GR5243 Project 1

This is an R Markdown Notebook.

1. Introduction

The *American National Election Studies* (ANES) are surveys of voters in the U.S. on the national scale. For each presidential election since 1948, ANES collects responses from respondents both before and after the election. The goal of ANES is to understand political behaviors using systematic surveys. ANES' data and results have been routinely used by news outlets, election campaigns and political researchers.

2. Explorary Data Analysis

Import R packages for later use

```
library(usethis)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(haven)
library(devtools)
library(RColorBrewer)
library(DT)
library(ggplot2)
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(stringr)
```

Import raw ANES data

```
anes_ts_cdf <- read_sav("D:/Columbia University/GR 5243/Project 1/Fall2020-Project1-zlj-0131/data/anes_

# Take a look at the data sets
head(anes_ts_cdf)
```

```
## # A tibble: 6 x 1,029
##   Version VCF0004 VCF0006 VCF0006a VCF0009x VCF0010x VCF0011x VCF0009y VCF0010y
##   <chr>     <dbl>   <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 ANES_C~    1948    1001 19481001        1        1        1        1        1
## 2 ANES_C~    1948    1002 19481002        1        1        1        1        1
## 3 ANES_C~    1948    1003 19481003        1        1        1        1        1
## 4 ANES_C~    1948    1004 19481004        1        1        1        1        1
## 5 ANES_C~    1948    1005 19481005        1        1        1        1        1
## 6 ANES_C~    1948    1006 19481006        1        1        1        1        1
## # ... with 1,020 more variables: VCF0011y <dbl>, VCF0009z <dbl>,
## #   VCF0010z <dbl>, VCF0011z <dbl>, VCF0012 <dbl>, VCF0013 <dbl+lbl>,
## #   VCF0014 <dbl+lbl>, VCF0015a <dbl+lbl>, VCF0015b <dbl+lbl>,
## #   VCF0016 <dbl+lbl>, VCF0017 <dbl+lbl>, VCF0018a <dbl+lbl>,
## #   VCF0018b <dbl+lbl>, VCF0019 <dbl+lbl>, VCF0050a <dbl+lbl>,
## #   VCF0050b <dbl+lbl>, VCF0070a <dbl+lbl>, VCF0070b <dbl+lbl>,
## #   VCF0071a <dbl+lbl>, VCF0071b <dbl+lbl>, VCF0071c <dbl+lbl>,
## #   VCF0071d <dbl+lbl>, VCF0072a <dbl+lbl>, VCF0072b <dbl+lbl>,
## #   VCF0101 <dbl+lbl>, VCF0102 <dbl+lbl>, VCF0103 <dbl+lbl>, VCF0104 <dbl+lbl>,
## #   VCF0105a <dbl+lbl>, VCF0105b <dbl+lbl>, VCF0106 <dbl+lbl>,
## #   VCF0107 <dbl+lbl>, VCF0108 <dbl+lbl>, VCF0109 <dbl+lbl>, VCF0110 <dbl+lbl>,
## #   VCF0111 <dbl+lbl>, VCF0112 <dbl+lbl>, VCF0113 <dbl+lbl>, VCF0114 <dbl+lbl>,
## #   VCF0115 <dbl+lbl>, VCF0116 <dbl+lbl>, VCF0117 <dbl+lbl>, VCF0118 <dbl+lbl>,
## #   VCF0119 <dbl+lbl>, VCF0120 <dbl+lbl>, VCF0121 <dbl+lbl>, VCF0122 <dbl+lbl>,
## #   VCF0123 <dbl+lbl>, VCF0124 <dbl+lbl>, VCF0125 <dbl+lbl>, VCF0126 <dbl+lbl>,
## #   VCF0126a <dbl+lbl>, VCF0126b <dbl+lbl>, VCF0126c <dbl+lbl>,
## #   VCF0127 <dbl+lbl>, VCF0127a <dbl+lbl>, VCF0127b <dbl+lbl>,
## #   VCF0128 <dbl+lbl>, VCF0128a <dbl+lbl>, VCF0128b <dbl+lbl>,
## #   VCF0129 <dbl+lbl>, VCF0130 <dbl+lbl>, VCF0130a <dbl+lbl>,
## #   VCF0131 <dbl+lbl>, VCF0132 <dbl+lbl>, VCF0133 <dbl+lbl>, VCF0134 <dbl+lbl>,
## #   VCF0135 <dbl+lbl>, VCF0136 <dbl+lbl>, VCF0137 <dbl+lbl>, VCF0138 <dbl+lbl>,
## #   VCF0138a <dbl+lbl>, VCF0138b <dbl+lbl>, VCF0138c <dbl+lbl>,
## #   VCF0138d <dbl+lbl>, VCF0138e <dbl+lbl>, VCF0139 <dbl+lbl>,
## #   VCF0140 <dbl+lbl>, VCF0140a <dbl+lbl>, VCF0141 <dbl+lbl>,
## #   VCF0142 <dbl+lbl>, VCF0143 <dbl+lbl>, VCF0144 <dbl+lbl>, VCF0145 <dbl+lbl>,
## #   VCF0146 <dbl+lbl>, VCF0147 <dbl+lbl>, VCF0148 <dbl+lbl>,
## #   VCF0148a <dbl+lbl>, VCF0149 <dbl+lbl>, VCF0150 <dbl+lbl>,
## #   VCF0151 <dbl+lbl>, VCF0152 <dbl+lbl>, VCF0153a <dbl+lbl>,
## #   VCF0153b <dbl+lbl>, VCF0153c <dbl+lbl>, VCF0154a <dbl+lbl>,
```

```
## #   VCF0154b <dbl+lbl>, VCF0155 <dbl+lbl>, VCF0156 <dbl+lbl>,
## #   VCF0157 <dbl+lbl>, ...
```

Process variables for analysis

For this project, I would start from the question of is there a relationship between voters' religious background and the party they would like to vote. I would concentrate on the two major parties, Democrat and republican, because they are always the candidates of the President election.

