In this project, we will available a database of cars get in kaggle available in follow link:https://www.kaggle.com/datasets/gagandeep16/car-sales. The database has a lot informations as manufacturer, sales, vehicle type and others. Our mission is extract relevant informations by graphs, and tables using the software R.

First, we will verify wichs brands has most/less sales in this period. The analysis show us that the Ford has done more sales in this period, whereas Porsche has done less in the same period. This has done by the code below.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#import the base
dados <- read.csv('Car_sales.csv')
#agroup the sale by brands
dados_agrupados = dados %>% group_by(Manufacturer) %>%
summarise(total_Sales = sum(Sales_in_thousands))
#get the brand that sales max
dados_agrupados$Manufacturer[which.max(dados_agrupados$total_Sales)]
```

```
## [1] "Ford"
```

```
#get the brand that sales min
dados_agrupados$Manufacturer[which.min(dados_agrupados$total_Sales)]
```

```
## [1] "Porsche"
```

Now, we will analyze only passenger cars. We will verify which car is the the most/less expensive in this database. The analysis show us that the model CL500 of Mercedes Benz has the most expensive car, where as the model Metro of Chevrolet has the least expensive car. This has done by code below.

```
#date cars of type passenger
dados_passenger = dados[dados$Vehicle_type == "Passenger",]
#get brand and model of the car with most value
dados_passenger[which.max(dados_passenger$Price_in_thousands),1:2]
```

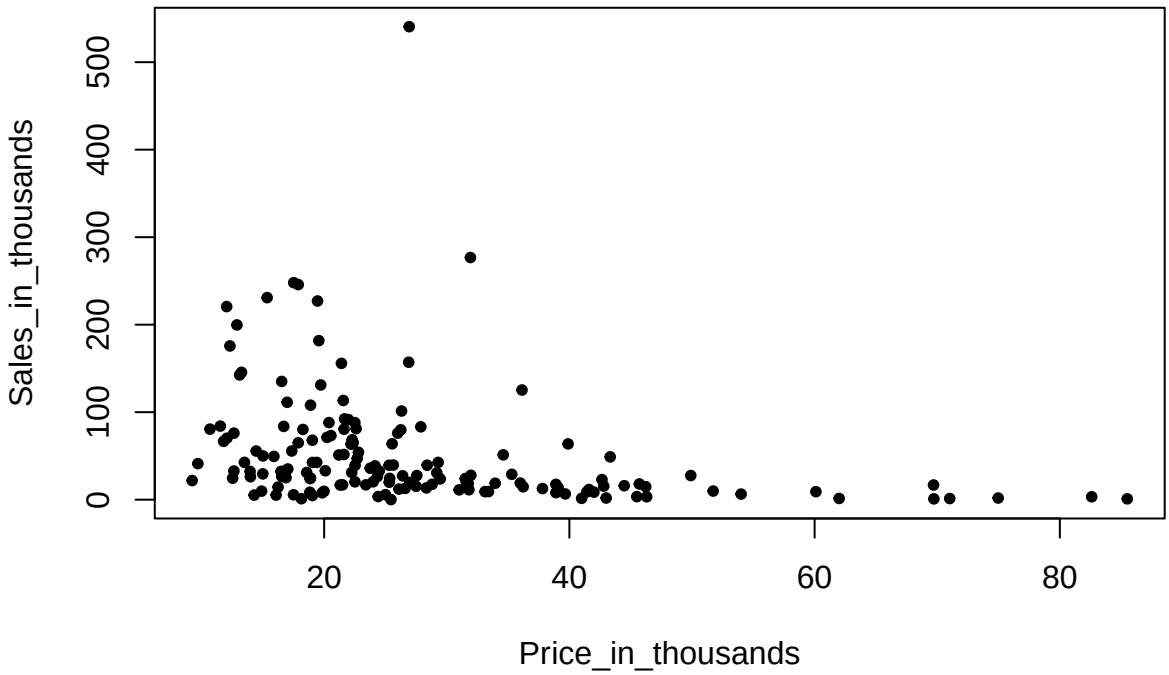
```
##     Manufacturer Model
## 100   Mercedes-B CL500
```

