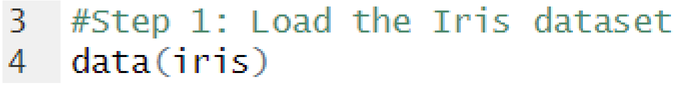
Creating Scatterplots and Boxplots in R

This tutorial will use the common Iris dataset that is included in R. We'll learn how to open the dataset, view the structure, view a summary of the data, and create visuals to understand the data.

Step 1: Load the Iris Dataset

Iris is included in base R so there is no need to download it from another source.

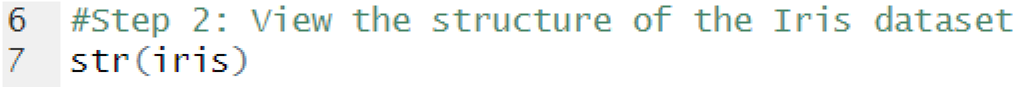


```
#Step 1: Load the Iris dataset
data(iris)
```

The 'data()' function will open the dataset in your workspace.

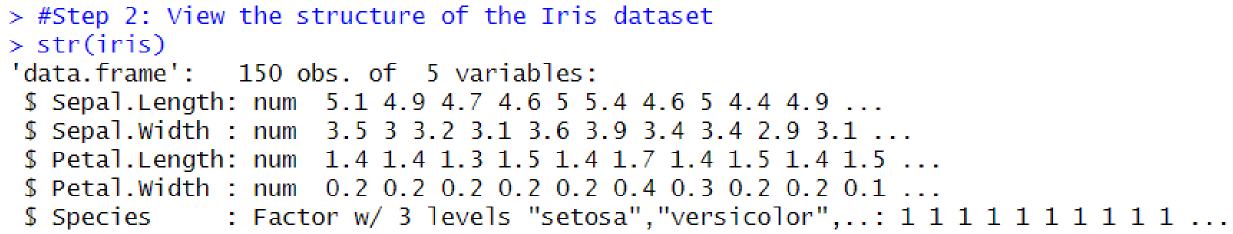
Step 2: View the Iris Dataset

Examine the structure of the Iris dataset.



```
#Step 2: View the structure of the Iris dataset
str(iris)
```

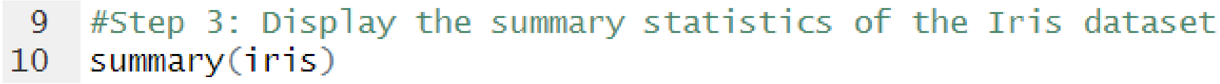
The 'str()' function will display information about the variables and their types. This is what will appear when you run the line:



```
> #Step 2: View the structure of the Iris dataset
> str(iris)
'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 ...
```

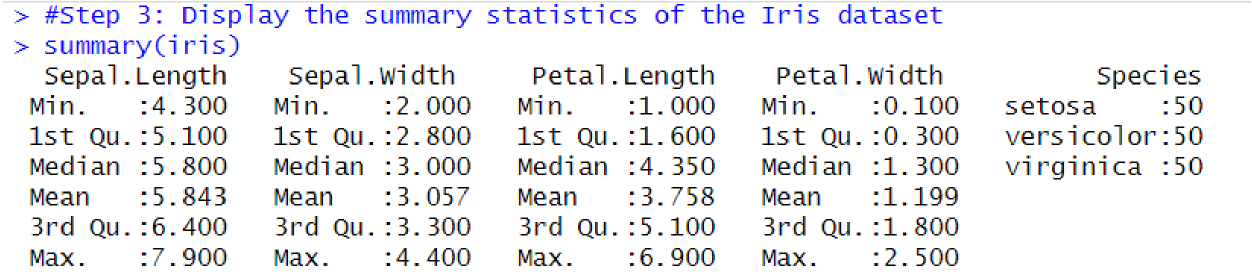
Step 3: Summarize the Iris Dataset

Load the summary statistics for each variable of the dataset.



```
#Step 3: Display the summary statistics of the Iris dataset
summary(iris)
```

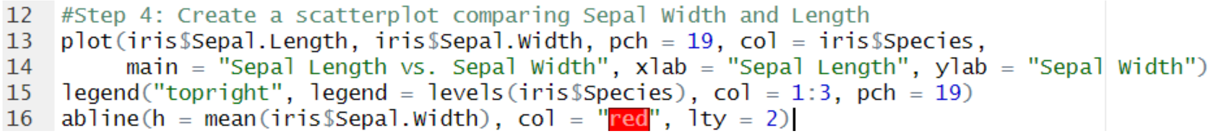
The 'summary()' function will display the summary statistics of your variables. This includes information such as minimum and maximum, median, and mean. This is what will appear when you run the line:



```
> #Step 3: Display the summary statistics of the Iris dataset
> summary(iris)
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width          Species  
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa    :50  
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50  
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50  
 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                  
```

