Lecture Notes 2/6/2023

Prof. Bisbee, Vanderbilt University

2023-02-06

```
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✓ ggplot2 3.3.6     ✓ purrr   0.3.4
## ✓ tibble  3.1.7     ✓ dplyr   1.0.9
## ✓ tidyr   1.2.0     ✓ stringr 1.4.0
## ✓ readr   2.1.2     ✓ forcats 0.5.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
mi_ep <- read_rds('https://github.com/jbisbee1/DS1000_S2023/blob/main/Lectures/4_Uni_Mul
tivariate/data/MI2020_ExitPoll_small.rds?raw=true')
```

Preparing data “Wrangling”

```
MI_final_small <- mi_ep %>%
  filter(preschoice == 'Donald Trump, the Republican' | preschoice == 'Joe Biden, the De
mocrat') %>%
  mutate(BidenVoter = ifelse(preschoice == 'Joe Biden, the Democrat',1,0),
         TrumpVoter = ifelse(BidenVoter == 1,0,1),
         AGE10 = ifelse(AGE10 == 99,NA,AGE10))
```

```
MI_final_small %>%
  select(BidenVoter,TrumpVoter,preschoice)
```

```
## # A tibble: 1,182 × 3
##    BidenVoter TrumpVoter preschoice
##         <dbl>      <dbl> <chr>
##  1          1          0 Joe Biden, the Democrat
##  2          1          0 Joe Biden, the Democrat
##  3          1          0 Joe Biden, the Democrat
##  4          1          0 Joe Biden, the Democrat
##  5          1          0 Joe Biden, the Democrat
##  6          1          0 Joe Biden, the Democrat
##  7          1          0 Joe Biden, the Democrat
##  8          1          0 Joe Biden, the Democrat
##  9          0          1 Donald Trump, the Republican
## 10          1          0 Joe Biden, the Democrat
## # … with 1,172 more rows
```

Analysis

```
MI_final_small %>%
  count(SEX,preschoice) %>% # count total respondents by gender + choice
  mutate(prop_overall = n / sum(n)) %>% # calculate proportion of all
  group_by(SEX) %>% # 40+41: calculate total by gender
  mutate(tot_resp = sum(n)) %>%
  mutate(prop_gender = n / tot_resp) %>% # Calculate proportion gender
  select(SEX,preschoice,n,tot_resp,prop_gender) # Just make it visible in class
```

```
## # A tibble: 4 × 5
## # Groups:   SEX [2]
##     SEX preschoice                          n tot_resp prop_gender
##   <dbl> <chr>                           <int>    <int>       <dbl>
## 1     1 Donald Trump, the Republican   247      551       0.448
## 2     1 Joe Biden, the Democrat        304      551       0.552
## 3     2 Donald Trump, the Republican   212      631       0.336
## 4     2 Joe Biden, the Democrat        419      631       0.664
```

Multivariate: Age X Sex

```
toplot <- MI_final_small %>%
  group_by(SEX,AGE10) %>%
  summarise(prop_trump = mean(TrumpVoter,na.rm=T)) %>%
  # ungroup() %>%
  mutate(prop_trump = round(prop_trump,digits = 2)) %>%
  spread(key = SEX,value = prop_trump) %>%
  rename(Male = `1`,Female = `2`) %>%
  mutate(genderGap = Female - Male)
```

```
## `summarise()` has grouped output by 'SEX'. You can override using the `.groups`
## argument.
```

