Guide to the CCES Cumulative Common Content (2006 - 2018)

Shiro Kuriwaki*

Guide last updated: 2020-03-07

Cite this dataset as:

> Kuriwaki, Shiro, 2019, "Cumulative CCES Common Content (2006-2019)", doi:10.7910/DVN/II2DB6, Harvard Dataverse, V4.

This dataset combines thirteen years (2006 - 2019) of the Cooperative Congressional Election Study (CCES). The CCES is an online survey conducted around November of each year, asking a range of questions on political behavior and public opinion. Its principal investigators are Stephen Ansolabehere, Sam Luks, and Brian Schaffner.

CCES questions change from year to year, so this cumulative file only includes a subset of those questions that are standard. Users can still merge in year-specific questions easily into this cumulative file and take advantage of its harmonized variables.

This dataset was constructed from CCES datasets from each year. The final product is a `tibble`-style data frame (built in R) that is also available as a Stata `dta` file. In addition, the same dataset is available on `Crunch`, an analytics interface optimized for survey datasets.

Please note that this cumulative dataset makes some 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 repository. Alternatively, please contact me by email.
- **To obtain the individual year's CCES datasets,** search the CCES dataverse or access the CCES homepage. Sign-up to the Crunch dataset from the homepage as well.
- **To understand the survey methodology,** consult the Frequently Asked Questions page of the CCES homepage or the methodology section of a recent Common Content's codebook.

*Department of Government, Harvard University. ORCID: https://orcid.org/0000-0002-5687-2647. Bug reports welcome. My thanks to Alexander Agadjanian, Steve Ansolabehere, Stephen DiMauro, Bernard Fraga, Nathan Kaplan, Mayya Komisarchik, Stephen Pettigrew, and Boris Shor for their suggestions and contributions. Thanks to Joe Williams at YouGov, and Jon Keane, Mike Malecki, and Gordon Shotwell at Crunch for their help.

Contents

Getting Started . . . 4
Labelled variables (for analysis in R) . . . 4
Adding more variables . . . 5
Features of the Cumulative Dataset . . . 6
Unified Variable Names . . . 6
Chosen Candidate Names and Identifiers . . . 6
Crunch . . . 6
Variables . . . 8
Administration . . . 8
year: CCES year . . . 8
starttime: Start time . . . 8
tookpost: Took post-election wave . . . 9
Weights . . . 9
weight: Survey weight (Year-Specific) . . . 9
weight_cumulative: Survey weight (Cumulative) . . . 9
rvweight: Survey weights to validated registered voters . . . 10
rvweight_post: Survey weights to validated registered voters, post-election wave . . . 10
weight_post: Survey weight for post-election wave . . . 10
Geography . . . 11
Demographics . . . 12
gender: Gender . . . 12
birthyr: Year of birth . . . 12
age: Age . . . 12
educ: Education . . . 12
race: Race . . . 12
hispanic: Hispanic . . . 13
faminc: Family Income . . . 13
marstat: Marital Status . . . 14
citizen: Citizenship . . . 14
Validations . . . 15
vv_regstatus: Validated registration status . . . 15
vv_party_gen: Validated registered party . . . 15
vv_party_prm: Validated registered Primary party . . . 16
Turnout . . . 16
vv_turnout_gvm: Validated turnout General Election . . . 16
vv_turnout_pvm: Validated turnout Primary Election (Congressional) . . . 16
Partisan Identity . . . 17
pid3: Partisan identity (3 point) . . . 17
pid7: Partisan identity (7 point) . . . 17
pid3_leaner: Partisan identity (including leaners) . . . 17
ideo5: Ideology (5 point) . . . 18
Economy . . . 18
economy_retro: Retrospective economy . . . 18
News Interest . . . 18
newsint: News Interest . . . 18
Approval . . . 19
approval_pres: President approval . . . 19
approval_rep: House Representative approval . . . 19
approval_sen1: Senator 1 approval . . . 20

approval_sen2: Senator 2 approval . . . 20
approval_gov: Governor approval . . . 20
Presidential Vote . . . 22
intent_pres_08: 2008 President preference (before voting) . . . 22
intent_pres_12: 2012 President preference (before voting) . . . 22
intent_pres_16: 2016 President preference (before voting) . . . 22
voted_pres_08: 2008 President vote choice (after voting) . . . 23
voted_pres_12: 2012 President vote choice (after voting) . . . 23
voted_pres_16: 2016 President vote choice (after voting) . . . 23
House, Senate and Governor Vote . . . 24
intent_rep: House preference (before voting) . . . 24
intent_sen: Senate preference (before voting) . . . 25
intent_gov: Governor preference (before voting) . . . 25
voted_rep: House vote choice (after voting) . . . 25
voted_sen: Senate vote choice (after voting) . . . 26
voted_gov: Governor vote choice (after voting) . . . 27
Metadata and Identifiers . . . 28
Identifiers . . . 28
Current Representatives' Name and Party . . . 28
ICPSR Identifiers . . . 28
Name of Chosen Candidate . . . 29
Party of Chosen Candidate . . . 30
Version History of Dataverse Releases . . . 31
Version 5.0 . . . 31
Version 4.0 . . . 31
Version 3.0 . . . 31
Version 2.0 . . . 31
Version 1.0 . . . 31

