CCES Cumulative Common Content (2006 - 2017)

Shiro Kuriwaki*

Guide last updated: 2019-03-20

This dataset combines twelve years (2006 - 2017) of the Cooperative Congressional Election Study (Principal Investigators: Stephen Ansolabehere, Sam Luks, Brian Schaffner).

The Cooperative Congressional Election Study (CCES) is an online survey conducted around November of each year, asking a range of questions on political behavior and public opinion. Questions can change from year to year; this cumulative file includes standard questions asked multiple years.

This dataset was constructed based off CCES datasets from each year. A set of R scripts formatted, merged, and standardized these datasets to generate a `tibble`-style data frame. In addition, the same dataset is available on `Crunch`, an accessible analytics interface optimized for survey datasets.

Please note that this cumulative dataset makes modifications to the original CCES datasets for comparability. These modifications are only made when differences are deemed sufficiently minor, and are documented in source code (see below). However, for details on the survey methodology and a list of all questions, readers should consult the guides for each year.

To see the source code,

report a bug, or ask a question about the data, please feel free to file an issue from the source code page: https://github.com/kuriwaki/cces_cumulative. Alternatively, please contact me by email.

To obtain the individual year's CCES datasets,

search the CCES dataverse (https://dataverse.harvard.edu/dataverse/cces) or access the CCES homepage (https://cces.gov.harvard.edu/). Sign-up to the Crunch dataset from the homepage as well.

To examine the survey methodology,

consult the Methodology section of the most recent Common Content's codebook: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GDF6Z0.

Getting Started

The `.Rds` format can be read into R. This format preserves dataset properties such as the distinction between integers and doubles, and labelled variables. Unlike a `.Rdata` file, an `.Rds` file must be assigned to an object when loading.

```
df <- readRDS("cumulative_2006_2017.Rds")
```

The dataset in R is best viewed in `dplyr`, although it can be analyzed as a standard data frame.

```
library(tidyverse)
df
```

```
# A tibble: 394,755 x 77
    year case_id weight weight_cumulati~ state st    cd     dist dist_up
   <int>   <int>  <dbl>            <dbl> <chr> <chr> <chr> <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
 7  2006  439251   0.777           0.565 Minn~ MN    MN-3      3       3
 8  2006  439254   0.839           0.610 Neva~ NV    NV-2      2       2
 9  2006  439255   0.331           0.241 Texas TX    TX-24    24      24
10  2006  439263   1.10            0.802 Mary~ MD    MD-2      2       2
# ... with 394,745 more rows, and 68 more variables: cong <int>,
#   cong_up <int>, zipcode <chr>, county_fips <chr>, tookpost <int+lbl>,
#   weight_post <dbl>, starttime <dttm>, pid3 <int+lbl>,
#   pid3_leaner <int+lbl>, pid7 <int+lbl>, ideo5 <fct>, gender <int+lbl>,
#   birthyr <int>, age <int>, race <int+lbl>, hispanic <int+lbl>,
#   educ <int+lbl>, faminc <fct>, economy_retro <int+lbl>,
#   newsint <int+lbl>, approval_pres <int+lbl>, approval_rep <fct>,
#   approval_sen1 <fct>, approval_sen2 <fct>, approval_gov <int+lbl>,
#   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_rep_party <fct>, voted_rep <fct>, voted_rep_party <fct>,
#   intent_gov <fct>, intent_gov_party <fct>, voted_gov <fct>,
#   voted_gov_party <fct>, intent_sen <fct>, intent_sen_party <fct>,
#   voted_sen <fct>, voted_sen_party <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>
```

A Stata dta file is provided as well. `cumulative_2006_2017.dta` can be read by Stata, or in R by the haven package

```
library(haven)
df <- read_dta("cumulative_2006_2017.dta")
```

A note on variable types. The R dataset stores variables in `numeric`, `character`, `factor`, or `labelled` format. The first three classes are commonly used, but the `lablelled` format is more recent. Essentially it is similar to a `factor` but more compatible with Stata and SPSS. It is built around R's haven package, which includes more documentation.

A labelled variable's labels are not shown immediately in the Console:

```
select(df, year, case_id, pid3)
```

```
# A tibble: 394,755 x 3
     year case_id pid3
    <int>   <int> <int+lbl>
 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 394,745 more rows
```

But labels can be displayed by transforming the labelled vector into a factor.

```
library(haven)
select(df, year, case_id, pid3) %>%
  mutate(pid3_fct = as_factor(pid3))
```

```
# A tibble: 394,755 x 4
     year case_id pid3      pid3_fct
    <int>   <int> <int+lbl> <fct>
 1   2006  439219 1         Democrat
 2   2006  439224 4         Other
 3   2006  439228 1         Democrat
 4   2006  439237 1         Democrat
 5   2006  439238 1         Democrat
 6   2006  439242 3         Independent
 7   2006  439251 2         Republican
 8   2006  439254 1         Democrat
 9   2006  439255 1         Democrat
10   2006  439263 1         Democrat
# ... with 394,745 more rows
```

Features of the 2006 - 2017 Cumulative Dataset

Unified Variable Names

Most variables in this dataset come straight from each year's CCES. However, it renames and standardizes variable names, making them accessible in one place. Please see the rest of this guide or the Crunch dataset for a full list and description of variables.

Chosen Candidate Names and Identifiers

One addition to this cumulative dataset are variables of candidate names and identifiers that a respondent chose. In the individual year's CCES datasets, typically the response values for a vote choice question is a generic label, e.g. `Candidate1` and `Candidate2`. Then, separate variables of names and parties correspond to each `Candidate1` and `Candidate2`.

