

State Legislative Election Returns (1967-2016)

Codebook

Carl E. Klarner

Codebook Last Updated: September 27, 2018

Authorship

William D. Berry, Thomas M. Carsey, Malcolm Jewell, Richard G. Niemi, Lynda W. Powell and James Snyder were responsible for compiling previous editions (ICPSR studies 21480, 3938, 8907 and 34297) and have generously allowed Klarner to be the sole author of this edition. Klarner collected the data for the 2007-2016 elections, and corrected errors in pre-2007 observations.

Citation

Please use the following citation.

Klarner, Carl, 2018, "State Legislative Election Returns, 1967-2016", <https://doi.org/10.7910/DVN/3WZFK9>, Harvard Dataverse, V3, UNF:6:pV4h1CP/B8pHthjjQThTTw==

Overview

This codebook describes the data contained in the Stata file "slers1967to2016_20180908.dta" archived at Harvard's Dataverse. This data set is an updated version of previous editions of State Legislative Election Returns data (ICPSR studies #8907, #3938, #21480 and #34297). The dataset has been updated through 2016, numerous errors have been corrected, new variables have been added, and existing variable names and codes have been changed.

Coverage

Each of the more than 350,000 observations refers to an individual candidate who ran for state legislative office, with general election returns for 2013 to 2015, and some states in 2016 being broken down further by county. For some states before 1990, there are observations for candidates competing in primaries and special elections. For years between 1990 and 2012, almost all observations are for general elections, although all primary elections in 2013 to 2015 have been added. Some observations in the dataset have missing values for one or more variables.

This data set contains information on state legislative election returns from 1967 through 2016. The only elections it contains from 1967 are primaries from Louisiana. The following table reports the first year in which elections were held in a state for both the state house and state senate. Column four is useful because it indicates after which election a nearly complete list of state legislators can be generated. For example, the California State Senate had about half of its seats up in 1968, and the other half up in 1970. As a result, the database can produce a complete list of sitting legislators for the 1971 legislative session, but not the 1969 legislative session.

| State | State Houses (sen = 0) | State Senates (sen = 1) | State Senates (sen=1), Year by Which All Seats Have Been Up |
|---|------------------------|-------------------------|---|
| AZ, CT, DE, FL, GA, ID, KS, LA, ME, MA, NH, NY, RI, SC, SD and TN | 1968 | 1968 | 1968 |
| MN | 1968 | 1970 | 1970 |
| NM | 1968 | 1970 | 1972 |
| KY | 1969 | 1969 | 1971 |
| NJ and VA | 1969 | 1971 | 1971 |
| AL, MD, MI and NC | 1970 | 1970 | 1970 |
| MS | 1971 | 1971 | 1971 |
| VT | 1986 | 1986 | 1986 |
| Remaining states | 1968 | 1968 | 1970 |

An effort was made for this edition to include all general elections that occurred between 1968 and 2016, but general elections in North Carolina in 1968 and Vermont from 1968 to 1984 are not covered. Ongoing efforts are being made to obtain these missing elections.

The above table indicates that New Mexico does not have all of its state senators accounted for until the 1972 elections. This is because eight senators were elected to six-year terms in 1966.

Eight cases containing the names of these New Mexico State Senators, as well cases containing all the sitting legislators in Vermont from 1968 to present, have been collected and will be added to SLERs so that it can yield a comprehensive list of all state legislators in the country from the 1971 legislative session and on, when time permits.

In earlier editions, some state legislative contests were not reported, especially for states in which uncontested elections were not put on the ballot (AR, FL (1970 and later), HI (1990 and later), LA (1979 and later), and OK (1972 and later)), but there is now total coverage of such elections. South Dakota in the 1976 and 1978 elections may be another instance of such a state, although it would be strange for such a law to only be operative for two elections. There are clusters of missing votes in other state-years as well. For 178 elections (not including the 1966 New Mexico Senators and Vermont legislators before 1986), data are available only for the winner (173 of which are in NC from 1984 to 1990). These cases are identified in the variable “uncert” with a code of “incompleteelect.”

Sources

The data comes from five sources. The data from 1967 through 1989 comes from ICPSR Study #8907 (directed by Malcolm Jewell), as updated by William Berry and Thomas Carsey in ICPSR Study #3938. Most of the observations from 1990 through 2000 were provided by Richard Niemi, Lynda Powell, and James Snyder. A portion of the 2000 data and data for 2001 to 2003 were entered by research assistants working for Carsey and Berry. Data for 2004 to 2006 was provided by Richard Niemi and Lynda Powell with some assistance from Carl Klarner, and that for 2007 to 2010 was collected from state Web sites by Carl Klarner and contained in ICPSR Study #34297. Data for 2011 to 2016 was collected from state Web sites by Carl Klarner and included in this edition. Improvements, corrections and other changes for all years were made by Carl Klarner in 2010-2018.

Acknowledgments

This project received support from a number of sources. Carsey and Berry received three separate grants from the National Science Foundation that contributed to producing this data set. NSF Grant # SES-0136526 and NSF Grant # SES-0136526, Amendment 001 supported their efforts to clean the candidate name variable in ICPSR Study #8907 that covers the years 1967–1989. They released that data as ICPSR Study #3938. NSF Grant # SES-0317924

supported extending the data set through 2003 and cleaning the name variable in the updated data to permit its merging with their previous efforts. Numerous students and others were involved in both data entry and name cleaning. Richard Niemi, Lynda Powell, and Jim Snyder graciously provided the election returns they had already entered covering much of the period from 1990-2000. Carl Klarner received financial assistance from the Department of Political Science at Indiana State University as well as the Indiana State University Research Committee.

Numerous people helped by providing information, including countless state reference librarians, Belinda Davis (Louisiana State University), Peter Granda (ICPSR), Will Bullock (Princeton), Laura Hatafsky (Georgetown), Curtis Reynolds (Kent State), Jonathan Backer (Columbia University), Ronald Weber (University of Wisconsin-Milwaukee), and my workers at Indiana State University (Brent Ellis, Marcel Oliveira, Thomas Estabrook, Pei-Shiue Hsieh, Charlynn Turner, Michael Melshen, Tina Mason and Steven Flowers), Karen Brown (Eugene V. Debs Foundation) and Kristin Phillips (FSU). The following individuals assisted with the collection of the 2011-2012 update: Magan Bell (Duke), David Brockman (UCB), Daniel Butler (Yale), Nicholas Carnes (Duke), Marion Johnson (Duke), Erik Juenke (Michigan State University), Thad Kousser (UCSD), and Paul Pena (Yale). Since 2014 numerous people have contributed help who I haven't kept good track of. But these include Curtis Lock Reynolds, Alexander Fourniaies (University of Chicago), and my workers at the University of Florida (Eduardo Santana, Dillon Farneti, Dillon Boatner, Sicheng Zhu, Laura Uribe, Ana Colic, Dareen Odige and Thuy Pham). Of special mention is Adam S. Myers (Providence College) who, in addition to other assistance, diverted two valuable days of research time at the Library of Congress to obtain Vermont results from 1968 to 1984 (which I haven't added yet). Also of special mention is Peter Koppstein (Princeton) who has alerted me to numerous errors concerning the identities of candidates.

Errors

If you identify errors, please alert Carl Klarner (Carl.Klarner@gmail.com).

User Aids

There are many aspects of this data set that make it challenging to work with. A check list of questions that may aid in the use of these data appear in Appendix F. Code that transforms the structure of the dataset for a variety of purposes are included in other Appendices.

Variable List

The variable names for this edition of the data set were changed from earlier editions. Appendix A contains a conversion chart between the variable names of ICPSR #21480, subsequent editions, and this edition.

All strings in the dataset have been changed to lower case letters with the exception of variables that store exact replicas of variables from earlier versions.

| Variable Name | Brief Variable Description |
|---------------|--|
| caseid | Unique Identifier |
| uncert | Specifies Variable or Variables There is Uncertainty About |
| year | Year |
| month | Month |
| day | Day |
| sid | Number of State in Alpha List |
| sab | State Abbreviation |
| sfips | State FIPS Code |
| sicpsr | State ICPSR Code |

| | |
|------------|--|
| cname | County Name |
| cfips | County FIPS Code |
| sen | Senate |
| ddez | District Designation Given on Ballot |
| dname | Name of District for Use |
| dno | District Number for Use |
| geopost | Geographic Post for Use |
| mmdpost | MMD Post for Use |
| specpost | Tracks Rare Cases of Candidates Switching Posts or Other Related Changes |
| dtype | District Type Code |
| dseats | Number of Seats in District |
| popnum | Numerator in Ideal Pop Formula |
| redist | Redistricting Code |
| redist1 | District Name of Former District |
| redist2 | District Number of Former District |
| redist3 | Geographic Post of Former District |
| regime | Redistricting Regime |
| flot | Floterial District Code |
| flot1 | District Name of Covering Floterial District |
| flot2 | District Number of Covering Floterial District |
| flot3 | Geo Post of Covering Floterial District |
| nest | Number of House Districts Nested Inside a Senate District |
| nest1 | Name of Nesting Senate District |
| nest2 | Number of Nesting Senate District |
| nest3 | Geographic Post of Nesting Senate District |
| etype | Election Type Code |
| deter | Dummy: Election Sent a Candidate to Sit in the Legislature for a Full Term |
| eseats | Number of Seats Up for Election |
| term | Anticipated Term Length |
| termz | Actual Term Length |
| cando | Original Candidate Name from Returns |
| candformat | Format of Original Candidate Name |
| cand | Standardized Candidate Name |

| | |
|--------------|---|
| candid | Candidate Id |
| partyo | Original Party Designation from Returns |
| party | Detailed Party |
| partyz | Party for Use |
| partyt | True Party |
| partytsource | Source for True Party |
| firstcase | Early Case in Dataset Code |
| prior1 | Service in the Chamber before Dataset Began |
| prior2 | Service in the Legislature before Dataset Began |
| exper | Prior Legislative Experience of Candidate |
| vote | Number of Votes for Candidate |
| outcome | Winner Status of Candidate |
| hold | Holdover Legislator Code |
| hold1 | New District Name of Holdover Legislator |
| hold2 | New District Number of Holdover Legislator |
| hold3 | New Geographic Post of Holdover Legislator |
| tenure1 | Years of Continuous Tenure in Chamber |
| tenure2 | Years of Continuous Tenure in Legislature |
| limit | Term Limit Code |
| last | Candidate Last Name |
| first | Candidate First Name |
| middle | Candidate Middle Name |
| middle2 | Candidate Extra Middle Name Components |
| nick | Candidate Nickname |
| suffix | Candidate Name Suffix |
| prefix | Candidate Name Prefix |
| namenum | Number in Candidate's Name |
| last1 | Alternative Candidate Last Name #1 |
| last2 | Alternative Candidate Last Name #2 |
| last3 | Alternative Candidate Last Name #3 |
| last4 | Alternative Candidate Last Name #4 |
| ltype1 | Type of Alternative Candidate Last Name #1 |
| ltype2 | Type of Alternative Candidate Last Name #2 |

| | |
|--------------|--|
| ltype3 | Type of Alternative Candidate Last Name #3 |
| ltype4 | Type of Alternative Candidate Last Name #4 |
| lastb | Candidate Last Name from 2nd Source |
| firstb | Candidate First Name from 2nd Source |
| middleb | Candidate Middle Name from 2nd Source |
| middle2b | Candidate Extra Middle Name Components from 2nd Source |
| nickb | Candidate Nickname from 2nd Source |
| suffixb | Candidate Name Suffix from 2nd Source |
| last1b | Alternative Candidate Last Name #1 from 2nd Source |
| last2b | Alternative Candidate Last Name #2 from 2nd Source |
| ltype1b | Type of Alternative Candidate Last Name #1 from 2nd Source |
| ltype2b | Type of Alternative Candidate Last Name #2 from 2nd Source |
| namebsource | Candidate 2nd Source Name |
| v38 | Candidate Name from Icpsr8907 |
| v39 | Candidate Name from Icpsr21480 |
| v56 | District Number from Icpsr21480 |
| v57 | Geographic Post from Icpsr21480 |
| v58 | District Name from Icpsr21480 |
| v18_20171211 | Candidate Id from December 11, 2017 |
| v19_20171211 | Standardized Candidate Name from December 11, 2017 |
| v19_20160217 | Standardized Candidate Name from February 17, 2016 |

Unit of Analysis

In some instances, the same candidate name appears more than once for the same race in the same district in the same year. This occurs for two main reasons.