```
#get brand and model of the car with less value
dados_passenger[which.min(dados_passenger$Price_in_thousands),1:2]
```

```
##    Manufacturer Model
## 27    Chevrolet Metro
```

Even though Porsche and Ford not appear in the unless/most expensive cars, we suspect that exists any relation between sales and pricing of the car. For check this, we will generate a scatter graph sales versus price. We perceptive that in general cars with lowest price has more sales that cars with biggest price. This has done by code below.

```
attach(dados)
plot(Price_in_thousands,Sales_in_thousands,pch = 20)
```



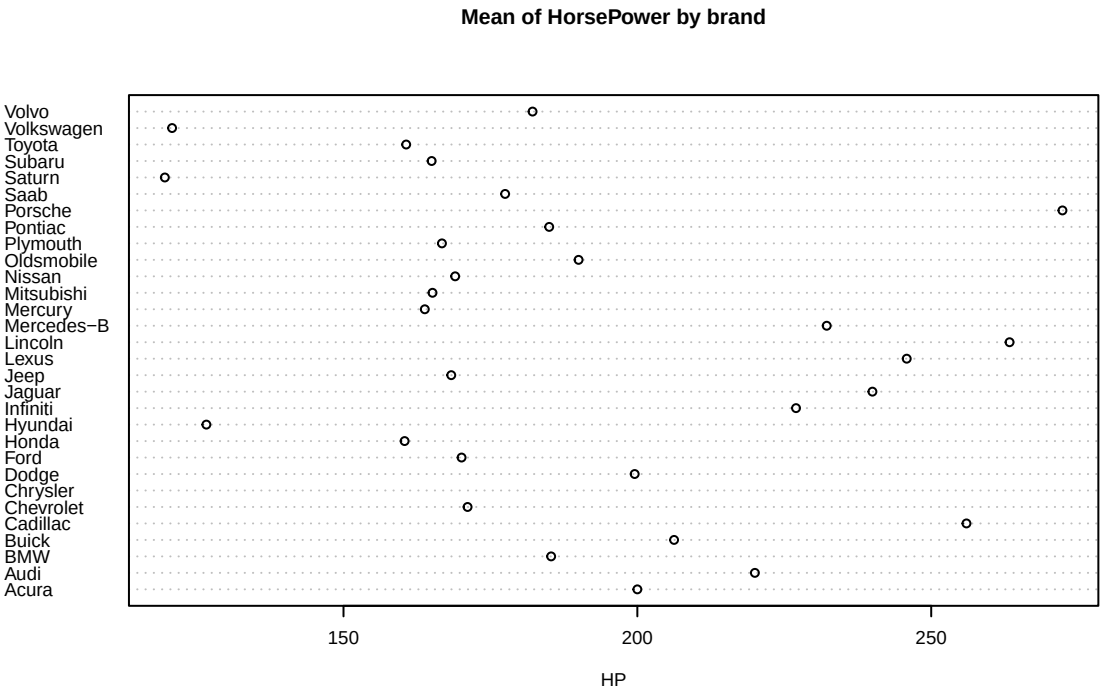
Now, our mission is get a dotted graph of the mean of the horsepower by brand. The analysis show us that Porsche and Lincon are the brands that has biggest mean horse power, whereas Saturn and Volksvagen are the brands that has a lowest mean horse power. We got this table by the code below.