 Instead, the cumulative dataset shows both the generic label *and* the chosen candidate's name, party, and identifier, which will vary across individuals.

```
select(df, year, case_id, st, matches("voted_sen"))
```

```
# A tibble: 394,755 x 7
    year case_id st    voted_sen voted_sen_party voted_sen_chosen
   <int>   <int> <chr> <fct>     <fct>           <chr>
 1  2006  439219 NC    <NA>      <NA>            <NA>
 2  2006  439224 OH    [Democra~ Democratic      Sherrod C. Brow~
 3  2006  439228 NJ    [Democra~ Democratic      Robert Menendez~
 4  2006  439237 IL    <NA>      <NA>            <NA>
 5  2006  439238 NY    [Democra~ Democratic      Hillary Rodham ~
 6  2006  439242 TX    I Did No~ <NA>            <NA>
 7  2006  439251 MN    [Republi~ Republican      Mark Kennedy (R)
 8  2006  439254 NV    [Democra~ Democratic      Jack Carter (D)
 9  2006  439255 TX    [Democra~ Democratic      Barbara Ann Rad~
10  2006  439263 MD    I Did No~ <NA>            <NA>
# ... with 394,745 more rows, and 1 more variable: voted_sen_fec <chr>
```

Crunch

A version of the dataset is also included in Crunch, a database platform that makes it easy to view and analyze survey data either with our without any programming experience. Crunch is in beta at the time of writing.

1. Obtain Access: For View access to the dataset (free), please sign up here: https://harvard.az1.qualtrics.com/jfe/form/SV_066hQi4Eeco3Kap. For questions and more access, please contact the CCES Team.

2. Browse: Crunch offers a web GUI for quickly browsing variables:
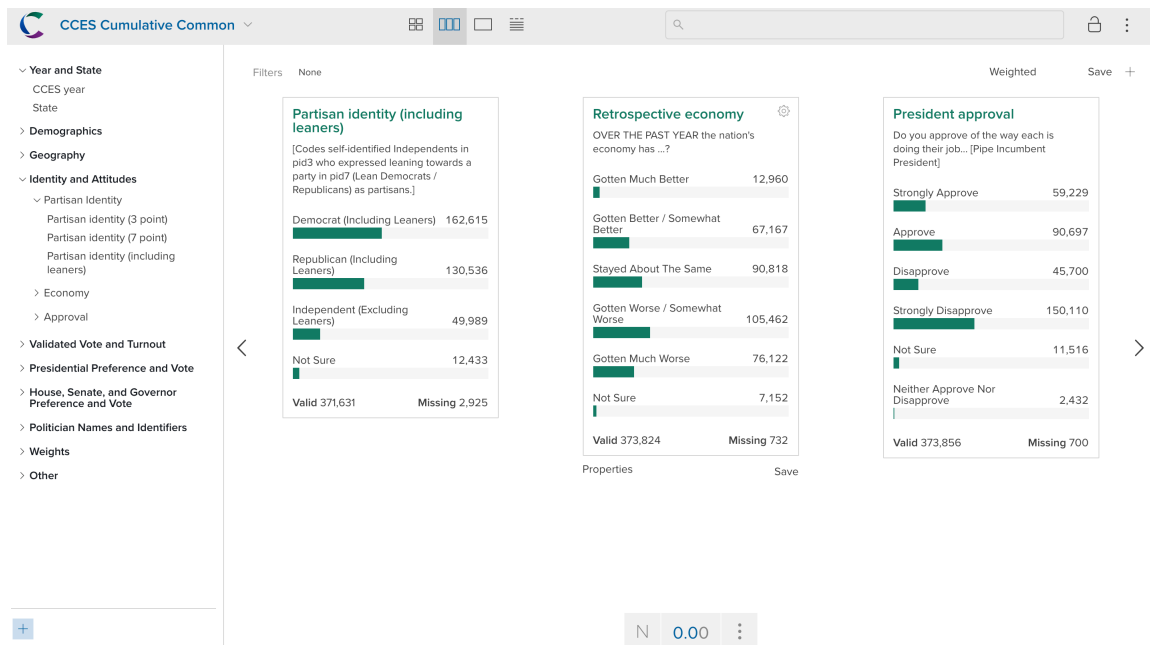
3. Analyze: The crunch interface allows Viewers to make cross-tabs and bar graphs quickly.
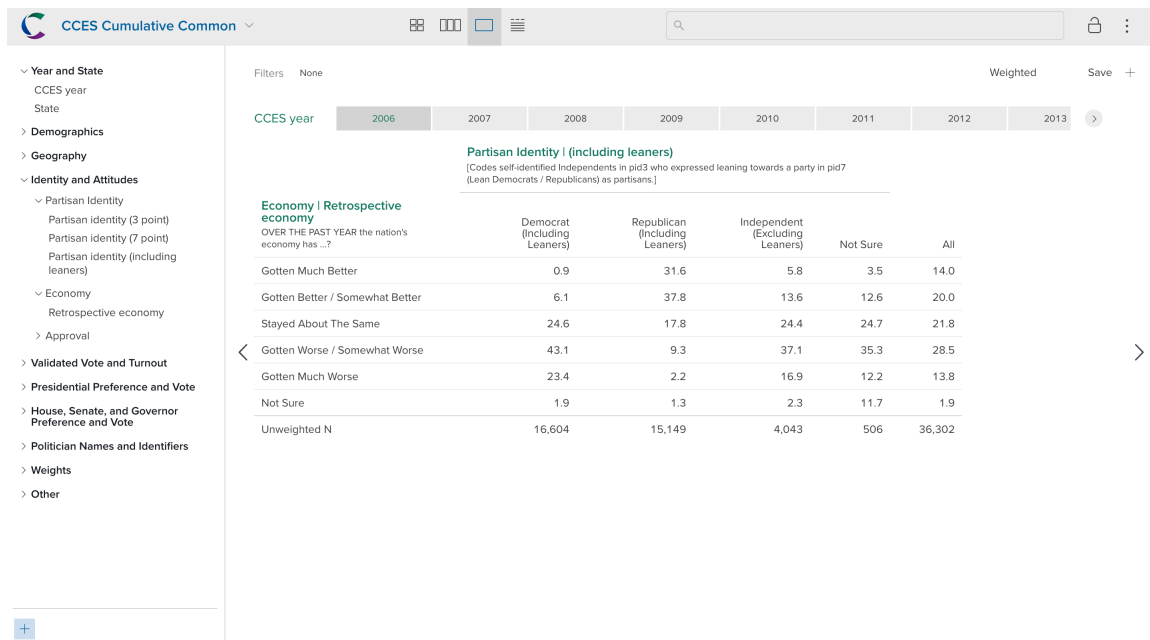
Crunch datasets can also be manipulated from a R package, `crunch` https://github.com/Crunch-io/rcrunch.

