

WhiteWine

Harshita Ved

4/24/2020

```
library(rmarkdown)
library(ggplot2)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.3
```

```
## -- Attaching packages ----- tidyverse
```

```
## v tibble  2.1.3      v dplyr    0.8.4
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
## v purrr   0.3.3
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
## Warning: package 'forcats' was built under R version 3.6.3
```

```
## -- Conflicts ----- tidyverse
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(dplyr)
library(readr)
library(DBI)
library(RSQLite)
```

```
## Warning: package 'RSQLite' was built under R version 3.6.3
```

```
library(purrr)
library(mlbench)
```

```
## Warning: package 'mlbench' was built under R version 3.6.3
```

```
library(modelr)
```

```
## Warning: package 'modelr' was built under R version 3.6.3
```

```
library(tibble)
```

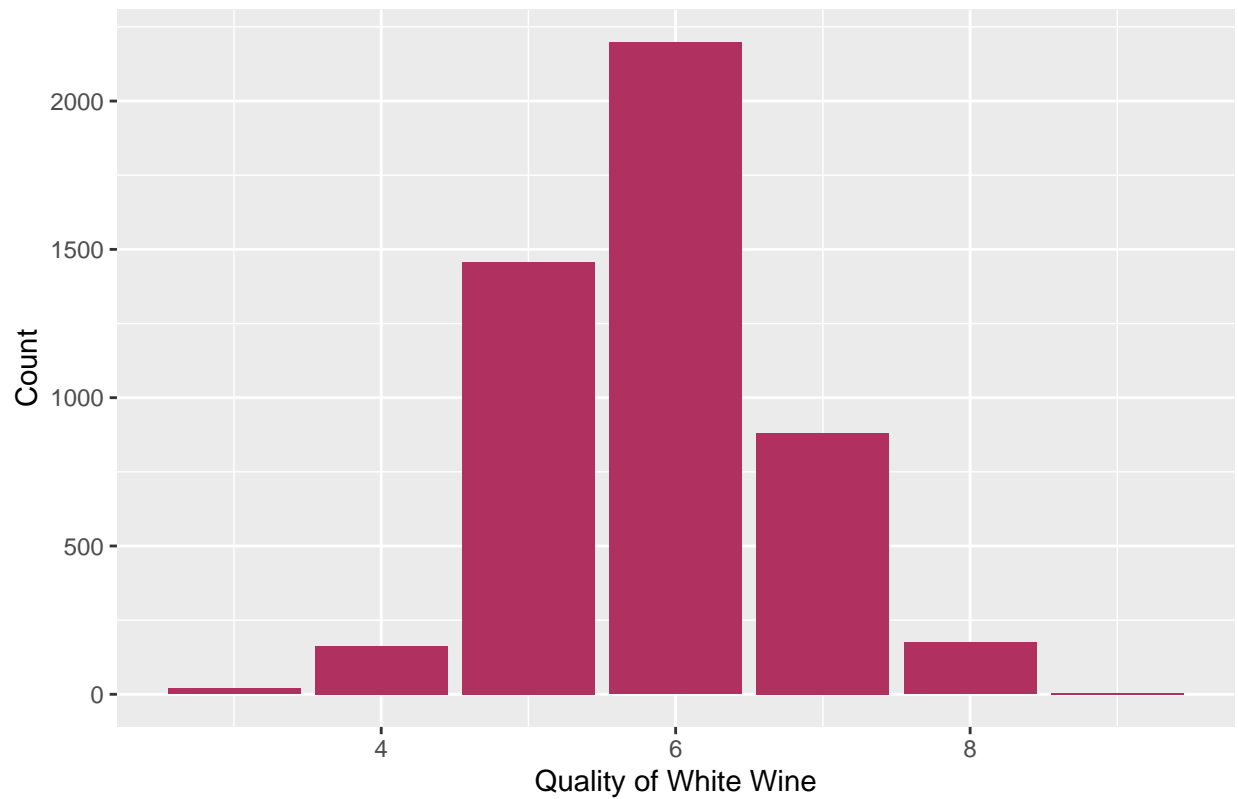
```
#The data set contains various different attributes of White Wine like pH, sugar content, density, alcohol content, etc.
#White Wine Dataset - https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/
#4898 observations
```

```
WhiteWineData <- read.csv("D:\\Coursework\\Semester 1 Fall 2019\\DS 5110 - Introduction to Data Management\\WhiteWineData.csv")
WhiteWineData <- filter(WhiteWineData, `fixed.acidity` != "" | `volatile.acidity` != "" | `citric.acidity` != "" | `residual.sugar` != "" | `chlorides` != "" | `free.sulfur.dioxide` != "" | `total.sulfur.dioxide` != "" | `density` != "" | `pH` != "" | `sulphates` != "" | `alcohol` != "")
head(WhiteWineData, n =10)
```

```
##      fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1             7.0             0.27           0.36           20.7       0.045
## 2             6.3             0.30           0.34           1.6       0.049
## 3             8.1             0.28           0.40           6.9       0.050
## 4             7.2             0.23           0.32           8.5       0.058
## 5             7.2             0.23           0.32           8.5       0.058
## 6             8.1             0.28           0.40           6.9       0.050
## 7             6.2             0.32           0.16           7.0       0.045
## 8             7.0             0.27           0.36           20.7       0.045
## 9             6.3             0.30           0.34           1.6       0.049
## 10            8.1             0.22           0.43           1.5       0.044
##      free.sulfur.dioxide total.sulfur.dioxide density    pH sulphates alcohol
## 1                    45                   170 1.0010 3.00      0.45      8.8
## 2                    14                   132 0.9940 3.30      0.49      9.5
## 3                    30                    97 0.9951 3.26      0.44     10.1
## 4                    47                   186 0.9956 3.19      0.40      9.9
## 5                    47                   186 0.9956 3.19      0.40      9.9
## 6                    30                    97 0.9951 3.26      0.44     10.1
## 7                    30                   136 0.9949 3.18      0.47      9.6
## 8                    45                   170 1.0010 3.00      0.45      8.8
## 9                    14                   132 0.9940 3.30      0.49      9.5
## 10                   28                   129 0.9938 3.22      0.45     11.0
##      quality
## 1           6
## 2           6
## 3           6
## 4           6
## 5           6
## 6           6
## 7           6
## 8           6
## 9           6
## 10          6
```

