Tugas Besar IF2220 - Probabilitas dan Statistika

Part 2: Data Visualization

Anggota:
13521116 - Juan Christopher Santoso
13521162 - Antonio Natthan Krishna

Data Preparation and Data Description

In [1]:
```
# Import Dataset
df <- read.csv("..\\test\\anggur.csv")

# Data Statistics
properties <- c("Rows", "Columns")
value <- c(nrow(df), ncol(df))
cbind(properties, value)

# List of Columns
columns_index <- c(1:ncol(df))
columns_name <- colnames(df)

# Display List
cbind(columns_index, columns_name)
```

A matrix: 2 × 2 of type chr

properties	value
Rows	1000
Columns	12

A matrix: 12 × 2 of type chr

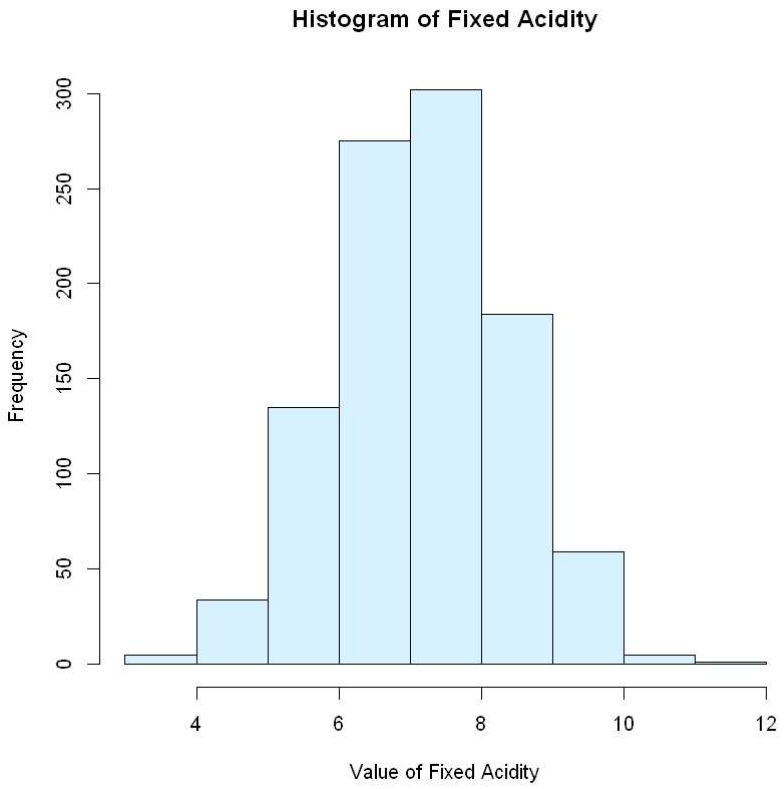
columns_index	columns_name
1	fixed.acidity
2	volatile.acidity
3	citric.acid
4	residual.sugar
5	chlorides
6	free.sulfur.dioxide
7	total.sulfur.dioxide
8	density
9	pH
10	sulphates
11	alcohol
12	quality

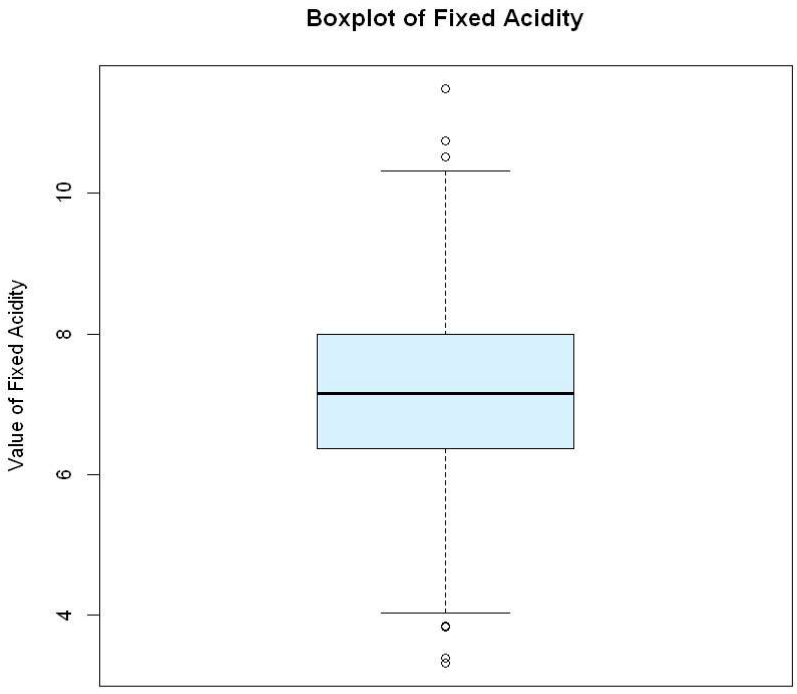
Global Functions used for Data Visualization

In [2]:
```
getHist <- function(v, name, color) {
    hist(v,
    main = paste("Histogram of", name),
    xlab = paste("Value of", name),
    ylab = "Frequency",
    col = color)
}

getBoxPlot <- function(v, name, color){
    boxplot(v,
    main = paste("Boxplot of", name),
    ylab = paste("Value of", name),
    col = "#D8F2FF"
    )
}
```

1. Kolom *fixed.acidity*

In [3]:
```
fixed_acidity <- df$fixed.acidity

getHist(fixed_acidity,"Fixed Acidity", "#D8F2FF")
getBoxPlot(fixed_acidity, "Fixed Acidity", "#D8F2FF")
```



