**Wine Quality data - Red wine**

1594014 김미현

1585063 박효선

**Ⅰ. Purpose**

1. 화학적 특징과 실제 등급과의 관련성을 구분

2. Quality 에 영향을 미치는 요인 찾기 (높은/ 낮은 품질에 영향을 미치는 요인?)

3. Classification

(1) Quality (등급) - 'good', 'average', 'bad’

- quality 화학적 특성과 전문가가 분류한 등급의 상관성이 있는지 파악하기

(2) Sweetness (당도) - 'Dry', 'Medium Dry', 'Medium Sweet', 'Sweet'

- residual sugar….. etc

(3) Body (무게감) – light bodied, medium bodied, full- bodied

- Density….. etc

4. 새로운 데이터가 들어왔을 때 등급 예측

**Ⅱ. EDA**

1. Variables

Table variable

|  |  |  |
| --- | --- | --- |
| 1 | fixed acidity | 불휘발산 |
| 2 | volatile acidity | 휘발성산 |
| 3 | citric acid | 구연산 |
| 4 | residual sugar | 잔당(桟糖) |
| 5 | Chlorides | 염화물 |
| 6 | free sulfur dioxide | free 이산화황 |
| 7 | total sulfur dioxide | total 이산화황 |
| 8 | Density | 농도 |
| 9 | pH | 수소이온농도지수 |
| 10 | Sulphates | 황산염 |
| 11 | Alcohol | 알코올 |
| 12 | quality (score between 0 and 10) | 등급( 0~10 점) secondary data |

pH: 0에서 14까지 있으며, 7미만은 산성을, 7이상은 알칼리성

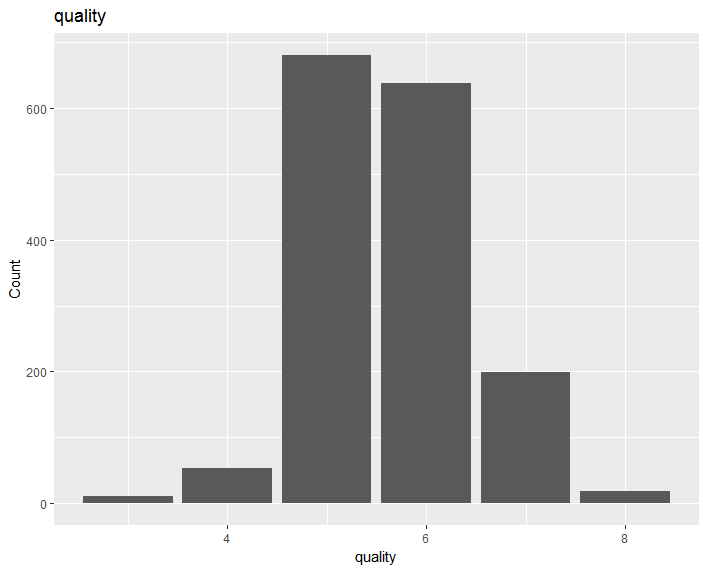
1. **Fixed acidity**: most acids involved with wine or fixed or nonvolatile (do not evaporate readily)
2. **Volatile acidity**: the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste
3. **Citric acid**: found in small quantities, citric acid can add 'freshness' and flavor to wines
4. **Residual sugar**: the amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and wines with greater than 45 grams/liter are considered sweet
5. **Chlorides**: the amount of salt in the wine
6. **Free sulfur dioxide**: the free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine
7. **Total sulfur dioxide**: amount of free and bound forms of S02; in low concentrations, SO2 is mostly undetectable in wine, but at free SO2 concentrations over 50 ppm, SO2 becomes evident in the nose and taste of wine
8. **Density**: the density of water is close to that of water depending on the percent alcohol and sugar content
9. **pH**: describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale

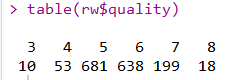
더 높으면 시큼하며, 낮을 경우는 너무 부드럽고 밋밋한 맛이 난다. 레드 와인은 3.3 ~ 3.6이

1. **Sulphates**: a wine additive which can contribute to sulfur dioxide gas (S02) levels, wich acts as an antimicrobial and antioxidant
2. **Alcohol**: the percent alcohol content of the wine
3. **Quality**: output variable (based on sensory data, score between 0 and 10)

