

Data Visualization_Midterm

20143311 김희수

2018-10-26

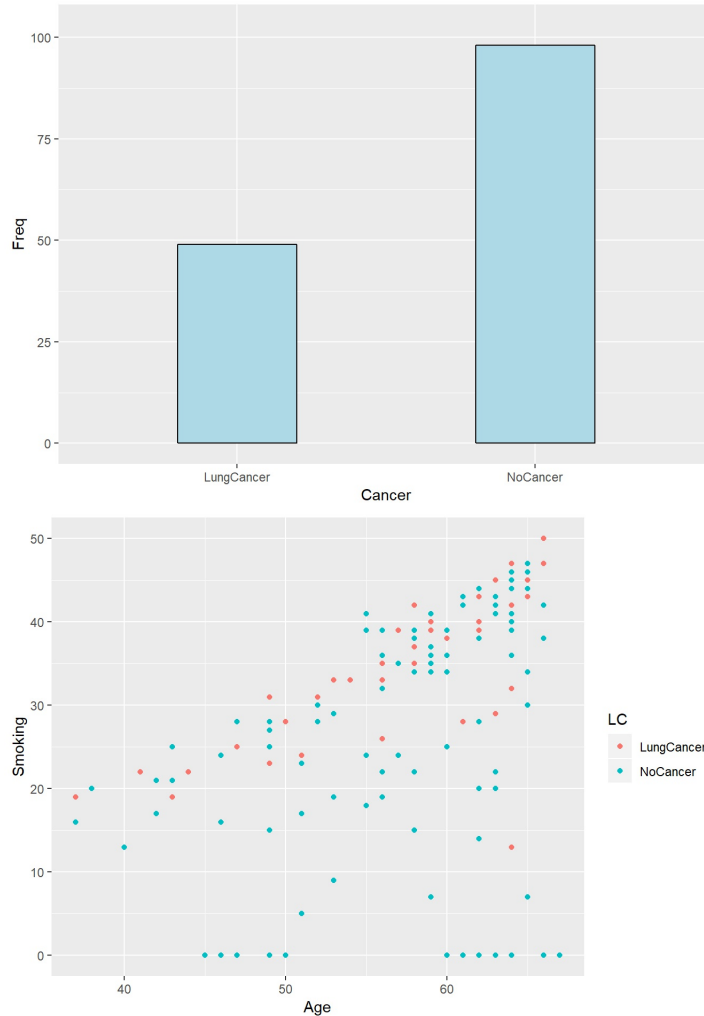
Read Data File

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

폐암여부를 시각화 하시오.



Problem 2

흡연기간과 폐암여부의 관련성을 살펴볼 수 있는 시각화를 시도해보시오.

• LungCancer:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	27	36	33.63	41	50

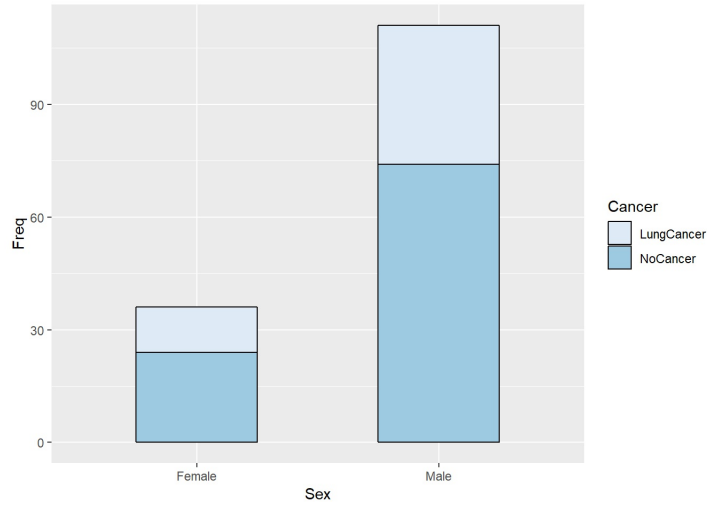
• NoCancer:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	16	25	24.96	38.75	47

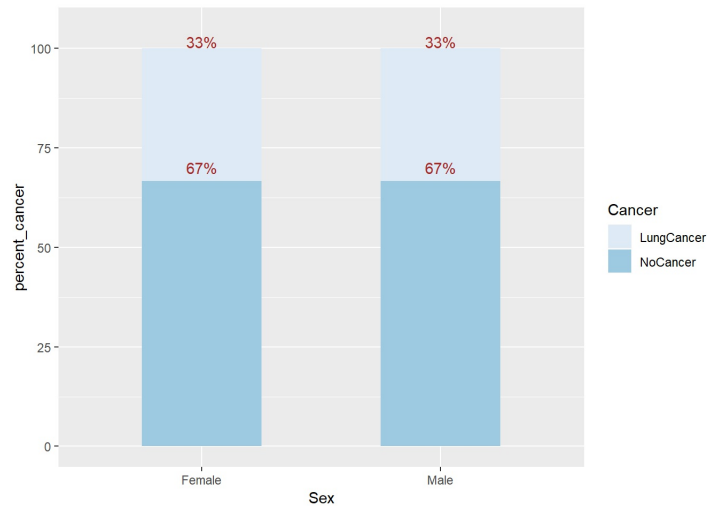


Problem 3

성별과 폐암여부의 관련성을 살펴볼 수 있는 시각화를 시도해보시오.



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



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

