

# CCES Cumulative Common Content (2006 - 2018)

Shiro Kuriwaki\*

Guide last updated: 2019-03-26

*Please cite dataset as*

Kuriwaki, Shiro, 2018, "Cumulative CCES Common Content (2006-2017)", doi:10.7910/DVN/II2DB6, Harvard Dataverse

This dataset combines twelve years (2006 - 2018) 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](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 code-book: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GDF6Z0>.

---

\*Department of Government, Harvard University. Email: [kuriwaki@g.harvard.edu](mailto:kuriwaki@g.harvard.edu). Bug reports welcome. My thanks to Alexander Agadjanian, Steve Ansolabehere, Stephen DiMauro, Bernard Fraga, Nathan Kaplan, Mayya Komisarchik, and Stephen Pettigrew for their contributions. Thanks to Joe Williams, Jon Keane and Gordon Shotwell of YouGov and Crunch for their help.

## 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_2018.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: 452,755 x 78
```

	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 452,745 more rows, and 69 more variables: cong <int>,  
#   cong_up <int>, zipcode <chr>, county_fips <chr>, tookpost <int+lbl>,  
#   weight_post <dbl>, starttime <dtm>, 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>, marstat <int+lbl>,  
#   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\_2018.dta can be read by Stata, or in R by the haven package

```
library(haven)
df <- read_dta("cumulative_2006_2018.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 labelled 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: 452,755 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 452,745 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: 452,755 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 452,745 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"))
```

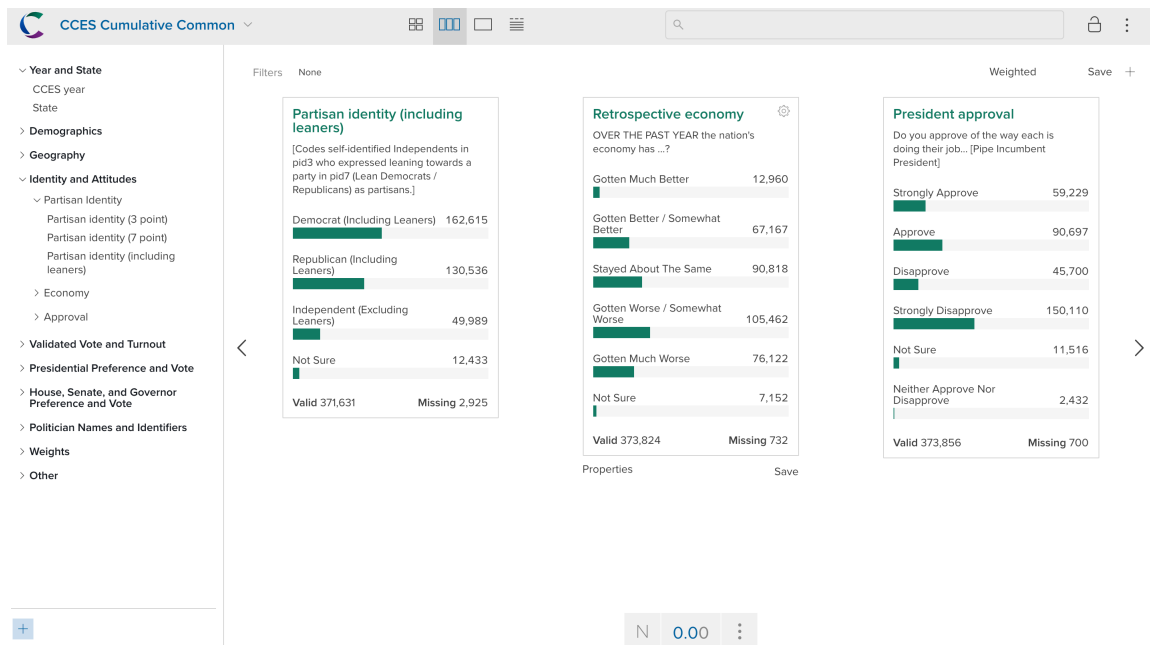
```
# A tibble: 452,755 x 7
  year case_id st   voted_sen voted_sen_party voted_sen_chosen
  <int>   <int> <chr>   <fct>      <fct>          <chr>
1  2006  439219 NC     <NA>       <NA>           <NA>
2  2006  439224 OH     [Democra~ Democratic     Sherrod C. Brow~
3  2006  439228 NJ     [Democra~ Democratic     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 452,745 more rows, and 1 more variable: voted_sen_fec <chr>
```

### Crunch

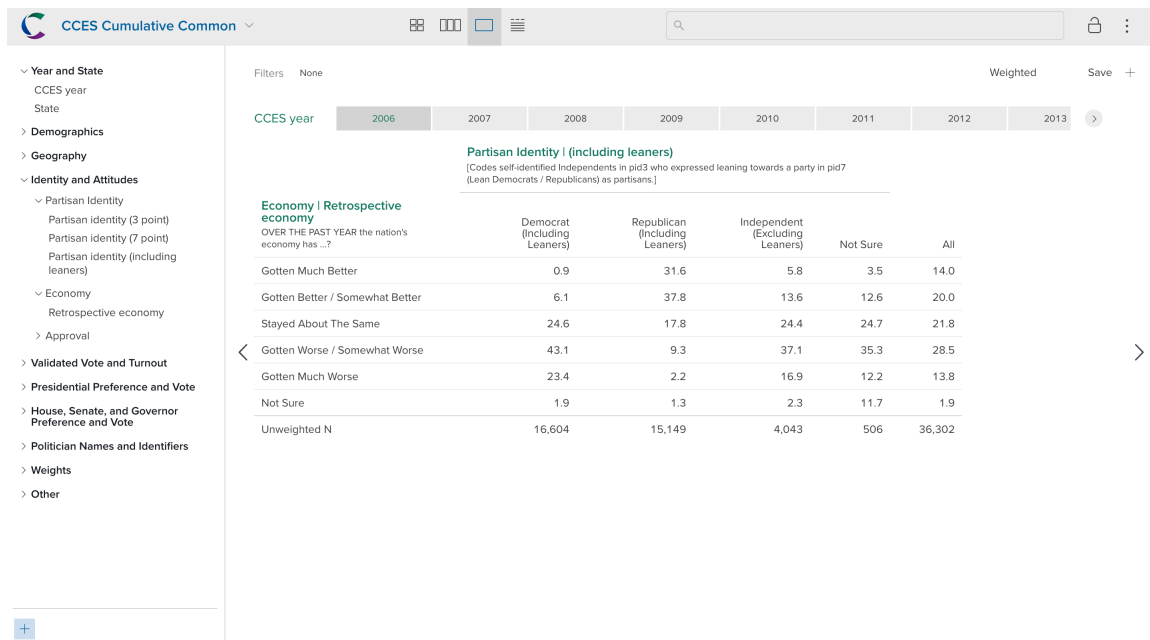
A version of the dataset is also included in Crunch, a database platform that makes it easy to view and analyze survey data either with or without any programming experience. Crunch is in beta at the time of writing.

1. Obtain Access: For View access to the dataset (free), please sign up here: [https://harvard.az1.qualtrics.com/jfe/form/SV\\_066hQi4Eeco3Kap](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	60000

### starttime: Start time

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

Min.	1st Qu.	Median
"2006-10-07 00:02:34"	"2010-10-11 15:40:49"	"2013-11-12 10:50:17"
Mean	3rd Qu.	Max.
"2013-06-29 00:06:52"	"2016-10-10 09:04:56"	"2018-11-05 23:27:49"

- 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		59064
Took Post-Election Survey		300892
	<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.
0.0000	0.4259	0.7281	1.0000	1.1741	15.0006

- Years: All of 2006-2018
- 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
0.00	0.25	0.48	0.81	0.97	19.40	60000

- 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	0.45	0.70	1.00	1.10	15.00	303170

- Years: 2012, 2016, 2018
- 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: 452,755

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 210102
Female 242653
```

