wine assignment

Trishla Thakur

2023-09-05

library(readr)  
df<- read.table(file="data/wine+quality/winequality-red.csv",sep=';',header =T)  
attach(df)

1. **What is the sample size?**

str(df)

## 'data.frame': 1599 obs. of 12 variables:  
## $ fixed.acidity : num 7.4 7.8 7.8 11.2 7.4 7.4 7.9 7.3 7.8 7.5 ...  
## $ volatile.acidity : num 0.7 0.88 0.76 0.28 0.7 0.66 0.6 0.65 0.58 0.5 ...  
## $ citric.acid : num 0 0 0.04 0.56 0 0 0.06 0 0.02 0.36 ...  
## $ residual.sugar : num 1.9 2.6 2.3 1.9 1.9 1.8 1.6 1.2 2 6.1 ...  
## $ chlorides : num 0.076 0.098 0.092 0.075 0.076 0.075 0.069 0.065 0.073 0.071 ...  
## $ free.sulfur.dioxide : num 11 25 15 17 11 13 15 15 9 17 ...  
## $ total.sulfur.dioxide: num 34 67 54 60 34 40 59 21 18 102 ...  
## $ density : num 0.998 0.997 0.997 0.998 0.998 ...  
## $ pH : num 3.51 3.2 3.26 3.16 3.51 3.51 3.3 3.39 3.36 3.35 ...  
## $ sulphates : num 0.56 0.68 0.65 0.58 0.56 0.56 0.46 0.47 0.57 0.8 ...  
## $ alcohol : num 9.4 9.8 9.8 9.8 9.4 9.4 9.4 10 9.5 10.5 ...  
## $ quality : int 5 5 5 6 5 5 5 7 7 5 ...

**b. Any outliers? Do you have any concerns about the data quality?**

summary(df)

## fixed.acidity volatile.acidity citric.acid residual.sugar   
## Min. : 4.60 Min. :0.1200 Min. :0.000 Min. : 0.900   
## 1st Qu.: 7.10 1st Qu.:0.3900 1st Qu.:0.090 1st Qu.: 1.900   
## Median : 7.90 Median :0.5200 Median :0.260 Median : 2.200   
## Mean : 8.32 Mean :0.5278 Mean :0.271 Mean : 2.539   
## 3rd Qu.: 9.20 3rd Qu.:0.6400 3rd Qu.:0.420 3rd Qu.: 2.600   
## Max. :15.90 Max. :1.5800 Max. :1.000 Max. :15.500   
## chlorides free.sulfur.dioxide total.sulfur.dioxide density   
## Min. :0.01200 Min. : 1.00 Min. : 6.00 Min. :0.9901   
## 1st Qu.:0.07000 1st Qu.: 7.00 1st Qu.: 22.00 1st Qu.:0.9956   
## Median :0.07900 Median :14.00 Median : 38.00 Median :0.9968   
## Mean :0.08747 Mean :15.87 Mean : 46.47 Mean :0.9967   
## 3rd Qu.:0.09000 3rd Qu.:21.00 3rd Qu.: 62.00 3rd Qu.:0.9978   
## Max. :0.61100 Max. :72.00 Max. :289.00 Max. :1.0037   
## pH sulphates alcohol quality   
## Min. :2.740 Min. :0.3300 Min. : 8.40 Min. :3.000   
## 1st Qu.:3.210 1st Qu.:0.5500 1st Qu.: 9.50 1st Qu.:5.000   
## Median :3.310 Median :0.6200 Median :10.20 Median :6.000   
## Mean :3.311 Mean :0.6581 Mean :10.42 Mean :5.636   
## 3rd Qu.:3.400 3rd Qu.:0.7300 3rd Qu.:11.10 3rd Qu.:6.000   
## Max. :4.010 Max. :2.0000 Max. :14.90 Max. :8.000

which(is.na(df))

## integer(0)

For outliers, I have compared medium and mean for initial analysis.

If medium and mean is not equal data distribution is not symmetric, so i did further analysis checking the 3rd quantile of each variable to understand if the graph for each variable is centre, left or right skew.

I also checked for any NULL values and found the data has no NULL value.

Attributes like fixed\_acidity,volatile\_acidity, citric\_acid,residual\_sugar,free\_sulfur\_dioxide, total\_sulfur\_dioxide, have unusual range of main and max value which help me to understand these variable have outliers.

