Iris.R

RMBIO11

2024-10-28

```r
# install.packages("randomForest")
# install.packages("e1071")
# install.packages("reshape2")
# install.packages("Iris")
# Load necessary libraries
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.4.1
```

```
## randomForest 4.7-1.2
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```r
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.4.1
```

```r
library(class)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.1
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:randomForest':
##
##     margin
```

```r
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.4.1
```

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.1
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:randomForest':
##
##      combine
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```r
setwd("C:/Users/RMBIO11/Documents/RSTUDIO/Iris")

# Load the Iris data
iris_data <- read.csv("iris.csv", header = TRUE)
str(iris_data)
```

```
## '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 ...
##  $ variety     : chr  "Setosa" "Setosa" "Setosa" "Setosa" ...
```
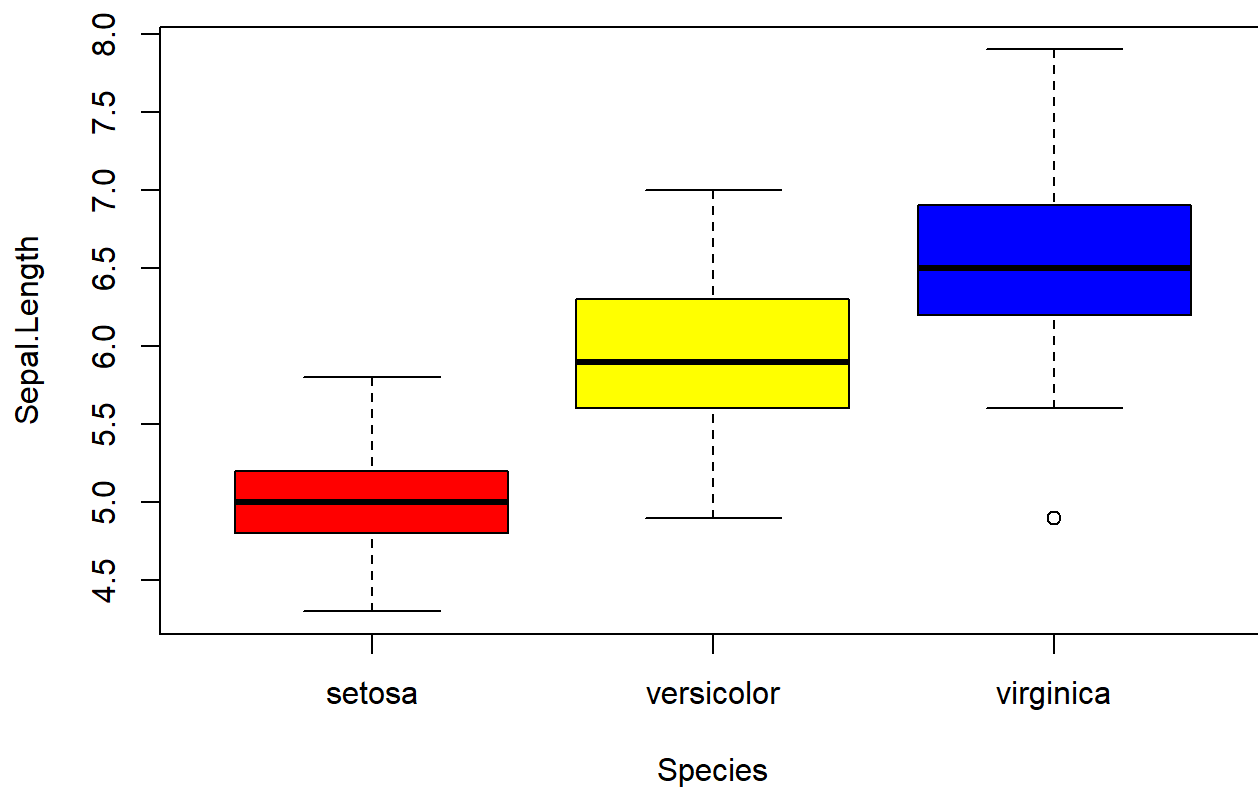
```r
# Graph Boxplot
boxplot(formula = Petal.Length ~ Species, data = iris, col=c("red","yellow", "blue"))
```
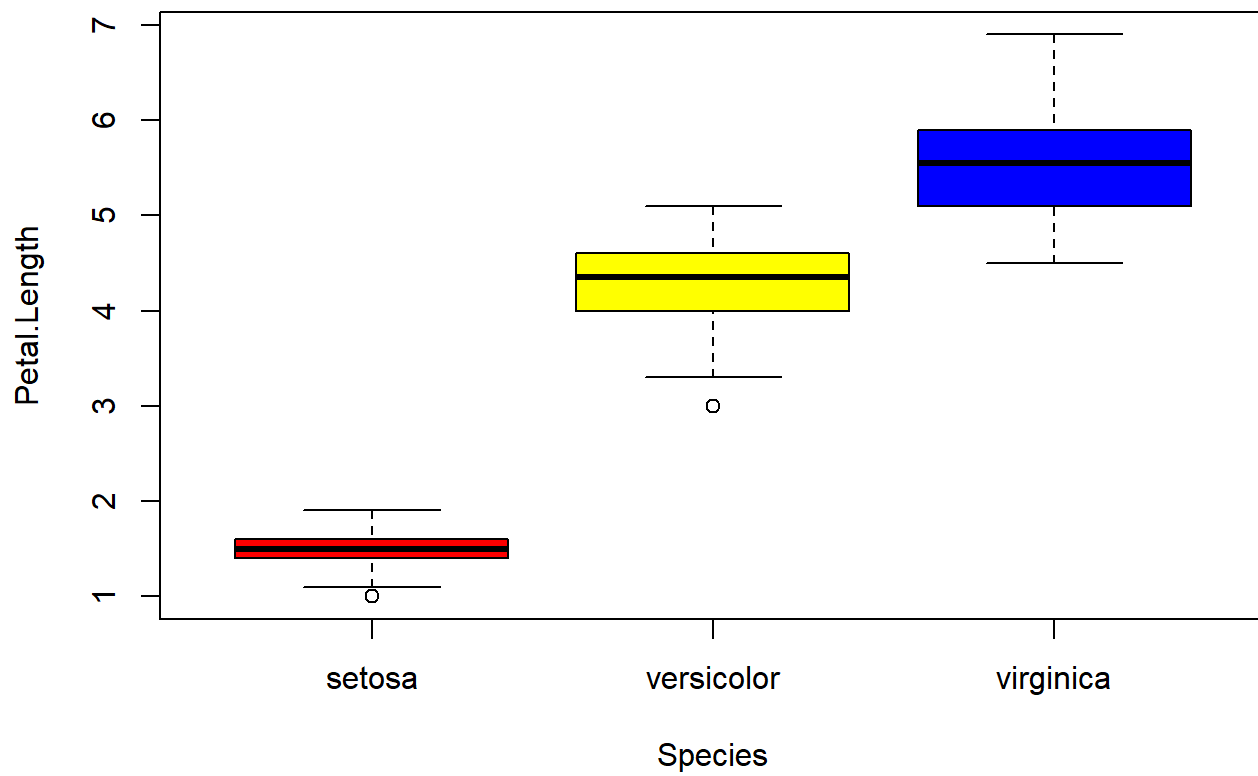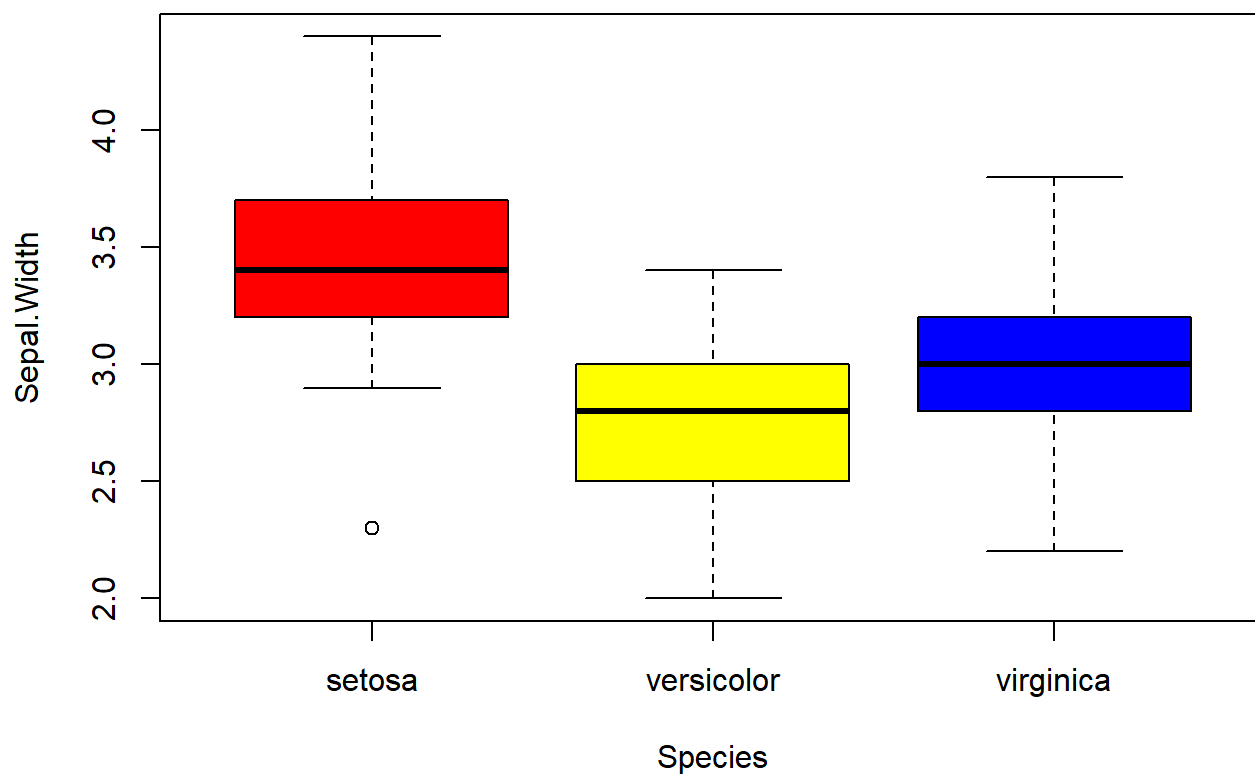
```
boxplot(formula = Petal.Width ~ Species, data = iris, col=c("red","yellow", "blue"))
```
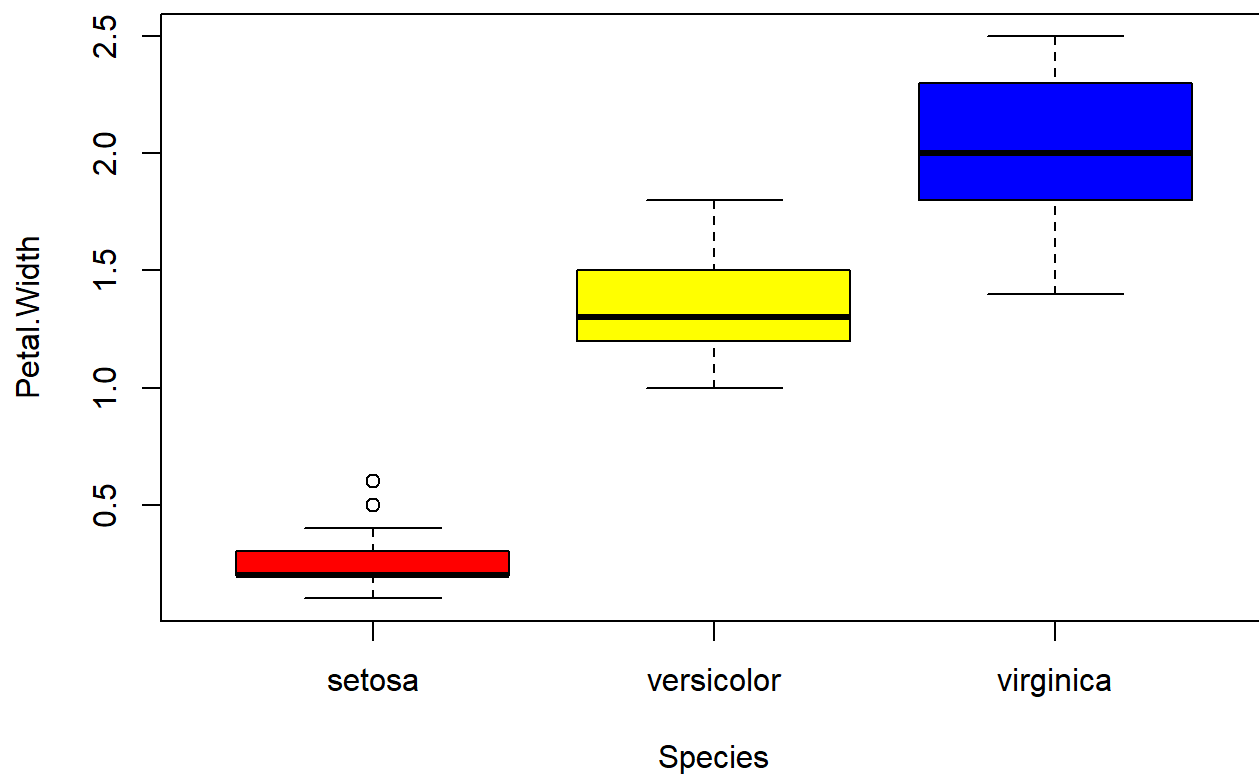
```
boxplot(formula = Sepal.Length ~ Species, data = iris, col=c("red","yellow", "blue"))
```