Variables

The sections below provide summary more information on each variable.

- The title shows the name as used in the dataset, suitable for coding ("alias" in Crunch terminology). followed by a more descriptive. name suitable for presentation ("name" in Crunch terminology).
- Question wordings, where applicable, immediately follow. Otherwise an description is provide in square brackets ([   ]). All square brackets, both in the description and the response options, indicate descriptions that are summaries of what the respondent saw rather than the question verbatim.
- A tabulation of response options (or summary statistics for numeric variables) follow. Numbers are unweighted counts.
- The "Years" bullet lists the years of the CCES in which data on the variable is available at all. If a year is not listed, either the question was not asked in the year or was not incorporated in the creation of this dataset.
- Finally, the "Limitations" bullet notes some of the caveats required when interpreting this variable. As this dataset is combinations of different surveys, some year-specific details on implementation are inevitably lost. For example, for all 2016 responses "Not Asked" and "Skipped" are both coded as a NA (missing) to stay consistent with past years that did not make that finer distinction.

Administration

year: CCES year

[Year of CCES Common Content]

```
year      n
2006 36421
2007   9999
2008 32800
2009 13800
2010 55400
2011 20150
2012 54535
2013 16400
2014 56200
2015 14250
2016 64600
2017 18200
2018   2000
```

starttime: Start time

[Pre-election wave start time (up to second)]

```
              Min.                1st Qu.                 Median
"2006-10-07 00:02:34" "2010-10-06 17:21:07" "2012-10-10 11:58:30"
              Mean                3rd Qu.                   Max.
"2012-09-17 22:09:37" "2015-11-11 02:25:15" "2018-11-05 22:18:53"
```

- Years: All of 2006-2018

tookpost: Took post-election wave

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

```
Warning: Factor 'tookpost' contains implicit NA, consider using
'forcats::fct_explicit_na'

                              tookpost      n
 Did Not Take Post-Election Survey  51143
         Took Post-Election Survey 250813
                             <NA>  92799
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018 (Post-election wave only exists for even years)

Weights

weight: Survey weight (Year-Specific)

[weights from pre-election survey of each year]

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
0.0000  0.4076  0.7219  1.0000  1.1887 15.0006   2000
```

- Years: All of 2006-2017
- In even years, they are re-computed after vote validation has been computed and those re-computed weights are taken here when available. The weights applied to the sample (which is originally drawn from a matched sample) are constructed to make each year's respondents' pool representative of the national adult population. See the methodology section of the 2016 Guide for details.
- Limitations: Only specific to each year. Built off of the entire pre-election wave sample, but not necessarily to adjust post-election wave respondents. See `weight_post`

weight_cumulative: Survey weight (Cumulative)

[weight variable with simple adjustment: multiplied a constant within year to make years comparable]

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
0.0000  0.2464  0.4803  0.8089  0.9657 19.3960   2000
```

- Years: All of 2006-2017
- Limitations: Only a simple transformation of `weight`. Specifically, `weight_cumulative` is `weight` divided by the following factor, for a given year.

```
# A tibble: 12 x 3
    year  size size_factor
   <dbl> <int>       <dbl>
 1  2006 36421        1.38
 2  2007  9999        0.378
 3  2008 32800        1.24
 4  2009 13800        0.521
 5  2010 55400        2.09
 6  2011 20150        0.761
```

```
 7  2012 54535      2.06
 8  2013 16400      0.619
 9  2014 56200      2.12
10  2015 14250      0.538
11  2016 64600      2.44
12  2017 18200      0.687
```

For example, all weights in the 2012 common content are divided by about 2.06, because it has about twice as many observations as the other datasets.

weight_post: Survey weight for post-election wave

[weight for post-election wave respondents. Only available for some of the even years.]

```
Min. 1st Qu.  Median   Mean 3rd Qu.    Max.    NA's
0.00    0.43    0.71   1.00    1.13   15.00  296978
```

- Years: 2012, 2016
- Limitations: Only available for some even years.

Geography

A series of variables for the respondent's location

- state: State (FIPS): [State (Imputed from input zipcode)]
- st: State abbreviation (FIPS): [State (Imputed from input zipcode)]
- dist: Congressional district number in current Congress: [Current Congressional District Number (Imputed from input zipcode)]
- dist_up: Congressional district number for upcoming Congress: [Upcoming Congressional District Number (Imputed from input zipcode)]
- cd: Congressional district in current Congress: [Current Congressional District (Imputed from input zipcode)]
- zipcode: Zipcode of residence: So that we can ask you about the news and events in your area, in what zip code do you currently reside?
- county_fips: County of residence: [County (Imputed from input zipcode)]

```
Observations: 394,755
Variables: 7
$ state       <chr> "California", "Pennsylvania", "Texas", "Texas", "T...
$ st          <chr> "CA", "PA", "TX", "TX", "TX", "NY", "NC", "NC", "M...
$ cd          <chr> "CA-2", "PA-5", "TX-16", "TX-19", "TX-6", "NY-28",...
$ dist        <int> 2, 5, 16, 19, 6, 28, 11, 7, 1, 17, 15, 1, 2, 6, 1,...
$ dist_up     <int> 1, 3, 16, 19, 6, 27, 11, 7, 2, 20, 12, 1, 2, 8, 1,...
$ zipcode     <chr> "95969", "16255", "79924", "79423", "76123", "1413...
$ county_fips <chr> "06007", "42031", "48141", "48303", "48439", "3606...
```

- Years: All of 2006-2018
- Limitations: Some years do not provide the variable relevant to dist_up, in which case the current district (dist) is assigned automatically. Thus, dist_up may not reflect, for example, district changes in off-cycle redistricting. Only residence (not registration) geographies included here; see individual years' for registration geographies.

Demographics

gender: Gender

Are you male or female?