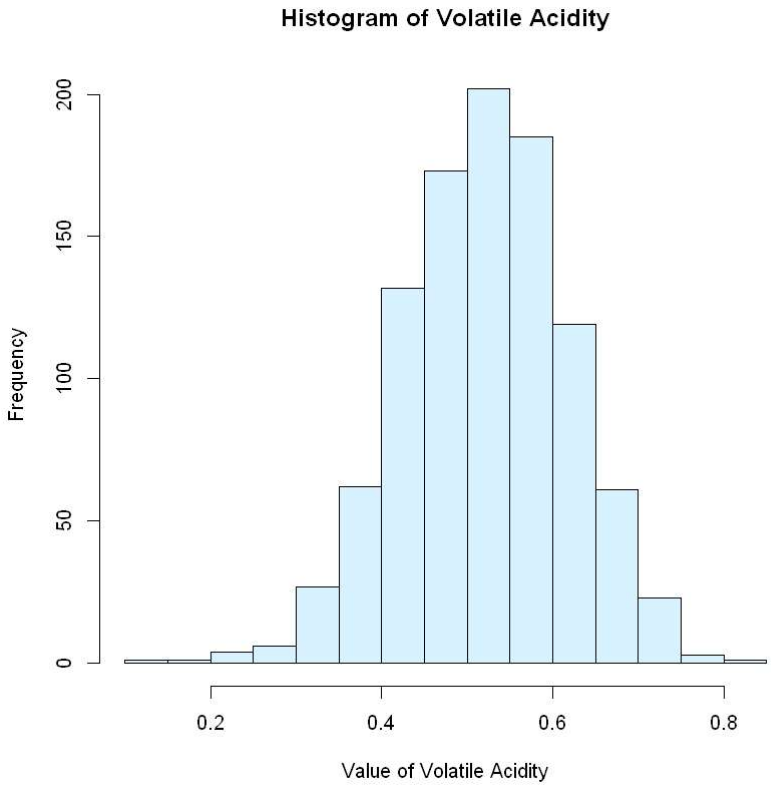


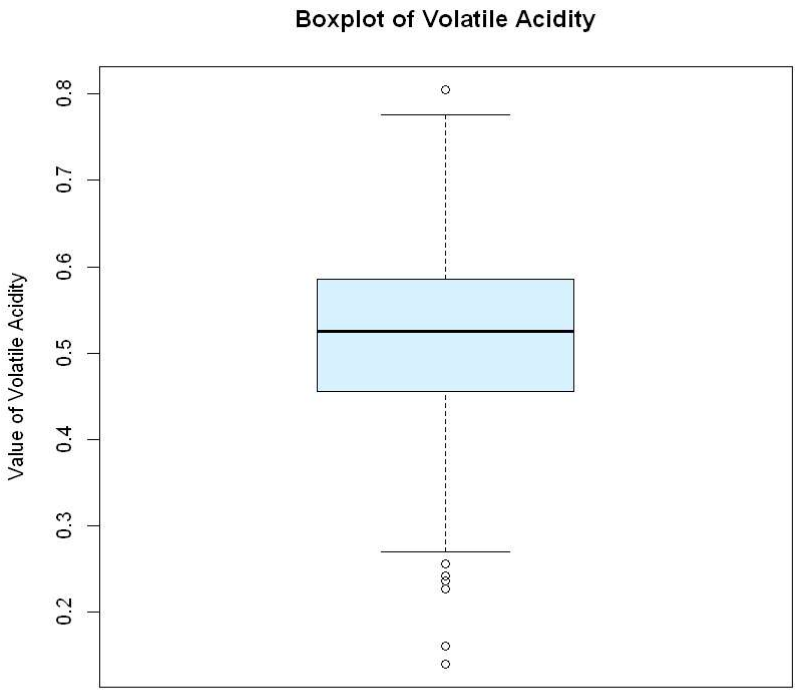
Deskripsi

Kolom Fixed Acidity secara kualitatif hampir terdistribusi normal, terdapat beberapa outlier (nilai tidak wajar) berdasarkan boxplot yang terlalu besar dan terlalu kecil

2. Kolom *volatile.acidity*

In [4]:
```
volatile_acidity <- df$volatile.acidity

getHist(volatile_acidity,"Volatile Acidity", "#D8F2FF")
getBoxPlot(volatile_acidity, "Volatile Acidity", "#D8F2FF")
```



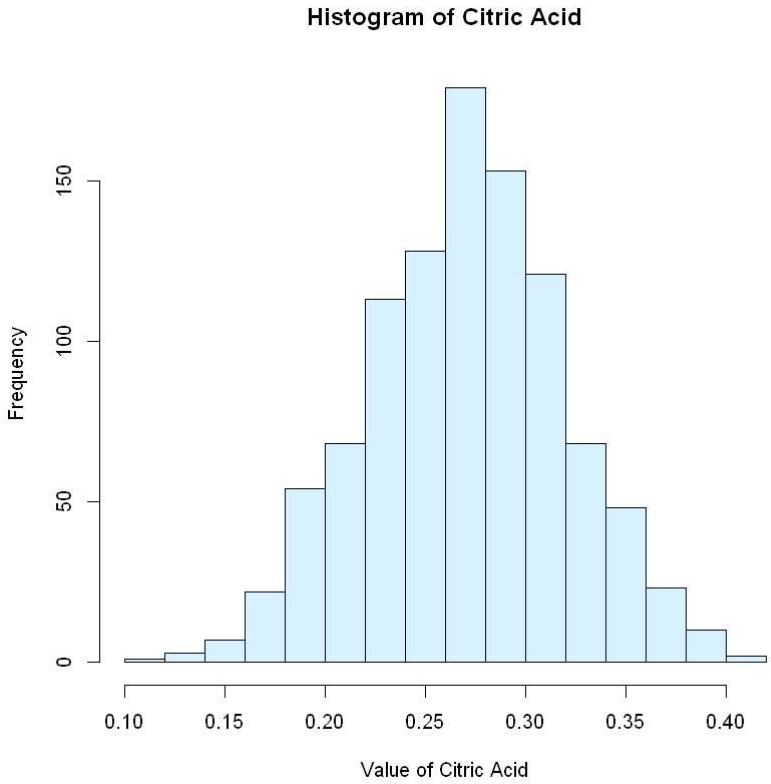


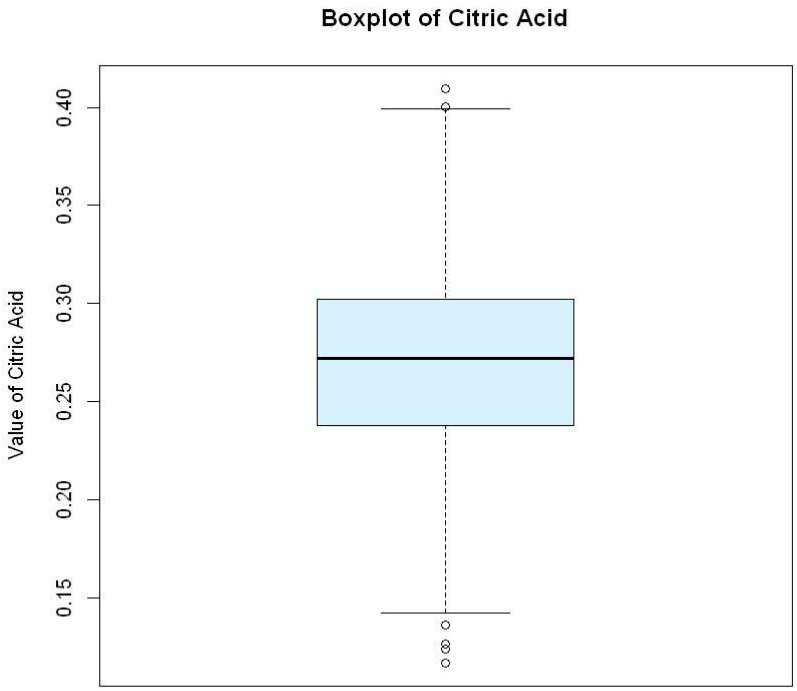
Deskripsi

Kolom Volatile Acidity secara kualitatif hampir terdistribusi normal. Terdapat satu nilai outlier atas dan cukup banyak data yang berada pada outlier yang bawah.