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 is assigned to an object.

```
cc <- readRDS("cumulative_2006_2019.Rds")
```

The dataset in R is best viewed with `dplyr`, although it can also be used without tidyverse.

```
library(tidyverse)
cc
```

A Stata dta version is provided as well. `cumulative_2006_2019.dta` can be read by Stata, or in R by the `haven` package

```
library(haven)
cc <- read_dta("cumulative_2006_2019.dta")
```

Labelled variables (for analysis in R)

A note on variable types. The R dataset stores variables in `numeric`, `character`, `factor`, or `labelled` class.[1] The first three classes are commonly used, but the `lablelled` format is more novel. `labelled` classes are numeric integers where each integer is associated with a label (See vignette here). This makes it equivalent to a `factor` but referenceable by its numeric value. It is essentially the labels in Stata and SPSS.

A labelled variable's labels are usually not shown. But recent versions of the `haven` package (version 2.1.0 or above) will display the associated labels in the Console if selected within tidyverse. This makes it immediately obvious which value is associated with which label:

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

```
# A tibble: 470,755 x 3
    year case_id            pid3
   <int>   <int>       <int+lbl>
 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 470,745 more rows
```

[1] Technically, this is now called a `labelled_haven` class, to disambiguate from an unrelated but older use of `labelled` in the Hmisc package.

Labels can be made explicit by coercing the labelled vector into a factor. However, this removes the numerical value codes of the labelled class.

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

```
# A tibble: 470,755 x 4
   year case_id         pid3 pid3_fct
  <int>   <int>    <int+lbl> <fct>
1  2006  439219 1 [Democrat] Democrat
2  2006  439224 4 [Other]    Other
3  2006  439228 1 [Democrat] Democrat
4  2006  439237 1 [Democrat] Democrat
5  2006  439238 1 [Democrat] Democrat
# ... with 4.708e+05 more rows
```

Unlike factors, labelled variables can be referenced by their underlying numeric value. It is sometimes useful to treat survey values as numbers rather than as raw text, and the labelled calss allows you to do that.

```
select(cc, year, case_id, pid3) %>%
  filter(pid3 == 1)
```

```
# A tibble: 167,172 x 3
   year case_id         pid3
  <int>   <int>    <int+lbl>
1  2006  439219 1 [Democrat]
2  2006  439228 1 [Democrat]
3  2006  439237 1 [Democrat]
4  2006  439238 1 [Democrat]
5  2006  439254 1 [Democrat]
# ... with 1.672e+05 more rows
```

In this cumulative (R) dataset, some variables are of class labelled, and some are of factor class. This is because the latter variables were different enough in their value codings across years that summarizing them into a single numeric value was difficult.

Adding more variables

As noted, the cumulative dataset only uses key variables from each year's common content. However,common content variables can be merged in.

In R, we recommend using the `left_join` or `inner_join` functions (or the base-R `merge` function). In Stata, use `merge 1:1`. In all cases, the combination of `year` and `case_id` **uniquely identifies each row** in the cumulative common content, so any merges should merge on year and the case identifier.

Features of the Cumulative Dataset

Beyond stacking together each year's common content, the cumulative dataset provides several additional features to facilitate analysis.

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 and party, which will vary across individuals.

