

# project1

Yunuo Ma

9/21/2020

## 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's data and results have been routinely used by news outlets, election campaigns and political researchers.

The *Time Series Cumulative Data* of ANES include answers, from respondents from different years, on selected questions that have been asked in three or more ANES' *Time Series* studies. Tremendous amount of efforts have been put into data consolidation as variables are often named differently in different years.

A rule of thumb for analyzing any data set is to understand its study design and data collection process first. You are strongly encouraged to read the *codebooks*.

## 2. Data processing for this R Notebook.

The following code blocks prepare a processed data set and save it in the `output` folder. The data folder should be only used for storing raw data. All processed data should be saved in the `output` folder. The notion here is that one can delete files from the output folder and reproduce them by re-running the codes.

### Step 3.1 Checking R packages for data processing

From the packages' descriptions:

- `tidyverse` is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures;
- `haven` enables R to read and write various data formats used by other statistical packages. `haven` is part of the `tidyverse`.
- `devtools` provides a collection of package development tools.
- `RColorBrewer` provides ready-to-use color palettes.
- `DT` provides an R interface to the JavaScript library DataTables;
- `ggplot2` a collection of functions for creating graphics, based on The Grammar of Graphics.

### Step 3.2 Import raw ANES data

Working with the DTA format of the raw ANES data, downloaded from this page.

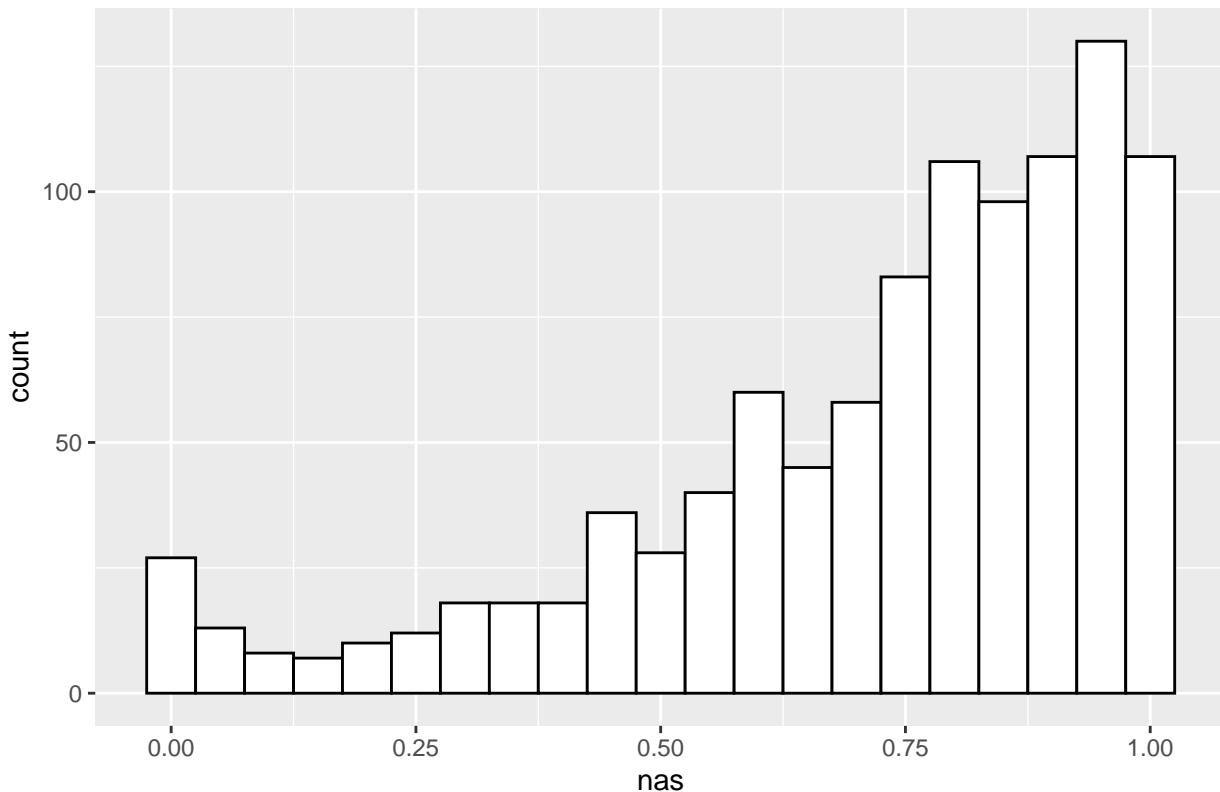
```
anes_dat <- read_dta("../data/anes_timeseries_cdf.dta")
dim(anes_dat)
```

```
## [1] 59944 1029
```

This data contains 59944 rows with 1029 columns.

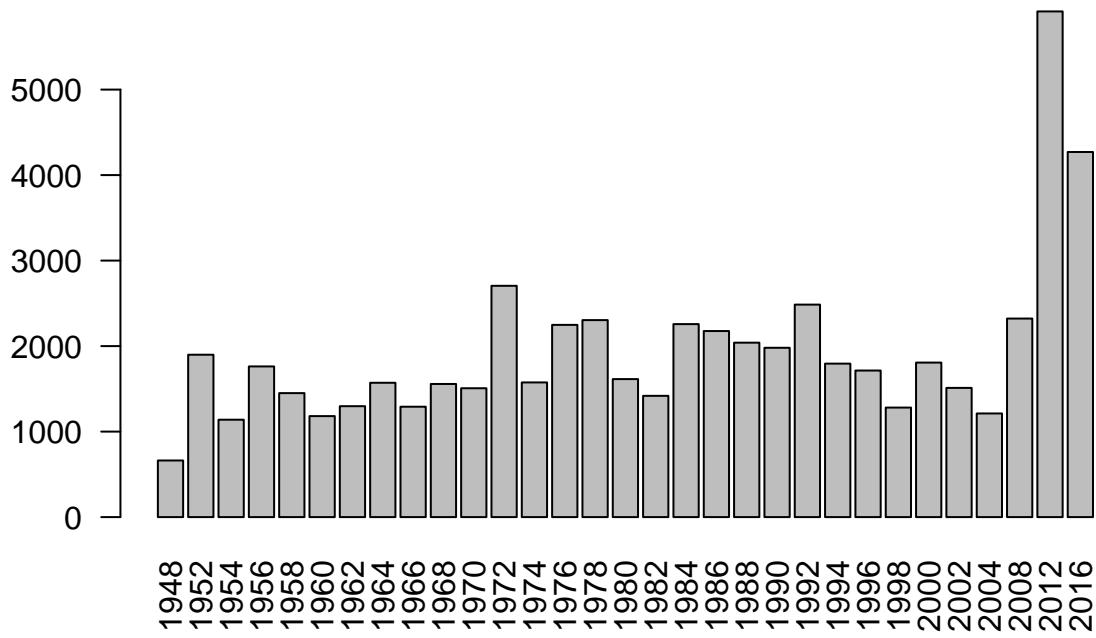
```
anes_NAs=anes_dat%>%
  summarise_all(list(na.mean=function(x){
    mean(is.na(x))}))
anes_NAs=data.frame(nas=unlist(t(anes_NAs)))
ggplot(anes_NAs, aes(x=nas)) +
  geom_histogram(color="black",
                 fill="white",
                 binwidth=0.05) +
  labs(title="Fractions of missing values")
```

## Fractions of missing values



```
barplot(table(anes_dat$VCF0004),
       las=2,
       main="number of respondents over the years")
```

## number of respondents over the years



Some variables are asked nearly all the years and some are asked only a few years.

### Step 3.3 Process variables for analysis

Some variables were selected based on their description in the ANES codebook.

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

anes = anes_dat %>%
  mutate(year = as_factor(VCF0004), #0 NA
        turnout = as_factor(VCF0703), #4903 NA
        #vote = as_factor(VCF0706), #4896 NA
        region = as_factor(VCF0112), #0 NA
        income = as_factor(VCF0114), #2517 NA
        work = as_factor(VCF0151), #13162 NA
        education = as_factor(VCF0110), #398 NA
        race = as_factor(VCF0105a), #287 NA
        religion = as_factor(VCF0128), #333 NA
        gender = as_factor(VCF0104), #141 NA
        # PARTISANSHIP VARIABLE
        partisanship_strength = as_factor(VCF0305), #1169 NA
        intended_actual_votes = as_factor(VCF0734), #2472 NA
        care_party_win = as_factor(VCF0311), #26115 NA #missing 2016
        # INFLUENCE VARIABLE
        try_influence = as_factor(VCF0717), #6373 NA
        days_discuss = VCF0733, #33342 NA
        #COSIDERED ELECTION RESULT
        considered_result = as_factor(VCF0700), #27600 NA #missing 2012
        # INTERESTED
        interest = as_factor(VCF0310)
```

```

) %>%

select(year, turnout, region, income,
       work, education, race, religion, gender,
       partisanship_strength, intended_actual_votes,
       care_party_win, try_influence,
       days_discuss, considered_result,
       interest) %>%
filter(year %in% Election_years) %>%
replace_na(list(days_discuss = mean(na.omit(anes_dat$VCF0733)))) %>%
na.omit()

#change region factor levels
anes$region = as.factor(as.character(anes$region))
levels(anes$region) <- c("Northeast", "North Central", "West", "South")

#deleted rows with ambiguous meaning of intended_actual_votes variable
l=levels(anes$intended_actual_votes)
index <- c(which(anes$intended_actual_votes == l[5]),
            which(anes$intended_actual_votes == l[6]),
            which(anes$intended_actual_votes == l[7]))
anes = anes[-index,]

#add intend and actual variables corresponding to
#intended party to vote and actual party to vote
anes = anes %>%
  mutate(intend = substring(as.character(intended_actual_votes), 13,22),
         actual = substring(as.character(intended_actual_votes), 31,40))
anes$intend = gsub("undecided:", "others", anes$intend)
anes$actual = gsub("emocratic;", "Democratic", anes$actual)
anes$actual = gsub("epublican;", "Republican", anes$actual)
anes = anes %>% mutate(intend = as_factor(intend),
                           actual = as_factor(actual))

# classified considered_result
anes$considered_result = as.character(anes$considered_result)
anes$considered_result = str_sub(anes$considered_result, 4,13)
anes$considered_result = gsub("DK; depend", "others", anes$considered_result)
anes$considered_result = gsub("Other cand", "others", anes$considered_result)
anes$considered_result = as.factor(anes$considered_result)

# changed votes or not
anes$changed_votes = ifelse(as.character(anes$intended_actual_votes) ==
"1. INTENDED Democratic: voted Democratic" | anes$intended_actual_votes ==
"9. INTENDED Republican: voted Republican" , 0,1)

# whether care party wins or not
anes$care_party_win = ifelse(as.character(anes$care_party_win) ==
"1. Don't care very much or DK, pro-con, depends, and", 0,1)

# drop redundant levels
anes$year = as_factor(as.character(anes$year))

```

```

anes$turnout = as_factor(as.character(anes$turnout))
anes$region = as_factor(as.character(anes$region))
anes$income = as_factor(as.character(anes$income))
anes$work = as_factor(as.character(anes$work))
anes$education = as_factor(as.character(anes$education))
anes$race = as_factor(as.character(anes$race))
anes$religion = as_factor(as.character(anes$religion))
anes$gender = as_factor(as.character(anes$gender))
anes$partisanship_strength = as_factor(as.character(anes$partisanship_strength))
anes$care_party_win = as_factor(as.character(anes$care_party_win))
anes$try_influence = as_factor(as.character(anes$try_influence))
anes$interest = as_factor(as.character(anes$interest))