- 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   1961   1963   1977   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.
18.00   36.00   51.00   49.61   62.00   109.00
```

- Years: All of 2006-2018

### 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  13939
High School Graduate 125323
Some College 112290
2-Year  43606
4-Year 103011
Post-Grad 54519
<NA>    67
```

- Years: All of 2006-2018

**race: Race**

What racial or ethnic group best describes you?

race	n
White	337793
Black	49162
Hispanic	35924
Asian	9134
Native American	3543
Mixed	8919
Other	7554
Middle Eastern	726

- 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	11113
No	321994
<NA>	119648

- 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	18780
10k - 20k	32995
20k - 30k	46019
30k - 40k	46728
40k - 50k	41768
50k - 60k	40863
60k - 70k	30063
70k - 80k	32459
80k - 100k	37809
100k - 120k	27566
120k - 150k	22108
150k+	25075
Prefer not to say	48955

```

Skipped    12
<NA>    1555

```

- Years: All of 2006-2018
- 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.

### **marstat: Marital Status**

What is your marital status?

Warning: Factor ‘marstat’ contains implicit NA, consider using ‘forcats::fct\_explicit\_na’

```

      marstat      n
Married 250629
Separated  7606
Divorced  49651
Widowed  21287
Single 101416
Domestic Partnership 20611
      <NA>    1555

```

- Years: All of 2006-2018

## Validations

All even years include indicators for validated voters, which means that YouGov has matched survey respondents' personal identifiable information to public voter files, which in turn officially record whether a person has voted or not. For more information, see Ansolabehere and Hersh (2012). Validation is often completed in the summer following the election; 2018 validation data is not available as of March 2019.