Data process for Question 1 (Religious background - Vote Party):

```
Election_years = as.character(seq(1952, 2016, 4))

Q1_data <- anes_ts_cdf %>%
  mutate(
  year = as_factor(VCF0004),
  vote = as_factor(VCF0704),
  religion = as_factor(VCF0128)
  ) %>%
  filter(year %in% Election_years) %>%
  filter(str_detect(vote, c("Democrat", "Republican")))

data.table(Q1_data %>%
             select(year, vote, religion) %>%
             sample_n(30))
```

```
##     year          vote                                        religion
##  1: 1988 2. Republican                                  1. Protestant
##  2: 2008   1. Democrat 4. Other and none (also includes DK preference)
##  3: 1968 2. Republican                                  1. Protestant
##  4: 1992 2. Republican                    2. Catholic [Roman Catholic]
##  5: 2016   1. Democrat 4. Other and none (also includes DK preference)
##  6: 2012 2. Republican                                  1. Protestant
##  7: 2012 2. Republican                    2. Catholic [Roman Catholic]
##  8: 1972 2. Republican                                  1. Protestant
##  9: 1964   1. Democrat                                  1. Protestant
## 10: 1984 2. Republican                                  1. Protestant
## 11: 1980 2. Republican                                  1. Protestant
## 12: 2008 2. Republican                    2. Catholic [Roman Catholic]
## 13: 1972 2. Republican                                  1. Protestant
## 14: 1952 2. Republican                                  1. Protestant
## 15: 2012 2. Republican                                  1. Protestant
## 16: 1984 2. Republican                    2. Catholic [Roman Catholic]
## 17: 2016 2. Republican 4. Other and none (also includes DK preference)
## 18: 1996   1. Democrat                                  1. Protestant
## 19: 2016 2. Republican                    2. Catholic [Roman Catholic]
## 20: 1952 2. Republican                                       3. Jewish
## 21: 1988 2. Republican                                  1. Protestant
## 22: 1972   1. Democrat                                  1. Protestant
## 23: 2012   1. Democrat                                  1. Protestant
## 24: 2008 2. Republican                    2. Catholic [Roman Catholic]
## 25: 2004 2. Republican                                  1. Protestant
## 26: 1996 2. Republican                                  1. Protestant
```

```
## 27: 2016 2. Republican                                    1. Protestant
## 28: 1976   1. Democrat                                    1. Protestant
## 29: 2016 2. Republican                     2. Catholic [Roman Catholic]
## 30: 2016   1. Democrat                                    1. Protestant
##     year          vote                                         religion
```

```
Q1_data %>% select(year, vote, religion)
```

```
## # A tibble: 12,183 x 3
##    year  vote          religion
##    <fct> <fct>         <fct>
##  1 1952  2. Republican 1. Protestant
##  2 1952  2. Republican 1. Protestant
##  3 1952  2. Republican 1. Protestant
##  4 1952  2. Republican 1. Protestant
##  5 1952  1. Democrat   1. Protestant
##  6 1952  2. Republican 1. Protestant
##  7 1952  1. Democrat   1. Protestant
##  8 1952  1. Democrat   2. Catholic [Roman Catholic]
##  9 1952  2. Republican 1. Protestant
## 10 1952  1. Democrat   1. Protestant
## # ... with 12,173 more rows
```

```
Q1_data <- Q1_data %>% select(year, vote, religion)

save(Q1_data, file = "../output/data_use.RData")
```

I am also interested in the relationship between Voters' education background and the party they would like to vote.

Data process for Question 2 (Education Background - Vote Party):

```
Q2_data <- anes_ts_cdf %>%
  mutate(
  year = as_factor(VCF0004),
  vote = as_factor(VCF0704),
  education = as_factor(VCF0110)
  ) %>%
  filter(year %in% Election_years) %>%
  filter(str_detect(vote, c("Democrat", "Republican")))

data.table(Q2_data %>%
             select(year, vote, education) %>%
             sample_n(30))
```

```
##     year          vote                                            education
##  1: 1964 2. Republican 2. High school (12 grades or fewer, incl. non-college
##  2: 1992   1. Democrat     3. Some college (13 grades or more but no degree;
##  3: 1988 2. Republican 2. High school (12 grades or fewer, incl. non-college
##  4: 1976   1. Democrat 2. High school (12 grades or fewer, incl. non-college
##  5: 1984 2. Republican         4. College or advanced degree (no cases 1948)
##  6: 2008 2. Republican 2. High school (12 grades or fewer, incl. non-college
##  7: 2016 2. Republican     3. Some college (13 grades or more but no degree;
```

```
##  8: 2008 2. Republican     3. Some college (13 grades or more but no degree;
##  9: 2012   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 10: 2016   1. Democrat     3. Some college (13 grades or more but no degree;
## 11: 2008   1. Democrat         4. College or advanced degree (no cases 1948)
## 12: 1956   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 13: 2016 2. Republican     3. Some college (13 grades or more but no degree;
## 14: 1996 2. Republican 2. High school (12 grades or fewer, incl. non-college
## 15: 1972 2. Republican                 1. Grade school or less (0-8 grades)
## 16: 2012   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 17: 2004 2. Republican     3. Some college (13 grades or more but no degree;
## 18: 1980 2. Republican         4. College or advanced degree (no cases 1948)
## 19: 1992   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 20: 1968   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 21: 1996 2. Republican     3. Some college (13 grades or more but no degree;
## 22: 1952 2. Republican 2. High school (12 grades or fewer, incl. non-college
## 23: 2000 2. Republican         4. College or advanced degree (no cases 1948)
## 24: 2012   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 25: 1992 2. Republican         4. College or advanced degree (no cases 1948)
## 26: 1992   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 27: 1964 2. Republican                 1. Grade school or less (0-8 grades)
## 28: 2012 2. Republican     3. Some college (13 grades or more but no degree;
## 29: 1984 2. Republican 2. High school (12 grades or fewer, incl. non-college
## 30: 2012   1. Democrat         4. College or advanced degree (no cases 1948)
##     year          vote                                              education
```

```
Q2_data %>% select(year, vote, education)
```

```
## # A tibble: 12,183 x 3
##    year  vote          education
##    <fct> <fct>         <fct>
##  1 1952  2. Republican 2. High school (12 grades or fewer, incl. non-college
##  2 1952  2. Republican 2. High school (12 grades or fewer, incl. non-college
##  3 1952  2. Republican 1. Grade school or less (0-8 grades)
##  4 1952  2. Republican 4. College or advanced degree (no cases 1948)
##  5 1952  1. Democrat   2. High school (12 grades or fewer, incl. non-college
##  6 1952  2. Republican 1. Grade school or less (0-8 grades)
##  7 1952  1. Democrat   2. High school (12 grades or fewer, incl. non-college
##  8 1952  1. Democrat   1. Grade school or less (0-8 grades)
##  9 1952  2. Republican 2. High school (12 grades or fewer, incl. non-college
## 10 1952  1. Democrat   2. High school (12 grades or fewer, incl. non-college
## # ... with 12,173 more rows
```

```
Q2_data <- Q2_data %>% select(year, vote, education)

save(Q2_data, file = "../output/data_use.RData")
```

I am also interested in the relationship between Voters' occupation and the party they would like to vote.