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

fixed\_acidity various calculations

mean(df$fixed.acidity) # 8.319637

## [1] 8.319637

median(df$fixed.acidity) # 7.9

## [1] 7.9

var(df$fixed.acidity) # 3.031416

## [1] 3.031416

sd(df$fixed.acidity) # 1.741096

## [1] 1.741096

#Standard Deviation is 2  
  
mean(df$fixed.acidity) -2\*sd(df$fixed.acidity) #4.837445 (Lower end)

## [1] 4.837445

mean(df$fixed.acidity) +2\*sd(df$fixed.acidity) # 11.80183 (Upper end)

## [1] 11.80183

#fixed.acidity>4.837445 &fixed.acidity<11.80183  
  
sum((df$fixed.acidity>4.837445 & df$fixed.acidity<11.80183) == TRUE)/ 1599 # 0.9499687

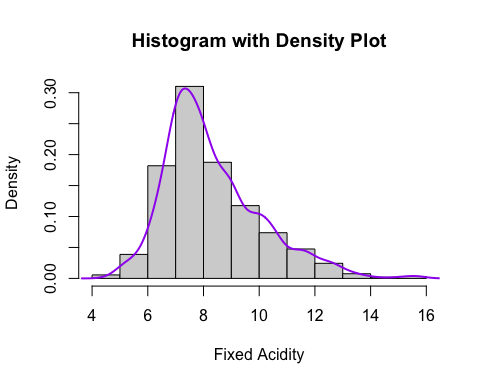
## [1] 0.9499687

#94% of the value lies in the 2 SD of the mean value, which show good data coverage   
  
quantile(df$fixed.acidity, p=c(0.25,0.5,0.75))

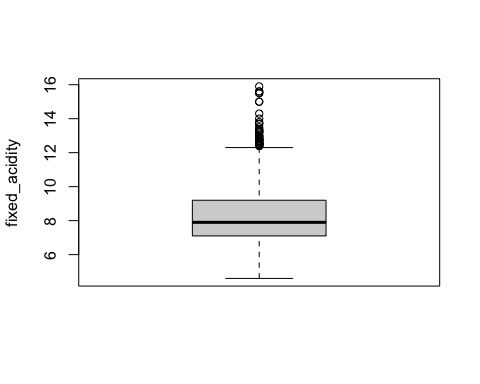
## 25% 50% 75%   
## 7.1 7.9 9.2

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
  
  
hist(df$fixed.acidity, freq = FALSE,   
 xlab = "Fixed Acidity", ylab = "Density", main = "Histogram with Density Plot")  
lines(density(df$fixed.acidity), lwd = 2, col = 'purple')



boxplot(df$fixed.acidity ,ylab ="fixed\_acidity")



**e. Do you see any skewed distributions?**

The right tail of the histogram of the distribution is longer and the median < mean. It is “right-skewed” or +ve skew

[Skewness](https://www.investopedia.com/terms/s/skewness.asp) measures the degree of symmetry of a distribution.

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

volatile.acidity various calculations

mean(df$volatile.acidity) # 0.5278205

## [1] 0.5278205

median(df$volatile.acidity) # 0.52

## [1] 0.52

var(df$volatile.acidity) # 0.03206238

## [1] 0.03206238

sd(df$volatile.acidity) # 0.1790597

## [1] 0.1790597

#Standard Deviation is 2  
  
mean(df$volatile.acidity) -2\*sd(df$volatile.acidity) #0.1697011 (Lower end)

## [1] 0.1697011

mean(df$volatile.acidity) +2\*sd(df$volatile.acidity) # 0.8859399 (Upper end)

## [1] 0.8859399

#df$volatile.acidity>0.1697011 & df$volatile.acidity< 0.8859399  
  
sum((df$volatile.acidity>0.1697011 & df$volatile.acidity< 0.8859399) == TRUE)/ 1599

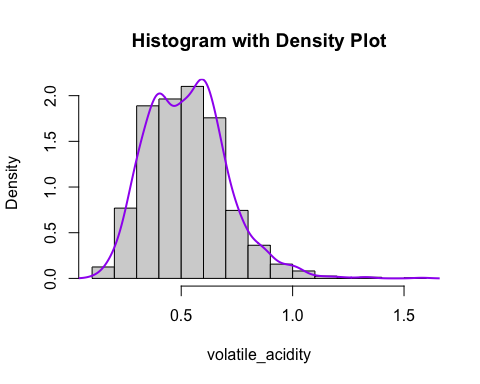
## [1] 0.9649781

# 0.9649781  
  
#the data coverage is more than 95% so it is possible to have data skewness and normal distribution is not symmetric.  
  
############# percentile ################################################  
quantile(df$volatile.acidity, p=c(0.25,0.5,0.75))

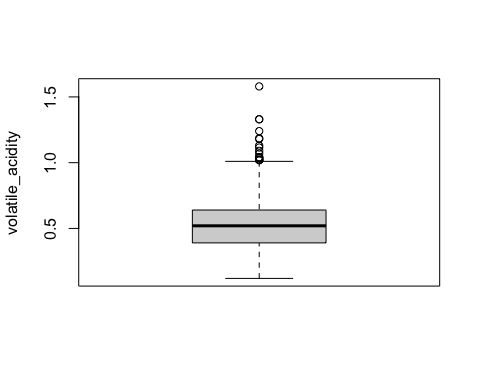
## 25% 50% 75%   
## 0.39 0.52 0.64

d. How can you visualize the distribution of each variable?

library(ggplot2)  
  
  
hist(df$volatile.acidity, freq = FALSE,   
 xlab = "volatile\_acidity", ylab = "Density", main = "Histogram with Density Plot")  
lines(density(df$volatile.acidity), lwd = 2, col = 'purple')



boxplot(df$volatile.acidity ,ylab ="volatile\_acidity")



e. Do you see any skewed distributions?

