Deliverable 3

Lab 3 - Numeric and Binary targets Forecasting Models

Othman Benmoussa & Eloi Cruz

May 29, 2022

Contents

1 Load Required Packages: to be increased over the course **2**

2 Load data **2**

3 Introduction **2**

4 Description of Model Building process for prediction of numeric response (price). **3**

4.1 Multiple regresion using covariates . . . 3

4.1.1 Model 1: price ~ mgp+mileage+tax+years_sell2 . . . 4

4.1.2 Model 2: log(price) ~ mileage+tax+years_sell2 . . . 8

4.1.2.1 Model validation . . . 10

4.2 Adding factor variables . . . 14

4.2.1 Model 3: price ~ mileage+tax+years_sell2 + model+engineSize . . . 15

4.2.2 Model 4: log(price) ~ mileage+tax+years_sell2 + model+engineSize . . . 18

4.3 Adding interactions . . . 23

4.3.1 Model 5 . . . 23

4.4 Influent data and outliers . . . 24

4.4.1 Model 6 . . . 25

4.5 Conclusion . . . 25

5 Description of Model Building process for prediction of binary response (Audi). **26**

5.1 Split into train and test . . . 26

5.2 Binary Models: Using numerical explanatory variables . . . 27

5.2.1 Model 1: Audi ~ mgp+mileage+tax+years_sell2 . . . 27

5.2.2 Model 2: Audi ~ mgp+mileage+tax . . . 28

5.2.3 Model 3: Audi ~ mgp+mileage+years_sell2 . . . 29

5.2.4 Understanding the model chosen (model 2) . . . 31

5.3 Binary Models: Adding fators . . . 33

5.3.1 Model 4: Audi~mileage+tax+mpg+fuelType+transmission+engineSize . . . 33

5.3.2 Model 5: Audi~mileage+tax+mpg+transmission+engineSize . . . 34

5.4 Binary model: Adding interactions . . . 35

5.4.1 Model 6 . . . 35

5.5 Influent data and outliers . . . 36

5.6 Confusion table analysis . . . 40

6 Annex **41**
6.1 A1 . . . 41
6.2 A2 . . . 43
6.3 A3 . . . 47
6.4 A4 . . . 56

1 Load Required Packages: to be increased over the course

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

requiredPackages <- c("effects","FactoMineR","car", "factoextra","RColorBrewer","ggplot2","dplyr","ggmap
install.packages("moments",repos = "http://cran.us.r-project.org")
#use this function to check if each package is on the local machine
#if a package is installed, it will be loaded
#if any are not, the missing package(s) will be installed and loaded
package.check <- lapply(requiredPackages, FUN = function(x) {
  if (!require(x, character.only = TRUE)) {
    install.packages(x, dependencies = TRUE)
    library(x, character.only = TRUE)
  }
})
#verify they are loaded
search()
```

2 Load data

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

# Clean workspace
rm(list=ls())
setwd("C:/Users/Eloi/Documents/ADEI/ADEI/Lab4 - Modeling")
load("EloiOthman_del2.RData")

## Adding multivarioant ouliers column
library(chemometrics)
res.mout <- Moutlier( df[ ,c(3,5,8,14)], quantile = 0.9995, tol=1e-40 )
llmout <- which((res.mout$md>res.mout$cutoff)&(res.mout$rd>res.mout$cutoff))
df$mout <- 0
df$mout[llmout] <- 1
df$mout <- factor( df$mout, labels = c("MvOut.No","MvOut.Yes"))
```

3 Introduction

Before we begin we would like to make a brief summary of all the work done so far. At the beginning of the course we were presented with a data set containing a list of vehicles and concrete characteristics of each one of them. Our goal is to create a mathematical model that, given the characteristics of a new instance, allows us to predict its price or to predict whether the car is Audi or not.

The first tasks consisted of analyzing and improving the data to understand the data we had in front of us and prepare it to be as optimal as possible before creating the corresponding mathematical models. Once this was done, it was time to create the linear models that would allow us to predict the data.

4 Description of Model Building process for prediction of numeric response (price).

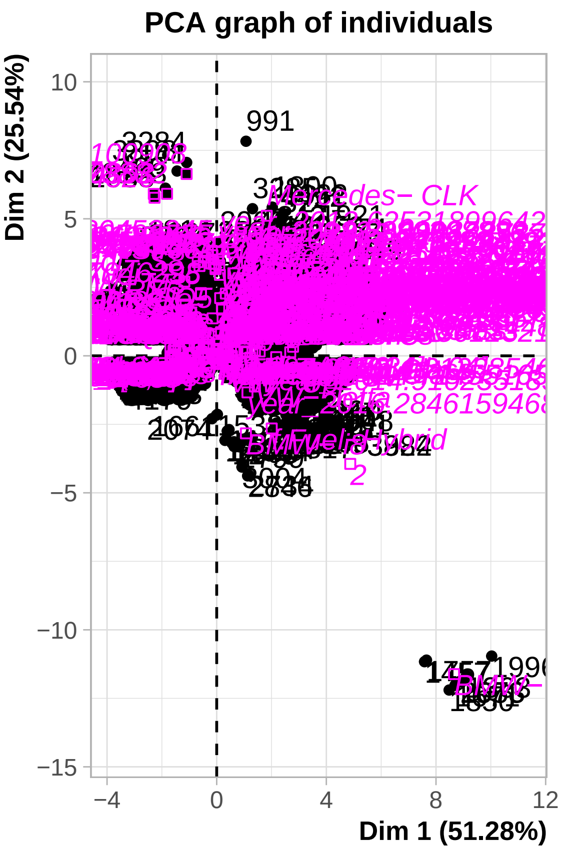
We will start by going through the process of creating a forecasting model for the prediction of the target numerical variable price.

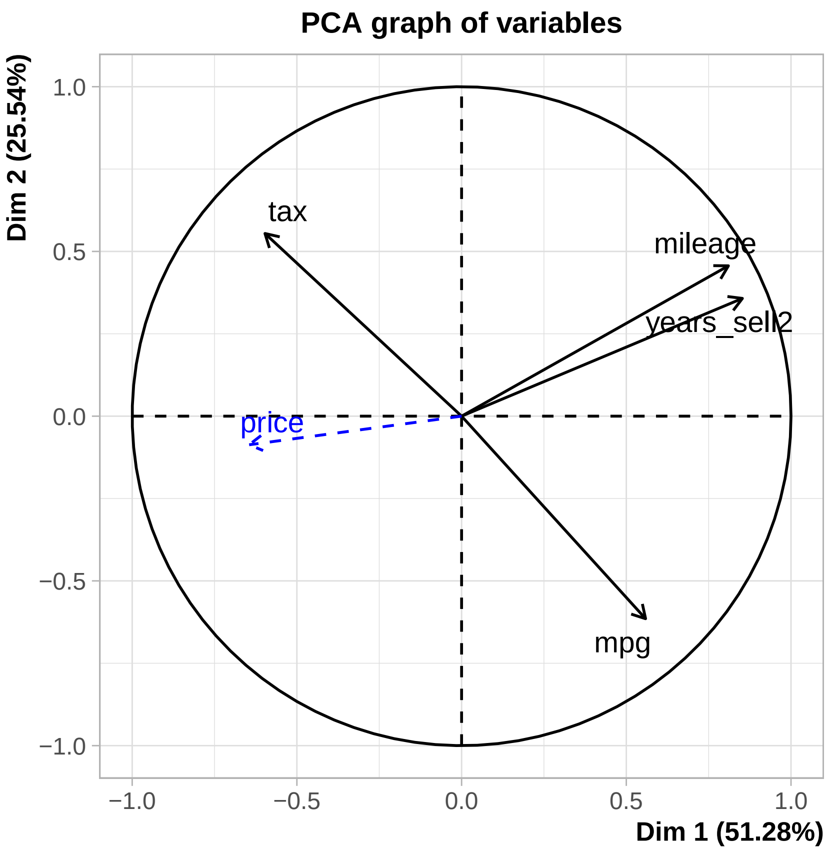
4.1 Multiple regresion using covariates

To begin with, we will start creating the best model possible using only the numeric variables available (mpg, millage, tax and years_sell2) to understand the relation between them and the target price.

Using the principal component analysis method, in the previous assignment, we saw that exists a strong negative correlation between the variable price and millage and years_sell2. This gives us a clue of which numeric variables will have more impact in the model creation process. We can see that there exists a positive correlation between price and tax and price and mpg but this relation is less strong. The condes method output shows that the correlation between price and mpg is really weak because it does not appear on the output.

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






```r
res.con <- condes(df[c(5,7,8,14)],num.var=which(names(df)=="price"))
res.con$quanti
```

```
##              correlation       p.value
## years_sell2    0.2421877  3.679975e-67
## mileage        0.2099942  1.431271e-50
## tax           -0.3526690 2.848857e-145
```

4.1.1 Model 1: price ~ mgp+mileage+tax+years_sell2

The frist model that we have created, includes all the covariates. The next steps will have the objective to analize the statistical influence of them in the creation of the model.

```r
# Preparing data
ll<-which(df$year==0);ll
df$year[ll]<-0.5
ll<-which(df$tax==0);ll
df$tax[ll]<-0.5
ll<-which(df$mileage==0);ll
df$mileage[ll]<-0.5
ll<-which(df$mpg==0);ll
df$mpg[ll]<-0.5
```

```r
#1st linear model with my numeric variables:
m1<-lm(price~mileage+tax+mpg+years_sell2,data=df)
summary(m1)
```

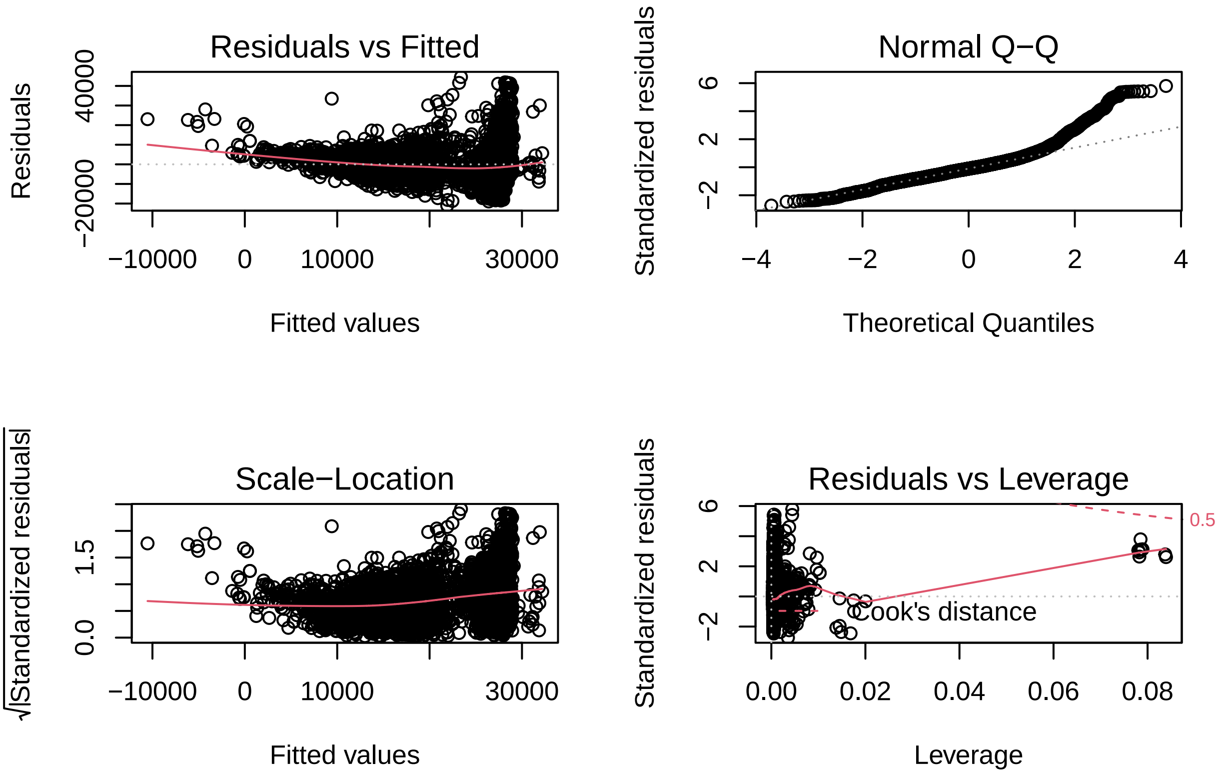
```
##
## Call:
## lm(formula = price ~ mileage + tax + mpg + years_sell2, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -21009  -4632   -763   3129  44542
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.226e+04  5.954e+02  54.193   <2e-16 ***
## mileage     -1.210e-01  7.178e-03 -16.855   <2e-16 ***
## tax          2.797e+01  2.081e+00  13.442   <2e-16 ***
## mpg         -4.608e+01  5.338e+00  -8.631   <2e-16 ***
## years_sell2 -6.002e+03  2.862e+02 -20.967   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7699 on 4957 degrees of freedom
## Multiple R-squared:  0.4249, Adjusted R-squared:  0.4244
## F-statistic: 915.5 on 4 and 4957 DF,  p-value: < 2.2e-16
```

```
vif(m1) #Variance inflation factor: multicorrelation
```

```
##     mileage         tax         mpg years_sell2
##    2.048586    1.214392    1.173044    2.161328
```

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



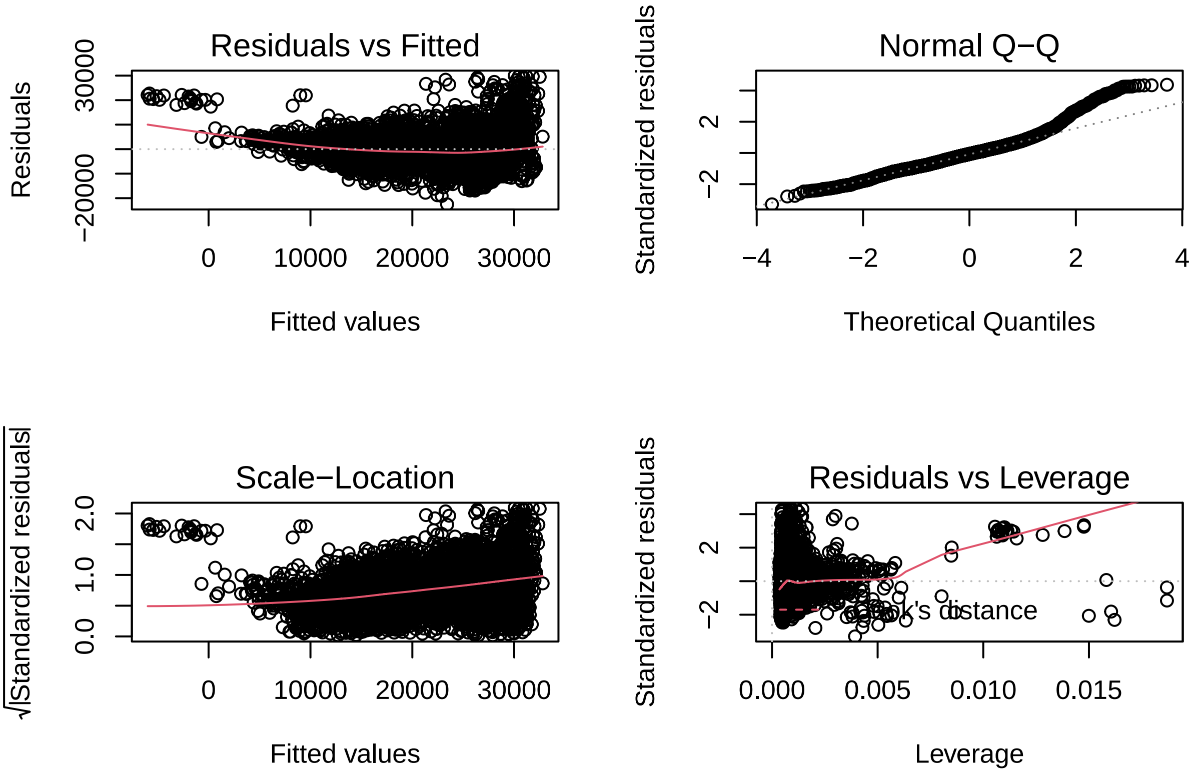

```
# Basic graphs for model validation
par(mfrow=c(1,1))
```

After the execution of the first model we can get some conclusions. Model 1 explains 42.49% of the variability of the target, which is really not sufficient. We should try to look at the correlated continuous variables in order to eliminate the redundancy and add factors to this regression.

From the point of view of the residuals we can see that the distribution of the residuals is not normal so they are not independent and we have to try find why. The residuals vs leverage plot shows us that there are some outliers that might be causing this non normal distribution.

```r
m1<-lm(price~mileage+tax+mpg+years_sell2,data=df[df$mout=="MvOut.No",])

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

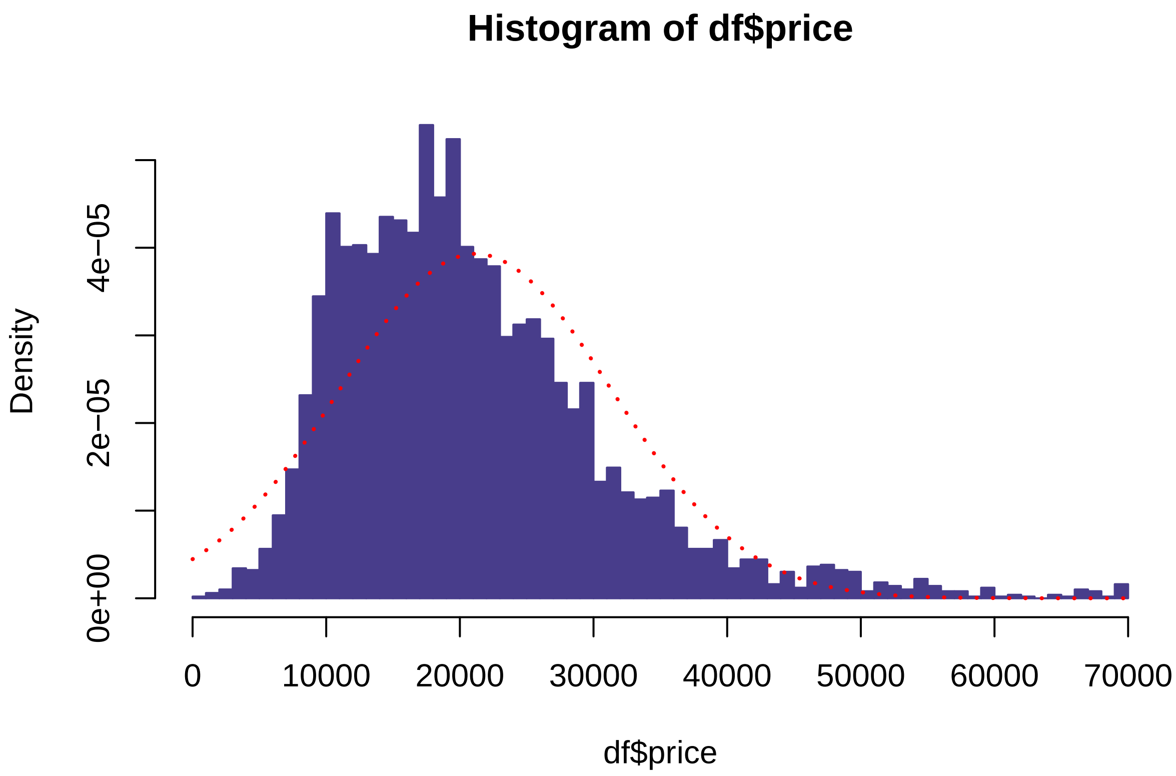



```r
par(mfrow=c(1,1))
```

We can see that extracting the multivariant outliers from the analysis helps to improve the normal distribution of the residuals but is not sufficient. We will do this process at the end of the analysis.

We will check if our variable target is normal to apply a transformation to improve the normal distribution of the residuals.

```r
hist(df$price,50,freq=F,col="darkslateblue",border = "darkslateblue")
mm<-mean(df$price);ss<-sd(df$price)
curve(dnorm(x,mean=mm,sd=ss),col="red",lwd=2,lty=3, add=T)
```




```r
shapiro.test(df$price)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$price
## W = 0.92211, p-value < 2.2e-16
```

```r
# skewness
library(e1071)
skewness(df$price)
```

```
## [1] 1.275432
```

```r
# kurtosis
library(moments)
```

```
##
## Attaching package: 'moments'

## The following objects are masked from 'package:e1071':
##
##     kurtosis, moment, skewness
```

```r
kurtosis(df$price)
```

```
## [1] 5.540289
```

We can see that our histogram is a bit skewed at the right and not completely symmetrical.It is not thus totally following a normal shape

The p-value is too small, we can thus reject the H0 hypothesis that indicates that the price variable is following a normal distribution.

Normal data should have 0 skewness: we see that our data is right skewed at 1.27

Normal data should be 0. We have 5.54, so, in this case, our data is not normal.

```
vif(m1)
```

```
##     mileage         tax         mpg years_sell2
##    2.136659    1.446020    1.549190    2.263210
```

The values given are not superior to 3 so we can say that correlation is not that impactful in this regression model

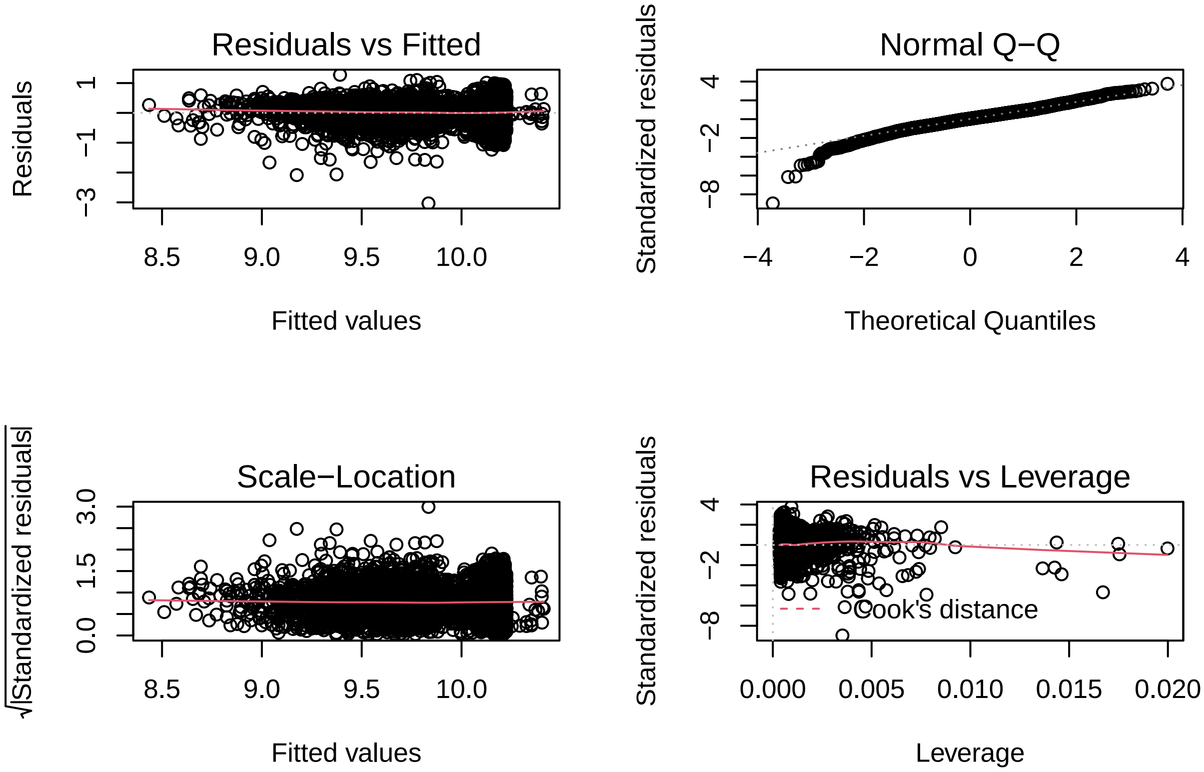
4.1.2 Model 2: log(price) ~ mileage+tax+years_sell2

As we know that the relation of the variable price and mpg is really weak we will compute another model extracting mpg from the analysis. What is more we will apply a logarithmic function on the variable price to make normal, as we saw on the lab.