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

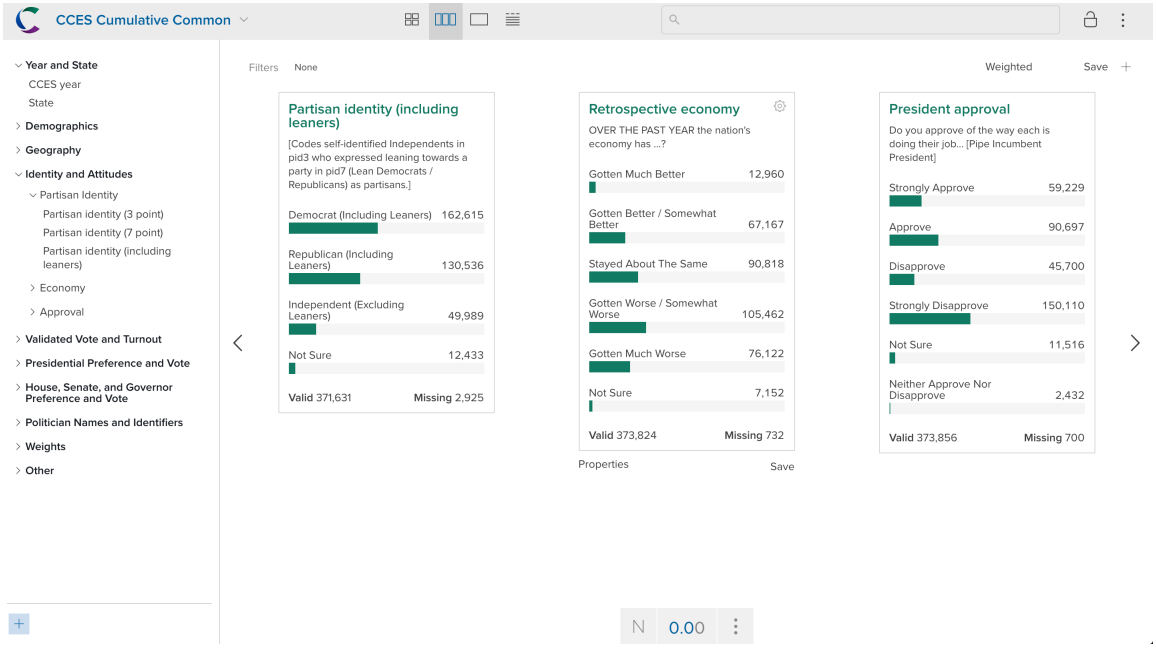
```
# A tibble: 470,755 x 6
    year case_id st    voted_sen           voted_sen_party voted_sen_chosen
   <int>   <int> <chr> <fct>               <fct>           <chr>
 1  2006  439219 NC    <NA>                <NA>            <NA>
 2  2006  439224 OH    [Democrat / Candid~ Democratic      Sherrod C. Brown (D)
 3  2006  439228 NJ    [Democrat / Candid~ Democratic      Robert Menendez (D)
 4  2006  439237 IL    <NA>                <NA>            <NA>
 5  2006  439238 NY    [Democrat / Candid~ Democratic      Hillary Rodham Clint~
 6  2006  439242 TX    I Did Not Vote In ~ <NA>            <NA>
 7  2006  439251 MN    [Republican / Cand~ Republican      Mark Kennedy (R)
 8  2006  439254 NV    [Democrat / Candid~ Democratic      Jack Carter (D)
 9  2006  439255 TX    [Democrat / Candid~ Democratic      Barbara Ann Radnofsk~
10  2006  439263 MD    I Did Not Vote In ~ <NA>            <NA>
# ... with 470,745 more rows
```

Crunch

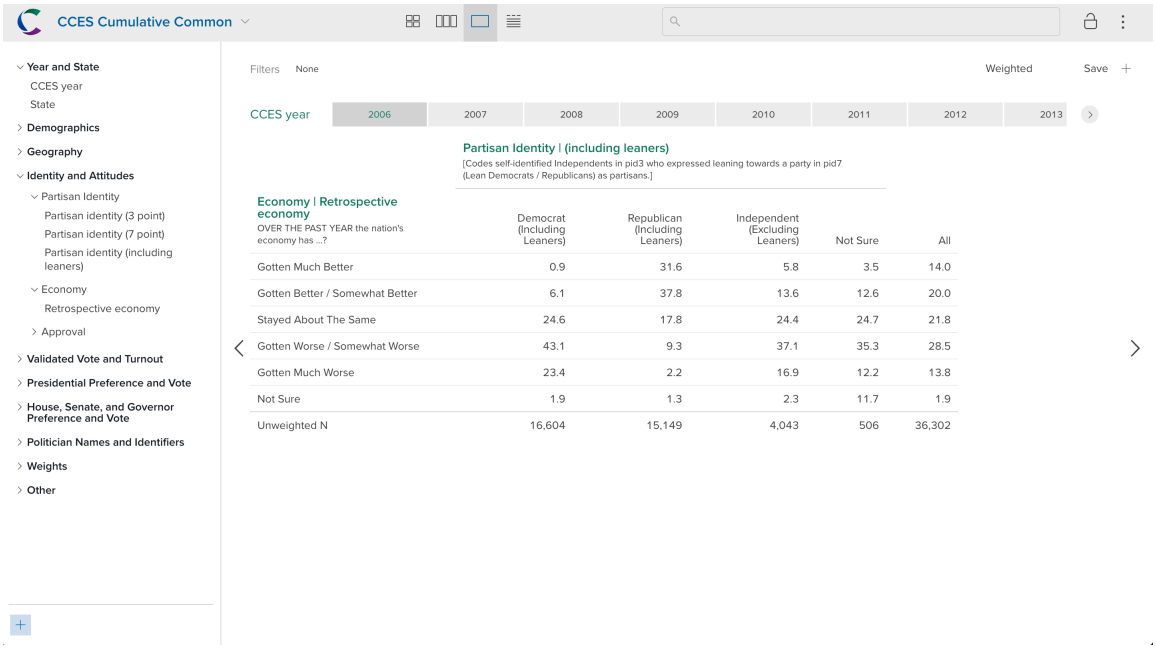
A version of the dataset is also included in Crunch, a database platform that makes it easy to view and analyze survey data either with our without any programming experience.

