Candidates in American General Elections

Jeremiah Cha[1], Shiro Kuriwaki[2], and James M. Snyder, Jr.[1]

[1]Department of Government, Harvard University.
[2]Department of Political Science and Institution of Social and Policy Studies, Yale University

Version 3.0 (October 2024)

Please cite as:
Cha, Jeremiah, Shiro Kuriwaki, and James M. Snyder, Jr., 2024, "Candidates in American General Elections", Harvard Dataverse. DOI: 10.7910/DVN/DGDRDT.

Between 2006 and 2022, there were over 4,000 races for Congress and Governor in the general election, in which 8,738 Democratic and Republican candidates ran for office. While many existing efforts compile these election results, few include standardized candidate names across years and offices, candidate incumbency status, and elections for different offices and years combined into a single file.

We provide a candidate-level dataset that is comprehensive for Presidential, Congressional and gubernatorial contests between 2006 and 2022 with standardized names for each unique candidate. We also record party, incumbency, vote totals, and election results. For example:

year	office	state	dist	party	name_snyder	inc	candidatevotes	won
2016	H	MO	5	D	CLEAVER, EMANUEL, II	1	190,766	1
2016	H	MO	5	R	TURK, JACOB	0	123,771	0
2016	H	MO	5	Lbt	WELBORN, ROY	0	9,733	0
2008	H	MO	7	D	MONROE, RICHARD	0	91,010	0
2008	H	MO	7	R	BLUNT, ROY D.	1	219,016	1
2008	H	MO	7	Lbt	CRAIG, KEVIN	0	6,971	0
2008	H	MO	7	Const	MADDOX, TRAVIS	0	6,166	0
2016	S	MO	3	D	KANDER, JASON	0	1,300,200	0
2016	S	MO	3	R	BLUNT, ROY D.	1	1,378,458	1
2016	S	MO	3	Const	RYMAN, FRED	0	25,407	0
2016	S	MO	3	Grn	MCFARLAND, JOHNATHAN	0	30,743	0
2016	S	MO	3	Lbt	DINE, JONATHAN	0	67,738	0

1 Data Sources and Coverage

Most of the candidate-level data – names and incumbency status – is a subset of the data collected and manually coded by James M. Snyder, Jr. Vote counts and candidate listings of Congress are entered from the Official Election Statistics compiled by the House Clerk (https://history.house.gov/Institution/Election-Statistics/Election-Statistics/).

Election results for the state office of the Governor are collected from statements of votes by the Secretary of State of each state. We include the odd-year elections in Virginia, New Jersey, Kentucky, Mississippi, and Louisiana. Recall elections (Wisconsin 2012, California 2021) are not included.

Presidential vote counts are taken from the data compiled by the MIT Election Data Science Lab (https://doi.org/10.7910/DVN/42MVDX), which in turn mostly sources from the Official Election Statistics by the House Clerk.

Incumbency is determined by a programmatic count of whether the name is repeated across years, and by manual inspection of the biography of each candidate.

This dataset excludes candidates who receive less than ten votes in their general election and those who identify Blank or Miscellaneous as their party. We also exclude overvotes, undervotes, and invalid votes. Although these removals make the `totalvotes` be a few votes less than the actual turnout, we made this choice uniformly because states vary in their reporting protocol for such votes. That is, given that some states report overvotes while others exclude them entirely in their official counts, it is more internally consistent to exclude them for all states.

Our code for recoding and editing the data is open-source in a Github repository (https://github.com/kuriwaki/cces_candidates).

2 Unique Identifiers

To identify a single election, group by `year`, `state`, `office`, `dist`, and `type` (General vs. Special). To isolate the set of candidates who won a seat in an election cycle, group by subset by `year`, `state`, `office`, and `type`, then subset by `won`.

To identify a single candidate across office and time, group by `state` and `name_snyder`. As we discuss in detail in this documentation, the candidate names are standardized (by Snyder's coding) so that the same candidate has the same spelling across years and office. In very rare instances, a candidate for a major party runs for office two different states. For example, Mitt Romney ran for Governor of Massachusetts, President, and U.S. Senator from Utah. For these candidates, group by `name_snyder`.

3 Variable Descriptions and Tabulations

`office`

The office the candidate is running for. Following the Snyder data, we use `H` for the US House of Representatives, `S` for the US Senate, and `G` for Governor.

year	office			
	P	S	H	G
2006	0	138	1,120	156
2007	0	0	17	5
2008	304	115	1,164	37
2009	0	0	16	14
2010	0	180	1,324	178
2011	0	0	13	20
2012	273	132	1,197	39
2013	0	11	21	11
2014	0	139	1,060	130
2015	0	0	7	8
2016	277	148	1,135	45
2017	0	2	21	10
2018	0	141	1,130	160
2019	0	0	10	9
2020	301	155	1,155	44
2021	0	0	25	8
2022	0	174	1,147	135
2023	0	0	0	20

`dist`

The district the candidate is running for. U.S. House races in at-large districts are numbered 1 instead of 0 or "-AL".