```
m2<-lm(log(price)~tax+mileage+years_sell2,data=df)
summary(m2)
```

```
##
## Call:
## lm(formula = log(price) ~ tax + mileage + years_sell2, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.03330 -0.19811  0.01832  0.21412  1.27487
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.025e+01  2.179e-02  470.30   <2e-16 ***
## tax          1.642e-03  8.735e-05   18.79   <2e-16 ***
## mileage     -8.114e-06  3.156e-07  -25.71   <2e-16 ***
## years_sell2 -2.706e-01  1.258e-02  -21.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3391 on 4958 degrees of freedom
## Multiple R-squared:  0.5057, Adjusted R-squared:  0.5054
## F-statistic:  1691 on 3 and 4958 DF,  p-value: < 2.2e-16
```

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



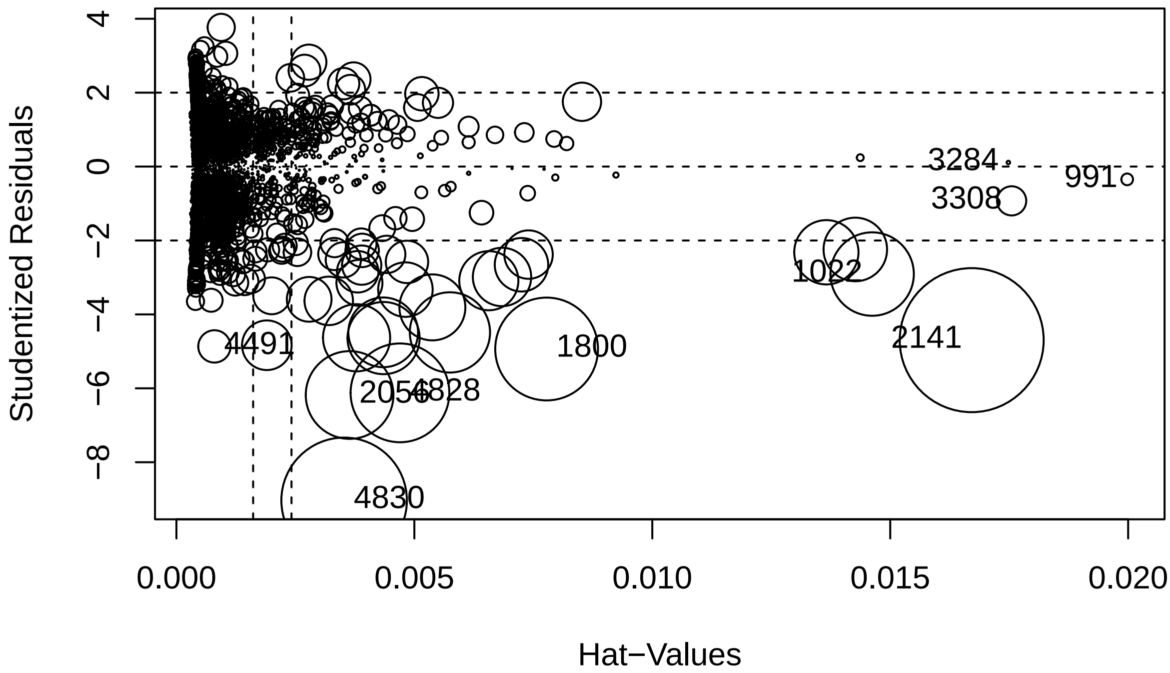

```
par(mfrow=c(1,1))
```

Model 2 now explains 41.6% of the variability of the target, We can confirm now that extracting the variable mpg from the analysis does not make a big effect in terms of geting the maximum variance possible (around 1% only).

Looking at the graphs, we can clearly see that model m2 is better suited than m1 for this regression, we will further analyze the plots for the m2 model and try to optimize the m2 model with Boxcox and BoxTidwell.

What is more, now the plots shows that the residuals are distributed in a normal way so we will choose this model as the valid one. We can see homeosticity too. We can see that we have a better normality, however the residuals vs leverage plot doesn't seem to have gotten better as more residuals with greater leverage have appeared, we will consider removing them after (especially number 4830, 4828, 2141, 2056 and 2050). We will take them out at the end of the analysis too.

```
influencePlot( m2, id=c(list="noteworthy",n=5))
```




```
##          StudRes          Hat       CookD
## 991  -0.3466855 0.0199792316 6.126777e-04
## 1022 -2.9083215 0.0146198023 3.132634e-02
## 1800 -4.9362429 0.0077790889 4.753477e-02
## 2056 -6.1810294 0.0036388632 3.462293e-02
## 2141 -4.6967756 0.0167121790 9.333644e-02
## 3284  0.1106055 0.0174828698 5.443174e-05
## 3308 -0.9275848 0.0175511191 3.842858e-03
## 4491 -4.8657713 0.0007985407 4.708750e-03
## 4828 -6.1226879 0.0046975805 4.390955e-02
## 4830 -9.0324372 0.0035242196 7.098132e-02
```

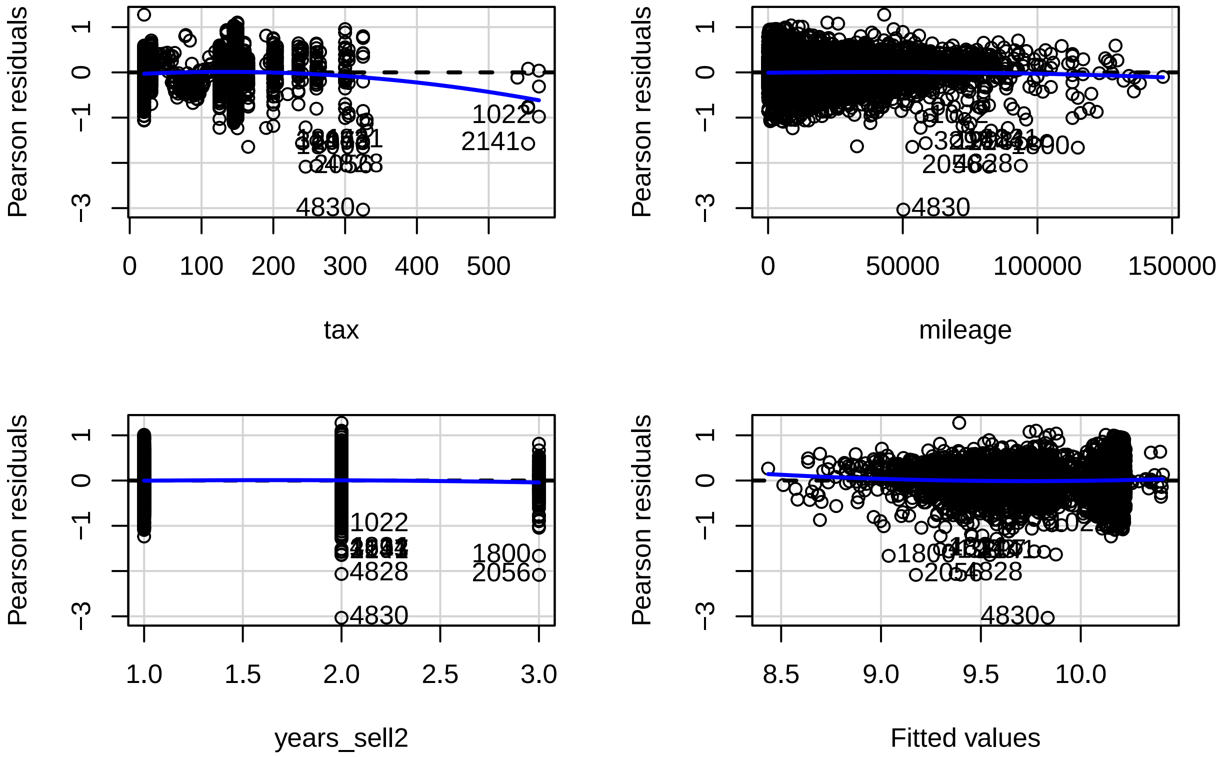
4.1.2.1 Model validation We have to check that our asumptions asociated with the multiple regression: *Linearity: The relationship between X and the mean of Y is linear.* Homoscedasticity: The variance of residual is the same for any value of X. *Independence: Observations are independent of each other.* Normality: For any fixed value of X, Y is normally distributed.

In multiple regression , two or more predictor variables might be correlated with each other (collinearity). In the presence of collinearity, the solution of the regression model can not be accurate. We can see that that there are not variables that are very correlated between them so we don't have much redundace.

```
vif(m2)
```

```
##         tax     mileage years_sell2
##    1.103037    2.040934    2.151738
```

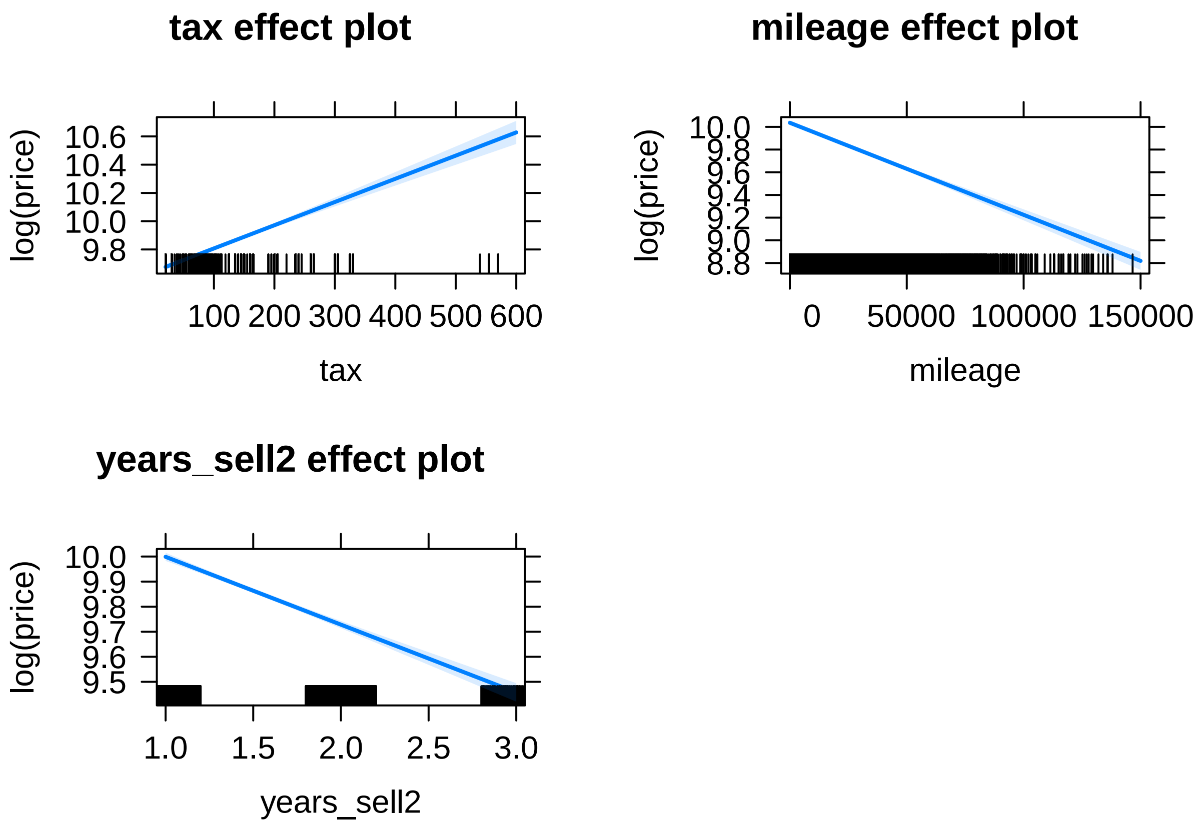
```
residualPlots(m2,id=list(method=cooks.distance(m2),n=10))
```



```
##              Test stat Pr(>|Test stat|)
## tax             -6.6872        2.527e-11 ***
## mileage         -1.8301         0.067293 .
## years_sell2     -1.8975         0.057821 .
## Tukey test       2.8441         0.004454 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we saw in the previous page, these graphics show that thee residuals are independent in this model so they not take part of the model explanation. By the way some extrem values affect in a negative way in the perarsons graphic for the tax variable. We can see great linearity in all four graphics.

```
library(effects)
plot(allEffects(m2))
```



We can see that years_sell2 and mileage have a negative correlation with the variable target log(price). When cars are older or have been driven for more miles the price of them decreases. What is the same when they are more used they are cheaper. In the other hand, the variable tax is directly correlated so more expensive cars pay more taxes but it has two extreme values that seem to reduce the quality of the prediction.

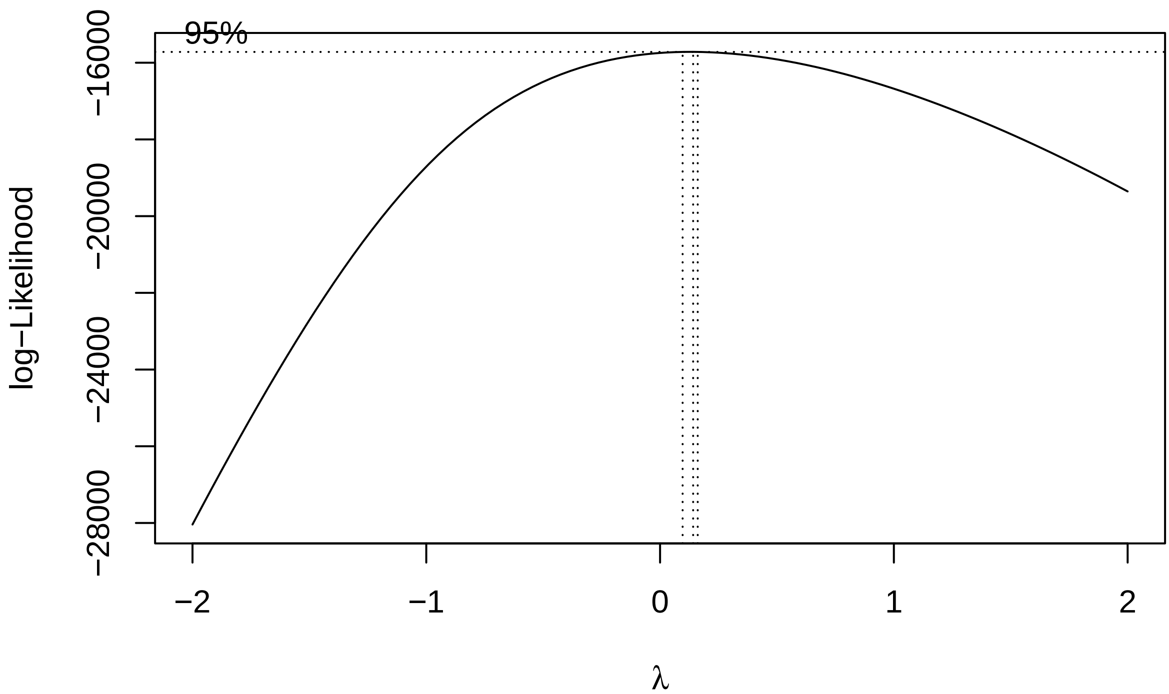
We will use the boxcox and boxTidwell methods to try to understand better the relation between variables and target and apply transformations if necessary.

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

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

```
boxcox(price~tax+mileage+years_sell2,data=df)
```



As we can see int the original model the lambda got by the boxcox method has a value near 0. As it is far from one this means that the lambda=0, so we had a good intuition by choosing the to put the target in log because it is far from the 1 value (value that determinates that data has not to be changed).

We will try the BoxTidwell method in order to see if it will make our model better by improving the normality of the residuals and adding variability explanation.

```
boxTidwell(log(price)~tax+mileage+years_sell2,data=df)
```

```
##              MLE of lambda Score Statistic (z)  Pr(>|z|)
## tax               0.079787             -6.7285 1.715e-11 ***
## mileage           0.786973              1.9430  0.052014 .
## years_sell2       1.899968             -3.1189  0.001815 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## iterations =  8
```

We will apply the transformations according to the output of the boxTidwell result. Mileage will not be transformed but tax and years_sell2 yes because they have value lambdas different from 1.

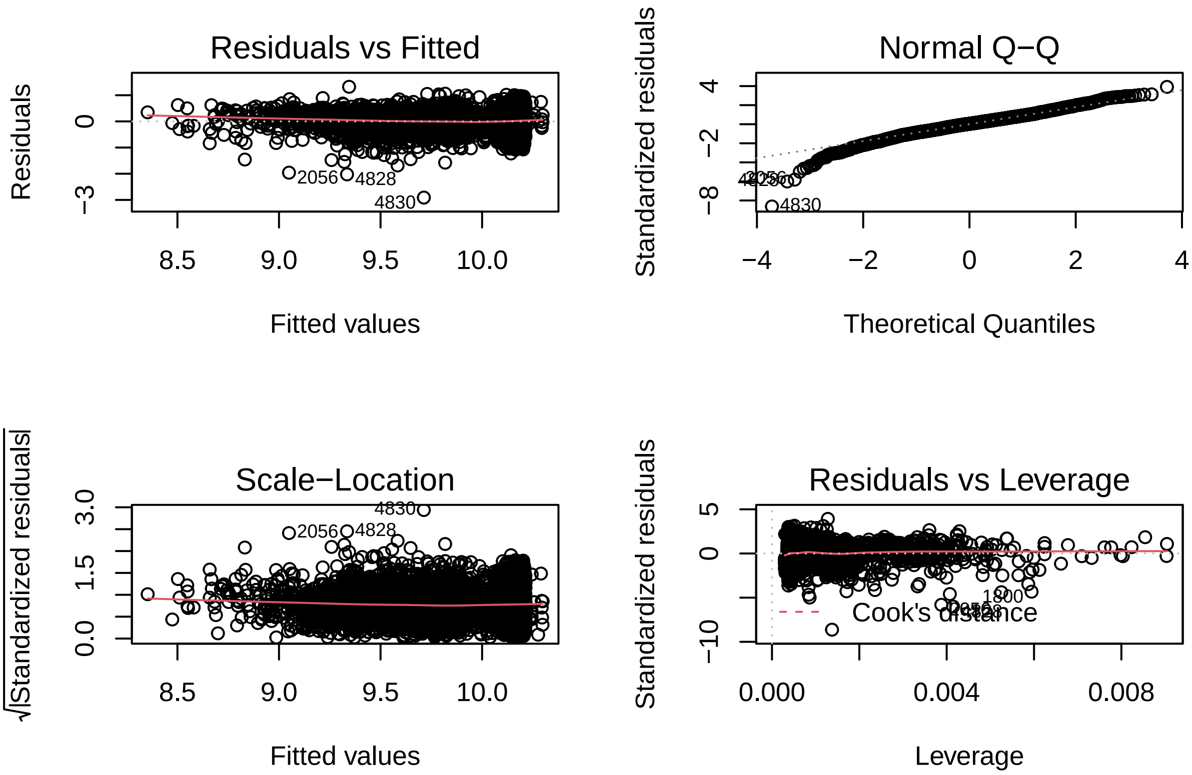
```
m2aux<-lm(log(price)~log(tax)+mileage+I(years_sell2^2),data=df)
summary(m2aux)
```

```
##
## Call:
## lm(formula = log(price) ~ log(tax) + mileage + I(years_sell2^2),
##     data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.91199 -0.19954  0.01567  0.20628  1.32125
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)        9.526e+00  3.670e-02  259.54   <2e-16 ***
## log(tax)           1.522e-01  7.095e-03   21.45   <2e-16 ***
## mileage           -7.892e-06  3.112e-07  -25.36   <2e-16 ***
## I(years_sell2^2) -7.403e-02  3.670e-03  -20.17   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3377 on 4958 degrees of freedom
## Multiple R-squared:  0.5098, Adjusted R-squared:  0.5095
## F-statistic:  1719 on 3 and 4958 DF,  p-value: < 2.2e-16
```

The explanatory of the variables hasn't changed, we will plot the residuals in order to see if we carry on with this new model.

```
par(mfrow=c(2,2))
plot(m2aux)
```



The new model doesn't improve the residuals nor the explanatory of the variables. For the residuals vs leverage plots, we can see that this model adds too many residuals with high leverage, which makes the model less strong. We will thus stick with m2.

4.2 Adding factor variables

Now we have to try to imporve it because a variance of 40% is not enough to get a good model so we will proceed adding factor variables.

```
condes(df,3)$quali
```

```
##                      R2      p.value
## model        0.51825594 0.000000e+00
## year         0.42133943 0.000000e+00
## transmission 0.26061149 0.000000e+00
## engineSize   0.26766925 0.000000e+00
## years_sell   0.35603346 0.000000e+00
```

```
## aux          0.33869720  0.000000e+00
## f.price      0.78269048  0.000000e+00
## f.miles      0.34006130  0.000000e+00
## f.tax        0.27581933  0.000000e+00
## mpg_d        0.31011346  0.000000e+00
## hcpck        0.36823004  0.000000e+00
## claKM        0.35925569  0.000000e+00
## manufacturer 0.09962391 1.847626e-112
## mout         0.01443004  2.058993e-17
## fuelType     0.01013366  1.076655e-11
## Audi         0.00361113  2.277616e-05
```

Now we have to choose the factors that we will use in our analysis. Using the previus result we will chose variables most correlated to the variable target price. The ones that has less correlation will not be used. We won't put the factor year as a predictor as it will induct a high correlation with the continuous variable years_sell2 which will make this regression impossible and show an error when calling VIF.

The factors used will be manufacturer, model, aux, transmission, engineSize and fuelType but we will start adding only the highest correlated: model and engineSize.

4.2.1 Model 3: price ~ mileage+tax+years_sell2 + model+engineSize

```
m3<-lm(price~tax+mileage+years_sell2+model+engineSize,data=df[!df$mout=="MvOut.Yes",]);

summary(m3)
```

```
##
## Call:
## lm(formula = price ~ tax + mileage + years_sell2 + model + engineSize,
##     data = df[!df$mout == "MvOut.Yes", ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26199.6  -2501.7   -236.7   1990.8  24092.0
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              2.804e+04  4.230e+02  66.299  < 2e-16 ***
## tax                     -8.172e+00  1.231e+00  -6.640 3.48e-11 ***
## mileage                 -1.682e-01  4.203e-03 -40.029  < 2e-16 ***
## years_sell2             -5.166e+03  1.576e+02 -32.780  < 2e-16 ***
## modelAudi- A3            2.590e+03  4.373e+02   5.924 3.37e-09 ***
## modelAudi- A4            2.746e+03  5.083e+02   5.403 6.85e-08 ***
## modelAudi- A5            4.960e+03  5.687e+02   8.721  < 2e-16 ***
## modelAudi- A6            4.455e+03  6.095e+02   7.309 3.14e-13 ***
## modelAudi- A7            4.405e+03  1.330e+03   3.312 0.000933 ***
## modelAudi- A8            9.609e+03  1.473e+03   6.525 7.52e-11 ***
## modelAudi- Q2            2.990e+03  5.509e+02   5.427 6.00e-08 ***
## modelAudi- Q3            5.455e+03  4.984e+02  10.945  < 2e-16 ***
## modelAudi- Q5            1.030e+04  5.475e+02  18.807  < 2e-16 ***
## modelAudi- Q7            1.798e+04  8.020e+02  22.424  < 2e-16 ***
## modelAudi- Q8            2.545e+04  2.339e+03  10.881  < 2e-16 ***
## modelAudi- RS3           1.568e+04  2.837e+03   5.527 3.43e-08 ***
## modelAudi- RS4           1.460e+04  4.038e+03   3.616 0.000303 ***
## modelAudi- RS5           3.179e+04  2.837e+03  11.207  < 2e-16 ***
## modelAudi- RS6           2.955e+04  4.007e+03   7.374 1.94e-13 ***
## modelAudi- S3            9.855e+03  3.994e+03   2.467 0.013644 *
## modelAudi- S4            1.112e+04  4.002e+03   2.778 0.005483 **
## modelAudi- S5           -3.846e+02  4.003e+03  -0.096 0.923461
## modelAudi- S8            1.119e+04  4.006e+03   2.794 0.005234 **
## modelAudi- SQ5           1.210e+04  2.036e+03   5.942 3.01e-09 ***
## modelAudi- TT            6.590e+03  8.398e+02   7.847 5.21e-15 ***
```

```
## modelBMW- 1 Series       -1.325e+02  4.427e+02  -0.299 0.764663    
## modelBMW- 2 Series       -1.380e+02  4.807e+02  -0.287 0.774022    
## modelBMW- 3 Series        2.010e+03  4.437e+02   4.531 6.01e-06 ***
## modelBMW- 4 Series        1.568e+03  5.417e+02   2.895 0.003807 ** 
## modelBMW- 5 Series        3.914e+03  5.300e+02   7.385 1.79e-13 ***
## modelBMW- 6 Series        4.585e+03  1.173e+03   3.908 9.42e-05 ***
## modelBMW- 7 Series        1.443e+04  1.331e+03  10.841  < 2e-16 ***
## modelBMW- M2              6.068e+03  2.339e+03   2.595 0.009497 ** 
## modelBMW- M3              1.195e+04  2.340e+03   5.109 3.37e-07 ***
## modelBMW- M4              1.630e+04  1.230e+03  13.257  < 2e-16 ***
## modelBMW- X1              2.724e+03  5.482e+02   4.969 6.97e-07 ***
## modelBMW- X2              3.634e+03  8.616e+02   4.218 2.51e-05 ***
## modelBMW- X3              8.504e+03  6.679e+02  12.732  < 2e-16 ***
## modelBMW- X4              9.844e+03  9.879e+02   9.964  < 2e-16 ***
## modelBMW- X5              1.397e+04  7.712e+02  18.119  < 2e-16 ***
## modelBMW- X6              1.316e+04  1.832e+03   7.184 7.81e-13 ***
## modelBMW- Z3             -9.709e+02  2.847e+03  -0.341 0.733088    
## modelBMW- Z4              6.450e+03  1.207e+03   5.342 9.59e-08 ***
## modelMercedes- A Class    2.224e+03  4.089e+02   5.438 5.65e-08 ***
## modelMercedes- B Class    1.136e+03  6.230e+02   1.823 0.068425 .  
## modelMercedes- C Class    4.656e+03  4.033e+02  11.545  < 2e-16 ***
## modelMercedes- CL Class   4.897e+03  7.660e+02   6.393 1.78e-10 ***
## modelMercedes- CLA Class  5.434e+03  1.303e+03   4.170 3.10e-05 ***
## modelMercedes- CLK        4.415e+03  2.862e+03   1.543 0.123008    
## modelMercedes- CLS Class  5.651e+03  1.096e+03   5.158 2.59e-07 ***
## modelMercedes- E Class    5.196e+03  4.622e+02  11.240  < 2e-16 ***
## modelMercedes- GL Class   3.855e+03  1.251e+03   3.080 0.002081 ** 
## modelMercedes- GLA Class  2.716e+03  5.954e+02   4.561 5.23e-06 ***
## modelMercedes- GLC Class  1.168e+04  5.191e+02  22.491  < 2e-16 ***
## modelMercedes- GLE Class  1.868e+04  7.062e+02  26.450  < 2e-16 ***
## modelMercedes- GLS Class  1.899e+04  2.036e+03   9.324  < 2e-16 ***
## modelMercedes- M Class    5.790e+03  1.476e+03   3.923 8.85e-05 ***
## modelMercedes- S Class    1.346e+04  1.224e+03  11.000  < 2e-16 ***
## modelMercedes- SL CLASS   5.865e+03  1.086e+03   5.398 7.06e-08 ***
## modelMercedes- SLK       -7.946e+02  1.547e+03  -0.514 0.607610    
## modelMercedes- V Class    9.513e+03  8.503e+02  11.188  < 2e-16 ***
## modelMercedes- X-CLASS    5.745e+03  1.264e+03   4.546 5.61e-06 ***
## modelVW- Amarok           2.524e+03  1.280e+03   1.971 0.048726 *  
## modelVW- Arteon           3.102e+03  9.591e+02   3.234 0.001229 ** 
## modelVW- Beetle          -2.296e+03  1.371e+03  -1.675 0.094000 .  
## modelVW- Caddy Maxi Life -1.182e+03  2.022e+03  -0.585 0.558898    
## modelVW- California       3.618e+04  2.842e+03  12.730  < 2e-16 ***
## modelVW- Caravelle        1.441e+04  1.380e+03  10.436  < 2e-16 ***
## modelVW- CC              -1.408e+03  1.374e+03  -1.025 0.305534    
## modelVW- Eos             -1.965e+03  3.997e+03  -0.492 0.622975    
## modelVW- Golf             3.613e+02  3.770e+02   0.958 0.337988    
## modelVW- Golf SV         -1.586e+03  8.446e+02  -1.878 0.060447 .  
## modelVW- Jetta           -7.731e+03  2.833e+03  -2.729 0.006383 ** 
## modelVW- Passat           1.038e+03  5.264e+02   1.973 0.048578 *  
## modelVW- Polo            -4.053e+03  3.849e+02 -10.531  < 2e-16 ***
## modelVW- Scirocco        -4.822e+02  8.285e+02  -0.582 0.560616    
## modelVW- Sharan           1.087e+03  9.033e+02   1.203 0.228948    
## modelVW- Shuttle          3.806e+03  1.666e+03   2.285 0.022379 *  
## modelVW- T-Cross         -1.408e+03  7.509e+02  -1.875 0.060803 .  
## modelVW- T-Roc            1.939e+03  5.915e+02   3.278 0.001053 ** 
## modelVW- Tiguan           2.718e+03  4.626e+02   5.875 4.51e-09 ***
## modelVW- Tiguan Allspace  5.288e+03  1.306e+03   4.048 5.25e-05 ***
## modelVW- Touareg          9.181e+03  8.481e+02  10.826  < 2e-16 ***
## modelVW- Touran           2.731e+03  7.152e+02   3.819 0.000136 ***
## modelVW- Up              -7.350e+03  5.283e+02 -13.912  < 2e-16 ***
## engineSizeMitjà           3.717e+03  1.678e+02  22.157  < 2e-16 ***
## engineSizeGran            9.133e+03  2.997e+02  30.476  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3977 on 4789 degrees of freedom
## Multiple R-squared:  0.831,  Adjusted R-squared:  0.828
## F-statistic: 273.9 on 86 and 4789 DF,  p-value: < 2.2e-16
```

We can see that adding only two factors we have captured 84% of the variability thanks to this model. It's a really good result.

