

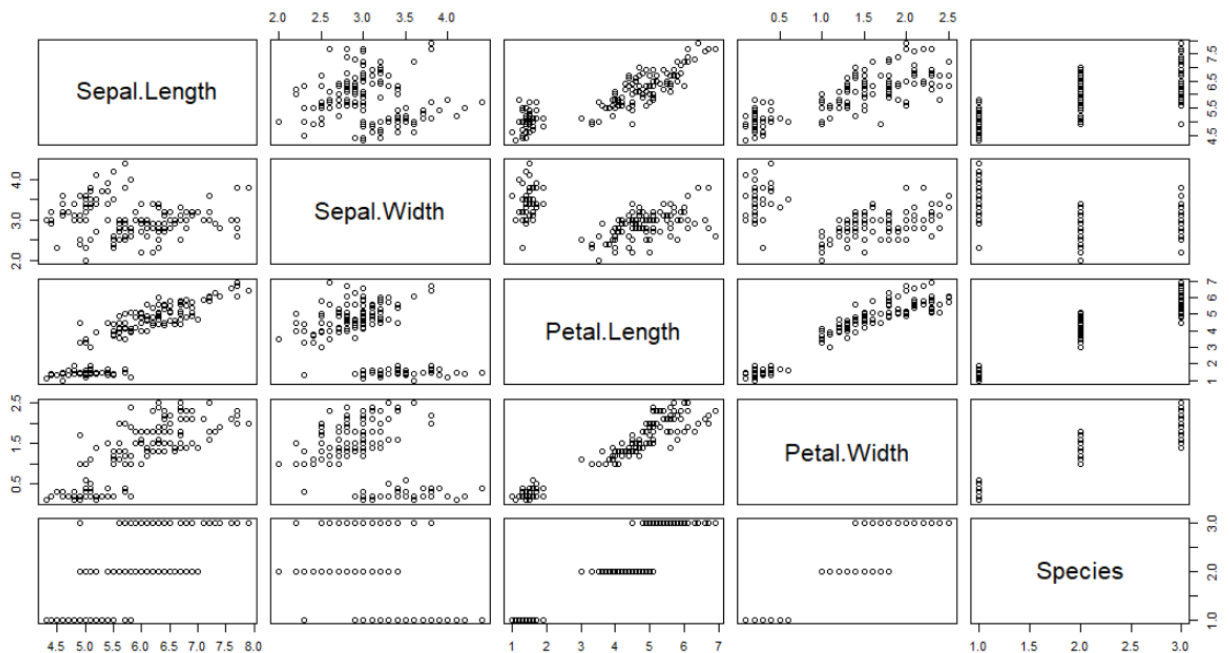
## #DataScience

### Exploratory Data Analysis in R: Data Visualization

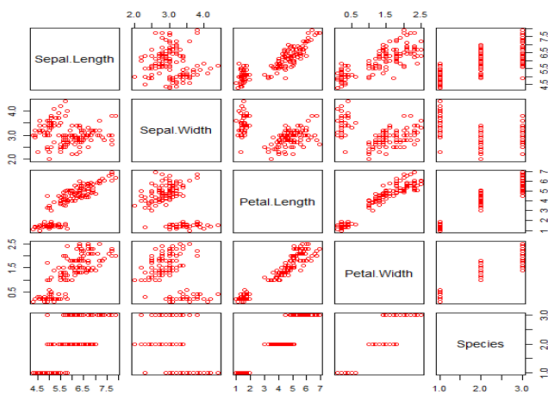
- Recall: in part 1, we used `summary()` and `skim()` to provide an overview of the dataset
- Aim: use `plot()` to visually have a look of the distribution of the data itself

`plot()`

- `plot(iris)`
- Output: 5 by 5 scatter plot in a pair-wise manner
  - Between various combinations of the 5 variables (sepal length, sepal width, petal length, petal width, species)



