

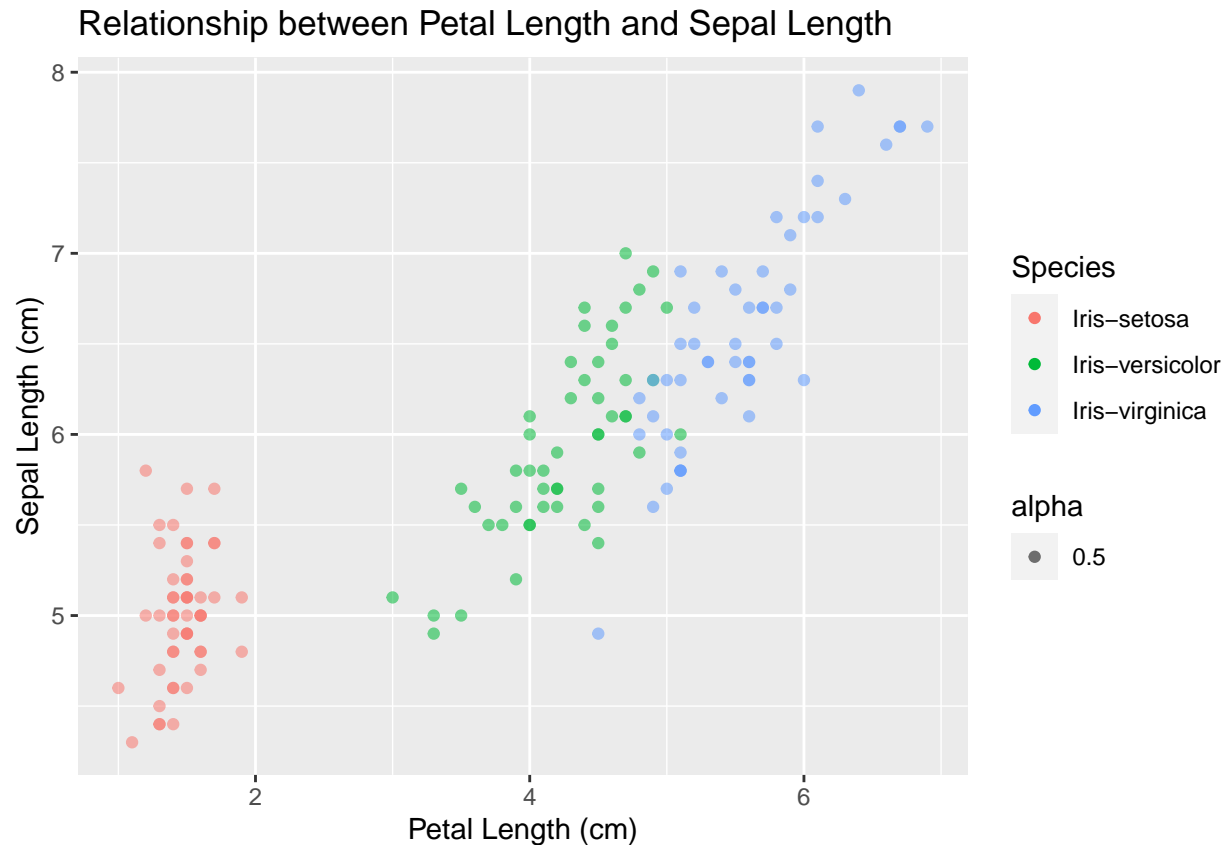
## Assignment 1 Question 2

Sheen Thusoo

f. Using R, generate a scatterplot of the sepal length (y-axis) vs. the petal length (x-axis) while colouring the points based on the iris species (use three visible colours, e.g. black, red, and green). Make sure to add a legend and axis labels. Comment on the patterns you observe.

```
iris_data <- read.csv("Iris.csv")

library("ggplot2")
ggplot(iris_data) +
  geom_point(
    aes(x = PetalLength,
        y = SepalLength,
        colour = Species,
        alpha = 0.5)
  ) +
  labs(
    title = 'Relationship between Petal Length and Sepal Length',
    x = 'Petal Length (cm)',
    y = 'Sepal Length (cm)'
  )
```



From the Figure above, it is clear that the each Iris species has a pattern. The Iris\_setosa species (red color) has both low petal length (between 0 to 2cm) and sepal length (0 to 6cm). Thus, its points are clustered toward the bottom left of the scatter plot. The Iris-versicolor species (green color) has points in the middle of the graph where its petal length ranges from about 3 to 5cm and it sepal length ranges from about 5 to 7cm. Thus, most of its points are clustered in the center of the scatter plot. Lastly, the Iris-virginica species (blue color) has larger petal lengths (between 4.5 to 7cm) and larger sepal lengths (between 4.5 to 8cm). Thus, most of its point are clustered to the right of the scatter plot. It seem to have the greatest range of sepal lengths as the points are spread out vertically from the bottom to the top of the scatter plot. It is also important to note that the points for Iris-versicolor (green) and Iris-virginica (blue) overlap slightly.