

Scatter/ dot Plot, Histogram, and Boxplot

DATA621 Blog 01

Zhi Ying Chen

27 November 2020

Data Visualization

Today's world is full of data, and we must perform proper analysis on it to obtain meaningful insights. Data visualization is not only one of most important tools can show us the graph from the data, but also as a vital tool that can dig out possible key insights from the data. If the analysis result cannot be displayed correctly, it will not be effectively communicated to the desired audience.

Pro: When you can plot the data you can has an insight into how your data its distributed or find missing values easier by just look at it on a graph. Selecting the right plot graph its also important because this can make easier to see through the data.

Con: if not selecting the best visualization graph can also lead to the wrong assumptions in the analysis.

Today, we are going to use Scatter/ dot Plot, Histogram, and Boxplot to visualize data, which can be to see the relationship between variables, the distribution of variables or the comparison between variables.

Load Packages

```
library(recommenderlab)
```

```
## Warning: package 'recommenderlab' was built under R version 3.5.3
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 3.5.3
```

```
## Loading required package: arules
```

```
## Warning: package 'arules' was built under R version 3.5.3
```

```
##
```

```
## Attaching package: 'arules'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      abbreviate, write
```

```
## Loading required package: proxy
```

```
## Warning: package 'proxy' was built under R version 3.5.3
```

```
##
```

```
## Attaching package: 'proxy'
```

```
## The following object is masked from 'package:Matrix':
```

```
##
```

```

##      as.matrix
## The following objects are masked from 'package:stats':
##
##      as.dist, dist
## The following object is masked from 'package:base':
##
##      as.matrix
## Loading required package: registry
library(ggplot2)
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.5.3
## -- Attaching packages ----- tidyverse 1.3.0
## v tibble  3.0.1      v dplyr    0.8.5
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
## v purrr   0.3.4
## Warning: package 'tibble' was built under R version 3.5.3
## Warning: package 'tidyr' was built under R version 3.5.3
## Warning: package 'readr' was built under R version 3.5.3
## Warning: package 'purrr' was built under R version 3.5.3
## Warning: package 'dplyr' was built under R version 3.5.3
## Warning: package 'forcats' was built under R version 3.5.3
## -- Conflicts ----- tidyverse_conflicts()
## x tidyr::expand() masks Matrix::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## x tidyr::pack()   masks Matrix::pack()
## x dplyr::recode() masks arules::recode()
## x tidyr::unpack() masks Matrix::unpack()
library(dplyr)
library(kableExtra)

## Warning: package 'kableExtra' was built under R version 3.5.3
##
## Attaching package: 'kableExtra'
## The following object is masked from 'package:dplyr':
##
##      group_rows
library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.5.3
##
## Attaching package: 'gridExtra'

```

```
## The following object is masked from 'package:dplyr':
##
##      combine
library(rmdformats)

## Warning: package 'rmdformats' was built under R version 3.5.3
library(formattable)

## Warning: package 'formattable' was built under R version 3.5.3
##
## Attaching package: 'formattable'
## The following object is masked from 'package:recommenderlab':
##
##      normalize
library(scales)

##
## Attaching package: 'scales'
## The following objects are masked from 'package:formattable':
##
##      comma, percent, scientific
## The following object is masked from 'package:purrr':
##
##      discard
## The following object is masked from 'package:readr':
##
##      col_factor
```

Read Data

```
set.seed(3)
data(iris)
summary(iris)
```

```
##      Sepal.Length      Sepal.Width      Petal.Length      Petal.Width
## Min.      :4.300      Min.      :2.000      Min.      :1.000      Min.      :0.100
## 1st Qu.:5.100      1st Qu.:2.800      1st Qu.:1.600      1st Qu.:0.300
## Median :5.800      Median :3.000      Median :4.350      Median :1.300
## 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
##      Species
## setosa      :50
## versicolor:50
## virginica   :50
##
##
##
```

Data Exploration

This data is about iris. The iris dataset (included with R), which introduced by Ronald Fisher in his 1936 paper The use of multiple measurements in taxonomic problems. It contains four measurements for 150 flowers representing three species of iris (Iris setosa, versicolor and virginica).

