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

Shiro Kuriwaki*

2024-06-17

Cite this dataset as:

> Kuriwaki, Shiro, 2024, "Cumulative CES Common Content", doi:10.7910/DVN/II2DB6, Harvard Dataverse, V9.

This dataset combines 18 years (2006 – 2023) 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 Stephen Ansolabehere, Jeremy Pope, Brian Schaffner, 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 (harmonizes) 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.

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, 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) . . . 6
Features of the Cumulative Dataset . . . 8
Unified Variable Names . . . 8
Chosen Candidate Names and Identifiers . . . 8
Variables . . . 8
Administration . . . 9
year: CCES year . . . 9
starttime: Start time . . . 9
tookpost: Took post-election wave . . . 9
Weights . . . 10
weight: Survey weight (Year-Specific) . . . 10
weight_cumulative: Survey weight (Cumulative) . . . 10
weight_post: Survey weight for post-election wave . . . 11
rvweight: Survey weights to validated registered voters . . . 12
rvweight_post: Survey weights to validated registered voters, post-election wave . . . 12
Geography . . . 12
Demographics . . . 13
gender: Sex (standardized) . . . 13
gender4: Gender . . . 14
sex: Sex . . . 14
birthyr: Year of birth . . . 14
age: Age . . . 14
educ: Education . . . 15
race: Race . . . 15
hispanic: Hispanic . . . 15
race_h: Race (any-part Hispanic) . . . 16
hisp_origin: Hispanic origin . . . 16
citizen: Citizenship . . . 17
religion: Religion . . . 17
relig_imp: Importance of religion . . . 17
relig_bornagain: Evangelical Christian . . . 18
relig_protestant: Branch of Protestantism . . . 18
relig_church: Church Attendance . . . 18
Family Status . . . 19
marstat: Marital Status . . . 19
ownhome: Home Ownership . . . 19
has_child: Parent of Young Children . . . 20
no_milstat: Military Status (None) . . . 20
Validated Turnout and Registration . . . 20
vv_regstatus: Validated registration status . . . 20
vv_party_gen: Validated registered party . . . 21
vv_party_prm: Validated registered Primary party . . . 21
Turnout . . . 22
vv_turnout_gvm: Validated turnout General Election . . . 22
vv_turnout_pvm: Validated turnout Primary Election (Congressional) . . . 22

intent_turnout_self: Self-reported turnout (pre-election wave) 23
voted_turnout_self: Self-reported turnout (post-election wave) 23
Partisan Identity 23
pid3: Partisan identity (3 point) 23
pid7: Partisan identity (7 point) 24
pid3_leaner: Partisan identity (including leaners) 24
ideo5: Ideology (5 point) 24
Economics and Income 25
faminc: Family Income 25
employ: Employment Status 25
no_healthins: Uninsured 26
union: Union membership 26
union_hh: Union membership in household 27
economy_retro: Retrospective economy 27
News Interest 27
newsint: News Interest 27
Approval 28
approval_pres: President approval 28
approval_rep: House Representative approval 28
approval_sen1: Senator 1 approval 29
approval_sen2: Senator 2 approval 29
approval_gov: Governor approval 29
Vote Choice Variables 31
Presidential Vote 31
intent_pres_party: President preference party 31
voted_pres_party: President vote in last election 31
intent_pres_08: 2008 President preference (before voting) 32
intent_pres_12: 2012 President preference (before voting) 32
intent_pres_16: 2016 President preference (before voting) 32
intent_pres_20: 2020 President preference (before voting) 33
voted_pres_08: 2008 President vote choice (after voting) 33
voted_pres_12: 2012 President vote choice (after voting) 34
voted_pres_16: 2016 President vote choice (after voting) 34
voted_pres_20: 2020 President vote choice (after voting) 34
House, Senate and Governor Vote 35
intent_rep: House preference (before voting) 35
intent_sen: Senate preference (before voting) 35
intent_gov: Governor preference (before voting) 36
voted_rep: House vote choice (after voting) 36
voted_sen: Senate vote choice (after voting) 37
voted_gov: Governor vote choice (after voting) 37
Metadata and Identifiers 39
Identifiers 39
Current Representatives' Name and Party 39
ICPSR Identifiers 39
Name of Chosen Candidate 40
Party of Chosen Candidate 41
Version History of Dataverse Releases 42
Version 9.0 (released 2024-06-17) 42
Version 8.0 (released 2023-05-12) 42
Version 7.0 (released 2022-03-24) 42