First, it occurs because of “fusion” voting. This is routine in New York and Connecticut, but sometimes occurs in other states as well. In New York and Connecticut, the same candidate can be nominated by more than one party and will thus appear on the ballot more than once with each different party label, in a practice sometimes referred to as “fusion.” Listing a candidate multiple times, once for each party that nominated him or her, permits users of the dataset to determine how many votes a candidate received from each party.

Second, it occurs when county breakdowns are reported, which is the case for general elections from 2013 through 2015, and 21 out of 45 states in 2016.

The variables in SLERs are measured at many different units of analysis, and aggregation is necessary to compute many quantities of interest. Combinations of variables specify these groups, and some key examples are listed here.

| | |
|------------------|--|
| The variables... | Together, the variables identify / define cases for one... |
|------------------|--|

| | |
|---|--|
| sid, sen, dname, dno, geopost, regime | district / unique geographic locale in one redistricting period |
| sid, sen, dname, dno, geopost, mmdpost, regime | post in one redistricting period or a district in one redistricting period if it is a district with no mmd-posts |
| sid, sen, dname, dno, geopost | district in one state chamber. Note that districts change the area they represent over time (sometimes by moving to the other side of the state each decade, as in AR or MT), so these variables do not define a unique geographic locale if a time variable is not specified. |
| sid, sen, dname, dno, geopost, mmdpost, year, month, day, etype | one election for one post / aka “an election” / aka “a contest” |
| sid, sen, dname, dno, geopost, mmdpost, year | “an election cycle,” in other words, the general election and its associated primary, and any associated runoffs, as long as special elections do not occur in that post and year. This also makes an exception of Louisiana before 1975 when its primaries were in a different calendar year than the general election they were for. |
| sid, sen, dname, dno, geopost, mmdpost, year | election in one post if only general elections are being examined |
| sid, sen, dname, dno, geopost, mmdpost, year, etype, candid | candidate in one election |
| sid, sen, dname, dno, geopost, mmdpost, year, etype, candid, party | one party a candidate ran on in one election |
| sid, sen, dname, dno, geopost, mmdpost, year, etype, candid, party, cname | For the most part, defines one row in the dataset. However, 186 cases are not identified by those. 113 of those are caused by missing parties. 73 of those need to be looked into. |

VARIABLE DESCRIPTIONS

Coverage for each variable listed below is complete unless otherwise stated, with the exception of missing information identified in the variable “uncert.”

BASIC VARIABLES

caseid

Unique identification number

For each case that was included in ICPSR #21480, this number was identical in the next edition of the dataset. This number is almost always the same from edition to edition for the same case, although very rarely cases are dropped and then added with a different caseid.

uncert

Name or names of variable or variables there is uncertainty about, “uncertainty” also including situations where there is missing data for a variable. The following codes are also possible entries for this variable.

identity: indicates there is uncertainty about the identity of the candidate.

incompleteelect: indicates that only the winner of an election is contained in the dataset. Implies that “vote” is missing.

writeinstatus: indicates that the writein status of a candidate is unknown.

generalproblem: indicates that the case in question should be checked for all variables.

year

Year of Election (four digit)

Coverage: complete.

month

Month of Election (1-12)

Coverage: missing for three cases

day

Day of Election (1-31)

Coverage: all years for general elections, but is often missing for special elections and primaries.

sid

Number of state in alphabetical list between 1 and 50. See Appendix B. The District of Columbia is not included.

Coverage: complete.

sab

State Postal Abbreviation

A string variable that records the two-letter state postal abbreviation for each state, but in lower case letters. See Appendix B.

Coverage: complete.

sfips

State FIPS Code

Records the FIPS code for each state, which is a common identifier for states used by the U. S. Census Bureau. See Appendix B.

Coverage: complete.

sicpsr

ICPSR state code (See ICPSR Study #8907 for details). See Appendix B.

Coverage: complete.

cfips

Census county FIPS code.

Format: one, two or three digit number, with the exception of the code “9999.”

Coverage: all 2013, 2014 and 2015 general elections, except for AK, and some states in 2016.

There are some voters who contribute to totals for state legislative elections, but are not assigned to a particular county, and they are given special codes for this variable. Values of “cname” of “emergencyutilityworkers” in Maine are given a code of “9999” for cfips, and values of “cname” of “stateuocava” (uniformed service and overseas voters), also in Maine, are given a code of “9998.”

Note: this code does not include the two digit prefix that represents the Census’s state FIPS code.

cname

Name of county.

The word “county” is not included, except in the case of Virginia, which has both counties and cities and where those designations are sometimes necessary to distinguish two locales. For VA, those designations always end with either “county” or “city.”

Counties with “Saint” in their name are spelled “st” Ste. Genevieve county Missouri is left “ste” instead of “st.” as that is the official name.

Format: all lower case with all non-alphabetic characters removed, such as spaces, apostrophes (i.e., o’Brien), dashes (i.e., miami-dade) and periods (i.e., “st.”).

Coverage: all 2013, 2014 and 2015 general elections, except for AK, and also includes 19 states in 2016.

Two other designations appear in this field that are not counties: emergencyutilityworkers and stateuocava, which are emergency workers and overseas/uniformed voters in Maine.

sen

Dummy variable indicating state senate.

1 = State Senate / Upper House (and Unicameral in Nebraska)

0 = State Lower House (House of Representatives, Assembly, House of Delegates, etc.)

Coverage: complete.

DISTRICT & POST IDENTIFIERS

A “duel system” of district designations is utilized in this dataset, where one variable (“ddez”) is made to appear as similar to how district designations are presented on returns as possible, and another battery of district designation variables has been altered to facilitate analysis. Unfortunately, variables were extensively standardized before this system was implemented, although efforts are being made to restore the original presentation of the district designations by comparing them to original sources.

ddez

District designation in the returns.

This variable reports a district designation along with any MMD-posts as closely to how the district designation was presented in the returns as possible. This means that all components of a district and post designation (alphabetical, numeric or other) appear in this variable, and alpha components are in proper format. Letters are not converted to numbers (i.e., “A” to “1,” etc.).

Unlike the other district designation variables that are about to be described, the goal of this variable is not to facilitate ease of analysis.

In prior editions, characters were first changed to all capital letters, then all lower case letters, and have now been converted back to proper format. Similarly, letters were often changed to numbers in earlier editions, and many of these have been changed back to letters for this variable.

dname

Alphabetical component of district designation, altered for ease of use.

Several states use names or letters rather than numbers to identify specific districts. If a district is solely named with a number, this variable is coded as missing.

When there is a mixture of a number (including the words “first,” “second,” etc., which were converted to numbers) and an alpha component, the number is placed in either “dist,” “geopost,” or “mmdpost” as appropriate and described in those variables, and the alpha portion of the district designation is placed in “dname.”

This variable was altered where necessary to make a district’s name identical within a redistricting period/regime along with other changes to prevent a variety of problems, ease merging data by adopting easy to remember rules, and make manipulating the data easier. Accordingly, all letters were made lowercase and anything that wasn’t a letter (spaces, dashes, the word “and” or the symbol “&,” for example) were removed. Furthermore, if the alpha component of a name is comprised of two components (often two place names, e.g., “bristol and norfolk”), these were alphabetized with the exception of districts in Vermont, Wyoming and the Massachusetts Senate. The order of the alpha component of names is important in these instances (i.e., “bristol and norfolk” and “norfolk and bristol” are two distinct districts).

Note that district designations are also referred to in the historical record inconsistently (i.e., Nevada), making it impossible to have district designations that are both appropriate for analysis and match the historic record.

UOA: “dname,” “dno” and “geopost” for a specific time period together define a unique geographic locale.

dno

District number

The district number for each office. Some states solely use names or letters to identify districts. In those cases, “dist” is coded as missing.

This variable was altered where necessary to standardize a district’s designation within a redistricting period/regime, and to make other comparisons possible.

Values of “dist” are also altered when necessary so that the same geographic locale has the same district number. For example, up until the 1990s, Florida numbered seats sequentially from “1” and up, whether the district had one seat or was a post-MMD. In other words, one seat might be assigned district number “40” and then another post in the exact same geographic locale might be assigned district number “41.” Both districts would be assigned a value of “40” here, and mmdpost would be coded “1” for what was originally district 40, and “2” for what was originally district 41. The same practice was also used by AR, GA, TN and WV. The convention adopted here facilitates the comparison of post-MMDs within the same district.

UOA: “dname,” “dno” and “geopost” for a specific time period together define a unique geographic locale.

geopost

Post position that identifies a geographic locale. “Post” here refers only to the type of post that is nested geographically inside of a larger geographic area denoted by “dname” or “dist.” For example, Minnesota State district 67 is divided into two areas, one being Minnesota House district 67A and the other being Minnesota House district 67B.

Posts represented by letters in the returns were converted to numbers, for example, the letter “A” was converted to a “1,” “K” to an “11,” etc. When returns report a district designation as having two numbers separated by a dash (which occurs in VT, for example), and an MMD-post is not being referred to, the second number is put into this variable. Rarely, the representation of a geographic post with an alpha component is converted into a number and placed here (OR 1968 and 1970). In the case of the MS 1971 House, post designations that referred to both geographic posts and MMD-posts were reassigned numbers and put into geopost and mmdpost.

When a district does not use posts, this variable is system missing.

mmdpost

Multimember district post position. An example of this type of post is any Washington State House seat in 2008 (i.e., “District 7, Position 1”) or any Idaho State House seat in 2016 (i.e., “District 23, Post A”).

When MMD-posts are represented in the returns by a letter, they are converted into a number (A=1, B=2, etc.).

MMD-post designations that do not appear in returns are added to facilitate analysis. The first reason this is necessary is in chamber-years that number all seats in the chamber sequentially from “1” and up, even if those seats are multimember districts with posts, as explained above.

A second reason this is necessary is to track the same seat over time in a district with multiple seats that have an alternating component (i.e., “dtype”=4, 5 or 6). It also allows different seats in the same district to be differentiated over time. For example, WV Senate District #1 was assigned an “mmdpost” score of “2” in 1998, but when the other seat for that district was up in 2000, it was assigned a score of “1.” Sometimes in several states, due to vacancy, both seats in an alternating arrangement are up at the same. However, in 2000, WV Senate district #11 was assigned a score of “1” for one of its seats, but a score of “2” for the other seat up that year. WV Senate #17 (i.e., “dist”=17) is given values of 3 and 4 for “mmdpost” when WV Senate #8 and #17 were the same geographic locale, before the 2012 elections. This system enables the appropriate lagging of variables.

When a district does not use posts, this variable is blank.

The county distribution rule applied to MD HS #36 and MD HS #37-2 in 2014. However, #36 has four counties (at least), and three seats are in that district. #37-2 has three counties (at least), and two seats are in that district. That means the counties are not equivalent to MMD posts, although similar, and cannot simply be put into “mmdpost.” The county distribution rule did not change who won in the 2014 primaries.

specpost

Special mmd-post codes were added to track two situations.