The right tail of the histogram of the distribution is longer and the median is slightly less than mean.

The box plot show few outliers showing unusual values for volatile\_acidity

It is “right-skewed” or +ve skew

[Skewness](https://www.investopedia.com/terms/s/skewness.asp) measures the degree of symmetry of a distribution.

The normal distribution has several key features and properties that define it.

First, its [mean](https://www.investopedia.com/terms/m/mean.asp) (average), [median](https://www.investopedia.com/terms/m/median.asp) (midpoint), and [mode](https://www.investopedia.com/terms/m/mode.asp) (most frequent observation) are all equal to one another. Moreover, these values all represent the peak, or highest point, of the distribution. The distribution then falls symmetrically around the mean, the width of which is defined by the [standard deviation](https://www.investopedia.com/terms/s/standarddeviation.asp).

Normal Distribution was higher than 95% so few data fall outside the 2 SD.

Also, means that data falling outside of three standard deviations (“3-sigma”) would signify rare occurrences.

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

citric.acid various calculations

mean(df$citric.acid) # 0.271

## [1] 0.2709756

median(df$citric.acid) # 0.260

## [1] 0.26

var(df$citric.acid) # 0.03794748

## [1] 0.03794748

sd(df$citric.acid) # 0.1790597

## [1] 0.1948011

#Standard Deviation is 2  
  
mean(df$citric.acid) -2\*sd(df$citric.acid) # -0.1186267 (Lower end)

## [1] -0.1186267

mean(df$citric.acid) +2\*sd(df$citric.acid) # 0.6605779 (Upper end)

## [1] 0.6605779

#df$citric.acid>-0.1186267 & df$citric.acid< 0.6605779  
  
sum((df$citric.acid>-0.1186267 & df$citric.acid< 0.6605779) == TRUE)/ 1599 # 0.9781113

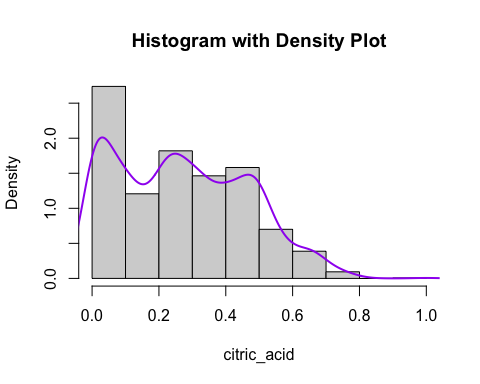
## [1] 0.9781113

############# percentile ################################################  
quantile(df$citric.acid, p=c(0.25,0.5,0.75))

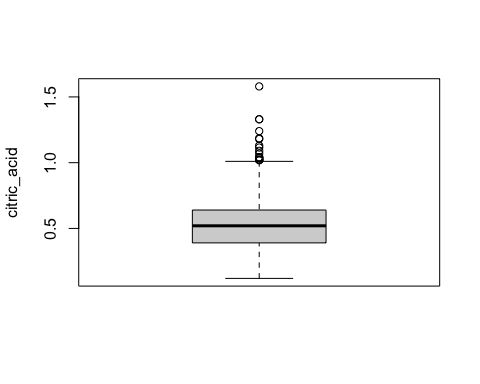
## 25% 50% 75%   
## 0.09 0.26 0.42

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
  
  
hist(df$citric.acid, freq = FALSE,   
 xlab = "citric\_acid", ylab = "Density", main = "Histogram with Density Plot")  
lines(density(df$citric.acid), lwd = 2, col = 'purple')



boxplot(df$volatile.acidity ,ylab ="citric\_acid")



**e. Do you see any skewed distributions?**

The right tail of the histogram of the distribution is longer and the median is slightly less than mean.

The box plot show few outliers showing unusual values for citric\_acid

It is “right-skewed” or +ve skew

the data coverage is more than 95% showing normal distribution is not symmetric.