1. Obtain Access: For View access to the dataset (free), please sign up here. 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`. To learn more about the features, please take a look at their homepage crunch.io or their 5-minute demo video.

Variables

The sections below provide summary statistics and more information on each variable.

- The title shows the name of the variable as it appears in the dataset ("alias" in Crunch terminology), followed by a more descriptive name suitable for presentation ("name" in Crunch terminology).
- Question wordings, where applicable, immediately follow. Otherwise a description is provided in square brackets ([ ]). All square brackets, both in the description and the response options, indicate descriptions that are summaries 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]

	n
2006	36,421
2007	9,999
2008	32,800
2009	13,800
2010	55,400
2011	20,150
2012	54,535
2013	16,400
2014	56,200
2015	14,250
2016	64,600
2017	18,200
2018	60,000
2019	18,000

`starttime`: Start time

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

```
                 Min.               1st Qu.                Median
"2006-10-07 00:02:34" "2010-10-12 08:17:58" "2013-11-23 15:04:56"
                 Mean               3rd Qu.                  Max.
"2013-09-26 03:40:41" "2016-10-12 11:20:19" "2019-12-05 18:34:04"
```

– Years: All of 2006-2019

`tookpost`: Took post-election wave

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

	n
Did Not Take Post-Election Survey	77,064
Took Post-Election Survey	300,892
(Missing)	92,799

– Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2019 (Post-election wave only exists for even years)

Weights

`weight`: Survey weight (Year-Specific)

[weights for pre-election survey of each year]

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000  0.4321  0.7302  1.0000  1.1734 15.0006
```

– Years: All of 2006-2019
– 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.
0.0000  0.2484  0.4686  0.7873  0.9277 19.3959
```

– Years: All of 2006-2019
– Limitations: Only a simple transformation of `weight`. Specifically, `weight_cumulative` is `weight` divided by the year-specific factors shown in the following table. For example, all weights in the 2016 common content are divided by about 2.44, because it has about twice as many observations as the other datasets.

Year	Observations	Factor
2006	36,421	1.38
2007	9,999	0.38
2008	32,800	1.24
2009	13,800	0.52
2010	55,400	2.09
2011	20,150	0.76
2012	54,535	2.06
2013	16,400	0.62
2014	56,200	2.12
2015	14,250	0.54
2016	64,600	2.44
2017	18,200	0.69
2018	60,000	2.27
2019	18,000	0.68

`rvweight`: Survey weights to validated registered voters

[weights to validated registered voter population]

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.0     0.6     0.8     1.0     1.2    15.0  430738
```

- Years: 2018
- In 2018, YouGov computed weights after vote validation to weight to the target population of registered voters. See the methodology section of the 2018 Guide for details. For this reason, and to distinguish it from previous year's post-validation weights, the cumulative renames the 2018 `vvweight` into `rvweight`.
- 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 `rvweight_post`

`rvweight_post`: Survey weights to validated registered voters, post-election wave

[weights to validated registered voter population, post-election wave]

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.0     0.5     0.8     1.0     1.2    15.0  433806
```

- Years: 2018
- Limitations: Only available for some even years.

`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
 0.0     0.4     0.7     1.0     1.1    15.0  321170
```

- 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: 470,755
Variables: 7
$ state       <chr> "California", "Pennsylvania", "Texas", "Texas", "Texas"...
$ st          <chr> "CA", "PA", "TX", "TX", "TX", "NY", "NC", "NC", "MA", "...
$ cd          <chr> "CA-02", "PA-05", "TX-16", "TX-19", "TX-06", "NY-28", "...
$ dist        <int> 2, 5, 16, 19, 6, 28, 11, 7, 1, 17, 15, 1, 2, 6, 1, 1, 1...
$ dist_up     <int> 1, 3, 16, 19, 6, 27, 11, 7, 2, 20, 12, 1, 2, 8, 1, 1, 1...
$ zipcode     <chr> "95969", "16255", "79924", "79423", "76123", "14131", "...
$ county_fips <chr> "06007", "42031", "48141", "48303", "48439", "36063", "...
```

- Years: All of 2006-2019
- Note the distinction between `dist` and `dist_up`, especially in 2012. The former should generally be used for linking respondents to their representatives at the time of the survey, where as the latter can be used for the district in which they will vote for. New districts were drawn in 2010-2012 and candidates ran in new district maps in the 2012 CCES. Because respondents would not be *represented* in the new district lines until January 2013, in the 2012 CCES `dist` is the old district line and `dist_up` is the new district line for the General Election.
- 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 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?”

	n
Male	217,949
Female	252,806

- Years: All of 2006-2019

`birthyr`: Year of birth

“In what year were you born?”

