

Guide to the CCES Cumulative Common Content (2006 - 2020)

 Shiro Kuriwaki*

Guide last updated: 2021-04-06

Cite this dataset as:

Kuriwaki, Shiro, 2021, "Cumulative CCES Common Content", doi:10.7910/DVN/II2DB6, Harvard Dataverse, V6.

This dataset combines 15 years (2006 – 2020) of the Cooperative Congressional Election Study (CCES). The CCES is an online survey conducted around November of each year, asking a range of questions on political behavior and public opinion. Its principal investigators are Stephen Ansolabehere, Sam Luks, and Brian Schaffner.

Each year's CCES 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, 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. In addition, the same dataset is available on Crunch, an analytics interface optimized for survey datasets. The source code is open-source.

Please note that this cumulative dataset makes some modifications to the original CCES 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 datasets**, search the [CCES dataverse](#) or access the [CCES homepage](#). Sign-up to the Crunch dataset from the homepage as well.
- **To understand the survey methodology**, consult the [Frequently Asked Questions](#) page of the CCES homepage or the methodology section of a [recent Common Content's](#) codebook.

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

Contents

| | |
|--|----|
| Getting Started | 4 |
| Data Read-in | 4 |
| Data Download | 4 |
| Unique identifiers and how to add more variables | 4 |
| Labelled variables (for analysis in R) | 5 |
| Features of the Cumulative Dataset | 7 |
| Unified Variable Names | 7 |
| Chosen Candidate Names and Identifiers | 7 |
| Crunch | 7 |
| Variables | 9 |
| Administration | 9 |
| year: CCES year | 9 |
| starttime: Start time | 9 |
| tookpost: Took post-election wave | 10 |
| Weights | 10 |
| weight: Survey weight (Year-Specific) | 10 |
| weight_cumulative: Survey weight (Cumulative) | 11 |
| 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 | 14 |
| gender: Gender | 14 |
| birthyr: Year of birth | 14 |
| age: Age | 14 |
| educ: Education | 14 |
| race: Race | 14 |
| hispanic: Hispanic | 15 |
| citizen: Citizenship | 15 |
| religion: Religion | 15 |
| Family Status | 16 |
| marstat: Marital Status | 16 |
| ownhome: Home Ownership | 16 |
| has_child: Parent of Young Children | 17 |
| no_milstat: Military Status (None) | 17 |
| Validations | 18 |
| vv_regstatus: Validated registration status | 18 |
| vv_party_gen: Validated registered party | 18 |
| vv_party_prm: Validated registered Primary party | 19 |
| Turnout | 19 |
| vv_turnout_gvm: Validated turnout General Election | 19 |
| vv_turnout_pvm: Validated turnout Primary Election (Congressional) | 19 |
| Partisan Identity | 20 |
| pid3: Partisan identity (3 point) | 20 |
| pid7: Partisan identity (7 point) | 20 |
| pid3_leaner: Partisan identity (including leaners) | 20 |
| ideo5: Ideology (5 point) | 21 |
| Economics and Income | 21 |
| faminc: Family Income | 21 |

| | |
|---|----|
| employ: Employment Status | 21 |
| no_healthins: Uninsured | 22 |
| union: Union membership | 22 |
| union_hh: Union membership in household | 23 |
| economy_retro: Retrospective economy | 23 |
| News Interest | 23 |
| newsint: News Interest | 23 |
| Approval | 24 |
| approval_pres: President approval | 24 |
| approval_rep: House Representative approval | 24 |
| approval_sen1: Senator 1 approval | 25 |
| approval_sen2: Senator 2 approval | 25 |
| approval_gov: Governor approval | 25 |
| Vote Choice Variables | 27 |
| Presidential Vote | 27 |
| intent_pres_party: President preference party | 27 |
| voted_pres_party: President vote in last election | 27 |
| intent_pres_08: 2008 President preference (before voting) | 28 |
| intent_pres_12: 2012 President preference (before voting) | 28 |
| intent_pres_16: 2016 President preference (before voting) | 28 |
| intent_pres_20: 2020 President preference (before voting) | 29 |
| voted_pres_08: 2008 President vote choice (after voting) | 29 |
| voted_pres_12: 2012 President vote choice (after voting) | 29 |
| voted_pres_16: 2016 President vote choice (after voting) | 30 |
| voted_pres_20: 2020 President vote choice (after voting) | 30 |
| House, Senate and Governor Vote | 30 |
| intent_rep: House preference (before voting) | 30 |
| intent_sen: Senate preference (before voting) | 31 |
| intent_gov: Governor preference (before voting) | 32 |
| voted_rep: House vote choice (after voting) | 32 |
| voted_sen: Senate vote choice (after voting) | 33 |
| voted_gov: Governor vote choice (after voting) | 33 |
| Metadata and Identifiers | 34 |
| Identifiers | 34 |
| Current Representatives' Name and Party | 34 |
| ICPSR Identifiers | 34 |
| Name of Chosen Candidate | 35 |
| Party of Chosen Candidate | 36 |
| Version History of Dataverse Releases | 37 |
| Version 6.0 | 37 |
| Version 5.0 | 37 |
| Version 4.0 | 37 |
| Version 3.0 | 37 |
| Version 2.0 | 37 |
| Version 1.0 | 38 |

