

Data Visualization_Midterm

20143311 김희수

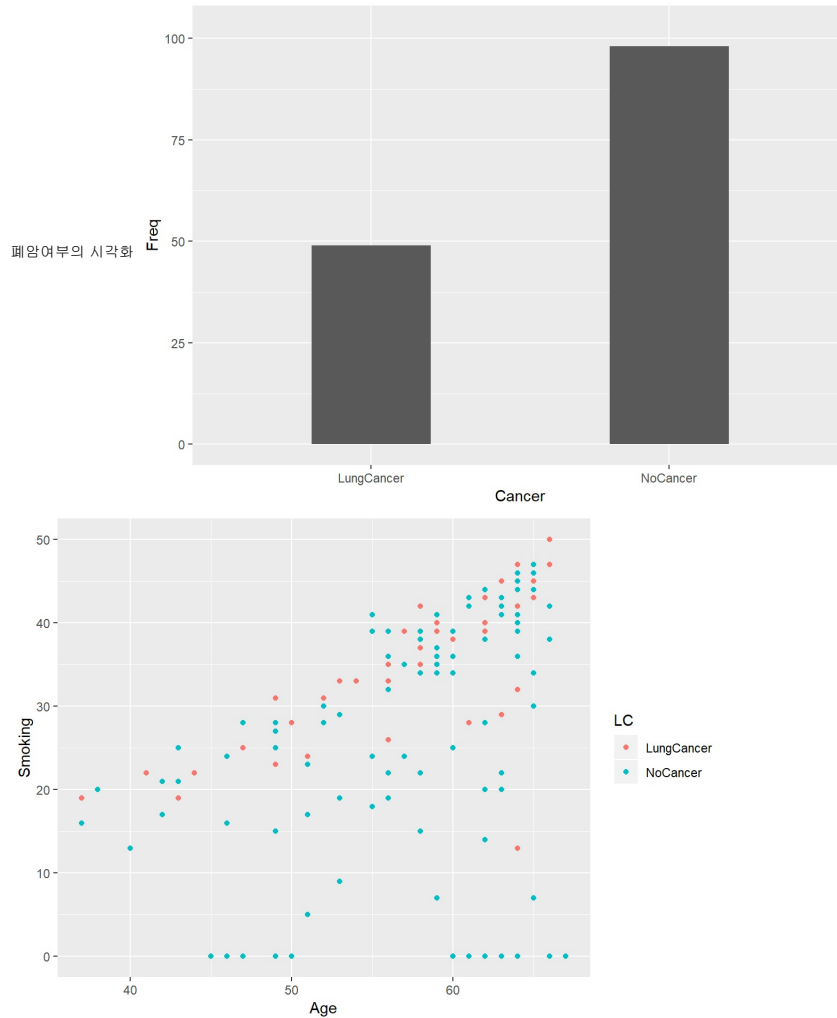
2018-10-21

Read Data File

read.csv를 이용해 lung-cancer.csv 파일을 불러온다.

```
## 'data.frame': 147 obs. of 4 variables:  
## $ LC : chr "LungCancer" "LungCancer" "LungCancer" "LungCancer" ...  
## $ Sex : chr "Male" "Male" "Male" "Male" ...  
## $ Age : int 37 41 43 46 49 51 52 53 56 56 ...  
## $ Smoking: int 19 22 19 24 31 24 31 33 33 26 ...
```

Problem 1



Problem 2

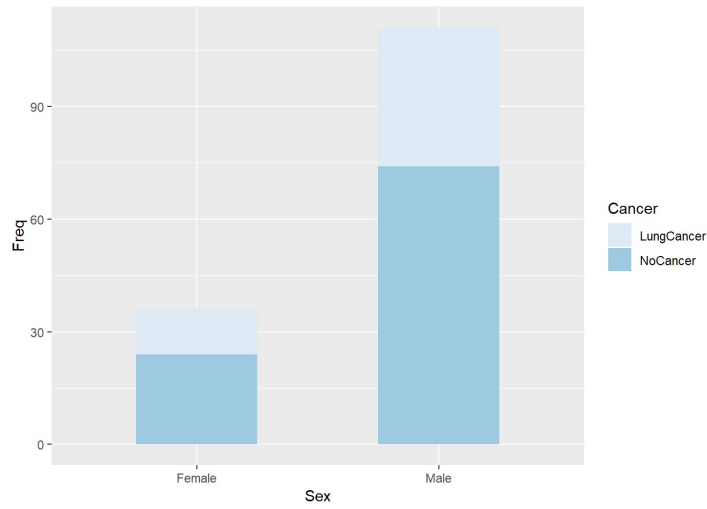
흡연여부와 암발생에 대한 그래프이다.

```
## LungCancer NoCancer  
## Min. : 0.00 Min. : 0.00  
## 1st Qu.:27.00 1st Qu.:16.00  
## Median :36.00 Median :25.00  
## Mean :33.63 Mean :24.96  
## 3rd Qu.:41.00 3rd Qu.:38.75  
## Max. :50.00 Max. :47.00
```

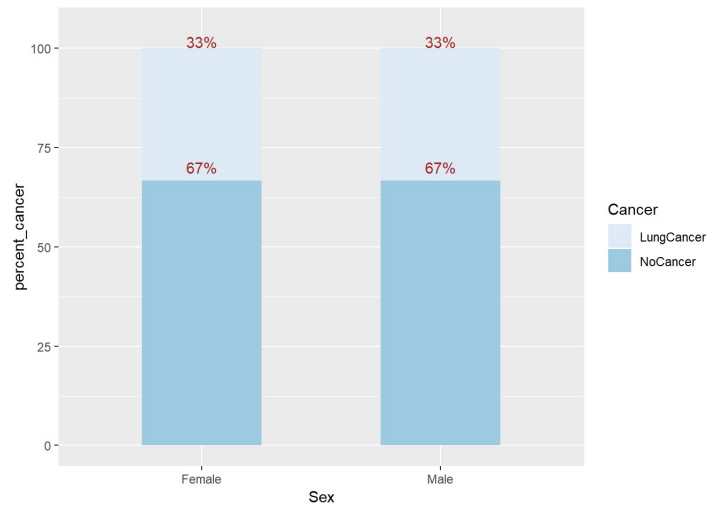


Problem 3

성별과 암발생간 차이를 나타낸 그래프이다.



`ddply("Sex", transform, percent_cancer = Freq / sum(Freq) * 100)` 를 사용해 성별간 암환자 비율을 계산해 퍼센트로 나타낸후 `geom_bar` 로 그린 그래프이다.



`ggplot`안에 `aes(x좌표)`에서 `interaction` 함수를 사용해 성별과 암발생을 나눠 그린 그래프이다.