Data process for Question 3 (Occupations - Vote Party):

```
Q3_data <- anes_ts_cdf %>%
  mutate(
  year = as_factor(VCF0004),
```

```
  vote = as_factor(VCF0704),
  occupation = as_factor(VCF0115)
  ) %>%
  filter(year %in% Election_years) %>%
  filter(str_detect(vote, c("Democrat", "Republican")))

data.table(Q3_data %>%
             select(year, vote, occupation) %>%
             sample_n(30))
```

```
##     year          vote                                              occupation
##  1: 2012 2. Republican                                                   <NA>
##  2: 1952   1. Democrat                                                   <NA>
##  3: 1992 2. Republican 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  4: 2016   1. Democrat                                                   <NA>
##  5: 1972 2. Republican                        1. Professional and managerial
##  6: 2012 2. Republican                                                   <NA>
##  7: 1984   1. Democrat                        1. Professional and managerial
##  8: 1996   1. Democrat                        1. Professional and managerial
##  9: 2012   1. Democrat                                                   <NA>
## 10: 2008 2. Republican                                                   <NA>
## 11: 1992 2. Republican                        1. Professional and managerial
## 12: 2012 2. Republican                                                   <NA>
## 13: 2016   1. Democrat                                                   <NA>
## 14: 1968   1. Democrat          3. Skilled, semi-skilled and service workers
## 15: 1956 2. Republican 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## 16: 1988   1. Democrat 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## 17: 1976 2. Republican                        1. Professional and managerial
## 18: 2012   1. Democrat                                                   <NA>
## 19: 2016   1. Democrat                                                   <NA>
## 20: 2000   1. Democrat                        1. Professional and managerial
## 21: 2012 2. Republican                                                   <NA>
## 22: 2012 2. Republican                                                   <NA>
## 23: 1952 2. Republican 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## 24: 2000   1. Democrat                        1. Professional and managerial
## 25: 1968   1. Democrat                         2. Clerical and sales workers
## 26: 2008   1. Democrat                                                   <NA>
## 27: 1952 2. Republican          3. Skilled, semi-skilled and service workers
## 28: 2000 2. Republican                        1. Professional and managerial
## 29: 2012   1. Democrat                                                   <NA>
## 30: 2008   1. Democrat                                                   <NA>
##     year          vote                                              occupation
```

```
Q3_data %>% select(year, vote, occupation)
```

```
## # A tibble: 12,183 x 3
##    year  vote          occupation
##    <fct> <fct>         <fct>
##  1 1952  2. Republican 3. Skilled, semi-skilled and service workers
##  2 1952  2. Republican 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  3 1952  2. Republican 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  4 1952  2. Republican 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
```

```
##  5 1952  1. Democrat   6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  6 1952  2. Republican 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  7 1952  1. Democrat   3. Skilled, semi-skilled and service workers
##  8 1952  1. Democrat   6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  9 1952  2. Republican 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## 10 1952  1. Democrat   6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## # ... with 12,173 more rows
```

```
Q3_data <- Q3_data %>% select(year, vote, occupation)

save(Q3_data, file = "../output/data_use.RData")
```

Data process for new derivative question:

```
newQ_data <- anes_ts_cdf %>%
  mutate(
  year = as_factor(VCF0004),
  vote = as_factor(VCF0704),
  education = as_factor(VCF0110),
  religion = as_factor(VCF0128),
  occupation = as_factor(VCF0115)
  ) %>%
  filter(year %in% Election_years) %>%
  filter(str_detect(vote, c("Democrat", "Republican")))

data.table(newQ_data %>%
             select(year, vote, education, religion, occupation) %>%
             sample_n(30))
```

```
##     year          vote                                              education
##  1: 1980 2. Republican 2. High school (12 grades or fewer, incl. non-college
##  2: 2012   1. Democrat 2. High school (12 grades or fewer, incl. non-college
##  3: 1988 2. Republican 2. High school (12 grades or fewer, incl. non-college
##  4: 1972   1. Democrat 2. High school (12 grades or fewer, incl. non-college
##  5: 1972   1. Democrat                  1. Grade school or less (0-8 grades)
##  6: 1980   1. Democrat                  1. Grade school or less (0-8 grades)
##  7: 2008   1. Democrat 2. High school (12 grades or fewer, incl. non-college
##  8: 1964 2. Republican                  1. Grade school or less (0-8 grades)
##  9: 2012   1. Democrat     3. Some college (13 grades or more but no degree;
## 10: 1996 2. Republican         4. College or advanced degree (no cases 1948)
## 11: 2012 2. Republican     3. Some college (13 grades or more but no degree;
## 12: 2012   1. Democrat         4. College or advanced degree (no cases 1948)
## 13: 1992 2. Republican         4. College or advanced degree (no cases 1948)
## 14: 1956 2. Republican 2. High school (12 grades or fewer, incl. non-college
## 15: 2012   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 16: 1980 2. Republican     3. Some college (13 grades or more but no degree;
## 17: 1952 2. Republican 2. High school (12 grades or fewer, incl. non-college
## 18: 1952   1. Democrat                  1. Grade school or less (0-8 grades)
## 19: 1968   1. Democrat                  1. Grade school or less (0-8 grades)
## 20: 1976   1. Democrat                  1. Grade school or less (0-8 grades)
## 21: 1956 2. Republican                  1. Grade school or less (0-8 grades)
## 22: 1968 2. Republican 2. High school (12 grades or fewer, incl. non-college
## 23: 2012 2. Republican     3. Some college (13 grades or more but no degree;
```

```
## 24: 1984 2. Republican 2. High school (12 grades or fewer, incl. non-college
## 25: 2012   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 26: 1956   1. Democrat                  1. Grade school or less (0-8 grades)
## 27: 1972   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 28: 2016   1. Democrat 2. High school (12 grades or fewer, incl. non-college
## 29: 1984   1. Democrat         4. College or advanced degree (no cases 1948)
## 30: 2016 2. Republican         4. College or advanced degree (no cases 1948)
##     year          vote                                                education
##                                         religion
##  1:                    2. Catholic [Roman Catholic]
##  2: 4. Other and none (also includes DK preference)
##  3:                                    1. Protestant
##  4:                                    1. Protestant
##  5:                                    1. Protestant
##  6:                                    1. Protestant
##  7:                    2. Catholic [Roman Catholic]
##  8:                                    1. Protestant
##  9:                    2. Catholic [Roman Catholic]
## 10:                    2. Catholic [Roman Catholic]
## 11:                                    1. Protestant
## 12:                                    1. Protestant
## 13:                    2. Catholic [Roman Catholic]
## 14:                                    1. Protestant
## 15:                                    1. Protestant
## 16:                                        3. Jewish
## 17:                                    1. Protestant
## 18:                                    1. Protestant
## 19:                                    1. Protestant
## 20:                                    1. Protestant
## 21:                                    1. Protestant
## 22:                    2. Catholic [Roman Catholic]
## 23:                                    1. Protestant
## 24:                                    1. Protestant
## 25:                                    1. Protestant
## 26:                                    1. Protestant
## 27:                                    1. Protestant
## 28: 4. Other and none (also includes DK preference)
## 29:                    2. Catholic [Roman Catholic]
## 30:                                    1. Protestant
##                                              religion
##                                                  occupation
##  1:          3. Skilled, semi-skilled and service workers
##  2:                                                 <NA>
##  3:          3. Skilled, semi-skilled and service workers
##  4: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  5:                               4. Laborers, except farm
##  6: 5. Farmers, farm managers, farm laborers and foremen;
##  7:                                                 <NA>
##  8: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  9:                                                 <NA>
## 10:                       1. Professional and managerial
## 11:                                                 <NA>
## 12:                                                 <NA>
## 13:                       1. Professional and managerial
```

```
## 14: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## 15:                                                  <NA>
## 16:                          2. Clerical and sales workers
## 17:                          2. Clerical and sales workers
## 18:          3. Skilled, semi-skilled and service workers
## 19:                         1. Professional and managerial
## 20: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## 21: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## 22:          3. Skilled, semi-skilled and service workers
## 23:                                                  <NA>
## 24:                          2. Clerical and sales workers
## 25:                                                  <NA>
## 26:                          2. Clerical and sales workers
## 27:          3. Skilled, semi-skilled and service workers
## 28:                                                  <NA>
## 29:                         1. Professional and managerial
## 30:                                                  <NA>
##                                                 occupation
```

```
newQ_data %>% select(year, vote, education, religion, occupation)
```

```
## # A tibble: 12,183 x 5
##    year  vote     education                religion       occupation
##    <fct> <fct>    <fct>                    <fct>          <fct>
##  1 1952  2. Repu~ 2. High school (12 gra~ 1. Protestant  3. Skilled, semi-skill~
##  2 1952  2. Repu~ 2. High school (12 gra~ 1. Protestant  6. Homemakers (1972-19~
##  3 1952  2. Repu~ 1. Grade school or les~ 1. Protestant  6. Homemakers (1972-19~
##  4 1952  2. Repu~ 4. College or advanced~ 1. Protestant  6. Homemakers (1972-19~
##  5 1952  1. Demo~ 2. High school (12 gra~ 1. Protestant  6. Homemakers (1972-19~
##  6 1952  2. Repu~ 1. Grade school or les~ 1. Protestant  6. Homemakers (1972-19~
##  7 1952  1. Demo~ 2. High school (12 gra~ 1. Protestant  3. Skilled, semi-skill~
##  8 1952  1. Demo~ 1. Grade school or les~ 2. Catholic [~ 6. Homemakers (1972-19~
##  9 1952  2. Repu~ 2. High school (12 gra~ 1. Protestant  6. Homemakers (1972-19~
## 10 1952  1. Demo~ 2. High school (12 gra~ 1. Protestant  6. Homemakers (1972-19~
## # ... with 12,173 more rows
```