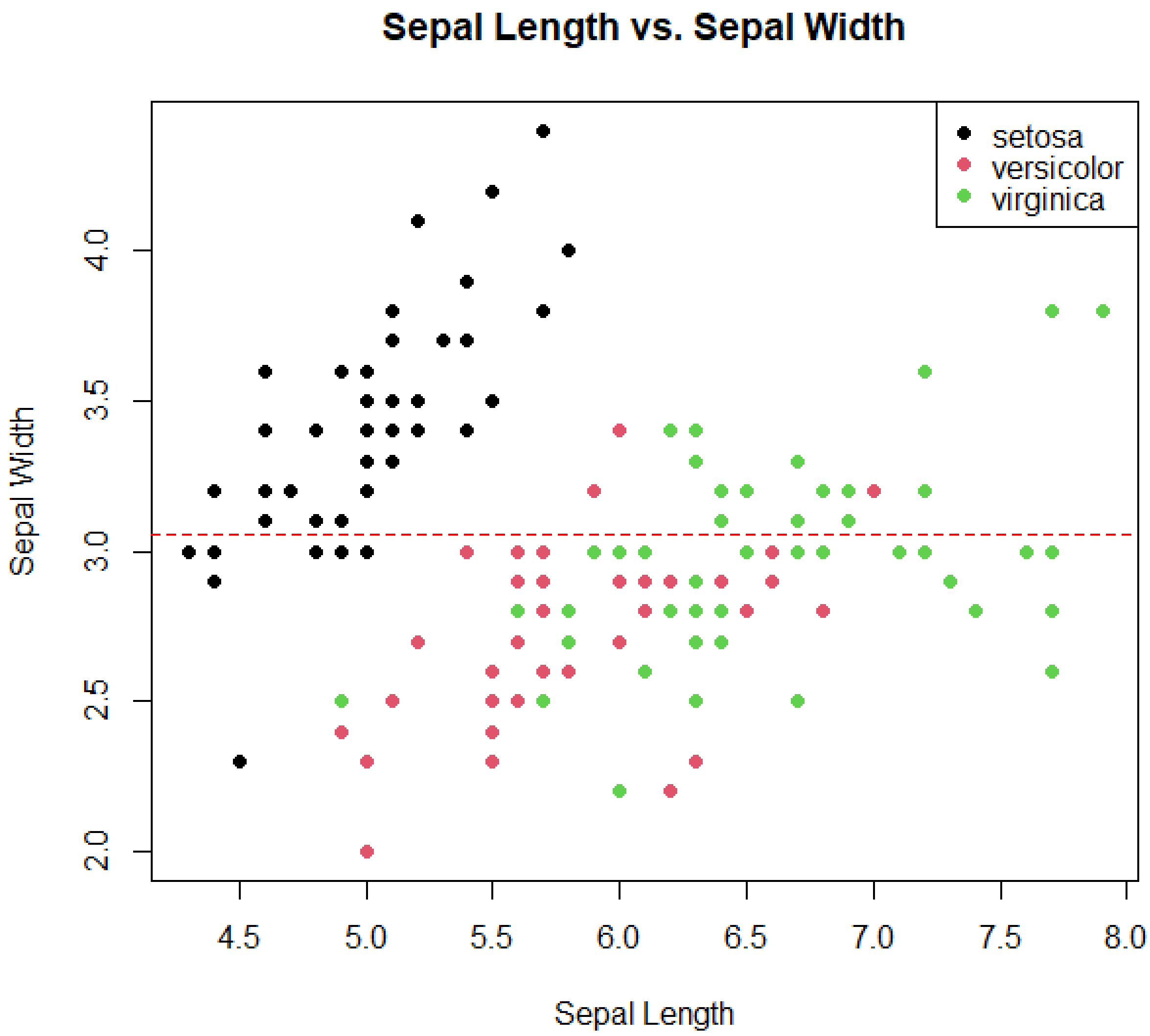
Step 4: Visualize the Data with a Scatterplot

This scatterplot will create a visualization of the relationship between sepal length and sepal width for the three species in the dataset.



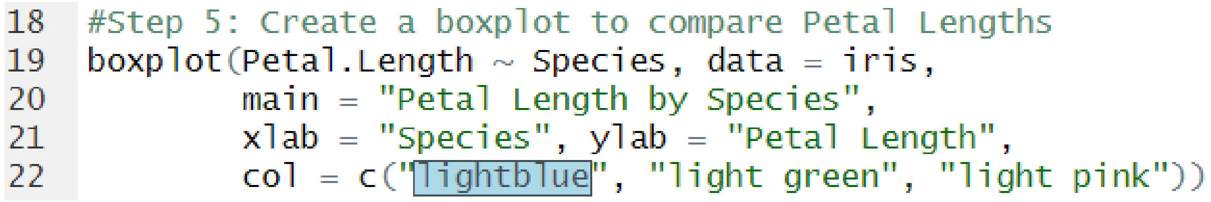
```
#Step 4: Create a scatterplot comparing Sepal Width and Length
plot(iris$Sepal.Length, iris$Sepal.Width, pch = 19, col = iris$Species,
     main = "Sepal Length vs. Sepal Width", xlab = "Sepal Length", ylab = "Sepal Width")
legend("topright", legend = levels(iris$Species), col = 1:3, pch = 19)
abline(h = mean(iris$Sepal.Width), col = "red", lty = 2)
```