3. Kolom *citric.acid*

In [5]:
```
citric_acid <- df$citric.acid

getHist(citric_acid,"Citric Acid", "#D8F2FF")
getBoxPlot(citric_acid, "Citric Acid", "#D8F2FF")
```



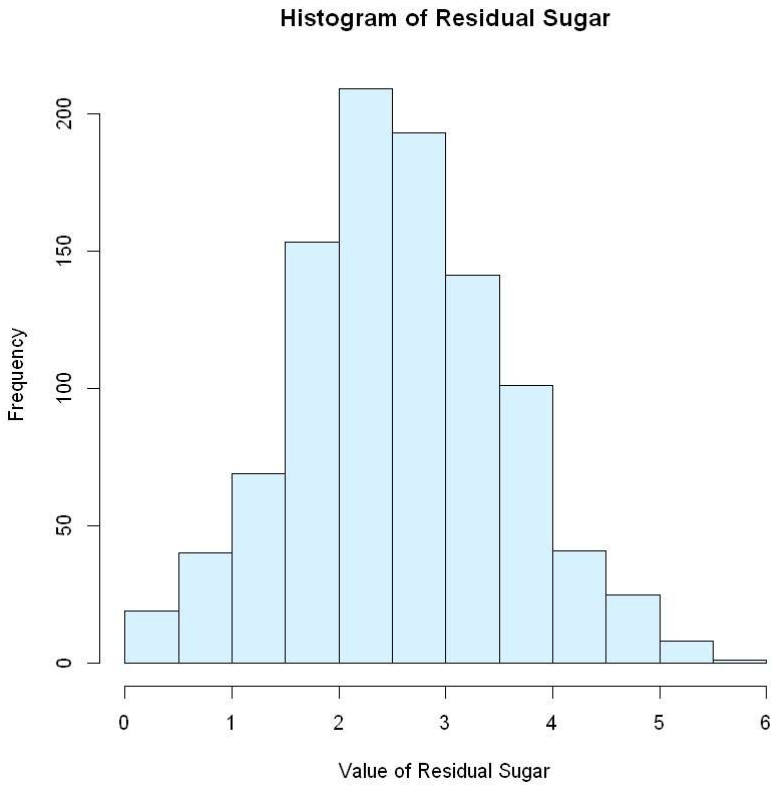


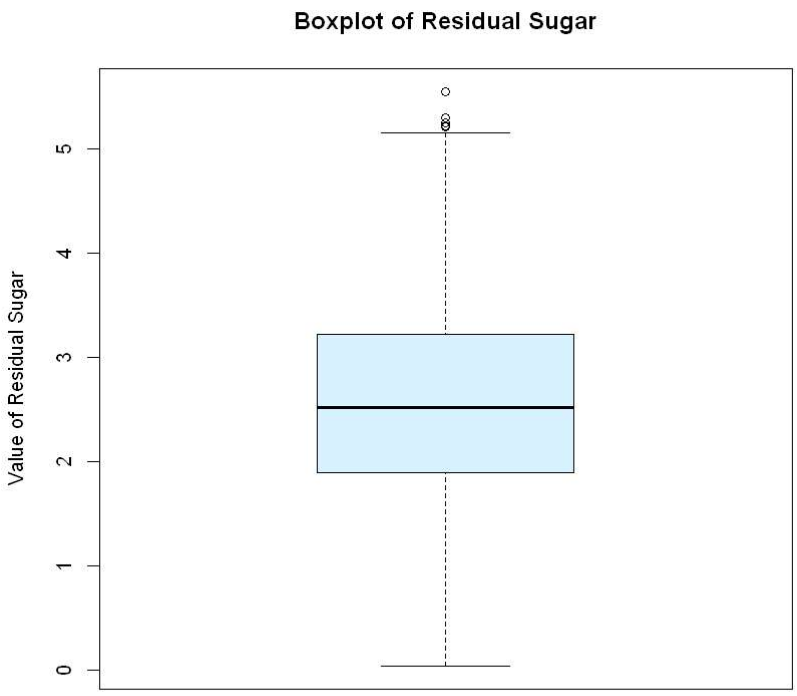
Deskripsi

Kolom Citric Acid secara kualitatif hampir terdistribusi normal, terdapat beberapa outlier (nilai tidak wajar), terdapat 2 buah data yang berada pada outlier atas dan empat data yang terdapat pada outlier bawah

4. Kolom *residual.sugar*

In [6]:
```
residual_sugar <- df$residual.sugar

getHist(residual_sugar,"Residual Sugar", "#D8F2FF")
getBoxPlot(residual_sugar, "Residual Sugar", "#D8F2FF")
```





