Sri Lanka Institute of Information Technology

Lab Submission
Lab Sheet 04

IT24101219
Piyarathna S.G.D.V
Probability and Statistics| IT2120

B.Sc.(Hons) in Information Technology

Question

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

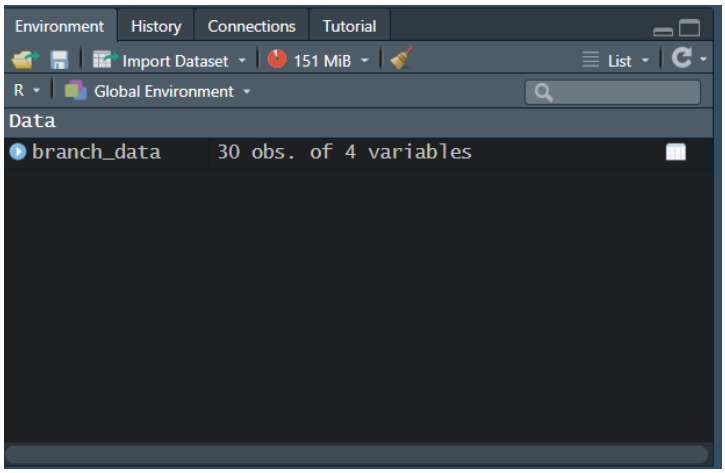
Code

```r
setwd("C:\\Users\\dinin\\OneDrive - Sri Lanka Institute of Information Technology\\Desktop\\IT24101219")
getwd()
branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")

```

Output(s)

```
Console  Terminal  Background Jobs
R 4.5.1 · C:/Users/dinin/OneDrive - Sri Lanka Institute of Information Technology/Desktop/IT24101219/
> setwd("C:\\Users\\dinin\\OneDrive - Sri Lanka Institute of Information Technology\\Desktop\\IT24101219")
> getwd()
[1] "C:/Users/dinin/OneDrive - Sri Lanka Institute of Information Technology/Desktop/IT24101219"
> branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")
>
```



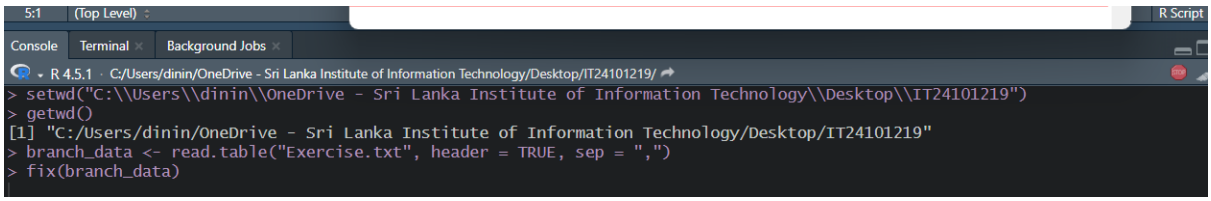
Question

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

Code

```r
fix(branch_data)
typeof(branch_data)
```

Output(s)



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			

1. Branch: Integer (Categorical)
2. Sales_X1: Floating Point (Continuous)
3. Advertising_X2: Integer (Categorical)
4. Years_X3: Integer (Categorical)

Question

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

Code

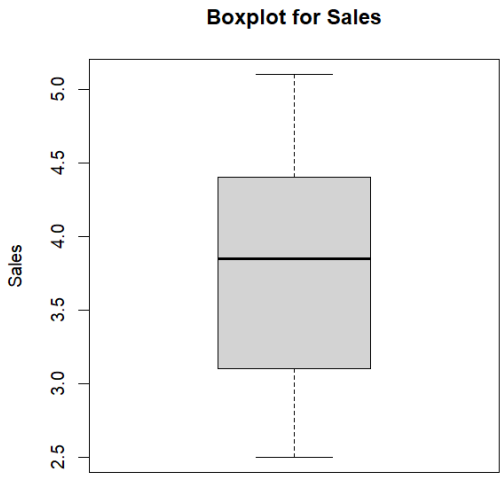
```
attach(branch_data)

boxplot(branch_data['Sales_X1'], main = "Boxplot for Sales", ylab = "Sales")

```

Output(s)

```
> attach(branch_data)
> boxplot(branch_data['Sales_X1'], main = "Boxplot for Sales", ylab = "Sales")
>
```



Interpretation: The sales box plot looks like to have a symmetrical distribution.

Question

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

Code

```
12  summary(Advertising_X2)
13
14  IQR(Advertising_X2)
```

Output(s)

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

Question

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

Code

```r
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)
```

Output(s)

```
> 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)
[1] "Upper Bound =  14.5"
[1] "Lower Bound =  -3.5"
[1] "Outliers: "
>
```