

lab-07-simpsons.Rmd

Walaa Ali

3 March 2021

Packages

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

Exercises

1.

```
?Whickham
```

Your answer: The data is observational as the description states that is based on age, smoking, and mortality, which are all observable events and not produced via experiments.

2.

```
nrow(Whickham)
```

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

Your answer; obs 1314 var 3 Each row represents age, whether or not it is a smoker, and is it alive or not

3.

```
names(Whickham)
```

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

Your answer: there are 3 variables, “Outcome”, “smoker”, and “age” outcome (factor) age(integer) smoker(factor)

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

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

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

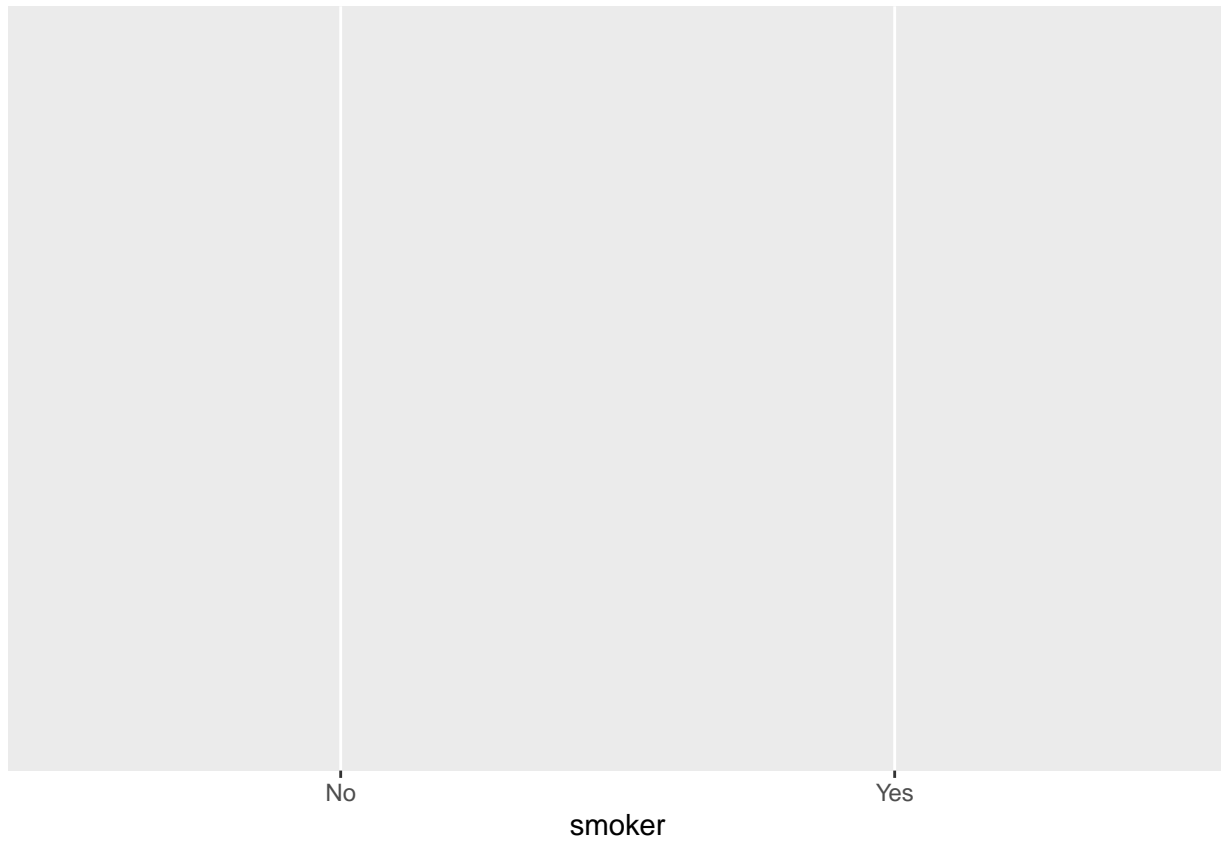
```
## [1] "integer"
```

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

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

Your answer: I expect that health will deteriorate with the passage of time, too, the incidence of cancer, and if he continues to smoke, he may lead to death.