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

```

10 variables represent basic information about election and demographic characteristics are included: year, turnout, region, income, work, education, race, religion, gender variables. Then I chose other 7 variables that indicates partisanship (partisanship\_strength), reported pre vote intention/reported post vote for president (intended\_actual\_votes), whether respondent care a good deal of which party wins presidential election (care\_party\_win) and implies expression of political opinions(try\_influence: respondent try to influence the vote of others during the campaign, days\_discuss: how many days in the past week did respondent talk about politics with family or friend) to some extent together with respondents' opinions of which party will win eventually for president election (considered\_result) and interest variable demonstrate the degree respondents pay attention to political campaigns in elections (interest).

First I replaced NAs in days\_discuss variable with the mean value of rest of valid data, and removed all the rows contain with NA values. Then I deleted rows with ambiguous meaning of intended\_actual\_votes variable that could not help to decide whether there was a changed between intend vote and actual vote. Since intended\_actual\_votes combined intended votes and actual votes, I separated it and added two columns corresponding to each of them (intend, actual). In order to have a clear explaination, I futhur classified considered\_result into three categories: Democratic, Republican, and others. I also added a new column named changed\_votes that classified several cases of intended versus actual votes into 2 cases: changed or remain the same. Finally, I dropped redundant factor levels after all these as last step of data processing and cleaning. There are 12989 rows with 19 columns with my data.

Biases in our data: 1. Selection bias: Bias that occurs because the actual probabilities with which units are sampled differ from the selection probabilities specified by the investigator. 1) Failing to obtain responses from all the chosen sample. From the chosen sample, some people did not participate in this survey, which causes non response issues. Some of the respondents who participate in the survey did not answer all of the questions that missing data related to response bias with partial responses. 2) Using a sample selection procedure that is unknown to investigators, depends on some characteristic associated with properties of interest. There might exist survey data quality issue that investigator might took convenience sample that are easier to select or most likely to respond, and these are often not representative of nonresponding units or harder-to-select units.

2. Measurement Bias: when response has a tendency to differ from the true value in one direction.

- Obtaining accurate responses is challenging particularly in surveys of people
  - 1) People sometimes do not tell the truth Among the variables in my data, respondent might lie about their actual income level and interest that the degree respondent pay attention to political campaigns in elections so that they might get psychological comfort or make theirselves look more successful/active.
  - 2) People forgot For example, the days\_discuss variable in my data is the survey result of asking "how many days in the past week did respondent talk about politics with family or friend". However, respondent might include some days with discussions that occured more than a week ago.

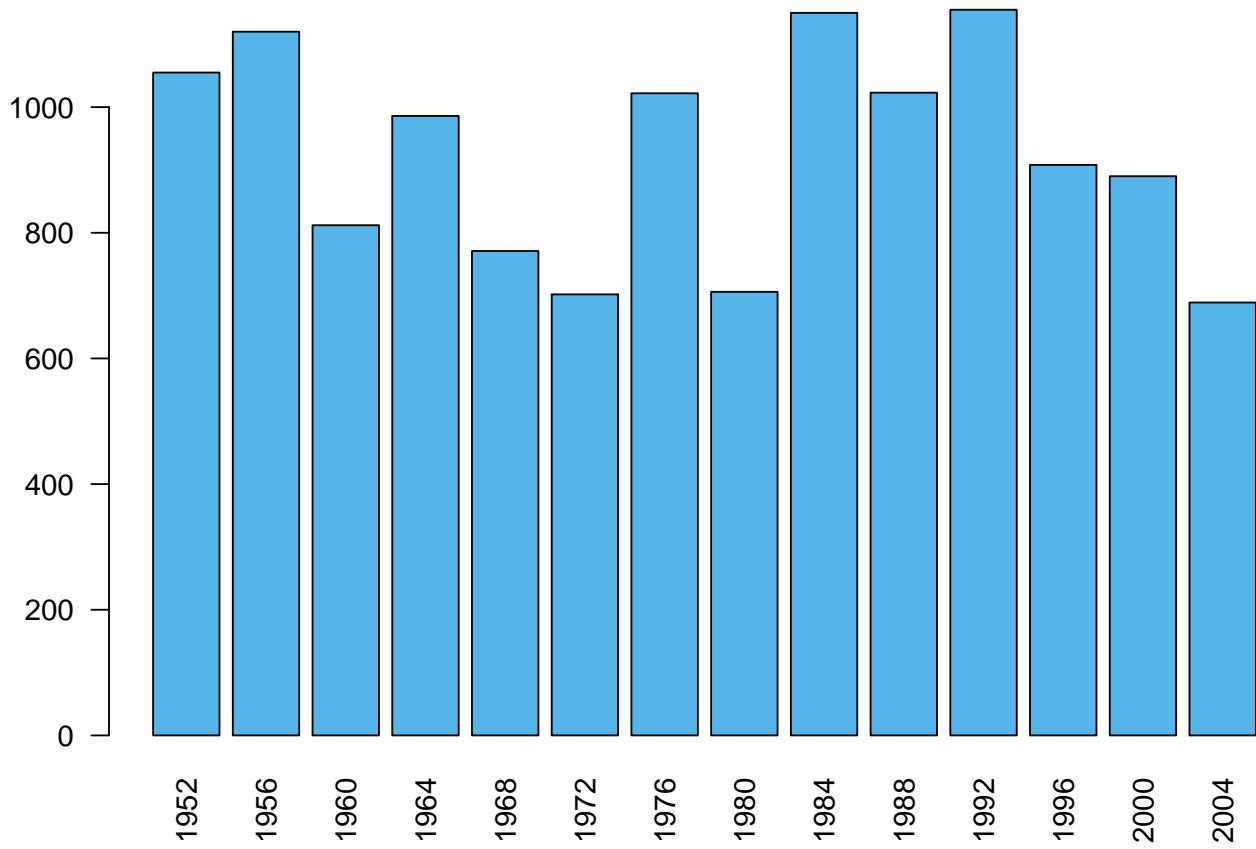
- 3) People do not always understand questions For example, VCF9088 (not in my data) is asking “where would you place [the Democratic Presidential Candidate] on the scale” regarding to political views people might hold are arranged from extremely liberal to extremely conservative, providing with 7 answers that respondent could choose from. However, respondent might not have enough detailed knowledge about [the Democratic Presidential Candidate] to identify where is best position to put on liberal-conservative scale.
- 4) People give different answers using different interviewing method The ANES 2012 and 2016 Times Series Study included both face-to-face (in-person) interviews and Web interviews. (codebook app)
- 5) Question wording varies Question wording are keep changing over the years, some questions are not worded identically in successive surveys that incompleteness of same question asked in different years exists. For example care\_party\_win variable in my data, which was asking “whether respondent care a good deal of which party wins presidential election” and 2016 version of this question is not comparable. Even if a question is worded identically in successive surveys, its replacement in the survey instrument may be different with unknown effect. (codebook intro)
- 6) The orders of questions asking each year differ Questions are not necessarily coded the same way in this dataset as they are in the election study datasets from which they came, question order effects might exist. (codebook intro)

## 4. Analysis

### 4.1 Descriptive Statistics - Interesting Facts about my data

```
barplot(table(as.character(anes$year)),
       las=2,
       main="Number of Respondents over the Years", col="#56B4E9")
```

## Number of Respondents over the Years



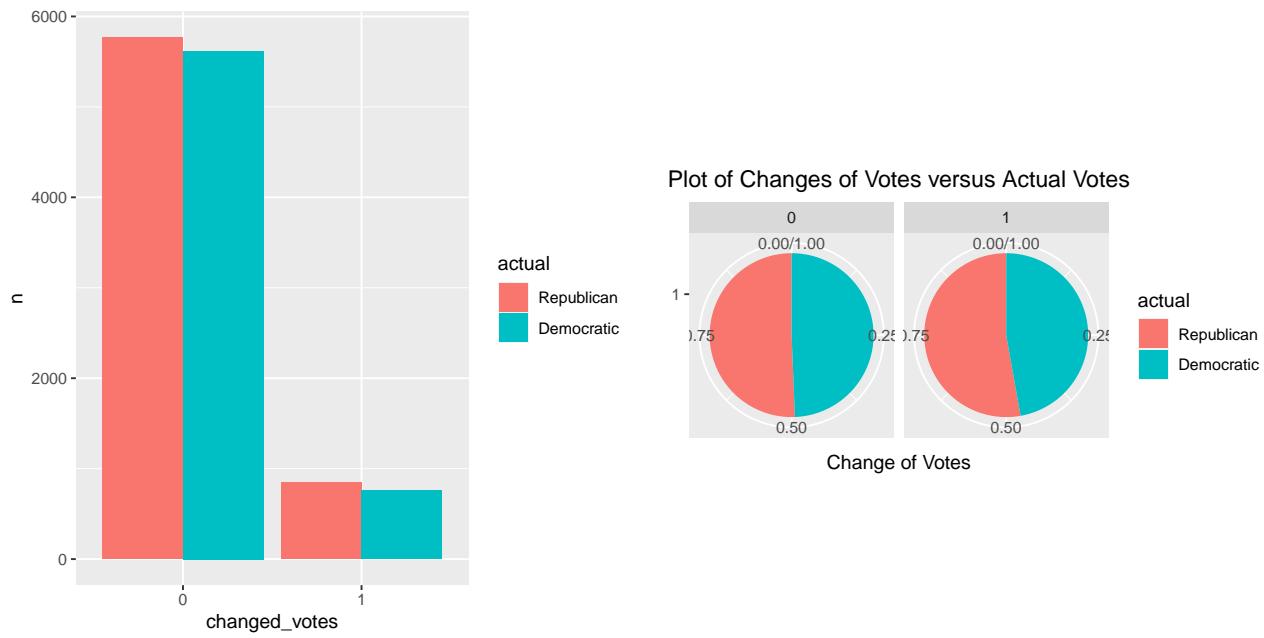
As a result of election conditions and political circumstances varies each year, some survey questions keeps changing among all the years in original dataset based on their detailed description in the ANES code-book. Here, variable year is not integrated after data processing since some of questions were not asked or comparable in some specific years.

```

cv = anes %>% count(changed_votes,actual)
agg_ord <- mutate(cv,
                    changed_votes = reorder(changed_votes, -n, sum),
                    actual = reorder(actual, -n, sum))
p1 <- ggplot(agg_ord) + geom_col(aes(x = changed_votes,
                                         y = n, fill = actual),
                                         position = "dodge")

p2 <- ggplot(data=anes, aes(x=factor(1), stat="bin",
                             fill=actual)) +
  geom_bar(position="fill")+
  ggtitle("Plot of Changes of Votes versus Actual Votes") +
  xlab("") + ylab("Change of Votes")+
  facet_grid(facets=. ~ changed_votes)+ 
  coord_polar(theta="y")+
  theme(plot.title = element_text(hjust = 0.5))
grid.arrange(p1, p2, nrow = 1)

```



From the plot we could observe that there are more Republican respondents than Democratic respondents in my dataset, about half of respondents who did not change their intention voted for Republican and half of respondents who did not change their intention voted for Democratic. For those who change their intention, more than half of them changed from Democratic to Republican and less than half of them changed from Republican to Democratic.

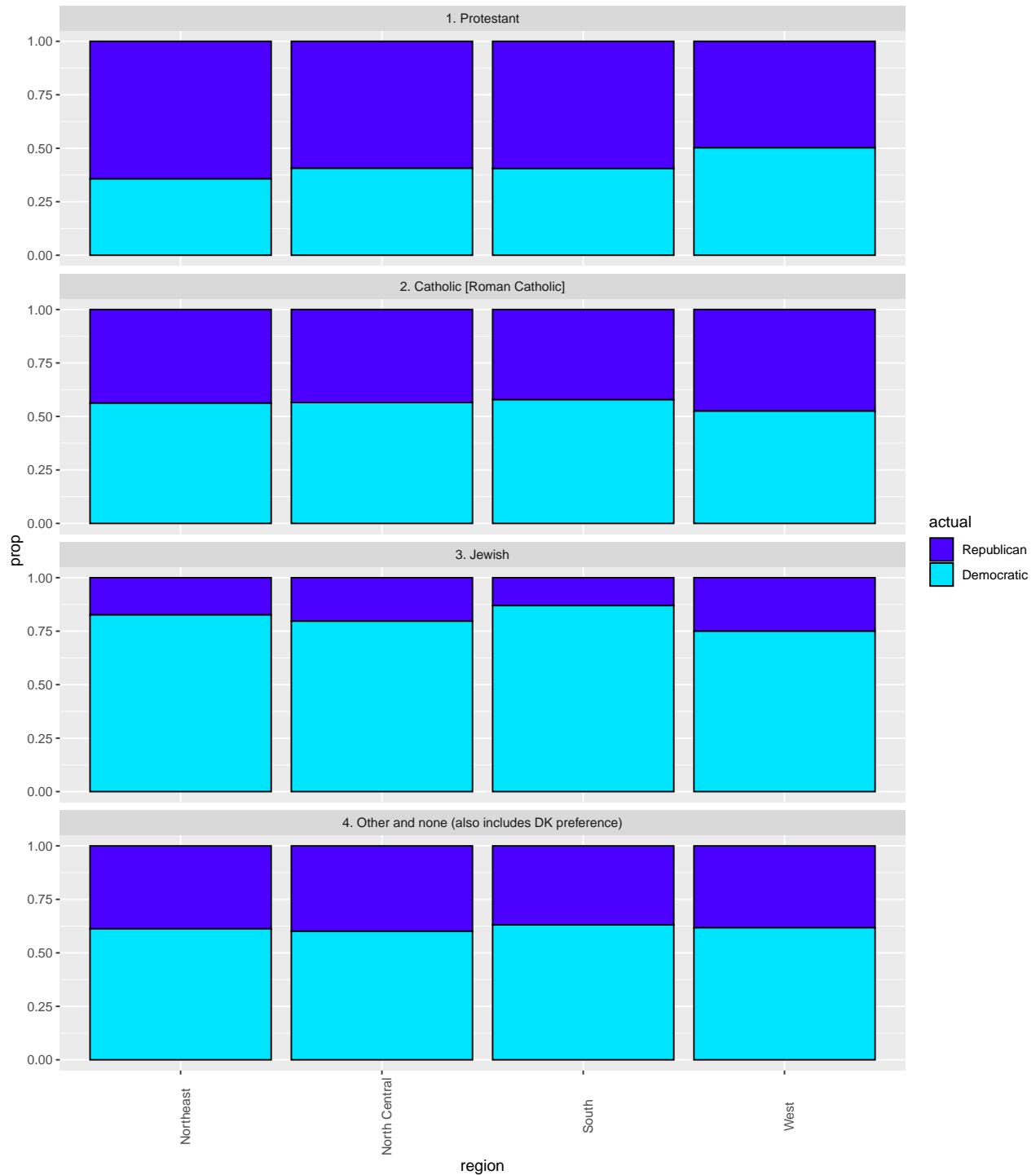
```

