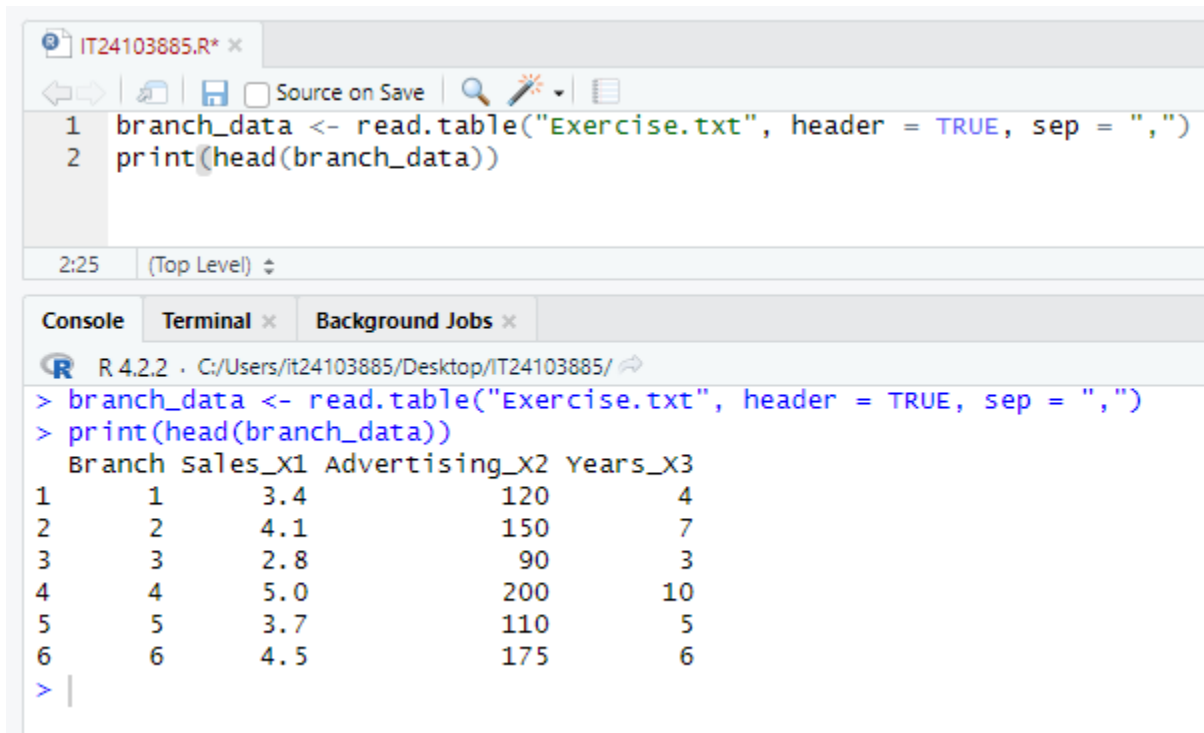


## IT2120 - Lab Sheet 04

### IT24103885 - Senarathna Y.M.C.S

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



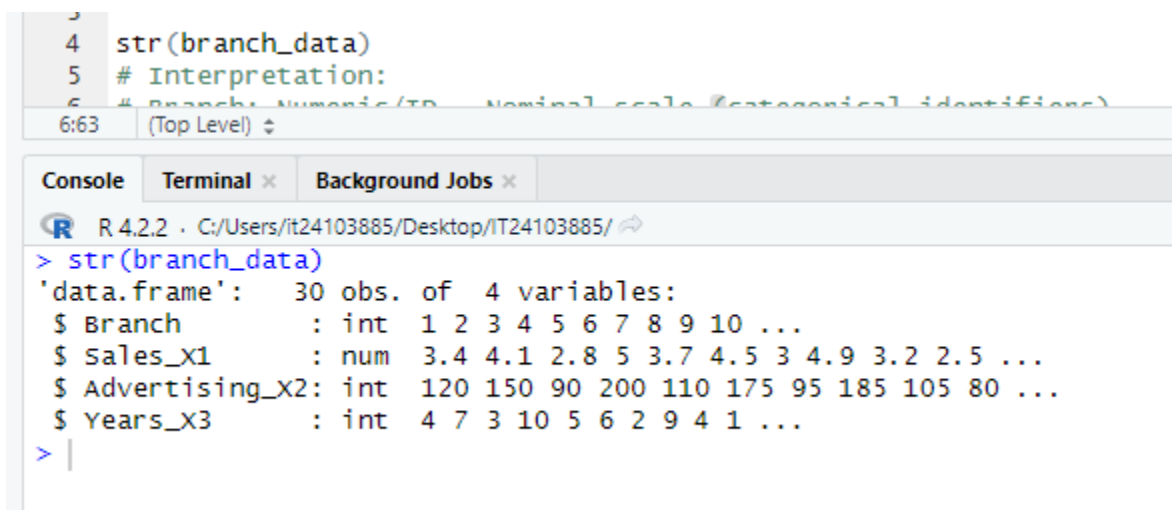
The screenshot shows the R Studio interface. The script editor contains the following code:

```
1 branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")
2 print(head(branch_data))
```

The console output shows the result of the code execution:

```
> branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")
> print(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.



The screenshot shows the R Studio interface. The script editor contains the following code:

```
4 str(branch_data)
5 # Interpretation:
6 # Branch: Numeric (ID) - Nominal scale (categorical identifiers)
```

The console output shows the result of the code execution:

