IT2120 - Probability and Statistics

Lab Sheet 03

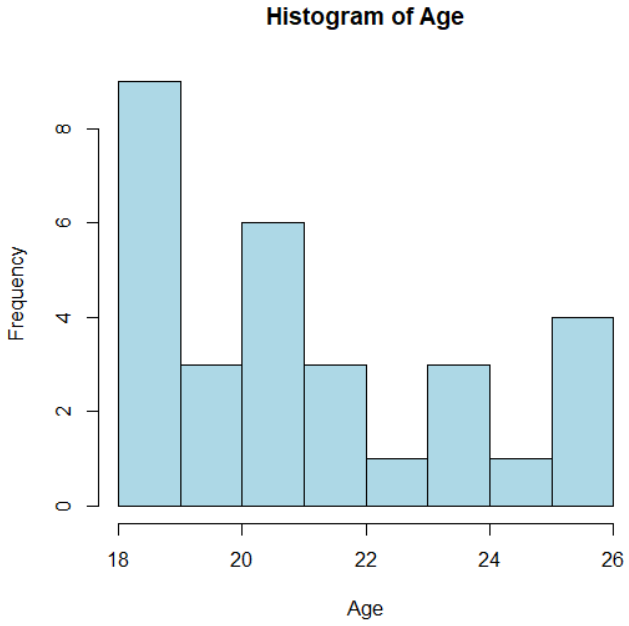
Exercise

1. Import the dataset (‘Exercise.csv‘) into R and store it in a data frame called “student data”.

```
> setwd('C:/Users/it24100387/Desktop/IT24100387')
> getwd()
[1] "C:/Users/it24100387/Desktop/IT24100387"
> # Import the CSV file into a data frame called student_data
> student_data <- read.csv("Exercise.csv", header = TRUE)
```

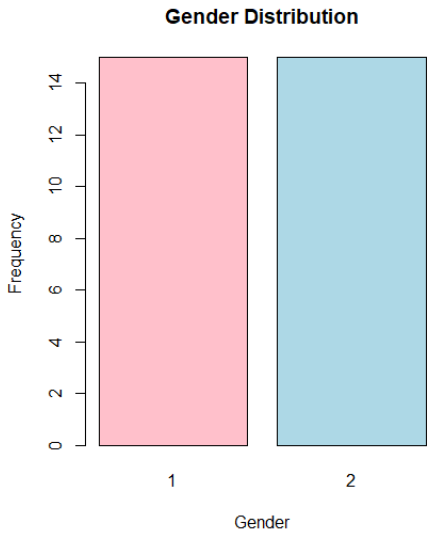
2. Produce the summary statistics and histogram for the variable “X1” (Age).

```
> # Summary statistics for Age
> summary(student_data$X1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   19.00   21.00   21.40   23.75   26.00
> hist(student_data$X1,
+      main = "Histogram of Age",
+      xlab = "Age",
+      col = "lightblue",
+      border = "black")
```



3. Create a bar chart and frequency table for “X2” (Gender).

```
> # Bar chart for Gender
> barplot(table(student_data$X2),
+         main = "Gender Distribution",
+         xlab = "Gender",
+         ylab = "Frequency",
+         col = c("pink", "lightblue"))
```




```
> # Frequency table for Gender
> table(student_data$X2)

 1  2
15 15
>
```

4. How does the age (X1) change according to the accommodation (X3)? Analyze it using a suitable graph and interpret the results. (Note that accommodation has three levels which are type 1, type 2 and type 3)

```
> # Boxplot: Age vs. Accommodation
> boxplot(X1 ~ X3, data = student_data,
+         main = "Age by Accommodation Type",
+         xlab = "Accommodation Type",
+         ylab = "Age",
+         col = c("lightgreen", "lightblue", "lightpin
k"))
>
```