
summary(res.ca)
```

```
##
## Call:
## CA(X = tt)
##
## The chi square of independence between the two variables is equal to 54.86447 (p-value =  4.936962e-1
##
## Eigenvalues
##                        Dim.1   Dim.2
## Variance               0.010   0.001
## % of var.             93.524   6.476
## Cumulative % of var.  93.524 100.000
##
## Rows
##                    Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## f.price-[0,15]   |     5.065 | -0.129 47.400  0.968 |  0.023 22.525  0.032 |
## f.price-(15,20] |     3.100 |  0.111 29.333  0.979 |  0.016  9.212  0.021 |
## f.price-(20,26] |     1.039 | -0.051  5.348  0.532 | -0.048 67.822  0.468 |
## f.price-(26,90] |     1.857 |  0.086 17.919  0.998 | -0.004  0.440  0.002 |
##
## Columns
##                    Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## f.Fuel-Diesel    |     2.097 |  0.058 18.591  0.917 | -0.017 24.232  0.083 |
## f.Fuel-Petrol    |     4.045 | -0.097 37.670  0.963 |  0.019 20.637  0.037 |
## f.Fuel-Hybrid    |     4.920 |  0.633 43.740  0.920 |  0.187 55.131  0.080 |
```

```
chisq.test(tt)
```

```
##
##  Pearson's Chi-squared test
##
## data:  tt
## X-squared = 54.864, df = 6, p-value = 4.937e-10
```

```
res.ca$eig
```

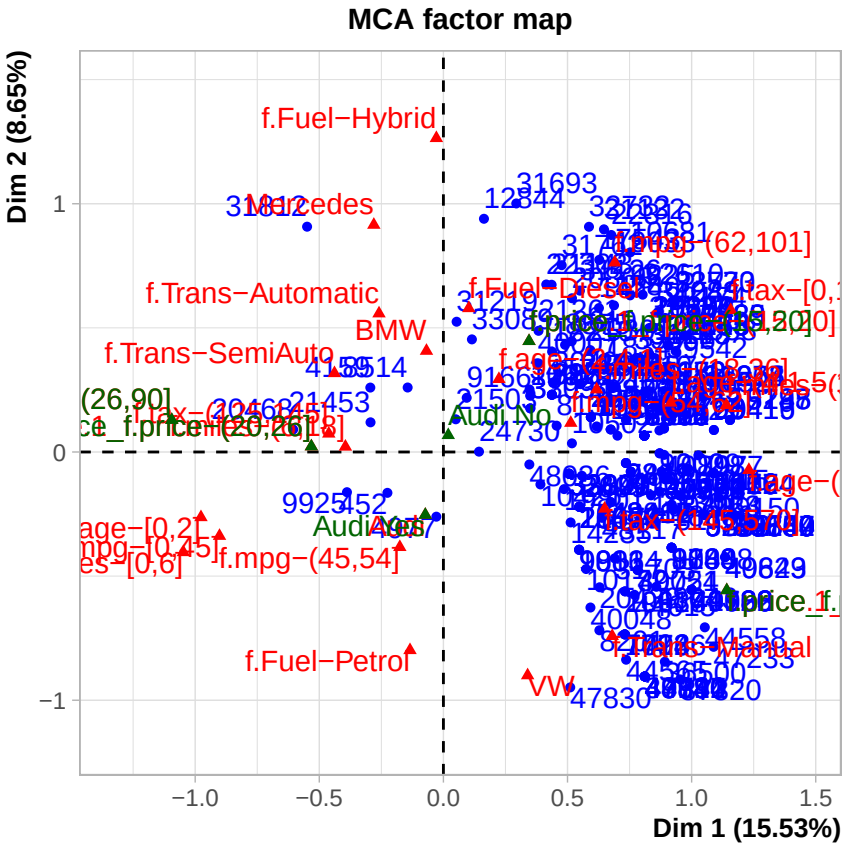
```
##         eigenvalue percentage of variance cumulative percentage of variance
## dim 1 0.0103450057              93.523594                           93.52359
## dim 2 0.0007163802               6.476406                          100.00000
```

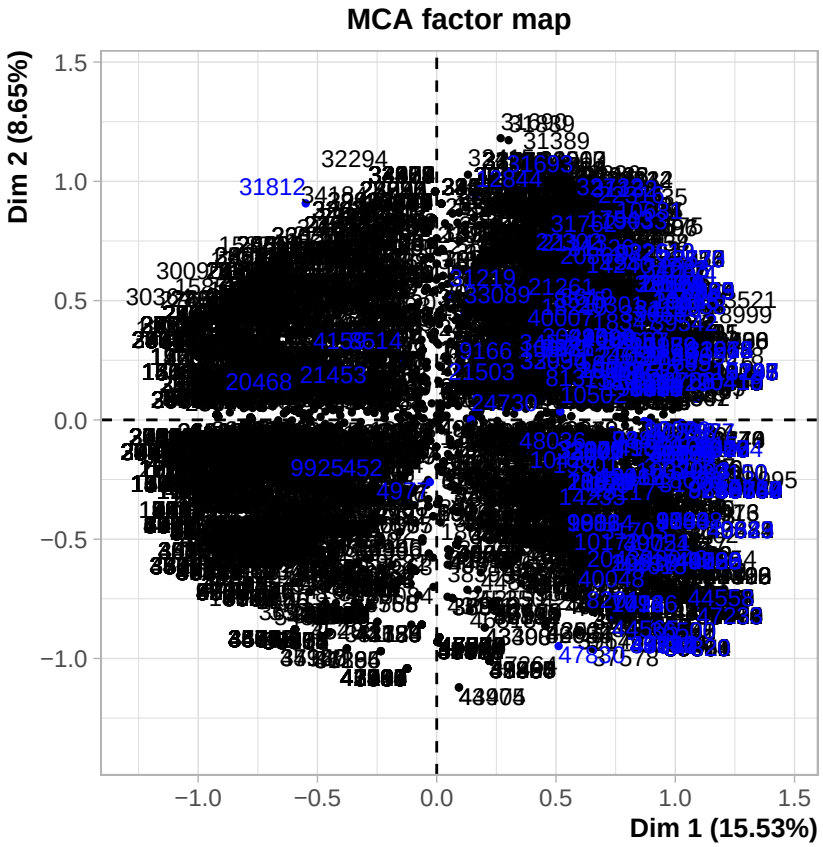
9 MCA analysis

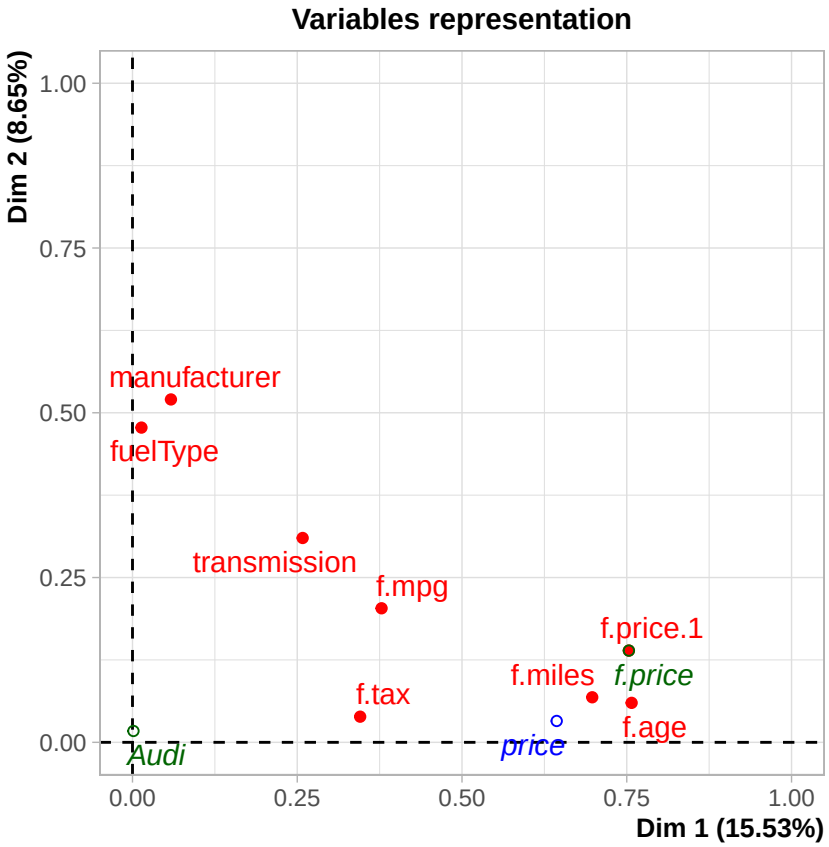
9.1 Eigenvalues and dominant axes analysis

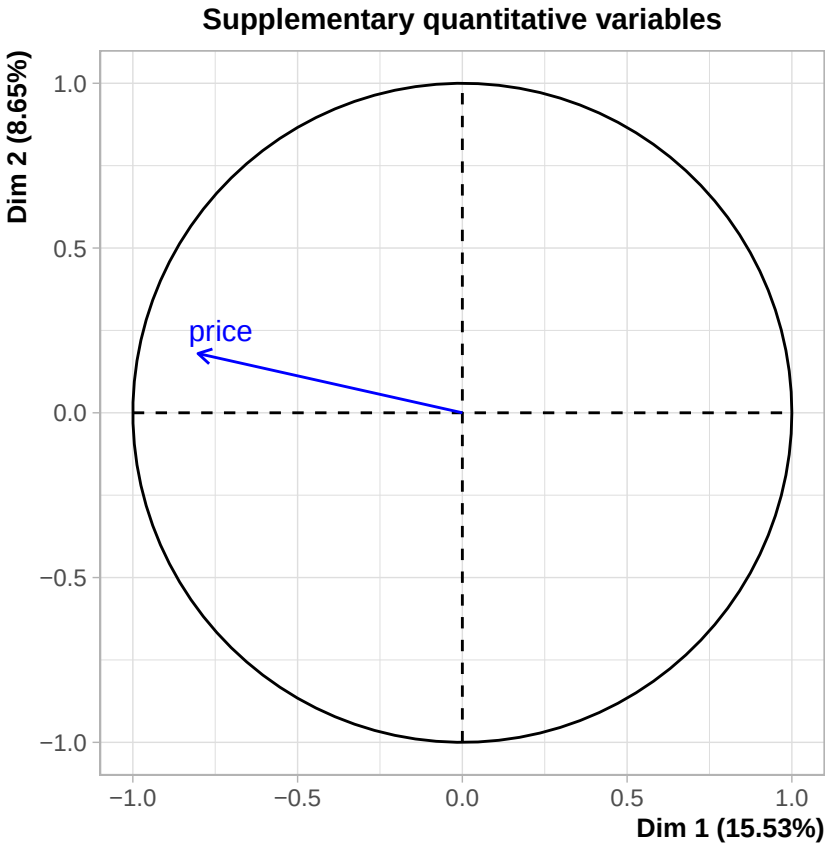
El nombre de dimensions que hem d'agafar per HC han de ser totes aquelles les quals tenen un eigenvalue més gran que la mitjana d'aquest, que en són 9.

```
par(mfrow=c(1,1))
llvout<-which(df$mout=="YesMOut")
res.mca<-MCA(df[,c("f.price","Audi",vars_dis[c(3:4, 6:11)],"price") ],quali.sup=c(1,2),quanti.sup=11 , i
```


MCA factor map
Dim 2 (8.65%)
Dim 1 (15.53%)
f.Fuel−Hybrid
Mercedes
f.Trans−Automatic
BMW
f.Trans−SemiAuto
(26,90]
f.mpg−(45,54]
f.Fuel−Petrol
VW
Audi
(62,101]
Manual


MCA factor map
Dim 2 (8.65%)
Dim 1 (15.53%)






```
summary(res.mca,nbelements=50, nbind=0)
```

```
##
## Call:
## MCA(X = df[, c("f.price", "Audi", vars_dis[c(3:4, 6:11)], "price")],
##       ind.sup = llvout, quanti.sup = 11, quali.sup = c(1, 2))
##
##
```

```
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6   Dim.7
## Variance               0.408   0.227   0.207   0.160   0.152   0.143   0.136
## % of var.             15.532   8.654   7.904   6.086   5.809   5.437   5.192
## Cumulative % of var.  15.532  24.186  32.090  38.176  43.984  49.422  54.614
##                        Dim.8   Dim.9  Dim.10  Dim.11  Dim.12  Dim.13  Dim.14
## Variance               0.133   0.127   0.120   0.117   0.115   0.110   0.098
## % of var.              5.077   4.842   4.555   4.457   4.367   4.178   3.724
## Cumulative % of var.  59.690  64.532  69.087  73.544  77.911  82.088  85.812
##                       Dim.15  Dim.16  Dim.17  Dim.18  Dim.19  Dim.20  Dim.21
## Variance               0.091   0.072   0.069   0.054   0.047   0.021   0.019
## % of var.              3.455   2.752   2.633   2.045   1.774   0.799   0.730
## Cumulative % of var.  89.267  92.019  94.652  96.696  98.471  99.270 100.000
##
## Categories
##                              Dim.1     ctr    cos2  v.test     Dim.2     ctr
## f.Trans-Manual             |   0.680   5.019   0.253  34.635 |  -0.742  10.712
## f.Trans-SemiAuto           |  -0.439   2.405   0.132 -25.027 |   0.317   2.254
## f.Trans-Automatic          |  -0.259   0.493   0.021 -10.003 |   0.558   4.092
## f.Fuel-Diesel              |   0.102   0.180   0.013   7.988 |   0.579  10.395
## f.Fuel-Petrol              |  -0.135   0.236   0.013  -7.970 |  -0.799  14.934
## f.Fuel-Hybrid              |  -0.028   0.000   0.000  -0.204 |   1.264   0.947
## Audi                       |  -0.073   0.034   0.001  -2.599 |  -0.253   0.746
## BMW                        |  -0.068   0.031   0.001  -2.468 |   0.407   1.975
## Mercedes                   |  -0.280   0.638   0.028 -11.573 |   0.915  12.194
## VW                         |   0.339   1.086   0.051  15.564 |  -0.900  13.710
## f.price.1_f.price-[0,15]   |   1.142  10.840   0.485  47.933 |  -0.557   4.635
## f.price.1_f.price-(15,20]  |   0.345   0.926   0.040  13.839 |   0.447   2.782
## f.price.1_f.price-(20,26]  |  -0.532   1.892   0.079 -19.334 |   0.022   0.006
## f.price.1_f.price-(26,90]  |  -1.094   9.434   0.414 -44.282 |   0.129   0.234
## f.miles-[0,6]              |  -1.048   9.142   0.409 -44.032 |  -0.404   2.440
## f.miles-(6,18]             |  -0.395   1.266   0.056 -16.312 |   0.021   0.006
## f.miles-(18,36]            |   0.617   3.103   0.138  25.546 |   0.252   0.926
## f.miles-(36,195]           |   1.140   7.875   0.320  38.934 |   0.189   0.388
## f.tax-[0,125]              |   1.158   2.273   0.078  19.276 |   0.571   0.994
## f.tax-(125,145]            |  -0.462   3.986   0.333 -39.721 |   0.074   0.182
## f.tax-(145,570]            |   0.650   4.332   0.212  31.717 |  -0.229   0.963
## f.mpg-[0,45]               |  -0.902   6.162   0.267 -35.550 |  -0.338   1.556
## f.mpg-(45,54]              |  -0.176   0.278   0.013  -7.803 |  -0.383   2.379
## f.mpg-(54,62]              |   0.513   1.919   0.082  19.720 |   0.117   0.179
## f.mpg-(62,101]             |   0.690   3.228   0.135  25.296 |   0.763   7.077
## f.age-[0,2]                |  -0.977  11.788   0.644 -55.227 |  -0.263   1.534
## f.age-(2,4.1]              |   0.224   0.441   0.020   9.781 |   0.294   1.364
## f.age-(4.1,5.1]            |   0.916   4.177   0.163  27.750 |   0.198   0.352
## f.age-(5.1,15]             |   1.231   6.814   0.260  35.120 |  -0.073   0.043
##                              cos2  v.test     Dim.3     ctr    cos2  v.test
## f.Trans-Manual              0.301 -37.771 |  -0.187   0.744   0.019  -9.516 |
## f.Trans-SemiAuto            0.069  18.088 |  -0.058   0.082   0.002  -3.295 |
## f.Trans-Automatic           0.098  21.512 |   0.375   2.026   0.044  14.465 |
## f.Fuel-Diesel               0.433  45.272 |   0.167   0.943   0.036  13.031 |
## f.Fuel-Petrol               0.473 -47.304 |  -0.235   1.411   0.041 -13.895 |
## f.Fuel-Hybrid               0.017   9.076 |   0.550   0.196   0.003   3.949 |
## Audi                        0.017  -9.020 |   0.156   0.310   0.007   5.558 |
## BMW                         0.046  14.726 |   0.363   1.722   0.036  13.143 |
## Mercedes                    0.301  37.778 |  -0.037   0.021   0.000  -1.515 |
## VW                          0.360 -41.280 |  -0.331   2.034   0.049 -15.193 |
## f.price.1_f.price-[0,15]    0.116 -23.395 |   0.339   1.880   0.043  14.239 |
## f.price.1_f.price-(15,20]   0.068  17.910 |  -0.600   5.496   0.122 -24.055 |
## f.price.1_f.price-(20,26]   0.000   0.809 |  -0.406   2.170   0.046 -14.768 |
## f.price.1_f.price-(26,90]   0.006   5.210 |   0.579   5.185   0.116  23.419 |
## f.miles-[0,6]               0.061 -16.979 |   0.459   3.450   0.079  19.295 |
## f.miles-(6,18]              0.000   0.871 |  -0.501   4.008   0.091 -20.702 |
## f.miles-(18,36]             0.023  10.416 |  -0.785   9.867   0.223 -32.495 |
## f.miles-(36,195]            0.009   6.449 |   1.096  14.303   0.296  37.430 |
```

```
## f.tax-[0,125]               0.019   9.514 |   0.380   0.481   0.008   6.327 |
## f.tax-(125,145]             0.008   6.342 |  -0.197   1.420   0.060 -16.913 |
## f.tax-(145,570]             0.026 -11.164 |   0.295   1.759   0.044  14.416 |
## f.mpg-[0,45]                0.038 -13.333 |   0.389   2.251   0.050  15.329 |
## f.mpg-(45,54]               0.061 -17.032 |  -0.117   0.241   0.006  -5.183 |
## f.mpg-(54,62]               0.004   4.498 |  -0.295   1.248   0.027 -11.344 |
## f.mpg-(62,101]              0.165  27.958 |   0.038   0.019   0.000   1.398 |
## f.age-[0,2]                 0.047 -14.873 |   0.269   1.761   0.049  15.226 |
## f.age-(2,4.1]               0.035  12.839 |  -1.043  18.850   0.439 -45.608 |
## f.age-(4.1,5.1]             0.008   6.011 |  -0.040   0.016   0.000  -1.219 |
## f.age-(5.1,15]              0.001  -2.075 |   1.350  16.106   0.313  38.516 |
##
## Categorical variables (eta2)
##                              Dim.1 Dim.2 Dim.3
## transmission               | 0.258 0.310 0.047 |
## fuelType                   | 0.014 0.478 0.042 |
## manufacturer               | 0.058 0.520 0.068 |
## f.price.1                  | 0.753 0.139 0.245 |
## f.miles                    | 0.698 0.068 0.525 |
## f.tax                      | 0.345 0.039 0.061 |
## f.mpg                      | 0.378 0.203 0.062 |
## f.age                      | 0.757 0.060 0.610 |
##
## Supplementary categories
##                                 Dim.1    cos2  v.test     Dim.2    cos2  v.test
## f.price_f.price-[0,15]    |   1.142   0.485  47.933 |  -0.557   0.116 -23.395 |
## f.price_f.price-(15,20]   |   0.345   0.040  13.839 |   0.447   0.068  17.910 |
## f.price_f.price-(20,26]   |  -0.532   0.079 -19.334 |   0.022   0.000   0.809 |
## f.price_f.price-(26,90]   |  -1.094   0.414 -44.282 |   0.129   0.006   5.210 |
## Audi No                   |   0.020   0.001   2.599 |   0.068   0.017   9.020 |
## Audi Yes                  |  -0.073   0.001  -2.599 |  -0.253   0.017  -9.020 |
##                              Dim.3    cos2  v.test
## f.price_f.price-[0,15]       0.339   0.043  14.239 |
## f.price_f.price-(15,20]     -0.600   0.122 -24.055 |
## f.price_f.price-(20,26]     -0.406   0.046 -14.768 |
## f.price_f.price-(26,90]      0.579   0.116  23.419 |
## Audi No                     -0.042   0.007  -5.558 |
## Audi Yes                     0.156   0.007   5.558 |
##
## Supplementary categorical variables (eta2)
##                              Dim.1 Dim.2 Dim.3
## f.price                    | 0.753 0.139 0.245 |
## Audi                       | 0.001 0.017 0.007 |
##
## Supplementary continuous variable
##                               Dim.1    Dim.2    Dim.3
## price                      | -0.802 |  0.180 |  0.162 |
```

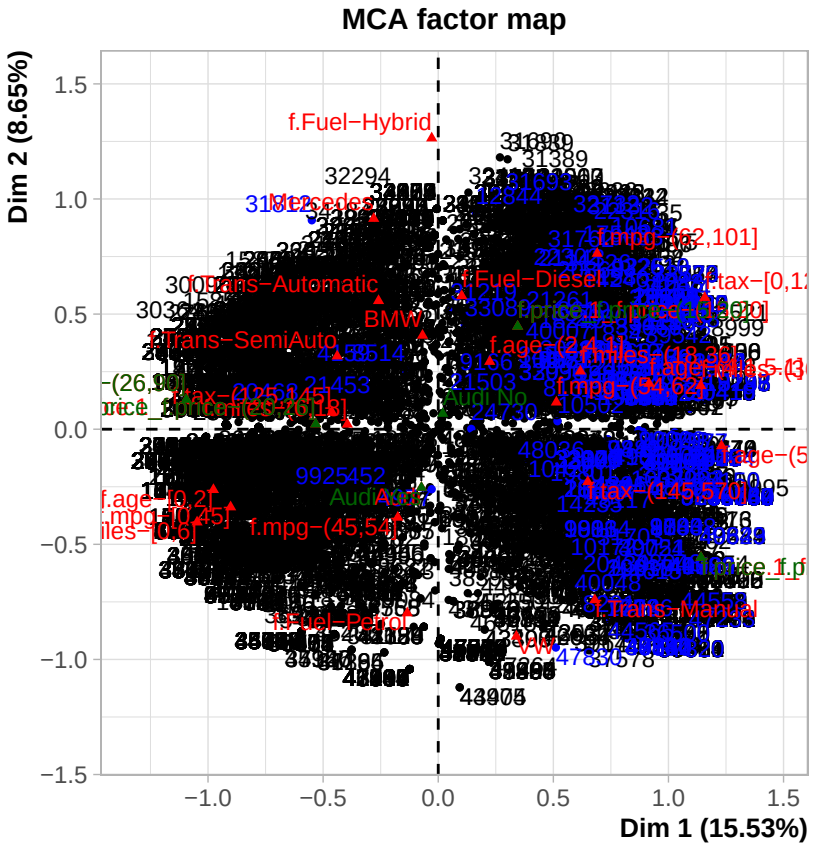
```
which(res.mca$eig[,1] > mean(res.mca$eig[,1]))
```