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

residual.sugar various calculations

mean(df$residual.sugar) #2.538806

## [1] 2.538806

median(df$residual.sugar) #2.2

## [1] 2.2

var(df$residual.sugar) # 1.987897

## [1] 1.987897

sd(df$residual.sugar) # 1.409928

## [1] 1.409928

#Standard Deviation is 2  
  
mean(df$residual.sugar) -2\*sd(df$residual.sugar) # -0.2810506 (Lower end)

## [1] -0.2810506

mean(df$residual.sugar) +2\*sd(df$residual.sugar) # 5.358662 (Upper end)

## [1] 5.358662

#df$residual.sugar>-0.2810506 & df$residual.sugar< 5.358662  
  
sum((df$residual.sugar>-0.2810506 & df$residual.sugar< 5.358662) == TRUE)/ 1599 # 0.9530957

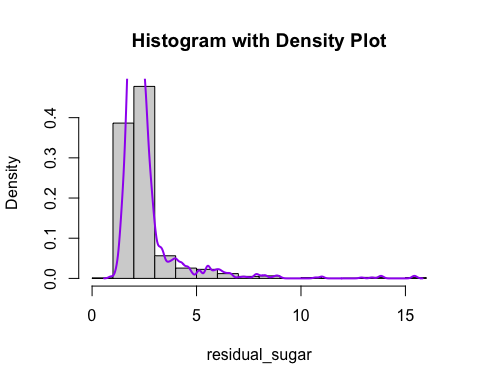
## [1] 0.9530957

############# percentile ################################################  
  
quantile(df$residual.sugar, p=c(0.25,0.5,0.75))

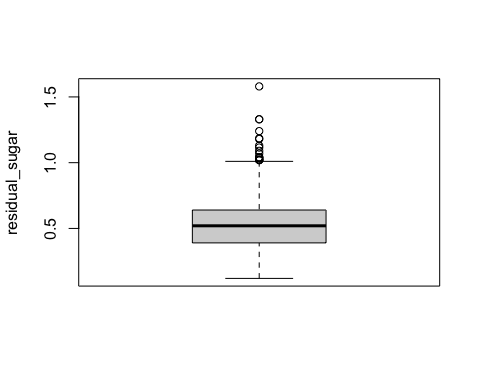
## 25% 50% 75%   
## 1.9 2.2 2.6

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
  
  
hist(df$residual.sugar, freq = FALSE,   
 xlab = "residual\_sugar", ylab = "Density", main = "Histogram with Density Plot")  
lines(density(df$residual.sugar), lwd = 2, col = 'purple')



boxplot(df$volatile.acidity ,ylab ="residual\_sugar")



**e. Do you see any skewed distributions?**

The right tail of the histogram of the distribution is longer and the median is less than mean.

The box plot show few outliers showing unusual values for residual\_sugar

It is “right-skewed” or +ve skew

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

chlorides various calculations

mean(df$chlorides) # 0.08746654

## [1] 0.08746654

median(df$chlorides) # 0.079

## [1] 0.079

var(df$chlorides) # 0.002215143

## [1] 0.002215143

sd(df$chlorides) # 0.0470653

## [1] 0.0470653

#Standard Deviation is 2  
  
mean(df$chlorides) -2\*sd(df$chlorides) # -0.006664062 (Lower end)

## [1] -0.006664062

mean(df$chlorides) +2\*sd(df$chlorides) # 0.1815971 (Upper end)

## [1] 0.1815971

#df$chlorides>-0.006664062 & df$chlorides< 0.1815971  
  
sum((df$chlorides>-0.006664062 & df$chlorides< 0.1815971) == TRUE)/ 1599 # 0.9718574

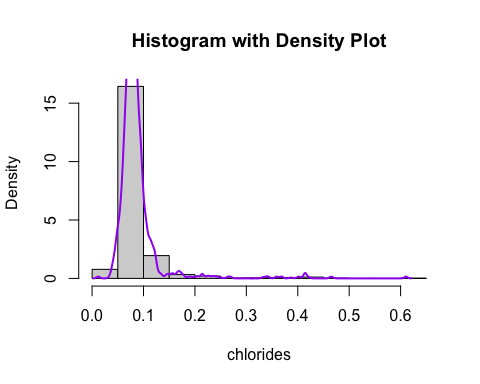
## [1] 0.9718574

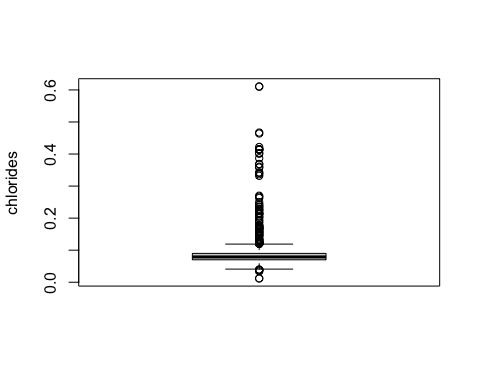
############# percentile ################################################  
  
quantile(df$chlorides, p=c(0.25,0.5,0.75))

## 25% 50% 75%   
## 0.070 0.079 0.090

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
  
  
hist(df$chlorides, freq = FALSE,   
 xlab = "chlorides", ylab = "Density", main = "Histogram with Density Plot")  
lines(density(df$chlorides), lwd = 2, col = 'purple')





**e. Do you see any skewed distributions?**

The right tail of the histogram of the distribution is longer and the median is significantly less than mean.

