# Experiment-2.3

## Student Name: Nabha Varshney UID: 20BCS4995

**Branch: CSE Section/Group: 20BCS-DM-704 (A)**

## Semester: 6th Date of Performance: 21thApr 2023 Subject Name: Data Mining Subject Code: 20CSP- 351

**Aim** – To perform the cluster analysis by k-means method using R.

## Objective-

* Represent the reading of file using R studio
* Displaying the graph using clusterR, caret and cluster.
* Demonstration of Cluster analysis by K-means method.

**Script and Output-**

str(iris)

# Installing Packages

install.packages("ClusterR")

install.packages("cluster")

# Loading package

library(ClusterR)

library(cluster)

library(caret)

# Removing initial label of Species from original dataset

iris\_data <- iris[, -5]

iris\_data

# Fitting K-Means clustering Model to training dataset

set.seed(240) # Setting seed

km <- kmeans(iris\_data, centers = 3, nstart = 20)

km

#Cluster identification for each observation

km$cluster

# Confusion Matrix

cm <- table(iris$Species, kmeans.re$cluster)

cm

# Model Evaluation and visualization

plot(iris\_data[c("Sepal.Length", "Sepal.Width")])

plot(iris\_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster)

plot(iris\_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster,main = "K-means with 3 clusters")

## Plotting cluster centers

km$centers

km$centers[, c("Sepal.Length", "Sepal.Width")]

# cex is font size, pch is symbol

points(km$centers[, c("Sepal.Length", "Sepal.Width")],col = 1:3, pch = 8, cex = 3)

