

Ex no: 10

Visualize data using any plotting framework

Aim:

To visualize data using any plotting framework

Procedure:

a) SCATTER PLOT:

1. Install and load the ggplot2 package for data visualization.
2. Use the iris dataset for plotting.
3. Set Sepal.Length as the x-axis and Sepal.Width as the y-axis in the `'aes()'` function.
4. Assign colors to the points based on the categorical `**Species**` variable.
5. Add scatter plot points to the graph using `'geom_point()'`, setting the size of the points.
6. Add a plot title and label the x-axis and y-axis using `'labs()'`.
7. Apply a minimal theme to the plot for a clean appearance using `'theme_minimal()'`.

b) BAR CHART:

1. Install and load the ggplot2 package for data visualization.
2. Use the iris dataset for plotting.
3. Set Species as the x-axis variable in the `'aes()'` function.
4. Use `'geom_bar()'` to create a bar plot representing the count of each species, filling the bars with a steelblue color.
5. Add a plot title and label the x-axis as Species and the y-axis as Count using the `'labs()'` function.
6. Apply the `theme_minimal()` function to give the plot a clean and minimalistic appearance.
7. Display the bar plot to visualize the count of different species in the iris dataset.

c) HISTOGRAM:

1. Install and load the ggplot2 package for data visualization.
2. Use the iris dataset for plotting.
3. Set Sepal.Length as the x-axis variable in the `'aes()'` function.
4. Create a histogram using `'geom_histogram()'`, specifying a bin width of 0.3, with bars filled in orange and outlined in black.

5. Add a plot title and label the x-axis as Sepal Length (cm) and the y-axis as Frequency using the ``labs()`` function.
6. Apply the `theme_minimal()` function to give the plot a clean and minimalistic appearance.
7. Display the histogram to visualize the distribution of Sepal Length in the iris dataset.

d) BOX PLOT:

1. Install and load the `ggplot2` package for data visualization.
2. Use the iris dataset for plotting.
3. Set Species as the x-axis variable and Sepal.Length as the y-axis variable in the ``aes()`` function, with bars filled based on Species.
4. Create a box plot using ``geom_boxplot()``.
5. Add a plot title and label the x-axis as Species and the y-axis as Sepal Length (cm) using the ``labs()`` function.
6. Apply the `theme_minimal()` function to give the plot a clean and minimalistic appearance.
7. Display the box plot to visualize the distribution of Sepal Length for each Species.

Program:

a) SCATTER PLOT:

```
# Install ggplot2 (if not already installed)
install.packages("ggplot2")

# Load the ggplot2 package
library(ggplot2)

# Scatter plot of Sepal.Length vs Sepal.Width, colored by Species
ggplot(data = iris, aes(x = Sepal.Length, y = Sepal.Width, color = Species)) +
  geom_point(size = 3) + # Adds points
  labs(title = "Scatter Plot of Sepal Dimensions",
        x = "Sepal Length (cm)",
        y = "Sepal Width (cm)") + # Adds axis labels and title
  theme_minimal() # Applies a minimal theme
```

b) BAR CHART:

```
# Install ggplot2 (if not already installed)
install.packages("ggplot2")

# Load the ggplot2 package
library(ggplot2)

# Bar plot of Species counts
ggplot(data = iris, aes(x = Species)) +
  geom_bar(fill = "steelblue") + # Adds bars filled with steel blue color
  labs(title = "Count of Different Species in Iris Dataset",
        x = "Species",
        y = "Count") +
  theme_minimal()
```

c) HISTOGRAM:

```
# Install ggplot2 (if not already installed)
install.packages("ggplot2")

# Load the ggplot2 package
library(ggplot2)

# Histogram of Sepal Length
ggplot(data = iris, aes(x = Sepal.Length)) +
  geom_histogram(binwidth = 0.3, fill = "orange", color = "black") + # Adds
  histogram bars
  labs(title = "Histogram of Sepal Length",
        x = "Sepal Length (cm)",
        y = "Frequency") +
  theme_minimal()
```

d) BOX PLOT:

```
# Install ggplot2 (if not already installed)
install.packages("ggplot2")
```

```
# Load the ggplot2 package
library(ggplot2)

# Box plot of Sepal Length for each Species
ggplot(data = iris, aes(x = Species, y = Sepal.Length, fill = Species)) +
  geom_boxplot() + # Adds box plot
  labs(title = "Box Plot of Sepal Length by Species",
        x = "Species",
        y = "Sepal Length (cm)") +
  theme_minimal()
```

