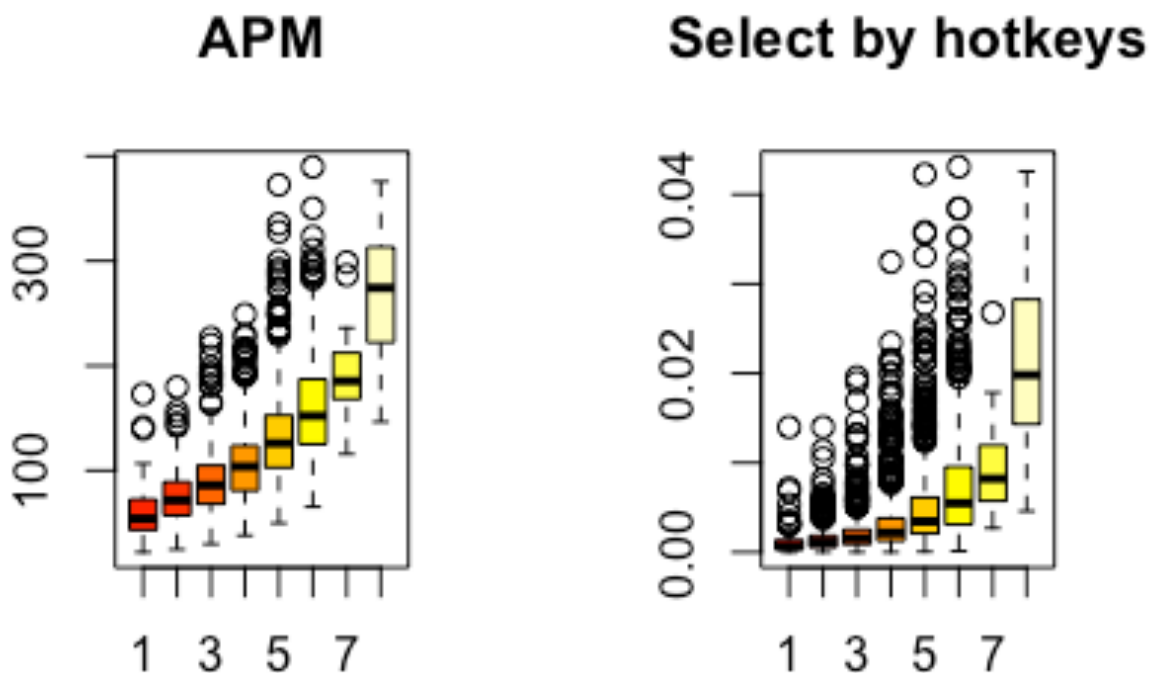


탐색적 자료분석과 시각화 과제(1)

Skillcraft

등급별로 Skillcraft 데이터의 탐색적 자료분석을 실시하였으며 어떤 변수가 다음 레벨로 승급하기 위한 필요 변수인지 탐색하는 것이 이 과제의 목표이다.

(1) Action Per Minute / Select By Hot keys



등급별로 APM 및 Select by Hotkeys의 분포는 등급이 높을 수록 높아진다. 따라서 APM과 Select by Hotkeys는 모든 레벨에서 다음 레벨로 승급하기 위한 위한 조건이라고 볼 수 있다. 또한 APM과 SBH 간에 상관관계가 있다는 것을 알 수 있다.

```
> cor.test(a$APM, a$SelectByHotkeys)
```

Pearson's product-moment correlation

data: a\$APM and a\$SelectByHotkeys

t = 90.432, df = 3393, p-value < 2.2e-16

2017. 8. 31.

