In [1]:
```
#Experiment No: 8 - Write a R program to K-Nearest Neighbour Algorithm.
```

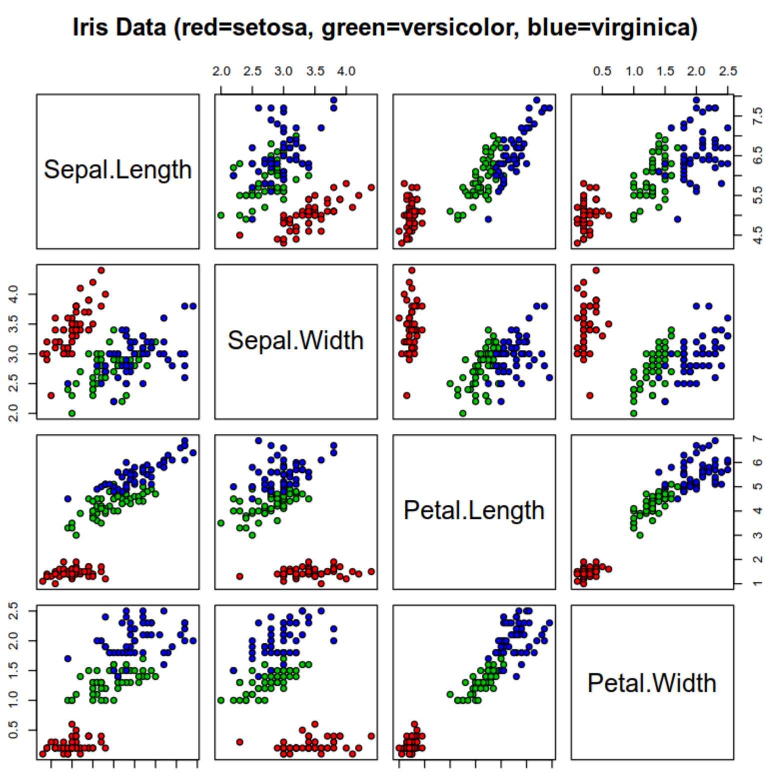
In [2]:
```
# Loading libraries
library(e1071)
library(caTools)
library(class)
```

In [3]:
```
# Show the dataset
print("Displaying the iris dataset:")
print(head(iris))  # Show the first few rows of the dataset

# Show the help page, with information about the dataset
?iris  # Access the help page for the iris dataset

# Create scatterplots of all pairwise combinations of the 4 variables in the
pairs(iris[1:4], main="Iris Data (red=setosa, green=versicolor, blue=virgini
      pch=21, bg=c("red","green3","blue")[unclass(iris$Species)])
```

```
[1] "Displaying the iris dataset:"
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
```



Edgar Anderson's Iris Data

Description

This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*.

Usage

```
iris
iris3
```

Format

`iris` is a data frame with 150 cases (rows) and 5 variables (columns) named `Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`, and `Species`.

`iris3` gives the same data arranged as a 3-dimensional array of size 50 by 4 by 3, as represented by S-PLUS. The first dimension gives the case number within the species subsample, the second the measurements with names `Sepal L.`, `Sepal W.`, `Petal L.`, and `Petal W.`, and the third the species.

Source

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, Part II, 179–188.

The data were collected by Anderson, Edgar (1935). The irises of the Gaspe Peninsula, *Bulletin of the American Iris Society*, **59**, 2–5.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole. (has `iris3` as `iris`.)

See Also

`matplot` some examples of which use `iris`.

Examples

```
dni3 <- dimnames(iris3)
ii <- data.frame(matrix(aperm(iris3, c(1,3,2)), ncol = 4,
                        dimnames = list(NULL, sub(" L.",".Length",
                                        sub(" W.",".Width", dni3
[[2]])))),
    Species = gl(3, 50, labels = sub("S", "s", sub("V", "v", dni3
[[3]]))))
all.equal(ii, iris) # TRUE
```

[Package *datasets* version 4.3.1]

In [4]:

```
# Split the data into training and testing sets (70% train, 30% test)
set.seed(42)  # For reproducibility
split <- sample.split(iris$Species, SplitRatio = 0.7)

# Create training and testing datasets
train_cl <- subset(iris, split == TRUE)
test_cl <- subset(iris, split == FALSE)

# Feature Scaling
train_scale <- scale(train_cl[, 1:4])
test_scale <- scale(test_cl[, 1:4])
```

In [5]:

```
# K-NN Model fitting and predictions using different values for K
results <- data.frame(K = integer(), Accuracy = numeric(), stringsAsFactors
```

In [6]:

```
# Loop through different K values
for (k in c(1, 3, 5, 7, 15, 19)) {
  # Fit the K-NN model
  classifier_knn <- knn(train = train_scale,
                        test = test_scale,
                        cl = train_cl$Species,
                        k = k)

  # Confusion Matrix
  cm <- table(test_cl$Species, classifier_knn)
  print(paste('Confusion Matrix for k =', k))
  print(cm)

  # Model Evaluation
  misClassError <- mean(classifier_knn != test_cl$Species)
  accuracy <- 1 - misClassError
  results <- rbind(results, data.frame(K = k, Accuracy = accuracy))

  print(paste('Accuracy for k =', k, '=', accuracy))
}
```

```
[1] "Confusion Matrix for k = 1"
            classifier_knn
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          2        13
[1] "Accuracy for k = 1 = 0.955555555555556"
[1] "Confusion Matrix for k = 3"
            classifier_knn
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          1        14
[1] "Accuracy for k = 3 = 0.977777777777778"
[1] "Confusion Matrix for k = 5"
            classifier_knn
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         14         1
  virginica       0          3        12
[1] "Accuracy for k = 5 = 0.911111111111111"
[1] "Confusion Matrix for k = 7"
            classifier_knn
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          2        13
[1] "Accuracy for k = 7 = 0.955555555555556"
[1] "Confusion Matrix for k = 15"
            classifier_knn
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          2        13
[1] "Accuracy for k = 15 = 0.955555555555556"
[1] "Confusion Matrix for k = 19"
            classifier_knn
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         14         1
  virginica       0          3        12
[1] "Accuracy for k = 19 = 0.911111111111111"
```

In [7]:

```
# Display all results
print("Summary of Accuracies:")
print(results)
```

```
[1] "Summary of Accuracies:"
   K  Accuracy
1  1 0.9555556
2  3 0.9777778
3  5 0.9111111
4  7 0.9555556
5 15 0.9555556
6 19 0.9111111
```