Causality

Claire Liow

University of Tokyo

5/26/2022

Table of Contents

Load packages and data

```
## load packages
library(tidyverse)
library(qss)

## load data
resume <- read_csv("data/resume.csv")
```

Subset data

- `select`: Return columns by name/number/etc.
- `filter`: Return rows by name/number/etc.

```
## subset data with first name
resumeN <- resume %>%
  select(firstname)

## subset data with black names
resumeB <- resume %>%
  filter(race == "black")

## subset data with black, female-sounding names
resumeBf <- resume %>%
  filter(race == "black" & sex == "female")
```

Summarize data

```
## callback rate for black female names
Bf_callback <- resume %>%
  filter(race == "black" & sex =="female") %>%
  summarize(callback_rate = mean(call, na.rm = TRUE))

## callback rate for white female names
Wf_callback <- resume %>%
  filter(race == "white" & sex == "female") %>%
  summarize(callback_rate = mean(call, na.rm = TRUE))

## difference between white and black women
Wf_callback - Bf_callback
```

```
##   callback_rate
## 1    0.03264689
```

Add new variable

calculate target values

The way we did previously with `filter()` and `summarise()`.

create factor variable

```
## create a factor variable that takes one of the four values
resume <- resume %>%
  mutate(type = case_when(race == "black" & sex == "female" ~ "BlackFemale",
                          race == "black" & sex == "male" ~ "BlackMale",
                          race == "white" & sex == "female" ~ "WhiteFemale",
                          race == "white" & sex == "male" ~ "WhiteMale",
                          TRUE ~ "Other"))

head(resume)
```

```
## # A tibble: 6 x 5
##   firstname sex    race   call type
##   <chr>     <chr>  <chr> <dbl> <chr>
## 1 Allison   female white     0 WhiteFemale
## 2 Kristen   female white     0 WhiteFemale
## 3 Lakisha   female black     0 BlackFemale
## 4 Latonya   female black     0 BlackFemale
## 5 Carrie    female white     0 WhiteFemale
## 6 Jay       male   white     0 WhiteMale
```