Data cleaning

Jenny

January 2019

Introduction

This file documents the data cleaning for the Representation Project

Data cleaning - Cumulative

Read in the data.

```
tbl <- read_rds("../Data/cumulative_2006_2017.Rds")
```

Take a look at the data

```
head(tbl)
```

```
## # A tibble: 6 x 70
##    year case_id weight weight_cumulati~ state st      cd    dist dist_up
##   <int>   <int>  <dbl>            <dbl> <chr> <chr> <S3:> <int>   <int>
## 1  2006  439219  1.85             1.35  Nort~ NC    NC-10    10      10
## 2  2006  439224  0.968            0.704 Ohio  OH    OH-3      3       3
## 3  2006  439228  1.59             1.16  New ~ NJ    NJ-1      1       1
## 4  2006  439237  1.40             1.02  Illi~ IL    IL-9      9       9
## 5  2006  439238  0.903            0.656 New ~ NY    NY-22    22      22
## 6  2006  439242  0.839            0.610 Texas TX    TX-11    11      11
## # ... with 61 more variables: cong <int>, cong_up <int>, zipcode <chr>,
## #   county_fips <chr>, tookpost <int>, weight_post <dbl>,
## #   starttime <dttm>, pid3 <int>, pid7 <int>, pid3_leaner <int>,
## #   ideo5 <fct>, gender <int>, birthyr <int>, age <int>, race <int>,
## #   hispanic <int>, educ <int>, faminc <fct>, economy_retro <int>,
## #   approval_pres <int>, approval_rep <fct>, approval_sen1 <fct>,
## #   approval_sen2 <fct>, approval_gov <int>, intent_pres_08 <fct>,
## #   intent_pres_12 <fct>, intent_pres_16 <fct>, voted_pres_08 <fct>,
## #   voted_pres_12 <fct>, voted_pres_16 <fct>, vv_regstatus <fct>,
## #   vv_party_gen <fct>, vv_party_prm <fct>, vv_turnout_gvm <fct>,
## #   vv_turnout_pvm <fct>, intent_rep <fct>, intent_sen <fct>,
## #   intent_gov <fct>, voted_rep <fct>, voted_sen <fct>, voted_gov <fct>,
## #   intent_rep_chosen <chr>, intent_rep_fec <chr>,
## #   intent_sen_chosen <chr>, intent_sen_fec <chr>,
## #   intent_gov_chosen <chr>, intent_gov_fec <chr>, voted_rep_chosen <chr>,
## #   voted_rep_fec <chr>, voted_sen_chosen <chr>, voted_sen_fec <chr>,
## #   voted_gov_chosen <chr>, voted_gov_fec <chr>, rep_current <chr>,
## #   rep_icpsr <int>, sen1_current <chr>, sen1_icpsr <int>,
## #   sen2_current <chr>, sen2_icpsr <int>, gov_current <chr>, gov_fec <chr>
```

Another way to look at the data, where we can clearly see all variable names and types.

```
tbl %>% glimpse()
```

```
## Observations: 392,755
## Variables: 70
```

```
## $ year            <int> 2006, 2006, 2006, 2006, 2006, 2006, 2006, 20...
## $ case_id         <int> 439219, 439224, 439228, 439237, 439238, 4392...
## $ weight          <dbl> 1.8516757, 0.9683084, 1.5934412, 1.3985290, ...
## $ weight_cumulative <dbl> 1.3460123, 0.7038787, 1.1582975, 1.0166128, ...
## $ state           <chr> "North Carolina", "Ohio", "New Jersey", "Ill...
## $ st              <chr> "NC", "OH", "NJ", "IL", "NY", "TX", "MN", "N...
## $ cd              <S3: glue> "NC-10", "OH-3", "NJ-1", "IL-9", "NY-22...
## $ dist            <int> 10, 3, 1, 9, 22, 11, 3, 2, 24, 2, 9, 5, 1, 1...
## $ dist_up         <int> 10, 3, 1, 9, 22, 11, 3, 2, 24, 2, 9, 5, 1, 1...
## $ cong            <int> 109, 109, 109, 109, 109, 109, 109, 109, 109,...
## $ cong_up         <int> 110, 110, 110, 110, 110, 110, 110, 110, 110,...
## $ zipcode         <chr> "28645", "45409", "08030", "60613", "12783",...
## $ county_fips     <chr> "37027", "39113", "34007", "17031", "36105",...
## $ tookpost        <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0,...
## $ weight_post     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ starttime       <dttm> 2006-10-07 00:02:34, 2006-10-07 00:02:53, 2...
## $ pid3            <int> 1, 4, 1, 1, 1, 3, 2, 1, 1, 1, 2, 2, 3, 3, 2,...
## $ pid7            <int> 1, 3, 1, 1, 1, 3, 7, 1, 1, 1, 7, 6, 5, 5, 6,...
## $ pid3_leaner     <int> 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 2, 2, 2, 2, 2,...
## $ ideo5           <fct> Liberal, Moderate, Liberal, Liberal, Liberal...
## $ gender          <int> 2, 1, 2, 2, 1, 2, 1, 2, 2, 1, 1, 2, 1, 2, 1,...
## $ birthyr         <int> 1974, 1957, 1952, 1972, 1986, 1979, 1959, 19...
## $ age             <int> 32, 49, 54, 34, 20, 27, 47, 20, 77, 19, 53, ...
## $ race            <int> 1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1,...
## $ hispanic        <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ educ            <int> 2, 6, 2, 5, 3, 3, 4, 3, 3, 3, 3, 2, 4, 2, 2,...
## $ faminc          <fct> 10k - 20k, 150k+, 30k - 40k, Less than 10k, ...
## $ economy_retro   <int> 4, 5, 5, 3, 3, 3, 1, 4, 5, 3, 1, 4, 1, 1, 5,...
## $ approval_pres   <int> 4, 4, 4, 4, 4, 4, 1, 4, 4, 3, 2, 4, 2, 1, 4,...
## $ approval_rep    <fct> Strongly Disapprove, Disapprove / Somewhat D...
## $ approval_sen1   <fct> Disapprove / Somewhat Disapprove, Strongly D...
## $ approval_sen2   <fct> Strongly Disapprove, Disapprove / Somewhat D...
## $ approval_gov    <int> 2, 3, 1, 2, 3, 5, 2, 2, 4, 2, 2, 3, 3, 1, 2,...
## $ intent_pres_08  <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ intent_pres_12  <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ intent_pres_16  <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ voted_pres_08   <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ voted_pres_12   <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ voted_pres_16   <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ vv_regstatus    <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ vv_party_gen    <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ vv_party_prm    <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ vv_turnout_gvm  <fct> Voted, Voted, No Record Of Voting, Voted, No...
## $ vv_turnout_pvm  <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ intent_rep      <fct> [Democrat / Candidate 1], [Democrat / Candid...
## $ intent_sen      <fct> NA, [Democrat / Candidate 1], [Democrat / Ca...
## $ intent_gov      <fct> NA, [Democrat / Candidate 1], NA, [Democrat ...
## $ voted_rep       <fct> [Democrat / Candidate 1], [Democrat / Candid...
## $ voted_sen       <fct> NA, [Democrat / Candidate 1], [Democrat / Ca...
## $ voted_gov       <fct> NA, [Democrat / Candidate 1], NA, [Democrat ...
## $ intent_rep_chosen <chr> "Richard C. Carsner (D)", "Stephanie Studeba...
## $ intent_rep_fec  <chr> "H6NC10141", "H6OH03142", "H0NJ01066", "H8IL...
## $ intent_sen_chosen <chr> NA, "Sherrod C. Brown (D)", "Robert Menendez...
## $ intent_sen_fec  <chr> NA, "S6OH00163", "S6NJ00289", NA, NA, NA, "S...
```

```
## $ intent_gov_chosen <chr> NA, "Ted Strickland (D)", NA, "Rod Blagojevi...
## $ intent_gov_fec    <chr> NA, "OH19691", NA, "IL7", "NY19490", NA, "MN...
## $ voted_rep_chosen  <chr> "Richard C. Carsner (D)", "Stephanie Studeba...
## $ voted_rep_fec     <chr> "H6NC10141", "H6OH03142", "H0NJ01066", "H8IL...
## $ voted_sen_chosen  <chr> NA, "Sherrod C. Brown (D)", "Robert Menendez...
## $ voted_sen_fec     <chr> NA, "S6OH00163", "S6NJ00289", NA, "S0NY00188...
## $ voted_gov_chosen  <chr> NA, "Ted Strickland (D)", NA, "Rod Blagojevi...
## $ voted_gov_fec     <chr> NA, "OH19691", NA, "IL7", "NY19490", NA, "MN...
## $ rep_current       <chr> "Patrick T. McHenry (R)", "Michael R. Turner...
## $ rep_icpsr         <int> 20522, 20342, 29132, 29911, 29380, 20531, 29...
## $ sen1_current      <chr> "Elizabeth Dole (R)", "Mike DeWine (R)", "Ro...
## $ sen1_icpsr        <int> 40303, 15020, 29373, 15021, 14858, 49306, 40...
## $ sen2_current      <chr> "Richard Burr (R)", "George V. Voinovich (R)...
## $ sen2_icpsr        <int> 29548, 49903, 14914, 40502, 40105, 40305, 40...
## $ gov_current       <chr> "Michael Easley (D)", "Bob Taft (R)", "Jon C...
## $ gov_fec           <chr> "NC5998", NA, "NJ6395", "IL7", NA, "TX3156",...
```

Select variables (from Kuriwaki Guide)

```
tbl %>%
  select(year, case_id, pid3)
```

```
## # A tibble: 392,755 x 3
##     year case_id  pid3
##    <int>   <int> <int>
## 1   2006  439219     1
## 2   2006  439224     4
## 3   2006  439228     1
## 4   2006  439237     1
## 5   2006  439238     1
## 6   2006  439242     3
## 7   2006  439251     2
## 8   2006  439254     1
## 9   2006  439255     1
## 10  2006  439263     1
## # ... with 392,745 more rows
```

Display frequencies of validated turnout General Election over time - shows only first few rows in console

```
tbl %>%
  group_by(year) %>%
  count(vv_turnout_gvm)
```

```
## # A tibble: 21 x 3
## # Groups:   year [12]
##     year vv_turnout_gvm          n
##    <int> <fct>               <int>
## 1   2006 Voted               15575
## 2   2006 No Record Of Voting 20489
## 3   2006 No Voter File         357
## 4   2007 <NA>                 9999
## 5   2008 Voted               22235
## 6   2008 No Record Of Voting 10520
## 7   2008 No Voter File          45
## 8   2009 <NA>                13800
## 9   2010 Voted               33854
```

```
## 10  2010 No Record Of Voting 20215
## # ... with 11 more rows
```

Display frequencies of validated turnout General Election over time - shows all rows in console in `kable` format to print in a clean formatted table.

```r
tbl %>%
  group_by(year) %>%
  count(vv_turnout_gvm) %>%
  kable()
```

year	vv_turnout_gvm	n
2006	Voted	15575
2006	No Record Of Voting	20489
2006	No Voter File	357
2007	NA	9999
2008	Voted	22235
2008	No Record Of Voting	10520
2008	No Voter File	45
2009	NA	13800
2010	Voted	33854
2010	No Record Of Voting	20215
2010	No Voter File	1331
2011	NA	20150
2012	Voted	36402
2012	No Record Of Voting	18133
2013	NA	16400
2014	Voted	25309
2014	No Record Of Voting	30891
2015	NA	14250
2016	Voted	35829
2016	No Record Of Voting	28771
2017	NA	18200

Recode of vv_turnout_gvm to dichotomous

```r
tbl <- tbl %>%
  mutate(vv_turnout_gvm_binary = as.numeric(vv_turnout_gvm == "Voted"))
```