2. EDA

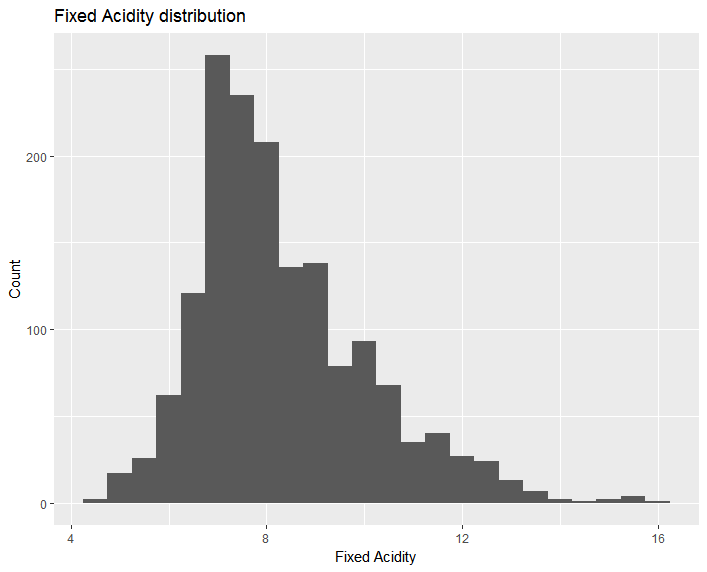
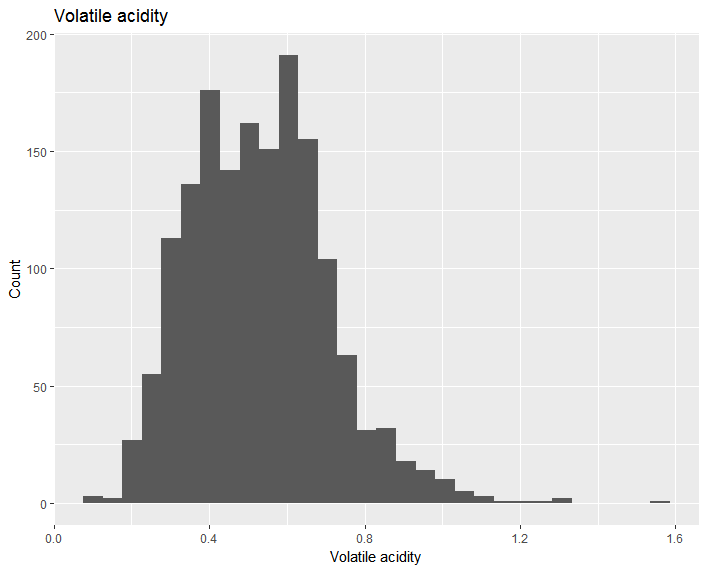
(1) Bar plot