Situation #1

Very rarely, a winner in a post-MMD election will switch posts even though there is no redistricting. This happens about 30 times in the time period covered. This variable reports the MMD-post the incumbent will switch to in the next election (i.e., at time t+1). If there are two posts in the district, the other post in the same year (i.e., time t) will also have a value, whether it has an incumbent who will switch posts or not. The value for “specpost” in this post will be the same as the value of the post the incumbent is switching from that year.

For example, if candidate A wins the 2012 election in post 1 of district 1, but then runs in post 2 of district 1 in 2014, “specpost” will receive a value of “2” for cases with values of “1” for “dno,” “1” for “mmdpost” and “2012” for “year.” Likewise, “specpost” will receive a value of “1” for cases with values of “1” for “dno,” “2” for “mmdpost,” and “2012” for “year.” This system make it easy to avoid using one post’s value for two post’s lagged values in the future, and to enable appropriate lagging generally.

Situation #2

Around 115 cases are in an FFA-MMD that elects legislators who will serve different lengths of terms in office. Some are known to have been caused by vacancies, some have been confirmed to not be caused by vacancies, while information has not been obtained for other cases. Evidence indicates that the length of these terms are sometimes not announced in advance for the second type of case mentioned above. These cases appear in Alaska (1974 to 1980), Hawaii (1974 to 1980), Nevada (1974 to 1980), North Dakota (2002 to 2010), and Wyoming (1972 to 1990).

To deal with these situations, letters were assigned to mmdpost when two conditions held for a sid-sen-dname-dno-geopost-regime. First, a set of cases defined by sid-sen-dname-dno-geopost-regime had to have at least one election that had an FFA-MMD component (“dtype”=3, 6, 9 or 10). Second, a “seat” up for election in that period had to appear in two or more distinct sets of seats. An example of this would be a district with three seats,

where seats A and B are up in 1972 and 1976, seat C is up in 1974, and seats A and C are up in 1978. When conditions one and two are met, the code prefix “zzz” is input into “mmdpost” for the entire regime for this district. Furthermore, letters are arbitrarily assigned to each seat, and a string of all seats up in the FFA-MMD appears for each case within the election. For example, if seats A, B, C and D are up, then “zzzabcd” would be the entire entry for every case in the election. All seats appear for all cases in the FFA-MMD because there are no posts, and differentiating seats with this variable would prevent it from being used for its purpose. Another example is if seats A and B are up in 1972, and seat A is up in 1974. Seat A appears in two sets of seats, so condition two is observed. Note that if seats A and B were in a post-MMD, condition one would fail. It is also the case that when more than two seats are up, and one has “termz”=2 and one has “termz”=4, an “a” always appears in the next election. If two seats in the first election have “termz”=2, then “ab” appears in the second election.

UOA: this variable is constant at the election (i.e., sid sen dname dno geopost mmdpost year month day etype) level.

DISTRICT ATTRIBUTES

dtype

District type code.

- 1: Single-Member District
- 2: Multimember District with Positions / Posts
- 3: Multimember District Free-for-All
- 4: Multimember District Alternating Years (Staggered Terms), but only one seat up in that year.
- 5: Multimember District Alternating Years (Staggered Terms), but multiple seats up in that year, with posts.
- 6: Multimember District Alternating Years (Staggered Terms), but multiple seats up in that year, without

posts.

- 7: Floterial District (Single-Member)
- 8: Floterial District (Multimember with Positions / Posts)
- 9: Floterial District (Multimember Free-for-All)
- 10: Multimember district with cumulative voting (this only includes the Illinois House in 1980 and before).

This variable is coded “1” for top two primaries (i.e., when “etype”=ttp), even though “eseats” is coded “2” (see below).

A district being defined as having an alternating component (i.e., have a “dtype” score of “4,” “5,” or “6”) is defined as follows. A sufficient condition for a district to be defined as having an alternating component is if, in any election in a district-regime (i.e., sid sen dname dno geopost regime), not all of its seats are up for election. If no post-mmd is involved (i.e., dtype is neither “2,” “5” or “8”), then the variable “eseats” not being the same as the variable “dseats” would mean this condition is true (with the seldom observed exception of a runoff being involved). A second sufficient condition for there to be an alternating component is if any election within a district-regime, or any two elections held at the same time in the same district, have different term lengths (recorded in the variable “termz”) for any of its winning candidates.

Sometimes districts that utilize free-for-all multimember rules to designate winners have special elections for only one of the seats. When such special elections are held at a time other than election day of even numbered years (or odd numbered years for states having regularly scheduled elections in odd numbered years), “dtype” is given the same code as the district has in general elections.

UOA: An election. A district-regime is almost always the unit of analysis, with the exception of district-regimes that have a mix of 4, 5 and/or 6 and Maryland for the districts affected by the resident delegate rule. These exceptions mean that the unit of analysis is an election. Aside from these exceptions, whether a district has an alternating component or not is an attribute of a district-regime.

dseats

Number of seats in a district.

UOA: a district-regime (dname-dno-geopost-regime).

This is not the same as the number of seats reported in “eseats” because “eseats” is an attribute of an mmd-post within a district if there are multiple mmd-posts, or if not all seats in the election are up because of an alternating component.

Note: For floterial districts, this variable does not count the districts that it “floats” over / are nested within it. Accordingly, if a district has “7” as a value of “dtype,” this variable will be coded “1.”

popnum

Short for “population numerator” and enables computation of a district's ideal population.

UOA: a district-regime (dname-dno-geopost-regime).

It is X in the following formula.

$(X / (\text{number of seats in a chamber})) * \text{state population} = \text{"ideal population"} / \text{district population}$ if there was no malapportionment.

This variable is the same as dseats when the district in question is not a floterial district or encompassed by a floterial district.

For floterial districts, the formula is as follows.

$X=A+B$, where A is the number of seats in the floterial district, and B is the number of seats in the non-floterial districts that are encompassed by the floterial district.

For non-floterial districts that are encompassed by a floterial district, the formula is as follows.

$X=[(A+B)/B]*C$, where A and B are defined as before, and C is the number of seats in the non-floterial district in question.

The situation would become more complex if a floterial district does not neatly nest the non-floterial districts that have some area inside of it, and a simple formula will not do. I am not aware of any such floterial districts.

Note that if these fractions get rounded to a different number of decimals from each other, problems may result when collapsing the data.

REDISTRICTING VARIABLES

redist

Code indicates whether a district been redrawn since the last time the seat up for election had an election.

1: boundaries of the district were altered since the last time the seat or seats up for election in that district in that election cycle had an election, but it is not the result of "holdover redistricting" (codes 6 and 7 below).

For some years, if a district still had 99% or more of its previous population after having its boundaries changed, the district was not coded as having been redistricted (for example, some altered districts in the GA 2014 Senate, the GA 2016 House, and the TX 2014 House).

2: This code identifies two situations. In the first, the boundaries of a district are identical to what they were before the election in question, but the district designation has changed. For example, Alaska Senate District T in 2004 was the same geographic area as District S in 2000. In the second situation, districts in a chamber generally have been redrawn, but not the district in question, and its district designation has not changed. However, a code of "2" is not given if a situation of "holdover redistricting" is involved, which would result in a code of "7" (see below).

3: district represents the same geographic locale as it did in the past, but district type has changed (dtype) or the number of legislators elected from the geographic locale has changed (dseats). However, a dtype change between 4 and 5, 5 and 4, 4 and 6 or 6 and 4 would not be enough to bring about a code of "3."

4: situations that involve the elements of the situations that bring about both a code of "2" and a code of "3."

5: Nothing about the district has changed since the last election, but it is in a relationship with another district—such as a nested relationship or a floterial relationship—that has either been redrawn, changed its number of seats, or has switched to or from an FFA method of determining winners.

6: Holdover redistricting: District has been redistricted since the last time the seat or seats up in this election were up. However, district was redistricted because of a new map put in place before the last election. For example, a new map was created in 1981, half the state senate seats were up in 1982, and the other half were up in 1984, while the map was not changed between the 1982 and 1984 elections. The districts having elections in 1984 would be coded "6."

7: Same as a code of "2" except has the properties of holdover redistricting described for code "6."

8: Holdover redistricting for alternating seats: District has been redistricted since the last time the seat or seats up in this election were up. However, district was redistricted because of a new map put in place before the last election and the district in question has alternating seats, one of which has already had an election immediately after the district was redrawn. For example, a new map was put in place for the 1982 elections to the WV State Senate, where each district has two state senators with alternating seats. Elections in 1984 then receive codes of "8" for redist (assuming that seat was not up two years before, which might be the case if a vacancy is involved).

Codes of "8" also result because of etype="gs" elections in the redistricting year for one seat in a district that is usually an FFA-MMD. For example, in the ND Sen 1972 there was one seat up in district #5 in a etype="gs," a four member FFA-MMD, and then all four seats were up in 1974. In 1974, all four seats received a score of "2."

9: Same as a code of "2" except has the properties of holdover redistricting described for code "8."

0: else.

Note: this definition means that the primary and general elections for one seat would both receive the same redist code in the same year (or more accurately, in the same election cycle), as would the runoff after such a general election if one occurred.

UOA: a district-post-election period (i.e., sid sen dname dno geopost mmdpost year if no special elections are involved, and excepting early Louisiana elections).

redist1, redist2 and redist3

These record the district designation the district was called before redistricting if its borders were unchanged. In other words, if the variable “redist” has a code of “2,” “4” or “5,” this set of variables records a district designation. “redist1” records the “dname” of such a district, “redist2” records the “dno,” and “redist3” records the “geopost,” as appropriate.

Be forewarned: sometimes a district changes its district designation, represents the same geographic locale as a prior differently designated district, but in between when those two districts were up, another district with the same name as the district designation that was earlier in time was up. For example, in 1972, district F was up in the AK senate. In 1976, district K, that represented the exact same geographic locale as 1972 district F was up. But District F was also up in 1974, although representing a different geographic locale than District F in 1972 and District K in 1976. If year+termz is used in the code when creating lagged variables, this should not cause problems.

regime

Short for “redistricting regime.”

Generally, gives the year of the first general election held under a new map the district in question was redrawn in. The same year appears for this variable in every election the district has until it is redistricted again. Note the specific district may not have had an election in that year. For example, senate district #14 in CA was up in 1990 and 1994, but since redistricting occurred in CA for the 1992 elections, it is given a score of 1992 for 1994 and 1998 cases.

Specifically, if the variable “redist” has scores of between “1” and “5” for a state chamber in an election year, that election year is assigned as the score for “regime” for the rest of the decade for those districts, if there is no further redistricting. If districts in the chamber are up in the next election that were not up in the first election after a map has been redrawn, and they have “redist” scores of “6,” “7” or “8,” these are assigned the “regime” year that the districts up two years prior have.

Note: when there is a small amount of redistricting in a state between Censuses (i.e., “re-redistricting” that isn’t statewide), the newly drawn district takes on a value of “regime” that does not match the other districts in the same chamber and year.

Note: an implication of these coding rules is that state senates with four-year non-staggered terms that have two year terms for the state house (i.e., KS, MI, NM, SC, VA), when the state senate is up two years after the first election in which the entire state house has been redistricted, the entire house gets a value for “regime” that is different than the value the senate receives. For example, for the 2000s, the VA HS gets a value of 2001 while the VA Sen gets a value of 2003.

