IT2120 - Probability and Statistics

Lab Sheet 04

IT24103279

Dushmanthi W. D. H.

Exercise

1. Import the dataset ('Exercise.txt') into R and store it in a data frame called" branch data".

```
setwd("C:\\Users\\IT24103279\\Desktop\\IT24103279")

branch_data <-read.csv("Exercise.txt")
head(branch_data)
```

```
>
> branch_data <-read.csv("Exercise.txt")
> head(branch_data)
  Branch Sales_X1 Advertising_X2 Years_X3
1      1      3.4            120        4
2      2      4.1            150        7
3      3      2.8             90        3
4      4      5.0            200       10
5      5      3.7            110        5
6      6      4.5            175        6
>
```

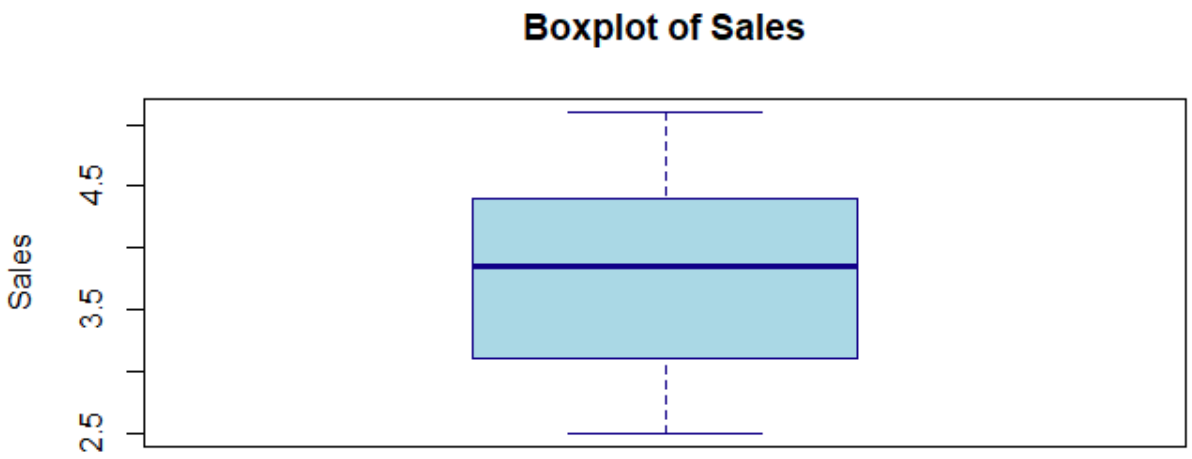
2. Identify the variable type and scale of measurement for each variable.

```
str(branch_data)
```

```
> str(branch_data)
'data.frame':   30 obs. of  4 variables:
 $ Branch        : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Sales_X1      : num  3.4 4.1 2.8 5 3.7 4.5 3 4.9 3.2 2.5 ...
 $ Advertising_X2: int  120 150 90 200 110 175 95 185 105 80 ...
 $ Years_X3      : int  4 7 3 10 5 6 2 9 4 1 ...
>
```

3. Obtain boxplot for sales and interpret the shape of the sales distribution.

```
boxplot(branch_data$Sales_X1,
        main = "Boxplot of Sales",
        ylab = "Sales",
        col = "lightblue",
        border = "darkblue")
```



4. Calculate the five number summary and IQR for advertising variable.

```
summary(branch_data$Advertising)

IQR_advertising <- IQR(branch_data$Advertising)
IQR_advertising
```

```
> summary(branch_data$Advertising)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   80.0   101.2   132.5   134.8   158.8   210.0
>
> IQR_advertising <- IQR(branch_data$Advertising)
> IQR_advertising
[1] 57.5
>
```

5. Write an R function to find the outliers in a numeric vector and check for outliers in years variables.

```
find_outliers <- function(x) {
  Q1 <- quantile(x, 0.25)
  Q3 <- quantile(x, 0.75)
  IQR_value <- IQR(x)
  lower_bound <- Q1 - 1.5 * IQR_value
  upper_bound <- Q3 + 1.5 * IQR_value
  outliers <- x[x < lower_bound | x > upper_bound]
  return(outliers)
}

outliers_years <- find_outliers(branch_data$Years_X3)
outliers_years
```

```
> outliers_years <- find_outliers(branch_data$Years_X3)
> outliers_years
integer(0)
>
```