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, “0utcome”,“smoker”, and “age” outcome (factor) age(integer) smoker(factor)

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

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

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

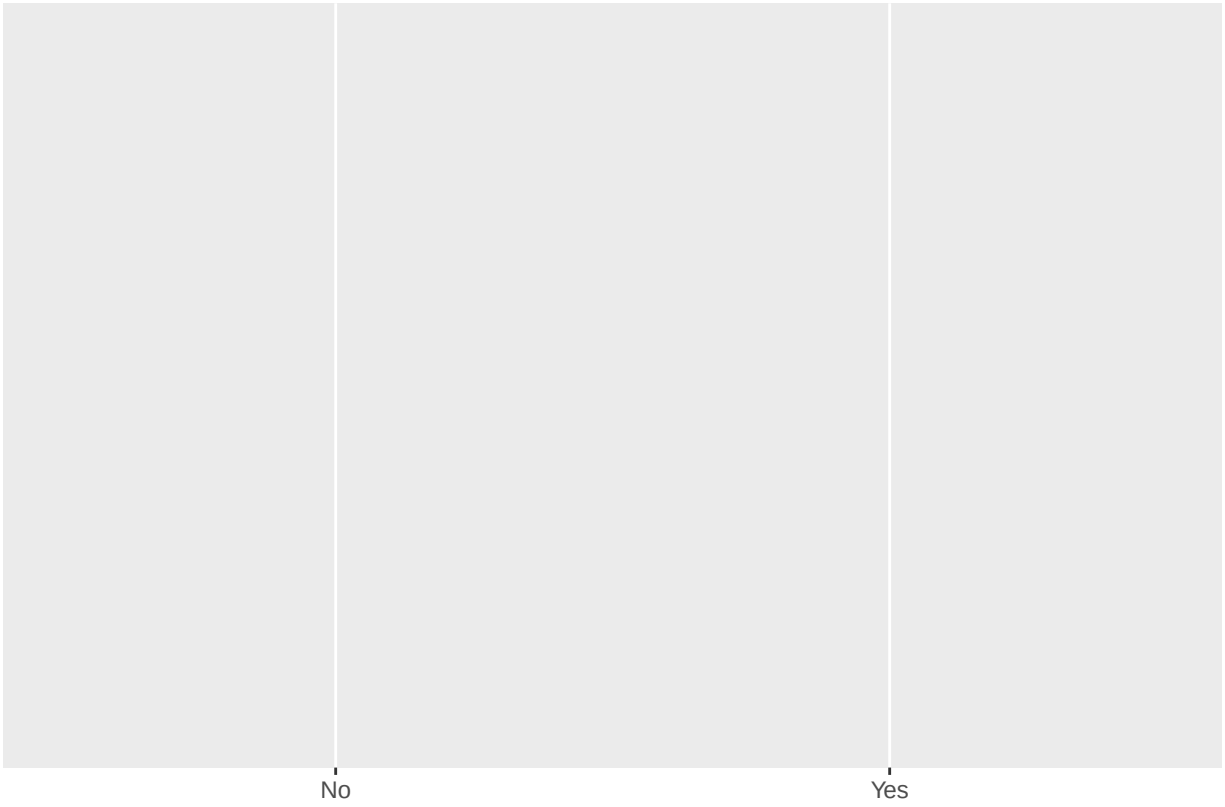
```
## [1] "integer"
```

```
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.

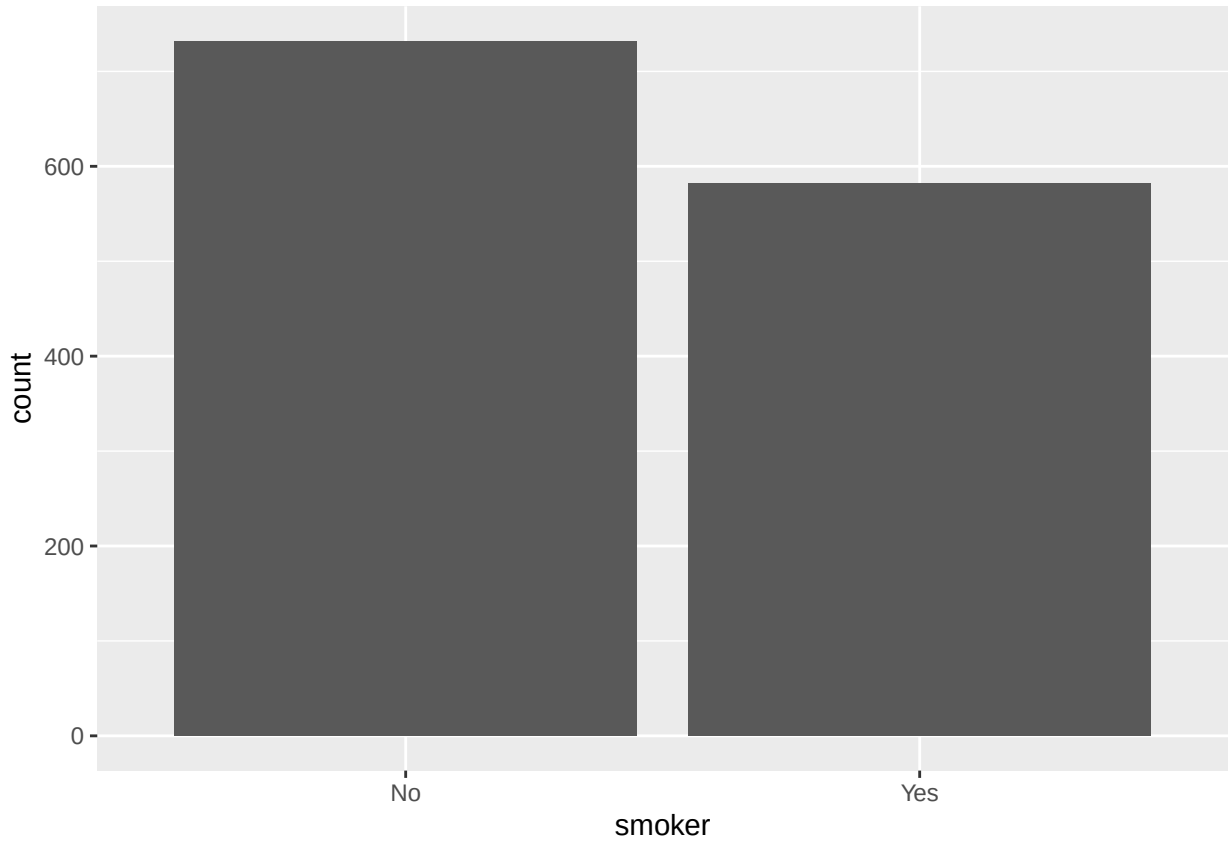
```
ggplot(Whickham, aes(x = smoker))
```



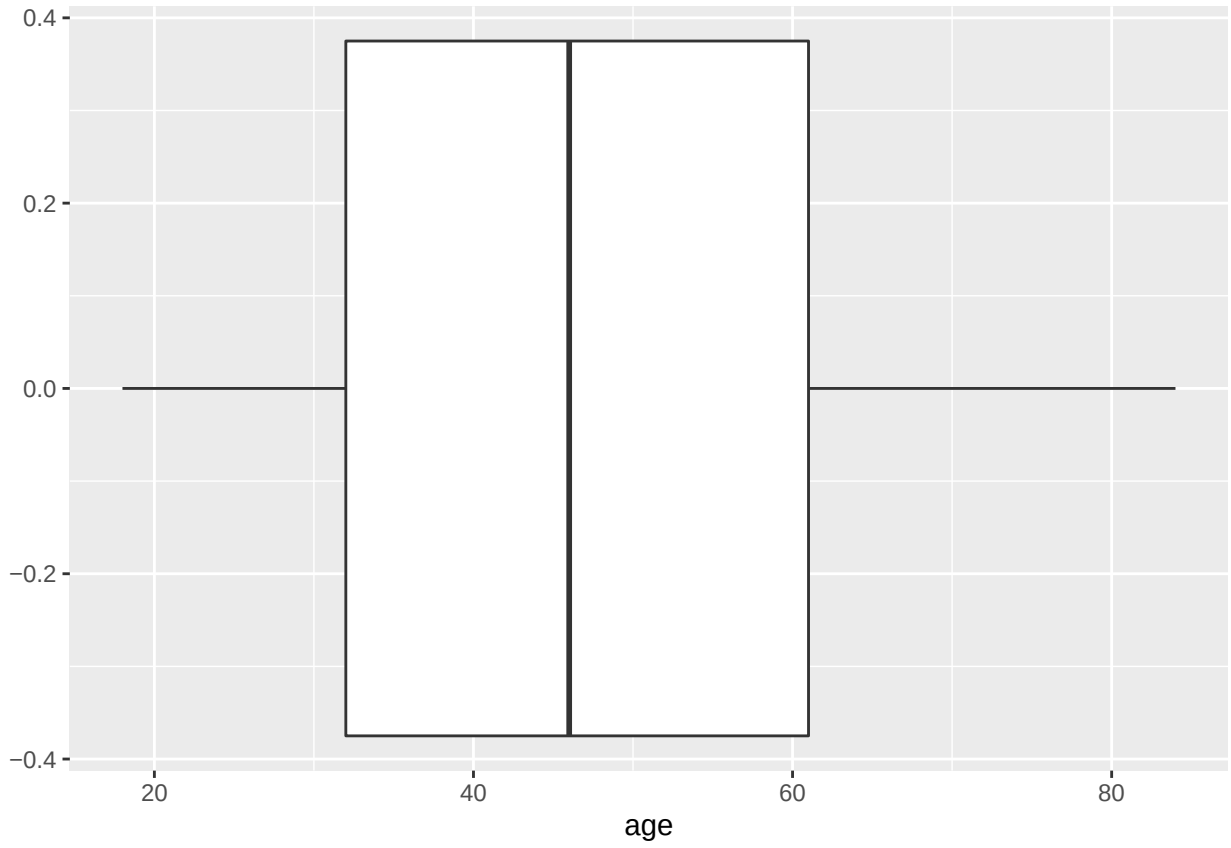