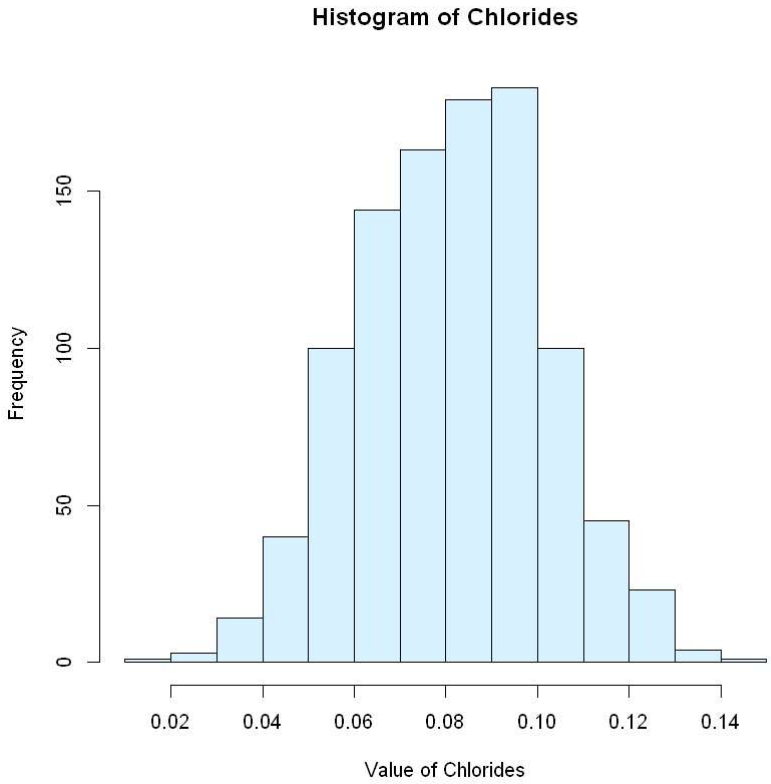
Deskripsi

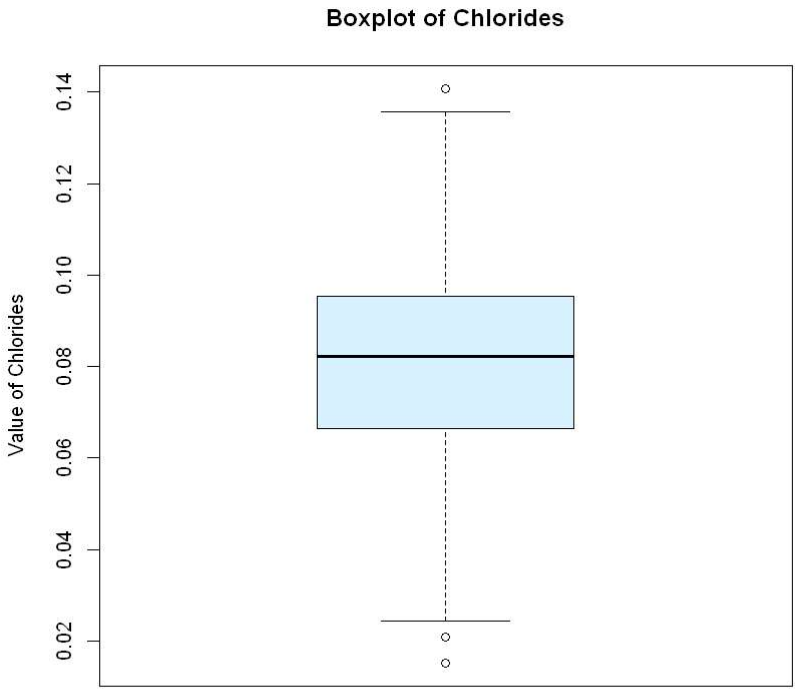
Deskripsi

Kolom Residual Sugar secara kualitatif hampir terdistribusi normal, terdapat cukup banyak data yang merupakan outlier atas berdasarkan boxplot. Tidak ada data yang menjadi outlier bawah

5. Kolom *chlorides*

In [7]:
```
chlorides <- df$chlorides

getHist(chlorides,"Chlorides", "#D8F2FF")
getBoxPlot(chlorides, "Chlorides", "#D8F2FF")
```



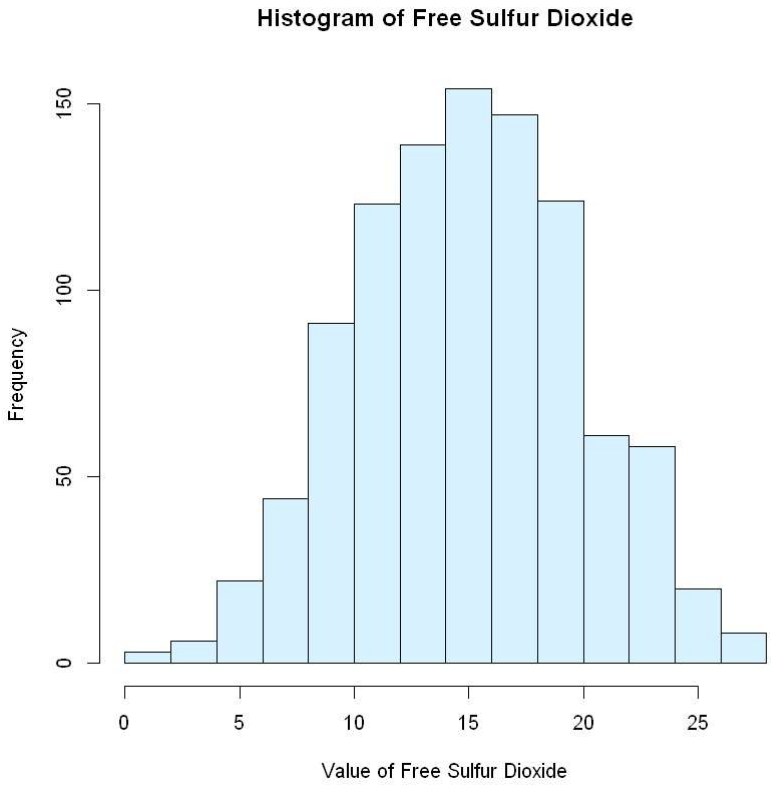


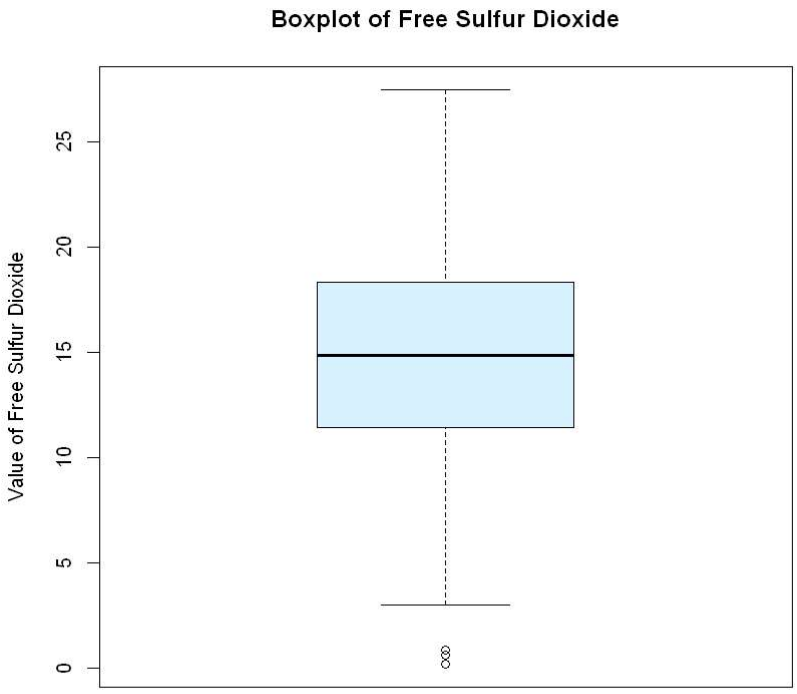
Deskripsi

Kolom Chlorides secara kualitatif hampir terdistribusi normal, terdapat 1 outlier atas dan 2 outlier bawah.