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1900    1950    1961    1963    1978    2001
```

- Years: All of 2006-2019

`age`: Age

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

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
18.0    36.0    51.0    49.6    62.0   109.0
```

- Years: All of 2006-2019

`educ`: Education

“What is the highest level of education you have completed?”

	n
No HS	14,840
High School Graduate	130,123
Some College	116,433
2-Year	45,942
4-Year	106,727
Post-Grad	56,623
(Missing)	67

- Years: All of 2006-2019

`race`: Race

“What racial or ethnic group best describes you?”

	n
White	350,311
Black	51,102
Hispanic	38,156
Asian	9,598
Native American	3,790
Mixed	9,354
Other	7,708
Middle Eastern	736

- Years: All of 2006-2019
- 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]"

	n
Yes	11,399
No	337,449
(Missing)	121,907

- Years: 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
- In years in which this question was fielded, this question supplements the `race` variable by asking those who did *not* respond "Hipsanic" in the `race` question.

`faminc`: Family Income

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

	n
Less than 10k	19,803
10k - 20k	34,469
20k - 30k	47,793
30k - 40k	48,521
40k - 50k	43,368
50k - 60k	42,308
60k - 70k	31,180
70k - 80k	33,792
80k - 100k	39,258
100k - 120k	28,577
120k - 150k	22,930
150k+	26,150
Prefer not to say	51,038
Skipped	12
(Missing)	1,556

- Years: All of 2006-2019

- Limitations: The income brackets provided changed slightly over time. The brackets in this cumulative dataset coarsens certain brackets, losing some granularity. In particular, from 2011-2016, respondents answering "over 150k" were asked a follow-up question to select one of several brackets above 150k. Here, 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?"

	n
Married	259,450
Separated	7,885
Divorced	51,680
Widowed	22,188
Single / Never Married	106,418
Domestic Partnership	21,579
(Missing)	1,555

- Years: All of 2006-2019
- The option "Single" was used till 2016, which was then replaced by "Never Married" in 2017 and 2018.
- The option "Domestic Partnership" was used till 2016, which was then replaced by "Domestic / Civl Partnership" in 2017 and 2018.

`citizen`: Citizenship

[Based on self-report for immigration status]

	n
Citizen	437,930
Non-Citizen	7,045
(Missing)	25,780

- Years: 2006, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2016, 2017, 2018, 2019
- These come from the immigration status questions, which asks respondents between first, second, and third generation citizens, and other foreign-born citizens. Here we mark anyone who does not answer the last category to be a citizen.

Validations

Observations in even years include indicators for validated voting, 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. Validation is often completed in the summer following the election. For more information, see Ansolabehere and Hersh (2012).

`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.]

	n
Active	218,373
No Record Of Registration	77,521
Unregistered	15,869
Dropped	6,607
Inactive	3,565
Multiple Appearances	1,600
(Missing)	147,220

- Years: 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: Collapses some response options

`vv_party_gen`: Validated registered party

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

	n
No Record Of Party Registration	79,041
Unknown	68,895
Democratic Party	37,600
Republican Party	29,494
No Party Affiliation	13,874
Declined To State	2,376
Other	1,635
Independent Party	1,511
Liberatarian Party	537
Green Party	265
Cns	44
Constitution Party	38
Reform Party	11
Wor	9
Socialist Party	5
(Missing)	235,420

- Years: 2012, 2014, 2016, 2018
- Limitations: Not available for some even years

`vv_party_prm`: Validated registered Primary party

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

	n
No Record Of Party Registration	208,027
Republican Party	14,486
Democratic Party	12,783
No Party Affiliation	16
Liberatarian Party	11
Other	8
Green Party	4
(Missing)	235,420

- Years: 2012, 2014, 2016, 2018
- 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.]

	n
Voted	202,966
No Record Of Voting	155,257
No Voter File	1,733
(Missing)	110,799

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: Collapses most response options. For example, 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. Congressional primaries.]

	n
No Record Of Voting	225,737
Voted	96,435
No Voter File	1,363
(Missing)	147,220

- Years: 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: See `vv_turnout_gvm`

