```r
rm(list = ls())

library(ISLR)
library(class)
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```r
library(leaps)
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```r
library(car)
```

```
## Loading required package: carData
```

```r
require(e1071)
```

```
## Loading required package: e1071
```

```r
library(bootstrap)
library(rpart)
library(gbm)
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:caret':
##
##     cluster
```

```
## Loading required package: splines
```

```
## Loading required package: parallel
```

```
## Loaded gbm 2.1.3
```

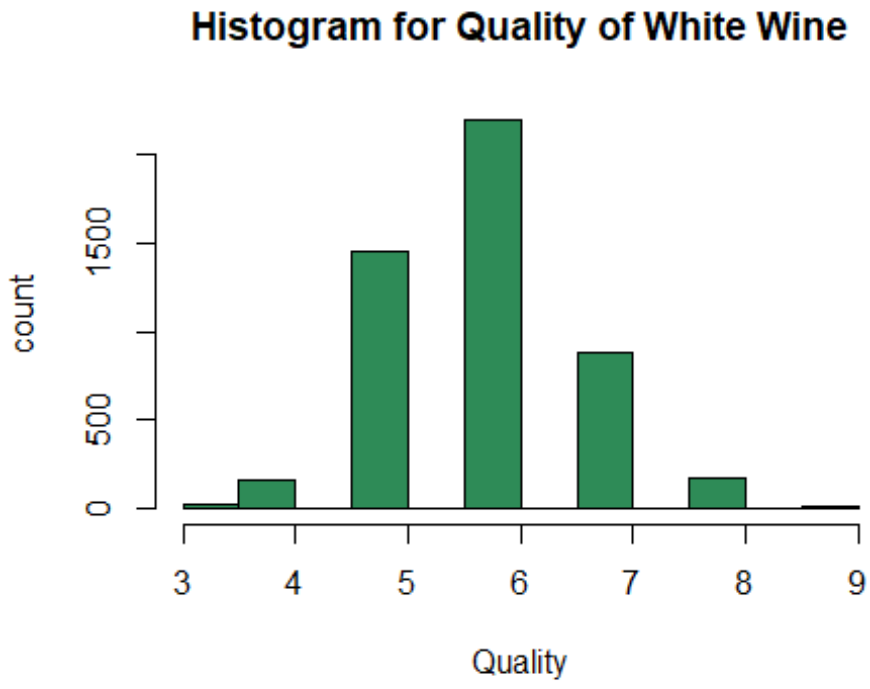
```r
white_wine = read.csv2("https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv")

summary(white_wine)
```

```
##  fixed.acidity  volatile.acidity  citric.acid   residual.sugar
##  6.8    : 308   0.28   : 263     0.3    : 307   1.2    : 187
##  6.6    : 290   0.24   : 253     0.28   : 282   1.4    : 184
##  6.4    : 280   0.26   : 240     0.32   : 257   1.6    : 165
##  6.9    : 241   0.25   : 231     0.34   : 225   1.3    : 147
##  6.7    : 236   0.22   : 229     0.29   : 223   1.1    : 146
##  7      : 232   0.27   : 218     0.26   : 219   1.5    : 142
##  (Other):3311   (Other):3464     (Other):3385   (Other):3927
##    chlorides    free.sulfur.dioxide total.sulfur.dioxide    density
##  0.044  : 201   29     : 160        111    :  69         0.992  :  64
##  0.036  : 200   31     : 132        113    :  61         0.9928 :  61
##  0.042  : 184   26     : 129        117    :  57         0.9932 :  53
```

```
##  0.04   : 182   35     : 129        118    :  55         0.993  :  52
##  0.046  : 181   34     : 128        114    :  54         0.9934 :  50
##  0.048  : 174   36     : 127        122    :  54         0.9938 :  49
##  (Other):3776   (Other):4093        (Other):4548         (Other):4569
##        pH         sulphates       alcohol        quality
##  3.14   : 172   0.5    : 249   9.4    : 229   Min.   :3.000
##  3.16   : 164   0.46   : 225   9.5    : 228   1st Qu.:5.000
##  3.22   : 146   0.44   : 216   9.2    : 199   Median :6.000
##  3.19   : 145   0.38   : 214   9      : 185   Mean   :5.878
##  3.18   : 138   0.42   : 181   10     : 162   3rd Qu.:6.000
##  3.2    : 137   0.48   : 179   10.5   : 160   Max.   :9.000
##  (Other):3996   (Other):3634   (Other):3735
```

The white wine dataset contain 12 variables and 4898 observations. The 12 variables are: fixed.acidity, volatile.acidity, citric.acid, residual.sugar, chlorides, free.sulfur.dioxide, total.sulfur.dioxide, density, pH, sulphates, alcohol, quality.