Display frequencies of recoded Validated turnout General Election over time

```r
tbl %>%
  group_by(year) %>%
  count(vv_turnout_gvm_binary)
```

```
## # A tibble: 18 x 3
## # Groups:   year [12]
##     year vv_turnout_gvm_binary     n
##    <int>                 <dbl> <int>
## 1  2006                     0 20846
## 2  2006                     1 15575
## 3  2007                    NA  9999
## 4  2008                     0 10565
## 5  2008                     1 22235
## 6  2009                    NA 13800
```

```
##  7  2010                    0 21546
##  8  2010                    1 33854
##  9  2011                   NA 20150
## 10  2012                    0 18133
## 11  2012                    1 36402
## 12  2013                   NA 16400
## 13  2014                    0 30891
## 14  2014                    1 25309
## 15  2015                   NA 14250
## 16  2016                    0 28771
## 17  2016                    1 35829
## 18  2017                   NA 18200
```

Display table, frequencies of recoded Validated turnout General Election over time

```
tbl %>%
  group_by(year) %>%
  count(vv_turnout_gvm_binary) %>%
  kable(align = c("l", "c", "c"))
```

year	vv_turnout_gvm_binary	n
2006	0	20846
2006	1	15575
2007	NA	9999
2008	0	10565
2008	1	22235
2009	NA	13800
2010	0	21546
2010	1	33854
2011	NA	20150
2012	0	18133
2012	1	36402
2013	NA	16400
2014	0	30891
2014	1	25309
2015	NA	14250
2016	0	28771
2016	1	35829
2017	NA	18200

Descriptive Statistics

Year:

```
tbl %>%
  count(year)
```

```
## # A tibble: 12 x 2
##     year     n
##    <int> <int>
## 1   2006 36421
## 2   2007  9999
## 3   2008 32800
## 4   2009 13800
```

```
## 5   2010 55400
## 6   2011 20150
## 7   2012 54535
## 8   2013 16400
## 9   2014 56200
## 10  2015 14250
## 11  2016 64600
## 12  2017 18200
```

tookpost - "Whether or not the respondent took the post-election wave of the survey (in even years)"

```r
tbl %>%
  group_by(year) %>%
  count(tookpost) %>%
  filter(!is.na(tookpost)) %>%
  mutate(percent = round((n / sum(n)) * 100, 2))
```

```
## # A tibble: 12 x 4
## # Groups:   year [6]
##     year tookpost     n percent
##    <int>    <int> <int>   <dbl>
## 1  2006        0  7664    21.0
## 2  2006        1 28757    79.0
## 3  2008        0  5779    17.6
## 4  2008        1 27021    82.4
## 5  2010        0  8716    15.7
## 6  2010        1 46684    84.3
## 7  2012        0  9700    17.8
## 8  2012        1 44835    82.2
## 9  2014        0  7312    13.0
## 10 2014        1 48888    87.0
## 11 2016        0 11701    18.1
## 12 2016        1 52899    81.9
```

Weight - year-specific for all years, see Kuriwaki p.7 for notes

```r
tbl %>%
  group_by(year) %>%
  summarize(mean = mean(weight, na.rm = TRUE),
            sd   = sd(weight, na.rm = TRUE),
            min  = min(weight),
            max  = max(weight)
  )
```

```
## # A tibble: 12 x 5
##     year  mean    sd       min   max
##    <int> <dbl> <dbl>     <dbl> <dbl>
## 1  2006 1.000 0.530 0.331      2.92
## 2  2007 1.000 0.959 0.379      3.41
## 3  2008 1.    0.775 0.299      6.49
## 4  2009 1.000 0.675 0.115      5.03
## 5  2010 1.000 1.26  0          7.03
## 6  2011 1     1.09  0.0863     7.00
## 7  2012 1     1.17  0.0000100 15.0
## 8  2013 1     1.02  0.000100  12.0
## 9  2014 1     1.17  0.000100  15.0
```

```
## 10   2015 1.000 0.946 0.000112   10.0
## 11   2016 1     0.979 0.000100   15.0
## 12   2017 1     1.07  0.000100   10.0
```

Weight - cumulative. Includes simple adjustment of multiplying a constant within year to make years comparable.

```
tbl %>%
  group_by(year) %>%
  summarize(mean = mean(weight_cumulative, na.rm = TRUE),
            sd   = sd(weight_cumulative, na.rm = TRUE),
            min  = min(weight_cumulative),
            max  = max(weight_cumulative)
  )
```

```
## # A tibble: 12 x 5
##     year  mean    sd          min   max
##    <int> <dbl> <dbl>        <dbl> <dbl>
##  1  2006 0.727 0.385 0.241         2.12
##  2  2007 2.65  2.54  1.00          9.03
##  3  2008 0.807 0.626 0.241         5.24
##  4  2009 1.92  1.29  0.220         9.65
##  5  2010 0.478 0.600 0             3.36
##  6  2011 1.31  1.44  0.113         9.20
##  7  2012 0.485 0.570 0.00000485    7.28
##  8  2013 1.61  1.64  0.000162     19.4
##  9  2014 0.471 0.550 0.0000472     7.07
## 10  2015 1.86  1.76  0.000208     18.6
## 11  2016 0.410 0.401 0.0000410     6.15
## 12  2017 1.45  1.56  0.000145     14.5
```

Geographic variables *state - imputed from input zipcode* st - same data as "State", Var name = 2-letter State abbreviation

```
tbl %>%
  count(state, st)
```

```
## # A tibble: 51 x 3
##    state                st        n
##    <chr>                <chr> <int>
##  1 Alabama              AL     5106
##  2 Alaska               AK      881
##  3 Arizona              AZ     9928
##  4 Arkansas             AR     3703
##  5 California           CA    37821
##  6 Colorado             CO     6824
##  7 Connecticut          CT     4622
##  8 Delaware             DE     1455
##  9 District of Columbia DC      878
## 10 Florida              FL    28879
## # ... with 41 more rows
```

Geographic variables *cd: Congressional district in current Congress

```
tbl %>%
  count(st, cd)
```

```
## # A tibble: 448 x 3
```

```
##     st    cd              n
##     <chr> <S3: glue> <int>
##   1 AK    AK-1         881
##   2 AL    AL-1         723
##   3 AL    AL-2         754
##   4 AL    AL-3         692
##   5 AL    AL-4         628
##   6 AL    AL-5         828
##   7 AL    AL-6         765
##   8 AL    AL-7         716
##   9 AR    AR-1         836
## 10 AR    AR-2         962
## # ... with 438 more rows
```

Gender

```r
tbl %>%
  count(gender)
```

```
## # A tibble: 2 x 2
##   gender      n
##    <int>  <int>
## 1      1 184273
## 2      2 208482
```

Age

```r
tbl %>%
  summarize(mean = mean(age),
            sd   = sd(age),
            min  = min(age),
            max  = max(age)
  ) %>%
  mutate(variable = "age") %>%
  select(variable, everything())
```

```
## # A tibble: 1 x 5
##   variable  mean    sd   min   max
##   <chr>    <dbl> <dbl> <dbl> <dbl>
## 1 age       49.8  16.2    18   109
```
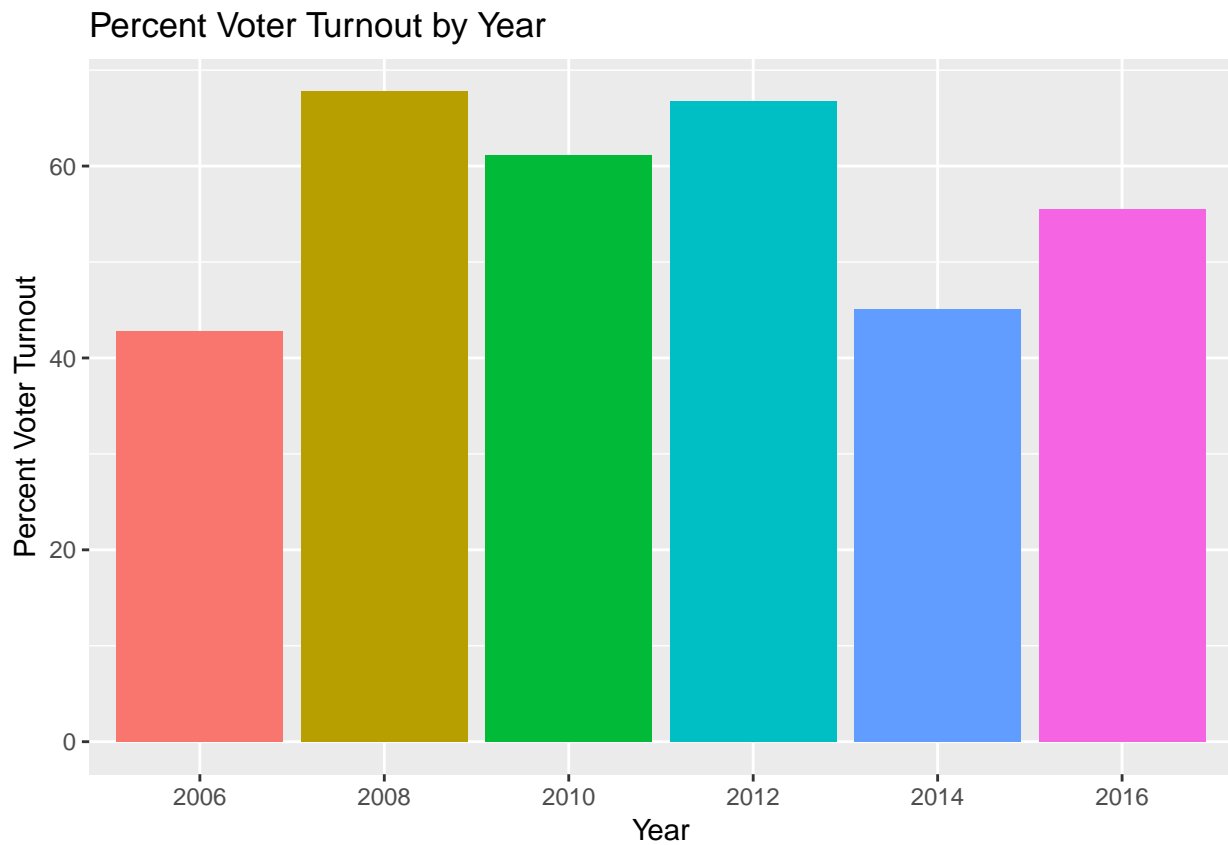
Figures

Plot turnout dichotomous by year via bar charts:

```r
# counts
tbl %>%
  filter(year %% 2 != 1) %>%          # filter to even years
  mutate(year = as.character(year)) %>%
  ggplot(aes(x = year, y = vv_turnout_gvm_binary, colour = year)) +
  geom_bar(stat = "identity") +
  labs(x = "Year", y = "Voter Turnout", title = "Voter Turnout by Year") +
  theme(legend.position = "none")
```

