

# lab-07-simpsons.Rmd

Renad Faris Al-Shammari

3 March 2021

## Packages

```
library(tidyverse)
library(mosaicData)
```

## Exercises

1.

```
view(Whickham)
```

Your answer: Observation1 , because we just watching the people and write the notes also we didn't involving in any situation or controlling any variables.

2.

```
nrow(Whickham)
```

```
## [1] 1314
```

Your answer; 1314 , represent recorded participants' age, smoking status at baseline .

3.

```
ncol(Whickham)
```

```
## [1] 3
```

Your answer: 3 outcome  
smoker age

```
class(Whickham$age)
```

```
## [1] "integer"
```

```
class(Whickham$smoker)
```

```
## [1] "factor"
```

```
class(Whickham$outcome)
```

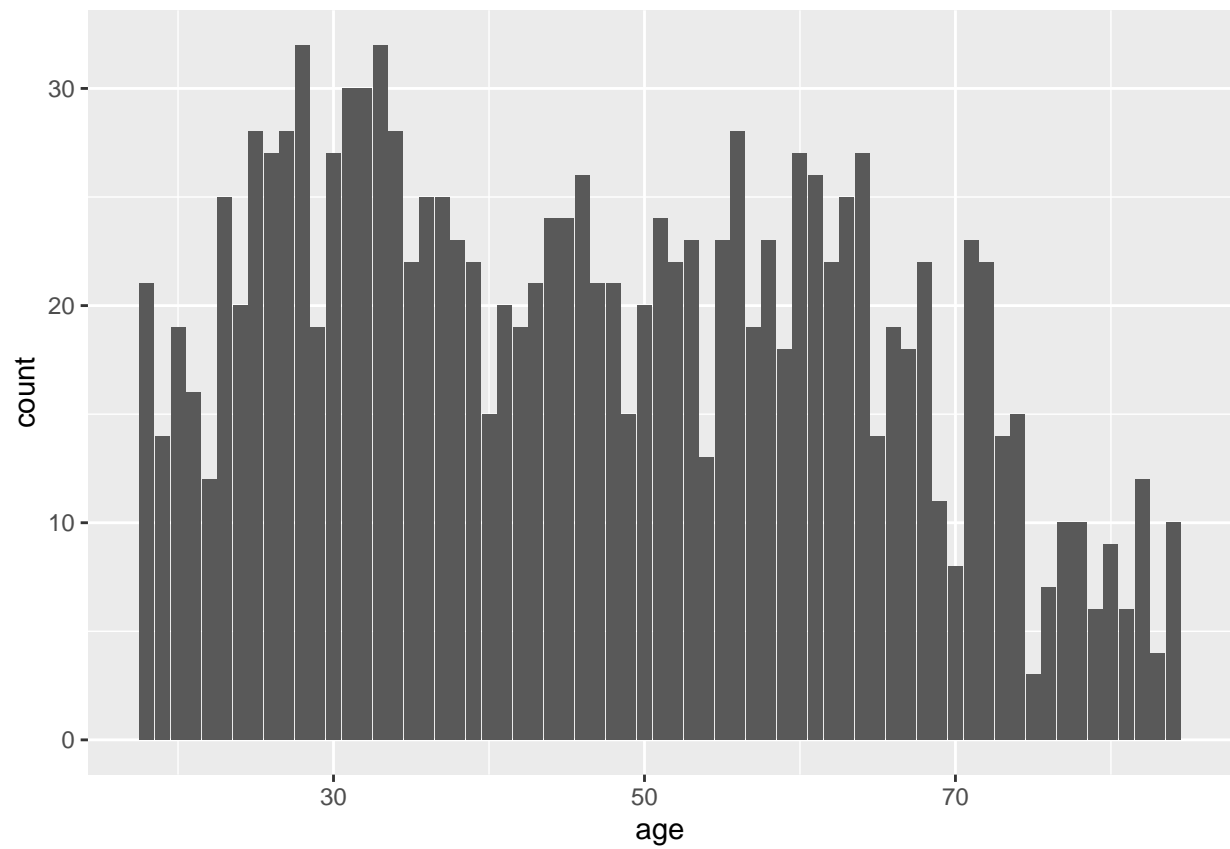
```
## [1] "factor"
```

Your answer: outcome and smoker and age  
outcome and smoker are (Categorical) and are (Factor) age it's (Numerical) and (Integer)