The box plot show very high outliers values for chlorides

It is “right-skewed” or +ve skew

the data coverage is more than 95% showing normal distribution is not symmetric.

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

free.sulfur.dioxide various calculations

mean(df$free.sulfur.dioxide) #15.87492

## [1] 15.87492

median(df$free.sulfur.dioxide) # 14

## [1] 14

var(df$free.sulfur.dioxide) # 109.4149

## [1] 109.4149

sd(df$free.sulfur.dioxide) # 10.46016

## [1] 10.46016

#Standard Deviation is 2  
  
mean(df$free.sulfur.dioxide) -2\*sd(df$free.sulfur.dioxide) #-5.045392 (Lower end)

## [1] -5.045392

mean(df$free.sulfur.dioxide) +2\*sd(df$free.sulfur.dioxide) # 36.79524 (Upper end)

## [1] 36.79524

#df$free.sulfur.dioxide>-5.045392 & df$free.sulfur.dioxide< 36.79524  
  
sum((df$free.sulfur.dioxide>-5.045392 & df$free.sulfur.dioxide< 36.79524) == TRUE)/ 1599

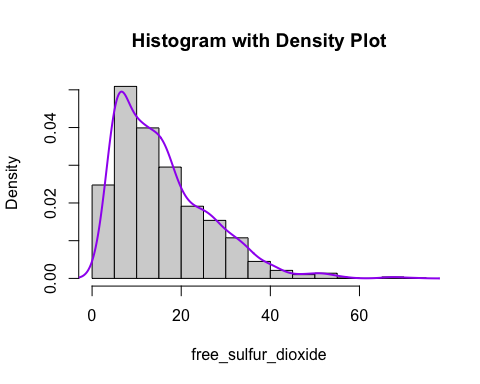
## [1] 0.9587242

#0.9587242  
  
############# percentile ################################################  
  
quantile(df$free.sulfur.dioxide, p=c(0.25,0.5,0.75))

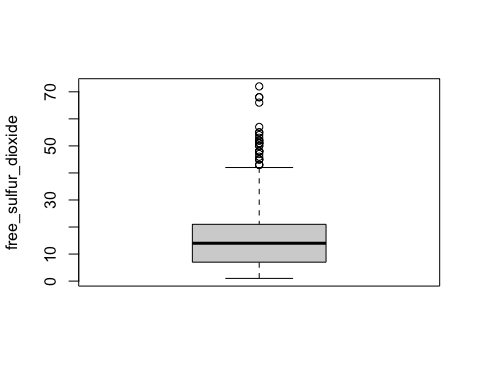
## 25% 50% 75%   
## 7 14 21

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
  
  
hist(df$free.sulfur.dioxide, freq = FALSE,   
 xlab = "free\_sulfur\_dioxide", ylab = "Density", main = "Histogram with Density Plot")  
lines(density(df$free.sulfur.dioxide), lwd = 2, col = 'purple')



boxplot(df$free.sulfur.dioxide ,ylab ="free\_sulfur\_dioxide")



**e. Do you see any skewed distributions?**

The right tail of the histogram of the distribution is longer and the median is less than mean.

The box plot show outliers values for free\_sulfur\_dioxide

It is “right-skewed” or +ve skew

----------------------------------------------------------------------------------------

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

total.sulfur.dioxide various calculations

mean(df$total.sulfur.dioxide) # 46.46779

## [1] 46.46779

median(df$total.sulfur.dioxide) # 38

## [1] 38

var(df$total.sulfur.dioxide) # 1082.102

## [1] 1082.102

sd(df$total.sulfur.dioxide) # 32.89532

## [1] 32.89532

#Standard Deviation is 2  
  
mean(df$total.sulfur.dioxide) -2\*sd(df$total.sulfur.dioxide) # -19.32286 (Lower end)

## [1] -19.32286

mean(df$total.sulfur.dioxide) +2\*sd(df$total.sulfur.dioxide) # 112.2584 (Upper end)

## [1] 112.2584

#df$total.sulfur.dioxide>-19.32286 & df$total.sulfur.dioxide< 112.2584  
  
sum((df$total.sulfur.dioxide>-19.32286 & df$total.sulfur.dioxide< 112.2584) == TRUE)/ 1599

