Project 1

Xiaoli Sun(xs2338)

9/20/2020

Load data:

```
library(haven)
anes_dat <-
    read_sav("../data/anes_timeseries_cdf.sav")
```

Use Library:

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.2
```

```
library(haven)
library(RColorBrewer)
library(ggplot2)
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.0.2
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.0.2
```

Clean data:

```
election_years=as.character(seq(1948, 2016, 4))
data<-anes_dat%>%
  transmute(
    year=as_factor(VCF0004),
    turnout=as_factor(VCF0703),
    vote=as_factor(VCF0706),
    age = as_factor(VCF0102),
    gender=as_factor(VCF0104),
    # share with the family
    party=as_factor(VCF0302),
    partyoffather = as_factor(VCF0306),
    partyofmother = as_factor(VCF0307),
    #info on president
```

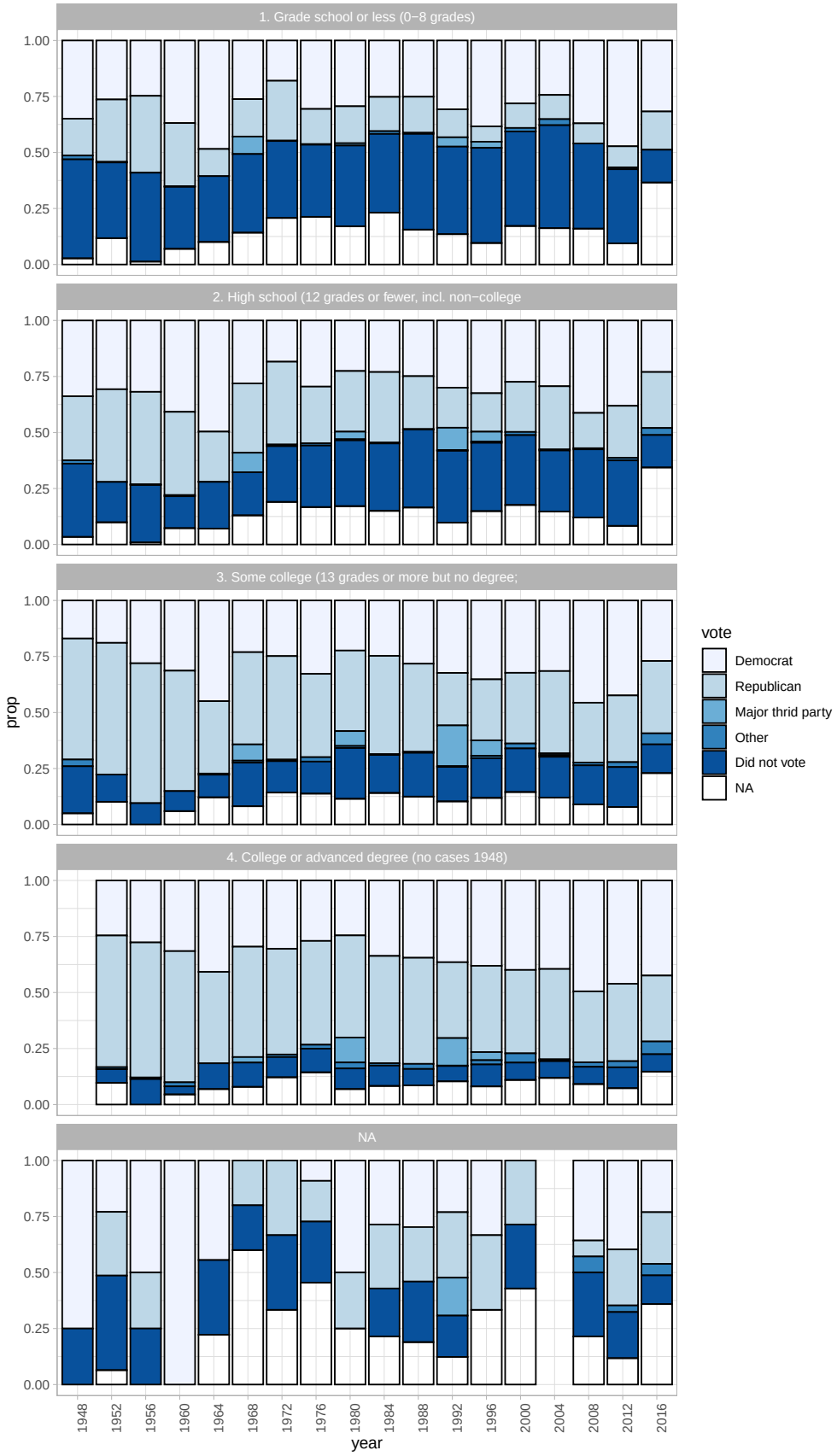
```
    intelligent = as_factor(VCF0338),
    compassionate = as_factor(VCF0339),
    decent = as_factor(VCF0340),
    inspring = as_factor(VCF0341),
    knowledgeable = as_factor(VCF0342),
    moral = as_factor(VCF0343),
    leadership = as_factor(VCF0344),
    cares = as_factor(VCF0345),
    #behavior
    influenceothers = as_factor(VCF0717),
    talkwithothers = as_factor(VCF0732),
    #opinion
    womenequal = as_factor(VCF0834),
    healthinsurance = as_factor(VCF0806),
    job = as_factor(VCF0809),
    occupation = as_factor(VCF0151),
    mostimportant = as_factor(VCF0875),
    #parties
    republican_healthinsurance = as_factor(VCF9093),
    republican_womenequal = as_factor(VCF9091), #VCF0538
    republican_defense = as_factor(VCF9089),
    democratic_healthinsurance = as_factor(VCF9085),
    democratic_womenqual = as_factor(VCF9081),
    democratic_defense = as_factor(VCF9087),
    #environment
    education = as_factor(VCF0110),
    region = as_factor(VCF0112),
    income = as_factor(VCF0114),