배주연

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

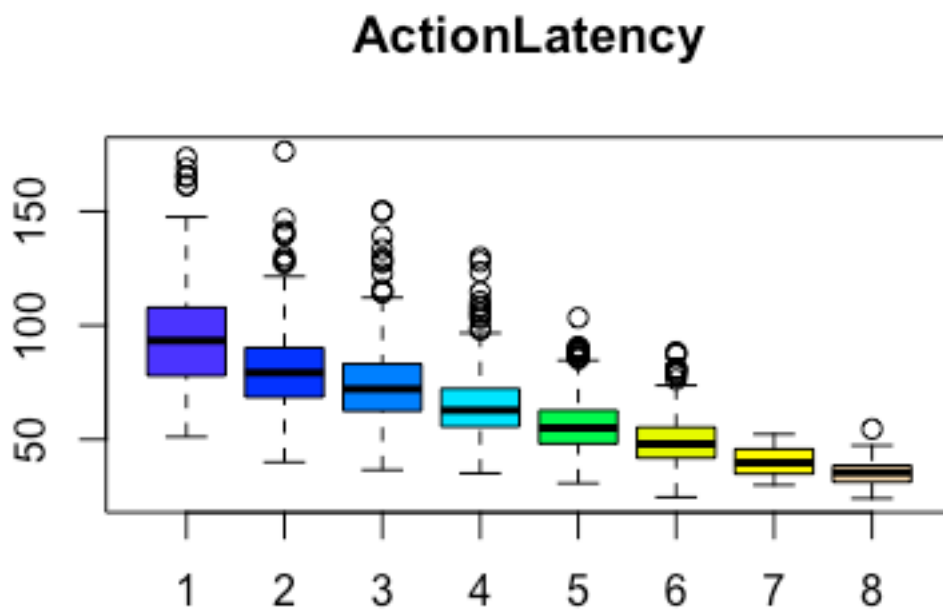
0.8305412 0.8502859

sample estimates:

cor

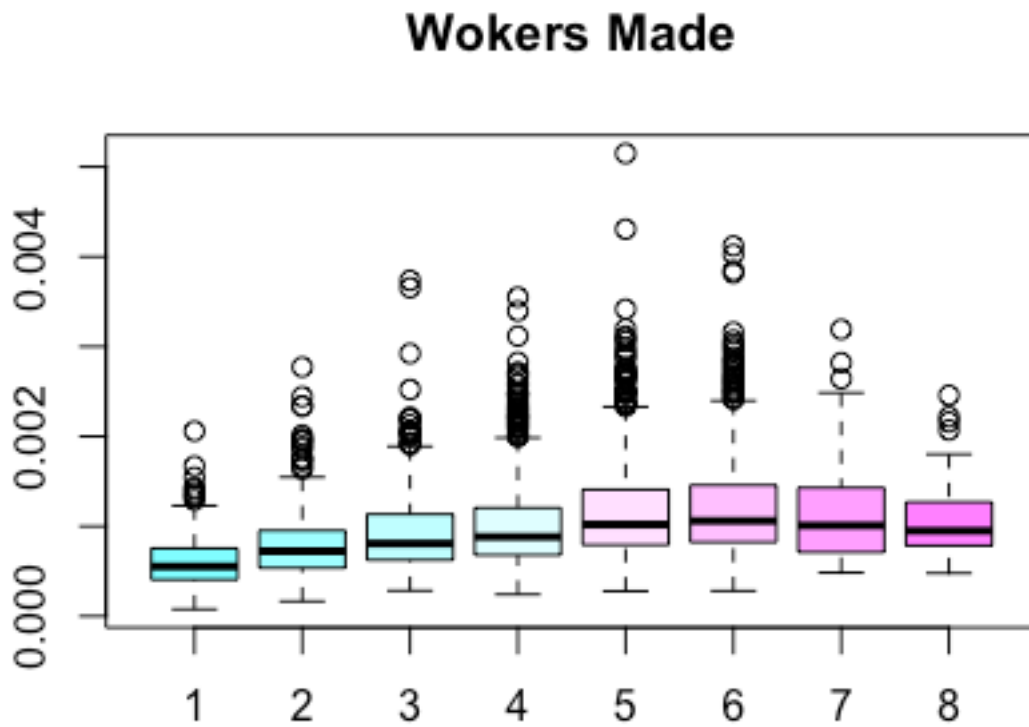
0.8406927

(2) Action Latency



Action Latency의 경우 등급과 음의 상관관계를 보이는데 선형적이라는 것이 특징이다. 등급이 높을 수록 PAC의 첫번째 동장에서 완료할 동작까지 걸리는 평균시간이 적다.