### **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>	189220

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

### **vv\_party\_gen: Validated registered party**

[Validation results. Only available for some states 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>	277420

- 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>	277420

- 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>	152799

- 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>	189220

- 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	160637
Republican	118907
Independent	126270
Other	17975
Not Sure	20012
<NA>	8954

- 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	107733
Not Very Strong Democrat	54838
Lean Democrat	45411
Independent	60946
Lean Republican	47739
Not Very Strong Republican	43810
Strong Republican	75782
Not Sure	13480
<NA>	3016

- 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)		207982
Republican (Including Leaners)		167331
Independent (Excluding Leaners)		60946
	Not Sure	13480
	<NA>	3016

- 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		40142
Liberal		79591
Moderate		141633
Conservative		105268
Very Conservative		52243
	Not Sure	32104
	<NA>	1774

- 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		29702
Gotten Better / Somewhat Better		100766
Stayed About The Same		118229
Gotten Worse / Somewhat Worse		114965
Gotten Much Worse		77949
	Not Sure	10294
	<NA>	850

- 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	225200
Some Of The Time	105337
Only Now And Then	50141
Hardly At All	24621
Don't Know	10449
<NA>	37007

- 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	92872
Somewhat Approve	104677
Somewhat Disapprove	47005
Strongly Disapprove	194516
Not Sure	12507
Neither Approve Nor Disapprove	443
<NA>	735

- 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	65102
Approve / Somewhat Approve	142503
Disapprove / Somewhat Disapprove	79595



Strongly Disapprove	70663
Never Heard / Not Sure	85772
Neither Approve Nor Disapprove	1798
<NA>	7322

- 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	58761
Approve / Somewhat Approve	144415
Disapprove / Somewhat Disapprove	90846
Strongly Disapprove	89017
Never Heard / Not Sure	63783
Neither Approve Nor Disapprove	1413
<NA>	4520

- 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	63135
Approve / Somewhat Approve	139337
Disapprove / Somewhat Disapprove	88539
Strongly Disapprove	89232
Never Heard / Not Sure	66076
Neither Approve Nor Disapprove	1158
<NA>	5278

- 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	67404
Somewhat Approve	139940
Somewhat Disapprove	84271
Strongly Disapprove	117282
Not Sure	40318
Neither Approve Nor Disapprove	1414
<NA>	2126

- 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>	422560

- 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>	400512

- 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>	389753

- 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>	286153

- 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>	294399

- 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	43891
Hilary Clinton	51342
Other / Someone Else	10091
Did Not Vote	627
Not Sure / Don't Recall	527
<NA>	346277

- 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]	128231
[Republican / Candidate 2]	115292
[Other / Candidate 3]	4401
\$HouseCand4Name (\$HouseCand4Party)	37
Other	2259
I'm Not Sure	70460
No One	19235
\$HouseCand5Name (\$HouseCand5Party)	23
I Won't Vote In This Election	2269
\$HouseCand6Name (\$HouseCand6Party)	41
\$HouseCand7Name (\$HouseCand7Party)	20
\$HouseCand8Name (\$HouseCand8Party)	14
\$HouseCand9Name (\$HouseCand9Party)	1

```
$HouseCand10Name ($HouseCand10Party)      1
$HouseCand11Name ($HouseCand11Party)      3
<NA> 110468
```

- 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] 97220
[Republican / Candidate 2] 82433
[Other / Candidate 3]    4477
$SenCand4Name ($SenCand4Party) 19
          Other      1713
          I'm Not Sure 38112
          No One      12419
I Won't Vote In This Election 1145
          <NA> 215217
```

- 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] 74561
[Republican / Candidate 2] 66292
[Other / Candidate 3]    4055
          Other      1390
```

I'm Not Sure	24296
No One	7991
I Won't Vote In This Election	466
<NA>	273704

- 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]		115799
[Republican / Candidate 2]		109342
[Other / Candidate 3]		2786
\$HouseCand4Name (\$HouseCand4Party)		27
Other		3124
I Did Not Vote In This Race		12539
\$HouseCand5Name (\$HouseCand5Party)		24
Not Sure		4493
\$HouseCand6Name (\$HouseCand6Party)		39
\$HouseCand7Name (\$HouseCand7Party)		15
\$HouseCand8Name (\$HouseCand8Party)		16
\$HouseCand9Name (\$HouseCand9Party)		2
\$HouseCand10Name (\$HouseCand10Party)		2
\$HouseCand11Name (\$HouseCand11Party)		3
<NA>		204544

- 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]		84885
[Republican / Candidate 2]		75584
[Other / Candidate 3]		2974

Other	1967
Not Sure	2094
\$SenCand4Name (\$SenCand4Party)	11
I Did Not Vote In This Race	4789
<NA>	280451

- 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]		67899
[Republican / Candidate 2]		62887
[Other / Candidate 3]		2805
Other		1820
I Did Not Vote In This Race		10290
Not Sure		1091
<NA>		305963

- 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: 452,755

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: 452,755

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: 452,755

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/05PX0B](https://doi.org/10.7910/DVN/05PX0B), 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: 452,755

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: 452,755

Variables: 6

```
$ intent_rep_fec <chr> "H6NC10141", "H60H03142", "H0NJ01066", "H8IL090...
$ intent_sen_fec <chr> NA, "S60H00163", "S6NJ00289", NA, NA, NA, "S6MN...
$ intent_gov_fec <chr> NA, "OH19691", NA, "IL7", "NY19490", NA, "MN472...
$ voted_rep_fec <chr> "H6NC10141", "H60H03142", "H0NJ01066", "H8IL090...
$ voted_sen_fec <chr> NA, "S60H00163", "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.