

State Legislative Primary Election Returns Database, 1990–2026

Andrew C. W. Myers¹
Steven Rogers²
Alexander Fouirnaies³
Andrew B. Hall⁴
Cassandra Handan-Nader⁵
Jason Windett⁶

Abstract

State legislative primary elections play a central role in selecting lawmakers in the American states, yet the absence of a comprehensive national database has limited systematic research on their outcomes. We introduce a new database of state legislative primary election returns covering all 50 states and both legislative chambers from 1990 through 2026. This dataset contains 159,573 candidate-election observations spanning 90,860 primary races and records candidates' names, primary and primary-runoff vote totals, and incumbency statuses. To facilitate applied research, we also link this dataset to existing databases on candidate ideology, campaign finance filings, and general-election returns. This database will enable systematic research on candidate entry, primary competition, ideological extremism, electoral accountability, and legislative representation in the American states.

Keywords: election data, state legislatures, primary elections, American politics

¹Hoover Fellow, Hoover Institution, Stanford University, United States of America. Corresponding author. Email: myersa@stanford.edu.

²Associate Professor of Political Science, Saint Louis University, United States of America. Email: steven.rogers@slu.edu.

³Associate Professor of Political Economy, Harris School, University of Chicago, United States of America. Email: fouirnaies@uchicago.edu.

⁴Davies Family Professor of Political Economy, Graduate School of Business, Stanford University, and Senior Fellow, Hoover Institution, United States of America. Email: andrewbhall@stanford.edu.

⁵Assistant Professor of Political Science, New York University, United States of America. Email: cassandra.handan-nader@nyu.edu.

⁶Professor, University of North Carolina – Charlotte, United States of America. Email: jwindett@charlotte.edu.

1 Background & Summary

State legislatures occupy a central place in American democracy, with nearly 7,400 legislators enacting policy in key areas including education, healthcare, criminal justice, and election administration. Primary elections are the first—and often decisive—stage in selecting these legislators. Yet despite their importance, state legislative primary elections have received far less scholarly attention than general elections or congressional primaries, in large part because comprehensive data on their outcomes has not been available.

In this paper, we introduce a new database of state legislative primary election returns covering all ninety-nine state legislative chambers for the years 1990–2026. Our database is the first comprehensive source of state legislative primary returns over the past three decades. It contains 159,573 candidate-election observations across 90,860 primary races involving approximately 95,645 distinct candidates. The database records each candidate’s vote total and win indicator, primary runoff vote totals and outcomes where applicable, candidate names, incumbency status, and district characteristics. We also link each candidate to Klarner’s state legislative election returns database (SLERs),¹ the Database on Ideology, Money in Politics, and Elections (DIME),² the National Institute on Money in State Politics (NIMSP),³ and Shor and McCarty’s legislator ideology (NP-score) estimates.^{4,5}

This new database will enable researchers to study a wide range of questions about primary elections, candidate selection, and legislative representation that have previously been difficult to address due to data limitations.

2 Methods

2.1 Data Collection

We constructed our database by collecting certified primary election returns for every state legislative primary election from 1990 through 2026. Our primary source for each state-chamber-year was the official Secretary of State or state election commission website, from which we downloaded machine-readable returns whenever available. When official digital returns were unavailable—most commonly for earlier years—we obtained data from state Blue Books and legislative manuals, or submitted direct requests to state election offices.

The resulting raw data varied substantially in format across states and years, including structured tabular files (CSV, JSON, and Excel), scanned PDFs, images of official canvass documents, and printed volumes. We developed custom parsing scripts to extract and standardize data from each format. Where automated extraction was not feasible, the data was entered manually. In a small number of cases where official sources could not be located, we supplemented our data with secondary sources (e.g., newspaper reports).

2.2 Standardization

Candidate names. The formatting of candidates’ names varies widely in our raw data across states and years. We standardized all names by converting them to uppercase; removing extraneous text such as titles (e.g., “Representative,” “Senator”), winner asterisks, and

Table 1 – District Naming Conventions Used in American State Legislatures as of 2024. This table reports how states label their legislative districts. **Numeric** indicates numbers, **Alpha** indicates alphabetic characters, and **Post** indicates a position letter. U and L stand for upper and lower chambers, respectively.