Getting Started

Data Read-in

The .Rds format can be read into R. This format preserves dataset properties such as the distinction between integers and doubles, and labelled variables.

```
cc <- readRDS("cumulative_2006-2020.Rds")
```

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

```
library(tidyverse)
cc
```

A Stata dta version is provided as well. `cumulative_2006-2020.dta` can be read by Stata, or in R by the `haven` package

```
library(haven)
cc <- read_dta("cumulative_2006-2020.dta")
```

R is free software and, if necessary, the `haven` and `readr` package can be used to export the CCES datasets in other formats such as plain-text, or SPSS sav files. Plain-text formats are somewhat less convenient because they do not preserve value labels.

Data Download

Downloading the data via R. 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` (> 0.3.0) allows this by the function:

```
library(dataverse)
cc <- get_dataframe_by_name(
  filename = "cumulative_2006-2019.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 dataset¹.

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:

¹Dagonel, Angelo, 2021, "Cumulative CCES Policy Preferences", doi:10.7910/DVN/05XDQ0, Harvard Dataverse.

```
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 labelled 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: 531,755 x 3
   year case_id pid3
  <int> <int>   <int+lbl>
1  2006  439219 1 [Democrat]
2  2006  439224 4 [Other]
3  2006  439228 1 [Democrat]
4  2006  439237 1 [Democrat]
5  2006  439238 1 [Democrat]
6  2006  439242 3 [Independent]
7  2006  439251 2 [Republican]
8  2006  439254 1 [Democrat]
9  2006  439255 1 [Democrat]
10 2006  439263 1 [Democrat]
# ... with 531,745 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))
```

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

```
# A tibble: 531,755 x 4
  year case_id      pid3 pid3_fct
  <int>   <int>   <int+lbl> <fct>
1  2006  439219 1 [Democrat] Democrat
2  2006  439224 4 [Other]    Other
3  2006  439228 1 [Democrat] Democrat
4  2006  439237 1 [Democrat] Democrat
5  2006  439238 1 [Democrat] Democrat
# ... with 531,750 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: 189,911 x 3
  year case_id      pid3
  <int>   <int>   <int+lbl>
1  2006  439219 1 [Democrat]
2  2006  439228 1 [Democrat]
3  2006  439237 1 [Democrat]
4  2006  439238 1 [Democrat]
5  2006  439254 1 [Democrat]
# ... with 189,906 more rows
```

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

Features of the Cumulative Dataset

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

Unified Variable Names

Most variables in this dataset come straight from each year's CCES. However, it renames and standardizes variable names, making them accessible in one place. Please see the rest of this guide or the Crunch dataset for a full list and description of variables.

Chosen Candidate Names and Identifiers

One addition to this cumulative dataset are variables of candidate names and identifiers that a respondent chose. In the individual year's CCES datasets, typically the response values for a vote choice question is a generic label, e.g., Candidate1 and Candidate2. Then, separate variables of names and parties correspond to each Candidate1 and Candidate2.

Instead, the cumulative dataset shows both the generic label *and* the chosen candidate's name and party, which will vary across individuals.

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