Version 6.0 (released 2021-04-06) 42
Version 5.0 (released 2020-10-04) 42
Version 4.0 (released 2019-09-09) 43
Version 3.0 (released 2019-04-29) 43
Version 2.0 (released 2018-04-16) 43
Version 1.0 (released 2018-01-24) 43

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_2006-2023.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_2006-2023.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_2006-2023.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_2006-2023.dta",
  dataset = "10.7910/DVN/II2DB6",
  original = TRUE,
  .f = haven::read_dta,
  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"))
```

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: 641,955 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]
```

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

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

```
 8  2006  439254 1 [Democrat]
 9  2006  439255 1 [Democrat]
10  2006  439263 1 [Democrat]
# i 641,945 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: 641,955 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 641,950 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.

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

```
# A tibble: 234,545 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 234,540 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: 641,955 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 641,945 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

`starttime`: Start time

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

- Years: All of 2006-2023
- 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)]

	n
Did Not Take Post-Election Survey	77,532
Took Post-Election Survey	403,424
(Missing)	160,999

	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

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022 (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  0.4505  0.7349  1.0000  1.1611 15.0614
```

- Years: All of 2006-2023
- 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  0.2747  0.4960  0.8201  0.9574 21.4290
```

- Years: All of 2006-2023

- 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 2.21, because it has about twice as many observations as the other datasets.
- This transformation means that this simple cumulative weight makes sense when all years should be re-weighted to have the same sample size, even if that means doubling the weights. For analyses that are done year by year, using `weight` and `weight_cumulative` are equivalent.

::: .cell layout-align="center" ::: .cell-output .cell-output-stdout
“‘

Year	Observations	Factor
2006	36,421	1.25
2007	9,999	0.34
2008	32,800	1.12
2009	13,800	0.47
2010	55,400	1.89
2011	20,150	0.69
2012	54,535	1.86
2013	16,400	0.56
2014	56,200	1.92
2015	14,250	0.49
2016	64,600	2.21
2017	18,200	0.62
2018	60,000	2.05
2019	18,000	0.62
2020	61,000	2.09
2021	25,700	0.88
2022	60,000	2.05
2023	24,500	0.84

“‘
::: :::

`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    15.5  389839
```

- 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. When applying the weights to a subset of the data, make sure that they are rescaled to mean 1 in the non-missing subset (most statistical software will do this automatically).
- To use the `weight_post` for years it is available but `weight` for where it is not, one could coalesce into a new variable.

`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    16.3  558871
```

- Years: 2018, 2020
- 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 the 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  565992
```

- Years: 2018, 2020
- 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: 641,955
Columns: 14
$ 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~
```

- Years: All of 2006-2023
- 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	293,081
Female	347,810
(Missing)	1,064

- Years: All of 2006-2023
- 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	49,341
Woman	59,795
Non-Binary	842
Other	222
(Missing)	531,755

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

`sex`: Sex

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

	n
Male	243,740
Female	288,015
(Missing)	110,200

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

`birthyr`: Year of birth

"In what year were you born?"

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1900    1952    1963    1966    1980    2005
```

- Years: All of 2006-2023

`age`: Age

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

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   35.00   51.00   49.54   63.00  109.00
```

- Years: All of 2006-2023

`educ`: Education

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

	n
No HS	21,662
High School Graduate	177,252
Some College	154,019
2-Year	64,370
4-Year	145,621
Post-Grad	78,964
(Missing)	67

- Years: All of 2006-2023

`race`: Race

"What racial or ethnic group best describes you?"

	n
White	468,573
Black	72,695
Hispanic	55,240
Asian	15,080
Native American	5,209
Mixed	13,422
Other	10,675
Middle Eastern	1,061

- Years: All of 2006-2023
- 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	22,731
No	485,348
(Missing)	133,876

- Years: 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023
- In years in which this question was fielded, this question adds to the `race` variable by asking those who did *not* respond "Hispanic" in the `race` question. For a combined version of `race` and `hispanic`, see `race_h` (any-part Hispanic).
- 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	458,542
Black	71,033
Hispanic	72,709
Asian	14,693
Native American	4,721
Mixed	9,477
Other	9,779
Middle Eastern	1,001

- Years: All of 2006-2023
- 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	17,828
United States	17,223
Spain	7,666
Puerto Rico	7,036
South America	3,591
Cuba	2,639
Central America	2,278
Dominican Republic	1,307
Caribbean	657