For U.S. Senate candidates, the district variable denotes the class of the particular Senate seat (e.g. Dianne Feinstein has a `dist` value of 1 since her seat is part of Senate Class I).

Office	Minimum	Maximum	Distinct Values
H	1	53	53
S	1	3	3
G			1
P			1

type

The type of election the candidate is running in. We use `G` to denote general election candidates and `S` to denote special election candidates. Candidates are said to run in special elections if they are running to serve out the term of another legislator.

Senate

year	type	
	G	S
2006	138	0
2008	111	4
2010	162	18
2012	132	0
2013	0	11
2014	130	9
2016	148	0
2017	0	2
2018	133	8
2020	135	20
2022	170	4

House

year	type	
	G	S
2006	1,111	9
2007	0	17
2008	1,138	26
2009	0	16
2010	1,298	26
2011	0	13
2012	1,179	18
2013	0	21
2014	1,044	16
2015	0	7
2016	1,128	7
2017	0	21
2018	1,122	8
2019	0	10
2020	1,153	2
2021	0	25
2022	1,125	22

Governor

year	type	
	G	S
2006	156	0
2007	5	0
2008	37	0
2009	14	0
2010	174	4
2011	15	5
2012	39	0
2013	11	0
2014	130	0
2015	8	0
2016	45	0
2017	10	0
2018	160	0
2019	9	0
2020	44	0
2021	8	0
2022	135	0
2023	20	0

runoff

The candidate's election round. In some states (e.g. Louisiana), if no candidate receives over 50 percent of the vote, the top two candidates must run in a runoff election again to determine the winner. In such a case, *only the deciding contest is included*. For example, in the 2020 Georgia Senate elections, both the general and special election went to a runoff because no candidate got over 50 percent of the vote in either race. Therefore, we exclude the November election results and include the January runoff. A runoff variable with a value of 1 denotes that the election the candidate is running in is a runoff election, while a 0 denotes that the election is not. A missing value is given for the vast majority of states in which no such runoff option is available, and a candidate can win by plurality.

Senate

year	runoff	
	0	1
2006	0	0
2008	5	2
2010	15	0
2012	0	0
2013	0	0
2014	3	2
2016	3	2
2017	0	0
2018	0	0
2020	15	4
2022	14	2

House

year	runoff	
	0	1
2006	35	4
2007	0	2
2008	52	0
2009	0	0
2010	40	2
2011	0	0
2012	30	2
2013	0	2
2014	25	4
2015	0	0
2016	29	2
2017	0	2
2018	54	0
2019	0	0
2020	52	2
2021	4	0
2022	51	0

Governor

year	runoff	
	0	1
2006	3	0
2007	1	0
2008	0	0
2009	0	0
2010	3	0
2011	1	0
2012	0	0
2013	0	0
2014	3	0
2015	2	0
2016	0	0
2017	0	0
2018	3	0
2019	4	2
2020	0	0
2021	0	0
2022	0	0
2023	0	0

`name_snyder`

Standardized candidate name following Snyder coding rules. The specifics of these coding rules are important for overtime tracking because **name and state uniquely identify a unique person** across all his Congressional candidate data (at least 1950 to the present). The syntax is

`[Last name], [First Name] [Middle name] ([Nickname]), [JR./SR./I/II/III]`

For example:

- `SEWELL, TERRYCINA ANDREA (TERRI)`: commonly known as Terri Sewell (AL)
- `GRASSLEY, CHARLES ERNEST (CHUCK)`: commonly known as Chuck Grassley (IA)
- `CORNYN, JOHN, III`: commonly known as John Cornyn (TX)
- `KENNEDY, JOSEPH P. (JOE), III`: commonly known as Joe Kennedy (MA)
- `KELLY, GEORGE J. (MIKE), JR.`: commonly known as Mike Kelly (PA). Note the use of periods for abbreviation and comma before the JR.
- `WASSERMAN SCHULTZ, DEBBIE`: note the last name is not hyphenated and is two words

- `JACKSON LEE, SHEILA`: note that the last name is not Lee, so we use this instead of "Lee, Sheila Jackson"

Note that these names are *different* from other names (1) on the ballot, (2) on the Clerk of the House report, or the (3) FEC. Any of these sources can disagree on the formatting of the names or whether to abbreviate a middle name. The same candidate in the same election year may even have different formats in each of these sources.