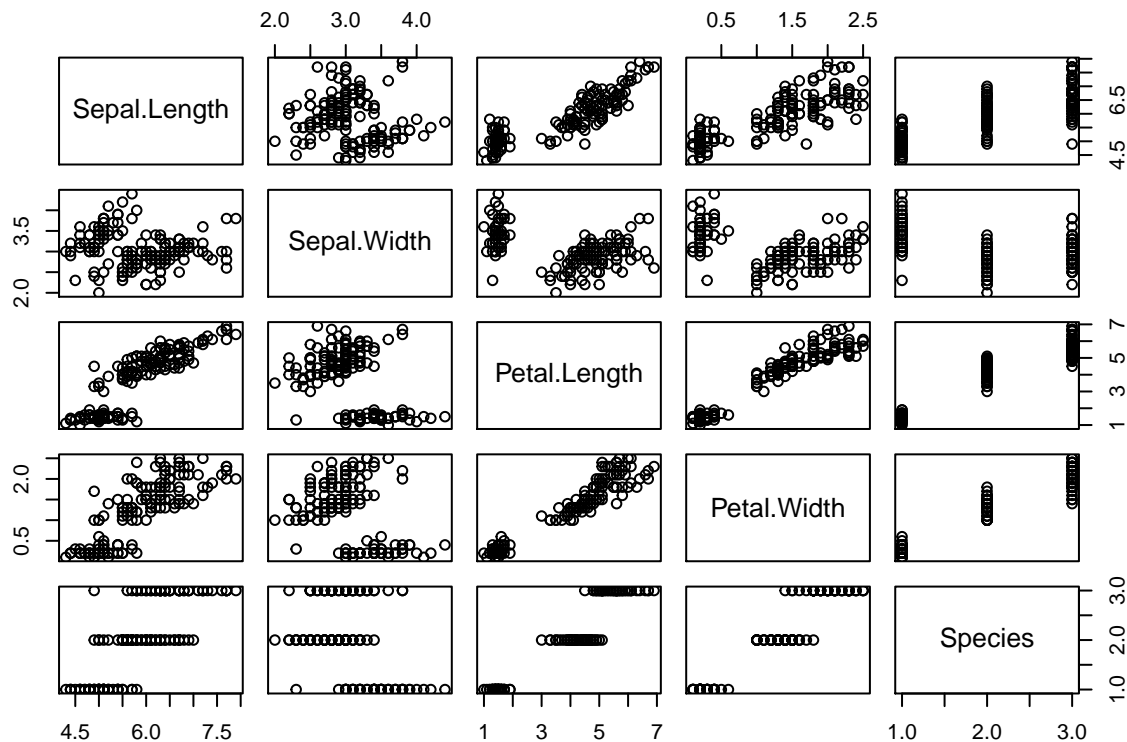
It is a 150 rows x 5 columns rating matrix of class 'realRatingMatrix' with 750 ratings.

```
dim(iris)
```

