```r
data = read.csv("/Users/srikanthgembali/Downloads/car data.csv")
head(data)
```

```
##        Car_Name Year Selling_Price Present_Price Kms_Driven Fuel_Type
## 1          ritz 2014          3.35          5.59      27000    Petrol
## 2           sx4 2013          4.75          9.54      43000    Diesel
## 3          ciaz 2017          7.25          9.85       6900    Petrol
## 4       wagon r 2011          2.85          4.15       5200    Petrol
## 5         swift 2014          4.60          6.87      42450    Diesel
## 6 vitara brezza 2018          9.25          9.83       2071    Diesel
##   Seller_Type Transmission Owner
## 1      Dealer       Manual     0
## 2      Dealer       Manual     0
## 3      Dealer       Manual     0
## 4      Dealer       Manual     0
## 5      Dealer       Manual     0
## 6      Dealer       Manual     0
```

Data source - https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardekho?resource=download&select=car+data.csv

```r
summary(data) #Data Summary
```

```
##    Car_Name              Year      Selling_Price    Present_Price
##  Length:301         Min.   :2003   Min.   : 0.100   Min.   : 0.320
##  Class :character   1st Qu.:2012   1st Qu.: 0.900   1st Qu.: 1.200
##  Mode  :character   Median :2014   Median : 3.600   Median : 6.400
##                     Mean   :2014   Mean   : 4.661   Mean   : 7.628
##                     3rd Qu.:2016   3rd Qu.: 6.000   3rd Qu.: 9.900
##                     Max.   :2018   Max.   :35.000   Max.   :92.600
##    Kms_Driven      Fuel_Type         Seller_Type        Transmission
##  Min.   :   500   Length:301         Length:301         Length:301
##  1st Qu.: 15000   Class :character   Class :character   Class :character
##  Median : 32000   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 36947
##  3rd Qu.: 48767
##  Max.   :500000
##      Owner
##  Min.   :0.00000
##  1st Qu.:0.00000
##  Median :0.00000
##  Mean   :0.04319
##  3rd Qu.:0.00000
##  Max.   :3.00000
```

```r
str(data) #Data Structure
```

```
## 'data.frame':    301 obs. of  9 variables:
##  $ Car_Name     : chr  "ritz" "sx4" "ciaz" "wagon r" ...
##  $ Year         : int  2014 2013 2017 2011 2014 2018 2015 2015 2016 2015 ...
##  $ Selling_Price: num  3.35 4.75 7.25 2.85 4.6 9.25 6.75 6.5 8.75 7.45 ...
##  $ Present_Price: num  5.59 9.54 9.85 4.15 6.87 9.83 8.12 8.61 8.89 8.92 ...
```

```
##  $ Kms_Driven   : int  27000 43000 6900 5200 42450 2071 18796 33429 20273 42367 ...
##  $ Fuel_Type    : chr  "Petrol" "Diesel" "Petrol" "Petrol" ...
##  $ Seller_Type  : chr  "Dealer" "Dealer" "Dealer" "Dealer" ...
##  $ Transmission : chr  "Manual" "Manual" "Manual" "Manual" ...
##  $ Owner        : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
#Average Selling Price (in thousands)

avg_selling_price = mean(data$Selling_Price)
avg_selling_price
```

```
## [1] 4.661296
```

```
#Range of KMS driven

kms_driven_range = range(data$Kms_Driven)
kms_driven_range
```

```
## [1]    500 500000
```

```
#Fuel Types

data$Fuel_Type = as.factor(data$Fuel_Type)

types = levels(data$Fuel_Type)
types
```

```
## [1] "CNG"    "Diesel" "Petrol"
```

```
#Seller Types

data$Seller_Type = as.factor(data$Seller_Type)

seller_types = levels(data$Seller_Type)
seller_types
```

```
## [1] "Dealer"     "Individual"
```

```
class(data$Transmission)
```

```
## [1] "character"
```

