Guide to the Cumulative Common Content of the Cooperative (Congressional) Election Study

Shiro Kuriwaki*

2025-10-08

Cite this dataset as:

> Kuriwaki, Shiro, 2025, "Cumulative CES Common Content", doi:10.7910/DVN/II2DB6, Harvard Dataverse, V11.

This dataset combines 19 years (2006 – 2024) of the Cooperative Congressional Election Study (CCES), renamed the Cooperative Congressional Election Study (CES) from 2020. The CCES/CES is an online survey conducted around November of each year, asking a range of questions on political behavior and public opinion. Its current principal investigators are Brian Schaffner, Jeremy Pope, and Marissa Shih.

Each year's CCES/CES asks hundreds of questions, many of which change from year to year. This cumulative file only includes a subset of those questions that are standard and important. It standardizes its values across years and creates a few new variables too. Users can still merge in their year-specific questions of interest easily into this cumulative file and take advantage of its standardized variables.

I constructed this dataset from each year's full CCES/CES, all of them publicly available as separate datasets on Dataverse.The final product is a `tibble`-style data frame built in R that is also available as a Stata `dta` file and a `feather` file.

Please note that this cumulative dataset makes some modifications to the original CCES/CES datasets to maintain comparability across years. These modifications are only made when differences are deemed sufficiently minor. Still, 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/CES datasets,** search the CES dataverse or access the CES 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 CES homepage or the methodology section of a recent Common Content's codebook.

*Department of Political Science, Yale University. Website: https://www.shirokuriwaki.com. ORCID: https://orcid.org/0000-0002-5687-2647. My thanks to Alexander Agadjanian, Jeremiah Cha, Steve Ansolabehere, Valerie Bradley, Stephen DiMauro, Bernard Fraga, Nathan Kaplan, Silvia Kim, Mayya Komisarchik, Stephen Pettigrew, Boris Shor, Brian Schaffner, Caroline Soler, and Gerlad Wright for finding errors and providing suggestions. Thanks to Joe Williams at YouGov, and Jon Keane, Mike Malecki, and Gordon Shotwell at Crunch for their help.

Contents

Getting Started . . . 5
Data Read-in . . . 5
Data Download . . . 5
Unique identifiers and how to add more variables . . . 6
Labelled variables (for analysis in R) . . . 7
Features of the Cumulative Dataset . . . 9
Unified Variable Names . . . 9
Chosen Candidate Names and Identifiers . . . 9
Variables . . . 9
Administration . . . 10
`year`: CCES year . . . 10
`starttime`: Start time . . . 10
`tookpost`: Took post-election wave . . . 10
Weights . . . 11
`weight`: Survey weight (Year-Specific) . . . 11
`weight_cumulative`: Survey weight (Cumulative) . . . 11
`weight_post`: Survey weight for post-election wave . . . 12
`vvweight`: Survey weights to validated registered voters . . . 13
`vvweight_post`: Survey weights to validated registered voters, post-election wave . . . 13
Geography . . . 13
Demographics . . . 14
`gender`: Sex (standardized) . . . 14
`gender4`: Gender . . . 15
`sex`: Sex . . . 15
`sexuality`: Sexual Orientation . . . 15
`birthyr`: Year of birth . . . 16
`age`: Age . . . 16
`educ`: Education . . . 16
`race`: Race . . . 16
`hispanic`: Hispanic . . . 17
`race_h`: Race (any-part Hispanic) . . . 17
`hisp_origin`: Hispanic origin . . . 17
`citizen`: Citizenship . . . 18
`religion`: Religion . . . 18
`relig_imp`: Importance of religion . . . 19
`relig_bornagain`: Evangelical Christian . . . 19
`relig_protestant`: Branch of Protestantism . . . 19
`relig_church`: Church Attendance . . . 20
Family Status . . . 20
`marstat`: Marital Status . . . 20
`ownhome`: Home Ownership . . . 21
`has_child`: Parent of Young Children . . . 21
`no_milstat`: Military Status (None) . . . 21
Validated Turnout and Registration . . . 22
`vv_regstatus`: Validated registration status . . . 22
`vv_party_gen`: Validated registered party . . . 22
`vv_party_prm`: Validated registered Primary party . . . 23
`vv_state`: Validated state of registration . . . 23
Turnout . . . 25

vv_turnout_gvm: Validated turnout General Election . . . 25
vv_turnout_pvm: Validated turnout Primary Election (Congressional) . . . 25
intent_turnout_self: Self-reported turnout (pre-election wave) . . . 25
voted_turnout_self: Self-reported turnout (post-election wave) . . . 26
Partisan Identity . . . 26
pid3: Partisan identity (3 point) . . . 26
pid7: Partisan identity (7 point) . . . 26
pid3_leaner: Partisan identity (including leaners) . . . 27
ideo5: Ideology (5 point) . . . 27
Economics and Income . . . 27
faminc: Family Income . . . 27
employ: Employment Status . . . 28
no_healthins: Uninsured . . . 28
union: Union membership . . . 29
union_hh: Union membership in household . . . 29
investor: Investor . . . 30
economy_retro: Retrospective economy . . . 30
News Interest . . . 30
newsint: News Interest . . . 30
Approval . . . 31
approval_pres: President approval . . . 31
approval_rep: House Representative approval . . . 31
approval_sen1: Senator 1 approval . . . 31
approval_sen2: Senator 2 approval . . . 32
approval_gov: Governor approval . . . 32
Vote Choice Variables . . . 33
Presidential Vote . . . 33
intent_pres_party: President preference party . . . 33
voted_pres_party: President vote in last election . . . 33
intent_pres_08: 2008 President preference (before voting) . . . 34
intent_pres_12: 2012 President preference (before voting) . . . 34
intent_pres_16: 2016 President preference (before voting) . . . 35
intent_pres_20: 2020 President preference (before voting) . . . 35
intent_pres_24: 2024 President preference (before voting) . . . 35
voted_pres_08: 2008 President vote choice (after voting) . . . 36
voted_pres_12: 2012 President vote choice (after voting) . . . 36
voted_pres_16: 2016 President vote choice (after voting) . . . 37
voted_pres_20: 2020 President vote choice (after voting) . . . 37
voted_pres_24: 2024 President vote choice (after voting) . . . 37
House, Senate and Governor Vote . . . 38
intent_rep: House preference (before voting) . . . 38
intent_sen: Senate preference (before voting) . . . 39
intent_gov: Governor preference (before voting) . . . 39
voted_rep: House vote choice (after voting) . . . 39
voted_sen: Senate vote choice (after voting) . . . 40
voted_gov: Governor vote choice (after voting) . . . 40
voted_rep_party: House vote choice party . . . 41
voted_sen_party: Senate vote choice party . . . 42
voted_gov_party: Governor vote choice party . . . 43
Metadata and Identifiers . . . 44
Identifiers . . . 44