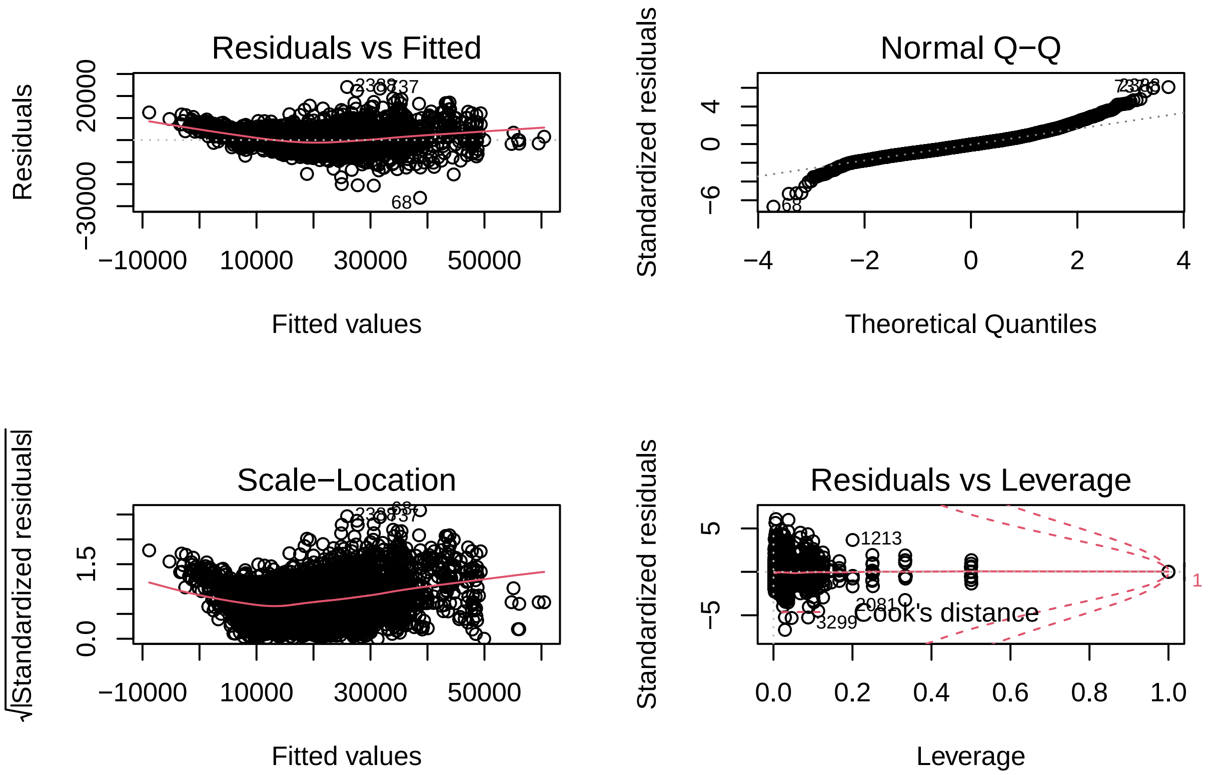
Let's check the plots to see how are the residuals' normality and leverage.

```
par(mfrow=c(2,2))
plot(m3)
```

```
## Warning: not plotting observations with leverage one:
##   464, 745, 969, 4876
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



We can see that even though the variability retention of this model is excellent, the residuals don't have a good behaviour as: *The extreme quantiles don't follow a normal distribution* There are some extreme values that need to be removed (number 2388, 1015, 2741 etc) which have a big leverage and affect the regression. *The scale location graph's red line is not exactly horizontal.

Let's check for correlated variables:

```
vif(m3)
```

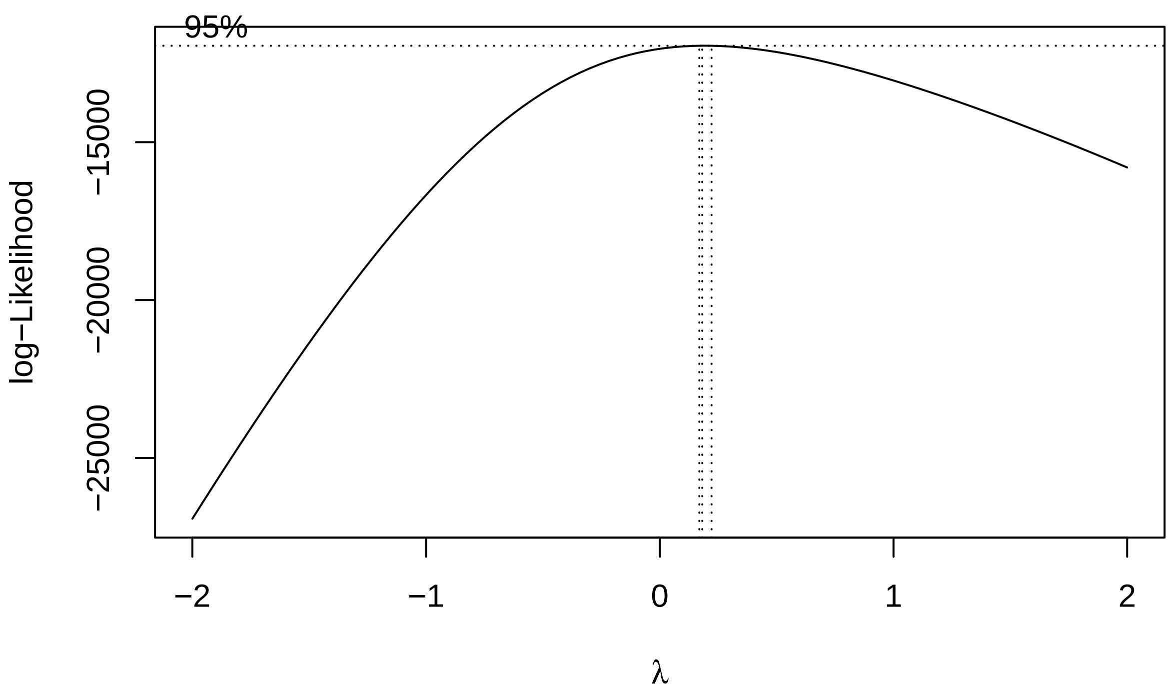
```
##                  GVIF Df GVIF^(1/(2*Df))
## tax          1.518413  1        1.232239
## mileage      2.262666  1        1.504216
## years_sell2  2.386401  1        1.544798
## model        4.737313 81        1.009648
## engineSize   3.647713  2        1.381991
```

The variables which have the highest correlations in the model are model, engineSize and mileage

```
Anova(m3)
```

```
## Anova Table (Type II tests)
##
## Response: price
##                 Sum Sq   Df  F value    Pr(>F)
## tax         6.9739e+08    1   44.095 3.476e-11 ***
## mileage     2.5342e+10    1 1602.352 < 2.2e-16 ***
## years_sell2 1.6994e+10    1 1074.539 < 2.2e-16 ***
## model       6.6014e+10   81   51.531 < 2.2e-16 ***
## engineSize  1.5579e+10    2  492.533 < 2.2e-16 ***
## Residuals   7.5741e+10 4789
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
boxcox(price~tax+mileage+years_sell2+model+engineSize,data=df[!df$mout=="MvOut.Yes",])
```



As we can see in the coxbox plot the lambda is a value near 0 so a log function will have to be applied to the target value to make a better relation with the variables.

4.2.2 Model 4: log(price) ~ mileage+tax+years_sell2 + model+engineSize

```
m4<-lm(log(price)~tax+mileage+years_sell2+model+engineSize,data=df[!df$mout=="MvOut.Yes",])
summary(m4)
```

```
##
## Call:
## lm(formula = log(price) ~ tax + mileage + years_sell2 + model +
##     engineSize, data = df[!df$mout == "MvOut.Yes", ])
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max 
## -1.97723 -0.09981 -0.00060  0.10078  0.73755 
## 
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)    
## (Intercept)              1.015e+01  1.817e-02 558.654  < 2e-16 ***
## tax                     -6.575e-05  5.286e-05  -1.244 0.213577    
## mileage                 -1.006e-05  1.805e-07 -55.754  < 2e-16 ***
## years_sell2             -2.158e-01  6.768e-03 -31.886  < 2e-16 ***
## modelAudi- A3            1.502e-01  1.878e-02   7.996 1.60e-15 ***
## modelAudi- A4            1.613e-01  2.183e-02   7.387 1.76e-13 ***
## modelAudi- A5            2.604e-01  2.443e-02  10.661  < 2e-16 ***
## modelAudi- A6            2.604e-01  2.618e-02   9.946  < 2e-16 ***
## modelAudi- A7            2.600e-01  5.713e-02   4.552 5.44e-06 ***
## modelAudi- A8            3.714e-01  6.325e-02   5.872 4.59e-09 ***
## modelAudi- Q2            1.938e-01  2.366e-02   8.192 3.26e-16 ***
## modelAudi- Q3            2.837e-01  2.141e-02  13.253  < 2e-16 ***
## modelAudi- Q5            4.469e-01  2.352e-02  19.003  < 2e-16 ***
## modelAudi- Q7            5.750e-01  3.444e-02  16.693  < 2e-16 ***
## modelAudi- Q8            6.652e-01  1.004e-01   6.623 3.90e-11 ***
## modelAudi- RS3           5.533e-01  1.218e-01   4.542 5.71e-06 ***
## modelAudi- RS4           7.700e-01  1.734e-01   4.440 9.18e-06 ***
## modelAudi- RS5           8.912e-01  1.218e-01   7.316 2.98e-13 ***
## modelAudi- RS6           9.611e-01  1.721e-01   5.585 2.47e-08 ***
## modelAudi- S3            5.141e-01  1.715e-01   2.997 0.002742 ** 
## modelAudi- S4            3.659e-01  1.719e-01   2.129 0.033306 *  
## modelAudi- S5            8.319e-02  1.719e-01   0.484 0.628496    
## modelAudi- S8            5.209e-01  1.720e-01   3.028 0.002473 ** 
## modelAudi- SQ5           5.722e-01  8.744e-02   6.544 6.62e-11 ***
## modelAudi- TT            3.331e-01  3.607e-02   9.236  < 2e-16 ***
## modelBMW- 1 Series       6.241e-04  1.901e-02   0.033 0.973817    
## modelBMW- 2 Series       3.626e-02  2.065e-02   1.757 0.079058 .  
## modelBMW- 3 Series       9.485e-02  1.905e-02   4.978 6.66e-07 ***
## modelBMW- 4 Series       1.324e-01  2.326e-02   5.690 1.34e-08 ***
## modelBMW- 5 Series       2.153e-01  2.276e-02   9.457  < 2e-16 ***
## modelBMW- 6 Series       2.490e-01  5.038e-02   4.941 8.02e-07 ***
## modelBMW- 7 Series       4.674e-01  5.716e-02   8.178 3.67e-16 ***
## modelBMW- M2             2.372e-01  1.004e-01   2.361 0.018262 *  
## modelBMW- M3             5.238e-01  1.005e-01   5.212 1.94e-07 ***
## modelBMW- M4             4.778e-01  5.281e-02   9.047  < 2e-16 ***
## modelBMW- X1             1.693e-01  2.354e-02   7.191 7.43e-13 ***
## modelBMW- X2             1.766e-01  3.700e-02   4.772 1.87e-06 ***
## modelBMW- X3             3.786e-01  2.869e-02  13.197  < 2e-16 ***
## modelBMW- X4             4.166e-01  4.243e-02   9.818  < 2e-16 ***
## modelBMW- X5             5.106e-01  3.312e-02  15.418  < 2e-16 ***
## modelBMW- X6             5.288e-01  7.867e-02   6.722 2.00e-11 ***
## modelBMW- Z3            -6.256e-01  1.223e-01  -5.116 3.24e-07 ***
## modelBMW- Z4             1.468e-01  5.186e-02   2.830 0.004674 ** 
## modelMercedes- A Class   1.494e-01  1.756e-02   8.509  < 2e-16 ***
## modelMercedes- B Class   1.029e-01  2.676e-02   3.847 0.000121 ***
## modelMercedes- C Class   2.481e-01  1.732e-02  14.321  < 2e-16 ***
## modelMercedes- CL Class  2.953e-01  3.290e-02   8.977  < 2e-16 ***
## modelMercedes- CLA Class  3.207e-01  5.597e-02   5.730 1.07e-08 ***
## modelMercedes- CLK      -6.373e-01  1.229e-01  -5.184 2.26e-07 ***
## modelMercedes- CLS Class  2.848e-01  4.705e-02   6.053 1.53e-09 ***
## modelMercedes- E Class   2.698e-01  1.985e-02  13.590  < 2e-16 ***
## modelMercedes- GL Class  2.517e-01  5.375e-02   4.683 2.90e-06 ***
## modelMercedes- GLA Class  1.962e-01  2.557e-02   7.673 2.02e-14 ***
## modelMercedes- GLC Class  4.918e-01  2.230e-02  22.057  < 2e-16 ***
## modelMercedes- GLE Class  6.180e-01  3.033e-02  20.378  < 2e-16 ***
## modelMercedes- GLS Class  6.035e-01  8.745e-02   6.901 5.86e-12 ***
## modelMercedes- M Class   3.533e-01  6.338e-02   5.575 2.62e-08 ***
## modelMercedes- S Class   4.473e-01  5.257e-02   8.509  < 2e-16 ***
## modelMercedes- SL CLASS  2.997e-01  4.666e-02   6.424 1.46e-10 ***
```

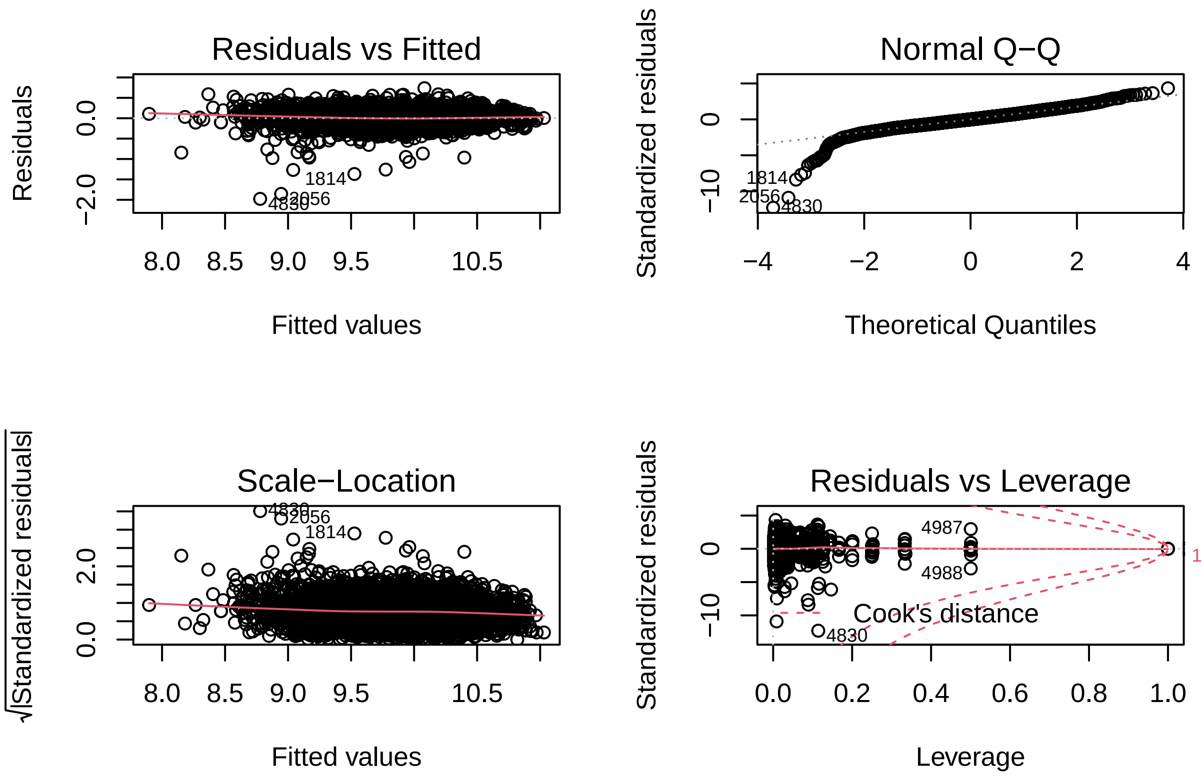
```
## modelMercedes- SLK       -1.288e-01  6.646e-02  -1.938 0.052628 .
## modelMercedes- V Class    3.481e-01  3.652e-02   9.533  < 2e-16 ***
## modelMercedes- X-CLASS    2.520e-01  5.428e-02   4.643 3.53e-06 ***
## modelVW- Amarok           1.424e-01  5.498e-02   2.590 0.009634 **
## modelVW- Arteon           1.763e-01  4.119e-02   4.279 1.91e-05 ***
## modelVW- Beetle          -6.023e-01  5.887e-02 -10.230  < 2e-16 ***
## modelVW- Caddy Maxi Life  1.094e-02  8.685e-02   0.126 0.899729
## modelVW- California       9.342e-01  1.220e-01   7.654 2.34e-14 ***
## modelVW- Caravelle        5.471e-01  5.929e-02   9.228  < 2e-16 ***
## modelVW- CC              -1.132e-01  5.903e-02  -1.918 0.055117 .
## modelVW- Eos             -4.566e-01  1.717e-01  -2.660 0.007849 **
## modelVW- Golf             2.691e-02  1.619e-02   1.662 0.096576 .
## modelVW- Golf SV         -9.225e-02  3.627e-02  -2.543 0.011017 *
## modelVW- Jetta           -6.482e-01  1.217e-01  -5.327 1.05e-07 ***
## modelVW- Passat           5.978e-02  2.261e-02   2.644 0.008218 **
## modelVW- Polo            -2.908e-01  1.653e-02 -17.593  < 2e-16 ***
## modelVW- Scirocco        -5.804e-03  3.558e-02  -0.163 0.870434
## modelVW- Sharan           1.035e-01  3.879e-02   2.668 0.007657 **
## modelVW- Shuttle          2.146e-01  7.155e-02   2.999 0.002718 **
## modelVW- T-Cross         -3.169e-03  3.225e-02  -0.098 0.921743
## modelVW- T-Roc            1.357e-01  2.540e-02   5.342 9.64e-08 ***
## modelVW- Tiguan           1.594e-01  1.987e-02   8.021 1.31e-15 ***
## modelVW- Tiguan Allspace  2.705e-01  5.610e-02   4.822 1.47e-06 ***
## modelVW- Touareg          3.746e-01  3.642e-02  10.284  < 2e-16 ***
## modelVW- Touran           1.667e-01  3.072e-02   5.428 5.99e-08 ***
## modelVW- Up              -6.218e-01  2.269e-02 -27.406  < 2e-16 ***
## engineSizeMitjà           1.909e-01  7.205e-03  26.493  < 2e-16 ***
## engineSizeGran            3.889e-01  1.287e-02  30.219  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1708 on 4789 degrees of freedom
## Multiple R-squared:  0.868,  Adjusted R-squared:  0.8656
## F-statistic:   366 on 86 and 4789 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(2,2))
plot(m4)
```

```
## Warning: not plotting observations with leverage one:
##   464, 745, 969, 4876
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



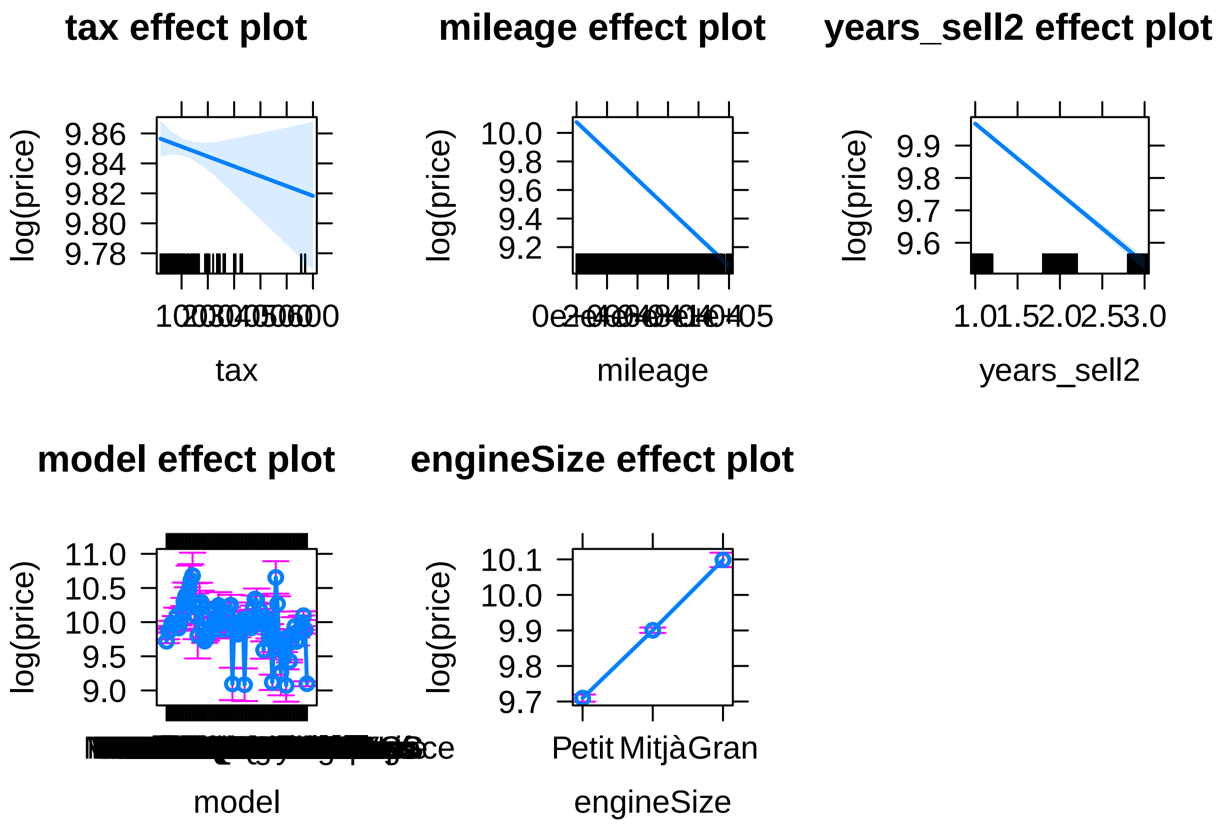
The normality of the regression has improved thanks to the transfromation but there is a bad normal distribution for lower quantiles. Residuals are linear distributed. There is some influent data that will be removed at the end.

```
vif(m4)
```

```
##                   GVIF Df GVIF^(1/(2*Df))
## tax          1.518413  1        1.232239
## mileage      2.262666  1        1.504216
## years_sell2 2.386401  1        1.544798
## model        4.737313 81        1.009648
## engineSize   3.647713  2        1.381991
```

We have good values for VIF, correlation doesn't have a big effect on our regression

```
plot(allEffects(m4))
```




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

```
## Warning in AIC.default(m1, m2, m3, m4): models are not all fitted to the same
## number of observations
```

```
##    df        AIC
## m1  6 99971.871
## m2  5  3356.272
## m3 88 94752.731
## m4 88 -3308.528
```

AIC function shows that the best fitted model is model 4 the last one because its AIC is the lower. We can see the positive linear relation between engine size and price too.

```
anova(m4)
```

```
## Analysis of Variance Table
##
## Response: log(price)
##                Df Sum Sq Mean Sq  F value    Pr(>F)
## tax             1 169.90 169.901  5823.84 < 2.2e-16 ***
## mileage         1 310.19 310.185 10632.52 < 2.2e-16 ***
## years_sell2     1  47.63  47.633  1632.78 < 2.2e-16 ***
## model          81 359.23   4.435   152.02 < 2.2e-16 ***
## engineSize      2  31.36  15.679   537.46 < 2.2e-16 ***
## Residuals    4789 139.71   0.029
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value of the test is less than 0.05 for all variables we can reject null hypothesis and for all chosen variables have effect on the prediciton of the target value.

Finally we will check if adding the variables manufacturer and transmission can improve the model in a significant way. We can see that they does not add variability so we will not consider them to make the model more robust.