A potential complication for code is caused by the following oddity. Sometimes a state senate district is redistricted that has nested districts inside of it mid-decade. Sometimes one or all of those nested districts are not redistricted. When this happens, the variable “redist” is given a code of “5” for the district that was not altered, as explained in the description of “redist,” which gives them both the same regime score. This practice facilitates comparison between nested house and senate districts in this situation. A similar oddity occurs once with a floter district.

UOA: the highest level of aggregation that are never split by a district-regime are groups of cases defined by an election (i.e., by year month day sid sen dname dno geopost mmdpost etype). District-regimes are in turn nested inside of districts (i.e., groups of cases defined by sid, sen, dname, dno, geopost).

DISTRICT RELATIONSHIP VARIABLES

flot

Codes of “1” mean that the district an election is in is a floter district. Codes of “2” mean that a district is encompassed by a floter district. Other cases are coded “0.”

UOA: a district-regime.

flot1, flot2 and flot3

Relevant district designation of floterial district the district at hand is nested inside of. “flot1” records the “dname” of such a district, “flot2” records the “dno,” and “flot3” records the “geopost.” These variables are only observed for cases in districts that are nested inside of a floterial district (i.e., the variable “flot” has a code of “2”).
UOA: a district-regime.

nest

Code for the relationship between state house and senate districts, involving nesting or identical districts.

A code of “1” means the district in question is either a house or senate district that is identical to a district in the other chamber.

A code of “2” means that a district is part of a nested relationship between state house districts and a state senate district. In a nested relationship, two or more state house districts are inside of the state senate district, and together define an area equal to the senate district. According to this definition, a house district is not nested even if it is completely included inside a senate district, if any of the other house districts that appear in that senate district are not entirely encompassed by the senate district.

A code of “3” means one of two things. The first is that a state senate district is identical to a state house district, but that it also nests two or more other state house districts inside of it. This is only observed in the New Jersey State Senate in 1971. A second meaning of a code of “3” is that a state house district is identical to one state senate district, but along with one or more other state house districts is nested inside of a different state senate district. This is only observed in the Nevada Assembly in 1968 and 1970. Both situations are only possible when a floterial district in the other chamber is involved. In the latter case, which State Senate district nests state Assembly districts can only be inferred from examining flot1-flot2-flot3 along with nest1-nest2-nest3.

A code of “0” indicates that there is no identical or nested relationship between the district in question and districts in the other chamber.

nest1, nest2 and nest3

Relevant district designation of state senate district a state house district is nested inside of or identical to. These variables are only observed if nest has a value. “nest1” records the “dname” of such a district, “nest2” records the “dno”, and “nest3” records the “geopost”.

When the variable “nest” is “3” and the case in question is in the state house, only the designation of the state senate district the state house district is identical to is reported. Which state senate district those districts are nested inside of can be inferred by examining flot1, flot2 and flot3.

ELECTION ATTRIBUTES

etype

Election Type code

UOA: an election / a district-post-regime-year-month-day (i.e., year month day regime sid sen dname dno geopost mmdpost).

It might seem unnecessary that “regime” must also be used to define what the unit of analysis of this variable is. However, sometimes special elections from a previous map are held concurrently with the first general election of a new redistricting regime to fill a vacancy. The variable “regime” differentiates the two groups of cases in this instance.

By design, all codes involving a general election start with the letter “g,” with the exception of codes of “ssg” and no code associated with a non-general election starts with a “g.” Similarly, by design, all codes involving a special election start with the letter “s,” with the exception of codes of “gs” and no code associated with a non-special election starts with an “s.”

The most common codes for this variable are as follows

g: general election

s: special election

gs: Special election held concurrently with general elections in other districts for the remainder of a term that would not normally be up (i.e., a senate seat with a four-year term has an election two years after the last election). Note: this would also include a state senate election up after two years in an even year November election when there are four year non-staggered terms present.

ssg: special statewide general election. All seats in a chamber are up, but not at a normally scheduled time. Usually by court order.

ttp: Top two primary. Top two vote getters go on to general election, regardless of party. Observed in CA, NE and WA.

nps: non-partisan special election.

The following codes are combined with other codes for the same case. The prefixes to such codes are as follows.

dp: Democratic primary

rp: Republican primary

la: System of elections utilized in Louisiana

For special election primaries, the code for the special election is placed in front of the party's primary for that special election. For example, "sdp" is a Democratic primary for a special election.

The following codes sometimes follow the code prefix, but only appear for 1) primaries elections in states where primaries can have a runoff, 2) special elections in states where special elections can have a runoff 3) a few general elections in Georgia that had runoffs, 4) a general election in Vermont that had a tie for second place, resulting in a runoff and 5) in Louisiana.

fsettle: first round settled / first round where a runoff is possible, but a runoff did not occur.

funset: first round unsettled / first round where a runoff occurred after this election.

fpartunset: first round partially unsettled / first round where a runoff occurred after this election, but where some candidates in the current election did obtain enough votes to "place" (i.e., usually be put on the general election ticket of a party). For example, in a primary for a free-for-all multimember district with two seats, one candidate in the primary might obtain enough votes to qualify to run in the general election for their party, but no other candidate does. In that case, the second and third top vote getters would go on to a runoff, while the top vote getter would not.

runoff: primary, special or general election runoff

Non-major party primaries are assigned codes as follows. The codes start with the name of the party, all lower case, but sometimes abbreviated (see below), followed by "p," followed by "first" (if appropriate) and followed by "settled" (if appropriate). Coding non-major party primaries in this way is necessary because sometimes there are two or more of these primaries at the same time in the same district. The following abbreviations were used for this purpose.

independence indep

independent indep

american amer

constitution const

constitutional const

libertarian liber

progressive prog

workingfamilies work

upf. Primary of unknown type (observed in LA in 1967 and 1971), first round. Probably a Democratic primary.

uprunoff. Primary of unknown type (observed in LA in 1967 and 1971), runoff. Probably a Democratic primary.

Minutia

When a candidate was unopposed in Louisiana for the year 1975, the state of Louisiana reported them in the returns as running in the second round, and so are given codes of "larunoff." When a candidate was unopposed in Louisiana for the years 1979 to present, they were not put on the ballot, and are given codes of "lafsettle."

Some states list independent candidates among the primary results merely because they filed to run, not because they had to compete in a primary to obtain access to the general election. If two or more such candidates appear in a district-post's "primary," numbers are given to these candidates' codes to differentiate their "primaries."

Codes of "dpf" and "rpf" are associated with cases that need to be looked into more.

Some primaries have no filed candidates but despite that write-in candidates are reported, even if none of those write-ins win the right to run on the party's label the primary is for. When this occurs, the string "doesnotcount" is added to the primary designation (i.e., "dpdoesnotcount" or "rpdoesnotcount"). These cases appear in New Hampshire and Tennessee.

deter

Dummy: A "1" is assigned to elections whose winner goes to sit in the legislature at the beginning of the legislative session. For other cases this variable is coded "0." By design, a score of "1" is assigned to elections in such a way that all seats in the legislature are accounted for when both "year," "termz" and "outcome" are taken into account.

If a candidate was running in a general election (etype =g) this variable is almost always coded "1." It is also generally coded "1" when "etype" is coded "gs."

Rarely, a special election (etype=s) will be coded “1” for this variable. This is the case when the election was used to fill a seat that was vacant because the results of a general election were indecisive (for example, if the normally scheduled general election was not held because of a death, if the results of the general election were thrown out, or if there was a tied vote).

For cases that represent the first round of an election in Louisiana from 1975 through present (“etype”=laf), this variable is coded “1” when the first round did not result in a runoff and “0” when it did. The second round of Louisiana elections (“etype”=lar) are always coded “1.” The same goes for any general election for which the first round was not decisive (a small number of Georgia and Vermont elections).

UOA: an election

eseats

For general elections outside of Louisiana, this reports the number of legislative seats to be filled in an election. If it is a partisan primary, it is the number of positions on the ballot to be filled by that party, unless a runoff situation is involved.

If this is a post-MMD district (“dtype”=2), this variable is coded “1” as there is only one winner for that particular post.

For top two primaries (i.e., “etype”=ttp), this variable is coded “2” because the top two vote getters go on to the general election.

Some states require runoffs for primaries, with different rules about the conditions that require a runoff.

Georgia and Vermont also have provisions for runoff elections for general elections, although this only occurs twice in Georgia and one in Vermont. In Vermont, runoffs only occur in the case of a tie. In the 1986 Vermont general election to Orleans #3, an FFA-MMD election with two seats, there was a tie for second place. In this situation, “eseats” is coded “3” for the first round, and “1” for the second round.

For the first-round election for an SMD in a state where primaries have a runoff if the top candidate fails to obtain a sufficient percentage of votes, a “1” is put in if the first round is decisive and a “2” is put in if the first round results in a runoff. Louisiana’s elections are coded in the same way. This is a corollary to the more general rule that if you add the “r” and “w” candidates together in a round, that will equal “eseats.”

Some primaries have no filed candidates but despite that write-in candidates are reported, even if none of those write-ins win the right to run on the party’s label the primary is for. When this occurs, “eseats” is coded “0.” These cases appear mostly in New Hampshire and Tennessee.

UOA: an election

term

Anticipated Length of Term

Measures the length of the term, in years, for which a candidate is running. This variable is based on the normal expected term length for that particular seat at that particular time.

For special elections, the term length is still recorded as the full length of the term for the seat being filled and not as the fraction of the term remaining. However, for instances where “etype” is coded “gs,” this variable is coded “2,” and not “4.”

A code of “9” is given for this variable when term length is not known during an election. This occurs in states where all senators are up in years that end in “2” and then term lengths are randomly assigned after the election (Arkansas). Nor is it accurately measured in states where a senator’s term length is uncertain when they are elected in years that end in “0” because their term length may be contingent on whether their district will be redrawn for the elections in the year ending in “2” (Alaska). Nor is it accurate for state senators elected in years ending in “0” where odd numbered districts are up in one election and even numbered districts are up in the next election, and senators who are assigned to a district with the opposite type of number (odd or even) must run after a two-year term (Iowa in 2002 and before). They are also assigned a “9.”

For a few cases where a general election does not fill a seat—for example, in the case of an election where only one candidate is running, but that candidate dies and no replacement is provided for the general election—and the seat is filled in the next calendar year, .5 is subtracted from the standard term length. For example, in the election for Florida House district #13, the single candidate for that seat did not run in the general election of November of 2014. A special election was held to fill that seat on February 17, 2015. “termz” is coded “1.5” for that election instead of “2.”

UOA: a district-post-regime-year-month-day

termz

Actual length of term, in years.

This variable represents the number of years between the election represented by the case and either: 1) the next time the winning candidate is up for election after continuously serving in office; or 2) the time at which the winning candidate leaves office before their term is up but only when a “etype”=“gs” election fills that vacant seat (this is only a small proportion of times in reality where a legislator leaves office early). Records of legislators leaving office early for the vast majority of cases have not been brought into the database yet.

Note: This term length merely takes that seat to the next time it was up. If, say, a legislator left office one year after election, that isn’t registered in this variable. In the case of a state with four-year state senate terms that do not have “etype”=“gs” elections, a state senator that leaves before two years are up, and is replaced by gubernatorial appointment or a special election that does not take place in November of an even numbered year (or an odd-year in the case of odd-year election states) is still given a “4” for this variable.

By design, “year,” “termz,” “deter” and “outcome” used in conjunction to create lists of state legislators in each calendar year will always result in a number of legislators that matches the number of seats in the legislative chamber.

As with “term,” .5 is subtracted from “termz” when a general election does not fill a seat and a special election fills that seat early in the next calendar year.

UOA: an election-candidate

CANDIDATE ATTRIBUTES

cando

Candidate name on returns.