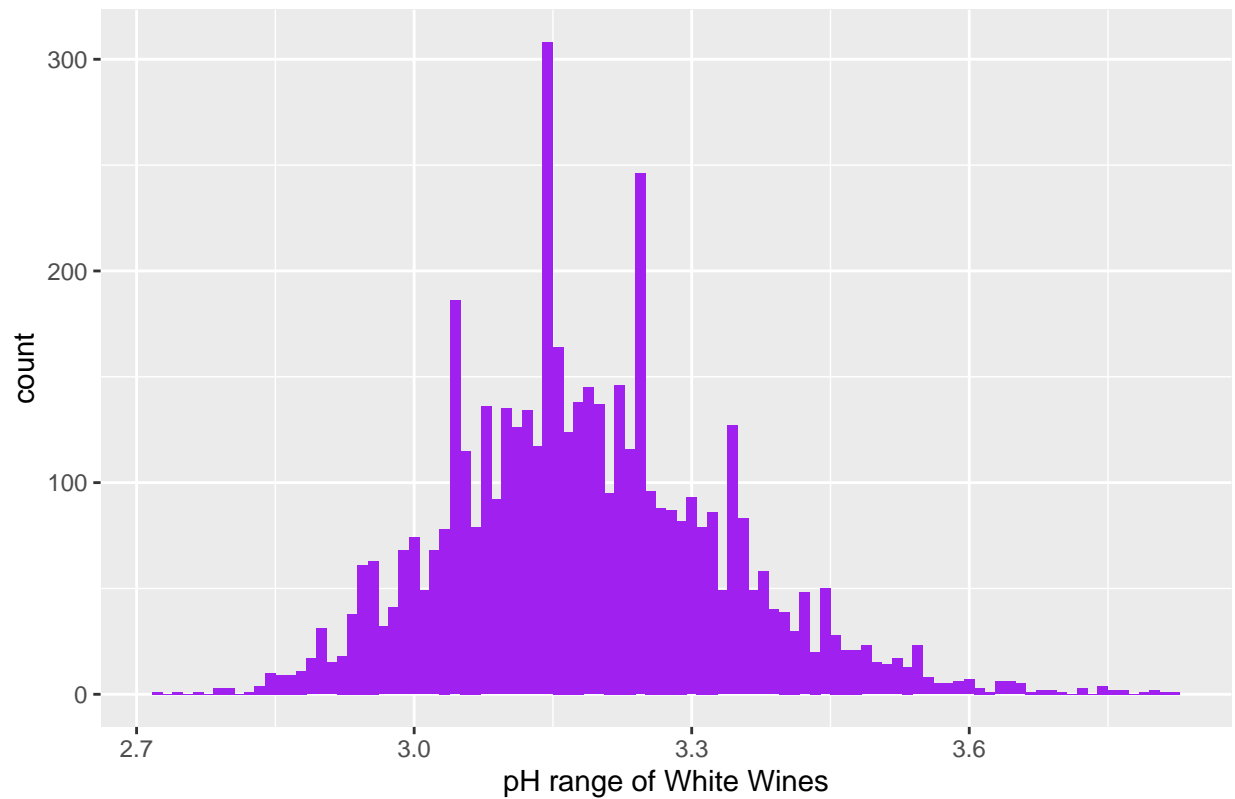
```
#Quality of White Wine
ggplot(data=WhiteWineData) +
  geom_bar(aes(x=`quality`), fill="maroon") +
  xlab("Quality of White Wine") + ylab("Count") +
  ggtitle("Quality Distribution of the dataset of White Wines")
```

