

lab-07-simpsons.Rmd

Fatimah-Albrahim

17 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; there are 1314 observations . as we know every row is an observation

3.

```
names(Whickham)
```

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

Your answer:

there are 3, “outcome”, “smoker”, “age”

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

```
## [1] Alive Dead
```

```
## Levels: Alive Dead
```

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

```
## [1] Yes No
```

```
## Levels: No Yes
```

```
unique(Whickham$age)
```

```
## [1] 23 18 71 67 64 38 45 76 28 27 34 20 72 48 66 30 33 68 61 43 47 22 39 80 59
```

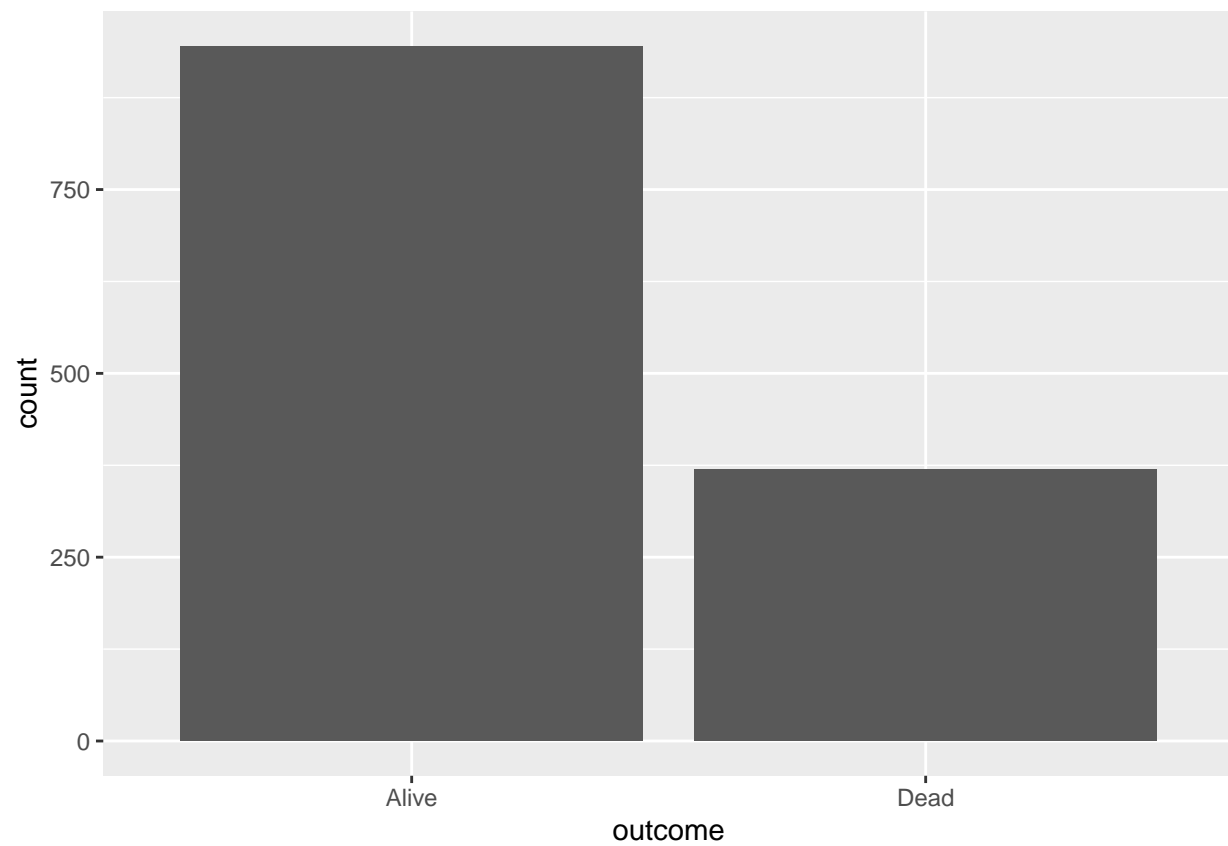
```
## [26] 56 62 51 32 60 37 36 50 55 73 52 25 53 31 54 69 79 75 21 29 24 26 49 84 40
```

```
## [51] 44 74 46 35 77 57 42 81 19 63 78 83 82 70 58 41 65
```

