STAT40830-Adv Data Prog with R
Assignment 2

Sai Sampath Vadrevu - 23200373

```r
# Import necessary libraries
library(ggplot2)
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
# Load the iris dataset
data("iris")

# Display the first few rows of the dataset
head(iris)
```

```
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
```

```r
# Calculate and print summary statistics for Sepal Length and Sepal Width
summary_stats <- iris %>%
  summarize(
    min_sepal_length = min(Sepal.Length),
    max_sepal_length = max(Sepal.Length),
    mean_sepal_length = mean(Sepal.Length),
    median_sepal_length = median(Sepal.Length),
    first_quartile_sepal_length = quantile(Sepal.Length, 0.25),
    third_quartile_sepal_length = quantile(Sepal.Length, 0.75),
    min_sepal_width = min(Sepal.Width),
    max_sepal_width = max(Sepal.Width),
    mean_sepal_width = mean(Sepal.Width),
    median_sepal_width = median(Sepal.Width),
    first_quartile_sepal_width = quantile(Sepal.Width, 0.25),
    third_quartile_sepal_width = quantile(Sepal.Width, 0.75)
  )

print("Summary Statistics:")
```

```
[1] "Summary Statistics:"
```

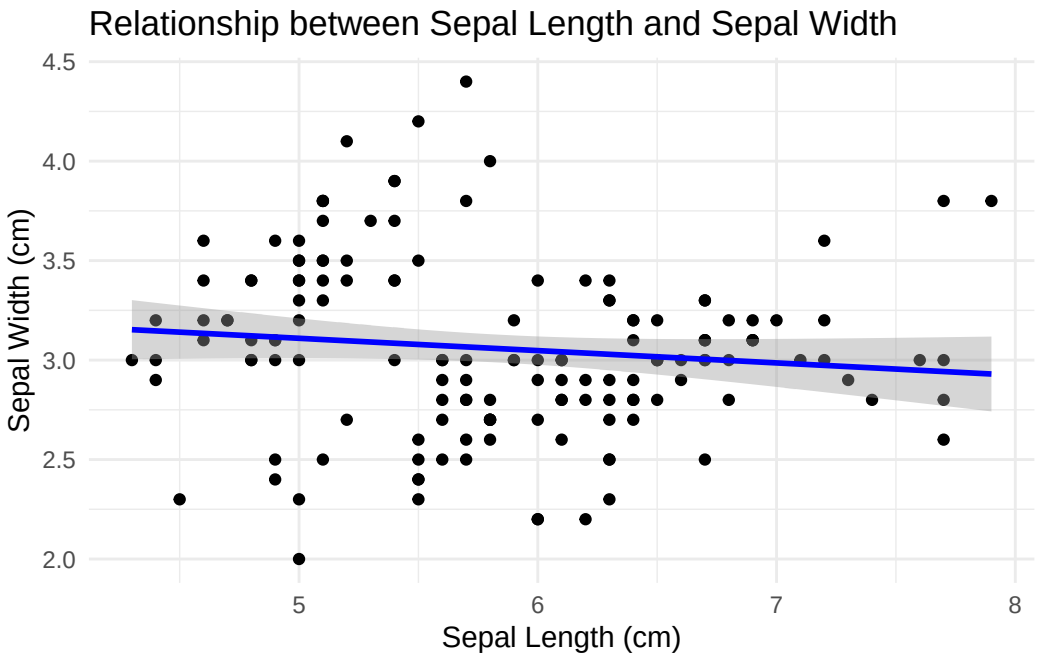
```r
print(summary_stats)
```

```
  min_sepal_length max_sepal_length mean_sepal_length median_sepal_length
1              4.3              7.9          5.843333                 5.8
  first_quartile_sepal_length third_quartile_sepal_length min_sepal_width
1                         5.1                         6.4               2
  max_sepal_width mean_sepal_width median_sepal_width
1             4.4         3.057333                  3
  first_quartile_sepal_width third_quartile_sepal_width
1                        2.8                        3.3
```

```r
# Create a scatter plot of Sepal Length vs Sepal Width with a regression line
ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width)) +
  geom_point() +
  geom_smooth(method = "lm", col = "blue") +
  labs(
    title = "Relationship between Sepal Length and Sepal Width",
    x = "Sepal Length (cm)",
```

```
    y = "Sepal Width (cm)"
  ) +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```




```
# Calculate and display the correlation between Sepal Length and Sepal Width
correlation <- cor(iris$Sepal.Length, iris$Sepal.Width)
print(paste("Correlation between Sepal Length and Sepal Width:", correlation))
```

```
[1] "Correlation between Sepal Length and Sepal Width: -0.117569784133002"
```

```
# Interpret the results
cat("Interpretation:\n")
```

```
Interpretation:
```

```r
cat("The summary statistics indicate that Sepal Length ranges from", summary_stats$min_sep
    "with an average of", round(summary_stats$mean_sepal_length, 2), "cm. Sepal Width rang
    summary_stats$min_sepal_width, "to", summary_stats$max_sepal_width, "with an average o
```

```
The summary statistics indicate that Sepal Length ranges from 4.3 to 7.9 with an average of 5
```

```r
cat("The scatter plot with a regression line shows a weak negative correlation between Sep
    "suggesting that as Sepal Length increases, Sepal Width slightly decreases.\n")
```

```
The scatter plot with a regression line shows a weak negative correlation between Sepal Lengt
```

```r
cat("This is supported by the correlation coefficient of", round(correlation, 2), "indicat
```

```
This is supported by the correlation coefficient of -0.12 indicating a weak negative correlat
```