Output:

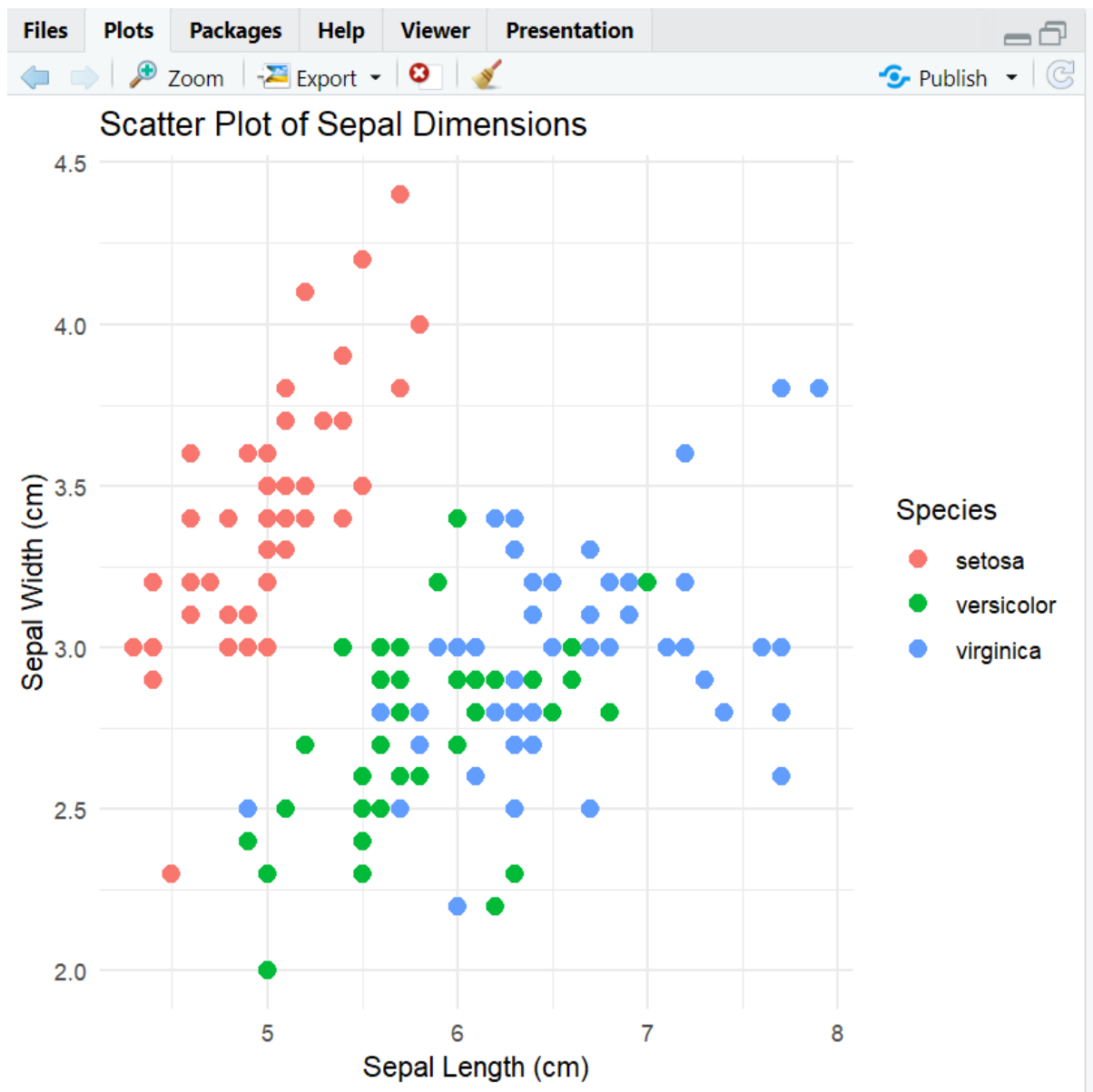
a) SCATTER PLOT:

```
> # Install ggplot2 (if not already installed)
> install.packages("ggplot2")
Error in install.packages : Updating loaded packages

> # Load the ggplot2 package
> library(ggplot2)

> # Scatter plot of Sepal.Length vs Sepal.Width, colored by Species
> ggplot(data = iris, aes(x = Sepal.Length, y = Sepal.Width, color = Species)) +
+ .... [TRUNCATED]
```