6. Kolom *free.sulfur.dioxide*

In [8]:
```
free_sulfur_dioxide <- df$free.sulfur.dioxide

getHist(free_sulfur_dioxide,"Free Sulfur Dioxide", "#D8F2FF")
getBoxPlot(free_sulfur_dioxide, "Free Sulfur Dioxide", "#D8F2FF")
```



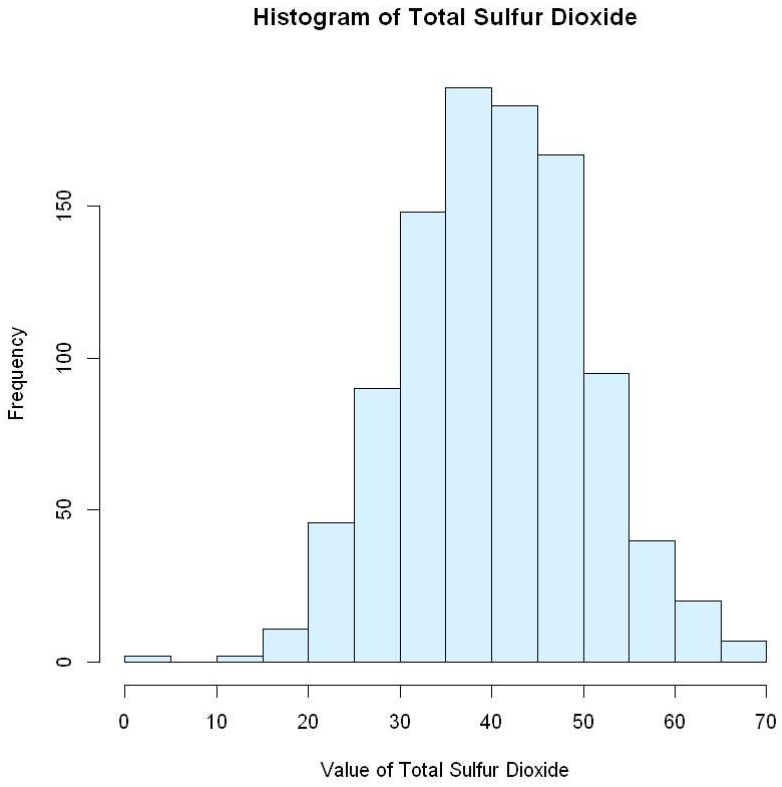


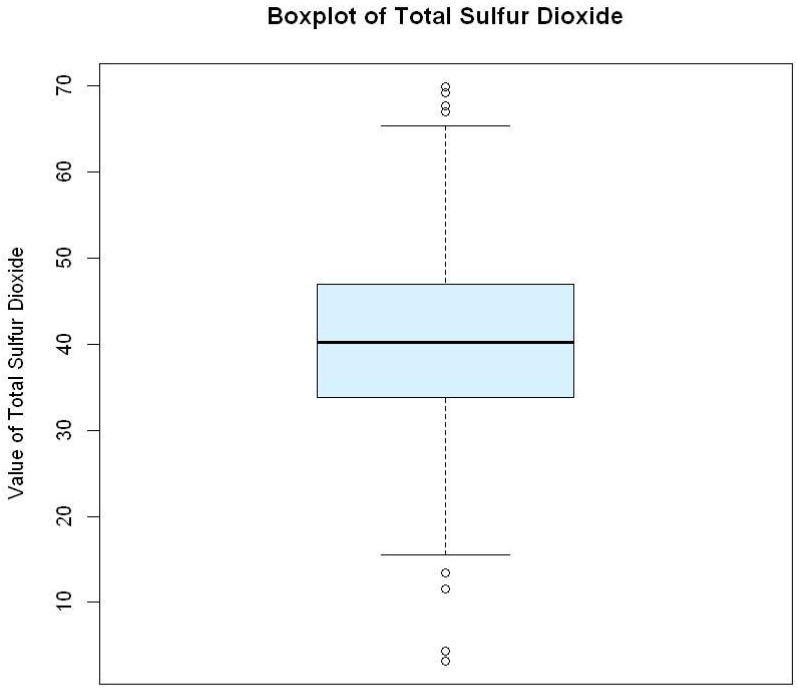
Deskripsi

Kolom Free Sulfur Dioxide secara kualitatif hampir terdistribusi normal, tidak terdapat outlier atas namun terdapat beberapa data yang merupakan outlier bawah.

7. Kolom *total.sulfur.dioxide*

In [9]:
```
total_sulfur_dioxide <- df$total.sulfur.dioxide

getHist(total_sulfur_dioxide,"Total Sulfur Dioxide", "#D8F2FF")
getBoxPlot(total_sulfur_dioxide, "Total Sulfur Dioxide", "#D8F2FF")
```



