# Assignment 2

**Student Name and ID**

A ECSE205 Homework Assignment

The Iris flower data set or Fisher's Iris data set is a multivariate data set used and made famous by the British statistician and biologist Ronald Fisher in his 1936 paper who initiated the Linear Discriminant Analysis, one of the early and still widely-used Machine Learning methods. The dataset can be downloaded from https://archive.ics.uci.edu/dataset/53/iris though there are many packages that include this dataset as a demo. The dataset consists of 5 variables, 4 continuous variables, and 1 categorical variable as follows.

- sepal length
- sepal width
- petal length
- petal width
- class (Setosa, Versicolor, Virginica)

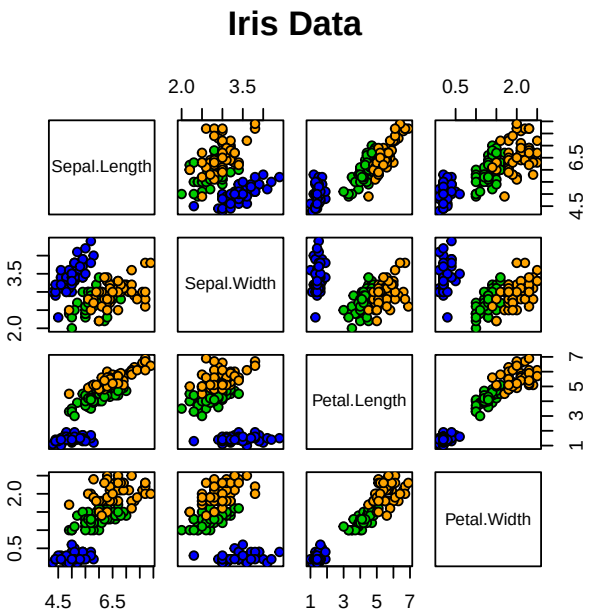
Include the codes before each plot or generated results.

### 0.0.1 Scatter Plot

**Problem 1: Scatter plot**

Use different colors to visualize the relationship between a pair of variable in different scatterplots similar to the scatterplot below. (20 points)

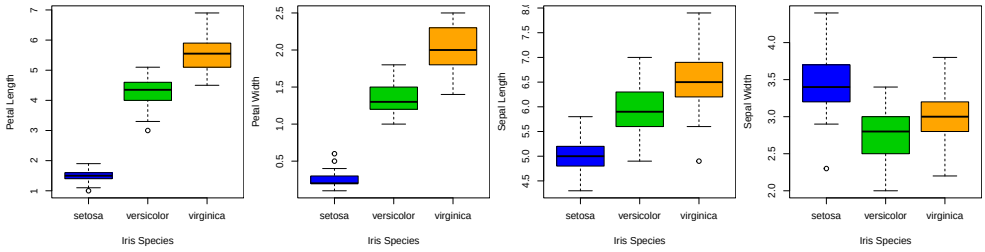
```
print("Hello World!")
```



## Problem 2: Box plot

Generate the boxplot for each variable in each category and explain what you can say about the distribution of each variable in each class. (20 points)

```
print("Hello World!")
```



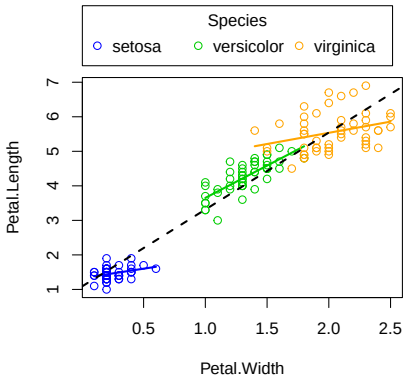

## Problem 3: Simple Regression

- Use the linear regression to predict the petal length using the petal width, i.e. to find $\beta_0, \beta_1$. Petal.Length = $\beta_0$ + $\beta_1$Petal.Width(20 points)

```
print("Hello World!")
```

```
    Coefficients:
(Intercept)  Petal.Width
      1.084        2.230
```

- Plot the regression line for each category on top of the overall line.(20 points)



**Problem 4: Descriptive Statistics**

Generate the descriptive statistics that include min, max, mean, first quartile Q1, second quartile or median Q2, and the 3rd quartile Q3, for each of the continuous variables.(20 points)

```
      Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
 Median :5.800   Median :3.000   Median :4.350   Median :1.300
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
```