Environment	History	Connections	Tutorial	
<div> Import Dataset 211 MiB </div> <div> R Global Environment <input type="text" value=""/> </div>				
Data				
data	7 obs. of 2 variables			
hc_complete	List of 7			
iris	150 obs. of 6 variables			
iris_data	150 obs. of 4 variables			
iris_scaled	num [1:150, 1:4] -0.898 -1.139 -1.381 -1.50...			
kmeans_result	List of 9			
linear_model	List of 12			
logistic_model	List of 30			
mtcars	32 obs. of 11 variables			
svm_model	List of 31			
test_data	45 obs. of 5 variables			
train_data	105 obs. of 5 variables			
tree_model	List of 14			
Values				
accuracy	0.9777777777777778			
clusters	int [1:150] 1 1 1 1 1 1 1 1 1 1 ...			
confusion_matrix	'table' int [1:3, 1:3] 14 0 0 0 18 0 0 1 12			
distance_matrix	'dist' num [1:11175] 1.172 0.843 1.1 0.259 1...			
heights	num [1:7] 150 160 165 170 175 180 185			
k	3			
predicted_probs	Named num [1:32] 0.461 0.461 0.598 0.492 0.2...			
predictions	Factor w/ 3 levels "setosa","versicolor",...:...			
sample_indices	int [1:105] 14 50 118 43 150 148 90 91 143 9...			
weights	num [1:7] 55 60 62 68 70 75 80			



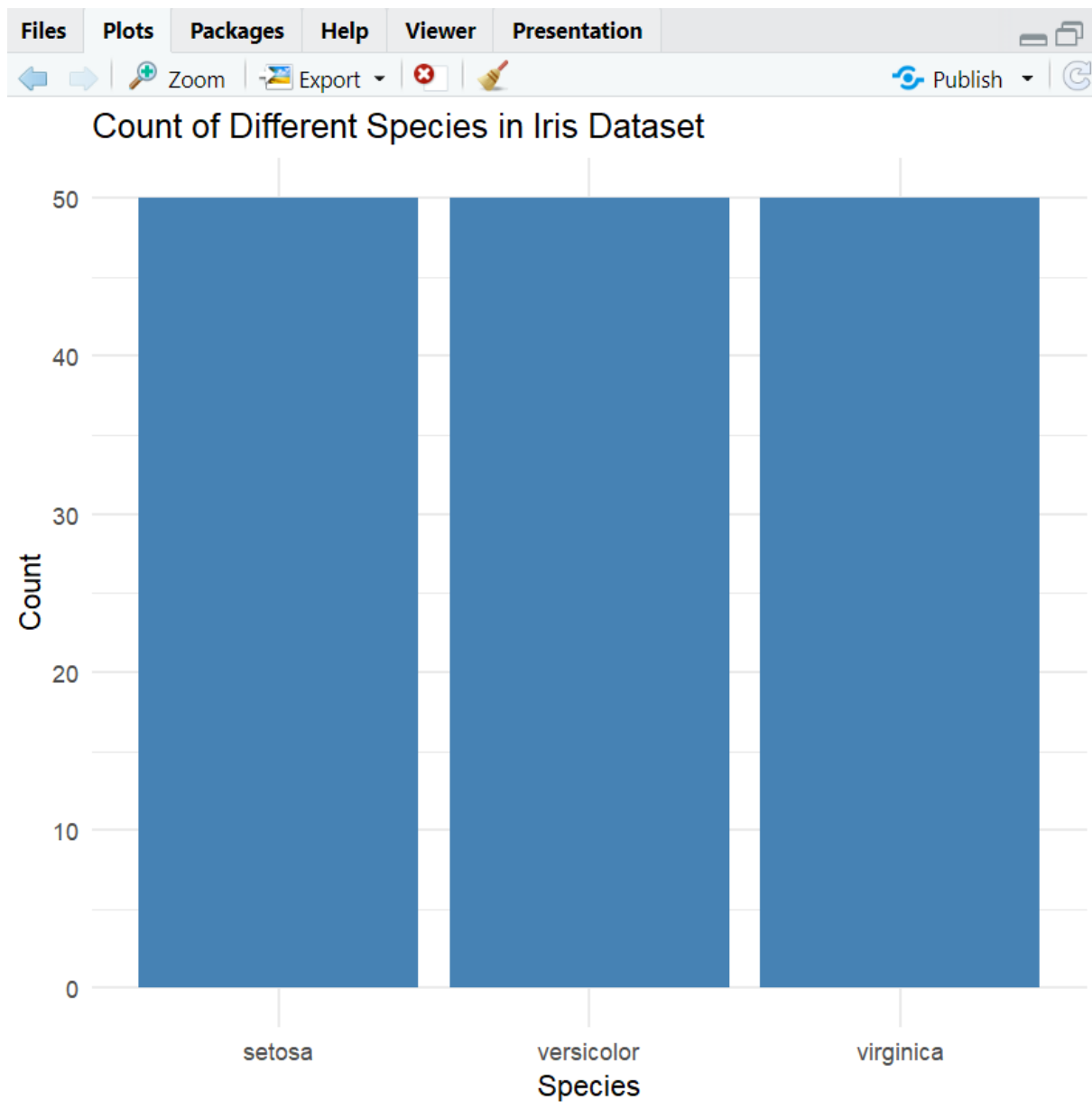
b) BAR CHART:

```
> # Install ggplot2 (if not already installed)
> install.packages("ggplot2")
Error in install.packages : Updating loaded packages

> # Load the ggplot2 package
> library(ggplot2)

> # Bar plot of Species counts
> ggplot(data = iris, aes(x = Species)) +
+   geom_bar(fill = "steelblue") + # Adds bars filled with steel blue color
+ .... [TRUNCATED]
```