```
geom_bar()
```

```
## 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()
```

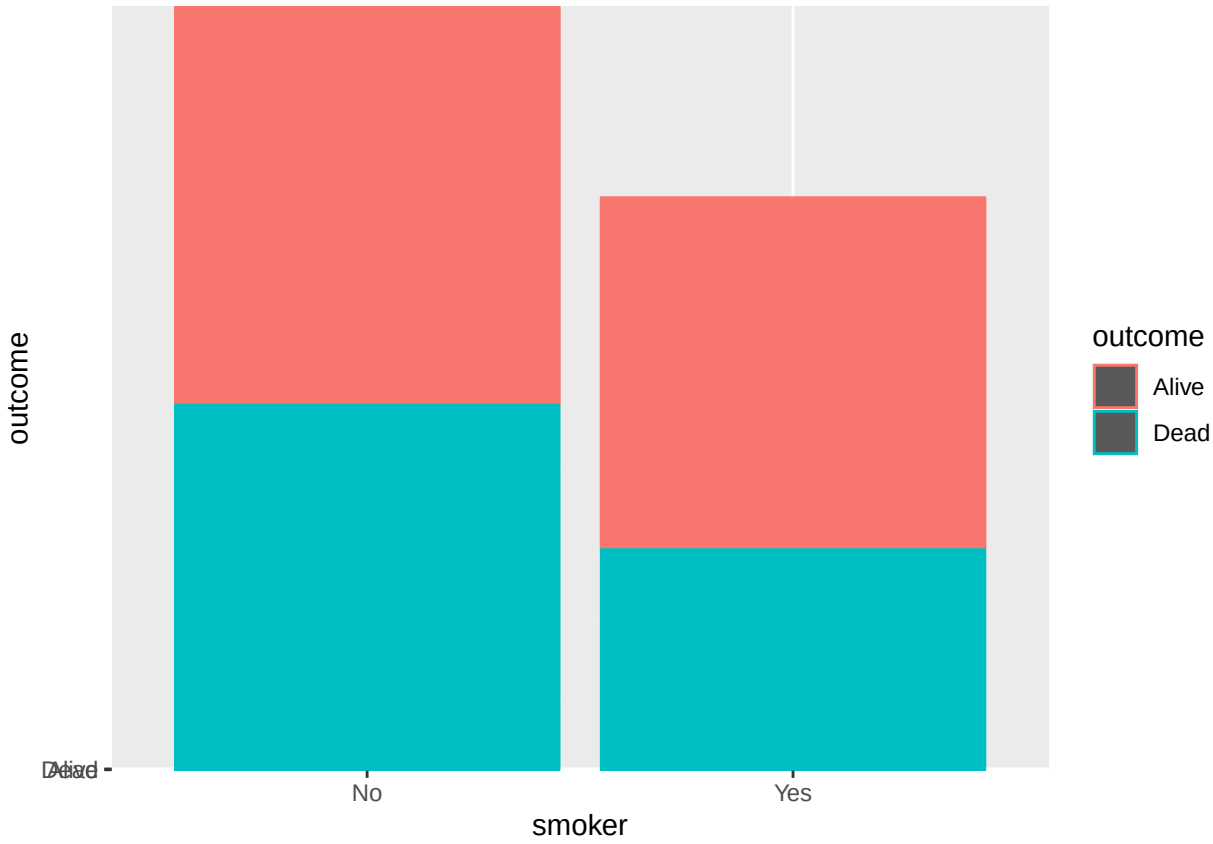



