Chapter 2: Causality
Data Transformation with Tidyverse Functions

Claire Liow

University of Tokyo

5/31/2022

1. Aggregate data with `group_by`

2. Summarize data with `summarize`

3. Reshape the data with `pivot_wider`

Section 1

Aggregate data with `group_by`

Aggregate data with group_by

```r
## count the observation in each subset
race.call.summary <- resume %>%
  group_by(race, call) %>%
  count()

## calculate callback rates by race
callback_by_race <- resume %>%
  group_by(race, sex) %>%
  summarize(callback_rate = mean(call))
```

Section 2

Summarize data with `summarize`

Summarize data with `summarize`

```
resume %>%
group_by(race, sex) %>%
summarize(callback = mean(call))

resume %>%
  group_by(firstname) %>%
  summarize(callback = mean(call))
```

Section 3

Reshape the data with `pivot_wider`

Reshape the data with pivot_wider

```
## tidyverse
## without pivoting the data
resume %>%
group_by(race, sex) %>%
summarize(callback = mean(call))

## after pivoting the data
resume %>%
group_by(race, sex) %>%
summarize(callback = mean(call)) %>%
pivot_wider(names_from = race, values_from = callback)
```