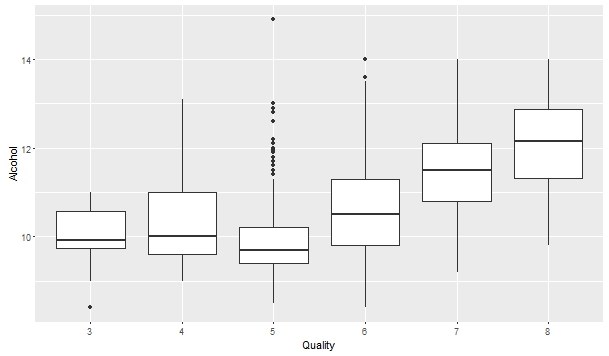
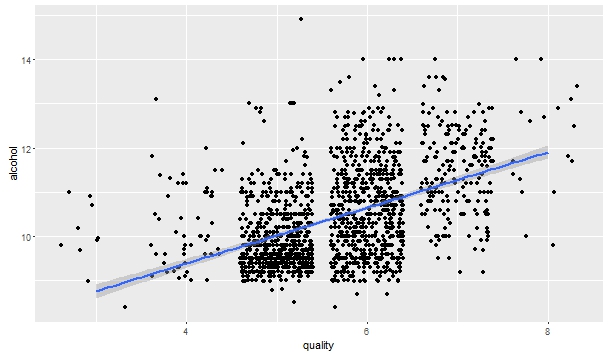
1. **Create visualizations to depict how residual sugar, density and alcohol affect the quality of the Wine.**
2. **R Code for Visualizations:**
3. path<-"C:/Users/Rahul Kumar/Desktop/Singulariti"
4. setwd(path)
5. data<-read.csv("winequality-red.csv",sep=";",header=TRUE)
6. ggplot(data=data, aes(as.factor(quality),alcohol))+geom\_boxplot() +xlab("Quality") +ylab("Alcohol")
7. ggplot(data=data, aes(quality,alcohol))+geom\_jitter() +geom\_smooth(method='lm')
8. ggplot(data=data, aes(as.factor(quality),residual.sugar))+geom\_boxplot() +xlab("Quality")+ylab("Residual Sugar")
9. ggplot(data=data, aes(quality,residual.sugar))+geom\_jitter() +geom\_smooth(method='lm')
10. ggplot(data=data, aes(as.factor(quality),density))+geom\_boxplot() +xlab("Quality")+ylab("Density")
11. ggplot(data=data, aes(quality,density))+geom\_jitter() +geom\_smooth(method='lm')
12. **Alcohol and Quality:**

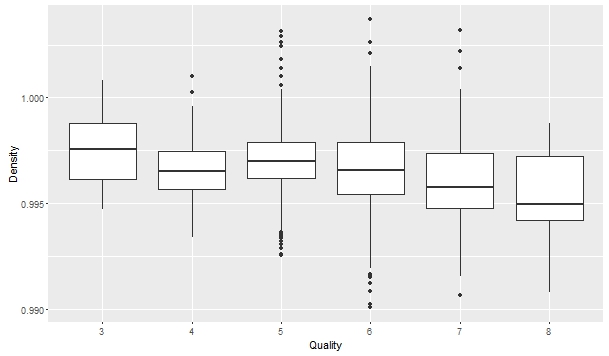


This boxplot shows that quality 6, 7 and 8 have higher alcohol content. This means alcohol content increase with quality.

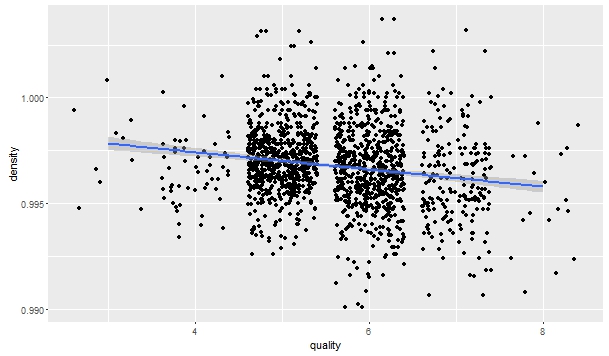


Scatterplot shows that quality and alcohol are positively correlated.