(3) Workers Made

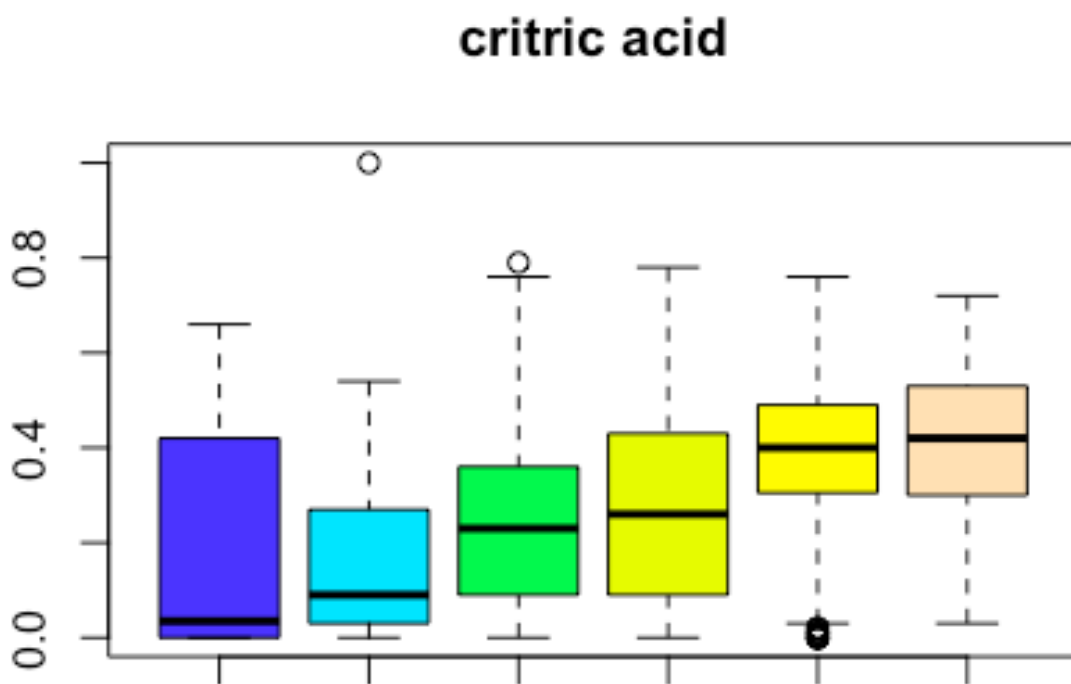


하나의 타임프레임 내의 평균 일꾼 수는 Bronze, Silver, Gold, Platinum 등급 까지 소량 증가하지만 Diamond, Master, GrandMaster, Professional 등급에서는 큰 차이가 없음을 알 수 있다. 따라서 일꾼 수를 무작정 늘리는 것은 등급을 올리기 위한 효과적인 방안이 아니다.