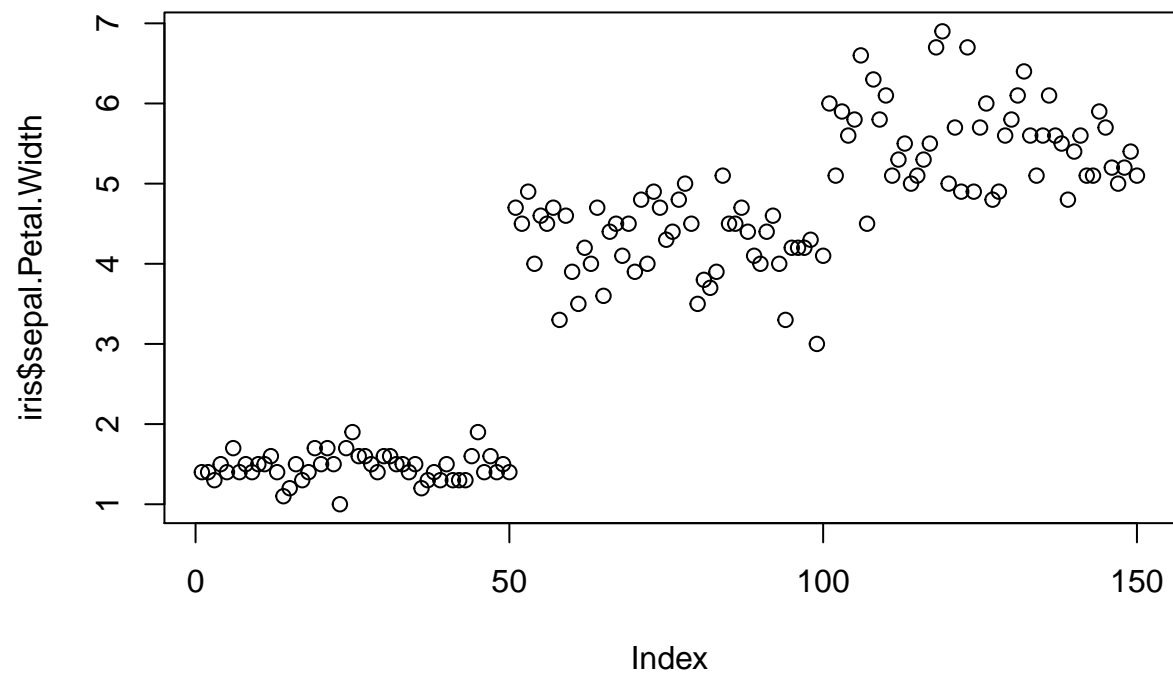
```
## [1] 150  5
```

Scatter / dot plot

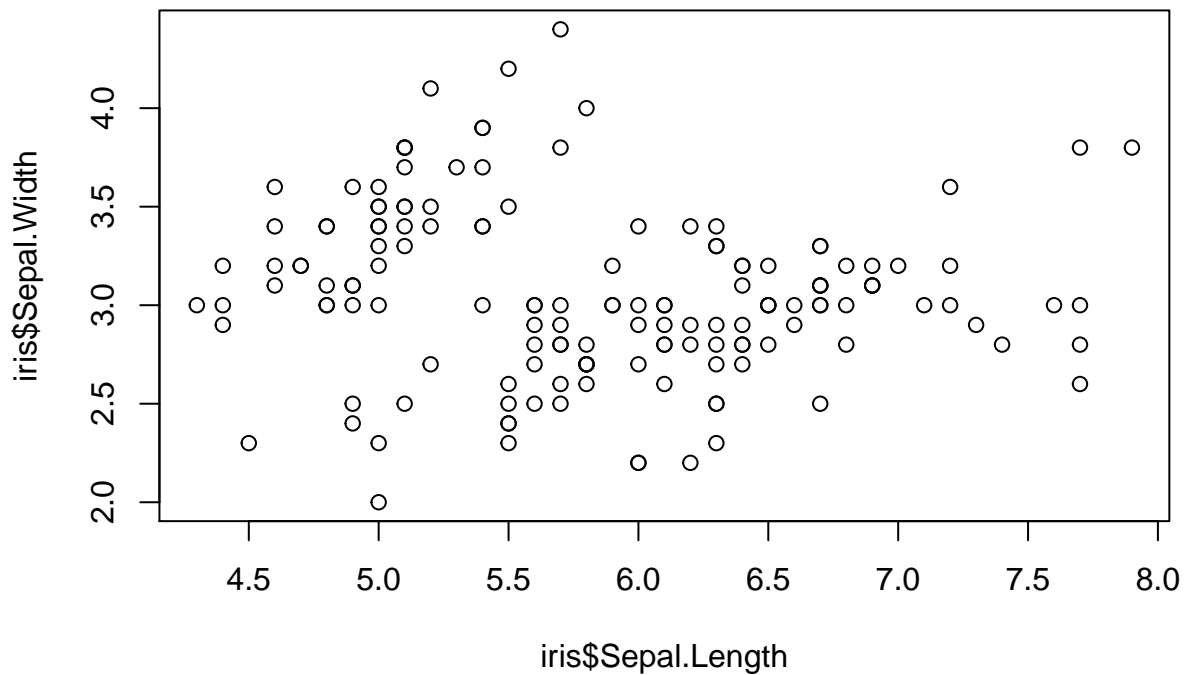
```
plot(iris)
```



```
plot(iris$Petal.Length, iris$sepal.Petal.Width)
```



```
plot(iris$Sepal.Length, iris$Sepal.Width)
```



