How Did People with Different Occupations and Income Levels Vote?

Zhenglei Chen

2020/9/21

1. Introduction

I would like to use ANES Cumulative Data File to answer a question: how did people with different occupations and income levels vote? This question can be split into three major parts: 1. Which income level group participate more in voting? high income or low income? 2. Are there any party perferances between high income level people and low income people? 3. Does certain occupation type prefer to vote for a particular party?

2. Load Packages for Data Processing

3. Import raw ANES data

```
library(haven)
anes_dat <-
    read_sav("C:/Users/Charlie/Desktop/project_1/anes_timeseries/anes_timeseries_cdf.sav")
```

4 A Closer Look At Variables

4.1 Variable "Income Level"

First, let's take a look how does ANES cumulative data file interpret different income levels. In questionaires from 1948 to 2004, respondents were asked to answer the question similiar to this: About what do you think your total income will be this year for yourself and your immediate family? Then the study collected the income data and transform it into categorical variable:Income groups. *VCF0114* is the variable corresponding to income ranges(1952 - 2004). It has five classes: 1. 0 to 16 percentile 2. 17 to 33 percentile 3. 34 to 67 percentile 4. 68 to 95 percentile 5. 96 to 100 percentile

Following is the detailed graph that shows specific income ranges corresponding to percentiles for each 4 years in *VCF0114*

Respondents who were classified into 1(0 to 16 percentile) means they had the lowest income level. Respondents who were classified into 5(96 to 100 percentile) means they had the highest income level.

Let's take a look the number of respondents with their income level:

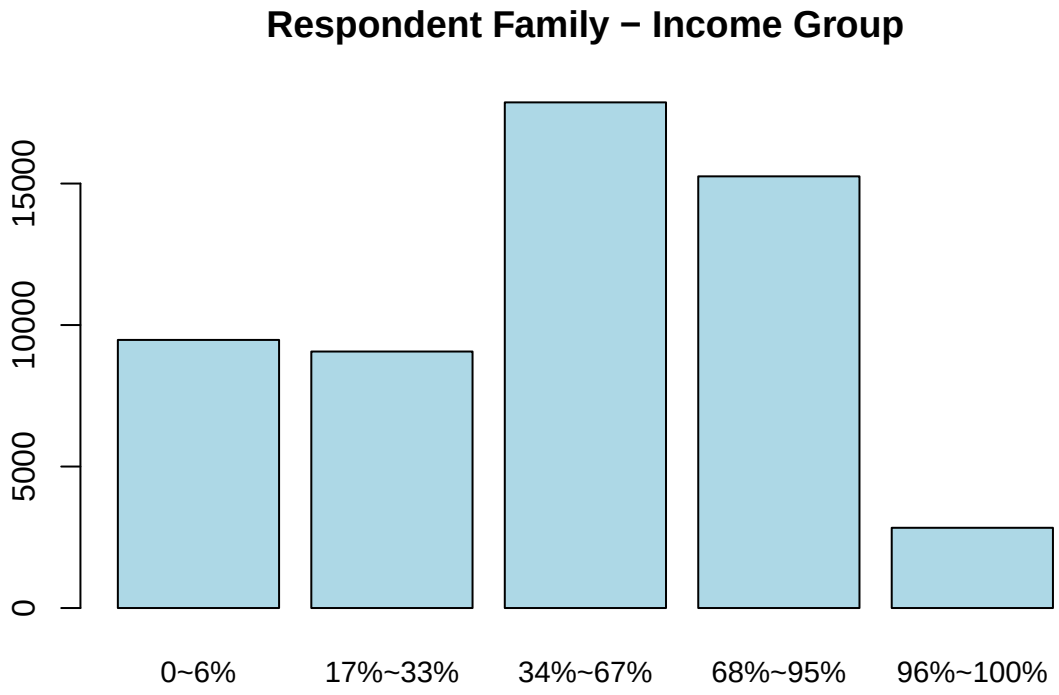
```
incomelevel = c("0~6%",
                "17%~33%",
                "34%~67%",
                "68%~95%",
                "96%~100%")
```

These are the income ranges corresponding to percentiles in VCF0114.