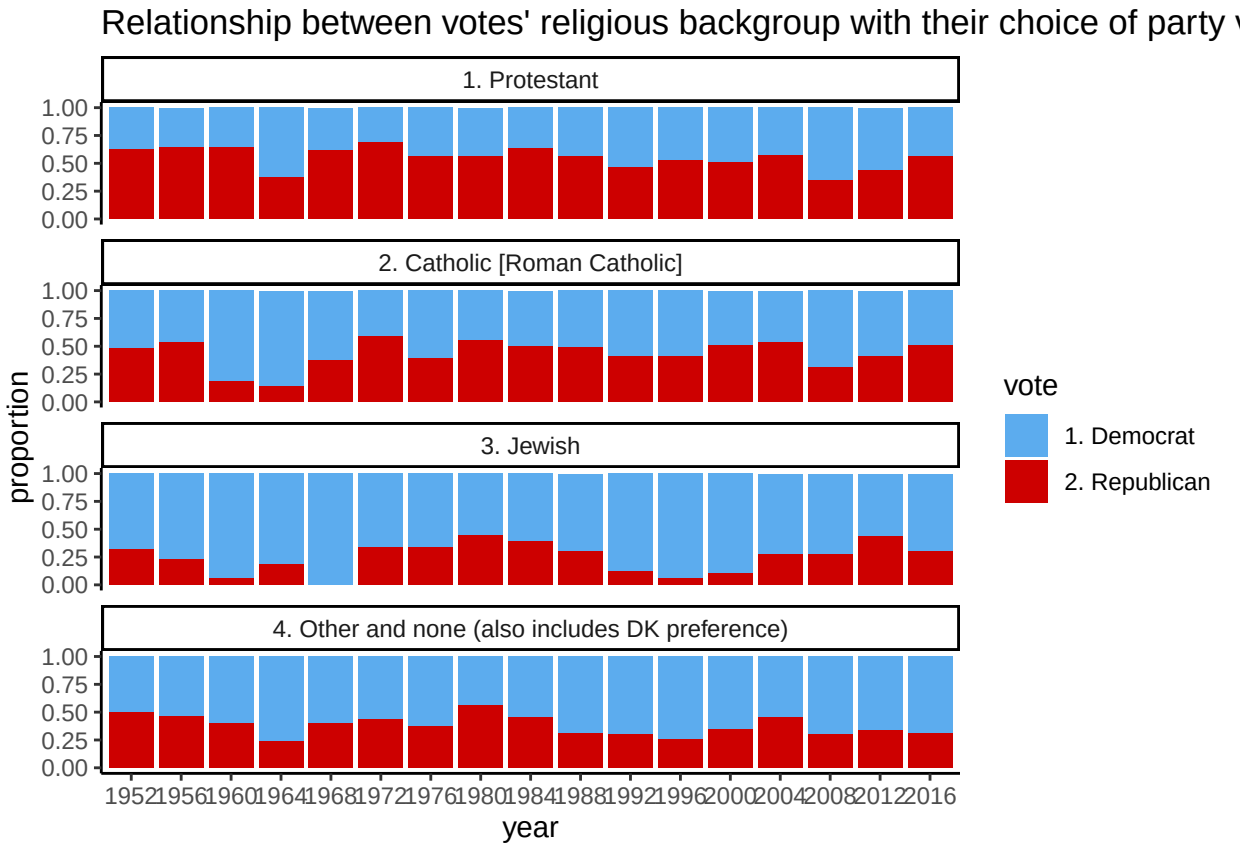
```
newQ_data <- newQ_data %>% select(year, vote, education, religion, occupation)

save(newQ_data, file = "../output/data_use.RData")
```

3. Data Visualization

Q1: Is there a relationship between voters' religious background and the party they would like to vote?

```
load(file = "../output/data_use.RData")
Q1_plotdata <- Q1_data %>%
  filter(!(str_detect(religion, "NA"))) %>%
  group_by(year, religion) %>%
  count(vote) %>%
```

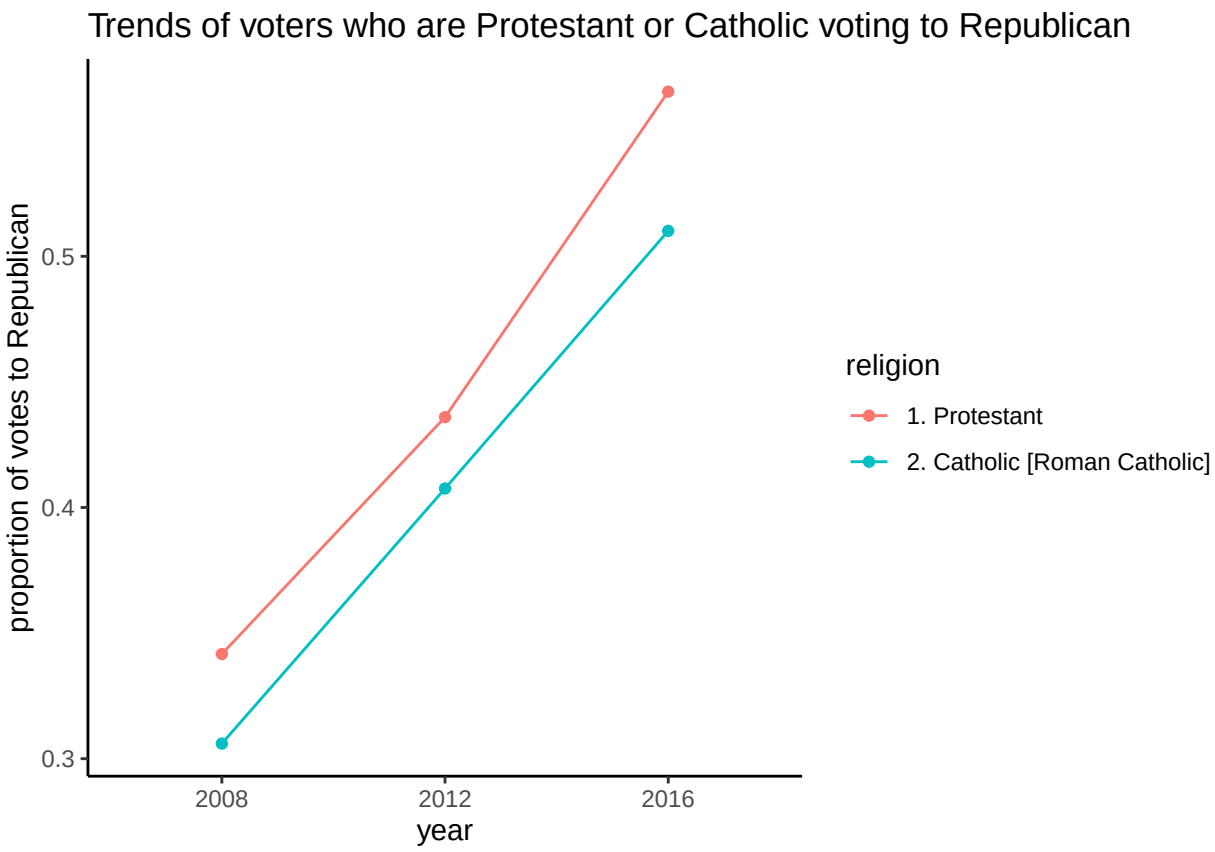
```
  group_by(year, religion) %>%
  mutate(
    prop = n/sum(n)
  )

ggplot(Q1_plotdata,
       aes(x = year, y = prop, fill = vote)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values=c("steelblue2", "red3")) +
  facet_wrap(~ religion, ncol = 1) +
  theme_classic() +
  labs(
    title = "Relationship between votes' religious backgroup with their choice of party voting",
    y = "proportion"
  )
```