Partisan Identity

`pid3`: Partisan identity (3 point)

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

	n
Democrat	167,172
Republican	123,840
Independent	131,094
Other	18,765
Not Sure	20,930
(Missing)	8,954

- Years: All of 2006-2019
- 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]

	n
Strong Democrat	112,344
Not Very Strong Democrat	56,761
Lean Democrat	47,170
Independent	63,612
Lean Republican	49,281
Not Very Strong Republican	45,343
Strong Republican	79,181
Not Sure	14,047
(Missing)	3,016

- Years: All of 2006-2019
- 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.]

	n
Democrat (Including Leaners)	216,275
Republican (Including Leaners)	173,805
Independent (Excluding Leaners)	63,612
Not Sure	14,047
(Missing)	3,016

- Years: All of 2006-2019

– Limitations: See `pid3`

`ideo5`: Ideology (5 point)

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

	n
Very Liberal	42,728
Liberal	82,662
Moderate	146,842
Conservative	108,598
Very Conservative	54,765
Not Sure	33,386
(Missing)	1,774

– Years: All of 2006-2019

Economy

`economy_retro`: Retrospective economy

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

	n
Gotten Much Better	31,757
Gotten Better / Somewhat Better	104,958
Stayed About The Same	126,496
Gotten Worse / Somewhat Worse	117,032
Gotten Much Worse	78,787
Not Sure	10,837
(Missing)	888

– Years: All of 2006-2019
– 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 .."

	n
Most Of The Time	234,647
Some Of The Time	109,997
Only Now And Then	52,172
Hardly At All	25,855
Don't Know	11,077
(Missing)	37,007

- Years: 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
- Limitations: Not asked in 2006. Similar questions about watching TV news was asked in 2006, but not included in this cumulative file.

Approval

`approval_pres`: President approval

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

	n
Strongly Approve	97,838
Approve / Somewhat Approve	107,126
Disapprove / Somewhat Disapprove	48,372
Strongly Disapprove	203,121
Never Heard / Not Sure	13,057
Neither Approve Nor Disapprove	443
(Missing)	798

- Years: All of 2006-2019
- 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]"

	n
Strongly Approve	67,519
Approve / Somewhat Approve	148,751
Disapprove / Somewhat Disapprove	82,834
Strongly Disapprove	73,328
Never Heard / Not Sure	89,022
Neither Approve Nor Disapprove	1,798
(Missing)	7,503

- Years: All of 2006-2019
- 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]"

	n
Strongly Approve	60,977
Approve / Somewhat Approve	150,506
Disapprove / Somewhat Disapprove	94,700
Strongly Disapprove	92,858
Never Heard / Not Sure	65,717
Neither Approve Nor Disapprove	1,413
(Missing)	4,584

- Years: All of 2006-2019
- 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]"

	n
Strongly Approve	65,698
Approve / Somewhat Approve	145,164
Disapprove / Somewhat Disapprove	92,000
Strongly Disapprove	93,440
Never Heard / Not Sure	67,946
Neither Approve Nor Disapprove	1,158
(Missing)	5,349

- See `approval_sen2`

approval_gov: Governor approval

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

	n
Strongly Approve	70,572
Approve / Somewhat Approve	145,296
Disapprove / Somewhat Disapprove	87,011
Strongly Disapprove	121,724
Never Heard / Not Sure	42,490
Neither Approve Nor Disapprove	1,414
(Missing)	2,248

- Years: All of 2006-2019

– Limitations: See approval_pres
– To see who the Governor refers to for a particular respondent, see gov_inc.

Presidential Vote

A note on the terms "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. In 2018, it also coalesces the responses to the straight ticket party option (`CC18_409`), so that those who selected the Republican straight party ticket in the applicable states will appear to have voted for the Republican candidate in all offices. The straight ticket party option was not asked in other years.

`intent_pres_08`: 2008 President preference (before voting)

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

	n
John Mccain	13,322
Barack Obama	12,897
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	1,697
(Missing)	440,560

- Years: 2008

`intent_pres_12`: 2012 President preference (before voting)

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

	n
Mitt Romney (Republican)	20,738
Barack Obama (Democratic)	24,401
Other	1,781
I Will Not Vote In This Election	1,467
I'm Not Sure	3,856
(Missing)	418,512

- Years: 2012

`intent_pres_16`: 2016 President preference (before voting)

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

	n
Donald Trump (Republican)	19,227
Hillary Clinton (Democrat)	27,502
Gary Johnson (Libertarian)	3,145
Jill Stein (Green)	1,400
Other	1,880
I Won't Vote In This Election	3,312
I'm Not Sure	6,536
(Missing)	407,753

- 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]"

	n
Barack Obama	73,986
John McCain	68,398
Other / Someone Else	4,204
Did Not Vote	18,227
Not Sure / Don't Recall	1,787
(Missing)	304,153

- 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]"

	n
Barack Obama	82,681
Mitt Romney	64,956
Other / Someone Else	5,890
Did Not Vote	2,839
Not Sure / Don't Recall	1,990
(Missing)	312,399

- Years: 2012, 2013, 2014, 2015, 2016
- Limitations: Response options offer slightly by year; some are collapsed into one.