The Snyder formalization solves some shortcomings in the official data above. Candidates may report different style names on the ballot vs. the FEC or they may have multiple FEC IDs at the same time. More commonly, a House representative who then runs for higher office like the Senate will often get a separate FEC ID, even though they are the same person. In theses cases, the Snyder formalization of identifying people by their full name and state is the most reliable.

Therefore, it is important in our formalization to spell out nicknames and distinguish between hyphenated last names to the extent that it uniquely identifies a person within a state. The same person's name should be formatted exactly the same way, and conversely, two different people with near-identical last names should somehow be given different names. For example, if a John Smith ran for Congress in 1980 in Ohio and then a *separate, unrelated* John Smith ran again in Ohio in 2020. To distinguish the two different people, we would need to find something like a middle name or suffix to distinguish the name in characters.

One shortcoming of the Snyder formalization is that it may be too rigid. Whether or not something is abbreviated can be determined on a coding rule decided early on. If a person's name changes, e.g. through marriage or divorce, the old name must be kept. For example:

- `BIDEN, JOSEPH R., JR.` has his middle name abbreviated, even though it is known (Robinette).
- `MACK, MARY BONO` (R-CA, 1998-2012) married to Connie Mack IV (R-FL, 2001-2003) during 2007-2013. They separated in 2013 and she now goes by Mary Bono, but in the data she will be coded by the above name.

party

The short party affiliation of the candidate. We use the "short" or colloquial party name. For example, the Democrat-Farmer-Labor Party in Minnesota is given a `D` instead of `DFL`. Candidates who ran on third party tickets in Connecticut and New York are simply given the major party name.

Senate

year	party				
	D	R	Lbt	Grn	Oth
2006	31	33	15	10	49
2008	35	34	13	5	28
2010	37	38	19	10	76
2012	32	33	13	6	48
2013	2	2	0	0	7
2014	34	36	20	6	43
2016	35	33	20	8	52
2017	1	1	0	0	0
2018	36	35	17	6	47
2020	38	36	17	4	60
2022	38	38	21	3	74

House

year	party				
	D	R	Lbt	Grn	Oth
2006	427	393	111	37	152
2007	4	6	1	1	5
2008	430	402	128	52	152
2009	5	5	1	2	3
2010	420	441	157	51	255
2011	4	4	0	1	4
2012	422	425	134	57	159
2013	5	7	1	2	6
2014	411	403	102	31	113
2015	3	3	0	1	0
2016	421	408	116	54	136
2017	7	5	4	1	4
2018	445	402	113	30	140
2019	3	3	2	1	1
2020	438	420	117	17	163
2021	6	15	1	0	3
2022	427	438	86	5	191

Governor

year	party				
	D	R	Lbt	Grn	Oth
2006	36	36	14	15	55
2007	2	3	0	0	0
2008	11	11	5	0	10
2009	2	2	0	0	10
2010	37	37	19	12	73
2011	7	4	1	0	8
2012	11	11	9	1	7
2013	2	2	2	0	5
2014	35	36	17	5	37
2015	3	3	0	0	2
2016	12	12	10	2	9
2017	2	2	2	1	3
2018	36	36	22	8	58
2019	3	3	1	0	2
2020	10	11	9	1	13
2021	2	2	2	1	1
2022	36	37	24	0	38
2023	4	8	0	0	8

party_formal

The formal name of the candidate's party affiliation, for example DFL in Minnesota. Third party candidate names follow what is given in the House Clerk document or state SOS reports (for Governors), with some minor abbreviations and standardizations.