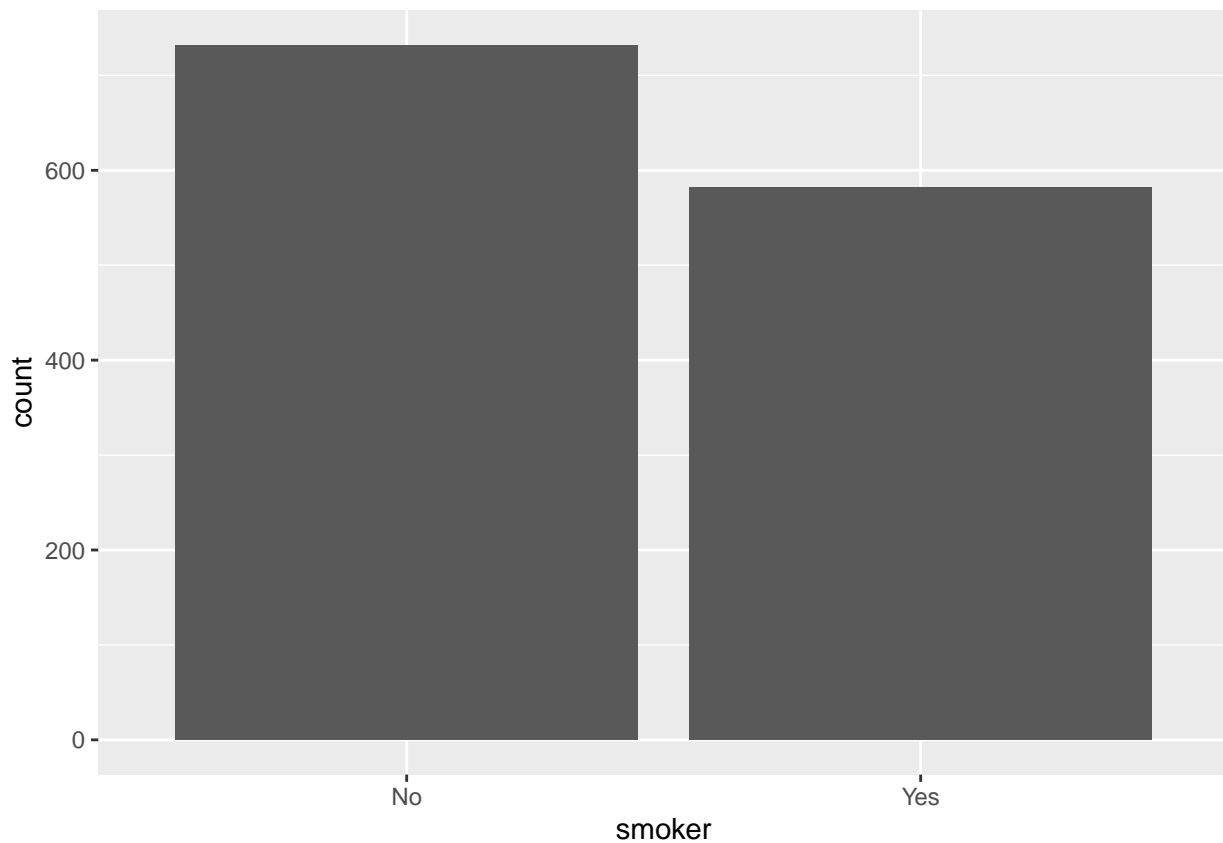
```
ggplot(Whickham, aes(x = outcome)) +  
  geom_bar()
```



```
ggplot(Whickham, aes(x = age)) +  
  geom_bar()
```



```
ggplot(Whickham, aes(x = smoker)) +  
  geom_bar()
```



4. I think his health is not good. it is exposed to many substances harmful to the lungs, which lead to the loss of the lungs ability to filter, as well as toxins become confined to the lung and smokers are more susceptible to respiratory infections. And+ I expect the health will be worser and may be the person will be died after while , if he keeping smoke

```
ggplot(data=Whickham, aes(x=smoker, y=outcome, color=outcome)) + geom_bar(stat="identity")
```



Knit, commit, and push to github.