- Adding `col =` argument
  - `plot(iris, col = 'red')`

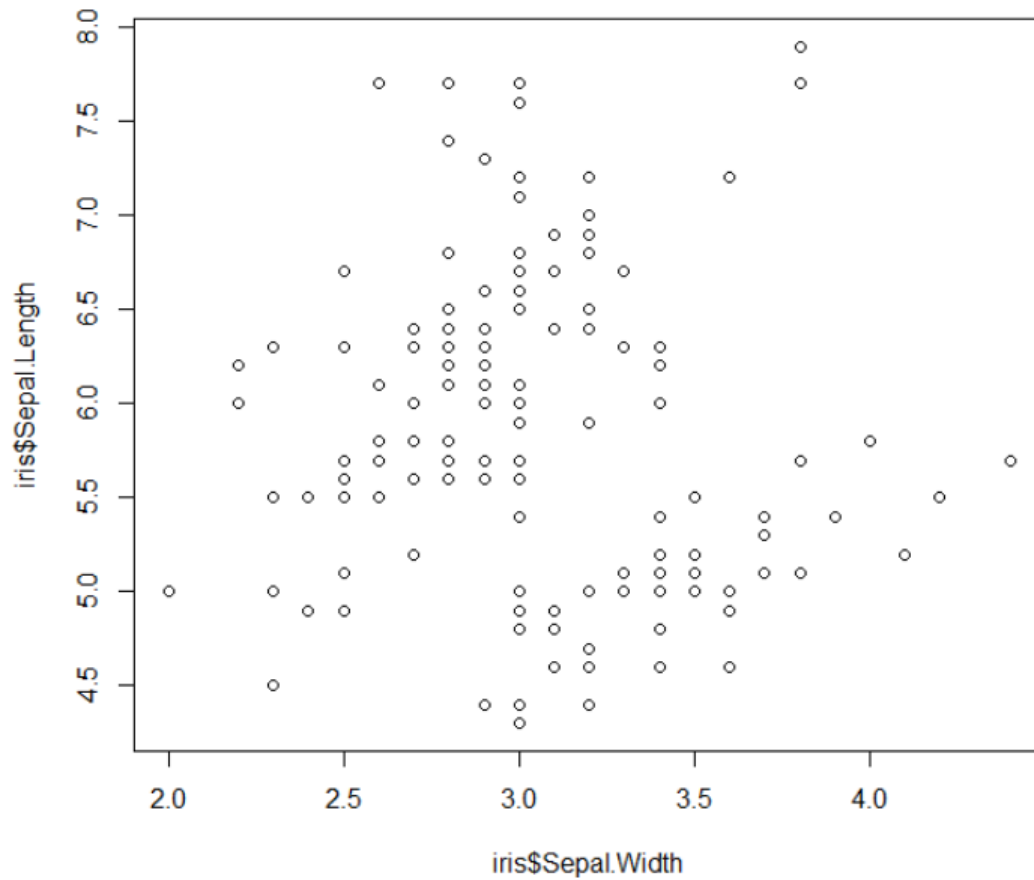


- just adds colour

## Scatter Plot

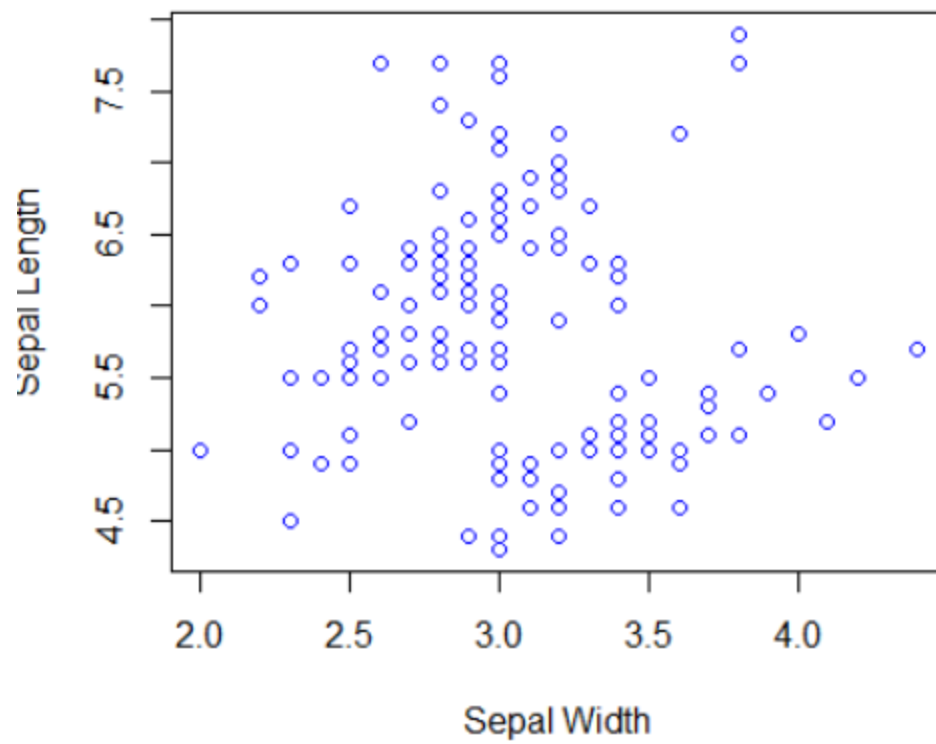
- To obtain a scatter plot as output
- `plot(var1, var2)`
- `plot(iris$Sepal.Width, iris$Sepal.Length)`

```
82  
83 # Scatter Plot.  
84 plot(iris$Sepal.Width, iris$Sepal.Length)  
