BA - Assignment – Setting Up R

Gaurav Kudeshia

2023-10-01

Summary

Assignment Overview: This assignment focuses on analyzing the "Carseats" dataset, involving various operations and visualizations.

Steps Taken:

- Library Upload: We imported the "ISLR" library to facilitate our analysis.

- Data set Selection: The data set selected for analysis is "Carseats," which forms the basis of our exploration.

- Summary Analysis: Utilizing the summary function, we delved into the dataset, extracting essential statistics such as mean, median, quartiles, as well as minimum and maximum values.

- Maximum Value Extraction: Employing the MAX function, we determined the maximum value present within the dataset.

- Interquartile Range Calculation: The Interquartile Range (IQR) was calculated, providing insights into the spread of the data.

- Data Visualization: To visually comprehend the relationship between Price and Sales, we created a Scatter plot, offering a graphical representation of the correlation between these attributes.

- The correlation coefficient quantifies the strength and direction of this relationship.The negative correlation value suggests that Sales and Price exhibit an inverse relationship. In other words, as one attribute increases, the other tends to decrease.

Installed ISLR package by going to-> packages -> install option Took ISLR library as an input

```
library(ISLR)
```

#Summarized carseats dataset

```
A <- summary(Carseats)
A
```

```
##      Sales          CompPrice         Income        Advertising
## Min.   : 0.000   Min.   : 77    Min.   : 21.00   Min.   : 0.000
## 1st Qu.: 5.390   1st Qu.:115    1st Qu.: 42.75   1st Qu.: 0.000
## Median : 7.490   Median :125    Median : 69.00   Median : 5.000
## Mean   : 7.496   Mean   :125    Mean   : 68.66   Mean   : 6.635
## 3rd Qu.: 9.320   3rd Qu.:135    3rd Qu.: 91.00   3rd Qu.:12.000
## Max.   :16.270   Max.   :175    Max.   :120.00   Max.   :29.000
##   Population         Price         ShelveLoc        Age          Education
## Min.   : 10.0   Min.   : 24.0   Bad   : 96   Min.   :25.00   Min.   :10.0
## 1st Qu.:139.0   1st Qu.:100.0   Good  : 85   1st Qu.:39.75   1st Qu.:12.0
## Median :272.0   Median :117.0   Medium:219   Median :54.50   Median :14.0
## Mean   :264.8   Mean   :115.8                Mean   :53.32   Mean   :13.9
## 3rd Qu.:398.5   3rd Qu.:131.0                3rd Qu.:66.00   3rd Qu.:16.0
## Max.   :509.0   Max.   :191.0                Max.   :80.00   Max.   :18.0
## Urban         US
## No :118   No :142
## Yes:282   Yes:258
##
##
##
##
```

#Determined the total count of rows present in the carseats dataset

```
B <- nrow(Carseats)
B
```

```
## [1] 400
```

#Got the maximum value of the advertising attribute

```
C <- max(Carseats$Advertising)
C
```

```
## [1] 29
```