```
gender       n
  Male 185146
Female 209609
```

- Years: All of 2006-2018

birthyr: Year of birth

In what year were you born?

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1900    1950    1960    1962    1975    2000
```

- Years: All of 2006-2018

age: Age

[Approximate age computed from the year of survey minus Year of Birth]

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 18.00   37.00   52.00   49.84   62.00  109.00    2000
```

- Years: All of 2006-2017

educ: Education

What is the highest level of education you have completed?

```
Warning: Factor 'educ' contains implicit NA, consider using
'forcats::fct_explicit_na'
```

```
               educ      n
              No HS  11846
High School Graduate 109264
       Some College 100070
             2-Year  37777
             4-Year  89245
          Post-Grad  46486
               <NA>     67
```

- Years: All of 2006-2018

race: Race

What racial or ethnic group best describes you?

```
           race       n
          White 294271
          Black  43718
       Hispanic  31123
          Asian   7394
Native American   3112
          Mixed   7498
          Other   7024
 Middle Eastern    615
```

- Years: All of 2006-2018
- Limitations: The "Hispanic" value may undercount self-identified Hispanics. See `hispanic`

hispanic: Hispanic

Are you of Spanish, Latino, or Hispanic origin or descent? [Asked if response to race is not Hispanic]

```
Warning: Factor 'hispanic' contains implicit NA, consider using
'forcats::fct_explicit_na'

 hispanic      n
      Yes   9170
       No 265938
     <NA> 119647
```

- Years: 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018
- In years in which this question was fielded, this question supplements the `race` variable by asking does who did *not* respnd "Hipsanic" in the `race` question.

faminc: Family Income

Thinking back over the last year, what was your family's annual income? [Brackets coarsened]

```
Warning: Factor 'faminc' contains implicit NA, consider using
'forcats::fct_explicit_na'

          faminc     n
   Less than 10k 15880
        10k - 20k 28630
        20k - 30k 40176
        30k - 40k 40586
        40k - 50k 36487
        50k - 60k 35620
        60k - 70k 26004
        70k - 80k 28193
       80k - 100k 32741
      100k - 120k 23952
      120k - 150k 18936
            150k+ 21259
 Prefer not to say 43063
```

```
     Skipped    12
        <NA>  3216
```

- Years: All of 2006-2017
- Limitations: The income brackets provided changed slightly over time. The brackets in this cumulative dataset coarsens some brackets, with loss of some granularity. In particular, from 2011-2016, respondents answering "over 150k" were asked in a follow-up question when the respondent initially selected "over 150k". These are top-coded and only labelled as "over 150k."
- The 2009 CCES did not have an option for 60-70k.

Validations

vv_regstatus: Validated registration status

[Validation results. Missing if validation was not conducted in the year. Categories are aggregated. Both Matched-not registered and unmatched are labeled as a no record.]

```
Warning: Factor 'vv_regstatus' contains implicit NA, consider using
'forcats::fct_explicit_na'
```

```
              vv_regstatus       n
                    Active  178356
 No Record Of Registration   61861
              Unregistered   13826
                   Dropped    5294
                  Inactive    3047
       Multiple Appearances    1151
                      <NA>  131220
```

- Years: 2008, 2010, 2012, 2014, 2016
- Limitations: Collapses some response options

vv_party_gen: Validated registered party

[Validation results. Only available for some staets and years]

```
Warning: Factor 'vv_party_gen' contains implicit NA, consider using
'forcats::fct_explicit_na'
```

```
                   vv_party_gen       n
 No Record Of Party Registration   60890
                        Unknown   51379
                Democratic Party   27058
                Republican Party   21522
             No Party Affiliation    9835
                Declined To State    1579
                           Other    1286
                Independent Party    1176
                Liberatarian Party     376
                     Green Party     194
              Constitution Party      27
                    Reform Party       9
                 Socialist Party       3
                            Cns       1
                           <NA>  219420
```

- Years: 2012, 2014, 2016
- Limitations: Not available for some even years

vv_party_prm: Validated registered Primary party

[Validation results. Only available for some staets and years]

```
Warning: Factor 'vv_party_prm' contains implicit NA, consider using
'forcats::fct_explicit_na'
```

```
               vv_party_prm       n
No Record Of Party Registration 157120
               Republican Party   10010
               Democratic Party    8202
                          Other       3
                           <NA> 219420
```

- Years: 2012, 2014, 2016
- Limitations: Not available for some even years

Turnout

vv_turnout_gvm: Validated turnout General Election

[Validation results. All vote methods (polling, mail, early, unknown, etc..) are aggregated as a vote.]

Warning: Factor 'vv_turnout_gvm' contains implicit NA, consider using
'forcats::fct_explicit_na'

```
     vv_turnout_gvm       n
             Voted 169204
No Record Of Voting 129019
     No Voter File    1733
             <NA>   94799
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016
- Limitations: Collapses most response options. In particular, the particular voting method is collapsed into one category, even though gvm stands for General Election voting method. Also, the result of not matching to a voter file is collapsed with the result of matching to a voter file and having no indication of turning out to vote. The distinction is unclear in earlier years, and is thus collapsed for all years here. For finer distinctions, see the individual year's CCES.

vv_turnout_pvm: Validated turnout Primary Election (Congressional)

[Validation results]

Warning: Factor 'vv_turnout_pvm' contains implicit NA, consider using
'forcats::fct_explicit_na'

```
     vv_turnout_pvm       n
No Record Of Voting 185927
             Voted  76245
     No Voter File    1363
             <NA> 131220
```

- Years: 2008, 2010, 2012, 2014, 2016
- Limitations: See vv_turnout_gvm

Identity and Attitudes

Partisan Identity

pid3: Partisan identity (3 point)

Generally speaking, do you think of yourself as a . . . ?

```
Warning: Factor 'pid3' contains implicit NA, consider using
'forcats::fct_explicit_na'

        pid3      n
    Democrat 139919
  Republican 103398
 Independent 110162
       Other  15535
    Not Sure  16841
        <NA>   8900
```

