Assignment_4, Div: CSAI-B, Roll No.: 37

Atharva Salitri

25-01-2025

Exercise 1

```
data <- data.frame(
  Name = c("Tendulkar", "Ponting", "Kallis", "Dravid", "Cook"),
  Matches = c(200, 168, 166, 164, 161),
  Innings = c(329, 287, 280, 286, 291),
  HighestScore = c(248, 257, 224, 270, 294),
  Average = c(53.78, 51.85, 55.37, 52.31, 45.35)
)
```

```
# 1. What is the highest score of Tendulkar?
tendulkar_highest_score <- data$HighestScore[data$Name == "Tendulkar"]
cat("Highest score of Tendulkar:", tendulkar_highest_score, "\n")
```

```
## Highest score of Tendulkar: 248
```

```
# 2. Display the name and the average of the player who is having maximum highest score
max_highest_score <- max(data$HighestScore)
player_max_highest <- data[data$HighestScore == max_highest_score, c("Name", "Average")]
cat("Player with the maximum highest score:\n")
```

```
## Player with the maximum highest score:
```

```
print(player_max_highest)
```

```
##   Name Average
## 5 Cook   45.35
```

```
# 3. Display the name, matches, innings, and average of the players having score above 250
players_above_250 <- data[data$HighestScore > 250, c("Name", "Matches", "Innings", "Averag
e")]
cat("Players with a highest score above 250:\n")
```

```
## Players with a highest score above 250:
```

```
print(players_above_250)
```

```
##      Name Matches Innings Average
## 2 Ponting     168     287   51.85
## 4  Dravid     164     286   52.31
## 5    Cook     161     291   45.35
```

```
# 4. Find the row number of the data for which the highest score is equal or greater than 270
row_highest_270 <- which(data$HighestScore >= 270)
cat("Row number(s) with the highest score >= 270:", row_highest_270, "\n")
```

```
## Row number(s) with the highest score >= 270: 4 5
```

```
# 5. Modify Tendulkar's number of matches to 201
data$Matches[data$Name == "Tendulkar"] <- 201
cat("Updated data:\n")
```

```
## Updated data:
```

```
print(data)
```

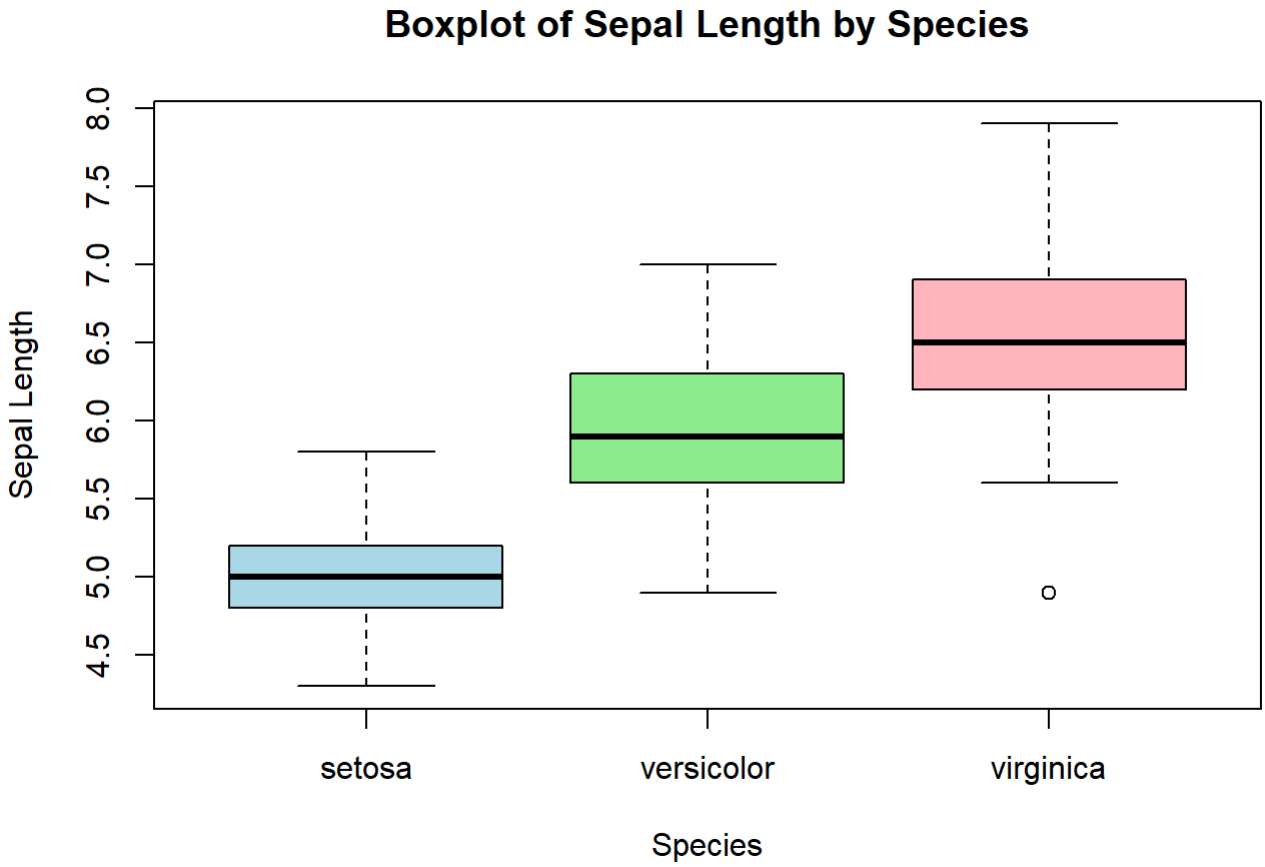
```
##        Name Matches Innings HighestScore Average
## 1 Tendulkar     201     329          248   53.78
## 2   Ponting     168     287          257   51.85
## 3    Kallis     166     280          224   55.37
## 4    Dravid     164     286          270   52.31
## 5      Cook     161     291          294   45.35
```