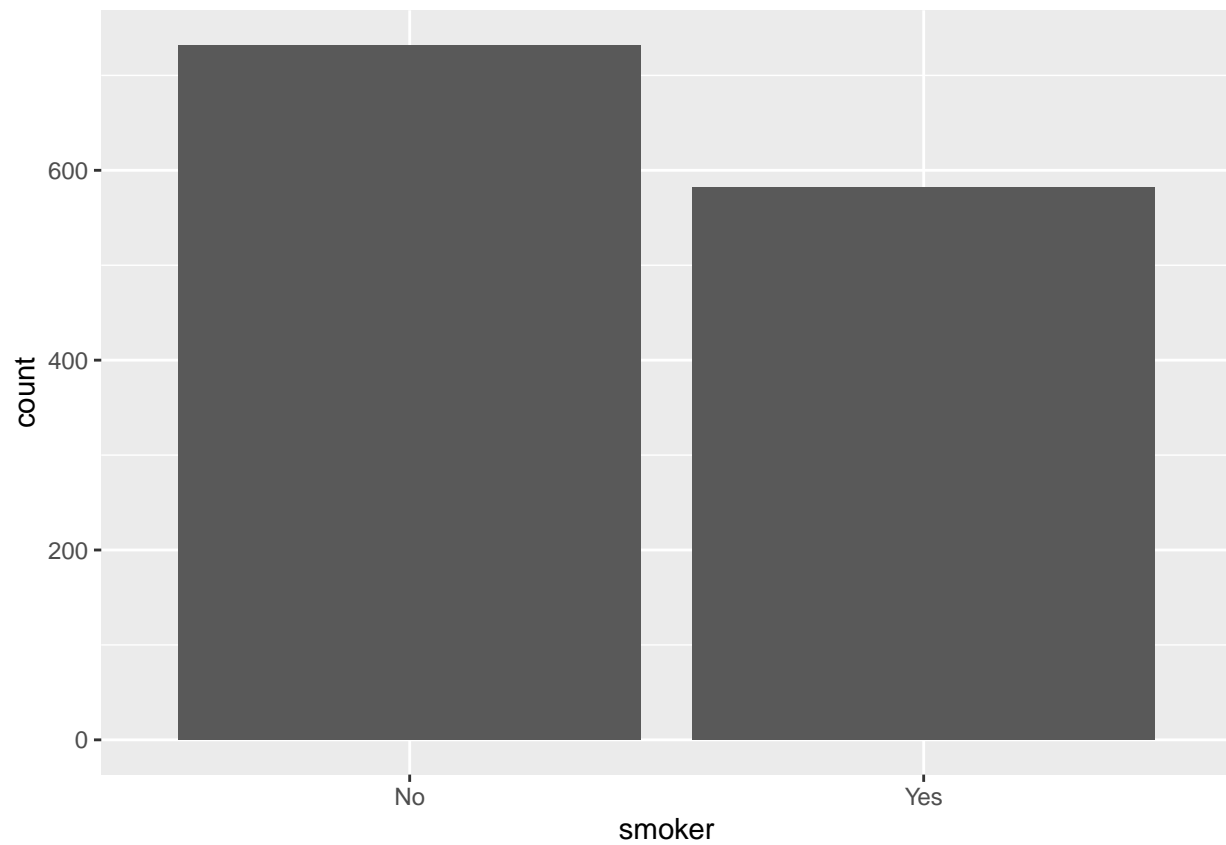
Your answer: using the ‘unique()’ funtion on the 3 variables we could see that “outcome” only takes Alive or Dead value, which makes it categorical non-ordinal. “smoker” only takes Yes or No, which also makes it

categorical non-ordinal. Age is numerical continuous data.