```
# A tibble: 531,755 x 6
  year case_id st      voted_sen      voted_sen_party voted_sen_chosen
  <int>   <int> <chr>   <fct>          <fct>          <chr>
1  2006  439219 NC      <NA>          <NA>          <NA>
2  2006  439224 OH      [Democrat / Candid~ Democratic      Sherrod C. Brown (D)
3  2006  439228 NJ      [Democrat / Candid~ Democratic      Robert Menendez (D)
4  2006  439237 IL      <NA>          <NA>          <NA>
5  2006  439238 NY      [Democrat / Candid~ Democratic      Hillary Rodham Clint~
6  2006  439242 TX      I Did Not Vote In ~ <NA>          <NA>
7  2006  439251 MN      [Republican / Cand~ Republican      Mark Kennedy (R)
8  2006  439254 NV      [Democrat / Candid~ Democratic      Jack Carter (D)
9  2006  439255 TX      [Democrat / Candid~ Democratic      Barbara Ann Radnofsk~
10 2006  439263 MD      I Did Not Vote In ~ <NA>          <NA>
# ... with 531,745 more rows
```

Crunch

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

1. Obtain Access: The Crunch interface currently does not allow users to sign up to a particular dataset on their own. For free view access to the dataset, please sign up via the sign-up link in the CCES [website](#). For questions and more access, please contact the CCES Team.

2. Browse: Crunch offers a web GUI for quickly browsing variables:



3. Analyze: The crunch interface allows Viewers to make cross-tabs and bar graphs quickly.



Crunch datasets can also be manipulated from a R package, `crunch`. To learn more about the features, please take a look at their homepage crunch.io or their [5-minute demo video](#).

Variables

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

- The title shows the name of the variable as it appears in the dataset (“alias” in Crunch terminology), followed by a more descriptive name suitable for presentation (“name” in Crunch terminology).
- Question wordings, where applicable, immediately follow. Otherwise a description is provided in square brackets ([]). All square brackets, both in the description and the response options, indicate descriptions that are summaries rather than the question verbatim.
- A tabulation of response options (or summary statistics for numeric variables) 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 |

starttime: Start time

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

| | 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 |

- Years: All of 2006-2020

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 | 68,513 |
| Took Post-Election Survey | 352,443 |
| (Missing) | 110,799 |

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020 (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 | 0.4479 | 0.7332 | 1.0000 | 1.1611 | 15.0006 |

- Years: All of 2006-2020
- 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`
- In Version 6, the 2020 weights are preliminary weights that were computed without the vote validation in hand. In Version 7, the same variable weight will be updated with YouGov's weight.

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 | 0.3034 | 0.5515 | 0.9252 | 1.0586 | 24.0297 |

- Years: All of 2006-2020
- 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.
- 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.

| 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 |

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 | 0.5 | 0.7 | 1.0 | 1.1 | 15.0 | 330619 |

- Years: 2012, 2016, 2018, 2020
- 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 | 15.0 | 491738 |

- Years: 2018
- In 2018, YouGov computed weights after vote validation to weight to the target population of registered voters. See the methodology section of the [2018 Guide](#) for details. For this reason, and to distinguish it from 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 | 494806 |

- Years: 2018
- 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: 531,755

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 2006-2020
- 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: Gender

“Are you male or female?”

| | n |
|--------|---------|
| Male | 243,741 |
| Female | 288,014 |

– Years: All of 2006-2020

birthyr: Year of birth

“In what year were you born?”

| | | | | | |
|------|---------|--------|------|---------|------|
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 1900 | 1951 | 1962 | 1964 | 1979 | 2002 |

– Years: All of 2006-2020

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 | 49.46 | 62.00 | 109.00 |

– Years: All of 2006-2020

educ: Education

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

| | n |
|----------------------|---------|
| No HS | 16,820 |