Exercise 2

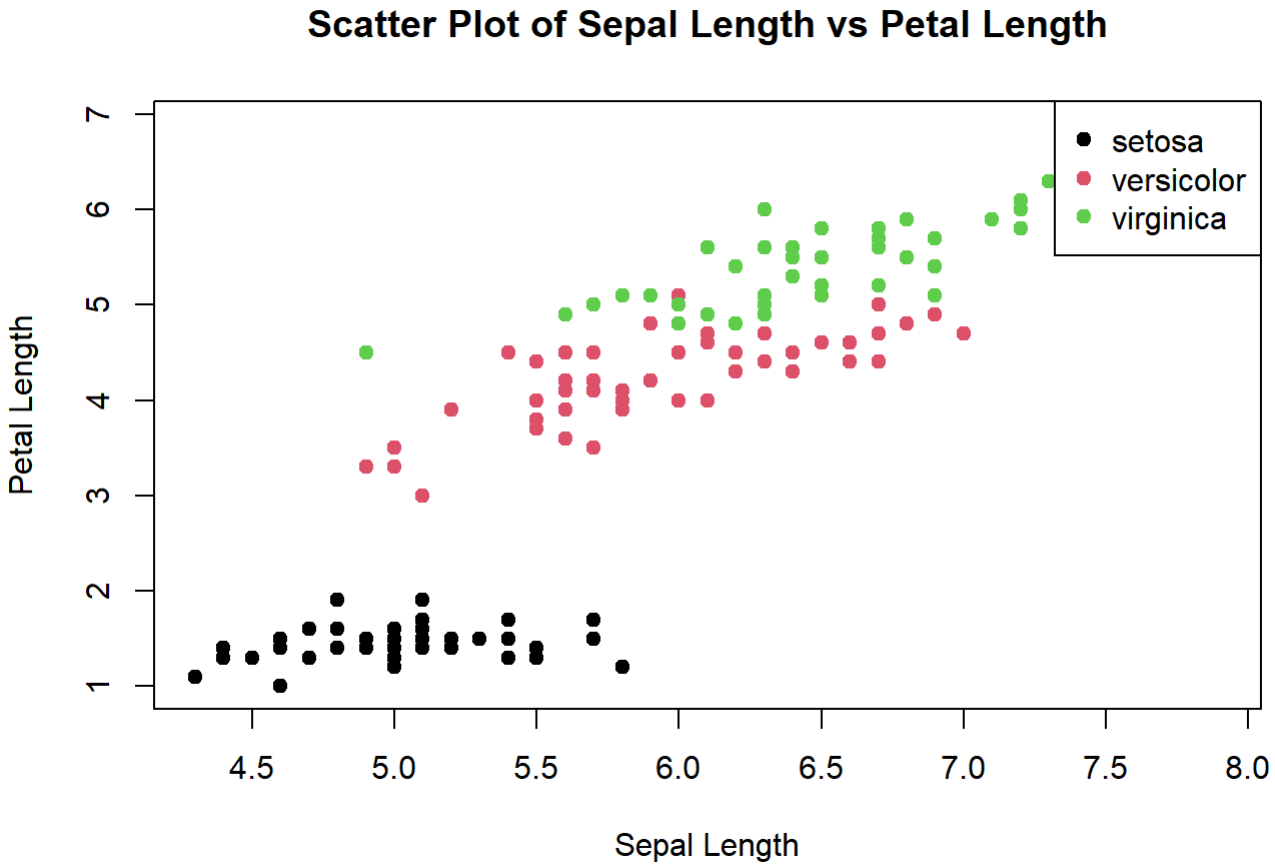
```
data("iris")
data = iris
str(data)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

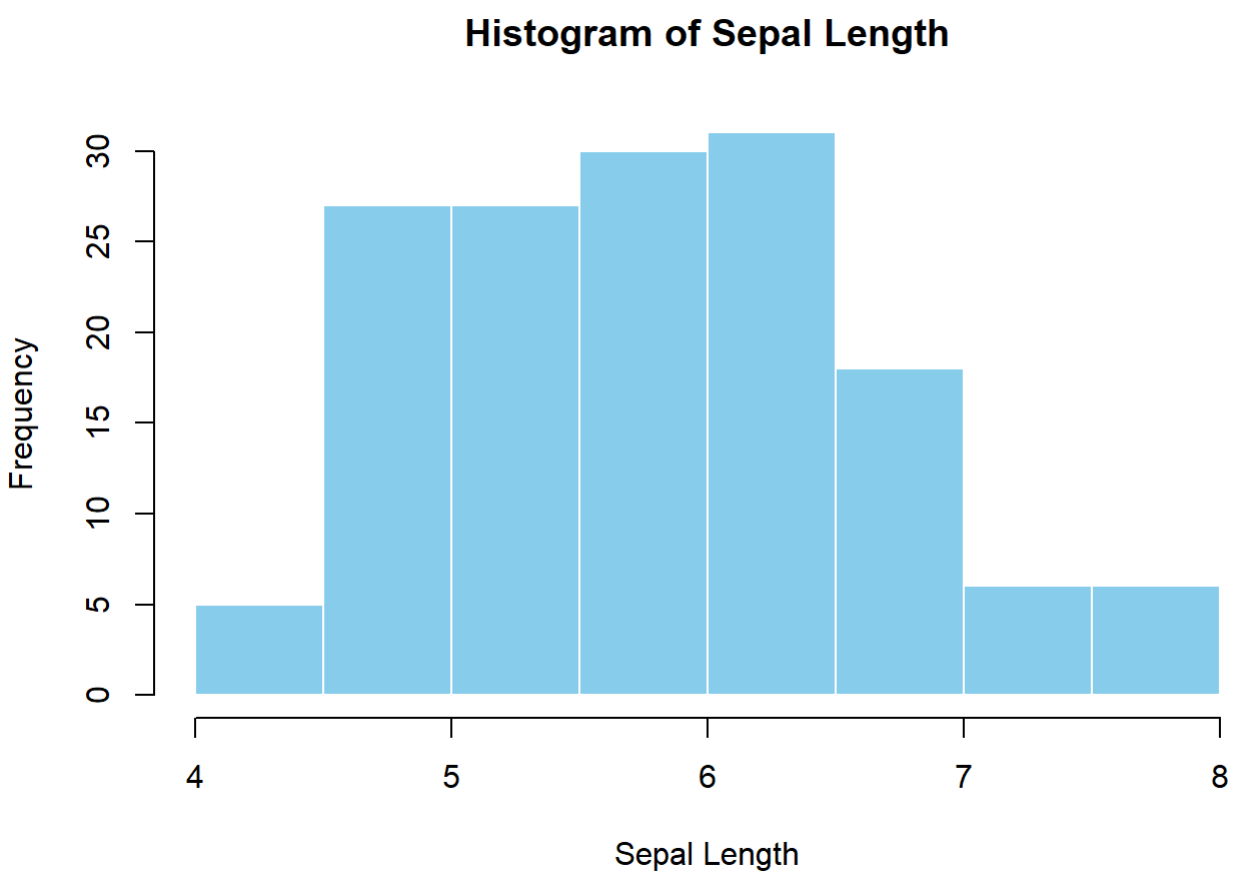
```
# Boxplot
boxplot(data$Sepal.Length ~ data$Species, main = "Boxplot of Sepal Length by Species",
        xlab = "Species", ylab = "Sepal Length", col = c("lightblue", "lightgreen", "lightpin
k"))
```




```
# Scatter Plot
species_colors <- as.numeric(factor(data$Species))
plot(data$Sepal.Length, data$Petal.Length, main = "Scatter Plot of Sepal Length vs Petal Leng
th",
     xlab = "Sepal Length", ylab = "Petal Length", col = species_colors, pch = 19)

legend("topright", legend = levels(factor(data$Species)), col = 1:3, pch = 19)
```




