Set-II

1. (i)Write a function to find the factorial of a given number using “for” Loop

(ii) Create a 3x4 matrix with 12 random numbers between 1-100; have the matrix be filled our row

by-row, instead of column-by-column. Name the columns of the matrix *uno*, *dos*, *tres*, *cuatro*, and

the rows *x*, *y*, *z*. Scale the matrix by 10 and save the result.

(iii) Extract the column called “uno” as a vector from the original matrix and save the result

**Input:**

i.)# Function to calculate the factorial of a number

factorial <- function(n) {

result <- 1

for (i in 1:n) {

result <- result \* i

}

return(result)

}

# Example usage

number <- 5

factorial(number)

ii.)

# Create a matrix filled row-by-row with random numbers

set.seed(123)

random\_matrix <- matrix(sample(1:100, 12, replace = TRUE), nrow = 3, ncol = 4, byrow = TRUE)

iii.)

colnames(random\_matrix) <- c("uno", "dos", "tres", "cuatro")

rownames(random\_matrix) <- c("x", "y", "z")

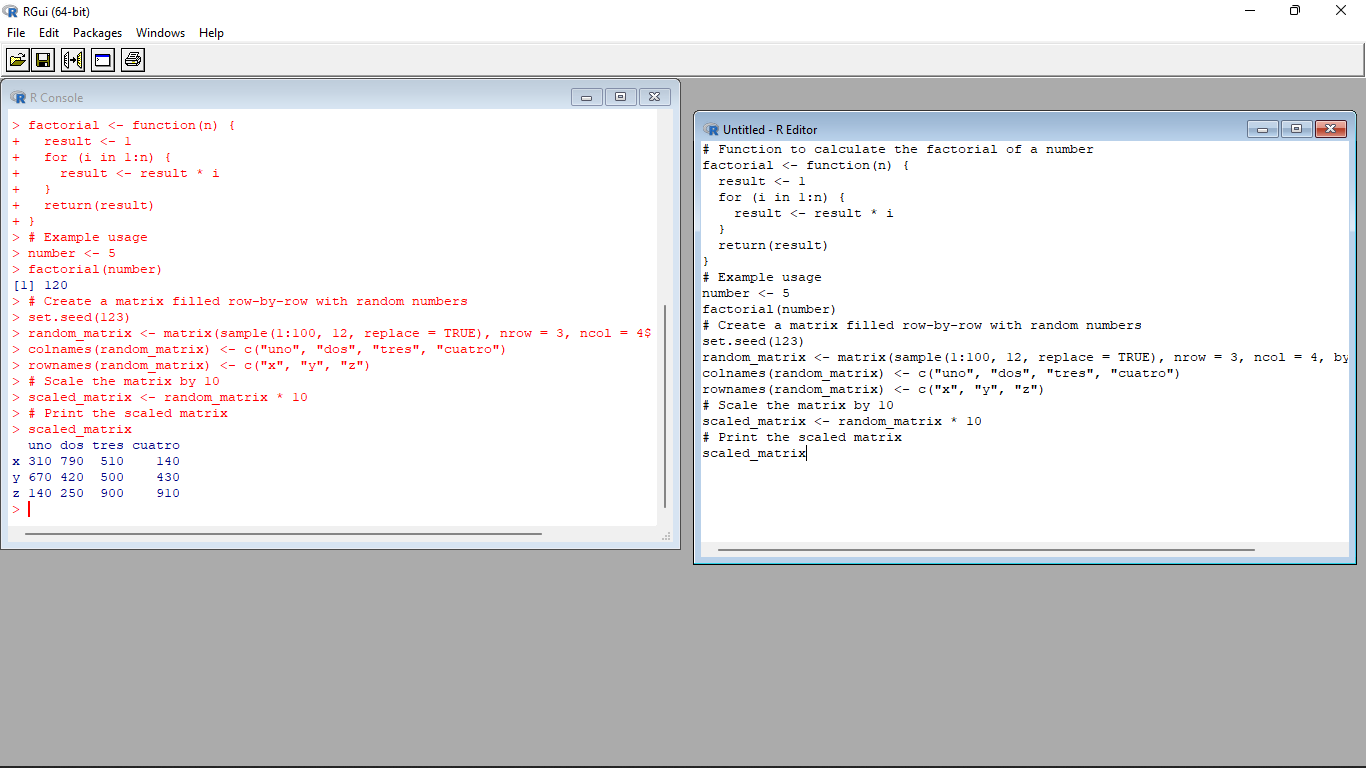
# Scale the matrix by 10

scaled\_matrix <- random\_matrix \* 10

# Print the scaled matrix

scaled\_matrix

**Output** :



2. In 1936, Edgar Anderson collected data to quantify the geographic variations of iris flowers. The data

set consists of 50 samples from each of the three sub-species ( *iris setosa*, *iris virginica,* and *iris*

