CCES Cumulative Common Content (2006 - 2016)

Shiro Kuriwaki*

Guide last updated: 2018-01-17

Kuriwaki, Shiro, 2018, "Cumulative CCES Common Content (2006-2016)", doi:10.7910/DVN/II2DB6, Harvard Dataverse

This dataset combines eleven years (2006 - 2016) of the Cooperative Congressional Election Study (Principal Investigators: Stephen Ansolabehere, Brian Schaffner, and Sam Luks).

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. While questions can change from year to year, some standard questions like the ones this cumulative file includes allow for interesting comparisons.

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/).

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.

*Department of Government, Harvard University. Email: kuriwaki@g.harvard.edu, Website: www.shirokuriwaki.com. My thanks to Steve Ansolabehere, Stephen DiMauro, Nathan Kaplan, Stephen Pettigrew, Joe Williams, and the Crunch Team for their contributions and guidance.

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.

```
df <- readRDS("cumulative_2006_2016.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: 374,556 x 67
    year case_id weight weight_cumulative state  st    cd     dist dist_up
   <int>   <int>  <dbl>             <dbl> <chr>  <chr> <chr> <int>   <int>
 1  2006  439219  1.85              1.67  North~ NC    NC-10    10      10
 2  2006  439224  0.968             0.872 Ohio   OH    OH-3      3       3
 3  2006  439228  1.59              1.44  New J~ NJ    NJ-1      1       1
 4  2006  439237  1.40              1.26  Illin~ IL    IL-9      9       9
 5  2006  439238  0.903             0.813 New Y~ NY    NY-22    22      22
 6  2006  439242  0.839             0.756 Texas  TX    TX-11    11      11
 7  2006  439251  0.777             0.700 Minne~ MN    MN-3      3       3
 8  2006  439254  0.839             0.756 Nevada NV    NV-2      2       2
 9  2006  439255  0.331             0.299 Texas  TX    TX-24    24      24
10  2006  439263  1.10              0.993 Maryl~ MD    MD-2      2       2
# ... with 374,546 more rows, and 58 more variables: cong <int>, cong_up
#   <int>, zipcode <chr>, county_fips <chr>, tookpost <int+lbl>, weight_vv
#   <dbl>, weight_vv_post <dbl>, starttime <dttm>, pid3 <int+lbl>, pid7
#   <int+lbl>, gender <int+lbl>, birthyr <int>, age <int>, race <int+lbl>,
#   educ <int+lbl>, economy_retro <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_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>
```

A Stata dta file is provided as well. `cumulative_2006_2016.dta` can be read by Stata, or in R by the haven package

```
library(haven)
df <- read_dta("cumulative_2006_2016.dta")
```

