lab-07-simpsons.Rmd

rana alshammari/2201000999

17 March 2021

Packages

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

Exercises

1.

```
?Whickham
```

Your answer: he 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; the are 1314 observations. as we knaw evrey row observation

3.

```
names(Whickham)
```

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

Your answer:
obs:1314.Represent outcome , smoker and age

```
unique(Whickham$outcome)
```

```
## [1] Alive Dead
## Levels: Alive Dead
```

```
unique(Whickham$smoker)
```

```
## [1] Yes No
## Levels: No Yes
```

```
unique(Whickham$ega)
```

```
## NULL
```

Your answer: Using the unique() function on the 3 variables we could see that "outcome" only Dead value, which makes it categorical non-ordinal. "smoker" only takes Yes or No, makes it categorical non-ordinal. Age is numerical continous data
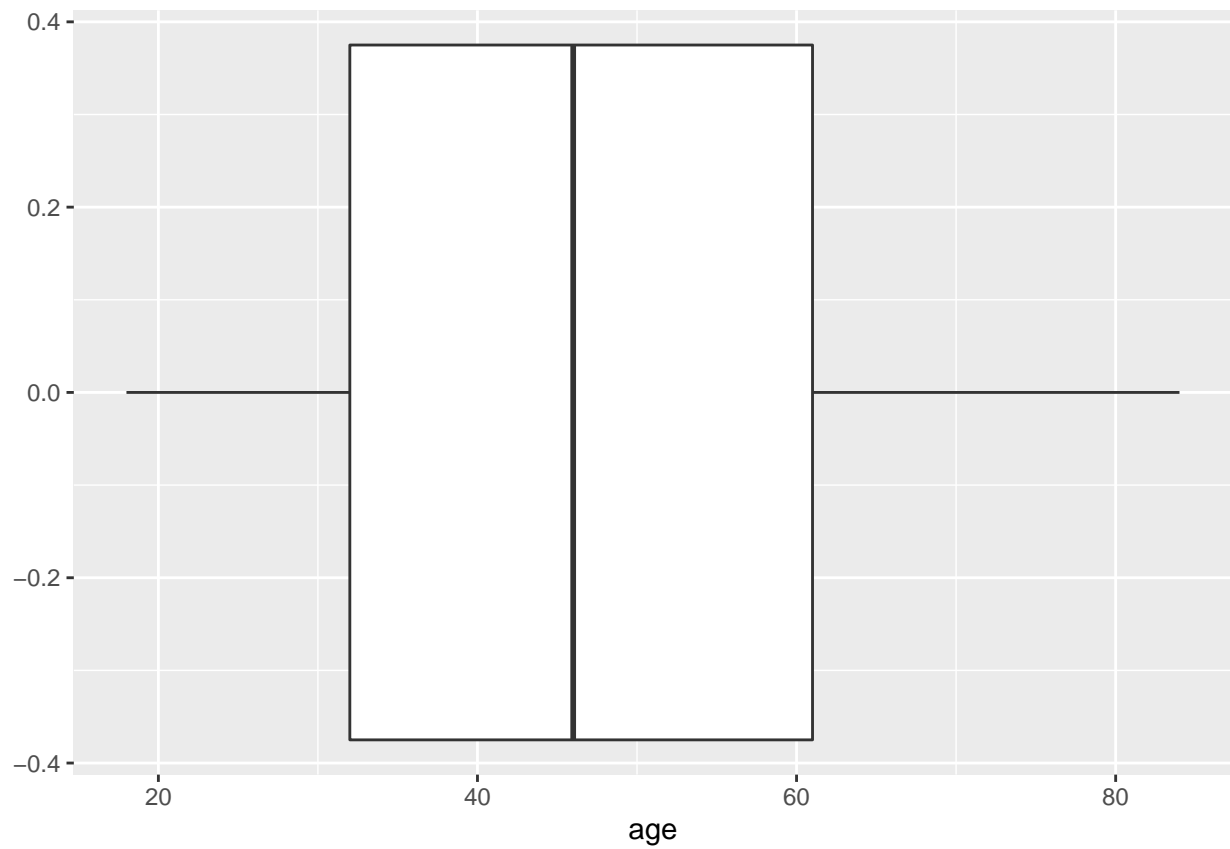
```
ggplot(Whickham, aes(x = outcome)) +
geom_bar()
```