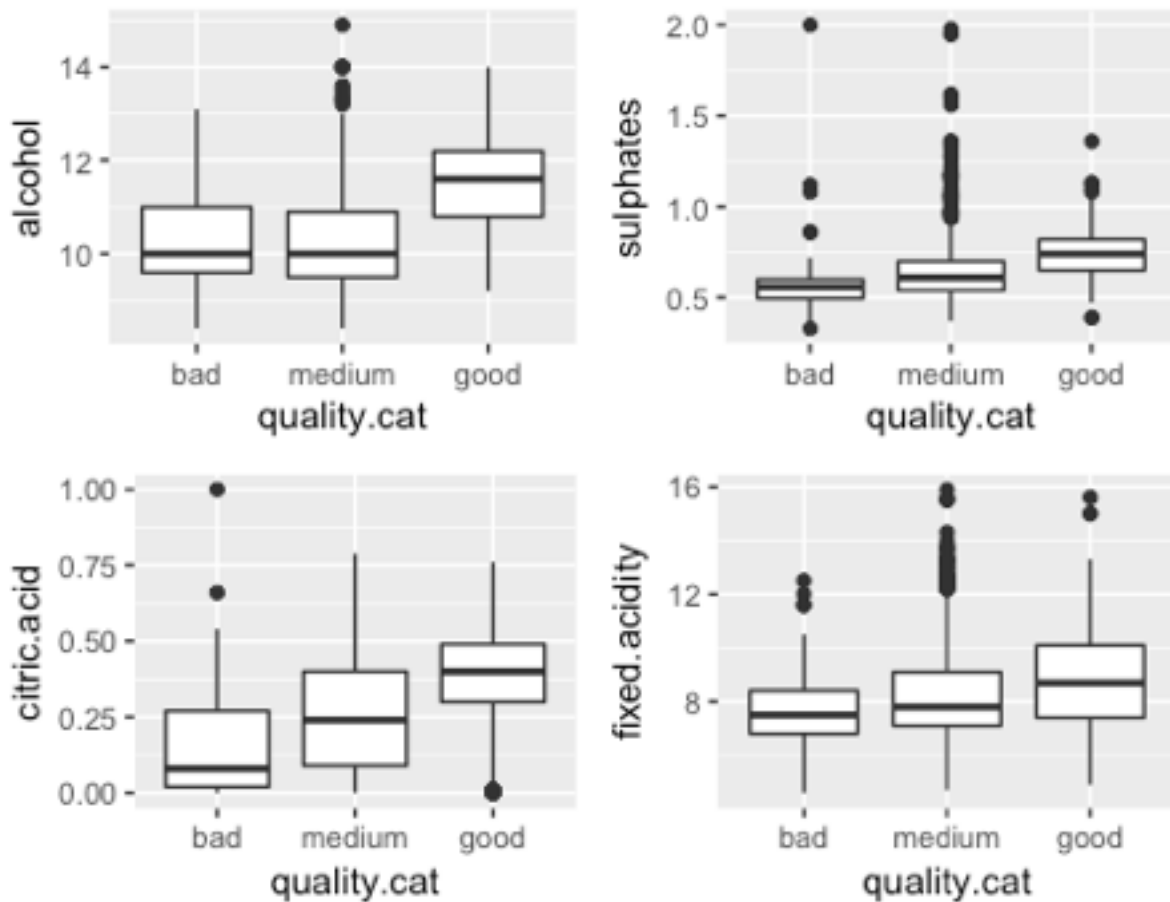
Wine Quality data

등급별로 Wine Quality data(red, white) 데이터의 탐색적 자료분석을 실시하였다. red wine , white wine을 각각 살펴보며 wine quility와 변수들의 관계를 파악하는 것이 목표이다.

1. Red wine



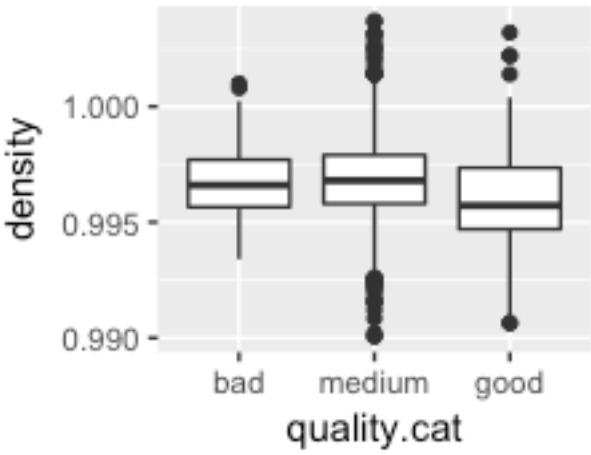
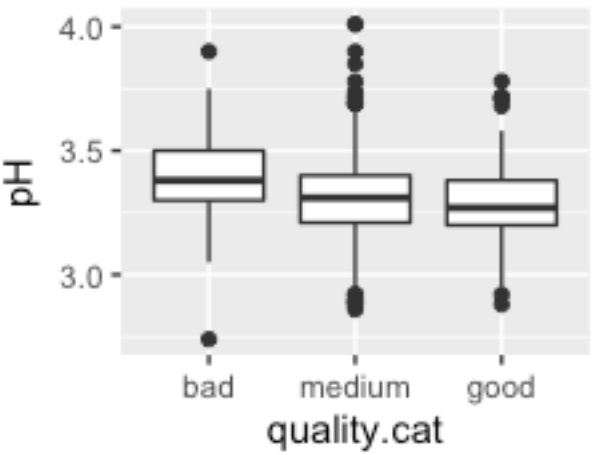
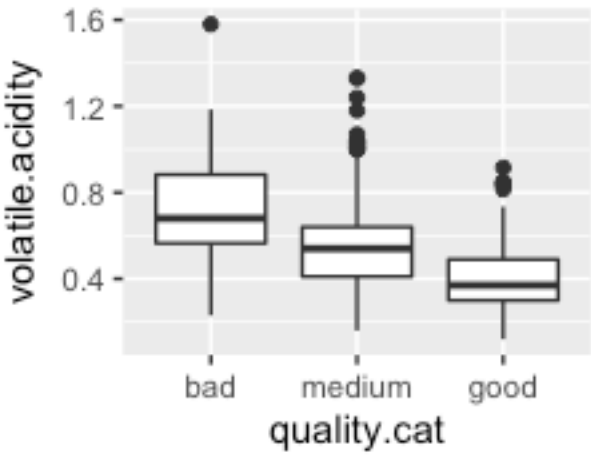
quality-critic acid boxplot에서도 확인할 수 있듯이 3번과 4번, 5번과 6번, 7번과 8번 등급은 비슷한 분포를 가지는 것으로 보인다. 이후의 분석에서는 3번 4번 와인을 'bad'로, 5번 6번 와인을 'medium', 7번 8번 와인을 'good'으로 범주화했다.