- Years: All of 2006-2018
- Limitations: Response options offer slightly by year. For example, the Not Sure option is not a response option in years 2006 and 2010. Open-text responses not included. 2010 values are from the post-election wave.

pid7: Partisan identity (7 point)

[Based on branching from Partisan Identity question]

```
Warning: Factor 'pid7' contains implicit NA, consider using
'forcats::fct_explicit_na'

                       pid7     n
           Strong Democrat 94021
  Not Very Strong Democrat 47849
              Lean Democrat 39451
               Independent 52706
           Lean Republican 42211
Not Very Strong Republican 38070
         Strong Republican 66030
                  Not Sure 11491
                      <NA>  2926
```

- Years: All of 2006-2018
- Limitations: See pid3

pid3_leaner: Partisan identity (including leaners)

[Codes self-identified Independents in pid3 who expressed leaning towards a party in pid7 (Lean Democrats / Republicans) as partisans.]

```
Warning: Factor 'pid3_leaner' contains implicit NA, consider using
'forcats::fct_explicit_na'
```

```
                  pid3_leaner        n
      Democrat (Including Leaners) 181321
    Republican (Including Leaners) 146311
  Independent (Excluding Leaners)  52706
                         Not Sure  11491
                            <NA>    2926
```

- Years: All of 2006-2018
- Limitations: See pid3

ideo5: Ideology (5 point)

In general, how would you describe your own political viewpoint?

Warning: Factor 'ideo5' contains implicit NA, consider using
'forcats::fct_explicit_na'

```
            ideo5      n
     Very Liberal  32975
          Liberal  69032
         Moderate 124892
     Conservative  93629
Very Conservative  45250
         Not Sure  27609
             <NA>   1368
```

- Years: All of 2006-2018

Economy

economy_retro: Retrospective economy

OVER THE PAST YEAR the nation's economy has . . . ?

Warning: Factor 'economy_retro' contains implicit NA, consider using
'forcats::fct_explicit_na'

```
                    economy_retro       n
             Gotten Much Better   18179
  Gotten Better / Somewhat Better   85983
           Stayed About The Same 101850
    Gotten Worse / Somewhat Worse 105662
               Gotten Much Worse  74666
                       Not Sure   7644
                            <NA>    771
```

- Years: All of 2006-2018
- Limitations: Response options varies by year. Some are collapsed into one category (e.g. Gotten Better, presented in some years, and Gotten Somewhat Better, presented in other years, are collapsed into Gotten Better / Somewhat Better). Some are left as is. For example, Not Sure was not an option in 2009.

News Interest

`newsint`: News Interest

Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs ..

```
Warning: Factor 'newsint' contains implicit NA, consider using
'forcats::fct_explicit_na'

          newsint      n
 Most Of The Time 196583
 Some Of The Time  90164
Only Now And Then  42629
   Hardly At All   20303
      Don't Know    8441
            <NA>   36635
```

- Years: 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018
- Limitations: Not asked in 2006.

Approval

`approval_pres`: President approval

Do you approve of the way each is doing their job. . . [Pipe Incumbent President]

```
Warning: Factor 'approval_pres' contains implicit NA, consider using
'forcats::fct_explicit_na'

                     approval_pres      n
             Strongly Approve  78352
             Somewhat Approve  95536
          Somewhat Disapprove  41953
          Strongly Disapprove 166967
                    Not Sure   10800
Neither Approve Nor Disapprove    443
                         <NA>    704
```

- Years: All of 2006-2018
- Limitations: `Neither approve nor disapprove` only included in 2007.
- This question is asked in a grid format, along with Governors, Congress, and Courts.

`approval_rep`: House Representative approval

Do you approve of the way each is doing their job. . . [Pipe Incumbent Representative's Name]

```
Warning: Factor 'approval_rep' contains implicit NA, consider using
'forcats::fct_explicit_na'

                     approval_rep      n
              Strongly Approve  57686
      Approve / Somewhat Approve 122975
 Disapprove / Somewhat Disapprove  68619
```

```
          Strongly Disapprove  61684
        Never Heard / Not Sure  75650
Neither Approve Nor Disapprove   1798
                          <NA>   6343
```

- Years: All of 2006-2018
- Limitations: `Neither approve nor disapprove` only included in 2007.
- This question is asked in a grid format, along with Senators (`approval_sen1`, `approval_sen2`).
- To see who [Representative] refers to for a particular respondent, see `rep_inc` (incumbent identifier in `rep_icpsr`)

approval_sen1: Senator 1 approval

Do you approve of the way each is doing their job. . . [Pipe Incumbent Senator 1's Name]

```
Warning: Factor 'approval_sen1' contains implicit NA, consider using
'forcats::fct_explicit_na'
```

```
                 approval_sen1      n
               Strongly Approve  51711
       Approve / Somewhat Approve 124869
 Disapprove / Somewhat Disapprove  78618
             Strongly Disapprove  76862
           Never Heard / Not Sure  56948
   Neither Approve Nor Disapprove   1413
                            <NA>   4334
```

- Years: All of 2006-2018
- Limitations: : Response options varies by year. Some are collapsed into one category (e.g. `Approve`, presented in some years, and `Somewhat Approve`, presented in other years, are collapsed into `Approve / Somewhat Approve`). `Neither approve nor disapprove` only included in 2007.
- To see who [Senator 1] refers to for a particular respondent, see `sen1_inc` (incumbent identifier in `sen1_icpsr`)

approval_sen2: Senator 2 approval

Do you approve of the way each is doing their job. . . [Pipe Incumbent Senator 2's Name]

```
Warning: Factor 'approval_sen2' contains implicit NA, consider using
'forcats::fct_explicit_na'
```

```
                 approval_sen2      n
               Strongly Approve  54229
       Approve / Somewhat Approve 120459
 Disapprove / Somewhat Disapprove  77087
             Strongly Disapprove  77507
           Never Heard / Not Sure  59207
   Neither Approve Nor Disapprove   1158
                            <NA>   5108
```

- See `approval_sen2`

approval_gov: Governor approval

Do you approve of the way each is doing their job... Governor of [Pipe State]

Warning: Factor 'approval_gov' contains implicit NA, consider using
'forcats::fct_explicit_na'

```
                   approval_gov      n
               Strongly Approve  57801
               Somewhat Approve 122423
            Somewhat Disapprove  74506
            Strongly Disapprove 102470
                       Not Sure  34168
 Neither Approve Nor Disapprove   1414
                           <NA>   1973
```

- Years: All of 2006-2018
- Limitations: See `approval_pres`
- To see who the Governor refers to for a particular respondent, see `gov_inc` (incumbent identifier in `gov_fec`, if applicable).