*versicolor*).Four features were measured in centimeters (cm): the lengths and the widths of both

sepals and petals

(i)Find dimension, Structure, Summary statistics, Standard Deviation of all features.

(ii)Find mean and standard deviation of features groped by three species of Iris

flowers (Iris setosa, Iris virginica and Iris versicolor)

(iii)Find quantile value of sepal width and length

(iv)create new data frame named iris1 which have a new column name

**Sepal.Length.Cate** that categorizes “Sepal.Length” by quantile

(v) Average value of numerical varialbes by two categorical variables: Species and

Sepal.Length.Cate.

**Input:**

i.)# Load the iris dataset

data(iris)

# Dimension of the dataset

dim(iris)

# Structure of the dataset

str(iris)

# Summary statistics of all features

summary(iris)

# Standard deviation of all features

sapply(iris[, 1:4], sd)

ii.)# Mean and standard deviation grouped by species

aggregate(. ~ Species, data = iris, FUN = function(x) c(mean = mean(x), sd = sd(x)))

iii.)# Quantile values of sepal width and length

quantile(iris$Sepal.Width)

quantile(iris$Sepal.Length)

iv.)# Create new column with categorical values based on quantiles

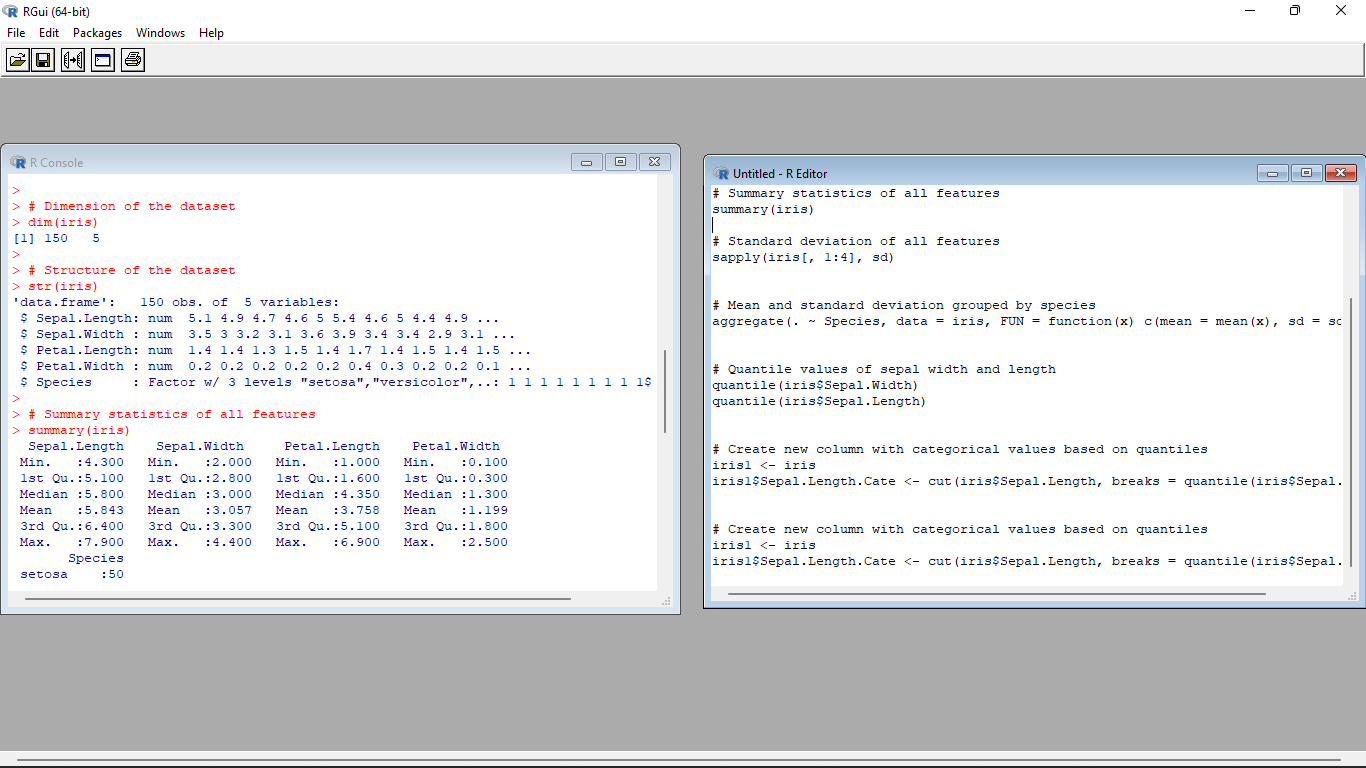
iris1 <- iris

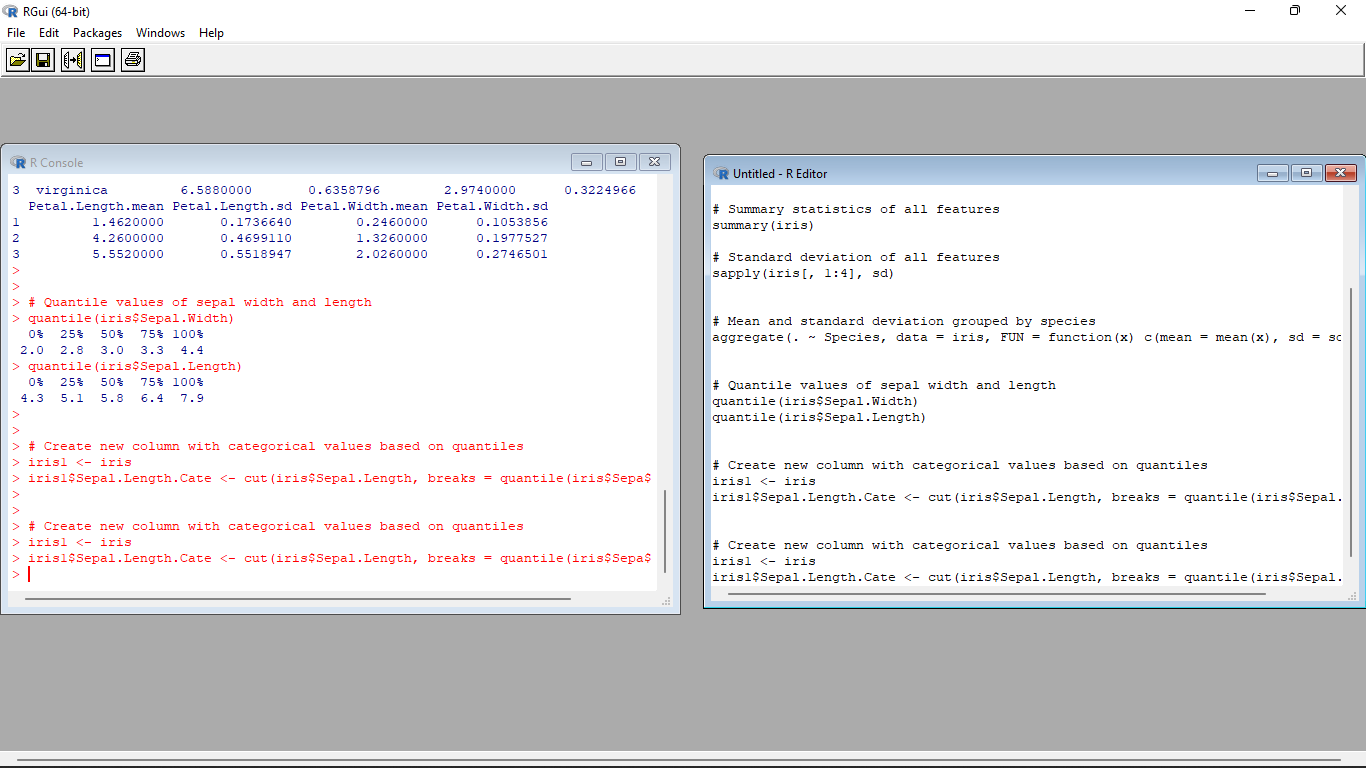
iris1$Sepal.Length.Cate <- cut(iris$Sepal.Length, breaks = quantile(iris$Sepal.Length))