Environment	History	Connections	Tutorial
R ▾ Global Environment ▾			
Data			
data	7 obs. of 2 variables		
hc_complete	List of 7		
iris	150 obs. of 6 variables		
iris_data	150 obs. of 4 variables		
iris_scaled	num [1:150, 1:4] -0.898 -1.139 -1.381 -1.50...		
kmeans_result	List of 9		
linear_model	List of 12		
logistic_model	List of 30		
mtcars	32 obs. of 11 variables		
svm_model	List of 31		
test_data	45 obs. of 5 variables		
train_data	105 obs. of 5 variables		
tree_model	List of 14		
Values			
accuracy	0.977777777777778		
clusters	int [1:150] 1 1 1 1 1 1 1 1 1 1 ...		
confusion_matrix	'table' int [1:3, 1:3] 14 0 0 0 18 0 0 1 12		
distance_matrix	'dist' num [1:11175] 1.172 0.843 1.1 0.259 1...		
heights	num [1:7] 150 160 165 170 175 180 185		
k	3		
predicted_probs	Named num [1:32] 0.461 0.461 0.598 0.492 0.2...		
predictions	Factor w/ 3 levels "setosa","versicolor",...:...		
sample_indices	int [1:105] 14 50 118 43 150 148 90 91 143 9...		
weights	num [1:7] 55 60 62 68 70 75 80		