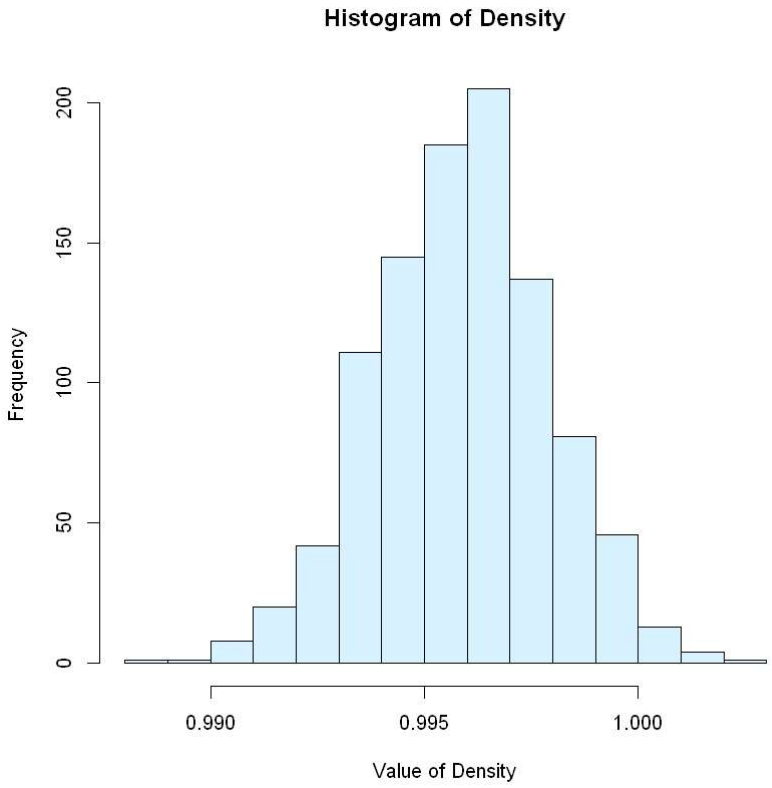


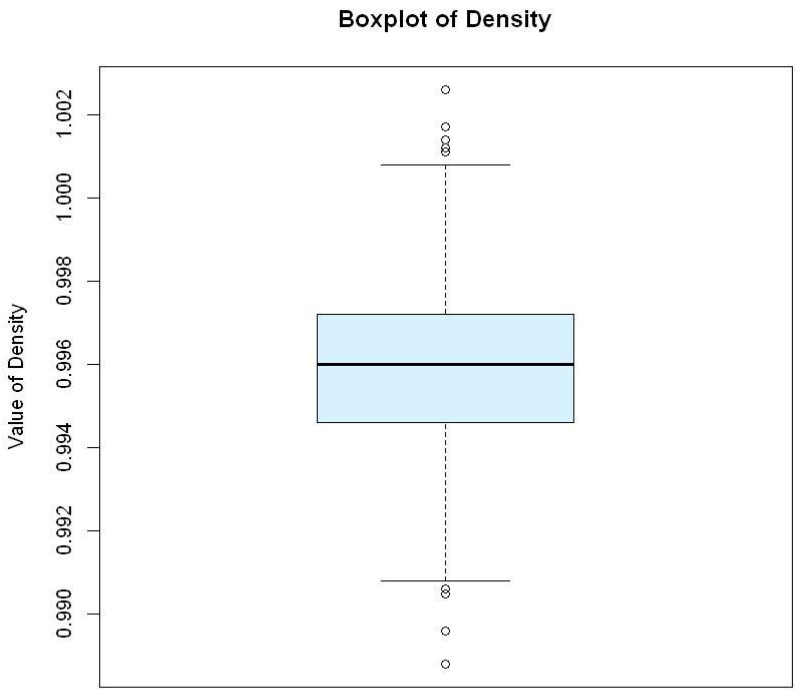
Deskripsi

Kolom Total Sulfur Dioxide secara kualitatif hampir terdistribusi normal, terdapat beberapa outlier atas dan bawah yang jumlahnya relatif lebih banyak dibandingkan kolom lain

8. Kolom *density*

In [10]:
```
density <- df$density

getHist(density,"Density", "#D8F2FF")
getBoxPlot(density, "Density", "#D8F2FF")
```



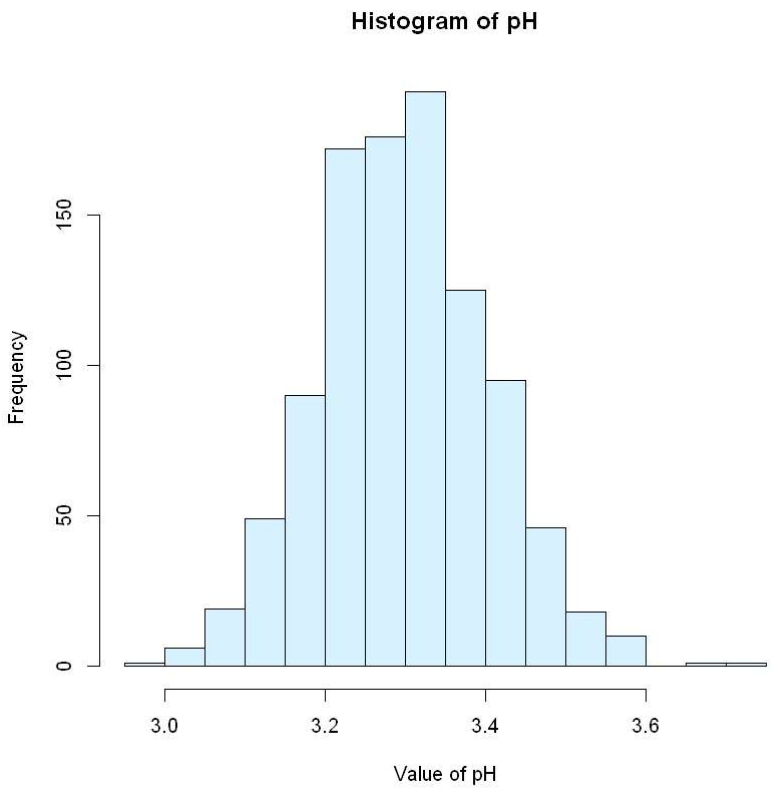


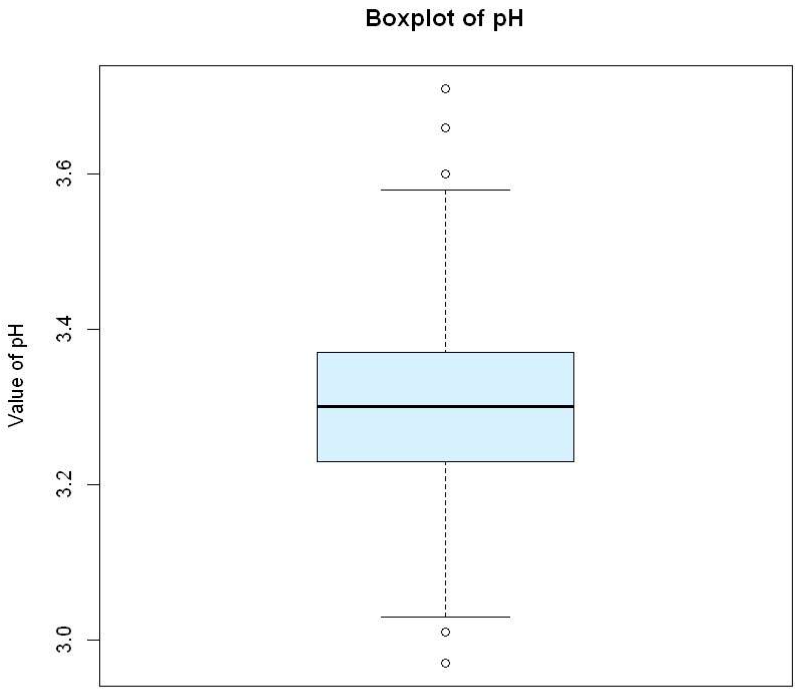
Deskripsi

Kolom Density secara kualitatif hampir terdistribusi normal, terdapat beberapa outlier atas dan bawah yang jumlahnya relatif lebih banyak dibandingkan kolom lain