party-formal	Office		
	G	H	S
D	238	3,598	302
R	250	3,603	312
Lbt	139	1,075	155
Grn	46	342	58
I	148	613	182
	1	0	0
12 Visions Pty	0	0	1
A New Direction	0	1	0
A-Bushist Cand	0	0	1
Abundant Amer	0	1	0
ACL	1	0	0
Action No Talk	0	1	0
ADB	0	1	0
Agent of Change	0	1	0
AK Indep	2	1	2
Allen 4 Cong	0	1	0
Alliance	0	1	0
Aloha Aina	0	0	1
Aloha Aina Pty	0	1	0
Amer Cong Pty	0	1	0
Amer Elect	1	0	0
Amer Indep	2	11	2
Amer Labor	0	1	0
Amer Pty of SC	1	8	2
Amer Renaiss Move	0	1	0
Amer Shopping Pty	0	1	0
Amer Third Posit	1	0	0
Amer Values	0	1	0
America First	0	0	1
American Values	0	1	0
Amers Elect	0	2	0
An Indep Voice	0	1	0
Anti-Prohib	1	0	1
Anti-Prohib,Lbt	0	0	1
Approval Voting	0	0	1
Approval Voting Pty	0	1	1
BBA	0	1	0
Be Determined	0	1	0
BFS	0	0	1
Blue Enigma	1	1	0
Bob's for Jobs	0	1	0
Bring Home Troops	0	1	0
Building Your Legacy	0	1	0
Bully Breaks Hearts	0	1	0
By Petition	0	1	0
C	5	43	3
C4C 2018	0	1	0
Cannon Fire	0	1	0
Cannot Be Bought	0	1	0
Change Change Change	0	1	0
Change is Needed	0	1	0
Check This Column	0	1	0
Cit Legisl	0	1	0
Clear Water Pty	1	0	0
CO Center P	0	1	0
Coal on Gov Reform	0	1	0
Common Sense	1	2	1
Communist	0	1	0
Compassionate C	0	1	0
Concerned Cit	1	0	1
Conservative	0	2	0
Const	15	96	29
Const C	0	1	0
Constitution	4	0	0
CSV	1	0	0
CT for Lieberman	0	0	1
CTRY	0	0	2
D-R	0	2	1
D, Griebel Frank, Wk Fam	1	0	0
D, I, Reform, Wk Fam, Women's Equality	0	1	0
D, I, Wk Fam, Women's Equality	1	3	1
D, Independence	0	1	0
D, Independence, Wk Fam	0	6	0
D, Moderate	0	1	0
D, Reform, Wk Fam	0	2	0
D, Reform, Wk Fam, Women's Equality	0	1	0
D, Save America Movement, Wk Fam	0	1	0
D, Wk Fam	2	43	3
D, Wk Fam, Women's Equality	0	10	0
D, Women's Equality	0	1	0
D,C,Indep,WF	0	1	0
D,I,WF	0	1	0
D,IDP,WF	0	1	0
D,Indep,WF	1	10	1
D,Progressive	1	0	0
D,WF	0	90	2
D,WF,IDP	0	7	0
D,WF,Indep	2	6	2
D,Wk Fam	1	21	1
Defend Amer Const	0	1	0
DFL	5	73	7
Dvrsity is Strngth	0	1	0
Earth Rights	1	0	0
Ecology Dem	1	0	0
Econ Growth	0	1	2
Ed the Barber	0	1	0
Educ Community Law	0	1	0
Energy Indep	0	1	0
ETP	0	1	0
Financial Indep	0	1	0
FL Whig Pty	0	1	0
Flrsh Every Person	0	1	0
For Amers	0	1	0
For People	0	1	0
For Polit Revolut	0	1	0
For the People	0	1	1
Free Energy	1	0	0
Freedom	1	1	0
Freedom, Respons, Action	0	1	0
FTP	0	1	0
Future Vision	0	1	0
Glass-Steagall Now	1	0	0
God We Trust	0	0	2
Going Away P	0	1	0
Gov Save Amer	0	1	0
GPP	0	1	0
Grassroots	2	0	1
Grassroots Cannabis	0	1	0
Grassroots-Cannabis	0	1	0
Grassroots-Legal Cannabis	0	4	1
Grav Buoy Solution	0	1	0
Green	4	0	0
Green Tea Patriots	0	1	0
Grn-I	1	0	0
Grn-Rainbow	1	0	0
Grn,Pg	0	3	1
Haris Bhatti Pty	0	1	0
HFC	0	1	0