(1) Increasing quality of red wine

양의 상관관계를 가지는 변수들을 보여주는 boxplot의 set이다.

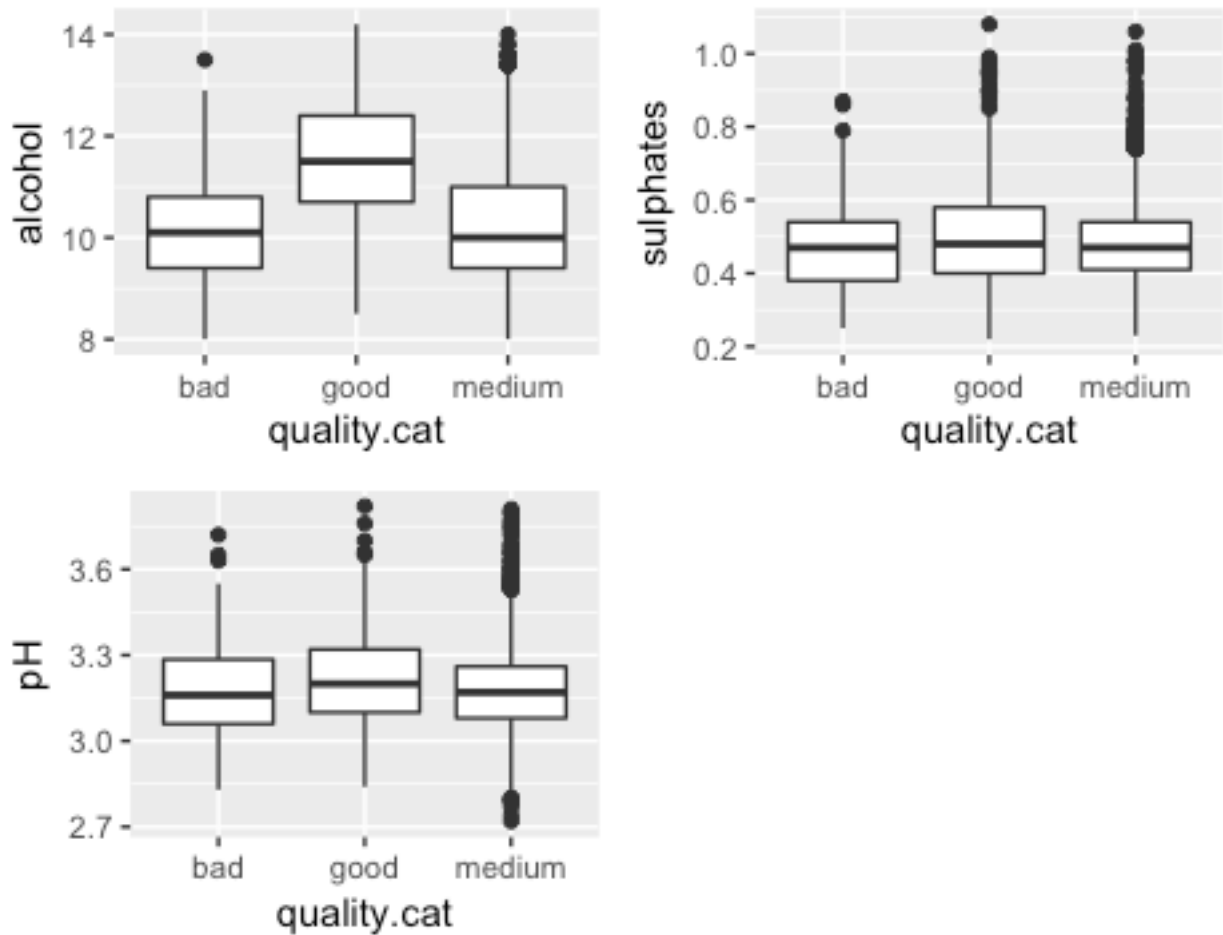
(2) Decreasing quality of red wine

다음은 음의 상관관계를 보이는 변수들이다. volatile.acidity, pH, density는 등급이 올라갈수록 수치가 낮아지는 경향이 있다.