Numeric	Alpha	Post	State-chambers
✓	—	—	AK (L), AL, AR, AZ, CA, CO, CT, DE, FL, GA, HI, IA, ID (U), IL, IN, KS, KY, LA, MD (U), ME, MI, MN (U), MO, MS, MT, NC, ND (L) [†] , ND (U), NE, NH (U), NJ, NM, NV, NY, OH, OK, OR, PA, RI, SC, SD, TN, TX, UT, VA, WA (U), WI, WV, WY
✓	✓	—	MA, NH (L), VT (L)
—	✓	—	AK (U), VT (U)
✓	—	✓	ID (L), MD (L), MN (L), WA (L)

[†]Districts 4 and 9 introduced candidate filing by position beginning in 2022.

parenthetical notes; and stripping special characters. We then extracted first, last, middle, and suffix name components from the standardized name string using the Stata name-parsing routine `extrname`.

District identifiers. States use a variety of naming conventions to identify legislative districts. We standardized district identifiers by decomposing each into up to three components, provided as separate variables: `district_num` (the district number), `district_alpha` (an alphabetic name component, e.g., a county name in Massachusetts or a town name in Vermont), and `district_post` (a position designator used when candidates file for a specific seat within a multi-member district). Table 1 documents the resulting conventions by state-chamber.

The post component required careful treatment in multi-member districts. In some chambers, candidates file for—and compete within—a specific post within a multi-member district (Idaho, Maryland, Minnesota, and Washington House). In these chambers the post variable defines a separate filing constituency (or race). In other multi-member chambers (North and South Dakota House and West Virginia Senate), candidates compete at-large within the district, with the top- k finishers winning k seats; since these positions are assigned post hoc and do not define separate races, we drop the post identifier for these races.

2.3 Data Coverage

After collection, standardization, and deduplication, our final dataset contains 159,573 candidate-election observations drawn from 90,860 distinct primary races involving approximately 95,645 distinct candidates. Table 2 reports the composition of the dataset across candidate-

Table 2 – Dataset Composition Across Candidate- and Race-Level Characteristics.

	Observations	Share
<i>Total</i>		
All candidate-elections	159,573	100.0%
<i>Chamber</i>		
House	127,828	80.1%
Senate	31,745	19.9%
<i>Incumbency</i>		
Incumbents	44,377	27.8%
Non-incumbents	102,089	64.0%
<i>Party</i>		
Democratic	80,347	50.4%
Republican	78,522	49.2%
<i>Election timing</i>		
Even-year	150,403	94.3%
Odd-year	9,170	5.7%
<i>Runoffs</i>		
No primary runoff	144,083	90.3%
Participated in primary runoff	2,489	1.6%

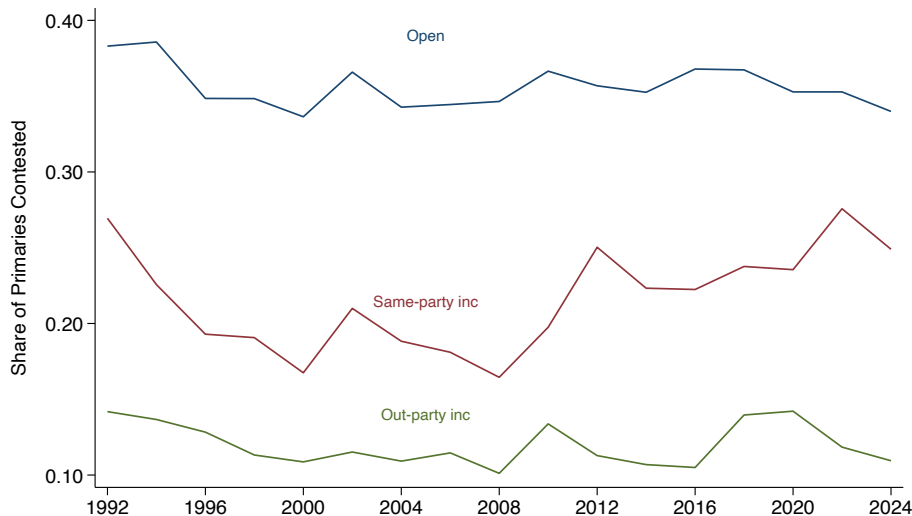
and race-level characteristics. Roughly four in five observations cover elections in states’ lower legislative chambers, three in ten correspond to incumbents, and the sample is evenly split between Democratic and Republican candidates.