| High School Graduate | 146,753 |
| Some College | 129,764 |
| 2-Year | 52,478 |
| 4-Year | 120,876 |
| Post-Grad | 64,997 |
| (Missing) | 67 |

– Years: All of 2006-2020

race: Race

“What racial or ethnic group best describes you?”

| | n |
|-----------------|---------|
| White | 394,442 |
| Black | 58,050 |
| Hispanic | 43,337 |
| Asian | 11,429 |
| Native American | 4,261 |
| Mixed | 10,702 |
| Other | 8,714 |
| Middle Eastern | 820 |

- Years: All of 2006-2020
- Limitations: The “Hispanic” value may undercount self-identified Hispanics. See hispanic

hispanic: Hispanic

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

| | n |
|-----------|---------|
| Yes | 13,197 |
| No | 391,401 |
| (Missing) | 127,157 |

- Years: 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020
- In years in which this question was fielded, this question supplements the race variable by asking those who did *not* respond “Hispanic” in the race question.

citizen: Citizenship

[Based on self-report for immigration status]

| | n |
|-------------|---------|
| Citizen | 497,777 |
| Non-Citizen | 8,055 |
| (Missing) | 25,923 |

- Years: 2006, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2016, 2017, 2018, 2019, 2020
- 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 | 188,308 |
| Roman Catholic | 102,108 |
| Mormon | 7,453 |
| Eastern or Greek Orthodox | 2,541 |
| Jewish | 12,332 |
| Muslim | 2,475 |
| Buddhist | 4,185 |
| Hindu | 1,410 |
| Atheist | 24,985 |
| Agnostic | 28,368 |
| Nothing in Particular | 87,590 |
| Something Else | 32,550 |
| (Missing) | 37,450 |

- Years: 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020
- The response options have stayed largely consistent and follow that of Pew.

Family Status

marstat: Marital Status

“What is your marital status?”

| | n |
|------------------------|---------|
| Married | 288,146 |
| Separated | 8,983 |
| Divorced | 58,312 |
| Widowed | 25,402 |
| Single / Never Married | 124,197 |
| Domestic Partnership | 25,129 |
| (Missing) | 1,586 |

- Years: All of 2006-2020
- 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 | 315,799 |
| Rent | 149,830 |
| Other | 21,084 |
| (Missing) | 45,042 |

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

has_child: Parent of Young Children

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

| | n |
|-----------|---------|
| Yes | 122,375 |
| No | 362,059 |
| (Missing) | 47,321 |

- Years: 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020

no_milstat: Military Status (None)

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

| | n |
|-----------|---------|
| Yes | 219,461 |
| No | 302,200 |
| (Missing) | 10,094 |

- Years: 2006, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020
- 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.

Validations

Observations in even years include indicators for validated voting, which means that YouGov has matched survey respondents' personal identifiable information to public voter files, which in turn officially record whether a person has voted or not. Validation is often completed in the summer following the election. For more information, see [Ansolabehere and Hersh \(2012\)](#).

vv_regstatus: Validated registration status

[Validation results. Missing if validation was not conducted in the year. Categories are aggregated. Both Matched-not registered and unmatched are labeled as a no record.]

| | n |
|---------------------------|---------|
| Active | 218,373 |
| No Record of Registration | 77,521 |
| Unregistered | 15,869 |
| Dropped | 6,607 |
| Inactive | 3,565 |
| Multiple Appearances | 1,600 |
| (Missing) | 208,220 |

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

vv_party_gen: Validated registered party

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

| | n |
|---------------------------------|---------|
| No Record of Party Registration | 79,041 |
| Unknown | 68,895 |
| Democratic Party | 37,600 |
| Republican Party | 29,494 |
| No Party Affiliation | 13,874 |
| Declined to State | 2,376 |
| Other | 1,635 |
| Independent Party | 1,511 |
| Libertarian Party | 537 |
| Green Party | 265 |
| Cns | 44 |
| Constitution Party | 38 |
| Reform Party | 11 |
| Wor | 9 |
| Socialist Party | 5 |
| (Missing) | 296,420 |

- Years: 2012, 2014, 2016, 2018
- 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 | 208,027 |
| Republican Party | 14,486 |
| Democratic Party | 12,783 |
| No Party Affiliation | 16 |
| Libertarian Party | 11 |
| Other | 8 |
| Green Party | 4 |
| (Missing) | 296,420 |

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

Turnout**vv_turnout_gvm: Validated turnout General Election**