Voter Turnout by Year
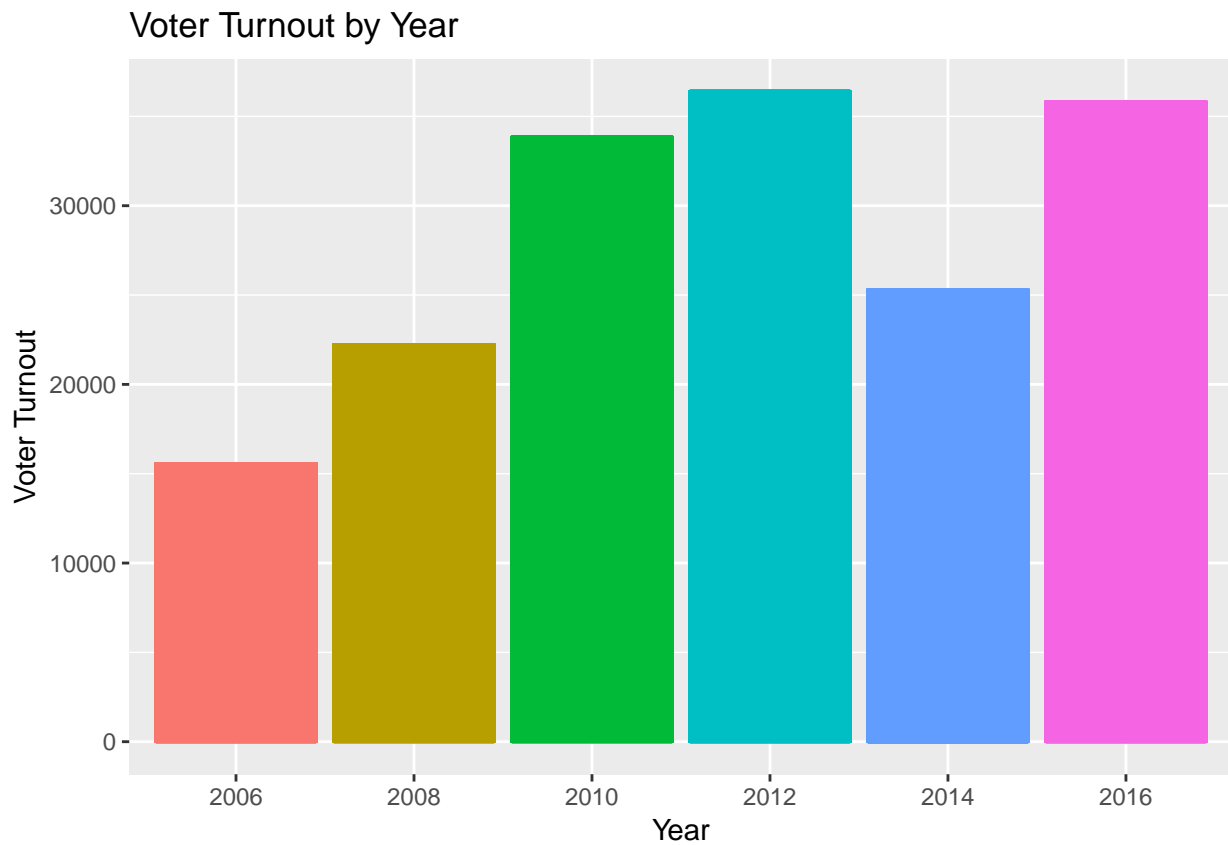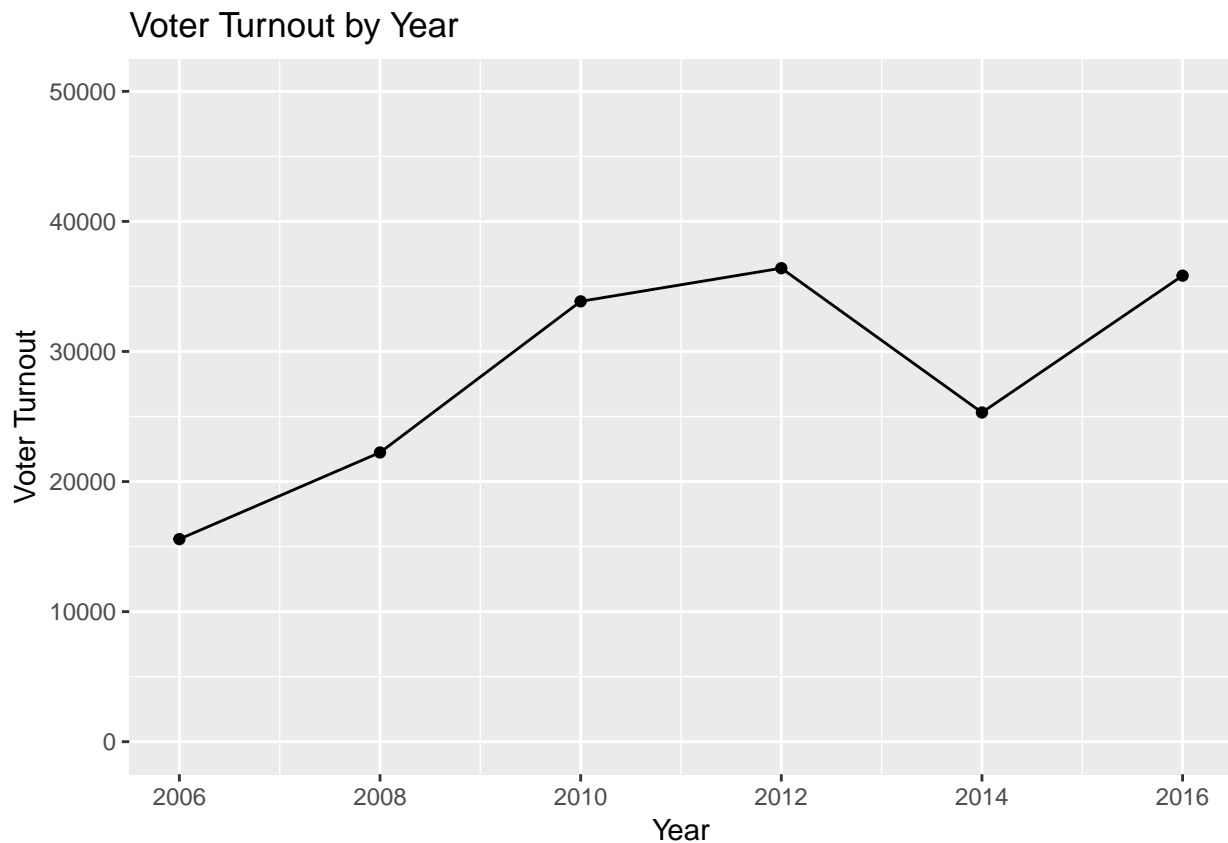


```
# percent
tbl %>%
  filter(year %% 2 != 1) %>%          # filter to even years
  mutate(year = as.character(year)) %>%
  group_by(year) %>%
  summarize(percent = mean(vv_turnout_gvm_binary, na.omit = TRUE) * 100) %>%
  ggplot(aes(x = year, y = percent, fill = year)) +
  geom_bar(stat = "identity") +
  labs(x = "Year", y = "Percent Voter Turnout", title = "Percent Voter Turnout by Year") +
  theme(legend.position = "none")
```

Percent Voter Turnout by Year

Plot turnout dichotomous by year via line plots:

```r
# counts
tbl %>%
  group_by(year) %>%
  count(vv_turnout_gvm_binary) %>%
  filter(vv_turnout_gvm_binary == 1) %>%
  ggplot(aes(x = year, y = n)) +
  geom_line() +
  geom_point() +
  ylim(0, 50000) +
  labs(x = "Year", y = "Voter Turnout", title = "Voter Turnout by Year")
```

Voter Turnout by Year


```r
# percents
tbl %>%
  filter(year %% 2 != 1) %>%        # filter to even years
  mutate(year = as.character(year)) %>%
  group_by(year) %>%
  summarize(percent = mean(vv_turnout_gvm_binary, na.omit = TRUE) * 100) %>%
  ungroup() %>%
  ggplot(aes(x = year, y = percent, group = 1)) +
  geom_point() +
  geom_line() +
  ylim(0, 100) +
  labs(x = "Year", y = "Percent Voter Turnout", title = "Percent Voter Turnout by Year")
```

Percent Voter Turnout by Year
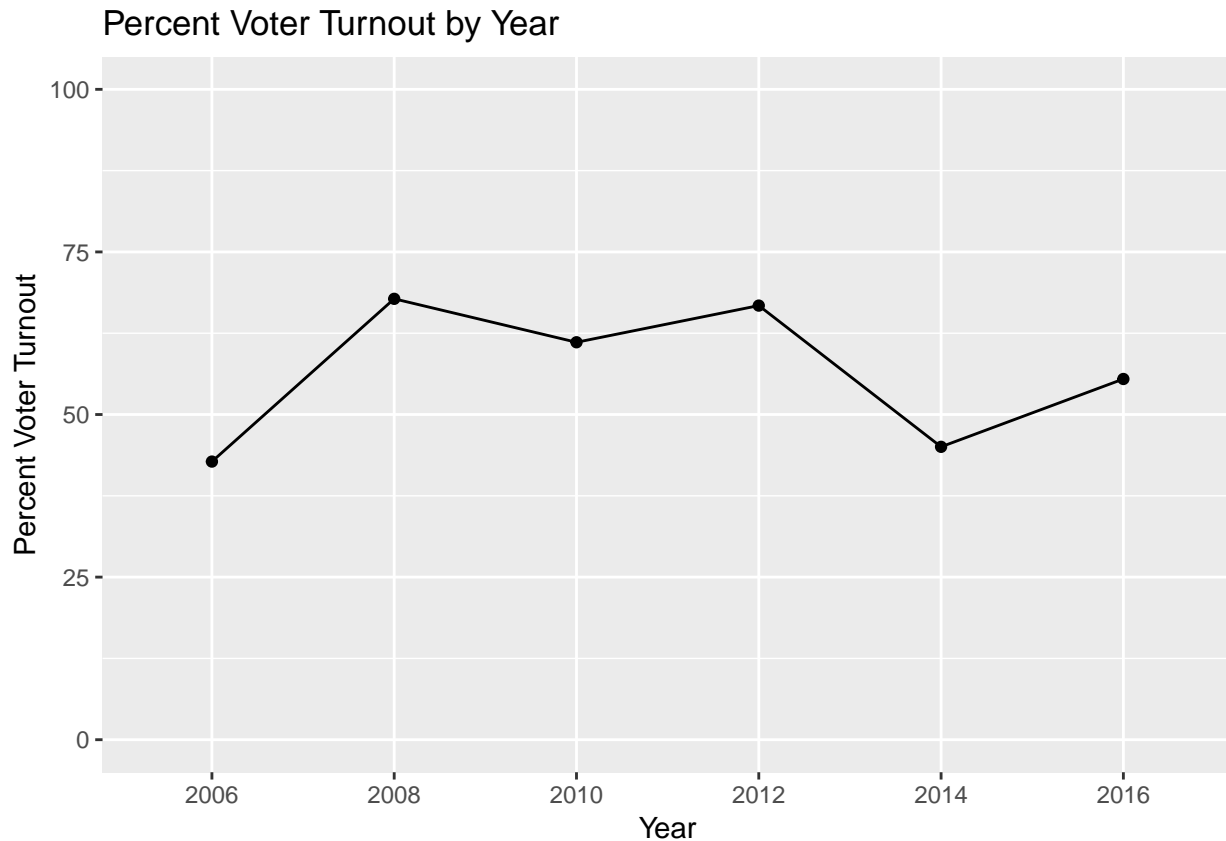
Data cleaning - 2012, replicating Perspectives findings with focus on Independents

JO dropbox link to Nov 2018 work on this topic in stata: CCES 2012_Russ Sage: https://www.dropbox.com/sh/fc7cn2fmaxocsyu/AADt5fFQ9jFXUQgcW65pVYvRa?dl=0

Read in 2012 cleaned data for Perspectives paper.

```
tbl2012 <- read_dta("../Data/CCES_foranalysis.dta")
```

Examine the variables we are interested in.

```
tbl2012 %>%
  select(healthcarerepeal1_House1,
         ryanbudget_House1,
         koreafreetrade_House1,
         simpsonbowles_House1,
         keystonepipeline_House1,
         dontaskdonttell_House1) %>%
  glimpse()
```

```
## Observations: 54,535
## Variables: 6
## $ healthcarerepeal1_House1 <chr> "For", "For", "For", "For", "For", "F...
## $ ryanbudget_House1        <chr> "For", "For", "For", "For", "For", "F...
## $ koreafreetrade_House1    <chr> "For", "For", "For", "For", "For", "F...
## $ simpsonbowles_House1     <chr> "For", "For", "For", "For", "For", "F...
## $ keystonepipeline_House1  <chr> "For", "For", "For", "For", "For", "F...
```

```
## $ dontaskdonttell_House1    <chr> "Yea", "Yea", "Yea", "Yea", "Yea", "Y...
```