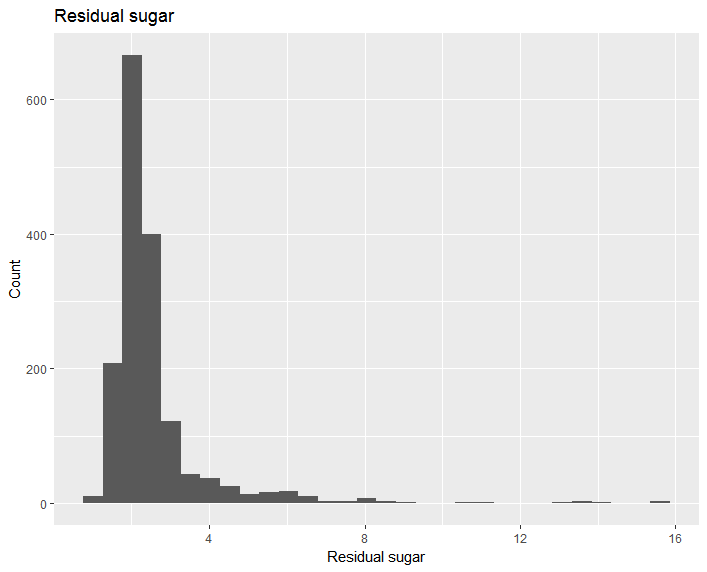
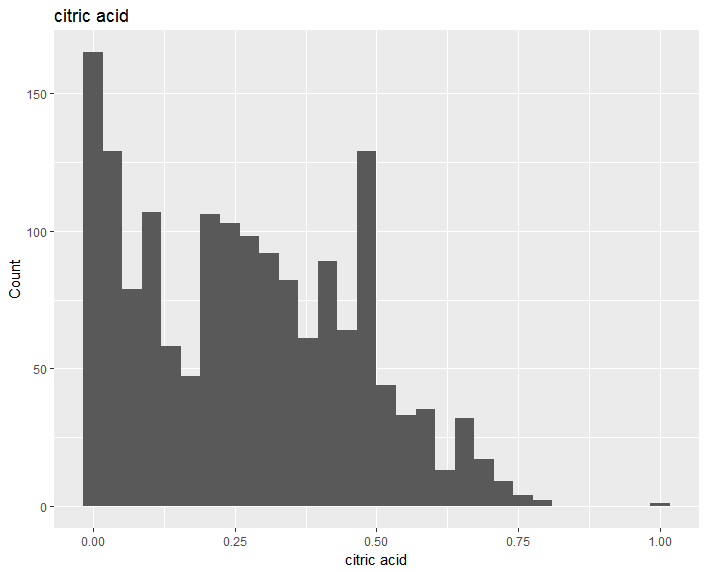
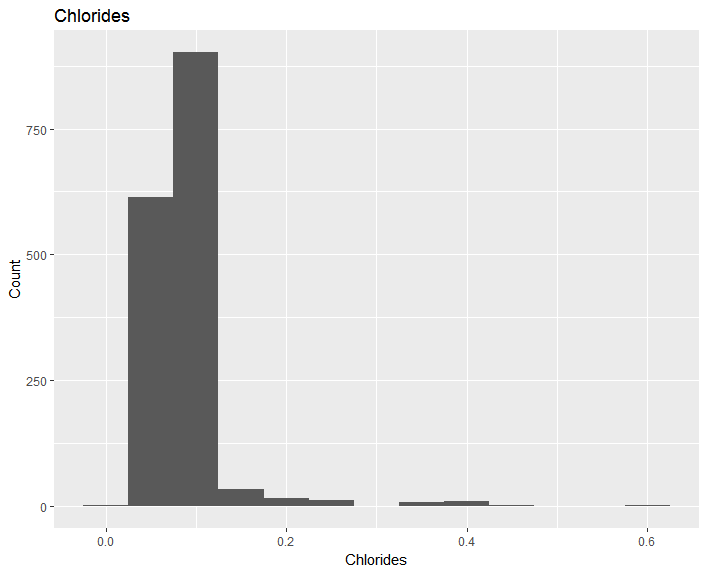
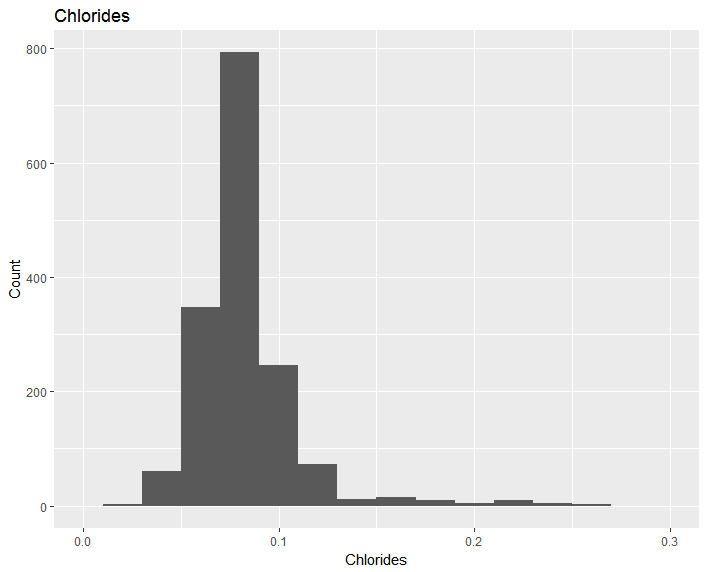