Features of the 2006 - 2016 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"))
```

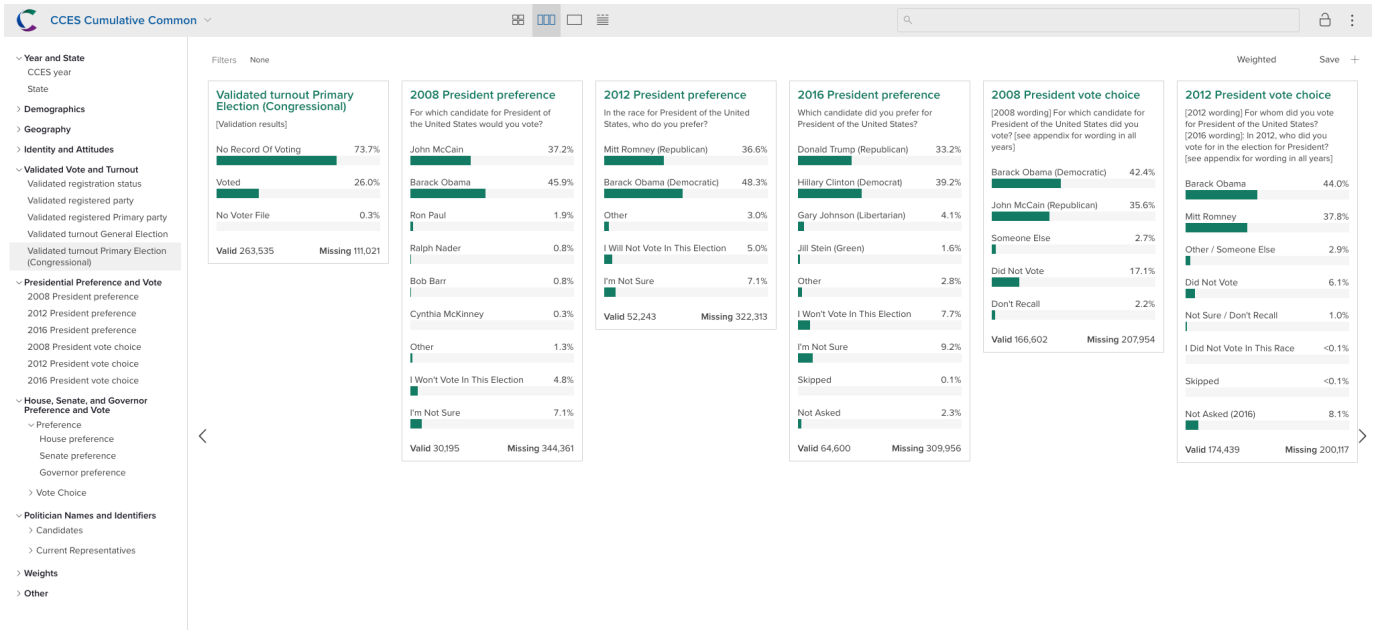
```
# A tibble: 374,556 x 6
    year case_id st    voted_sen         voted_sen_chosen    voted_sen_fec
   <int>   <int> <chr> <fct>             <chr>               <chr>
 1  2006  439219 NC    <NA>              <NA>                <NA>
 2  2006  439224 OH    [Democrat / Cand~ Sherrod C. Brown (~ S6OH00163
 3  2006  439228 NJ    [Democrat / Cand~ Robert Menendez (D) S6NJ00289
 4  2006  439237 IL    <NA>              <NA>                <NA>
 5  2006  439238 NY    [Democrat / Cand~ Hillary Rodham Cli~ S0NY00188
 6  2006  439242 TX    I Did Not Vote I~ <NA>                <NA>
 7  2006  439251 MN    [Republican / Ca~ Mark Kennedy (R)    S6MN00275
 8  2006  439254 NV    [Democrat / Cand~ Jack Carter (D)     S6NV00150
 9  2006  439255 TX    [Democrat / Cand~ Barbara Ann Radnof~ S6TX00180
10  2006  439263 MD    I Did Not Vote I~ <NA>                <NA>
# ... with 374,546 more rows
```

Crunch

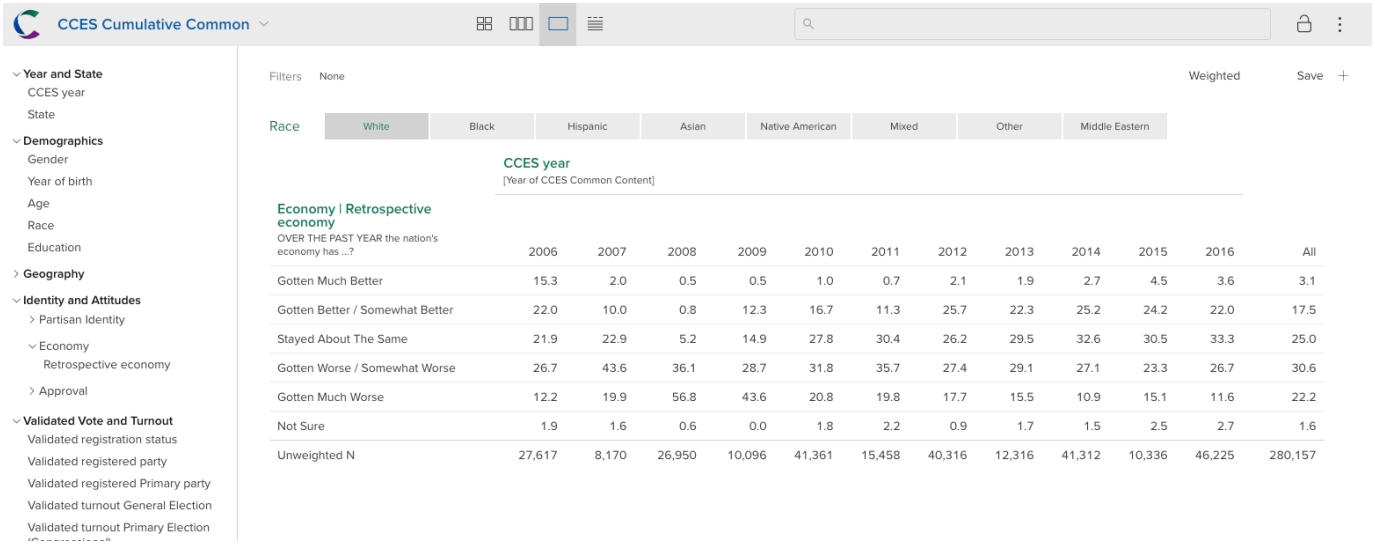
A version of the dataset is also included in Crunch, a database platform that makes it easy to view and analyze survey data either with our without any programming experience. Crunch is in beta at the time of writing.

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 is followed by the name of the variable in the dataset ("alias" in Crunch) followed by the variable label ("name" in Crunch).
- 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 verbatim response).
- A tabulation of response options (or summary statistics for numeric variables) follow. Numbers are unweighted counts. Response options are ordered by the ordering in the dataset.
- 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 incorporate 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.

Administration

`year`: CCES year

[Year of CCES Common Content]

year	n
2006	36421
2007	10000
2008	32800
2009	13800
2010	55400
2011	20150
2012	54535
2013	16400
2014	56200
2015	14250
2016	64600

`starttime`: Start time

[Pre-election wave start time]

