#######################################

# Loading & Visualising Iris data set #

#######################################

# Load data from R datasets

data()

library(datasets)

data(iris)

# View the data

View(iris)

# Display summary statistics

head(iris, 5)

tail(iris, 5)

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

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

146 6.7 3.0 5.2 2.3 virginica

147 6.3 2.5 5.0 1.9 virginica

148 6.5 3.0 5.2 2.0 virginica

149 6.2 3.4 5.4 2.3 virginica

150 5.9 3.0 5.1 1.8 virginica

# summary()

summary(iris)

summary(iris$Sepal.Length)

summary(iris$Petal.Length)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100 setosa :50

1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300 versicolor:50

Median :5.800 Median :3.000 Median :4.350 Median :1.300 virginica :50

Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199

3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800

Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500

# Any missing data? Check the total number of

# missing values in the iris data-set

sum(is.na(iris))

[1] 0

# Install package "skimr"

# "skimr" - Compact & Flexible Summaries of data

# - expands on summary() by providing larger set of statistics

# install.packages("skimr")

library(skimr)

# Perform skim to display compact summary statistics of data

skim(iris)

--- Data Summary ------------------------

Values

Name iris

Number of rows 150

Number of columns 5

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Column type frequency:

factor 1

numeric 4

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Group variables None

- Variable type: factor -----------------------------------------------------------

# A tibble: 1 x 6

skim\_variable n\_missing complete\_rate ordered n\_unique top\_counts

\* <chr> <int> <dbl> <lgl> <int> <chr>

1 Species 0 1 FALSE 3 set: 50, ver: 50, vir: 50

- Variable type: numeric ----------------------------------------------------------

# A tibble: 4 x 11

skim\_variable n\_missing complete\_rate mean sd p0 p25 p50 p75 p100

\* <chr> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>

1 Sepal.Length 0 1 5.84 0.828 4.3 5.1 5.8 6.4 7.9

2 Sepal.Width 0 1 3.06 0.436 2 2.8 3 3.3 4.4

3 Petal.Length 0 1 3.76 1.77 1 1.6 4.35 5.1 6.9

4 Petal.Width 0 1 1.20 0.762 0.1 0.3 1.3 1.8 2.5

hist

<chr>

1 ▆▇▇▅▂

2 ▁▆▇▂▁

3 ▇▁▆▇▂

4 ▇▁▇▅▃

# Group data by Species & then perform skim to summarise

# using pipeline operator

iris %>%

dplyr::group\_by(Species) %>%

skim()

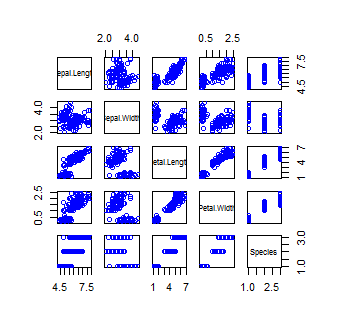
# Data Visualization

# R base plot()

# Panel plots

plot(iris)

plot(iris, col = "blue")



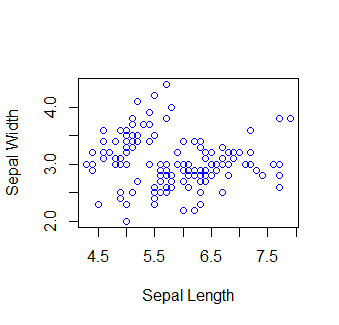
# Scatter plot

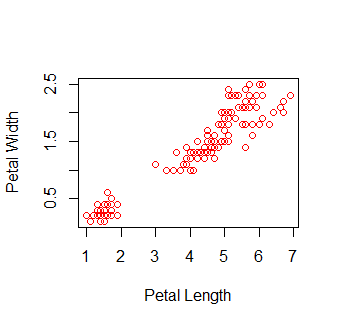
plot(iris$Sepal.Length, iris$Sepal.Width, col = "blue",

xlab = "Sepal Length", ylab = "Sepal Width")

plot(iris$Petal.Length, iris$Petal.Width, col = "red",

xlab = "Petal Length", ylab = "Petal Width")