```
## dim 1 dim 2 dim 3 dim 4 dim 5 dim 6 dim 7 dim 8 dim 9
##     1     2     3     4     5     6     7     8     9
```

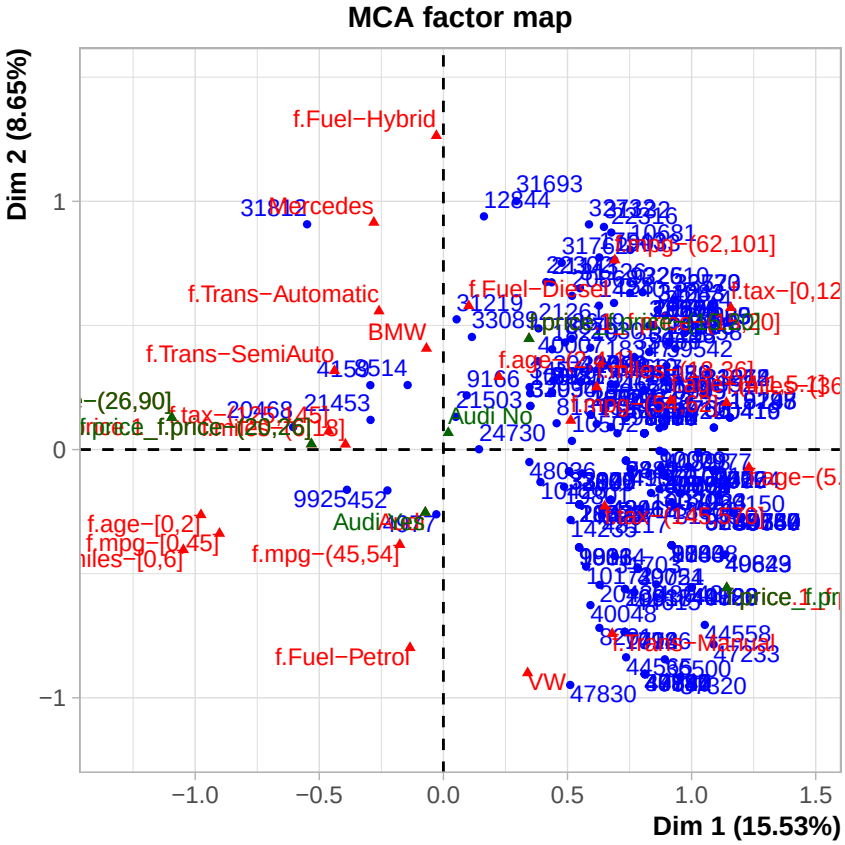
9.2 Individuals point of view

Podem veure que els individus més significatius són tots Volswaggen, manuals, dièsel i de fa entre 7 i 9 anys i que comparteixen casi tots les categories de taxa, mpg, preu, mileage, ...

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




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




```
inds <- res.mca$ind$coord
inds <- as.data.frame(inds)
rang<-inds[order(inds$`Dim 1`, decreasing = TRUE),]

rang[1,]
```

```
##          Dim 1     Dim 2     Dim 3     Dim 4    Dim 5
## 48995 1.234215 -0.213821 0.6910406 0.3573846 1.133276
```

```
res.pca$ind$coord[row.names(rang)[1:10],1]
```

```
##    48995    35322    35749    35755    35812    35864    36538    36601
## 2.179962 2.720048 3.908535 4.338991 3.061369 2.671092 3.467632 2.598045
##    36835    36988
## 2.474084 3.264550
```

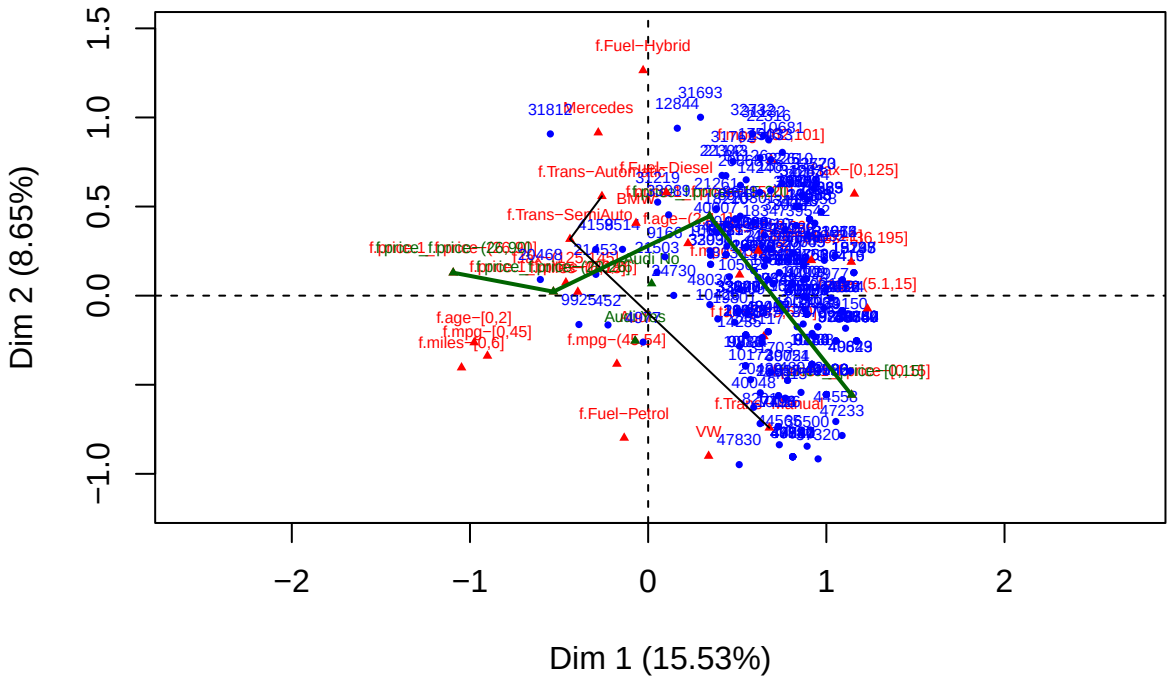
```
df[which(row.names(df) %in% row.names(res.pca$ind$coord[row.names(rang)[1:10],])),1:18]
```

```
##            model year price   transmission mileage      fuelType      tax  mpg
## 35322   VW- Golf 2014  8385 f.Trans-Manual   45703 f.Fuel-Diesel 148.0509 68.9
## 35749   VW- Golf 2013  6998 f.Trans-Manual   68931 f.Fuel-Diesel 149.7279 68.9
## 35755   VW- Golf 2013  6498 f.Trans-Manual   81025 f.Fuel-Diesel 149.6457 68.9
## 35812   VW- Golf 2015 11998 f.Trans-Manual   67125 f.Fuel-Diesel 146.5921 67.3
## 35864   VW- Golf 2014 13989 f.Trans-Manual   45995 f.Fuel-Diesel 148.5779 67.3
## 36538   VW- Golf 2014  7490 f.Trans-Manual   66707 f.Fuel-Diesel 147.9080 68.9
## 36601   VW- Golf 2014  8795 f.Trans-Manual   42275 f.Fuel-Diesel 148.0754 68.9
## 36835   VW- Golf 2015 13570 f.Trans-Manual   50624 f.Fuel-Diesel 146.7085 67.3
## 36988   VW- Golf 2014  9164 f.Trans-Manual   61000 f.Fuel-Diesel 147.9536 68.9
## 48995 VW- Touran 2014  7995 f.Trans-Manual   43000 f.Fuel-Diesel 125.0000 61.4
##       engineSize manufacturer age        f.price          f.miles
## 35322     Medium           VW   7 f.price-[0,15] f.miles-(36,195]
## 35749     Medium           VW   8 f.price-[0,15] f.miles-(36,195]
## 35755     Medium           VW   8 f.price-[0,15] f.miles-(36,195]
## 35812     Medium           VW   6 f.price-[0,15] f.miles-(36,195]
## 35864     Medium           VW   7 f.price-[0,15] f.miles-(36,195]
## 36538     Medium           VW   7 f.price-[0,15] f.miles-(36,195]
## 36601     Medium           VW   7 f.price-[0,15] f.miles-(36,195]
## 36835     Medium           VW   6 f.price-[0,15] f.miles-(36,195]
## 36988     Medium           VW   7 f.price-[0,15] f.miles-(36,195]
## 48995     Medium           VW   7 f.price-[0,15] f.miles-(36,195]
##                 f.tax          f.mpg          f.age    Audi   mout
## 35322 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi No NoMOut
## 35749 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi No NoMOut
## 35755 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi No NoMOut
## 35812 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi No NoMOut
## 35864 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi No NoMOut
## 36538 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi No NoMOut
## 36601 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi No NoMOut
## 36835 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi No NoMOut
## 36988 f.tax-(145,570] f.mpg-(62,101] f.age-(5.1,15] Audi No NoMOut
## 48995   f.tax-[0,125]  f.mpg-(54,62] f.age-(5.1,15] Audi No NoMOut
```

9.3 Interpreting map of categories

Les categories més característiques del a primera dimensió són l'edat (entre 5 i 15 anys), el preu (entre 15.000 i 20.000 i entre 26.000 i 90.000) i el mileage (entre 0 i 6.000).

```
# Representation of categories
plot.MCA(res.mca,choix=c("ind"),invisible=c("ind"),axes=c(1,2), graph.type = "classic", cex = 0.5)
lines(res.mca$var$coord[1:3,1],res.mca$var$coord[1:3,2],lwd=1,col="black") # Transmission
lines(res.mca$quali.sup$coord[1:4,1],res.mca$quali.sup$coord[1:4,2],lwd=2,col="darkgreen")
```




```
names(res.mca)
```

```
## [1] "eig"        "call"       "ind"        "var"        "svd"
## [6] "ind.sup"    "quali.sup"  "quanti.sup"
```

```
res.mca$var
```

```
## $coord
##                                 Dim 1       Dim 2       Dim 3       Dim 4
## f.Trans-Manual              0.68011555 -0.74169709 -0.18685448 -0.33759922
## f.Trans-SemiAuto           -0.43892518  0.31723413 -0.05779316  0.10887774
## f.Trans-Automatic          -0.25938685  0.55782413  0.37508243  0.31439930
## f.Fuel-Diesel               0.10215748  0.57898235  0.16665696 -0.29438057
## f.Fuel-Petrol              -0.13452824 -0.79850057 -0.23455212  0.34542656
## f.Fuel-Hybrid              -0.02841956  1.26418146  0.54999326  1.75140779
## Audi                       -0.07300006 -0.25336443  0.15611771  0.45267039
## BMW                        -0.06823329  0.40720436  0.36342728  0.15816848
## Mercedes                   -0.28026484  0.91484090 -0.03669891 -0.14650597
## VW                          0.33932141 -0.89995232 -0.33123558 -0.29586451
## f.price.1_f.price-[0,15]    1.14154005 -0.55716712  0.33910499 -0.34615107
## f.price.1_f.price-(15,20]   0.34502693  0.44654324 -0.59975764  0.39181018
## f.price.1_f.price-(20,26]  -0.53196777  0.02225183 -0.40634927 -0.37511555
## f.price.1_f.price-(26,90]  -1.09427576  0.12873885  0.57873280  0.29723455
## f.miles-[0,6]              -1.04753759 -0.40392795  0.45901963 -0.14880164
## f.miles-(6,18]             -0.39478841  0.02108902 -0.50104763 -0.15237304
## f.miles-(18,36]             0.61729554  0.25169733 -0.78521241  0.37950007
## f.miles-(36,195]            1.14004545  0.18883900  1.09599756 -0.10114873
## f.tax-[0,125]               1.15760755  0.57137529  0.37995297  2.00024788
## f.tax-(125,145]            -0.46175101  0.07372564 -0.19661568 -0.30564900
## f.tax-(145,570]             0.64958622 -0.22863807  0.29525751  0.22613452
## f.mpg-[0,45]               -0.90195099 -0.33827736  0.38891091  0.52681370
## f.mpg-(45,54]              -0.17556910 -0.38323597 -0.11661600 -0.10056552
## f.mpg-(54,62]               0.51313855  0.11704546 -0.29517731  0.62587849
## f.mpg-(62,101]              0.69047033  0.76311879  0.03815162 -1.12898936
```

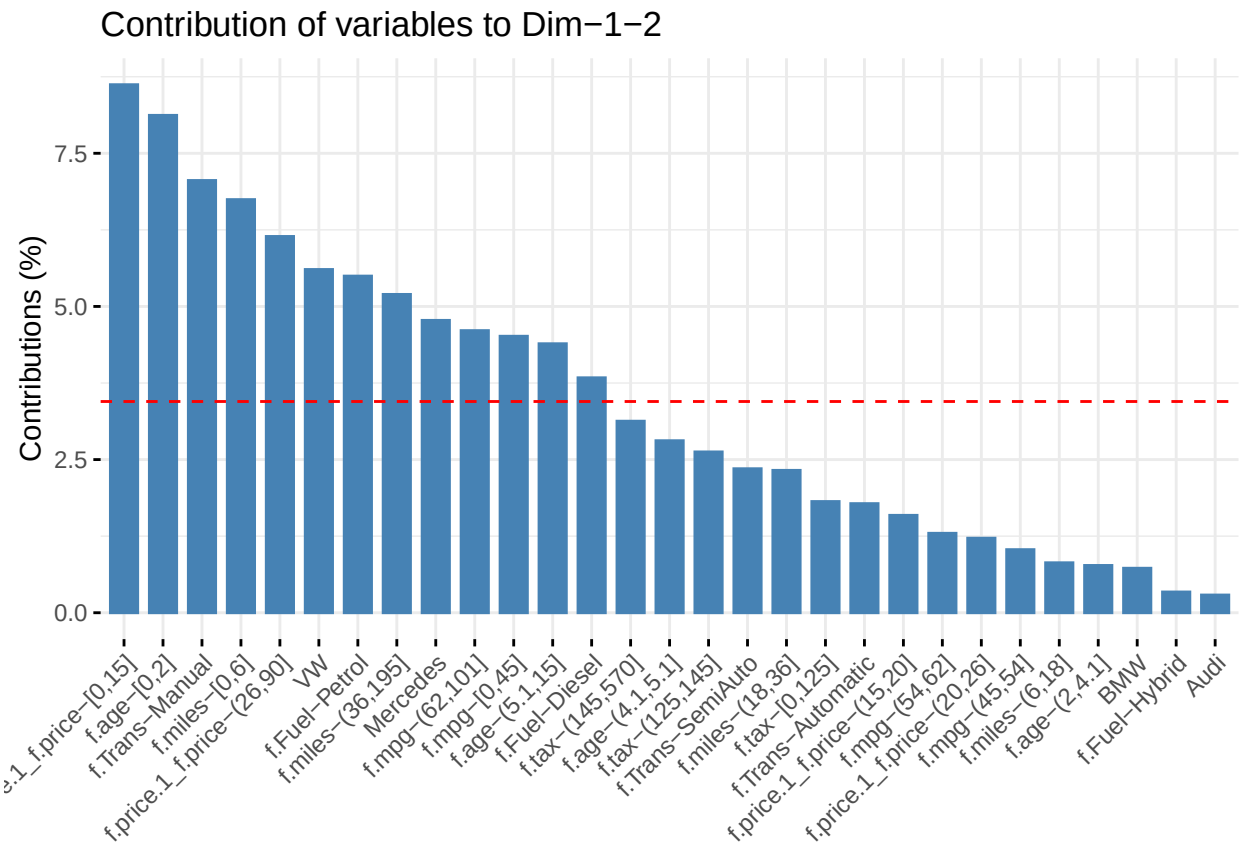
```
## f.age-[0,2]               -0.97667355 -0.26302020  0.26927149 -0.19602325
## f.age-(2,4.1]              0.22361595  0.29352039 -1.04265678 -0.01078729
## f.age-(4.1,5.1]            0.91594786  0.19841430 -0.04022280  0.60355320
## f.age-(5.1,15]             1.23067245 -0.07272421  1.34968813 -0.10823158
##                                 Dim 5
## f.Trans-Manual             0.08364964
## f.Trans-SemiAuto           0.02717541
## f.Trans-Automatic         -0.17013338
## f.Fuel-Diesel              0.13771410
## f.Fuel-Petrol             -0.16044910
## f.Fuel-Hybrid             -0.86457938
## Audi                      -0.31003161
## BMW                        0.29628001
## Mercedes                  -0.05734586
## VW                         0.05371057
## f.price.1_f.price-[0,15]  -0.04790321
## f.price.1_f.price-(15,20]  0.00629641
## f.price.1_f.price-(20,26]  0.52765175
## f.price.1_f.price-(26,90] -0.40350914
## f.miles-[0,6]              0.18654298
## f.miles-(6,18]             0.10916311
## f.miles-(18,36]           -0.49666426
## f.miles-(36,195]           0.26466144
## f.tax-[0,125]              2.36333425
## f.tax-(125,145]            0.11502863
## f.tax-(145,570]           -0.59987154
## f.mpg-[0,45]              -0.83203725
## f.mpg-(45,54]              0.68653873
## f.mpg-(54,62]              0.55643958
## f.mpg-(62,101]            -0.58327186
## f.age-[0,2]                0.21344259
## f.age-(2,4.1]             -0.17302023
## f.age-(4.1,5.1]           -0.33939927
## f.age-(5.1,15]             0.12857944
##
## $contrib
##                                  Dim 1        Dim 2       Dim 3        Dim 4
## f.Trans-Manual            5.018623e+00 10.712053901  0.74440461  3.155905434
## f.Trans-SemiAuto          2.404546e+00  2.254303724  0.08191962  0.377600314
## f.Trans-Automatic         4.930460e-01  4.092480405  2.02594615  1.848658351
## f.Fuel-Diesel             1.803159e-01 10.395007478  0.94302561  3.821324842
## f.Fuel-Petrol             2.361906e-01 14.934358432  1.41090337  3.974194173
## f.Fuel-Hybrid             2.666551e-04  0.946966452  0.19625148  2.584591000
## Audi                      3.449781e-02  0.745824869  0.31005046  3.385410045
## BMW                       3.089311e-02  1.974668910  1.72221330  0.423653531
## Mercedes                  6.376464e-01 12.193665360  0.02148483  0.444687880
## VW                        1.085993e+00 13.710211167  2.03358128  2.107135187
## f.price.1_f.price-[0,15]  1.084003e+01  4.634679205  1.87974747  2.543791840
## f.price.1_f.price-(15,20] 9.255378e-01  2.782374830  5.49569481  3.046081626
## f.price.1_f.price-(20,26] 1.892414e+00  0.005942613  2.16983806  2.401472104
## f.price.1_f.price-(26,90] 9.433857e+00  0.234344845  5.18531079  1.776381595
## f.miles-[0,6]             9.142457e+00  2.439674873  3.44960273  0.470805030
## f.miles-(6,18]            1.266245e+00  0.006484884  4.00801899  0.481401085
## f.miles-(18,36]           3.103216e+00  0.925941311  9.86695098  2.993308348
## f.miles-(36,195]          7.875268e+00  0.387797342 14.30286535  0.158213606
## f.tax-[0,125]             2.272841e+00  0.993778952  0.48115942 17.318729441
## f.tax-(125,145]           3.986190e+00  0.182381123  1.42023927  4.457488508
## f.tax-(145,570]           4.332334e+00  0.963265168  1.75886745  1.339935890
## f.mpg-[0,45]              6.161648e+00  1.555526104  2.25120028  5.364720269
## f.mpg-(45,54]             2.781660e-01  2.378706657  0.24116099  0.232920690
## f.mpg-(54,62]             1.919344e+00  0.179222748  1.24805109  7.287281282
## f.mpg-(62,101]            3.228248e+00  7.077201647  0.01936804 22.027158677
## f.age-[0,2]               1.178827e+01  1.534368693  1.76081896  1.211903995
## f.age-(2,4.1]             4.412112e-01  1.364325499 18.84979472  0.002620399
```