anes_actual_region_religion= anes %>%
  group_by(region, religion)%>%
  count(actual)%>%
  group_by(region, religion)%>%
  mutate(
    prop=n/sum(n)
  )
ggplot(anes_actual_region_religion,
       aes(x=region, y=prop, fill=actual)) +
  geom_bar(stat="identity", colour="black")+
  scale_fill_manual(values=c(topo.colors(2)))+
  facet_wrap(~religion, ncol=1) +
  theme(axis.text.x = element_text(angle = 90))+
  labs(title="Which party candidate did religious groups more intend to
        \n vote for in the election with different regions?")+
  theme(plot.title = element_text(hjust = 0.5))

```

Which party candidate did religious groups more intend to

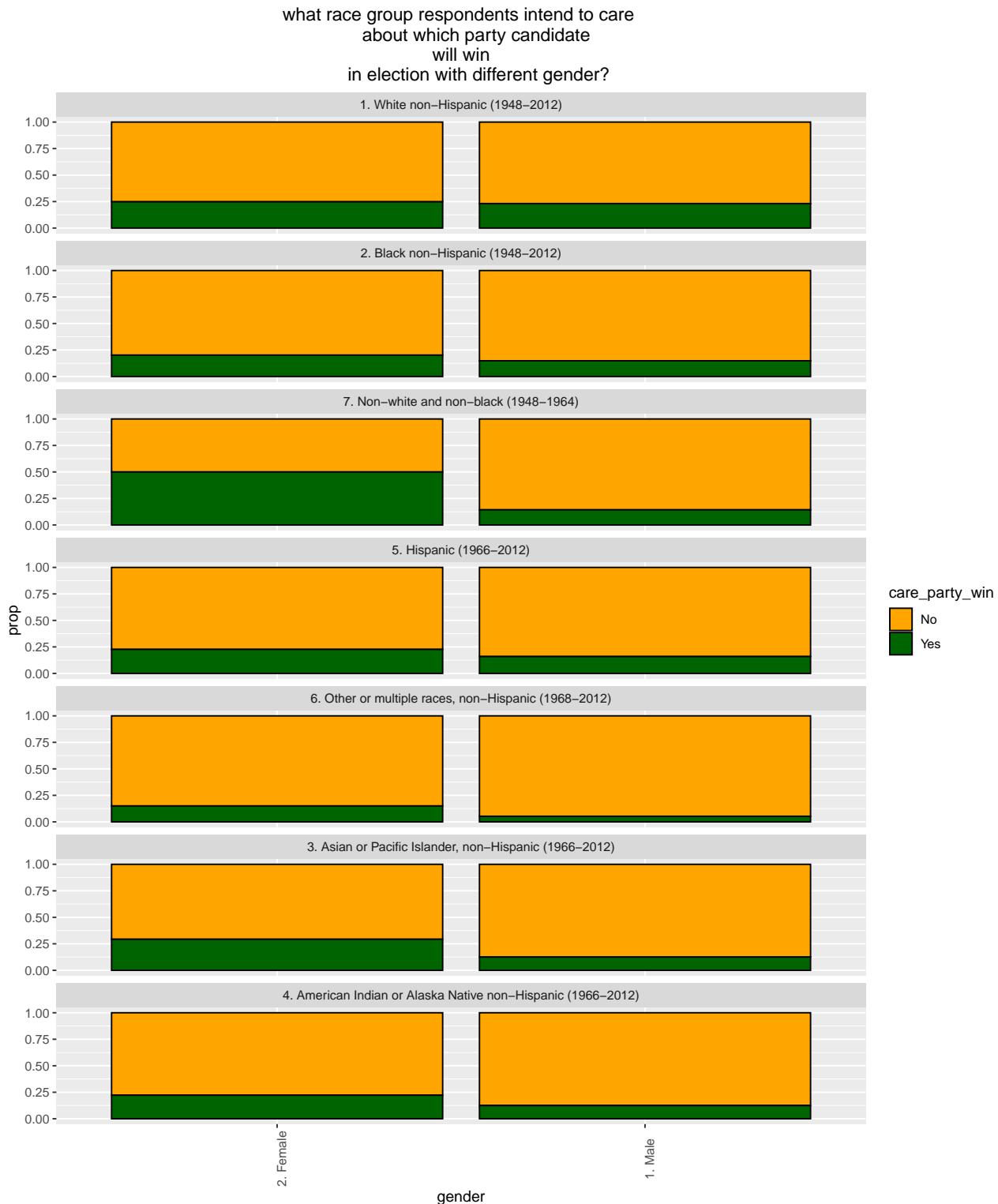
vote for in the election with different regions?



Various information could be shown in this plot. For respondents from protestant religious group and actually voted for Democratic candidates, larger proportion of them located in West region and less proportion of them located in Northeast region; more respondents from protestant religious group actually voted for Republican candidates rather than Democratic candidates. Slightly more respondents from Catholic[Roman Catholic] religious group actually voted for Democratic candidates rather than Republican candidates; Catholic[Roman Catholic] religious respondents from West actually voted less for Democratic comparing to Catholic[Roman

Catholic] religious respondents from other regions. Overwhelmingly more respondents from Jewish religious group actually voted for Democratic candidates rather than Republican candidates; Jewish religious respondents from South actually voted more for Democratic comparing to Jewish religious respondents from other regions; Among those respondents who belong to other or none of religious groups, larger portion of them actually voted for Democratic candidates rather than Republican candidates.

```
anes_cpw = anes %>% mutate(care_party_win = as_factor(anes$care_party_win))
levels(anes_cpw$care_party_win) = c("No", "Yes")
anes_care_race_gender= anes_cpw %>%
  group_by(gender, race)%>%
  count(care_party_win)%>%
  group_by(gender, race)%>%
  mutate(
    prop=n/sum(n)
  )
ggplot(anes_care_race_gender,
       aes(x=gender, y=prop, fill=care_party_win)) +
  geom_bar(stat="identity", colour="black")+
  scale_fill_manual(values=c('orange','dark green'))+
  facet_wrap(~race, ncol=1) +
  theme(axis.text.x = element_text(angle = 90))+
  labs(title="what race group respondents intend to care
        about which party candidate \n will win
        in election with different gender?")+
  theme(plot.title = element_text(hjust = 0.5))
```



Generally speaking, it is interesting that bigger portion of female respondents tends to care about which party wins presidential election and most portion of respondents from each group seems to not care about which party wins the presidential election. There is a large proportion gap regarding to whether respondents care about which party wins presidential election between the two genders for race non-white and non-black (1948-1964) group and tiny differences in proportion exists regarding to whether respondents care about which party wins presidential election between the two genders for White non-Hispanic (1948-2012) group.

```

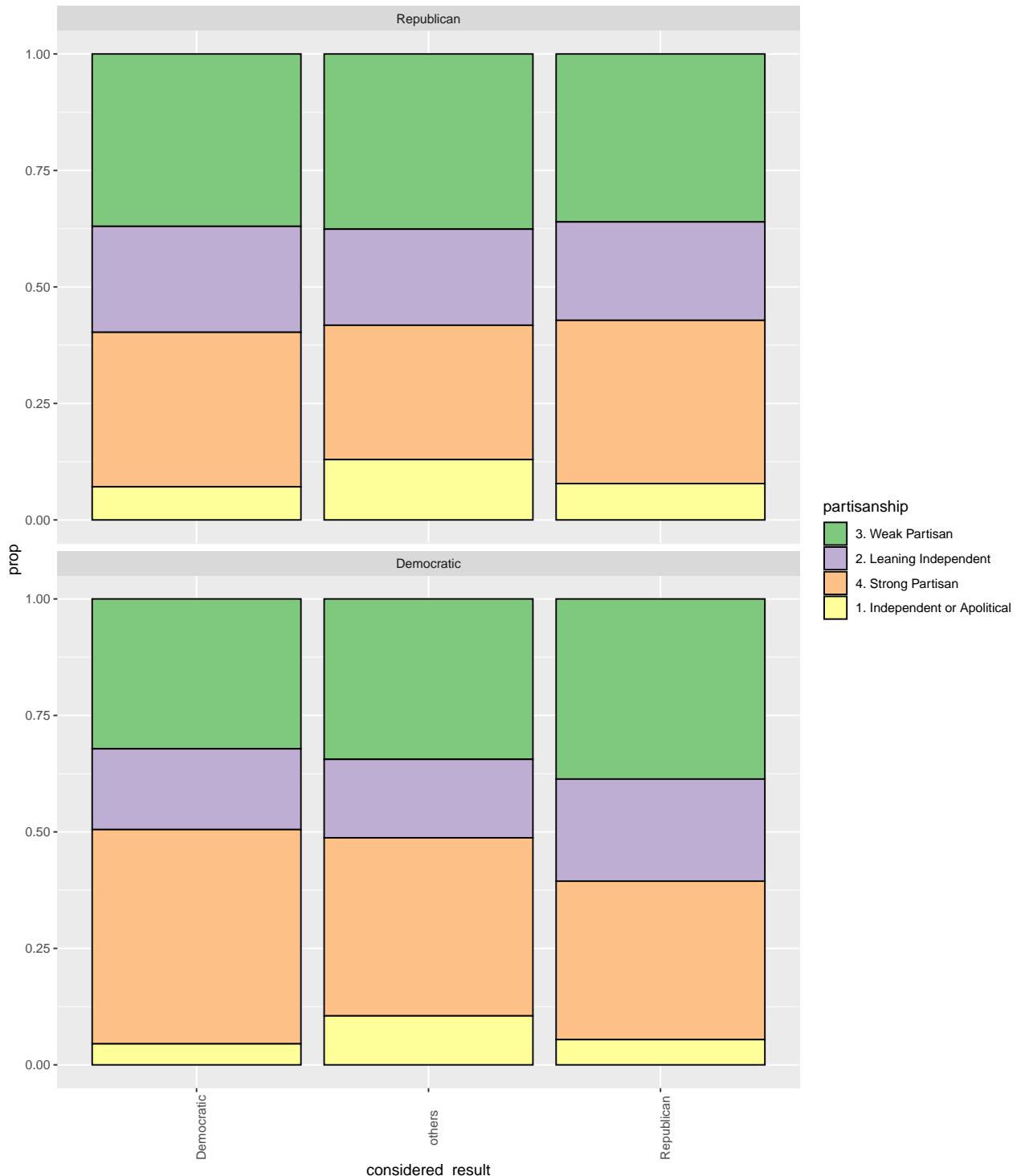
anes_cpw2 = anes %>% mutate(partisanship = as_factor(anes$partisanship_strength))
names(anes_cpw2)

## [1] "year"                  "turnout"                "region"
## [4] "income"                 "work"                   "education"
## [7] "race"                   "religion"               "gender"
## [10] "partisanship_strength" "intended_actual_votes" "care_party_win"
## [13] "try_influence"          "days_discuss"           "considered_result"
## [16] "interest"               "intend"                 "actual"
## [19] "changed_votes"          "partisanship"

anes_partisanship_actual_considered_result= anes_cpw2 %>%
  group_by(considered_result, actual)%>%
  count(partisanship)%>%
  group_by(considered_result, actual)%>%
  mutate(
    prop=n/sum(n)
  )
ggplot(anes_partisanship_actual_considered_result,
       aes(x=considered_result, y=prop, fill=partisanship)) +
  geom_bar(stat="identity", colour="black")+
  scale_fill_manual(values=brewer.pal(4, "Accent"))+
  facet_wrap(~actual, ncol=1) +
  theme(axis.text.x = element_text(angle = 90))+
  labs(title="Difference between actual votes among different
       considered vote results in November with various partisanship strength")+
  theme(plot.title = element_text(hjust = 0.5))

```

Difference between actual votes among different  
considered vote results in November with various partisanship strength



For respondents who both actually voted for Republican and actually voted for Democratic, strong partisanship consists smallest proportion in others comparing to Democratic and Republican regarding to respondent's opinion about who will be elected president in November. For respondents who actually voted for Democratic, weak partisanship consists bigger proportion in Republican comparing to Democratic and Republican regarding to respondent's opinion about who will be elected president in November. However, for respondents who actually voted for Republic, weak partisanship did not consists bigger proportion in

Democratic comparing to Others and Republican regarding to respondent's opinion about who will be elected president in November.

**4.2 It is quite interesting that some respondents intend to vote for a particular party but changed their mind eventually when they actually voted. What might be some significant factors behind this?**