```
df_horse_brand = dados %>% group_by(Manufacturer) %>%
summarise(mean_horse = mean(Horsepower))

attach(df_horse_brand)
```

```
## The following object is masked from dados:
##
##     Manufacturer
```

```
#graph dotchart
dotchart(mean_horse, labels = Manufacturer, cex = 0.55,
main = 'Mean of HorsePower by brand', xlab = 'HP')
```



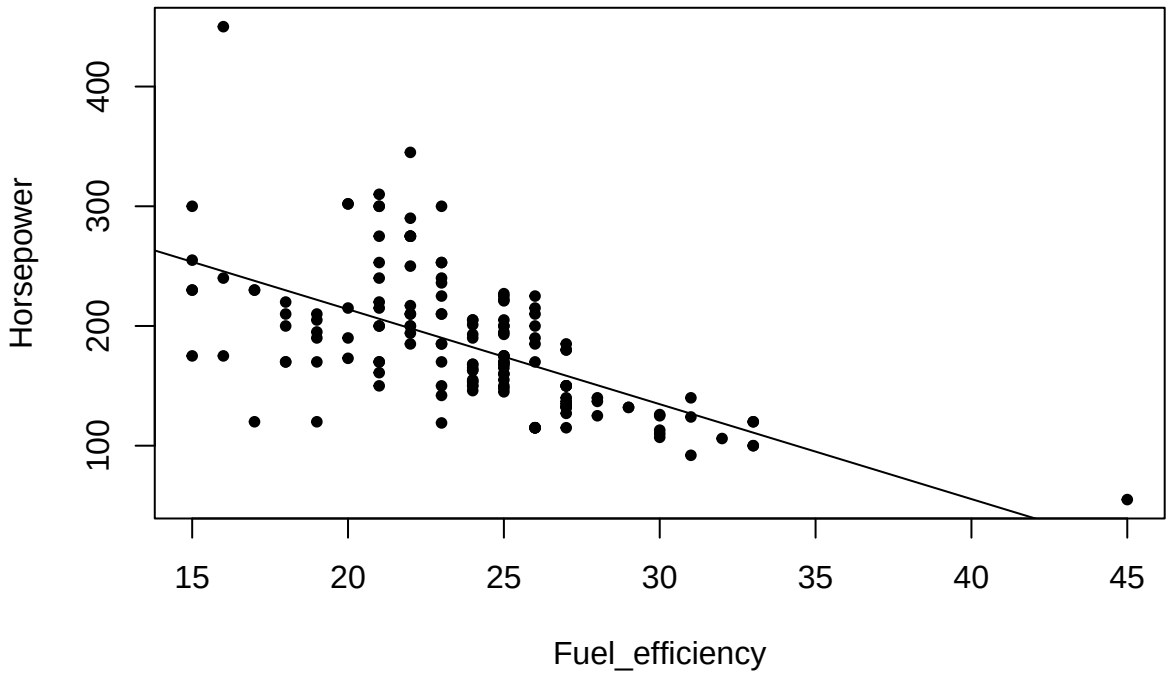
Now, we will inspect if exists any relation between horsepower and fuel efficiency. We will generate a graph contemplating all cars about this two aspects and the regression line about this points. The results show us that in general cars with more fuel efficiency has a loss horsepower.

```
attach(dados)
```

```
## The following object is masked from df_horse_brand:
##
##     Manufacturer
```

```
## The following objects are masked from dados (pos = 4):
##
##     Curb_weight, Engine_size, Fuel_capacity, Fuel_efficiency,
##     Horsepower, Latest_Launch, Length, Manufacturer, Model,
##     Power_perf_factor, Price_in_thousands, Sales_in_thousands,
##     Vehicle_type, Wheelbase, Width, X__year_resale_value
```