Current Representatives' Name and Party 44
ICPSR Identifiers 44
Name of Chosen Candidate 45
Version History of Dataverse Releases 46
Version 11.0 (released 2025-10-08) 46
Version 10.0 (released 2025-05-21) 46
Version 9.0 (released 2024-06-17) 46
Version 8.0 (released 2023-05-12) 46
Version 7.0 (released 2022-03-24) 46
Version 6.0 (released 2021-04-06) 47
Version 5.0 (released 2020-10-04) 47
Version 4.0 (released 2019-09-09) 47
Version 3.0 (released 2019-04-29) 47
Version 2.0 (released 2018-04-16) 47
Version 1.0 (released 2018-01-24) 47

Getting Started

Data Read-in

The dataset in R is best viewed with the tidyverse suite of packages, as well as the package for converting and reading Stata files, `haven`. Although we provide three file formats of the same content, we recommend using the Stata (`.dta`) file along with the `haven` package's `read_dta` function most of the time, and using the `.feather` file when speed is a priority.

```
library(tidyverse)
library(haven)
cc <- read_dta("cumulative_2006-2024.dta")
```

Loading the `haven` package allows using labelled variables in versatile ways (see the section "Labelled variables"). Plain-text formats are less useful because they do not preserve value labels.

We also provide a `.rds` format that is specific to R:

```
cc <- read_rds("cumulative_2006-2024.rds")
```

This format preserves dataset properties such as the distinction between integers and doubles, as well as the labelled variables. Between the `.rds` and `.dta` versions, there is some difference in whether variables are saved as characters (rds) or labelled variables (dta), but the content is the same.

From version 8.0 and onwards, we also include a `.feather` format. This is a format that is optimized for speed: it reads and writes faster than the others. Its content is identical to the `.rds` version. For example, in R, use the function `read_feather` from the `arrow` package.

```
library(arrow)
cc <- read_feather("cumulative_2006-2024.feather")
```

In one example, I found that R read the `.feather` file in 0.3 seconds, and it took the `.rds` format more than 10 times slower (3.5 seconds) and the `.dta` format to be an additional 5 times slower than the `.rds` fomrat (16 seconds).

Data Download

Downloading the data via the dataverse package. In some cases, it may be convenient to download the dataset directly into an R environment without downloading the file to one's computer. The recent version of `dataverse` (version 0.3.0 or later) allows this by the function:

```
library(dataverse)
cc <- get_dataframe_by_name(
  filename = "cumulative_2006-2024.dta",
  dataset = "10.7910/DVN/II2DB6",
  original = TRUE,
  .f = haven::read_dta,
  server = "dataverse.harvard.edu"
)
```

To download the `arrow` dataset, replace the `filename` argument with `cumulative_2006-2024.feather` and the `.f` argument with `arrow::read_feather`.

Downloading datasets that are hundreds of megabytes large can take multiple minutes each time. The total time it takes to download is a combination of the (1) download time, and (2) time for reading the downloaded file into R. As of 2025, R users can skip the step 1 after the first time by asking the dataverse package to save a cache of the data to the computer.

To do this:

- Make sure you are using the `dataverse` package that is version 0.3.15 or above (releasd on CRAN in May 2025)
- Specify the version of the dataset you want to download using the `version` argument of `get_dataframe_by_name`.

For example, to download version 10 of the feather verison of the dataset, the R code is the following:

```
library(dataverse)
cc <- get_dataframe_by_name(
  filename = "cumulative_2006-2024.feather",
  version = "11",
  dataset = "10.7910/DVN/II2DB6",
  original = TRUE,
  .f = arrow::read_feather,
  server = "dataverse.harvard.edu"
)
```

Unique identifiers and how to add more variables

The cumulative dataset only uses key variables from each year's common content. But users can still merge in other common content variables, or variables from other CCES datasets like the policy preferences dataset[1].

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. For example, suppose we have separately downloaded the 2016 Common Content and read it in as follows:

```
cc16 <- read_dta("CCES16_Common_OUTPUT_Feb2018_VV.dta")
```

Suppose we want to merge in the 2016-specific issue questions that ask respondent's views about key votes in Congress. This variable all start with `"CC16_351"` and the case-identifier is called `V101`, so we can merge this into the cumulative file as follows:

```
# subset
cc16_rc <- select(cc16, V101, matches("CC16_351"))

# join on case ID
cc_rc <- cc |>
  filter(year == 2016) |>
  left_join(cc16_rc, by = c("case_id" = "V101"))
```

[1] Dagonel, Angelo, 2021, "Cumulative CCES Policy Preferences", doi:10.7910/DVN/OSXDQO, Harvard Dataverse.

Labelled variables (for analysis in R)

A note on variable types. The R dataset stores variables in `numeric`, `character`, `factor`, or `labelled` class.[2] 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: 701,955 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]
# i 701,945 more rows
```

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: 701,955 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
# i 701,950 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 class allows you to do that.

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

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

```
# A tibble: 257,527 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]
# i 257,522 more rows
```

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

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/CES. 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/CES 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: 701,955 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 / Candidate 1] Democratic      Sherrod C. Brow~
 3  2006  439228 NJ    [Democrat / Candidate 1] Democratic      Robert Menendez~
 4  2006  439237 IL    <NA>                      <NA>            <NA>
 5  2006  439238 NY    [Democrat / Candidate 1] Democratic      Hillary Rodham ~
 6  2006  439242 TX    I Did Not Vote In This ~ <NA>            <NA>
 7  2006  439251 MN    [Republican / Candidate~ Republican      Mark Kennedy (R)
 8  2006  439254 NV    [Democrat / Candidate 1] Democratic      Jack Carter (D)
 9  2006  439255 TX    [Democrat / Candidate 1] Democratic      Barbara Ann Rad~
10  2006  439263 MD    I Did Not Vote In This ~ <NA>            <NA>
# i 701,945 more rows
```

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) follows. 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 a combination 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
2020	61,000
2021	25,700
2022	60,000
2023	24,500
2024	60,000

`starttime`: Start time

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

- Years: All of 2006-2024
- Times are recorded in UTC (Greenwich mean time) for all respondents regardless of the location of the interview, at least for 2021.

`tookpost`: Took post-election wave

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

	Earliest Date	Latest Date
2006	2006-10-07	2006-11-08
2007	2007-11-09	2007-12-10
2008	2008-10-08	2008-11-03
2009	2009-11-05	2009-12-14
2010	2010-10-01	2010-11-01
2011	2011-11-09	2012-01-07
2012	2012-10-01	2012-11-05
2013	2013-11-06	2013-12-03
2014	2014-10-01	2014-11-03
2015	2015-11-06	2015-12-03
2016	2016-09-28	2016-11-07
2017	2017-11-08	2017-12-12
2018	2018-09-27	2018-11-05
2019	2019-11-06	2019-12-05
2020	2020-09-29	2020-11-02
2021	2021-11-03	2021-12-07
2022	2022-09-29	2022-11-08
2023	2023-11-08	2023-12-11
2024	2024-10-01	2024-11-04

	n
Did Not Take Post-Election Survey	88,100
Took Post-Election Survey	452,856
(Missing)	160,999

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024 (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  0.4340  0.7302  1.0000  1.1615 15.0614
```

- Years: All of 2006-2024
- 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 data representative of the national adult population**. See the methodology section of the 2016 Guide and the FAQ on the CCES website 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  0.2985  0.5406  0.8878  1.0312 24.0297
```

- Years: All of 2006-2024
- 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 1.97, because it has about twice as many observations as the other datasets.
- Therefore, this cumulative weight simply *re-weights all years to have the same sample size*. This means that, if both odd years and even years are included in the weighted average, every even-year observation is down-weighted by a factor of about 2. For analyses that are done year by year, using `weight` and `weight_cumulative` are equivalent.