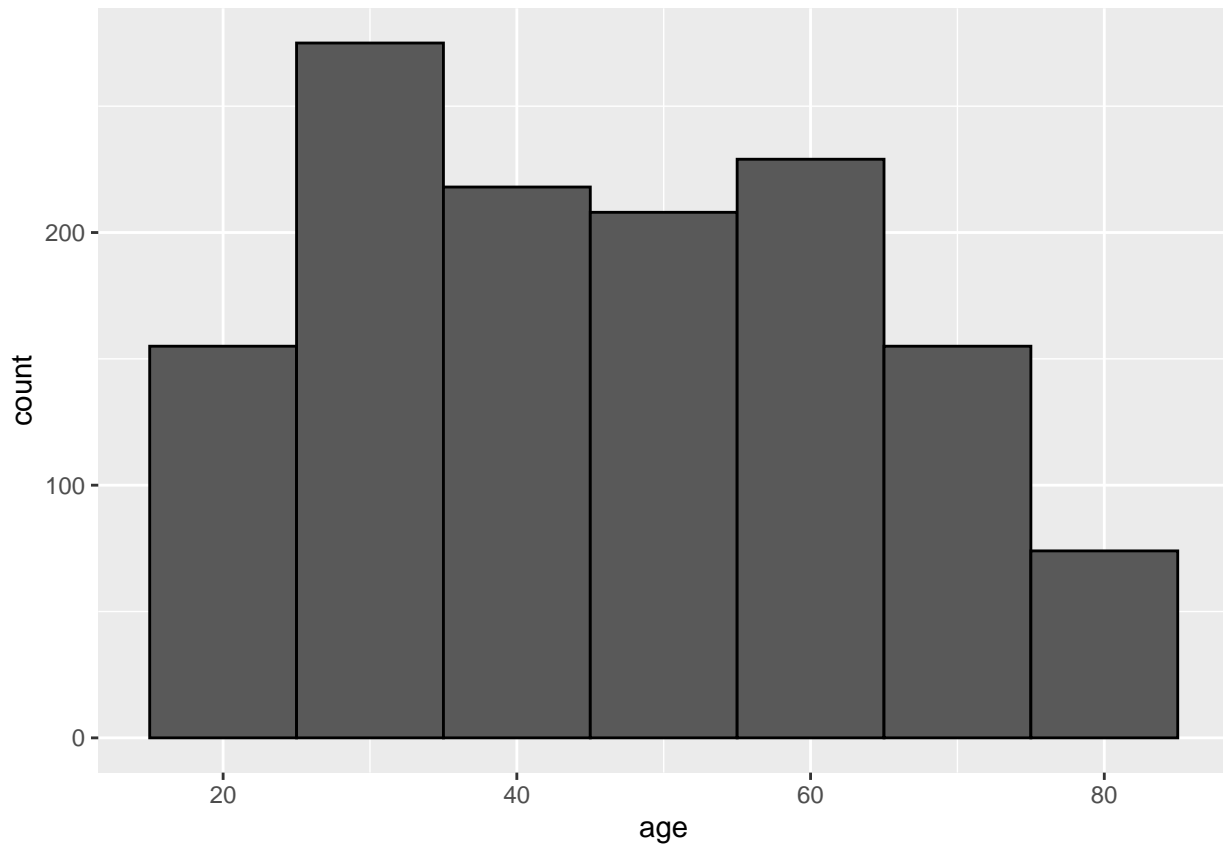
```
ggplot(Whickham, aes(x = outcome)) +  
  geom_bar()
```



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

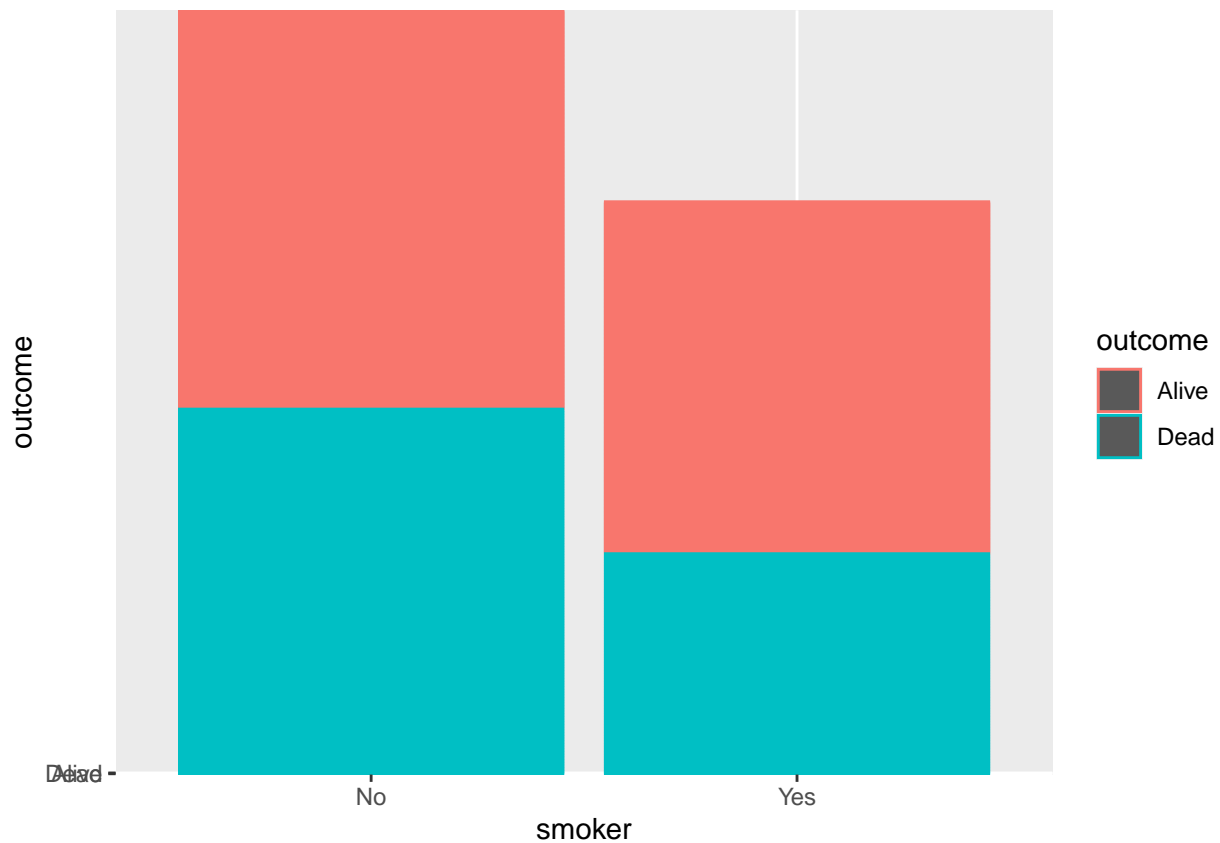


```
ggplot(Whickham, aes(x = age )) +  
  geom_histogram(binwidth = 10 , color = "black")
```