```
                Min.               1st Qu.                Median
"2006-10-07 00:02:34" "2010-10-01 19:36:34" "2012-10-07 10:30:42"
                Mean               3rd Qu.                  Max.
"2012-06-06 07:11:46" "2014-10-18 06:01:53" "2016-11-07 16:46:42"
```

- Years: All of 2006-2016

`tookpost`: Took post-election wave

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

```
                          tookpost      n
 Did NOT Take Post-Election Survey  50872
         Took Post-Election Survey 249084
                             <NA>  74600
```

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

Weights

`weight`: Survey weight (Year-Specific)

[weights from pre-survey of each year]

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.3859  0.6986  1.0000  1.1686 20.0004
```

- Years: All of 2006-2016
- 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.

`weight_cumulative`: Survey weight (Cumulative)

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

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.2892  0.5687  0.9633  1.1382 24.0297
```

- Years: All of 2006-2016
- Limitations: Only a simple transformation of `weight`

`weight_vv`: Survey weight for validated voters

[weight among validated voters to construct a representative sample of the voters. Only available for some years.]

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00    0.43    0.74    1.00    1.19   15.00  255421
```

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

`weight_vv_post`: Survey weight for post-election wave validated voters

[weight among validated voters and post-election wave respondents to construct a representative sample of the voters. Only available for some years.]

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00    0.32    0.64    0.89    1.06   15.00  265121
```

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

Geography

A series of variables for the respondent's location

- `state`: State: [State (Imputed from input zipcode)]
- `st`: State abbreviation: [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: 374,556
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 2006-2016
- 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 176528
 Female 198028
```

- Years: All of 2006-2016

`birthyr`: Year of birth

In what year were you born?

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1900    1949    1960    1962    1975    1998
```

- Years: All of 2006-2016

`age`: Age

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

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   37.00   52.00   49.92   62.00  109.00
```

- Years: All of 2006-2016

`educ`: Education

What is the highest level of education you have completed?

```
                educ      n
               No HS  11128
 High School Graduate 103364
         Some College  95472
              2-Year  35368
              4-Year  85044
           Post-Grad  44113
                <NA>     67
```

- Years: All of 2006-2016

`race`: Race

What racial or ethnic group best describes you?

```
            race      n
           White 280670
           Black  41334
        Hispanic  28449
           Asian   6991
 Native American   2937
           Mixed   6901
           Other   6707
  Middle Eastern    567
```

- Years: All of 2006-2016

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 labelled as a no record.]

vv_regstatus	n
Active	176387
No Record Of Registration	64474
Unregistered	13397
Dropped	5168
Inactive	2983
Multiple Appearances	1126
<NA>	111021