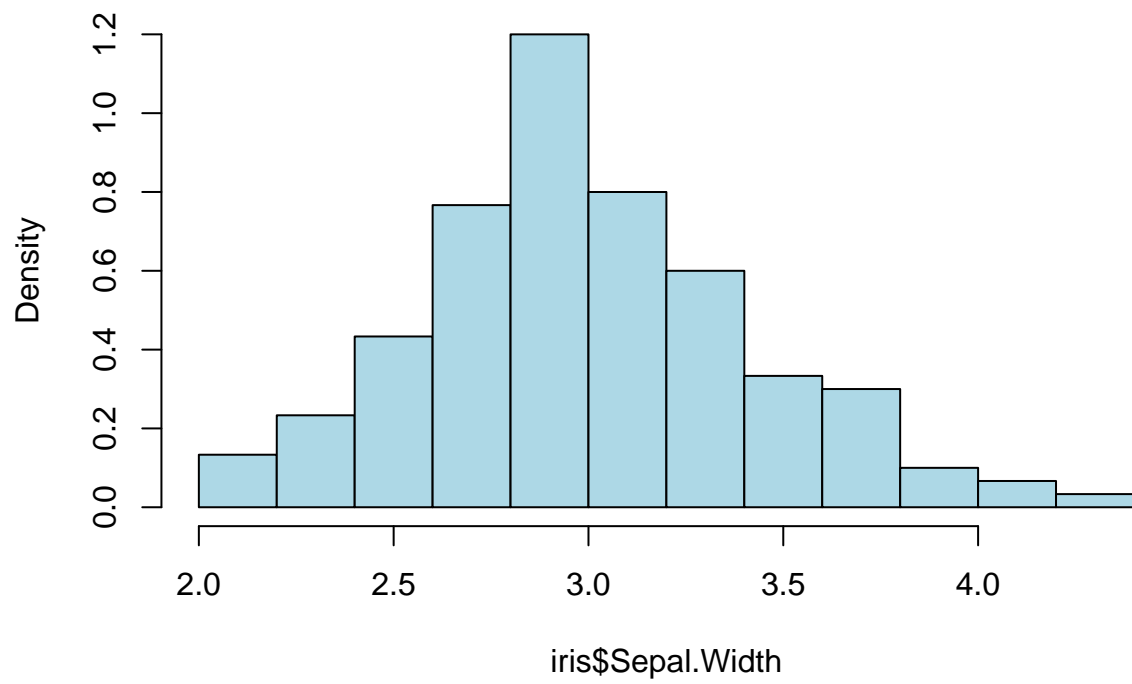
We get a matrix of scatterplots which is a correlation matrix of all the columns. And we can quickly discover the relationship between variables by merely looking at the plots drawn between them.

Histogram

A histogram is quite similar to a bar chart except that it groups values into continuous ranges. A histogram represents the frequencies of values of a variable bucketed into ranges.

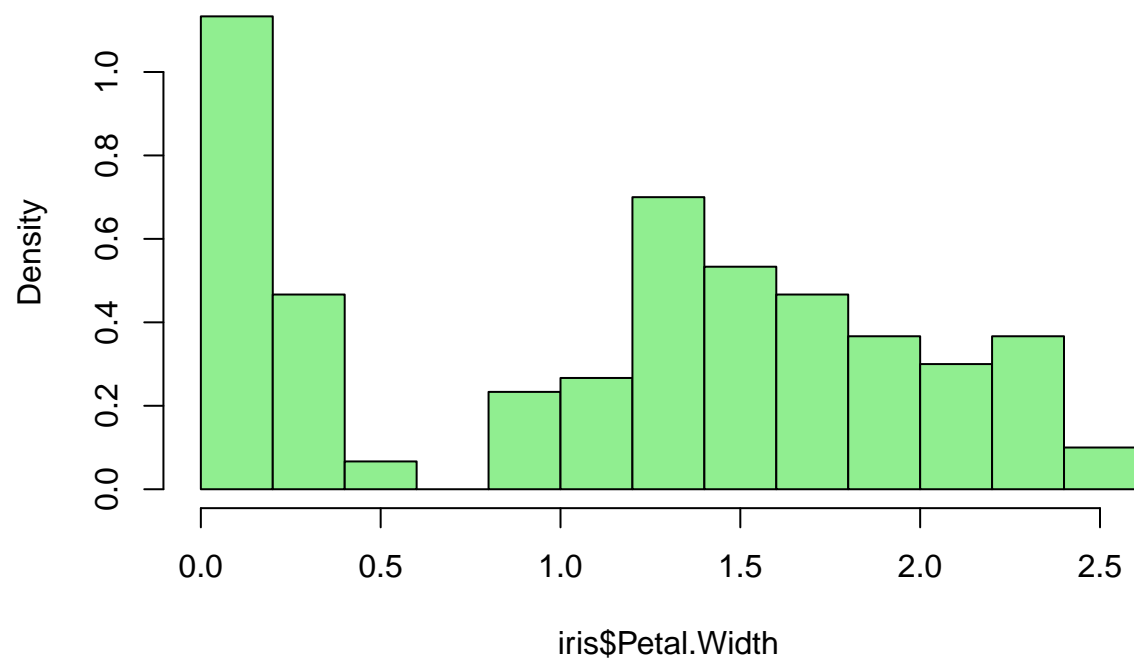
```
hist(iris$Sepal.Width, main="Histogram of Sepal Width", col="light blue", freq=FALSE)
```

Histogram of Sepal Width



```
hist(iris$Petal.Width, main="Histogram of Petal Width", col="light green", freq=FALSE)
```

Histogram of Petal Width



Boxplot

```
boxplot(iris[,0:4], main='Multiple Box plots')
```


Multiple Box plots