Honest, Integ, Compassion	0	1	0
I Cit Const Gov	0	1	0
I Grn	0	13	2
I No War No Bail	0	1	0
I Pg	0	2	0
I Reform	0	1	1
I,Lbt	0	1	0
I,Vote Pop Chng	0	1	0
IAP	5	32	10
Imp	0	1	0
Indep Const Cand	0	1	0
Indep for Lbty	0	0	1
Indep Pg	0	1	0
Indep Pty of DE	0	1	1
Independence	0	2	0
Independence-Alliance	1	0	0
Independent	27	0	0
Independent Party of Delaware	1	0	0
Indpendent	2	0	0
Iowa Pty	1	0	0
Itkis Campaign	0	1	0
Jersey Strong Ind	0	0	1
Jobs & Justice	0	1	0
Justic,Peace,Secur	0	1	0
Justice Mercy Humility	0	1	0
Keystone	1	0	0
Keystone P	0	0	1
Labor	0	1	0
Labour P	0	1	0
LaRouche	0	0	1
LaRouche Was Right	0	0	1
Lbt, Serve America Movement	0	1	0
Lbt,Const	0	0	1
Legal Marijuana Now	0	14	3
Legalize Cannabis	1	0	0
Legalize Marijuana Now	1	0	0
LFC	0	1	0
Liberty	0	1	0
Liberty Caucus	0	1	0
Liberty Union	7	6	4
Mahali P	0	1	0
Make Change Happen	0	1	0
Make Govt Work	0	1	0
Make It Simple	0	0	1
marklovett.us	0	1	0
Medical Freedom	0	1	0
MN Open Pg	0	0	1
Moderate	1	2	0
Moderate Choice	0	1	0
Mountain	4	1	5
Move Ever Forward	0	1	0
Mr Smith to Wash	0	1	0
Natural Law	2	5	3
Nebraska	1	0	1
Never Give Up	0	1	0
New Beginnings	0	1	0
New Day NJ	0	0	1
New Indep Pty	1	0	1
New Way Forward	0	1	0
NJ First	0	1	0
no affiliation	0	2	0
No Medical Neglig	0	1	0
No Slogan	0	1	0
No Slogan Filed	0	2	0
None of Them	0	1	0
Nonpartisan	1	1	0
NOPTY	0	0	1
Not For Sale	0	1	0
NP Delaware	0	1	0
NPA	8	26	19
NSA DID 911	1	1	0
Of the People	0	1	0
Of, By, For!	0	0	1
Ohter	0	1	0
Opp Cng Gridlock	0	1	0
Oth	0	1	0
Other	4	15	2
Overthrow Incumbs	0	1	0
P&F	3	22	2
Pacific Green	1	15	2
Pacific Grn	1	5	1
Patriot Movement	0	1	0
Peace & Prosp	0	0	1
People's Agenda	0	1	0
People's Choice Pty	0	1	0
Peoples Indep Pg	0	1	0
Personal Choice	0	0	1
Petition	1	0	1
Pirate	0	1	0
PNW	0	1	0
Pols are Crooks	0	2	0
Poor People Campgn	0	0	1
Populist	1	1	0
PPC	0	0	1
Preserve Green Space	0	1	0
Pro Life Cons	0	1	0
Prog	0	1	1
Progress With Pam	0	1	0
Progressive	3	5	1
Pty Free	0	1	0
Quast Term Limits	0	0	1
Quit Raising Taxes	1	0	0
R, C	1	16	0
R, C, I	0	2	0
R, C, I, Reform	0	6	0
R, C, Independence	0	4	0
R, C, Lbt, Save Our City	0	1	0
R, C, Lbt, Serve America Movement	0	1	0
R, C, Reform	1	5	1
R, C, Serve America Movement	0	1	0
R, Conservative	0	19	1
R, Conservative, Medical Freedom	0	1	0
R, Conservative, Parent	0	1	0
R, Const, Independence	0	1	0
R, I	1	5	0
R, Reform	0	1	0
R,C	1	43	3
R,C,I	0	1	0
R,C,Indep	0	32	0
R,C,Lbt	0	2	0
R,C,SCC	1	0	0
R,C,Taxp	1	0	1
R,CR	0	1	0
R,CRV	0	15	0
R,CRV,IDP	0	5	0
R,I	0	5	1
R,IDP	0	1	0
R,Indep	0	3	0
R,Lbt	0	1	0
R,Tax	0	1	0