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

| | n |
|---------------------|---------|
| Voted | 202,966 |
| No Record of Voting | 155,257 |
| No Voter File | 1,733 |
| (Missing) | 171,799 |

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018
- Limitations: Collapses most response options. For example, the particular voting method is collapsed into one category, even though gvm stands for General Election voting *method*. Also, the result of not matching to a voter file is collapsed with the result of matching to a voter file and having no indication of turning out to vote. The distinction is unclear in earlier years, and is thus collapsed for all years here. For finer distinctions, see the individual year's CCES.

vv_turnout_pvm: Validated turnout Primary Election (Congressional)

[Validation results. Congressional primaries.]

| | n |
|---------------------|---------|
| No Record of Voting | 225,737 |
| Voted | 96,435 |
| No Voter File | 1,363 |
| (Missing) | 208,220 |

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

Partisan Identity

pid3: Partisan identity (3 point)

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

| | n |
|-------------|---------|
| Democrat | 189,911 |
| Republican | 139,064 |
| Independent | 148,207 |
| Other | 21,163 |
| Not Sure | 24,456 |
| (Missing) | 8,954 |

- Years: All of 2006-2020
- 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 | 128,356 |
| Not Very Strong Democrat | 63,487 |
| Lean Democrat | 53,786 |
| Independent | 72,472 |
| Lean Republican | 54,473 |
| Not Very Strong Republican | 50,344 |
| Strong Republican | 89,402 |
| Not Sure | 16,419 |
| (Missing) | 3,016 |

- Years: All of 2006-2020
- 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) | 245,629 |
| Republican (Including Leaners) | 194,219 |
| Independent (Excluding Leaners) | 72,472 |
| Not Sure | 16,419 |
| (Missing) | 3,016 |

- Years: All of 2006-2020

- Limitations: See pid3

ideo5: Ideology (5 point)

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

| | n |
|-------------------|---------|
| Very Liberal | 50,863 |
| Liberal | 93,992 |
| Moderate | 166,063 |
| Conservative | 119,316 |
| Very Conservative | 61,564 |
| Not Sure | 38,143 |
| (Missing) | 1,814 |

- Years: All of 2006-2020

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 | 23,479 |
| 10k - 20k | 39,108 |
| 20k - 30k | 53,657 |
| 30k - 40k | 54,254 |
| 40k - 50k | 48,423 |
| 50k - 60k | 47,190 |
| 60k - 70k | 34,946 |
| 70k - 80k | 38,117 |
| 80k - 100k | 44,287 |
| 100k - 120k | 32,287 |
| 120k - 150k | 26,490 |
| 150k+ | 30,819 |
| Prefer not to say | 57,111 |
| Skipped | 12 |
| (Missing) | 1,575 |

- Years: All of 2006-2020
- 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 | 208,990 |
| Part-Time | 54,391 |
| Temporarily Laid Off | 4,640 |
| Unemployed | 34,754 |
| Retired | 113,241 |
| Permanently Disabled | 31,166 |
| Homemaker | 38,008 |
| Student | 23,590 |
| Other | 12,682 |
| (Missing) | 10,293 |

- Years: 2006, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020

no_healthins: Uninsured

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

| | n |
|-----------|---------|
| Yes | 50,154 |
| No | 402,317 |
| (Missing) | 79,284 |

- Years: 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020
- 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 | 34,336 |
| Yes, Formerly | 99,147 |
| No, Never | 354,261 |
| (Missing) | 44,011 |

- Years: 2006, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020
- 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 | 44,390 |
| Yes, Formerly | 70,116 |
| No, Never | 367,050 |
| Not Sure | 3,576 |
| (Missing) | 46,623 |

- Years: 2006, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020
- 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 | 36,745 |
| Gotten better / somewhat better | 111,201 |
| Stayed about the same | 131,927 |
| Gotten worse / somewhat worse | 133,660 |
| Gotten much worse | 103,991 |
| Not sure | 13,233 |
| (Missing) | 998 |

- Years: All of 2006-2020
- 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 | 266,841 |
| Some of the time | 125,397 |
| Only now and then | 59,403 |
| Hardly at all | 29,944 |
| Don't Know | 13,117 |
| (Missing) | 37,053 |

- Years: 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020
- 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 | 112,590 |
| Approve / Somewhat Approve | 115,740 |
| Disapprove / Somewhat Disapprove | 52,585 |
| Strongly Disapprove | 235,043 |
| Never Heard / Not Sure | 14,524 |
| Neither Approve nor Disapprove | 443 |
| (Missing) | 830 |

- Years: All of 2006-2020