```
## f.age-(4.1,5.1]           4.176505e+00  0.351736855  0.01582701  4.628119403
## f.age-(5.1,15]            6.814200e+00  0.042705952 16.10570291  0.134505454
##                                  Dim 5
## f.Trans-Manual            2.030018e-01
## f.Trans-SemiAuto          2.464661e-02
## f.Trans-Automatic         5.671831e-01
## f.Fuel-Diesel             8.761998e-01
## f.Fuel-Petrol             8.983858e-01
## f.Fuel-Hybrid             6.598994e-01
## Audi                      1.663833e+00
## BMW                       1.557494e+00
## Mercedes                  7.138376e-02
## VW                        7.275735e-02
## f.price.1_f.price-[0,15]  5.104228e-02
## f.price.1_f.price-(15,20] 8.241889e-04
## f.price.1_f.price-(20,26] 4.978435e+00
## f.price.1_f.price-(26,90] 3.430008e+00
## f.miles-[0,6]             7.752364e-01
## f.miles-(6,18]            2.588769e-01
## f.miles-(18,36]           5.371607e+00
## f.miles-(36,195]          1.134893e+00
## f.tax-[0,125]             2.533083e+01
## f.tax-(125,145]           6.614650e-01
## f.tax-(145,570]           9.879102e+00
## f.mpg-[0,45]              1.402068e+01
## f.mpg-(45,54]             1.137341e+01
## f.mpg-(54,62]             6.034929e+00
## f.mpg-(62,101]            6.159869e+00
## f.age-[0,2]               1.505448e+00
## f.age-(2,4.1]             7.062971e-01
## f.age-(4.1,5.1]           1.533366e+00
## f.age-(5.1,15]            1.988958e-01
##
## $cos2
##                                  Dim 1        Dim 2        Dim 3        Dim 4
## f.Trans-Manual            2.533483e-01 0.3013045820 0.0191231577 6.242456e-02
## f.Trans-SemiAuto          1.322790e-01 0.0690986780 0.0022933101 8.139320e-03
## f.Trans-Automatic         2.113282e-02 0.0977363778 0.0441890934 3.104737e-02
## f.Fuel-Diesel             1.347561e-02 0.4328513165 0.0358636911 1.118990e-01
## f.Fuel-Petrol             1.341370e-02 0.4725764545 0.0407755984 8.843681e-02
## f.Fuel-Hybrid             8.792153e-06 0.0173972022 0.0032928755 3.339144e-02
## Audi                      1.426394e-03 0.0171824230 0.0065237527 5.484756e-02
## BMW                       1.285954e-03 0.0457991848 0.0364811031 6.909916e-03
## Mercedes                  2.828825e-02 0.3014115697 0.0004850373 7.729999e-03
## VW                        5.116119e-02 0.3598795267 0.0487519617 3.889591e-02
## f.price.1_f.price-[0,15]  4.852220e-01 0.1155923607 0.0428180147 4.461589e-02
## f.price.1_f.price-(15,20] 4.044451e-02 0.0677455279 0.1222095559 5.215609e-02
## f.price.1_f.price-(20,26] 7.894366e-02 0.0001381269 0.0460622965 3.925335e-02
## f.price.1_f.price-(26,90] 4.141187e-01 0.0057317937 0.1158317189 3.055413e-02
## f.miles-[0,6]             4.094724e-01 0.0608825974 0.0786226839 8.262297e-03
## f.miles-(6,18]            5.619122e-02 0.0001603440 0.0905102416 8.370588e-03
## f.miles-(18,36]           1.378280e-01 0.0229143898 0.2230105298 5.209245e-02
## f.miles-(36,195]          3.201375e-01 0.0087836626 0.2958771523 2.520073e-03
## f.tax-[0,125]             7.847440e-02 0.0191182540 0.0084540534 2.343004e-01
## f.tax-(125,145]           3.332046e-01 0.0084943926 0.0604131563 1.459961e-01
## f.tax-(145,570]           2.124547e-01 0.0263202322 0.0438929261 2.574698e-02
## f.mpg-[0,45]              2.669134e-01 0.0375448504 0.0496254860 9.105811e-02
## f.mpg-(45,54]             1.285738e-02 0.0612616734 0.0056724749 4.218466e-03
## f.mpg-(54,62]             8.212974e-02 0.0042730733 0.0271767261 1.221831e-01
## f.mpg-(62,101]            1.351436e-01 0.1650781591 0.0004126019 3.613142e-01
## f.age-[0,2]               6.441381e-01 0.0467152130 0.0489621945 2.594749e-02
## f.age-(2,4.1]             2.020622e-02 0.0348141718 0.4393010693 4.702231e-05
## f.age-(4.1,5.1]           1.626319e-01 0.0076315032 0.0003136236 7.061472e-02
## f.age-(5.1,15]            2.604839e-01 0.0009096074 0.3133017422 2.014670e-03
```

```
##                                Dim 5
## f.Trans-Manual            3.832487e-03
## f.Trans-SemiAuto          5.070634e-04
## f.Trans-Automatic         9.091608e-03
## f.Fuel-Diesel             2.448865e-02
## f.Fuel-Petrol             1.908078e-02
## f.Fuel-Hybrid             8.137113e-03
## Audi                      2.572794e-02
## BMW                       2.424586e-02
## Mercedes                  1.184330e-03
## VW                        1.281851e-03
## f.price.1_f.price-[0,15]  8.544513e-04
## f.price.1_f.price-(15,20] 1.346913e-05
## f.price.1_f.price-(20,26] 7.766787e-02
## f.price.1_f.price-(26,90] 5.630903e-02
## f.miles-[0,6]             1.298504e-02
## f.miles-(6,18]            4.296269e-03
## f.miles-(18,36]           8.922301e-02
## f.miles-(36,195]          1.725336e-02
## f.tax-[0,125]             3.270812e-01
## f.tax-(125,145]           2.067793e-02
## f.tax-(145,570]           1.811795e-01
## f.mpg-[0,45]              2.271381e-01
## f.mpg-(45,54]             1.966013e-01
## f.mpg-(54,62]             9.657555e-02
## f.mpg-(62,101]            9.643782e-02
## f.age-[0,2]               3.076396e-02
## f.age-(2,4.1]             1.209688e-02
## f.age-(4.1,5.1]           2.232986e-02
## f.age-(5.1,15]            2.843407e-03
##
## $v.test
##                                Dim 1       Dim 2      Dim 3       Dim 4
## f.Trans-Manual             34.6353028 -37.7713806  -9.515679 -17.1924480
## f.Trans-SemiAuto          -25.0268074  18.0881796  -3.295273   6.2080337
## f.Trans-Automatic         -10.0031956  21.5123627  14.464970  12.1247385
## f.Fuel-Diesel               7.9879296  45.2719669  13.031292 -23.0182967
## f.Fuel-Petrol              -7.9695584 -47.3038002 -13.895052  20.4633406
## f.Fuel-Hybrid              -0.2040364   9.0761089   3.948641  12.5741188
## Audi                       -2.5988414  -9.0199098   5.557875  16.1153094
## BMW                        -2.4675885  14.7261380  13.142984   5.7200046
## Mercedes                  -11.5734542  37.7780860  -1.515471  -6.0499210
## VW                         15.5643260 -41.2798929 -15.193437 -13.5710026
## f.price.1_f.price-[0,15]   47.9325162 -23.3950813  14.238796 -14.5346557
## f.price.1_f.price-(15,20]  13.8385245  17.9101947 -24.055400  15.7149319
## f.price.1_f.price-(20,26] -19.3338620   0.8087216 -14.768377 -13.6332174
## f.price.1_f.price-(26,90] -44.2815106   5.2096107  23.419291  12.0280422
## f.miles-[0,6]             -44.0323926 -16.9787838  19.294518  -6.2547563
## f.miles-(6,18]            -16.3115127   0.8713373 -20.701836  -6.2956121
## f.miles-(18,36]            25.5463385  10.4163158 -32.495459  15.7053415
## f.miles-(36,195]           38.9339337   6.4490807  37.429645  -3.4543518
## f.tax-[0,125]              19.2763143   9.5144591   6.326922  33.3078398
## f.tax-(125,145]           -39.7205690   6.3419988 -16.913199 -26.2924209
## f.tax-(145,570]            31.7170738 -11.1636150  14.416414  11.0413754
## f.mpg-[0,45]              -35.5504572 -13.3332242  15.328949  20.7643966
## f.mpg-(45,54]              -7.8025441 -17.0315596  -5.182583  -4.4692768
## f.mpg-(54,62]              19.7201506   4.4981110 -11.343800  24.0527983
## f.mpg-(62,101]             25.2963391  27.9579163   1.397737 -41.3620924
## f.age-[0,2]               -55.2267489 -14.8726774  15.226161 -11.0842841
## f.age-(2,4.1]               9.7814342  12.8392018 -45.608010  -0.4718587
## f.age-(4.1,5.1]            27.7499889   6.0112534  -1.218609  18.2855329
## f.age-(5.1,15]             35.1196718  -2.0753291  38.516019  -3.0886021
##                                Dim 5
## f.Trans-Manual              4.2599092
```

```
## f.Trans-SemiAuto            1.5494983
## f.Trans-Automatic          -6.5611558
## f.Fuel-Diesel              10.7681835
## f.Fuel-Petrol              -9.5051306
## f.Fuel-Hybrid              -6.2071917
## Audi                      -11.0372919
## BMW                        10.7146698
## Mercedes                   -2.3680803
## VW                          2.4636491
## f.price.1_f.price-[0,15]   -2.0114241
## f.price.1_f.price-(15,20]   0.2525398
## f.price.1_f.price-(20,26]  19.1770005
## f.price.1_f.price-(26,90] -16.3286029
## f.miles-[0,6]               7.8411829
## f.miles-(6,18]              4.5103033
## f.miles-(18,36]           -20.5540983
## f.miles-(36,195]            9.0385090
## f.tax-[0,125]              39.3539016
## f.tax-(125,145]             9.8949488
## f.tax-(145,570]           -29.2896756
## f.mpg-[0,45]              -32.7948028
## f.mpg-(45,54]              30.5107720
## f.mpg-(54,62]              21.3842290
## f.mpg-(62,101]            -21.3689741
## f.age-[0,2]                12.0692735
## f.age-(2,4.1]              -7.5682702
## f.age-(4.1,5.1]           -10.2826006
## f.age-(5.1,15]              3.6692686
##
## $eta2
##                    Dim 1      Dim 2      Dim 3      Dim 4       Dim 5
## transmission 0.25820269 0.31002175 0.04734248 0.06878570 0.009695397
## fuelType     0.01359386 0.47753748 0.04232834 0.13266097 0.029695978
## manufacturer 0.05835270 0.52020996 0.06784221 0.08129407 0.041052157
## f.price.1    0.75318509 0.13916202 0.24450090 0.12483453 0.103199308
## f.miles      0.69758447 0.06833117 0.52495768 0.05244690 0.091980805
## f.tax        0.34545791 0.03888122 0.06075373 0.29543149 0.437561176
## f.mpg        0.37794568 0.20337535 0.06240548 0.44618703 0.458511292
## f.age        0.75737121 0.05984840 0.60968647 0.07638979 0.048109215
```

```
fviz_contrib(res.mca, choice="var", axes = 1:2)
```



9.4 Interpreting the axes association to factor map

Com hem dit a l'apartat anterior i es pot veure seguidament, price, miles i age són les variables factor més associades a la primera dimensió, encara que la taxa i els mpg també son grans contribuidors.

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

```
##                            Dim 1 Dim 2 Dim 1 Dim 2 Dim 1 Dim 2
## f.Trans-Manual              0.68 -0.74  0.25  0.30  5.02 10.71
## f.Trans-SemiAuto           -0.44  0.32  0.13  0.07  2.40  2.25
## f.Trans-Automatic          -0.26  0.56  0.02  0.10  0.49  4.09
## f.Fuel-Diesel               0.10  0.58  0.01  0.43  0.18 10.40
## f.Fuel-Petrol              -0.13 -0.80  0.01  0.47  0.24 14.93
## f.Fuel-Hybrid              -0.03  1.26  0.00  0.02  0.00  0.95
## Audi                       -0.07 -0.25  0.00  0.02  0.03  0.75
## BMW                        -0.07  0.41  0.00  0.05  0.03  1.97
## Mercedes                   -0.28  0.91  0.03  0.30  0.64 12.19
## VW                          0.34 -0.90  0.05  0.36  1.09 13.71
## f.price.1_f.price-[0,15]    1.14 -0.56  0.49  0.12 10.84  4.63
## f.price.1_f.price-(15,20]   0.35  0.45  0.04  0.07  0.93  2.78
## f.price.1_f.price-(20,26]  -0.53  0.02  0.08  0.00  1.89  0.01
## f.price.1_f.price-(26,90]  -1.09  0.13  0.41  0.01  9.43  0.23
## f.miles-[0,6]              -1.05 -0.40  0.41  0.06  9.14  2.44
## f.miles-(6,18]             -0.39  0.02  0.06  0.00  1.27  0.01
## f.miles-(18,36]             0.62  0.25  0.14  0.02  3.10  0.93
## f.miles-(36,195]            1.14  0.19  0.32  0.01  7.88  0.39
## f.tax-[0,125]               1.16  0.57  0.08  0.02  2.27  0.99
## f.tax-(125,145]            -0.46  0.07  0.33  0.01  3.99  0.18
## f.tax-(145,570]             0.65 -0.23  0.21  0.03  4.33  0.96
## f.mpg-[0,45]               -0.90 -0.34  0.27  0.04  6.16  1.56
## f.mpg-(45,54]              -0.18 -0.38  0.01  0.06  0.28  2.38
## f.mpg-(54,62]               0.51  0.12  0.08  0.00  1.92  0.18
## f.mpg-(62,101]              0.69  0.76  0.14  0.17  3.23  7.08
## f.age-[0,2]                -0.98 -0.26  0.64  0.05 11.79  1.53
```

```
## f.age-(2,4.1]              0.22  0.29  0.02  0.03  0.44  1.36
## f.age-(4.1,5.1]            0.92  0.20  0.16  0.01  4.18  0.35
## f.age-(5.1,15]             1.23 -0.07  0.26  0.00  6.81  0.04
```

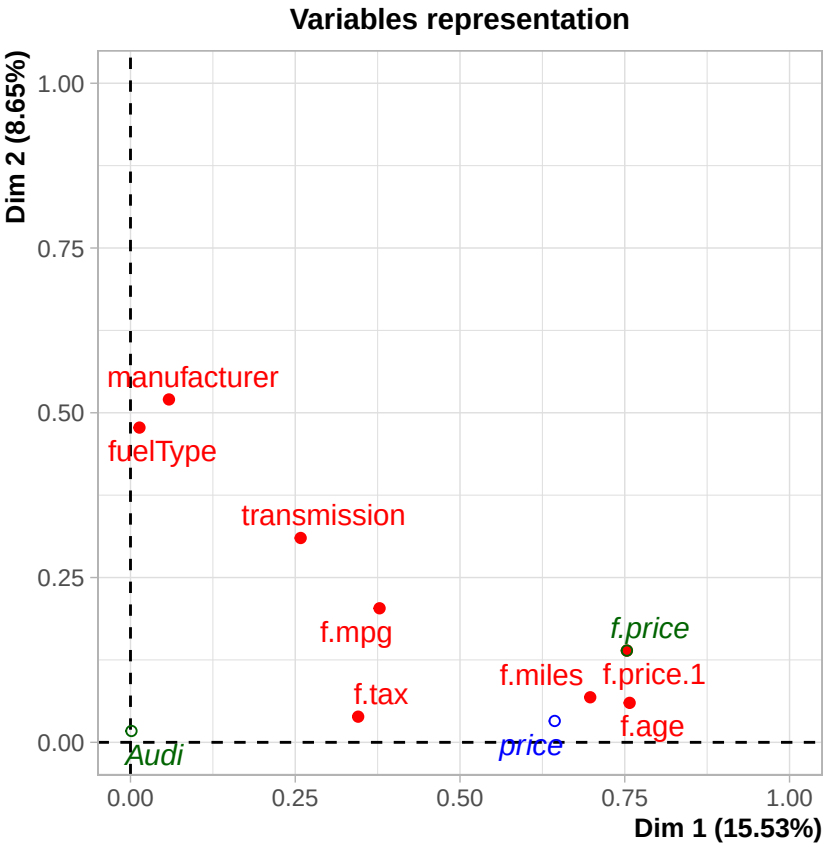
```
round(cbind(res.mca$var$cos2[,1:2],res.mca$var$contrib[,1:2]),2)
```

```
##                           Dim 1 Dim 2 Dim 1 Dim 2
## f.Trans-Manual             0.25  0.30  5.02 10.71
## f.Trans-SemiAuto           0.13  0.07  2.40  2.25
## f.Trans-Automatic          0.02  0.10  0.49  4.09
## f.Fuel-Diesel              0.01  0.43  0.18 10.40
## f.Fuel-Petrol              0.01  0.47  0.24 14.93
## f.Fuel-Hybrid              0.00  0.02  0.00  0.95
## Audi                       0.00  0.02  0.03  0.75
## BMW                        0.00  0.05  0.03  1.97
## Mercedes                   0.03  0.30  0.64 12.19
## VW                         0.05  0.36  1.09 13.71
## f.price.1_f.price-[0,15]   0.49  0.12 10.84  4.63
## f.price.1_f.price-(15,20]  0.04  0.07  0.93  2.78
## f.price.1_f.price-(20,26]  0.08  0.00  1.89  0.01
## f.price.1_f.price-(26,90]  0.41  0.01  9.43  0.23
## f.miles-[0,6]              0.41  0.06  9.14  2.44
## f.miles-(6,18]             0.06  0.00  1.27  0.01
## f.miles-(18,36]            0.14  0.02  3.10  0.93
## f.miles-(36,195]           0.32  0.01  7.88  0.39
## f.tax-[0,125]              0.08  0.02  2.27  0.99
## f.tax-(125,145]            0.33  0.01  3.99  0.18
## f.tax-(145,570]            0.21  0.03  4.33  0.96
## f.mpg-[0,45]               0.27  0.04  6.16  1.56
## f.mpg-(45,54]              0.01  0.06  0.28  2.38
## f.mpg-(54,62]              0.08  0.00  1.92  0.18
## f.mpg-(62,101]             0.14  0.17  3.23  7.08
## f.age-[0,2]                0.64  0.05 11.79  1.53
## f.age-(2,4.1]              0.02  0.03  0.44  1.36
## f.age-(4.1,5.1]            0.16  0.01  4.18  0.35
## f.age-(5.1,15]             0.26  0.00  6.81  0.04
```

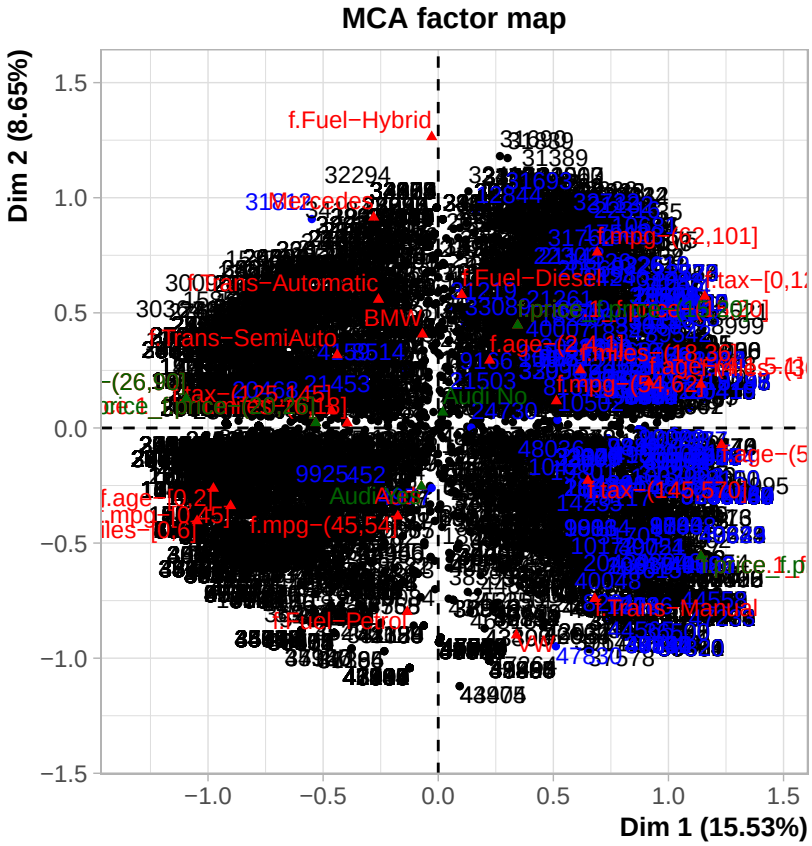
```
res.des<-dimdesc(res.mca)
res.des$`Dim 1`$quali
```

```
##                        R2       p.value
## f.price       0.753185092  0.000000e+00
## f.price.1     0.753185092  0.000000e+00
## f.miles       0.697584474  0.000000e+00
## f.tax         0.345457908  0.000000e+00
## f.mpg         0.377945676  0.000000e+00
## f.age         0.757371210  0.000000e+00
## transmission 0.258202690 1.071651e-307
## manufacturer 0.058352704  2.206312e-61
## fuelType      0.013593864  8.569755e-15
## Audi          0.001426394  9.339857e-03
```

```
plot.MCA(res.mca,choix=c("var"),axes=c(1,2))
```



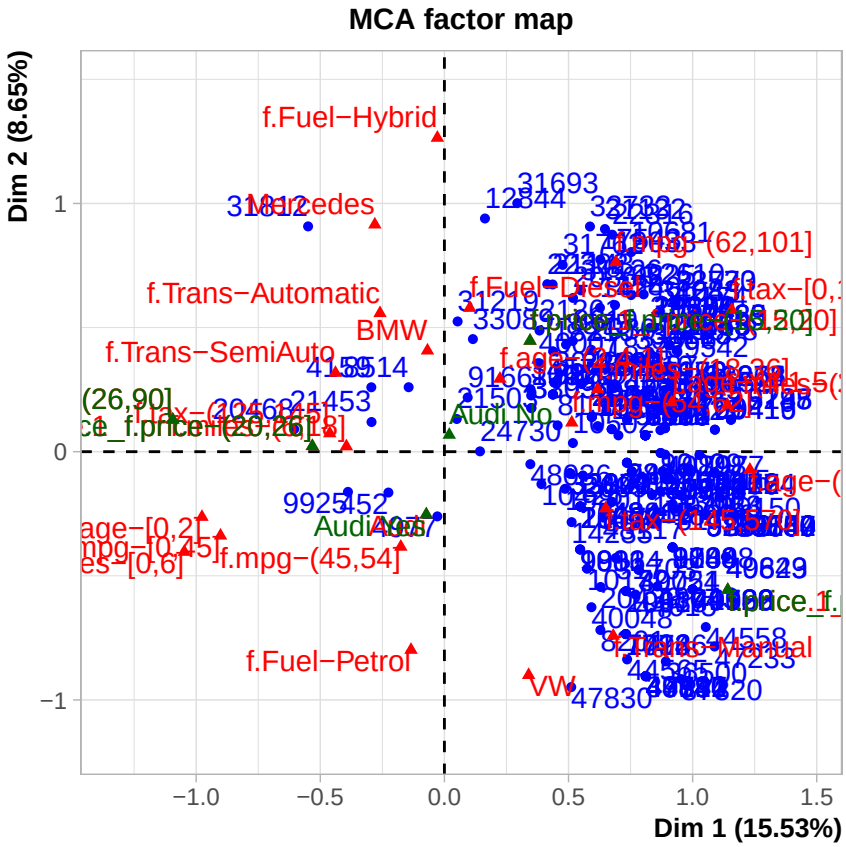