YEAR	0-16 PERCENTILE	17-33 PERCENTILE	34-67 PERCENTILE	68-95 PERCENTILE	96-100 PERCENTILE
1948	none-$999	$1000-1999	$2000-2999	$3000-4999	$5000+
1952	none-$1999	$2000-2999	$3000-3999	$4000-9999	$10000 +
1954	none-$1999	$2000-2999	$4000-5999	$4000-9999	$10000 +
1956	none-$1999	$2000-3999	$4000-5999	$6000-9999	$10000 +
1958	none-$1999	$2000-3999	$4000-5999	$6000-14999	$15000 +
1960	none-$1999	$2000-3999	$4000-5999	$6000-14999	$15000 +
1962	none-$2999	$3000-3999	$4000-7499	$7500-14999	$15000 +
1964	none-$2999	$3000-4999	$5000-7499	$7500-14999	$15000 +
1966	none-$2999	$3000-3999	$4000-7499	$7500-14999	$15000 +
1968	none-$2999	$3000-5999	$6000-9999	$10000-19999	$20000 +
1970	none-$2999	$3000-4999	$5000-9999	$10000-24999	$25000 +
1972	none-$3999	$4000-5999	$6000-11999	$12000-24999	$25000 +
1974	none-$3999	$4000-6999	$7000-14999	$15000-34999	$35000 +
1976	none-$3999	$4000-7999	$8000-14999	$15000-34999	$35000 +
1978	none-$5999	$6000-10999	$11000-19999	$20000-34999	$35000 +
1980	none-$6999	$7000-11999	$12000-24999	$25000-49999	$50000 +
1982	none-$6999	$7000-12999	$13000-24999	$25000-49999	$50000 +
1984	none-$6999	$7000-12999	$13000-29999	$30000-59999	$60000 +
1986	none-$8999	$9000-14999	$15000-34999	$35000-74999	$75000 +
1988	none-$9999	$10000-14999	$15000-34999	$35000-89999	$90000 +
1990	none-$9999	$10000-16999	$17000-34999	$35000-89999	$90000 +
1992	none-$9999	$10000-19999	$20000-39999	$40000-89999	$90000 +
1994	none-$11999	$12000-21999	$22000-44999	$45000-104999	$105000 +
1996	none-$11999	$12000-21999	$22000-49999	$50000-104999	$105000 +
1998	none-$8999	$9999-21999	$22000-49999	$50000-104999	$105000 +
2000	none-$14999	$15000-34999	$35000-64999	$65000-124999	$125000 +
2004	none-$16999	$17000-34999	$35000-69999	$70000-119999	$120000 +

Figure 1: A caption

```r
barplot(table(anes_dat$VCF0114), names = incomelevel,

        col = "lightblue",
        cex.names=0.9,
        main = "Respondent Family - Income Group")
```



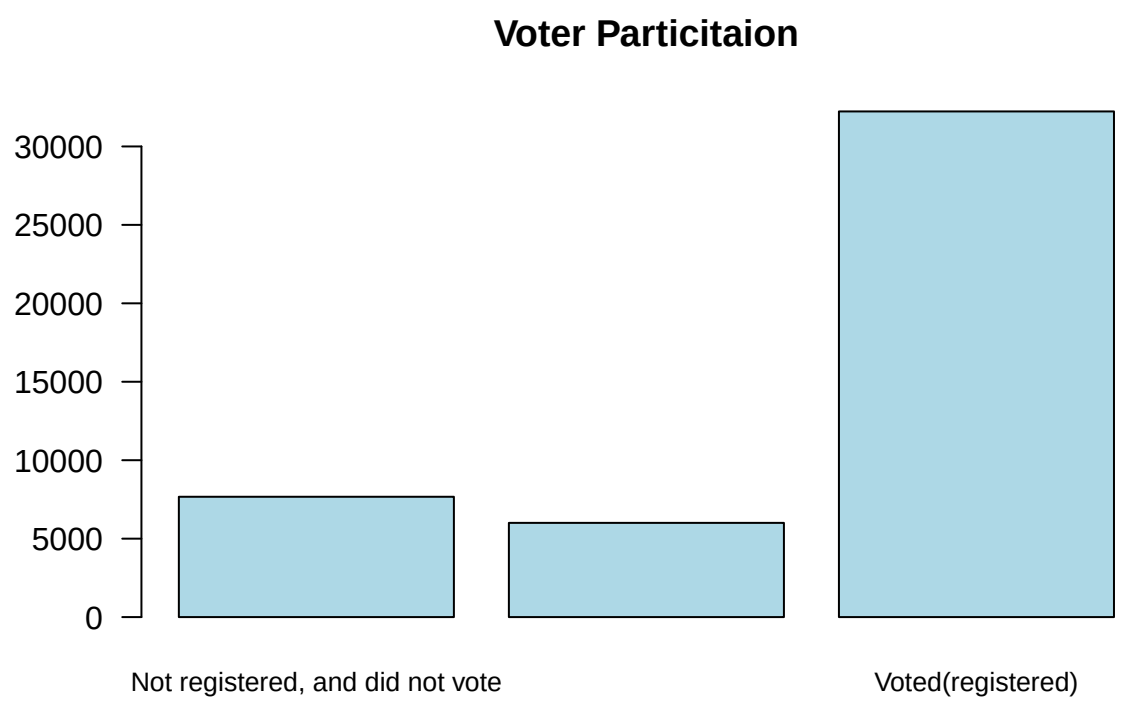
It seems that income levels are a little bit left-skewed. 34 to 67 percentile has the the highest number of people. The number of respondents who belonged to the top 4% income level is much smaller than other groups.