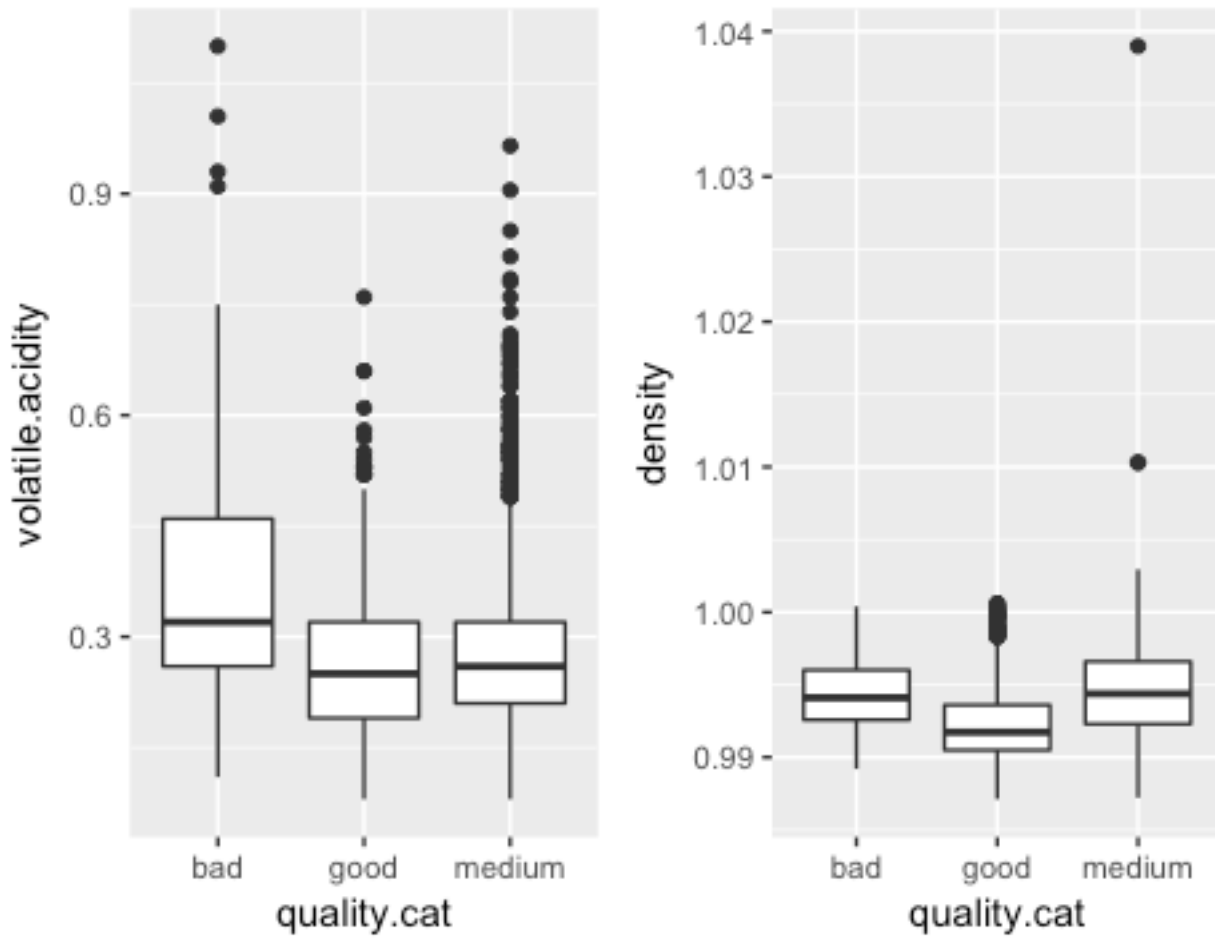
2. white wine

(1) Increasing quality of white wine



alcohol, sulphatates, pH는 white wine quality와 양의 상관관계가 있다고 해석할 수 있다.

(2) Decreasing quality of white wine



volatile acidity는 white wine quality와 음의 상관관계가 있다. density는 good 판정을 받은 와인에서 다른 등급의 와인에 비해 낮은 수치를 보이고 있다.