Iris DataSet

Skyler Preston

MIS6240

Data Warehousing

Hands-On Lab 1

3-26-2023

Professor

Dr. Amitava Karmaker

**Dataset Name**: iris

**Description**- The dataset provides measurements in centimeters of sepal length and width and petal length and width for 50 flowers with thee species types Iris setosa , versicolor, and virgincia.

**Format:** The dataset contains 150 rows and 5 variables

**Details:** The data set is great for showing various analysis techniques of graphical methos and diagrams or plots.

In order to generate the basic information in R you need to type a command. See Below

Library(datasets) will load the data set manually.

Below are three commands to get basic information about a data set.

**Head(iris);**

A picture containing text

Description automatically generated

This command shows the general information of the dataset broken into variables and operations.

**Summary(iris);**

Text

Description automatically generated

This command shows the general statistical summary of the data.

**Str(iris);**

Text

Description automatically generated with medium confidence

This command shows the structure of the table. This command also breaks out the dataset into a list.

**Two Data Preparation Task:**

1. **Extracting a variable**

**Command:**

select(iris, Species, Petal.Width) # by name

**Useful:** When selecting data in a table format this command allows you to organize the table in a row format. This is helpful for organizing all rows in the dataset to follow an order of 1-150. The reason why this is necessary when performing analysis you can make quick comparison when the data is organized by the variables.

Table

Description automatically generated

Text

Description automatically generated

A picture containing table

Description automatically generated

**Command:**

select(iris, 5, 4) # by column index

**Useful:** This allows you to extract the index of the object of the dataset. In this case the vector organized the data in a numeric order. The index of 5 and will break it out by the 4 columns. This is helpful for extracting a variable from the dataset.

Text

Description automatically generated with medium confidence

**Graphs:**

**Command:** boxplot(iris)

**Graph Type**: Box Plot

The box plot is used to show patterns of a group. In this case the data shows the measurements of iris flowers by the length and the width. The graph indicates the sepal. Length is just under 6.0 centimeters. On the other side of the box plot the average species is around 2.0 centimeters. The petal length is just over 4.0 centimeters. The additional analysis that this can give you about iris flowers is the median, first quartile and third quartile.

Chart, box and whisker chart

Description automatically generated

**Bar Char**

**Command :**

par(mfrow=c(3,1))

barplot(iris$Petal.Length) #Creating simple Bar Graph

**Graph Type:** Bar Chart

The graph is sorting all the iris flowers by the petal.length. This type of graph can be useful to see all three-vector petal. Length side by side to help determine what species is consider the taller petal flower type of the iris species.

Chart

Description automatically generated

**Scatter Plot**

**Command:**

plot(x=iris$Petal.Length) #Simple Scatter Plot

**Graph Type :** Scatter Plot

The usefulness of a scatter plot with the iris dataset would allow you to see petal and length and see if it numerically matches the visual check between the relation of the other vectors.

Text

Description automatically generated

**References**

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