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



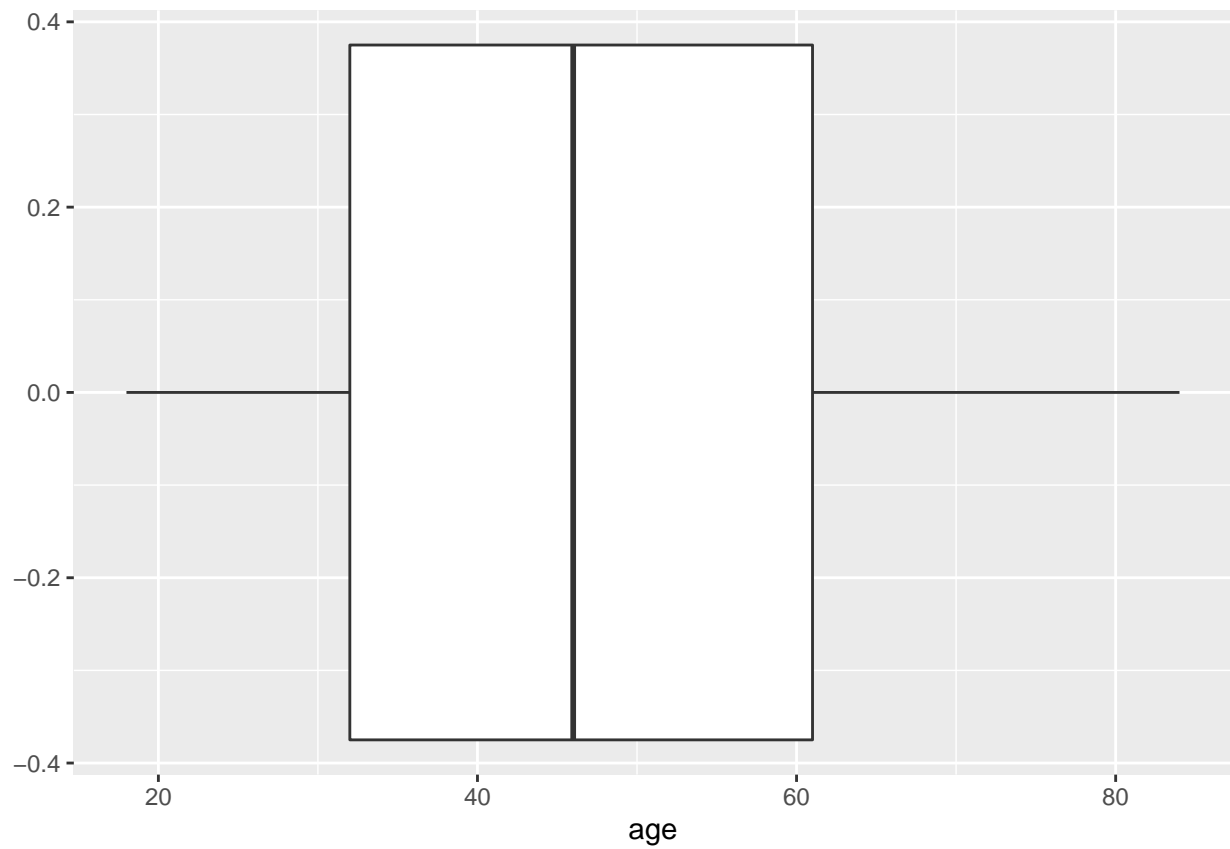
```
geom_bar()
```

```
## geom_bar: width = NULL, na.rm = FALSE, orientation = NA  
## stat_count: width = NULL, na.rm = FALSE, orientation = NA  
## position_stack
```

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

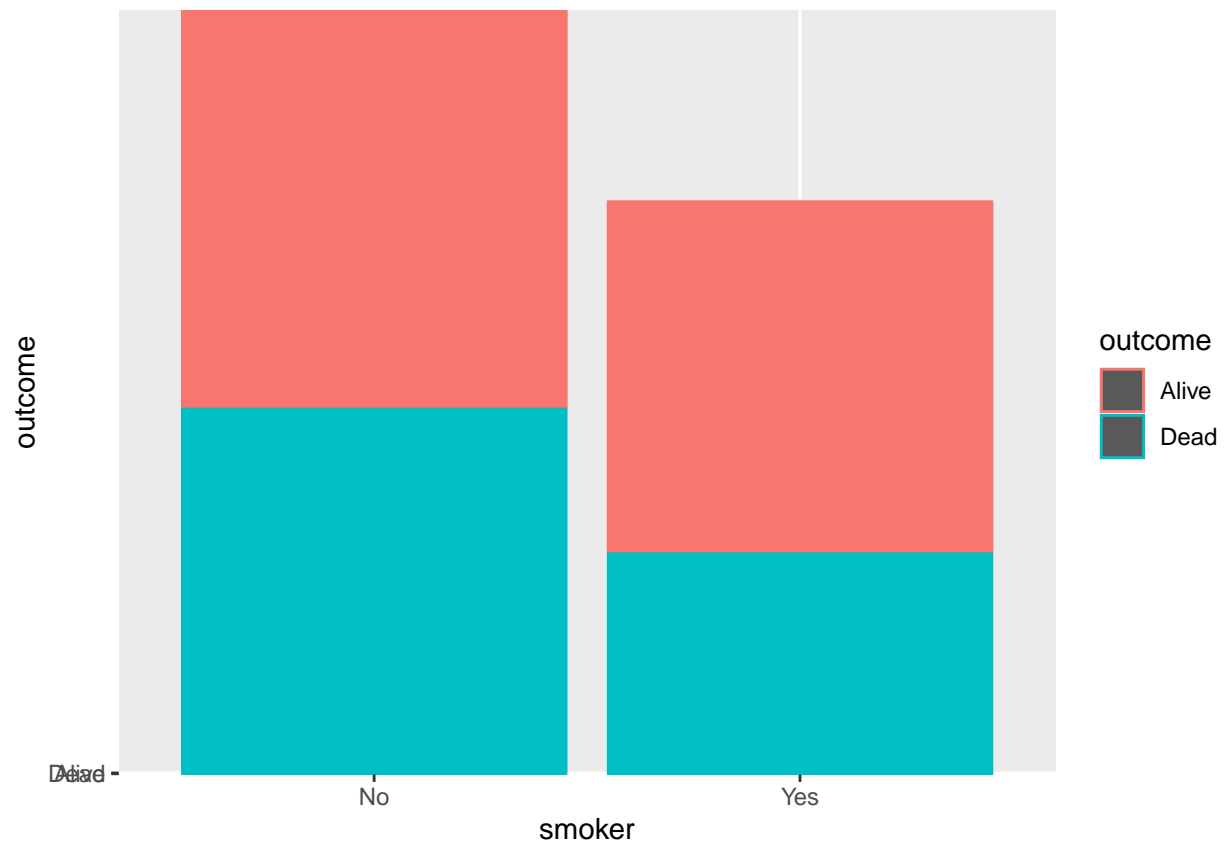


```
ggplot(Whickham, aes(x = age)) +  
  geom_boxplot()
```



4.

```
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
```

6.

7.

Knit, commit, and push to github.