V.)# Create new column with categorical values based on quantiles

iris1 <- iris

iris1$Sepal.Length.Cate <- cut(iris$Sepal.Length, breaks = quantile(iris$Sepal.Length))

**Output:**



3. (i)Plot Scatter plot between sepals width and length grouped by Species

(ii) Plot Scatter plot between petals width and length grouped by Species

(iii)Draw the Box plot for Sepals length grouped by Species

(iv) Draw the Box plot for petals length grouped by Species

(v)Find the correlation among the four features

**Input :**

i.)library(ggplot2)

# Scatter plot for Sepal Width vs. Length grouped by Species

ggplot(iris, aes(x = Sepal.Width, y = Sepal.Length, color = Species)) +

geom\_point() +

labs(x = "Sepal Width", y = "Sepal Length", title = "Scatter plot of Sepal Width vs. Length grouped by Species")

ii.)# Scatter plot for Petal Width vs. Length grouped by Species

ggplot(iris, aes(x = Petal.Width, y = Petal.Length, color = Species)) +

geom\_point() +

labs(x = "Petal Width", y = "Petal Length", title = "Scatter plot of Petal Width vs. Length grouped by Species")

iii).# Box plot for Sepal Length grouped by Species

ggplot(iris, aes(x = Species, y = Sepal.Length)) +

geom\_boxplot() +

labs(x = "Species", y = "Sepal Length", title = "Box plot of Sepal Length grouped by Species")

iv.)# Box plot for Petal Length grouped by Species

ggplot(iris, aes(x = Species, y = Petal.Length)) +

geom\_boxplot() +

labs(x = "Species", y = "Petal Length", title = "Box plot of Petal Length grouped by Species")

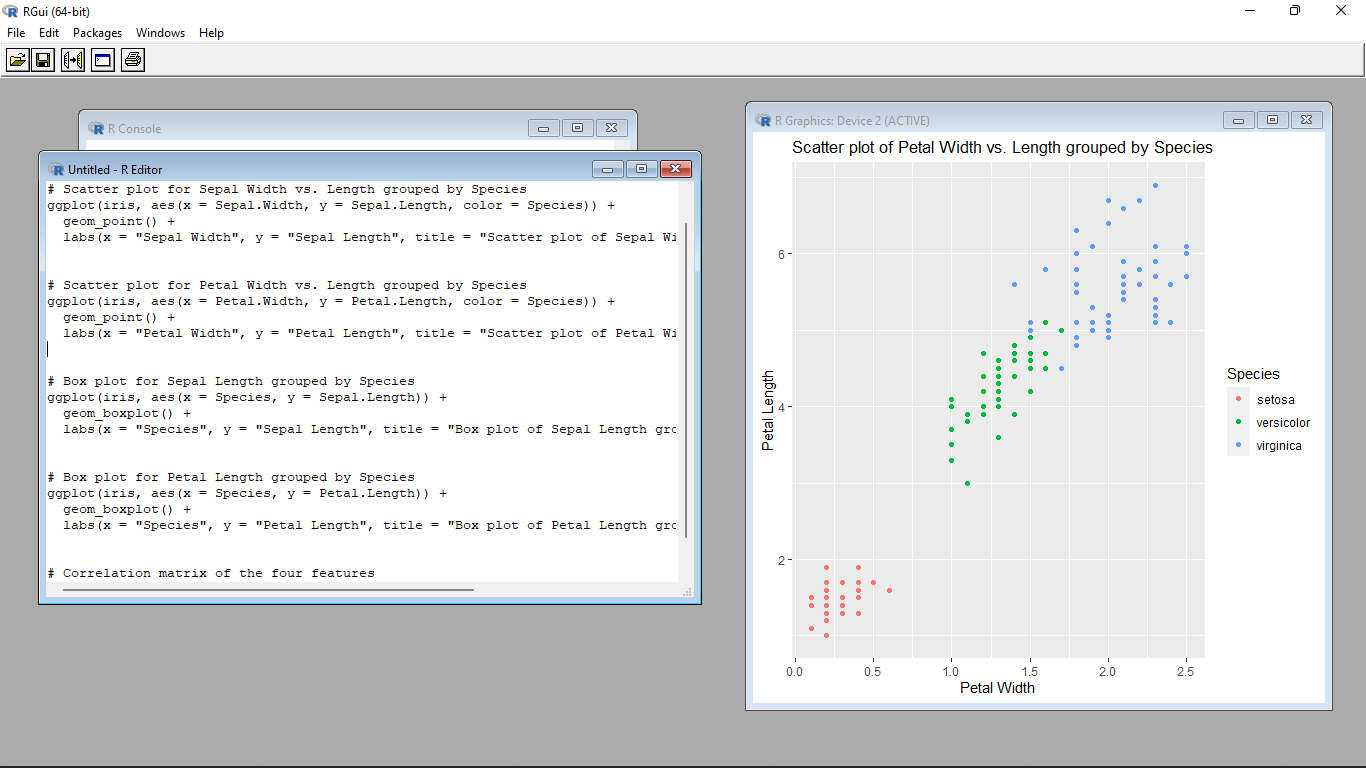
V.)