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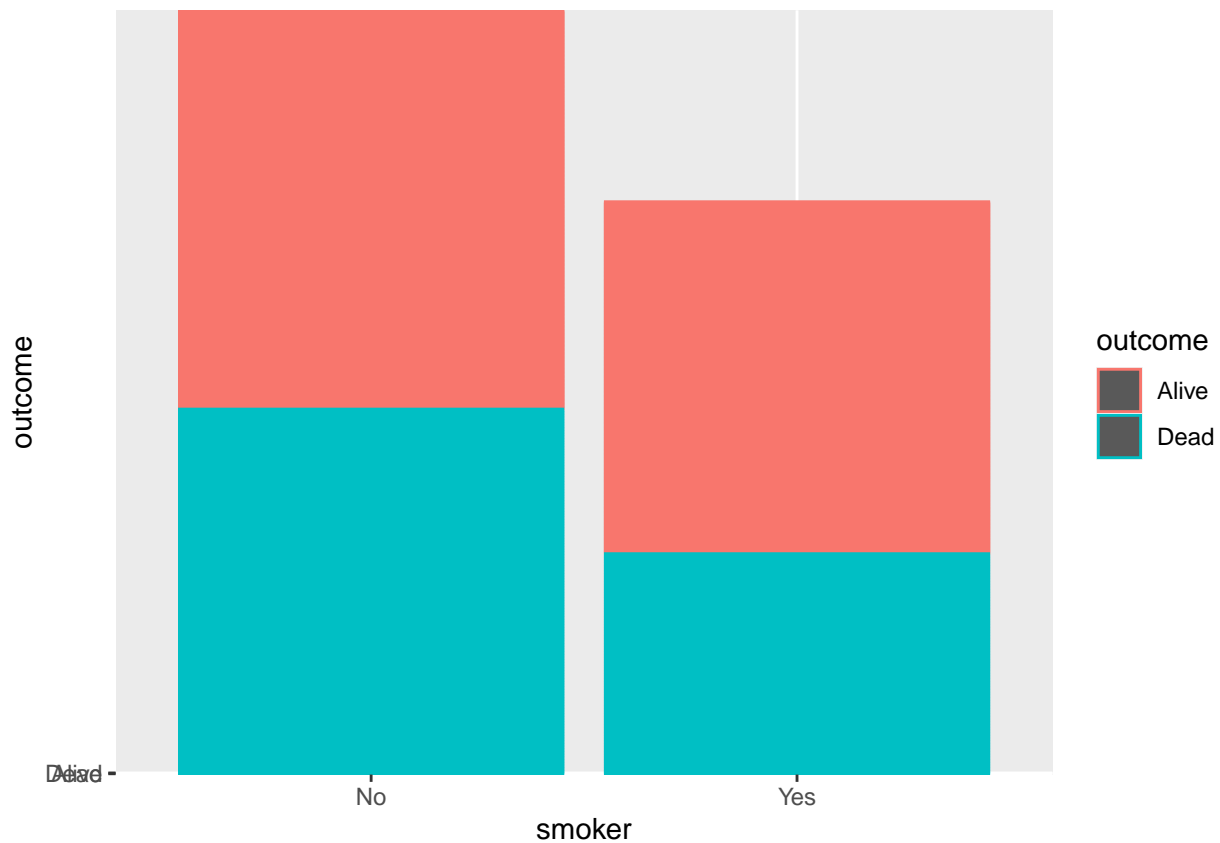
```
ggplot (Whickham, aes(x = age)) +geom_boxplot()
```

4.

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

Smokers if they continue to smoke harm them and they die 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
```
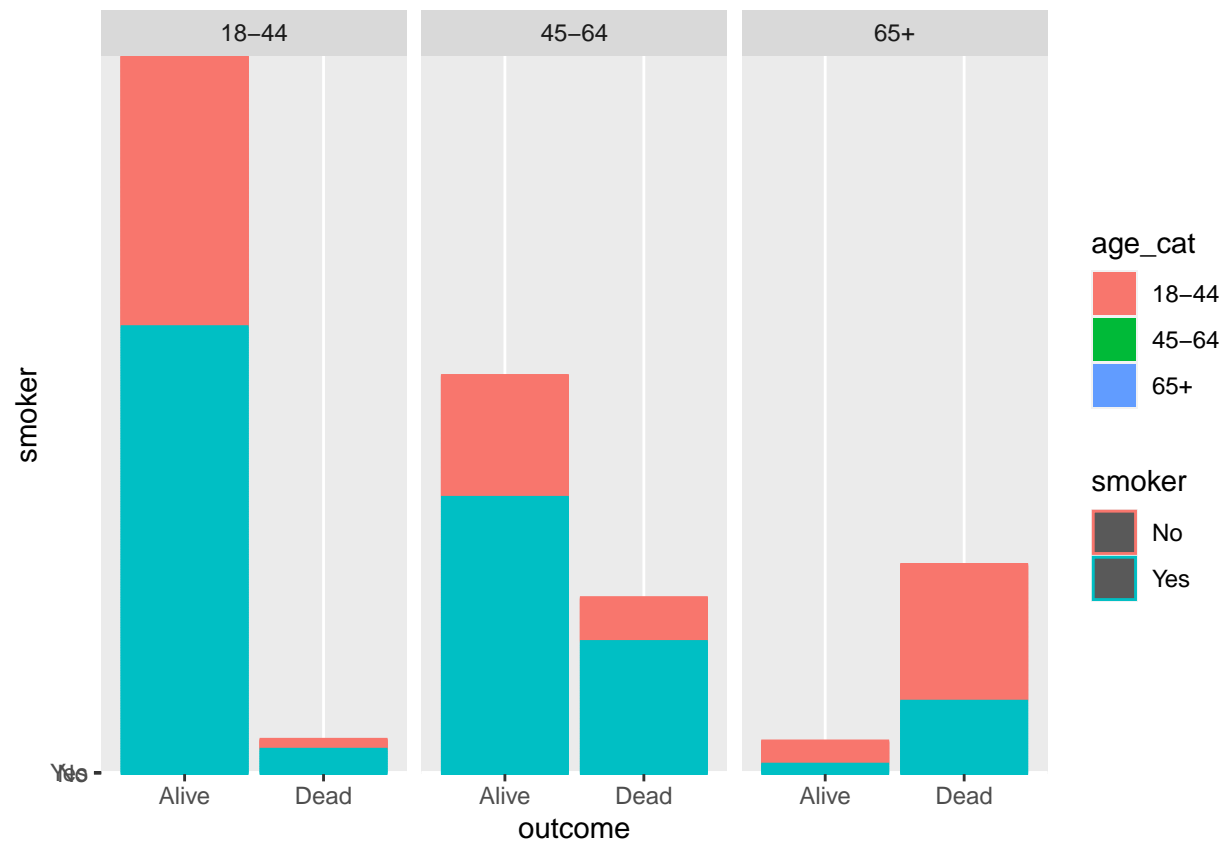
```
230/732
```

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

Smokers die more 6.

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

7.

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


Knit, commit, and push to github.