One of the most widely noted features of state legislative primaries is how rarely they are contested, yet scholars have lacked the data to systematically document this pattern. Figure 1 uses our database to measure contestation in state legislative primaries over time. We find that fewer than 40% of primaries since 1990 were contested. This average masks substantial heterogeneity across electoral contexts, however. Open-seat primaries (races without an incumbent) are the most likely to be contested (35–40%). Same-party incumbent primaries are contested at about half that rate, while out-party primaries—where a challenger would face the opposing party’s incumbent in November—attract competition in only around 10–13% of races. These patterns have remained stable across our sample, although all three show modest upticks in major redistricting years.

2.4 Linking to External Datasets

To facilitate future research, we link our primary-election database to four existing datasets on legislative politics. We conduct this merging using fuzzy name matching within defined state-chamber-district-year blocks using the Stata package `megamerge`. Table 3 summarizes the four external datasets and our match rates. Note that match rates are below 100%

Figure 1 – Share of Contested Primaries by Electoral Context. Each line shows the average share of primary elections that are contested in even years, stratified across electoral environments. Open races are primary elections without an incumbent, same-party incumbent races feature an incumbent from the party holding the primary, and out-party incumbent races feature an incumbent from the opposing party.



because many primary-election candidates do not run in general elections, raise contributions, and/or have an estimated ideology score.

We first link to Klarner’s State Legislative Election Returns database¹, which provides the corresponding general-election returns. Matched candidates receive a `klarner_id` that can be used to pull any variable from the SLERs. We also incorporate the redistricting regime (`regime`) and single-member district indicator (`smd`) from the SLERs. Second, we match to the Database on Ideology, Money in Politics, and Elections (DIME)², linking each matched candidate to a `bonica_id`. Third, we merge in identifiers from the National Institute on Money in State Politics (Follow the Money), which provides state-level campaign finance disclosures. Linked candidates receive an `ftm_id`. Finally, we merge in identifiers (`sm_id`) from Shor and McCarty’s common-space ideology estimates (NP-scores)^{4;5}. At the time of writing, all four external datasets only extend through the 2024 election cycle.

3 Data Records

The database is available in three formats at <https://github.com/AndyCWMyers/state-legislative-primary-returns>: a Stata binary file (`.dta`), a comma-separated values file (`.csv`), and an R data file (`.rds`). All three files contain identical data. Each row represents an individual candidate in a single primary election. Each file contains 159,573 rows and 26 variables. Table 4 describes each variable.

Table 3 – Match Rates to External Datasets. Each row reports the number and share of candidate-election observations linked to an external database. SLERs stand for State Legislative Election Returns¹. DIME stands for Database on Ideology, Money in Politics, and Elections². NP-scores are the Shor-McCarty common-space legislator ideology estimates^{4;5}.

Dataset	Content	Matched	Share
Klarner SLERs	General-election returns	94,647	59.3%
DIME	Ideology estimates	90,290	56.6%
Follow the Money	Campaign contributions	104,531	65.5%
Shor-McCarty	Legislator ideology (NP-scores)	46,136	28.9%

4 Technical Validation

4.1 Randomized Spot-Checks

After our collection was complete, the data underwent two systematic checks. First, research assistants verified the completeness of contested primary election records by cross-referencing the collected data against a comprehensive list of expected contested races. For the portion of the database spanning 1990 to 2016, we deliberately withheld at least four contested races from each state’s checklist that research assistants were required to identify and add. Second, research assistants verified the accuracy of candidate names and vote totals for a random sample of at least 15% (1990–2016) and 5% (2017–2024) of races per state.

4.2 Internal Consistency Checks

We also performed several internal consistency checks on our primary-election vote data.

Win indicator consistency. In single-member district primaries, exactly one candidate should have `win_primary = 1`, and that candidate should have the highest vote total in the race. In multi-member districts, the top- k vote-getters receive `win_primary = 1`, where k equals `district_seats`. We verified both conditions across all races with vote data.

Runoff consistency. For state-years that use a primary runoff system, we verified that runoff variables are populated for the appropriate candidates.

