Project proposal 2

Approximately Normal: Jenny Huang, Bella Larsen, Jessie Bierschenk, Aidan Gildea

16 October 2020

```
library(tidyverse)
library(broom)
library(patchwork)
```

```
elections <- read_csv("data/data_full.csv")
```

```
elections_clean <- elections %>%
  mutate(metro = if_else(METRO == 2, "Metro", "Not Metro/Unknown")) %>%
  mutate(sex = case_when(SEX == 1 ~ "Male",
                         SEX == 2 ~ "Female")) %>%
  mutate(marst = case_when(MARST == 1 ~ "Married",
                           MARST == 2 ~ "Married",
                           MARST == 4 ~ "Divorced/Separated",
                           MARST == 3 ~ "Divorced/Separated",
                           TRUE ~ "Not Married/Other")) %>%
  mutate(veteran = if_else(VETSTAT == 2, "Yes", "No/Uknnown")) %>%
  mutate(citizen = case_when(CITIZEN == 5 | CITIZEN == 9 ~ "No/Unknown",
                             CITIZEN == 1 | CITIZEN == 2 | CITIZEN == 3 ~
                               "Native Born",
                             CITIZEN == 4 ~ "Naturalized")) %>%
  mutate(ethnicity = if_else(HISPAN == 0 | HISPAN == 901 | HISPAN == 902,
                             "Not Hispanic/Unknown", "Hispanic/Latinx")) %>%
  mutate(employed = if_else(LABFORCE == 2, "Yes", "No/Unknown")) %>%
  mutate(highest_education = case_when(EDUC99 == 0 | EDUC99 == 1 ~
                                         "None/Unknown",
                               EDUC99 == 4 | EDUC99 == 5 | EDUC99 == 6 |
                                 EDUC99 == 7 | EDUC99 == 8 | EDUC99 == 9
                               ~ "Some High School",
                               EDUC99 == 10 ~ "High School Degree/GED",
                               EDUC99 == 11 ~ "Some College",
                               EDUC99 == 12 | EDUC99 == 13 | EDUC99 == 14
                               ~ "Associate Degree",
                               EDUC99 == 15 ~ "Bachelors Degree",
                               EDUC99 == 16 ~ "Masters Degree",
                               EDUC99 == 17 ~ "Professional Degree",
                               EDUC99 == 18 ~ "Doctorate Degree")) %>%
  mutate(current_student = case_when(SCHLCOLL == 5 | SCHLCOLL == 0 ~
                                       "No/Unknown",
                                     SCHLCOLL == 1  ~ "High School Full Time",
                                     SCHLCOLL == 2 ~ "High School Part Time",
                                     SCHLCOLL == 3 ~ "College Full Time",
                                     SCHLCOLL == 4 ~ "College Part Time")) %>%
#of people 16-24
```

```
  mutate(race = case_when(RACE == 820 | RACE == 830 | RACE == 830 | RACE == 819
                          | RACE == 804 | RACE == 805 | RACE == 806 |
                            RACE == 807 | RACE == 808 | RACE == 810 |
                            RACE == 811 | RACE == 812 | RACE == 813 |
                            RACE == 814 | RACE == 815 | RACE == 816 |
                            RACE == 817 | RACE == 818 | RACE == 803 |
                            RACE == 801 | RACE == 802 | RACE == 830 ~
                            "2 or more races",
                          RACE == 100 ~ "White",
                          RACE == 200 ~ "Black",
                          RACE == 651 | RACE == 809 | RACE == 652 | RACE == 650
                          ~ "Asian or Pacific Islander",
                          RACE ==  300 ~ "Native American",
                          RACE == 999 | RACE == 700 ~ "Other/Unknown")) %>%
  #Need to go over these race categorizations as a group!