    #political issue
    self.consider = as_factor(VCF0301),
    interest = as_factor(VCF0310),
    care.win = as_factor(VCF0311)
    ) %>% filter(year %in% election_years)
levels(data$age) <- c('NA','17-24','25-34','35-44','45-54','55-64','65-74','75 and over')
levels(data$vote) <- c('NA','Democrat','Republican','Major thrid party','Other','Did not vote')
# save data
save(data, file="../output/mydata2.RData")
```

How do Education Affects Voting?

Here we use VCF0110 to describe the education level.

After learning from the education level, the first thing to check is whether there exist a kind of tendency for under-educated or well-educated. As the following Figure shows, people who took less than 8 grades may have a good chance voting for nothing. Also, the more one is educated, the less the probability of not voting for president is.

How did Different Education Groups Acted in the Election?



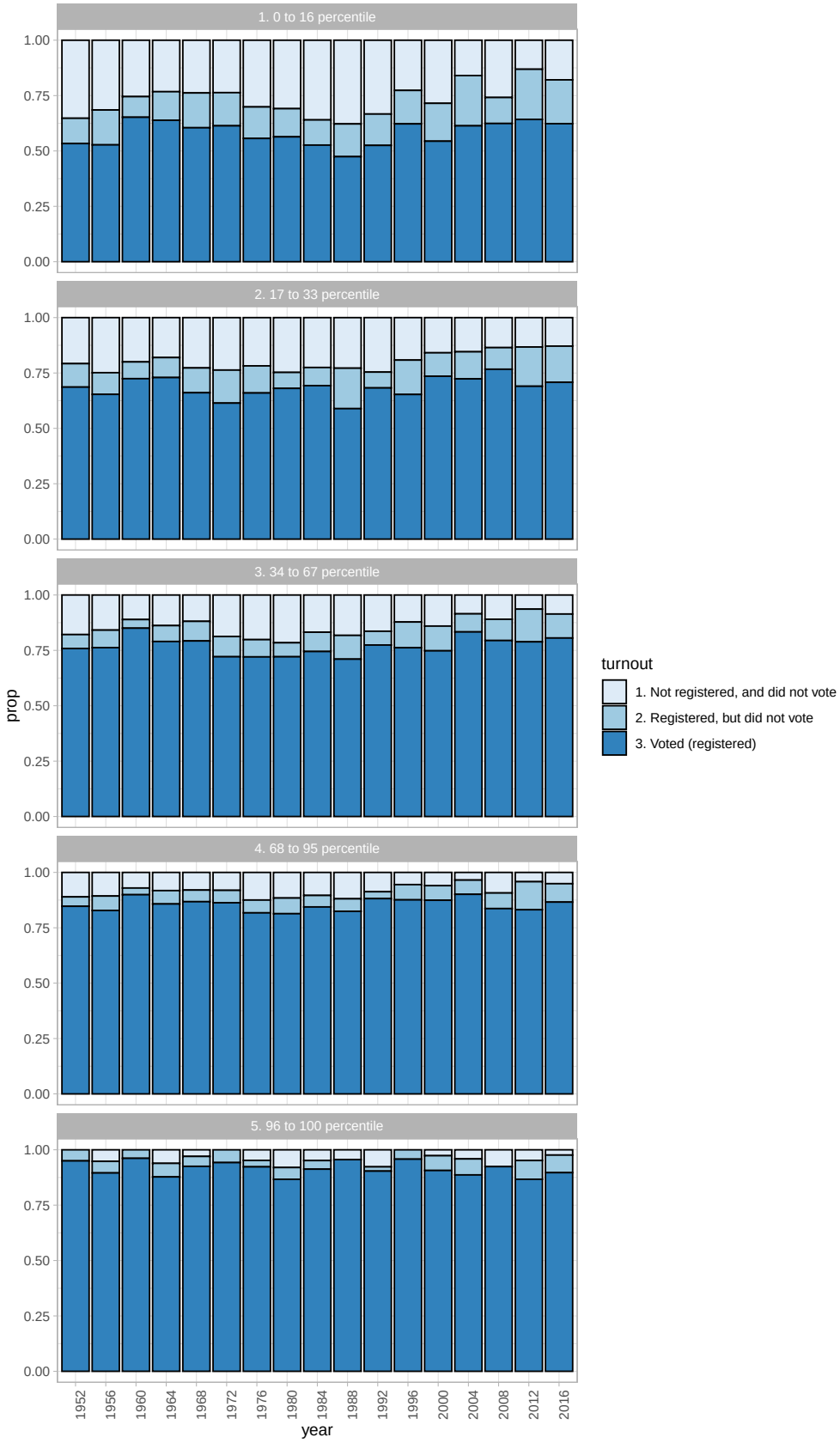
It is quite clear that the education level of people keeps improving for decades. People are more willing to vote when there education level increase.

Income and Turnout

According to the conclution from education analysis, we know that people are more willing to vote when they have a higher education level. From experience, people who have a higher education level earn more. So let's look into the relationship between people's income and the turnout.