```
boxplot(formula = Sepal.Width ~ Species, data = iris, col=c("red","yellow", "blue"))
```

```
setosa <- iris[iris$Species == "setosa", ]
virginica <- iris[iris$Species == "virginica", ]
dim(setosa)
```

```
## [1] 50  5
```

```
dim(virginica)
```

```
## [1] 50  5
```

```
# Shapiro test on normality:
setosaN<- as.numeric(iris[iris$Species == "setosa", "Sepal.Length"])
table(duplicated(setosaN))
```

```
##
## FALSE   TRUE
##    15     35
```

```
ShapiroTestSetosa <- shapiro.test(setosaN)
print(ShapiroTestSetosa)
```

```
##
##   Shapiro-Wilk normality test
##
## data:  setosaN
## W = 0.9777, p-value = 0.4595
```

```
summary(ShapiroTestSetosa)
```

```
##           Length Class  Mode
## statistic 1      -none- numeric
## p.value   1      -none- numeric
## method    1      -none- character
## data.name 1      -none- character
```

```
# Shapiro test on normality:
virginicaN<- as.numeric(iris[iris$Species == "virginica", "Sepal.Length"])
table(duplicated(virginicaN))
```

```
##
## FALSE   TRUE
##    21     29
```

```
ShapiroTestVirg <- shapiro.test(virginicaN)
print(ShapiroTestVirg)
```

```
##
##   Shapiro-Wilk normality test
##
## data:  virginicaN
## W = 0.97118, p-value = 0.2583
```

```
summary(ShapiroTestVirg)
```

```
##           Length Class  Mode
## statistic 1      -none- numeric
## p.value   1      -none- numeric
## method    1      -none- character
## data.name 1      -none- character
```

```
# Comparition T-test Petal.Length of setosa and virginica
PairTTest<-t.test(x = setosa$Petal.Length, y = virginica$Petal.Length)
print(PairTTest)
```

```
##
##   Welch Two Sample t-test
##
## data:  setosa$Petal.Length and virginica$Petal.Length
## t = -49.986, df = 58.609, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.253749 -3.926251
## sample estimates:
## mean of x mean of y
##     1.462     5.552
```

```
summary(PairTTest)
```

```
##            Length Class  Mode
## statistic  1      -none- numeric
## parameter  1      -none- numeric
## p.value    1      -none- numeric
## conf.int   2      -none- numeric
## estimate   2      -none- numeric
## null.value 1      -none- numeric
## stderr     1      -none- numeric
## alternative 1     -none- character
## method     1      -none- character
## data.name  1      -none- character
```

```
# Comparition AOV Petal.Length of setosa and virginica
PetLenAOV<-aov(formula = Petal.Length ~ Species, data = iris)
print(PetLenAOV)
```

```
## Call:
##    aov(formula = Petal.Length ~ Species, data = iris)
##
## Terms:
##                 Species Residuals
## Sum of Squares  437.1028   27.2226
## Deg. of Freedom        2       147
##
## Residual standard error: 0.4303345
## Estimated effects may be unbalanced
```

```
summary(object = PetLenAOV)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Species       2  437.1  218.55    1180 <2e-16 ***
## Residuals   147   27.2    0.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```