R,Tax,C,Indep	0	2	0
R,Taxp	0	1	0
Reduce Property Tax	1	0	0
Reform	5	33	6
Reform Nation	0	0	1
Renewable Energy	1	0	0
Rent 2 Damn High	2	0	1
Represent the 99%	0	1	0
Resource	1	0	0
Respbl Fair Integ	0	0	1
Restor Amer Prom	0	1	0
Rfm	0	1	0
RIN	0	1	0
RTB	0	1	0
Sapient	1	0	0
Save America Movement	1	0	0
Serve America Movement	0	3	0
Soc	0	1	4
Soc Action	0	1	0
Socialist Equality	0	0	1
Socialist Workers	1	0	0
Solid Defend Life	0	0	1
Stop Boss Politics	0	2	0
SUS	0	2	0
SWP	3	14	3
Tax Equity Rebell	1	0	0
Tax Revolt	0	1	1
Tax Revolt I	0	1	0
Taxpayers	1	0	0
Tea Pty	0	5	0
Tea Pty of NV	0	0	1
Teddy Roosevelt Pg	0	1	0
The Inclusion Candidate	0	1	0
THI	0	1	0
Time for Change	0	1	0
Time For Truth	0	2	0
TLP	0	1	0
Together We Stand	0	1	0
Totally Ind Cand	0	0	1
Towne for Cong	0	1	0
Trade, Health, Environm	0	1	0
Transp Govt Pty	0	1	0
Trump Conservative	0	1	0
Truth & Merit	0	1	0
Truth Matters	1	0	0
Truth Vision Hope	0	3	0
Unaffiliated	3	6	0
UNI	0	1	0
United Citizens	2	1	0
United Utah	0	6	0
Unity	1	0	0
Unity is Strength	0	1	0
Unity P of CO	0	2	1
UPA	1	8	3
Upstate Jobs	0	1	0
US Marij	2	0	3
US Taxpayers	3	43	6
UT Justice Pty	0	0	1
Veteran for Change	0	1	0
Vets Pty of Amer	0	1	0
Vote Pop Chng	0	1	0
VoteKISS	0	1	1
VT Grn	1	1	1
VT Localist	1	0	0
W-I	15	165	90
W-I (C)	2	0	0
W-I (CO Center)	0	1	0
W-I (Comm Sense Mod)	0	1	0
W-I (Const)	0	1	1
W-I (D)	1	14	11
W-I (Grn)	0	4	0
W-I (I)	0	9	7
W-I (Indep R Pty)	0	0	1
W-I (Lbt)	1	2	1
W-I (Logic P)	0	0	1
W-I (None)	0	0	1
W-I (NP)	0	6	1
W-I (R)	1	11	9
W-I (Soc)	0	0	1
W-I (Unity Pty)	0	0	1
We Deserve Better	0	4	0
We the people	1	0	0
We the People	0	5	0
Willie Wilson Pty	0	0	1
Withdr Troops Now	0	1	0
Wk Fam	0	3	2
Wk Fam, Women's Equality	0	1	0
Women of Power	0	1	0
Women's Equality	0	1	0
Working Class	1	20	0
Your Country Again	0	1	0
your voice hard	0	1	0

inc

The incumbency status of the candidate. A 0 means the candidate is not an incumbent, a 1 means the candidate is an incumbent, a 2 means the candidate is an incumbent that was elected in a special election, and 3 indicates a U.S. Senator who was appointed by the Governor.

- District numberings do not matter for determining incumbency. For example, Conor Lamb who won a special election to PA-18 in March 2018 is still an incumbent when he ran in November 2018 for the re-drawn PA-17.
- This may differ from **the CES incumbency variable** in some redistricting cases when two incumbents were forced to run against each other. For example, in 2012 Betty Sutton (OH-13) and Jim Renacci (OH-16), both House incumbents, were forced to run in OH-16. While the CES incumbency variable (HouseCandIncumbent) lists only Sutton as the incumbent, this dataset will list both Sutton and Renacci as incumbents.
- If a special election and general election are held on the same day (e.g., NY-29's House election in 2010), the special election results on incumbency (inc = 2) do not carry over to the general election incumbency status.

Senate

year	inc			
	0	1	2	3
2006	109	28	0	1
2008	85	28	0	2
2010	158	20	0	2
2012	111	18	3	0
2013	11	0	0	0
2014	112	25	1	1
2016	119	28	1	0
2017	2	0	0	0
2018	109	30	0	2
2020	126	26	1	2
2022	145	27	2	0

House

year	inc		
	0	1	2
2006	718	398	4
2007	17	0	0
2008	766	390	8
2009	16	0	0
2010	931	384	9
2011	13	0	0
2012	819	374	4
2013	21	0	0
2014	673	380	7
2015	7	0	0
2016	746	384	5
2017	21	0	0
2018	758	362	10
2019	10	0	0
2020	775	376	4
2021	25	0	0
2022	779	356	12

Governor

year	inc			
	0	1	2	3
2006	130	23	1	2
2007	3	2	0	0
2008	29	8	0	0
2009	13	1	0	0
2010	165	10	0	3
2011	18	2	0	0
2012	34	4	1	0
2013	10	1	0	0
2014	102	28	0	0
2015	7	1	0	0
2016	40	4	0	1
2017	10	0	0	0
2018	142	15	1	2
2019	7	2	0	0
2020	35	8	0	1
2021	6	2	0	0
2022	108	26	0	1
2023	18	2	0	0

nextup

The year that an incumbent candidate is up for re-election. This variable can be useful in determining tenure in office for special election candidates, who may not have served a full term. See `type` for designation of special election candidates.