4.2 Variable "Voter Particitaion"

VCF0703 is the variable that can easily tell what percentage of respondents registered and voted and what percentage did not. It has three classes: 1. Not registered, and did not vote 2. Registered, but did not vote 3. Voted(registered)

followed is the graph shows the number of respondents in each class

```r
vote_turnout = c("Not registered, and did not vote",
            "Registered, but did not vote",
            "Voted(registered)")
barplot(table(anes_dat$VCF0703), names = vote_turnout,
        las=1,
        col = "lightblue",
        cex.names=0.8,
        main="Voter Particitaion")
```



4.3 Variable "Occupation Group"

VCF0115 is the variable that reveals respondents' occupations, it has six classes: 1. Professional and managerial 2. Clerical and sales workers 3. Skilled, semi-skilled and service workers 4. Laborers, except farm 5. Farmers, farm managers, farm laborers and foremen; forestry and fishermen 6. Homemakers

followed is the graph shows the number of respondents in each class

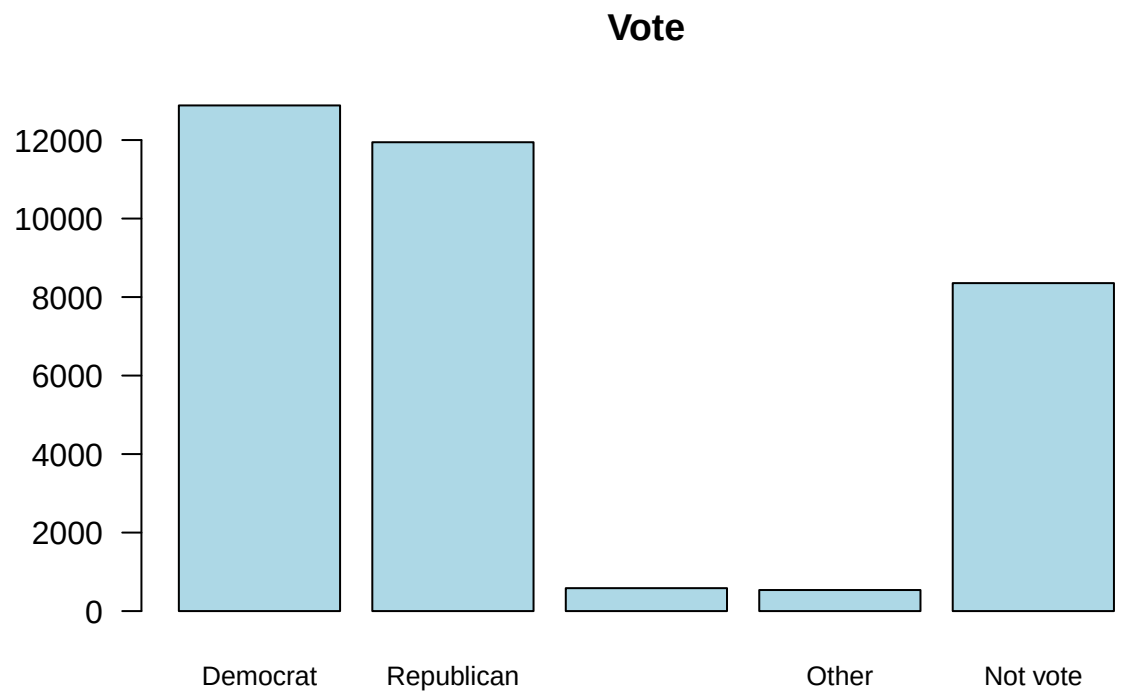
```
barplot(table(anes_dat$VCF0115),
        las=1,
        col = "lightblue",
        main="Occupation Group")
```

4.4 Variable "Vote"

VCF0706 is the variable that reveals respondents' choices of presidential vote and nonvote, it has five classes: 1. Democrat 2. Republican 3. Major third party candidate 4. Other 7. Did not vote or voted but not for president

followed is the graph shows the number of respondents in each class

```
vote_choice = c("Democrat",
                "Republican",
                "Third party candidate",
                "Other",
                "Not vote")
barplot(table(anes_dat$VCF0706), names = vote_choice,
        las=1,
        col = "lightblue",
        cex.names=0.8,
        main="Vote")
```



5. Data processing

5.1 Process variables for analysis

First, I selected five variables for further analysis and filtered data was saved to the `output` folder.

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

ANES_data = anes_dat%>%
  mutate(
    year = as_factor(VCF0004),
    income = as_factor(VCF0114),
    turnout = as_factor(VCF0703),
    occupation = as_factor(VCF0115),
    vote=as_factor(VCF0706)
  )%>%
  filter(year %in% Presidential_Election_Year)

library(data.table)

data.table(ANES_data%>%
             select(year, income, turnout, occupation, vote)%>%
             filter(!is.na(turnout))%>%
             sample_n(20))
```