5 Usage Notes

Three features of our dataset warrant special attention. First, each row represents a candidate-election. Since many states do not publish vote totals for uncontested primaries, however, candidates winning uncontested primary elections often do not appear in our database. We include data on uncontested primary elections whenever available, but researchers wanting to

Table 4 – Variables Included in Database.

Variable	Description
Identifiers	
year	Election year.
state	State (2-letter USPS abbreviation).
chamber	Legislative chamber (U = upper/senate, L = lower/house).
party	Party (dem = Democratic, rep = Republican, NP = Nebraska nonpartisan).
district_num	District number component (missing if absent).
district_alpha	Named district component.
district_post	Within-district position designator.
primary_race_id	Unique primary race ID (state $\times$ year $\times$ chamber $\times$ district $\times$ party).
klarner_id	Klarner Politics general election data candidate ID.
bonica_rid	DIME recipient ID ² .
ftm_id	Follow the Money / NIMSP candidate ID.
sm_id	Shor-McCarty legislator ID for NP-score ideal points ^{4;5} .
Candidate name	
first_name	Candidate first name (standardized).
middle_name	Candidate middle name (standardized).
last_name	Candidate last name (standardized).
name_suffix	Candidate name suffix (Jr., Sr., III, etc.).
Primary election	
votes_primary	Candidate's raw vote total in the primary election.
win_primary	=1 if the candidate won a plurality (or was a top- k finisher in a k -seat district).
Primary runoff	
ran_primary_runoff	=1 if the candidate participated in a primary runoff.
votes_primary_runoff	Candidate's raw vote total in the primary runoff.
win_primary_runoff	=1 if the candidate won the primary runoff.
win_primary_final	=1 if the candidate won the nomination, accounting for any primary runoff.
Incumbency	
inc	=1 if the candidate held a seat in the same chamber at the time of the election.
District characteristics	
regime	Redistricting regime (increments with each post-Census redistricting cycle).
smd	=1 if the election is in a single-member district.
district_seats	Number of seats available in the district.

recover the identity of winners of uncontested primary elections should merge our database to Klarner’s SLERs.

Second, Louisiana uses a jungle primary in which all candidates compete on a single ballot regardless of party, with a runoff between the top two finishers if no candidate receives a majority. We treat the first-stage contest as the primary election and include it in the dataset. And finally, since Nebraska’s unicameral legislature is formally nonpartisan, we set these candidates’ party to NP (for no party).

6 Data Availability

Our database is available at <https://github.com/AndyCWMyers/state-legislative-primary-returns>.

7 Code Availability

The replication code for the analyses presented in this paper is available at <https://github.com/AndyCWMyers/state-legislative-primary-returns>.

8 Acknowledgements

The authors thank Dan Carter, Abby Faust, Kunal Goel, Tegan Hoover, Emily Johansson, Adam Kneepkens, Elise Kostial, Laniah Lewis, Sequoyah Lopez, Janet Malzahn, Patrick Monahan, Jeffrey Seib, Anish Sundar, and Lucie Wood for excellent research assistance.

9 Author Contributions

A.M. conceived and designed the database, coordinated data collection, wrote the data cleaning and assembly pipeline, and wrote the manuscript. S.R. contributed to data collection and to writing the manuscript. A.F., A.H., C.H.N., and J.W. contributed to data collection.

10 Competing Interests

The authors declare no competing interests.

11 Funding Statement

This project was not supported by external funding.

References

1. Klarner, C. State Legislative Election Returns, 1967–2024 (2025). URL <https://doi.org/10.7910/DVN/3WZFK9>.
2. Bonica, A. Database on Ideology, Money in Politics, and Elections (2025). URL <https://data.stanford.edu/dime>. Version 4.
3. National Institute on Money in State Politics. FollowTheMoney.org (2025). URL <https://www.followthemoney.org/>.
4. Shor, B. & McCarty, N. The ideological mapping of American legislatures. *Am. Polit. Sci. Rev.* **105**, 530–551 (2011). URL <https://doi.org/10.1017/S0003055411000153>.
5. Shor, B. Individual State Legislator Shor-McCarty Ideology Data, January 2025 Update. Harvard Dataverse (2025). URL <https://doi.org/10.7910/DVN/SG0Q7G>.