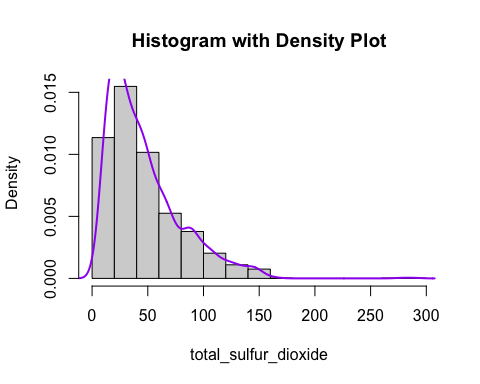
## [1] 0.9499687

#0.9499687  
  
############# percentile ################################################  
  
quantile(df$total.sulfur.dioxide, p=c(0.25,0.5,0.75))

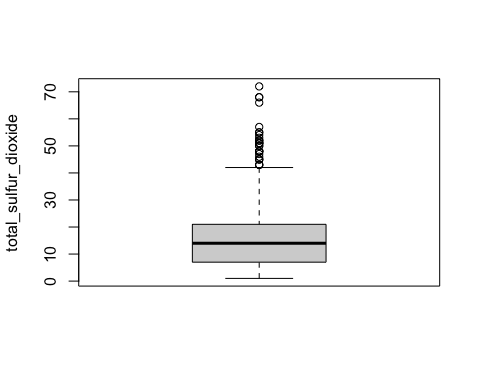
## 25% 50% 75%   
## 22 38 62

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
  
  
hist(df$total.sulfur.dioxide, freq = FALSE,   
 xlab = "total\_sulfur\_dioxide", ylab = "Density", main = "Histogram with Density Plot")  
lines(density(df$total.sulfur.dioxide), lwd = 2, col = 'purple')



boxplot(df$free.sulfur.dioxide ,ylab ="total\_sulfur\_dioxide")



**e. Do you see any skewed distributions?**

The right tail of the histogram of the distribution is longer and the median is significantly less than mean.

The box plot show outliers values for total\_sulfur\_dioxide

It is “right-skewed” or +ve skew

also the Normal distribution is less than 95% of 2 standard deviation showing asymmetric curve.

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

density various calculations

mean(df$density) # 0.9967467

## [1] 0.9967467

median(df$density) # 0.99675

## [1] 0.99675

var(df$density) # 3.562029

## [1] 3.562029e-06

sd(df$density) #0.001887334

## [1] 0.001887334

#Standard Deviation is 2  
  
mean(df$density) -2\*sd(df$density) # 0.992972 (Lower end)

## [1] 0.992972

mean(df$density) +2\*sd(df$density) # 1.000521 (Upper end)

## [1] 1.000521

#df$density>0.992972 & df$density< 1.000521  
  
sum((df$density>0.992972 & df$density<1.000521) == TRUE)/ 1599 # 0.9493433

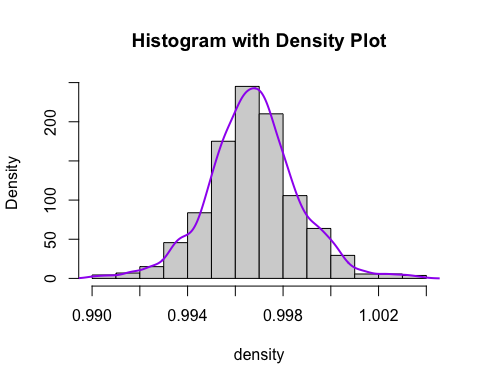
## [1] 0.9493433

############# percentile ################################################  
  
quantile(df$density, p=c(0.25,0.5,0.75))

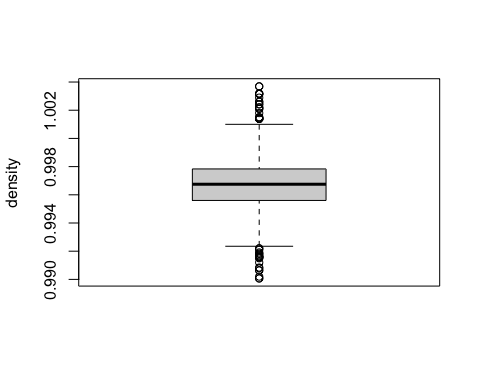
## 25% 50% 75%   
## 0.995600 0.996750 0.997835

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
  
  
hist(df$density, freq = FALSE,   
 xlab = "density", ylab = "Density", main = "Histogram with Density Plot")  
lines(density(df$density), lwd = 2, col = 'purple')



boxplot(df$density ,ylab ="density")



**e. Do you see any skewed distributions?**

The curve is following symmetry and there is no skewness also the median = means

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

pH various calculations

mean(df$pH) # 3.311113

## [1] 3.311113

median(df$pH) # 3.31

## [1] 3.31

var(df$pH) # 0.02383518

## [1] 0.02383518

sd(df$pH) # 0.1543865

## [1] 0.1543865

#Standard Deviation is 2  
  
mean(df$pH) -2\*sd(df$pH) # 3.00234 (Lower end)

## [1] 3.00234

mean(df$pH) +2\*sd(df$pH) # 3.619886 (Upper end)