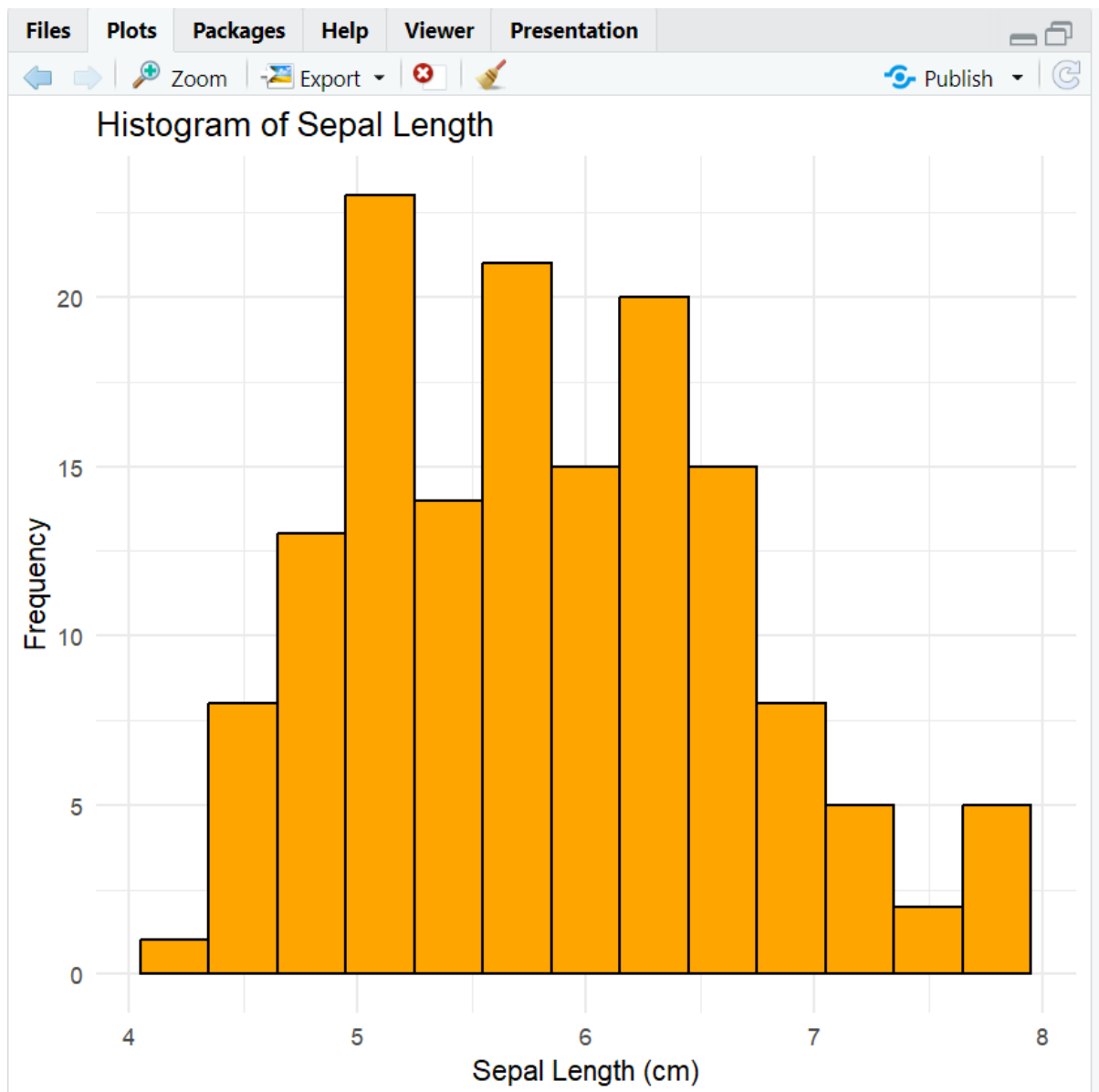
c) HISTOGRAM:

```
> # Install ggplot2 (if not already installed)
> install.packages("ggplot2")
Error in install.packages : Updating loaded packages

> # Load the ggplot2 package
> library(ggplot2)

> # Histogram of Sepal Length
> ggplot(data = iris, aes(x = Sepal.Length)) +
+   geom_histogram(binwidth = 0.3, fill = "orange", color = "black") +
+   .... [TRUNCATED]
```


Environment	History	Connections	Tutorial	
<div> Import Dataset 204 MiB </div> <div> R Global Environment </div> <div> List </div>				
Data				
data	7 obs. of 2 variables			
hc_complete	List of 7			
iris	150 obs. of 6 variables			
iris_data	150 obs. of 4 variables			
iris_scaled	num [1:150, 1:4] -0.898 -1.139 -1.381 -1.50...			
kmeans_result	List of 9			
linear_model	List of 12			
logistic_model	List of 30			
mtcars	32 obs. of 11 variables			
svm_model	List of 31			
test_data	45 obs. of 5 variables			
train_data	105 obs. of 5 variables			
tree_model	List of 14			
Values				
accuracy	0.9777777777777778			
clusters	int [1:150] 1 1 1 1 1 1 1 1 1 1 ...			
confusion_matrix	'table' int [1:3, 1:3] 14 0 0 0 18 0 0 1 12			
distance_matrix	'dist' num [1:11175] 1.172 0.843 1.1 0.259 1...			
heights	num [1:7] 150 160 165 170 175 180 185			
k	3			
predicted_probs	Named num [1:32] 0.461 0.461 0.598 0.492 0.2...			
predictions	Factor w/ 3 levels "setosa","versicolor",...:...			
sample_indices	int [1:105] 14 50 118 43 150 148 90 91 143 9...			
weights	num [1:7] 55 60 62 68 70 75 80			



d) BOX PLOT:

```
> # Install ggplot2 (if not already installed)
> install.packages("ggplot2")
Error in install.packages : Updating loaded packages

> # Load the ggplot2 package
> library(ggplot2)

> # Box plot of Sepal Length for each species
> ggplot(data = iris, aes(x = Species, y = Sepal.Length, fill = Species)) +
+   geom_boxplot() + # Adds .... [TRUNCATED]
```