```
plot(Fuel_efficiency,Horsepower,pch = 20)
abline(372.464,-7.926)
```



Now, we will investigate the relation between engine size and horsepower, we hope that relation is proportional. In fact, we can perceptive in the graph that how bigger engine size, bigger horsepower, that is make sense.

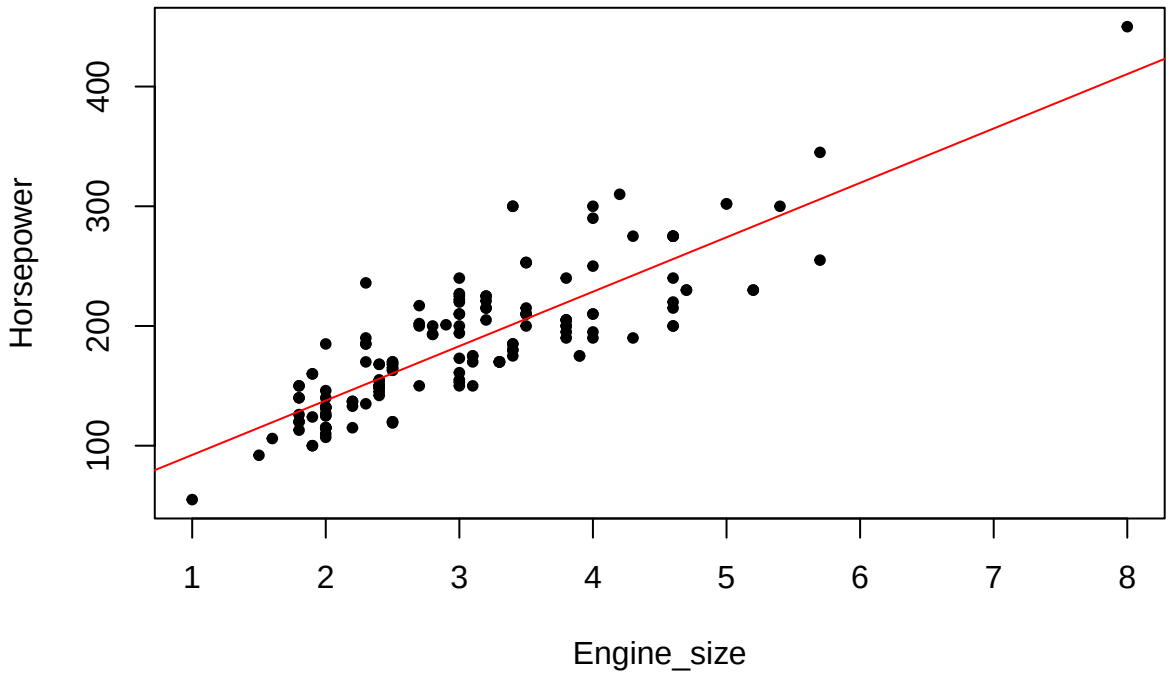
```
attach(dados)
```

```
## The following objects are masked from dados (pos = 3):
##
##     Curb_weight, Engine_size, Fuel_capacity, Fuel_efficiency,
##     Horsepower, Latest_Launch, Length, Manufacturer, Model,
##     Power_perf_factor, Price_in_thousands, Sales_in_thousands,
##     Vehicle_type, Wheelbase, Width, X__year_resale_value
```

```
## The following object is masked from df_horse_brand:
##
##     Manufacturer
```

```
## The following objects are masked from dados (pos = 5):
##
##     Curb_weight, Engine_size, Fuel_capacity, Fuel_efficiency,
##     Horsepower, Latest_Launch, Length, Manufacturer, Model,
##     Power_perf_factor, Price_in_thousands, Sales_in_thousands,
##     Vehicle_type, Wheelbase, Width, X__year_resale_value
```

```
plot(Engine_size,Horsepower,pch = 20)
abline(lm(Horsepower ~ Engine_size),col = 'red')
```



Finally, we will find the top 5 cars in power performance fator.

```
dados_power = dados %>% arrange(desc(Power_perf_factor))
dados_power[1:5,1:2]
```

```
##   Manufacturer          Model
## 1        Dodge          Viper
## 2    Chevrolet       Corvette
## 3   Mercedes-B          CL500
## 4   Mercedes-B       SL-Class
## 5      Porsche Carrera Cabrio
```