R_K_Means_Clustering

Ulises Jose Bustamante Mora

2023-10-27

Importing the libraries

```
library(ggplot2)
```

Importing the Data

```
df1 = read.csv("winequality-red.csv", sep =";")
df2 = read.csv("winequality-white.csv", sep =";")
```

We are going to combine the data set and create an extra column to differentiate it.

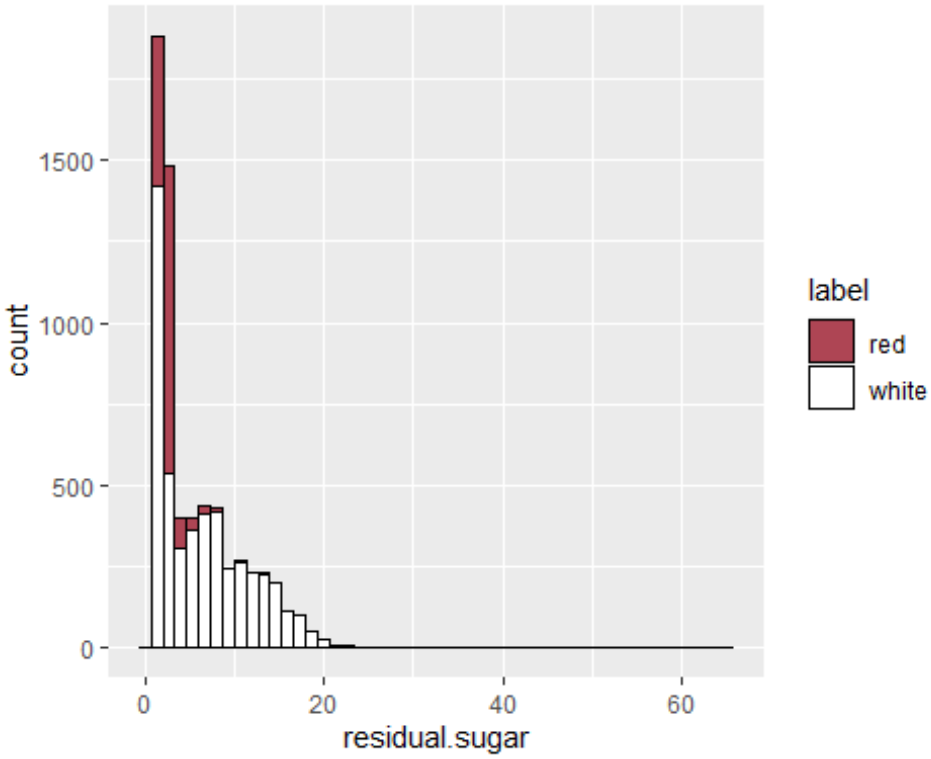
```
df1$label = sapply(df1$pH, function(x){"red"})
df2$label = sapply(df2$pH, function(x){"white"})

wine = rbind(df1, df2)
colnames(wine)
```

```
##  [1] "fixed.acidity"        "volatile.acidity"     "citric.acid"
##  [4] "residual.sugar"       "chlorides"
"free.sulfur.dioxide"
##  [7] "total.sulfur.dioxide" "density"              "pH"
## [10] "sulphates"            "alcohol"              "quality"
## [13] "label"
```

EDA

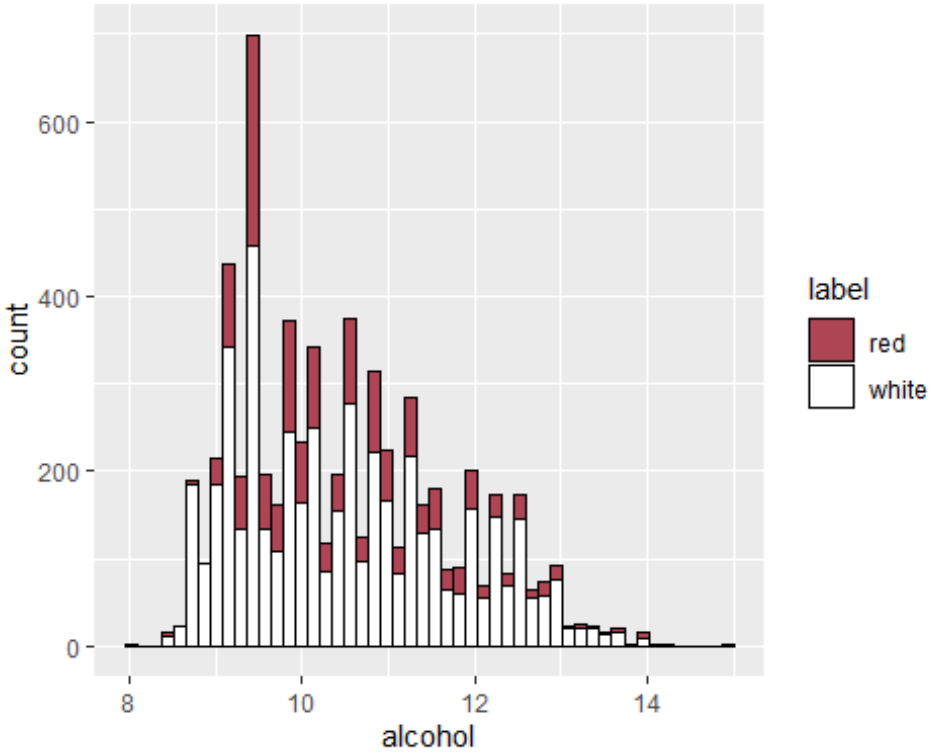
Distribution of sugar residuals

```
ggplot(wine, aes(residual.sugar)) + geom_histogram(aes(fill = label),
color = "black", bins = 50) + scale_fill_manual(values=c("#ae4554",
"white"))
```