```
set.seed(5243)
n <- nrow(anes)
index <- sample.int(n, n*0.8)
anes_train <- anes[index,]
anes_test <- anes[-index,]

anes_2 = anes[,-c(2,11,17)]
anes_2$days_discuss = as.numeric(as_factor(anes$days_discuss))
anes_2$changed_votes = as.factor(anes$changed_votes)
anes_train <- anes_2[index,]
anes_test <- anes_2[-index,]

anes_glm <- glm(as.factor(changed_votes) ~ ., family = binomial("logit"), data=anes_train)
summary(anes_glm)
```

```
##
## Call:
## glm(formula = as.factor(changed_votes) ~ ., family = binomial("logit"),
##      data = anes_train)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.7103  -0.5121  -0.3684  -0.2598   2.8291
##
## Coefficients:
##                               Estimate
## (Intercept)                -2.245367
## year1956                  -0.120684
## year1960                   0.109937
## year1964                  -0.176303
## year1968                   0.395662
## year1972                  -0.056744
## year1976                   -0.716013
## year1980                   0.229573
## year1984                  -0.405948
## year1988                   0.035468
## year1992                  -0.157202
## year1996                  -0.229534
## year2000                   0.074512
## year2004                  -0.229893
## regionNorth Central        -0.087651
## regionSouth                 0.045107
## regionWest                  0.050827
## income3. 34 to 67 percentile 0.044675
```

## income1. 0 to 16 percentile	-0.021755
## income2. 17 to 33 percentile	0.027505
## income5. 96 to 100 percentile	-0.030928
## work6. Homemakers (1980-later: no other occupation (any	-0.033487
## work3. Skilled, semi-skilled and service workers	0.044529
## work4. Laborers, except farm	0.085963
## work5. Farmers, farm managers, farm laborers and foremen;	0.221786
## work1. Professional and managerial	-0.017678
## education1. Grade school or less (0-8 grades)	0.012759
## education4. College or advanced degree (no cases 1948)	-0.096623
## education3. Some college (13 grades or more but no degree;	0.107786
## race2. Black non-Hispanic (1948-2012)	-0.032012
## race7. Non-white and non-black (1948-1964)	-11.598033
## race5. Hispanic (1966-2012)	0.114693
## race6. Other or multiple races, non-Hispanic (1968-2012)	-0.090421
## race3. Asian or Pacific Islander, non-Hispanic (1966-2012)	0.089924
## race4. American Indian or Alaska Native non-Hispanic (1966-2012)	1.071776
## religion2. Catholic [Roman Catholic]	0.145282
## religion3. Jewish	-0.270844
## religion4. Other and none (also includes DK preference)	-0.235925
## gender1. Male	-0.129146
## partisanship_strength2. Leaning Independent	0.255233
## partisanship_strength4. Strong Partisan	-0.746425
## partisanship_strength1. Independent or Apolitical	0.824448
## care_party_win0	1.026569
## try_influence2. Yes	-0.444503
## days_discuss	0.002211
## considered_resultothers	0.761118
## considered_resultRepublican	-0.024440
## interest2. Somewhat interested	0.151547
## interest1. Not much interested	0.290474
## interest9. DK	-12.440398
## actualDemocratic	0.041378
##	Std. Error
## (Intercept)	0.208052
## year1956	0.152453
## year1960	0.160693
## year1964	0.167131
## year1968	0.158986
## year1972	0.175196
## year1976	0.177804
## year1980	0.166158
## year1984	0.169878
## year1988	0.161359
## year1992	0.168984
## year1996	0.186964
## year2000	0.180287
## year2004	0.211739
## regionNorth Central	0.090244
## regionSouth	0.104056
## regionWest	0.096588
## income3. 34 to 67 percentile	0.080560
## income1. 0 to 16 percentile	0.119073
## income2. 17 to 33 percentile	0.101651

## income5. 96 to 100 percentile	0.149049
## work6. Homemakers (1980-later: no other occupation (any	0.110723
## work3. Skilled, semi-skilled and service workers	0.100230
## work4. Laborers, except farm	0.212489
## work5. Farmers, farm managers, farm laborers and foremen;	0.187456
## work1. Professional and managerial	0.104767
## education1. Grade school or less (0-8 grades)	0.103495
## education4. College or advanced degree (no cases 1948)	0.109088
## education3. Some college (13 grades or more but no degree;	0.089071
## race2. Black non-Hispanic (1948-2012)	0.131267
## race7. Non-white and non-black (1948-1964)	193.665388
## race5. Hispanic (1966-2012)	0.181573
## race6. Other or multiple races, non-Hispanic (1968-2012)	0.570404
## race3. Asian or Pacific Islander, non-Hispanic (1966-2012)	0.408404
## race4. American Indian or Alaska Native non-Hispanic (1966-2012)	0.438965
## religion2. Catholic [Roman Catholic]	0.079063
## religion3. Jewish	0.218771
## religion4. Other and none (also includes DK preference)	0.135524
## gender1. Male	0.077907
## partisanship_strength2. Leaning Independent	0.082013
## partisanship_strength4. Strong Partisan	0.089833
## partisanship_strength1. Independent or Apolitical	0.101731
## care_party_win0	0.069748
## try_influence2. Yes	0.075418
## days_discuss	0.020790
## considered_resultothers	0.106329
## considered_resultRepublican	0.088086
## interest2. Somewhat interested	0.077335
## interest1. Not much interested	0.098761
## interest9. DK	535.411236
## actualDemocratic	0.074047
##	z value
## (Intercept)	-10.792
## year1956	-0.792
## year1960	0.684
## year1964	-1.055
## year1968	2.489
## year1972	-0.324
## year1976	-4.027
## year1980	1.382
## year1984	-2.390
## year1988	0.220
## year1992	-0.930
## year1996	-1.228
## year2000	0.413
## year2004	-1.086
## regionNorth Central	-0.971
## regionSouth	0.433
## regionWest	0.526
## income3. 34 to 67 percentile	0.555
## income1. 0 to 16 percentile	-0.183
## income2. 17 to 33 percentile	0.271
## income5. 96 to 100 percentile	-0.208
## work6. Homemakers (1980-later: no other occupation (any	-0.302

## work3. Skilled, semi-skilled and service workers	0.444
## work4. Laborers, except farm	0.405
## work5. Farmers, farm managers, farm laborers and foremen;	1.183
## work1. Professional and managerial	-0.169
## education1. Grade school or less (0-8 grades)	0.123
## education4. College or advanced degree (no cases 1948)	-0.886
## education3. Some college (13 grades or more but no degree;	1.210
## race2. Black non-Hispanic (1948-2012)	-0.244
## race7. Non-white and non-black (1948-1964)	-0.060
## race5. Hispanic (1966-2012)	0.632
## race6. Other or multiple races, non-Hispanic (1968-2012)	-0.159
## race3. Asian or Pacific Islander, non-Hispanic (1966-2012)	0.220
## race4. American Indian or Alaska Native non-Hispanic (1966-2012)	2.442
## religion2. Catholic [Roman Catholic]	1.838
## religion3. Jewish	-1.238
## religion4. Other and none (also includes DK preference)	-1.741
## gender1. Male	-1.658
## partisanship_strength2. Leaning Independent	3.112
## partisanship_strength4. Strong Partisan	-8.309
## partisanship_strength1. Independent or Apolitical	8.104
## care_party_win0	14.718
## try_influence2. Yes	-5.894
## days_discuss	0.106
## considered_resultothers	7.158
## considered_resultRepublican	-0.277
## interest2. Somewhat interested	1.960
## interest1. Not much interested	2.941
## interest9. DK	-0.023
## actualDemocratic	0.559
##	Pr(> z )
## (Intercept)	< 2e-16 ***
## year1956	0.42858
## year1960	0.49388
## year1964	0.29148
## year1968	0.01282 *
## year1972	0.74602
## year1976	5.65e-05 ***
## year1980	0.16708
## year1984	0.01686 *
## year1988	0.82602
## year1992	0.35222
## year1996	0.21956
## year2000	0.67939
## year2004	0.27759
## regionNorth Central	0.33142
## regionSouth	0.66466
## regionWest	0.59874
## income3. 34 to 67 percentile	0.57920
## income1. 0 to 16 percentile	0.85503
## income2. 17 to 33 percentile	0.78671
## income5. 96 to 100 percentile	0.83562
## work6. Homemakers (1980-later: no other occupation (any	0.76232
## work3. Skilled, semi-skilled and service workers	0.65685
## work4. Laborers, except farm	0.68581

```

## work5. Farmers, farm managers, farm laborers and foremen;          0.23676
## work1. Professional and managerial                                0.86601
## education1. Grade school or less (0-8 grades)                      0.90188
## education4. College or advanced degree (no cases 1948)              0.37576
## education3. Some college (13 grades or more but no degree;          0.22624
## race2. Black non-Hispanic (1948-2012)                                 0.80733
## race7. Non-white and non-black (1948-1964)                            0.95225
## race5. Hispanic (1966-2012)                                         0.52761
## race6. Other or multiple races, non-Hispanic (1968-2012)             0.87405
## race3. Asian or Pacific Islander, non-Hispanic (1966-2012)            0.82573
## race4. American Indian or Alaska Native non-Hispanic (1966-2012)      0.01462 *
## religion2. Catholic [Roman Catholic]                                    0.06613 .
## religion3. Jewish                                                 0.21571
## religion4. Other and none (also includes DK preference)                0.08171 .
## gender1. Male                                              0.09738 .
## partisanship_strength2. Leaning Independent                         0.00186 **
## partisanship_strength4. Strong Partisan                           < 2e-16 ***
## partisanship_strength1. Independent or Apolitical                 5.31e-16 ***
## care_party_win0                                                 < 2e-16 ***
## try_influence2. Yes                                            3.77e-09 ***
## days_discuss                                                0.91531
## considered_resultothers                                     8.18e-13 ***
## considered_resultRepublican                                0.78143
## interest2. Somewhat interested                               0.05004 .
## interest1. Not much interested                             0.00327 **
## interest9. DK                                              0.98146
## actualDemocratic                                         0.57629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 7709.0 on 10390 degrees of freedom
## Residual deviance: 6725.3 on 10340 degrees of freedom
## AIC: 6827.3
##
## Number of Fisher Scoring iterations: 12

```

```

#AIC
step(anes_glm)

```

```

## Start:  AIC=6827.3
## as.factor(changed_votes) ~ year + region + income + work + education +
##   race + religion + gender + partisanship_strength + care_party_win +
##   try_influence + days_discuss + considered_result + interest +
##   actual
##
##                               Df Deviance    AIC
## - work                  5   6727.4 6819.4
## - income                 4   6725.9 6819.9
## - race                   6   6732.8 6822.8
## - region                 3   6728.5 6824.5
## - education               3   6729.0 6825.0
## - days_discuss            1   6725.3 6825.3

```

```

## - actual           1  6725.6 6825.6
## <none>            6725.3 6827.3
## - gender          1  6728.0 6828.0
## - interest         3  6734.8 6830.8
## - religion         3  6735.8 6831.8
## - try_influence    1  6761.2 6861.2
## - year             13 6789.8 6865.8
## - considered_result 2  6790.3 6888.3
## - partisanship_strength 3  6929.2 7025.2
## - care_party_win   1  6939.7 7039.7
##
## Step: AIC=6819.39
## as.factor(changed_votes) ~ year + region + income + education +
##     race + religion + gender + partisanship_strength + care_party_win +
##     try_influence + days_discuss + considered_result + interest +
##     actual
##
##                               Df Deviance     AIC
## - income                  4  6728.0 6812.0
## - race                     6  6735.0 6815.0
## - region                   3  6730.4 6816.4
## - days_discuss             1  6727.4 6817.4
## - actual                   1  6727.7 6817.7
## - education                3  6731.9 6817.9
## - gender                   1  6729.1 6819.1
## <none>                    6727.4 6819.4
## - interest                 3  6737.3 6823.3
## - religion                 3  6738.0 6824.0
## - try_influence             1  6763.5 6853.5
## - year                      13 6791.6 6857.6
## - considered_result         2  6793.1 6881.1
## - partisanship_strength     3  6931.2 7017.2
## - care_party_win            1  6941.6 7031.6
##
## Step: AIC=6812.03
## as.factor(changed_votes) ~ year + region + education + race +
##     religion + gender + partisanship_strength + care_party_win +
##     try_influence + days_discuss + considered_result + interest +
##     actual
##
##                               Df Deviance     AIC
## - race                     6  6735.6 6807.6
## - region                   3  6731.1 6809.1
## - days_discuss              1  6728.0 6810.0
## - actual                   1  6728.4 6810.4
## - education                 3  6733.1 6811.1
## - gender                   1  6729.8 6811.8
## <none>                    6728.0 6812.0
## - interest                 3  6738.0 6816.0
## - religion                 3  6738.7 6816.7
## - try_influence              1  6764.4 6846.4
## - year                      13 6792.4 6850.4
## - considered_result          2  6794.2 6874.2
## - partisanship_strength      3  6932.5 7010.5

```

