

Review 1 Notes

2024-01-25

Testing Pset 3 Data

```
require(tidyverse)
```

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

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —  
## ✓ dplyr      1.1.2      ✓ readr      2.1.4  
## ✓ forcats    1.0.0      ✓ stringr    1.5.0  
## ✓ ggplot2     3.4.4      ✓ tibble     3.2.1  
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0  
## ✓ purrr      1.0.1
```

```
## — Conflicts — tidyverse_conflicts() —  
## ✗ dplyr::filter() masks stats::filter()  
## ✗ dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to  
o become errors
```

```
dat <- read_rds("https://github.com/jbisbee1/DS1000_S2024/raw/main/data/MI2020_ExitPoll.  
rds")
```

```
dat <- dat %>%  
  mutate(FAVTRUMP = as.numeric(haven::as_factor(FAVTRUMP))) %>%  
  mutate(fav_trump_text = ifelse(FAVTRUMP == 1,  
                                "Approve",  
                                ifelse(FAVTRUMP == 2,  
                                        "Disapprove",  
                                        "Refuse")))  
  
dat %>%  
  mutate(na_trump = ifelse(is.na(FAVTRUMP), 'This is NA', 'This is not NA')) %>%  
  count(FAVTRUMP, na_trump)
```

```
## # A tibble: 4 × 3
##   FAVTRUMP na_trump      n
##   <dbl> <chr>      <int>
## 1      1 This is not NA    220
## 2      2 This is not NA    385
## 3      3 This is not NA     10
## 4     NA This is NA     616
```

```
dat
```

```
## # A tibble: 1,231 × 64
##   ID WEIGHT LALVOTERID GROUP ZIP DISTRICT Z1 S1
##   <dbl> <dbl> <chr>      <dbl+lbl> <dbl> <dbl> <dbl> <dbl+lbl>
## 1      9  0.405 LALMI6290066 3 [3]  49327      2  NA 1 [Yes]
## 2     66  1.81  LALMI2492492 1 [1]  48234     14  NA 1 [Yes]
## 3    225  0.860 LALMI548981440 4 [4]  48301      9 48322 1 [Yes]
## 4    243  0.199 LALMI505377239 1 [1]  48130      7 48130 1 [Yes]
## 5    286  0.177 LALMI6831689 1 [1]  49946      1  NA 1 [Yes]
## 6    293  0.492 LALMI4019782 1 [1]  48615      4  NA 1 [Yes]
## 7    365  1.37  LALMI4151378 1 [1]  48906      4 48813 1 [Yes]
## 8    367  1.15  LALMI5912584 1 [1]  49442      2  NA 1 [Yes]
## 9    388  1.50  LALMI6635050 1 [1]  48451      5  NA 1 [Yes]
## 10   417  1.30  LALMI3567125 1 [1]  48197     12  NA 1 [Yes]
## # i 1,221 more rows
## # i 56 more variables: S2A <dbl+lbl>, S2B <dbl+lbl>, S3 <dbl+lbl>,
## #   S4 <dbl+lbl>, VERSION <dbl+lbl>, PRSMI20 <dbl+lbl>, SENMI20 <dbl+lbl>,
## #   TIME16 <dbl+lbl>, ISSUE20 <dbl+lbl>, QLT20 <dbl+lbl>, TEMPBIDEN <dbl+lbl>,
## #   TEMPTRUMP <dbl+lbl>, CONTROLSEN <dbl+lbl>, FINSIT <dbl+lbl>,
## #   ECONVCORONA20 <dbl+lbl>, FAVBIDEN2 <dbl+lbl>, FAVTRUMP <dbl>,
## #   FORCAND <dbl+lbl>, NEWVOTER <dbl+lbl>, NEC <dbl+lbl>, ...
```

Digging into research-ey questions

```
dat %>%
  count(fav_trump_text)
```

```
## # A tibble: 4 × 2
##   fav_trump_text      n
##   <chr>      <int>
## 1 Approve      220
## 2 Disapprove   385
## 3 Refuse        10
## 4 <NA>        616
```