```
ggplot(Whickham, aes(x = age)) +
  geom_boxplot()
```



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
```

```
502+230
```

```
## [1] 732
```

```
230/732
```

```
## [1] 0.3142077
```

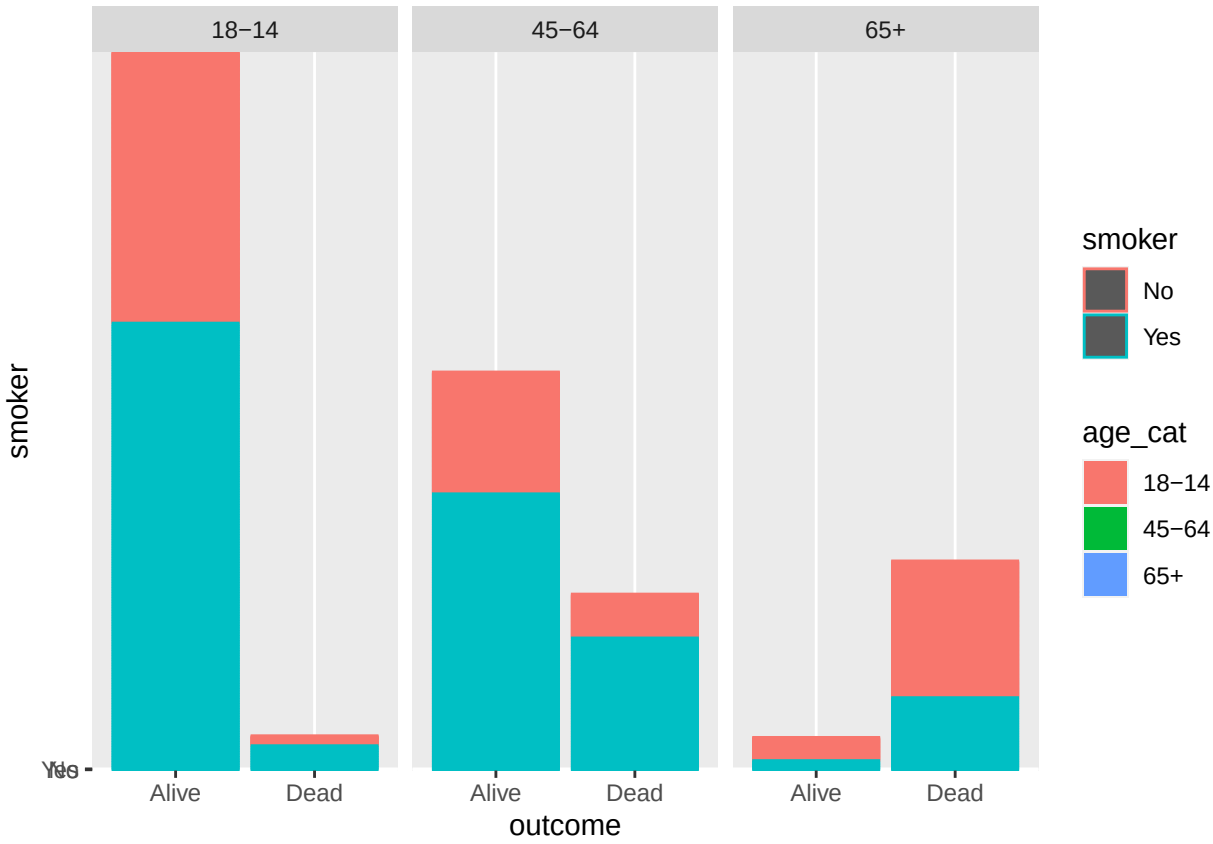
smoker (732) N0 — 31,4 (Dead) »(68.6) Alive smoker (582) Yes — 23,8 (Dead) »(76.2) Alive

i does not expected this result because now the most died pepole 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") +
```



The percentage of live nonsmokers between 18-64 is higher than that of non-smokers, unlike the 65+ category, the percentage of live nonsmokers is higher than the percentage of smokers.

what changes > the category of the age it's appear to us and we see the most of dead pepole not smoker in age (65+) . . . but in age (45-64) and (18-44) the most dead pepole are smoker that is relationship between the smoking and helath not clearly but can say that your helath will be change to worst if you be smoker .

Knit, commit, and push to github.