```
plot.MCA(res.mca,choix=c("ind"),cex=0.8)
```

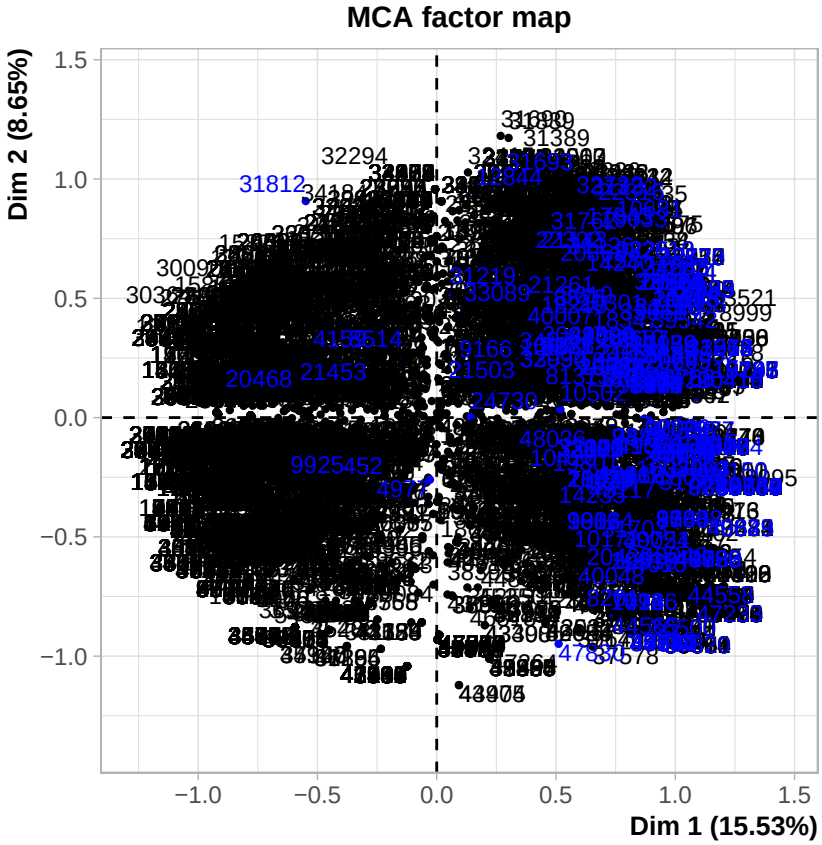


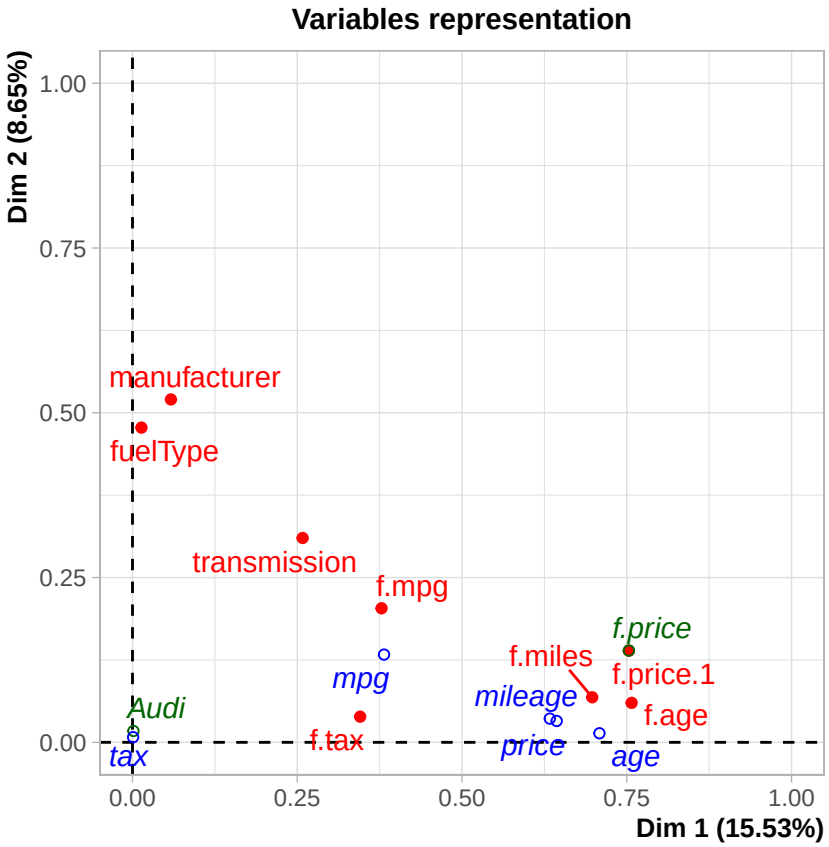
9.5 MCA with supplementary variables

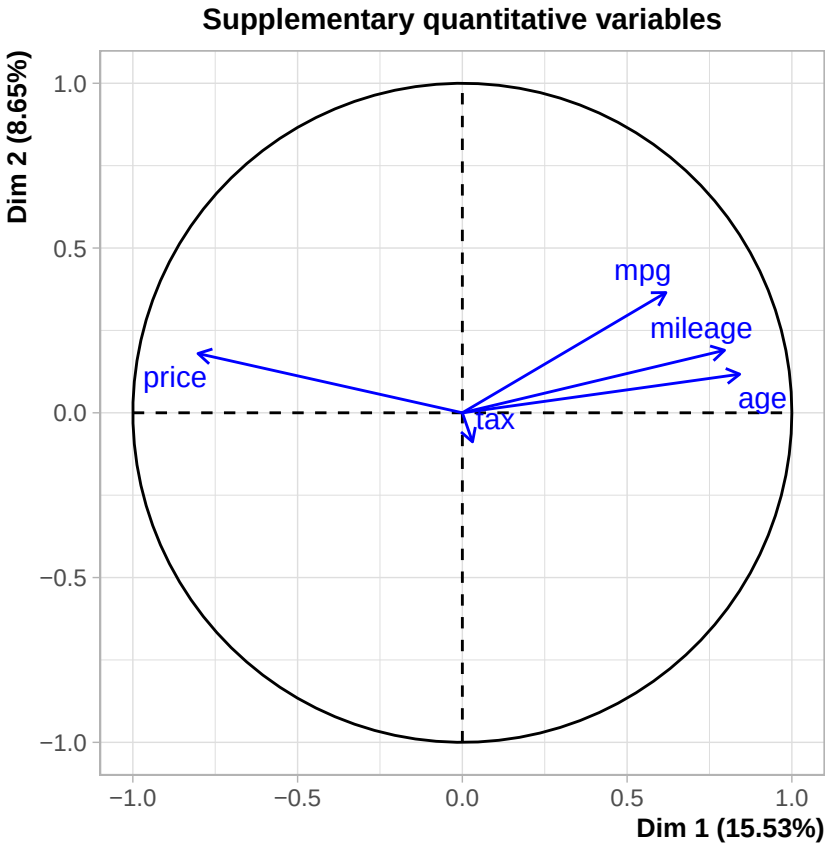
Fent MCA amb variables suplementaries numèriques i qualitatives podem veure que hi han alguns factors que aporten més al les dues primeres dimensions, com es veu en el gràfic de punts.

```
ll <- which( df$mout == "YesMOut")
res.mcaS<-MCA(df[,c("f.price","Audi",vars_dis[c(3:4, 6:11)],vars_con,"price") ],quali.sup=c(1,2),quanti.
```

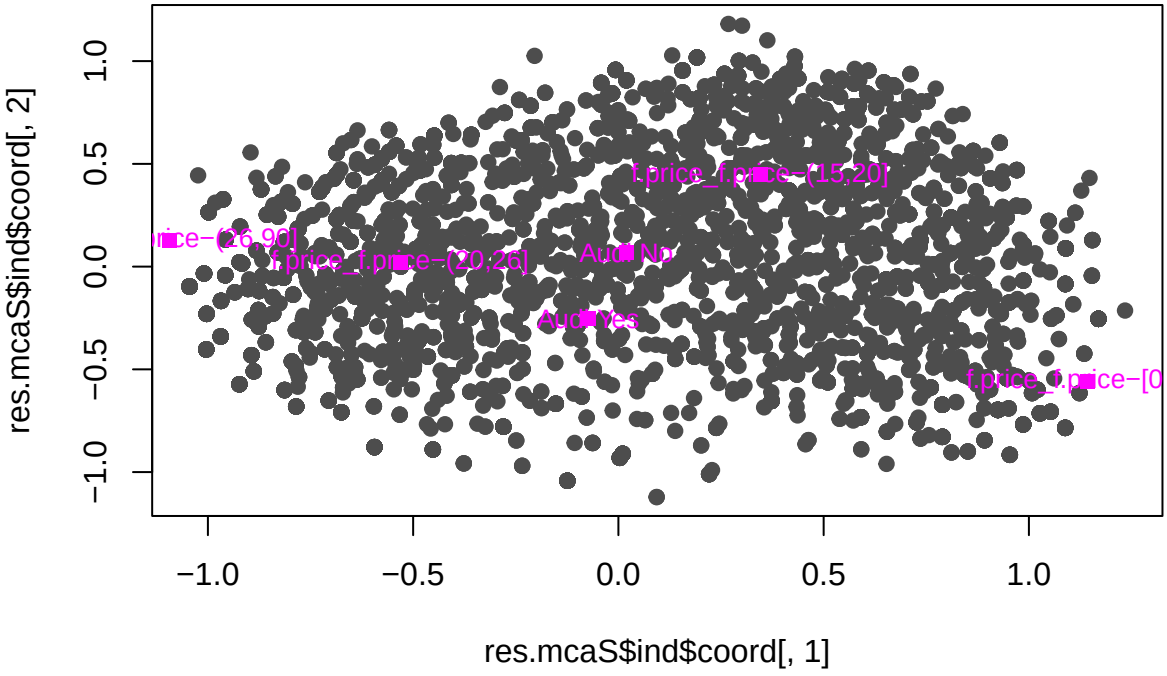









```
plot(res.mcaS$ind$coord[,1],res.mcaS$ind$coord[,2],pch=19,col="grey30")
points(res.mcaS$quali.sup$coord[,1],res.mcaS$quali.sup$coord[,2],pch=15,col="magenta")
text(res.mcaS$quali.sup$coord[,1],res.mcaS$quali.sup$coord[,2],labels=names(res.mcaS$quali.sup$coord[,1]
```




```
which(res.mcaS$eig[,1] > mean(res.mcaS$eig[,1]))
```

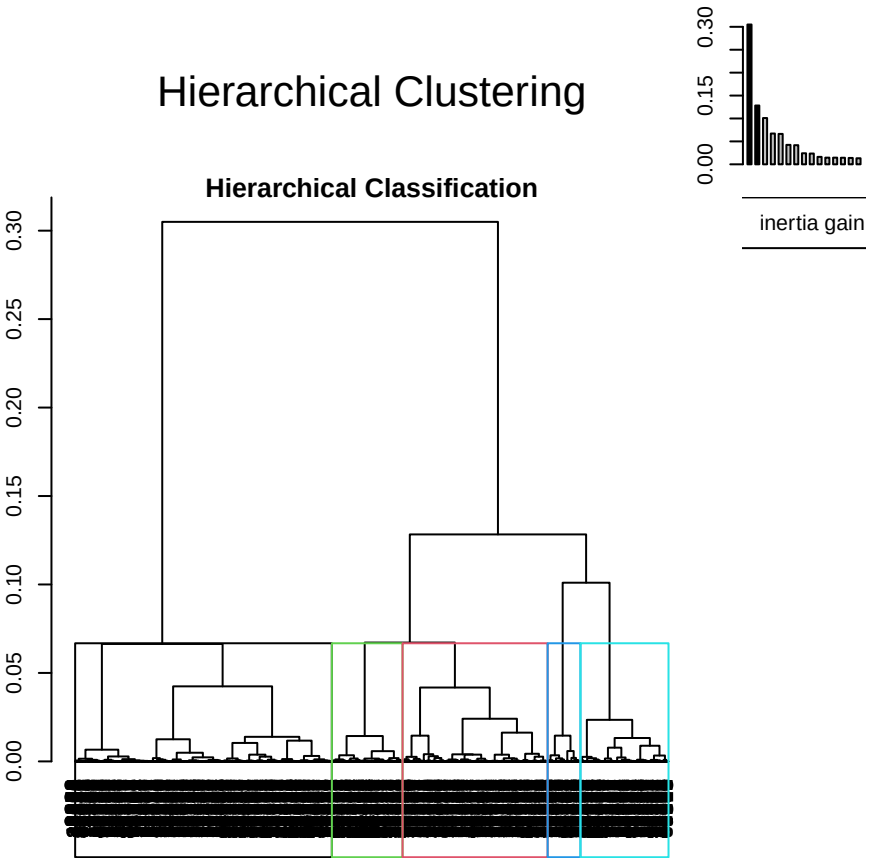
```
## dim 1 dim 2 dim 3 dim 4 dim 5 dim 6 dim 7 dim 8 dim 9
##     1     2     3     4     5     6     7     8     9
```

10 Hierarchical Clustering (from MCA)

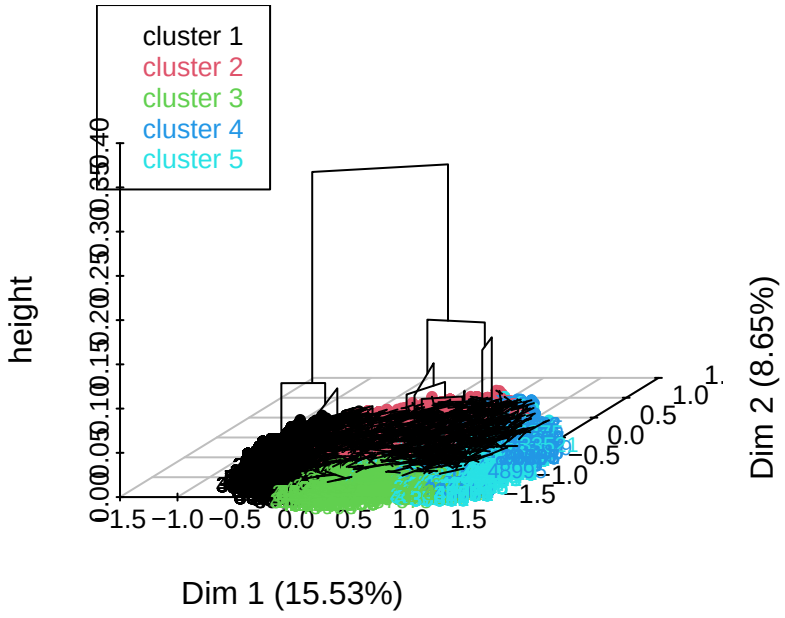
10.1 Description of clusters

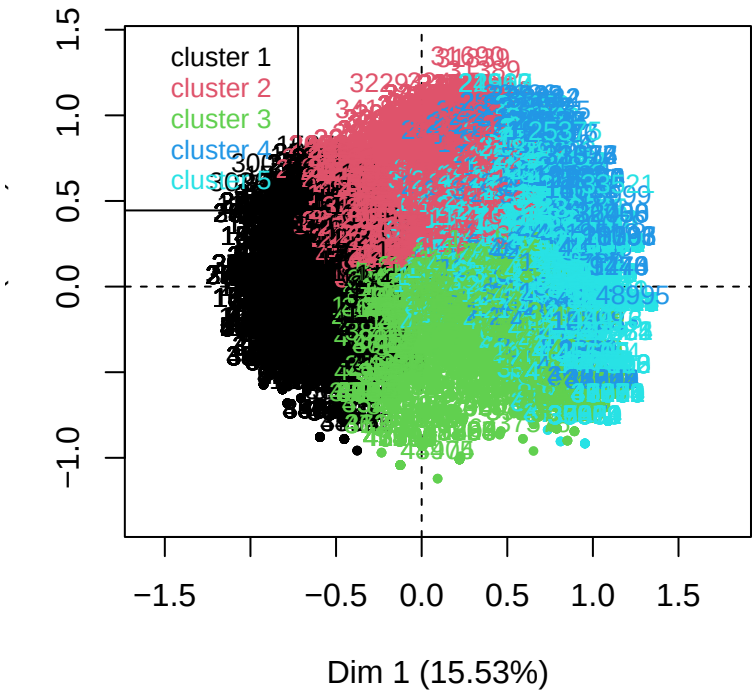
Els clusters es descriuen sobretot per les variables categoriques f.price, transmission fuelType, ... i per les variables numèriques price, mileage i age. Podem veure per cada un dels clusters quines categories el descriuen millor.

```
res.hcmc<-HCPC(res.mcaS,nb.clust=5,order=TRUE)
```



Hierarchical clustering on the factor map






```
df$claHCMC<-6
df[row.names(res.hcmc$data.clust),"claHCMC"]<-res.hcmc$data.clust$clust
df$claHCMC<-factor(df$claHCMC)
levels( df$claHCMC ) <- paste0( "f.claHCMC-",levels( df$claHCMC ))
summary(res.hcmc$data.clust$clust)
```

```
##    1    2    3    4    5
## 1768  997  968  272  731
```

```
table(df$claHCMC)
```

```
##
## f.claHCMC-1 f.claHCMC-2 f.claHCMC-3 f.claHCMC-4 f.claHCMC-5 f.claHCMC-6
##        1768         997         968         272         731         224
```

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

```
##                    p.value df
## f.price       0.000000e+00 12
## transmission  0.000000e+00  8
## manufacturer  0.000000e+00 12
## f.price.1     0.000000e+00 12
## f.miles       0.000000e+00 12
## f.tax         0.000000e+00  8
## f.mpg         0.000000e+00 12
## f.age         0.000000e+00 12
## fuelType     2.933954e-259  8
## Audi          2.575025e-11  4
```

```
res.hcmc$desc.var$category
```

```
## $`1`
##                                    Cla/Mod    Mod/Cla    Global
```

```
## f.age=f.age-[0,2]                   84.9135673 91.6855204 40.308277
## f.miles=f.miles-[0,6]               89.5881896 65.2149321 27.174831
## f.price.1=f.price.1_f.price-(26,90] 87.3459326 60.1244344 25.696791
## f.price=f.price_f.price-(26,90]     87.3459326 60.1244344 25.696791
## f.mpg=f.mpg-[0,45]                  76.8376068 50.8484163 24.704392
## f.tax=f.tax-(125,145]               52.4930748 85.7466063 60.979730
## transmission=f.Trans-SemiAuto       50.7261411 55.3167421 40.709459
## f.price.1=f.price.1_f.price-(20,26] 55.3727009 32.3529412 21.811655
## f.price=f.price_f.price-(20,26]     55.3727009 32.3529412 21.811655
## transmission=f.Trans-Automatic      46.9964664 30.0904977 23.902027
## f.miles=f.miles-(6,18]              46.0557769 32.6923077 26.499155
## fuelType=f.Fuel-Petrol              42.7083333 48.6990950 42.567568
## f.mpg=f.mpg-(45,54]                 43.8307030 34.5588235 29.434122
## manufacturer=Audi                   43.6000000 24.6606335 21.114865
## Audi=Audi Yes                       43.6000000 24.6606335 21.114865
## manufacturer=BMW                    42.9268293 24.8868778 21.642736
## Audi=Audi No                        35.6531049 75.3393665 78.885135
## fuelType=f.Fuel-Diesel              33.3458224 50.3393665 56.355574
## manufacturer=VW                     27.2477694 22.4547511 30.764358
## f.tax=f.tax-[0,125]                  0.0000000  0.0000000  5.532095
## f.mpg=f.mpg-(54,62]                 15.8969805 10.1244344 23.775338
## f.tax=f.tax-(145,570]               15.8890290 14.2533937 33.488176
## f.price.1=f.price.1_f.price-(15,20] 10.9075770  7.4095023 25.358953
## f.price=f.price_f.price-(15,20]     10.9075770  7.4095023 25.358953
## transmission=f.Trans-Manual         15.3937947 14.5927602 35.388514
## f.mpg=f.mpg-(62,101]                 7.5525813  4.4683258 22.086149
## f.age=f.age-(5.1,15]                 1.0071942  0.3959276 14.674831
## f.age=f.age-(4.1,5.1]                1.1703511  0.5090498 16.237331
## f.age=f.age-(2,4.1]                  9.6111519  7.4095023 28.779561
## f.miles=f.miles-(36,195]             0.9615385  0.5090498 19.763514
## f.miles=f.miles-(18,36]              2.2257552  1.5837104 26.562500
## f.price.1=f.price.1_f.price-[0,15]   0.1556420  0.1131222 27.132601
## f.price=f.price_f.price-[0,15]       0.1556420  0.1131222 27.132601
##                                          p.value     v.test
## f.age=f.age-[0,2]                    0.000000e+00        Inf
## f.miles=f.miles-[0,6]                0.000000e+00        Inf
## f.price.1=f.price.1_f.price-(26,90]  0.000000e+00        Inf
## f.price=f.price_f.price-(26,90]      0.000000e+00        Inf
## f.mpg=f.mpg-[0,45]                  5.109820e-226  32.095408
## f.tax=f.tax-(125,145]               1.616117e-173  28.082201
## transmission=f.Trans-SemiAuto        5.791765e-56  15.760776
## f.price.1=f.price.1_f.price-(20,26]  9.205409e-41  13.368767
## f.price=f.price_f.price-(20,26]      9.205409e-41  13.368767
## transmission=f.Trans-Automatic       2.264258e-14   7.634655
## f.miles=f.miles-(6,18]               1.430579e-13   7.393459
## fuelType=f.Fuel-Petrol               4.893262e-11   6.574148
## f.mpg=f.mpg-(45,54]                  2.892793e-09   5.937569
## manufacturer=Audi                    4.628374e-06   4.580964
## Audi=Audi Yes                        4.628374e-06   4.580964
## manufacturer=BMW                     3.225957e-05   4.156900
## Audi=Audi No                         4.628374e-06  -4.580964
## fuelType=f.Fuel-Diesel               1.248972e-10  -6.433258
## manufacturer=VW                      3.328105e-22  -9.689913
## f.tax=f.tax-[0,125]                  7.338051e-56 -15.745814
## f.mpg=f.mpg-(54,62]                  2.999329e-71 -17.847933
## f.tax=f.tax-(145,570]               4.140289e-112 -22.500111
## f.price.1=f.price.1_f.price-(15,20] 3.115088e-121 -23.413484
## f.price=f.price_f.price-(15,20]     3.115088e-121 -23.413484
## transmission=f.Trans-Manual         2.476421e-127 -24.004821
## f.mpg=f.mpg-(62,101]                3.478383e-134 -24.652283
## f.age=f.age-(5.1,15]                1.709335e-141 -25.324663
## f.age=f.age-(4.1,5.1]               4.281448e-156 -26.617296
## f.age=f.age-(2,4.1]                 7.231742e-159 -26.855722
## f.miles=f.miles-(36,195]            4.431045e-199 -30.102966
```