Some descriptive statistics.

```r
tbl2012_vars <- tbl2012 %>%
  select(healthcarerepeal1_House1,
         ryanbudget_House1,
         koreafreetrade_House1,
         simpsonbowles_House1,
         keystonepipeline_House1,
         dontaskdonttell_House1) %>%
  colnames()

tbl2012_vars %>%
  map(~ tbl2012 %>%
        count(!!sym(.x)) %>%
        mutate(percent = round((n / sum(n)) * 100, 2)))
```

```
## [[1]]
## # A tibble: 4 x 3
##   healthcarerepeal1_House1     n percent
##   <chr>                    <int>   <dbl>
## 1 ""                         109     0.2
## 2 Against                  21699    39.8
## 3 Did Not Vote               216     0.4
## 4 For                      32511    59.6
##
## [[2]]
## # A tibble: 4 x 3
##   ryanbudget_House1     n percent
##   <chr>             <int>   <dbl>
## 1 ""                  439     0.8
## 2 Against           22282    40.9
## 3 Did Not Vote        557    1.02
## 4 For               31257    57.3
##
## [[3]]
## # A tibble: 4 x 3
##   koreafreetrade_House1     n percent
##   <chr>                 <int>   <dbl>
## 1 ""                      428    0.78
## 2 Against               17447    32.0
## 3 Did Not Vote            520    0.95
## 4 For                   36140    66.3
##
## [[4]]
## # A tibble: 5 x 3
##   simpsonbowles_House1     n percent
##   <chr>                <int>   <dbl>
## 1 ""                     679    1.25
## 2 Against              47854    87.8
## 3 Did Not Vote           965    1.77
## 4 For                   4774    8.75
## 5 Present                263    0.48
##
```

```
## [[5]]
## # A tibble: 4 x 3
##   keystonepipeline_House1     n percent
##   <chr>                   <int>   <dbl>
## 1 ""                        679    1.25
## 2 Against                 15051   27.6
## 3 Did Not Vote             1380    2.53
## 4 For                     37425   68.6
##
## [[6]]
## # A tibble: 7 x 3
##   dontaskdonttell_House1     n percent
##   <chr>                  <int>   <dbl>
## 1 ""                       109    0.2
## 2 .                        368    0.67
## 3 Announced Against        269    0.49
## 4 Announced For            145    0.27
## 5 Did not vote             926    1.7
## 6 Nay                    21403   39.2
## 7 Yea                    31315   57.4
```

```r
# counts of healthcare repeal votes in the House grouped by state district and party
tbl2012 %>%
  group_by(st_dist, party_House1) %>%
  count(healthcarerepeal1_House1)
```

```
## # A tibble: 436 x 4
## # Groups:   st_dist, party_House1 [436]
##    st_dist party_House1 healthcarerepeal1_House1     n
##    <chr>   <chr>        <chr>                    <int>
##  1 AK1     Republican   For                        128
##  2 AL1     Republican   For                        108
##  3 AL2     Republican   For                        126
##  4 AL3     Republican   For                        106
##  5 AL4     Republican   For                         85
##  6 AL5     Republican   For                        112
##  7 AL6     Republican   For                        113
##  8 AL7     Democratic   Against                     99
##  9 AR1     Republican   For                         97
## 10 AR2     Republican   For                        141
## # ... with 426 more rows
```

These are pretty messy, let's recode them.

```r
tbl2012 <- tbl2012 %>%
  mutate_at(tbl2012_vars, list(~ na_if(., y = ""))) %>%
  mutate(
    healthcarerepeal1_House1 = case_when(
      healthcarerepeal1_House1 == "For"     ~ "For",
      healthcarerepeal1_House1 == "Against" ~ "Against",
      TRUE ~ NA_character_
    ),
    ryanbudget_House1 = case_when(
      ryanbudget_House1 == "For"     ~ "For",
      ryanbudget_House1 == "Against" ~ "Against",
      TRUE ~ NA_character_
```

```r
  ),
  koreafreetrade_House1 = case_when(
    koreafreetrade_House1 == "For"     ~ "For",
    koreafreetrade_House1 == "Against" ~ "Against",
    TRUE ~ NA_character_
  ),
  simpsonbowles_House1 = case_when(
    simpsonbowles_House1 == "For"     ~ "For",
    simpsonbowles_House1 == "Against" ~ "Against",
    TRUE ~ NA_character_
  ),
  keystonepipeline_House1 = case_when(
    keystonepipeline_House1 == "For"     ~ "For",
    keystonepipeline_House1 == "Against" ~ "Against",
    TRUE ~ NA_character_
  ),
  dontaskdonttell_House1 = case_when(
    dontaskdonttell_House1 == "Announced For"     ~ "For",
    dontaskdonttell_House1 == "Announced Against" ~ "Against",
    dontaskdonttell_House1 == "Yea"     ~ "For",
    dontaskdonttell_House1 == "Nay" ~ "Against",
    TRUE ~ NA_character_
  )
)

# check recodes
tbl2012_vars %>%
  map(~ tbl2012 %>%
        count(!!sym(.x)) %>%
        mutate(percent = round((n / sum(n)) * 100, 2)))
```

```
## [[1]]
## # A tibble: 3 x 3
##   healthcarerepeal1_House1     n percent
##   <chr>                    <int>   <dbl>
## 1 Against                  21699    39.8
## 2 For                      32511    59.6
## 3 <NA>                       325     0.6
##
## [[2]]
## # A tibble: 3 x 3
##   ryanbudget_House1     n percent
##   <chr>             <int>   <dbl>
## 1 Against           22282    40.9
## 2 For               31257    57.3
## 3 <NA>                996    1.83
##
## [[3]]
## # A tibble: 3 x 3
##   koreafreetrade_House1     n percent
##   <chr>                 <int>   <dbl>
## 1 Against               17447    32.0
## 2 For                   36140    66.3
## 3 <NA>                    948    1.74
```

```
##
## [[4]]
## # A tibble: 3 x 3
##   simpsonbowles_House1     n percent
##   <chr>               <int>   <dbl>
## 1 Against             47854   87.8
## 2 For                  4774    8.75
## 3 <NA>                 1907    3.5
##
## [[5]]
## # A tibble: 3 x 3
##   keystonepipeline_House1     n percent
##   <chr>                   <int>   <dbl>
## 1 Against                 15051   27.6
## 2 For                     37425   68.6
## 3 <NA>                     2059    3.78
##
## [[6]]
## # A tibble: 3 x 3
##   dontaskdonttell_House1     n percent
##   <chr>                  <int>   <dbl>
## 1 Against                21672   39.7
## 2 For                    31460   57.7
## 3 <NA>                    1403    2.57
```

Data cleaning - Panel 2010-2012-2014

Panel data and Guide downloaded 1Feb2019: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:
10.7910/DVN/TOE8I1]

As reminder of prior work, link to prior panel data analyses in stata: CCES panel 2010-2012-2014: https:
//www.dropbox.com/sh/gj4vfv9xzvdgkwd/AABpmEqKQBT6EQDVCFvqeV7Ka?dl=0

I. Cleaning respondent data

Read in the data.

```
tbl_panel <- haven::read_dta("../Data/CCES_Panel_Full3waves_VV_V4.dta", encoding = "latin1")
```

Recode of relevant respondent measures Begin with 1 indicator as example: CC10_330A = Roll Call,
American Recovery and Reinvestment Act

```
tbl_panel %>%
  count(CC10_330A)
```

```
## # A tibble: 3 x 2
##   CC10_330A     n
##   <dbl+lbl> <int>
## 1 1          4616
## 2 2          4775
## 3 NA          109
```

See head of dataset in in console

```
head(tbl_panel)
```

```
## # A tibble: 6 x 1,631
```

```
##    caseid weight CC10_301_1 CC10_301_2 CC10_301_3 CC10_301_4 CC10_301_5
##     <dbl>  <dbl> <dbl+lbl>  <dbl+lbl>  <dbl+lbl>  <dbl+lbl>  <dbl+lbl>
## 1  25233  0.538 1          1          1          1          2
## 2  38716  3.47  1          2          1          1          2
## 3   7796  0.602 2          1          1          1          2
## 4  78113  0.264 1          1          1          1          2
## 5  63800  0.197 1          1          1          2          2
## 6  15002  0.204 1          1          1          1          2
## # ... with 1,624 more variables: CC10_301b <dbl+lbl>, CC10_301c <dbl+lbl>,
## #   CC10_302 <dbl+lbl>, CC10_304 <dbl+lbl>, CC10_305 <dbl+lbl>,
## #   CC10_308a <dbl+lbl>, CC10_308b <dbl+lbl>, CC10_308c <dbl+lbl>,
## #   CC10_308d <dbl+lbl>, CC10_308e <dbl+lbl>, CC10_309a <dbl+lbl>,
## #   CC10_309b <dbl+lbl>, CC10_309c <dbl+lbl>, CC10_309d <dbl+lbl>,
## #   CC10_310a <dbl+lbl>, CC10_310b <dbl+lbl>, CC10_310c <dbl+lbl>,
## #   CC10_310d <dbl+lbl>, CC10_315a <dbl+lbl>, CC10_315b <dbl+lbl>,
## #   CC10_315c <dbl+lbl>, CC10_316 <dbl+lbl>, CC10_317 <dbl+lbl>,
## #   CC10_317_t <chr>, CC10_320 <dbl+lbl>, CC10_321 <dbl+lbl>,
## #   CC10_322_1 <dbl+lbl>, CC10_322_2 <dbl+lbl>, CC10_322_3 <dbl+lbl>,
## #   CC10_322_4 <dbl+lbl>, CC10_322_7 <dbl+lbl>, CC10_322_8 <dbl+lbl>,
## #   CC10_324 <dbl+lbl>, CC10_325 <dbl+lbl>, CC10_326 <dbl+lbl>,
## #   CC10_327 <dbl+lbl>, CC10_328 <dbl+lbl>, CC10_329 <dbl+lbl>,
## #   CC10_330A <dbl+lbl>, CC10_330B <dbl+lbl>, CC10_330C <dbl+lbl>,
## #   CC10_330D <dbl+lbl>, CC10_330E <dbl+lbl>, CC10_330F <dbl+lbl>,
## #   CC10_330G <dbl+lbl>, CC10_330H <dbl+lbl>, CC10_330I <dbl+lbl>,
## #   CC10_330J <dbl+lbl>, CC10_334E <dbl+lbl>, CC10_334Hb <dbl+lbl>,
## #   CC10_334Ib <dbl+lbl>, CC10_335a1 <dbl+lbl>, CC10_335a2 <dbl+lbl>,
## #   CC10_335b1 <dbl+lbl>, CC10_335b2 <dbl+lbl>, CC10_335c1 <dbl+lbl>,
## #   CC10_335c2 <dbl+lbl>, CC10_341A <dbl+lbl>, CC10_341B <dbl+lbl>,
## #   CC10_341C <dbl+lbl>, CC10_341F <dbl+lbl>, CC10_341G <dbl+lbl>,
## #   CC10_341H <dbl+lbl>, CC10_341I <dbl+lbl>, CC10_341J <dbl+lbl>,
## #   CC10_341K <dbl+lbl>, CC10_341L <dbl+lbl>, CC10_341M <dbl+lbl>,
## #   CC10_341R <dbl+lbl>, CC10_350 <dbl+lbl>, CC10_351 <dbl+lbl>,
## #   CC10_352a <dbl+lbl>, CC10_352b <dbl+lbl>, CC10_354 <dbl+lbl>,
## #   CC10_355 <dbl+lbl>, CC10_355_t <chr>, CC10_355a <dbl+lbl>,
## #   CC10_355a_t <chr>, CC10_355b <dbl+lbl>, CC10_355b_t <chr>,
## #   CC10_356 <dbl+lbl>, CC10_356_t <chr>, CC10_356a <dbl+lbl>,
## #   CC10_356a_t <chr>, CC10_390 <dbl+lbl>, CC10_390_t <chr>,
## #   CC10_390a <dbl+lbl>, CC10_390a_t <chr>, CC10_401 <dbl+lbl>,
## #   CC10_402a <dbl+lbl>, CC10_402a_t <chr>, CC10_402b <dbl+lbl>,
## #   CC10_402b_t <chr>, CC10_403 <dbl+lbl>, CC10_405 <dbl+lbl>,
## #   CC10_406a <dbl+lbl>, CC10_406b_1 <dbl+lbl>, CC10_406b_2 <dbl+lbl>,
## #   CC10_406b_3 <dbl+lbl>, CC10_406c <dbl+lbl>, ...
```

View 1st 6 rows of data for all vars as dataset JO NOTE: I'm aware this command causes problems with Rmarkdown, but I'm retaining it during this preliminary data cleaning phase because it is important to view the stata meta-data that describe each variable but does not get reported in R (see e.g. 3rd column "CC10_301_1" is labeled underneath the variable name "Media use - Blog 2010"). This information is important to reference during data cleaning for this dataset because the pdf Guide does not always correctly document the variables that are actually in the data.

```
head(tbl_panel) %>% View
```