UOA: an election-candidate

candformat

Format of candidate name in returns: “first name first” or “last name first.”

UOA: an election-candidate

cand

Full standardized candidate name, always identical and unique for one individual within one state.

“writein” indicates votes cast for candidates identified in the official returns only as write-ins.

The same values of this variable may be held by separate individuals, but only if they are in separate states.

By design, this value is made unique for one candidate within one state.

UOA: a candidate

The variables “cand,” v38 and v39 have candidates’ full names. v38 is the name of the candidate as originally entered in the ICPSR data set #8907. It has observations for 1967 through 1989 only. v38 attempts to record the name as it appeared on the election returns or newspaper accounts used for ICPSR’s original data gathering process for Study #8907. The name as recorded by ICPSR in v38 included a high percentage of cases in which an individual who appears in the data set more than once had his/her name recorded differently across observations (e.g., Mary Jones & Mary S. Jones). These inconsistencies made it impossible to reliably track candidates over time. v39 is the candidate name variable adjusted by Berry and Carsey for a prior release of the dataset so that each person who appears more than once has his/her name recorded in an identical manner across observations. This “clean” version of the candidate name variable exists for virtually all observations in the data set. A major portion of their work centered on the construction of v39 in order to provide a reliable indicator of whether each candidate appears elsewhere in the data set. A more detailed discussion of the name cleaning process is provided in the codebook for ICPSR Study #3938. In 2010-11, Klarner also engaged in a number of routines to check whether candidates across time in the earlier dataset were the same individual. Part of this process involved assigning the various name components of candidates (i.e., last names, first names) to separate variables. These name components can be found in variables with titles like “first,” “last,” “middle,” etc. “cand” is the variable containing the full name of a candidate for this version of the dataset, standardized over time so they are the same for a candidate who was judged to be the same individual. These values are also unique to the individual they represent within one state. The name component variables are often not standardized over time. The original name components from the returns should have been saved without alteration alongside the standardized name components, but this was often not done. Recollecting the entire dataset “from the ground up” would solve these problems, and with today’s technologies would be much easier than doing the same task in the past.

candid

Unique identification number given to each candidate.

This variable is generally equal to the minimum of caseid for an individual. However, this rule does not always hold. Since the last edition of SLERs, corrections have been made to whether two entries pertained to the same individual or two distinct individuals. The strict application of this rule would result in values for “candid” that differed more than necessary between the version of SLERs (i.e., 1967-2010 versus 1967-2015). Changes in “candid” between the two versions of SLERs were minimized. This was done to minimize the time necessary to merge candidate information for scholars whose data is linked to SLERs by the old version of “candid.” Note that this procedure caused a small number of candidates to have a value of “candid” that is equal to caseid for a case associated with a different candidate (although “candid” is never the same for two individuals). Scholars utilizing the old version of “candid” for merging external data are cautioned to be aware that corrections to “cand” may cause errors in matching.

For “scattering” and “writein,” it is the minimum of caseid within a group of cases defined by sid sen dname dno geopost mmdpost year etype. This is almost always equal to caseid for the case in question, the exceptions being where “scattering” or “writein” appear when county breakdowns exist.

UOA: a candidate

partyo

Party designation of candidate on ballot, mostly unaltered from returns.

When keys were available, I substituted party names for codes.

Years available for: 2013-2016

UOA: a candidate-election-party line

party

Detailed Party Name

In earlier editions of SLERs, a code was assigned to parties. Many of these codes were ICPSR party codes. Now that memory is less of an issue, the party names associated have replaced the codes to remove the need to look them up. Some codes with lengthy explanations have been left in, and the descriptions of these codes appear in Appendix C.

These party designations were changed to all lower case, all spaces and all other non-alphabetic or non-numeric characters were removed (with the exception of forward slashes), and the string “party” was often removed. When votes for two or more parties were reported on the same line for a candidate, the word “fused” was removed if it appeared, and forward slashes were put between the parties.

Furthermore, for the 2016 elections and on, detailed party codes stopped being assigned to new non-major party designations that appear in the returns.

UOA: a candidate-election-party line

partyz

Simplified Party Code

The parties in the variable “party” were reduced to one of seven codes. Sometimes the same value in “party” would result in a different value in “partyz” based on extra research. The codes are as follows.

d = Democratic (includes instances where non-major parties fuse with a Democrat and the votes aren’t reported separately)

r = Republican (includes instances where non-major parties fuse with a Republican and the votes aren’t reported separately)

b = Fused Democratic and Republican (NH and VT, some CA and MA)

nonmaj = non-major party (third party or independent)

nonpart = non-partisan election (i.e., Nebraska, Minnesota (1972 and before only), some special elections)

writein = write-in / scattering

partymiss = unidentified party

UOA: a candidate-election-party line

partyt

Assigns just one party to a candidate in one election season (i.e., the primary and general election in one year), using the same seven codes used in partyz.

For example, a candidate running in NY with fusion as a Democrat and a Republican is assigned the party they are expected to caucus with in the state legislature, measured by how they end up caucusing in the state legislature.

For example, a candidate who files in a state primary as a Democrat, and then is written in by voters in the Republican primary, has “d” designated as their “true” party.

UOA: a candidate-election

partytsource

Brief description of decision to code “trueparty” the way it was, when a candidate ran as both a Democrat and a Republican in the same election. Also gives source of external information and “d,” “r,” “dr,” or “rd” when the Council of State Governments Directory I (or Supplement #1 to the Book of the States) identifies the candidate as such.

Values only exist for this variable for candidates who ran as both a Democrat and a Republican (fusion).

UOA: a candidate-election

firstcase

Indicates whether incumbency and tenure in the other chamber of the legislature cannot be observed because the case is in the beginning of the dataset.

This variable is coded “1” if it is not possible to observe whether a candidate is an incumbent and it is also not possible to observe whether a candidate has served in the other chamber immediately before the election because the case is in the beginning of the dataset. A score of “2” is given to situations where it is possible to observe whether a candidate served in the other chamber immediately before the election, but it is still not possible to observe whether they were an incumbent or not. All other cases are system missing.

All cases in the first year a chamber has general elections are coded “1” for this variable. For the second year in which a chamber with staggered seats has elections, cases are often “2,” are sometimes “1” and are sometimes “0.” Chambers that do not have staggered seats are all “0” in the second year they have elections. A score of “2” generally results in the second year a state senate with staggered seats has elections.

If a candidate runs in a general election, and then runs again in another seat in that chamber, the case will be coded “0” for them even if that is the first time the later seat was up for election. For example, if a state senate candidate loses an election in 1968, and then runs again in a different district in 1970, even though that district has not been up for election yet, that candidate will be coded “0” for this variable in 1970, although their opponent may not be.

Cases are often coded “2” in the case of posts for alternating seats with four-year terms the second time a post is up in the dataset. For example, what is designated “post 1” (mmdpost) of district #1 of the WV state senate was up in 1968. “Post 2” of that same district was up in 1970 and is coded “2” because it is possible one of the two candidates is an incumbent and this cannot be known from SLERs, although whether they served in the West Virginia House of Delegates via election in 1968 is known.

UOA: a candidate-election

prior1

This variable reports the election year before SLERs begins that resulted in continuous tenure in the legislative chamber a candidate was running for. If they had no prior tenure in the chamber, they are given a value of “0.”

This variable is only observed for candidates who appear in the beginning of the dataset. In other words, this variable only has values if firstcase is “1” or “2” and is system missing when firstcase has a value of “system missing.”

Note that if there is a gap in service in the chamber before SLERs begins, the first election year in which the legislator won after the last such gap is what is recorded.

To code this variable, it was assumed that a state senator with a four-year term would not run for the state house two years into their term. For example, if a candidate ran for the state house in 1968, lost that election, and then ran for the state senate in 1970, it was assumed they were not an incumbent state senator in 1970.

UOA: a candidate-election

prior2

This variable is only observed for candidates who appear in the beginning of the dataset (i.e., have values of “2” for “firstcase”).

It reports the election year before SLERs begins that resulted in continuous tenure in the legislature up to when the election the case in question was in occurred.

Note that if there is a gap in legislative service before SLERs begins, the first election year in which the legislator won after the last such gap is what is recorded.

To code this variable, it was assumed that a state senator with a four-year term would not run for the state house two years into their term.

If this variable reports a year, and prior1 has a value of "0," the candidate it is associated with has an "exper" value of "other."

UOA: a candidate-election

exper

Code for prior state legislative experience of candidate.

Brief descriptions of this variables five possible codes are as follows.

inc: incumbent: served in chamber running for immediately before the election

oth: served in the other chamber immediately before the election.

pastinc: served in the chamber before the election, but not immediately before.

pastoth: served in the other chamber before the election, but not immediately before.

pastboth: served in both chambers of the legislature before the election, but neither one immediately before.

More detailed descriptions of these codes follow.

inc: For most cases, it measures whether someone won a general election, and then ran in the same chamber in an election that occurred when the term of their current office was served (i.e., "year" of past election + "termz" = "year" of current election).

There are some instances when an incumbent chooses to run in an election to the same chamber before their entire term has been served. For example, a candidate for the California State Senate might win in 1990, and rather than serve out their four-year term, decide to run before their term is up to take advantage of a new district they wish to run in that is having an election in 1992. Such candidates are coded "inc" for this variable.

For the most part, legislators who were appointed to the term preceding a general election, or legislators elected in a special election, are not coded "inc" for this variable. This is because most instances of such legislators are not recorded in SLERs. For the purposes of this variable, legislators who came to office through "etype"="gs" elections in their last election are considered to be incumbents. Legislators who came to office in a general special election (i.e., the Rhode Island State Senate in 1983, the Alabama State Senate and House in 1983) are also considered to be incumbents.

For the beginning years of a state in the dataset (cases for which the variable "firstcase" is "1"), codes of "inc" are assigned to candidates who were listed as serving in the same chamber they are running for immediately before the election in *Council of State Governments, Book of the States Supplement I: State Elected Officials and the Legislatures*, Lexington, KY: Council of State Governments (various years). Codes of "oth" are assigned similarly from this source for early years in the dataset. Some of these incumbents may not have come to office through general elections, but since most did, they are coded "inc" anyway.

Occasionally a voter writes in the name of an individual who is in fact an incumbent—but often the person is an incumbent in a neighboring district. Or a voter may write in the name of an incumbent who has retired. Calling these candidates incumbents would be problematic when assessing trends in incumbency, or assessing the impact of incumbency on vote share. For example, of the 161 write-in candidates who were coded as incumbents in ICPSR #21480, 100 received less than ten votes. Very rarely, a case can be made that a write-in candidate constitutes a "serious candidate" and should thus be counted as an incumbent. For this edition of SLERs, write-ins are coded "inc," "oth," etc. for this variable using the same method as every other candidate, to enable scholars to deal with them as they wish.

UOA: a candidate-election

vote

Vote total for a candidate in an election. Sometimes the vote total for a candidate in an election for one party line, or for a candidate in one county within a legislative district, or both.

This numeric variable records the vote total received by the candidate for the party designated in the variable "party." In other words, for instances where a candidate is running on multiple parties' lines, and where the votes cast for these separate parties for the same candidate are tabulated separately, only the vote total for that party is given for this variable.

Votes identified as "scattering" or unidentified write-ins are widely but inconsistently reported in returns, and the inclusion of these amounts in this dataset is also inconsistent. For the data entered in 2002 and 2003, every name that appeared on the election returns was entered regardless of the vote total, and all "scattering" was collected. For 2009 to 2010 as well, vote totals of "scattering" and "write-ins" were always collected if reported in the returns on official state Web sites. These are reported in party as "writein" or "scattering."