Presidential Vote

> ### A note on `intent` and `voted`
>
> In this dataset we make the distinction between "intent" / "preference" vs. "voted" / "vote choice". "Intent" (or "preference") refers to the response to the prospective question of the sort "who would you vote for?" in the *pre-election* wave. "Vote choice" refers to the response to the retrospective question of the sort "in the election this November, who did you vote for?" Response to the vote choice questions coalesces both *post-election* wave responses (the bulk of the responses) and pre-election respondents who reported having already voted early.

`intent_pres_08`: 2008 President preference (before voting)

For which candidate for President of the United States would you vote?

Warning: Factor 'intent_pres_08' contains implicit NA, consider using 'forcats::fct_explicit_na'

```
            intent_pres_08      n
               John McCain  13322
             Barack Obama  12897
                 Ron Paul    535
               Ralph Nader    209
                 Bob Barr    258
          Cynthia McKinney     74
                    Other    352
 I Won't Vote In This Election    851
             I'm Not Sure   1697
                     <NA> 364560
```

- Years: 2008

`intent_pres_12`: 2012 President preference (before voting)

In the race for President of the United States, who do you prefer?

Warning: Factor 'intent_pres_12' contains implicit NA, consider using 'forcats::fct_explicit_na'

```
                intent_pres_12      n
      Mitt Romney (Republican)  20738
     Barack Obama (Democratic)  24401
                        Other   1781
 I Will Not Vote In This Election   1467
               I'm Not Sure   3856
                        <NA> 342512
```

- Years: 2012

`intent_pres_16`: 2016 President preference (before voting)

Which candidate did you prefer for President of the United States?

```
Warning: Factor 'intent_pres_16' contains implicit NA, consider using
'forcats::fct_explicit_na'

                intent_pres_16       n
    Donald Trump (Republican)    19227
   Hillary Clinton (Democrat)    27502
   Gary Johnson (Libertarian)     3145
           Jill Stein (Green)     1400
                        Other     1880
 I Won't Vote In This Election     3312
              I'm Not Sure        6536
                         <NA>   331753
```

- Years: 2016

`voted_pres_08`: 2008 President vote choice (after voting)

2008: For which candidate for President of the United States did you vote? [see guide for wording in all years]

```
Warning: Factor 'voted_pres_08' contains implicit NA, consider using
'forcats::fct_explicit_na'

            voted_pres_08       n
 Barack Obama (Democratic)  73986
  John McCain (Republican)  68398
             Someone Else    4204
             Did Not Vote   18227
             Don't Recall    1787
                     <NA> 228153
```

- Years: 2008, 2009, 2010, 2011, 2012
- Limitations: Response options offer slightly by year; some are collapsed into one.

`voted_pres_12`: 2012 President vote choice (after voting)

2012: For whom did you vote for President of the United States? 2016: In 2012, who did you vote for in the election for President? [see guide for wording in all years]

```
Warning: Factor 'voted_pres_12' contains implicit NA, consider using
'forcats::fct_explicit_na'

              voted_pres_12       n
             Barack Obama   82681
              Mitt Romney   64956
      Other / Someone Else    5890
              Did Not Vote    2758
     Not Sure / Don't Recall   1990
 I Did Not Vote In This Race     81
                      <NA> 236399
```

- Years: 2012, 2013, 2014, 2015, 2016
- Limitations: Response options offer slightly by year; some are collapsed into one.
- This variable coalesces two variables: Either the response to the early vote question in the pre-election wave if the respondent indicates they have already voted, or if not, the response in the post-election wave.

`voted_pres_16`: 2016 President vote choice (after voting)

2017: In the election for U.S. President, who did you vote for? [If reported voting] 2016: For whom did you vote for President of the United States? [Post-election]

```
Warning: Factor 'voted_pres_16' contains implicit NA, consider using
'forcats::fct_explicit_na'
```

voted_pres_16	n
Donald Trump	25199
Hilary Clinton	30148
Other / Someone Else	5630
Did Not Vote	241
Not Sure / Don't Recall	331
<NA>	333206

- Years: 2016, 2017, 2018
- This variable coalesces two variables in the CCES: Either the response to the early vote question in the pre-election wave if the respondent indicates they have already voted, or if not, the response in the post-election wave.

House, Senate and Governor Voting

Preference

`intent_rep`: House preference (before voting)

In the general election for U.S. House of Representatives in your area, who do you prefer? See `intent_rep_party` for party affiliation.

```
Warning: Factor 'intent_rep' contains implicit NA, consider using
'forcats::fct_explicit_na'
```

intent_rep	n
[Democrat / Candidate 1]	104693
[Republican / Candidate 2]	97632
[Other / Candidate 3]	4085
$HouseCand4Name ($HouseCand4Party)	18
Other	1737
I'm Not Sure	60920
No One	15965
$HouseCand5Name ($HouseCand5Party)	20
I Won't Vote In This Election	2269
$HouseCand6Name ($HouseCand6Party)	19
$HouseCand7Name ($HouseCand7Party)	15
$HouseCand8Name ($HouseCand8Party)	14
$HouseCand9Name ($HouseCand9Party)	1

```
$HouseCand10Name ($HouseCand10Party)        1
$HouseCand11Name ($HouseCand11Party)        3
                                <NA> 107363
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: Only available for even years. The third party candidate not specified for early years. The fourth candidate and onwards not shown for most years. Response options differ by year.
- Note that it is not always the case that 1 is a Democrat and 2 is a Republican. When two Democrats are on the general ballot (e.g. in top-two primary states like California), both candidates are Democrats. Use `intent_rep_party` for party affiliation.
- Note that for each respondent, a name (and party affiliation) is shown in place of the square bracket values. To see the candidate chosen, see `intent_rep_chosen`. [Other / Candidate 3] refers to the third option presented, whereas `Other` refers to the unnamed choice after all numbered candidates.

`intent_sen`: Senate preference (before voting)

In the race for U.S. Senator in your state, who do you prefer? See `intent_sen_party` for party affiliation.