Recode CC10_330A as factor - converts the labeled numeric column into a factor with strings as the values.

```
tbl_panel <- tbl_panel %>%
    mutate(CC10_330A_fac = as_factor(CC10_330A))

# check successful mutate
tbl_panel %>%
  count(CC10_330A, CC10_330A_fac)
```

```
## # A tibble: 3 x 3
##   CC10_330A CC10_330A_fac      n
##   <dbl+lbl> <fct>          <int>
## 1 1         Support         4616
## 2 2         Oppose          4775
## 3 NA        <NA>             109
```

Count all 2010 vars noted in pdf Guide to verify they're in dataset (p.49 of Guide)

```
CC10_330_vars <- tbl_panel %>%
  select(CC10_330B, CC10_330C, CC10_330D, CC10_330E,
         CC10_330F, CC10_330G, CC10_330H, CC10_330I, CC10_330J) %>%
  colnames()

CC10_330_vars %>%
  map(~ tbl_panel %>% count(!!sym(.x)))
```

```
## [[1]]
## # A tibble: 3 x 2
##   CC10_330B     n
##   <dbl+lbl> <int>
## 1 1          6376
## 2 2          3055
## 3 NA           69
##
## [[2]]
## # A tibble: 3 x 2
##   CC10_330C     n
##   <dbl+lbl> <int>
## 1 1          4952
## 2 2          4386
## 3 NA          162
##
## [[3]]
## # A tibble: 3 x 2
##   CC10_330D     n
##   <dbl+lbl> <int>
## 1 1          4664
## 2 2          4764
## 3 NA           72
##
## [[4]]
## # A tibble: 3 x 2
##   CC10_330E     n
##   <dbl+lbl> <int>
## 1 1          4798
## 2 2          4447
## 3 NA          255
```

```
## 
## [[5]]
## # A tibble: 3 x 2
##    CC10_330F      n
##    <dbl+lbl> <int>
## 1  1          6365
## 2  2          3009
## 3 NA           126
## 
## [[6]]
## # A tibble: 3 x 2
##    CC10_330G      n
##    <dbl+lbl> <int>
## 1  1          5651
## 2  2          3725
## 3 NA           124
## 
## [[7]]
## # A tibble: 3 x 2
##    CC10_330H      n
##    <dbl+lbl> <int>
## 1  1          6191
## 2  2          3199
## 3 NA           110
## 
## [[8]]
## # A tibble: 3 x 2
##    CC10_330I      n
##    <dbl+lbl> <int>
## 1  1          1100
## 2  2           328
## 3 NA          8072
## 
## [[9]]
## # A tibble: 3 x 2
##    CC10_330J      n
##    <dbl+lbl> <int>
## 1  1           385
## 2  2          1070
## 3 NA          8045
```

Recode all remaining 2010 vars as factor

```r
tbl_panel <- tbl_panel %>%
  mutate_at(vars(CC10_330B, CC10_330C, CC10_330D, CC10_330E,
                 CC10_330F, CC10_330G, CC10_330H, CC10_330I, CC10_330J),
            funs(fac = as_factor(.)))
```

Confirm that mutate performed correctly

```r
fac_vars <- CC10_330_vars %>%
  paste0("_fac")

map2(CC10_330_vars, fac_vars, ~ tbl_panel %>% count(!!sym(.x), !!sym(.y)))
```

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

```
## # A tibble: 3 x 3
##   CC10_330B CC10_330B_fac      n
##   <dbl+lbl> <fct>          <int>
## 1 1         Support         6376
## 2 2         Oppose          3055
## 3 NA        <NA>              69
##
## [[2]]
## # A tibble: 3 x 3
##   CC10_330C CC10_330C_fac      n
##   <dbl+lbl> <fct>          <int>
## 1 1         Support         4952
## 2 2         Oppose          4386
## 3 NA        <NA>             162
##
## [[3]]
## # A tibble: 3 x 3
##   CC10_330D CC10_330D_fac      n
##   <dbl+lbl> <fct>          <int>
## 1 1         Support         4664
## 2 2         Oppose          4764
## 3 NA        <NA>              72
##
## [[4]]
## # A tibble: 3 x 3
##   CC10_330E CC10_330E_fac      n
##   <dbl+lbl> <fct>          <int>
## 1 1         Support         4798
## 2 2         Oppose          4447
## 3 NA        <NA>             255
##
## [[5]]
## # A tibble: 3 x 3
##   CC10_330F CC10_330F_fac      n
##   <dbl+lbl> <fct>          <int>
## 1 1         Support         6365
## 2 2         Oppose          3009
## 3 NA        <NA>             126
##
## [[6]]
## # A tibble: 3 x 3
##   CC10_330G CC10_330G_fac      n
##   <dbl+lbl> <fct>          <int>
## 1 1         Support         5651
## 2 2         Oppose          3725
## 3 NA        <NA>             124
##
## [[7]]
## # A tibble: 3 x 3
##   CC10_330H CC10_330H_fac      n
##   <dbl+lbl> <fct>          <int>
## 1 1         Support         6191
## 2 2         Oppose          3199
## 3 NA        <NA>             110
```

```
##
## [[8]]
## # A tibble: 3 x 3
##   CC10_330I CC10_330I_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Support        1100
## 2 2         Oppose          328
## 3 NA        <NA>           8072
##
## [[9]]
## # A tibble: 3 x 3
##   CC10_330J CC10_330J_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Support         385
## 2 2         Oppose         1070
## 3 NA        <NA>           8045
```

2012: Count all 2012 vars noted in pdf Guide to verify they're in dataset (p. 143 of Guide)

```
CC12_vars <- tbl_panel %>%
  select(CC12_330A, CC12_330B, CC12_330C, CC12_330D, CC12_330E,
         CC12_330F, CC12_330G, CC12_330H, CC12_332A, CC12_332B,
         CC12_332C, CC12_332D, CC12_332E, CC12_332F) %>%
  colnames()

CC12_vars %>%
  map(~ tbl_panel %>% count(!!sym(.x)))
```

```
## [[1]]
## # A tibble: 3 x 2
##   CC12_330A     n
##   <dbl+lbl> <int>
## 1 1          4633
## 2 2          4806
## 3 NA           61
##
## [[2]]
## # A tibble: 3 x 2
##   CC12_330B     n
##   <dbl+lbl> <int>
## 1 1          6705
## 2 2          2733
## 3 NA           62
##
## [[3]]
## # A tibble: 3 x 2
##   CC12_330C     n
##   <dbl+lbl> <int>
## 1 1          5025
## 2 2          4346
## 3 NA          129
##
## [[4]]
## # A tibble: 3 x 2
##   CC12_330D     n
```

```
##     <dbl+lbl> <int>
## 1  1          4828
## 2  2          4602
## 3 NA            70
##
## [[5]]
## # A tibble: 3 x 2
##    CC12_330E     n
##     <dbl+lbl> <int>
## 1  1          4272
## 2  2          4898
## 3 NA           330
##
## [[6]]
## # A tibble: 3 x 2
##    CC12_330F     n
##     <dbl+lbl> <int>
## 1  1          6417
## 2  2          2910
## 3 NA           173
##
## [[7]]
## # A tibble: 3 x 2
##    CC12_330G     n
##     <dbl+lbl> <int>
## 1  1          5962
## 2  2          3434
## 3 NA           104
##
## [[8]]
## # A tibble: 3 x 2
##    CC12_330H     n
##     <dbl+lbl> <int>
## 1  1          6227
## 2  2          3144
## 3 NA           129
##
## [[9]]
## # A tibble: 3 x 2
##    CC12_332A     n
##     <dbl+lbl> <int>
## 1  1          2436
## 2  2          6919
## 3 NA           145
##
## [[10]]
## # A tibble: 3 x 2
##    CC12_332B     n
##     <dbl+lbl> <int>
## 1  1          3975
## 2  2          5363
## 3 NA           162
##
## [[11]]
```

```
## # A tibble: 3 x 2
##   CC12_332C      n
##   <dbl+lbl> <int>
## 1 1          6077
## 2 2          3188
## 3 NA          235
##
## [[12]]
## # A tibble: 3 x 2
##   CC12_332D      n
##   <dbl+lbl> <int>
## 1 1          3073
## 2 2          6188
## 3 NA          239
##
## [[13]]
## # A tibble: 3 x 2
##   CC12_332E      n
##   <dbl+lbl> <int>
## 1 1          4269
## 2 2          5119
## 3 NA          112
##
## [[14]]
## # A tibble: 3 x 2
##   CC12_332F      n
##   <dbl+lbl> <int>
## 1 1          4967
## 2 2          4223
## 3 NA          310
```

Note: the final 2 vars in the list in the Guide don't appear in the data: CC12_332G = Repeal ACA CC12_332H = Keystone

Recode all 2012 roll call votes as factors

```r
tbl_panel <- tbl_panel %>%
  mutate_at(vars(CC12_330A, CC12_330B, CC12_330C, CC12_330D, CC12_330E,
                 CC12_330F, CC12_330G, CC12_330H, CC12_332A, CC12_332B,
                 CC12_332C, CC12_332D, CC12_332E, CC12_332F),
            funs(fac = as_factor(.)))

# Confirm that mutate performed correctly
CC12_fac_vars <- CC12_vars %>%
  paste0("_fac")

map2(CC12_vars, CC12_fac_vars, ~ tbl_panel %>% count(!!sym(.x), !!sym(.y)))
```

```
## [[1]]
## # A tibble: 3 x 3
##   CC12_330A CC12_330A_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Yes            4633
## 2 2         No             4806
## 3 NA        <NA>             61
##
```

```
## [[2]]
## # A tibble: 3 x 3
##   CC12_330B CC12_330B_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Yes            6705
## 2 2         No             2733
## 3 NA        <NA>             62
##
## [[3]]
## # A tibble: 3 x 3
##   CC12_330C CC12_330C_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Yes            5025
## 2 2         No             4346
## 3 NA        <NA>            129
##
## [[4]]
## # A tibble: 3 x 3
##   CC12_330D CC12_330D_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Yes            4828
## 2 2         No             4602
## 3 NA        <NA>             70
##
## [[5]]
## # A tibble: 3 x 3
##   CC12_330E CC12_330E_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Yes            4272
## 2 2         No             4898
## 3 NA        <NA>            330
##
## [[6]]
## # A tibble: 3 x 3
##   CC12_330F CC12_330F_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Yes            6417
## 2 2         No             2910
## 3 NA        <NA>            173
##
## [[7]]
## # A tibble: 3 x 3
##   CC12_330G CC12_330G_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Yes            5962
## 2 2         No             3434
## 3 NA        <NA>            104
##
## [[8]]
## # A tibble: 3 x 3
##   CC12_330H CC12_330H_fac     n
##   <dbl+lbl> <fct>         <int>
## 1 1         Yes            6227
## 2 2         No             3144
```

```
## 3 NA          <NA>              129
##
## [[9]]
## # A tibble: 3 x 3
##    CC12_332A CC12_332A_fac       n
##    <dbl+lbl> <fct>           <int>
## 1  1         Yes              2436
## 2  2         No               6919
## 3 NA         <NA>              145
##
## [[10]]
## # A tibble: 3 x 3
##    CC12_332B CC12_332B_fac       n
##    <dbl+lbl> <fct>           <int>
## 1  1         Yes              3975
## 2  2         No               5363
## 3 NA         <NA>              162
##
## [[11]]
## # A tibble: 3 x 3
##    CC12_332C CC12_332C_fac       n
##    <dbl+lbl> <fct>           <int>
## 1  1         Yes              6077
## 2  2         No               3188
## 3 NA         <NA>              235
##
## [[12]]
## # A tibble: 3 x 3
##    CC12_332D CC12_332D_fac       n
##    <dbl+lbl> <fct>           <int>
## 1  1         Yes              3073
## 2  2         No               6188
## 3 NA         <NA>              239
##
## [[13]]
## # A tibble: 3 x 3
##    CC12_332E CC12_332E_fac       n
##    <dbl+lbl> <fct>           <int>
## 1  1         Yes              4269
## 2  2         No               5119
## 3 NA         <NA>              112
##
## [[14]]
## # A tibble: 3 x 3
##    CC12_332F CC12_332F_fac       n
##    <dbl+lbl> <fct>           <int>
## 1  1         Yes              4967
## 2  2         No               4223
## 3 NA         <NA>              310
```