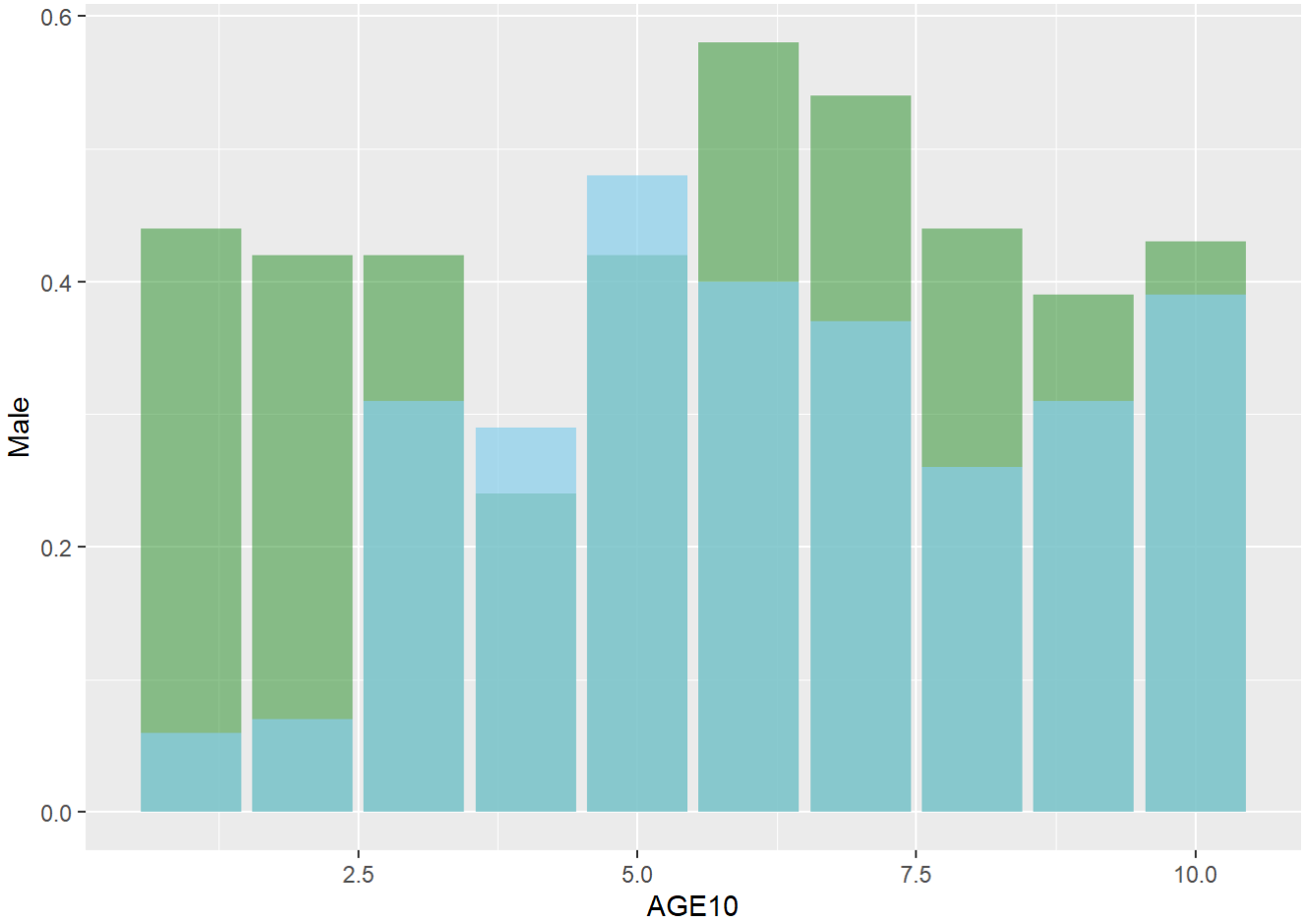
```
toplot
```

```
## # A tibble: 11 × 4
##    AGE10  Male Female genderGap
##    <dbl> <dbl>  <dbl>     <dbl>
##  1     1  0.44   0.06   -0.38
##  2     2  0.42   0.07   -0.35
##  3     3  0.42   0.31   -0.11
##  4     4  0.24   0.29    0.05
##  5     5  0.42   0.48    0.06
##  6     6  0.58   0.4    -0.18
##  7     7  0.54   0.37   -0.17
##  8     8  0.44   0.26   -0.18
##  9     9  0.39   0.31   -0.08
## 10    10  0.43   0.39   -0.0400
## 11    NA  0.67   0.57   -0.100
```

Visualize it!