Year	Observations	Factor
2006	36,421	1.11
2007	9,999	0.30
2008	32,800	1.00
2009	13,800	0.42
2010	55,400	1.69
2011	20,150	0.61
2012	54,535	1.66
2013	16,400	0.50
2014	56,200	1.71
2015	14,250	0.43
2016	64,600	1.97
2017	18,200	0.55
2018	60,000	1.83
2019	18,000	0.55
2020	61,000	1.86
2021	25,700	0.78
2022	60,000	1.83
2023	24,500	0.75
2024	60,000	1.83

`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    0.69    1.00    1.07   15.50  449839
```

- Years: 2012, 2016, 2018, 2020, 2022
- Limitations: Only available for some even years.
- To analyze the post-election wave responses for years that do not have a post-specific weight, the second-best option is to use the normal `weight` variable. These weights will not be designed for the post-wave subset but they tend to correlate.

`vvweight`: Survey weights to validated registered voters

[weights to validated registered voter population]

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
0.00    0.53    0.79    1.00    1.16   16.31  537396
```

- Years: 2018, 2020, 2022, 2024
- Note these weight are missing for anyone not validated as a actively registered voters (`vv_regstatus`). Not all actively registered voters turn out (`vv_turnout_gvm`). Starting in 2018, YouGov computed weights after vote validation and weighted to the target population of *validated* registered voters. See the methodology section of the 2018 Guide or the 2020, 2022, 2024 guide for details. Until Dataverse V10 of this dataset, the cumulative called this variable `rvweight`, but from V11 onwards, we switched to calling it `vvweight`.
- 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 `vvweight_post`

`vvweight_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
0.00    0.50    0.75    1.00    1.14   15.18  551380
```

- Years: 2018, 2020, 2022, 2024
- Limitations: Only available for some even years.

Geography

A series of variables for the respondent's location:

- `state`: State (FIPS): [State]
- `state_post`: State (FIPS), post-election: [State, post-election]
- `st`: State abbreviation (FIPS): [State Abbreviation]
- `st_post`: State abbreviation (FIPS), post-election: [State, post-election]
- `dist`: Congressional district number in current Congress: [Current Congressional District Number]
- `dist_post`: Congressional district number in current Congress, post-election: [Current Congressional District Number, post-election]
- `dist_up`: Congressional district number for upcoming Congress: [Upcoming Congressional District Number]
- `dist_up_post`: Congressional district number for upcoming Congress, post-election: [Upcoming Congressional District Number, post-election]
- `cd`: Congressional district in current Congress: [Current Congressional District]
- `cd_post`: Congressional district in current Congress, post-election: [Current Congressional District, post-election]
- `cd_up`: Congressional district in upcoming Congress: [Upcoming Congressional District]
- `cd_up_post`: Congressional district in upcoming Congress, post-election: [Upcoming Congressional District, post-election]
- `zipcode`: Zipcode (lookupzip): [lookupzip in most years.] 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)]

```
Rows: 701,955
Columns: 15
$ state        <chr> "California", "Pennsylvania", "Texas", "Texas", "Texas", ~
$ st           <chr> "CA", "PA", "TX", "TX", "TX", "NY", "NC", "NC", "MA", "CA~
$ state_post   <chr> NA, "Pennsylvania", NA, "Texas", "Texas", "New York", NA,~
$ st_post      <chr> NA, "PA", NA, "TX", "TX", "NY", NA, NA, "MA", NA, "MI", "~
$ dist         <int> 2, 5, 16, 19, 6, 28, 11, 7, 1, 17, 15, 1, 2, 6, 1, 1, 16,~
$ dist_up      <int> 1, 3, 16, 19, 6, 27, 11, 7, 2, 20, 12, 1, 2, 8, 1, 1, 15,~
$ cd           <chr> "CA-02", "PA-05", "TX-16", "TX-19", "TX-06", "NY-28", "NC~
$ cd_up        <chr> "CA-01", "PA-03", "TX-16", "TX-19", "TX-06", "NY-27", "NC~
$ dist_post    <int> NA, 5, NA, 19, 6, 28, NA, NA, 1, NA, 15, 1, 2, NA, NA, NA~
$ dist_up_post <int> NA, 3, NA, 19, 6, 27, NA, NA, 2, NA, 12, 1, 2, NA, NA, NA~
$ cd_post      <chr> NA, "PA-05", NA, "TX-19", "TX-06", "NY-28", NA, NA, "MA-0~
$ cd_up_post   <chr> NA, "PA-03", NA, "TX-19", "TX-06", "NY-27", NA, NA, "MA-0~
$ zipcode      <chr> "95969", "16255", "79924", "79423", "76123", "14131", "28~
$ county_fips  <chr> "06007", "42031", "48141", "48303", "48439", "36063", "37~
$ vv_state     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```

- Years: All of 2006-2024
- 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, whereas 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.
- `zipcode` mostly relies on the variable often called `lookupzip` in each year's CCES. This is the zipcode of voter registration, or if not available, the residential zipcode, of the respondent. It is called lookup because it is used to look up the congressional district and other geographies of the respondent. For more information on zipcodes, see the CCES question.
- 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`: Sex (standardized)

"Are you. . . ? <1> Male <2> Female [2018-2020] Are you male or female? [2006-2016]"

	n
Male	320,535
Female	379,802
(Missing)	1,618

- Years: All of 2006-2024
- Although named `gender` historically, this question's wording is such that it asks people's sex. Until 2020, only two choices were allowed. From 2021, the question was discontinued in favor of the question now called `gender4` (see below).

- The cumulative file has, as of V8, kept the `gender` variable and relabelled it as a "standardized" binary variable for all years by making the following custom recodes for 2021 onwards: `Man` in `gender4` is recoded to `Male`, `Woman` in `gender4` is recoded to `Female`, and other categories are coded as missing.
- The cumulative file has created a variable `sex` to more accurately capture what the `gender` question has asked 2006–2020 (see below).
- Some years ask respondents if they identify as transsexual. These are currently not included in the cumulative common content.