```
m4aux<-lm(log(price)~tax+mileage+years_sell2+model+engineSize+manufacturer+transmission,data=df[!df$mout

# Summary -> Anex A1
```

4.3 Adding interactions

Once we have selected the model with covariates and factors, we will proceed to add interaction between all variables (including factors) and all factors and we will proceed to check which ones have more impact in the resulting model.

4.3.1 Model 5

```
m5<-lm(log(price)~(tax+mileage+years_sell2+engineSize+model+transmission+fuelType)*(engineSize+model+tra
# output of step(m5) -> Annex - A2
```

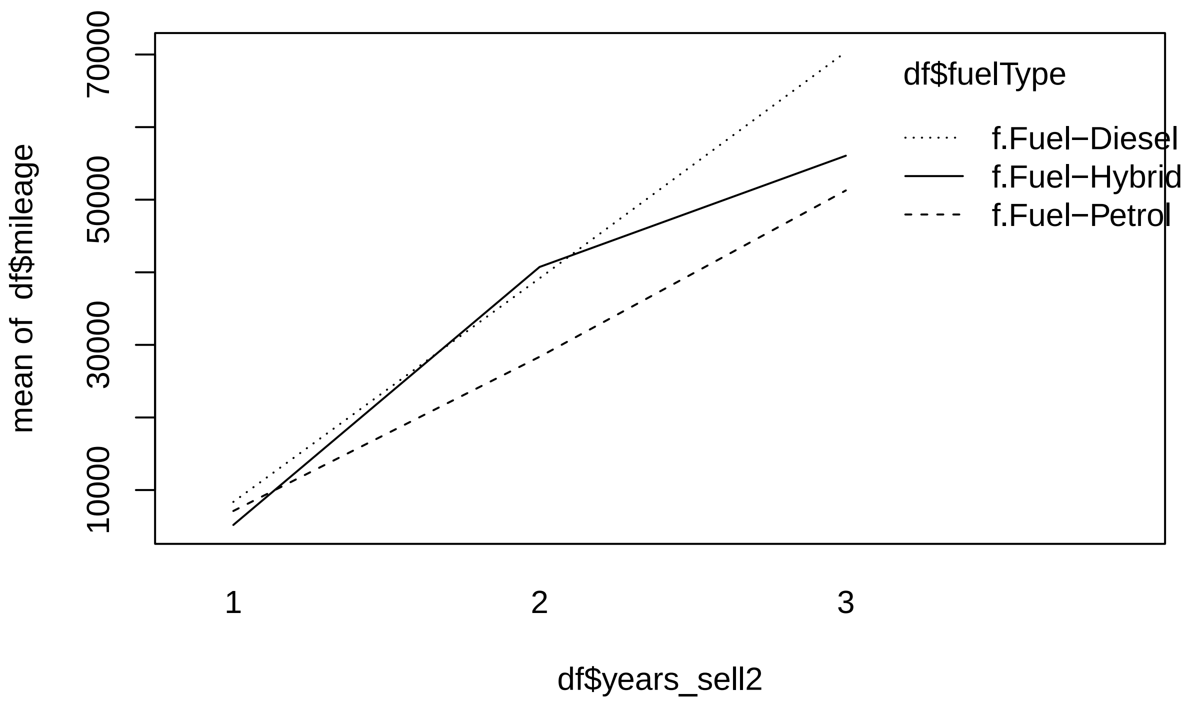
Our first model adding interactions is the model m5. At the end of the output of the step function we can see that the most important interactions are the next ones;

mileage:engineSize
model:fuelType *tax:transmission* years_sell2:fuelType *mileage:transmission* engineSize:model
engineSize:fuelType mileage:fuelType

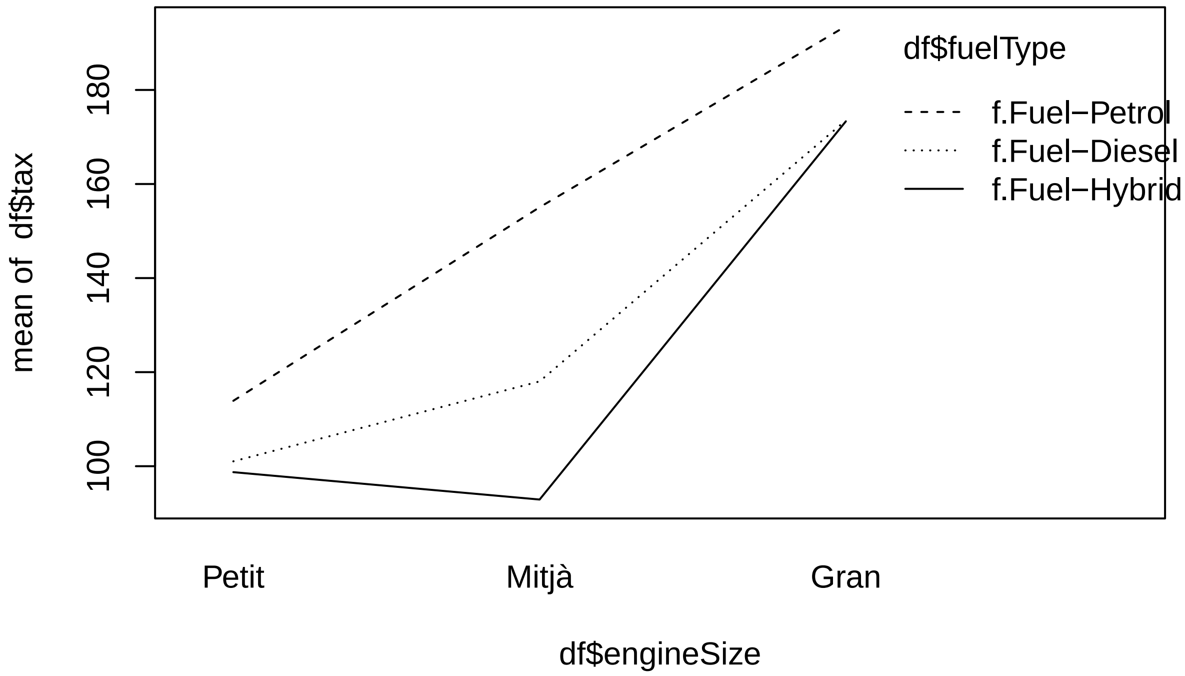
We can see that there is a correlation between factors and a correlation between a factor and a covariate as the statement asked. The resultant model, the proposed by the step function is the next one:

log(price) ~ tax + mileage + years_sell2 + engineSize + model + transmission + fuelType + tax:transmission + mileage:engineSize + mileage:transmission + mileage:fuelType + years_sell2:fuelType + engineSize:model + engineSize:fuelType + model:fuelType

```
interaction.plot(df$years_sell2,df$fuelType,df$mileage)
```



```
interaction.plot(df$engineSize,df$fuelType,df$tax)
```



We can see that the mean of mileage is bigger for older cars. Cars of the three different types behave in the same way. How much older they are more mileage they have.

We can see that that the mean of tax grows with the engineSize for all types of typeFuel. Interaction is thus present between these three variables. We can see that for the hybrid cars the tax decreases a little bit for the medium engine cars but then for the big engine cars it increases to the same level of Diesel cars.

```
AIC(m1,m2,m3,m4,m5)
```

```
## Warning in AIC.default(m1, m2, m3, m4, m5): models are not all fitted to the
## same number of observations

##      df       AIC
## m1    6 99971.871
## m2    5  3356.272
## m3   88 94752.731
## m4   88 -3308.528
## m5  547 -4940.884
```

We can see that the most explicative value of all the ones that we have created is the number 5 as it includes covariates, factors and the interactions proposed by the step method. Before finishing with the model selection we will analise influent data to try to make the residual linearity a little better.

log(price) ~ tax + mileage + years_sell2 + engineSize + model + transmission + fuelType + tax:transmission + mileage:engineSize + mileage:transmission + mileage:fuelType + years_sell2:fuelType + engineSize:model + engineSize:fuelType + model:fuelType

4.4 Influent data and outliers

During the realization of the analysis we have seen (in the first model) that multivariant outliers of the dataset have a negative impact on the independence on the residuals. What is more we have discovered some influential data that hava an impact on the data distributed in the first quantile. We will now proceed to remove this data from the analysis.

For the model 5, our last model, we can see that there are some values that have a big impact on the studentized values. We will proceed to extract them from the analysis and then check the normality of the residuals another time.

4.4.1 Model 6

```
m6<-lm(log(price) ~ tax + mileage + years_sell2 + engineSize + model + transmission + fuelType + tax:tra
```

```
#output in Annex - A3
```

Influential observations are those whose leverage is over 0.06. 117 observations satisfy this condition. Observations 1616, 1776, 2329, 3102, 3228 and 4742 are outliers for Cook's distance (over 0.05).

4.5 Conclusion

The bestfitted model found for our data is the model 6 described in the previous section.

```
AIC(m5, m6)
```

```
## Warning in AIC.default(m5, m6): models are not all fitted to the same number of
## observations
```

```
##     df       AIC
## m5 547 -4940.884
## m6 223 -4599.409
```

```
Anova(m5,m6)
```

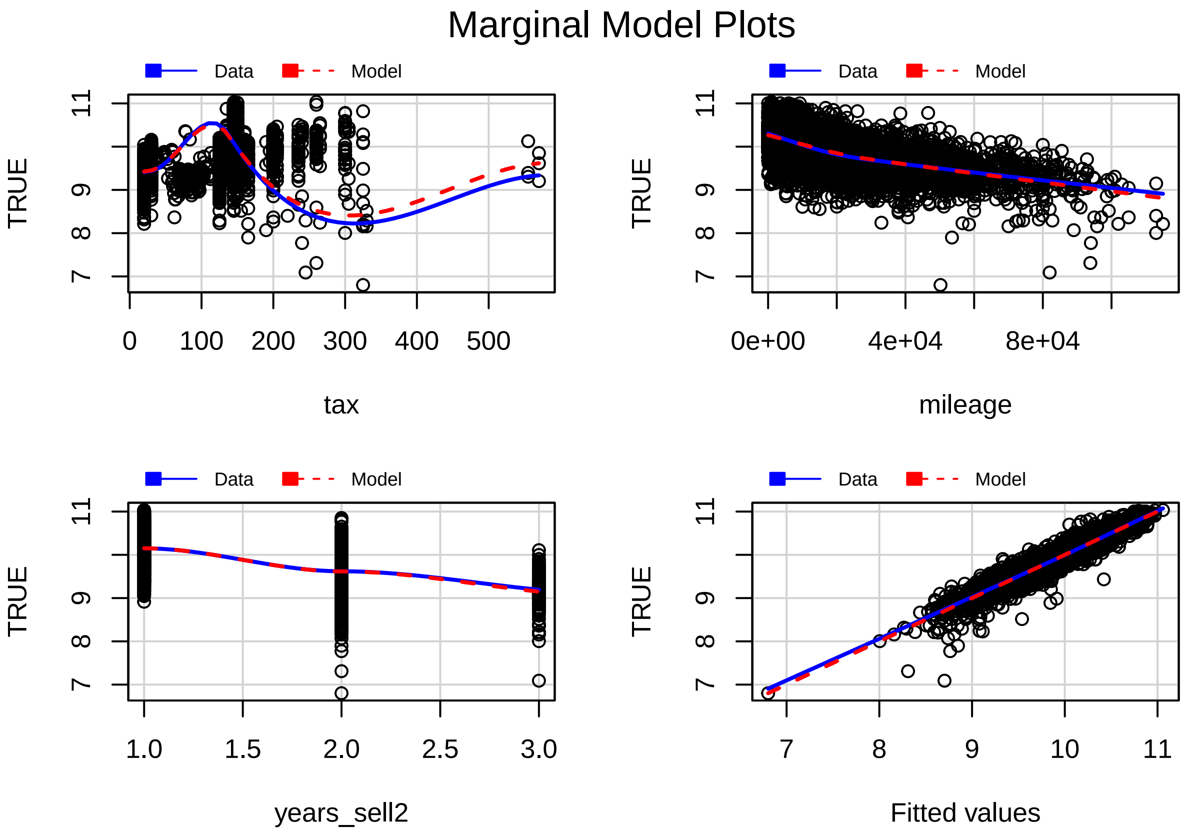
```
## Note: model has aliased coefficients
##       sums of squares computed by model comparison
```

```
## Anova Table (Type II tests)
##
## Response: log(price)
##                            Sum Sq  Df   F value    Pr(>F)
## tax                         0.083   1    4.2796  0.038631 *
## mileage                    62.640   1 3212.9058 < 2.2e-16 ***
## years_sell2                19.343   1  992.1103 < 2.2e-16 ***
## engineSize                 21.325   2  546.9064 < 2.2e-16 ***
## model                     135.549  86   80.8431 < 2.2e-16 ***
## transmission                5.399   2  138.4506 < 2.2e-16 ***
## fuelType                    3.402   2   87.2576 < 2.2e-16 ***
## tax:engineSize              0.076   2    1.9469  0.142838
## tax:model                   3.576  62    2.9582 1.432e-13 ***
## tax:transmission            0.234   2    6.0086  0.002478 **
## tax:fuelType                0.256   2    6.5763  0.001407 **
## mileage:engineSize          0.012   2    0.3129  0.731338
## mileage:model               3.704  69    2.7537 6.927e-13 ***
## mileage:transmission        0.236   2    6.0439  0.002392 **
## mileage:fuelType            1.381   2   35.4069 5.558e-16 ***
## years_sell2:engineSize      0.053   2    1.3667  0.255046
## years_sell2:model           2.338  52    2.3058 3.667e-07 ***
## years_sell2:transmission    0.016   2    0.4102  0.663551
## years_sell2:fuelType        0.075   2    1.9287  0.145464
## engineSize:model            7.120  56    6.5209 < 2.2e-16 ***
## engineSize:transmission     0.277   4    3.5554  0.006678 **
## engineSize:fuelType         0.982   3   16.7881 7.663e-11 ***
## model:transmission          2.504 100    1.2845  0.030911 *
```

```
## model:fuelType              5.224   57    4.7009 < 2.2e-16 ***
## transmission:fuelType       0.147    3    2.5052  0.057298 .
## Residuals                  86.096 4416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
marginalModelPlots(m6)
```

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



The shape of the model follows the data and the fitted values shows us a stronger model. Blue and red data are nearly superposed and follow the same function.

5 Description of Model Building process for prediction of binary response (Audi).

In the second part of the assignament we will go through the process of creating a forecasting model for the prediction of the binary variable Audi. So our objective is to create a model that helps us to predict the probability of a certain input of data corresponds to an audi car or not.

5.1 Split into train and test

```
# 80% train sample and 20% test sample
set.seed(1234)
llwork <- sample(1:nrow(df),round(0.80*nrow(df),0))

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

5.2 Binary Models: Using numerical explanatory variables

```
res.cat <- catdes(df, num.var = which(names(df)=="Audi"))
res.cat$quanti.var
```

```
##                      Eta2      P-value
## mpg         0.007593209 7.829241e-10
## price       0.003611130 2.277616e-05
## mileage     0.002087672 1.284525e-03
## years_sell2 0.001174984 1.574831e-02
```

Before starting with the model building process, we have executed the catdes method to try to visualize if there is a high correlation between the target variable and the numeric explanatory variables. We can reject the null hypothesis so there is correlation with the binary variable with all the variables.

```
ll<-which(df_train$years_sell2==0);ll
df$years_sell2[ll]<-0.5

ll<-which(df_train$tax==0);ll
df$tax[ll]<-0.5

ll<-which(df_train$mpg==0);ll
df$mpg[ll]<-0.5

ll<-which(df_train$mileage==0);ll
df$mileage[ll]<-0.5
```

5.2.1 Model 1: Audi ~ mgp+mileage+tax+years_sell2

```
bm1<-glm(Audi~mileage+tax+mpg+years_sell2,family="binomial"(link = logit),data=df)
summary(bm1)
```

```
##
## Call:
## glm(formula = Audi ~ mileage + tax + mpg + years_sell2, family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3314  -0.7169  -0.6311  -0.4695   2.2024
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.902e-01  2.820e-01   2.093 0.036354 *
## mileage      7.328e-06  2.243e-06   3.267 0.001087 **
## tax         -2.599e-03  7.353e-04  -3.535 0.000408 ***
## mpg         -4.029e-02  4.206e-03  -9.579  < 2e-16 ***
## years_sell2  2.125e-01  9.257e-02   2.295 0.021721 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5041.6  on 4961  degrees of freedom
## Residual deviance: 4910.9  on 4957  degrees of freedom
## AIC: 4920.9
##
## Number of Fisher Scoring iterations: 5
```

```r
vif(bm1)
```

```
##     mileage         tax         mpg years_sell2
##    2.129238    1.576111    1.829975    2.253959
```

```r
Anova(bm1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Audi
##             LR Chisq Df Pr(>Chisq)
## mileage       10.426  1  0.0012429 **
## tax           12.639  1  0.0003779 ***
## mpg          114.242  1  < 2.2e-16 ***
## years_sell2    5.241  1  0.0220613 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

First of all we can see in the logistic regression output that according to the p-value, all four numeric variables are statistically significant. What is more the VIF value for the 4 variables is small and this support the idea that the four variables are significant to the model. We can see that all these continuous variables are important for this binary regression. The anova test suports the idea that the 4 variables are statistically significant for the prediction model construction.

5.2.2 Model 2: Audi ~ mgp+mileage+tax

We can see that years_sell2 has the biggest p-value, we will see if omitting this variable would change our model.

```r
bm2<-glm(Audi~mileage+tax+mpg,family="binomial"(link = logit),data=df);
summary(bm2)
```

```
##
## Call:
## glm(formula = Audi ~ mileage + tax + mpg, family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2717  -0.7129  -0.6319  -0.4811   2.2086
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  7.673e-01  2.701e-01   2.841 0.004500 **
## mileage      1.078e-05  1.641e-06   6.570 5.03e-11 ***
## tax         -2.709e-03  7.370e-04  -3.676 0.000237 ***
## mpg         -3.861e-02  4.101e-03  -9.414  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5041.6  on 4961  degrees of freedom
## Residual deviance: 4916.2  on 4958  degrees of freedom
## AIC: 4924.2
##
## Number of Fisher Scoring iterations: 5
```

```r
AIC(bm1,bm2)
```

```
##      df      AIC
## bm1  5 4920.910
## bm2  4 4924.151
```

```
anova(bm1,bm2)
```

```
## Analysis of Deviance Table
##
## Model 1: Audi ~ mileage + tax + mpg + years_sell2
## Model 2: Audi ~ mileage + tax + mpg
##   Resid. Df Resid. Dev Df Deviance
## 1      4957     4910.9
## 2      4958     4916.2 -1  -5.2409
```

We can see that the model bm2 is approximately as strong as bm1 with one less variable, we will then carry on with this model.

5.2.3 Model 3: Audi ~ mgp+mileage+years_sell2

In order to see if removing years_sell2 instead of mpg was a goof idea, we check the results with the following regression bm4.

```
bm3<-glm(Audi~mileage+tax+years_sell2,family="binomial"(link = logit),data=df);
summary(bm3)
```

```
##
## Call:
## glm(formula = Audi ~ mileage + tax + years_sell2, family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0413  -0.6842  -0.6538  -0.6419   1.8676
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.757e+00  1.538e-01 -11.428   <2e-16 ***
## mileage      4.470e-06  2.201e-06   2.031   0.0422 *
## tax          1.533e-03  6.072e-04   2.525   0.0116 *
## years_sell2  6.740e-02  9.000e-02   0.749   0.4539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5041.6  on 4961  degrees of freedom
## Residual deviance: 5025.2  on 4958  degrees of freedom
## AIC: 5033.2
##
## Number of Fisher Scoring iterations: 4
```

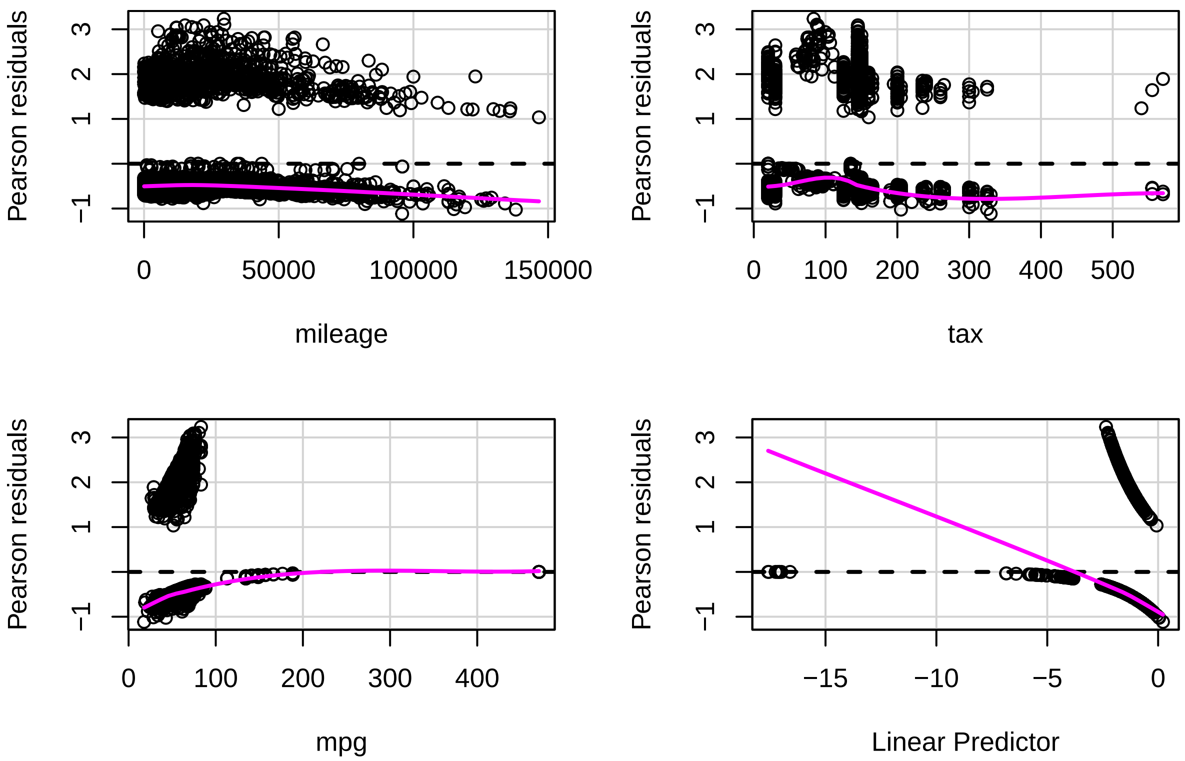
```
AIC(bm2,bm3)
```

```
##      df      AIC
## bm2  4 4924.151
## bm3  4 5033.152
```

This shows us that we made the right choice at the beginning as the AIC value is better for bm2 and the p-value of years_sell2 is too high in bm3 which makes the years_sell2 variable not having a big role (the smallest role) in this regression

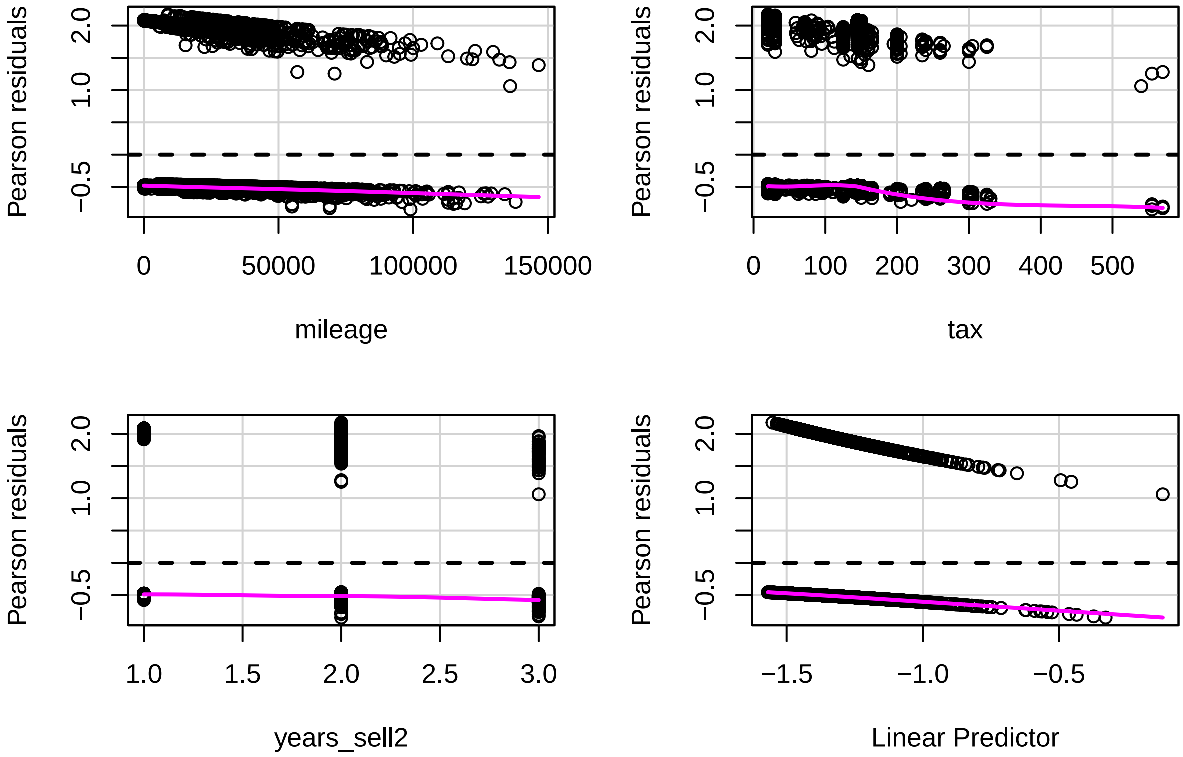
In order to definitively validate our model, we are going the plot the residual plots. This will make us take our final decision