## Visualizing clusters

y\_kmeans <- km$cluster

clusplot(iris\_data[, c("Sepal.Length", "Sepal.Width")],

         y\_kmeans,

         lines = 0,

         shade = TRUE,

         color = TRUE,

         labels = 2,

         plotchar = FALSE,

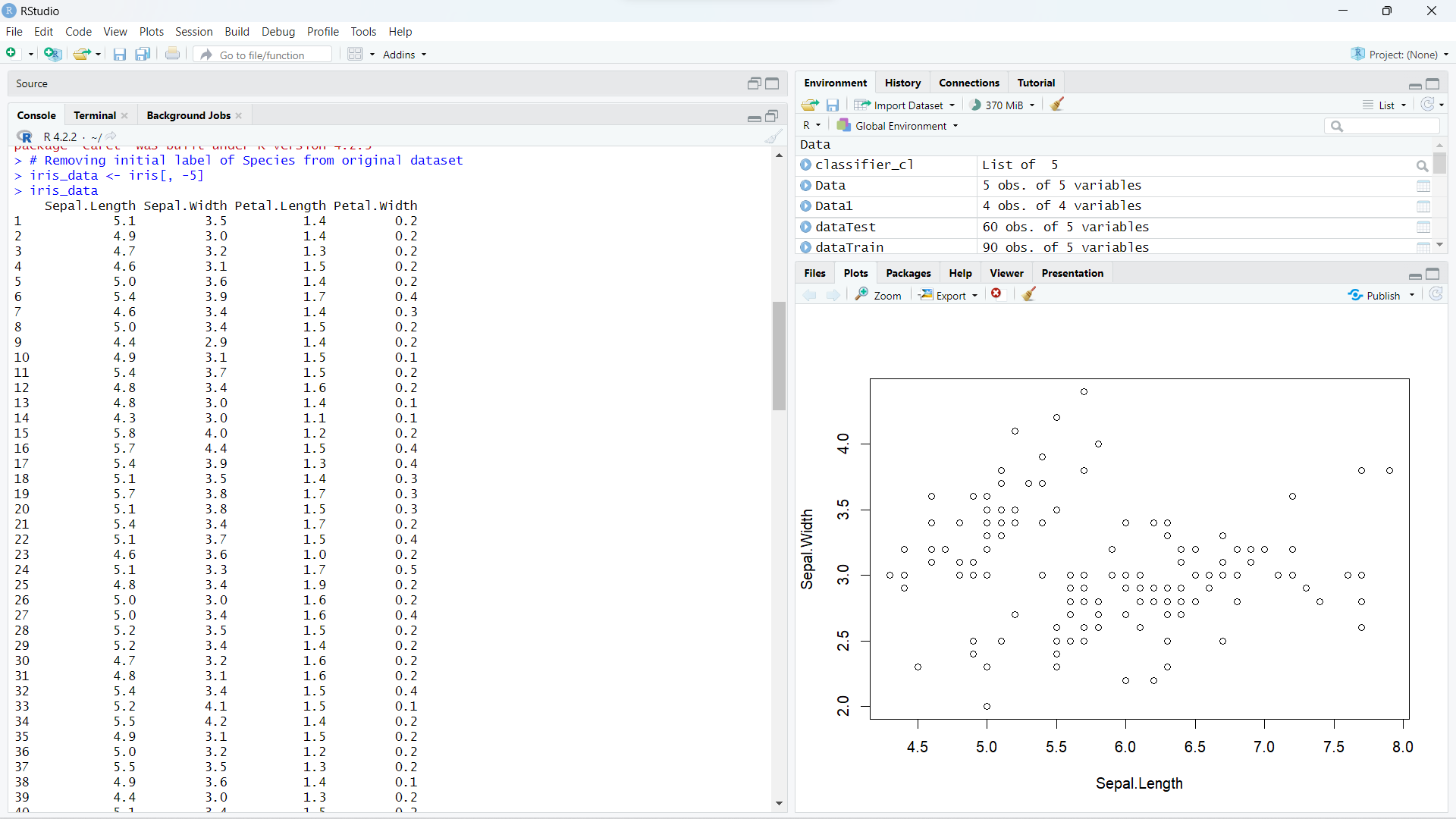
         span = TRUE,

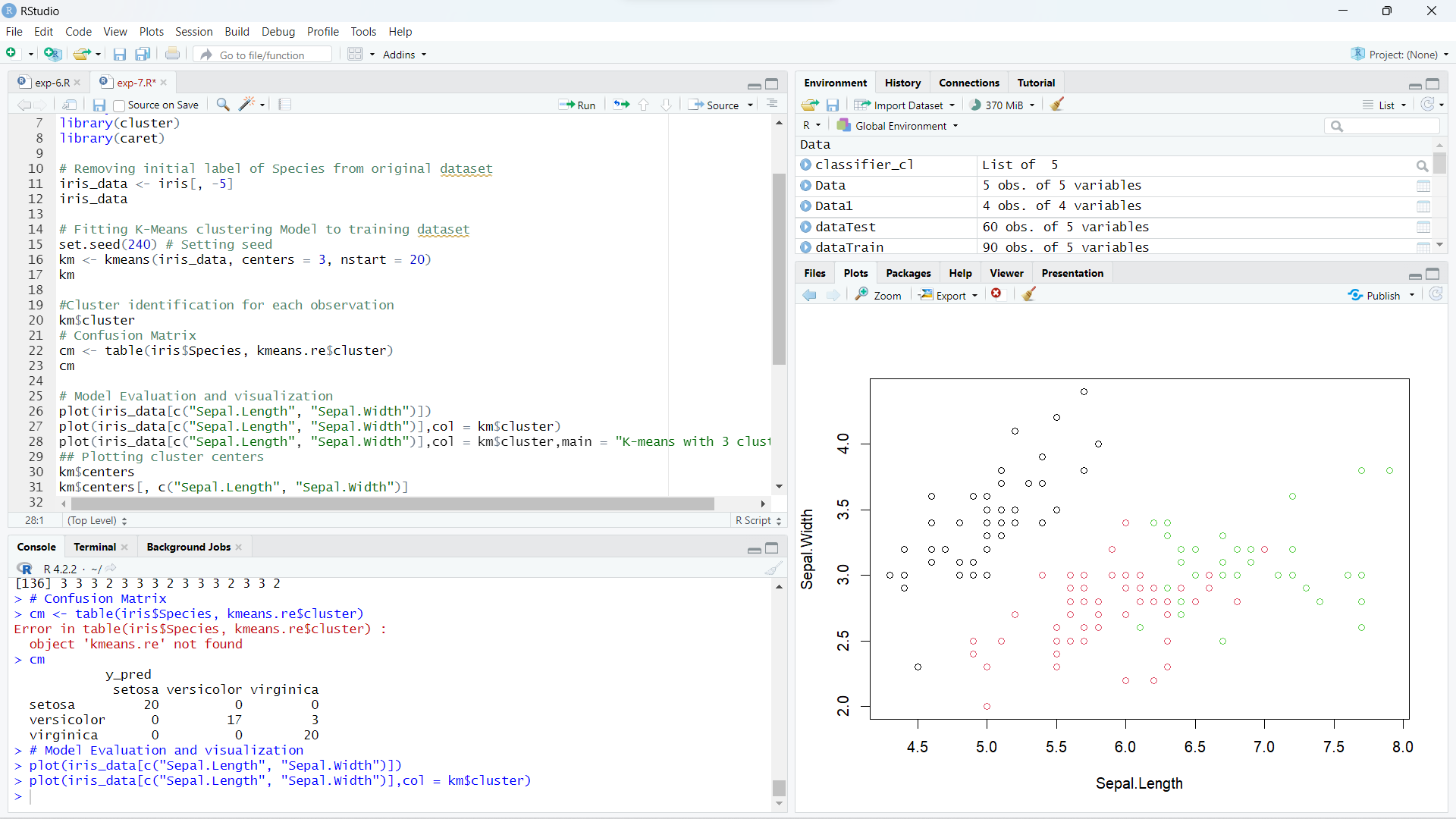
         main = paste("Cluster iris"),