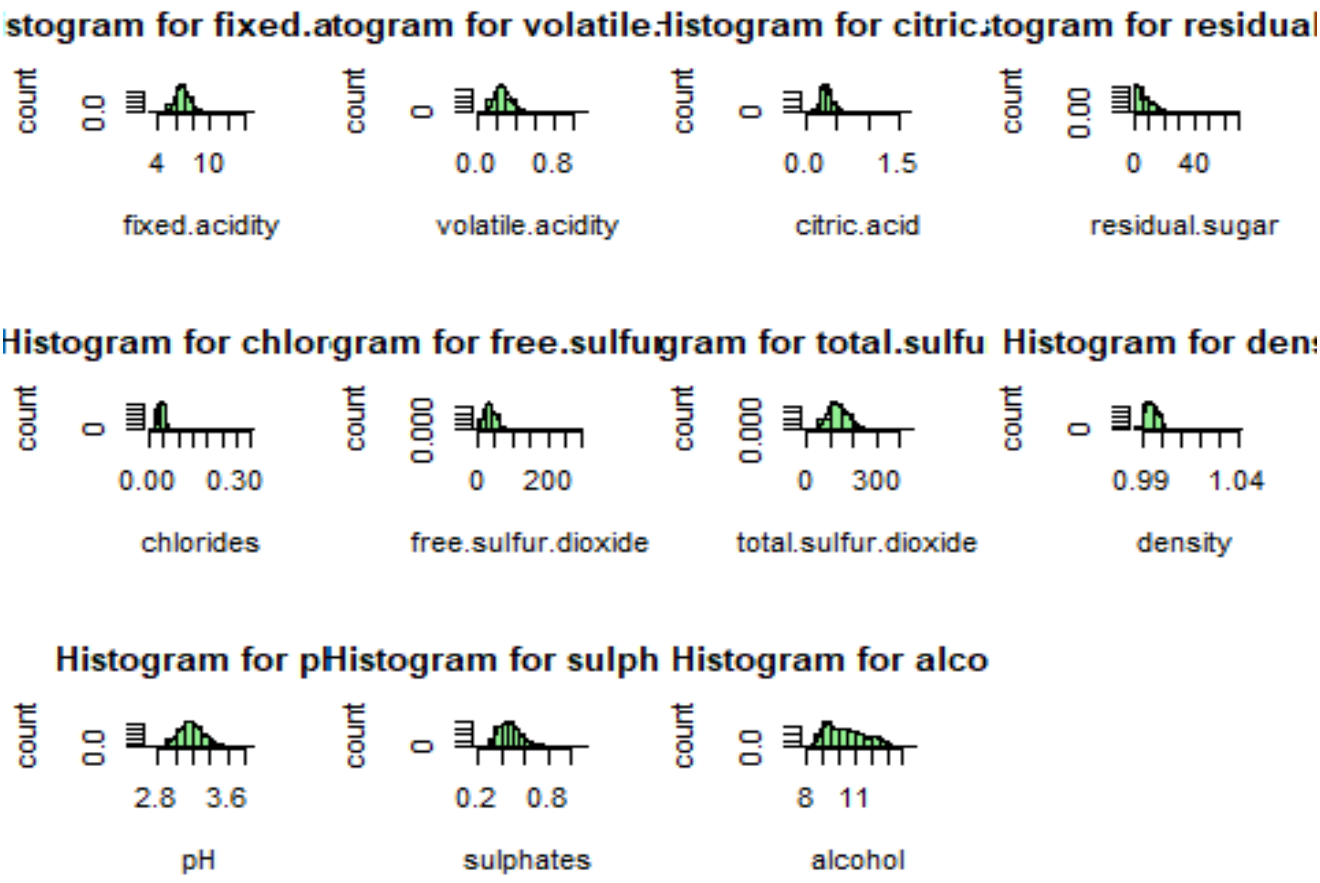
```
white_wine = data.frame(lapply(white_wine, function(x) as.numeric(as.character(x))))

hist(white_wine$quality, main = "Histogram for Quality of White Wine",
     xlab ="Quality", ylab = "count", col ="seagreen")
```