```
# Manually
220 / (220+385+10)
```

```
## [1] 0.3577236
```

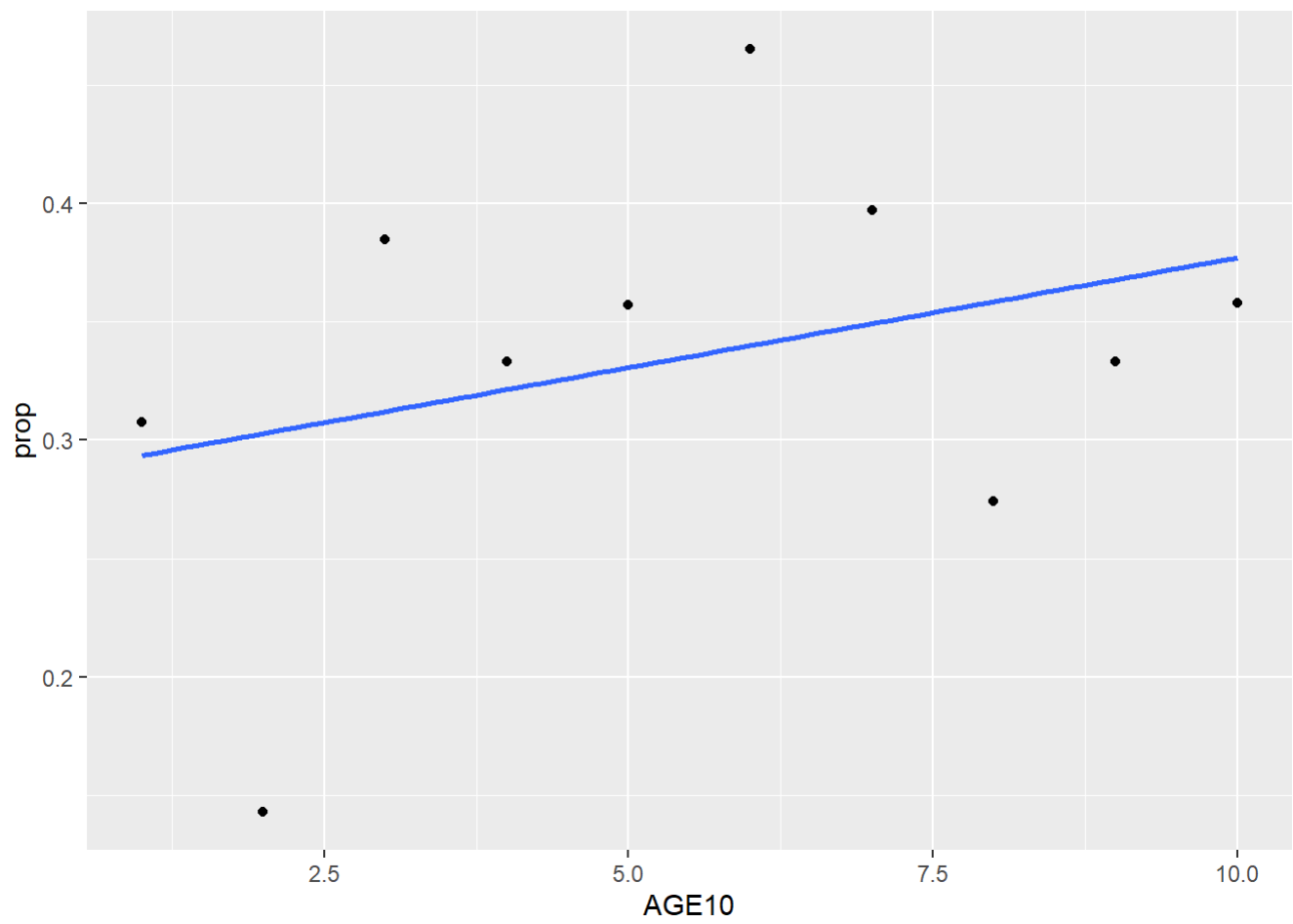
```
# Using code
dat %>%
  count(fav_trump_text) %>%
  filter(!is.na(fav_trump_text)) %>%
  mutate(totVoters = sum(n)) %>%
  mutate(prop = n / totVoters)
```

```
## # A tibble: 3 × 4
##   fav_trump_text      n totVoters  prop
##   <chr>          <int>    <int> <dbl>
## 1 Approve        220      615 0.358
## 2 Disapprove     385      615 0.626
## 3 Refuse         10      615 0.0163
```

RQ: Relationship between Age and Trump Approval

```
dat %>%
  filter(!is.na(fav_trump_text)) %>%
  select(AGE10, fav_trump_text) %>%
  count(AGE10, fav_trump_text) %>%
  group_by(AGE10) %>%
  mutate(totResp = sum(n)) %>%
  ungroup() %>%
  mutate(prop = n / totResp) %>%
  filter(fav_trump_text == "Approve") %>%
  filter(AGE10 < 99) %>%
  ggplot(aes(x = AGE10,
             y = prop)) +
  geom_point() +
  geom_smooth(method = 'lm', se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
#geom_bar(stat = 'identity')
```