`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]"

	n
Hilary Clinton	57,694
Donald Trump	49,751
Other / Someone Else	11,478
Did Not Vote	5,026
Not Sure / Don't Recall	527
(Missing)	346,279

- Years: 2016, 2017, 2018, 2019
- 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 Vote

`intent_rep`: House preference (before voting)

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

	n
[Democrat / Candidate 1]	128,231
[Republican / Candidate 2]	115,292
[Other / Candidate 3]	4,401
$HouseCand4Name ($HouseCand4Party)	37
Other	2,259
I'm Not Sure	70,460
No One	19,235
$HouseCand5Name ($HouseCand5Party)	23
I Won't Vote In This Election	2,269
$HouseCand6Name ($HouseCand6Party)	41
$HouseCand7Name ($HouseCand7Party)	20
$HouseCand8Name ($HouseCand8Party)	14
$HouseCand9Name ($HouseCand9Party)	1
$HouseCand10Name ($HouseCand10Party)	1
$HouseCand11Name ($HouseCand11Party)	3
(Missing)	128,468

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: Only available for even years. The third party candidate is not specified for early years. The fourth candidate and below are 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` to see the party affiliation of the chosen candidate.
- Note that for each respondent, a name (and party affiliation) is shown in place of the square bracket values. To see the name of 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?"

	n
[Democrat / Candidate 1]	97,220
[Republican / Candidate 2]	82,433
[Other / Candidate 3]	4,477
$SenCand4Name ($SenCand4Party)	19
Other	1,713
I'm Not Sure	38,112
No One	12,419
I Won't Vote In This Election	1,145
(Missing)	233,217

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: See `intent_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.
- See `intent_sen_party` for the party affiliation of the chosen candidate.

`intent_gov`: Governor preference (before voting)

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

	n
[Democrat / Candidate 1]	74,561
[Republican / Candidate 2]	66,292
[Other / Candidate 3]	4,055
Other	1,390
I'm Not Sure	24,296
No One	7,991
I Won't Vote In This Election	466
(Missing)	291,704

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: See `intent_rep`. For governor elections in odd years, see individual year's CCES.
- See `intent_gov_party` for the party affiliation of the chosen candidate.

`voted_rep`: House vote choice (after voting)

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

	n
[Democrat / Candidate 1]	117,581
[Republican / Candidate 2]	111,255
[Other / Candidate 3]	2,786
$HouseCand4Name ($HouseCand4Party)	27
Other	3,120
I Did Not Vote In This Race	12,535
$HouseCand5Name ($HouseCand5Party)	24
Not Sure	4,493
$HouseCand6Name ($HouseCand6Party)	39
$HouseCand7Name ($HouseCand7Party)	15
$HouseCand8Name ($HouseCand8Party)	16
$HouseCand9Name ($HouseCand9Party)	2
$HouseCand10Name ($HouseCand10Party)	2
$HouseCand11Name ($HouseCand11Party)	3
(Missing)	218,857

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- This variable coalesces two variables in the CCES for years 2012 and onward: 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
- See `voted_rep_party` for party affiliation.

`voted_sen`: Senate vote choice (after voting)

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

	n
[Democrat / Candidate 1]	86,668
[Republican / Candidate 2]	77,569
[Other / Candidate 3]	2,974
Other	1,967
Not Sure	2,094
$SenCand4Name ($SenCand4Party)	11
I Did Not Vote In This Race	4,789
(Missing)	294,683

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- This variable coalesces two variables in the CCES for years 2012 and onward: 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.
- See `voted_sen_party` for party affiliation.
- Senate Special elections where both senate seats are up for election is often recorded as different columns in the year-specific CCES, but these are not collected in the cumulative.

`voted_gov`: Governor vote choice (after voting)

"For whom did you vote for Governor?"

	n
[Democrat / Candidate 1]	68,445
[Republican / Candidate 2]	63,434
[Other / Candidate 3]	2,800
Other	1,817
I Did Not Vote In This Race	10,116
Not Sure	1,091
(Missing)	323,052

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- This variable coalesces two variables in the CCES for years 2012 and onward: 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.
- See `voted_gov_party` for party affiliation.