```
## f.miles=f.miles-(18,36]             2.002494e-255 -34.138182
## f.price.1=f.price.1_f.price-[0,15]  3.058477e-315 -37.956119
## f.price=f.price_f.price-[0,15]      3.058477e-315 -37.956119
##
## $`2`
##                                       Cla/Mod   Mod/Cla    Global       p.value
## f.age=f.age-(2,4.1]                 48.495965 66.298897 28.779561 1.339454e-175
## f.miles=f.miles-(18,36]             44.117647 55.667001 26.562500 1.814524e-110
## fuelType=f.Fuel-Diesel              31.734732 84.954865 56.355574 1.322666e-102
## manufacturer=Mercedes               41.945774 52.758275 26.478041  6.691717e-91
## f.mpg=f.mpg-(62,101]                44.646272 46.840522 22.086149  7.884440e-89
## f.price.1=f.price.1_f.price-(15,20] 41.465445 49.949850 25.358953  1.057925e-81
## f.price=f.price_f.price-(15,20]     41.465445 49.949850 25.358953  1.057925e-81
## transmission=f.Trans-SemiAuto       28.941909 55.967904 40.709459  6.924401e-28
## f.price.1=f.price.1_f.price-(20,26] 30.687318 31.795386 21.811655  9.211184e-17
## f.price=f.price_f.price-(20,26]     30.687318 31.795386 21.811655  9.211184e-17
## f.tax=f.tax-(125,145]               24.619114 71.313942 60.979730  2.191543e-14
## transmission=f.Trans-Automatic      29.151943 33.099298 23.902027  8.595853e-14
## Audi=Audi No                        22.938972 85.957874 78.885135  1.856679e-10
## f.miles=f.miles-(6,18]              25.976096 32.698094 26.499155  9.031700e-07
## f.age=f.age-(4.1,5.1]               25.747724 19.859579 16.237331  6.175207e-04
## manufacturer=BMW                    24.682927 25.376128 21.642736  1.480437e-03
## fuelType=f.Fuel-Hybrid              39.215686  2.006018  1.076858  3.110071e-03
## f.tax=f.tax-(145,570]               18.032787 28.686058 33.488176  2.665223e-04
## manufacturer=Audi                   14.000000 14.042126 21.114865  1.856679e-10
## Audi=Audi Yes                       14.000000 14.042126 21.114865  1.856679e-10
## f.mpg=f.mpg-(45,54]                 13.414634 18.756269 29.434122  8.318605e-18
## f.miles=f.miles-(36,195]            10.683761 10.030090 19.763514  3.277744e-20
## f.tax=f.tax-[0,125]                  0.000000  0.000000  5.532095  1.703821e-28
## f.price.1=f.price.1_f.price-(26,90] 10.435497 12.738215 25.696791  9.083323e-29
## f.price=f.price_f.price-(26,90]     10.435497 12.738215 25.696791  9.083323e-29
## f.age=f.age-(5.1,15]                 5.467626  3.811434 14.674831  1.460354e-34
## f.mpg=f.mpg-[0,45]                   8.119658  9.528586 24.704392  2.317729e-41
## f.price.1=f.price.1_f.price-[0,15]   4.280156  5.516550 27.132601  3.447018e-83
## f.price=f.price_f.price-[0,15]       4.280156  5.516550 27.132601  3.447018e-83
## manufacturer=VW                      5.353466  7.823470 30.764358  1.539874e-83
## transmission=f.Trans-Manual          6.503580 10.932798 35.388514  7.298834e-85
## fuelType=f.Fuel-Petrol               6.448413 13.039117 42.567568 9.800876e-112
## f.age=f.age-[0,2]                    5.238345 10.030090 40.308277 5.271623e-123
## f.miles=f.miles-[0,6]                1.243201  1.604814 27.174831 1.972013e-128
##                                         v.test
## f.age=f.age-(2,4.1]                  28.252148
## f.miles=f.miles-(18,36]              22.331804
## fuelType=f.Fuel-Diesel               21.507565
## manufacturer=Mercedes                20.218780
## f.mpg=f.mpg-(62,101]                 19.982095
## f.price.1=f.price.1_f.price-(15,20]  19.145366
## f.price=f.price_f.price-(15,20]      19.145366
## transmission=f.Trans-SemiAuto        10.946265
## f.price.1=f.price.1_f.price-(20,26]   8.314536
## f.price=f.price_f.price-(20,26]       8.314536
## f.tax=f.tax-(125,145]                 7.638859
## transmission=f.Trans-Automatic        7.460862
## Audi=Audi No                          6.372750
## f.miles=f.miles-(6,18]                4.911642
## f.age=f.age-(4.1,5.1]                 3.423800
## manufacturer=BMW                      3.178491
## fuelType=f.Fuel-Hybrid                2.956644
## f.tax=f.tax-(145,570]                -3.645838
## manufacturer=Audi                    -6.372750
## Audi=Audi Yes                        -6.372750
## f.mpg=f.mpg-(45,54]                  -8.595109
## f.miles=f.miles-(36,195]             -9.209458
## f.tax=f.tax-[0,125]                 -11.072601
```

```
## f.price.1=f.price.1_f.price-(26,90] -11.128817
## f.price=f.price_f.price-(26,90]     -11.128817
## f.age=f.age-(5.1,15]                -12.261356
## f.mpg=f.mpg-[0,45]                  -13.470979
## f.price.1=f.price.1_f.price-[0,15]  -19.322904
## f.price=f.price_f.price-[0,15]      -19.322904
## manufacturer=VW                     -19.364451
## transmission=f.Trans-Manual         -19.520868
## fuelType=f.Fuel-Petrol              -22.461856
## f.age=f.age-[0,2]                   -23.586749
## f.miles=f.miles-[0,6]               -24.109820
##
## $`3`
##                                        Cla/Mod    Mod/Cla    Global
## transmission=f.Trans-Manual         47.3150358 81.9214876 35.388514
## manufacturer=VW                     50.0343171 75.3099174 30.764358
## fuelType=f.Fuel-Petrol              39.8809524 83.0578512 42.567568
## f.price.1=f.price.1_f.price-[0,15]  42.6459144 56.6115702 27.132601
## f.price=f.price_f.price-[0,15]      42.6459144 56.6115702 27.132601
## f.miles=f.miles-(18,36]             36.8044515 47.8305785 26.562500
## f.age=f.age-(2,4.1]                 35.0696992 49.3801653 28.779561
## f.mpg=f.mpg-(54,62]                 36.3232682 42.2520661 23.775338
## f.tax=f.tax-(145,570]               31.7780580 52.0661157 33.488176
## f.age=f.age-(4.1,5.1]               31.9895969 25.4132231 16.237331
## f.price.1=f.price.1_f.price-(15,20] 27.2273106 33.7809917 25.358953
## f.price=f.price_f.price-(15,20]     27.2273106 33.7809917 25.358953
## f.mpg=f.mpg-(45,54]                 25.7532281 37.0867769 29.434122
## f.miles=f.miles-(6,18]              25.2589641 32.7479339 26.499155
## Audi=Audi No                        21.0653105 81.3016529 78.885135
## manufacturer=Audi                   18.1000000 18.6983471 21.114865
## Audi=Audi Yes                       18.1000000 18.6983471 21.114865
## fuelType=f.Fuel-Hybrid               0.0000000  0.0000000  1.076858
## f.age=f.age-(5.1,15]                 8.2014388  5.8884298 14.674831
## f.tax=f.tax-[0,125]                  1.5267176  0.4132231  5.532095
## f.tax=f.tax-(125,145]               15.9279778 47.5206612 60.979730
## f.mpg=f.mpg-(62,101]                 9.7514340 10.5371901 22.086149
## f.price.1=f.price.1_f.price-(20,26]  8.6156825  9.1942149 21.811655
## f.price=f.price_f.price-(20,26]      8.6156825  9.1942149 21.811655
## f.miles=f.miles-(36,195]             7.5854701  7.3347107 19.763514
## f.miles=f.miles-[0,6]                9.0909091 12.0867769 27.174831
## f.mpg=f.mpg-[0,45]                   8.3760684 10.1239669 24.704392
## f.age=f.age-[0,2]                    9.7957046 19.3181818 40.308277
## manufacturer=BMW                     4.0000000  4.2355372 21.642736
## transmission=f.Trans-SemiAuto        7.5207469 14.9793388 40.709459
## transmission=f.Trans-Automatic       2.6501767  3.0991736 23.902027
## manufacturer=Mercedes                1.3556619  1.7561983 26.478041
## f.price.1=f.price.1_f.price-(26,90]  0.3286771  0.4132231 25.696791
## f.price=f.price_f.price-(26,90]      0.3286771  0.4132231 25.696791
## fuelType=f.Fuel-Diesel               6.1446235 16.9421488 56.355574
##                                         p.value     v.test
## transmission=f.Trans-Manual         6.002397e-249  33.698884
## manufacturer=VW                     2.322745e-233  32.617419
## fuelType=f.Fuel-Petrol              8.805263e-186  29.069187
## f.price.1=f.price.1_f.price-[0,15]  6.312058e-108  22.068744
## f.price=f.price_f.price-[0,15]      6.312058e-108  22.068744
## f.miles=f.miles-(18,36]              2.149757e-58  16.110586
## f.age=f.age-(2,4.1]                  5.121694e-53  15.326054
## f.mpg=f.mpg-(54,62]                  2.211753e-47  14.458624
## f.tax=f.tax-(145,570]                4.305031e-41  13.425186
## f.age=f.age-(4.1,5.1]                1.060788e-16   8.297776
## f.price.1=f.price.1_f.price-(15,20]  4.151632e-11   6.598563
## f.price=f.price_f.price-(15,20]      4.151632e-11   6.598563
## f.mpg=f.mpg-(45,54]                  8.274706e-09   5.762765
## f.miles=f.miles-(6,18]               1.184915e-06   4.858144
```

```
## Audi=Audi No                           3.745143e-02   2.080809
## manufacturer=Audi                      3.745143e-02  -2.080809
## Audi=Audi Yes                          3.745143e-02  -2.080809
## fuelType=f.Fuel-Hybrid                 8.042290e-06  -4.464055
## f.age=f.age-(5.1,15]                   4.580119e-21  -9.418400
## f.tax=f.tax-[0,125]                    1.714199e-21  -9.521061
## f.tax=f.tax-(125,145]                  1.665242e-21  -9.524071
## f.mpg=f.mpg-(62,101]                   5.075929e-25 -10.331502
## f.price.1=f.price.1_f.price-(20,26]    2.243089e-30 -11.454089
## f.price=f.price_f.price-(20,26]        2.243089e-30 -11.454089
## f.miles=f.miles-(36,195]               2.232884e-32 -11.846904
## f.miles=f.miles-[0,6]                  2.729262e-36 -12.579740
## f.mpg=f.mpg-[0,45]                     1.040603e-36 -12.655688
## f.age=f.age-[0,2]                      3.717927e-54 -15.495545
## manufacturer=BMW                       9.913083e-63 -16.716654
## transmission=f.Trans-SemiAuto          1.944006e-82 -19.233414
## transmission=f.Trans-Automatic         7.234165e-86 -19.638617
## manufacturer=Mercedes                 1.435105e-117 -23.051067
## f.price.1=f.price.1_f.price-(26,90] 4.633631e-134 -24.640667
## f.price=f.price_f.price-(26,90]     4.633631e-134 -24.640667
## fuelType=f.Fuel-Diesel              4.410941e-176 -28.291387
##
## $`4`
##                                        Cla/Mod    Mod/Cla    Global
## f.tax=f.tax-[0,125]                 98.09160305 94.4852941  5.532095
## f.mpg=f.mpg-(54,62]                 16.69626998 69.1176471 23.775338
## f.miles=f.miles-(36,195]            16.23931624 55.8823529 19.763514
## f.age=f.age-(4.1,5.1]               14.95448635 42.2794118 16.237331
## f.age=f.age-(5.1,15]                14.82014388 37.8676471 14.674831
## f.price.1=f.price.1_f.price-(15,20] 10.65778518 47.0588235 25.358953
## f.price=f.price_f.price-(15,20]     10.65778518 47.0588235 25.358953
## f.price.1=f.price.1_f.price-[0,15]   9.10505837 43.0147059 27.132601
## f.price=f.price_f.price-[0,15]       9.10505837 43.0147059 27.132601
## manufacturer=BMW                     9.07317073 34.1911765 21.642736
## fuelType=f.Fuel-Diesel               6.78156613 66.5441176 56.355574
## f.miles=f.miles-(18,36]              7.71065183 35.6617647 26.562500
## fuelType=f.Fuel-Hybrid              13.72549020  2.5735294  1.076858
## f.age=f.age-(2,4.1]                  3.96184886 19.8529412 28.779561
## fuelType=f.Fuel-Petrol               4.16666667 30.8823529 42.567568
## f.price.1=f.price.1_f.price-(20,26]  2.51694095  9.5588235 21.811655
## f.price=f.price_f.price-(20,26]      2.51694095  9.5588235 21.811655
## manufacturer=VW                      2.74536719 14.7058824 30.764358
## f.miles=f.miles-(6,18]               1.83266932  8.4558824 26.499155
## f.mpg=f.mpg-(62,101]                 0.09560229  0.3676471 22.086149
## f.price.1=f.price.1_f.price-(26,90]  0.08216927  0.3676471 25.696791
## f.price=f.price_f.price-(26,90]      0.08216927  0.3676471 25.696791
## f.mpg=f.mpg-[0,45]                   0.00000000  0.0000000 24.704392
## f.tax=f.tax-(145,570]                0.63051702  3.6764706 33.488176
## f.miles=f.miles-[0,6]                0.00000000  0.0000000 27.174831
## f.age=f.age-[0,2]                    0.00000000  0.0000000 40.308277
## f.tax=f.tax-(125,145]                0.17313019  1.8382353 60.979730
##                                         p.value     v.test
## f.tax=f.tax-[0,125]                  0.000000e+00        Inf
## f.mpg=f.mpg-(54,62]                  3.044012e-60  16.371737
## f.miles=f.miles-(36,195]             2.364314e-42  13.638482
## f.age=f.age-(4.1,5.1]                4.733059e-26  10.556639
## f.age=f.age-(5.1,15]                 1.718407e-22   9.757196
## f.price.1=f.price.1_f.price-(15,20]  1.700644e-15   7.961428
## f.price=f.price_f.price-(15,20]      1.700644e-15   7.961428
## f.price.1=f.price.1_f.price-[0,15]   6.450947e-09   5.804627
## f.price=f.price_f.price-[0,15]       6.450947e-09   5.804627
## manufacturer=BMW                     9.043695e-07   4.911381
## fuelType=f.Fuel-Diesel               4.298639e-04   3.521036
## f.miles=f.miles-(18,36]              6.726854e-04   3.400476
```

```
## fuelType=f.Fuel-Hybrid                3.304220e-02   2.131570
## f.age=f.age-(2,4.1]                   5.642804e-04  -3.448227
## fuelType=f.Fuel-Petrol                4.751535e-05  -4.067524
## f.price.1=f.price.1_f.price-(20,26]   4.893138e-08  -5.455151
## f.price=f.price_f.price-(20,26]       4.893138e-08  -5.455151
## manufacturer=VW                       3.508106e-10  -6.274477
## f.miles=f.miles-(6,18]                1.696124e-14  -7.671790
## f.mpg=f.mpg-(62,101]                  2.803362e-29 -11.233140
## f.price.1=f.price.1_f.price-(26,90]   5.089806e-35 -12.346467
## f.price=f.price_f.price-(26,90]       5.089806e-35 -12.346467
## f.mpg=f.mpg-[0,45]                    2.091714e-35 -12.417825
## f.tax=f.tax-(145,570]                 1.605193e-35 -12.438991
## f.miles=f.miles-[0,6]                 1.656206e-39 -13.152051
## f.age=f.age-[0,2]                     4.389419e-64 -16.901455
## f.tax=f.tax-(125,145]                3.576006e-106 -21.885433
##
## $`5`
##                                          Cla/Mod    Mod/Cla    Global
## f.miles=f.miles-(36,195]              64.52991453 82.6265390 19.763514
## f.age=f.age-(5.1,15]                  70.50359712 67.0314637 14.674831
## f.price.1=f.price.1_f.price-[0,15]    43.81322957 77.0177839 27.132601
## f.price=f.price_f.price-[0,15]        43.81322957 77.0177839 27.132601
## f.tax=f.tax-(145,570]                 33.66960908 73.0506156 33.488176
## f.mpg=f.mpg-(62,101]                  37.95411090 54.3091655 22.086149
## fuelType=f.Fuel-Diesel                21.99325590 80.3009576 56.355574
## transmission=f.Trans-Manual           24.46300716 56.0875513 35.388514
## f.age=f.age-(4.1,5.1]                 26.13784135 27.4965800 16.237331
## manufacturer=BMW                      19.31707317 27.0861833 21.642736
## manufacturer=Audi                     18.70000000 25.5813953 21.114865
## Audi=Audi Yes                         18.70000000 25.5813953 21.114865
## Audi=Audi No                          14.56102784 74.4186047 78.885135
## f.mpg=f.mpg-(45,54]                   11.04734577 21.0670315 29.434122
## manufacturer=Mercedes                 10.60606061 18.1942544 26.478041
## f.price.1=f.price.1_f.price-(15,20]    9.74188177 16.0054720 25.358953
## f.price=f.price_f.price-(15,20]        9.74188177 16.0054720 25.358953
## f.mpg=f.mpg-(54,62]                    9.05861456 13.9534884 23.775338
## f.miles=f.miles-(18,36]                9.14149444 15.7318741 26.562500
## f.tax=f.tax-[0,125]                    0.38167939  0.1367989  5.532095
## f.mpg=f.mpg-[0,45]                     6.66666667 10.6703146 24.704392
## transmission=f.Trans-SemiAuto          7.72821577 20.3830369 40.709459
## f.price.1=f.price.1_f.price-(20,26]    2.80735721  3.9671683 21.811655
## f.price=f.price_f.price-(20,26]        2.80735721  3.9671683 21.811655
## fuelType=f.Fuel-Petrol                 6.79563492 18.7414501 42.567568
## f.age=f.age-(2,4.1]                    2.86133529  5.3351573 28.779561
## f.price.1=f.price.1_f.price-(26,90]    1.80772391  3.0095759 25.696791
## f.price=f.price_f.price-(26,90]        1.80772391  3.0095759 25.696791
## f.miles=f.miles-(6,18]                 0.87649402  1.5047880 26.499155
## f.tax=f.tax-(125,145]                  6.78670360 26.8125855 60.979730
## f.miles=f.miles-[0,6]                  0.07770008  0.1367989 27.174831
## f.age=f.age-[0,2]                      0.05238345  0.1367989 40.308277
##                                           p.value     v.test
## f.miles=f.miles-(36,195]               0.000000e+00        Inf
## f.age=f.age-(5.1,15]                  4.073085e-307  37.460364
## f.price.1=f.price.1_f.price-[0,15]    1.310356e-212  31.119546
## f.price=f.price_f.price-[0,15]        1.310356e-212  31.119546
## f.tax=f.tax-(145,570]                 7.824320e-127  23.956932
## f.mpg=f.mpg-(62,101]                   1.581831e-98  21.067463
## fuelType=f.Fuel-Diesel                 4.013285e-49  14.732053
## transmission=f.Trans-Manual            1.449155e-35  12.447157
## f.age=f.age-(4.1,5.1]                  2.666093e-17   8.460336
## manufacturer=BMW                       1.430133e-04   3.802898
## manufacturer=Audi                      1.582989e-03   3.159023
## Audi=Audi Yes                          1.582989e-03   3.159023
## Audi=Audi No                           1.582989e-03  -3.159023
```

```
## f.mpg=f.mpg-(45,54]                  3.118762e-08  -5.534636
## manufacturer=Mercedes                1.262019e-08  -5.691129
## f.price.1=f.price.1_f.price-(15,20]  4.991801e-11  -6.571180
## f.price=f.price_f.price-(15,20]      4.991801e-11  -6.571180
## f.mpg=f.mpg-(54,62]                  1.187092e-12  -7.106864
## f.miles=f.miles-(18,36]              5.131135e-14  -7.528539
## f.tax=f.tax-[0,125]                  1.130279e-18  -8.821409
## f.mpg=f.mpg-[0,45]                   8.986680e-25 -10.276572
## transmission=f.Trans-SemiAuto        1.658751e-36 -12.619019
## f.price.1=f.price.1_f.price-(20,26]  3.651745e-48 -14.582088
## f.price=f.price_f.price-(20,26]      3.651745e-48 -14.582088
## fuelType=f.Fuel-Petrol               2.817281e-49 -14.755942
## f.age=f.age-(2,4.1]                  2.515280e-66 -17.203139
## f.price.1=f.price.1_f.price-(26,90]  7.911744e-71 -17.793674
## f.price=f.price_f.price-(26,90]      7.911744e-71 -17.793674
## f.miles=f.miles-(6,18]               2.071199e-88 -19.933823
## f.tax=f.tax-(125,145]                1.007892e-92 -20.424742
## f.miles=f.miles-[0,6]               2.779933e-109 -22.209503
## f.age=f.age-[0,2]                   6.141039e-181 -28.683430
```