85
```



```
92  
93 plot(iris$Sepal.Width, iris$Sepal.Length, col = 'blue',  
94       xlab = "Sepal width", ylab = "Sepal Length")  
95
```

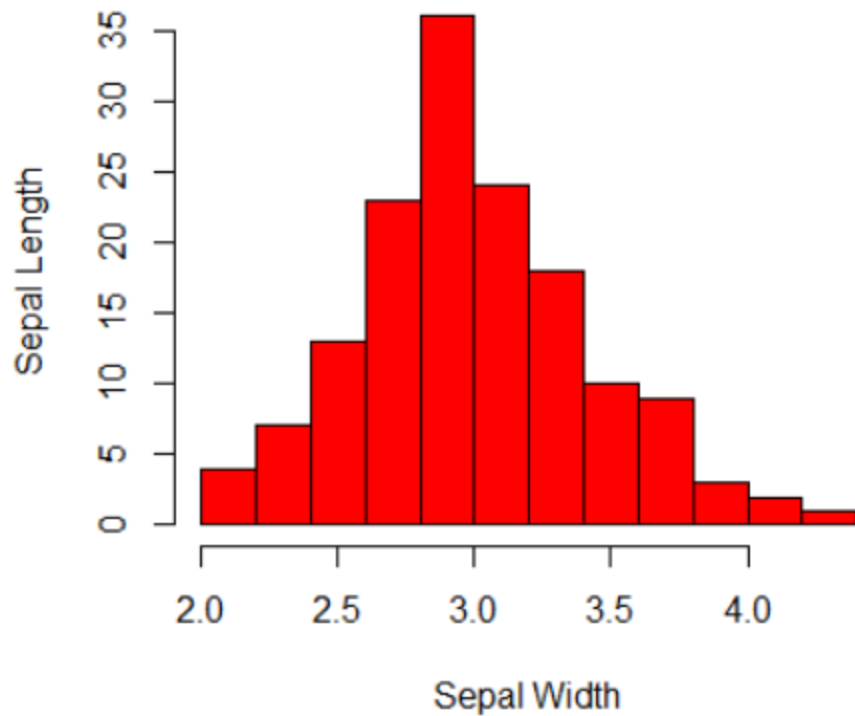
- Adds x and y axis titles



Histogram

```
97 # Histogram.  
98 hist(iris$Sepal.Width)  
99 hist(iris$Sepal.Width, col="red") # Adds red bars.  
100  
101
```