Metadata and Identifiers

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: 470,755
Variables: 2
$ year    <int> 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006,...
$ case_id <int> 439219, 439224, 439228, 439237, 439238, 439242, 439251, 439...
```

Current Representatives' Name and Party

The four names in the three offices are representatives of the respondent *at the time of the survey*. Names are printed as shown, and similarly parties are shown if the particular year's CCES did not show party. For example, Senator Shelby is presented as `Richard Craig Shelby`, `Richard C. Shelby (R)`, `Richard Shelby (R)`, `Richard C. Shelby (R)`, depending on the year. Party names are abbreviated down to initials (`D` for Democrat, `R` for Republican, `I` for Independent) in this dataset.

```
Observations: 470,755
Variables: 4
$ rep_current  <chr> "Patrick T. McHenry (R)", "Michael R. Turner (R)", "Ro...
$ sen1_current <chr> "Elizabeth Dole (R)", "Mike DeWine (R)", "Robert Menen...
$ sen2_current <chr> "Richard Burr (R)", "George V. Voinovich (R)", "Frank ...
$ gov_current  <chr> "Michael Easley (D)", "Bob Taft (R)", "Jon Corzine (D)...
```

ICPSR Identifiers

Unique identifiers (ICPSR / Nominate for Congress, FEC for Governor) for the current representatives. Identifiers are not part of the individual year's CCES. Instead, I attempt to merge in these identifiers through a series of name and district merges.

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

```
Observations: 470,755
Variables: 3
$ rep_icpsr  <dbl> 20522, 20342, 29132, 29911, 29380, 20531, 29126, 29739, ...
$ sen1_icpsr <dbl> 40303, 15020, 29373, 15021, 14858, 49306, 40101, 15054, ...
$ sen2_icpsr <dbl> 29548, 49903, 14914, 40502, 40105, 40305, 40302, 29537, ...
```

- Years: All of 2006-2019

- Limitations: Please note there may be some incorrect merges, especially for nontraditional names and representatives who were elected in special elections and may not be in some datasets.

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/

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 this particular respondent preferred voting for Maxine Waters, one of the two Democrats in the race.

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

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

The name and party are those as provided in the CCES datasets (e.g. in the form `HouseCand1Name`).

Name of Chosen Candidate

```
Observations: 470,755
Variables: 6
$ intent_rep_chosen <chr> "Richard C. Carsner (D)", "Stephanie Studebaker (...
$ intent_sen_chosen <chr> NA, "Sherrod C. Brown (D)", "Robert Menendez (D)"...
$ intent_gov_chosen <chr> NA, "Ted Strickland (D)", NA, "Rod Blagojevich (D...
$ voted_rep_chosen  <chr> "Richard C. Carsner (D)", "Stephanie Studebaker (...
$ voted_sen_chosen  <chr> NA, "Sherrod C. Brown (D)", "Robert Menendez (D)"...
$ voted_gov_chosen  <chr> NA, "Ted Strickland (D)", NA, "Rod Blagojevich (D...
```

- 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)

Party of Chosen Candidate

```
Observations: 470,755
Variables: 8
$ intent_pres_party <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ voted_pres_party  <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ intent_rep_party  <fct> Democratic, Democratic, Democratic, Democratic, D...
$ voted_rep_party   <fct> Democratic, Democratic, Democratic, Democratic, D...
$ intent_gov_party  <fct> NA, Democratic, NA, Democratic, Democratic, NA, R...
$ voted_gov_party   <fct> NA, Democratic, NA, Democratic, Democratic, NA, R...
$ intent_sen_party  <fct> NA, Democratic, Democratic, NA, NA, NA, Republica...
$ voted_sen_party   <fct> NA, Democratic, Democratic, NA, Democratic, NA, R...
```

- Years: varies by office
- Early years may mislabel the candidate's party, especially when the two candidates are of the same party (as in top-two primary states)

Version History of Dataverse Releases

Dataverse assigns version numbers by incremeting a full number if any of the datasets change, and an incrementing decimal when the description changes.

Version 5.0

- Released `2020-XX-XX`
- Enter 2019 common content (up to n = 470,755)

Version 4.0

- Released `2019-09-09`
- Enter 2018 vote validation
- Coalesce straight party ticket vote into vote choice entries
- Remove FEC identifiers

Version 3.0

- Released `2019-04-29`
- Add 2018 Common Content before vote validation (up to n = 452,755)

Version 2.0

- Released `2018-04-16`
- Add 2017 Common Content (up to n = 392,755)
- Corrects 2016 validated vote entries inherited from Common Content.
- Consolidates weights to a single column, using post-vote validation weights for even years.
- Adds `hispanic` and `faminc` variables

Version 1.0

- Released `2018-01-24`
- First upload, covering 2006 - 2016 (n = 374,556)