# Plotting with corrplot: Correlation Analysis

# Visualising the Correlation Matrix

install.packages("corrplot")

library(corrplot)

# Compute Correlation Matrix

CorMat <- cor(iris[,1:4])

head(round(CorMat,2))

Sepal.Length Sepal.Width Petal.Length Petal.Width

Sepal.Length 1.00 -0.12 0.87 0.82

Sepal.Width -0.12 1.00 -0.43 -0.37

Petal.Length 0.87 -0.43 1.00 0.96

Petal.Width 0.82 -0.37 0.96 1.00

# Visualise Correlation Matrix

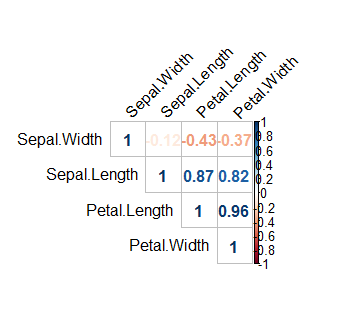
corrplot(CorMat, method = "circle")

corrplot(CorMat, method = "number",

type = "upper", order = "hclust",

tl.col = "black", tl.srt = 45,

sig.level = 0.05, insig = "blank")



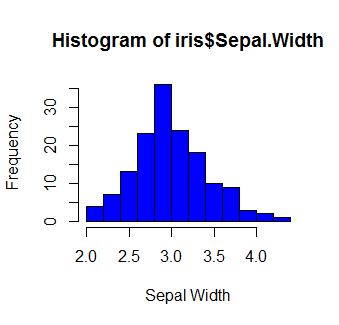
# Histogram

hist(iris$Sepal.Width, col = "blue", xlab = "Sepal Width")

hist(iris$Sepal.Length, col = "blue", xlab = "Sepal Length")

hist(iris$Petal.Width, col = "red", xlab = "Petal Width")

hist(iris$Petal.Length, col = "red", xlab = "Petal Length")



# Box-plot: Using ggplot2 package

# Uncomment the lines if packages are not installed

# install.packages("ggplot2")

# install.packages("dplyr")

library(ggplot2)

library(dplyr)

ggplot2::ggplot(iris, aes(x=iris$Sepal.Length,y= iris$Species))+

geom\_boxplot(outlier.color = "red", outlier.size = 3) +

coord\_flip()

ggplot2::ggplot(iris, aes(x=iris$Sepal.Width,y= iris$Species))+

geom\_boxplot(outlier.color = "red", outlier.size = 3) +

coord\_flip()

ggplot2::ggplot(iris, aes(x=iris$Petal.Length,y= iris$Species))+

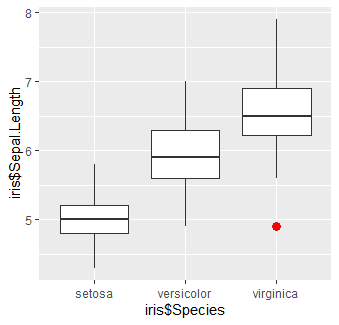
geom\_boxplot(outlier.color = "red", outlier.size = 3) +

coord\_flip()

ggplot2::ggplot(iris, aes(x=iris$Petal.Width,y= iris$Species))+

geom\_boxplot(outlier.color = "red", outlier.size = 3) +

coord\_flip()



# Feature plots:

# Integrating all features in single graph

# Box-plot: Features in X-axis, Factor(Species) in Y-axis

install.packages("caret")

library(caret)

featurePlot(x = iris[,1:4],

y = iris$Species,

plot = "box",

strip=strip.custom(par.strip.text=list(cex=.7)),

scales = list(x = list(rot=90),

y = list(relation="free")),

layout = c(4,1),

auto.key = list(columns = 2))