```
anes_income_turnout = data %>%
  filter(!is.na(income) & !is.na(turnout))%>%
  group_by(year, income)%>%
  count(turnout)%>%
  group_by(year, income)%>%
  mutate(
    prop=n/sum(n)
  )

ggplot(anes_income_turnout,
       aes(x=year, y=prop, fill=turnout)) +
  geom_bar(stat="identity", colour="black") + facet_wrap(~income, ncol=1) + theme_light()+
  theme(axis.text.x = element_text(angle = 90))+
  scale_fill_brewer(palette="Blues")+
  labs(title="How did different racial groups participated in the election \n over the years?")
```

How did different racial groups participated in the election over the years?



```
ggsave("../figs/plot2.png")
```

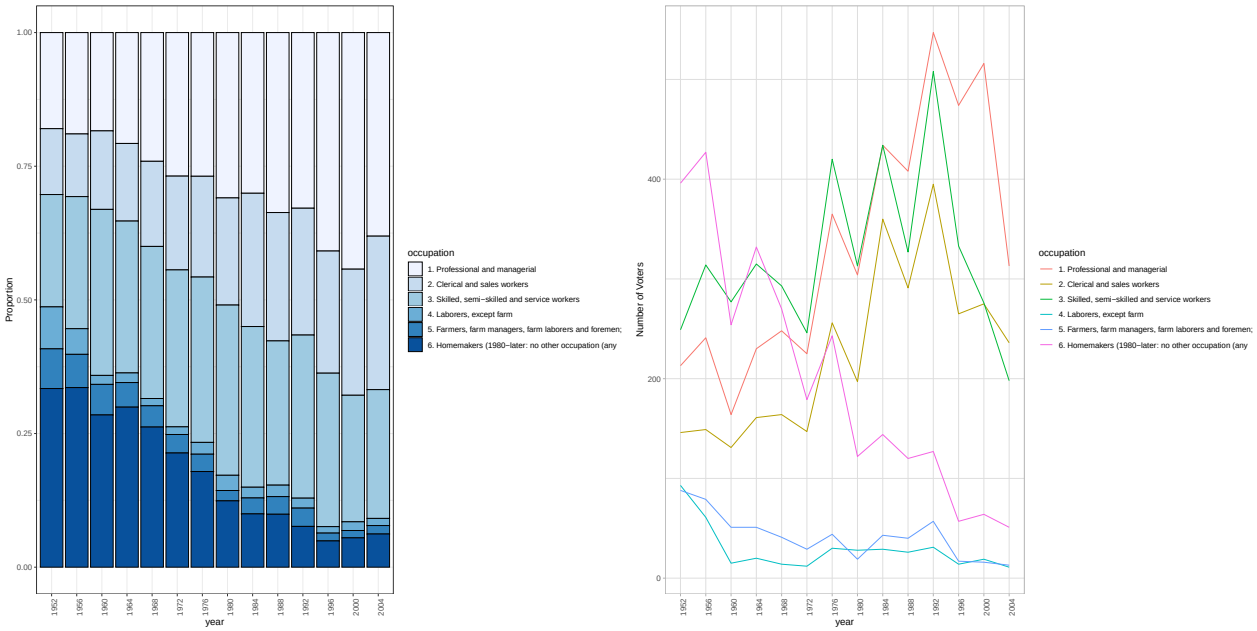
```
## Saving 8 x 14 in image
```

Occupation