Senate

year	nextup							
	2012	2014	2016	2018	2020	2022	2026	2028
2006	138	0	0	0	0	0	0	0
2008	4	111	0	0	0	0	0	0
2010	14	4	162	0	0	0	0	0
2012	0	0	0	132	0	0	0	0
2013	0	11	0	0	0	0	0	0
2014	0	0	11	0	128	0	0	0
2016	0	0	0	0	0	148	0	0
2017	0	0	0	2	0	0	0	0
2018	0	0	0	0	4	0	0	0
2020	0	0	0	0	0	2	135	0
2022	0	0	0	0	0	0	4	170

House

year	nextup									
	2006	2008	2010	2012	2014	2016	2018	2020	2022	2024
2006	9	1,111	0	0	0	0	0	0	0	0
2007	0	17	0	0	0	0	0	0	0	0
2008	0	26	1,138	0	0	0	0	0	0	0
2009	0	0	16	0	0	0	0	0	0	0
2010	0	0	26	1,298	0	0	0	0	0	0
2011	0	0	0	13	0	0	0	0	0	0
2012	0	0	0	18	1,179	0	0	0	0	0
2013	0	0	0	0	21	0	0	0	0	0
2014	0	0	0	0	16	1,044	0	0	0	0
2015	0	0	0	0	0	7	0	0	0	0
2016	0	0	0	0	0	7	1,128	0	0	0
2017	0	0	0	0	0	0	21	0	0	0
2018	0	0	0	0	0	0	8	1,122	0	0
2019	0	0	0	0	0	0	0	10	0	0
2020	0	0	0	0	0	0	0	2	1,153	0
2021	0	0	0	0	0	0	0	0	25	0
2022	0	0	0	0	0	0	0	0	22	1,125

Governor

(Varies, with NJ, VA, MS, LA, KY having odd year elections and NH, VT having two year terms)

candidatevotes

The number of total votes the candidate received.

- For candidates running on multiple party tickets, this will be the *sum* of all of their votes. For example, in 2016, Rep. Rosa L. DeLauro (CT-03) ran as a Democrat and also ran as a Working Families Party candidate. She won 192,274 votes in the former and 21,298 votes in the latter, so her `candidatevotes` is the total, 213,572.
- Florida and Oklahoma do not report the vote count for a House candidate if she won unopposed. In these cases, we have the vote count as `NA` but have the candidate winning (`won == 1`).

Therefore, the data for such corner cases look like:

year	office	state	dist	party	party_formal	name_snyder	candidatevotes
2016	H	CT	3	D	D, Wk Fam	DELAURO, ROSA L.	213572
2016	H	OK	1	R	R	BRIDENSTINE, JAMES FREDERICK (JIM)	NA
2018	H	FL	10	D	D	DEMINGS, VALDEZ B. (VAL)	NA

The total distribution for a candidate's votes looks as follows:

Office	Minimum	10th Quantile	25th	50th	75th	90th	Maximum
H	1	3,356	14,061	93,987	148,864	196,388	387,109
S	2	1,186	9,372	88,246	688,350	1,563,543	7,864,624
G	14	2,548	10,277	101,868	588,529	1,469,642	7,721,410
P	24	765	3,106	19,162	416,495	1,611,998	11,110,250

totalvotes

Total votes for all candidates in each election. This is the sum of `candidatevotes`. It does not include undervotes, overvotes, scattered votes or votes for candidates with less than 10 votes. Also note that only deciding elections are included in the data (see `runoff`). Therefore for runoff elections, `totalvotes` reflects the total turnout in the second round, which often has lower turnout than the November election.

won

Candidate won the general election. A 1 means the candidate won the general election, and a 0 means the candidate lost the general election.

Presidential candidates win electoral college votes at the state level, but only win office with a majority of the college votes. Because of this ambiguity, we have left this variable missing for Presidential candidates.

Senate

year	won	
	0	1
2006	105	33
2008	80	35
2010	142	38
2012	99	33
2013	9	2
2014	103	36
2016	114	34
2017	1	1
2018	106	35
2020	120	35
2022	139	35

House

year	won	
	0	1
2006	683	437
2007	12	5
2008	721	443
2009	11	5
2010	883	441
2011	9	4
2012	756	441
2013	15	6
2014	620	440
2015	4	3
2016	697	438
2017	15	6
2018	690	437
2019	7	3
2020	719	436
2021	18	7
2022	704	443

Governor

year	won	
	0	1
2006	120	36
2007	2	3
2008	26	11
2009	12	2
2010	141	37
2011	16	4
2012	28	11
2013	9	2
2014	94	36
2015	5	3
2016	33	12
2017	8	2
2018	124	36
2019	6	3
2020	33	11
2021	6	2
2022	99	36
2023	17	3

The above table has more than 435 candidates winning in the House each year because they count special elections held that year too. Here are the counts for only general elections. For 2018, the House winners are not 435 because we coded the results in NC-09 to be missing, due to a voter fraud case where the election was not certified.