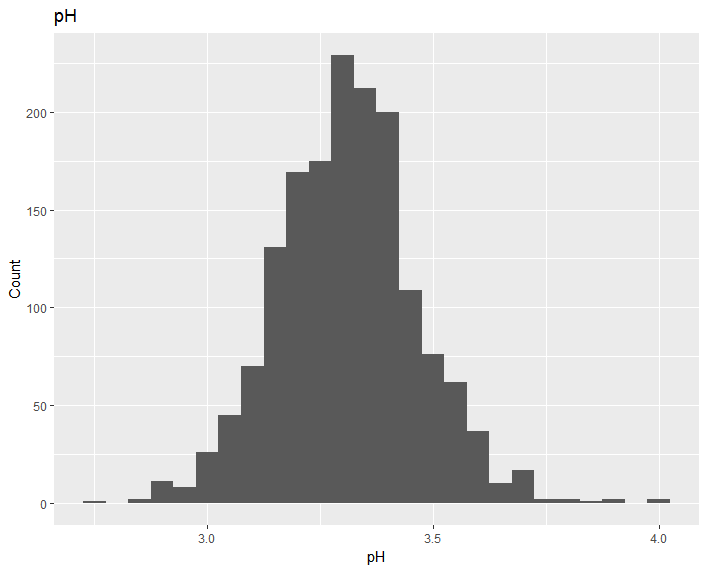
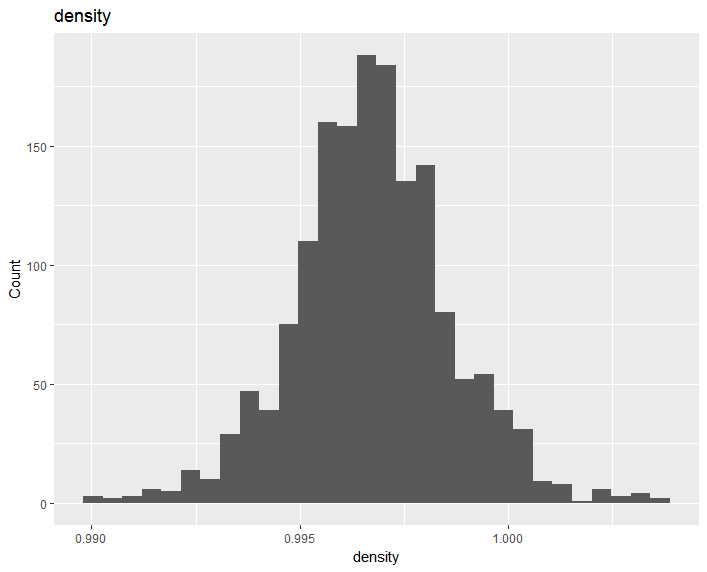
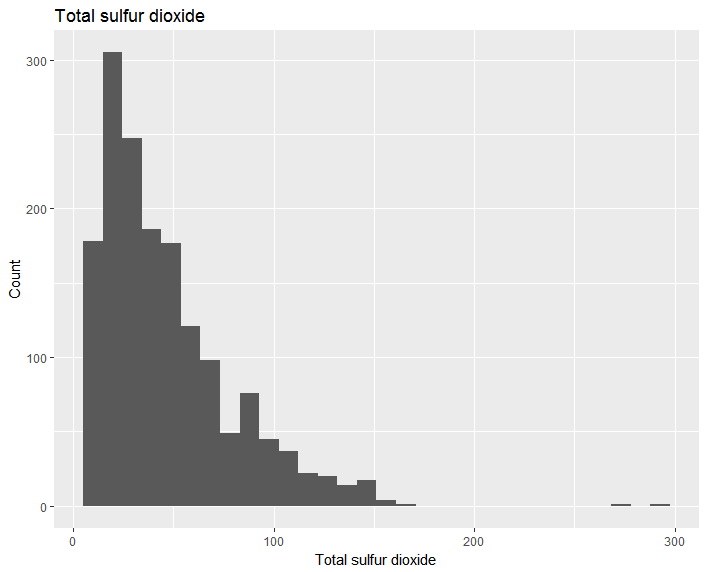
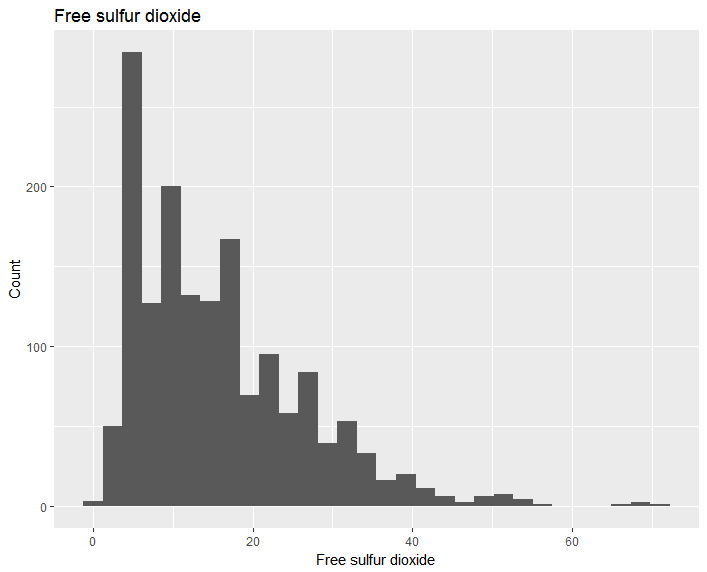