By ignoring the fourth category of other and none, we focus on the three major regions in States. It is hard to find a trend in Jewish, but we could see in most of the time, the majority of voters who are Jewish are with the Democrat. Also, we would find a clearly upward trend of the proportions of voters who are Protestant or Catholic voting to Republican are increasing during the most three recent elections. Let us take a closer look of the line charts of the votes.

```
ggplot(Q1_plotdata %>%
         filter(vote == "2. Republican") %>%
         filter(year %in% c(2008, 2012, 2016)) %>%
         filter(religion %in% c("1. Protestant", "2. Catholic [Roman Catholic]")),
```

```
      aes(x = year, y = prop, group = religion, color = religion)) +
  geom_point() +
  geom_line() +
  theme_classic() +
  labs(
    title = "Trends of voters who are Protestant or Catholic voting to Republican ",
    y = "proportion of votes to Republican"
  )
```

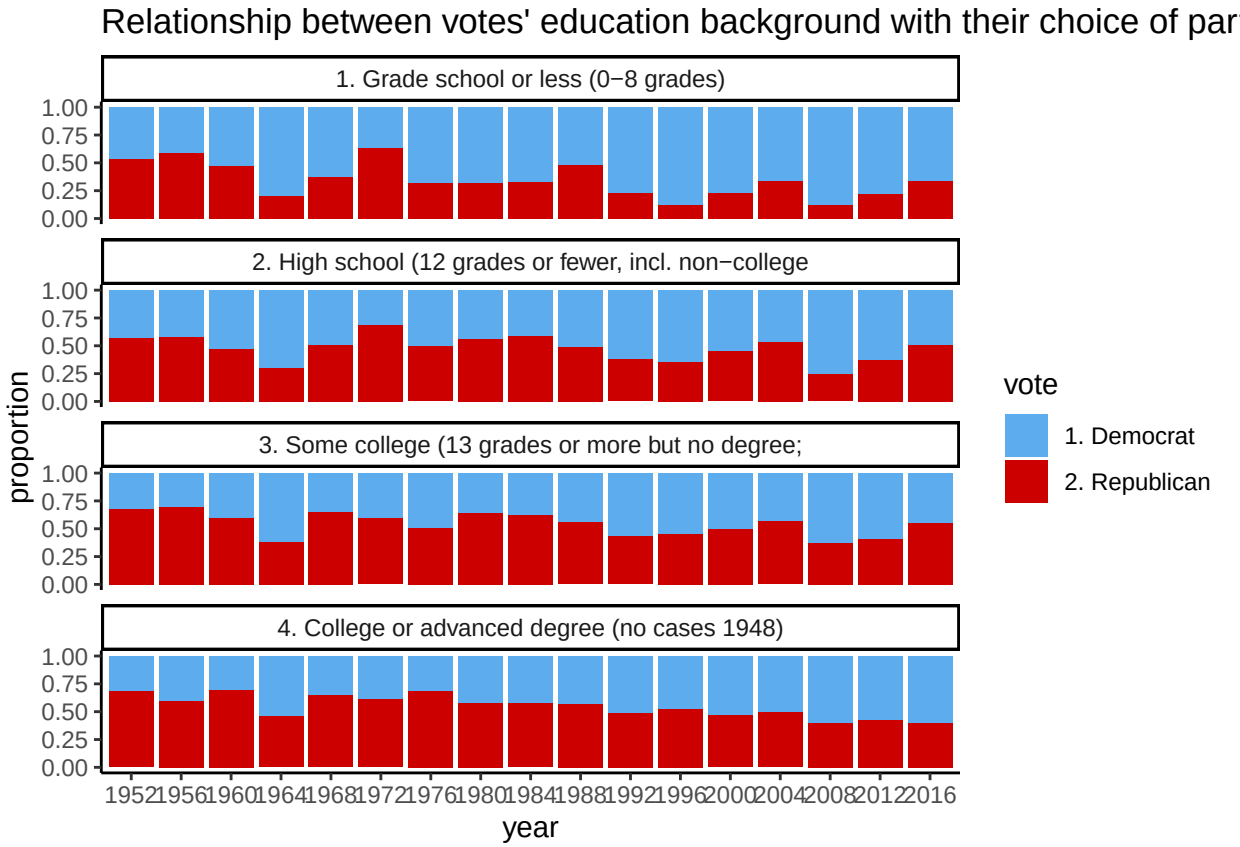


Let us put this interesting finding aside for a moment, because I am also interested in if there exist any relationship between voters' education background and the party they support.

Q2: Is there a relationship between voters' education background and the party they would like to vote?

```
Q2_plotdata <- Q2_data %>%
  filter(!(str_detect(education, "NA"))) %>%
  group_by(year, education) %>%
  count(vote) %>%
  group_by(year, education) %>%
  mutate(
    prop = n/sum(n)
  )
```