```
Warning: Factor 'intent_sen' contains implicit NA, consider using
'forcats::fct_explicit_na'

                      intent_sen      n
      [Democrat / Candidate 1]  78964
    [Republican / Candidate 2]  69174
         [Other / Candidate 3]   4131
 $SenCand4Name ($SenCand4Party)     19
                         Other   1213
                  I'm Not Sure  31884
                        No One   9586
   I Won't Vote In This Election   1145
                          <NA> 198639
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: See `intente_rep`. When both senate seats are up for re-election in the same year, only responses to the first senate seat is incorporated. For the second senate seat, see individual year's CCES.

`intent_gov`: Governor preference (before voting)

In the race for Governor in your state, who do you prefer? See `intent_gov_party` for party affiliation.

```
Warning: Factor 'intent_gov' contains implicit NA, consider using
'forcats::fct_explicit_na'

                   intent_gov      n
    [Democrat / Candidate 1]  56231
  [Republican / Candidate 2]  50783
       [Other / Candidate 3]   3696
                       Other    900
```

```
                I'm Not Sure   18532
                     No One    5793
I Won't Vote In This Election     466
                       <NA>  258354
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: See `intente_rep`. For governor elections in odd years, see individual year's CCES.

Vote Choice

voted_rep: House vote choice (after voting)

For whom did you vote for U.S. House? See `voted_rep_party` for party affiliation.

```
Warning: Factor 'voted_rep' contains implicit NA, consider using
'forcats::fct_explicit_na'

                               voted_rep       n
              [Democrat / Candidate 1]   95401
            [Republican / Candidate 2]   94629
                 [Other / Candidate 3]    2578
    $HouseCand4Name ($HouseCand4Party)      15
                                 Other    2459
            I Did Not Vote In This Race   11630
    $HouseCand5Name ($HouseCand5Party)      22
                              Not Sure    4032
    $HouseCand6Name ($HouseCand6Party)      17
    $HouseCand7Name ($HouseCand7Party)      13
    $HouseCand8Name ($HouseCand8Party)      16
    $HouseCand9Name ($HouseCand9Party)       2
  $HouseCand10Name ($HouseCand10Party)       2
  $HouseCand11Name ($HouseCand11Party)       3
                                  <NA>  183936
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- This variable coalesces two variables in the CCES for years 2012 and onwards: Either the response to the early vote question in the pre-election wave if the respondent indicates they have already voted, or if not, the response in the post-election wave.
- Note that it is not always the case that 1 is a Democrat and 2 is a Republican. When two Democrats are on the general ballot (e.g. in top-two primary states like California), both candidates are Democrats. Use `voted_rep_party` for party affiliation.

voted_sen: Senate vote choice (after voting)

For whom did you vote for U.S. Senator? See `voted_sen_party` for party affiliation.

```
Warning: Factor 'voted_sen' contains implicit NA, consider using
'forcats::fct_explicit_na'

                   voted_sen       n
      [Democrat / Candidate 1]   69365
    [Republican / Candidate 2]   64243
         [Other / Candidate 3]    2755
```

```
                Other    1638
             Not Sure    1856
$SenCand4Name ($SenCand4Party)     11
  I Did Not Vote In This Race    4133
                      <NA> 250754
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- This variable coalesces two variables in the CCES for years 2012 and onwards: Either the response to the early vote question in the pre-election wave if the respondent indicates they have already voted, or if not, the response in the post-election wave.

voted_gov: Governor vote choice (after voting)

For whom did you vote for Governor? See `voted_gov_party` for party affiliation.

```
Warning: Factor 'voted_gov' contains implicit NA, consider using
'forcats::fct_explicit_na'

                 voted_gov      n
   [Democrat / Candidate 1]  47215
 [Republican / Candidate 2]  45662
     [Other / Candidate 3]   2478
                    Other   1181
I Did Not Vote In This Race   4682
                Not Sure    918
                    <NA> 292619
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- This variable coalesces two variables in the CCES for years 2012 and onwards: Either the response to the early vote question in the pre-election wave if the respondent indicates they have already voted, or if not, the response in the post-election wave.

Text

Identifiers

The case identifier `case_id` is unique within the year and is identical to the case identifiers in the individual year's CCES. It should be used in conjunction with `year` for a unique identifier for the whole dataset. Some individuals across years may be the same YouGov panel respondent with different identifiers; for example the 2007 CCES draws from the 2006 CCES respondents.

```
Observations: 394,755
Variables: 2
$ year    <int> 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, ...
$ case_id <int> 439219, 439224, 439228, 439237, 439238, 439242, 439251...
```

Current Representatives

Name and Party

The four names in the three offices that represent the respondent *at the time of the survey*. Parties are not shown if the particular year's CCES did not show party. Party names are also abbreviated down to initials (`D` for Democrat, `R` for Republican, `I` for Independent) in this dataset.

```
Observations: 394,755
Variables: 4
$ rep_current  <chr> "Patrick T. McHenry (R)", "Michael R. Turner (R)"...
$ sen1_current <chr> "Elizabeth Dole (R)", "Mike DeWine (R)", "Robert ...
$ sen2_current <chr> "Richard Burr (R)", "George V. Voinovich (R)", "F...
$ gov_current  <chr> "Michael Easley (D)", "Bob Taft (R)", "Jon Corzin...
```

Incumbent Identifiers

Unique identifiers (ICPSR / Nominate for Congress, FEC for Governor) for the current representatives. Identifiers are not part of the individual year's CCES but merged on for this cumulative dataset only.