`gender4`: Gender

"What is your gender?"

	n
Man	76,795
Woman	91,787
Non-Binary	1,290
Other	328
(Missing)	531,755

- Years: 2021, 2022, 2023, 2024
- See the explanation under `gender`

`sex`: Sex

"Are you male or female? [2006-2016]"

	n
Male	243,740
Female	288,015
(Missing)	170,200

- Years: All of 2006-2020
- See the explanation under `gender`

`sexuality`: Sexual Orientation

"Which of the following describes your sexuality?"

	n
Heterosexual / Straight	331,854
Lesbian / Gay Woman	5,417
Gay Man	11,113
Bisexual	18,348
Other	5,248
Prefer not to Say	8,797
(Missing)	321,178

- Years: 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024

`birthyr`: Year of birth

"In what year were you born?"

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1900    1953    1964    1966    1981    2006
```

- Years: All of 2006-2024

`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    52.0    49.8    63.0   109.0
```

- Years: All of 2006-2024

`educ`: Education

"What is the highest level of education you have completed?"

	n
No HS	23,795
High School Graduate	193,235
Some College	167,980
2-Year	71,036
4-Year	158,918
Post-Grad	86,924
(Missing)	67

- Years: All of 2006-2024

`race`: Race

"What racial or ethnic group best describes you?"

	n
White	510,016
Black	80,423
Hispanic	60,390
Asian	17,029
Native American	5,791
Mixed	15,369
Other	11,710
Middle Eastern	1,227

- Years: All of 2006-2024
- Limitations: The "Hispanic" value may undercount self-identified Hispanics. See `hispanic` and `race_h`.

`hispanic`: Hispanic

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

	n
Yes	30,378
No	537,700
(Missing)	133,877

- Years: 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024
- Different years have treated this question differently. Until 2017, this question was *not* asked to those who responded "Hispanic" in the `race` question. From 2018-2022, this question was asked to all respondents. From 2024, the value of these questions for all those who said they were Hispanic appear to have been set to Yes. For a combined version of `race` and `hispanic`, see `race_h` (any-part Hispanic), which is more stable version.
- See `hisp_origin` for more details on hispanic origin

`race_h`: Race (any-part Hispanic)

[race (What racial or ethnic group best describes you?) combined with hispanic ethnicity]

	n
White	498,544
Black	78,624
Hispanic	80,356
Asian	16,593
Native American	5,226
Mixed	10,777
Other	10,674
Middle Eastern	1,161

- Years: All of 2006-2024
- This variable combines the `race` and `hispanic` variables in a single variable by coding "any-part Hispanic" as Hispanic. For example, White Hispanics and Black Hispanics are coded as "Hispanic" in this definition and "White" voters consist of non-Hispanic Whites only. If a respondent identified as Hispanic in *either* the `race` or `hispanic` question, they get coded as "Hispanic" in `race_h`.
- For years that do not have a `hispanic` question, this variable falls back to the `race` values.

`hisp_origin`: Hispanic origin

"From which country or region do you trace your heritage or ancestry? (Check all that apply) [asked if any-part Hispanic]"

	n (choose all that apply)
Mexico	21,059
United States	20,039
Spain	9,338
Puerto Rico	8,151
South America	4,174
Cuba	3,088
Central America	2,669
Dominican Republic	1,511
Caribbean	805

- Years: 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024
- This is a multichoice question that is then concatenated into a single character. The sequence `"!!"` is used as a delimiter. For example, `"United States!!Cuba"` indicates that the respondent chose both "United States" and "Cuba" as responses in this multi-choice question.
- In the above table, we separately count the occurrence of each of the options. For example, the respondent in the above example will be counted twice in the table.
- This question is asked only to those who identify as any-part Hispanic in the `race` and `hispanic` questions, with a few exceptions in 2016 and 2018 for multi-racial respondents.
- To make the text manageable, the following response options are *not* considered and extracted from each year's common content when making this variable: `"Other"`, `"Not Hispanic or Latino"`, `"No country in particular"`. Open-text responses in the `"Other"` responses are not used either. See each year's common content for these data.

`citizen`: Citizenship

[Based on self-report for immigration status]

	n
Citizen	665,258
Non-Citizen	10,472
(Missing)	26,225

- Years: 2006, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024
- These come from the immigration status questions (often called `immstat`), which ask 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.
- Limitation: Most of the missingness comes from 2007 and 2015, when the immigration status question does not appear to have been asked.

`religion`: Religion

"What is your present religion, if any?"

	n
Protestant	240,966
Roman Catholic	134,146
Mormon	9,332
Eastern or Greek Orthodox	3,547
Jewish	16,614
Muslim	3,968
Buddhist	5,749
Hindu	2,023
Atheist	37,378
Agnostic	40,175
Nothing in Particular	125,279
Something Else	45,190
(Missing)	37,588

- Years: All of 2007-2024
- The response options have stayed largely consistent and follow that of Pew.

`relig_imp`: Importance of religion

"How important is religion in your life?"

	n
Very Important	249,412
Somewhat Important	166,912
Not Too Important	96,553
Not at All Important	135,819
(Missing)	53,259

- Years: All of 2008-2024
- In common content, often named `pew_religimp`

`relig_bornagain`: Evangelical Christian

"Would you describe yourself as a born-again or evangelical Christian, or not?"

	n
Yes	199,689
No	482,355
8	39
(Missing)	19,872

- Years: All of 2006-2024 except 2007
- In common content, often named `pew_bornagain`

`relig_protestant`: Branch of Protestantism

"To which Protestant church or group do you belong?"

	n
Baptist	76,357
Methodist	33,706
Nondenominational or Independent Church	54,041
Lutheran	26,262
Presbyterian	16,348
Pentecostal	16,562
Episcopalian	11,437
Church of Christ or Disciples of Christ	9,725
Congregational or United Church of Christ	5,866
Holiness	2,413
Reformed	2,218
Adventist	2,438
Jehovah's Witness	2,843
Something Else	21,399
(Missing)	420,340

- Years: All of 2007-2024
- In common content, often named `religpew_protestant`

`relig_church`: Church Attendance

"Aside from weddings and funerals, how often do you attend religious services?"

	n
More Than Once a Week	54,292
Once a Week	115,467
Once or Twice a Month	50,451
A Few Times a Year	90,019
Seldom	149,846
Never	185,092
Don't Know	9,597
8	28
(Missing)	47,163

- Years: All of 2008-2024
- In common content, often named `pew_churatd`

Family Status

`marstat`: Marital Status

"What is your marital status?"

	n
Married	365,667
Separated	12,326
Divorced	78,199
Widowed	35,573
Single / Never Married	173,290
Domestic Partnership	35,260
(Missing)	1,640

- Years: All of 2006-2024
- 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 / Civil Partnership” in 2017 and 2018.