```
## Warning: package 'cowplot' was built under R version 4.0.2
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggpubr':
##
##     get_legend
```



Over the past presidential elections, the proportion of homemakers has drastically decreased, whereas the proportion of professional and managerial has remarkably increased. The proportion of skilled, semi-skilled and service workers has remained approximately the same over time.

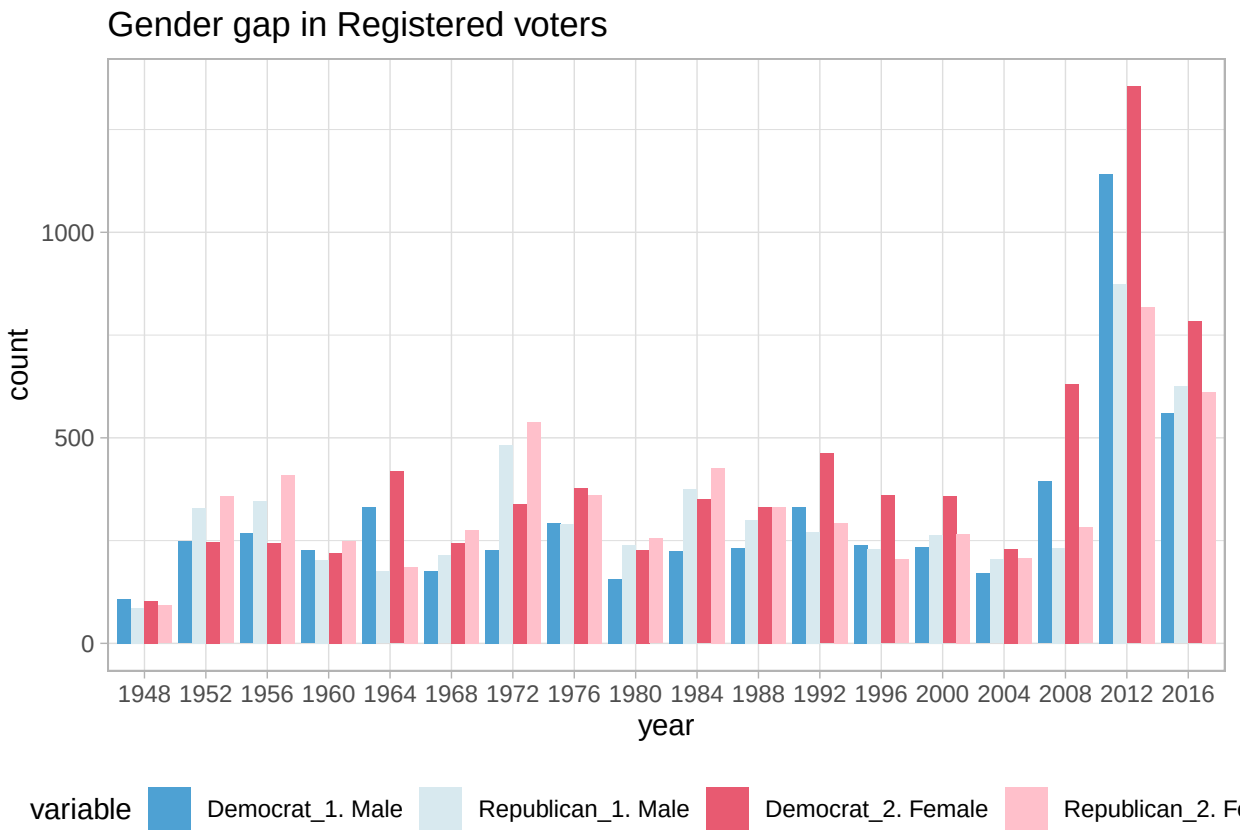
Is there a gender gap?

gender gaps

```
gap <- data %>% select(gender,year,vote) %>% filter(!is.na(gender) & gender != '3. Other (2016)') %>%
  group_by(gender,vote,year) %>% count() %>%
  filter(vote %in% c('Democrat','Republican')) %>%
  pivot_wider(names_from = c(vote,gender), values_from = n) %>%
  pivot_longer(-year, names_to = 'variable', values_to = 'count') %>%
  mutate(variable = as_factor(variable))
```

```
## Warning: Factor `vote` contains implicit NA, consider using
## `forcats::fct_explicit_na`

## Warning: Factor `vote` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
#levels(gap$variable)
ggplot(gap) +geom_bar(stat = 'identity',aes(x=year,y=count,fill = variable),position ='dodge')+
  scale_fill_manual(values=c('#4ea1d3', "#d8e9ef","#e85a71","pink"))+
  labs(title = 'Gender gap in Registered voters')+
  theme_light()+
  theme(legend.position = 'bottom')
```



The gender gap in voting refers to the difference in the percentage of women and the percentage of men voting for a given candidate. With blues representing male's votes and pink representing female's votes, the graph shows different preference between men and women. I would say there is a gender gap in 1980 as men voted more for republic while women vote two parties evenly. The gender gap is more obvious in recent years. In 2000, more women vote for democratic and men vote oppositely more to republic. Same scenario in 2004 and 2016.