The ‘plot()’ function creates a scatter plot using Sepal Length and Sepal Width variables. ‘pch = 19’ sets the point character and ‘col’ will differentiate the points by species. The ‘legend()’ function will add a legend to the plot. ‘abline()’ can be used to add a horizontal line at the mean of Sepal Width. This is how the plot should look:



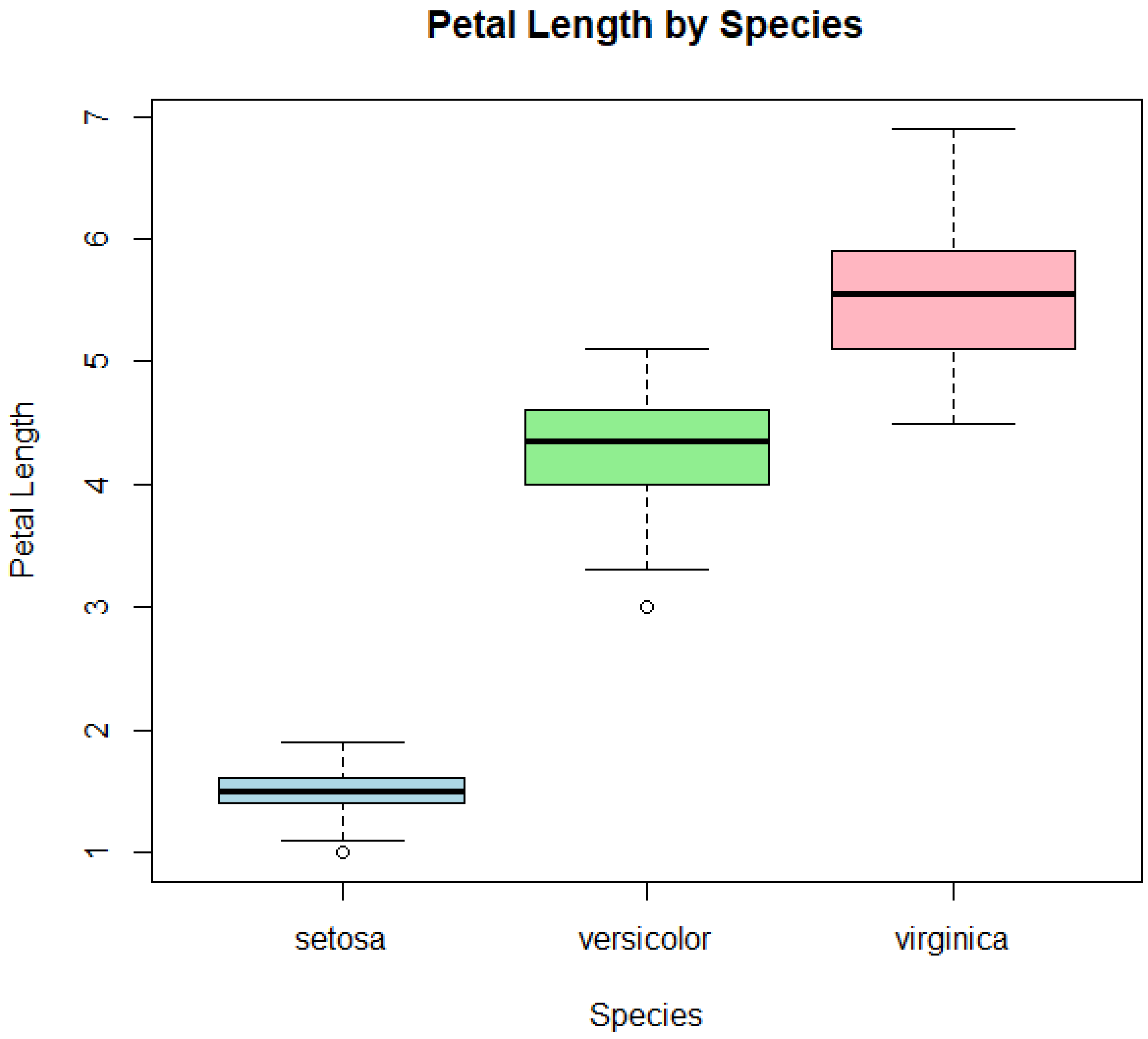
Step 5: Visualize the Data with a Boxplot

This boxplot will create a visual distribution of the petal lengths for each species.



```
#Step 5: Create a boxplot to compare Petal Lengths
boxplot(Petal.Length ~ Species, data = iris,
        main = "Petal Length by Species",
        xlab = "Species", ylab = "Petal Length",
        col = c("lightblue", "light green", "light pink"))
```

The ‘boxplot()’ function generates a boxplot using the provided variables. ‘iris$Petal.Length ~ iris$Species’ specifies that the variable being plotted against species will be petal length. ‘main’, ‘xlab’, and ‘ylab’ specify the titles and axis labels for the graph. ‘col’ selects a color for the box plot of each species. Here is how the graph will look:



References:

- Example Tutorial: http://betsymccall.net/edu/CDSE/coding/R/bar_graphs.pdf
- RDocumentation: https://www.rdocumentation.org/