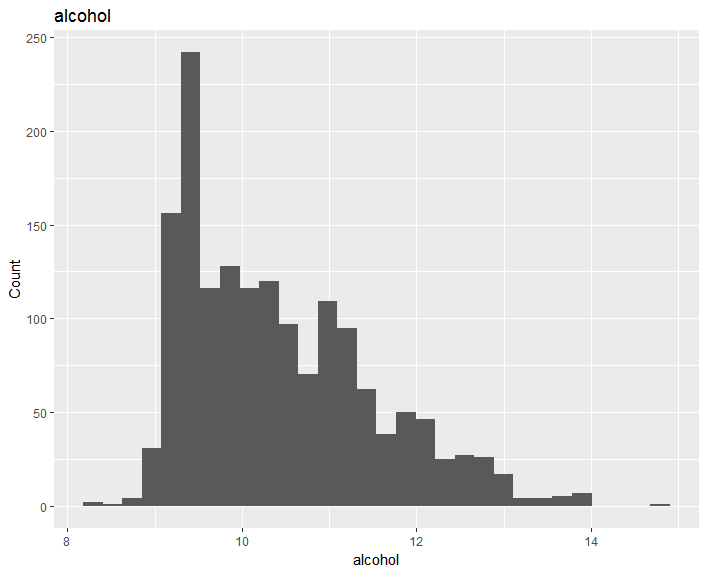
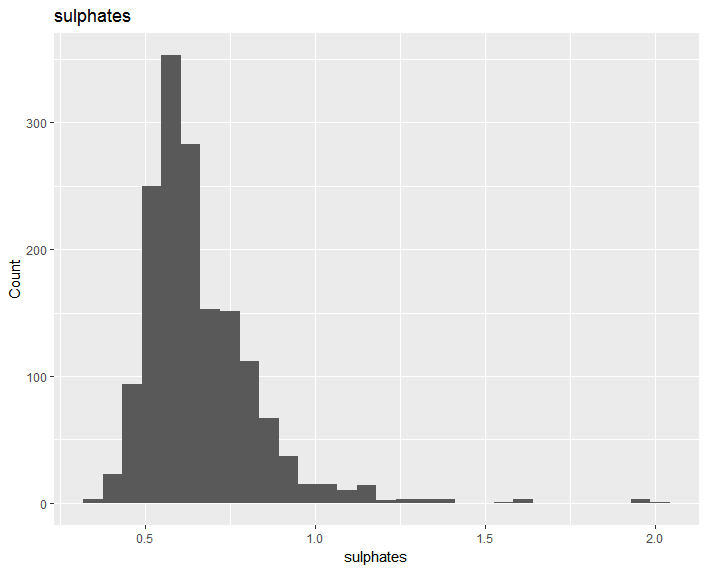
Quality 5,6,7 에 밀집되어 있기 때문에