`ownhome`: Home Ownership

“Do you own your home or pay rent?”

	n
Own	416,264
Rent	209,586
Other	30,909
(Missing)	45,196

- Years: 2006, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024

`has_child`: Parent of Young Children

“Are you the parent or guardian of any children under the age of 18?”

	n
Yes	159,067
No	495,387
(Missing)	47,501

- Years: All of 2008-2024

`no_milstat`: Military Status (None)

[Based on military household question; neither respondent nor immediate family has served]

	n
Yes	304,431
No	387,430
(Missing)	10,094

- Years: All of 2006-2024 except 2007
- The original question is of the form "We'd like to know whether you or someone in your immediate family is currently serving or has ever served in the U.S. military. Immediate family is defined as your parents, siblings, spouse, and children. Please check all boxes that apply.", where respondents can pick more than one of the options including the following: "I served personally", "Family served previously". The entry in the cumulative response only selects the "None" option. A value of `no_milstat == "Yes"` means that a respondent indicated they had neither served nor had an immediate family member who has served. To see the other responses, see the individual year's CCES.

Validated Turnout and Registration

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	342,915
No Record of Registration	128,866
Unregistered	18,070
Dropped	8,433
Inactive	4,086
Multiple Appearances	2,165
(Missing)	197,420

- Years: 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Limitations: Collapses some response options
- Note that the response categories changed from 2022, when CES started using TargetSmart instead of Catalist for their voter match. The TargetSmart matches do not distinguish between unregistered, dropped, vs. inactive registrants.

`vv_party_gen`: Validated registered party

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

	n
No Record of Party Registration	178,523
Unknown	87,810
Democratic Party	70,598
Republican Party	50,802
No Party Affiliation	18,380
Declined to State	3,093
Independent Party	2,829
Other	2,531
Libertarian Party	1,061
Green Party	479
Conservative Party	114
Working Family Party	48
Constitution Party	47
Reform Party	13
Socialist Party	7
(Missing)	285,620

- Years: 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Limitations: Note that the response categories changed from 2022, when CES started using TargetSmart instead of Catalist for their voter match. As of V11 of this dataverse replication, for example, TS did not report NPAs as a separate category in Florida. To identify NPA registrants, a good substitute may be to identify those in party registration states who are `vv_party_gen = "No Record of Party Registration"` but are active in `vv_regstatus`.

`vv_party_prm`: Validated registered Primary party

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

	n
No Record of Party Registration	358,949
Democratic Party	30,329
Republican Party	26,932
Libertarian Party	40
Other	44
No Party Affiliation	27
Green Party	7
Independent Party	7
(Missing)	285,620

- Years: 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Limitations: Not available for some even years

`vv_state`: Validated state of registration

[Validation results.]

	n
	110,960
MD	4,505
CA	21,083
IN	5,852
NY	14,060
OH	10,880
NJ	6,752
TX	17,423
FL	18,483
KS	2,157
GA	7,890
MO	6,024
PA	13,463
CO	4,523
IL	10,053
OK	2,435
AZ	6,313
MN	4,573
MA	5,132
NE	1,567
TN	5,361
NV	2,812
OR	4,514
CT	2,610
UT	1,776
VA	7,367
WI	5,373
NH	1,364
HI	761
AK	528
AL	3,523
SC	3,912
WA	6,072
NC	7,871
NM	1,710
MS	1,770
VT	588
LA	2,659
MI	8,843
IA	2,776
WV	1,813
AR	2,044
KY	3,818
ME	1,383
SD	704
RI	845
MT	1,038
ID	1,301
DE	958
DC	603
ND	566
WY	409
(Missing)	340,155

- Years: 2014, 2016, 2018, 2020, 2022, 2024
- 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	312,185
No Record of Voting	227,038
No Voter File	1,733
(Missing)	160,999

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024.
- 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	340,690
Voted	162,482
No Voter File	1,363
(Missing)	197,420

- Years: 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Limitations: See `vv_turnout_gvm`

`intent_turnout_self`: Self-reported turnout (pre-election wave)

"2020: Do you intend to vote in the 2020 general election on November 3rd?"

	n
Yes, definitely	282,042
Probably	32,331
I already voted (early or absentee)	36,668
Plan to vote early	8,209
I Plan to Vote Before November 5th	2,600
No	30,905
Undecided	22,407
(Missing)	286,793

- Years: 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Limitations: Collapses and simplifies response categories to standardize across years. See individual year's datasets for details. Some response categories are still not standardized. For example, "Plan to vote early" was not an option only in 2016. Additionally, the question is available in years 2010 and earlier but not included in this version of the cumulative dataset because those years rely on a 2006–2012 cumulative release.

`voted_turnout_self`: Self-reported turnout (post-election wave)

"2020: Which of the following statements best describes you?"

	n
Yes	387,495
No	51,124
(Missing)	263,336

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Limitations: Collapses reasons for not voting into a single "No". Actual responses distinguish between "Did not vote" and "Tried to vote but could not" in some years, for example. To standardize across years, the cumulative dataset lumps these levels and others together. See the individual year's datasets for details. The cumulative dataset also does not include responses for odd years even though in some odd years this question is asked.

Partisan Identity

`pid3`: Partisan identity (3 point)

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

	n
Democrat	257,527
Republican	184,380
Independent	197,155
Other	28,346
Not Sure	34,306
(Missing)	241

- Years: All of 2006-2024
- 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. 2020 values do not include the `Not Sure` option (option 5).

`pid7`: Partisan identity (7 point)

[Based on branching from Partisan Identity question]

	n
Strong Democrat	173,790
Not Very Strong Democrat	83,093
Lean Democrat	70,754
Independent	97,588
Lean Republican	69,137
Not Very Strong Republican	65,183
Strong Republican	118,558
Not Sure	20,817
Don't Know	3
(Missing)	3,032

- Years: All of 2006-2024
- 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)	327,637
Republican (Including Leaners)	252,878
Independent (Excluding Leaners)	97,588
Not Sure	20,817
(Missing)	3,035

- Years: All of 2006-2024
- Limitations: See `pid3`

`ideo5`: Ideology (5 point)

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

	n
Very Liberal	72,537
Liberal	125,017
Moderate	221,194
Conservative	150,837
Very Conservative	80,356
Not Sure	50,115
(Missing)	1,899

- Years: All of 2006-2024

Economics and Income

`faminc`: Family Income

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

	n
Less than 10k	33,744
10k - 20k	51,740
20k - 30k	70,700
30k - 40k	69,839
40k - 50k	62,076
50k - 60k	61,124
60k - 70k	45,417
70k - 80k	50,163
80k - 100k	57,837
100k - 120k	42,859
120k - 150k	37,330
150k+	45,847
Prefer not to say	71,619
Skipped	12
(Missing)	1,648

- Years: All of 2006-2024
- Limitations: The income brackets provided changed slightly over time. The brackets in this cumulative dataset coarsen certain original 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.