UOA: a candidate-election-party line

outcome

Election Outcome

This variable that reports the outcome of the election for each particular candidate. The variable is coded “w” for candidates in a standard general election who go on to sit in the legislature, and “w” for candidates in standard primary elections who go on to compete in the general election.

Louisiana’s two round elections, top two primaries, and party primaries that can have two rounds are coded as follows. In first round elections that are not followed by a runoff, the candidate receiving the most votes (or enough votes to go to the legislature in an FFA-MMD) is given a score of “w” (which stands for “winner”). If a first round is followed by a runoff, the candidates who go on to the runoff are given scores of “r” (which stands for “runoff”). The candidate receiving the most votes in the second round is then given a code of “w.” This is also how runoff general elections are coded in Georgia.

Note that there will always be two rounds for a top-two primary state, such as California, Washington, or Nebraska. Note that a trigger for a runoff is the same in Louisiana’s system and primaries in the states that have them (which are generally in the South). If no candidate receives 50 percent or more of the vote, there is a second round. What makes Louisiana’s system unique is that there is no partisan component, but a second round may not happen.

Note: A tie in the Vermont House’s Orleans 3 district in the 1986 election caused the two candidates who tied for second place to go to a special election. Therefore, although the district has two seats (an FFA MMD) only one candidate in that first round was coded as winning (“w”) while two that went on to the runoff were given “r.”

UOA: a candidate-election

hold

Dummy: 1 = holdover legislator, else = 0.

A holdover legislator is one who is elected to a four-year term (i.e., as defined by the variable “termz”) who is in a state that is either entirely or mostly redistricted midway through their term. For example, Kate Brown was elected to the Oregon State Senate to district #7 in the November 2000 election to a four-year term. The Oregon State Senate redistricted for the 2002 elections. Legislators in this situation are given a score of “1.” Moreover, Kate Brown was also assigned a new district number for the 2003-2004 session, district #21. Holdover legislators are always state senators with the exception of North Dakota state house members from 2000 on.

Many states do not assign a new district number to holdover state senators. However, such state senators are still considered “holdover senators” whether they are formally assigned a new district number or not and are given a score of “1.” Furthermore, all candidates in an election along with holdover legislators are coded “1.”

Oregon State Senators elected in the November 1970 election were said to not have a district number in the 1973-1974 Oregon Blue Book. As a consequence, the string “no district dez” is input into “hold1” and “333333” is assigned to “hold2” for these senators.

Note: Many holdover Senators who are not formally assigned new district numbers will run in a district with a different number than the district number they won under before redistricting. Senators also sometimes run in different districts than their holdover district.

Note: It is always the case that legislators’ holdover status is observed to start two years after their election, although this does not necessarily have to be the case. One unusual case, which is still consistent to the prior rule, is represented by the Kentucky Senators elected to five year terms in the general elections of November 1981, and they became holdover senators after districts were redrawn for the November 1983 general elections. This occurred when Kentucky switched from odd to even year elections for the state legislature.

UOA: an election contest

hold1 and hold2

District designation assigned to holdover state senator or North Dakota state house member. This is reported whether their holdover district designation is different than the one they were elected to or not. “hold1” records the “dname” of such a district and “hold2” records the “dno” value of such a district. The variable “hold3” is not included, as no holdover legislator district designation has a “geopost” value.

Furthermore, as long as only one district designation is given as a holdover district to the winner or winners in an election, all losers in that election are given the same holdover district designation as the winner or winners. There are 24 cases that have divergent holdover district designations for the winners, and so do not follow this rule.

UOA: usually an election but there are 24 cases where a candidate-election is the unit of analysis, as discussed above.

tenure1

Number of years a legislator has continuously served in the legislative chamber when the election in question was conducted. When a candidate first wins, they are given a score of “0.” In the next election, they are

given an amount equal to the number of years they have already served, etc. Candidates who were not serving in the chamber immediately before the election are given scores of “0.”

This variable takes legislative service before the dataset began into account.

UOA: a candidate-election

tenure2

Same as tenure1, except that it records continuous service in the legislature, not the legislative chamber.

UOA: a candidate-election

limit

Code indicates when a legislator faces a term limit.

1 = cannot run again, will be termed out at the end of their current term.

2 = can run again, but cannot run in the election after that, and the term served after the next election will be two years.

4 = can run again, but cannot run in the election after that, and the term served after the next election will be four years.

3 = can run again, but cannot run in the election after that, and the term served after the next election will be either two or four years. This situation is brought about by uncertainty about how long term in question will be in a state senate with four-year terms where all legislators are up in years ending in “2” such as Arkansas or Florida.

Coverage: candidates running in the 1967-2010 elections.

UOA: a candidate-election

NAME COMPONENTS

As a whole, the following name components have the property of not varying within one individual within one election (i.e., within a group of cases defined by sid-year-sen-dname-dno-geopost-mmdpost-etype).

Furthermore, the alternative candidate last name variables (last1, last2, last3 and last4) and the four variables describing those (des1, des2, des3 and des4) have the property of not varying within a group of cases defined by “candid.”

In earlier editions of SLERs, “last” was made to be constant within one individual (“candid”) which created some advantages. However, overall, this was a mistake to do. Much of the damage of that mistake has been done, but the practice is not being continued.

last

Last name of candidate.

Last names were standardized to always be the same for one individual for this variable, but not the other name components. Unfortunately, what a candidate’s last name was in which specific elections (such as, for which elections their maiden last name was run under) was often not retained when this standardization occurred.

Alternative last names the candidate was potentially known by are recoded in other variables.

UOA: a candidate-election: at first, this was made to not vary within an individual (i.e., candid) but were later allowed to vary.

first

First name of candidate.

UOA: a candidate-election

middle

Middle name of candidate.

UOA: a candidate-election

middle2

Extra middle name or names of candidate.

UOA: a candidate-election

For three cases, the candidate has a second, third and fourth middle name component. For ten other cases, the candidate has both a second and third middle name component.

nick

Nickname of candidate.

UOA: a candidate-election

suffix

Suffix in name of candidate.
UOA: a candidate-election

prefix

Prefix or title in name of candidate.
UOA: a candidate-election

namenum

Number used to differentiate an individual from other individuals with the same name in the same state.
UOA: a candidate

last1, last2, last3 and last4

First, second, third and fourth alternative last names that an individual might be known by.
These are either alternative spellings of a last name, maiden names, components of a compound last name, or alternative compound names.
It is often the case that how names are reported in election returns do not allow knowing whether a particular name component is a middle name, or a component of a compound last name.
UOA: candid: these were made to not vary within an individual.

ltype1, ltype2, ltype3 and ltype4

Description of alternative last name listed in last1, last2, last3 or last4, respectively.
These descriptions have been standardized and follow the following conventions.
Almost every entry ends with either “known,” “probable,” or “possible” to indicate the level of certainty I had when creating the entry.
Common entries are whether an alternative last name is a part of a non-hyphenated compound last name, is part of a hyphenated last name, is a maiden name the candidate has run under in the past (and what elections they ran in with that name), is a former married name (after divorce), or is a misspelling of the candidate’s last name. Sometimes two different ways a last name or last name component are alternative last names are attached together, each ending with one of the three strings indicating level of uncertainty.
UOA: candid: these were made to not vary within an individual.

lastb, firstb, middleb, middle2b, nickb, suffixb, last1b, last2b, ltype1b, ltype2b

Name component from a second source.
Note: “ltype1b” and “ltype2b” have not been standardized in a way that matches “ltype1,” etc.
UOA: a candidate-election

namebsource

Source of name component in “b” suffix name variables.
UOA: a candidate-election

ARCHIVED VARIABLES

v38 Candidate’s Name used in ICPSR #8907 (i.e., which covered 1967 to 1989)

v39 Candidate’s Name used in ICPSR #21480 (i.e., which covered 1967 to 2003).

v56 State legislative district number from ICPSR #21480.

This is the ICPSR #21480 variable named “v8” exactly as it appeared in that data set for the case in question.

v57 State legislative district post number from ICPSR #21480.

This is the ICPSR #21480 variable named “v9” exactly as it appeared in that data set for the case in question.

v58

State legislative district name from ICPSR #21480.

This is the ICPSR #21480 variable named “dname” exactly as it appeared in that data set for the case in question.

v18_20171211

Candidate ID code from the December 11, 2017 version of SLERs.

v19_20171211

Candidate standardized name from the December 11, 2017 version of SLERs.

v19_20160217

Candidate standardized name from the February 17, 2016 version of SLERs.

Appendix A: Key of Old (ICPSR #21480, covering 1967-2003) and New (1967-2010) Variable Numbers

| Variable Name in ICPSR #21480 (1967-2003) | Variable Name in ICPSR #34297 (1967-2010) | Variable Name in 1967-2015 Update | Variable Name in 1967-2016 Update |
|---|---|--------------------------------------|--------------------------------------|
| caseid | caseid | caseid | caseid |
| v1 | dropped | | |
| v2 | dropped | | |
| v3 | dropped | | |
| | v01 | same | sid |
| v34 | v02 | same | sab |
| v35 | v03 | same | sfips |
| v5 | v04 | dropped | sicpsr |
| | | countyname | cname |
| | | countyfips | cfips |
| v6 | v05 | same | year |
| v12 | v06 | same | month |
| | | | day |
| v7 | v07 | same | sen |
| distname | v08 / v58 | | |
| | v08 | v08 | dropped |
| | v08 | v08z | dname |
| v8 | v09 / v56 | | |
| | v09 | v09 | dropped |
| | v09 | v09z | dist |
| v9 | v10 / v57 | | |
| | v10 | dropped | |
| | v10a | v10a | dropped |
| | v10a | v10az | geopost |
| | v10b | v10b | dropped |
| | v10b | v10bz | mmdpost |
| | v10c | v10c | dropped |
| | v10d | v10bz | mmdpost |
| | | v08, v09, v10a v10b, v10c | ddez |
| distid | v11 | dropped | |

| | | | |
|------------|-----------|---------|---------|
| v10 | v12 | same | dtype |
| v21 | v13 | same | eseats |
| termlength | v14 | same | term |
| | v15 | same | termz |
| v11 | v16 | same | etype |
| | v17 | v17 | dropped |
| | | v17b | deter |
| v4 | v18 | v18 | candid |
| v32 | v19 / v39 | | |
| | v19 | same | name |
| v14 | v20 | same | party |
| | v21 | same | partyz |
| v15 | v22 | same | exper |
| v13 | v23 | same | vote |
| v31 | v24 | same | out |
| v17 | v25 | dropped | dropped |
| v18 | v26 | dropped | dropped |
| v19 | v27 | dropped | dropped |
| v20 | v28 | dropped | dropped |
| v22 | v29 | dropped | dropped |
| v23 | v30 | dropped | dropped |
| v24 | v31 | dropped | dropped |
| v25 | v32 | dropped | dropped |
| v26 | v33 | dropped | dropped |
| v27 | v34 | dropped | dropped |
| v28 | v35 | dropped | dropped |
| v29 | v36 | dropped | dropped |
| v30 | v37 | dropped | dropped |
| v38 | v38 | same | same |
| v39 | v39 | same | same |
| v33 | v40 | same | dropped |
| | v41 | dropped | |
| | v42 | dropped | |

| | | | |
|--|-----|---|--------------------------------------|
| | v43 | same | uncert |
| | v44 | same | last |
| | v45 | same | first |
| | v46 | same | middle |
| | v47 | same | middle2 |
| | v48 | same | middle3 |
| | v49 | same | middle4 |
| | v50 | same | nick |
| | v51 | same | suffix |
| | v52 | same | prefix |
| | v53 | same | namenum |
| | v54 | altlast1, altlast2, altlast3, altlast4 | last1, last2, last3, last4 |
| | v55 | altlastreason1, altlastreason2, altlastreason3, altlastreason4 | ltype1, ltype 2, ltype 3, ltype 4 |
| | v56 | same | same |
| | v57 | same | same |
| | v58 | same | same |