```
residualPlots(bm2)
```




```
##           Test stat Pr(>|Test stat|)
## mileage      0.7054           0.4010
## tax          2.1332           0.1441
## mpg          0.0452           0.8316
```

```
residualPlots(bm3)
```




```
##              Test stat Pr(>|Test stat|)
## mileage         2.7682          0.09615 .
## tax             2.1910          0.13881
## years_sell2     0.0252          0.87391
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can clearly see that the residuals in the bm3 model have a better shape

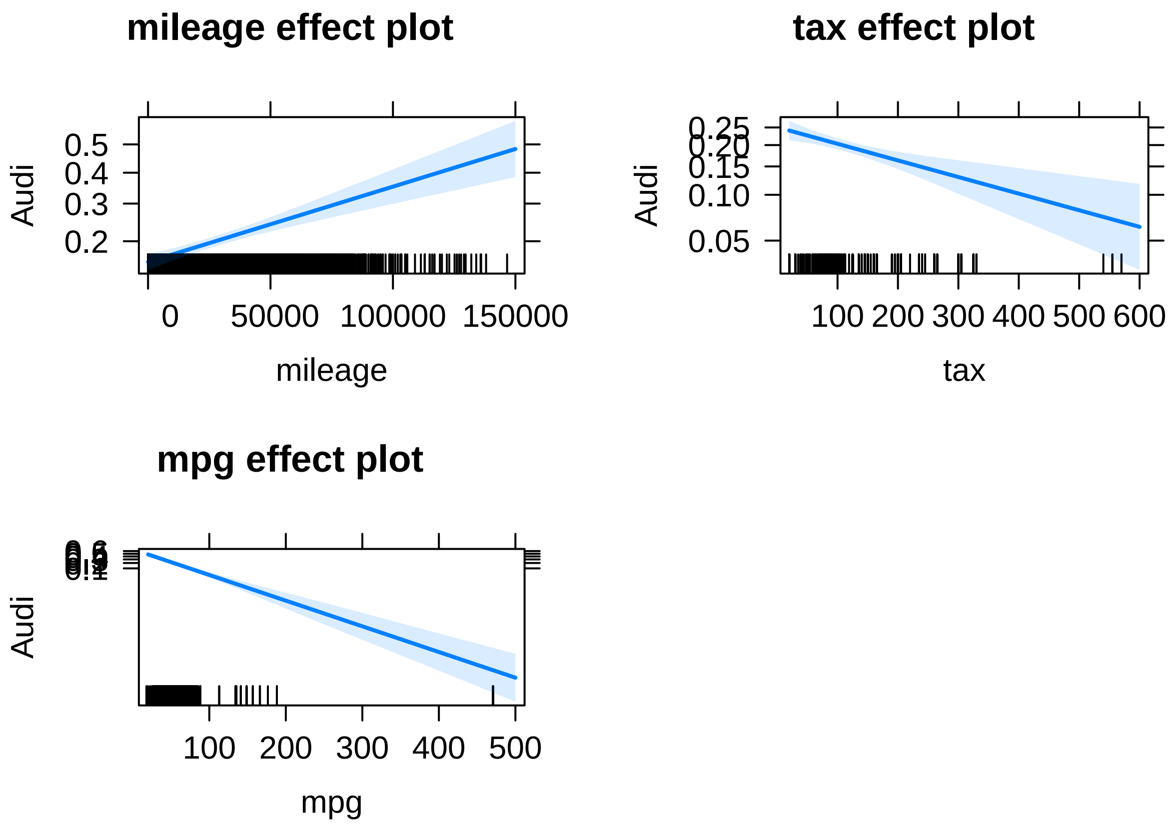
For • mileage: – we see that the smooth is plain, so it is ok. – the "weird" shapes that appear are because of the binary response model. • Tax: – we see that the smooth is plain, so it is ok. – the "weird" shapes that appear are because of the binary response model.
• Years_sell2 : – we see that the smooth is plain, so it is ok. – the "weird" shapes that appear are because of the binary response model

The overall shape of the linear predictor seems approximately plain, but as it was said in class, we can work with unfitted values in the model

Our chosen model is the binary model 2.

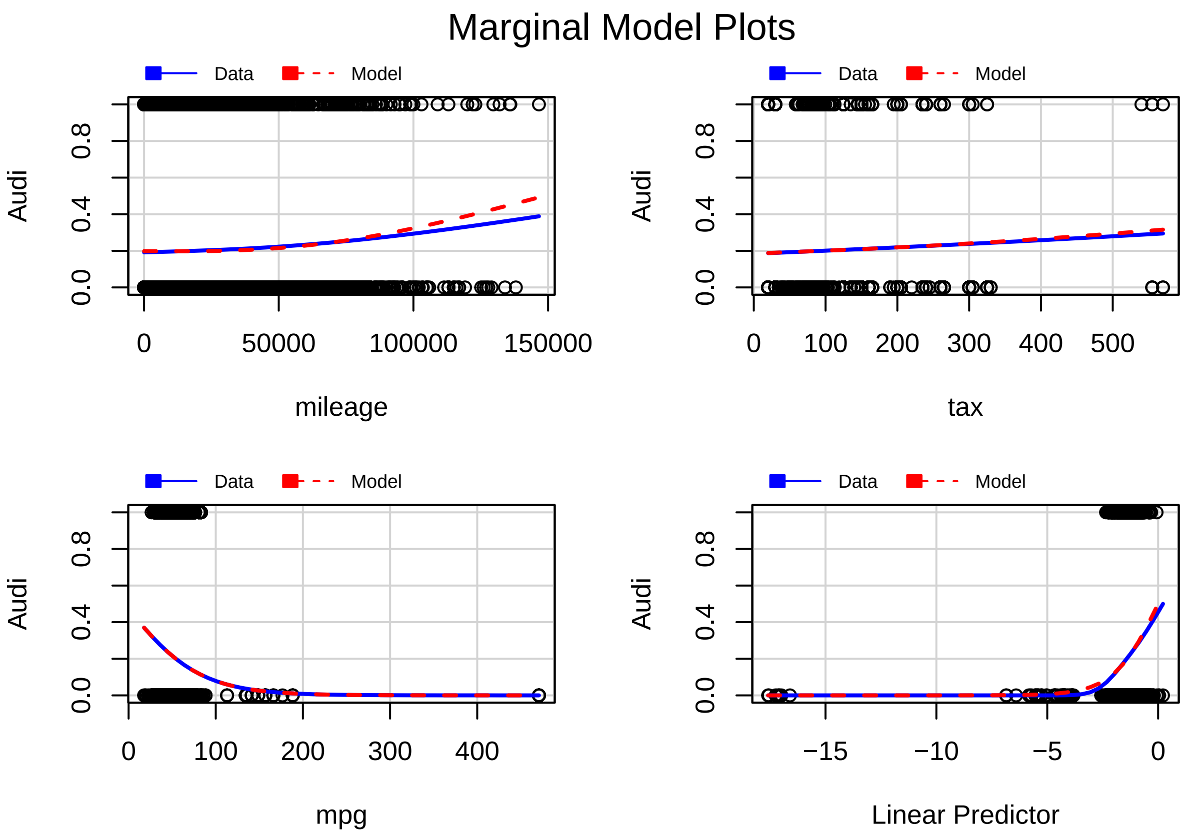
5.2.4 Understanding the model chosen (model 2)

```
plot(allEffects(bm2))
```



We can see that - As the mileage increases, the probability that tha car is an Audi between all the 4 brands increases. This show that Audi cars have a strong resistance to age. -As the tax price increases, the probability that the car is an Audi decreases, this show that the Audi cars are not that tax consuming, we must say though that the extreme values of tax aren't really too populated in order to give some shade to our interpretation -As the mpg variable increases, the probability of being an Audi decreases.

```
marginalModelPlots(bm2)
```



We can see that the data and the model are superposed.

5.3 Binary Models: Adding fators

We will now add factors to our bm4 linear model.

```
catdes(df,11)$test.chi2
```

```
##                  p.value df
## model        0.000000e+00 86
## manufacturer 0.000000e+00  3
## mpg_d        3.713009e-17  3
## fuelType     3.639271e-08  2
## f.price      3.230368e-05  3
## f.miles      4.712680e-04  3
## transmission 1.265959e-03  2
## f.tax        4.742923e-03  3
## aux          6.786383e-03  3
## years_sell   4.037010e-02  2
## hcpck        4.711223e-02  3
```

The factors most related to Audi are mpg_d, fuelType, f. miles (we won't use it because we already have the mileage variable) and transmission. We will try to include them in our new model bm4.

5.3.1 Model 4: Audi~mileage+tax+mpg+fuelType+transmission+engineSize

```
bm4<-glm(Audi~mileage+tax+mpg+fuelType+transmission+engineSize,family="binomial"(link = logit),data=df);
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(bm4)
```

```
##
## Call:
## glm(formula = Audi ~ mileage + tax + mpg + fuelType + transmission +
##     engineSize, family = binomial(link = logit), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2952  -0.7252  -0.6206  -0.4523   2.2240
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  1.777e+00  3.960e-01   4.488 7.20e-06 ***
## mileage                      1.065e-05  1.733e-06   6.146 7.95e-10 ***
## tax                         -2.337e-03  7.679e-04  -3.043 0.002342 **
## mpg                         -4.966e-02  5.454e-03  -9.105  < 2e-16 ***
## fuelTypef.Fuel-Petrol       -2.492e-01  1.049e-01  -2.375 0.017546 *
## fuelTypef.Fuel-Hybrid       -1.333e+01  1.846e+02  -0.072 0.942448
## transmissionf.Trans-SemiAuto -3.193e-01  9.632e-02  -3.315 0.000918 ***
## transmissionf.Trans-Automatic -3.091e-01  1.054e-01  -2.931 0.003377 **
## engineSizeMitjà             -2.403e-01  1.052e-01  -2.284 0.022344 *
## engineSizeGran              -4.030e-01  1.633e-01  -2.468 0.013596 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5041.6  on 4961  degrees of freedom
## Residual deviance: 4874.1  on 4952  degrees of freedom
## AIC: 4894.1
##
## Number of Fisher Scoring iterations: 15
```

We can see that the numeric variables continue to have a great impact on the model. Mileage is at the limit but we prefer to keep it instead of years_sell2 because gives our modal a better shape. We will not keep the factor fuelType because it does not have a good correlation with the target variable.

5.3.2 Model 5: Audi~mileage+tax+mpg+transmission+engineSize

```
bm5<-glm(Audi~mileage+tax+mpg+transmission+engineSize,family="binomial"(link = logit),data=df);
summary(bm5)
```

```
##
## Call:
## glm(formula = Audi ~ mileage + tax + mpg + transmission + engineSize,
##     family = binomial(link = logit), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3016  -0.7246  -0.6211  -0.4574   2.2309
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   1.191e+00  2.885e-01   4.126 3.69e-05 ***
## mileage                       1.060e-05  1.731e-06   6.123 9.21e-10 ***
## tax                          -2.097e-03  7.520e-04  -2.789 0.005291 **
## mpg                          -4.300e-02  4.371e-03  -9.837  < 2e-16 ***
## transmissionf.Trans-SemiAuto  -3.279e-01  9.623e-02  -3.408 0.000655 ***
## transmissionf.Trans-Automatic -3.072e-01  1.050e-01  -2.926 0.003436 **
## engineSizeMitjà              -9.568e-02  8.721e-02  -1.097 0.272565
## engineSizeGran               -2.089e-01  1.416e-01  -1.476 0.140059
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5041.6  on 4961  degrees of freedom
## Residual deviance: 4890.5  on 4954  degrees of freedom
## AIC: 4906.5
##
## Number of Fisher Scoring iterations: 5
```

```
AIC(bm4,bm5)
```

```
##     df      AIC
## bm4 10 4894.060
## bm5  8 4906.468
```

```
anova(bm4,bm5)
```

```
## Analysis of Deviance Table
##
## Model 1: Audi ~ mileage + tax + mpg + fuelType + transmission + engineSize
## Model 2: Audi ~ mileage + tax + mpg + transmission + engineSize
##   Resid. Df Resid. Dev Df Deviance
## 1      4952     4874.1
## 2      4954     4890.5 -2  -16.407
```

We will choose the model 5 as the good one using covariates and factors.

5.4 Binary model: Adding interactions

5.4.1 Model 6

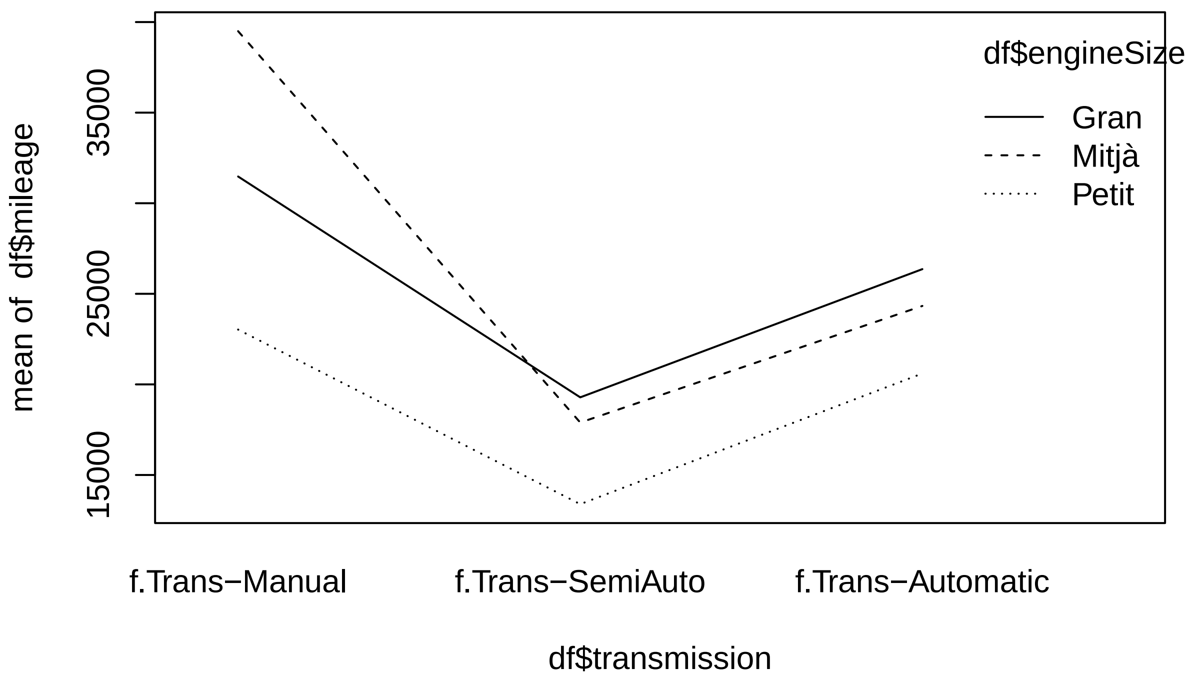
We will search for all the interactions between covariates and factors and between factors.

```
bm6<-glm(Audi~(mileage+tax+mpg+transmission+engineSize)*(transmission+engineSize),family="binomial"(link
# summary and step: Annex - A4
```

The final model obtained executing the functionstep is the next one in wich we can see interactions between transmission and engine size and beteween some other covariates: Audi ~ mileage + tax + mpg + transmission + engineSize + mileage:transmission + mileage:engineSize + tax:transmission + mpg:transmission + transmission:engineSize

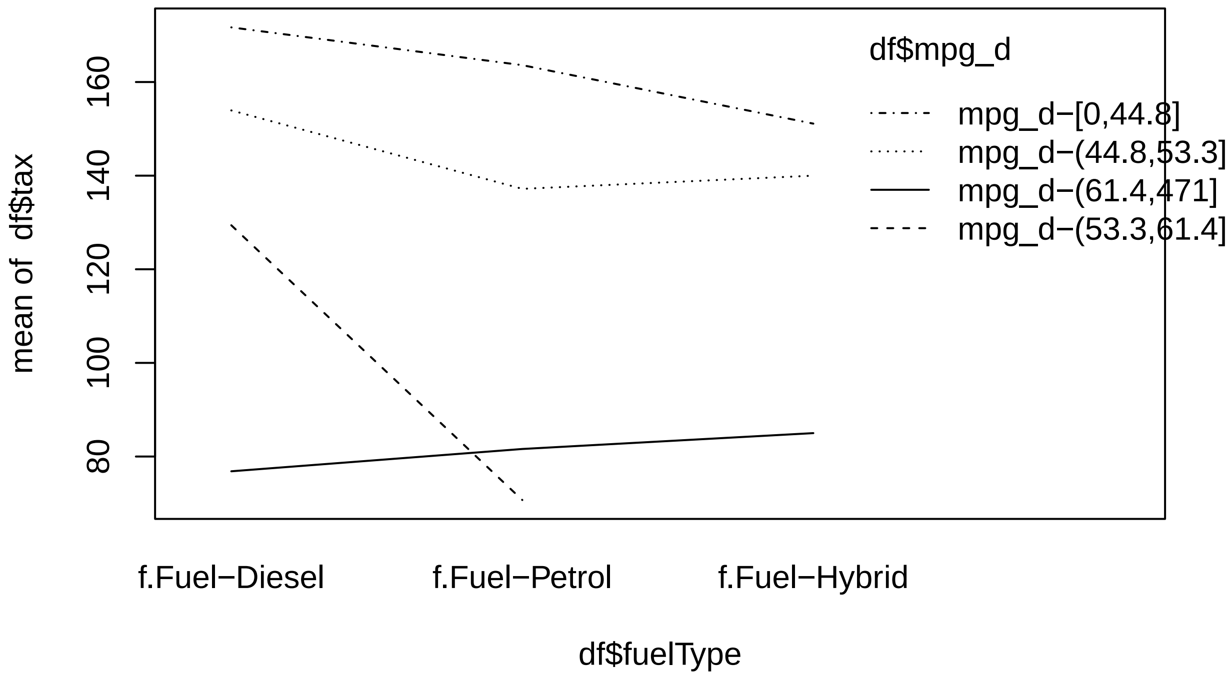
We can see that there is a high correlation between: -mileage:transmissionf.Trans-SemiAuto -mpg:transmissionf.Trans-SemiAuto -transmission:engineSize

```
interaction.plot(df$transmission,df$engineSize,df$mileage)
```



We can see that SemiAuto cars are the ones with less mileage, automatic cars are the second one and finally, manual cars are the ones that have run more kilometers.

```
interaction.plot(df$fuelType,df$mpg_d,df$tax)
```



We won't care about the hybrid cars because they represent only a small preoportion of cars

```
ll<-which(df$fuelType=="f.Fuel-Hybrid");length(ll)
```

```
## [1] 83
```

```
a<-length(ll)/nrow(df)
```

We can see that tax value decreases with the fueltype (Diesel to Petrol) for the most common cars (mpg_D between 0 and 53.3).

5.5 Influent data and outliers

```
#model with all data
bm8<-glm(Audi~(mileage+tax+mpg+transmission+engineSize)*(transmission+engineSize),family="binomial"(link
p <- length(bm8$coefficients)
n <- length(bm8$fitted.values)
llres <- which(abs(rstudent(bm8))>2.3);
llhat <- which(hatvalues(bm8)>(3*(p/n)));
llout<-which(abs(cooks.distance(bm8))>0.02);
llrem<-unique(c(llout,llres));llrem

#model without outliers and hish rstudent valies
dfaux = df[df$mout=="MvOut.No",]
bm9<-glm(Audi~(mileage+tax+mpg+transmission+engineSize)*(transmission+engineSize),family="binomial"(link
```

```
vif(bm9)
```

```
##                               GVIF Df GVIF^(1/(2*Df))
## mileage               4.971042e+00  1        2.229583
## tax                   4.589436e+00  1        2.142297
## mpg                   5.795611e+00  1        2.407408
```

```
## transmission           8.488963e+03  2        9.598728
## engineSize             5.817796e+08  2      155.306496
## mileage:transmission   1.335620e+01  2        1.911704
## mileage:engineSize     4.083181e+01  2        2.527840
## tax:transmission       3.131327e+02  2        4.206608
## tax:engineSize         3.798742e+02  2        4.414789
## mpg:transmission       2.669013e+03  2        7.187662
## mpg:engineSize         2.959291e+03  2        7.375593
## transmission:engineSize 2.138453e+08  4       10.996702
```

```
summary(bm9)
```