Quality ranges from 3 to 9 for white wine. It has most values concentrated in the categories 5, 6 and 7. Only a small proportion is in the categories 3,4,8 and 9

```
################################################################################
# Histogram for all predictors
################################################################################

par(mfrow = c(3,4))
hist(white_wine$fixed.acidity, main = "Histogram for fixed.acidity", prob = TRUE, xlab =
"fixed.acidity", ylab = "count", col = "lightgreen")
lines(density(white_wine$fixed.acidity), lwd = 1.5, col = "black")
```

```r
hist(white_wine$volatile.acidity, main = "Histogram for volatile.acidity", prob = TRUE,
xlab = "volatile.acidity", ylab = "count", col = "lightgreen")
lines(density(white_wine$volatile.acidity), lwd = 1.5, col = "black")

hist(white_wine$citric.acid, main = "Histogram for citric.acid", prob = TRUE,   xlab =
"citric.acid", ylab = "count", col = "lightgreen")
lines(density(white_wine$citric.acid), lwd = 1.5, col = "black")

hist(white_wine$residual.sugar, main = "Histogram for residual.sugar", prob = TRUE, xlab
= "residual.sugar", ylab = "count", col = "lightgreen")
lines(density(white_wine$residual.sugar), lwd = 1.5, col = "black")

hist(white_wine$chlorides, main = "Histogram for chlorides", prob = TRUE,       xlab =
"chlorides", ylab = "count", col = "lightgreen")
lines(density(white_wine$chlorides), lwd = 1.5, col = "black")

hist(white_wine$free.sulfur.dioxide, main = "Histogram for free.sulfur.dioxide", prob =
TRUE, xlab = "free.sulfur.dioxide", ylab = "count", col = "lightgreen")
lines(density(white_wine$free.sulfur.dioxide), lwd = 1.5, col = "black")

hist(white_wine$total.sulfur.dioxide, main = "Histogram for total.sulfur.dioxide", prob =
TRUE, xlab = "total.sulfur.dioxide", ylab = "count", col = "lightgreen")
lines(density(white_wine$total.sulfur.dioxide), lwd = 1.5, col = "black")

hist(white_wine$density, main = "Histogram for density", prob = TRUE, xlab = "density",
ylab = "count", col = "lightgreen")
lines(density(white_wine$density), lwd = 1.5, col = "black")

hist(white_wine$pH, main = "Histogram for pH", prob = TRUE, xlab = "pH", ylab = "count",
col = "lightgreen")
lines(density(white_wine$pH), lwd = 1.5, col = "black")

hist(white_wine$sulphates, main = "Histogram for sulphates", prob = TRUE, xlab =
"sulphates", ylab = "count", col = "lightgreen")
lines(density(white_wine$sulphates), lwd = 1.5, col = "black")

hist(white_wine$alcohol, main = "Histogram for alcohol", prob = TRUE, xlab = "alcohol",
ylab = "count", col = "lightgreen")
lines(density(white_wine$alcohol), lwd = 1.5, col = "black")
```


stogram for fixed.a
togram for volatile.
Histogram for citric.
togram for residual
count
0.0
4
10
fixed.acidity
0
0.0
0.8
volatile.acidity
0.0
1.5
citric.acid
0.00
0
40
residual.sugar
Histogram for chlor
gram for free.sulfu
gram for total.sulfu
Histogram for dens
0.00
0.30
chlorides
0.000
0
200
free.sulfur.dioxide
0.000
0
300
total.sulfur.dioxide
0.99
1.04
density
Histogram for pH
Histogram for sulph
Histogram for alco
2.8
3.6
pH
0.2
0.8
sulphates
8
11
alcohol