```

## - care_party_win      1  6942.4 7024.4
##
## Step: AIC=6807.6
## as.factor(changed_votes) ~ year + region + education + religion +
##   gender + partisanship_strength + care_party_win + try_influence +
##   days_discuss + considered_result + interest + actual
##
##                                     Df Deviance    AIC
## - region                      3  6739.2 6805.2
## - days_discuss                 1  6735.6 6805.6
## - actual                       1  6736.0 6806.0
## - education                     3  6740.8 6806.8
## - gender                        1  6737.5 6807.5
## <none>                         6735.6 6807.6
## - interest                      3  6745.9 6811.9
## - religion                      3  6746.3 6812.3
## - try_influence                  1  6771.8 6841.8
## - year                          13 6799.5 6845.5
## - considered_result              2  6802.2 6870.2
## - partisanship_strength          3  6940.9 7006.9
## - care_party_win                 1  6949.4 7019.4
##
## Step: AIC=6805.15
## as.factor(changed_votes) ~ year + education + religion + gender +
##   partisanship_strength + care_party_win + try_influence +
##   days_discuss + considered_result + interest + actual
##
##                                     Df Deviance    AIC
## - days_discuss                  1  6739.2 6803.2
## - actual                        1  6739.6 6803.6
## - education                      3  6744.1 6804.1
## - gender                         1  6741.1 6805.1
## <none>                          6739.2 6805.2
## - religion                       3  6749.0 6809.0
## - interest                       3  6749.4 6809.4
## - try_influence                   1  6775.0 6839.0
## - year                           13 6802.7 6842.7
## - considered_result               2  6806.1 6868.1
## - partisanship_strength          3  6942.7 7002.7
## - care_party_win                  1  6953.0 7017.0
##
## Step: AIC=6803.17
## as.factor(changed_votes) ~ year + education + religion + gender +
##   partisanship_strength + care_party_win + try_influence +
##   considered_result + interest + actual
##
##                                     Df Deviance    AIC
## - actual                        1  6739.6 6801.6
## - education                      3  6744.1 6802.1
## - gender                         1  6741.1 6803.1
## <none>                          6739.2 6803.2
## - religion                       3  6749.0 6807.0
## - interest                        3  6749.4 6807.4
## - try_influence                   1  6775.2 6837.2

```

```

## - year           13  6803.6 6841.6
## - considered_result    2   6806.2 6866.2
## - partisanship_strength 3   6942.7 7000.7
## - care_party_win      1   6953.0 7015.0
##
## Step: AIC=6801.6
## as.factor(changed_votes) ~ year + education + religion + gender +
##     partisanship_strength + care_party_win + try_influence +
##     considered_result + interest
##
##                                     Df Deviance     AIC
## - education                  3   6744.6 6800.6
## <none>                         6739.6 6801.6
## - gender                      1   6741.7 6801.7
## - religion                    3   6749.3 6805.3
## - interest                     3   6749.8 6805.8
## - try_influence                 1   6775.9 6835.9
## - year                        13  6803.9 6839.9
## - considered_result            2   6807.1 6865.1
## - partisanship_strength        3   6943.2 6999.2
## - care_party_win                1   6953.9 7013.9
##
## Step: AIC=6800.6
## as.factor(changed_votes) ~ year + religion + gender + partisanship_strength +
##     care_party_win + try_influence + considered_result + interest
##
##                                     Df Deviance     AIC
## <none>                         6744.6 6800.6
## - gender                      1   6746.9 6800.9
## - religion                    3   6754.8 6804.8
## - interest                     3   6755.9 6805.9
## - try_influence                 1   6782.1 6836.1
## - year                        13  6808.8 6838.8
## - considered_result            2   6813.0 6865.0
## - partisanship_strength        3   6949.2 6999.2
## - care_party_win                1   6959.2 7013.2
##
## Call: glm(formula = as.factor(changed_votes) ~ year + religion + gender +
##     partisanship_strength + care_party_win + try_influence +
##     considered_result + interest, family = binomial("logit"),
##     data = anes_train)
##
## Coefficients:
##                               (Intercept)
##                               -2.21272
##                               year1956
##                               -0.10561
##                               year1960
##                               0.12752
##                               year1964
##                               -0.17096
##                               year1968
##                               0.40292

```

```

##                                     year1972
##                                     -0.03091
##                                     year1976
##                                     -0.69329
##                                     year1980
##                                     0.25569
##                                     year1984
##                                     -0.35682
##                                     year1988
##                                     0.08510
##                                     year1992
##                                     -0.12091
##                                     year1996
##                                     -0.21025
##                                     year2000
##                                     0.10288
##                                     year2004
##                                     -0.19573
## religion2. Catholic [Roman Catholic]
##                                     0.13866
## religion3. Jewish
##                                     -0.29835
## religion4. Other and none (also includes DK preference)
##                                     -0.21037
## gender1. Male
##                                     -0.09907
## partisanship_strength2. Leaning Independent
##                                     0.24903
## partisanship_strength4. Strong Partisan
##                                     -0.73739
## partisanship_strength1. Independent or Apolitical
##                                     0.81466
## care_party_win0
##                                     1.02430
## try_influence2. Yes
##                                     -0.44761
## considered_resultothers
##                                     0.75499
## considered_resultRepublican
##                                     -0.05062
## interest2. Somewhat interested
##                                     0.16390
## interest1. Not much interested
##                                     0.31110
## interest9. DK
##                                     -10.35663
##
## Degrees of Freedom: 10390 Total (i.e. Null); 10363 Residual
## Null Deviance: 7709
## Residual Deviance: 6745 AIC: 6801

#BIC
output <- bic.glm(as.factor(changed_votes) ~ ., glm.family="binomial", data=anes_train, maxCol = 16)
summary(output)

```

```

## Call:
## bic.glm.formula(f = as.factor(changed_votes) ~ ., data = anes_train,      glm.family = "binomial", ma
## 
## 1 models were selected
## Best 1 models (cumulative posterior probability = 1 ):
## 
##          p!=0
## Intercept           100
## year                100
##     .1956
##     .1960
##     .1964
##     .1968
##     .1972
##     .1976
##     .1980
##     .1984
##     .1988
##     .1992
##     .1996
##     .2000
##     .2004
## region               0
##     .North Central
##     .South
##     .West
## income                0
##     .3. 34 to 67 percentile
##     .1. 0 to 16 percentile
##     .2. 17 to 33 percentile
##     .5. 96 to 100 percentile
## work                  0
##     .6. Homemakers (1980-later: no other occupation (any
##     .3. Skilled, semi-skilled and service workers
##     .4. Laborers, except farm
##     .5. Farmers, farm managers, farm laborers and foremen;
##     .1. Professional and managerial
## education              0
##     .1. Grade school or less (0-8 grades)
##     .4. College or advanced degree (no cases 1948)
##     .3. Some college (13 grades or more but no degree;
## race                   0
##     .2. Black non-Hispanic (1948-2012)
##     .7. Non-white and non-black (1948-1964)
##     .5. Hispanic (1966-2012)
##     .6. Other or multiple races, non-Hispanic (1968-2012)
##     .3. Asian or Pacific Islander, non-Hispanic (1966-2012)
##     .4. American Indian or Alaska Native non-Hispanic (1966-2012)
## religion                0
##     .2. Catholic [Roman Catholic]
##     .3. Jewish
##     .4. Other and none (also includes DK preference)

```

## gender	0
##       .1. Male	
## partisanship_strength	100
##           .2. Leaning Independent	
##           .4. Strong Partisan	
##           .1. Independent or Apolitical	
## care_party_win	100
##       .0	
## try_influence	100
##       .2. Yes	
## days_discuss	0
## considered_result	100
##       .others	
##       .Republican	
## interest	0
##       .2. Somewhat interested	
##       .1. Not much interested	
##       .9. DK	
## actual	0
##       .Democratic	
##	
## nVar	
## BIC	
## post prob	
##	EV
## Intercept	-2.07673
## year	
##       .1956	-0.09800
##       .1960	0.10077
##       .1964	-0.18937
##       .1968	0.37624
##       .1972	-0.03182
##       .1976	-0.72168
##       .1980	0.22401
##       .1984	-0.35172
##       .1988	0.07001
##       .1992	-0.15973
##       .1996	-0.22010
##       .2000	0.09429
##       .2004	-0.24199
## region	
##       .North Central	0.00000
##       .South	0.00000
##       .West	0.00000
## income	
##       .3. 34 to 67 percentile	0.00000
##       .1. 0 to 16 percentile	0.00000
##       .2. 17 to 33 percentile	0.00000
##       .5. 96 to 100 percentile	0.00000
## work	
##       .6. Homemakers (1980-later: no other occupation (any	0.00000
##       .3. Skilled, semi-skilled and service workers	0.00000
##       .4. Laborers, except farm	0.00000
##       .5. Farmers, farm managers, farm laborers and foremen;	0.00000

## .1. Professional and managerial	0.00000
## education	
## .1. Grade school or less (0-8 grades)	0.00000
## .4. College or advanced degree (no cases 1948)	0.00000
## .3. Some college (13 grades or more but no degree;	0.00000
## race	
## .2. Black non-Hispanic (1948-2012)	0.00000
## .7. Non-white and non-black (1948-1964)	0.00000
## .5. Hispanic (1966-2012)	0.00000
## .6. Other or multiple races, non-Hispanic (1968-2012)	0.00000
## .3. Asian or Pacific Islander, non-Hispanic (1966-2012)	0.00000
## .4. American Indian or Alaska Native non-Hispanic (1966-2012)	0.00000
## religion	
## .2. Catholic [Roman Catholic]	0.00000
## .3. Jewish	0.00000
## .4. Other and none (also includes DK preference)	0.00000
## gender	
## .1. Male	0.00000
## partisanship_strength	
## .2. Leaning Independent	0.20913
## .4. Strong Partisan	-0.76272
## .1. Independent or Apolitical	0.79601
## care_party_win	
## .0	1.08548
## try_influence	
## .2. Yes	-0.50932
## days_discuss	0.00000
## considered_result	
## .others	0.76732
## .Republican	-0.05687
## interest	
## .2. Somewhat interested	0.00000
## .1. Not much interested	0.00000
## .9. DK	0.00000
## actual	
## .Democratic	0.00000
##	
## nVar	
## BIC	
## post prob	
##	SD
## Intercept	0.12843
## year	
## .1956	0.15031
## .1960	0.15813
## .1964	0.16442
## .1968	0.15515
## .1972	0.17025
## .1976	0.17321
## .1980	0.15926
## .1984	0.16100
## .1988	0.15213
## .1992	0.15789
## .1996	0.17433

```

##      .2000          0.16209
##      .2004          0.19963
## region
##      .North Central          0.00000
##      .South           0.00000
##      .West            0.00000
## income
##      .3. 34 to 67 percentile          0.00000
##      .1. 0 to 16 percentile          0.00000
##      .2. 17 to 33 percentile          0.00000
##      .5. 96 to 100 percentile          0.00000
## work
##      .6. Homemakers (1980-later: no other occupation (any          0.00000
##      .3. Skilled, semi-skilled and service workers          0.00000
##      .4. Laborers, except farm          0.00000
##      .5. Farmers, farm managers, farm laborers and foremen;          0.00000
##      .1. Professional and managerial          0.00000
## education
##      .1. Grade school or less (0-8 grades)          0.00000
##      .4. College or advanced degree (no cases 1948)          0.00000
##      .3. Some college (13 grades or more but no degree;          0.00000
## race
##      .2. Black non-Hispanic (1948-2012)          0.00000
##      .7. Non-white and non-black (1948-1964)          0.00000
##      .5. Hispanic (1966-2012)          0.00000
##      .6. Other or multiple races, non-Hispanic (1968-2012)          0.00000
##      .3. Asian or Pacific Islander, non-Hispanic (1966-2012)          0.00000
##      .4. American Indian or Alaska Native non-Hispanic (1966-2012)          0.00000
## religion
##      .2. Catholic [Roman Catholic]          0.00000
##      .3. Jewish          0.00000
##      .4. Other and none (also includes DK preference)          0.00000
## gender
##      .1. Male          0.00000
## partisanship_strength
##      .2. Leaning Independent          0.08070
##      .4. Strong Partisan          0.08834
##      .1. Independent or Apolitical          0.10082
## care_party_win
##      .0          0.06731
## try_influence
##      .2. Yes          0.07247
## days_discuss
## considered_result
##      .others          0.10329
##      .Republican          0.08070
## interest
##      .2. Somewhat interested          0.00000
##      .1. Not much interested          0.00000
##      .9. DK          0.00000
## actual
##      .Democratic          0.00000
##
## nVar

```