Transformation

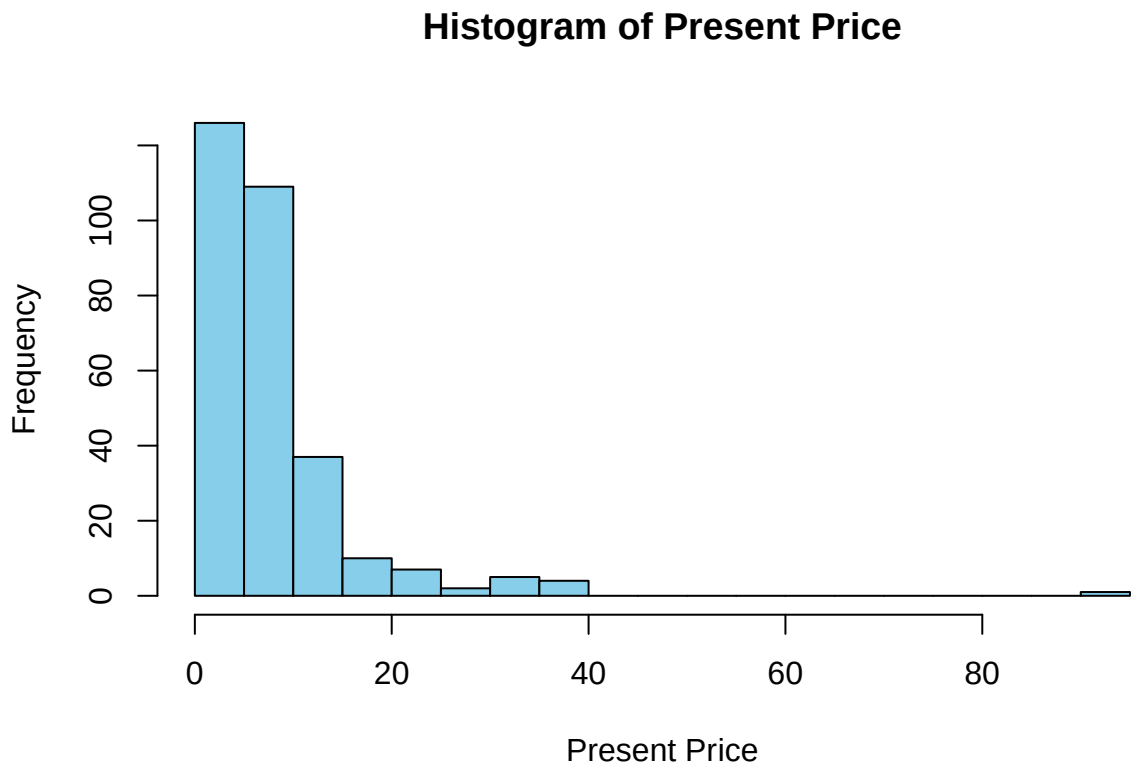
```
class_kms_driven = data$Kms_Driven
class(class_kms_driven)
```

```
## [1] "integer"
```

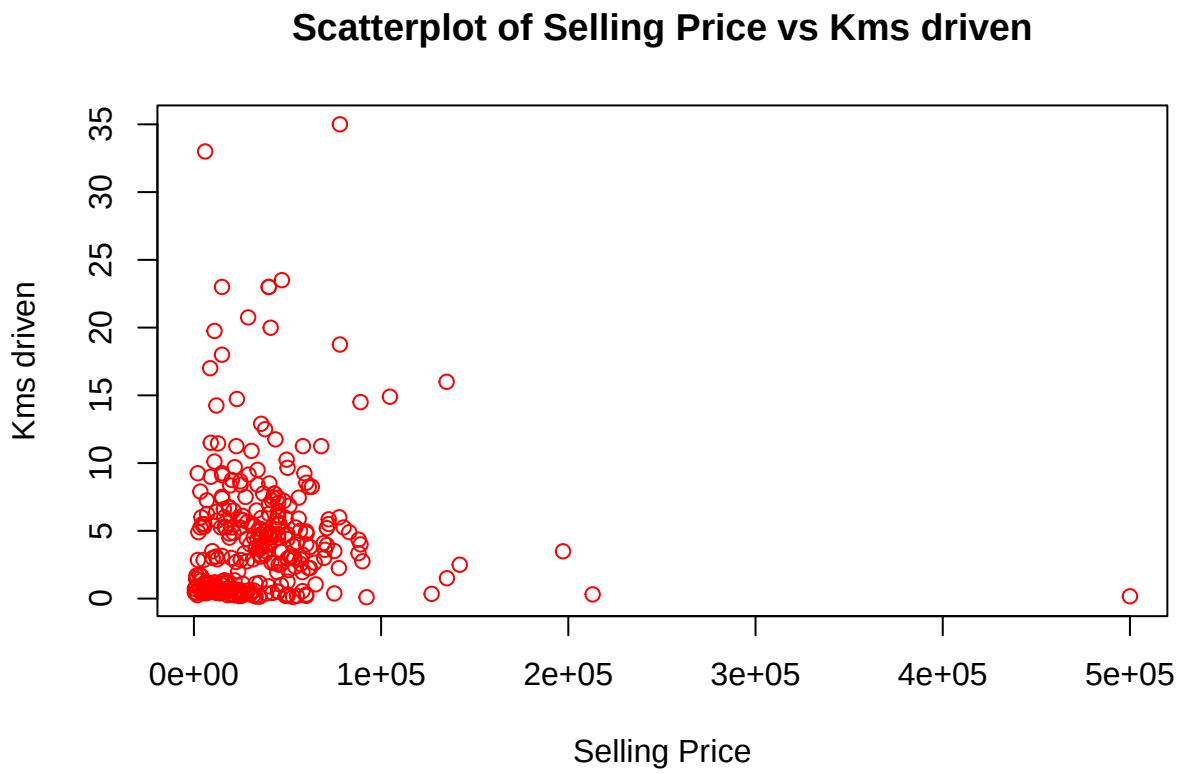
```r
transform_kms_driven = as.numeric(class_kms_driven)
class(transform_kms_driven)
```

```
## [1] "numeric"
```

```r
#Histogram

hist(data$Present_Price, n=20,
     main = "Histogram of Present Price",
     col = "skyblue",
     xlab = "Present Price")
```