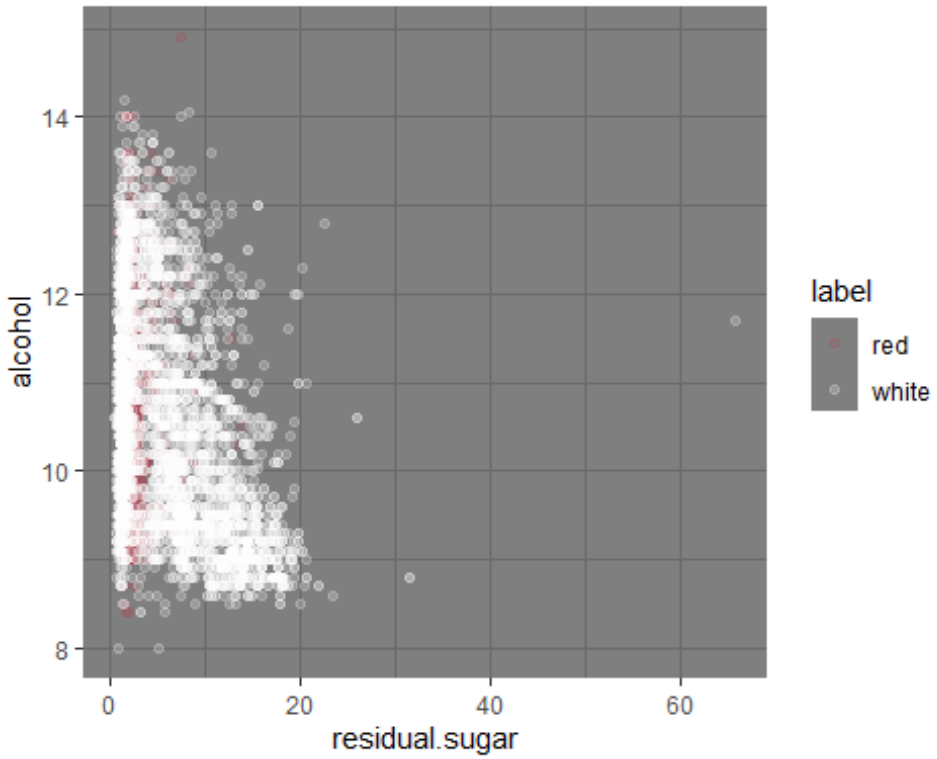
Distribution of alcohol

```
ggplot(wine, aes(alcohol)) + geom_histogram(aes(fill = label), color =
"black", bins = 50) + scale_fill_manual(values=c("#ae4554", "white"))
```



Scatter plot against alcohol and suggar

```
ggplot(wine, aes(residual.sugar, alcohol)) +
  geom_point(aes(color = label), alpha = 0.2) +
  scale_color_manual(values=c("#ae4554", "white")) + theme_dark()
```



Creating the model

```
wineClus = kmeans(wine[,1:12], 2)
print(wineClus$centers)
```

```
##   fixed.acidity volatile.acidity citric.acid residual.sugar  chlorides
## 1      6.904812        0.2871659   0.3397642       7.244809 0.04859257
## 2      7.623219        0.4086378   0.2908725       3.076425 0.06580983
##   free.sulfur.dioxide total.sulfur.dioxide   density       pH
sulphates
## 1            39.75590            155.69246 0.9947903 3.190808
0.4999485
## 2            18.39868             63.26318 0.9945736 3.254882
0.5724145
##    alcohol  quality
## 1 10.25932 5.824343
## 2 10.79722 5.810541
```

Confusion Matrix

```
table(wine$label, wineClus$cluster)

##
##             1    2
##   red      85 1514
##   white 3604 1294
```