```

## BIC
## post prob
##
## Intercept
## year
##    .1956          -9.800e-02
##    .1960          1.008e-01
##    .1964          -1.894e-01
##    .1968          3.762e-01
##    .1972          -3.182e-02
##    .1976          -7.217e-01
##    .1980          2.240e-01
##    .1984          -3.517e-01
##    .1988          7.001e-02
##    .1992          -1.597e-01
##    .1996          -2.201e-01
##    .2000          9.429e-02
##    .2004          -2.420e-01
## region
##    .North Central   .
##    .South           .
##    .West            .
## income
##    .3. 34 to 67 percentile   .
##    .1. 0 to 16 percentile   .
##    .2. 17 to 33 percentile   .
##    .5. 96 to 100 percentile   .
## work
##    .6. Homemakers (1980-later: no other occupation (any   .
##    .3. Skilled, semi-skilled and service workers   .
##    .4. Laborers, except farm   .
##    .5. Farmers, farm managers, farm laborers and foremen;   .
##    .1. Professional and managerial   .
## education
##    .1. Grade school or less (0-8 grades)   .
##    .4. College or advanced degree (no cases 1948)   .
##    .3. Some college (13 grades or more but no degree;   .
## race
##    .2. Black non-Hispanic (1948-2012)   .
##    .7. Non-white and non-black (1948-1964)   .
##    .5. Hispanic (1966-2012)   .
##    .6. Other or multiple races, non-Hispanic (1968-2012)   .
##    .3. Asian or Pacific Islander, non-Hispanic (1966-2012)   .
##    .4. American Indian or Alaska Native non-Hispanic (1966-2012)   .
## religion
##    .2. Catholic [Roman Catholic]   .
##    .3. Jewish   .
##    .4. Other and none (also includes DK preference)   .
## gender
##    .1. Male   .
## partisanship_strength
##    .2. Leaning Independent          2.091e-01
##    .4. Strong Partisan           -7.627e-01
##    .1. Independent or Apolitical  7.960e-01

```

```

## care_party_win .0 1.085e+00
## try_influence .2. Yes -5.093e-01
## days_discuss .
## considered_result
## .others 7.673e-01
## .Republican -5.687e-02
## interest
## .2. Somewhat interested .
## .1. Not much interested .
## .9. DK .
## actual
## .Democratic .
##
## nVar 5
## BIC -8.914e+04
## post prob 1

# Confusion Matrix and Model residual plots
coef(anes_glm)

```

```

## (Intercept) -2.245367432
## year1956 -0.120684299
## year1960 0.109937053
## year1964 0.109937053
## year1964 -0.176303276
## year1968 0.395661704
## year1972 0.395661704
## year1976 -0.056743747
## year1976 0.395661704
## year1980 -0.716012563
## year1980 0.395661704
## year1984 0.229573331
## year1984 0.229573331
## year1988 -0.405948121
## year1988 0.229573331
## year1992 0.035467630
## year1992 0.035467630
## year1996 -0.157202452
## year1996 0.035467630
## year2000 -0.229534472
## year2000 0.035467630
## year2004 0.074511762
## year2004 0.074511762
## -0.229892783
## regionNorth Central
## regionNorth Central -0.087651148
## regionSouth 0.045106900
## regionSouth 0.045106900

```

```

##                                     regionWest
##                                     0.050826506
##           income3. 34 to 67 percentile
##                                     0.044675049
##           income1. 0 to 16 percentile
##                                     -0.021755202
##           income2. 17 to 33 percentile
##                                     0.027505301
##           income5. 96 to 100 percentile
##                                     -0.030928142
## work6. Homemakers (1980-later: no other occupation (any
##                                     -0.033486927
## work3. Skilled, semi-skilled and service workers
##                                     0.044529473
##           work4. Laborers, except farm
##                                     0.085963195
## work5. Farmers, farm managers, farm laborers and foremen;
##                                     0.221785855
##           work1. Professional and managerial
##                                     -0.017677820
##           education1. Grade school or less (0-8 grades)
##                                     0.012759190
##           education4. College or advanced degree (no cases 1948)
##                                     -0.096622605
## education3. Some college (13 grades or more but no degree;
##                                     0.107785890
##           race2. Black non-Hispanic (1948-2012)
##                                     -0.032012295
##           race7. Non-white and non-black (1948-1964)
##                                     -11.598033100
##           race5. Hispanic (1966-2012)
##                                     0.114693062
##           race6. Other or multiple races, non-Hispanic (1968-2012)
##                                     -0.090421222
##           race3. Asian or Pacific Islander, non-Hispanic (1966-2012)
##                                     0.089923709
## race4. American Indian or Alaska Native non-Hispanic (1966-2012)
##                                     1.071776002
##           religion2. Catholic [Roman Catholic]
##                                     0.145281936
##           religion3. Jewish
##                                     -0.270844023
##           religion4. Other and none (also includes DK preference)
##                                     -0.235924674
##           gender1. Male
##                                     -0.129146220
##           partisanship_strength2. Leaning Independent
##                                     0.255232563
##           partisanship_strength4. Strong Partisan
##                                     -0.746425370
##           partisanship_strength1. Independent or Apolitical
##                                     0.824447657
##           care_party_win0
##                                     1.026568902

```

```

##                                try_influence2. Yes
##                                -0.444502817
##                                days_discuss
##                                0.002210831
##                                considered_resultothers
##                                0.761117661
##                                considered_resultRepublican
##                                -0.024439705
##                                interest2. Somewhat interested
##                                0.151547229
##                                interest1. Not much interested
##                                0.290473702
##                                interest9. DK
##                                -12.440398384
##                                actualDemocratic
##                                0.041377629

```

```
summary(anes_glm)$coef
```

	Estimate
##	-2.245367432
## (Intercept)	-0.120684299
## year1956	0.109937053
## year1960	-0.176303276
## year1964	0.395661704
## year1968	-0.056743747
## year1972	-0.716012563
## year1976	0.229573331
## year1980	-0.405948121
## year1984	0.035467630
## year1988	-0.157202452
## year1992	-0.229534472
## year1996	0.074511762
## year2000	-0.229892783
## year2004	-0.087651148
## regionNorth Central	0.045106900
## regionSouth	0.050826506
## regionWest	0.044675049
## income3. 34 to 67 percentile	-0.021755202
## income1. 0 to 16 percentile	0.027505301
## income2. 17 to 33 percentile	-0.030928142
## income5. 96 to 100 percentile	-0.033486927
## work6. Homemakers (1980-later: no other occupation (any	0.044529473
## work3. Skilled, semi-skilled and service workers	0.085963195
## work4. Laborers, except farm	0.221785855
## work5. Farmers, farm managers, farm laborers and foremen;	-0.017677820
## work1. Professional and managerial	0.012759190
## education1. Grade school or less (0-8 grades)	-0.096622605
## education4. College or advanced degree (no cases 1948)	0.107785890
## education3. Some college (13 grades or more but no degree;	-0.032012295
## race2. Black non-Hispanic (1948-2012)	-11.598033100
## race7. Non-white and non-black (1948-1964)	0.114693062
## race5. Hispanic (1966-2012)	-0.090421222
## race6. Other or multiple races, non-Hispanic (1968-2012)	0.089923709
## race3. Asian or Pacific Islander, non-Hispanic (1966-2012)	

## race4. American Indian or Alaska Native non-Hispanic (1966-2012)	1.071776002
## religion2. Catholic [Roman Catholic]	0.145281936
## religion3. Jewish	-0.270844023
## religion4. Other and none (also includes DK preference)	-0.235924674
## gender1. Male	-0.129146220
## partisanship_strength2. Leaning Independent	0.255232563
## partisanship_strength4. Strong Partisan	-0.746425370
## partisanship_strength1. Independent or Apolitical	0.824447657
## care_party_win0	1.026568902
## try_influence2. Yes	-0.444502817
## days_discuss	0.002210831
## considered_resulthothers	0.761117661
## considered_resultRepublican	-0.024439705
## interest2. Somewhat interested	0.151547229
## interest1. Not much interested	0.290473702
## interest9. DK	-12.440398384
## actualDemocratic	0.041377629
##	Std. Error
## (Intercept)	0.20805220
## year1956	0.15245253
## year1960	0.16069290
## year1964	0.16713104
## year1968	0.15898604
## year1972	0.17519629
## year1976	0.17780427
## year1980	0.16615807
## year1984	0.16987767
## year1988	0.16135942
## year1992	0.16898359
## year1996	0.18696438
## year2000	0.18028749
## year2004	0.21173855
## regionNorth Central	0.09024439
## regionSouth	0.10405645
## regionWest	0.09658849
## income3. 34 to 67 percentile	0.08055994
## income1. 0 to 16 percentile	0.11907285
## income2. 17 to 33 percentile	0.10165121
## income5. 96 to 100 percentile	0.14904882
## work6. Homemakers (1980-later: no other occupation (any	0.11072317
## work3. Skilled, semi-skilled and service workers	0.10023008
## work4. Laborers, except farm	0.21248928
## work5. Farmers, farm managers, farm laborers and foremen;	0.18745597
## work1. Professional and managerial	0.10476680
## education1. Grade school or less (0-8 grades)	0.10349492
## education4. College or advanced degree (no cases 1948)	0.10908803
## education3. Some college (13 grades or more but no degree;	0.08907135
## race2. Black non-Hispanic (1948-2012)	0.13126731
## race7. Non-white and non-black (1948-1964)	193.66538768
## race5. Hispanic (1966-2012)	0.18157271
## race6. Other or multiple races, non-Hispanic (1968-2012)	0.57040447
## race3. Asian or Pacific Islander, non-Hispanic (1966-2012)	0.40840395
## race4. American Indian or Alaska Native non-Hispanic (1966-2012)	0.43896542
## religion2. Catholic [Roman Catholic]	0.07906341

## religion3. Jewish	0.21877066
## religion4. Other and none (also includes DK preference)	0.13552402
## gender1. Male	0.07790729
## partisanship_strength2. Leaning Independent	0.08201346
## partisanship_strength4. Strong Partisan	0.08983347
## partisanship_strength1. Independent or Apolitical	0.10173131
## care_party_win0	0.06974821
## try_influence2. Yes	0.07541828
## days_discuss	0.02078994
## considered_resultothers	0.10632856
## considered_resultRepublican	0.08808619
## interest2. Somewhat interested	0.07733532
## interest1. Not much interested	0.09876147
## interest9. DK	535.41123611
## actualDemocratic	0.07404662
##	
## (Intercept)	-10.79232738
## year1956	-0.79161886
## year1960	0.68414380
## year1964	-1.05488051
## year1968	2.48865689
## year1972	-0.32388670
## year1976	-4.02697047
## year1980	1.38165624
## year1984	-2.38964969
## year1988	0.21980514
## year1992	-0.93028234
## year1996	-1.22769089
## year2000	0.41329413
## year2004	-1.08573889
## regionNorth Central	-0.97126418
## regionSouth	0.43348489
## regionWest	0.52621700
## income3. 34 to 67 percentile	0.55455663
## income1. 0 to 16 percentile	-0.18270497
## income2. 17 to 33 percentile	0.27058507
## income5. 96 to 100 percentile	-0.20750343
## work6. Homemakers (1980-later: no other occupation (any	-0.30243828
## work3. Skilled, semi-skilled and service workers	0.44427256
## work4. Laborers, except farm	0.40455309
## work5. Farmers, farm managers, farm laborers and foremen;	1.18313571
## work1. Professional and managerial	-0.16873495
## education1. Grade school or less (0-8 grades)	0.12328325
## education4. College or advanced degree (no cases 1948)	-0.88573059
## education3. Some college (13 grades or more but no degree;	1.21010723
## race2. Black non-Hispanic (1948-2012)	-0.24387103
## race7. Non-white and non-black (1948-1964)	-0.05988697
## race5. Hispanic (1966-2012)	0.63166466
## race6. Other or multiple races, non-Hispanic (1968-2012)	-0.15852124
## race3. Asian or Pacific Islander, non-Hispanic (1966-2012)	0.22018325
## race4. American Indian or Alaska Native non-Hispanic (1966-2012)	2.44159550
## religion2. Catholic [Roman Catholic]	1.83753697
## religion3. Jewish	-1.23802720
## religion4. Other and none (also includes DK preference)	-1.74083293

## gender1. Male	-1.65769099
## partisanship_strength2. Leaning Independent	3.11208141
## partisanship_strength4. Strong Partisan	-8.30898960
## partisanship_strength1. Independent or Apolitical	8.10416812
## care_party_win0	14.71821085
## try_influence2. Yes	-5.89383359
## days_discuss	0.10634142
## considered_resultothers	7.15816789
## considered_resultRepublican	-0.27745216
## interest2. Somewhat interested	1.95961213
## interest1. Not much interested	2.94116422
## interest9. DK	-0.02323522
## actualDemocratic	0.55880509
##	Pr(> z )
## (Intercept)	3.741913e-27
## year1956	4.285829e-01
## year1960	4.938844e-01
## year1964	2.914800e-01
## year1968	1.282266e-02
## year1972	7.460238e-01
## year1976	5.650013e-05
## year1980	1.670773e-01
## year1984	1.686445e-02
## year1988	8.260229e-01
## year1992	3.522249e-01
## year1996	2.195630e-01
## year2000	6.793911e-01
## year2004	2.775945e-01
## regionNorth Central	3.314167e-01
## regionSouth	6.646625e-01
## regionWest	5.987374e-01
## income3. 34 to 67 percentile	5.791980e-01
## income1. 0 to 16 percentile	8.550295e-01
## income2. 17 to 33 percentile	7.867102e-01
## income5. 96 to 100 percentile	8.356167e-01
## work6. Homemakers (1980-later: no other occupation (any	7.623180e-01
## work3. Skilled, semi-skilled and service workers	6.568455e-01
## work4. Laborers, except farm	6.858060e-01
## work5. Farmers, farm managers, farm laborers and foremen;	2.367554e-01
## work1. Professional and managerial	8.660051e-01
## education1. Grade school or less (0-8 grades)	9.018828e-01
## education4. College or advanced degree (no cases 1948)	3.757627e-01
## education3. Some college (13 grades or more but no degree;	2.262378e-01
## race2. Black non-Hispanic (1948-2012)	8.073307e-01
## race7. Non-white and non-black (1948-1964)	9.522457e-01
## race5. Hispanic (1966-2012)	5.276060e-01
## race6. Other or multiple races, non-Hispanic (1968-2012)	8.740461e-01
## race3. Asian or Pacific Islander, non-Hispanic (1966-2012)	8.257284e-01
## race4. American Indian or Alaska Native non-Hispanic (1966-2012)	1.462252e-02
## religion2. Catholic [Roman Catholic]	6.613066e-02
## religion3. Jewish	2.157060e-01
## religion4. Other and none (also includes DK preference)	8.171287e-02
## gender1. Male	9.737985e-02
## partisanship_strength2. Leaning Independent	1.857733e-03