```
> 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.

```
10  
11 boxplot(branch_data$Sales_X1, main = "Boxplot of Sales", ylab = "Sales",  
12         outline = TRUE, outpch = 8, horizontal = TRUE)
```

10:1 (Top Level) ↕

Console

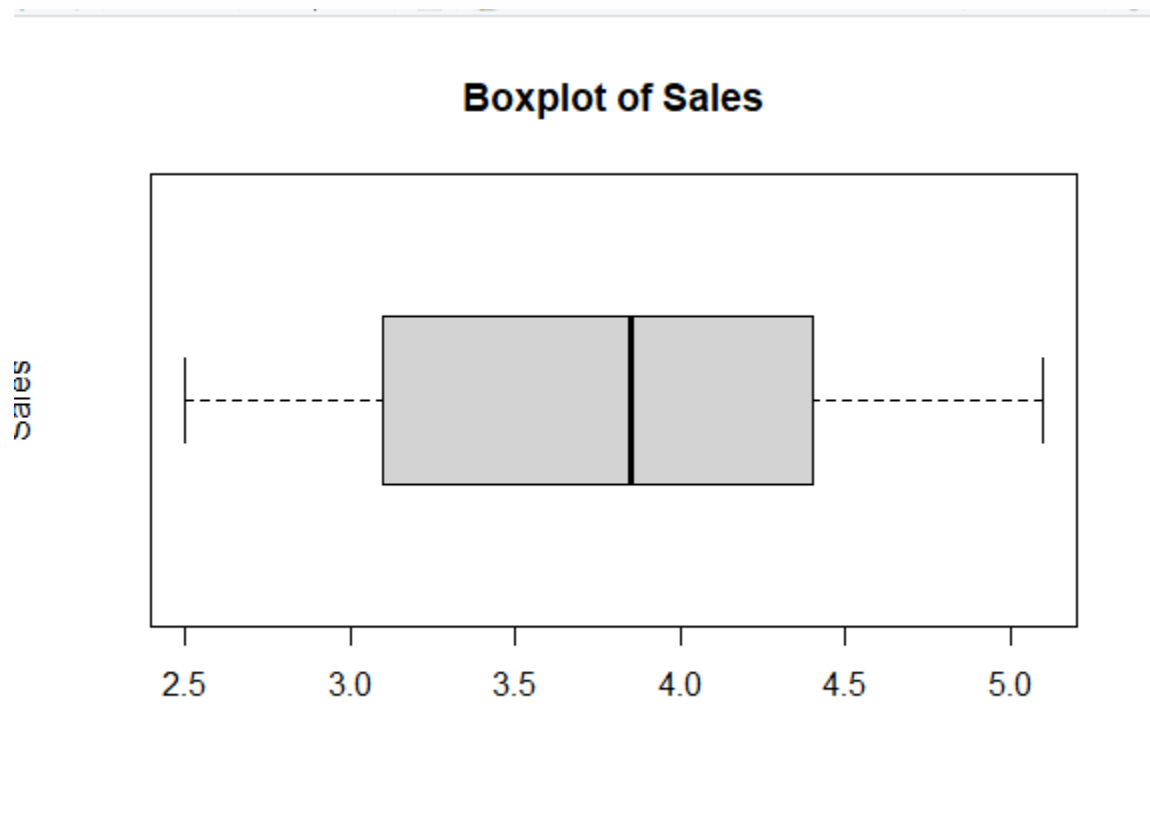
Terminal ×

Background Jobs ×



R 4.2.2 · C:/Users/it24103885/Desktop/IT24103885/ ↗

```
> boxplot(branch_data$Sales_X1, main = "Boxplot of Sales", ylab = "Sales",  
+         outline = TRUE, outpch = 8, horizontal = TRUE)  
> |
```



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

```
13  
14 summary(branch_data$Advertising_X2)  
15 iqr_advertising <- IQR(branch_data$Advertising_X2)  
16 cat("IQR for Advertising:", iqr_advertising, "\n")  
16:51 (Top Level) ⚡
```

Console	Terminal ×	Background Jobs ×
R 4.2.2 · C:/Users/it24103885/Desktop/IT24103885/ ⚡		
<pre>&gt; 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  &gt; iqr_advertising &lt;- IQR(branch_data\$Advertising_X2) &gt; cat("IQR for Advertising:", iqr_advertising, "\n") IQR for Advertising: 57.5 &gt;  </pre>		

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

```
19 find_outliers <- function(x) {  
20   Q1 <- quantile(x, 0.25, na.rm = TRUE)  
21   Q3 <- quantile(x, 0.75, na.rm = TRUE)  
22   IQR_val <- Q3 - Q1  
23   lower_bound <- Q1 - 1.5 * IQR_val  
24   upper_bound <- Q3 + 1.5 * IQR_val  
25   outliers <- x[x < lower_bound | x > upper_bound]  
26   return(outliers)  
27 }  
28  
29 outliers_years <- find_outliers(branch_data$Years_X3)  
30 print("Outliers in Years:")  
31 print(outliers_years)  
32  
29:44 (Top Level) ⚡
```

Console	Terminal ×	Background Jobs ×
R 4.2.2 · C:/Users/it24103885/Desktop/IT24103885/ ⚡		
<pre>&gt; outliers_years &lt;- find_outliers(branch_data\$Years_X3) &gt; print("Outliers in Years:") [1] "Outliers in Years:" &gt; print(outliers_years) integer(0) &gt;  </pre>		