- Years: 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023
- 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	606,084
Non-Citizen	9,686
(Missing)	26,185

- Years: 2006, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023
- These come from the immigration status questions, 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	221,928
Roman Catholic	123,285
Mormon	8,709
Eastern or Greek Orthodox	3,232
Jewish	15,131
Muslim	3,555
Buddhist	5,228
Hindu	1,819
Atheist	32,794
Agnostic	35,761
Nothing in Particular	112,632
Something Else	40,353
(Missing)	37,528

- Years: All of 2007-2023
- 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	229,087
Somewhat Important	152,331
Not Too Important	87,130
Not at All Important	120,159
(Missing)	53,248

- Years: All of 2008-2023
- 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	184,074
No	438,000
8	39
(Missing)	19,842

- Years: All of 2006-2023 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	69,727
Methodist	31,242
Nondenominational or Independent Church	48,570
Lutheran	24,171
Presbyterian	15,099
Pentecostal	15,124
Episcopalian	10,568
Church of Christ or Disciples of Christ	8,786
Congregational or United Church of Christ	5,434
Holiness	2,180
Reformed	2,038
Adventist	2,243
Jehovah's Witness	2,595
Something Else	17,474
(Missing)	386,704

- Years: All of 2007-2023
- 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	50,529
Once a Week	106,303
Once or Twice a Month	46,627
A Few Times a Year	82,411
Seldom	136,517
Never	163,772
Don't Know	8,630
8	28
(Missing)	47,138

- Years: All of 2008-2023
- In common content, often named `pew_churatd`

Family Status

`marstat`: Marital Status

"What is your marital status?"

	n
Married	338,976
Separated	11,074
Divorced	70,595
Widowed	31,562
Single / Never Married	156,255
Domestic Partnership	31,869
(Missing)	1,624

- Years: All of 2006-2023
- 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	380,990
Rent	188,659
Other	27,154
(Missing)	45,152

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

`has_child`: Parent of Young Children

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

	n
Yes	147,723
No	446,731
(Missing)	47,501

- Years: All of 2008-2023

`no_milstat`: Military Status (None)

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

	n
Yes	275,112
No	356,749
(Missing)	10,094

- Years: All of 2006-2023 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	303,115
No Record of Registration	108,666
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
- 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	136,093
Unknown	87,810
Democratic Party	50,181
Republican Party	37,148
No Party Affiliation	18,380
Dem	10,591
Rep	6,912
Declined to State	3,093
Other	2,259
Independent Party	1,851
Libertarian Party	726
Ind	504
Green Party	341
Libertarian	172
Cns	77
Green	76
Constitution Party	47
Wor	19
Cons	17
Reform Party	13
Peace and Freedom	9
Working Fam	9
Socialist Party	7
(Missing)	285,620

- Years: 2012, 2014, 2016, 2018, 2020, 2022
- Limitations: Note that if the state's voter roll does not record party registration, this value will be missing. 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	307,577
Democratic Party	21,325
Republican Party	18,723
Dem	4,500
Rep	4,121
Libertarian Party	29
No Party Affiliation	27
Other	19
Green Party	6
Ind	3
Libertarian	3
Independent Party	1
Green	1
(Missing)	285,620

- Years: 2012, 2014, 2016, 2018, 2020, 2022
- 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	275,737
No Record of Voting	203,486
No Voter File	1,733
(Missing)	160,999

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022 (2022 validated vote to be released in Version 9).
- 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	300,091
Voted	143,081
No Voter File	1,363
(Missing)	197,420

- Years: 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022
- 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	243,948
Probably	28,654
I already voted (early or absentee)	27,088
Plan to vote early	8,209
No	27,341
Undecided	19,922
(Missing)	286,793

- Years: 2012, 2014, 2016, 2018, 2020, 2022
- 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	343,700
No	47,904
(Missing)	250,351

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022
- 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	234,545
Republican	168,467
Independent	180,863
Other	25,975
Not Sure	31,864
(Missing)	241

- Years: All of 2006-2023
- 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	157,214
Not Very Strong Democrat	76,687
Lean Democrat	64,250
Independent	89,697
Lean Republican	63,689
Not Very Strong Republican	60,096
Strong Republican	107,732
Not Sure	19,555
Don't Know	3
(Missing)	3,032

- Years: All of 2006-2023
- 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)	298,151
Republican (Including Leaners)	231,517
Independent (Excluding Leaners)	89,697
Not Sure	19,555
(Missing)	3,035