```
##     year                 income                                turnout
##  1: 1952 3. 34 to 67 percentile                3. Voted (registered)
##  2: 2012 3. 34 to 67 percentile                3. Voted (registered)
##  3: 1988  1. 0 to 16 percentile                3. Voted (registered)
##  4: 1980 4. 68 to 95 percentile     2. Registered, but did not vote
##  5: 1952 4. 68 to 95 percentile                3. Voted (registered)
##  6: 2012  1. 0 to 16 percentile                3. Voted (registered)
##  7: 1952 4. 68 to 95 percentile                3. Voted (registered)
##  8: 1964 4. 68 to 95 percentile                3. Voted (registered)
##  9: 2016  1. 0 to 16 percentile                3. Voted (registered)
## 10: 2008 2. 17 to 33 percentile                3. Voted (registered)
## 11: 2016 4. 68 to 95 percentile 1. Not registered, and did not vote
## 12: 2016 3. 34 to 67 percentile     2. Registered, but did not vote
## 13: 1992 3. 34 to 67 percentile                3. Voted (registered)
## 14: 2000                   <NA>     2. Registered, but did not vote
## 15: 1976 3. 34 to 67 percentile 1. Not registered, and did not vote
## 16: 1976 3. 34 to 67 percentile     2. Registered, but did not vote
## 17: 2016  1. 0 to 16 percentile                3. Voted (registered)
## 18: 2012 3. 34 to 67 percentile                3. Voted (registered)
## 19: 1976 4. 68 to 95 percentile                3. Voted (registered)
## 20: 1984  1. 0 to 16 percentile     2. Registered, but did not vote
##                                                 occupation
##  1:                               4. Laborers, except farm
##  2:                                                   <NA>
##  3:          3. Skilled, semi-skilled and service workers
##  4: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  5:          3. Skilled, semi-skilled and service workers
##  6:                                                   <NA>
##  7: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  8: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
##  9:                                                   <NA>
## 10:                                                   <NA>
## 11:                                                   <NA>
## 12:                                                   <NA>
## 13: 5. Farmers, farm managers, farm laborers and foremen;
## 14: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## 15: 6. Homemakers (1972-1992: 7 IN VCF0116, 4 in VCF0118;
## 16:                          2. Clerical and sales workers
## 17:                                                   <NA>
## 18:                                                   <NA>
## 19:                          2. Clerical and sales workers
## 20:          3. Skilled, semi-skilled and service workers
##                                                       vote
##  1:                                             1. Democrat
##  2:                                           2. Republican
##  3:                                           2. Republican
##  4: 7. Did not vote or voted but not for president (exc.1972)
##  5:                                           2. Republican
##  6:                                           2. Republican
##  7:                                           2. Republican
##  8:                                             1. Democrat
##  9: 7. Did not vote or voted but not for president (exc.1972)
## 10:                                             1. Democrat
## 11:                                                    <NA>
```

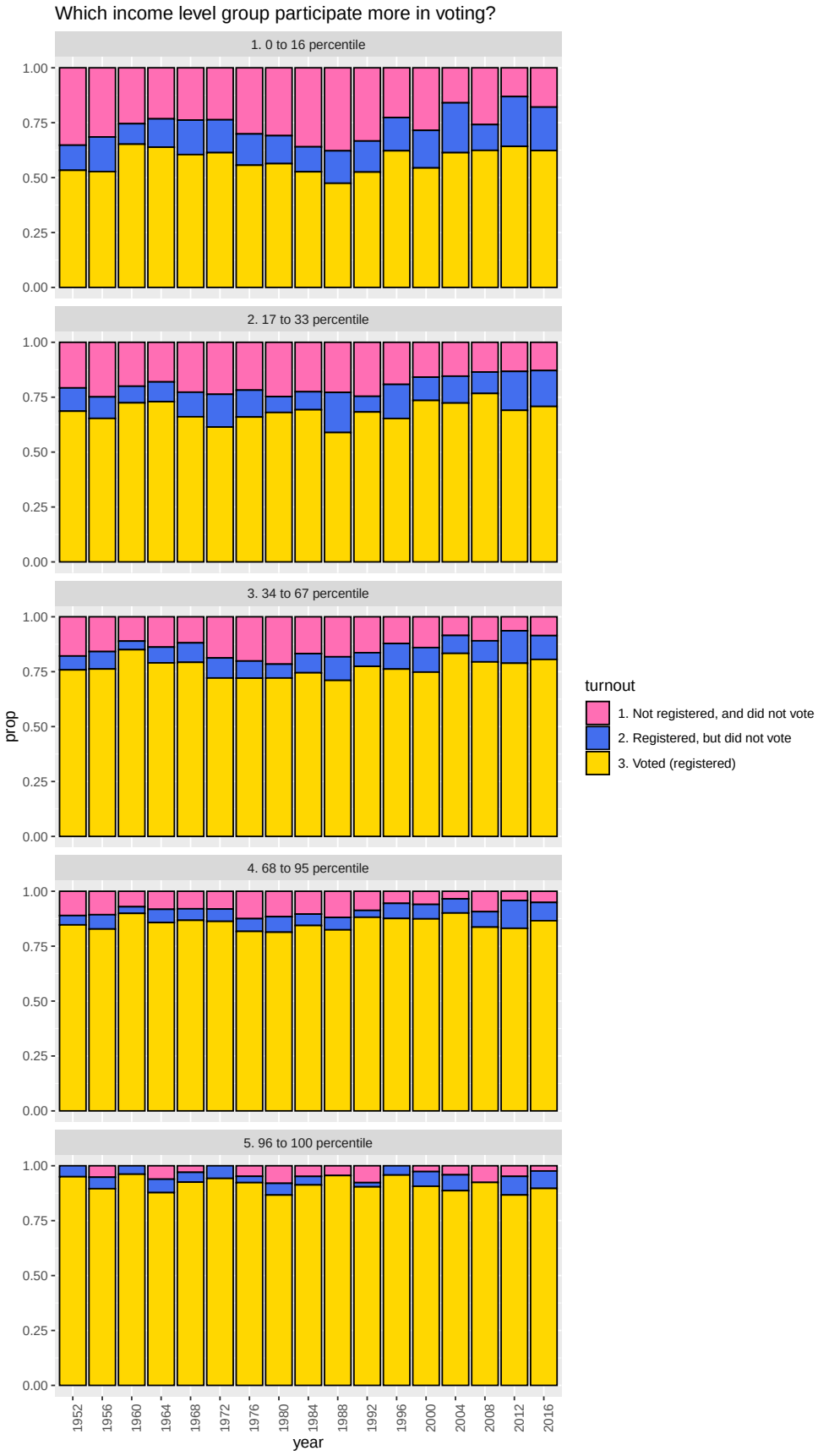
```
## 12: 7. Did not vote or voted but not for president (exc.1972)
## 13:     3. Major third party candidate (Wallace 1968/Anderson
## 14: 7. Did not vote or voted but not for president (exc.1972)
## 15: 7. Did not vote or voted but not for president (exc.1972)
## 16: 7. Did not vote or voted but not for president (exc.1972)
## 17:                                                 1. Democrat
## 18:                                                 1. Democrat
## 19:                                                 1. Democrat
## 20: 7. Did not vote or voted but not for president (exc.1972)
```