Appendix B: State Designation Codes (sid, v02, v03 and v04)

| State | sid | sab | sfips | sicpsr |
|---------------|-----|-----|-------|--------|
| Alabama | 1 | al | 1 | 41 |
| Alaska | 2 | ak | 2 | 81 |
| Arizona | 3 | az | 4 | 61 |
| Arkansas | 4 | ar | 5 | 42 |
| California | 5 | ca | 6 | 71 |
| Colorado | 6 | co | 8 | 62 |
| Connecticut | 7 | ct | 9 | 1 |
| Delaware | 8 | de | 10 | 11 |
| Florida | 9 | fl | 12 | 43 |
| Georgia | 10 | ga | 13 | 44 |
| Hawaii | 11 | hi | 15 | 82 |
| Idaho | 12 | id | 16 | 63 |
| Illinois | 13 | il | 17 | 21 |
| Indiana | 14 | in | 18 | 22 |
| Iowa | 15 | ia | 19 | 31 |
| Kansas | 16 | ks | 20 | 32 |
| Kentucky | 17 | ky | 21 | 51 |
| Louisiana | 18 | la | 22 | 45 |
| Maine | 19 | me | 23 | 2 |
| Maryland | 20 | md | 24 | 52 |
| Massachusetts | 21 | ma | 25 | 3 |
| Michigan | 22 | mi | 26 | 23 |
| Minnesota | 23 | mn | 27 | 33 |
| Mississippi | 24 | ms | 28 | 46 |
| Missouri | 25 | mo | 29 | 34 |
| Montana | 26 | mt | 30 | 64 |
| Nebraska | 27 | ne | 31 | 35 |
| Nevada | 28 | nv | 32 | 65 |
| New Hampshire | 29 | nh | 33 | 4 |
| New Jersey | 30 | nj | 34 | 12 |
| New Mexico | 31 | nm | 35 | 66 |

| | | | | |
|----------------|----|----|----|----|
| New York | 32 | ny | 36 | 13 |
| North Carolina | 33 | nc | 37 | 47 |
| North Dakota | 34 | nd | 38 | 36 |
| Ohio | 35 | oh | 39 | 24 |
| Oklahoma | 36 | ok | 40 | 53 |
| Oregon | 37 | or | 41 | 72 |
| Pennsylvania | 38 | pa | 42 | 14 |
| Rhode Island | 39 | ri | 44 | 5 |
| South Carolina | 40 | sc | 45 | 48 |
| South Dakota | 41 | sd | 46 | 37 |
| Tennessee | 42 | tn | 47 | 54 |
| Texas | 43 | tx | 48 | 49 |
| Utah | 44 | ut | 49 | 67 |
| Vermont | 45 | vt | 50 | 6 |
| Virginia | 46 | va | 51 | 40 |
| Washington | 47 | wa | 53 | 73 |
| West Virginia | 48 | wv | 54 | 56 |
| Wisconsin | 49 | wi | 55 | 25 |
| Wyoming | 50 | wy | 56 | 68 |

Appendix C: Party Codes

4001 = no party identified in the data

4002 = Nebraska non-partisan, and Minnesota non-partisan (1972 and before only for MN).

4003 = non-partisan by virtue of the fact that it is a non-partisan special election.

4004 = party unknown, and write-in status unknown. Could be a major party, could be a non-major party.

4005 = some type of non-major party candidate. In instances with this code, the evidence for this conclusion was because the candidate in question wasn't listed in either the Democratic or Republican primary.

4006 = combination of 4005 and 99997.

4007 = returns explicitly say that the candidate isn't affiliated with a party.

4008 = candidate assumed to be a non-major party candidate because no party was listed next to their name in the returns, and parties were generally listed for candidates in those returns.

4009 = candidate assumed to be a non-major party candidate because they were not listed in a list of candidates filing for the Democratic and Republican primaries.

4010 = candidate assumed to be a non-major party candidate because the candidate wasn't listed in the returns.

4011 = candidate received a code of "908" in SLERs, which could indicate a non-partisan election (i.e., Nebraska), but candidate is running in a state and year with partisan elections. Candidate was judged to be a non-major party candidate.

4012 = originally given a "term" code of 2668 (nominated by petition) but in a state and year with non-partisan elections.

5000 = non-major party candidate of some type. Code not supplied by ICPSR and after I stopped creating new codes to cover such candidates on December 12, 2015.

Codes in the 60,000s: these were judged to be write-ins. The code follows the convention of adding the party code that was given to the case in the 5th edition to SLERs to 60,000. The decision to do this is based on the following considerations: 1) The candidate received ten or fewer votes in a general election. 2) The case exists in the 1967 to 1989 period, when write-in candidates were far more likely to be reported than in the later period. 3) There are only 33 candidates in the 1989 to 2010 period who received ten or fewer votes who weren't write-ins. 4) The candidates receiving "60,000 codes" were all clustered in NH, and clustered in a few years, indicating that their appearance was because of sporadic reporting practices, not surges in the number of candidates getting ten or fewer votes. 5) 130 of these 169 cases only involve one vote. 10 involve 2, 9 involve 3. 7) For other cases that seemed to be write-ins, where returns could be consulted, candidates consistently turned out to be write-ins. The NH returns were consulted, but they did not say "write-in." My judgment is that this was a reporting problem.

70100 = party code was originally "200" but changed to "100." Evidence for decision: candidate switched parties twice in two succeeding elections, and Supplement #1 to the Book of the States indicates they never switched parties. Or the opposing major party candidate in the election, whose party code would be the same if the other candidate's party was switched, leading to two Democrats contesting one SMD, etc. A large proportion of candidates fitting these criteria were found to have the wrong party code when state specific returns were consulted, but such returns weren't available for this case. Other evidence used is too lengthy to relay here.

70200 = party code was originally "100" but changed to "200." For logic, see code 70100.

Codes that start with 80,000 follow the following conventions. All codes that start with 80,000 were judged to be write-in candidates.

1s place: 1 = candidate received a very small number of votes, 0 = did not receive a very small number of votes.

10s place: 1 = case originally had a missing party code. 2 = case originally had a minor party code. 3 = case originally had a major party code. 4 = case originally coded as "Democrat." 5 = case originally coded as "Republican."

100s place: 1 = district is fully contested with major-party candidates, 2 = district is over-contested with major-party candidates of the party the candidate in question was originally given.

1000s place: 1 = candidate in two places at the same time.

99983 = judged to be a write-in because candidate was in an "over-contested" election, and candidate did not appear in the Democratic primary in consulted returns. Candidate originally coded as a Democrat.

99984 = assumed to be write-in because candidate was missing a party code in the original SLERs database and the candidate in question was running in two places at the same time, and the candidate in question received far fewer votes than in the other district.

99985 = assumed to be a write-in because the candidate is running in another district at the same time where they are receiving many more votes than in this district, and furthermore, in this district they are getting a very small number of votes.

99986 = assumed to be write-in because candidate was missing a party code in the original SLERs database and one of the following conditions was satisfied, except in the case of being in a state-year in which 1) there are lots of fusion lines or 2) where it is known that a lot of major party candidates are missing their party labels. 1) candidate received less than 200 votes AND less than 3% of total votes cast in the election, or 2) candidate received less than 1% of the total vote.

99987 = write-in, non-major party. Evidence: candidate appears in other districts and a very small number of votes are obtained by the candidate in this election.

99988 = Republican and Democratic write-in. Assumed to be so, because 1) the election is “over-contested” with Democrats, and 2) the line has a very low vote total. Was previously coded as simply R&D or D&R in SLERs.

99989 = Republican write-in. Just because the vote total is below 10. Not an over-contested election. Was previously coded as 200 in SLERs.

99990 = Democratic write-in. Just because the vote total is below 10. Not an over-contested election. Was previously coded as 100 in SLERs.

99991 = Republican write-in. Assumed to be so, because 1) the election is “over-contested” with Republicans, and 2) the line has a very low vote total. Was previously coded as 200 in SLERs.

99992 = Democratic write-in. Assumed to be so, because 1) the election is “over-contested” with Democrats, and 2) the line has a very low vote total. Was previously coded as 100 in SLERs.

99993 = known to be non-major party from returns, but party not specified, and party not one of the parties with a code in “term.”

99994 = combination of code “99995” and “99997.”

99995 = assumed to be non-major party candidate because it was given a party code that did not have a reported value in the ICPSR codebook, but other candidates (52 or so) with that same code were found to be non-major party candidates. Only 11 candidates.

99996 = candidate received ten or fewer votes and it was assumed that the candidate was a write-in candidate. Had a missing party code in ICPSR #21480.

99997 = Assumed to be a non-major party candidate because the election is “fully contested,” and no party designation was given to this candidate in the prior database. “Fully contested” means that the number of winners in the district or post is equal to both the number of Democrats and the number of Republicans.

99998 = non-major party candidate, but party not specified

99999 = Write-in (codes 2683 and 9999 were also used).

Maine State representative returns for 1968 & 1970 were not reported by consecutive district numbers but by cities within each county. A district numbering system was imposed on these returns as follows. Note that ICPSR District 41 is composed of different cities for 1968 & 1970.

District Number

| | | | |
|-------|---|-----------|--|
| | Androscoggin County | | Sebago, Standish |
| 1 | Lewiston | 31 | Cumberland, Gray |
| 2 | Auburn | 32 | Harpwell, Yarmouth |
| 3 | Durham, Lisbon | 33 | Freeport, North Yarmouth, Pownal |
| 4 | Livermore, Livermore Falls, Turner | | Franklin County |
| 5 | Mechanic Falls, Minot, Poland | | |
| 6 | Greene, Leeds, Wales, Webster | 34 | Chester, Jay, New Sharon, Wilton |
| | Aroostook County | 35 | Carthage, Farmington, Industry, New Vineyard, Temple, Weld |
| 7 | Caribou | 36 | Avon, Eustis, Kingfield, Madrid, Phillips, Rangeley, Strong, Coplin Pl., Dallas Pl., Rangeley Pl., Sandy River Pl. |
| 8 | Houlton | | Hancock County |
| 9 | Presque Isle | | |
| 10 | Limestone | | |
| 11 | Easton, Fort Fairfield | 37 | Brooklin, Brooksville, Castine, Deer Isle, Sedgwick, Stonington, Swan's Island, Long Island Pl. |
| 12 | Amity, Bancroft, Benedicta, Crystal, Haynesville, Hodgdon, Island Falls, Linneus, New Limerick, Orient, Sherman, Weston, Cary Pl., Glenwood Pl., Macwahoc Pl., Reed Pl. | 38 | Bucksport, Dedham, Orland, Penobscott, Verona |
| 13 | Blaine, Bridgewater, Dyer Brook, Hersey, Littleton, Ludlow, Merrill, Monticello, Oakfield, Smyrna, Plantation E, Hammond Pl., Moro Pl. | 39 | Blue Hill, Ellsworth, Surry |
| 14 | Ashland, Castle Hill, Chapman, Mars Hill, Masardis, Westfield, Garfield Pl., Nashville Pl., Oxbow Pl. | 40 | Bar Harbor, Cranberry Isles, Mount Desert, Southwest Harbor, Tremont |
| 15 | Mapleton, New Sweden, Perham, Wade, Washburn, Woodland, Westmanland Pl. | 41 (1968) | Amherst, Aurora, Eastbrook, Franklin, Gouldsboro, Hancock, Lamoine, Mariaville, Otis, Sorrento, Sullivan, Trenton, Waltham, Winter Harbor, Osborn Pl., Plantation No. 33 |
| Eagle | Lake, Portage Lake, St. Agatha St. Francis, Allagash Pl., New Canada Pl, St. John Pl., Wallagrass Pl., Winterville Pl. | 41 (1970) | Amherst, Aurora, Eastbrook, Franklin, Gouldsboro, Hancock, Lamoine, Mariaville, Otis, Sorrento, Sullivan, Trenton, Waltham, Winter Harbor, Great Pond Pl., Osborn Pl. |
| 17 | Fort Kent, Frenchville | | Kennebec County |
| 18 | Grand Isle, Madawaska | 42 | Augusta |
| 19 | Stockholm, Van Buren, Caswell Pl., Cyr Pl., Hamlin Pl. | 43 | Waterville |
| | Cumberland County | 44 | Gardiner |
| 20 | Portland | 45 | Winslow |
| 21 | South Portland | 46 | Albion, Benton, Clinton, Sidney |
| 22 | Westbrook | 47 | China, Pittston, Vassalboro, Windsor |
| 23 | Brunswick | 48 | Chelsea, Farmingdale, Randolph |
| 24 | Scarborough | 49 | Hallowell, Litchfield, Manchester, West Gardiner |
| 25 | Falmouth | 50 | Monmouth, Readfield, Wayne, Winthrop |
| 26 | Cape Elizabeth | 51 | Belgrade, Fayette, Mount Vernon, Oakland, Rome, Vienna |
| 27 | Gorham | | Knox County |
| 28 | Windham | | |
| 29 | Bridgton, Casco, Harrison, Naples, Otisfield | | |
| 30 | Baldwin, New Gloucester, Raymond, | | |

