Exercise

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

```
1  setwd("C:/Users/IT24103554/Desktop/IT24103554")
2  getwd()
3
4  #1
5  branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")
6
```

3:1 (Top Level)

Console | Terminal | Background Jobs

R 4.2.2 · C:/Users/IT24103554/Desktop/IT24103554/

```
> setwd("C:/Users/IT24103554/Desktop/IT24103554")
> getwd()
[1] "C:/Users/IT24103554/Desktop/IT24103554"
>
> #1
> branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")
>
```

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

```
7  #2
8  head(branch_data)
9  str(branch_data)
10
```

9:17 (Top Level)

Console | Terminal | Background Jobs

R 4.2.2 · C:/Users/IT24103554/Desktop/IT24103554/

```
> #2
> 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
> 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.

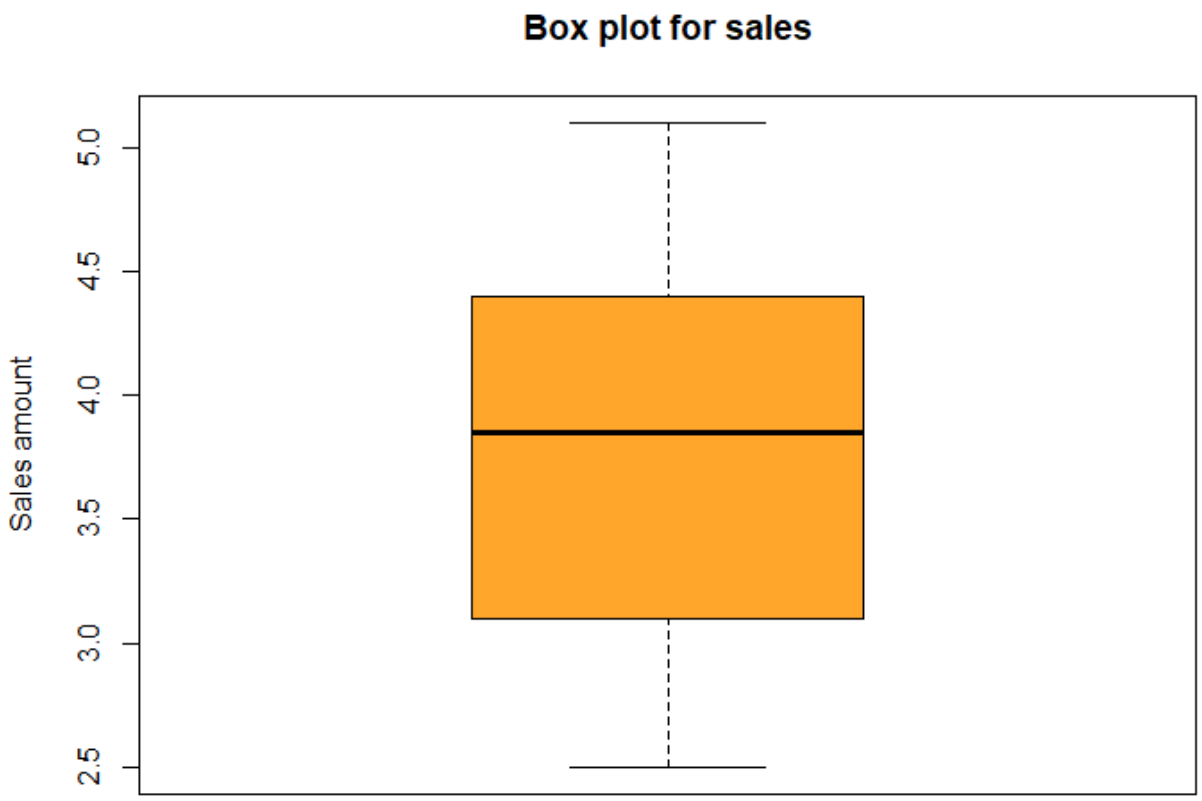
```
11  #3
12  boxplot(branch_data$Sales_X1,
13          main = "Box plot for sales",
14          ylab = "Sales amount",
15          col = "orange",
16          border = "black")
```

14:31 (Top Level)

Console Terminal Background Jobs

R 4.2.2 · C:/Users/IT24103554/Desktop/IT24103554/

```
> #3
> boxplot(branch_data$Sales_X1,
+         main = "Box plot for sales",
+         ylab = "Sales amount",
+         col = "orange",
+         border = "black")
>
```



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

```
17
18  #4
19  summary(branch_data$Advertising_X2)
20  IQR(branch_data$Advertising_X2)
```

16:26 (Top Level)

Console | Terminal | Background Jobs

R 4.2.2 · C:/Users/IT24103554/Desktop/IT24103554/

```
> #4
> summary(branch_data$Advertising_X2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   80.0   101.2   132.5   134.8   158.8   210.0
> IQR(branch_data$Advertising_X2)
[1] 57.5
>
```

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

```
#5
get.outliers <- function(z){
  q1 <- quantile(z)[2]
  q3 <- quantile(z)[4]

  IQR <- q3 - q1

  UB <- q3 + 1.5 * IQR
  LB <- q1 + 1.5 * IQR

  print(paste("Upper Bound =", UB))
  print(paste("Lower Bound =", LB))

  outliers <- sort(z[z < LB | z > UB])

  if (length(outliers) > 0) {
    print(paste("Outliers:", paste(outliers, collapse = ",")))
  } else {
    print("No outliers detected.")
  }
}
get.outliers(branch_data$Years_X3)
```

```
> #5
> get.outliers <- function(z){
+   q1 <- quantile(z)[2]
+   q3 <- quantile(z)[4]
+
+   IQR <- q3 - q1
+
+   UB <- q3 + 1.5 * IQR
+   LB <- q1 + 1.5 * IQR
+
+   print(paste("Upper Bound =", UB))
+   print(paste("Lower Bound =", LB))
+
+   outliers <- sort(z[z < LB | z > UB])
+
+   if (length(outliers) > 0) {
+     print(paste("Outliers:", paste(outliers, collapse = ",")))
+   } else {
+     print("No outliers detected.")
+   }
+ }
> get.outliers(branch_data$Years_X3)
[1] "Upper Bound = 14.5"
[1] "Lower Bound = 10"
[1] "Outliers: 1,1,2,2,2,3,3,3,4,4,4,5,5,5,5,6,6,6,6,7,7,7,8,8,9,9"
>
```