4. 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")
```



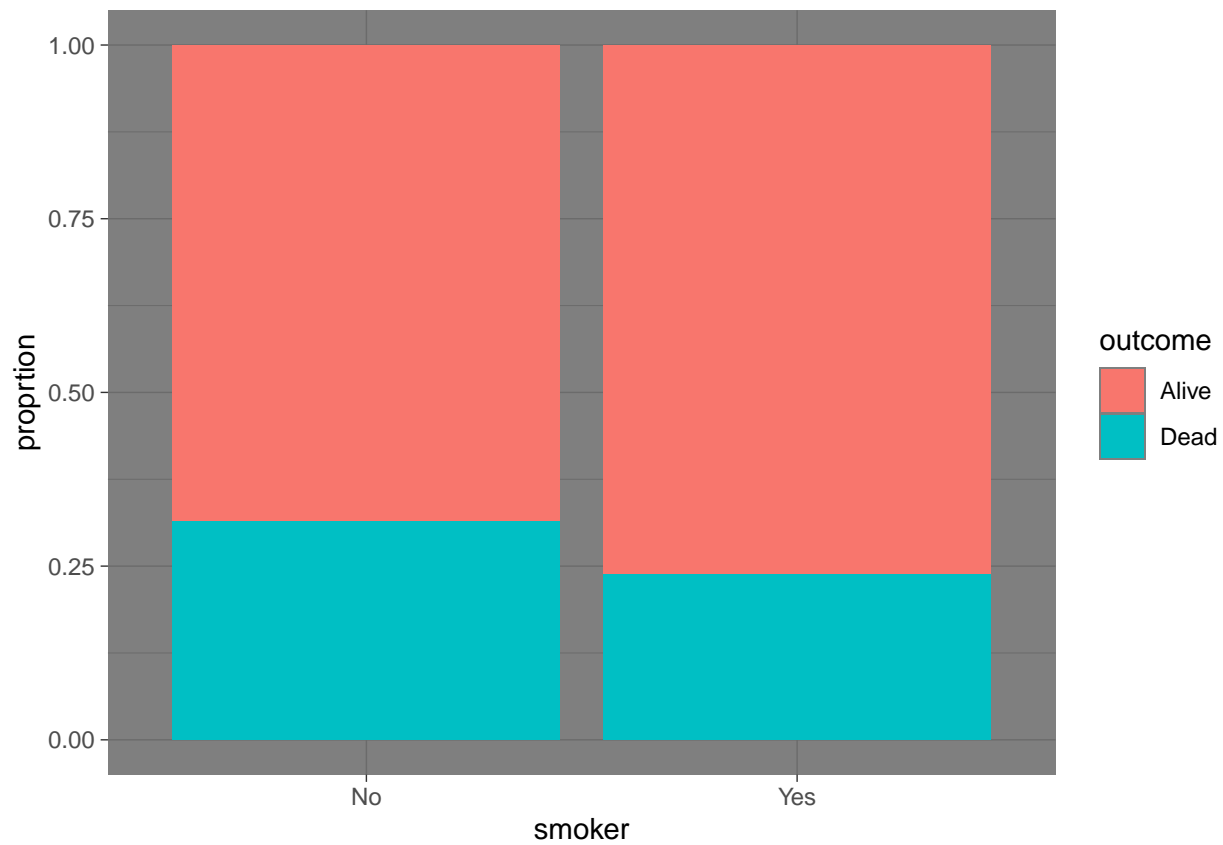
Knit, commit, and push to github.

5.