```

## partisanship_strength4. Strong Partisan          9.652038e-17
## partisanship_strength1. Independent or Apolitical 5.310763e-16
## care_party_win0                                4.925192e-49
## try_influence2. Yes                            3.773372e-09
## days_discuss                                 9.153115e-01
## considered_resultothers                      8.176239e-13
## considered_resultRepublican                 7.814329e-01
## interest2. Somewhat interested            5.004114e-02
## interest1. Not much interested           3.269811e-03
## interest9. DK                           9.814626e-01
## actualDemocratic                         5.762947e-01

```

```

glm_prob = predict(anes_glm,type="response",newdata=anes_test)
glm_pred = rep("0",dim(anes_test)[1])
glm_pred[glm_prob>0.5]="1"
cm = table(glm_pred,anes_test$changed_votes)
conf_mat = data.frame(matrix(c(2243, 17, 332, 16),ncol=2))
colnames(conf_mat) = c("True 0","True 1")
row.names(conf_mat) = c("Predicted 0", "Predicted 1")
print(conf_mat)

```

```

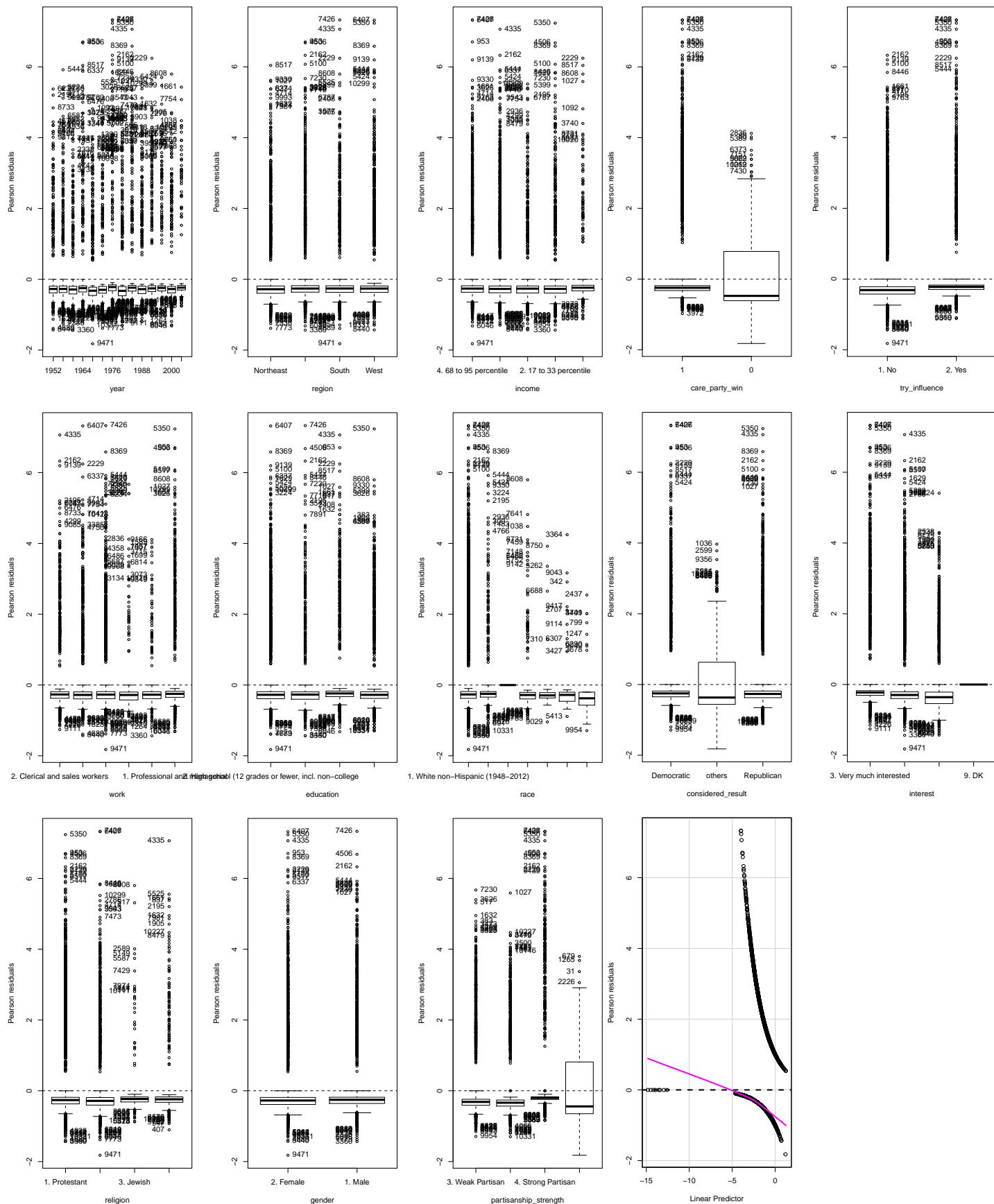
##           True 0 True 1
## Predicted 0    2243    332
## Predicted 1      17     16

```

```

par(mfrow=c(4,5))
library(car)
residualPlots(anes_glm,plot = TRUE)

```



```

##                                     Test stat Pr(>|Test stat|)
## year
## region
## income
## work
## education
## race
## religion
## gender
## partisanship_strength
## care_party_win
## try_influence
## days_discuss          0.1401      0.7082
## considered_result
## interest
## actual

```

```

# Hosmer Lemeshow Goodness of Fit test
hoslem.test(anes_glm$y, anes_glm$fitted)

```

```

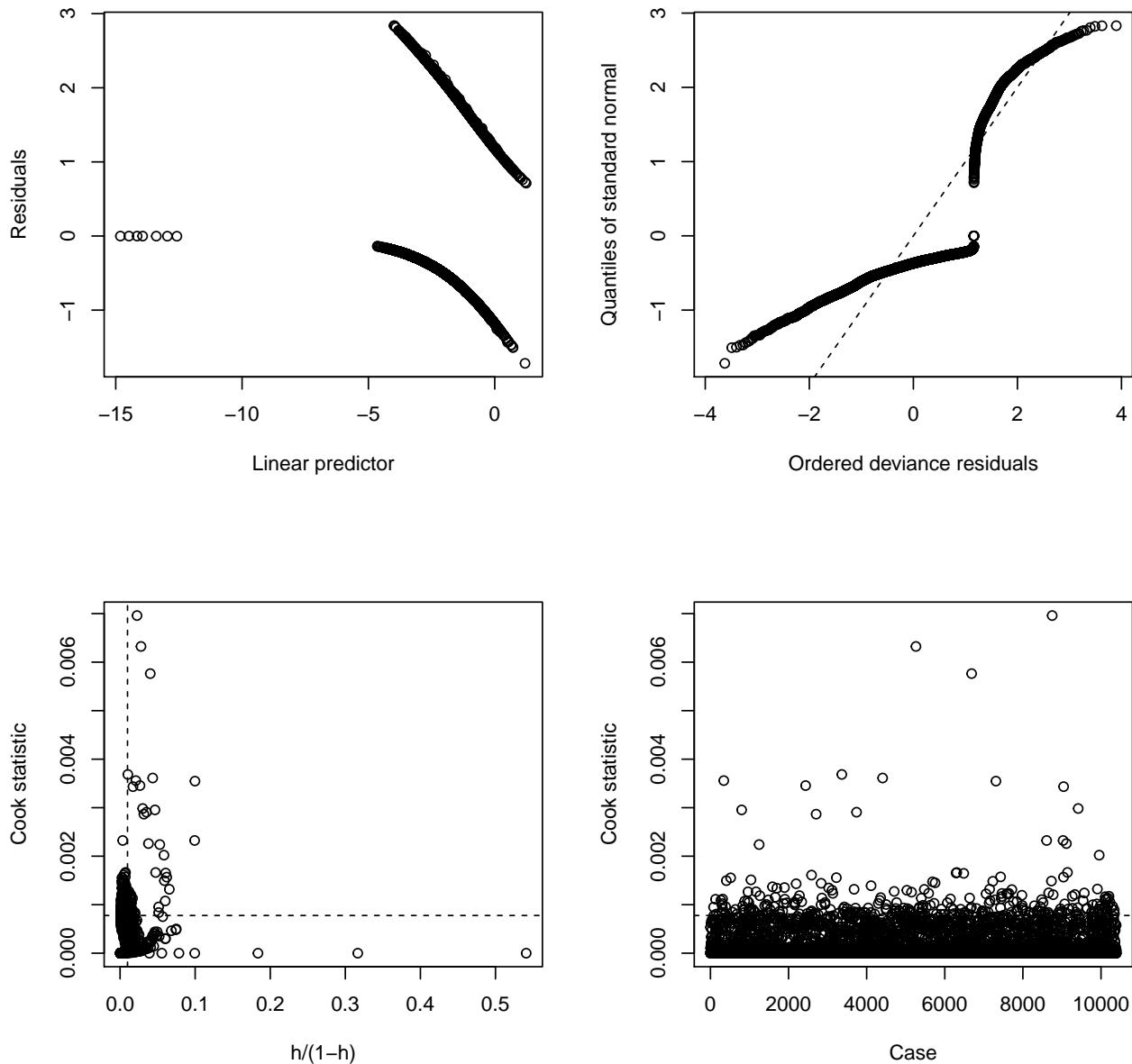
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data: anes_glm$y, anes_glm$fitted
## X-squared = 9.9127, df = 8, p-value = 0.2712

```