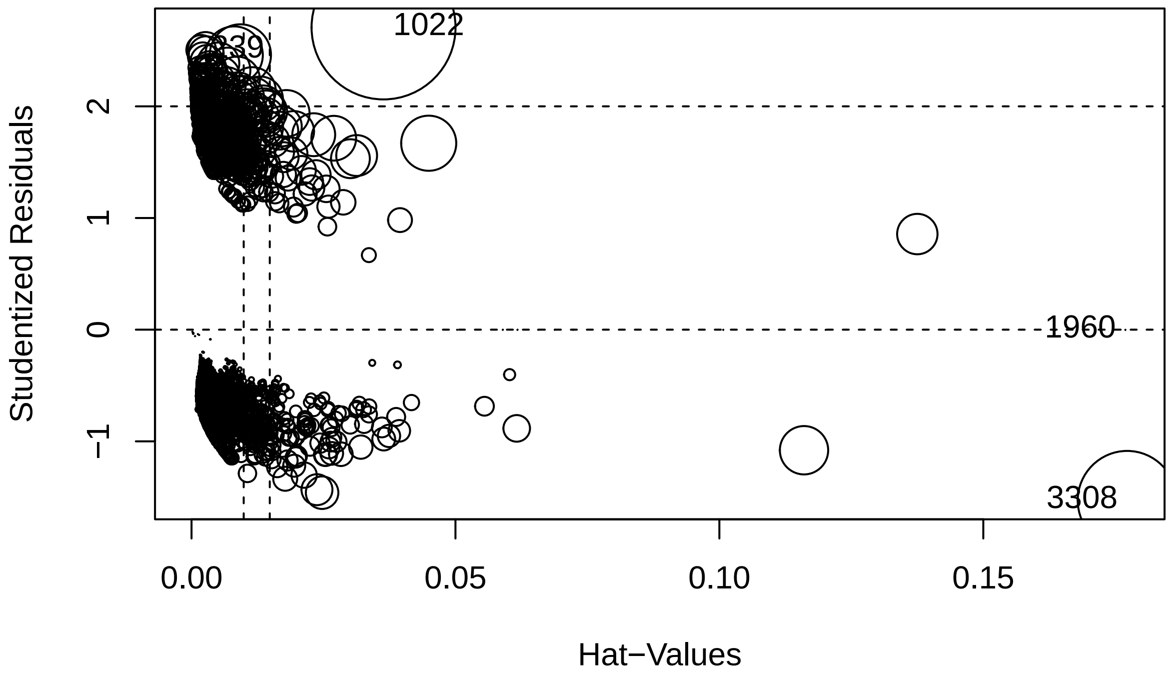
```
##
## Call:
## glm(formula = Audi ~ (mileage + tax + mpg + transmission + engineSize) *
##     (transmission + engineSize), family = binomial(link = logit),
##     data = dfaux[-llrem, ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4437  -0.7186  -0.6203  -0.3599   2.5351
##
## Coefficients:
##                                               Estimate Std. Error z value
## (Intercept)                                 -7.963e-01  5.066e-01  -1.572
## mileage                                      5.678e-06  3.785e-06   1.500
## tax                                         -5.766e-04  1.370e-03  -0.421
## mpg                                         -9.185e-03  7.764e-03  -1.183
## transmissionf.Trans-SemiAuto                 1.805e+00  8.573e-01   2.106
## transmissionf.Trans-Automatic                3.002e+00  9.755e-01   3.078
## engineSizeMitjà                              9.442e-01  8.363e-01   1.129
## engineSizeGran                              -1.339e+01  2.589e+02  -0.052
## mileage:transmissionf.Trans-SemiAuto         2.060e-05  5.324e-06   3.869
## mileage:transmissionf.Trans-Automatic       -1.012e-05  5.559e-06  -1.821
## mileage:engineSizeMitjà                      4.283e-06  4.843e-06   0.884
## mileage:engineSizeGran                       1.639e-05  9.391e-06   1.745
## tax:transmissionf.Trans-SemiAuto             2.577e-03  2.380e-03   1.083
## tax:transmissionf.Trans-Automatic           -2.808e-03  2.559e-03  -1.097
## tax:engineSizeMitjà                         -2.097e-03  2.089e-03  -1.004
## tax:engineSizeGran                          -4.024e-03  3.352e-03  -1.201
## mpg:transmissionf.Trans-SemiAuto            -4.808e-02  1.289e-02  -3.728
## mpg:transmissionf.Trans-Automatic           -5.537e-02  1.491e-02  -3.713
## mpg:engineSizeMitjà                         -1.387e-02  1.215e-02  -1.142
## mpg:engineSizeGran                           1.115e-02  2.021e-02   0.552
## transmissionf.Trans-SemiAuto:engineSizeMitjà  -5.945e-01  2.197e-01  -2.706
## transmissionf.Trans-Automatic:engineSizeMitjà  1.225e-01  2.665e-01   0.460
## transmissionf.Trans-SemiAuto:engineSizeGran    1.262e+01  2.589e+02   0.049
## transmissionf.Trans-Automatic:engineSizeGran   1.309e+01  2.589e+02   0.051
##                                             Pr(>|z|)
## (Intercept)                                 0.116008
## mileage                                     0.133613
## tax                                         0.673863
## mpg                                         0.236835
## transmissionf.Trans-SemiAuto                0.035244 *
## transmissionf.Trans-Automatic               0.002086 **
## engineSizeMitjà                             0.258874
## engineSizeGran                              0.958757
## mileage:transmissionf.Trans-SemiAuto        0.000109 ***
## mileage:transmissionf.Trans-Automatic       0.068641 .
## mileage:engineSizeMitjà                     0.376431
## mileage:engineSizeGran                      0.081014 .
## tax:transmissionf.Trans-SemiAuto            0.278911
## tax:transmissionf.Trans-Automatic           0.272496
```

```
## tax:engineSizeMitjà                           0.315398
## tax:engineSizeGran                            0.229869
## mpg:transmissionf.Trans-SemiAuto              0.000193 ***
## mpg:transmissionf.Trans-Automatic             0.000205 ***
## mpg:engineSizeMitjà                           0.253529
## mpg:engineSizeGran                            0.581178
## transmissionf.Trans-SemiAuto:engineSizeMitjà  0.006815 **
## transmissionf.Trans-Automatic:engineSizeMitjà 0.645788
## transmissionf.Trans-SemiAuto:engineSizeGran   0.961119
## transmissionf.Trans-Automatic:engineSizeGran  0.959673
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4912.5  on 4857  degrees of freedom
## Residual deviance: 4663.6  on 4834  degrees of freedom
## AIC: 4711.6
##
## Number of Fisher Scoring iterations: 13
```

```
Anova(bm9)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Audi
##                          LR Chisq Df Pr(>Chisq)
## mileage                    32.819  1  1.011e-08 ***
## tax                         4.798  1  0.0284856 *
## mpg                       108.247  1  < 2.2e-16 ***
## transmission               13.334  2  0.0012724 **
## engineSize                  8.441  2  0.0146947 *
## mileage:transmission       30.468  2  2.421e-07 ***
## mileage:engineSize          3.071  2  0.2153185
## tax:transmission            5.044  2  0.0802940 .
## tax:engineSize              1.705  2  0.4263829
## mpg:transmission           18.122  2  0.0001161 ***
## mpg:engineSize              2.806  2  0.2458338
## transmission:engineSize    14.521  4  0.0058040 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

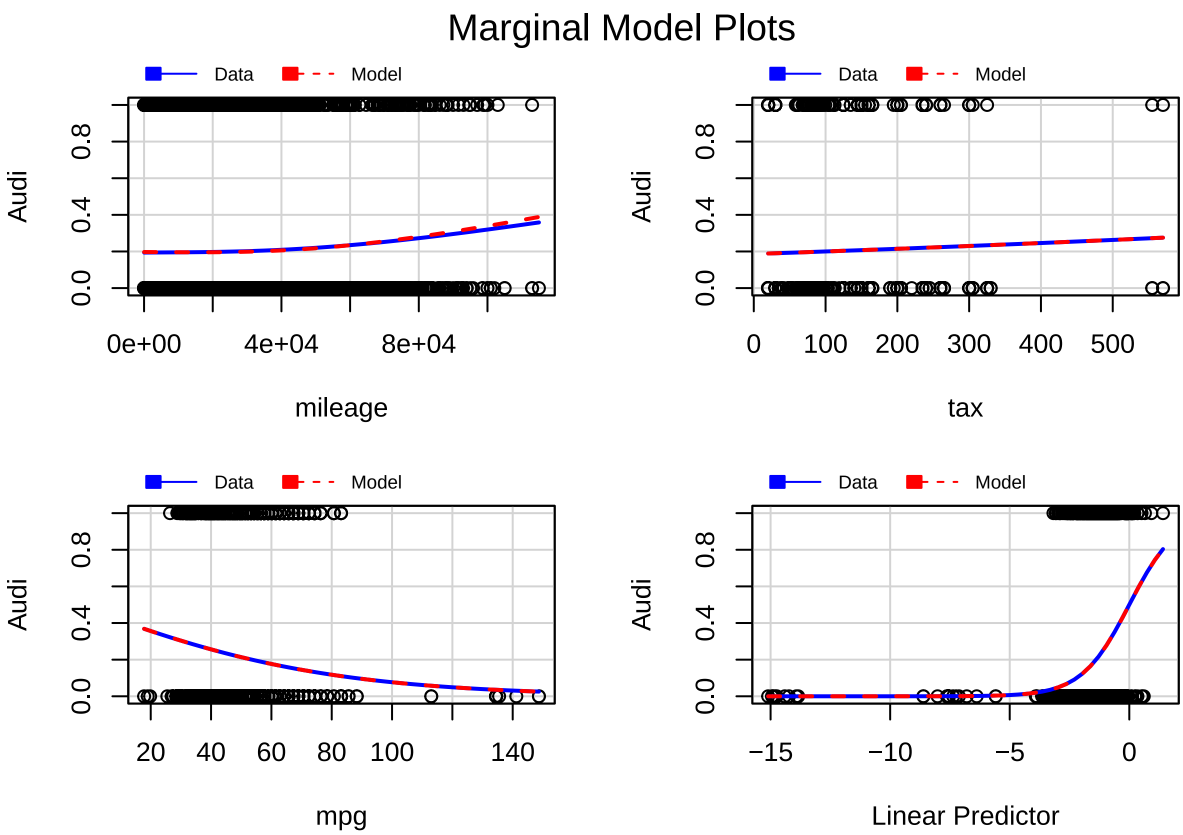
```
influencePlot(bm9)
```




```
##            StudRes         Hat       CookD
## 239   2.505561851 0.001836134 1.660482e-03
## 1022  2.706910052 0.036374351 3.894636e-02
## 1960 -0.001466548 0.176926834 1.056653e-08
## 3308 -1.528683257 0.177259345 1.836806e-02
```

```
marginalModelPlots(bm9)
```

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




```
outlierTest(bm9)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##       rstudent unadjusted p-value Bonferroni p
## 1022  2.70691          0.0067913           NA
```

One we have included interactions in the model we have proceed to remove all outliers and most influent data to imporve the results of the predictor output.

5.6 Confusion table analysis

```
bm7<-glm(Audi~(mileage+tax+mpg+transmission+engineSize)*(transmission+engineSize),family="binomial"(link

library(ResourceSelection)
pred_test <- predict(bm7, newdata=df_test, type="response")
ht <- hoslem.test(df_test$Audi, pred_test)
cbind(ht$observed, ht$expected)
# ROC Curve

library("ROCR")
library("AUC")

#dadesroc<-prediction(pred_test,df_test$Audi)
#performance(dadesroc,"auc",fpr.stop=0.05)
#par(mfrow=c(1,2))
#plot(performance(dadesroc,"err"))
#par(mfrow=c(1,1))
#plot(performance(dadesroc,"tpr","fpr"))
#abline(0,1,lty=2)

library(cvAUC)
AUC(pred_test,df_test$Audi)
```

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

```
##
## audi.est  No Yes
##        0 794 192
##        1   2   4
```

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

```
## [1] 80.44355
```

```r
100*(tt[2,2]/(tt[2,1]+ tt[2,2])) # precision
```

```
## [1] 66.66667
```

```r
prob.audi <- bm7$fit
audi.est <- ifelse(prob.audi<0.5,0,1)
tt<-table(audi.est,df$Audi);tt
```

```
##
## audi.est   No  Yes
##        0 3933  998
##        1    9   22
```

```r
100*tt[1,1]/sum(tt)
```

```
## [1] 79.26239
```

```r
100*(tt[2,2]/(tt[2,1]+ tt[2,2])) # precision
```

```
## [1] 70.96774
```

After applying our selected model with the test data, we can see the resultant confusion matrix. We can see that the model has an accuracy of 80%. This means that the 80% of the data predicted is correct. We can see that the model has nearly a 70% of precision. 70% of the times that a car is an Audi the model predicts it correctly. The model loses precision when predicting positives because there are much more non audi cars than audi cars. This means that the model is better predicting non adui cars than audi cars.

6 Annex

6.1 A1

```r
summary(m4aux)
```

```
##
## Call:
## lm(formula = log(price) ~ tax + mileage + years_sell2 + model +
##     engineSize + manufacturer + transmission, data = df[!df$mout ==
##     "MvOut.Yes", ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.01188 -0.09268  0.00000  0.09381  0.76374
```

```
##
## Coefficients: (3 not defined because of singularities)
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                  1.010e+01  1.769e-02 570.822  < 2e-16 ***
## tax                         -1.436e-04  5.109e-05  -2.811 0.004963 ** 
## mileage                     -9.637e-06  1.755e-07 -54.927  < 2e-16 ***
## years_sell2                 -2.070e-01  6.537e-03 -31.657  < 2e-16 ***
## modelAudi- A3                1.394e-01  1.811e-02   7.697 1.68e-14 ***
## modelAudi- A4                1.561e-01  2.105e-02   7.416 1.42e-13 ***
## modelAudi- A5                2.438e-01  2.359e-02  10.334  < 2e-16 ***
## modelAudi- A6                2.296e-01  2.530e-02   9.074  < 2e-16 ***
## modelAudi- A7                2.152e-01  5.509e-02   3.906 9.52e-05 ***
## modelAudi- A8                3.436e-01  6.099e-02   5.634 1.86e-08 ***
## modelAudi- Q2                1.900e-01  2.281e-02   8.330  < 2e-16 ***
## modelAudi- Q3                2.832e-01  2.063e-02  13.730  < 2e-16 ***
## modelAudi- Q5                4.098e-01  2.277e-02  17.997  < 2e-16 ***
## modelAudi- Q7                5.433e-01  3.325e-02  16.342  < 2e-16 ***
## modelAudi- Q8                6.601e-01  9.689e-02   6.813 1.08e-11 ***
## modelAudi- RS3               5.219e-01  1.174e-01   4.445 8.99e-06 ***
## modelAudi- RS4               7.263e-01  1.671e-01   4.346 1.41e-05 ***
## modelAudi- RS5               8.768e-01  1.175e-01   7.463 9.99e-14 ***
## modelAudi- RS6               9.201e-01  1.658e-01   5.548 3.04e-08 ***
## modelAudi- S3                4.836e-01  1.654e-01   2.924 0.003467 ** 
## modelAudi- S4                3.608e-01  1.657e-01   2.178 0.029453 *  
## modelAudi- S5                3.312e-02  1.657e-01   0.200 0.841589    
## modelAudi- S8                5.110e-01  1.658e-01   3.081 0.002071 ** 
## modelAudi- SQ5               5.501e-01  8.438e-02   6.518 7.84e-11 ***
## modelAudi- TT                3.455e-01  3.476e-02   9.939  < 2e-16 ***
## modelBMW- 1 Series           1.332e-03  1.832e-02   0.073 0.942039    
## modelBMW- 2 Series           2.397e-02  1.992e-02   1.204 0.228820    
## modelBMW- 3 Series           7.300e-02  1.843e-02   3.962 7.53e-05 ***
## modelBMW- 4 Series           1.030e-01  2.248e-02   4.583 4.69e-06 ***
## modelBMW- 5 Series           1.738e-01  2.209e-02   7.867 4.47e-15 ***
## modelBMW- 6 Series           2.129e-01  4.859e-02   4.382 1.20e-05 ***
## modelBMW- 7 Series           4.291e-01  5.511e-02   7.786 8.43e-15 ***
## modelBMW- M2                 2.068e-01  9.680e-02   2.137 0.032655 *  
## modelBMW- M3                 5.236e-01  9.684e-02   5.407 6.72e-08 ***
## modelBMW- M4                 4.613e-01  5.090e-02   9.062  < 2e-16 ***
## modelBMW- X1                 1.493e-01  2.271e-02   6.572 5.49e-11 ***
## modelBMW- X2                 1.487e-01  3.570e-02   4.164 3.18e-05 ***
## modelBMW- X3                 3.378e-01  2.774e-02  12.179  < 2e-16 ***
## modelBMW- X4                 3.799e-01  4.095e-02   9.277  < 2e-16 ***
## modelBMW- X5                 4.811e-01  3.205e-02  15.014  < 2e-16 ***
## modelBMW- X6                 4.860e-01  7.584e-02   6.409 1.60e-10 ***
## modelBMW- Z3                -5.825e-01  1.178e-01  -4.944 7.93e-07 ***
## modelBMW- Z4                 1.051e-01  5.001e-02   2.102 0.035631 *  
## modelMercedes- A Class       1.161e-01  1.703e-02   6.818 1.03e-11 ***
## modelMercedes- B Class       5.551e-02  2.592e-02   2.141 0.032306 *  
## modelMercedes- C Class       1.996e-01  1.693e-02  11.790  < 2e-16 ***
## modelMercedes- CL Class      2.662e-01  3.176e-02   8.383  < 2e-16 ***
## modelMercedes- CLA Class     2.837e-01  5.415e-02   5.239 1.68e-07 ***
## modelMercedes- CLK          -6.934e-01  1.186e-01  -5.848 5.30e-09 ***
## modelMercedes- CLS Class     2.426e-01  4.542e-02   5.342 9.63e-08 ***
## modelMercedes- E Class       2.240e-01  1.931e-02  11.599  < 2e-16 ***
## modelMercedes- GL Class      2.253e-01  5.199e-02   4.333 1.50e-05 ***
## modelMercedes- GLA Class     1.477e-01  2.478e-02   5.959 2.72e-09 ***
## modelMercedes- GLC Class     4.481e-01  2.162e-02  20.722  < 2e-16 ***
## modelMercedes- GLE Class     5.796e-01  2.930e-02  19.783  < 2e-16 ***
## modelMercedes- GLS Class     5.795e-01  8.429e-02   6.875 7.00e-12 ***
## modelMercedes- M Class       3.057e-01  6.114e-02   5.000 5.93e-07 ***
## modelMercedes- S Class       4.188e-01  5.074e-02   8.253  < 2e-16 ***
## modelMercedes- SL CLASS      2.852e-01  4.499e-02   6.339 2.52e-10 ***
## modelMercedes- SLK          -1.825e-01  6.411e-02  -2.846 0.004448 ** 
## modelMercedes- V Class       3.596e-01  3.524e-02  10.205  < 2e-16 ***
```

```
## modelMercedes- X-CLASS         2.443e-01  5.253e-02   4.651 3.39e-06 ***
## modelVW- Amarok                1.488e-01  5.315e-02   2.800 0.005136 **
## modelVW- Arteon                1.434e-01  3.976e-02   3.606 0.000314 ***
## modelVW- Beetle               -5.953e-01  5.673e-02 -10.493  < 2e-16 ***
## modelVW- Caddy Maxi Life      -1.039e-02  8.369e-02  -0.124 0.901178
## modelVW- California            9.132e-01  1.176e-01   7.764 9.97e-15 ***
## modelVW- Caravelle             5.277e-01  5.721e-02   9.225  < 2e-16 ***
## modelVW- CC                   -1.111e-01  5.688e-02  -1.953 0.050849 .
## modelVW- Eos                  -4.114e-01  1.654e-01  -2.487 0.012930 *
## modelVW- Golf                  2.862e-02  1.560e-02   1.834 0.066649 .
## modelVW- Golf SV              -1.033e-01  3.496e-02  -2.955 0.003139 **
## modelVW- Jetta                -6.058e-01  1.173e-01  -5.166 2.49e-07 ***
## modelVW- Passat                4.668e-02  2.179e-02   2.142 0.032271 *
## modelVW- Polo                 -2.764e-01  1.595e-02 -17.337  < 2e-16 ***
## modelVW- Scirocco              1.059e-02  3.430e-02   0.309 0.757604
## modelVW- Sharan                1.044e-01  3.738e-02   2.793 0.005248 **
## modelVW- Shuttle               2.081e-01  6.894e-02   3.018 0.002560 **
## modelVW- T-Cross               2.200e-02  3.113e-02   0.707 0.479850
## modelVW- T-Roc                 1.458e-01  2.449e-02   5.956 2.77e-09 ***
## modelVW- Tiguan                1.724e-01  1.915e-02   9.000  < 2e-16 ***
## modelVW- Tiguan Allspace       2.471e-01  5.408e-02   4.569 5.02e-06 ***
## modelVW- Touareg               3.427e-01  3.516e-02   9.747  < 2e-16 ***
## modelVW- Touran                1.451e-01  2.962e-02   4.899 9.97e-07 ***
## modelVW- Up                   -5.966e-01  2.190e-02 -27.243  < 2e-16 ***
## engineSizeMitjà                1.557e-01  7.193e-03  21.650  < 2e-16 ***
## engineSizeGran                 3.451e-01  1.262e-02  27.335  < 2e-16 ***
## manufacturerf.Man-BMW                 NA         NA      NA       NA
## manufacturerf.Man-Mercedes            NA         NA      NA       NA
## manufacturerf.Man-VW                  NA         NA      NA       NA
## transmissionf.Trans-SemiAuto   1.324e-01  6.889e-03  19.224  < 2e-16 ***
## transmissionf.Trans-Automatic  9.861e-02  7.544e-03  13.071  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1646 on 4787 degrees of freedom
## Multiple R-squared:  0.8775, Adjusted R-squared:  0.8752
## F-statistic: 389.5 on 88 and 4787 DF,  p-value: < 2.2e-16
```

6.2 A2

```
m5<-step( m5, k=log(nrow(df)))
```

```
## Start:  AIC=-15470.21
## log(price) ~ (tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType) * (engineSize + model + transmission +
##     fuelType)
##
##                             Df Sum of Sq    RSS    AIC
## - model:transmission       100    2.5043 88.601 -16179
## - mileage:model             69    3.7044 89.801 -15848
## - tax:model                 62    3.5758 89.672 -15796
## - years_sell2:model         52    2.3376 88.434 -15780
## - model:fuelType            57    5.2241 91.320 -15663
## - engineSize:model          56    7.1195 93.216 -15552
## - engineSize:transmission    4    0.2773 86.374 -15488
## - transmission:fuelType      3    0.1465 86.243 -15487
## - mileage:engineSize         2    0.0122 86.109 -15486
## - years_sell2:transmission   2    0.0160 86.112 -15486
## - years_sell2:engineSize     2    0.0533 86.150 -15484
## - years_sell2:fuelType       2    0.0752 86.172 -15483
## - tax:engineSize             2    0.0759 86.172 -15483
```

```
## - tax:transmission           2    0.2343 86.331 -15474
## - mileage:transmission       2    0.2357 86.332 -15474
## - tax:fuelType               2    0.2564 86.353 -15472
## <none>                                  86.096 -15470
## - engineSize:fuelType        3    0.9819 87.078 -15440
## - mileage:fuelType           2    1.3806 87.477 -15408
##
## Step:  AIC=-16178.89
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:engineSize + tax:model + tax:transmission +
##     tax:fuelType + mileage:engineSize + mileage:model + mileage:transmission +
##     mileage:fuelType + years_sell2:engineSize + years_sell2:model +
##     years_sell2:transmission + years_sell2:fuelType + engineSize:model +
##     engineSize:transmission + engineSize:fuelType + model:fuelType +
##     transmission:fuelType
##
##                           Df Sum of Sq    RSS    AIC
## - mileage:model           72    5.2586 93.859 -16506
## - years_sell2:model       55    2.7456 91.346 -16496
## - tax:model               65    4.9534 93.554 -16462
## - model:fuelType          57    5.9024 94.503 -16344
## - engineSize:model        56    8.0289 96.630 -16225
## - engineSize:transmission  4    0.1811 88.782 -16203
## - transmission:fuelType    3    0.1260 88.727 -16197
## - years_sell2:transmission  2    0.0141 88.615 -16195
## - mileage:engineSize        2    0.0238 88.624 -16195
## - tax:engineSize            2    0.0657 88.666 -16192
## - years_sell2:engineSize    2    0.0709 88.672 -16192
## - years_sell2:fuelType      2    0.1018 88.702 -16190
## - tax:transmission          2    0.1607 88.761 -16187
## - tax:fuelType              2    0.2222 88.823 -16184
## - mileage:transmission      2    0.2851 88.886 -16180
## <none>                                 88.601 -16179
## - engineSize:fuelType       3    1.0926 89.693 -16144
## - mileage:fuelType          2    1.4348 90.035 -16116
##
## Step:  AIC=-16505.48
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:engineSize + tax:model + tax:transmission +
##     tax:fuelType + mileage:engineSize + mileage:transmission +
##     mileage:fuelType + years_sell2:engineSize + years_sell2:model +
##     years_sell2:transmission + years_sell2:fuelType + engineSize:model +
##     engineSize:transmission + engineSize:fuelType + model:fuelType +
##     transmission:fuelType
##
##                           Df Sum of Sq     RSS    AIC
## - tax:model               68    5.8801  99.739 -16783
## - years_sell2:model       57    5.5008  99.360 -16708
## - model:fuelType          58    5.8713  99.731 -16698
## - engineSize:model        57    8.2179 102.077 -16574
## - engineSize:transmission  4    0.1351  93.994 -16532
## - transmission:fuelType    3    0.1374  93.997 -16524
## - years_sell2:transmission  2    0.0049  93.864 -16522
## - tax:engineSize            2    0.0539  93.913 -16520
## - tax:transmission          2    0.0871  93.946 -16518
## - mileage:engineSize        2    0.1300  93.989 -16516
## - years_sell2:engineSize    2    0.1469  94.006 -16515
## - years_sell2:fuelType      2    0.1733  94.033 -16513
## - tax:fuelType              2    0.2410  94.100 -16510
## <none>                                  93.859 -16506
## - mileage:transmission      2    0.3435  94.203 -16504
## - engineSize:fuelType       3    1.1666  95.026 -16470
## - mileage:fuelType          2    1.6444  95.504 -16436
##
```

```
## Step:  AIC=-16782.62
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:engineSize + tax:transmission +
##     tax:fuelType + mileage:engineSize + mileage:transmission +
##     mileage:fuelType + years_sell2:engineSize + years_sell2:model +
##     years_sell2:transmission + years_sell2:fuelType + engineSize:model +
##     engineSize:transmission + engineSize:fuelType + model:fuelType +
##     transmission:fuelType
##
##                           Df Sum of Sq     RSS    AIC
## - years_sell2:model        59    5.3304 105.070 -17026
## - engineSize:transmission   4    0.2460  99.985 -16804
## - transmission:fuelType     3    0.1400  99.879 -16801
## - years_sell2:transmission  2    0.0156  99.755 -16799
## - tax:fuelType              2    0.1499  99.889 -16792
## - years_sell2:engineSize    2    0.1651  99.905 -16791
## - mileage:engineSize        2    0.1687  99.908 -16791
## - tax:engineSize            2    0.2939 100.033 -16785
## - tax:transmission          2    0.3068 100.046 -16784
## - years_sell2:fuelType      2    0.3227 100.062 -16784
## <none>                                  99.739 -16783
## - mileage:transmission      2    0.5312 100.271 -16773
## - engineSize:model         59   11.4221 111.162 -16747
## - model:fuelType           58   11.3652 111.105 -16741
## - engineSize:fuelType       3    1.6855 101.425 -16725
## - mileage:fuelType          2    1.9877 101.727 -16702
##
## Step:  AIC=-17026.34
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:engineSize + tax:transmission +
##     tax:fuelType + mileage:engineSize + mileage:transmission +
##     mileage:fuelType + years_sell2:engineSize + years_sell2:transmission +
##     years_sell2:fuelType + engineSize:model + engineSize:transmission +
##     engineSize:fuelType + model:fuelType + transmission:fuelType
##
##                           Df Sum of Sq    RSS    AIC
## - engineSize:transmission   4    0.2668 105.34 -17048
## - transmission:fuelType     3    0.1451 105.22 -17045
## - years_sell2:engineSize    2    0.0134 105.08 -17043
## - years_sell2:transmission  2    0.0136 105.08 -17043
## - model:fuelType           58   10.7918 115.86 -17035
## - tax:fuelType              2    0.1990 105.27 -17034
## - mileage:engineSize        2    0.2267 105.30 -17033
## - tax:engineSize            2    0.2552 105.33 -17031
## - tax:transmission          2    0.3555 105.42 -17027
## <none>                                 105.07 -17026
## - years_sell2:fuelType      2    0.4548 105.53 -17022
## - mileage:transmission      2    0.5814 105.65 -17016
## - engineSize:fuelType       3    1.5211 106.59 -16981
## - engineSize:model         61   12.7004 117.77 -16979
## - mileage:fuelType          2    2.4876 107.56 -16927
##
## Step:  AIC=-17047.8
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:engineSize + tax:transmission +
##     tax:fuelType + mileage:engineSize + mileage:transmission +
##     mileage:fuelType + years_sell2:engineSize + years_sell2:transmission +
##     years_sell2:fuelType + engineSize:model + engineSize:fuelType +
##     model:fuelType + transmission:fuelType
##
##                           Df Sum of Sq    RSS    AIC
## - years_sell2:engineSize    2    0.0082 105.34 -17064
## - years_sell2:transmission  2    0.0198 105.36 -17064
## - transmission:fuelType     3    0.2011 105.54 -17064
```