Environment

History

Connections

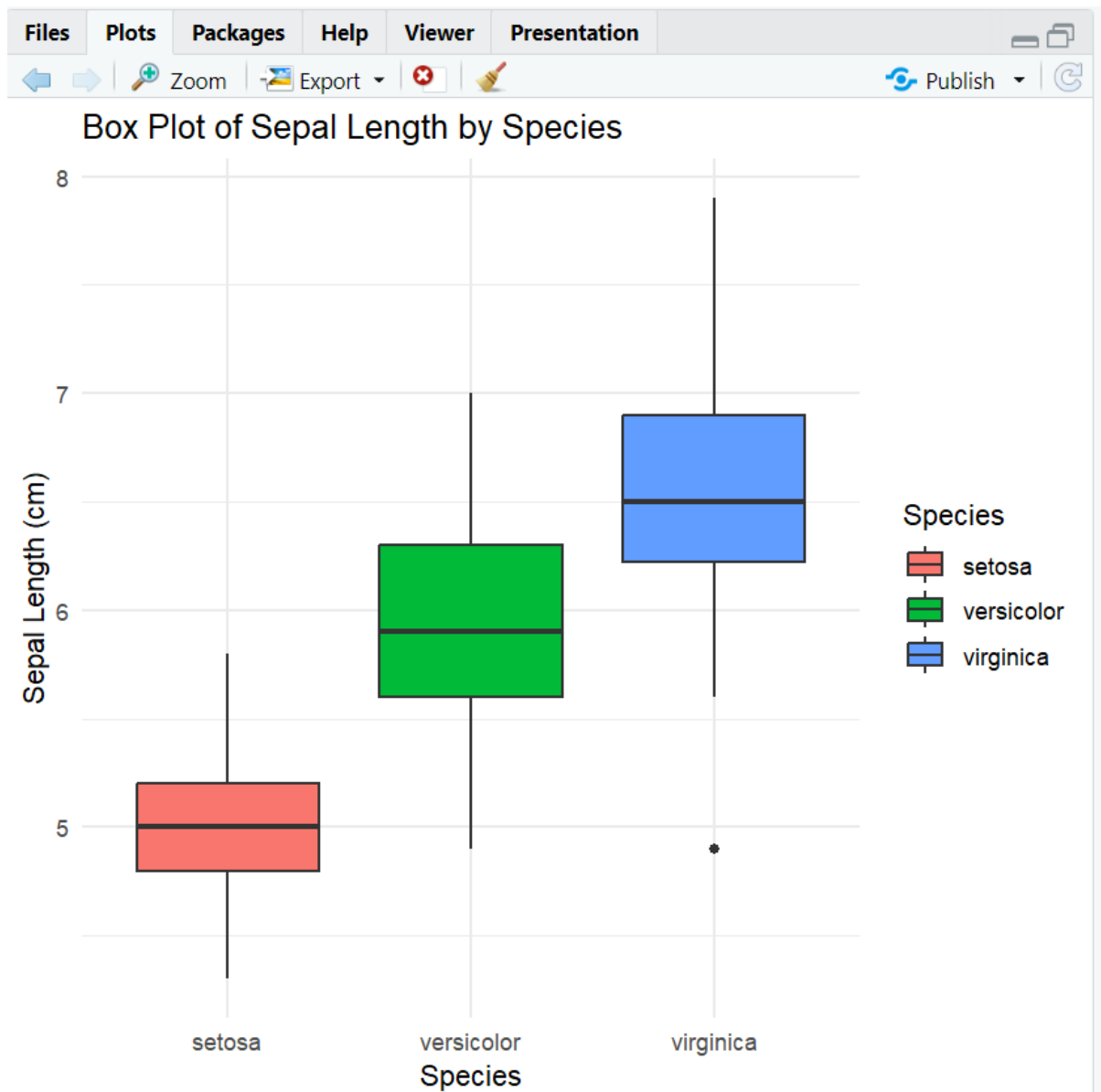
Tutorial

Import Dataset

208 MiB

R

Global Environment



Result:

Thus the implementation of clustering techniques – Hierarchical and K-Means using R programming has been executed successfully.