

# lab-07-simpsons.Rmd

Hanan Munthir Aldhafeeri

17 March 2021

## Packages

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

## Exercises

1.

```
?Whickham
```

Your answer: observational , Because they notice a person's health status after a period of time

2.

```
nrow(Whickham)
```

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

Your answer; obs 1314

Each row represents whether the person is alive or dead, is he a smoker or not, and how old he is

3.

```
names(Whickham)
```

```
## [1] "outcome" "smoker" "age"
```

Your answer: 3 var , “outcome” “smoker” “age”

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

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

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

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

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

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

Your answer: 3 var , “outcome” “smoker” “age”

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



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



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



4. I expect that the health will be worse after a period of time and that it may cause the person to die if he continues to 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 no (732) : 31,4 (dead) » (68,6) alive smoker yes (582) : 23,8 (dead) » (76,2) alive

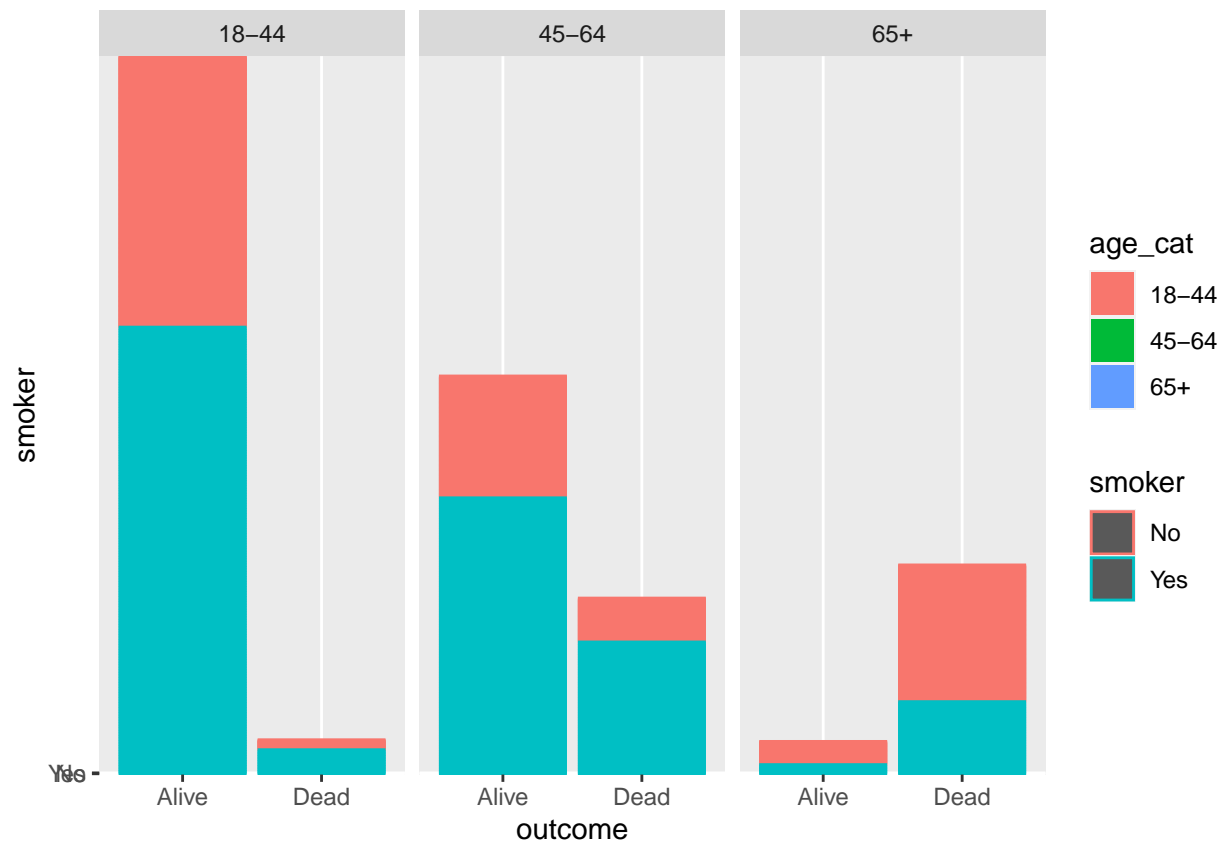
does not expected this result because now the most died people not smoker

6.

```
Whickham <- Whickham%>% mutate (age_cat = case_when (age <= 44 ~ "18-44", age > 44. & age <= 64 ~ "45-64", age > 64 ~ "65-99"))
```

7.

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



Knit, commit, and push to github.