5.

```
Whickham %>%
  count(smoker, outcome)
```

```
##   smoker outcome    n
## 1     No    Alive 502
## 2     No    Dead 230
## 3    Yes    Alive 443
## 4    Yes    Dead 139
```

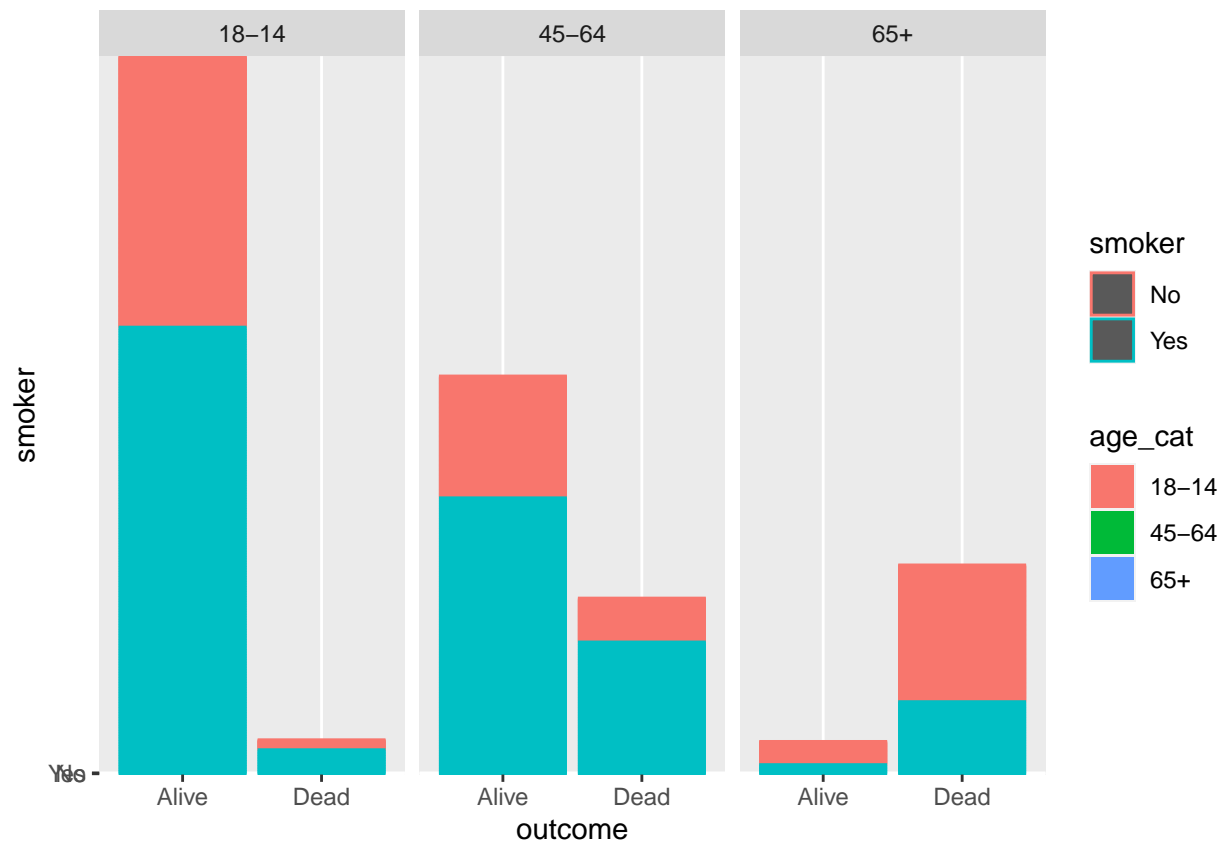
Smoker “732” (NO) = 0.3142077 (Dead) » (68.6) Alive Smoker “582” (YES) = 0.2388316 (Alive) » (76.2) Alive I doesn’t expected this result because now the most died people not smoker.

6.

```
Whickham <- Whickham %>% mutate(age_cat = case_when(age <= 44 ~ "18-14" , age > 44 & age <= 64 ~ "45-64"
```

7.

```
ggplot(data=Whickham, aes(x=outcome, y=smoker,color=smoker, fill=age_cat)) + geom_bar(stat="identity")
```



what changes : The category of the age it's appear to us and we see the most of dead people is smoker in age (65+) but in age (45-64) and (18-14) the most dead people is smoker that is relationship between the smoking and health not clearly but can say that your health will be change to worst if you be smoker and The category of the Alive it's appears in the youth's (18-14) age more than (45-64) and 65+ Also the smoker (YES) it's more Alive in the category age it's (18-14) I believe that the effects increase with age, unlike those who are young and can resist diseases and live more than the age of adults.

Knit, commit, and push to github.