Quality Distribution of the dataset of White Wines

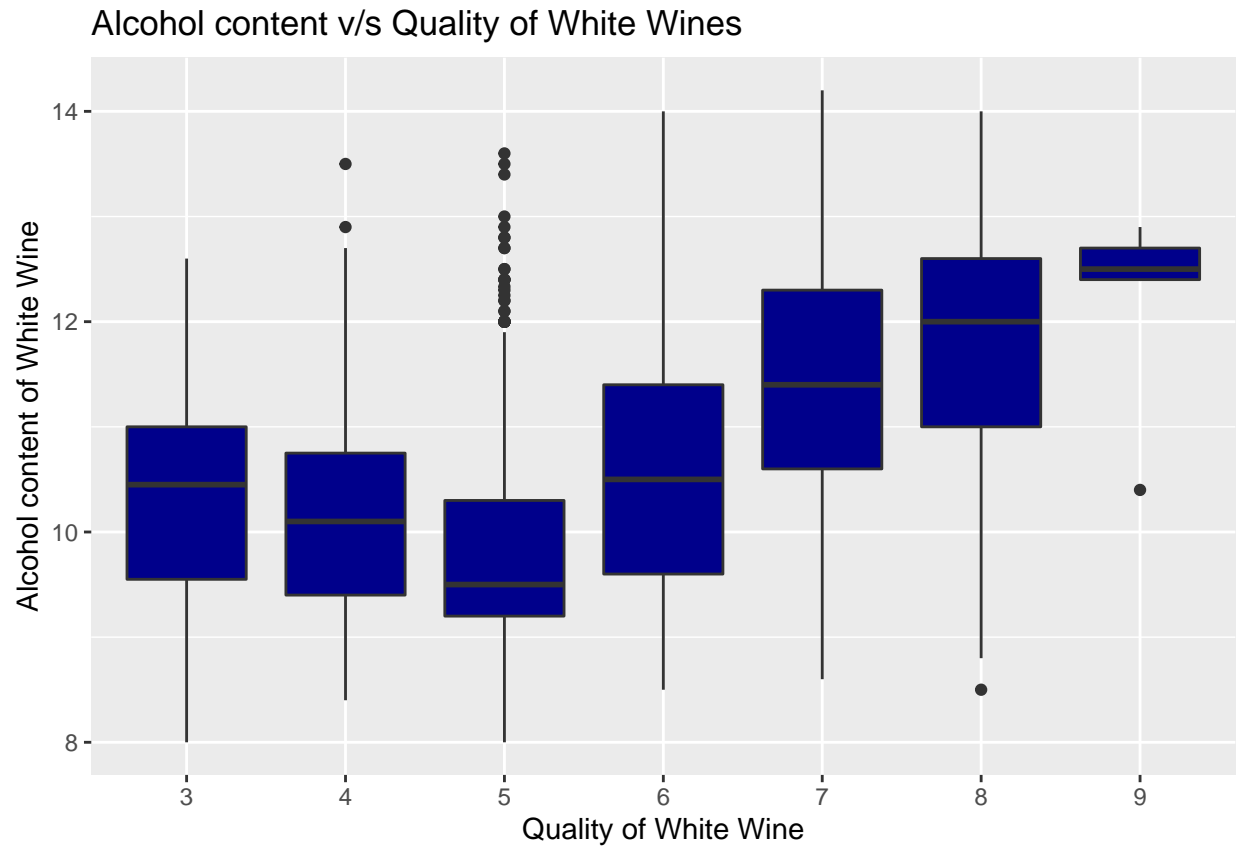


```
#pH range of White Wines  
ggplot(data=WhiteWineData) +  
  geom_histogram(aes(x=pH), bins=100, fill="purple") +  
  xlab("pH range of White Wines") +  
  ggtitle("pH range of White Wines with respect to Count")
```

