# lab-07-simpsons.Rmd

### Maram Nawaf

#### 17 March 2021

## **Packages**

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

#### **Exercises**

1.

Your answer: Observational, because we just watching the people and write the notes also we didn't involving in any situation or controlling any variables. 2.

```
nrow(Whickham)
```

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

Your answer; 1314, represent recorded participants' age, smoking status at baseline 3.

```
names (Whickham)
```

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

Your answer:

3, age (Numerical), smoker and outcome are categorical

unique(Whickham\$outcome)

```
## [1] Alive Dead
## Levels: Alive Dead
unique(Whickham$smoker)
```

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

unique(Whickham\$age)

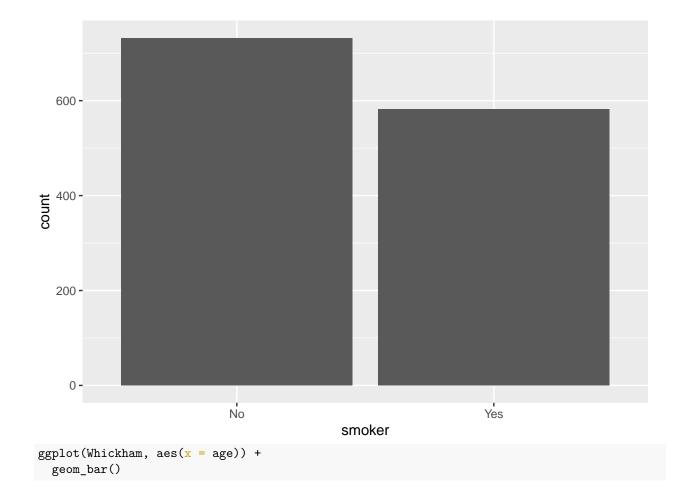
## Levels: No Yes

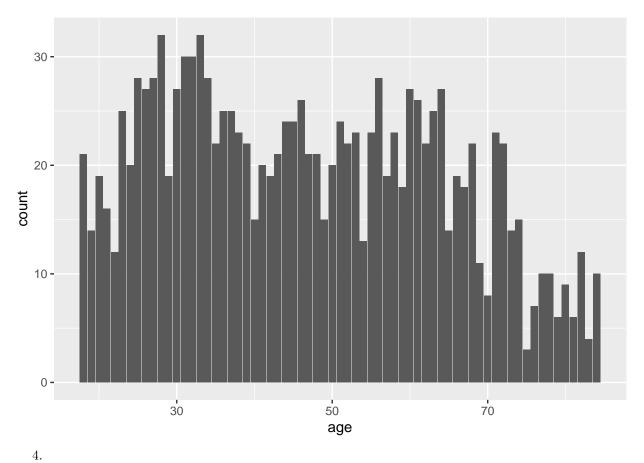
```
## [1] 23 18 71 67 64 38 45 76 28 27 34 20 72 48 66 30 33 68 61 43 47 22 39 80 59 ## [26] 56 62 51 32 60 37 36 50 55 73 52 25 53 31 54 69 79 75 21 29 24 26 49 84 40 ## [51] 44 74 46 35 77 57 42 81 19 63 78 83 82 70 58 41 65
```

Your answer: using the 'unique()' function on the 3 variables we cloud 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 continous data. continous one of the best ways to visualise categorical data is through the use ofbar charts.

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







Knit, commit, and push to github.

5.

# Whickham %>% count(smoker, outcome)

## smoker outcome ## 1 No Alive 502 ## 2 No Dead 230 ## 3 Yes Alive 443 ## 4 Yes Dead 139 6. 7.

Knit, commit, and push to github.