- 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 | 77,888 |
| Approve / Somewhat Approve | 162,016 |
| Disapprove / Somewhat Disapprove | 89,161 |
| Strongly Disapprove | 85,248 |
| Never Heard / Not Sure | 81,756 |
| Never Heard of this Person | 25,762 |
| Neither Approve nor Disapprove | 1,798 |
| (Missing) | 8,126 |

- Years: All of 2006-2020
- 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 | 70,716 |
| Approve / Somewhat Approve | 163,600 |
| Disapprove / Somewhat Disapprove | 102,114 |
| Strongly Disapprove | 108,806 |
| Never Heard / Not Sure | 63,069 |
| Never Heard of this Person | 17,210 |
| Neither Approve nor Disapprove | 1,413 |
| (Missing) | 4,827 |

- Years: All of 2006-2020
- 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 | 77,296 |
| Approve / Somewhat Approve | 157,288 |
| Disapprove / Somewhat Disapprove | 98,264 |
| Strongly Disapprove | 110,725 |
| Never Heard / Not Sure | 63,243 |
| Never Heard of this Person | 18,184 |
| Neither Approve nor Disapprove | 1,158 |
| (Missing) | 5,597 |

- 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 | 84,692 |
| Approve / Somewhat Approve | 163,731 |
| Disapprove / Somewhat Disapprove | 95,758 |
| Strongly Disapprove | 137,241 |
| Never Heard / Not Sure | 46,424 |
| Neither Approve nor Disapprove | 1,414 |
| (Missing) | 2,495 |

- Years: All of 2006-2020
- 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 coalesces both *post-election* wave responses (the bulk of the responses) and pre-election respondents who reported having already voted early. In 2018, it also coalesces the responses to the straight ticket party option (CC18_409), so that those who selected the Republican straight party ticket in the applicable states will appear to have voted for the Republican candidate in all offices. The straight ticket party option was not asked in other years.

The category "Did Not Vote" is recorded only when the respondent selects that option. Respondents who have missing values for intent or vote choice can also be non-voters for a variety of reasons. For general turnout, see the section on turnout.

Presidential Vote

intent_pres_party: President preference party

[Party of presidential candidate chosen in intent_pres]

| | n |
|-----------------|---------|
| Democratic | 87,198 |
| Republican | 72,267 |
| Third Party | 5,412 |
| Independent | 209 |
| Other Candidate | 5,403 |
| (Missing) | 361,266 |

– Years: 2008, 2012, 2016, 2020

voted_pres_party: President vote in last election

[Party of presidential candidate chosen in last election]

| | n |
|-----------------|---------|
| Democratic | 188,806 |
| Republican | 163,040 |
| Other Candidate | 20,206 |
| Did not Vote | 26,041 |
| (Missing) | 133,662 |

– Years: 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020

– 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 |
| I'm Not Sure | 1,697 |
| (Missing) | 501,560 |

- 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 |
| I'm Not Sure | 3,856 |
| (Missing) | 479,512 |

- 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 |
| I'm Not Sure | 6,536 |
| (Missing) | 468,753 |

- 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) | 18,980 |
| Joe Biden (Democrat) | 22,398 |
| Other | 1,390 |
| I Won't Vote in this Election | 2,390 |
| I'm Not Sure | 3,793 |
| (Missing) | 482,804 |

- 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 | 18,227 |
| Not Sure / Don't Recall | 1,787 |
| (Missing) | 365,153 |

- Years: 2008, 2009, 2010, 2011, 2012
- Limitations: Response options differ 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_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 | 2,827 |
| Not Sure / Don't Recall | 1,990 |
| (Missing) | 373,783 |

- Years: 2012, 2013, 2014, 2015, 2016
- Limitations: Response options differ 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 | 80,294 |
| Donald Trump | 67,320 |
| Other / Someone Else | 15,895 |
| Did Not Vote | 21,102 |
| Not Sure / Don't Recall | 516 |
| (Missing) | 346,628 |

- Years: 2016, 2017, 2018, 2019, 2020
- 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 | 26,187 |
| Donald Trump | 17,703 |
| Other / Someone Else | 1,458 |
| I Did Not Vote in this Race | 101 |
| Did Not Vote | 13 |
| Not Sure | 190 |
| (Missing) | 486,103 |

- Years: 2020
- 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] | 147,844 |
| [Republican / Candidate 2] | 131,754 |
| [Other / Candidate 3] | 4,449 |
| \$HouseCand4Name (\$HouseCand4Party) | 45 |
| Other | 2,664 |
| \$HouseCand5Name (\$HouseCand5Party) | 24 |
| I Won't Vote in this Election | 2,269 |
| \$HouseCand6Name (\$HouseCand6Party) | 45 |