```
## - tax:fuelType              2    0.1947 105.53 -17056
## - model:fuelType           58   10.8415 116.18 -17055
## - tax:engineSize            2    0.2469 105.58 -17053
## - mileage:engineSize        2    0.2496 105.59 -17053
## - tax:transmission          2    0.3311 105.67 -17049
## <none>                                 105.34 -17048
## - years_sell2:fuelType      2    0.3680 105.70 -17048
## - mileage:transmission      2    0.5888 105.92 -17037
## - engineSize:model         61   12.5175 117.85 -17010
## - engineSize:fuelType       3    1.9800 107.32 -16981
## - mileage:fuelType          2    2.5143 107.85 -16948
##
## Step:  AIC=-17064.43
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:engineSize + tax:transmission +
##     tax:fuelType + mileage:engineSize + mileage:transmission +
##     mileage:fuelType + years_sell2:transmission + years_sell2:fuelType +
##     engineSize:model + engineSize:fuelType + model:fuelType +
##     transmission:fuelType
##
##                           Df Sum of Sq    RSS    AIC
## - years_sell2:transmission  2    0.0175 105.36 -17081
## - transmission:fuelType     3    0.1997 105.55 -17081
## - tax:fuelType              2    0.1976 105.54 -17072
## - model:fuelType           58   10.8408 116.19 -17072
## - tax:engineSize            2    0.2637 105.61 -17069
## - tax:transmission          2    0.3251 105.67 -17066
## <none>                                 105.34 -17064
## - mileage:engineSize        2    0.4395 105.78 -17061
## - years_sell2:fuelType      2    0.4922 105.84 -17058
## - mileage:transmission      2    0.5855 105.93 -17054
## - engineSize:model         61   12.5787 117.92 -17024
## - engineSize:fuelType       3    2.0203 107.36 -16996
## - mileage:fuelType          2    2.6962 108.04 -16956
##
## Step:  AIC=-17080.63
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:engineSize + tax:transmission +
##     tax:fuelType + mileage:engineSize + mileage:transmission +
##     mileage:fuelType + years_sell2:fuelType + engineSize:model +
##     engineSize:fuelType + model:fuelType + transmission:fuelType
##
##                        Df Sum of Sq    RSS    AIC
## - transmission:fuelType  3    0.1977 105.56 -17097
## - model:fuelType        58   10.8367 116.20 -17088
## - tax:fuelType           2    0.2009 105.56 -17088
## - tax:engineSize         2    0.2623 105.62 -17085
## - tax:transmission       2    0.3612 105.72 -17081
## <none>                              105.36 -17081
## - mileage:engineSize     2    0.4393 105.80 -17077
## - years_sell2:fuelType   2    0.5008 105.86 -17074
## - mileage:transmission   2    0.9472 106.31 -17053
## - engineSize:model      61   12.5837 117.95 -17040
## - engineSize:fuelType    3    2.0176 107.38 -17012
## - mileage:fuelType       2    2.7187 108.08 -16971
##
## Step:  AIC=-17096.85
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:engineSize + tax:transmission +
##     tax:fuelType + mileage:engineSize + mileage:transmission +
##     mileage:fuelType + years_sell2:fuelType + engineSize:model +
##     engineSize:fuelType + model:fuelType
##
##                       Df Sum of Sq    RSS    AIC
```

```
## - tax:fuelType          2    0.2236 105.78 -17103
## - tax:engineSize        2    0.2434 105.80 -17102
## - model:fuelType       58   10.9980 116.56 -17099
## <none>                              105.56 -17097
## - tax:transmission      2    0.4432 106.00 -17093
## - mileage:engineSize    2    0.4591 106.02 -17092
## - years_sell2:fuelType  2    0.5092 106.07 -17090
## - mileage:transmission  2    0.8251 106.39 -17075
## - engineSize:model     61   12.6703 118.23 -17054
## - engineSize:fuelType   3    1.9976 107.56 -17029
## - mileage:fuelType      2    2.7735 108.33 -16985
##
## Step:  AIC=-17103.37
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:engineSize + tax:transmission +
##     mileage:engineSize + mileage:transmission + mileage:fuelType +
##     years_sell2:fuelType + engineSize:model + engineSize:fuelType +
##     model:fuelType
##
##                        Df Sum of Sq    RSS    AIC
## - tax:engineSize        2    0.2249 106.01 -17110
## <none>                              105.78 -17103
## - tax:transmission      2    0.3992 106.18 -17102
## - mileage:engineSize    2    0.4168 106.20 -17101
## - model:fuelType       58   11.3094 117.09 -17093
## - years_sell2:fuelType  2    0.6345 106.42 -17091
## - mileage:transmission  2    0.8728 106.66 -17080
## - engineSize:model     61   12.6804 118.46 -17061
## - engineSize:fuelType   3    1.8246 107.61 -17044
## - mileage:fuelType      2    2.8753 108.66 -16987
##
## Step:  AIC=-17109.85
## log(price) ~ tax + mileage + years_sell2 + engineSize + model +
##     transmission + fuelType + tax:transmission + mileage:engineSize +
##     mileage:transmission + mileage:fuelType + years_sell2:fuelType +
##     engineSize:model + engineSize:fuelType + model:fuelType
##
##                        Df Sum of Sq    RSS    AIC
## <none>                              106.01 -17110
## - mileage:engineSize    2    0.5099 106.52 -17103
## - model:fuelType       58   11.2555 117.26 -17103
## - tax:transmission      2    0.5337 106.54 -17102
## - years_sell2:fuelType  2    0.5907 106.60 -17099
## - mileage:transmission  2    0.8981 106.91 -17085
## - engineSize:model     61   12.7174 118.73 -17067
## - engineSize:fuelType   3    1.7313 107.74 -17055
## - mileage:fuelType      2    2.7879 108.80 -16998
```

6.3 A3

```
thhat <-3*length(coef(m6))/nrow(df);
llhat <- which( hatvalues(m6) > thhat);
df[llhat,]
```

```
##                     model            year price      transmission   mileage
## 4                Audi- A6            2018 16600 f.Trans-Automatic 22958.000
## 52               Audi- A8            2018 40990  f.Trans-SemiAuto 15157.084
## 154              Audi- Q7            2016 30790  f.Trans-SemiAuto 14727.000
## 378              Audi- Q2            2017 18989    f.Trans-Manual 28275.000
## 403              Audi- A4            2019 32000 f.Trans-Automatic  4500.000
## 467              Audi- Q7            2020 59990  f.Trans-SemiAuto  6000.000
```

```
## 482              Audi- A1                2019 22500  f.Trans-SemiAuto  3869.000
## 535              Audi- A3                2016 17211    f.Trans-Manual 31776.000
## 538              Audi- A6                2019 36780 f.Trans-Automatic  8231.000
## 572              Audi- Q3                2019 31490 f.Trans-Automatic  7753.000
## 651              Audi- A4                2017 19025 f.Trans-Automatic 11754.000
## 661              Audi- A1                2020 18500    f.Trans-Manual   641.000
## 738              Audi- A5                2015 18400  f.Trans-SemiAuto 31176.000
## 753              Audi- Q3                2017 25500 f.Trans-Automatic 21090.000
## 754              Audi- A3                2014 11650    f.Trans-Manual 22014.000
## 759              Audi- A3                2019 24200    f.Trans-Manual  6081.000
## 849              Audi- Q7                2020 55750 f.Trans-Automatic   875.000
## 889              Audi- Q3                2019 33990 f.Trans-Automatic  7500.000
## 941              Audi- A4                2019 26250 f.Trans-Automatic  8299.000
## 971              Audi- A6                2015 21950 f.Trans-Automatic 43000.000
## 978              Audi- A3                2016 15650    f.Trans-Manual 35437.000
## 979              Audi- Q3                2019 27190    f.Trans-Manual  3555.000
## 1016             Audi- A5                2020 31500 f.Trans-Automatic    11.000
## 1094       BMW- 1 Series                 2015 16314    f.Trans-Manual 17846.000
## 1106       BMW- 2 Series                 2018 16998    f.Trans-Manual  5898.000
## 1148       BMW- 3 Series                 2015 13990 f.Trans-Automatic 37087.000
## 1264       BMW- 3 Series                 2020 41990  f.Trans-SemiAuto   131.090
## 1309       BMW- 5 Series                 2019 35475  f.Trans-SemiAuto    15.000
## 1358       BMW- 4 Series                 2015 17980  f.Trans-SemiAuto 35255.000
## 1447       BMW- 3 Series                 2019 28998  f.Trans-SemiAuto  5568.000
## 1449       BMW- 5 Series                 2019 32780 f.Trans-Automatic  3774.000
## 1455       BMW- 3 Series                 2016 17547 f.Trans-Automatic 13969.000
## 1465       BMW- 1 Series                 2013 10995  f.Trans-SemiAuto 32514.000
## 1469       BMW- 4 Series                 2019 22980  f.Trans-SemiAuto  8672.000
## 1477       BMW- 3 Series                 2019 25480 f.Trans-Automatic  9839.000
## 1511       BMW- 3 Series                 2019 27995  f.Trans-SemiAuto  1501.000
## 1535             BMW- X2                 2019 27950  f.Trans-SemiAuto  7419.000
## 1540       BMW- 3 Series                 2019 30950  f.Trans-SemiAuto  4112.000
## 1542       BMW- 1 Series                 2019 16950  f.Trans-SemiAuto 11137.000
## 1583       BMW- 1 Series                 2017 16444  f.Trans-SemiAuto 20848.000
## 1589       BMW- 2 Series                 2017 13591    f.Trans-Manual 15001.000
## 1616       BMW- 3 Series                 2019 39995 f.Trans-Automatic   999.000
## 1629             BMW- X3                 2019 32950  f.Trans-SemiAuto  4953.000
## 1680       BMW- 4 Series                 2016 17127  f.Trans-SemiAuto 34479.000
## 1734       BMW- 4 Series                 2015 18290  f.Trans-SemiAuto 25000.000
## 1776       BMW- 4 Series                 2015 21149  f.Trans-SemiAuto 29627.000
## 1779       BMW- 3 Series                 2019 24995  f.Trans-SemiAuto  7130.000
## 1812       BMW- 2 Series                 2018 15995    f.Trans-Manual 28857.000
## 1813             BMW- X2                 2019 25950 f.Trans-Automatic  3078.000
## 1815             BMW- X4                 2015 23500 f.Trans-Automatic 31723.000
## 1840       BMW- 1 Series                 2013  9490 f.Trans-Automatic 63000.000
## 1854       BMW- 6 Series                 2019 32750 f.Trans-Automatic  9205.000
## 1989             BMW- M4                 2020 47488 f.Trans-Automatic    11.000
## 2006       BMW- 5 Series                 2017 19499 f.Trans-Automatic 21728.000
## 2011       BMW- 1 Series                 2016 17499 f.Trans-Automatic 26855.000
## 2029             BMW- X4                 2017 24999 f.Trans-Automatic 35351.000
## 2084       BMW- 5 Series                 2018 26790 f.Trans-Automatic 20000.000
## 2099       BMW- 4 Series                 2018 18500 f.Trans-Automatic 21387.000
## 2121       BMW- 5 Series                 2016 18500 f.Trans-Automatic 37933.000
## 2210   Mercedes- C Class                 2019 27000  f.Trans-SemiAuto  6406.000
## 2220   Mercedes- C Class                 2016 17699  f.Trans-SemiAuto 43236.000
## 2327   Mercedes- C Class                 2017 25232 f.Trans-Automatic 15104.000
## 2329   Mercedes- E Class                 2019 28995  f.Trans-SemiAuto 12630.000
## 2354   Mercedes- C Class                 2018 24791  f.Trans-SemiAuto 16052.000
## 2527   Mercedes- C Class                 2017 20990  f.Trans-SemiAuto 26675.000
## 2729   Mercedes- A Class                 2017 20000    f.Trans-Manual 13685.000
## 2754 Mercedes- GLC Class                 2020 49995  f.Trans-SemiAuto  5000.000
## 2824   Mercedes- E Class                 2019 37000  f.Trans-SemiAuto  2837.000
## 2908   Mercedes- E Class                 2018 27579  f.Trans-SemiAuto 13000.000
## 3039  Mercedes- CL Class                 2019 26299 f.Trans-Automatic  4413.000
```

```
## 3057   Mercedes- A Class            2017 14299    f.Trans-Manual 21008.000
## 3072   Mercedes- E Class            2018 21899 f.Trans-Automatic 22985.000
## 3102   Mercedes- C Class            2015 13990 f.Trans-Automatic 29000.000
## 3130   Mercedes- C Class            2017 21498 f.Trans-Automatic 26145.000
## 3140 Mercedes- GLC Class            2016 19970 f.Trans-Automatic 67286.000
## 3149   Mercedes- B Class            2015 10999 f.Trans-Automatic 54349.000
## 3168   Mercedes- C Class            2013  8799 f.Trans-Automatic 57172.000
## 3173   Mercedes- E Class            2015 15991 f.Trans-Automatic 36705.000
## 3192   Mercedes- C Class            2017 17400 f.Trans-Automatic 41677.000
## 3202  Mercedes- SL CLASS            2017 20900 f.Trans-Automatic 26560.000
## 3215   Mercedes- A Class            2017 13499    f.Trans-Manual 22298.000
## 3216  Mercedes- CL Class            2014 13299    f.Trans-Manual 47027.000
## 3224   Mercedes- A Class            2018 18816 f.Trans-Automatic  7855.000
## 3225   Mercedes- A Class            2014 11599    f.Trans-Manual 52598.000
## 3228   Mercedes- A Class            2016 18699 f.Trans-Automatic 13118.000
## 3290 Mercedes- GLC Class            2019 24250 f.Trans-Automatic 21252.000
## 3308   Mercedes- E Class            2013 14995  f.Trans-SemiAuto 55000.000
## 3310   Mercedes- C Class 2015.86700757686  9495 f.Trans-Automatic 39000.000
## 3348 Mercedes- GLC Class            2017 21600 f.Trans-Automatic 54609.000
## 3364   Mercedes- A Class            2016 10500    f.Trans-Manual 62528.000
## 3371   Mercedes- A Class            2016 13300    f.Trans-Manual 33723.000
## 3674            VW- Golf            2019 15446    f.Trans-Manual 11143.000
## 3738            VW- Golf            2019 17298    f.Trans-Manual  8908.000
## 4595          VW- Tiguan            2019 33950    f.Trans-Manual  8000.000
## 4599          VW- Tiguan            2018 23750  f.Trans-SemiAuto 16000.000
## 4602          VW- Tiguan            2017 19495    f.Trans-Manual 36118.000
## 4714              VW- Up            2018  8995  f.Trans-SemiAuto 11000.000
## 4721              VW- Up            2016  7495    f.Trans-Manual 28388.000
## 4729              VW- Up            2014  4091    f.Trans-Manual 78847.000
## 4736              VW- Up            2015  5510    f.Trans-Manual 51190.000
## 4739              VW- Up            2020 11780    f.Trans-Manual  1561.597
## 4743              VW- Up            2020 10790    f.Trans-Manual  1000.000
## 4744              VW- Up            2017  6290    f.Trans-Manual 43356.000
## 4781              VW- Up            2015  8159 f.Trans-Automatic 20818.000
## 4790              VW- Up            2017  7495    f.Trans-Manual 22000.000
## 4792              VW- Up            2017  7400    f.Trans-Manual 12314.000
## 4802        VW- Scirocco            2014 12600    f.Trans-Manual 37506.000
## 4874         VW- Touareg            2019 42995 f.Trans-Automatic  1445.000
## 4887          VW- Arteon            2019 26490 f.Trans-Automatic  5907.000
## 4888          VW- Arteon            2019 23990  f.Trans-SemiAuto  2239.000
## 4889          VW- Arteon            2019 29995  f.Trans-SemiAuto  6789.000
## 4892          VW- Arteon            2019 34000 f.Trans-Automatic  1000.000
## 4901          VW- Touran            2018 20072    f.Trans-Manual 10162.000
## 4902          VW- Touran            2016 14995    f.Trans-Manual 28136.000
## 4912          VW- Touran            2018 21450  f.Trans-SemiAuto  9156.000
## 4913          VW- Touran            2016 14950    f.Trans-Manual 31004.000
## 4914          VW- Touran            2019 20000 f.Trans-Automatic 20535.000
##           fuelType       tax  mpg engineSize   manufacturer Audi total
## 4    f.Fuel-Petrol 145.00000 50.4      Petit     f.Man-Audi  Yes     0
## 52   f.Fuel-Petrol 145.00000 37.7       Gran     f.Man-Audi  Yes     1
## 154  f.Fuel-Diesel 160.00000 48.7       Gran     f.Man-Audi  Yes     0
## 378  f.Fuel-Petrol 145.00000 50.4      Petit     f.Man-Audi  Yes     0
## 403  f.Fuel-Diesel 145.00000 44.1      Mitjà     f.Man-Audi  Yes     0
## 467  f.Fuel-Diesel 145.00000 33.2       Gran     f.Man-Audi  Yes     0
## 482  f.Fuel-Petrol 150.00000 44.1      Petit     f.Man-Audi  Yes     0
## 535  f.Fuel-Petrol  30.00000 58.9      Petit     f.Man-Audi  Yes     0
## 538  f.Fuel-Petrol 145.00000 34.0      Mitjà     f.Man-Audi  Yes     0
## 572  f.Fuel-Petrol 145.00000 31.7      Mitjà     f.Man-Audi  Yes     0
## 651  f.Fuel-Petrol 145.00000 51.4      Petit     f.Man-Audi  Yes     0
## 661  f.Fuel-Petrol 145.00000 48.7      Petit     f.Man-Audi  Yes     0
## 738  f.Fuel-Diesel 125.00000 58.9      Mitjà     f.Man-Audi  Yes     0
## 753  f.Fuel-Petrol 145.00000 40.4      Mitjà     f.Man-Audi  Yes     0
## 754  f.Fuel-Petrol 145.00000 48.7      Petit     f.Man-Audi  Yes     0
## 759  f.Fuel-Diesel 145.00000 55.4      Mitjà     f.Man-Audi  Yes     0
```

```
## 849  f.Fuel-Diesel 150.00000  33.2       Gran     f.Man-Audi  Yes     0
## 889  f.Fuel-Diesel 145.00000  47.1      Mitjà     f.Man-Audi  Yes     0
## 941  f.Fuel-Diesel 145.00000  50.4      Mitjà     f.Man-Audi  Yes     0
## 971  f.Fuel-Diesel 160.00000  50.4       Gran     f.Man-Audi  Yes     0
## 978  f.Fuel-Diesel  20.00000  67.3      Mitjà     f.Man-Audi  Yes     0
## 979  f.Fuel-Diesel 145.00000  42.8      Mitjà     f.Man-Audi  Yes     0
## 1016 f.Fuel-Diesel 145.00000  47.1      Mitjà     f.Man-Audi  Yes     1
## 1094 f.Fuel-Petrol 300.00000  35.3       Gran      f.Man-BMW   No     0
## 1106 f.Fuel-Petrol 145.00000  42.2      Petit      f.Man-BMW   No     0
## 1148 f.Fuel-Diesel 125.00000  61.4      Mitjà      f.Man-BMW   No     0
## 1264 f.Fuel-Petrol 145.00000  34.9       Gran      f.Man-BMW   No     1
## 1309 f.Fuel-Diesel 145.00000  53.3       Gran      f.Man-BMW   No     0
## 1358 f.Fuel-Petrol 160.00000  44.1      Mitjà      f.Man-BMW   No     0
## 1447 f.Fuel-Petrol 150.00000  42.2      Mitjà      f.Man-BMW   No     0
## 1449 f.Fuel-Diesel 145.00000  65.7      Mitjà      f.Man-BMW   No     0
## 1455 f.Fuel-Hybrid  49.46007 134.5      Mitjà      f.Man-BMW   No     1
## 1465 f.Fuel-Diesel  30.00000  64.2      Mitjà      f.Man-BMW   No     0
## 1469 f.Fuel-Diesel 150.00000  65.7      Mitjà      f.Man-BMW   No     0
## 1477 f.Fuel-Diesel 145.00000  57.7      Mitjà      f.Man-BMW   No     0
## 1511 f.Fuel-Petrol 145.00000  43.5      Mitjà      f.Man-BMW   No     0
## 1535 f.Fuel-Diesel 145.00000  58.9      Mitjà      f.Man-BMW   No     0
## 1540 f.Fuel-Diesel 145.00000  49.6      Mitjà      f.Man-BMW   No     0
## 1542 f.Fuel-Diesel 145.00000  72.4      Petit      f.Man-BMW   No     0
## 1583 f.Fuel-Petrol 145.00000  48.7      Mitjà      f.Man-BMW   No     0
## 1589 f.Fuel-Diesel 145.00000  74.3      Petit      f.Man-BMW   No     0
## 1616 f.Fuel-Petrol 145.00000  34.9       Gran      f.Man-BMW   No     0
## 1629 f.Fuel-Petrol 145.00000  30.4      Mitjà      f.Man-BMW   No     0
## 1680 f.Fuel-Diesel  30.00000  65.7      Mitjà      f.Man-BMW   No     0
## 1734 f.Fuel-Diesel 145.00000  56.5       Gran      f.Man-BMW   No     0
## 1776 f.Fuel-Diesel 160.00000  49.6       Gran      f.Man-BMW   No     0
## 1779 f.Fuel-Petrol 150.00000  47.9      Mitjà      f.Man-BMW   No     0
## 1812 f.Fuel-Diesel 150.00000  54.3      Petit      f.Man-BMW   No     0
## 1813 f.Fuel-Petrol 145.00000  36.2      Mitjà      f.Man-BMW   No     0
## 1815 f.Fuel-Diesel 200.00000  47.9       Gran      f.Man-BMW   No     0
## 1840 f.Fuel-Diesel 160.00000  51.4      Mitjà      f.Man-BMW   No     0
## 1854 f.Fuel-Diesel 145.00000  44.1      Mitjà      f.Man-BMW   No     0
## 1989 f.Fuel-Petrol 150.00000  34.0       Gran      f.Man-BMW   No     1
## 2006 f.Fuel-Diesel 145.00000  65.7      Mitjà      f.Man-BMW   No     0
## 2011 f.Fuel-Petrol 235.00000  37.7       Gran      f.Man-BMW   No     0
## 2029 f.Fuel-Diesel 200.00000  47.1       Gran      f.Man-BMW   No     0
## 2084 f.Fuel-Hybrid 140.00000 156.9      Mitjà      f.Man-BMW   No     0
## 2099 f.Fuel-Petrol 150.00000  48.7      Mitjà      f.Man-BMW   No     0
## 2121 f.Fuel-Diesel 165.00000  50.4       Gran      f.Man-BMW   No     0
## 2210 f.Fuel-Diesel 145.00000  64.2      Mitjà f.Man-Mercedes   No     0
## 2220 f.Fuel-Diesel  30.00000  61.4      Mitjà f.Man-Mercedes   No     0
## 2327 f.Fuel-Diesel 145.00000  58.9      Mitjà f.Man-Mercedes   No     0
## 2329 f.Fuel-Diesel 145.00000  61.4      Mitjà f.Man-Mercedes   No     0
## 2354 f.Fuel-Petrol 145.00000  44.1      Mitjà f.Man-Mercedes   No     0
## 2527 f.Fuel-Diesel 145.00000  58.9      Mitjà f.Man-Mercedes   No     0
## 2729 f.Fuel-Petrol 145.00000  41.5      Mitjà f.Man-Mercedes   No     0
## 2754 f.Fuel-Petrol 145.00000  27.4       Gran f.Man-Mercedes   No     0
## 2824 f.Fuel-Diesel 145.00000  57.7      Mitjà f.Man-Mercedes   No     0
## 2908 f.Fuel-Diesel 150.00000  57.7      Mitjà f.Man-Mercedes   No     0
## 3039 f.Fuel-Petrol 145.00000  38.2      Mitjà f.Man-Mercedes   No     0
## 3057 f.Fuel-Diesel 145.00000  72.4      Petit f.Man-Mercedes   No     0
## 3072 f.Fuel-Diesel 145.00000  65.7      Mitjà f.Man-Mercedes   No     0
## 3102 f.Fuel-Diesel  20.00000  64.2      Mitjà f.Man-Mercedes   No     0
## 3130 f.Fuel-Diesel  30.00000  61.4      Mitjà f.Man-Mercedes   No     0
## 3140 f.Fuel-Diesel 125.00000  56.5      Mitjà f.Man-Mercedes   No     0
## 3149 f.Fuel-Diesel  30.00000  60.1      Petit f.Man-Mercedes   No     0
## 3168 f.Fuel-Diesel  30.00000  64.2      Mitjà f.Man-Mercedes   No     0
## 3173 f.Fuel-Diesel 150.00000  54.3       Gran f.Man-Mercedes   No     0
## 3192 f.Fuel-Diesel  30.00000  64.2      Mitjà f.Man-Mercedes   No     0
## 3202 f.Fuel-Petrol 145.00000  47.9      Mitjà f.Man-Mercedes   No     0
```