```
# Histogram
hist(data$Sepal.Length, main = "Histogram of Sepal Length",
     xlab = "Sepal Length", col = "skyblue", border = "white")
```

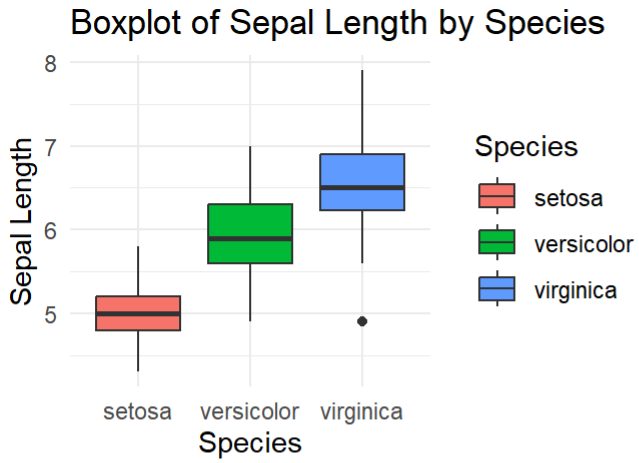


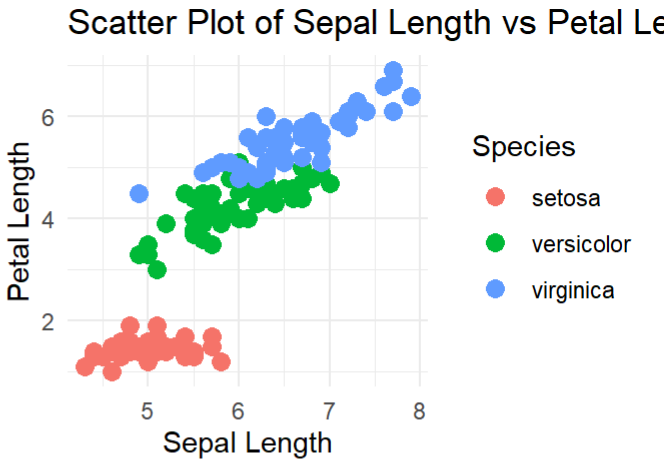

```
library(ggplot2)
library(ggpubr)

# Boxplot with ggplot
gg_boxplot <- ggplot(data, aes(x = Species, y = Sepal.Length, fill = Species)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Boxplot of Sepal Length by Species", x = "Species", y = "Sepal Length")
```

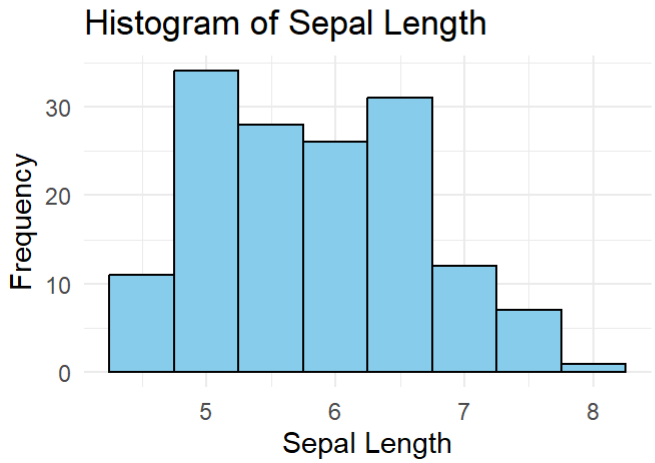
```
# Scatter plot with ggplot
gg_scatter <- ggplot(data, aes(x = Sepal.Length, y = Petal.Length, color = Species)) +
  geom_point(size = 3) +
  theme_minimal() +
  labs(title = "Scatter Plot of Sepal Length vs Petal Length", x = "Sepal Length", y = "Petal
Length")
```

```
# Histogram with ggplot
gg_hist <- ggplot(data, aes(x = Sepal.Length)) +
  geom_histogram(binwidth = 0.5, fill = "skyblue", color = "black") +
  theme_minimal() +
  labs(title = "Histogram of Sepal Length", x = "Sepal Length", y = "Frequency")
```

```
# Combine all plots using ggpubr
ggpubr::ggarrange(gg_boxplot, gg_scatter, gg_hist, ncol = 2, nrow = 2)
```


Boxplot of Sepal Length by Species
Sepal Length
8
7
6
5
setosa
versicolor
virginica
Species
Species
setosa
versicolor
virginica


Scatter Plot of Sepal Length vs Petal Le
Petal Length
6
4
2
5
6
7
8
Sepal Length
Species
setosa
versicolor
virginica


Histogram of Sepal Length
Frequency
30
20
10
0
5
6
7
8
Sepal Length