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

vv_party_gen: Validated registered party

[Validation results]

vv_party_gen	n
No Record Of Party Registration	63049
Unknown	51053
Democratic Party	26323
Republican Party	21131
No Party Affiliation	9187
Declined To State	1559
Other	1275
Independent Party	1169
Liberatarian Party	366
Green Party	185
Constitution Party	27
Reform Party	9
Socialist Party	2
<NA>	199221

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

vv_party_prm: Validated registered Primary party

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

vv_party_prm	n
No Record Of Party Registration	157651
Republican Party	9827
Democratic Party	7857
<NA>	199221

- 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.]

vv_turnout_gvm	n
Voted	167531
No Record Of Voting	130692
No Voter File	1733
<NA>	74600

- 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]

vv_turnout_pvm	n
No Record Of Voting	186905
Voted	75267
No Voter File	1363
<NA>	111021

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

Identity and Attitudes

Partisan Identity

`pid3`: Partisan identity

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

```
        pid3      n
    Democrat 132407
  Republican  98672
 Independent 104296
       Other  14597
   Not Sure  15684
    Skipped     24
       <NA>   8876
```

- Years: All of 2006-2016
- 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.

`pid7`: Partisan identity (7 point)

[based on branching from Partisan Identity question]

```
                      pid7     n
           Strong Democrat 89117
   Not Very Strong Democrat 45242
             Lean Democrat 37432
               Independent 49221
           Lean Republican 40464
 Not Very Strong Republican 36225
          Strong Republican 63147
                   Not Sure 10783
                    Skipped    40
                  Not Asked    24
                       <NA>  2861
```

- Years: All of 2006-2016
- Limitations: See `pid3`

Economy

`economy_retro`: Retrospective economy

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

```
                 economy_retro      n
             Gotten Much Better  15621
 Gotten Better / Somewhat Better  80090
          Stayed About The Same  95382
  Gotten Worse / Somewhat Worse 102542
```

```
Gotten Much Worse  73450
         Not Sure   6739
             <NA>    732
```

- Years: All of 2006-2016
- 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.

Approval

`approval_pres`: President approval

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

```
                approval_pres      n
             Strongly Approve  74158
             Somewhat Approve  92205
          Somewhat Disapprove  40113
          Strongly Disapprove 156820
                    Not Sure  10117
Neither Approve Nor Disapprove   443
                      Skipped    72
                         <NA>   628
```

- Years: All of 2006-2016
- 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]

```
                  approval_rep      n
               Strongly Approve  55069
      Approve / Somewhat Approve 116457
Disapprove / Somewhat Disapprove  65067
            Strongly Disapprove  58680
         Never Heard / Not Sure  71268
  Neither Approve Nor Disapprove   1798
                        Skipped     96
                      Not Asked    487
                           <NA>   5634
```

- Years: All of 2006-2016
- 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]

approval_sen1	n
Strongly Approve	49354
Approve / Somewhat Approve	118550
Disapprove / Somewhat Disapprove	74192
Strongly Disapprove	72873
Never Heard / Not Sure	53904
Neither Approve Nor Disapprove	1414
Skipped	243
Not Asked	230
<NA>	3796

- Years: All of 2006-2016
- 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]

approval_sen2	n
Strongly Approve	51255
Approve / Somewhat Approve	114286
Disapprove / Somewhat Disapprove	73083
Strongly Disapprove	73362
Never Heard / Not Sure	56368
Neither Approve Nor Disapprove	1158
Skipped	177
Not Asked	230
<NA>	4637

- See `approval_sen2`

approval_gov: Governor approval

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

approval_gov	n
Strongly Approve	54544
Somewhat Approve	116578
Somewhat Disapprove	71062
Strongly Disapprove	97473
Not Sure	31575
Neither Approve Nor Disapprove	1414

```
     Skipped     83
 Not Asked     230
      <NA>    1597
```

- Years: All of 2006-2016
- 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

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

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>	344361

- Years: 2008

`intent_pres_12`: 2012 President preference

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

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>	322313

- Years: 2012

`intent_pres_16`: 2016 President preference

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

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
                     Skipped     77
                  Not Asked   1521
                       <NA> 309956
```

- Years: 2016

`voted_pres_08`: 2008 President vote choice

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

```
            voted_pres_08      n
Barack Obama (Democratic)  73986
 John McCain (Republican)  68398
             Someone Else   4204
             Did Not Vote  18227
             Don't Recall   1787
                     <NA> 207954
```

- 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

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