```
res.hcmc$desc.var$quanti.var
```

```
##              Eta2       P-value
## mileage 0.6682386  0.000000e+00
## mpg     0.3311356  0.000000e+00
## age     0.6836936  0.000000e+00
## price   0.5293610  0.000000e+00
## tax     0.2585553 2.873711e-305
```

```
res.hcmc$desc.var$quanti
```

```
## $`1`
##             v.test Mean in category Overall mean sd in category   Overall sd
## price    46.540437      29905.35068 21486.968961   8802.6662898  9606.547493
## tax      -5.822917        145.80703   147.045162      3.1597187    11.292649
## mpg     -37.795503         44.99226    52.924126      8.9251065    11.145627
## mileage -44.636901       5593.12274 20567.651340   5400.0775177 17816.742468
## age     -47.813874          1.96776     3.595831      0.6561415     1.808375
##              p.value
## price   0.000000e+00
## tax     5.782921e-09
## mpg     0.000000e+00
## mileage 0.000000e+00
## age     0.000000e+00
##
## $`2`
##            v.test Mean in category Overall mean sd in category   Overall sd
## mpg     19.800247        59.134894    52.924126   9.905002e+00    11.145627
## mileage  5.561355     23356.205302 20567.651340   1.095282e+04 17816.742468
## age      5.484670         3.874963     3.595831   9.972515e-01     1.808375
## tax      4.096036       148.346918   147.045162   1.135756e+01    11.292649
## price   -1.992745     20948.216650 21486.968961   5.277327e+03  9606.547493
##              p.value
## mpg     2.962386e-87
## mileage 2.676886e-08
## age     4.142412e-08
## tax     4.202838e-05
## price   4.628942e-02
##
## $`3`
##           v.test Mean in category Overall mean sd in category  Overall sd
## tax     5.950041       148.971682   147.045162      10.689778   11.292649
```

```
## mpg     4.899594        54.489876    52.924126       7.562348   11.145627
## age     4.128975         3.809917     3.595831       1.263369    1.808375
## price -24.759453     14667.255165 21486.968961    4319.838528 9606.547493
##             p.value
## tax    2.680750e-09
## mpg    9.603469e-07
## age    3.643847e-05
## price 2.452527e-135
##
## $`4`
##             v.test Mean in category Overall mean sd in category   Overall sd
## age      18.291448         5.543227     3.595831       1.328805     1.808375
## mileage  17.433261     38853.898678 20567.651340   15485.023039 17816.742468
## mpg       5.287442        56.393633    52.924126       3.318060    11.145627
## price   -10.829053     15362.400735 21486.968961    4217.534184  9606.547493
## tax     -31.333978       126.213255   147.045162       5.036066    11.292649
##               p.value
## age      9.680826e-75
## mileage  4.613804e-68
## mpg      1.240385e-07
## price    2.507340e-27
## tax     1.608309e-215
##
## $`5`
##            v.test Mean in category Overall mean sd in category   Overall sd
## mileage  43.54534     46958.455017 20567.651340   13510.241703 17816.742468
## age      41.43552         6.144679     3.595831       1.238291     1.808375
## mpg      19.38367        60.273038    52.924126      11.490189    11.145627
## tax      16.71150       153.464553   147.045162      15.866504    11.292649
## price   -25.44948     13170.693570 21486.968961    4966.973255  9606.547493
##               p.value
## mileage  0.000000e+00
## age      0.000000e+00
## mpg      1.060129e-83
## tax      1.080847e-62
## price   7.153308e-143
```

10.2 Parangons and class-specific individuals

Donats els individus més representatius de cada classe, podem representar-los gràficament.

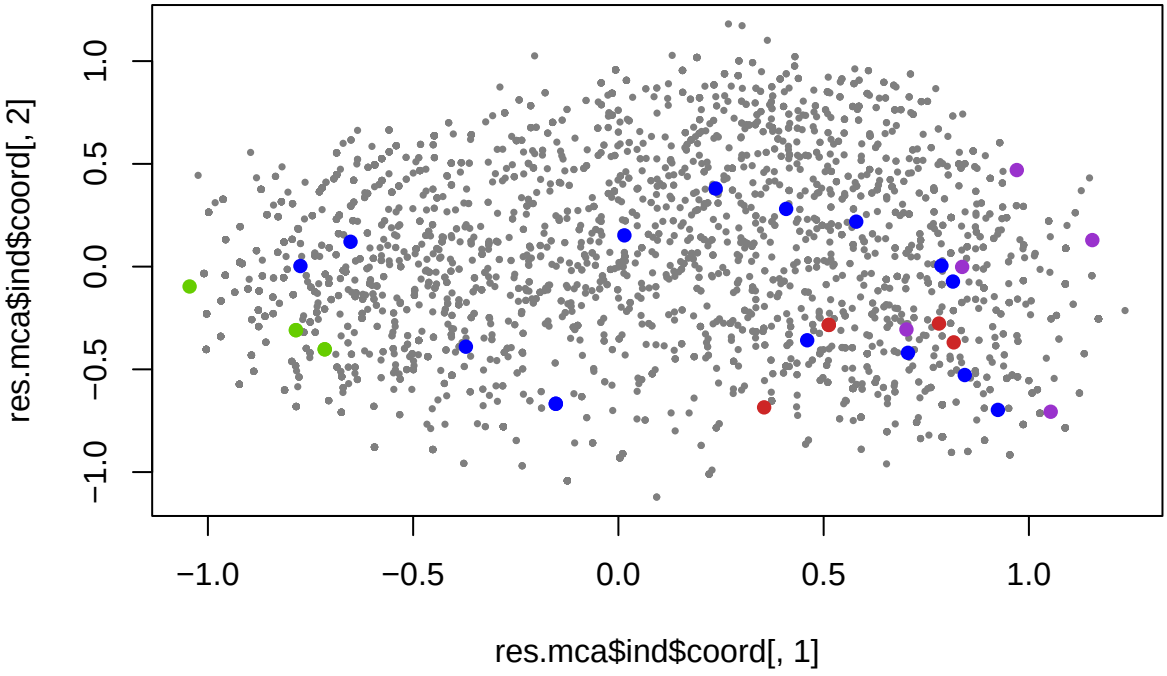
```
res.hcmc$desc.ind$para
```

```
## Cluster: 1
##     18928     22729     25630       443       476
## 0.1803942 0.2189477 0.2189477 0.2342737 0.2342737
## ------------------------------------------------------------
## Cluster: 2
##     34537     22123     32516     10713     11751
## 0.1214907 0.2578165 0.2769244 0.2833911 0.2833911
## ------------------------------------------------------------
## Cluster: 3
##     41491     41616     43758       109     16245
## 0.1682779 0.1682779 0.1682779 0.1919294 0.1936337
## ------------------------------------------------------------
## Cluster: 4
##     32079      1553     24116      5024      6437
## 0.2343599 0.2934321 0.3089505 0.3177295 0.3177295
## ------------------------------------------------------------
## Cluster: 5
##      8107      1018      1272      1738      2078
## 0.2823640 0.3062242 0.3062242 0.3062242 0.3062242
```

```
res.hcmc$desc.ind$dist
```

```
## Cluster: 1
##    13960     7482     7710     7854    10062
## 1.780210 1.721864 1.721864 1.721864 1.721864
## ------------------------------------------------------------
## Cluster: 2
##    21451    34317    31839    32212    32415
## 1.800565 1.800565 1.784303 1.769162 1.717596
## ------------------------------------------------------------
## Cluster: 3
##     4596     2019     5577     9058    36431
## 1.506109 1.499239 1.499239 1.489441 1.461653
## ------------------------------------------------------------
## Cluster: 4
##    19645    20127    20134    10855    11006
## 1.956803 1.956803 1.956803 1.892389 1.892389
## ------------------------------------------------------------
## Cluster: 5
##    11641    12372    13362    15153    16656
## 1.744774 1.744774 1.744774 1.744774 1.744774
```

```
#### Characteristic individuals
para1<-which(rownames(res.mca$ind$coord)%in%names(res.hcpc$desc.ind$para[[1]]))
dist1<-which(rownames(res.mca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[1]]))
para2<-which(rownames(res.mca$ind$coord)%in%names(res.hcpc$desc.ind$para[[2]]))
dist2<-which(rownames(res.mca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[2]]))
para3<-which(rownames(res.mca$ind$coord)%in%names(res.hcpc$desc.ind$para[[3]]))
dist3<-which(rownames(res.mca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[3]]))
plot(res.mca$ind$coord[,1],res.mca$ind$coord[,2],col="grey50",cex=0.5,pch=16)
points(res.mca$ind$coord[para1,1],res.mca$ind$coord[para1,2],col="blue",cex=1,pch=16)
points(res.mca$ind$coord[dist1,1],res.mca$ind$coord[dist1,2],col="chartreuse3",cex=1,pch=16)
points(res.mca$ind$coord[para2,1],res.mca$ind$coord[para2,2],col="blue",cex=1,pch=16)
points(res.mca$ind$coord[dist2,1],res.mca$ind$coord[dist2,2],col="darkorchid3",cex=1,pch=16)
points(res.mca$ind$coord[para3,1],res.mca$ind$coord[para3,2],col="blue",cex=1,pch=16)
points(res.mca$ind$coord[dist3,1],res.mca$ind$coord[dist3,2],col="firebrick3",cex=1,pch=16)
```



10.3 Comparison of clusters obtained after Hierarchical Clustering (based on PCA) focusing on f.price target

Comparem el price en les dues classificacions usant el Eta2. Podem veure que en el hcpc la variable price no caracteritza molt als clusters, doncs tax o age tenen un valor de Eta2 més gran. En canvi a hcmc te un valor relativament alt, i per tant més relacio amb els clusters.

```
res.hcpc$desc.var$quanti.var
```

```
##              Eta2 P-value
## price   0.3675013       0
## mileage 0.6025290       0
## tax     0.7078747       0
## mpg     0.3420681       0
## age     0.6478671       0
```

```
res.hcmc$desc.var$quanti.var
```

```
##              Eta2       P-value
## mileage 0.6682386  0.000000e+00
## mpg     0.3311356  0.000000e+00
## age     0.6836936  0.000000e+00
## price   0.5293610  0.000000e+00
## tax     0.2585553 2.873711e-305
```

10.4 Comparison of clusters obtained after Hierarchical Clustering (based on PCA) focusing on Audi binary target

Comparem la variable binaria Audi YES/NO mirant la seva representació. En el cluster 3 de hcpc hi ha una sobre-representació del Audi YES. En el cas de hcmc, en el cluster 1 està sobre-representat el Audi YES, en el cluster 2 el Audi NO i en el 3 el Audi NO.

```
res.hcpc$desc.var$category$`3`["Audi=Audi No",]
```

```
##       Cla/Mod       Mod/Cla        Global       p.value        v.test
##  2.408994e+00  6.293706e+01  7.888514e+01  9.344095e-06 -4.431820e+00
```

```
res.hcpc$desc.var$category$`3`["Audi=Audi Yes",]
```

```
##      Cla/Mod      Mod/Cla       Global      p.value       v.test
## 5.300000e+00 3.706294e+01 2.111486e+01 9.344095e-06 4.431820e+00
```

```
res.hcmc$desc.var$category$`1`["Audi=Audi No",]
```

```
##       Cla/Mod       Mod/Cla        Global       p.value        v.test
##  3.565310e+01  7.533937e+01  7.888514e+01  4.628374e-06 -4.580964e+00
```

```
res.hcmc$desc.var$category$`1`["Audi=Audi Yes",]
```

```
##      Cla/Mod      Mod/Cla       Global      p.value       v.test
## 4.360000e+01 2.466063e+01 2.111486e+01 4.628374e-06 4.580964e+00
```

```
res.hcmc$desc.var$category$`2`["Audi=Audi No",]
```

```
##      Cla/Mod      Mod/Cla       Global      p.value       v.test
## 2.293897e+01 8.595787e+01 7.888514e+01 1.856679e-10 6.372750e+00
```

```
res.hcmc$desc.var$category$`2`["Audi=Audi Yes",]
```

```
##       Cla/Mod       Mod/Cla        Global       p.value        v.test
##  1.400000e+01  1.404213e+01  2.111486e+01  1.856679e-10 -6.372750e+00
```

```
res.hcmc$desc.var$category$`3`["Audi=Audi No",]
```

```
##     Cla/Mod     Mod/Cla      Global     p.value      v.test
## 21.06531049 81.30165289 78.88513514  0.03745143  2.08080862
```

```
res.hcmc$desc.var$category$`3`["Audi=Audi Yes",]
```

```
##     Cla/Mod     Mod/Cla      Global     p.value      v.test
## 18.10000000 18.69834711 21.11486486  0.03745143 -2.08080862
```

11 General Regression Model

Segons el resultat de la Multivariate Analisys fet a les entregues anteriors, construim un model de regressió lineal inicial sense cap transformació de les variables. Tal com expressa l'anunciat, els models d'aquest apartat han de tenir dues variables numèriques a les que anomenarem covariables. També haurem d'afegir als models les variables factors més significants. L'últim pas per la modelització serà afegir transformacions al model tal que hi hagi una interacció entre dos factors i una altra entre un factor i una covariable.

11.1 Primer model amb covariables

Per construir un primer model mínimament efectiu, mirem els resultats de MVA de les entregues anteriors. Aquest ens diu que les variables numèriques (a partir d'ara covariables) més relacionades amb la variable target Price són age, mpg i, amb menys relació, tax i mileage.

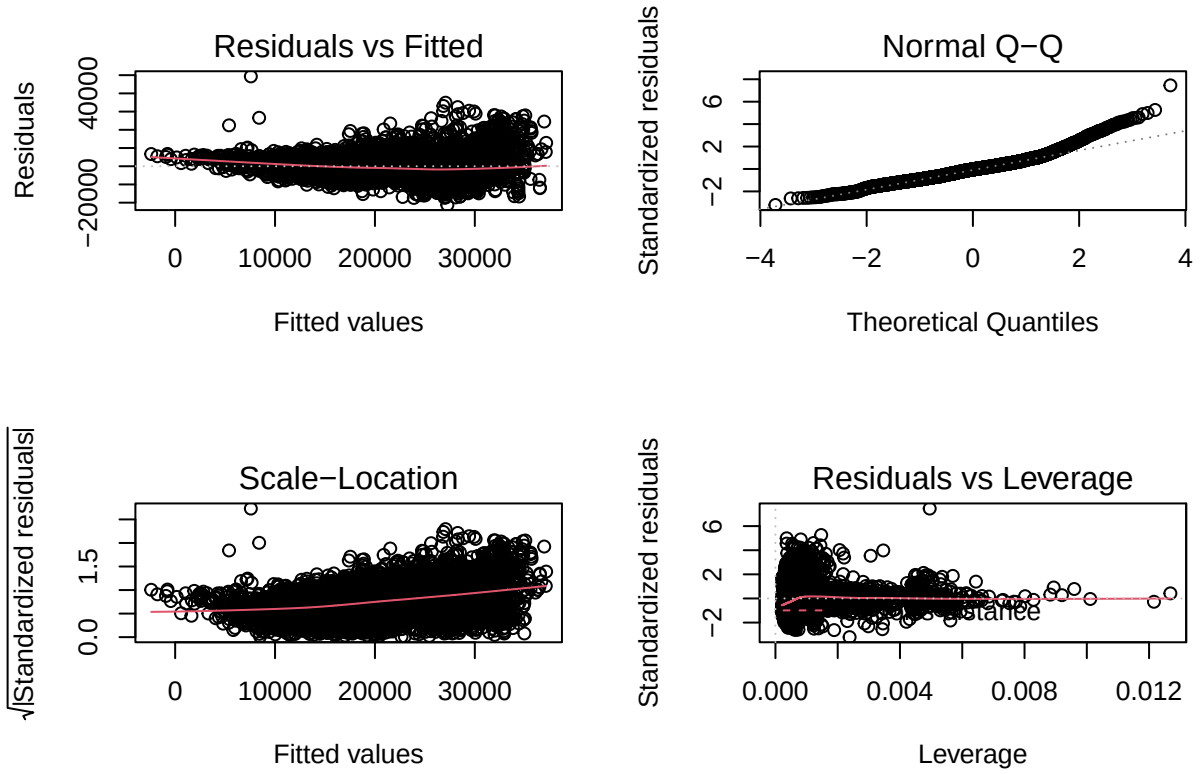
```r
m1<-lm(price~tax+mpg+age+mileage,data=df)
summary(m1)
```

```
##
## Call:
## lm(formula = price ~ tax + mpg + age + mileage, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -21256  -4539   -266   3270  49323
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.872e+04  1.450e+03  33.595  < 2e-16 ***
## tax          1.795e+00  8.711e+00   0.206    0.837
## mpg         -3.654e+02  9.185e+00 -39.777  < 2e-16 ***
## age         -2.051e+03  7.923e+01 -25.882  < 2e-16 ***
## mileage     -3.560e-02  7.345e-03  -4.847 1.29e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6633 on 4955 degrees of freedom
## Multiple R-squared:  0.5427, Adjusted R-squared:  0.5423
## F-statistic:  1470 on 4 and 4955 DF,  p-value: < 2.2e-16
```

```r
Anova(m1)
```

```
## Anova Table (Type II tests)
##
## Response: price
##               Sum Sq   Df   F value    Pr(>F)
## tax       1.8677e+06    1    0.0424    0.8368
## mpg       6.9620e+10    1 1582.2107 < 2.2e-16 ***
## age       2.9475e+10    1  669.8657 < 2.2e-16 ***
## mileage   1.0335e+09    1   23.4889 1.295e-06 ***
## Residuals 2.1803e+11 4955
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow=c(2,2))
plot(m1,id.n=0)
```




```
par(mfrow=c(1,1))
```

Com es pot veure, les variables seleccionades expliquen un 54% de la variablilitat del preu dels cotxes, la que menys contribueix al model es tax. Amb els plots podem veure que els residus sembla que segueixen una distribució normal bastant exacte i una certa homoscedasticitat, però provarem de transformar les variables per trobar una millor explicabilitat pel model.

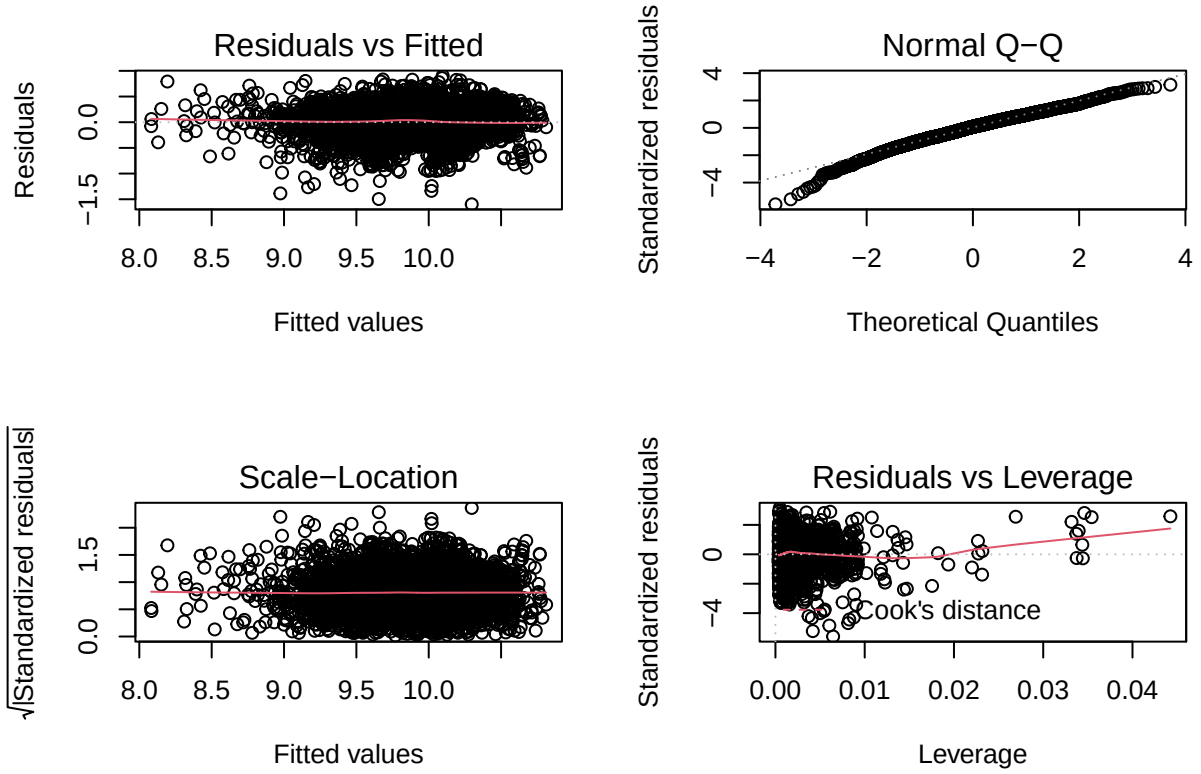
```
m2<-lm(log(price)~poly(tax,2)+poly(mpg,2)+poly(age,2)+sqrt(mileage),data=df)
summary(m2)
```

