Data Visualization

Siliang Gong

November 21, 2021

Description of the data

The data used in this assignment is a Breast Cancer Coimbra data set, which is publicly available at UCI machine learning repository (link). The data set includes clinical features measurement for 64 patients with breast cancer and 52 healthy controls. There are 10 predictors, all quantitative, and a binary dependent variable, indicating the presence or absence of breast cancer.

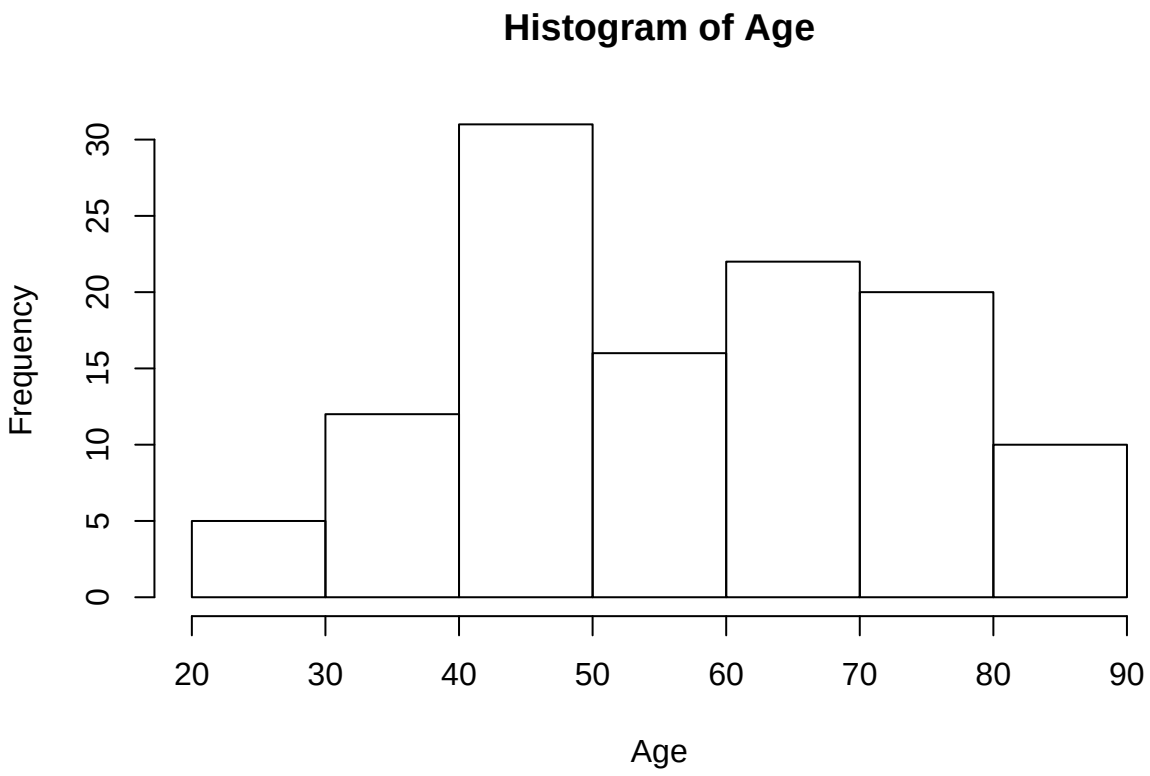
Read the data

```
# read the data into a dataframe

breast.data <- read.csv("dataR2.csv")
```

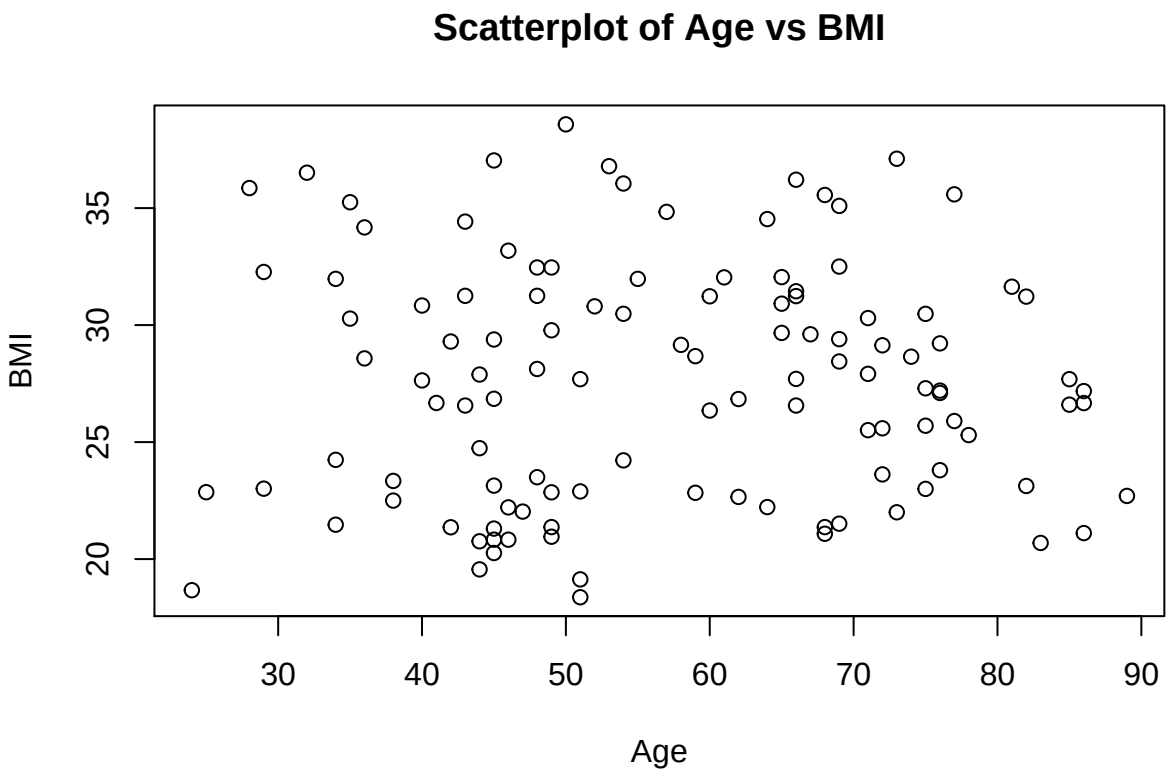
Visualization

In this section, I present the histogram of age and the scatterplot of age vs BMI.

```
# rename Classification to group and change group = 1 to "Healthy" and group =2 to "Patients"

breast.data = rename(breast.data, group = Classification)
breast.data$group[breast.data$group == 1] = "Healthy"
breast.data$group[breast.data$group == 2] = "Patients"

# Histogram of age

hist(breast.data$Age, xlab = "Age", main = "Histogram of Age")
```




```
# Scatterplot of Age vs BMI

plot(breast.data$Age, breast.data$BMI, xlab = "Age", ylab = "BMI", main = "Scatterplot of Age vs BMI")
```


Scatterplot of Age vs BMI
BMI
20
25
30
35
30
40
50
60
70
80
90
Age