pH range of White Wines with respect to Count



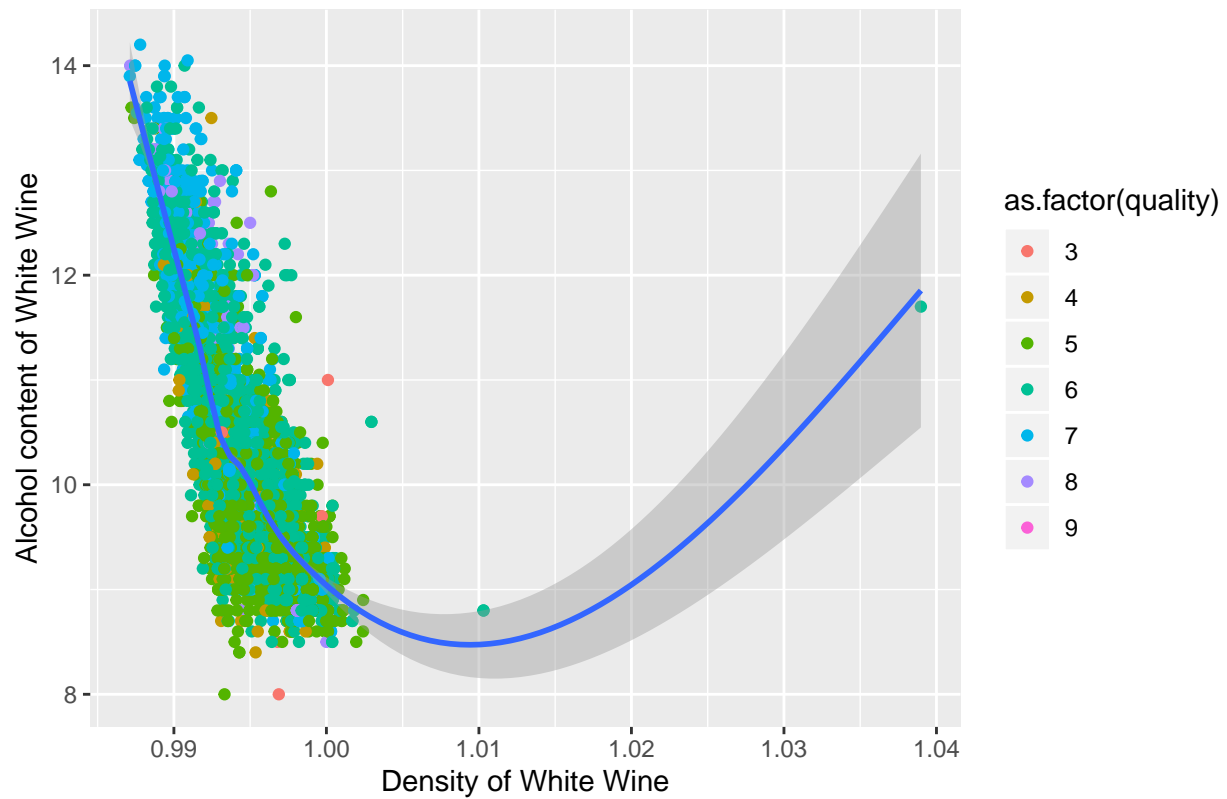
```
#Alcohol content v/s Quality of White Wines  
ggplot(data=WhiteWineData) +  
  geom_boxplot(aes(x=as.factor(quality), y=`alcohol`), fill="dark blue") +  
  xlab("Quality of White Wine") + ylab("Alcohol content of White Wine") +  
  ggtitle("Alcohol content v/s Quality of White Wines")
```



```
#Alcohol content v/s Density of White Wines
ggplot(data=WhiteWineData) +
  geom_point(aes(x=density, y=`alcohol`, color = as.factor(quality))) +
  xlab("Density of White Wine") + ylab("Alcohol content of White Wine") +
  ggtitle("Alcohol content v/s Density of White Wines") + geom_smooth(aes(x=density, y=`alcohol`))

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

Alcohol content v/s Density of White Wines



```
#Sulphur Dioxide v/s Quality of White Wines
ggplot(data=WhiteWineData) +
  geom_boxplot(aes(x=as.factor(quality), y=`total.sulfur.dioxide`), fill="yellow") +
  xlab("Quality of White Wine") + ylab("sulfur-dioxide content of White Wine") +
  ggtitle("Sulphur Dioxide v/s Quality of White Wines")
```

Sulphur Dioxide v/s Quality of White Wines