```
##
## Call:
## lm(formula = log(price) ~ poly(tax, 2) + poly(mpg, 2) + poly(age,
##     2) + sqrt(mileage), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.59870 -0.17975  0.02137  0.19780  0.90445
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     9.976e+00  1.506e-02 662.320  < 2e-16 ***
## poly(tax, 2)1   1.453e+00  3.095e-01   4.694 2.75e-06 ***
## poly(tax, 2)2   2.238e+00  2.970e-01   7.537 5.71e-14 ***
## poly(mpg, 2)1 -1.195e+01  3.401e-01 -35.126  < 2e-16 ***
## poly(mpg, 2)2  4.603e+00  2.915e-01  15.793  < 2e-16 ***
## poly(age, 2)1 -1.685e+01  5.398e-01 -31.213  < 2e-16 ***
## poly(age, 2)2 -3.457e+00  3.275e-01 -10.555  < 2e-16 ***
## sqrt(mileage) -9.382e-04  1.080e-04  -8.689  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2869 on 4952 degrees of freedom
## Multiple R-squared:  0.6311, Adjusted R-squared:  0.6306
## F-statistic:  1210 on 7 and 4952 DF,  p-value: < 2.2e-16
```

```
Anova(m2)
```

```
## Anova Table (Type II tests)
##
## Response: log(price)
##                Sum Sq   Df F value    Pr(>F)
## poly(tax, 2)     6.24    2  37.910 < 2.2e-16 ***
## poly(mpg, 2)   123.36    2 749.086 < 2.2e-16 ***
## poly(age, 2)   118.34    2 718.648 < 2.2e-16 ***
## sqrt(mileage)    6.22    1  75.503 < 2.2e-16 ***
## Residuals      407.74 4952
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(2,2))
plot(m2,id.n=0)
```




```
par(mfrow=c(1,1))
AIC(m1,m2)
```

```
##    df        AIC
## m1  6 101377.556
## m2  9   1701.126
```

Sembla que aquest model ens dona millors resultats, de fet podem veure que ara el tax si que contribueix al logaritme de la variable target, cosa que abans no passava.

11.2 Millora del model inicial afegint factors

Millorarem el primer model trobat a l'apartat anterior afegint les variables factor amb més relació amb el target, que són Year (que es una variable dependent de Age, per tant no aporta res nou al model), fuelType i transmission. Podriem afegir també model, que està molt relacionada, però ho rebutjem per evitar una complexitat del model molt alta.

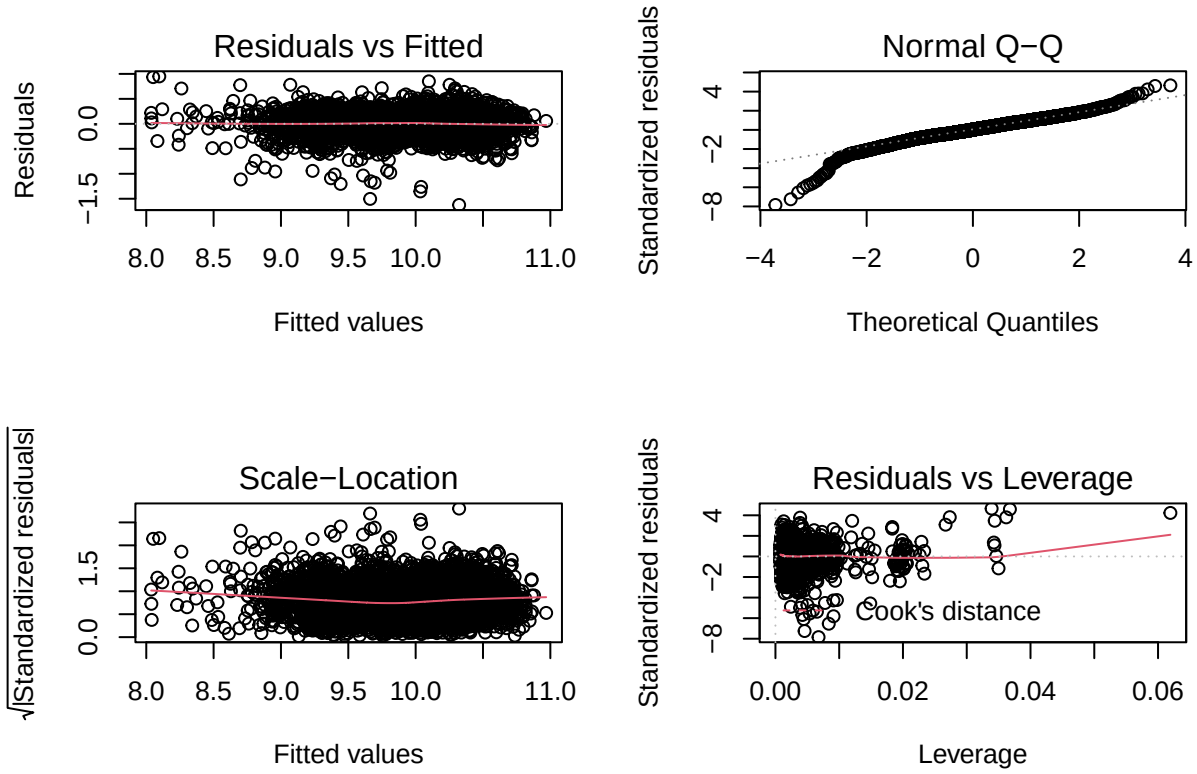
```
m3<-update(m2, ~.+fuelType+transmission,data=df)
summary(m3)
```

```
##
## Call:
## lm(formula = log(price) ~ poly(tax, 2) + poly(mpg, 2) + poly(age,
##     2) + sqrt(mileage) + fuelType + transmission, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6241 -0.1134  0.0058  0.1372  0.9496
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   1.003e+01  1.286e-02 779.465  < 2e-16 ***
## poly(tax, 2)1                 2.377e-01  2.249e-01   1.057    0.290
## poly(tax, 2)2                 1.054e+00  2.157e-01   4.885 1.07e-06 ***
## poly(mpg, 2)1                -1.525e+01  2.879e-01 -52.978  < 2e-16 ***
## poly(mpg, 2)2                 3.901e+00  2.114e-01  18.456  < 2e-16 ***
## poly(age, 2)1                -1.226e+01  3.973e-01 -30.861  < 2e-16 ***
## poly(age, 2)2                -5.202e+00  2.399e-01 -21.684  < 2e-16 ***
## sqrt(mileage)                -1.536e-03  7.883e-05 -19.480  < 2e-16 ***
## fuelTypef.Fuel-Petrol        -3.040e-01  7.235e-03 -42.013  < 2e-16 ***
## fuelTypef.Fuel-Hybrid         5.451e-02  2.816e-02   1.936    0.053 .
## transmissionf.Trans-SemiAuto  2.492e-01  7.485e-03  33.289  < 2e-16 ***
## transmissionf.Trans-Automatic 2.284e-01  8.465e-03  26.985  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2077 on 4948 degrees of freedom
## Multiple R-squared:  0.8069, Adjusted R-squared:  0.8065
## F-statistic:  1880 on 11 and 4948 DF,  p-value: < 2.2e-16
```

```
Anova(m3)
```

```
## Anova Table (Type II tests)
##
## Response: log(price)
##                Sum Sq   Df  F value    Pr(>F)
## poly(tax, 2)    1.058    2   12.261  4.88e-06 ***
## poly(mpg, 2)  136.696    2 1584.720 < 2.2e-16 ***
## poly(age, 2)   89.491    2 1037.466 < 2.2e-16 ***
## sqrt(mileage)  16.367    1  379.476 < 2.2e-16 ***
## fuelType       76.757    2  889.840 < 2.2e-16 ***
## transmission   52.687    2  610.798 < 2.2e-16 ***
## Residuals     213.405 4948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(2,2))
plot(m3,id.n=0)
```




```
par(mfrow=c(1,1))

m4 <- step(m3, k=log(nrow(df)))
```

```
## Start:  AIC=-15501.9
## log(price) ~ poly(tax, 2) + poly(mpg, 2) + poly(age, 2) + sqrt(mileage) +
##     fuelType + transmission
##
##                  Df Sum of Sq    RSS    AIC
## <none>                        213.40 -15502
## - poly(tax, 2)    2     1.058 214.46 -15494
## - sqrt(mileage)   1    16.367 229.77 -15144
## - transmission    2    52.687 266.09 -14424
## - fuelType        2    76.757 290.16 -13995
## - poly(age, 2)    2    89.491 302.90 -13782
## - poly(mpg, 2)    2   136.696 350.10 -13064
```

```
summary(m4)
```

```
##
## Call:
## lm(formula = log(price) ~ poly(tax, 2) + poly(mpg, 2) + poly(age,
##     2) + sqrt(mileage) + fuelType + transmission, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6241 -0.1134  0.0058  0.1372  0.9496
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  1.003e+01  1.286e-02 779.465  < 2e-16 ***
## poly(tax, 2)1                2.377e-01  2.249e-01   1.057    0.290
## poly(tax, 2)2                1.054e+00  2.157e-01   4.885 1.07e-06 ***
## poly(mpg, 2)1               -1.525e+01  2.879e-01 -52.978  < 2e-16 ***
```

```
## poly(mpg, 2)2                 3.901e+00  2.114e-01  18.456  < 2e-16 ***
## poly(age, 2)1                -1.226e+01  3.973e-01 -30.861  < 2e-16 ***
## poly(age, 2)2                -5.202e+00  2.399e-01 -21.684  < 2e-16 ***
## sqrt(mileage)                -1.536e-03  7.883e-05 -19.480  < 2e-16 ***
## fuelTypef.Fuel-Petrol        -3.040e-01  7.235e-03 -42.013  < 2e-16 ***
## fuelTypef.Fuel-Hybrid         5.451e-02  2.816e-02   1.936    0.053 .
## transmissionf.Trans-SemiAuto  2.492e-01  7.485e-03  33.289  < 2e-16 ***
## transmissionf.Trans-Automatic 2.284e-01  8.465e-03  26.985  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2077 on 4948 degrees of freedom
## Multiple R-squared:  0.8069, Adjusted R-squared:  0.8065
## F-statistic:  1880 on 11 and 4948 DF,  p-value: < 2.2e-16
```

```
AIC(m1,m2,m3, m4)
```

```
##    df        AIC
## m1  6 101377.556
## m2  9   1701.126
## m3 13  -1502.138
## m4 13  -1502.138
```

Podem veure que afegir els factors fuelTpe i transmission fa millorar molt l'explicabilitat de la variabilitat del target, que ara és d'un 80%. Per últim, hem fet un step per trobar un model millor respecte el m3 que ja es prou bo, i el resultat es un model exactament igual.

11.3 Interacció entre variables pel model final

Per acabar de trobar el millor model, afegim interaccions entre factors i covariables. En concret multipliquem els mpg amb el tipus de fuel i dividim el fuelType amb la transmissió. Aquestes modificacions ens donen que el model te una explicabilitat del 81%.

```
m5<- lm(log(price) ~ poly(tax, 2) + poly(mpg,2) + poly(age, 2) + mileage + age*fuelType + fuelType:trans
m6<- lm(log(price) ~ poly(tax, 2) + poly(mpg,2) + poly(age, 2) + mileage + tax*fuelType + fuelType:trans
m7<- lm(log(price) ~ poly(tax, 2) + poly(mpg,2) + poly(age, 2) + mileage + mpg*fuelType + fuelType:trans
m8<- lm(log(price) ~ poly(tax, 2) + poly(mpg,2) + poly(age, 2) + mileage + mileage*fuelType + fuelType:t
AIC(m1,m2,m3,m7)
```

```
##    df        AIC
## m1  6 101377.556
## m2  9   1701.126
## m3 13  -1502.138
## m7 18  -1668.921
```

```
summary(m7)
```

```
##
## Call:
## lm(formula = log(price) ~ poly(tax, 2) + poly(mpg, 2) + poly(age,
##     2) + mileage + mpg * fuelType + fuelType:transmission, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.64829 -0.11763  0.00221  0.13070  0.79740
##
## Coefficients: (2 not defined because of singularities)
##                                                  Estimate Std. Error
## (Intercept)                                     9.952e+00  1.012e-02
## poly(tax, 2)1                                   1.766e-01  2.212e-01
```

```
## poly(tax, 2)2                                          8.869e-01  2.133e-01
## poly(mpg, 2)1                                         -1.398e+01  3.471e-01
## poly(mpg, 2)2                                          2.474e+00  2.574e-01
## poly(age, 2)1                                         -1.221e+01  3.611e-01
## poly(age, 2)2                                         -4.208e+00  2.245e-01
## mileage                                               -5.215e-06  2.322e-07
## mpg                                                          NA         NA
## fuelTypef.Fuel-Petrol                                 -9.348e-03  4.366e-02
## fuelTypef.Fuel-Hybrid                                 -4.921e-01  1.554e-01
## mpg:fuelTypef.Fuel-Petrol                             -6.199e-03  7.898e-04
## mpg:fuelTypef.Fuel-Hybrid                              1.317e-02  2.471e-03
## fuelTypef.Fuel-Diesel:transmissionf.Trans-SemiAuto     2.283e-01  9.842e-03
## fuelTypef.Fuel-Petrol:transmissionf.Trans-SemiAuto     2.341e-01  1.127e-02
## fuelTypef.Fuel-Hybrid:transmissionf.Trans-SemiAuto     4.281e-02  5.824e-02
## fuelTypef.Fuel-Diesel:transmissionf.Trans-Automatic    2.144e-01  1.058e-02
## fuelTypef.Fuel-Petrol:transmissionf.Trans-Automatic    2.162e-01  1.402e-02
## fuelTypef.Fuel-Hybrid:transmissionf.Trans-Automatic           NA         NA
##                                                       t value Pr(>|t|)
## (Intercept)                                           983.211  < 2e-16 ***
## poly(tax, 2)1                                           0.798  0.42474
## poly(tax, 2)2                                           4.158 3.27e-05 ***
## poly(mpg, 2)1                                         -40.279  < 2e-16 ***
## poly(mpg, 2)2                                           9.612  < 2e-16 ***
## poly(age, 2)1                                         -33.813  < 2e-16 ***
## poly(age, 2)2                                         -18.742  < 2e-16 ***
## mileage                                               -22.458  < 2e-16 ***
## mpg                                                        NA       NA
## fuelTypef.Fuel-Petrol                                  -0.214  0.83048
## fuelTypef.Fuel-Hybrid                                  -3.167  0.00155 **
## mpg:fuelTypef.Fuel-Petrol                              -7.849 5.09e-15 ***
## mpg:fuelTypef.Fuel-Hybrid                               5.332 1.02e-07 ***
## fuelTypef.Fuel-Diesel:transmissionf.Trans-SemiAuto     23.192  < 2e-16 ***
## fuelTypef.Fuel-Petrol:transmissionf.Trans-SemiAuto     20.774  < 2e-16 ***
## fuelTypef.Fuel-Hybrid:transmissionf.Trans-SemiAuto      0.735  0.46240
## fuelTypef.Fuel-Diesel:transmissionf.Trans-Automatic    20.259  < 2e-16 ***
## fuelTypef.Fuel-Petrol:transmissionf.Trans-Automatic    15.415  < 2e-16 ***
## fuelTypef.Fuel-Hybrid:transmissionf.Trans-Automatic        NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2041 on 4943 degrees of freedom
## Multiple R-squared:  0.8137, Adjusted R-squared:  0.8131
## F-statistic:  1349 on 16 and 4943 DF,  p-value: < 2.2e-16
```

12 Logistic Regression Model

Per la construcció del model de regressió logística haurem de separar el dataset en dades d'entrenament i dades de test primer de tot. També haurem de fer el model inicial respecte els resultats de la Multivariate Analisys de les anteriors entregues, on s'hauran d'afegir dues variables numèriques. Haurem d'afegir els factors més significants i interaccions entre aquests i amb un variable numèrica. Després d'analitzar la correctesa del model final haurem de predir la variable binaria Audi i mostrar la matriu de confusió de les prediccions de l'entrenament i del resultat d'aplicar el model a les dades de test.

12.1 Separació del dataset

Creem dos conjunts de dades nous, train i test, a partir d'una divisió 70-30 del dataset original. El train dataset servirà per construir els millors models, mentres que el test dataset l'usarem per provar el millor model trobat i comprovar els resultats.

```
llwork <- sample(1:nrow(df),round(0.70*nrow(df),0))
```

```
dfall<-df
df_train <- dfall[llwork,]
df_test <-dfall[-llwork,]
```

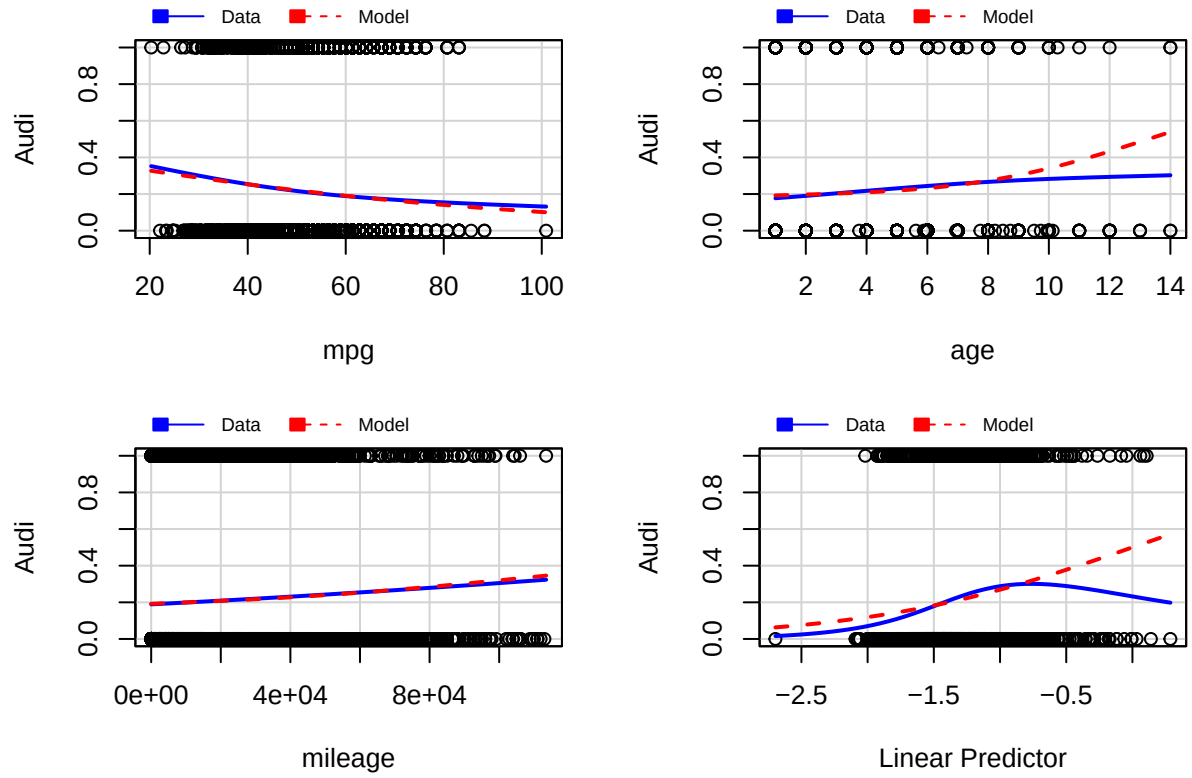
12.2 Primer model amb covariables

Segons el MVA del entregables anteriors, les variables numèriques (o covariables) més relacionades amb la variable target binaria Audi són mpg, age i mileage, i seran les que usarem per construir el nostre model inicial. Podem veure que la variables age no es molt representativa

```
mb1<-glm(Audi~mpg+age+mileage,family="binomial",data=df_train)
summary(mb1)
```

```
##
## Call:
## glm(formula = Audi ~ mpg + age + mileage, family = "binomial",
##     data = df_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3013  -0.7116  -0.6454  -0.5472   2.0701
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.422e-01  2.072e-01  -1.652   0.0986 .
## mpg         -2.511e-02  3.905e-03  -6.430 1.27e-10 ***
## age          4.849e-02  3.343e-02   1.450   0.1470
## mileage      6.870e-06  3.140e-06   2.188   0.0287 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3600.2  on 3471  degrees of freedom
## Residual deviance: 3544.4  on 3468  degrees of freedom
## AIC: 3552.4
##
## Number of Fisher Scoring iterations: 4
```

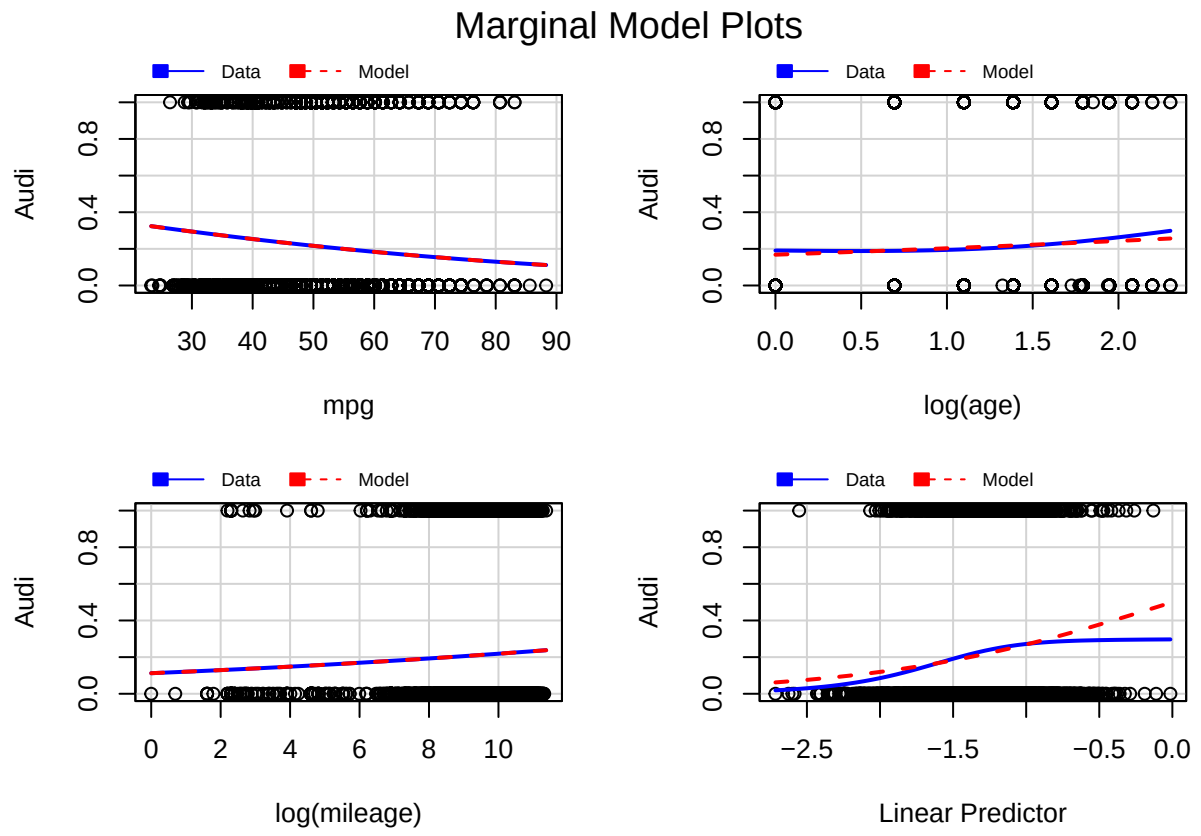
```
marginalModelPlots(mb1)
```