Senate

year	won (generals)	
	0	1
2006	105	33
2008	78	33
2010	128	34
2012	99	33
2014	97	33
2016	114	34
2018	100	33
2020	102	33
2022	136	34

House

year	won (generals)	
	0	1
2006	676	435
2008	703	435
2010	863	435
2012	744	435
2014	609	435
2016	693	435
2018	685	434
2020	718	435
2022	690	435

Governor

year	won (generals)	
	0	1
2006	120	36
2007	2	3
2008	26	11
2009	12	2
2010	138	36
2011	12	3
2012	28	11
2013	9	2
2014	94	36
2015	5	3
2016	33	12
2017	8	2
2018	124	36
2019	6	3
2020	33	11
2021	6	2
2022	99	36
2023	17	3

The set of tables below show the *party* of the candidate that *won*.

Senate

year	party (winners)		
	D	R	Oth
2006	22	9	2
2008	20	15	0
2010	13	24	1
2012	23	8	2
2013	2	0	0
2014	12	24	0
2016	12	22	0
2017	1	0	0
2018	22	11	2
2020	15	20	0
2022	15	20	0

House

year	party (winners)	
	D	R
2006	233	204
2007	2	3
2008	264	179
2009	5	0
2010	195	246
2011	2	2
2012	206	235
2013	2	4
2014	190	250
2015	1	2
2016	195	243
2017	1	5
2018	236	201
2019	0	3
2020	222	214
2021	3	4
2022	215	228

Governor

year	party (winners)		
	D	R	Oth
2006	20	16	0
2007	1	2	0
2008	7	4	0
2009	0	2	0
2010	13	23	1
2011	2	2	0
2012	7	4	0
2013	1	1	0
2014	11	24	1
2015	1	2	0
2016	6	6	0
2017	2	0	0
2018	16	20	0
2019	2	1	0
2020	3	8	0
2021	1	1	0
2022	18	18	0
2023	1	2	0

The NC-09 House race in 2018 is listed as having no winner because it was later overturned. A special election was held in 2019.

4 Related Work and Extensions

The MIT Election Data Science Lab (https://dataverse.harvard.edu/dataverse/medsl_election_returns) produces similar election results. Their data goes back as far as 1976. However, the data does not include incumbency status, and the names are taken as officially reported rather than standardized to match across years.

Our dataset does not contain information about the gender or race of the candidates. We hope to work collect that data and combine other sources of data in the future. In the meantime there are several related data sources users can rely on. First, Numerous years in the CES have data available on candidate race and gender for interested researchers. Please see the table below for further information regarding the availability of such information by year and where it can be located.

Table 1 – The Availability of Candidate Race and Gender Data by Year

CES	Candidate Race	Candidate Gender
2006		
2008	Variables for H, S, G	Variables for H, S, G
2010	Supplemental Data for H[1]	Variables for H, S, G
2012	Supplemental Data for H[2]	Variables for H and S
2014	Supplemental Data for H[3]	
2016	Supplemental Data for H, S[4]	
2018		Variables for current H post

1. https://doi.org/10.7910/DVN/KC9EQR
2. https://doi.org/10.7910/DVN/NI3BDE
3. https://doi.org/10.7910/DVN/D1N0GO
4. https://doi.org/10.7910/DVN/IAOZOU

Finally, other researchers have collected this data in forthcoming work. "Separating Race and Party in Congressional Elections" by Bernard Fraga (available at https://www.bernardfraga.com/research) uses hand-coded race data from Congressional general and primary elections from 2006 - 2020. "Partisanship and Nationalization in American Elections" by Sharif Almani and Carlos Algara (*Electoral Studies*, doi: 10.7910/DVN/DGUMFI) contains county-level partisan election data from the statewide offices of President, Governor, and Senator going back to the Civil War.

5 Reporting Errors and Suggesting Improvements

If you find any errors or ambiguities, please do not hesitate to contact us. The preferred method is to open a "Issue" at https://github.com/kuriwaki/cces_candidates/issues, which we monitor. Otherwise, please email us at jeremiahcha@g.harvard.edu, shirokuriwaki@gmail.com, or jsnyder@gov.harvard.edu.

6 Version History

- Dataverse v3.0: Added 2021, 2022, 2023 general elections (2024-10-18). These do not include House special elections held during this time.
- Dataverse v2.0: Minor Fixes (2023-06-16). Fixes coding for fusion candidates, incumbents, special elections, and additional name inconsistencies (Github issues 25, 26, 27, 28, 29, 30, 35). Thanks to Luca Bellodi and Bill Wilkerson for finding some of these errors.
- Dataverse v1.0: Initial Dataverse Release (2021-09-27).

7 Acknowledgements

Our thanks to Stephen Ansolabehere and Jaclyn Kaslovsky for their suggestions and comments.