  mutate(why_not_vote = case_when(VOWHYNOT == 10 ~ "Inconvenience",
                                  VOWHYNOT == 4 ~ "Interest",
                                  VOWHYNOT == 7 ~ "Political",
                                  VOWHYNOT == 9 | VOWHYNOT ==  6 |
                                    VOWHYNOT == 5 | VOWHYNOT == 2 ~
                                    "Logistical",
                                  VOWHYNOT == 1 ~ "Physically Unable",
                                  VOWHYNOT == 8 ~ "Registration Issues",
                                  VOWHYNOT == 3 ~ "Forgot")) %>%
#Reason why eligible voter did not vote
  mutate(why_not_reg = case_when(VOYNOTREG == 8 | VOYNOTREG == 3 ~
                                   "Not Eligible",
                                 VOYNOTREG == 7 | VOYNOTREG == 6 ~
                                   "Not Interested/Vote Won't Matter",
                                 VOYNOTREG == 5 ~ "Language Barrier",
                                 VOYNOTREG == 4 ~ "Physically Unable",
                                 VOYNOTREG == 2 | VOYNOTREG == 1 ~
                                   "Lacked Info/Missed Deadline")) %>%
  mutate(voting_method = case_when(VOTEHOW == 1 ~ "In Person",
                                   VOTEHOW == 2 ~ "Mail-In")) %>%
  mutate(voting_time = case_when(VOTEWHEN == 2 ~ "Early",
                                 VOTEWHEN == 1 ~"Voting Day")) %>%
  mutate(voted = case_when(VOTED == 1 ~ 0,
                           VOTED == 2 ~ 1)) %>%
  mutate(registered = case_when(VOREG == 1 ~ 0,
                           VOREG == 2 ~ 1)) %>%
  mutate(how_registered = case_when(VOREGHOW == 8 ~ "Online",
                                    VOREGHOW == 7 ~ "Polling Place",
                                    VOREGHOW == 6 ~ "Registration Drive",
                                    VOREGHOW == 5 | VOREGHOW == 2 |
                                      VOREGHOW == 1 ~
                                      "Govt Office/Public Agency",
                                    VOREGHOW == 4 ~ "School/College/Hospital",
                                    VOREGHOW == 3 ~ "By Mail")) %>%
  select(metro, sex, marst, veteran, citizen, ethnicity, employed,
         highest_education, current_student, race, why_not_vote, why_not_reg,
         voting_method, voting_time, voted, registered, how_registered, YEAR,
         STATEFIP, AGE)
```

```r
elections_clean <- elections_clean %>%
  mutate(metro = factor(metro),
         sex = factor (sex),
         marst = factor(marst),
         veteran = factor(veteran),
         citizen = factor(citizen),
         ethnicity = factor(ethnicity),
         employed = factor(employed),
         highest_education = factor(highest_education),
         current_student = factor(current_student),
         race = factor(race),
         why_not_vote = factor(why_not_vote),
         why_not_reg = factor(why_not_reg),
         voting_method = factor(voting_method),
         voting_time = factor(voting_time),
         voted = factor(voted),
         registered = factor(registered),
         how_registered = factor(how_registered))
```

```r
glimpse(elections_clean)
```

```
## Rows: 643,429
## Columns: 20
## $ metro             <fct> Not Metro/Unknown, Not Metro/Unknown, Not Metro/U...
## $ sex               <fct> Male, Male, Female, Male, Female, Female, Male, M...
## $ marst             <fct> Divorced/Separated, Married, Married, Married, Ma...
## $ veteran           <fct> No/Uknnown, No/Uknnown, No/Uknnown, No/Uknnown, N...
## $ citizen           <fct> Native Born, Native Born, Native Born, Native Bor...
## $ ethnicity         <fct> Not Hispanic/Unknown, Not Hispanic/Unknown, Not H...
## $ employed          <fct> Yes, No/Unknown, No/Unknown, Yes, Yes, No/Unknown...
## $ highest_education <fct> Some College, High School Degree/GED, Some High S...
## $ current_student   <fct> No/Unknown, No/Unknown, No/Unknown, No/Unknown, N...
## $ race              <fct> White, White, White, White, White, Black, Black, ...
## $ why_not_vote      <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ why_not_reg       <fct> NA, NA, NA, NA, NA, NA, Lacked Info/Missed Deadli...
## $ voting_method     <fct> In Person, NA, In Person, In Person, In Person, I...
## $ voting_time       <fct> Voting Day, NA, Voting Day, Voting Day, Voting Da...
## $ voted             <fct> 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1...
## $ registered        <fct> NA, 1, NA, NA, NA, NA, 0, 0, NA, NA, NA, NA, 0, N...
## $ how_registered    <fct> Govt Office/Public Agency, Govt Office/Public Age...
## $ YEAR              <dbl> 2004, 2004, 2004, 2004, 2004, 2004, 2004, 2004, 2...
## $ STATEFIP          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ AGE               <dbl> 38, 70, 61, 60, 60, 37, 50, 38, 34, 31, 80, 42, 1...
```