```
mb2<-glm(Audi~mpg+log(age)+log(mileage),family="binomial",data=df_train[!df_train$mout=="YesMOut",])
summary(mb2)
```

```
##
## Call:
## glm(formula = Audi ~ mpg + log(age) + log(mileage), family = "binomial",
##     data = df_train[!df_train$mout == "YesMOut", ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1717  -0.7133  -0.6344  -0.5125   2.2926
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.776406   0.334973  -2.318  0.02046 *
## mpg          -0.031288   0.004302  -7.274  3.5e-13 ***
## log(age)      0.348130   0.122048   2.852  0.00434 **
## log(mileage)  0.072416   0.040626   1.783  0.07466 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3406.8  on 3319  degrees of freedom
## Residual deviance: 3342.9  on 3316  degrees of freedom
## AIC: 3350.9
##
## Number of Fisher Scoring iterations: 4
```

```
marginalModelPlots(mb2)
```




```
AIC(mb1,mb2)
```

```
##     df      AIC
## mb1  4 3552.423
## mb2  4 3350.907
```

Després de provar diferents transformacions, hem comprovat que treure els multivariate Outliers i aplicar el logaritme a age i mileage ens fa obtenir un model millor.

12.3 Millora del model inicial afegint factors

Afegim variables factor per millorar el primer model. Les variables factor amb més relació amb la variable Audi i, per tant, les que afegirem al model són, apart de model i manufacturer que no són independents a Audi, engineSize, fuelType i transmission. Evitem ficar també variables factors construïdes a partir de variables numèriques, ja que no aportarien res al model si les numèriques ja estan considerades.

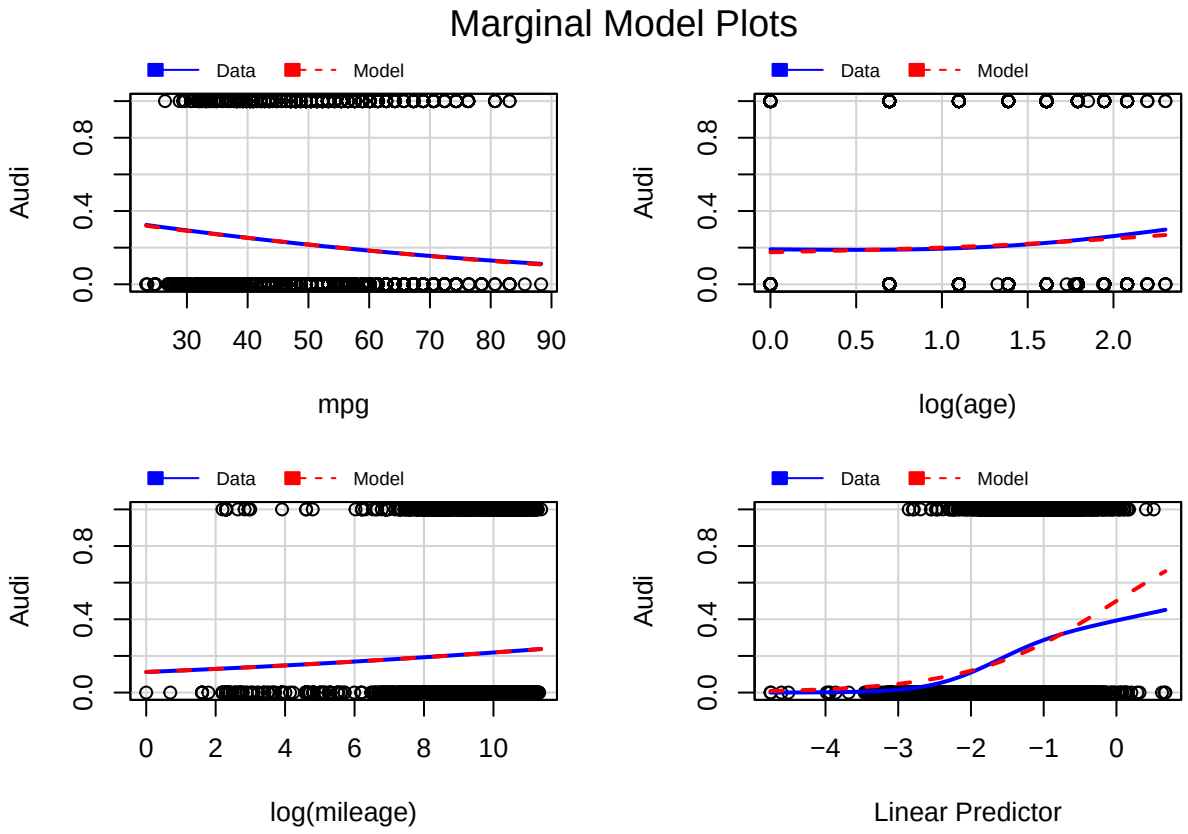
```
mb4 <- update(mb2, ~.+fuelType+transmission+engineSize,data=df_train[!df_train$mout=="YesMOut",])
summary(mb4)
```

```
##
## Call:
## glm(formula = Audi ~ mpg + log(age) + log(mileage) + fuelType +
##     transmission + engineSize, family = "binomial", data = df_train[!df_train$mout ==
##     "YesMOut", ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4733  -0.7237  -0.5802  -0.3638   2.4133
##
## Coefficients:
```

```
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    0.262888   0.475693   0.553 0.580508
## mpg                           -0.051091   0.006129  -8.336  < 2e-16 ***
## log(age)                       0.448229   0.128634   3.485 0.000493 ***
## log(mileage)                   0.089002   0.041017   2.170 0.030017 *
## fuelTypef.Fuel-Petrol         -0.233034   0.136333  -1.709 0.087396 .
## fuelTypef.Fuel-Hybrid         -1.308919   0.741208  -1.766 0.077408 .
## transmissionf.Trans-SemiAuto  -0.465569   0.113081  -4.117 3.84e-05 ***
## transmissionf.Trans-Automatic -0.434918   0.131442  -3.309 0.000937 ***
## engineSizeMedium               0.447397   0.137511   3.254 0.001140 **
## engineSizeBig                 -0.849629   0.206231  -4.120 3.79e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3406.8  on 3319  degrees of freedom
## Residual deviance: 3206.4  on 3310  degrees of freedom
## AIC: 3226.4
##
## Number of Fisher Scoring iterations: 5
```

```
marginalModelPlots(mb4)
```

```
## Warning in mmps(...): Interactions and/or factors skipped
```




```
Anova(mb4, test="LR")
```

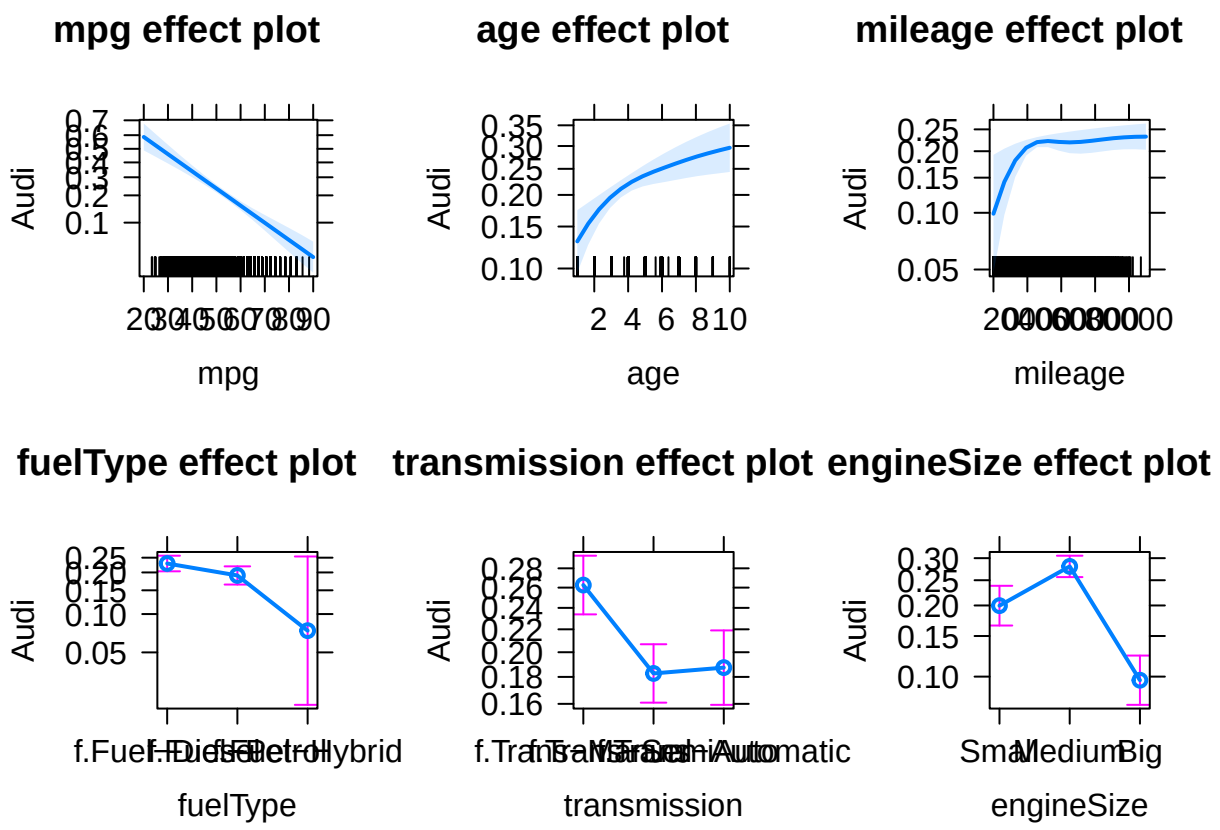
```
## Analysis of Deviance Table (Type II tests)
##
## Response: Audi
##              LR Chisq Df Pr(>Chisq)
## mpg            71.671  1  < 2.2e-16 ***
## log(age)       12.101  1  0.0005039 ***
```

```
## log(mileage)    5.014  1  0.0251482 *
## fuelType        6.708  2  0.0349435 *
## transmission   18.740  2  8.523e-05 ***
## engineSize     99.854  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova( mb4, mb2, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Audi ~ mpg + log(age) + log(mileage) + fuelType + transmission +
##     engineSize
## Model 2: Audi ~ mpg + log(age) + log(mileage)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      3310     3206.4
## 2      3316     3342.9 -6  -136.48 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(allEffects(mb4))
```




```
AIC(mb1,mb2,mb4)
```

```
## Warning in AIC.default(mb1, mb2, mb4): models are not all fitted to the same
## number of observations
```

```
##     df      AIC
## mb1  4 3552.423
## mb2  4 3350.907
## mb4 10 3226.431
```

12.4 Interacció entre variables pel model final

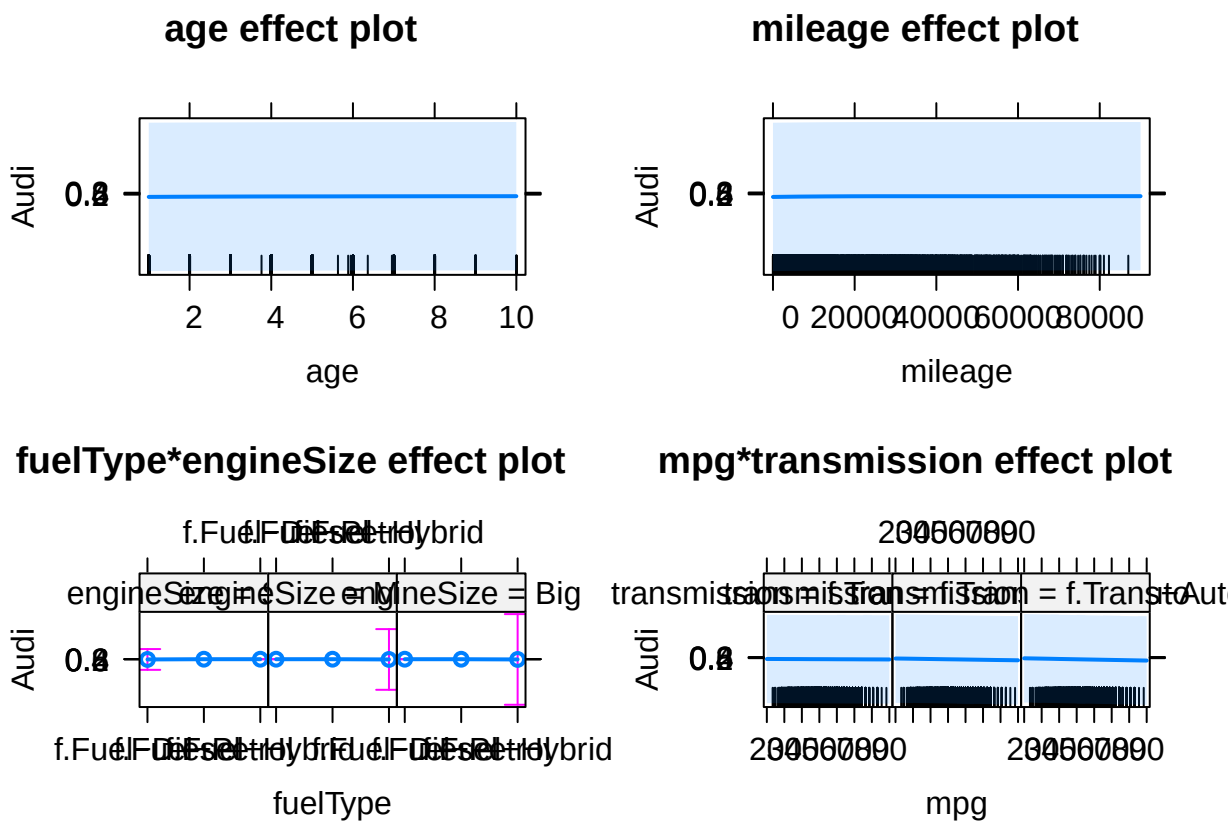
Per acabar de trobar el millor model, afegim interaccions entre factors i covariables.

```
mb5<- glm(formula = Audi ~ mpg+log(age)+log(mileage) + fuelType + transmission + engineSize + (fuelType*
          family = "binomial", data = df_train[!df_train$mout == "YesMOut", ])
summary(mb5)
```

```
##
## Call:
## glm(formula = Audi ~ mpg + log(age) + log(mileage) + fuelType +
##     transmission + engineSize + (fuelType * engineSize) + mpg *
##     transmission, family = "binomial", data = df_train[!df_train$mout ==
##     "YesMOut", ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4286  -0.7388  -0.5762  -0.0003   2.6197
##
## Coefficients:
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             -1.741e+01  3.335e+02  -0.052 0.958357
## mpg                                     -1.714e-02  8.476e-03  -2.022 0.043157
## log(age)                                 4.434e-01  1.291e-01   3.433 0.000596
## log(mileage)                             9.296e-02  4.113e-02   2.260 0.023806
## fuelTypef.Fuel-Petrol                    1.564e+01  3.335e+02   0.047 0.962600
## fuelTypef.Fuel-Hybrid                    1.589e+01  3.335e+02   0.048 0.961987
## transmissionf.Trans-SemiAuto             2.073e+00  5.717e-01   3.627 0.000287
## transmissionf.Trans-Automatic            2.888e+00  6.565e-01   4.398 1.09e-05
## engineSizeMedium                         1.620e+01  3.335e+02   0.049 0.961255
## engineSizeBig                            1.493e+01  3.335e+02   0.045 0.964302
## fuelTypef.Fuel-Petrol:engineSizeMedium -1.605e+01  3.335e+02  -0.048 0.961624
## fuelTypef.Fuel-Hybrid:engineSizeMedium -3.228e+01  1.030e+03  -0.031 0.974989
## fuelTypef.Fuel-Petrol:engineSizeBig    -1.645e+01  3.335e+02  -0.049 0.960661
## fuelTypef.Fuel-Hybrid:engineSizeBig    -2.992e+01  1.492e+03  -0.020 0.984006
## mpg:transmissionf.Trans-SemiAuto        -4.611e-02  1.055e-02  -4.369 1.25e-05
## mpg:transmissionf.Trans-Automatic       -6.255e-02  1.260e-02  -4.966 6.82e-07
##
## (Intercept)
## mpg                                     *
## log(age)                                ***
## log(mileage)                            *
## fuelTypef.Fuel-Petrol
## fuelTypef.Fuel-Hybrid
## transmissionf.Trans-SemiAuto            ***
## transmissionf.Trans-Automatic           ***
## engineSizeMedium
## engineSizeBig
## fuelTypef.Fuel-Petrol:engineSizeMedium
## fuelTypef.Fuel-Hybrid:engineSizeMedium
## fuelTypef.Fuel-Petrol:engineSizeBig
## fuelTypef.Fuel-Hybrid:engineSizeBig
## mpg:transmissionf.Trans-SemiAuto        ***
## mpg:transmissionf.Trans-Automatic       ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3406.8  on 3319  degrees of freedom
## Residual deviance: 3134.9  on 3304  degrees of freedom
## AIC: 3166.9
##
```

```
## Number of Fisher Scoring iterations: 16
```

```
plot(allEffects(mb5))
```




```
AIC(mb4,mb5,mb2)
```

```
##     df      AIC
## mb4 10 3226.431
## mb5 16 3166.898
## mb2  4 3350.907
```

```
Anova(mb5, test="LR")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Audi
##                       LR Chisq Df Pr(>Chisq)
## mpg                     65.612  1  5.490e-16 ***
## log(age)                11.760  1  0.0006051 ***
## log(mileage)             5.449  1  0.0195753 *
## fuelType                 6.954  2  0.0309068 *
## transmission            15.685  2  0.0003927 ***
## engineSize             103.290  2  < 2.2e-16 ***
## fuelType:engineSize     51.309  4  1.924e-10 ***
## mpg:transmission        30.804  2  2.047e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

12.5 Prediccions de la variable binaria Audi

Un cop trobat el millor model que intentarà predir la variable binaria Audi, provem de utilitzar-lo per fer la predicció amb el train dataset. Per visualitar els resultats mostrem la matriu de confusió. Per últim, fem el mateix amb el test dataset, que mostrarà realment si el model es poc explicatiu de la variable target, o si es genera overfitting, i mostrem la matriu de confusió dels resultats.

```
library(ResourceSelection)
```

```
## Warning: package 'ResourceSelection' was built under R version 4.1.2
```

```
## ResourceSelection 0.3-5   2019-07-22
```

```
pred_train <- predict(mb5, newdata=df_train, type="response")
htr <- hoslem.test(as.numeric(df_train$Audi)-1, pred_train)
htr
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  as.numeric(df_train$Audi) - 1, pred_train
## X-squared = 24.121, df = 8, p-value = 0.002187
```

```
cbind(htr$observed, htr$expected)
```

```
##                     y0  y1    yhat0      yhat1
## [1.27e-09,0.0702] 342   6 338.7783   9.221706
## (0.0702,0.118]    319  28 313.6972  33.302841
## (0.118,0.15]      306  41 300.8064  46.193636
## (0.15,0.183]      285  62 288.6040  58.396023
## (0.183,0.207]     278  69 279.3876  67.612430
## (0.207,0.236]     265  82 270.4755  76.524526
## (0.236,0.267]     250  97 259.8412  87.158791
## (0.267,0.314]     223 124 246.7398 100.260155
## (0.314,0.363]     232 115 229.3471 117.652859
## (0.363,0.735]     231 117 200.3202 147.679820
```

```
pred_test <- predict(mb5, newdata=df_test, type="response")
ht <- hoslem.test(as.numeric(df_test$Audi)-1, pred_test)
ht
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  as.numeric(df_test$Audi) - 1, pred_test
## X-squared = 34.948, df = 8, p-value = 2.733e-05
```

```
cbind(ht$observed, ht$expected)
```

```
##                     y0 y1     yhat0    yhat1
## [2.52e-09,0.0629] 149  0 145.69455  3.30545
## (0.0629,0.115]    141  8 135.40278 13.59722
## (0.115,0.147]     120 29 129.48240 19.51760
## (0.147,0.179]     120 28 123.89505 24.10495
## (0.179,0.205]     111 38 120.32353 28.67647
## (0.205,0.229]     112 37 116.73025 32.26975
## (0.229,0.266]      96 52 111.36578 36.63422
## (0.266,0.315]     104 45 105.76706 43.23294
## (0.315,0.369]     112 37  98.30878 50.69122
## (0.369,0.804]      97 52  84.80784 64.19216
```

```
# Confusion Table Analysis
audi.est <- ifelse(pred_train<0.5,0,1)
tt<-table(audi.est,df_train$Audi);
100*sum(diag(tt))/sum(tt)
```

```
## [1] 78.28341
```

```r
audi.est <- ifelse(pred_test<0.5,0,1)
tt<-table(audi.est,df_test$Audi);tt
```

```
##
## audi.est Audi No Audi Yes
##        0    1152      319
##        1      10        7
```

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

```
## [1] 77.88978
```

Segons el test de Hosmer i Lemeshow, com el p-valor es menor a 0.05, el model final trobat es un bon fit.

Com ens mostren els resultats expressats en les matrius de confusió, la predicció amb el dataset de train ens dona una accuracy del 78%, mentre que la del test un 77%, pel que podem suposar que el model fet pel train s'adapta correctament al test i que no es massa pobre ni pel contrari massa especific.