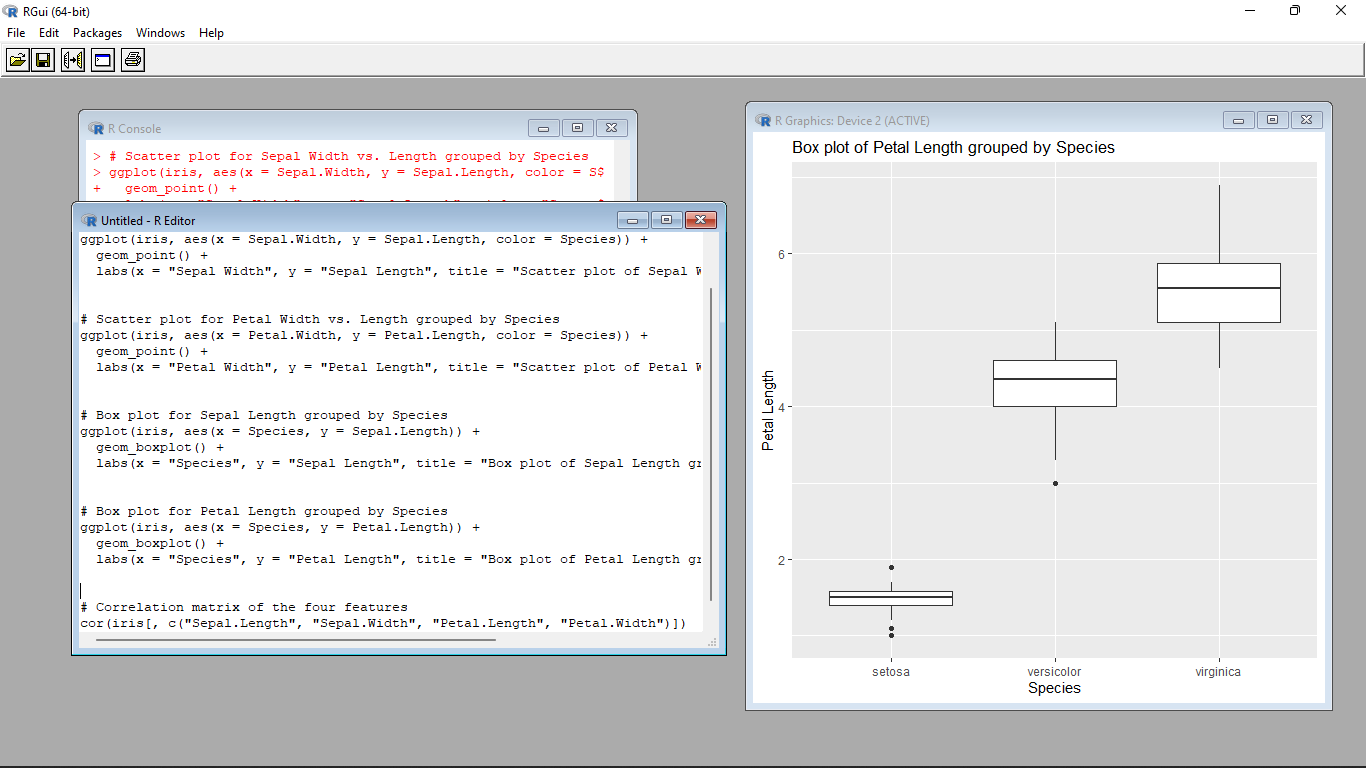
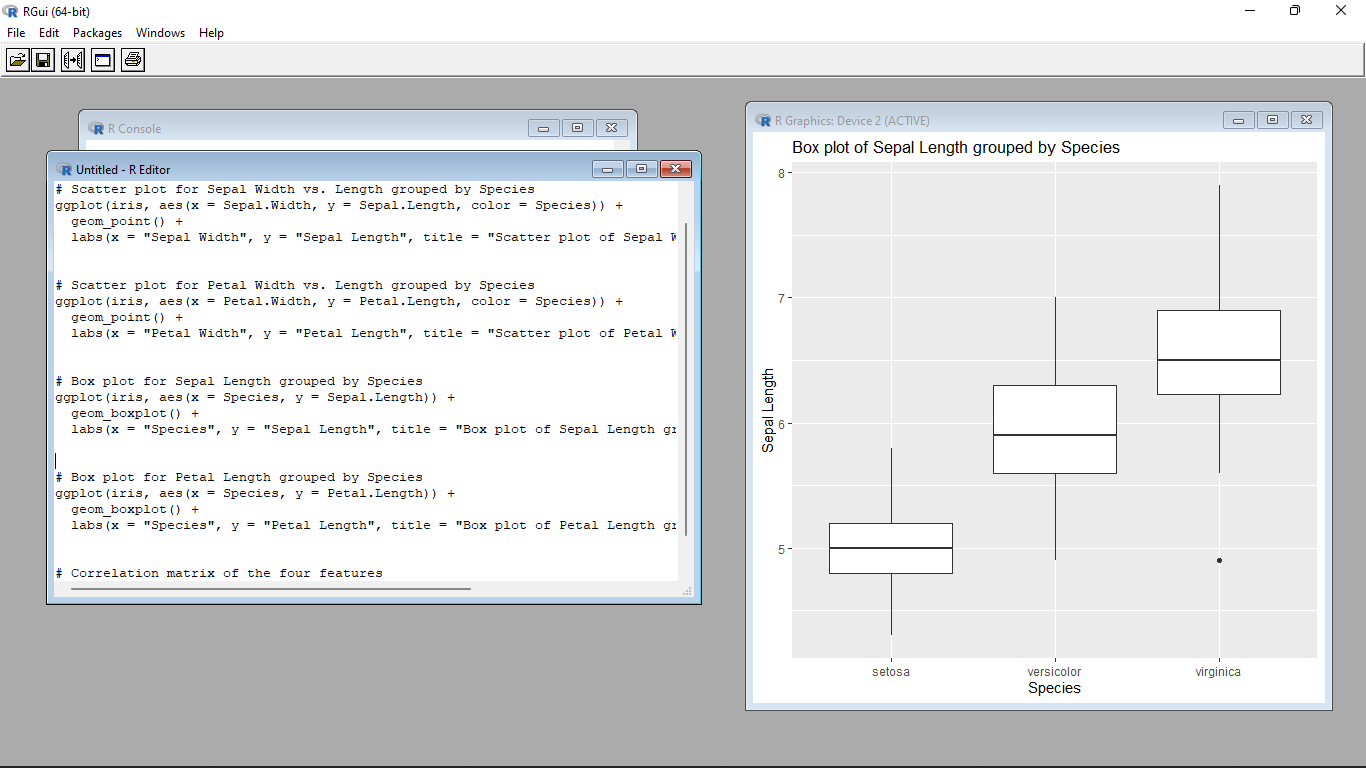
# Correlation matrix of the four features