Low quality 3,4 와 high quality 8 에 영향을 미치는 요인을 분석하는 것이 유의미할 것이다

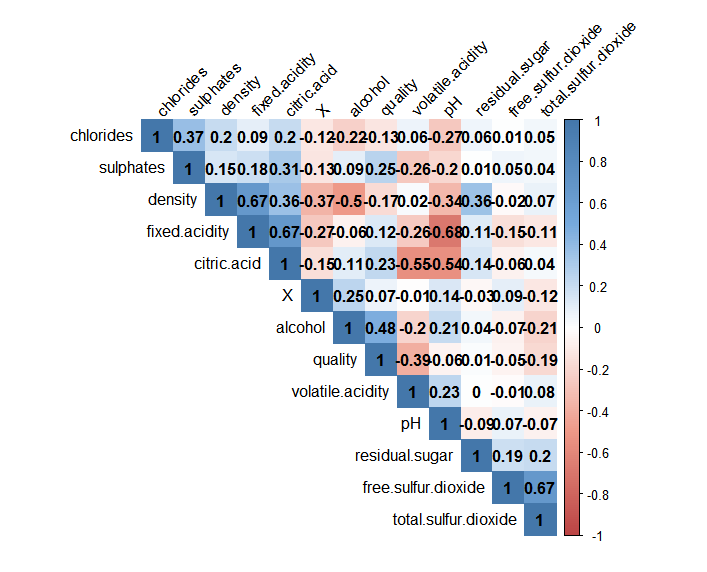
 

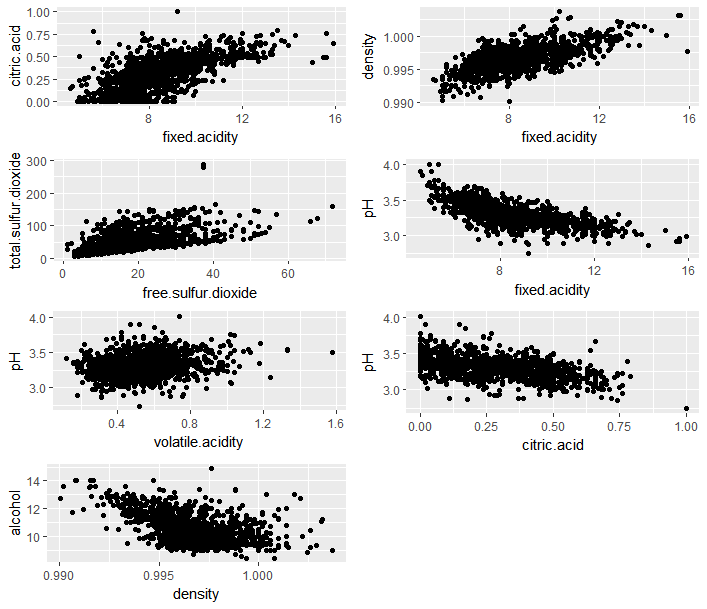
  





(2) Correlation

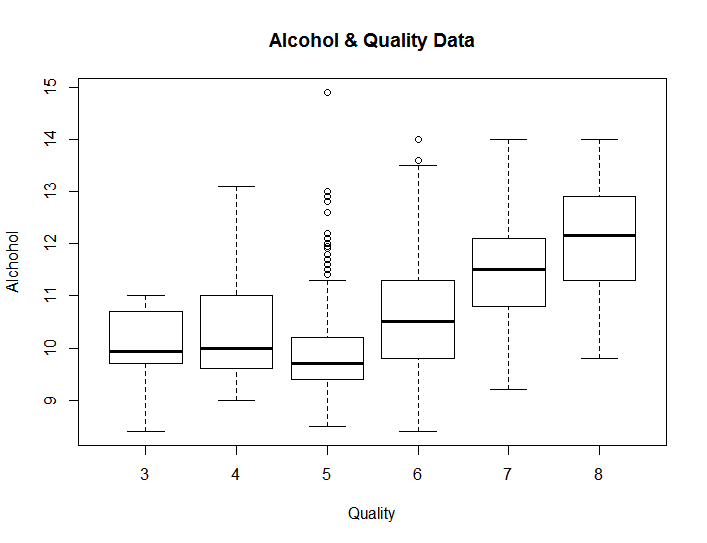


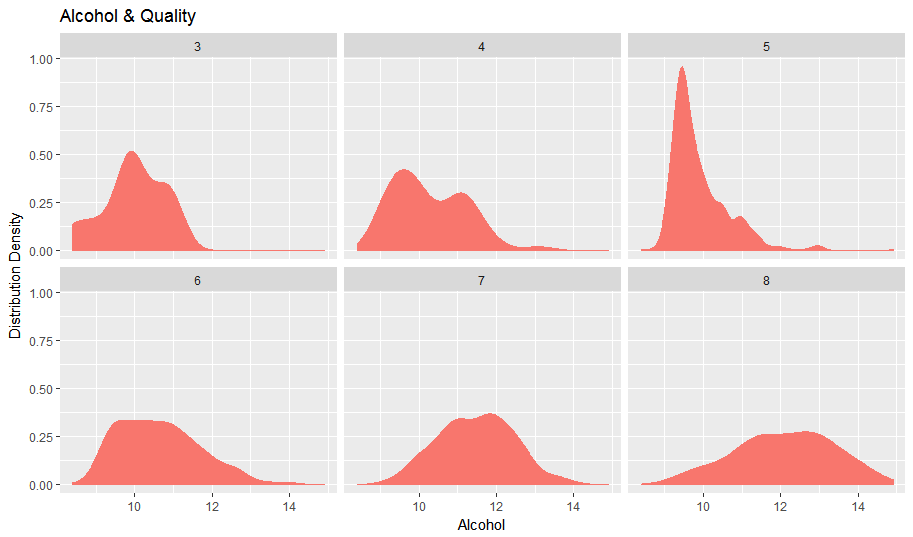


