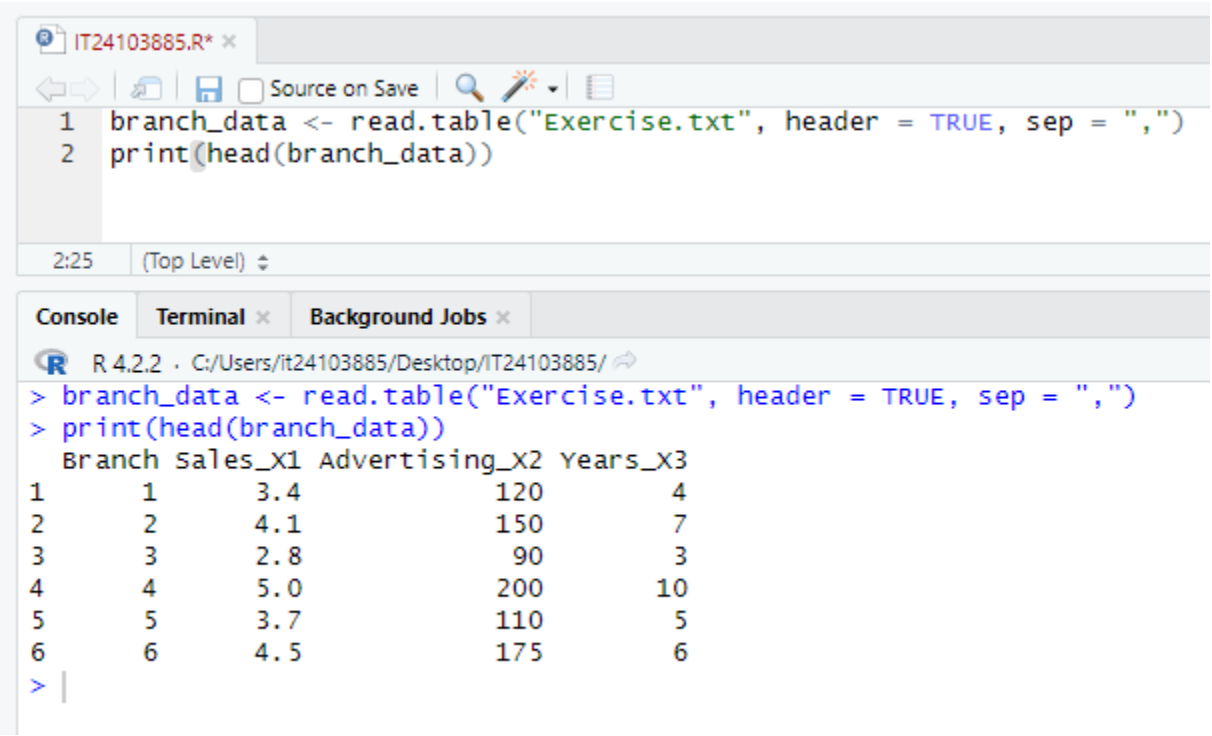
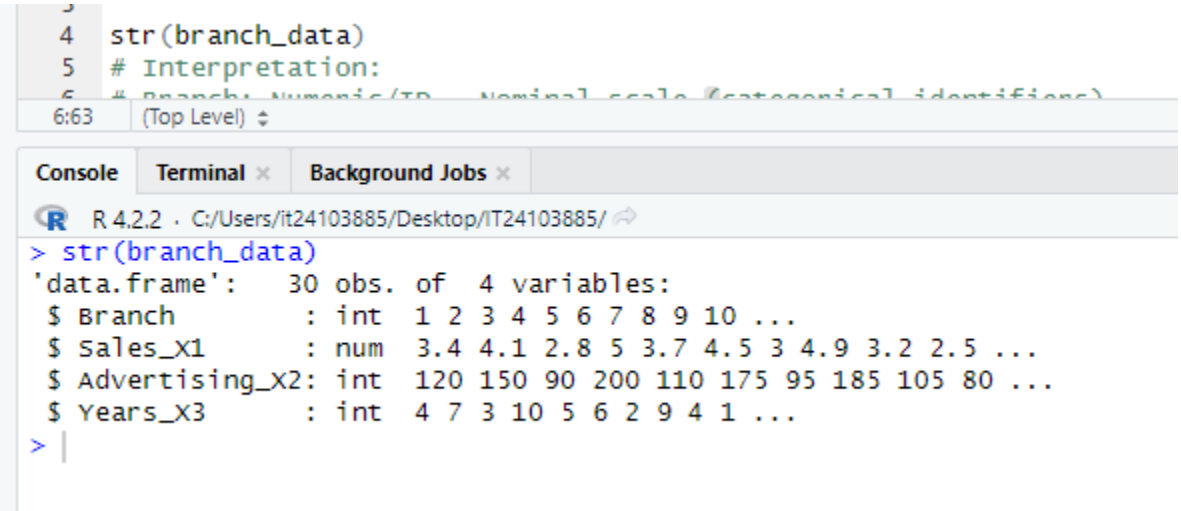
IT2120 - Lab Sheet 04