`employ`: Employment Status

"Which of the following best describes your current employment status?"

	n
Full-Time	272,710
Part-Time	73,479
Temporarily Laid Off	5,931
Unemployed	46,845
Retired	155,290
Permanently Disabled	42,552
Homemaker	48,333
Student	29,962
Other	16,334
(Missing)	10,519

- Years: All of 2006-2024 except 2007

`no_healthins`: Uninsured

[Based on health insurance question; respondent has none of the insurance options given]

	n
Yes	63,990
No	558,681
(Missing)	79,284

- Years: All of 2009-2024
- The original question is of the form “Do you currently have health insurance? (check all that apply)”, where respondents can pick more than one of the options including the following: “Yes, through my job or a family member’s employer”, “Yes, through a government program, such as Medicare or Medicaid”. The entry in the cumulative response only selects the “None” option. A value of `no_healthins == "Yes"` means that a respondent indicated they were not insured. To see the other responses, see the individual year’s CCES.

`union`: Union membership

“Are you a member of a union?”

	n
Yes, Currently	43,432
Yes, Formerly	125,908
No, Never	463,978
(Missing)	68,637

- Years: 2006, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2024
- Question wording and response options have been reworded to be harmonized across years. Please see each individual CCES for exact wording.
- The 2008 CCES in its common content has a union question that roughly combines both the `union` and `union_hh` question.

`union_hh`: Union membership in household

“Other than yourself, is any member of your household a union member?”

	n
Yes, Currently	54,524
Yes, Formerly	88,316
No, Never	474,290
Not Sure	13,357
(Missing)	71,468

- Years: 2006, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2024
- Question wording and response options have been reworded to be harmonized across years. Please see each individual CCES for exact wording.
- The 2008 CCES in its common content has a union question that roughly combines both the `union` and `union_hh` question.

`investor`: Investor

"Do you personally (or jointly with a spouse), have any money invested in the stock market right now, either in an individual stock or in a mutual fund?"

	n
Yes	215,180
No	283,957
(Missing)	202,818

- Years: 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2024

`economy_retro`: Retrospective economy

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

	n
Gotten much better	47,711
Gotten better / somewhat better	139,913
Stayed about the same	155,691
Gotten worse / somewhat worse	175,135
Gotten much worse	163,615
Not sure	18,686
(Missing)	1,204

- Years: All of 2006-2024
- Limitations: Response options vary 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	348,558
Some of the time	172,164
Only now and then	82,530
Hardly at all	43,518
Don't Know	17,998
(Missing)	37,187

- Years: All of 2007-2024
- 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	144,567
Approve / Somewhat Approve	161,915
Disapprove / Somewhat Disapprove	72,140
Strongly Disapprove	301,941
Never Heard / Not Sure	20,071
Neither Approve nor Disapprove	443
(Missing)	878

- Years: All of 2006-2024
- 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	105,671
Approve / Somewhat Approve	200,133
Disapprove / Somewhat Disapprove	107,892
Strongly Disapprove	119,160
Never Heard / Not Sure	158,161
Neither Approve nor Disapprove	1,798
(Missing)	9,140

- Years: All of 2006-2024
- 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	97,968
Approve / Somewhat Approve	200,775
Disapprove / Somewhat Disapprove	124,447
Strongly Disapprove	153,297
Never Heard / Not Sure	118,693
Neither Approve nor Disapprove	1,413
(Missing)	5,362

- Years: All of 2006-2024
- 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	106,848
Approve / Somewhat Approve	192,103
Disapprove / Somewhat Disapprove	116,303
Strongly Disapprove	156,110
Never Heard / Not Sure	123,306
Neither Approve nor Disapprove	1,158
(Missing)	6,127

- 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	121,028
Approve / Somewhat Approve	212,358
Disapprove / Somewhat Disapprove	120,686
Strongly Disapprove	183,536
Never Heard / Not Sure	59,902
Neither Approve nor Disapprove	1,414
(Missing)	3,031

- Years: All of 2006-2024
- Limitations: See `approval_pres`
- To see who the Governor refers to for a particular respondent, see `gov_inc`.

Vote Choice Variables

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 (in contrast to the intent questions) come from the post-election wave only. As of V5, I do not coalesce pre-election respondents who reported having already voted early.[3]

A note on missingness, undervotes, and not voting: As of V10 (2025), people **who self-report not turning out or not voting in a year's election at all have a missing (NA) value** for the variables starting with **voted_pres_**. In prior versions, these cases were given the value "Did not Vote", but it became too confusing to maintain this distinction because coding between years are inconsistent. Also as of V10, we provide the value **Undervoted** to indicate "I turned out to vote in the election but I skipped the office of President." However, a missing value can also mean that the respondent did not take the post-election wave, where this data comes from.

In short, respondents who have missing values for intent or vote choice can also be non-voters for a variety of reasons. When estimating voteshares, users should also look at the year-specific variables, turnout variables, and possibly the individual CES datasets. For turnout in the election, see the section on turnout.

Presidential Vote

intent_pres_party: President preference party

[Party of presidential candidate chosen in intent_pres]

	n
Democratic	111,184
Republican	92,029
Third Party	5,621
Other	25,937
(Missing)	467,184

- Years: 2008, 2012, 2016, 2020, 2024

voted_pres_party: President vote in last election

[Party of presidential candidate chosen in last election]

[3] 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.

	n
Democratic	251,401
Republican	209,625
Third Party	6,658
Other	19,954
Undervote	1,545
(Missing)	212,772

- Years: All of 2008-2024
- See the variables for separate years (e.g. `voted_pres_16`, `voted_pres_20`, etc..) for more detailed breakdowns between third-party candidates.
- Note that `NA` values can be a result of not answering the post-election wave, OR self-reporting that they did not turn out in the election (as of V10). See note above.
- If a respondent reports that they did not turn out to vote (`voted_turnout_self`) but gives a candidate in `voted_pres_**`, then we set `voted_pres_party` to NA, while leaving the `voted_pres_**` as is. This only happens, e.g. in about 20-30 cases in 2024.
- Note: In a presidential election year, this asks the vote of *that* year. The vote choice of the presidential election 4 years prior might be recorded separately. For example, for respondents in 2012, `voted_pres_party` corresponds to their 2012 vote, while `voted_pres_08` corresponds to their 2008 vote (which was asked in the same 2012 survey). For respondents in odd years, this represents their recollection of their vote in the last presidential election. For example, `voted_pres_party` for a respondent in 2023 indicates how they reported (in 2023) voting in the **2020** election.

`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
Not Sure	1,697
(Missing)	671,760

- Years: 2008
- Also see: `intent_pres_party` for vote choice in the most recent preceding presidential election into one party column.

`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
Not Sure	3,856
(Missing)	649,712

- Years: 2012
- Also see: `intent_pres_party` for vote choice in the most recent preceding presidential election into one party column.

`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
Not Sure	6,536
(Missing)	638,953

- Years: 2016
- Also see: `intent_pres_party` for vote choice in the most recent preceding presidential election into one party column.