Count all 2014 vars noted in pdf Guide to verify they're in dataset (p. 269ff of Guide)

```
CC14_vars <- tbl_panel %>%
  select(CC14_330A, CC14_330B, CC14_330C, CC14_330D, CC14_330E,
         CC14_330F, CC14_330G, CC14_330H, CC14_332A, CC14_332B,
```

```
          CC14_332C, CC14_332D, CC14_332E, CC14_332F) %>%
  colnames()

CC14_vars %>%
  map(~ tbl_panel %>% count(!!sym(.x)))
```

```
## [[1]]
## # A tibble: 3 x 2
##   CC14_330A      n
##   <dbl+lbl> <int>
## 1  1         4816
## 2  2         4577
## 3 NA          107
##
## [[2]]
## # A tibble: 3 x 2
##   CC14_330B      n
##   <dbl+lbl> <int>
## 1  1         6451
## 2  2         2939
## 3 NA          110
##
## [[3]]
## # A tibble: 3 x 2
##   CC14_330C      n
##   <dbl+lbl> <int>
## 1  1         5068
## 2  2         4281
## 3 NA          151
##
## [[4]]
## # A tibble: 3 x 2
##   CC14_330D      n
##   <dbl+lbl> <int>
## 1  1         4732
## 2  2         4703
## 3 NA           65
##
## [[5]]
## # A tibble: 3 x 2
##   CC14_330E      n
##   <dbl+lbl> <int>
## 1  1         4304
## 2  2         4836
## 3 NA          360
##
## [[6]]
## # A tibble: 3 x 2
##   CC14_330F      n
##   <dbl+lbl> <int>
## 1  1         6734
## 2  2         2572
## 3 NA          194
##
```

```
## [[7]]
## # A tibble: 3 x 2
##   CC14_330G       n
##   <dbl+lbl> <int>
## 1  1         5940
## 2  2         3463
## 3 NA           97
##
## [[8]]
## # A tibble: 3 x 2
##   CC14_330H       n
##   <dbl+lbl> <int>
## 1  1         6540
## 2  2         2857
## 3 NA          103
##
## [[9]]
## # A tibble: 3 x 2
##   CC14_332A       n
##   <dbl+lbl> <int>
## 1  1         1792
## 2  2         7587
## 3 NA          121
##
## [[10]]
## # A tibble: 3 x 2
##   CC14_332B       n
##   <dbl+lbl> <int>
## 1  1         3094
## 2  2         6271
## 3 NA          135
##
## [[11]]
## # A tibble: 3 x 2
##   CC14_332C       n
##   <dbl+lbl> <int>
## 1  1         5544
## 2  2         3758
## 3 NA          198
##
## [[12]]
## # A tibble: 3 x 2
##   CC14_332D       n
##   <dbl+lbl> <int>
## 1  1         2497
## 2  2         6805
## 3 NA          198
##
## [[13]]
## # A tibble: 3 x 2
##   CC14_332E       n
##   <dbl+lbl> <int>
## 1  1         4405
## 2  2         4982
```

```
## 3 NA            113
##
## [[14]]
## # A tibble: 3 x 2
##   CC14_332F      n
##   <dbl+lbl> <int>
## 1  1          4972
## 2  2          4233
## 3 NA           295
```

Recode all 2014 roll call votes as factors

```r
tbl_panel <- tbl_panel %>%
  mutate_at(vars(CC14_330A, CC14_330B, CC14_330C, CC14_330D, CC14_330E,
                 CC14_330F, CC14_330G, CC14_330H, CC14_332A, CC14_332B,
                 CC14_332C, CC14_332D, CC14_332E, CC14_332F),
            funs(fac = as_factor(.)))

# Confirm that mutate performed correctly
CC14_fac_vars <- CC14_vars %>%
  paste0("_fac")

map2(CC14_vars, CC14_fac_vars, ~ tbl_panel %>% count(!!sym(.x), !!sym(.y)))
```

```
## [[1]]
## # A tibble: 3 x 3
##   CC14_330A CC14_330A_fac     n
##   <dbl+lbl> <fct>         <int>
## 1  1        Yes            4816
## 2  2        No             4577
## 3 NA        <NA>            107
##
## [[2]]
## # A tibble: 3 x 3
##   CC14_330B CC14_330B_fac     n
##   <dbl+lbl> <fct>         <int>
## 1  1        Yes            6451
## 2  2        No             2939
## 3 NA        <NA>            110
##
## [[3]]
## # A tibble: 3 x 3
##   CC14_330C CC14_330C_fac     n
##   <dbl+lbl> <fct>         <int>
## 1  1        Yes            5068
## 2  2        No             4281
## 3 NA        <NA>            151
##
## [[4]]
## # A tibble: 3 x 3
##   CC14_330D CC14_330D_fac     n
##   <dbl+lbl> <fct>         <int>
## 1  1        Yes            4732
## 2  2        No             4703
## 3 NA        <NA>             65
```

```
## 
## [[5]]
## # A tibble: 3 x 3
##    CC14_330E CC14_330E_fac     n
##    <dbl+lbl> <fct>         <int>
## 1  1         Yes            4304
## 2  2         No             4836
## 3 NA         <NA>            360
## 
## [[6]]
## # A tibble: 3 x 3
##    CC14_330F CC14_330F_fac     n
##    <dbl+lbl> <fct>         <int>
## 1  1         Yes            6734
## 2  2         No             2572
## 3 NA         <NA>            194
## 
## [[7]]
## # A tibble: 3 x 3
##    CC14_330G CC14_330G_fac     n
##    <dbl+lbl> <fct>         <int>
## 1  1         Yes            5940
## 2  2         No             3463
## 3 NA         <NA>             97
## 
## [[8]]
## # A tibble: 3 x 3
##    CC14_330H CC14_330H_fac     n
##    <dbl+lbl> <fct>         <int>
## 1  1         Yes            6540
## 2  2         No             2857
## 3 NA         <NA>            103
## 
## [[9]]
## # A tibble: 3 x 3
##    CC14_332A CC14_332A_fac     n
##    <dbl+lbl> <fct>         <int>
## 1  1         Yes            1792
## 2  2         No             7587
## 3 NA         <NA>            121
## 
## [[10]]
## # A tibble: 3 x 3
##    CC14_332B CC14_332B_fac     n
##    <dbl+lbl> <fct>         <int>
## 1  1         Yes            3094
## 2  2         No             6271
## 3 NA         <NA>            135
## 
## [[11]]
## # A tibble: 3 x 3
##    CC14_332C CC14_332C_fac     n
##    <dbl+lbl> <fct>         <int>
## 1  1         Yes            5544
```

```
## 2  2         No                  3758
## 3 NA         <NA>                 198
##
## [[12]]
## # A tibble: 3 x 3
##   CC14_332D CC14_332D_fac      n
##   <dbl+lbl> <fct>          <int>
## 1  1        Yes             2497
## 2  2        No              6805
## 3 NA        <NA>             198
##
## [[13]]
## # A tibble: 3 x 3
##   CC14_332E CC14_332E_fac      n
##   <dbl+lbl> <fct>          <int>
## 1  1        Yes             4405
## 2  2        No              4982
## 3 NA        <NA>             113
##
## [[14]]
## # A tibble: 3 x 3
##   CC14_332F CC14_332F_fac      n
##   <dbl+lbl> <fct>          <int>
## 1  1        Yes             4972
## 2  2        No              4233
## 3 NA        <NA>             295
```

Recode Voting Participation

Validated vote variables, general election: `VV_general_10`, `VV_general_12`, `VV_general_14`

Recode as factor and binary variables .

```r
voting_vars <- c("VV_general_10", "VV_general_12", "VV_general_14")

tbl_panel <- tbl_panel %>%
  mutate_at(voting_vars, funs(fac = as_factor(.))) %>%
  mutate_at(voting_vars, funs(bin = case_when(
    . %in% c(1, 2, 3, 4, 5) ~ "vote",
    . == 6                  ~ "not vote",
    . == 9                  ~ NA_character_
  )))

# check recodes
voting_fac_vars <- voting_vars %>%
  paste0("_fac")

voting_bin_vars <- voting_vars %>%
  paste0("_bin")

pmap(list(voting_bin_vars, voting_fac_vars, voting_vars),
     ~ tbl_panel %>% count(!!sym(..1), !!sym(..2), !!sym(..3)))
```

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

```
## # A tibble: 6 x 4
##   VV_general_10_bin VV_general_10_fac           VV_general_10     n
##   <chr>             <fct>                       <dbl+lbl>     <int>
## 1 not vote          Confirmed Non-voter         6              1728
## 2 vote              Absentee voter              1              1281
## 3 vote              Early voter                 2               931
## 4 vote              Mail voter                  3               413
## 5 vote              In person or unknown method 4              4418
## 6 <NA>              Unmatched                   9               729
##
## [[2]]
## # A tibble: 7 x 4
##   VV_general_12_bin VV_general_12_fac           VV_general_12     n
##   <chr>             <fct>                       <dbl+lbl>     <int>
## 1 not vote          Confirmed Non-voter         6               879
## 2 vote              Absentee voter              1              1265
## 3 vote              Early voter                 2              1025
## 4 vote              Mail voter                  3               453
## 5 vote              In person                   4              2243
## 6 vote              Voted by unknown method     5              2148
## 7 <NA>              Unmatched                   9              1487
##
## [[3]]
## # A tibble: 7 x 4
##   VV_general_14_bin VV_general_14_fac           VV_general_14     n
##   <chr>             <fct>                       <dbl+lbl>     <int>
## 1 not vote          Confirmed Non-voter         6              1279
## 2 vote              Absentee voter              1               876
## 3 vote              Early voter                 2               861
## 4 vote              Mail voter                  3               798
## 5 vote              In person                   4              2278
## 6 vote              Voted by unknown method     5              2076
## 7 <NA>              Unmatched                   9              1332
```

Recode Political Participation

Attend meetings

Political activity: Attend meetings 2010 During the past year did you. . . Attend local political meetings (such as school board or city council)

Variables: `CC10_417a_1`, `CC12_417a_1`, `CC14_417a_1`

Recode all as binary factors.

```r
attend_meeting_vars <- c("CC10_417a_1", "CC12_417a_1", "CC14_417a_1")

tbl_panel <- tbl_panel %>%
  mutate_at(attend_meeting_vars, funs(fac = as_factor(.)))

# Confirm that mutate performed correctly
attend_meeting_fac_vars <- attend_meeting_vars %>%
  paste0("_fac")
```

```r
map2(attend_meeting_vars, attend_meeting_fac_vars,
     ~ tbl_panel %>% count(!!sym(.x), !!sym(.y)))
```

```
## [[1]]
## # A tibble: 2 x 3
##   CC10_417a_1 CC10_417a_1_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              2068
## 2 2           No               7432
##
## [[2]]
## # A tibble: 2 x 3
##   CC12_417a_1 CC12_417a_1_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              1822
## 2 2           No               7678
##
## [[3]]
## # A tibble: 2 x 3
##   CC14_417a_1 CC14_417a_1_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              1507
## 2 2           No               7993
```

Figures:

```r
attend_meeting_plot_tbl <- map2_df(attend_meeting_vars, attend_meeting_fac_vars,
                                   ~ tbl_panel %>%
                                     count(!!sym(.x), !!sym(.y)) %>%
                                     select(contains("_fac"), n) %>%
                                     mutate(year = .x) %>%
                                     rename(attend_meeting = .y) %>%
                                     select(year, attend_meeting, n)) %>%
  mutate(year = case_when(
    year == "CC10_417a_1" ~ "2010",
    year == "CC12_417a_1" ~ "2012",
    year == "CC14_417a_1" ~ "2014"
  )) %>%
  group_by(year) %>%
  mutate(percent = round((n / sum(n)) * 100, 1)) %>%
  ungroup() %>%
  filter(attend_meeting == "Yes")

attend_meeting_plot_tbl %>%
  knitr::kable()
```

year	attend_meeting	n	percent
2010	Yes	2068	21.8
2012	Yes	1822	19.2
2014	Yes	1507	15.9

```r
# bar chart
attend_meeting_plot_tbl %>%
```

```r
  ggplot(aes(x = year, y = percent, fill = year)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = percent + 1.5,      # nudge above top of bar
                label = paste0(percent, '%')),     # prettify
            position = position_dodge(width = .9),
            size = 4) +
  lims(y = c(0, 50)) +
  labs(x     = "Year",
       y     = "Percent",
       title = "Percent Attending Local Political Meetings by Year") +
  theme(legend.position = "none")
```