## [1] 3.619886

#df$pH>3.00234 & df$pH< 3.619886  
  
sum((df$pH>3.00234 & df$pH< 3.619886) == TRUE)/ 1599 # 0.9530957

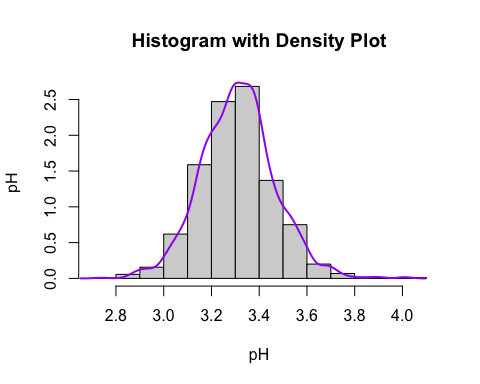
## [1] 0.9530957

############# percentile ################################################  
  
quantile(df$pH, p=c(0.25,0.5,0.75))

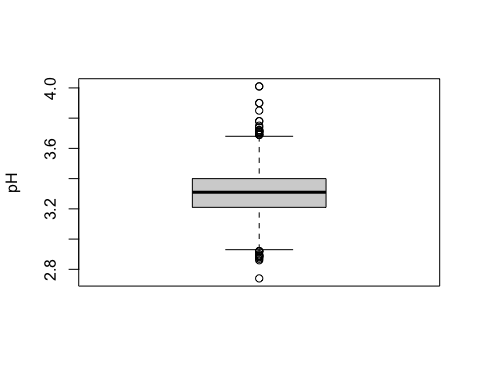
## 25% 50% 75%   
## 3.21 3.31 3.40

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
hist(df$pH, freq = FALSE,   
 xlab = "pH", ylab = "pH", main = "Histogram with Density Plot")  
lines(density(df$pH), lwd = 2, col = 'purple')



boxplot(df$pH ,ylab ="pH")



**e. Do you see any skewed distributions?**

The curve is following symmetry and there is no skewness also the median = means.

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

sulphates various calculations

mean(df$sulphates) # 0.6581488

## [1] 0.6581488

median(df$sulphates) # 0.62

## [1] 0.62

var(df$sulphates) # 0.02873262

## [1] 0.02873262

sd(df$sulphates) # 0.169507

## [1] 0.169507

#Standard Deviation is 2  
  
mean(df$sulphates) -2\*sd(df$sulphates) # 0.3191349 (Lower end)

## [1] 0.3191349

mean(df$sulphates) +2\*sd(df$sulphates) # 0.9971628 (Upper end)

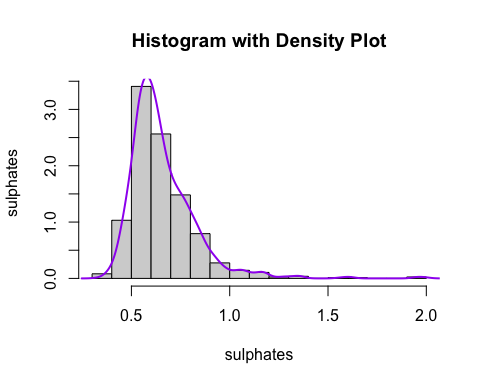
## [1] 0.9971628

#df$sulphates>0.3191349 & df$sulphates< 0.9971628  
  
sum((df$sulphates>0.3191349 & df$sulphates< 0.9971628) == TRUE)/ 1599 # 0.9631019

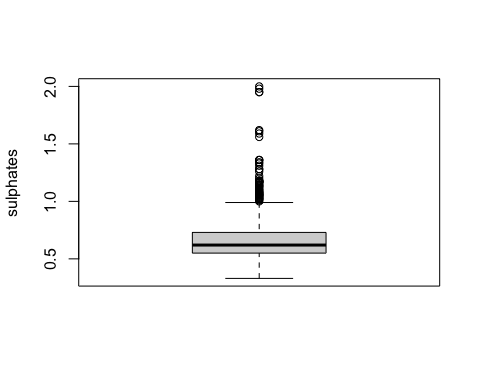
## [1] 0.9631019

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
hist(df$sulphates, freq = FALSE,   
 xlab = "sulphates", ylab = "sulphates", main = "Histogram with Density Plot")  
lines(density(df$sulphates), lwd = 2, col = 'purple')



boxplot(df$sulphates ,ylab ="sulphates")



**e. Do you see any skewed distributions?**

The right tail of the histogram of the distribution and the median is less than means.