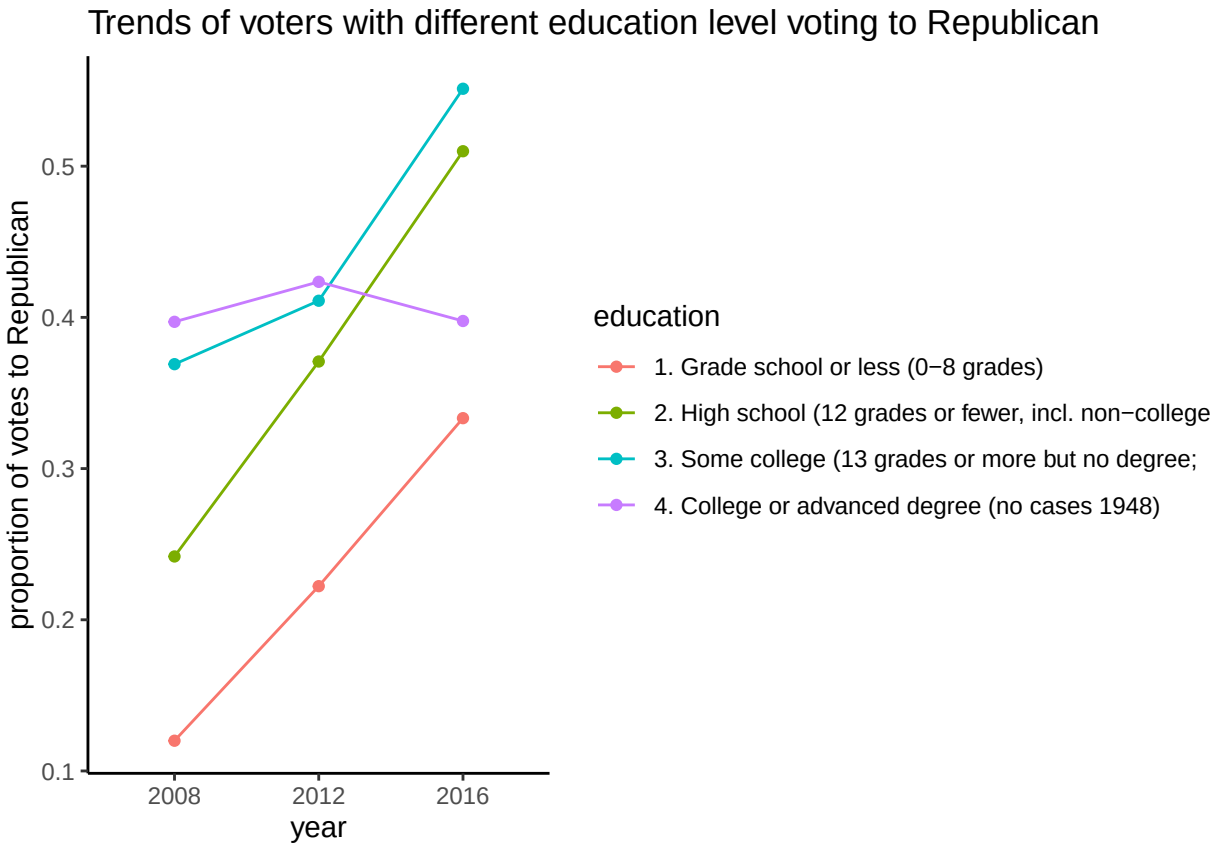
```
ggplot(Q2_plotdata,
       aes(x = year, y = prop, fill = vote)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values=c("steelblue2", "red3")) +
  facet_wrap(~ education, ncol = 1) +
  theme_classic() +
  labs(
    title = "Relationship between votes' education background with their choice of party voting",
    y = "proportion"
  )
```



By viewing the bar chart above, we cannot find a significant trend for education level under college or advanced degree. However, there is a obvious trend for the college or advanced degree, and this group of people are more and more likely to vote for the Democrat. Also, let us take a close look at the vote result in the most three recent elections.

```
ggplot(Q2_plotdata %>%
         filter(vote == "2. Republican") %>%
         filter(year %in% c(2008, 2012, 2016)),
       aes(x = year, y = prop, group = education, color = education)) +
  geom_point() +
  geom_line() +
  theme_classic() +
  labs(
    title = "Trends of voters with different education level voting to Republican ",
```

```
    y = "proportion of votes to Republican"
  )
```

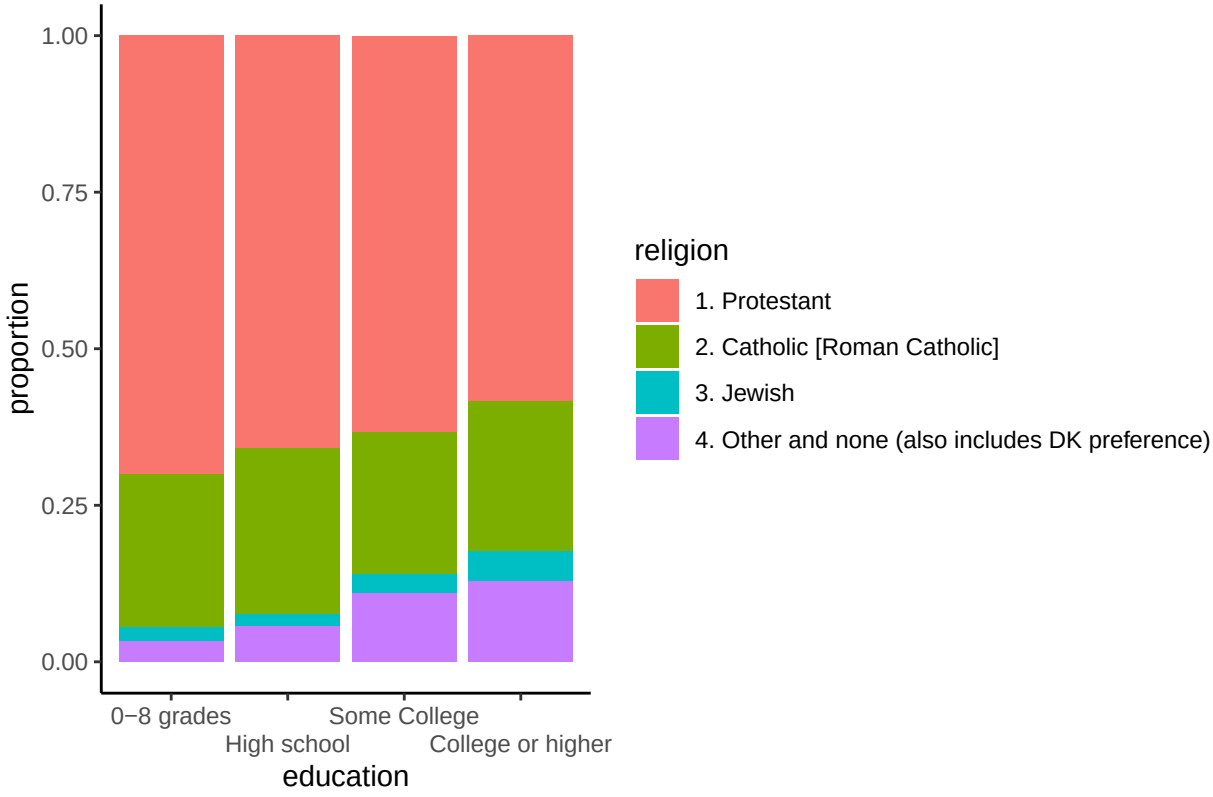


According to the line plot above, we can clearly see a upward trend of the proportion of voters voting to republican, who actually have weaker education background.

The similar upward trends of the support rating for Republican from people with weaker education background and people who are Protestant or Catholic make me wondering if there exist any correlation between these two group of people.