- Years: All of 2006-2023
- Limitations: See `pid3`

`ideo5`: Ideology (5 point)

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

	n
Very Liberal	64,986
Liberal	113,830
Moderate	201,224
Conservative	139,570
Very Conservative	73,947
Not Sure	46,512
(Missing)	1,886

– Years: All of 2006-2023

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	30,485
10k - 20k	47,367
20k - 30k	64,851
30k - 40k	64,603
40k - 50k	57,312
50k - 60k	56,307
60k - 70k	41,738
70k - 80k	46,105
80k - 100k	52,984
100k - 120k	38,930
120k - 150k	33,247
150k+	39,889
Prefer not to say	66,500
Skipped	12
(Missing)	1,625

– Years: All of 2006-2023
– 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	251,591
Part-Time	67,011
Temporarily Laid Off	5,478
Unemployed	42,536
Retired	138,999
Permanently Disabled	38,452
Homemaker	45,068
Student	27,581
Other	14,898
(Missing)	10,341

- Years: All of 2006-2023 except 2007

`no_healthins`: Uninsured

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

	n
Yes	59,633
No	503,038
(Missing)	79,284

- Years: All of 2009-2023
- 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	39,951
Yes, Formerly	114,947
No, Never	418,462
(Missing)	68,595

- Years: 2006, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022
- 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	50,728
Yes, Formerly	81,063
No, Never	430,152
Not Sure	8,578
(Missing)	71,434

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

`economy_retro`: Retrospective economy

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

	n
Gotten much better	42,043
Gotten better / somewhat better	127,393
Stayed about the same	146,589
Gotten worse / somewhat worse	162,165
Gotten much worse	145,423
Not sure	17,186
(Missing)	1,156

- Years: All of 2006-2023
- 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	318,598
Some of the time	155,682
Only now and then	74,887
Hardly at all	39,081
Don't Know	16,555
(Missing)	37,152

- Years: All of 2007-2023
- 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	132,935
Approve / Somewhat Approve	146,447
Disapprove / Somewhat Disapprove	65,318
Strongly Disapprove	277,458
Never Heard / Not Sure	18,478
Neither Approve nor Disapprove	443
(Missing)	876

- Years: All of 2006-2023
- 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	95,419
Approve / Somewhat Approve	186,599
Disapprove / Somewhat Disapprove	101,391
Strongly Disapprove	107,284
Never Heard / Not Sure	114,923
Never Heard of this Person	25,762
Neither Approve nor Disapprove	1,798
(Missing)	8,779

- Years: All of 2006-2023
- 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	87,318
Approve / Somewhat Approve	187,802
Disapprove / Somewhat Disapprove	116,898
Strongly Disapprove	137,938
Never Heard / Not Sure	88,189
Never Heard of this Person	17,210
Neither Approve nor Disapprove	1,413
(Missing)	5,187

- Years: All of 2006-2023
- 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	95,700
Approve / Somewhat Approve	180,171
Disapprove / Somewhat Disapprove	110,294
Strongly Disapprove	140,178
Never Heard / Not Sure	90,315
Never Heard of this Person	18,184
Neither Approve nor Disapprove	1,158
(Missing)	5,955

- 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	107,873
Approve / Somewhat Approve	195,622
Disapprove / Somewhat Disapprove	111,641
Strongly Disapprove	167,346
Never Heard / Not Sure	55,201
Neither Approve nor Disapprove	1,414
(Missing)	2,858

- Years: All of 2006-2023
- 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. It no longer coalesces pre-election respondents who reported having already voted early, as it did before V5 of this dataset. 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.

Be careful of the category "Did not Vote" when making cross-year comparisons. That category is recorded only when the respondent is *asked* a vote choice question and selects that option. In some years, the vote choice question is not asked to those who report not turning out at all in a preceding question. In those years, having a value of "Did not Vote" could mean that "I voted in the election but not for that office." In some years, the cumulative version changes the values of did not vote to missing values to be consistent with the same variable in other years. In short, respondents who have missing values for intent or vote choice can also be non-voters for a variety of reasons. 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	87,200
Republican	72,264
Third Party	5,412
Independent	209
Other Candidate	5,402
(Missing)	471,468

- Years: 2008, 2012, 2016, 2020

`voted_pres_party`: President vote in last election

[Party of presidential candidate chosen in last election]

	n
Democratic	239,508
Republican	199,701
Other Candidate	23,472
Did not Vote	22,695
(Missing)	156,579

- Years: All of 2008-2023

- 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_party_08` corresponds to their 2008 vote (which was asked in the same 2012 survey).

