IT24102724 – Lab4

```
1   #1
2   #set the working directry
3   setwd("C:\\Users\\IT24102724\\Desktop\\IT24102724\\Lab 04-20250822")
4
5
6   #read the text file with header
7   branch_data <- read.table("Exercise.txt",head =TRUE,sep = ",")
8
9   #view the file in a separate window
10  fix(branch_data)
11  |
12  #print the first rows
13  head(branch_data)
14
15
16  #Attach the file into R
17  attach(branch_data)
18
19  #2
20  #structure of the data
21  str(branch_data)
22
23  #Get a summery of the data
24  summary(branch_data)
25
26  #3
27  #boxplot for Sales
28  boxplot(branch_data$Sales_X1, outline = TRUE, outpch = 8, horizontal=TRUE,main= "Sales Distribution")
29
30  #4
31  #five number summery
32  summary(branch_data$Advertising_X2)
33
34  #calculate IQR
35  IQR(branch_data$Advertising_X2)
36
37
11:1    (Top Level)                                                                              R Script
```

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			

```
#print the first rows
head(branch_data)
```

```
> 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
#structure of the data
str(branch_data)

#Get a summery of the data
summary(branch_data)
```

```
> #2
> #structure of the 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 ...
> #Get a summery of the data
> summary(branch_data)
     Branch          Sales_X1      Advertising_X2      Years_X3
 Min.   : 1.00   Min.   :2.500   Min.   : 80.0   Min.   : 1.00
 1st Qu.: 8.25   1st Qu.:3.125   1st Qu.:101.2   1st Qu.: 3.25
 Median :15.50   Median :3.850   Median :132.5   Median : 5.50
 Mean   :15.50   Mean   :3.790   Mean   :134.8   Mean   : 5.70
 3rd Qu.:22.75   3rd Qu.:4.375   3rd Qu.:158.8   3rd Qu.: 7.75
 Max.   :30.00   Max.   :5.100   Max.   :210.0   Max.   :12.00
> |
```

```
#3
#boxplot for Sales
boxplot(branch_data$Sales_X1, outline = TRUE, outpch = 8, horizontal=TRUE,main= "Sales Distribution")
```

Sales Distribution

```
2.5        3.0        3.5        4.0        4.5        5.0
```
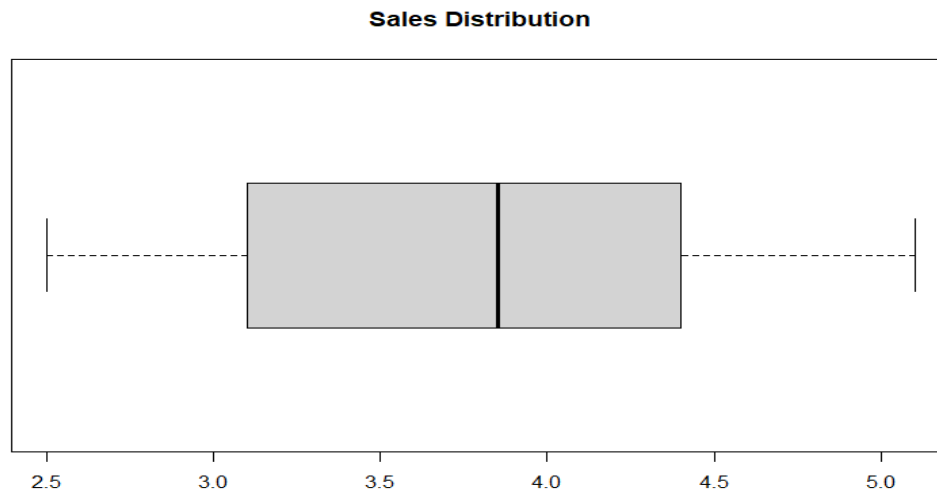
```
#4
#five number summery
summary(branch_data$Advertising_X2)

#calculate IQR
IQR(branch_data$Advertising_X2)
```

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

```
#5
#Function to find outliers
outliers <- function(years){
  Q1 <- quantile(years)[2]
  Q3 <- quantile(years)[4]
  iqr <- Q3-Q1

  lb <- Q1 - 1.5*iqr
  ub <- Q3 + 1.5*iqr

  outliers <- years[years < lb | years > ub]
  #sort the outlier
  outliers = sort(outliers)

  print(paste("Upper Bound : " , ub))
  print(paste("Lower Bound : " , lb))
  print(paste("IQR : ",iqr))
  print(paste("outliers: ", paste(outliers, collapse=",")))
}

#get outliers for the 'years' variable
outliers(branch_data$Years_X3)
```

```
> #5
> #Function to find outliers
> outliers <- function(years){
+   Q1 <- quantile(years)[2]
+   Q3 <- quantile(years)[4]
+   iqr <- Q3-Q1
+
+   lb <- Q1 - 1.5*iqr
+   ub <- Q3 + 1.5*iqr
+
+   outliers <- years[years < lb | years > ub]
+   #sort the outlier
+   outliers = sort(outliers)
+
+   print(paste("Upper Bound : " , ub))
+   print(paste("Lower Bound : " , lb))
+   print(paste("IQR : ",iqr))
+   print(paste("outliers: ", paste(outliers, collapse=",")))
+ }
> #get outliers for the 'years' variable
> outliers(branch_data$Years_X3)
[1] "Upper Bound :  14.5"
[1] "Lower Bound :  -3.5"
[1] "IQR :  4.5"
[1] "outliers:  "
```