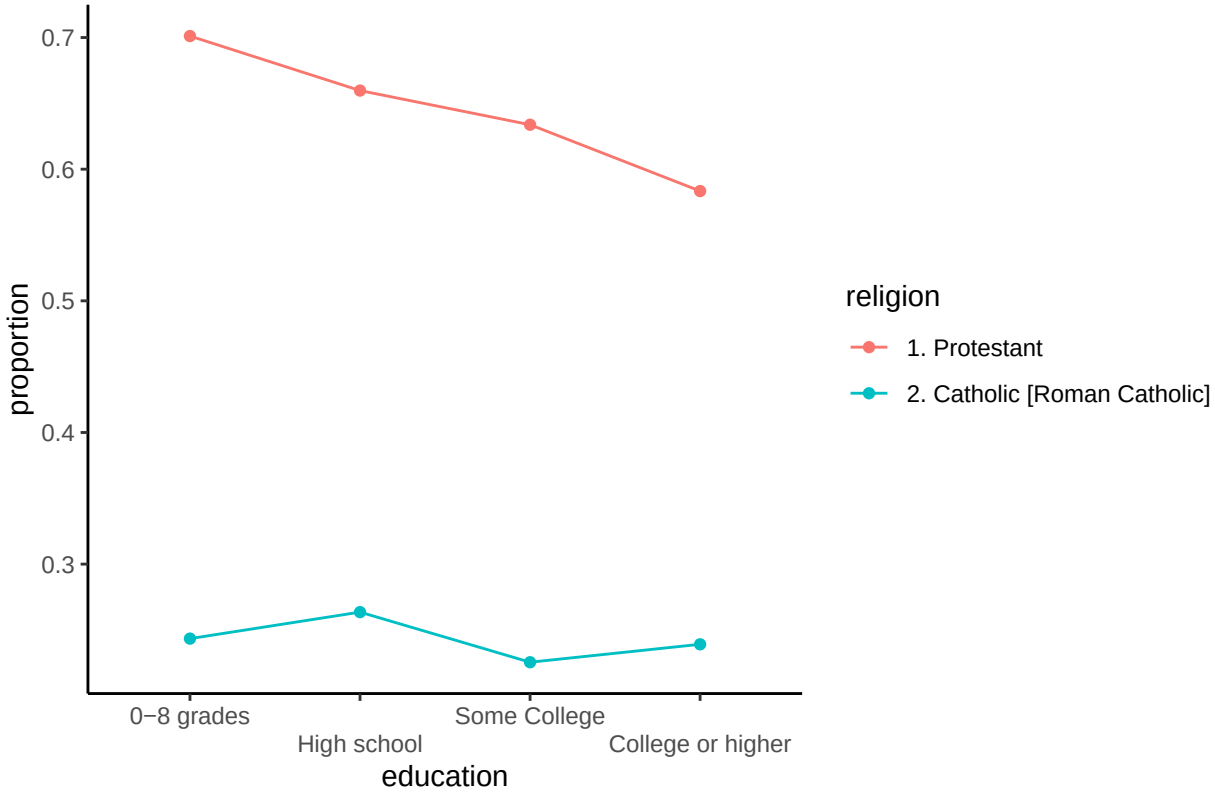
```
newQ_plotdata <- newQ_data %>%
  filter(!(str_detect(education, "NA"))) %>%
  filter(!(str_detect(religion, "NA"))) %>%
  filter(!(str_detect(occupation, "NA"))) %>%
  group_by(education) %>%
  count(religion) %>%
  group_by(education) %>%
  mutate(
    prop = n/sum(n)
  )

ggplot(newQ_plotdata,
       aes(x = education, y = prop, fill = religion)) +
  geom_bar(stat = "identity") +
  scale_x_discrete(labels = c("0-8 grades", "High school", "Some College", "College or higher"),
                   guide = guide_axis(n.dodge = 2)) +
```

```
  theme_classic() +
  labs(
    title = "",
    y = "proportion"
  )
```



Let's take a closer look of the line chart.

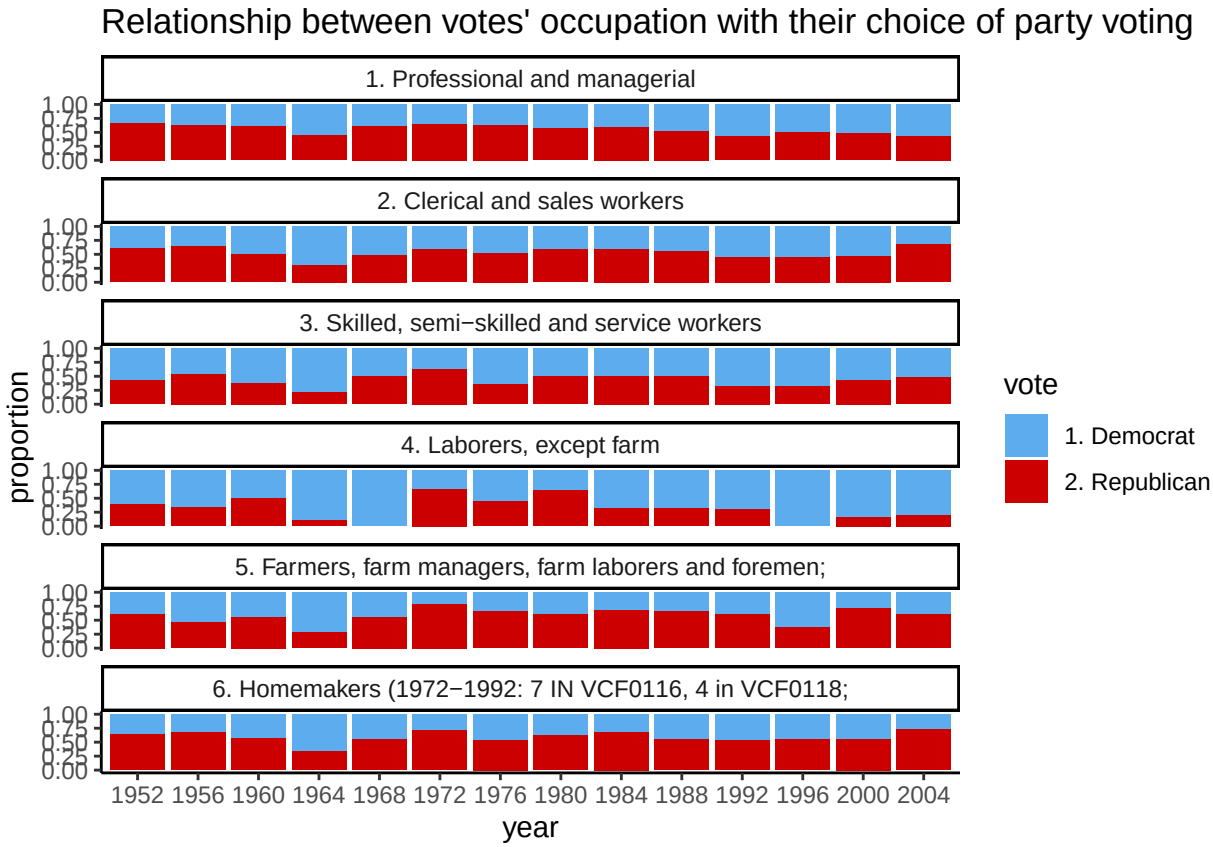
```
ggplot(newQ_plotdata %>%
         filter(religion %in% c("1. Protestant", "2. Catholic [Roman Catholic]")),
       aes(x = education, y = prop, group = religion, color = religion)) +
  geom_point() +
  geom_line() +
  scale_x_discrete(labels = c("0-8 grades", "High school", "Some College", "College or higher"),
                   guide = guide_axis(n.dodge = 2)) +
  theme_classic() +
  labs(
    title = "",
    y = "proportion"
  )
```



There is no obvious relationship between voters being Catholic and education, but we could find the obvious negative relationship between voters being a Protestant and the education background. People with weaker education background are tend to be a Protestant.