`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)	611,760

- Years: 2008
- 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)	589,712

- Years: 2012
- 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)	578,953

- Years: 2016
- 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)	593,008

- Years: 2020
- 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 / Someone Else	4,204
Did not Vote	251
Not Sure / Don't Recall	1,787
(Missing)	493,329

- 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 / Someone Else	5,872
Did not Vote	217
Did not Vote for this Office	2,602
Not Sure / Don't Recall	1,990
Other	8
(Missing)	483,983

- Years: 2012, 2013, 2014, 2015, 2016
- Limitations: Response options offer slightly by year; some are collapsed into one.
- 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	110,974
Donald Trump	92,760
Other / Someone Else	21,511
Did not Vote for this Office	15,068
Not Sure / Don't Recall	287
Other	229
(Missing)	401,126

- Years: 2016, 2017, 2018, 2019, 2020, 2021, 2022
- 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] 2020: For whom did you vote for President of the United States? [Post-election]

	n
Joe Biden	76,883
Donald Trump	54,357
Other / Someone Else	4,373
Did not Vote for President	19,135
Not Sure	190
Other	113
(Missing)	486,904

- Years: 2020, 2021, 2022, 2023
- 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]	169,617
[Republican / Candidate 2]	148,870
[Other / Candidate 3]	5,034
$HouseCand4Name ($HouseCand4Party)	100
Other	3,136
$HouseCand5Name ($HouseCand5Party)	29
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	27,070
Not Sure	88,739
(Missing)	196,950

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022
- 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]	125,646
[Republican / Candidate 2]	104,431
[Other / Candidate 3]	4,701
$SenCand4Name ($SenCand4Party)	89
Other	2,299
Not Sure	47,599
No One	16,789
I Won't Vote in this Election	1,145
(Missing)	339,256

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022
- 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]	97,089
[Republican / Candidate 2]	83,393
[Other / Candidate 3]	4,218
Other	1,929
Not Sure	29,620
No One	10,968
I Won't Vote in this Election	466
(Missing)	414,272

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022
- 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]	163,560
[Republican / Candidate 2]	144,795
[Other / Candidate 3]	3,163
$HouseCand4Name ($HouseCand4Party)	69
$HouseCand7Name ($HouseCand7Party)	37
Other	4,306
I Did Not Vote In This Race	12,362
$HouseCand5Name ($HouseCand5Party)	28
Not Sure	5,810
$HouseCand6Name ($HouseCand6Party)	66
I Did not Vote in this Race	2,045
I Did not Vote	345
$HouseCand8Name ($HouseCand8Party)	16
$HouseCand9Name ($HouseCand9Party)	2
$HouseCand10Name ($HouseCand10Party)	2
$HouseCand11Name ($HouseCand11Party)	3
(Missing)	305,346

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022
- 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]	116,544
[Republican / Candidate 2]	98,985
[Other / Candidate 3]	3,232
$SenCand4Name ($SenCand4Party)	64
Other	2,317
I Did not Vote	28
Not Sure	2,495
I Did Not Vote In This Race	5,564
(Missing)	412,726

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022
- 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]	84,672
[Republican / Candidate 2]	74,674
[Other / Candidate 3]	2,724
I Did not Vote in this Race	238
I Did not Vote	62
Other	2,289
I Did Not Vote In This Race	3,576
Not Sure	1,301
(Missing)	472,419

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022
- 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.

```
Rows: 641,955
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: 641,955
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

```
Rows: 641,955
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 2006-2023

- 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: 641,955
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
- 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

```
Rows: 641,955
Columns: 8
$ intent_pres_party <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ voted_pres_party  <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ intent_rep_party  <fct> Democratic, Democratic, Democratic, Democratic, Demo~
$ voted_rep_party   <fct> Democratic, Democratic, Democratic, Democratic, Demo~
$ intent_gov_party  <fct> NA, Democratic, NA, Democratic, Democratic, NA, Repu~
$ voted_gov_party   <fct> NA, Democratic, NA, Democratic, Democratic, NA, Repu~
$ intent_sen_party  <fct> NA, Democratic, Democratic, NA, NA, NA, Republican, ~
$ voted_sen_party   <fct> NA, Democratic, Democratic, NA, Democratic, NA, Repu~
```

- 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

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 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)