변수간 상관관계가 높은 변수 추출하여 scatter plot

(3) Quality에 영향을 미치는 요인 파악하기

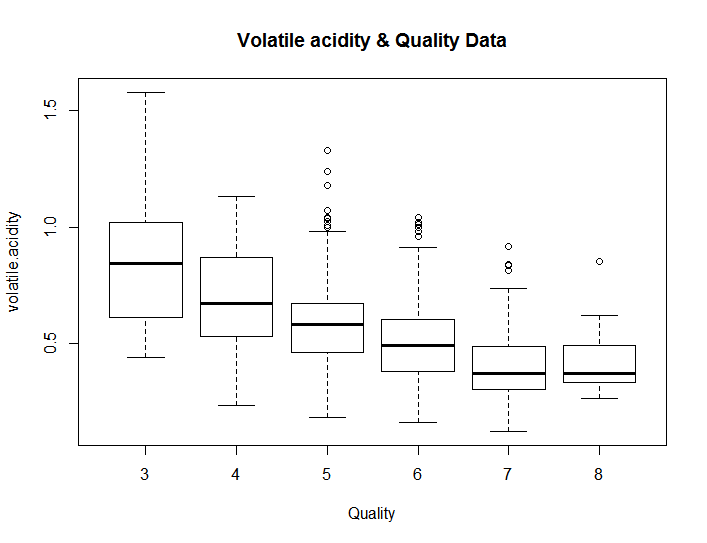
1) Alcohol - Quality 비교

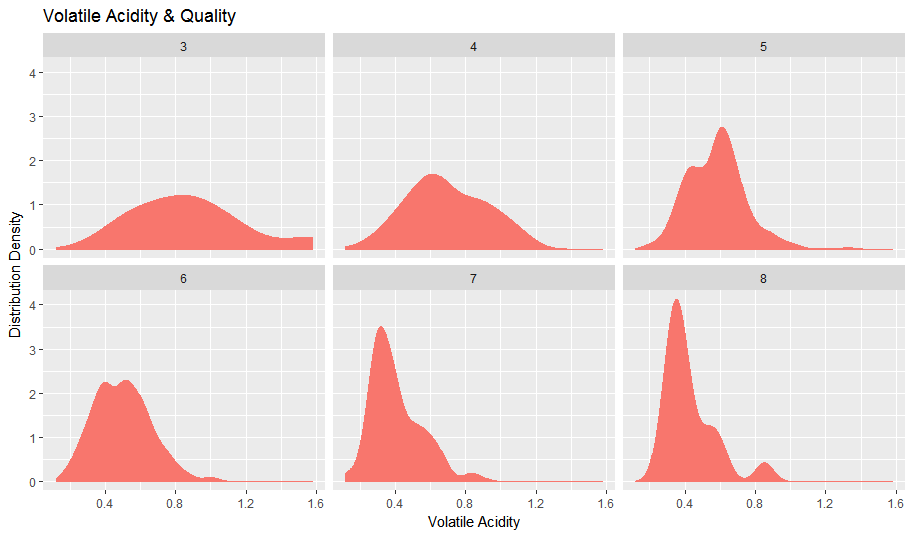




Quality 별 alcohol 의 분포

2) volatile acidity – Quality 비교





Quality 별 volatile acidity 의 분포