Q3: Is there a relationship between voters' occupation and the party they would like to vote?

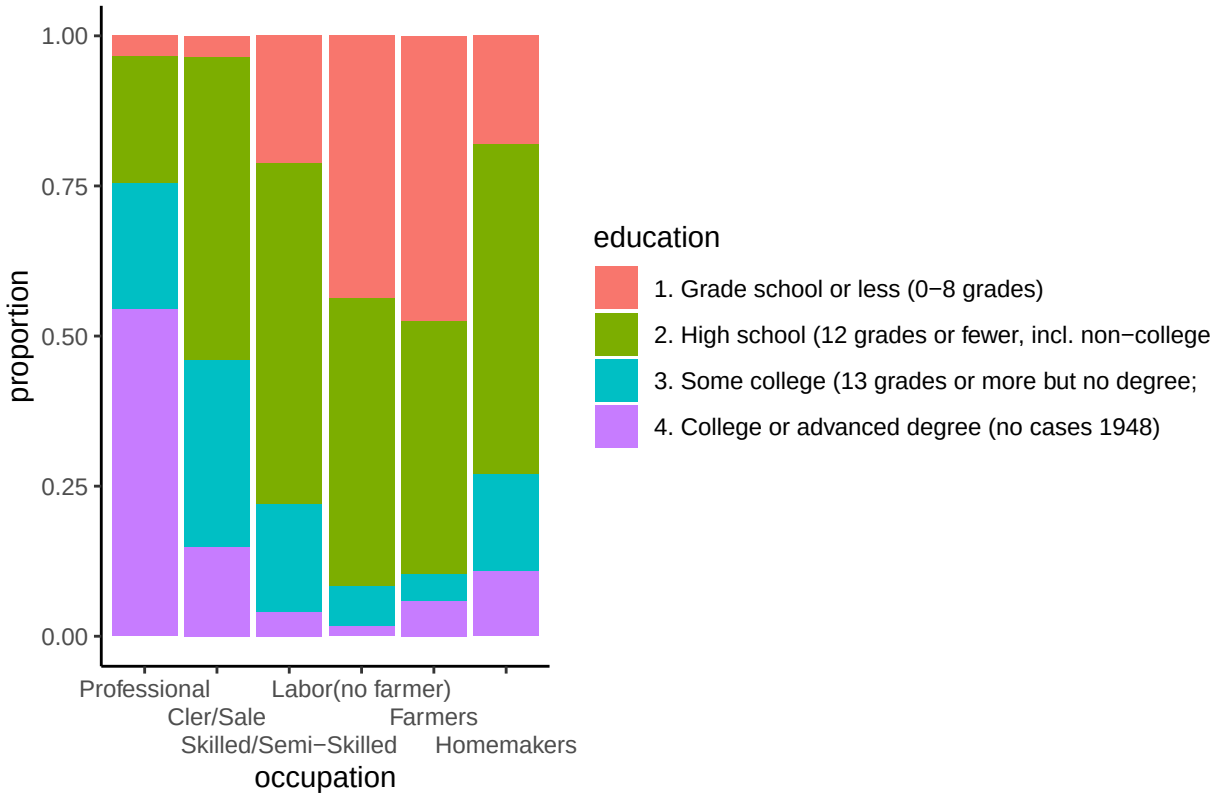
```
Q3_plotdata <- Q3_data %>%
  filter(!(str_detect(occupation, "NA"))) %>%
  group_by(year, occupation) %>%
  count(vote) %>%
  group_by(year, occupation) %>%
  mutate(
    prop = n/sum(n)
  )

ggplot(Q3_plotdata,
       aes(x = year, y = prop, fill = vote)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values=c("steelblue2", "red3")) +
  facet_wrap(~ occupation, ncol = 1) +
  theme_classic() +
  labs(
    title = "Relationship between votes' occupation with their choice of party voting",
    y = "proportion"
  )
```



By looking at the graph above, I find something interesting about the voters who are Laborers, who I expect having weaker educational background. In the previous finding, we know that people who have weaker educational background are tend to vote for Republican. Nevertheless, in this chart, Laborers are highly inclined to vote for Democrat. To validate my assumption of laborers having weaker educational background, let us construct another graph.

```
newQ_plotdata_2 <- newQ_data %>%
  filter(!(str_detect(education, "NA"))) %>%
  filter(!(str_detect(religion, "NA"))) %>%
  filter(!(str_detect(occupation, "NA"))) %>%
  group_by(occupation) %>%
  count(education) %>%
  group_by(occupation) %>%
  mutate(
    prop = n/sum(n)
  )

ggplot(newQ_plotdata_2,
       aes(x = occupation, y = prop, fill = education)) +
  geom_bar(stat = "identity") +
  scale_x_discrete(label = c("Professional", "Cler/Sale", "Skilled/Semi-Skilled", "Labor(no farmer)", "
                   guide = guide_axis(n.dodge = 3)) +
  theme_classic() +
  labs(
    title = "",
    y = "proportion"
```

```
  )
```




```
newQ_plotdata_2 %>%
  filter(occupation == "4. Laborers, except farm") %>%
  filter(education != "4. College or advanced degree (no cases 1948)") %>%
  mutate(sum_prop = sum(prop))
```

```
## # A tibble: 3 x 5
## # Groups:   occupation [1]
##   occupation          education                                    n   prop sum_prop
##   <fct>               <fct>                                    <int>  <dbl>    <dbl>
## 1 4. Laborers, excep~ 1. Grade school or less (0-8 grades)    79 0.436     0.983
## 2 4. Laborers, excep~ 2. High school (12 grades or fewer,~    87 0.481     0.983
## 3 4. Laborers, excep~ 3. Some college (13 grades or more ~    12 0.0663    0.983
```

As I expected, Laborers(except farm) are having weaker education background, and 98% of them are below college degree.

4. Conclusion

By starting with the question of if there exist a relationship between voters' religious background and the party they would like to vote. We observed that in most of the time, the majority of voters who are Jewish

are with the Democrat. At the same time, we would find a clearly upward trend of the proportions of voters who are Protestant or Catholic voting to Republican are increasing during the most three recent elections.

Then, we moved to the next question: if there exist any relationship between voters' education background and the party they support. People who have college or advanced degree are more and more likely to vote for the Democrat in the past few decades. Also, we could find a upward trend of the proportion of voters voting to republican, who actually have weaker education background. Under deeper investigation, we indeed find the obvious negative relationship between voters being a Protestant and the education background. People with weaker education background are tend to be a Protestant and are likely to vote to Republican.

Nevertheless, in our last question of investigating the relationship between voters' occupation and the party they support, we had a surprising finding. For laborers (except farm), who have weaker education background, according to our previous finding, we would expect that they like to support Republican. However, our data told us they actually like to vote for Democrat.

All of the above are the interesting findings I found through this data exploration journey.