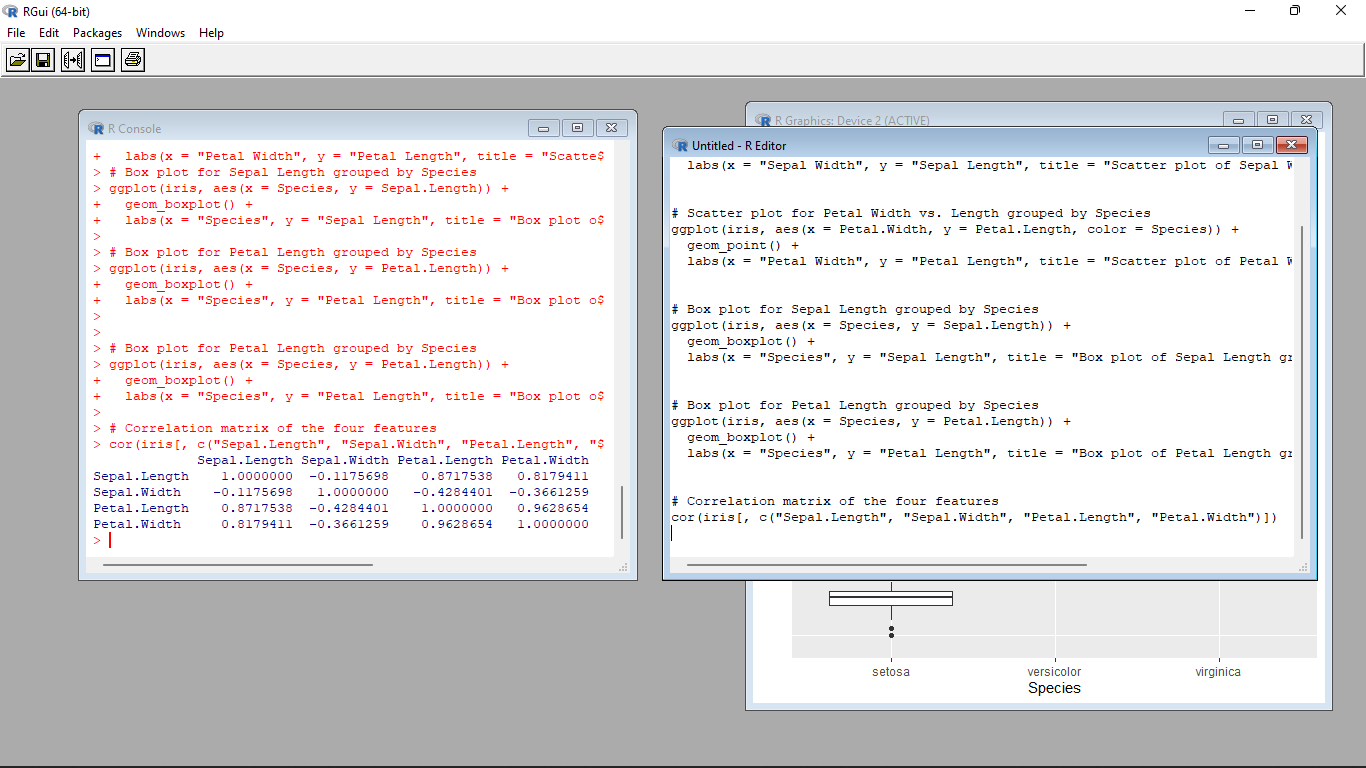
cor(iris[, c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width")])

**Output:**









4.(i) Randomly Sample the iris dataset such as 50% data for training and 50% for test

(ii)find summary statistics of above train and test dataset.

(iii)Create Logistics regression with train data

(iv)Predict the probability of the model using test data

(v)Create Confusion matrix for above test model

**Input :**

i.)# Set seed for reproducibility

set.seed(123)

# Randomly sample the iris dataset

train\_indices <- sample(1:nrow(iris), nrow(iris) \* 0.5)

train\_data <- iris[train\_indices, ]

test\_data <- iris[-train\_indices, ]

ii.)# Summary statistics of train dataset

summary(train\_data)

# Summary statistics of test dataset

summary(test\_data)

iii.)# Create logistic regression model

model <- glm(Species ~ ., data = train\_data, family = binomial)

print(model)

iv.)# Predict probabilities using the test data

probabilities <- predict(model, newdata = test\_data, type = "response")

print(probabilities)

V.)# Create confusion matrix

predicted\_classes <- ifelse(probabilities > 0.5, "versicolor", "setosa")

table(test\_data$Species, predicted\_classes)

**Output :**