```

glm.diag.plots(anes_glm)

```



```
mean(glm_pred$anes_test$changed_votes)
```

```
## [1] 0.869515
```

From the summary of logistic regression model we could observe that years 1968, 1976, 1984, American Indian or Alaska Native non-Hispanic race and partisanship, try\_influence, considered\_result others and not much interested about politics are relatively significant. After AIC model selection, it seems that 8 variables might have impacted changes of votes: gender, religion, interest, try\_influence, year, considered\_result, partisanship\_strength and care\_party\_win. There are 28 coefficients related to our final model after AIC model selection including intercept. However, it seems that only 5 variables: year, partisanship\_strength, care\_party\_win, try\_influence, considered\_result are significant with less variables and coefficient selected comparing to AIC for final model.

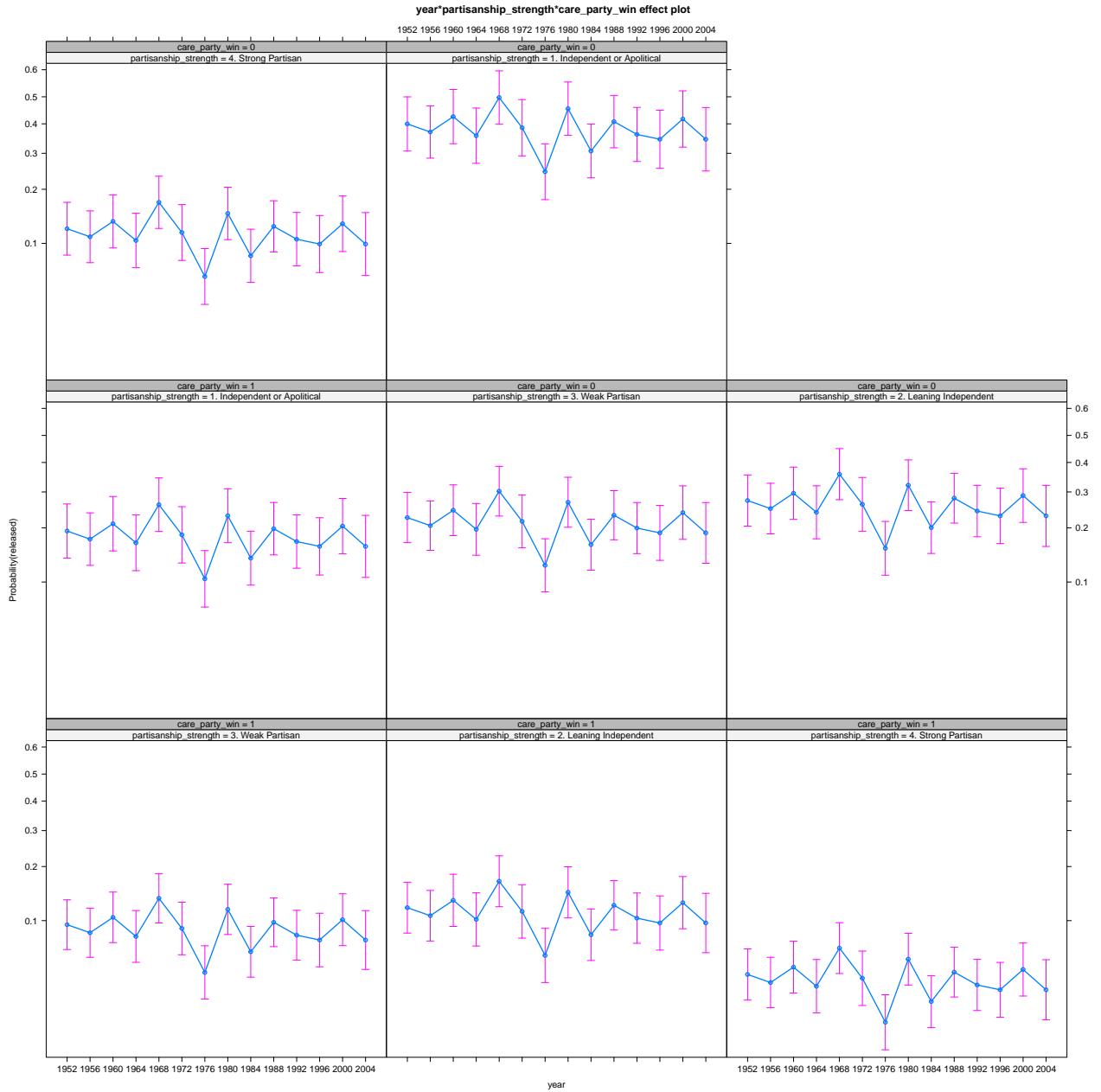
Since all variables in our model are categorical variables except days\_discuss, boxplots for all categorical variable seems difficult to interpret because of the discreteness in the distribution of the residuals. For subplot of partisanship\_strength, independent or apolitical seems to have a large IQR for Pearson Residuals than

other levels of partisanship strength and respondents who do not care about which party wins presidential election seems to have a large IQR for Pearson Residuals than respondents who care about which party wins presidential election, respondent's opinion of which party's candidate will be elected president in November with others have large IQR for Pearson Residuals than Democratic and Republican groups.

From the result of Hosmer Lemeshow Goodness of Fit test, our p-value reported as 0.2712, which is relatively large, so we can conclude that our logistic regression model is not a poor fit. Although diagnostic plots looks not very clean as usual since almost every variables in our model are categorical variables and data includes many biases as I mentioned, relatively speaking, our model is somewhat still valid to some extent that the accuracy of our model is around 0.87 which also implies that our model performance is not bad.

```
plot(effect("year:partisanship_strength:care_party_win",
            anes_glm,multiline=TRUE, ylab="Probability(released)",
            rug=FALSE),
     xaxt = "n", yaxt = "n",cex.lab=1.5, cex.axis=1.5, cex.main=1.5,
     cex.sub=1.5,ylab="Probability(released)")

## NOTE: year:partisanship_strength:care_party_win does not appear in the model
```



Here, we try to visualize if there is any interaction significant factors in our model including year, partisanship\_strength, and care\_party\_win in the effect plots. The general pattern of subplots are quite similar, with year 1976 has lowest probability and year 1968 has highest probability. Among respondents who care about which party wins presidential election, independent or apolitical in terms of partisanship strength have a higher probability among other partisanship strength groups, however, the probability is not large. Among respondents who do not care about which party wins presidential election, independent or apolitical in terms of partisanship strength also have a higher probability among other partisanship strength groups, but the probability for respondents who do not care about which party wins presidential election seems have a higher probability for all partisanship strength levels than corresponding partisanship strength levels with respondents who care about which party wins presidential election.

4.3 For those who intend to have a stronger strength of partisanship in election, do they have propensity for trying to influence the vote of others during the campaign?

```
# contingency table
print(table(anes$partisanship_strength, anes$try_influence))
```

```
##
##                                     1. No 2. Yes
##   3. Weak Partisan            3042 1543
##   2. Leaning Independent      1527 1072
##   4. Strong Partisan          2640 2292
##   1. Independent or Apolitical 598  275
```

```
# Chi-squared test
chisq.test(anes$partisanship_strength, anes$try_influence)
```

```
##
## Pearson's Chi-squared test
##
## data: anes$partisanship_strength and anes$try_influence
## X-squared = 191.1, df = 3, p-value < 2.2e-16
```

Here, we have a  $\chi^2$  value of 191.1 for Chi-squared test. Since we get a p-value of less than the significance level of 0.05, we can reject the null hypothesis and conclude that the two variables partisanship\_strength and try\_influence are, indeed, independent. However, problem with Pearson's  $\chi^2$  coefficient is that the range of its maximum value depends on the sample size and the size of the contingency table. These values may vary in different situations.

```
library(DescTools)
x1 = ContCoef(anes$partisanship_strength, anes$try_influence, correct = FALSE)
#Corrected contingency coefficient
x2 = ContCoef(anes$partisanship_strength, anes$try_influence, correct = TRUE)

library(lsr)
x3 = cramersV(anes$partisanship_strength, anes$try_influence)

library(rcompanion)
x4 = cramerV(anes$partisanship_strength, anes$try_influence, bias.correct = TRUE)
tbl = data.frame(matrix(c(x1,x2,x3,x4),ncol=2))
colnames(tbl) = c('Contingency Coefficient','Cramer's V')
row.names(tbl) = c('Original','Corrected')
print(tbl)

##           Contingency Coefficient Cramer's V
## Original             0.1204127  0.1212953
## Corrected            0.1702893  0.1203000
```

From above statistics we can see that the strength of association between the strength of partisanship from respondent's party identification and whether respondent try to influence the vote of others is very small.

```

df <- data.frame(
  partisanship = as.character(anes$partisanship_strength),
  influence = as.character(anes$try_influence),
  care = as.character(anes$care_party_win),
  years = as.character(anes$year),
  result = as.character(anes$considered_result)
)

# function to get chi square p value and Cramers V
f = function(x,y) {
  tbl = df %>% select(x,y) %>% table()
  cramV = round(cramersV(tbl), 4)
  data.frame(x, y, cramV) }

# create unique combinations of column names
# sorting will help getting a better plot (upper triangular)
df_comb = data.frame(t(combn(sort(names(df)), 2)), stringsAsFactors = F)

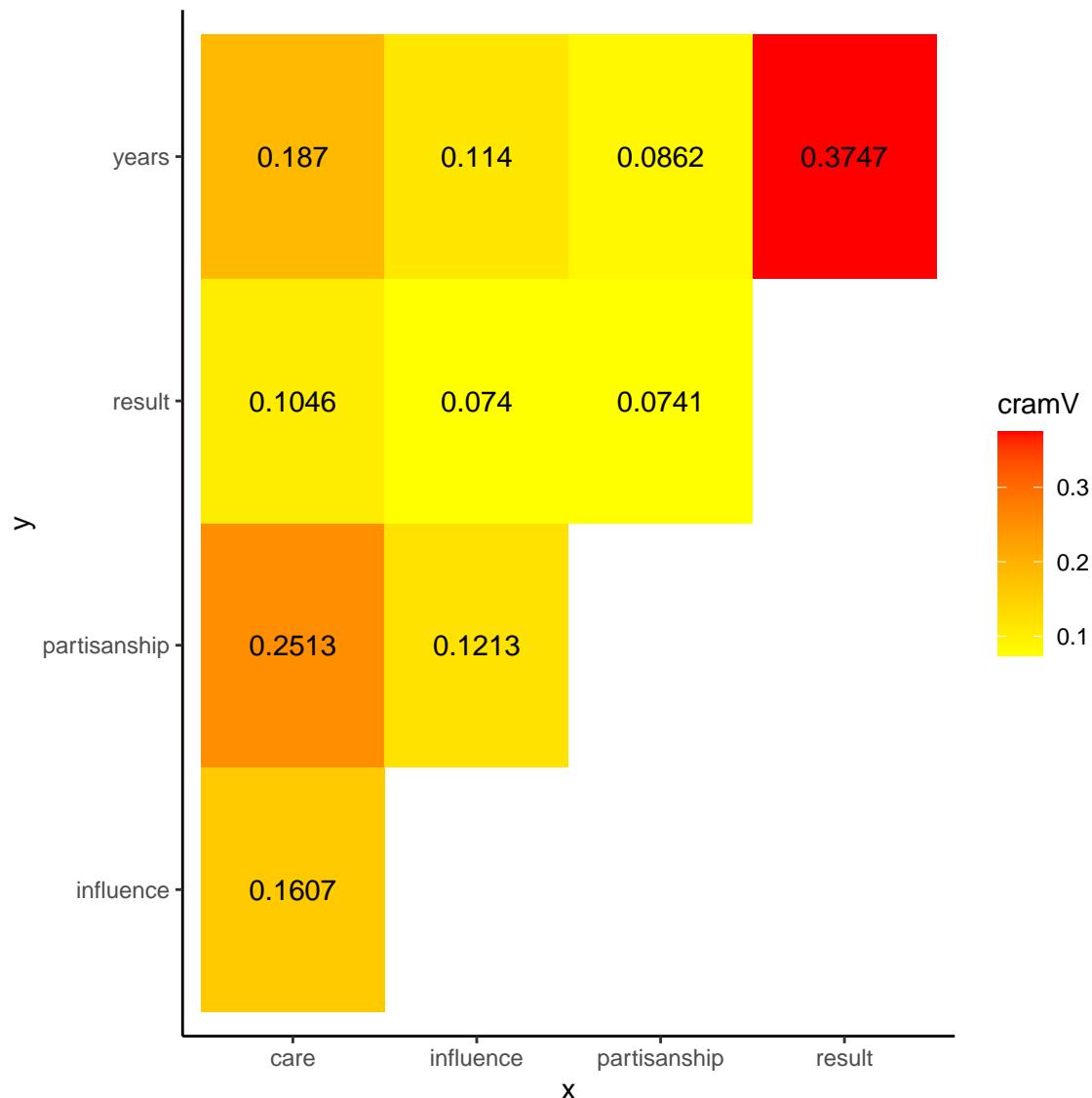
# apply function to each variable combination
df_res = map2_df(as.character(df_comb$X1), as.character(df_comb$X2), f)

## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(x)` instead of `x` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.

## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(y)` instead of `y` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.

# plot results
df_res %>%
  ggplot(aes(x,y,fill=cramV))+
  geom_tile()+
  geom_text(aes(x,y,label=cramV))+
  scale_fill_gradient(low="yellow",high="red")+
  theme_classic()

```



From the plot above with several significant factors from BIC result, we could observe that the year and whether respondent care which party wins presidential election have relatively stronger association but not strong enough in general, and respondent's opinion of which party's candidate will be elected president in November have relatively weak associations with both whether respondent care which party wins presidential election and if respondent try to influence the vote of others during the campaign.