9. Kolom *pH*

In [11]:
```
pH <- df$pH

getHist(pH,"pH", "#D8F2FF")
getBoxPlot(pH, "pH", "#D8F2FF")
```



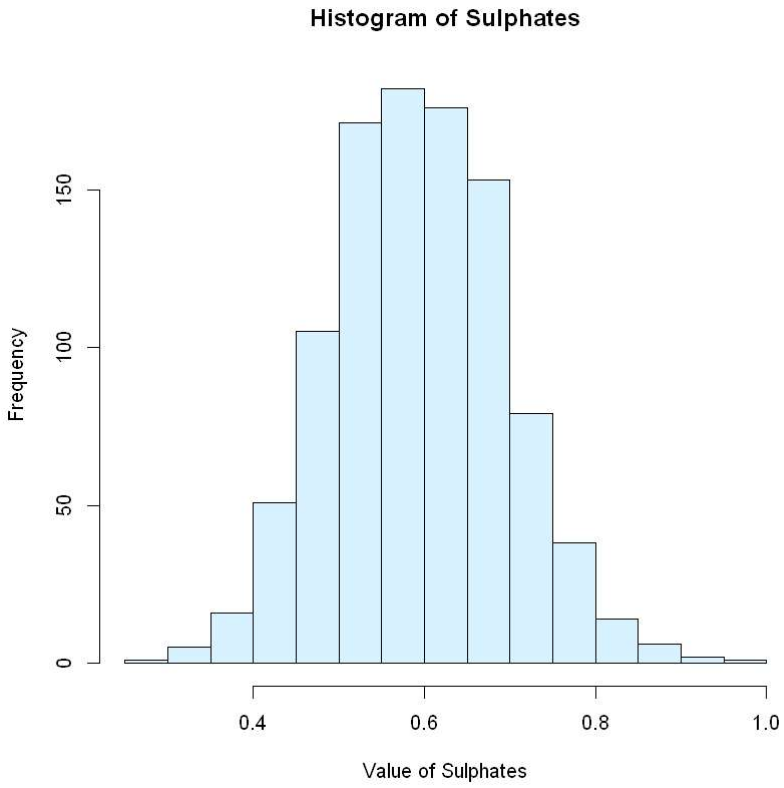


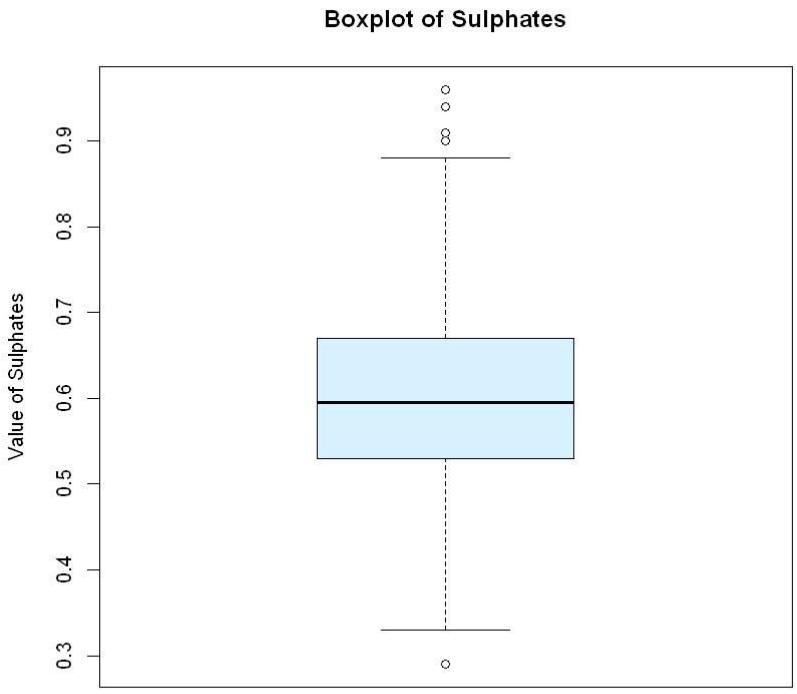
Deskripsi

Kolom Total Sulfur Dioxide secara kualitatif hampir terdistribusi normal, outlier atas dan bawah yang masing masing cukup memiliki jarak yang jauh dari range "kenormalan" data dan satu sama lain

10. Kolom *sulphates*

In [12]:
```
sulphates <- df$sulphates

getHist(sulphates,"Sulphates", "#D8F2FF")
getBoxPlot(sulphates, "Sulphates", "#D8F2FF")
```