```
## 3215 f.Fuel-Diesel 145.00000  72.4      Petit f.Man-Mercedes   No     0
## 3216 f.Fuel-Diesel  30.00000  64.2      Petit f.Man-Mercedes   No     0
## 3224 f.Fuel-Petrol 145.00000  28.5      Petit f.Man-Mercedes   No     0
## 3225 f.Fuel-Diesel  20.00000  70.6      Petit f.Man-Mercedes   No     0
## 3228 f.Fuel-Diesel 125.00000  58.9      Mitjà f.Man-Mercedes   No     0
## 3290 f.Fuel-Petrol 150.00000  37.2      Mitjà f.Man-Mercedes   No     0
## 3308 f.Fuel-Petrol 570.00000  19.8       Gran f.Man-Mercedes   No     1
## 3310 f.Fuel-Petrol 160.00000  43.5      Petit f.Man-Mercedes   No     1
## 3348 f.Fuel-Diesel 125.00000  56.5      Mitjà f.Man-Mercedes   No     0
## 3364 f.Fuel-Diesel  30.00000  64.2      Mitjà f.Man-Mercedes   No     0
## 3371 f.Fuel-Diesel  20.00000  68.9      Petit f.Man-Mercedes   No     0
## 3674 f.Fuel-Petrol 145.00000  49.6      Petit       f.Man-VW   No     0
## 3738 f.Fuel-Petrol 145.00000  47.1      Petit       f.Man-VW   No     0
## 4595 f.Fuel-Diesel 145.00000  42.8      Mitjà       f.Man-VW   No     0
## 4599 f.Fuel-Petrol 145.00000  40.4      Petit       f.Man-VW   No     0
## 4602 f.Fuel-Diesel 145.00000  58.9      Mitjà       f.Man-VW   No     0
## 4714 f.Fuel-Petrol 145.00000  68.9      Petit       f.Man-VW   No     0
## 4721 f.Fuel-Petrol  20.00000  64.2      Petit       f.Man-VW   No     0
## 4729 f.Fuel-Petrol  20.00000  62.8      Petit       f.Man-VW   No     0
## 4736 f.Fuel-Petrol  20.00000  62.8      Petit       f.Man-VW   No     0
## 4739 f.Fuel-Petrol 145.00000  50.4      Petit       f.Man-VW   No     1
## 4743 f.Fuel-Petrol 150.00000  54.3      Petit       f.Man-VW   No     0
## 4744 f.Fuel-Petrol  20.00000  64.2      Petit       f.Man-VW   No     0
## 4781 f.Fuel-Petrol  20.00000  64.2      Petit       f.Man-VW   No     0
## 4790 f.Fuel-Petrol 145.00000  68.9      Petit       f.Man-VW   No     0
## 4792 f.Fuel-Petrol 145.00000  68.9      Petit       f.Man-VW   No     0
## 4802 f.Fuel-Diesel  20.00000  55.4      Mitjà       f.Man-VW   No     0
## 4874 f.Fuel-Diesel 145.00000  34.0       Gran       f.Man-VW   No     0
## 4887 f.Fuel-Petrol 145.00000  32.8      Mitjà       f.Man-VW   No     0
## 4888 f.Fuel-Petrol 145.00000  40.4      Petit       f.Man-VW   No     0
## 4889 f.Fuel-Diesel 145.00000  50.4      Mitjà       f.Man-VW   No     0
## 4892 f.Fuel-Diesel 145.00000  37.7      Mitjà       f.Man-VW   No     0
## 4901 f.Fuel-Diesel 145.00000  51.4      Petit       f.Man-VW   No     0
## 4902 f.Fuel-Diesel  30.00000  64.2      Petit       f.Man-VW   No     0
## 4912 f.Fuel-Petrol 145.00000  51.4      Petit       f.Man-VW   No     0
## 4913 f.Fuel-Diesel  30.00000  64.2      Petit       f.Man-VW   No     0
## 4914 f.Fuel-Diesel 145.00000  49.6      Petit       f.Man-VW   No     0
##      years_sell years_sell2                aux       f.price          f.miles
## 4      Molt nou           1  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 52     Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 154    Semi nou           2 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 378    Semi nou           2  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 403    Molt nou           1         [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 467    Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A    f.miles-[0,6]
## 482    Molt nou           1         [0,5.89e+03] Segmento -  B    f.miles-[0,6]
## 535    Semi nou           2  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 538    Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 572    Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 651    Semi nou           2 (5.89e+03,1.69e+04] Segmento -  C   f.miles-(6,17]
## 661    Molt nou           1         [0,5.89e+03] Segmento -  C    f.miles-[0,6]
## 738    Semi nou           2  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 753    Semi nou           2  (1.69e+04,3.4e+04] Segmento -  B  f.miles-(17,34]
## 754    Semi nou           2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 759    Molt nou           1 (5.89e+03,1.69e+04] Segmento -  B   f.miles-(6,17]
## 849    Molt nou           1         [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 889    Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 941    Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 971    Semi nou           2  (3.4e+04,3.23e+05] Segmento -  B f.miles-(34,323]
## 978    Semi nou           2  (3.4e+04,3.23e+05] Segmento -  C f.miles-(34,323]
## 979    Molt nou           1         [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1016   Molt nou           1         [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1094   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 1106   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  C    f.miles-[0,6]
## 1148   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
```

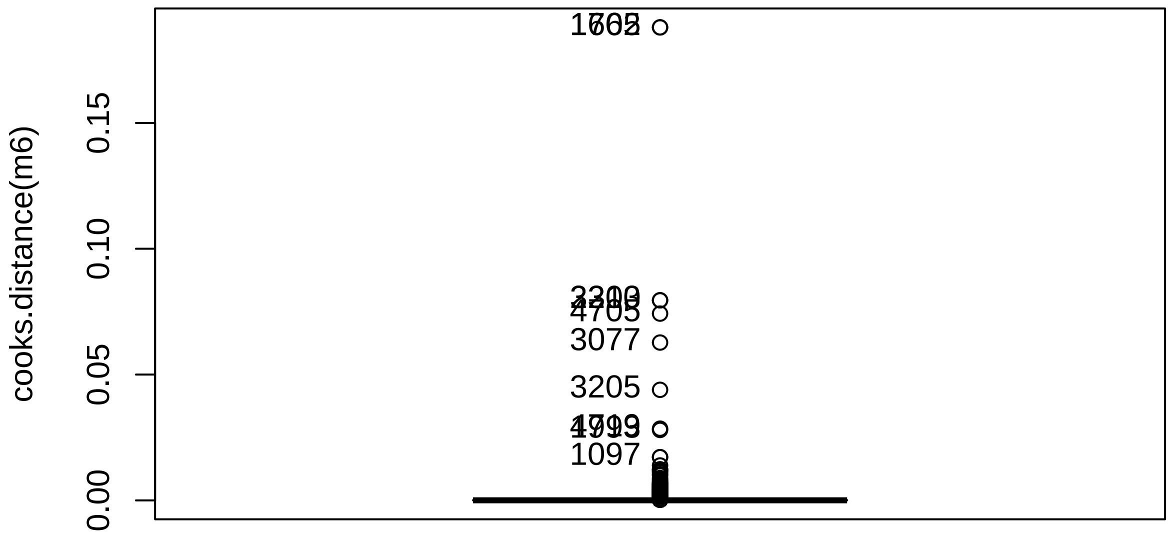
```
## 1264   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1309   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1358   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  C f.miles-(34,323]
## 1447   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1449   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1455   Semi nou           2 (5.89e+03,1.69e+04] Segmento -  C   f.miles-(6,17]
## 1465       Vell           3  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 1469   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  B   f.miles-(6,17]
## 1477   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  B   f.miles-(6,17]
## 1511   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1535   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 1540   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1542   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  C   f.miles-(6,17]
## 1583   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 1589   Semi nou           2 (5.89e+03,1.69e+04] Segmento -  D   f.miles-(6,17]
## 1616   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1629   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 1680   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  C f.miles-(34,323]
## 1734   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 1776   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  B  f.miles-(17,34]
## 1779   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  B   f.miles-(6,17]
## 1812   Molt nou           1  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 1813   Molt nou           1        [0,5.89e+03] Segmento -  B    f.miles-[0,6]
## 1815   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  B  f.miles-(17,34]
## 1840       Vell           3  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 1854   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 1989   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 2006   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 2011   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 2029   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  B f.miles-(34,323]
## 2084   Molt nou           1  (1.69e+04,3.4e+04] Segmento -  A  f.miles-(17,34]
## 2099   Molt nou           1  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
## 2121   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  C f.miles-(34,323]
## 2210   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 2220   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  C f.miles-(34,323]
## 2327   Semi nou           2 (5.89e+03,1.69e+04] Segmento -  B   f.miles-(6,17]
## 2329   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 2354   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  B   f.miles-(6,17]
## 2527   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  B  f.miles-(17,34]
## 2729   Semi nou           2 (5.89e+03,1.69e+04] Segmento -  C   f.miles-(6,17]
## 2754   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 2824   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 2908   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 3039   Molt nou           1        [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 3057   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 3072   Molt nou           1  (1.69e+04,3.4e+04] Segmento -  B  f.miles-(17,34]
## 3102   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 3130   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  B  f.miles-(17,34]
## 3140   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  C f.miles-(34,323]
## 3149   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 3168       Vell           3  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 3173   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  C f.miles-(34,323]
## 3192   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  C f.miles-(34,323]
## 3202   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  B  f.miles-(17,34]
## 3215   Semi nou           2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 3216   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 3224   Molt nou           1 (5.89e+03,1.69e+04] Segmento -  C   f.miles-(6,17]
## 3225   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 3228   Semi nou           2 (5.89e+03,1.69e+04] Segmento -  C   f.miles-(6,17]
## 3290   Molt nou           1  (1.69e+04,3.4e+04] Segmento -  B  f.miles-(17,34]
## 3308       Vell           3  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 3310   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 3348   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  B f.miles-(34,323]
## 3364   Semi nou           2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
```

```
## 3371   Semi nou          2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 3674   Molt nou          1 (5.89e+03,1.69e+04] Segmento -  C   f.miles-(6,17]
## 3738   Molt nou          1 (5.89e+03,1.69e+04] Segmento -  C   f.miles-(6,17]
## 4595   Molt nou          1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 4599   Molt nou          1 (5.89e+03,1.69e+04] Segmento -  B   f.miles-(6,17]
## 4602   Semi nou          2  (3.4e+04,3.23e+05] Segmento -  C f.miles-(34,323]
## 4714   Molt nou          1 (5.89e+03,1.69e+04] Segmento -  D   f.miles-(6,17]
## 4721   Semi nou          2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 4729   Semi nou          2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 4736   Semi nou          2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 4739   Molt nou          1         [0,5.89e+03] Segmento -  D    f.miles-[0,6]
## 4743   Molt nou          1         [0,5.89e+03] Segmento -  D    f.miles-[0,6]
## 4744   Semi nou          2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 4781   Semi nou          2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 4790   Semi nou          2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 4792   Semi nou          2 (5.89e+03,1.69e+04] Segmento -  D   f.miles-(6,17]
## 4802   Semi nou          2  (3.4e+04,3.23e+05] Segmento -  D f.miles-(34,323]
## 4874   Molt nou          1         [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 4887   Molt nou          1 (5.89e+03,1.69e+04] Segmento -  A    f.miles-[0,6]
## 4888   Molt nou          1         [0,5.89e+03] Segmento -  B    f.miles-[0,6]
## 4889   Molt nou          1 (5.89e+03,1.69e+04] Segmento -  A   f.miles-(6,17]
## 4892   Molt nou          1         [0,5.89e+03] Segmento -  A    f.miles-[0,6]
## 4901   Molt nou          1 (5.89e+03,1.69e+04] Segmento -  B   f.miles-(6,17]
## 4902   Semi nou          2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 4912   Molt nou          1 (5.89e+03,1.69e+04] Segmento -  B   f.miles-(6,17]
## 4913   Semi nou          2  (1.69e+04,3.4e+04] Segmento -  D  f.miles-(17,34]
## 4914   Molt nou          1  (1.69e+04,3.4e+04] Segmento -  C  f.miles-(17,34]
##                f.tax              mpg_d hcpck claKM      mout
## 4    f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 52   f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 154  f.tax-(150,570] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 378  f.tax-(145,150] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 403  f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 467  f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 482  f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 535    f.tax-(1,125] mpg_d-(53.3,61.4] kHP-3 kKM-1  MvOut.No
## 538  f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 572  f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 651  f.tax-(145,150] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 661  f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 738    f.tax-(1,125] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 753  f.tax-(145,150]    mpg_d-[0,44.8] kHP-2 kKM-2  MvOut.No
## 754  f.tax-(145,150] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 759  f.tax-(145,150] mpg_d-(53.3,61.4] kHP-1 kKM-3  MvOut.No
## 849  f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 889  f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 941  f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 971  f.tax-(150,570] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 978    f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 979  f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 1016 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 1094 f.tax-(150,570]    mpg_d-[0,44.8] kHP-2 kKM-2  MvOut.No
## 1106 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 1148   f.tax-(1,125] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 1264 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 1309 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 1358 f.tax-(150,570]    mpg_d-[0,44.8] kHP-2 kKM-2  MvOut.No
## 1447 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 1449 f.tax-(145,150]  mpg_d-(61.4,471] kHP-1 kKM-3  MvOut.No
## 1455   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 1465   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 1469 f.tax-(145,150]  mpg_d-(61.4,471] kHP-1 kKM-3  MvOut.No
## 1477 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-1 kKM-3  MvOut.No
## 1511 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
```

```
## 1535 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-1 kKM-3  MvOut.No
## 1540 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 1542 f.tax-(145,150]  mpg_d-(61.4,471] kHP-1 kKM-3  MvOut.No
## 1583 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 1589 f.tax-(145,150]  mpg_d-(61.4,471] kHP-2 kKM-2  MvOut.No
## 1616 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 1629 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 1680   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 1734 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 1776 f.tax-(150,570] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 1779 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 1812 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-1 kKM-3  MvOut.No
## 1813 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 1815 f.tax-(150,570] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 1840 f.tax-(150,570] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 1854 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 1989 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 2006 f.tax-(145,150]  mpg_d-(61.4,471] kHP-2 kKM-2  MvOut.No
## 2011 f.tax-(150,570]    mpg_d-[0,44.8] kHP-2 kKM-2  MvOut.No
## 2029 f.tax-(150,570] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 2084 f.tax-(125,145]  mpg_d-(61.4,471] kHP-3 kKM-1 MvOut.Yes
## 2099 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 2121 f.tax-(150,570] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 2210 f.tax-(145,150]  mpg_d-(61.4,471] kHP-1 kKM-3  MvOut.No
## 2220   f.tax-(1,125] mpg_d-(53.3,61.4] kHP-3 kKM-1  MvOut.No
## 2327 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 2329 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-1 kKM-3  MvOut.No
## 2354 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 2527 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 2729 f.tax-(145,150]    mpg_d-[0,44.8] kHP-2 kKM-2  MvOut.No
## 2754 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 2824 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-1 kKM-3  MvOut.No
## 2908 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-1 kKM-3  MvOut.No
## 3039 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 3057 f.tax-(145,150]  mpg_d-(61.4,471] kHP-2 kKM-2  MvOut.No
## 3072 f.tax-(145,150]  mpg_d-(61.4,471] kHP-1 kKM-3  MvOut.No
## 3102   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 3130   f.tax-(1,125] mpg_d-(53.3,61.4] kHP-3 kKM-1  MvOut.No
## 3140   f.tax-(1,125] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 3149   f.tax-(1,125] mpg_d-(53.3,61.4] kHP-3 kKM-1  MvOut.No
## 3168   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 3173 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 3192   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 3202 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-2 kKM-2  MvOut.No
## 3215 f.tax-(145,150]  mpg_d-(61.4,471] kHP-2 kKM-2  MvOut.No
## 3216   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 3224 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 3225   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 3228   f.tax-(1,125] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 3290 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 3308 f.tax-(150,570]    mpg_d-[0,44.8] kHP-2 kKM-2  MvOut.No
## 3310 f.tax-(150,570]    mpg_d-[0,44.8] kHP-2 kKM-2  MvOut.No
## 3348   f.tax-(1,125] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 3364   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 3371   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 3674 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 3738 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 4595 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 4599 f.tax-(145,150]    mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 4602 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-2 kKM-2  MvOut.No
## 4714 f.tax-(145,150]  mpg_d-(61.4,471] kHP-1 kKM-3  MvOut.No
## 4721   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 4729   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 4736   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
```

```
## 4739 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 4743 f.tax-(145,150] mpg_d-(53.3,61.4] kHP-1 kKM-3  MvOut.No
## 4744   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 4781   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 4790 f.tax-(145,150]  mpg_d-(61.4,471] kHP-2 kKM-2  MvOut.No
## 4792 f.tax-(145,150]  mpg_d-(61.4,471] kHP-2 kKM-2  MvOut.No
## 4802   f.tax-(1,125] mpg_d-(53.3,61.4] kHP-3 kKM-1  MvOut.No
## 4874 f.tax-(145,150]     mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 4887 f.tax-(145,150]     mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 4888 f.tax-(145,150]     mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 4889 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 4892 f.tax-(145,150]     mpg_d-[0,44.8] kHP-1 kKM-3  MvOut.No
## 4901 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 4902   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 4912 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
## 4913   f.tax-(1,125]  mpg_d-(61.4,471] kHP-3 kKM-1  MvOut.No
## 4914 f.tax-(145,150] mpg_d-(44.8,53.3] kHP-1 kKM-3  MvOut.No
```

```
Boxplot(cooks.distance(m6))
```




```
##  [1] 1762 1605 3203 2310 4705 3077 3205 4719 1993 1097
```

```
llcoo <- which( cooks.distance(m6) > 0.05);
df[llcoo,]
```

```
##                   model year price      transmission mileage      fuelType tax
## 1616     BMW- 3 Series 2019 39995 f.Trans-Automatic     999 f.Fuel-Petrol 145
## 1776     BMW- 4 Series 2015 21149  f.Trans-SemiAuto   29627 f.Fuel-Diesel 160
## 2329 Mercedes- E Class 2019 28995  f.Trans-SemiAuto   12630 f.Fuel-Diesel 145
## 3102 Mercedes- C Class 2015 13990 f.Trans-Automatic   29000 f.Fuel-Diesel  20
## 3228 Mercedes- A Class 2016 18699 f.Trans-Automatic   13118 f.Fuel-Diesel 125
## 4742            VW- Up 2019 10990    f.Trans-Manual    2000 f.Fuel-Petrol 150
##       mpg engineSize   manufacturer Audi total years_sell years_sell2
## 1616 34.9       Gran      f.Man-BMW   No     0   Molt nou           1
```

```
## 1776 49.6       Gran      f.Man-BMW   No     0   Semi nou          2
## 2329 61.4      Mitjà f.Man-Mercedes   No     0   Molt nou          1
## 3102 64.2      Mitjà f.Man-Mercedes   No     0   Semi nou          2
## 3228 58.9      Mitjà f.Man-Mercedes   No     0   Semi nou          2
## 4742 51.4      Petit       f.Man-VW   No     0   Molt nou          1
##                       aux       f.price         f.miles           f.tax
## 1616        [0,5.89e+03] Segmento -  A   f.miles-[0,6] f.tax-(145,150]
## 1776  (1.69e+04,3.4e+04] Segmento -  B f.miles-(17,34] f.tax-(150,570]
## 2329 (5.89e+03,1.69e+04] Segmento -  A  f.miles-(6,17] f.tax-(145,150]
## 3102  (1.69e+04,3.4e+04] Segmento -  D f.miles-(17,34]   f.tax-(1,125]
## 3228 (5.89e+03,1.69e+04] Segmento -  C  f.miles-(6,17]   f.tax-(1,125]
## 4742        [0,5.89e+03] Segmento -  D   f.miles-[0,6] f.tax-(145,150]
##                   mpg_d hcpck claKM     mout
## 1616    mpg_d-[0,44.8] kHP-1 kKM-3 MvOut.No
## 1776 mpg_d-(44.8,53.3] kHP-2 kKM-2 MvOut.No
## 2329 mpg_d-(53.3,61.4] kHP-1 kKM-3 MvOut.No
## 3102  mpg_d-(61.4,471] kHP-3 kKM-1 MvOut.No
## 3228 mpg_d-(53.3,61.4] kHP-2 kKM-2 MvOut.No
## 4742 mpg_d-(44.8,53.3] kHP-1 kKM-3 MvOut.No
```

6.4 A4

```
summary(bm6)
```

```
##
## Call:
## glm(formula = Audi ~ (mileage + tax + mpg + transmission + engineSize) *
##     (transmission + engineSize), family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5798  -0.7208  -0.6250  -0.3523   2.4804
##
## Coefficients:
##                                              Estimate Std. Error z value
## (Intercept)                                -7.463e-01  4.912e-01  -1.519
## mileage                                     4.965e-06  3.489e-06   1.423
## tax                                        -8.350e-04  1.332e-03  -0.627
## mpg                                        -9.312e-03  7.464e-03  -1.248
## transmissionf.Trans-SemiAuto                1.797e+00  8.429e-01   2.132
## transmissionf.Trans-Automatic               2.679e+00  9.438e-01   2.839
## engineSizeMitjà                             8.691e-01  8.215e-01   1.058
## engineSizeGran                             -1.382e+01  2.568e+02  -0.054
## mileage:transmissionf.Trans-SemiAuto        2.133e-05  4.941e-06   4.318
## mileage:transmissionf.Trans-Automatic      -1.180e-05  5.002e-06  -2.359
## mileage:engineSizeMitjà                     4.650e-06  4.308e-06   1.079
## mileage:engineSizeGran                      1.720e-05  8.854e-06   1.943
## tax:transmissionf.Trans-SemiAuto            2.296e-03  2.323e-03   0.988
## tax:transmissionf.Trans-Automatic          -2.158e-03  2.453e-03  -0.880
## tax:engineSizeMitjà                        -1.558e-03  2.040e-03  -0.764
## tax:engineSizeGran                         -2.373e-03  3.190e-03  -0.744
## mpg:transmissionf.Trans-SemiAuto           -4.728e-02  1.267e-02  -3.731
## mpg:transmissionf.Trans-Automatic          -4.934e-02  1.434e-02  -3.442
## mpg:engineSizeMitjà                        -1.370e-02  1.188e-02  -1.154
## mpg:engineSizeGran                          1.255e-02  1.937e-02   0.648
## transmissionf.Trans-SemiAuto:engineSizeMitjà  -5.783e-01  2.163e-01  -2.673
## transmissionf.Trans-Automatic:engineSizeMitjà  1.159e-01  2.619e-01   0.443
## transmissionf.Trans-SemiAuto:engineSizeGran    1.268e+01  2.568e+02   0.049
## transmissionf.Trans-Automatic:engineSizeGran   1.325e+01  2.568e+02   0.052
##                                              Pr(>|z|)
```

```
## (Intercept)                                   0.128698
## mileage                                       0.154645
## tax                                           0.530812
## mpg                                           0.212176
## transmissionf.Trans-SemiAuto                  0.032983 *
## transmissionf.Trans-Automatic                 0.004525 **
## engineSizeMitjà                               0.290061
## engineSizeGran                                0.957061
## mileage:transmissionf.Trans-SemiAuto          1.58e-05 ***
## mileage:transmissionf.Trans-Automatic         0.018307 *
## mileage:engineSizeMitjà                       0.280386
## mileage:engineSizeGran                        0.052042 .
## tax:transmissionf.Trans-SemiAuto              0.323050
## tax:transmissionf.Trans-Automatic             0.378995
## tax:engineSizeMitjà                           0.444963
## tax:engineSizeGran                            0.456959
## mpg:transmissionf.Trans-SemiAuto              0.000191 ***
## mpg:transmissionf.Trans-Automatic             0.000578 ***
## mpg:engineSizeMitjà                           0.248694
## mpg:engineSizeGran                            0.517016
## transmissionf.Trans-SemiAuto:engineSizeMitjà  0.007509 **
## transmissionf.Trans-Automatic:engineSizeMitjà 0.657930
## transmissionf.Trans-SemiAuto:engineSizeGran   0.960605
## transmissionf.Trans-Automatic:engineSizeGran  0.958831
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5041.6  on 4961  degrees of freedom
## Residual deviance: 4776.8  on 4938  degrees of freedom
## AIC: 4824.8
##
## Number of Fisher Scoring iterations: 13
```

```
step(bm6)
```

```
## Start:  AIC=4824.82
## Audi ~ (mileage + tax + mpg + transmission + engineSize) * (transmission +
##     engineSize)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

##                           Df Deviance    AIC
## - tax:engineSize           2   4777.6 4821.6
## - mpg:engineSize           2   4780.0 4824.0
## - tax:transmission         2   4780.7 4824.7
## - mileage:engineSize       2   4780.7 4824.7
## <none>                         4776.8 4824.8
## - transmission:engineSize  4   4791.2 4831.2
## - mpg:transmission         2   4793.9 4837.9
## - mileage:transmission     2   4814.7 4858.7
##
## Step:  AIC=4821.59
## Audi ~ mileage + tax + mpg + transmission + engineSize + mileage:transmission +
##     mileage:engineSize + tax:transmission + mpg:transmission +
##     mpg:engineSize + transmission:engineSize

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
##                           Df Deviance    AIC
## - mpg:engineSize           2   4781.5 4821.5
## <none>                         4777.6 4821.6
## - mileage:engineSize       2   4781.8 4821.8
## - tax:transmission         2   4782.0 4822.0
## - transmission:engineSize  4   4792.3 4828.3
## - mpg:transmission         2   4801.3 4841.3
## - mileage:transmission     2   4815.0 4855.0
##
## Step:  AIC=4821.51
## Audi ~ mileage + tax + mpg + transmission + engineSize + mileage:transmission +
##     mileage:engineSize + tax:transmission + mpg:transmission +
##     transmission:engineSize
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
##                           Df Deviance    AIC
## <none>                         4781.5 4821.5
## - tax:transmission         2   4786.5 4822.5
## - mileage:engineSize       2   4788.7 4824.7
## - transmission:engineSize  4   4796.7 4828.7
## - mpg:transmission         2   4814.0 4850.0
## - mileage:transmission     2   4819.4 4855.4
```

```
##
## Call:  glm(formula = Audi ~ mileage + tax + mpg + transmission + engineSize +
##     mileage:transmission + mileage:engineSize + tax:transmission +
##     mpg:transmission + transmission:engineSize, family = binomial(link = logit),
##     data = df)
##
## Coefficients:
##                                (Intercept)
##                                 -5.881e-01
##                                    mileage
##                                  5.431e-06
##                                        tax
##                                 -1.075e-03
##                                        mpg
##                                 -1.183e-02
##                  transmissionf.Trans-SemiAuto
##                                  2.236e+00
##                 transmissionf.Trans-Automatic
##                                  3.177e+00
##                              engineSizeMitjà
##                                 -5.281e-02
##                               engineSizeGran
##                                 -1.415e+01
##         mileage:transmissionf.Trans-SemiAuto
##                                  2.044e-05
##        mileage:transmissionf.Trans-Automatic
##                                 -1.276e-05
##                      mileage:engineSizeMitjà
##                                  3.827e-06
##                       mileage:engineSizeGran
##                                  1.939e-05
##             tax:transmissionf.Trans-SemiAuto
##                                  9.518e-04
##            tax:transmissionf.Trans-Automatic
##                                 -3.712e-03
##             mpg:transmissionf.Trans-SemiAuto
##                                 -5.250e-02
##            mpg:transmissionf.Trans-Automatic
##                                 -5.542e-02
```

```
##  transmissionf.Trans-SemiAuto:engineSizeMitjà
##                                 -5.151e-01
## transmissionf.Trans-Automatic:engineSizeMitjà
##                                  1.932e-01
##   transmissionf.Trans-SemiAuto:engineSizeGran
##                                  1.307e+01
##  transmissionf.Trans-Automatic:engineSizeGran
##                                  1.360e+01
##
## Degrees of Freedom: 4961 Total (i.e. Null);  4942 Residual
## Null Deviance:       5042
## Residual Deviance: 4782  AIC: 4822
```