

# Found data viz

Collin Ryan

## R Markdown

```
EUcars <- read_csv("car_prices.csv", na = c("", NA, "N/A"))
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,  
## e.g.:  
##   dat <- vroom(...)  
##   problems(dat)
```

```
## Rows: 12904 Columns: 18  
## -- Column specification -----  
## Delimiter: ","  
## chr (11): Brand, Model, Country, Gearbox, Fuel, Seller, Body Type, Type, Dri...  
## dbl  (7): Price, Kilometers, Year, Power, Seats, Doors, Cylinders  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
EUcars = EUcars |> drop_na() |> filter (Brand == "Fiat", Gearbox != "Semi-automatic", Drivetrain != "Re
```

```
EUcars <- replace(EUcars, EUcars == "Automatic", "0")  
EUcars <- replace(EUcars, EUcars == "Manual", "1")  
EUcars <- replace(EUcars, EUcars == "4WD", "1")  
EUcars <- replace(EUcars, EUcars == "Front", "0")  
EUcars <- replace(EUcars, EUcars == "AT", "1")  
EUcars <- replace(EUcars, EUcars == "BE", "2")  
EUcars <- replace(EUcars, EUcars == "DE", "3")  
EUcars <- replace(EUcars, EUcars == "IT", "4")  
EUcars <- replace(EUcars, EUcars == "NL", "5")
```

```
EUcars$Gearbox = as.numeric(EUcars$Gearbox)  
EUcars$Drivetrain = as.numeric(EUcars$Drivetrain)  
EUcars$Country = as.numeric(EUcars$Country)
```

## Visualize and summarize data

This data was found using Kaggle. The full dataset is of used cars sold in the EU, however it is pruned to just show the Fiat cars. The vizuization is of the correlation between the variables. (correlation matrix and correlation scatter plot)The data has some variables that are highly correlated. Outstanding correlation includes: Kilometers and Price with a strong negatth a strong pive correlation, Price and Power with a

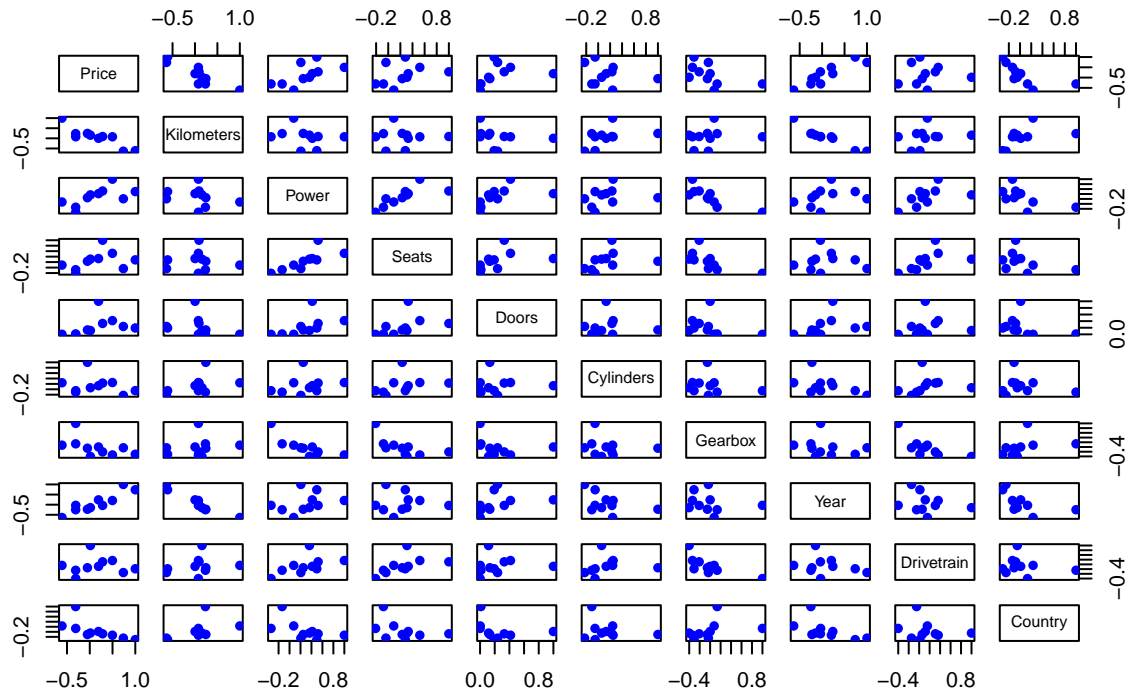
strong positive correlation, Price and Year wiositive correlation, Kilometers and Year with a strong negative correlation and Power and Seats with a strong positive correlation. Some variables such as Cylinders and Price as well as Drivetrain and Price have very little correlation. All the other variables have correlation somewhere in between.

```
cor_matrix<-cor(EUcars[,2:11])
cor_matrix
```

```
##           Price Kilometers      Power      Seats      Doors
## Price      1.00000000 -0.61030182  0.49745095  0.28191699  1.920660e-01
## Kilometers -0.61030182  1.00000000  0.07645695  0.09055902  1.810570e-03
## Power      0.49745095  0.07645695  1.00000000  0.52099967  4.149192e-01
## Seats      0.28191699  0.09055902  0.52099967  1.00000000  3.294652e-01
## Doors      0.19206598  0.00181057  0.41491923  0.32946520  1.000000e+00
## Cylinders  -0.05129488  0.24237880  0.24996127  0.23053116  1.331727e-01
## Gearbox    -0.30824765  0.07674345 -0.33754007 -0.20916237 -2.589792e-20
## Year       0.73774252 -0.63454640  0.20646219 -0.03813200  2.385605e-01
## Drivetrain  0.01373015  0.15656668  0.36348564  0.30888134  1.162643e-01
## Country    -0.30950166  0.23315770 -0.13494814 -0.07904203  1.148197e-02
##           Cylinders      Gearbox      Year Drivetrain      Country
## Price      -0.05129488 -3.082476e-01  0.73774252  0.01373015 -0.30950166
## Kilometers  0.24237880  7.674345e-02 -0.63454640  0.15656668  0.23315770
## Power      0.24996127 -3.375401e-01  0.20646219  0.36348564 -0.13494814
## Seats      0.23053116 -2.091624e-01 -0.03813200  0.30888134 -0.07904203
## Doors      0.13317267 -2.589792e-20  0.23856050  0.11626434  0.01148197
## Cylinders   1.00000000 -5.243618e-02 -0.22695567  0.05753373 -0.10660379
## Gearbox    -0.05243618  1.000000e+00 -0.03329068 -0.40148695  0.13408615
## Year       -0.22695567 -3.329068e-02  1.00000000 -0.14718764 -0.24991285
## Drivetrain  0.05753373 -4.014870e-01 -0.14718764  1.00000000 -0.04970134
## Country    -0.10660379  1.340862e-01 -0.24991285 -0.04970134  1.00000000
```

```
pairs(cor(EUcars[,2:11]), col = "blue", pch = 16, main="Correlation Scatter Plot")
```

## Correlation Scatter Plot



```
corrplot(cor_matrix, title="Correlation Plot")
```