```
Whickham %>%
  count(smoker, outcome) %>%
  group_by(smoker) %>%
  mutate(prop_outcome = n / sum(n)) %>%
  filter(outcome=="dead")

## # A tibble: 0 x 4
## # Groups:   smoker [0]
## # ... with 4 variables: smoker <fct>, outcome <fct>, n <int>,
## #   prop_outcome <dbl>

ggplot(Whickham, aes(x=smoker, fill=outcome)) +
  geom_bar(position = "fill") + labs(y="proprtion") + theme_dark()
```

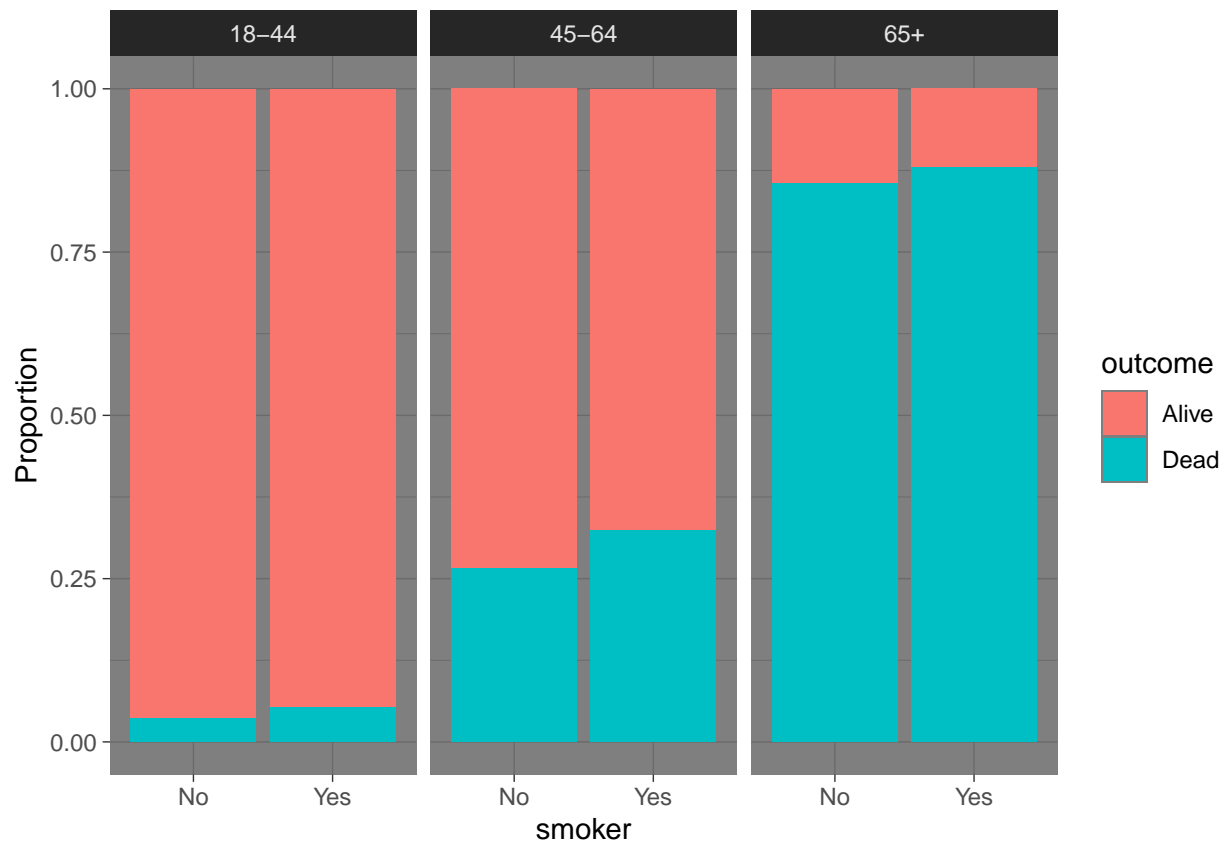


6.

```
Whickham <- Whickham%>% mutate (age_cat = case_when (age <= 44 ~ "18-44", age > 44. & age <= 64 ~ "45-64", age > 64 ~ "65-74"))
```

7.

```
ggplot(Whickham, aes(x=smoker, fill=outcome)) + geom_bar(position = "fill") +  
  labs(y="Proportion") + facet_grid(. ~ age_cat) + theme_dark()
```



Knit, commit, and push to github.