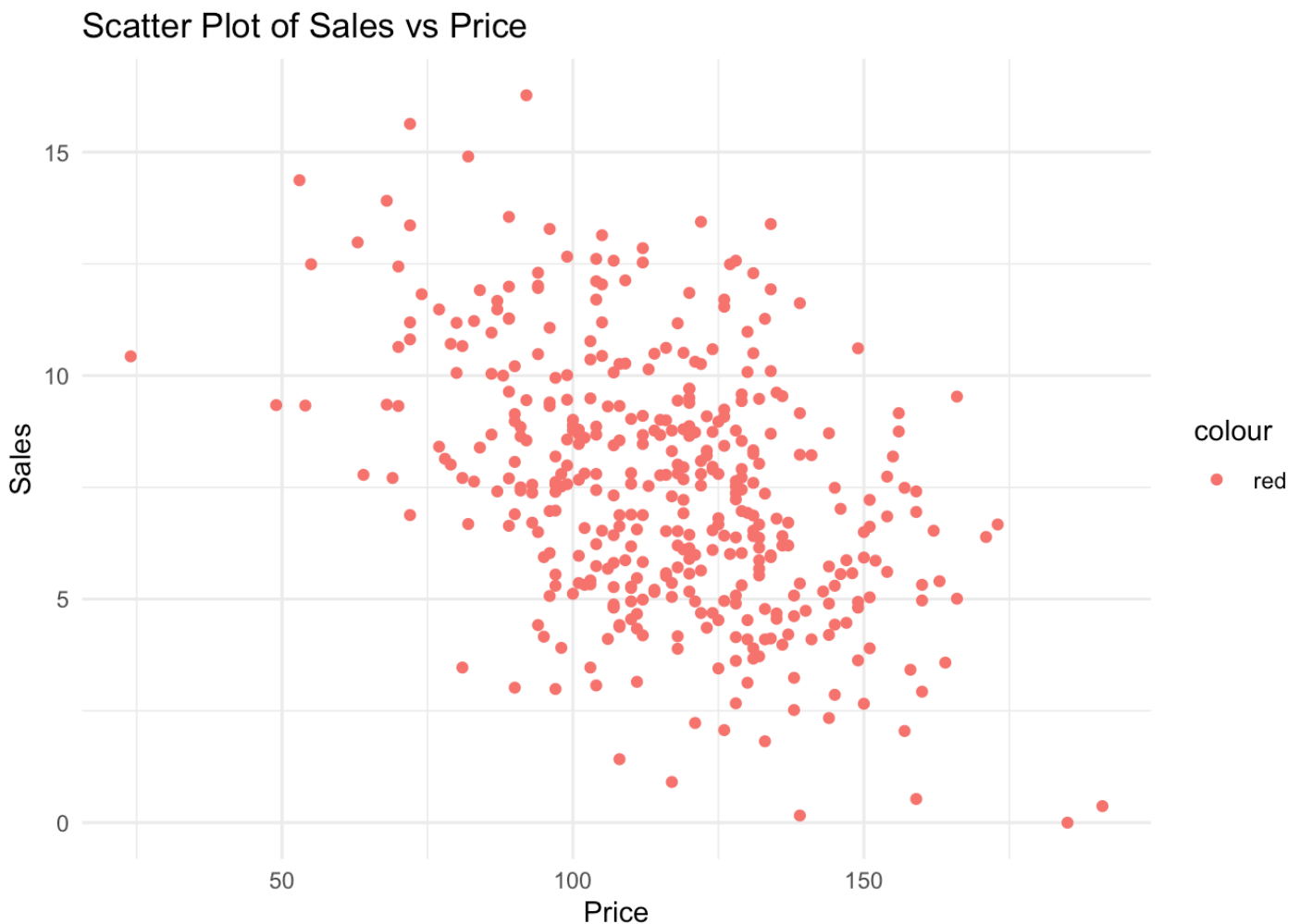#Calculated IQR of Price attribute

```
D <- IQR(Carseats$Price)
D
```

```
## [1] 31
```

#Loaded ggplot2

```
library(ggplot2)
```

Created an scatter plot of Sales against Price

```
ggplot(data = Carseats,aes(x=Price,y=Sales,col="red"))+
  geom_point()+
  labs(x = "Price", y = "Sales", title = "Scatter Plot of Sales vs Price")+
  theme_minimal()
```

Scatter Plot of Sales vs Price

#Correlation between Sales & Price

```
correlation <- cor(Carseats$Price, Carseats$Sales)
correlation
```

```
## [1] -0.4449507
```

The negative correlation value suggests that Sales and Price exhibit an inverse relationship. In other words, as one attribute increases, the other tends to decrease. This negative correlation indicates a reverse linear relationship between these attributes. When Sales rises, Price tends to fall, and vice versa.