library(dummies)

## dummies-1.5.6 provided by Decision Patterns

library(GGally)

## Loading required package: ggplot2

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(gplots)

##   
## Attaching package: 'gplots'

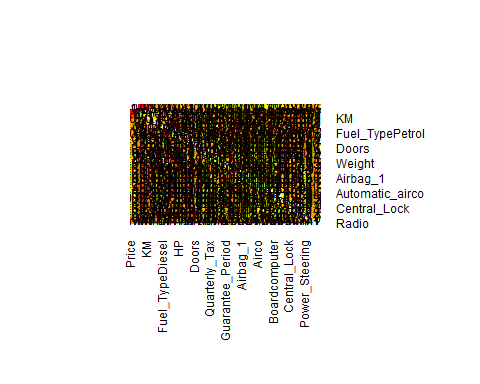
## The following object is masked from 'package:stats':  
##   
## lowess

toyota <- read.csv(file = "C:/Users/abhis/Documents/R Studio/datasets/ToyotaCorolla.csv", header = TRUE)

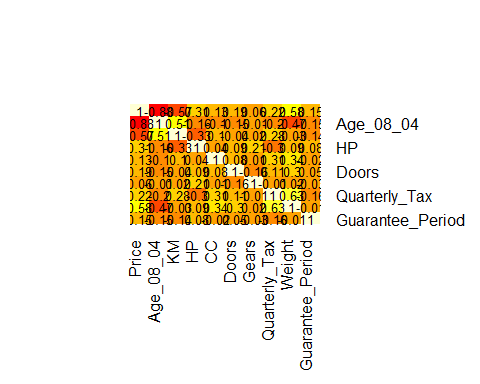
#a  
#The categorical variables in this dataset are: Model, FUel\_type, Color, Met\_Color, Mfr\_Month, Mfr\_Year, Automatic, Mfr\_Guarantee, BOVAG\_Guarantee, ABS, Airbag\_1 , Airbag\_2 , Airco , Automatic\_airco , Boardcomputer , CD\_Player , Central\_Lock , Powered\_Windows, Power\_Steering, Radio, Mistlamps, Sport\_Model, Backseat\_Divider, Metallic\_Rim, Radio\_cassette , Parking assistance system, Tow\_Bar  
  
#b  
# A Dummy variable or Indicator Variable is an artificial variable created to represent an attribute with two or more distinct categories/levels.A variable with N categories will be converted to N or N-1 dummy variables. We create dummy variables for categorical data as yes or no which are converted to 0 and 1.  
  
#c  
# N or N-1 variables are required to capture the information in a categorical variable  
  
#d  
  
toyota$Mfg\_Year <- factor(toyota$Mfg\_Year)  
  
toyotadummy <- dummy.data.frame(toyota[,-c(1,2,5,6,11,15)])

## Warning in model.matrix.default(~x - 1, model.frame(~x - 1), contrasts = FALSE):  
## non-list contrasts argument ignored

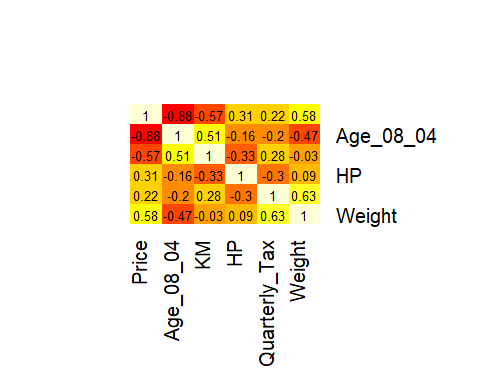
#One of the example for this is: The Fuel\_Type variable which has three category, Petrol, Diesel and CNG. This results in three different dummy variables Fuel\_Type\_Petrol, Fuel\_Type\_Diesel and Fuel\_Type\_CNG  
  
#e  
  
#Correlation Matrix  
Cor\_Matrix <- round(cor(toyotadummy),2)  
  
  
  
  
#After analysing Variables which are not important seeing the Correlation matrix excluded-  
#Radio\_Cassette,Tow\_Bar, Parking\_Assistant,Mettalic\_Rim, Backseat\_Divider  
#Sports\_Model,Mistlamps,BOVAG\_Guarantee, Mfr\_Guarantee, Automatic, Met\_Color  
#Radio and radio\_cassette have high correlation 0.99 correlation so only radio is considered  
Toyota\_a <- toyotadummy[,-c(8,9,15,16,29,30,31,32,33,34,35)]  
  
  
#Heatmap to analyse all data  
heatmap.2(cor(Toyota\_a), Rowv = FALSE, Colv = FALSE, dendrogram = "none",  
 cellnote = round(cor(Toyota\_a),2),  
 notecol = "black", key = FALSE, trace = 'none', margins = c(10,10))



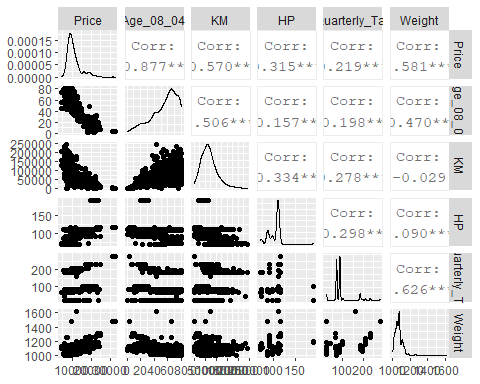
# All continuous variables  
Toyota\_Co.df <- toyotadummy[,c(1,2,3,7,10,11,12,13,14,17)]  
  
  
#Heat Map for only continuous variables  
  
  
heatmap.2(cor(Toyota\_Co.df), Rowv = FALSE, Colv = FALSE, dendrogram = "none",  
 cellnote = round(cor(Toyota\_Co.df),2),  
 notecol = "black", key = FALSE, trace = 'none', margins = c(10,10))



#After analysing data other highly Correalted variables which are dropped:-  
#CC, Doors, Gears, Guarantee period  
  
Toyota\_Drop.df <- Toyota\_Co.df[,-c(5,6,7,10)]  
  
  
  
#Heat Map of only relevant continuous variables  
#Continous Variables relevant- Price,Age,KM,Quaterly\_Tax,Weight,HP  
  
  
heatmap.2(cor(Toyota\_Drop.df), Rowv = FALSE, Colv = FALSE, dendrogram = "none",  
 cellnote = round(cor(Toyota\_Drop.df),2),  
 notecol = "black", key = FALSE, trace = 'none', margins = c(10,10))



y<-Toyota\_Drop.df  
  
#Only Continuous variables for Correlation matrix:-  
Cor\_Matrix <- round(cor(y),2)  
  
  
#Matrix scatter plot Continous variables   
ggpairs(Toyota\_Drop.df)



#After the analysis of several variables. The corelation between multiple variables in considered:  
#Age\_08\_04 is negatively correlated with price (-0.88): This means when the age of the car is more the price of the car goes down.KM is negatively correlated with price(-0.57): When the km increases similar to age the price of the car decreases. Weight is positively correlated with price(0.58): Weight of the Car is positively correlated with the Price (higher price higher weight). KM is positively correlated with Age\_08\_04( 0.51): Other than price, KM and Age of Car is positively stating a increasing relationship for both the factors when one increases. Weight is positively correlated with quarterly tax(0.63) : Quarterly Road Tax collected is positively related to Weight stating higher Car weight higher the Quarterly road tax.