Exercise

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

```
getwd()
setwd("C:\\Users\\it24102613\\Desktop\\IT24102613")

#Q1
branch_data <- read.table("Exercise.txt",header = TRUE,sep =",")

```

```
> getwd()
[1] "C:/Users/it24102613/Desktop/IT24102613"
> setwd("C:\\Users\\it24102613\\Desktop\\IT24102613")
>
> #Q1
> branch_data <- read.table("Exercise.txt",header = TRUE,sep =",")
>
```

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

```
#Q2
fix(branch_data)
str(branch_data)
attach(branch_data)

```

Data Editor

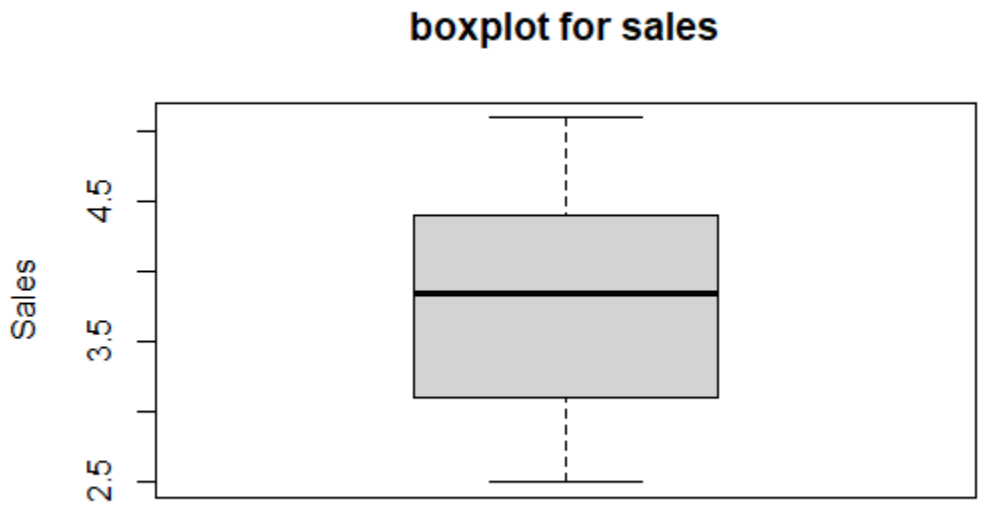
File Edit Help

	Branch	Sales_X1	Advertising_X2	Years_X3	var5	var6	var7
1	1	3.4	120	4			
2	2	4.1	150	7			
3	3	2.8	90	3			
4	4	5	200	10			
5	5	3.7	110	5			
6	6	4.5	175	6			
7	7	3	95	2			
8	8	4.9	185	9			
9	9	3.2	105	4			
10	10	2.5	80	1			
11	11	3.9	130	5			
12	12	4.2	140	7			
13	13	2.7	100	3			
14	14	3.6	125	4			
15	15	4.8	190	8			
16	16	3.3	115	5			
17	17	4	135	6			
18	18	5.1	210	12			
19	19	3.8	145	6			

```
> #Q2
> fix(branch_data)
> str(branch_data)
'data.frame':   30 obs. of  4 variables:
 $ Branch        : num  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: num  120 150 90 200 110 175 95 185 105 80 ...
 $ Years_X3      : num  4 7 3 10 5 6 2 9 4 1 ...
> attach(branch_data)
```

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

```
boxplot(branch_data$Sales_X1,main="boxplot for sales",
        ylab = "Sales"
        )
```



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

```
16  #Q4
17  #five number summary
18  summary(Advertising_X2)
19  #IQR
20  IQR(Advertising_X2)
21
```

```
> #Q4
> #five number summary
> summary(Advertising_X2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   80.0   101.2   132.5   134.8   158.8   210.0
> #IQR
> IQR(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.

```
#Q5
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))
  print(paste("Outliers: ",paste(sort(z[z<lb | z>ub],
                                      collapse = ","))))
}
get.outliers(Years_X3)
```

```
+ }
> get.outliers(Years_X3)
[1] "Upper bound =  14.5"
[1] "Lower bound =  -3.5"
```