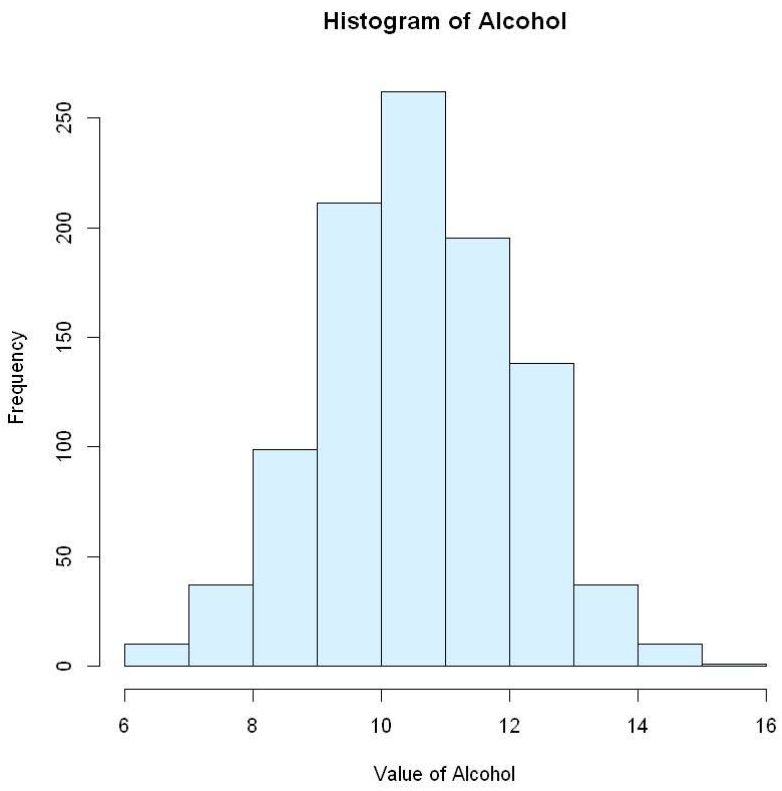


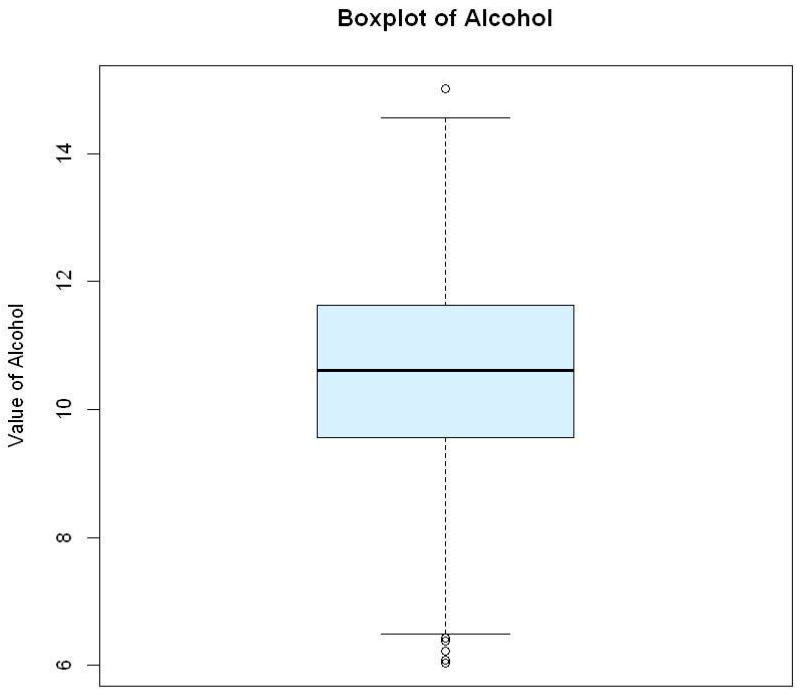
Deskripsi

Kolom Total Sulphates secara kualitatif hampir terdistribusi normal, terdapat beberapa outlier atas dan 1 outlier bawah

11. Kolom *alcohol*

In [13]:
```
alcohol <- df$alcohol

getHist(alcohol,"Alcohol", "#D8F2FF")
getBoxPlot(alcohol, "Alcohol", "#D8F2FF")
```



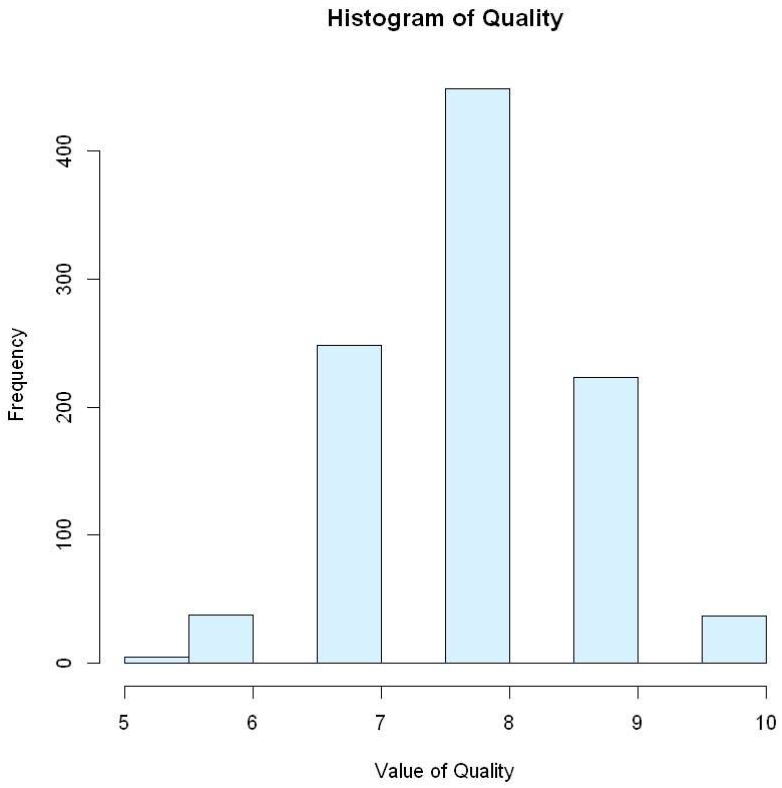


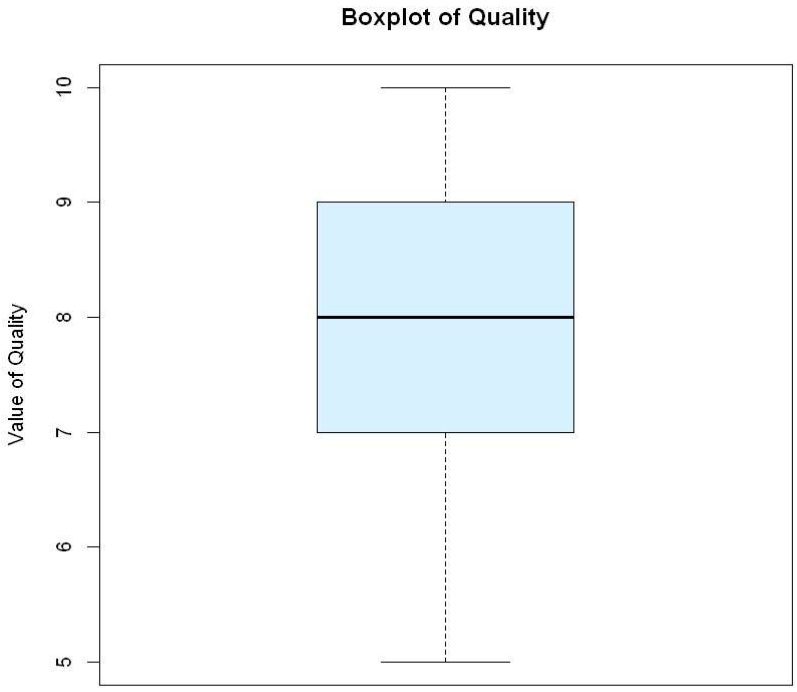
Deskripsi

Kolom Alcohol secara kualitatif hampir terdistribusi normal, terdapat cukup banyak outlier bawah, hanya terdapat 1 buah outlier atas

12. Kolom *quality*

In [14]:
```
quality <- df$quality

getHist(quality,"Quality", "#D8F2FF")
getBoxPlot(quality, "Quality", "#D8F2FF")
```





Deskripsi

Kolom Total Quality secara kualitatif hampir terdistribusi normal, tidak terdapat outlier yang menandakan bahwa data kolom quality merupakan instance yang bagus