| \$HouseCand9Name (\$HouseCand9Party) | 2 |
| \$HouseCand7Name (\$HouseCand7Party) | 35 |
| \$HouseCand8Name (\$HouseCand8Party) | 22 |
| \$HouseCand10Name (\$HouseCand10Party) | 1 |
| \$HouseCand11Name (\$HouseCand11Party) | 3 |
| I'm Not Sure | 79,615 |
| No One | 22,730 |
| (Missing) | 140,253 |

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020
- 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] | 108,119 |
| [Republican / Candidate 2] | 91,863 |
| [Other / Candidate 3] | 4,477 |
| \$SenCand4Name (\$SenCand4Party) | 19 |
| Other | 1,958 |
| I'm Not Sure | 41,844 |
| No One | 14,116 |
| I Won't Vote in this Election | 1,145 |
| (Missing) | 268,214 |

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020
- 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] | 77,357 |
| [Republican / Candidate 2] | 68,938 |
| [Other / Candidate 3] | 4,055 |
| Other | 1,529 |
| I'm Not Sure | 25,159 |
| No One | 8,349 |
| I Won't Vote in this Election | 466 |
| (Missing) | 345,902 |

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020
- 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] | 141,884 |
| [Republican / Candidate 2] | 128,606 |
| [Other / Candidate 3] | 2,819 |
| <code>\$HouseCand4Name (\$HouseCand4Party)</code> | 34 |
| <code>\$HouseCand7Name (\$HouseCand7Party)</code> | 34 |
| Other | 3,700 |
| I Did Not Vote In This Race | 12,361 |
| <code>\$HouseCand5Name (\$HouseCand5Party)</code> | 27 |
| Not Sure | 5,267 |
| <code>\$HouseCand6Name (\$HouseCand6Party)</code> | 44 |
| I Did Not Vote in this Race | 1,078 |
| I Did Not Vote | 260 |
| <code>\$HouseCand8Name (\$HouseCand8Party)</code> | 16 |
| <code>\$HouseCand9Name (\$HouseCand9Party)</code> | 2 |
| <code>\$HouseCand10Name (\$HouseCand10Party)</code> | 2 |
| <code>\$HouseCand11Name (\$HouseCand11Party)</code> | 3 |
| (Missing) | 235,618 |

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020
- 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] | 99,697 |
| [Republican / Candidate 2] | 87,238 |
| [Other / Candidate 3] | 2,963 |
| Other | 2,316 |
| Not Sure | 2,280 |
| \$SenCand4Name (\$SenCand4Party) | 11 |
| I Did Not Vote In This Race | 5,001 |
| (Missing) | 332,249 |

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020
- 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] | 66,407 |
| [Republican / Candidate 2] | 61,951 |
| [Other / Candidate 3] | 2,613 |
| I Did Not Vote in this Race | 39 |
| I Did Not Vote | 17 |
| Other | 1,861 |
| I Did Not Vote In This Race | 3,576 |
| Not Sure | 1,128 |
| (Missing) | 394,163 |

- Years: 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020
- 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: 531,755

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: 531,755

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

The matching of identifiers to respondent occurs through matching by district, by district and last name, or both:

- For House representatives, we join on `cong`, `st`, and `dist` to a NOMINATE database that only consists of unique observations according to the key. For duplicates with regards to these three variables (e.g., in the rare case where a new representative comes into office mid-session), we match on `cong`, `st`, `dist` and last name.
- For Senators, we join entirely on `cong`, `st`, and last name

Rows: 531,755

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 2006-2020

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

The unique identifiers can be used to join with other databases to append additional information such as committee membership and ideology scores, such as

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet (2017). Voteview: Congressional Roll-Call Votes Database. <https://voteview.com/>

The text responses that the respondent chose in each of the `intent_ / voted_` questions, if the respondent was a candidate. For example, respondent with `case_id = 163051575` in the 2012 CCES chose the first option in the House representative preference question. `intent_rep_chosen` shows that this particular respondent preferred voting for Maxine Waters, one of the two Democrats in the race.

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

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

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

Name of Chosen Candidate

Rows: 531,755

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
- 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: 531,755

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

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

Version 6.0

- 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)