```
ANES_data = ANES_data%>%select(year, income, turnout, occupation, vote)

save(ANES_data, file="C:/Users/Charlie/Documents/GitHub/Fall2020-Project1-charliechen1995/output/data_us
```

5.2 Answer the Question: Which income level group participate more in voting? high income or low income?

Which group tends to participate more in voting? I might guess home income group, is that true, let's find out:

```
anes_income_turnout_year = ANES_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_year,
       aes(x=year, y=prop, fill = turnout)) +
  geom_bar(stat="identity", colour="black")+
  scale_fill_manual(values=c("hotpink1", "royalblue2", "gold1"))+
  facet_wrap(~income, ncol=1) +
  theme(axis.text.x = element_text(angle = 90))+
  labs(title="Which income level group participate more in voting?")
```



The result is quite si-

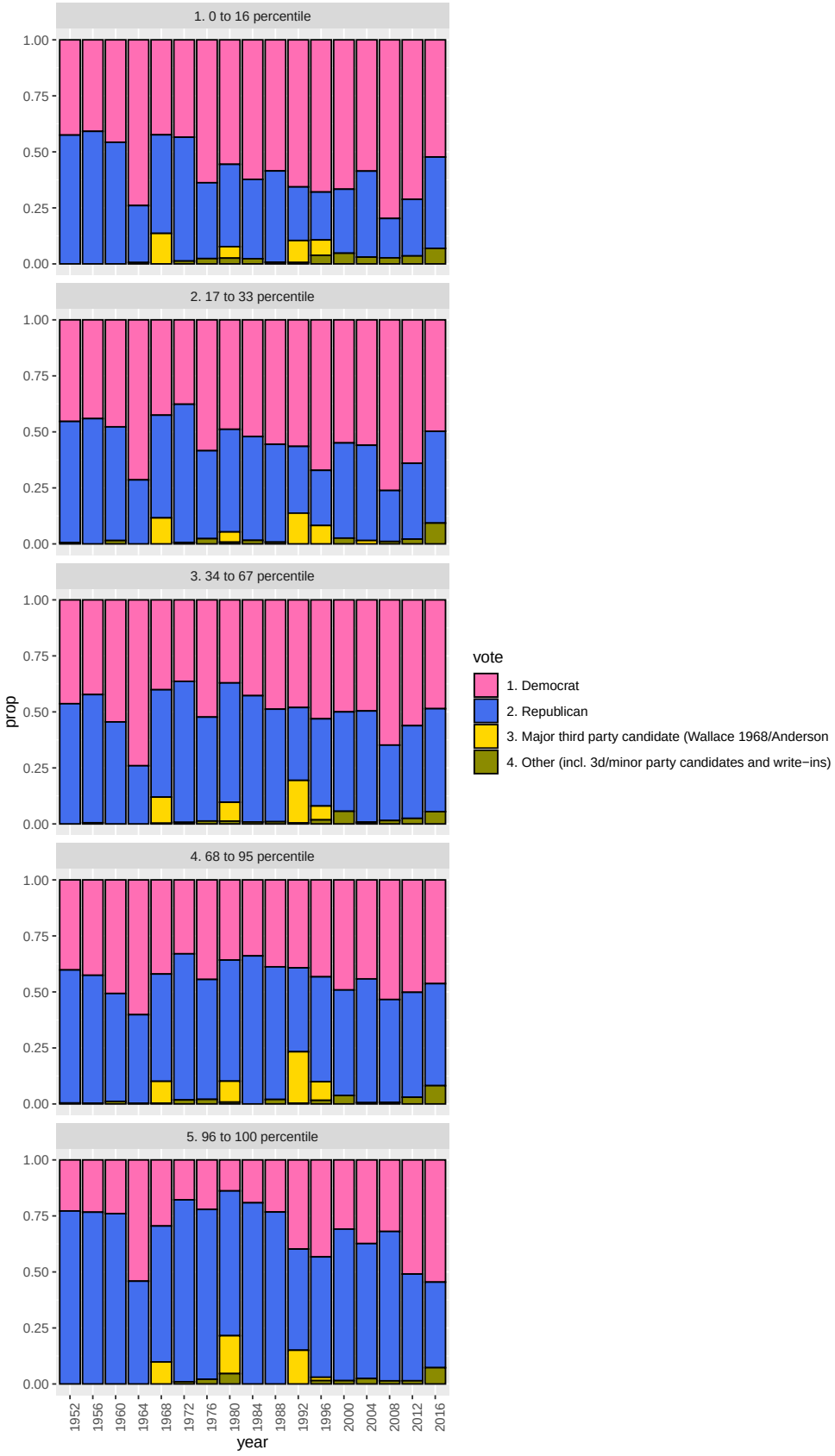
miliar to what I just guessed. We can see that high income groups vote more while low income groups tend to vote less. We can conclude that high level participate in voting more. However, we can see from the first plot (0 to 16 percentile), still more than half respondents registered and voted.

5.3 Answer the Question: Are there any party perferances between high income level people and low income people?

I really want to know that did rich people and poor people vote for the same party or not? Are there any party perferances between high income level people and low income people? Let's see the result:

```
anes_vote_income_year = ANES_data %>%
  filter(!is.na(income) & !is.na(vote))%>%
  filter(vote!="7. Did not vote or voted but not for president (exc.1972)")%>%
  group_by(year,income)%>%
  count(vote)%>%
  group_by(year, income)%>%
  mutate(
    prop=n/sum(n)
  )
#%>%
#  filter(vote == "1. Democrat" | vote == "2. Republican")

ggplot(anes_vote_income_year,
       aes(x=year, y=prop, fill=vote)) +
  geom_bar(stat="identity", colour="black")+
  scale_fill_manual(values=c("hotpink1", "royalblue2", "gold1", "yellow4"))+
  facet_wrap(~income, ncol=1) +
  theme(axis.text.x = element_text(angle = 90))+
  labs(title="Who did different income groups vote for in the election over the years?")
```

Who did different income groups vote for in the election over the years?



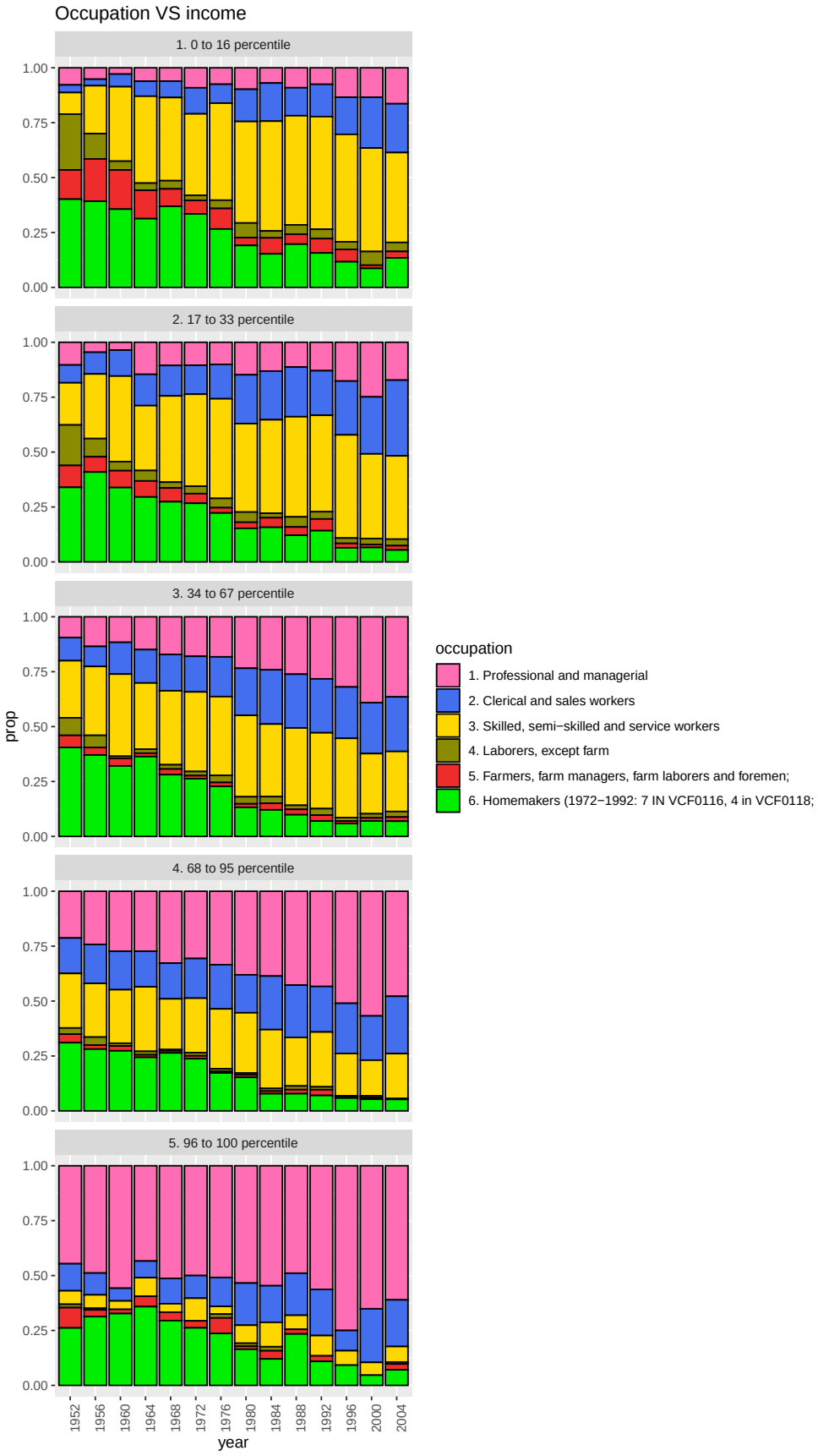
As we can see from

the plots, lowest income level group (0 to 16 percentile) and the second lowest income level group (17 to 33 percentile) were more likely to vote for Democrat Party. However, High income level respondents (96 to 100 percentile) voted for Republican Party more. Respondents with middle income levels did not have outstanding perferances.

5.4 Answer the Question: Does certain occupation type prefer to vote for a particular party?

Since we already know that different income level groups behave quite differently in voting, what about different occupations? Does certain occupation type prefer to vote for a particular party? First, I want to know whether occupations are associated with different income levels. There are several plots shows the result:

```
anes_occupation_income_year = ANES_data %>%
  filter(!is.na(income) & !is.na(occupation))%>%
  group_by(year,income)%>%
  count(occupation)%>%
  group_by(year, income)%>%
  mutate(
    prop=n/sum(n)
  )