`intent_pres_20`: 2020 President preference (before voting)

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

	n
Donald Trump (Republican)	18,977
Joe Biden (Democrat)	22,400
Other	1,389
I Won't Vote in this Election	2,390
Not Sure	3,791
(Missing)	653,008

- Years: 2020
- Also see: `intent_pres_party` for vote choice in the most recent preceding presidential election into one party column.

`intent_pres_24`: 2024 President preference (before voting)

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

	n
Kamala Harris (Democrat)	23,984
Donald Trump (Republican)	19,765
Other	1,656
I Won't Vote in this Election	1,992
Not Sure	2,999
(Missing)	651,559

- Years: 2024
- Also see: `intent_pres_party` for vote choice in the most recent preceding presidential election into one party column.

`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	4,204
Not Sure	1,787
(Missing)	553,580

- Years: 2008, 2009, 2010, 2011, 2012
- Limitations: Response options offer slightly by year; some are collapsed into one. In most years (2008-2010) this question is asked only to those who respondent saying they voted (those who said they did not vote are given a missing value). In 2011, there was no such branching.
- See `voted_pres_party` for vote choice in the most recent preceding presidential election into one party column.

`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	82,543
Mitt Romney	64,740
Other	5,872
Undervote	704
Not Sure	1,990
(Missing)	546,106

- Years: 2012, 2013, 2014, 2015, 2016
- Limitations: Response options offer slightly by year; some are collapsed into one.
- Note that `NA` values can be a result of not answering the post-election wave, OR self-reporting that they did not turn out in the election (as of V10). See note on Presidential vote.

- See `voted_pres_party` for vote choice in the most recent preceding presidential election into one party column.

`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	131,931
Donald Trump	110,839
Gary Johnson	8,370
Evan McMullin	1,282
Jill Stein	4,589
Other	10,892
Undervote	617
Not Sure	516
(Missing)	432,919

- Years: 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2024
- Note that `NA` values can be a result of not answering the post-election wave, OR self-reporting that they did not turn out in the election (as of V10). See note on Presidential vote.
- See `voted_pres_party` for vote choice in the most recent preceding presidential election into one party column.

`voted_pres_20`: 2020 President vote choice (after voting)

[If reported voting] 2024: For whom did you vote for President of the United States? [Post-election]

	n
Joe Biden	103,825
Donald Trump	73,790
Jo Jorgensen	2,182
Howie Hawkins	783
Other	3,068
Undervote	113
Not Sure	190
(Missing)	518,004

- Years: 2020, 2021, 2022, 2023, 2024

`voted_pres_24`: 2024 President vote choice (after voting)

[If reported voting] 2024: For whom did you vote for President of the United States? [Post-election]

	n
Kamala Harris	23,527
Donald Trump	18,501
Jill Stein	376
Robert F. Kennedy, Jr.	194
Cornel West	92
Chase Oliver	288
Other	574
Undervote	260
(Missing)	658,143

- Years: 2024
- Note that NA values can be a result of not answering the post-election wave, OR self-reporting that they did not turn out in the election (as of V10). See note on Presidential vote.
- Also see: `voted_pres_party` for vote choice in the most recent preceding presidential election into one party column.

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]	190,605
[Republican / Candidate 2]	165,769
[Other / Candidate 3]	5,488
$HouseCand4Name ($HouseCand4Party)	126
Other	3,568
$HouseCand5Name ($HouseCand5Party)	40
I Won't Vote in this Election	2,269
$HouseCand6Name ($HouseCand6Party)	74
$HouseCand9Name ($HouseCand9Party)	2
$HouseCand7Name ($HouseCand7Party)	37
$HouseCand8Name ($HouseCand8Party)	24
$HouseCand10Name ($HouseCand10Party)	1
$HouseCand11Name ($HouseCand11Party)	3
No One	30,443
Not Sure	97,076
(Missing)	206,430

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- 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]	142,855
[Republican / Candidate 2]	117,389
[Other / Candidate 3]	4,880
$SenCand4Name ($SenCand4Party)	89
Other	2,526
Not Sure	52,247
No One	18,837
I Won't Vote in this Election	1,145
(Missing)	361,987

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- 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]	99,966
[Republican / Candidate 2]	85,707
[Other / Candidate 3]	4,385
Other	1,992
Not Sure	30,548
No One	11,287
I Won't Vote in this Election	466
(Missing)	467,604

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Limitations: See `intent_rep`. Here we only record even-year races. 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]	186,079
[Republican / Candidate 2]	162,989
[Other / Candidate 3]	3,502
$HouseCand4Name ($HouseCand4Party)	91
$HouseCand7Name ($HouseCand7Party)	37
Other	4,774
I Did Not Vote In This Race	12,362
$HouseCand5Name ($HouseCand5Party)	42
Not Sure	6,395
$HouseCand6Name ($HouseCand6Party)	66
I Did not Vote in this Race	3,651
I Did not Vote	345
$HouseCand8Name ($HouseCand8Party)	16
$HouseCand9Name ($HouseCand9Party)	2
$HouseCand10Name ($HouseCand10Party)	2
$HouseCand11Name ($HouseCand11Party)	3
(Missing)	321,599

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- 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]	134,782
[Republican / Candidate 2]	112,426
[Other / Candidate 3]	3,361
$SenCand4Name ($SenCand4Party)	64
Other	2,317
I Did not Vote	28
Not Sure	2,733
I Did Not Vote In This Race	6,299
(Missing)	439,945

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- 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]	87,798
[Republican / Candidate 2]	77,113
[Other / Candidate 3]	2,851
I Did not Vote in this Race	327
I Did not Vote	62
Other	2,374
I Did Not Vote In This Race	3,576
Not Sure	1,344
(Missing)	526,510

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- See `voted_gov_party` for party affiliation.

`voted_rep_party`: House vote choice party

[Party of candidate chosen by respondent in voted_rep]

	n
Democratic	185,716
Republican	161,646
Libertarian	1,483
Independent	1,342
Green	254
Conservative	82
Reform	61
Constitution	37
Independence	32
No Party Preference	33
United Utah	22
Independent (Democratic)	19
Tea	16
Liberty Union	11
Socialist Workers	11
Working Class	10
Working Families	7
American	6
Green Independent	5
Independent American	6
Democratic-Republican	5
No Party Affiliation	5
D.C. Statehood Green	3
Unaffiliated	4
Liberty	3
Moderate	3
For Americans	2
Statehood Green	2
Unity	2
Alliance	1
Libertarian/Green	1
No Party Affiliation; (Democratic)	1
Wake Up USA	1
(Missing)	351,123

  - Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
  - Similar to `voted_pres_party`. See `voted_rep_chosen` for the name of the candidate, and `intent_rep_party` for the pre-election version of this variable.

`voted_sen_party`: Senate vote choice party

[Party of candidate chosen by respondent in voted_sen]

	n
Democratic	135,912
Republican	109,549
Independent	2,387
Libertarian	859
Green	203
Constitution	46
Write-In ONLY	16
Tea	12
Independence	11
Reform	7
Independent American Party of Utah	3
(Missing)	452,950

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Similar to `voted_pres_party`. See `voted_sen_chosen` for the name of the candidate, and `intent_sen_party` for the pre-election version of this variable.