The box plot show outliers values for sulphates

It is “right-skewed” or +ve skew

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

alcohol various calculations

mean(df$alcohol) #10.42298

## [1] 10.42298

median(df$alcohol) #10.2

## [1] 10.2

var(df$alcohol) # 1.135647

## [1] 1.135647

sd(df$alcohol) # 1.065668

## [1] 1.065668

#Standard Deviation is 2  
  
mean(df$alcohol) -2\*sd(df$alcohol) #8.291648 (Lower end)

## [1] 8.291648

mean(df$alcohol) +2\*sd(df$alcohol) #12.55432 (Upper end)

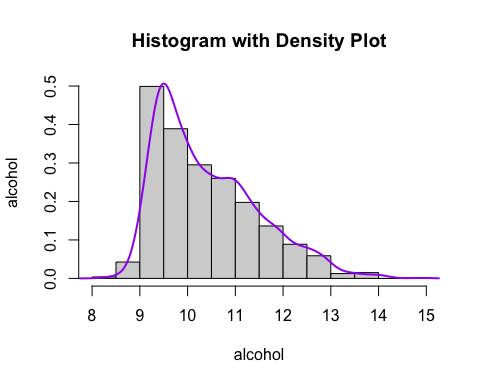
## [1] 12.55432

#df$alcohol>8.291648 & df$alcohol< 12.55432  
  
sum((df$alcohol>8.291648 & df$alcohol< 12.55432) == TRUE)/ 1599 #0.9562226

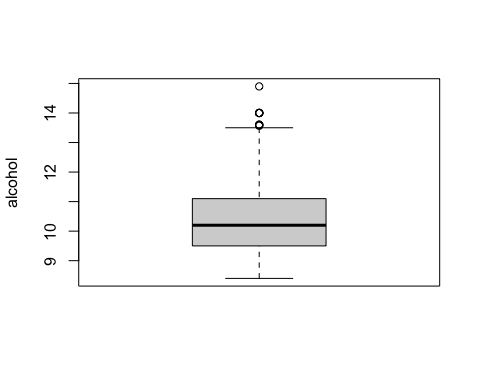
## [1] 0.9562226

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
  
hist(df$alcohol, freq = FALSE,   
 xlab = "alcohol", ylab = "alcohol", main = "Histogram with Density Plot")  
lines(density(df$alcohol), lwd = 2, col = 'purple')



boxplot(df$alcohol ,ylab ="alcohol")



**e. Do you see any skewed distributions?**

The right tail of the histogram of the distribution and the median is less than means.

The box plot show outliers values for sulphates

It is “right-skewed” or +ve skew

**c. How can you summarize the data of each variable in a concise way? What statistics are you going to present?**

Quality various calculations

mean(df$quality) # 5.636023

## [1] 5.636023

median(df$quality) # 6

## [1] 6

var(df$quality) # 0.6521684

## [1] 0.6521684

sd(df$quality) # 0.8075694

## [1] 0.8075694

#Standard Deviation is 2  
  
mean(df$quality) -2\*sd(df$quality) # 4.020884 (Lower end)

## [1] 4.020884

mean(df$quality) +2\*sd(df$quality) # 7.251161 (Upper end)

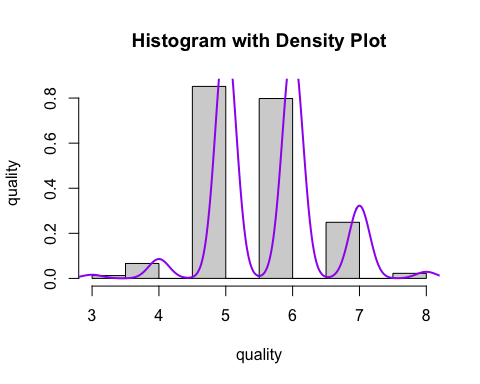
## [1] 7.251161

#df$quality>4.020884 & df$quality< 7.251161  
  
sum((df$quality>4.020884 & df$quality< 7.251161) == TRUE)/ 1599 # 0.9493433

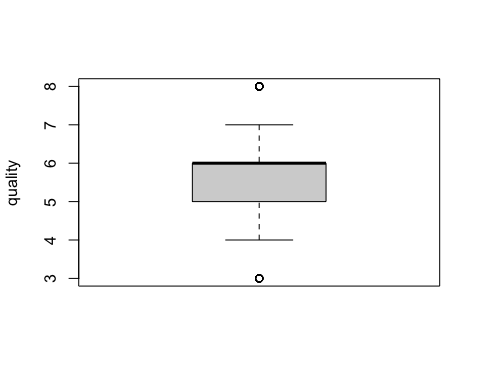
## [1] 0.9493433

**d. How can you visualize the distribution of each variable?**

library(ggplot2)  
  
hist(df$quality, freq = FALSE,   
 xlab = "quality", ylab = "quality", main = "Histogram with Density Plot")  
lines(density(df$quality), lwd = 2, col = 'purple')



boxplot(df$quality ,ylab ="quality")



e. **Do you see any skewed distributions?**

Quality is a discrete variable, we can’t find skewness for this variable and it is a final variable which depends on other variables.