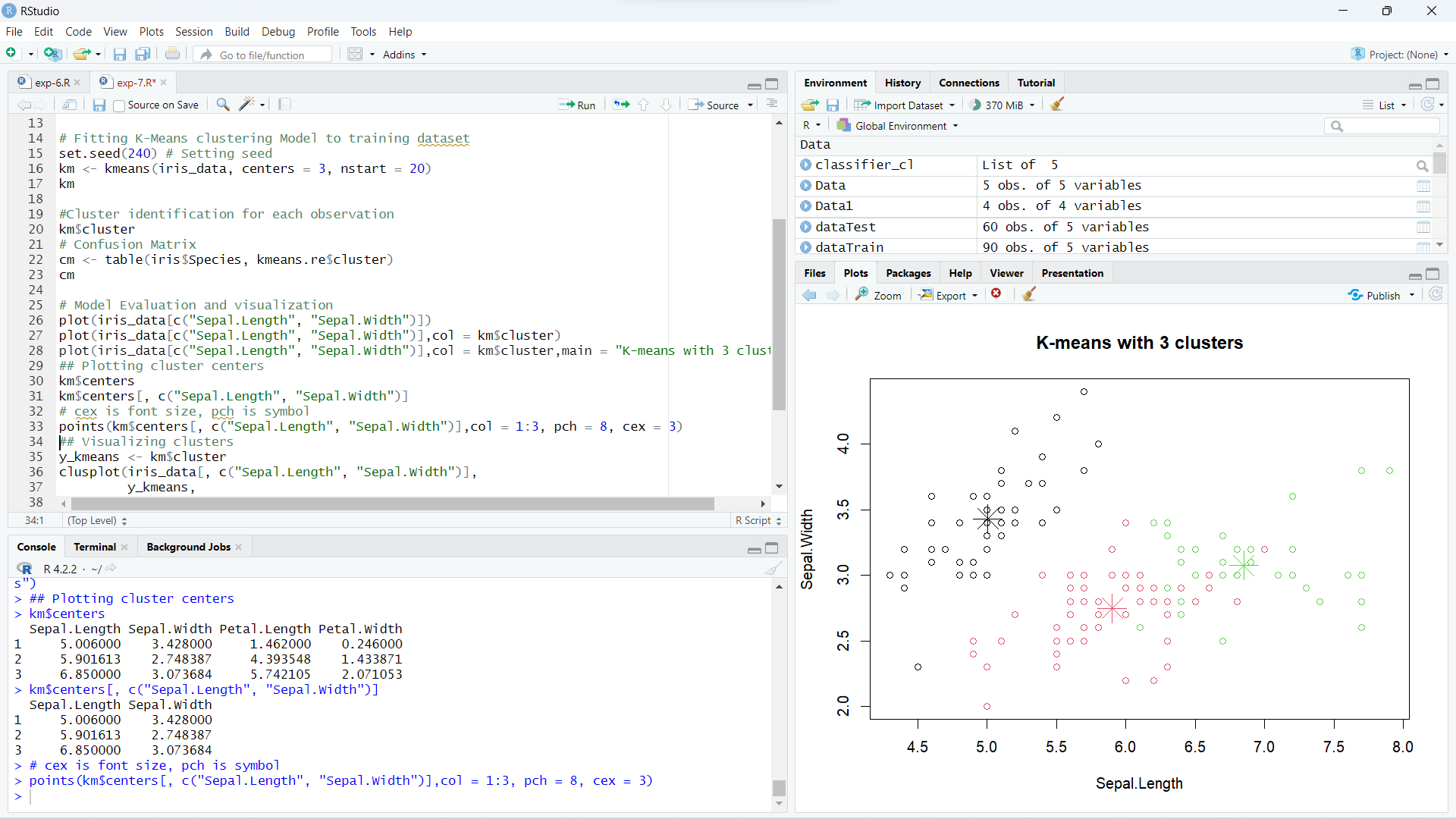
         xlab = 'Sepal.Length',

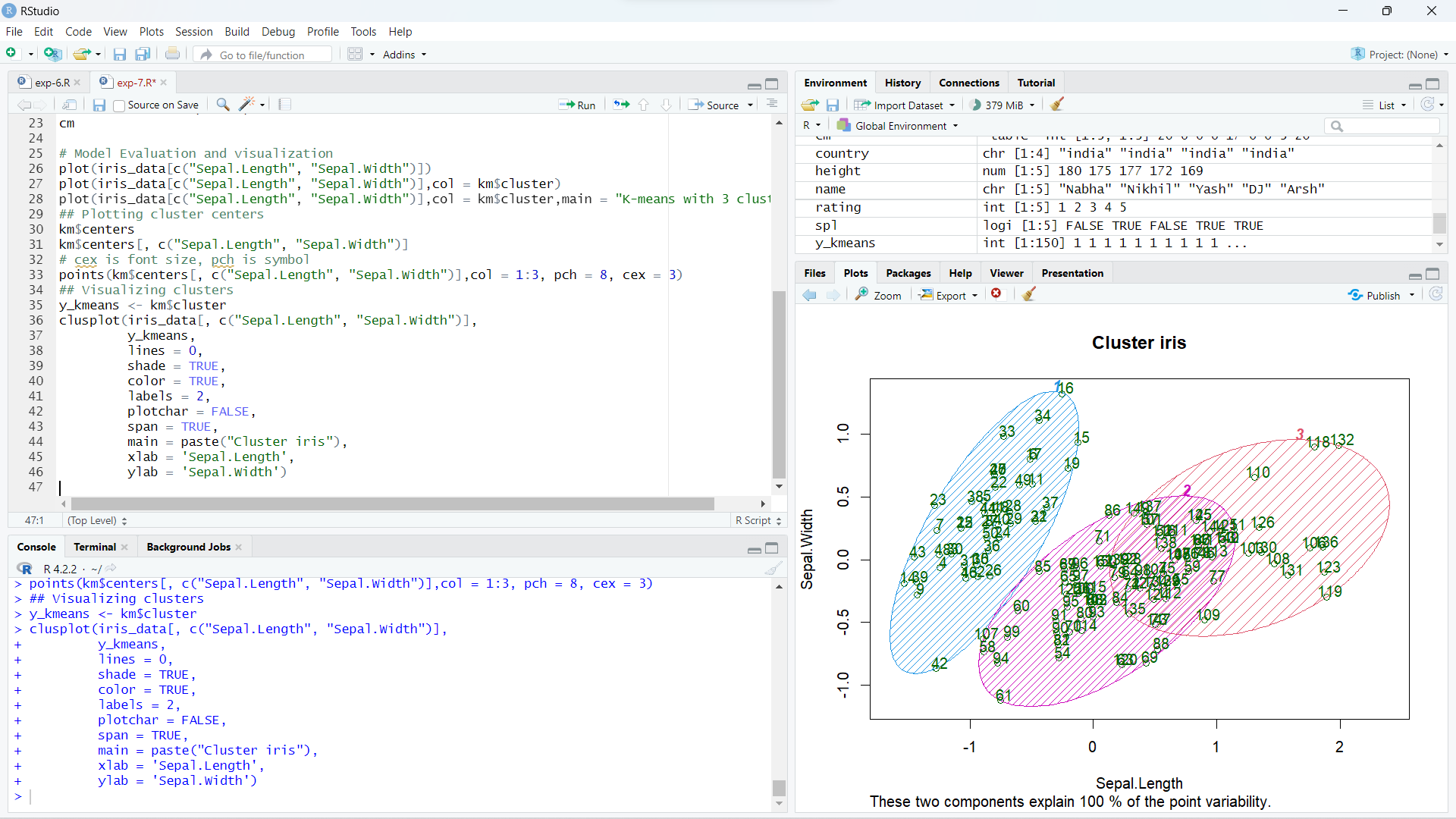
         ylab = 'Sepal.Width')

**Output-**

****





****