`voted_gov_party`: Governor vote choice party

[Party of candidate chosen by respondent in voted_gov]

	n
Democratic	87,341
Republican	76,588
Independent	438
Libertarian	411
American Constitution	351
Conservative	163
Independence	47
Green	39
Independent Party of Oregon	34
Alaska Independence	2
(Missing)	536,541

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Similar to `voted_pres_party`. See `voted_gov_chosen` for the name of the candidate, and `intent_gov_party` for the pre-election version of this variable.

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.

```
Rows: 701,955
Columns: 2
$ year    <int> 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, 20~
$ case_id <int> 439219, 439224, 439228, 439237, 439238, 439242, 439251, 439254~
```

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.

Because of the changes in naming by year, users should not assume that `rep_current` and `voted_rep_chosen` of a given respondent should be named the same way.

```
Rows: 701,955
Columns: 4
$ rep_current  <chr> "Patrick T. McHenry (R)", "Michael R. Turner (R)", "Rober~
$ sen1_current <chr> "Elizabeth Dole (R)", "Mike DeWine (R)", "Robert Menendez~
$ sen2_current <chr> "Richard Burr (R)", "George V. Voinovich (R)", "Frank R. ~
$ gov_current  <chr> "Michael Easley (D)", "Bob Taft (R)", "Jon Corzine (D)", ~
```

ICPSR Identifiers

Unique identifiers (ICPSR / Nominate for Congress) 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
- Missing for DC residents.

```
Rows: 701,955
Columns: 3
$ rep_icpsr  <dbl> 20522, 20342, 29132, 29911, 29380, 20531, 29126, 29739, 205~
$ sen1_icpsr <dbl> 40303, 15020, 29373, 15021, 14858, 49306, 40101, 15054, 493~
$ sen2_icpsr <dbl> 29548, 49903, 14914, 40502, 40105, 40305, 40302, 29537, 403~
```

- Years: All of 2006-2024
- 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, as_factor(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 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 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)
# i 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

```
Rows: 701,955
Columns: 6
$ intent_rep_chosen <chr> "Richard C. Carsner (D)", "Stephanie Studebaker (D)"~
$ intent_sen_chosen <chr> NA, "Sherrod C. Brown (D)", "Robert Menendez (D)", N~
$ intent_gov_chosen <chr> NA, "Ted Strickland (D)", NA, "Rod Blagojevich (D)",~
$ voted_rep_chosen  <chr> "Richard C. Carsner (D)", "Stephanie Studebaker (D)"~
$ voted_sen_chosen  <chr> NA, "Sherrod C. Brown (D)", "Robert Menendez (D)", N~
$ voted_gov_chosen  <chr> NA, "Ted Strickland (D)", NA, "Rod Blagojevich (D)",~
```

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024
- Early years may mislabel the candidate's party, especially when the two candidates are of the same party (as in top-two primary states)
- See the variables that replace `_chosen` with `_party` for the party of the candidate.

Version History of Dataverse Releases

Routine edits add new rows, add new variables, and change the customization of existing variables. The version history is explained below and old versions can be downloaded from Dataverse. Dataverse assigns version numbers by incrementing by a full number if any of the datasets change, and by incrementing by a decimal when only the description/metadata changes.

Version 11.0 (released `2025-10-08`)

- Includes vote validation variables for 2024 (`vv_regstatus`, `vv_turnout_gvm`, etc..). Value label numbering has changed as a result.
- Adds vote validated registered state as variable (`vv_state`)
- Renames what was previously `rvweight` to `vvweight` so that it is consistent with the Common content
- Fix: for `approval_*`, Standardize "Never Heard of this Person" in 2013-2014, 2018 to "Never Heard" like other years (issue 71)

Version 10.0 (released `2025-05-21`)

- Adds 2024 Common Content (before vote validation), up to n = 701,955
- Fixes a bug in `voted_pres_party` where past election vote was being filled in for some respondents who did not take the post-election wave in that year. Created a new value, "undervote", to distinguish between not turning out at all, and changed self-reported no turnout to NA
- Added self-reported registration (`reg_self`)
- Added sexual orientation (`sexuality`) and investor status (`investor`)

Version 9.0 (released `2024-06-17`)

- Adds 2023 Common Content, up to n = 641,955
- Adds 2022 vote validation
- Fixes and modifies issues in 2008 presidential vote "did not vote" (issue 63 on Github)
- Missing values in Stata are no longer large negative numbers (Thanks to Peter Wielhouwer)

Version 8.0 (released `2023-05-12`)

- Adds 2022 Common Content (before vote validation), up to n = 617,455
- Adds **`gender4`** (a variable for gender identification introduced in 2021), add back `gender` for 2021-2022, and add **`sex`** for 2006-2020.
- Adds a "any-part Hispanic" variable, **`race_h`**, that combines `race` and `hispanic`.
- Adds a "Hispanic origin" variable, **`hisp_origin`**, which concatenates the responses to the hispanic question.
- Added more religion variables (**`relig_imp`**, **`relig_bornagain`**, **`relig_protestant`**, **`relig_church`**).
- Added feather version of the data

Version 7.0 (released `2022-03-24`)

- Enter 2021 Common Content (up to n = 557,455)
- Enter 2020 validated vote variables
- Corrects error in 2010 3 point Party ID which had used the post-election wave rather than the pre-election wave used in the other years (Thanks to Gerald Wright).
- Adds variables for: **self reported turnout**

Version 6.0 (released `2021-04-06`)

- Enter 2020 common content (preliminary, before voter file match), pre-election (up to n = 531,755).
- Distinguished between third party Presidential vote (thanks to Valerie Bradley)
- Added further explanation to usage of weights where post-election weights are not available (thanks to Alexander Agadjanian)
- Added usage example of the R `dataverse` package

Version 5.0 (released `2020-10-04`)

- Enter 2019 common content (up to n = 470,755)
- Adds variables for: **employment, union membership, religion, citizenship, children, home-ownership, lack of insurance, and lack of military members in immediate family**. (thanks to Brian Schaffner)
- Add a separate variable for the post-election wave values of `state`, `st`, `dist`, `cd`, and `cd_up`. Between the pre-post waves, about 0.9 percent of respondents appear to move CDs and 0.4 percent move states.
- Undo coalescing pre-election wave already-voted vote choice, keeping `voted_*` variables with just post-election wave responses.
- Adds a separate variable for intent/voted party choice in presidential race
- Add leading zeros to congressional districts in the first digits (e.g., `"MA-1"` is now `"MA-01"`), and create a variable `cd_up` similar to `dist_up`.
- Modify prepositions of value labels to lower case (e.g., `Not At All` to `Not at All`)

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 (thanks to Bernard Fraga)

Version 1.0 (released `2018-01-24`)

- First upload, covering 2006 - 2016 (n = 374,556)