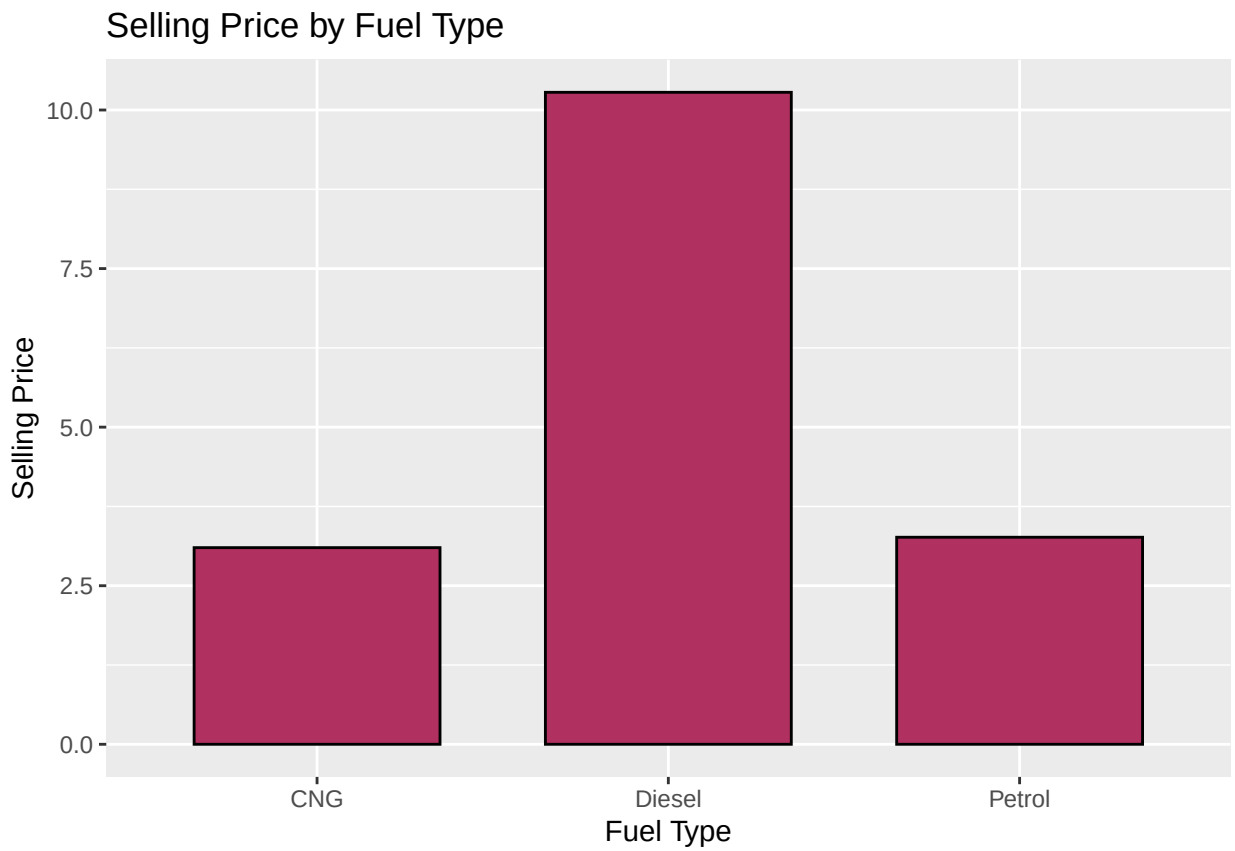



```r
#Scatterplot

plot(data$Kms_Driven, data$Selling_Price,
     main = "Scatterplot of Selling Price vs Kms driven",
     col = "red",
     xlab = "Selling Price",
     ylab = "Kms driven")
```




```
library(ggplot2)

ggplot(data, aes(x = Fuel_Type, y = Selling_Price)) +
  geom_bar(stat = "summary", fill = "maroon", color = "black", width = 0.7) +
  labs(title = "Selling Price by Fuel Type", x = "Fuel Type", y = "Selling Price")
```

```
## No summary function supplied, defaulting to `mean_se()`
```


Selling Price by Fuel Type
Selling Price
10.0
7.5
5.0
2.5
0.0
CNG
Diesel
Petrol
Fuel Type