IT24103885 - Senarathna Y.M.C.S

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



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



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

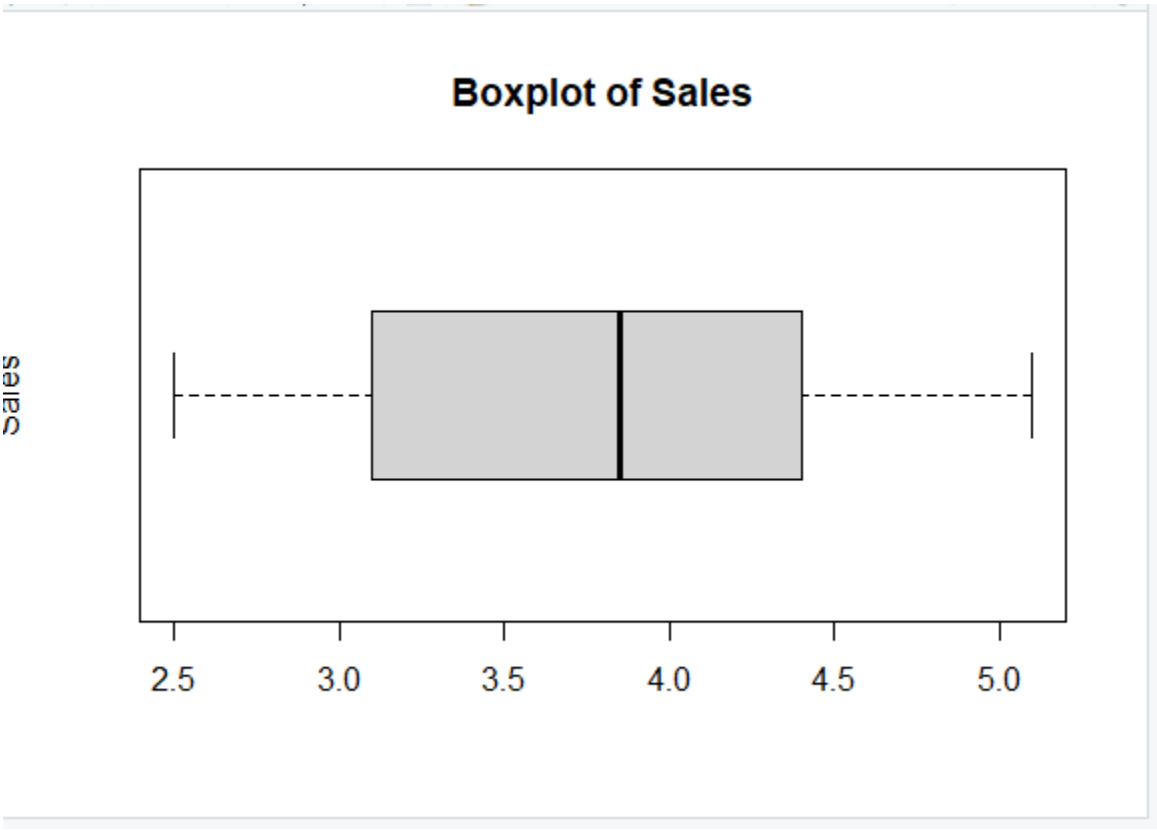
```
10
11  boxplot(branch_data$Sales_X1, main = "Boxplot of Sales", ylab = "Sales",
12          outline = TRUE, outpch = 8, horizontal = TRUE)
```

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.

```
summary(branch_data$Advertising_X2)
iqr_advertising <- IQR(branch_data$Advertising_X2)
cat("IQR for Advertising:", iqr_advertising, "\n")
```

```
> summary(branch_data$Advertising_X2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   80.0   101.2   132.5   134.8   158.8   210.0
> iqr_advertising <- IQR(branch_data$Advertising_X2)
> cat("IQR for Advertising:", iqr_advertising, "\n")
IQR for Advertising: 57.5
>
```

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

```
find_outliers <- function(x) {
  Q1 <- quantile(x, 0.25, na.rm = TRUE)
  Q3 <- quantile(x, 0.75, na.rm = TRUE)
  IQR_val <- Q3 - Q1
  lower_bound <- Q1 - 1.5 * IQR_val
  upper_bound <- Q3 + 1.5 * IQR_val
  outliers <- x[x < lower_bound | x > upper_bound]
  return(outliers)
}

outliers_years <- find_outliers(branch_data$Years_X3)
print("Outliers in Years:")
print(outliers_years)
```

```
> outliers_years <- find_outliers(branch_data$Years_X3)
> print("Outliers in Years:")
[1] "Outliers in Years:"
> print(outliers_years)
integer(0)
>
```