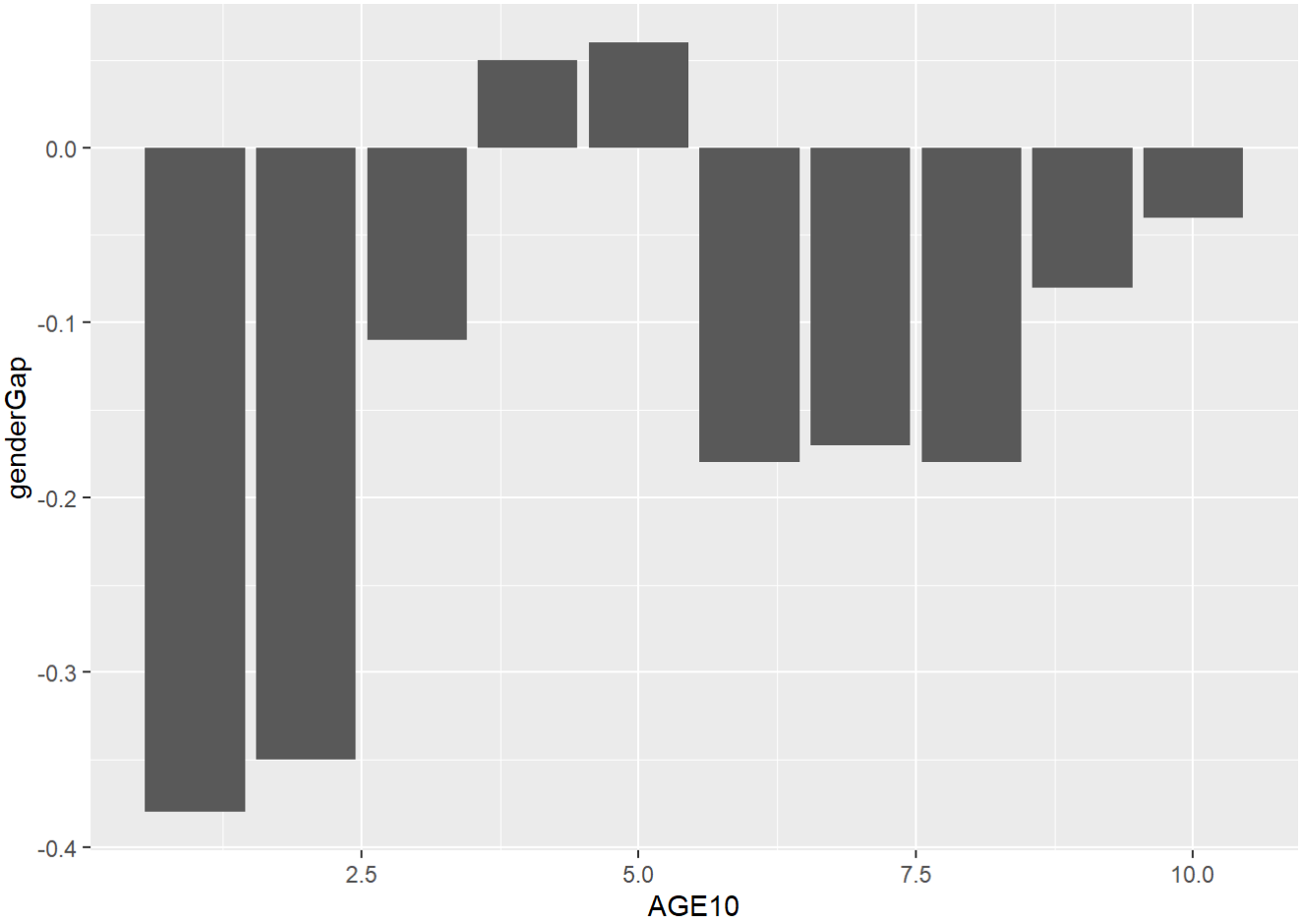
```
toplot %>%
  ggplot(aes(x = AGE10)) +
  geom_bar(aes(y = Male),fill = 'forestgreen',stat = 'identity',alpha = .5) +
  geom_bar(aes(y = Female),fill = 'skyblue',stat = 'identity',alpha = .7)
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
## Removed 1 rows containing missing values (position_stack).
```




```
toplot %>%
  ggplot(aes(x = AGE10,y = genderGap)) +
  geom_bar(stat = 'identity')
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```


0.0
-0.1
-0.2
-0.3
-0.4
genderGap
2.5
5.0
7.5
10.0
AGE10