ggplot(anes_occupation_income_year,
       aes(x=year, y=prop, fill=occupation)) +
  geom_bar(stat="identity", colour="black")+
  scale_fill_manual(values=c("hotpink1", "royalblue2", "gold1", "yellow4", "firebrick2", "green2"))+
  facet_wrap(~income, ncol=1) +
  theme(axis.text.x = element_text(angle = 90))+
  labs(title="Occupation VS income")
```



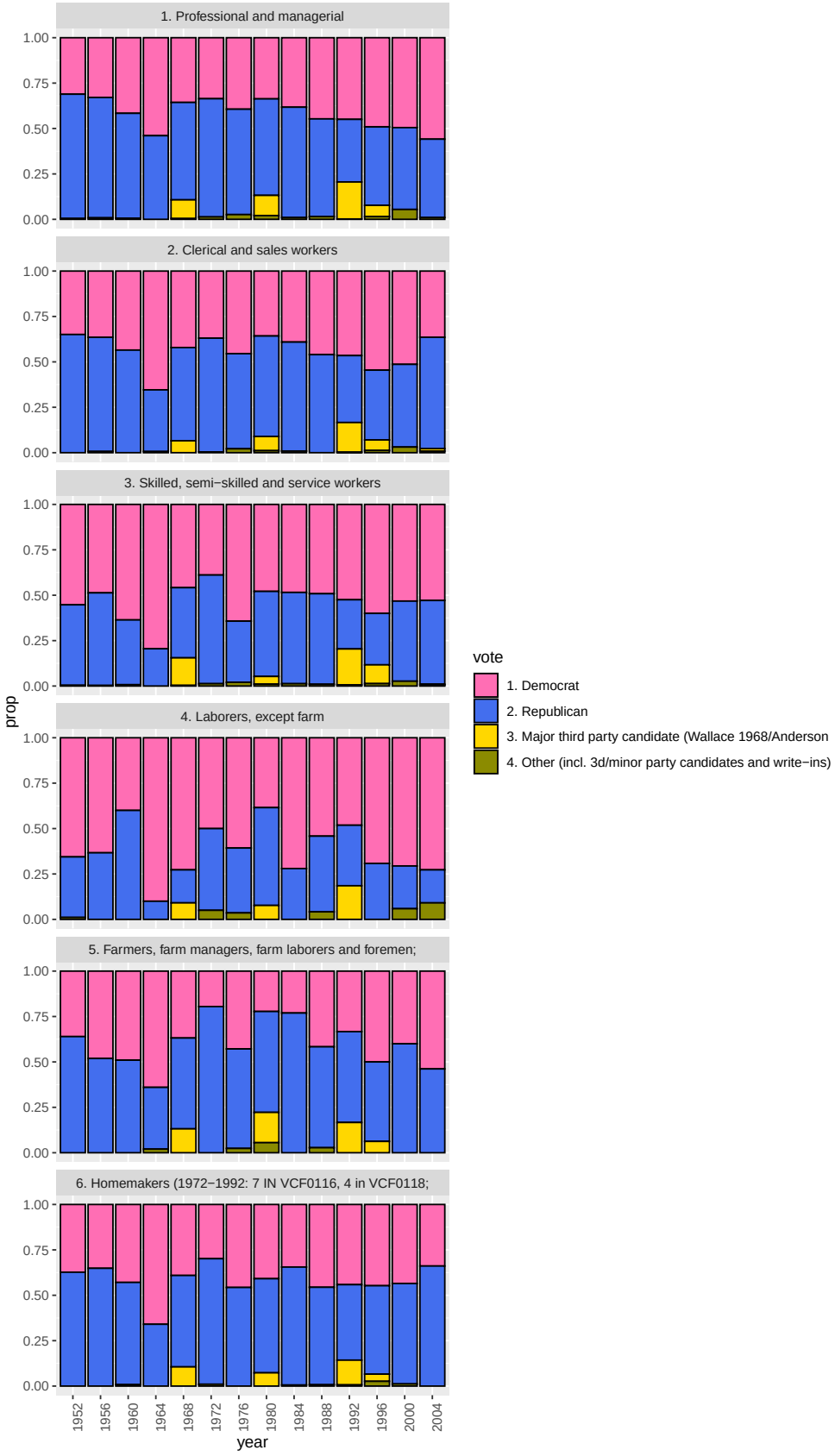
We can see from the

result that large proportion of highest income group (96 to 100 percentile) is Professional and managerial, however there is only small proportion in lowest income group (0 to 16 percentile). As income increases, proportion of categoria 1 (Professional and managerial) goes up. We can also see that large proportion of lowest income group (0 to 16 percentile) is categoria 3 (Skilled, semi-skilled and service workers), however there is only small proportion in highest income group (96 to 100 percentile). As income increases, proportion of categoria 3 (Skilled, semi-skilled and service workers) goes down.

Now, let's take a look what's the relationship between occupation groups and party perferances:

```
anes_occupation_vote_year = ANES_data %>%
  filter(!is.na(occupation) & !is.na(vote))%>%
  filter(vote!="7. Did not vote or voted but not for president (exc.1972)")%>%
  group_by(year,occupation)%>%
  count(vote)%>%
  group_by(year, occupation)%>%
  mutate(
    prop=n/sum(n)
  )
#%>%
#  filter(vote == "1. Democrat" | vote == "2. Republican")

ggplot(anes_occupation_vote_year,
       aes(x=year, y=prop, fill=vote)) +
  geom_bar(stat="identity", colour="black")+
  scale_fill_manual(values=c("hotpink1", "royalblue2", "gold1", "yellow4"))+
  facet_wrap(~occupation, ncol=1) +
  theme(axis.text.x = element_text(angle = 90))+
  labs(title="Does certain occupation type prefer to vote for a particular party")
```

Does certain occupation type prefer to vote for a particular party



We can see that cate-

gorial 3 (skilled, semi_skilled and service workers) and categorial 4 (laborers, except farmers) have clearly preferance on Democrat Party. Categorial 2 (clerical and sales workers), categorial 5 (farmers, farm managers, farm laboers and foremen), categorial 6 (homemakers) tend to vote for Republic Party more. Categorial 1 (Professional and managerial) tends to vote for Republican Party more before 1988, but after that, they prefer Democrat more and more.

6. Conclusion

1. High income groups participated more in voting while low income groups participated less in voting.
2. Lowest income level group (0 to 16 percentile) and the second lowest income level group (17 to 33 percentile) were more likely to vote for Democrat Party. High income level respondents (96 to 100 percentile) voted for Republican Party more. Respondents with middle income levels did not have outstanding perferances.
3. Categorial 3 (skilled, semi_skilled and service workers) and categorial 4 (laborers, except farmers) had clearly preferance on Democrat Party. Categorial 2 (clerical and sales workers), categorial 5 (farmers, farm managers, farm laboers and foremen), categorial 6 (homemakers) preferred to vote for Republic Party. Categorial 1 (Professional and managerial) liked to vote for Republican Party more before 1988, but after 1988, they preferred Democrat more and more.