The matching of identifiers to respondent occurs through matching by district, by district and last name, or both:

- For House representatives, we join on `cong`, `st`, and `dist` to a NOMINATE database that only consists of unique observations according to the key. For duplicates with regards to these three variables (e.g. in the rare case where a new representative comes into office mid-session), we match on `cong`, `st`, `dist` and last name.
- For Senators, we join entirely on `cong`, `st`, and last name
- For Governors, we join only on `st` and last name. In this period, there are no two governors in the same state that share the same last name.

```
Observations: 394,755
Variables: 4
$ rep_icpsr  <int> 20522, 20342, 29132, 29911, 29380, 20531, 29126, 29...
$ sen1_icpsr <int> 40303, 15020, 29373, 15021, 14858, 49306, 40101, 15...
$ sen2_icpsr <int> 29548, 49903, 14914, 40502, 40105, 40305, 40302, 29...
$ gov_fec    <chr> "NC5998", NA, "NJ6395", "IL7", NA, "TX3156", "MN472...
```

- Years: All of 2006-2018
- Limitations: Matching procedure may be incomplete or inaccurate.

The unique identifiers can be used to join with other databases to append additional information such as committee membership and ideology scores, such as

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet (2017). Voteview: Congressional Roll-Call Votes Database. https://voteview.com/

Candidates

The text responses that the respondent chose in each of the `intent_` / `voted_` questions, if the respondent was a candidate. For example, respondent with `case_id = 163051575` in the 2012 CCES chose the first option in the House representative preference question. `intent_rep_chosen` shows that for this particular respondent, the first option was `Maxine Waters (Democrat)` who has a FEC Identifier of `H4CA23011`.

```
df %>%
  filter(year == 2012, st == "CA", dist_up == 43) %>%
  select(matches("intent_rep"))
```

```
# A tibble: 91 x 4
   intent_rep           intent_rep_party intent_rep_chos~ intent_rep_fec
   <fct>                <fct>            <chr>            <chr>
 1 [Democrat / Candidate ~ Democratic    Maxine Waters (~ H4CA23011
 2 I'm Not Sure         <NA>             <NA>             <NA>
 3 No One               <NA>             <NA>             <NA>
 4 [Democrat / Candidate ~ Democratic    Maxine Waters (~ H4CA23011
 5 [Republican / Candidat~ Democratic    Bob Flores (D)   H2CA43385
 6 I'm Not Sure         <NA>             <NA>             <NA>
 7 Other                <NA>             <NA>             <NA>
 8 [Republican / Candidat~ Democratic    Bob Flores (D)   H2CA43385
 9 [Republican / Candidat~ Democratic    Bob Flores (D)   H2CA43385
10 [Democrat / Candidate ~ Democratic    Maxine Waters (~ H4CA23011
# ... with 81 more rows
```

The name and party are those as provided in the CCES datasets (e.g. in the form `HouseCand1Name`). The FEC ID is not part of the CCES but joined in this dataset.

For all three offices, the matching generally occurs by `year`, `st`, `dist_up` (not `dist`, because `dist_up`, refers to the district of the upcoming session) and `party`. `party` is the party affiliation as indicated in the CCES. For years 2008 and 2010, the first option is automatically labelled as a Democrat and the second option as a Republican, although these may be inaccurate at times.

The FEC database runs up until 2014 (thus more recent candidates do *not* get a FEC ID) originates from

Bonica, Adam , 2015, "Database on Ideology, Money in Politics, and Elections (DIME)", doi:10.7910/DVN/O5PX0B, Harvard Dataverse, V2

which helpfully includes candidates office sought, district (for House members), party affiliation, and cycle in which the candidate filed. The variable `cycle` in Bonica's data is used to join on the CCES dataset's `year` variable.

Only candidates who are unique within the district and party are considered for the first join. However, many candidates are not unique within the district-party, as many co-partisans may file in the same district. The second matching process thus considers the full name of the candidate

listed in the CCES and the candidates in the FEC database. *Within* the subset of year, district, and party, a Jaro-Winker string distance (that ranges from 0 to 1) is computed for both last name and the first name - middle name. If the sum of the two string distances are more than 0.2 for all possible combinations, no match is returned. If there is a unique combination that achieves a unique minimum that is below 0.2, that combination is declared a match. If there are multiple matches with the same minimum string distance, one is randomly chosen.

Chosen

```
Observations: 394,755
Variables: 6
$ intent_rep_chosen <chr> "Richard C. Carsner (D)", "Stephanie Studeba...
$ intent_sen_chosen <chr> NA, "Sherrod C. Brown (D)", "Robert Menendez...
$ intent_gov_chosen <chr> NA, "Ted Strickland (D)", NA, "Rod Blagojevi...
$ voted_rep_chosen  <chr> "Richard C. Carsner (D)", "Stephanie Studeba...
$ voted_sen_chosen  <chr> NA, "Sherrod C. Brown (D)", "Robert Menendez...
$ voted_gov_chosen  <chr> NA, "Ted Strickland (D)", NA, "Rod Blagojevi...
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- Early years may mislabel the candidate's party, especially when the two candidates are of the same party (as in top-two primary states)

Candidate Identifiers

```
Observations: 394,755
Variables: 6
$ intent_rep_fec <chr> "H6NC10141", "H6OH03142", "H0NJ01066", "H8IL090...
$ intent_sen_fec <chr> NA, "S6OH00163", "S6NJ00289", NA, NA, NA, "S6MN...
$ intent_gov_fec <chr> NA, "OH19691", NA, "IL7", "NY19490", NA, "MN472...
$ voted_rep_fec  <chr> "H6NC10141", "H6OH03142", "H0NJ01066", "H8IL090...
$ voted_sen_fec  <chr> NA, "S6OH00163", "S6NJ00289", NA, "S0NY00188", ...
$ voted_gov_fec  <chr> NA, "OH19691", NA, "IL7", "NY19490", NA, "MN472...
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016
- Limitations: Matching may be inaccurate (see previous section on matching methodology). In particular, a lack of a FEC ID may either indicate a failure of the matching procedure, or that the candidate in question did not file under the FEC. The match rate in the current procedure is upwards of 80 percent in the current procedure.