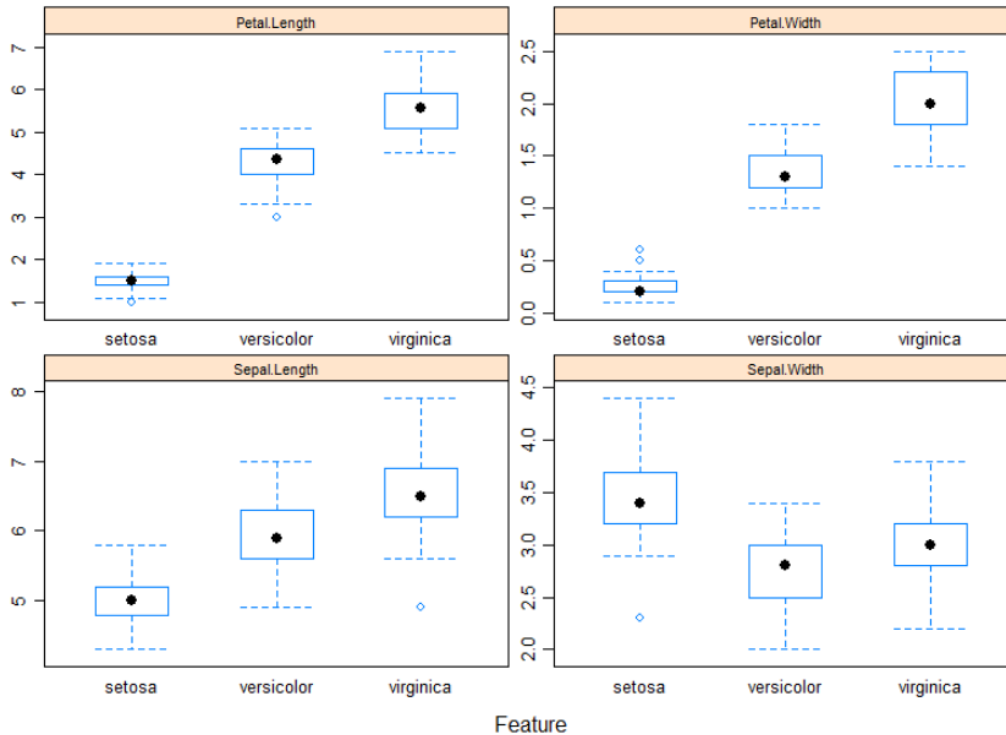
**Histogram of iris\$Sepal.Width**



- A histogram is a series of bars where each bar represents a range in the value distribution
- So the values of sepal width from 2.0 to about 2.2 will have a frequency of 5
  - Frequency is essentially the count
  - So, from the range of 2.0 to 2.2, there is a total of 5 samples
  - It's basically the frequency count

#### Feature Plot

- Feature plot shows the box plots for 4 variables as a function of 3 classes
- 4 variables: Sepal Length, Sepal Width, Petal Length, Petal Width
- 3 classes: Setosa, Virginica, Versicolor
  - For each variable, we get to see the relative distribution of the box plot



- Ex: For setosa, it has the lowest value amongst the 3 classes of flower for the Sepal.Length descriptor
  - As well as for Petal Length & Width
  - While setosa has the highest value for the Sepal Width
  - And so on...
  - The distribution is about the same for petal width and petal length with respect to virginica and versicolor
- Notice: Setosa has a very different distribution compared to the others
  - Versicolor and virginica roughly have similar distributions

edit this pdf cont..

## References

[https://www.youtube.com/watch?v=do9yrLrv7a4&list=PLtqF5YXg7GLk9QRC5kS5Am4Ijo4S9gqk\\_&index=2](https://www.youtube.com/watch?v=do9yrLrv7a4&list=PLtqF5YXg7GLk9QRC5kS5Am4Ijo4S9gqk_&index=2)

[https://www.youtube.com/watch?v=7XdoaQYwTeA&list=PLtqF5YXg7GLn0WWB\\_wQx7wHrlvbs0EH2e](https://www.youtube.com/watch?v=7XdoaQYwTeA&list=PLtqF5YXg7GLn0WWB_wQx7wHrlvbs0EH2e)

[https://github.com/mathstudent97/WebAppsInR\\_Part2/tree/main/4\\_WebApp](https://github.com/mathstudent97/WebAppsInR_Part2/tree/main/4_WebApp)