Percent Attending Local Political Meetings by Year
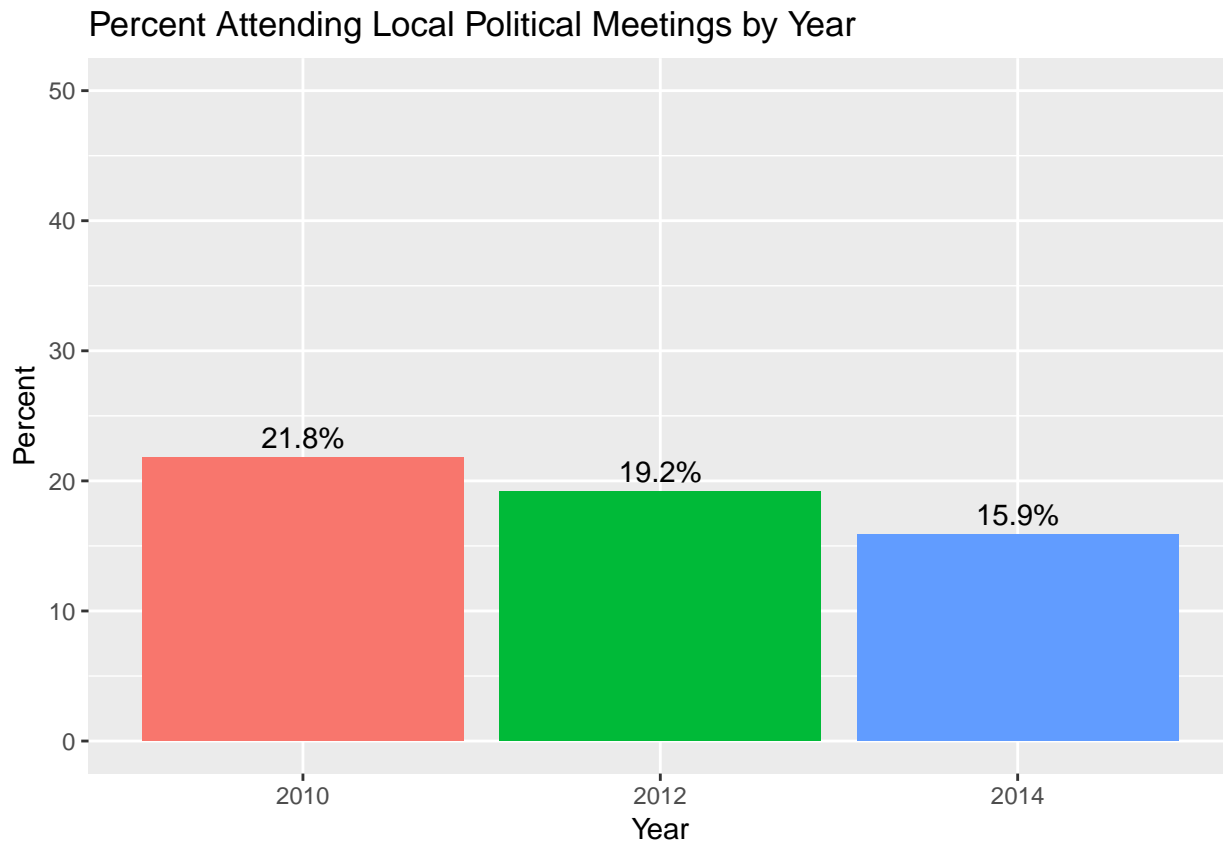
Political sign

During the past year did you... Put up a political sign (such as a lawn sign or bumper sticker)?

Variables: `CC10_417a_2`, `CC12_417a_2`, `CC14_417a_2`

Recode all as binary factors.

```r
pol_sign_vars <- c("CC10_417a_2", "CC12_417a_2", "CC14_417a_2")

tbl_panel <- tbl_panel %>%
  mutate_at(pol_sign_vars, funs(fac = as_factor(.)))

# Confirm that mutate performed correctly
pol_sign_fac_vars <- pol_sign_vars %>%
  paste0("_fac")
```

```
map2(pol_sign_vars, pol_sign_fac_vars,
     ~ tbl_panel %>% count(!!sym(.x), !!sym(.y)))
```

```
## [[1]]
## # A tibble: 2 x 3
##   CC10_417a_2 CC10_417a_2_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              2548
## 2 2           No               6952
##
## [[2]]
## # A tibble: 2 x 3
##   CC12_417a_2 CC12_417a_2_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              2958
## 2 2           No               6542
##
## [[3]]
## # A tibble: 2 x 3
##   CC14_417a_2 CC14_417a_2_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              1918
## 2 2           No               7582
```

Figure:

```
pol_sign_plot_tbl <- map2_df(pol_sign_vars, pol_sign_fac_vars,
                             ~ tbl_panel %>%
                               count(!!sym(.x), !!sym(.y)) %>%
                               select(contains("_fac"), n) %>%
                               mutate(year = .x) %>%
                               rename(political_sign = .y) %>%
                               select(year, political_sign, n)) %>%
  mutate(year = case_when(
    year == "CC10_417a_2" ~ "2010",
    year == "CC12_417a_2" ~ "2012",
    year == "CC14_417a_2" ~ "2014"
  )) %>%
  group_by(year) %>%
  mutate(percent = round((n / sum(n)) * 100, 1)) %>%
  ungroup() %>%
  filter(political_sign == "Yes")

pol_sign_plot_tbl %>%
  knitr::kable()
```

year	political_sign	n	percent
2010	Yes	2548	26.8
2012	Yes	2958	31.1
2014	Yes	1918	20.2

```
# bar chart
pol_sign_plot_tbl %>%
```

```
  ggplot(aes(x = year, y = percent, fill = year)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = percent + 1.5,     # nudge above top of bar
                label = paste0(percent, '%')),     # prettify
            position = position_dodge(width = .9),
            size = 4) +
  lims(y = c(0, 50)) +
  labs(x    = "Year",
       y    = "Percent",
       title = "Percent Who Put Up a Political Sign by Year") +
  theme(legend.position = "none")
```

Percent Who Put Up a Political Sign by Year
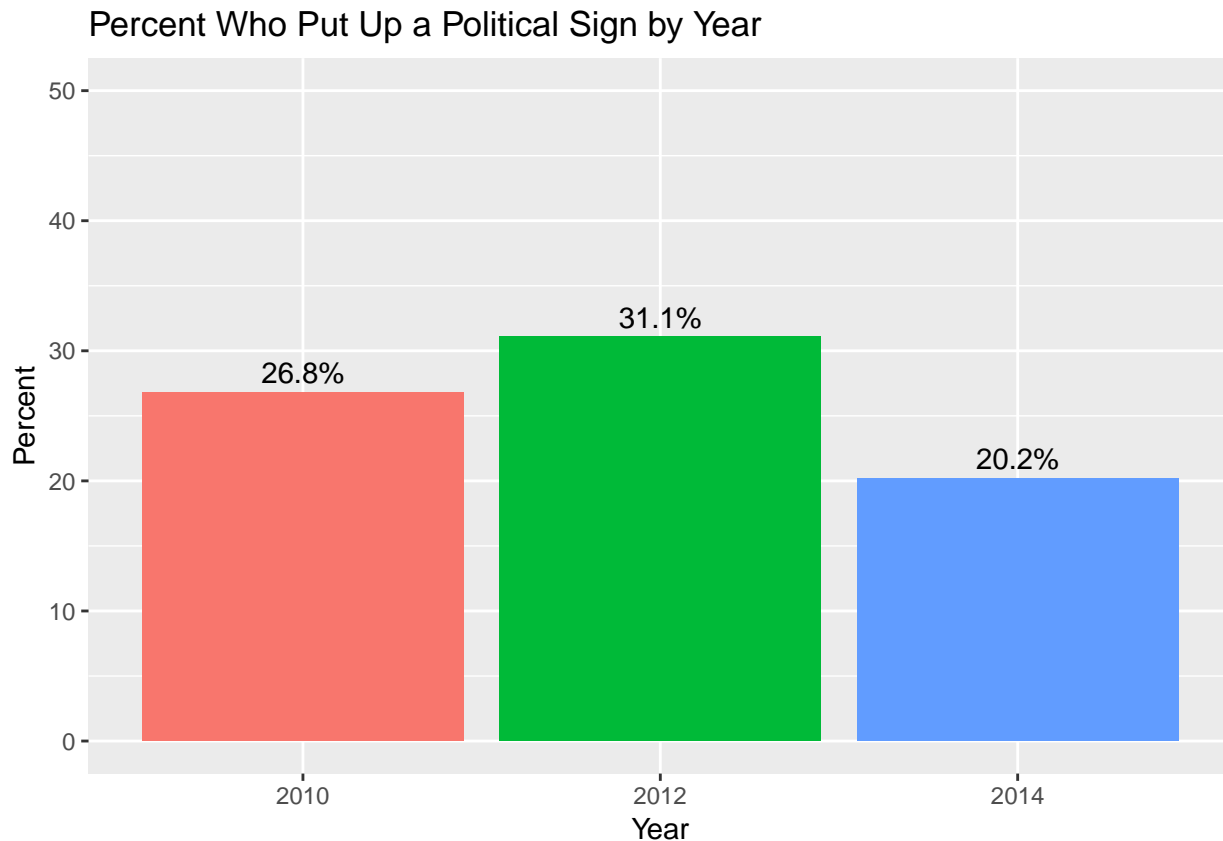
"Work for campaign"

During the past year did you. . . Work for a candidate or campaign?

Variables: `CC10_417a_3`, `CC12_417a_3`, `CC14_417a_3`

Recode all as binary factors.

```
work_campaign_vars <- c("CC10_417a_3", "CC12_417a_3", "CC14_417a_3")

tbl_panel <- tbl_panel %>%
  mutate_at(work_campaign_vars, funs(fac = as_factor(.)))

# Confirm that mutate performed correctly
work_campaign_fac_vars <- work_campaign_vars %>%
  paste0("_fac")
```

```
map2(work_campaign_vars, work_campaign_fac_vars,
     ~ tbl_panel %>% count(!!sym(.x), !!sym(.y)))
```

```
## [[1]]
## # A tibble: 2 x 3
##   CC10_417a_3 CC10_417a_3_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes               990
## 2 2           No               8510
##
## [[2]]
## # A tibble: 2 x 3
##   CC12_417a_3 CC12_417a_3_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              1144
## 2 2           No               8356
##
## [[3]]
## # A tibble: 2 x 3
##   CC14_417a_3 CC14_417a_3_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes               686
## 2 2           No               8814
```

Figure:

```
work_campaign_plot_tbl <- map2_df(work_campaign_vars, work_campaign_fac_vars,
                                  ~ tbl_panel %>%
                                    count(!!sym(.x), !!sym(.y)) %>%
                                    select(contains("_fac"), n) %>%
                                    mutate(year = .x) %>%
                                    rename(work_campaign = .y) %>%
                                    select(year, work_campaign, n)) %>%
  mutate(year = case_when(
    year == "CC10_417a_3" ~ "2010",
    year == "CC12_417a_3" ~ "2012",
    year == "CC14_417a_3" ~ "2014"
  )) %>%
  group_by(year) %>%
  mutate(percent = round((n / sum(n)) * 100, 1)) %>%
  ungroup() %>%
  filter(work_campaign == "Yes")

work_campaign_plot_tbl %>%
  knitr::kable()
```

year	work_campaign	n	percent
2010	Yes	990	10.4
2012	Yes	1144	12.0
2014	Yes	686	7.2

```
# bar chart
work_campaign_plot_tbl %>%
```

```
ggplot(aes(x = year, y = percent, fill = year)) +
geom_bar(stat = "identity") +
geom_text(aes(y = percent + 1.5,       # nudge above top of bar
              label = paste0(percent, '%')),     # prettify
          position = position_dodge(width = .9),
          size = 4) +
lims(y = c(0, 50)) +
labs(x     = "Year",
     y     = "Percent",
     title = "Percent Who Worked for a Campaign by Year") +
theme(legend.position = "none")
```