```
              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
                    Skipped    105
          Not Asked (2016)  15978
                       <NA> 200117
```

- 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

For whom did you vote for President of the United States?

voted_pres_16	n
Donald Trump (Republican)	18826
Hillary Clinton (Democrat)	22264
Gary Johnson (Libertarian)	1858
Jill Stein (Green)	924
Other	1144
I Didn't Vote In This Election	89
I'm Not Sure	238
Evan McMullin (Independent)	164
Skipped	27
Not Asked	7630
<NA>	321392

- Years: 2016
- 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

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

intent_rep	n
[Democrat / Candidate 1]	103873
[Republican / Candidate 2]	97039
[Other / Candidate 3]	4071
$HouseCand4Name ($HouseCand4Party)	18
Other	1720
I'm Not Sure	60579
No One	15860
$HouseCand5Name ($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>	89054

- Years: 2006, 2008, 2010, 2012, 2014, 2016

- 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 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

In the race for U.S. Senator in your state, who do you prefer?

intent_sen	n
[Democrat / Candidate 1]	78318
[Republican / Candidate 2]	68733
[Other / Candidate 3]	4113
$SenCand4Name ($SenCand4Party)	19
Other	1188
I'm Not Sure	31681
No One	9493
I Won't Vote In This Election	1145
<NA>	179866

- Years: 2006, 2008, 2010, 2012, 2014, 2016
- 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

In the race for Governor in your state, who do you prefer?

intent_gov	n
[Democrat / Candidate 1]	55600
[Republican / Candidate 2]	50244
[Other / Candidate 3]	3681
Other	882
I'm Not Sure	18342
No One	5723
I Won't Vote In This Election	466
<NA>	239618

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

Vote Choice

`voted_rep`: House vote choice

For whom did you vote for U.S. House?

```
                          voted_rep      n
           [Democrat / Candidate 1]  94644
         [Republican / Candidate 2]  94109
              [Other / Candidate 3]   2571
  $HouseCand4Name ($HouseCand4Party)     15
                               Other   2434
       I Did Not Vote In This Race  11576
  $HouseCand5Name ($HouseCand5Party)     22
                            Not Sure   4015
  $HouseCand6Name ($HouseCand6Party)     15
  $HouseCand7Name ($HouseCand7Party)     13
  $HouseCand8Name ($HouseCand8Party)     16
  $HouseCand9Name ($HouseCand9Party)      2
$HouseCand10Name ($HouseCand10Party)      2
$HouseCand11Name ($HouseCand11Party)      3
                                <NA> 165119
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016
- 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_sen`: Senate vote choice

For whom did you vote for U.S. Senator?

```
                      voted_sen      n
       [Democrat / Candidate 1]  68796
     [Republican / Candidate 2]  63838
          [Other / Candidate 3]   2743
                          Other   1624
                       Not Sure   1845
 $SenCand4Name ($SenCand4Party)     11
                        Skipped     10
    I Did Not Vote In This Race   4097
                           <NA> 231592
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016
- 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

For whom did you vote for Governor?

```
                    voted_gov      n
     [Democrat / Candidate 1]  46270
   [Republican / Candidate 2]  44908
        [Other / Candidate 3]   2451
                        Other   1158
```

```
I Did Not Vote In This Race   3928
                  Not Sure    911
                      <NA> 274930
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016
- 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: 374,556
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: 374,556
```

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: 374,556
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 2006-2016
- 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 3
   intent_rep                 intent_rep_chosen intent_rep_fec
   <fct>                      <chr>             <chr>
 1 [Democrat / Candidate 1]   Maxine Waters (D) H4CA23011
 2 I'm Not Sure               <NA>              <NA>
 3 No One                     <NA>              <NA>
 4 [Democrat / Candidate 1]   Maxine Waters (D) H4CA23011
 5 [Republican / Candidate 2] Bob Flores (D)    H2CA43385
 6 I'm Not Sure               <NA>              <NA>
 7 Other                      <NA>              <NA>
 8 [Republican / Candidate 2] Bob Flores (D)    H2CA43385
 9 [Republican / Candidate 2] Bob Flores (D)    H2CA43385
10 [Democrat / Candidate 1]   Maxine Waters (D) 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 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: 374,556
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
- 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: 374,556
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.