1. **Density and Quality:**

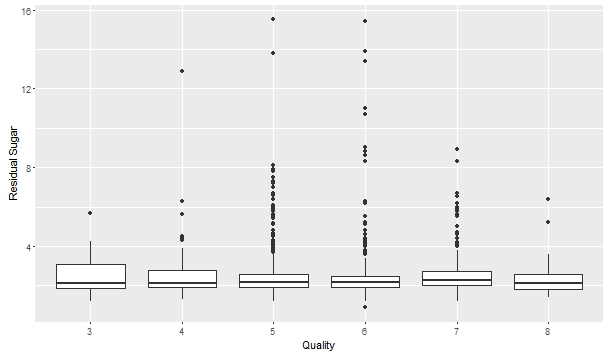
****

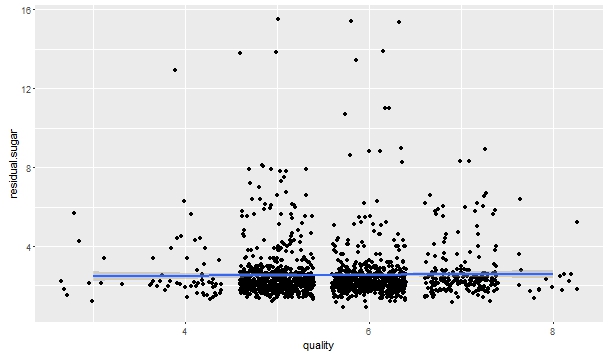
This boxplot shows that density decrease with quality increase.



Scatterplot shows that quality and density are negatively correlated.

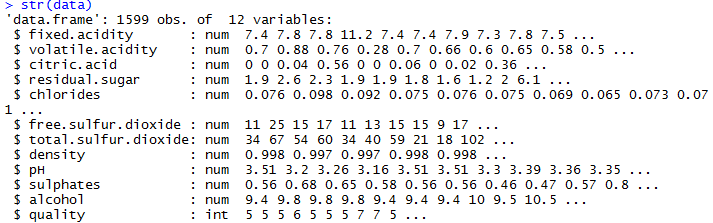
1. **Residual Sugar and Quality:**

****

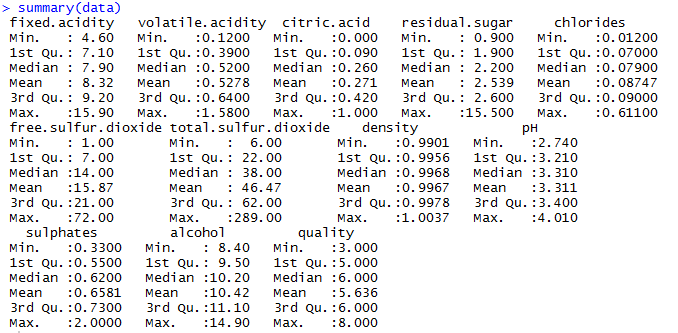
****

Scatterplot shows that quality and residual sugar are no correlated.

1. **Build a model based on this data to predict the Quality score for sample of red wine given all attributes from 1-11.**
2. **R Code for Modelling:**
3. **Loading Data in R:**
4. path<-"C:/Users/Rahul Kumar/Desktop/Singulariti"
5. setwd(path)
6. data<-read.csv("winequality-red.csv",sep=";",header=TRUE)
7. **Univariate Analysis:**
8. str(data)



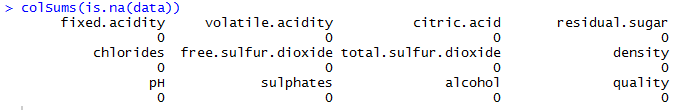
1. summary(data)



1. **Missing Value Checking:**
2. table(is.na(data))

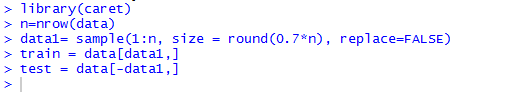


1. colSums(is.na(data))

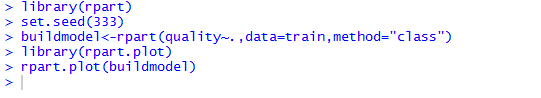


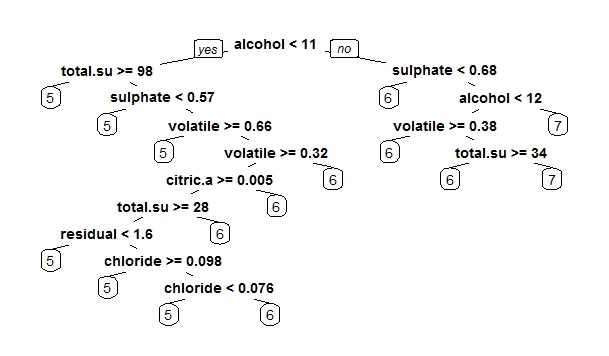
This shows that no missing values are present in data.

1. **Divide the data into two parts train and test data. 70 percent data will be train and 30 percent will be test data.**
2. library(caret)
3. n=nrow(data)
4. data1= sample(1:n, size = round(0.7\*n), replace=FALSE)
5. train = data[data1,]
6. test = data[-data1,]



1. **Building Model Using Decision Tree:**
2. library(rpart)
3. set.seed(333)
4. buildmodel<-rpart(quality~.,data=train,method="class")
5. library(rpart.plot)
6. rpart.plot(buildmodel)





1. **Make Prediction:**
2. prediction<-predict(buildmodel,newdata=test,type="class")



1. **Analyze Results:**
2. library(caret)
3. library(e1071)
4. confusionMatrix(prediction,test$quality)