Percent Who Worked for a Campaign by Year
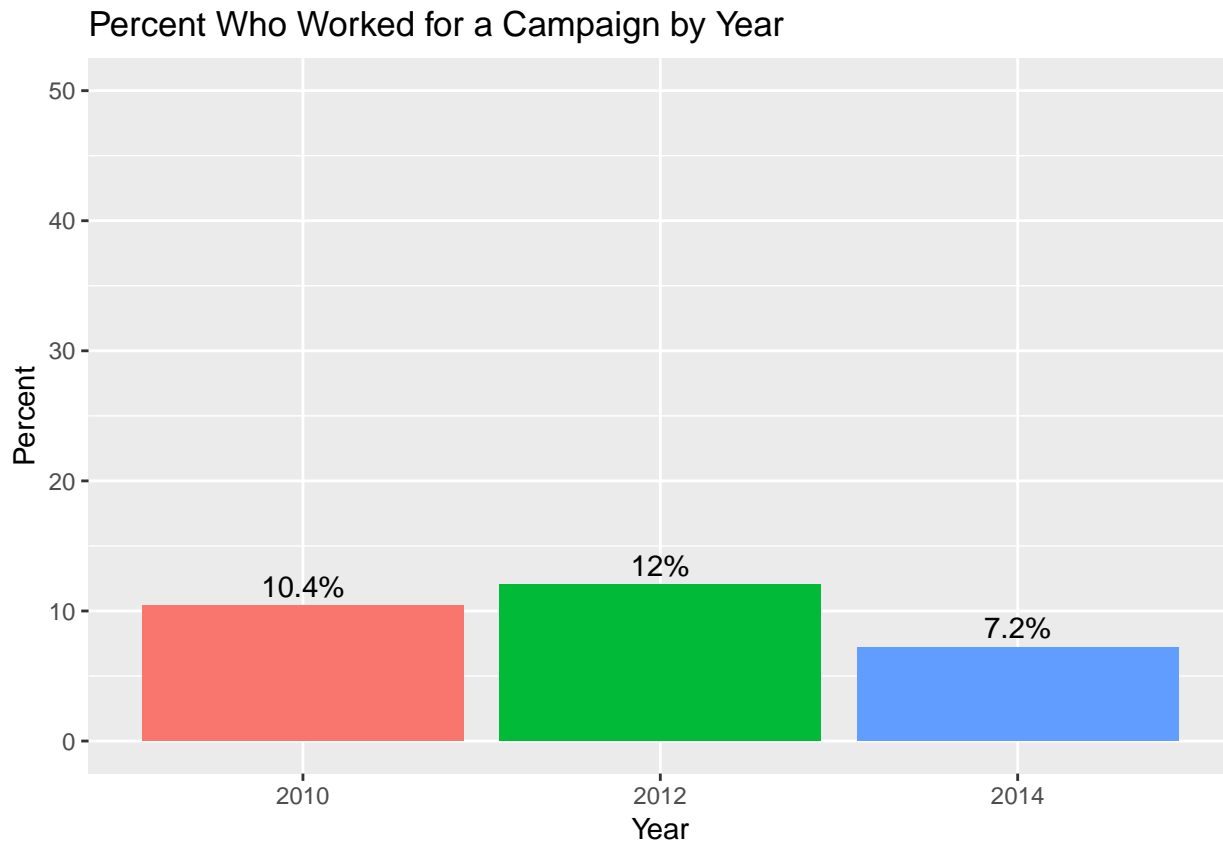
"Donate money"

During the past year did you... Donate money to a candidate, campaign, or political organization?

Variables: `CC10_417a_4`, `CC12_417a_4`, `CC14_417a_4`

```
donate_vars <- c("CC10_417a_4", "CC12_417a_4", "CC14_417a_4")

tbl_panel <- tbl_panel %>%
  mutate_at(donate_vars, funs(fac = as_factor(.)))

# Confirm that mutate performed correctly
donate_fac_vars <- donate_vars %>%
  paste0("_fac")
```

```
map2(donate_vars, donate_fac_vars,
     ~ tbl_panel %>% count(!!sym(.x), !!sym(.y)))
```

```
## [[1]]
## # A tibble: 2 x 3
##   CC10_417a_4 CC10_417a_4_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              3292
## 2 2           No               6208
##
## [[2]]
## # A tibble: 2 x 3
##   CC12_417a_4 CC12_417a_4_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              4030
## 2 2           No               5470
##
## [[3]]
## # A tibble: 2 x 3
##   CC14_417a_4 CC14_417a_4_fac     n
##   <dbl+lbl>   <fct>           <int>
## 1 1           Yes              2957
## 2 2           No               6543
```

Figure:

```
donate_plot_tbl <- map2_df(donate_vars, donate_fac_vars,
                           ~ tbl_panel %>%
                             count(!!sym(.x), !!sym(.y)) %>%
                             select(contains("_fac"), n) %>%
                             mutate(year = .x) %>%
                             rename(donate_money = .y) %>%
                             select(year, donate_money, n)) %>%
  mutate(year = case_when(
    year == "CC10_417a_4" ~ "2010",
    year == "CC12_417a_4" ~ "2012",
    year == "CC14_417a_4" ~ "2014"
  )) %>%
  group_by(year) %>%
  mutate(percent = round((n / sum(n)) * 100, 1)) %>%
  ungroup() %>%
  filter(donate_money == "Yes")

donate_plot_tbl %>%
  knitr::kable()
```

year	donate_money	n	percent
2010	Yes	3292	34.7
2012	Yes	4030	42.4
2014	Yes	2957	31.1

```
# bar chart
donate_plot_tbl %>%
```

```
ggplot(aes(x = year, y = percent, fill = year)) +
geom_bar(stat = "identity") +
geom_text(aes(y = percent + 1.5,      # nudge above top of bar
              label = paste0(percent, '%')),    # prettify
          position = position_dodge(width = .9),
          size = 4) +
lims(y = c(0, 50)) +
labs(x     = "Year",
     y     = "Percent",
     title = "Percent Who Donated Money to a Candidate, Campaign, or Political Organization by Year")
theme(legend.position = "none")
```


Percent Who Donated Money to a Candidate, Campaign, or Political Organiz
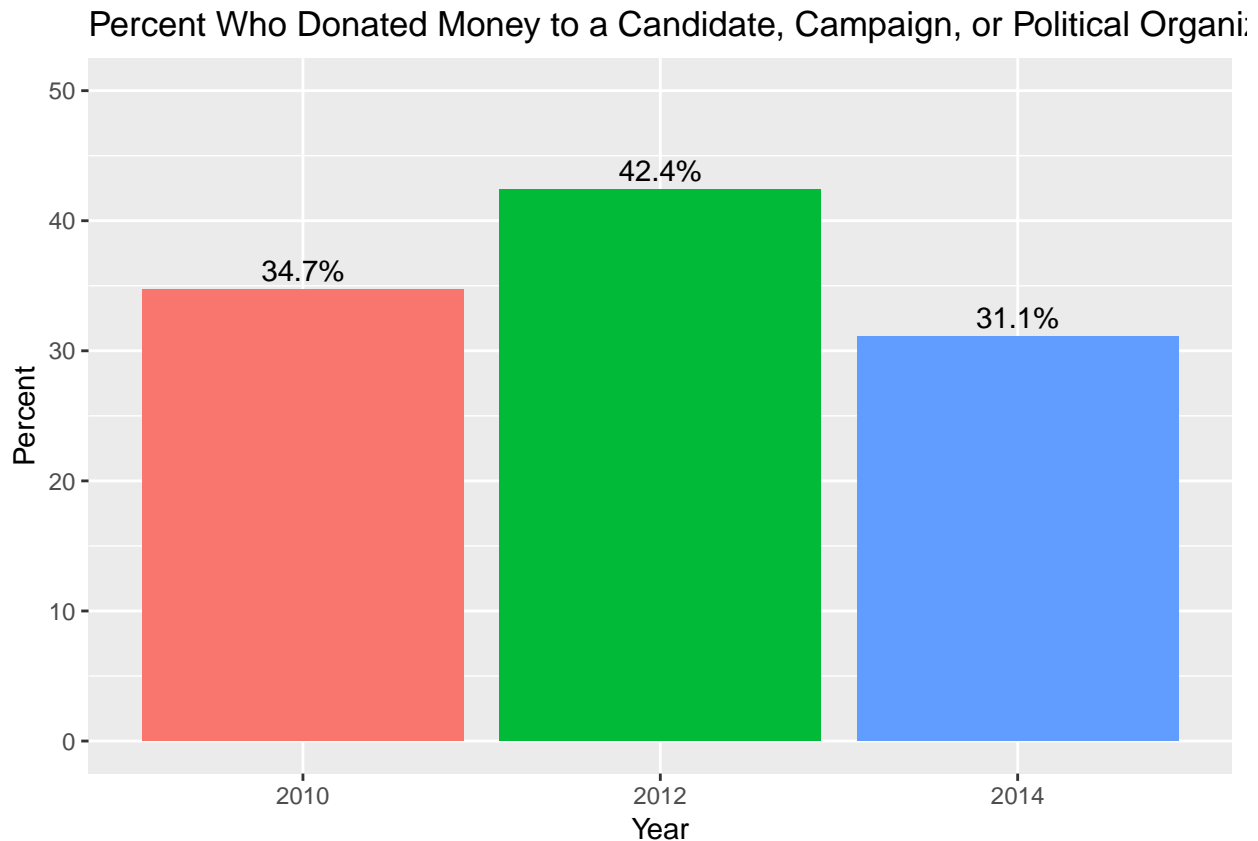
Merge available coded roll call votes

NOTE: 2010 exists in CCES dataverse, but only codes Congressional candidates by race and ethnicity variables.

2012 relevant folder in github: "CCES 2012 supplemental"

II. Create joined data file (respondents & roll call votes) In stata: Reshaped into wide format to join In R, relevant commands: spread, gather, mutate_at

Can reference prior stata work for joining procedure using stata: CCES panel 2010-2012-2014: https://www.dropbox.com/sh/gj4vfv9xzvdgkwd/AABpmEqKQBT6EQDVCFvqeV7Ka?dl=0

**NOTE: for 2012, CCES has two versions of roll call vote data, and I can't tell which one is the correct to use for merge:

1st: This is version we used in prior 2012 stata work, and this is the one referenced by Brian Schaffner in correspondence re: available roll call data. JA did significant recodes to get it to work. In github, this documentation is in folder "CCES 2010 supplemental" Cuevas-Molina, Ivelisse; Schaffner, Brian, 2015, "CCES 2014 Supplemental Data", https://doi.org/10.7910/DVN/D1N0GO, Harvard Dataverse, V3, UNF:6:M0Py+4hUexjh8WSlUevvDQ== [fileUNF]

2nd: I came across this next source randomly - it has an earlier publication date, so I wonder if there is a reason why they have two separate codings available, and why we used the 1st version in our prior work, and not this one. In github it's in the folder "CCES 2012 roll call votes_Pettigrew" Pettigrew, Stephen, 2013, "Roll Call Votes - 112th Congress- 2012 CCES", https://hdl.handle.net/1902.1/20540, Harvard Dataverse, V3

Link for CCES 2014 supplemental data: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/D1N0GO

Recode Socio-Dem Data

Socio-dem data - code below is a start

```
tbl_panel %>%
  select(gender_10, birthyr_10) %>%
  mutate(gender = as_factor(gender_10),
         age    = 2010 - birthyr_10)
```

```
## # A tibble: 9,500 x 4
##    gender_10 birthyr_10 gender age
##    <dbl+lbl> <dbl+lbl>  <fct>  <dbl+lbl>
## 1  1         1949       Male   61
## 2  2         1986       Female 24
## 3  2         1952       Female 58
## 4  2         1963       Female 47
## 5  1         1948       Male   62
## 6  1         1951       Male   59
## 7  1         1959       Male   51
## 8  2         1952       Female 58
## 9  2         1939       Female 71
## 10 1         1953       Male   57
## # ... with 9,490 more rows
```