- 52 Rockland
- 53 Cushing, Friendship, Isle Au Haut, North Haven, Owls Head, St. George, South Thomaston, Vinalhaven, Matinicus Isle Pl.
- 54 Camden, Hope, Rockport
- 55 Appleton, Thomaston, Union, Warren, Washington
- Lincoln County
- 56 Jefferson, Nobleboro, Waldoboro, Whitefield, Monhegan Pl., Somerville Pl.
- 57 Boothbay, Boothbay Harbor, Southport, Westport, Wiscasset
- 58 Alna, Bremen, Bristol, Damariscotta, Dresden Edgecomb, Newcastle, South Bristol
- Oxford County
- 59 Rumford
- 60 Mexico
- 61 Buckfield, Canton, Dixfield, Hartford, Peru, Sumner
- 62 Hebron, Paris, West Paris, Woodstock
- 63 Norway, Oxford, Waterford
- 64 Brownfield, Denmark, Fryeburg, Hiram, Lovell, Porter, Stow, Sweden
- 65 Andover, Bethel, Byron, Gilead, Greenwood, Hanover, Newry, Roxbury, Stoneham, Upton, Lincoln Pl, Magalloway Pl.
- Penobscot County
- 66 Bangor
- 67 Brewer
- 68 Old Town
- 69 Millinocket
- 70 Corinna, Exeter, Newport, Stetson
- 71 Hampden, Newburgh
- 72 Carmel, Dixmont, Etna, Hermon, Levant, Plymouth
- 73 Dexter, Garland
- 74 Alton, Bradford, Charleston, Corinth, Edinburg, Glenburn, Hudson, Kenduskeag, Veazie
- 75 Clifton, Eddington, Holden, Orrington
- 76 Orono
- 77 Bradley, Enfield, Howland, Lagrange Maxfield, Milford, Seboeis Pl.
- 78 Lincoln
- 79 Burlington, Chester, Greenbush, Greenfield, Indian Island Voting District, Lee, Lowell, Mattawamkeag, Passadumkeag, Springfield, Winn Woodville, Carroll Pl., Drew Pl., Grand Falls Pl., Lakeville Pl., Prentiss Pl., Webster Pl.
- 80 East Milinocket, Medway, Patten, Stacyville, Mount Chase Pl.
- Piscataquis County
- 81 Abbot, Greenville, Guilford, Monson, Shirley, Wellington, Willimantic, Blanchard Pl., Elliottsville Pl., Kingsbury Pl.
- 82 Atkinson, Bowerbank, Brownville, Medford, Milo, Sebec, Barnard Pl., Lake View Pl.
- 83 Dover-Foxcroft, Parkman, Sangerville Sagadahoc County
- 84 Bath
- 85 Bowdoin, Bowdoinham, Georgetown, Richmond, West Bath
- 86 Arrowsic, Phippsburg, Topsham, Woolwich
- Somerset County
- 87 Skowhegan
- 88 Fairfield
- 89 Canaan, Detroit, Palmyra, Pittsfield
- 90 Athens, Bingham, Cambridge, Harmony, Hartland, Moscow, Ripley, St. Albans, Brighton Pl.
- 91 Cornville, Madison, Mercer, Norridgewock, Smithfield
- 92 Anson, Embden, Jackman, Moose River, New Portland, Solon, Starks, Caratunk Pl., Dennistown Pl., Highland Pl., Pleasant Ridge Pl., The Forks Pl., West Forks Pl.
- Waldo County
- 93 Belfast
- 94 Belmont, Freedom, Islesboro, Knox, Liberty, Lincolnville, Montville, Morrill, Northport, Palermo, Searsmont
- 95 Brooks, Burnham, Jackson, Monroe, Swanville, Thorndike, Troy, Unity, Waldo
- 96 Frankfort, Prospect, Searsport, Stockton Springs, Winterport
- Washington County
- 97 Addison, Beals, Beddington, Centerville, Cherryfield, Columbia, Columbia Falls, Deblois, Harrington, Milbridge, Steuben, Whitneyville
- 98 East Machias, Jonesboro, Jonesport, Machias, Machiasport, Marshfield, Northfield, Roque Bluffs, Wesley
- 99 Cooper, Crawford, Cutler, Dennysville, Lubec, Pembroke, Whiting, Plantation No. 14
- 100 Alexander, Charlotte, Danforth, Eastport, Indian Township Voting District, Meddybemps, Perry, Pleasant Point Voting District, Robbinston, Talmadge, Topsfield, Vanceboro, Waite, Codyville Pl., Grand Lake Stream Pl., Plantation No. 21
- 101 Baileyville, Calais, Princeton, Baring Pl.

| | | | |
|-----|-------------------|-----|--|
| | York County | 109 | Eliot, South Berwick |
| 102 | Biddeford | 110 | North Berwick, Wells |
| 103 | Sanford | 111 | Acton, Berwick, Lebanon |
| 104 | Saco | 112 | Alfred, Arundel, Kennebunkport, Lyman |
| 105 | Kittery | 113 | Buxton, Dayton, Hollis, Limington |
| 106 | Kennebunk | 114 | Cornish, Limerick, Newfield, Parsons- Field, Shapleigh, Waterboro |
| 107 | Old Orchard Beach | | |
| 108 | York | | |

Appendix E-2: Maine House of Representatives 1972 District Designation Codes

Maine State representative returns for 1972 were not reported by consecutive district numbers but by cities within each county. A district numbering system was imposed on these returns as follows.

ICPSR

District Number

Maine Cities (by County)

| | | | |
|----|---|----|---|
| | Androscoggin County | | Cumberland County |
| 1 | Lewiston | 20 | Portland |
| 2 | Auburn | 21 | South Portland |
| 3 | Durham, Lisbon | 22 | Westbrook |
| 4 | Livermore, Livermore Falls, Turner | 23 | Brunswick |
| 5 | Mechanic Falls, Minot, Poland | 24 | Scarborough |
| 6 | Greene, Leeds, Wales, Sabattus | 25 | Falmouth |
| | Aroostook County | 26 | Cape Elizabeth |
| 7 | Caribou | 27 | Gorham |
| 8 | Houlton | 28 | Windham |
| 9 | Presque Isle | 29 | Bridgton, Casco, Harrison, Naples, Otisfield |
| 10 | Limestone | 30 | Baldwin, New Gloucester, Raymond, Sebago, Standish |
| 11 | Easton, Fort Fairfield | 31 | Cumberland, Gray |
| 12 | Amity, Bancroft, Benedicta, Crystal, Haynesville, Hodgdon, Island Falls, Linneus, New Limerick, Orient, Sherman, Weston, Cary Pl., Glenwood Pl., Macwahoc Pl., Reed Pl. | 32 | Harpswell, Yarmouth |
| 13 | Blaine, Bridgewater, Dyer Brook, Hersey, Littleton, Ludlow, Merrill, Monticello, Oakfield, Smyrna, Plantation E, Hammond Pl., Moro Pl. | 33 | Freeport, North Yarmouth, Pownal |
| 14 | Ashland, Castle Hill, Chapman, Mars Hill, Masardis, Westfield, Garfield Pl., Nashville Pl., Oxbow Pl.. | | Franklin County |
| 15 | Mapleton, New Sweden, Perham, Wade, Washburn, Woodland, Westmanland Pl. | 34 | Chesterfield, Jay, New Sharon, Wilton |
| 16 | Eagle Lake, Portage Lake, St. Agatha St. Francis, Allagash Pl., New Canada Pl., St. John Pl., Wallagrass Pl., Winterville Pl. | 35 | Carthage, Farmington, Industry, New Vineyard, Temple, Weld |
| 17 | Fort Kent, Frenchville | 36 | Avon, Carrabassett Valley, Eustis, Kingfield, Madrid, Phillips, Rangeley Strong, Coplin Pl., Dallas Pl., Rangeley Pl., Sandy River Pl. |
| 18 | Grand Isle, Madawaska, St. Agatha, Sinclair, T17, R4 District | | Hancock County |
| 19 | Stockholm, Van Buren, Caswell Pl., Caswell, Connor District, Cyr Pl., Hamlin Pl. | 37 | Brooklin, Brooksville, Castine, Deer Isle, Sedgwick, Stonington, Swan's Island, Long Island Pl. |
| | | 38 | Bucksport, Dedham, Orland, Penobscott, Verona |
| | | 39 | Blue Hill, Ellsworth, Surry |
| | | 40 | Bar Harbor, Cranberry Isles, Mount Desert, Southwest Harbor, Tremont |

- 41 Amherst, Aurora, Eastbrook, Franklin,
Gouldsboro, Hancock, Lamoine,
Mariaville, Otis, Sorrento, Sullivan,
Trenton, Waltham, Winter Harbor,
Great Pond Pl., Osborn Pl.
- Kennebec County
- 42 Augusta
- 43 Waterville
- 44 Gardiner
- 45 Winslow
- 46 Albion, Benton, Clinton, Sidney
- 47 China, Pittston, Vassalboro, Windsor
- 48 Chelsea, Farmingdale, Randolph
- 49 Hallowell, Litchfield, Manchester,
West Gardiner
- 50 Monmouth, Readfield, Wayne,
Winthrop
- 51 Belgrade, Fayette, Mount Vernon,
Oakland, Rome, Vienna
- Knox County
- 52 Rockland
- 53 Cushing, Friendship, Isle Au Haut, North
Haven, Owls Head, St. George, South
Thomaston, Vinalhaven, Matinicus Isle Pl.
- 54 Camden, Hope, Rockport
- 55 Appleton, Thomaston, Union, Warren,
Washington
- Lincoln County
- 56 Jefferson, Nobleboro, Waldoboro,
Whitefield, Monhegan Pl., Somerville Pl.
- 57 Boothbay, Boothbay Harbor, Southport,
Westport, Wiscasset
- 58 Alna, Bremen, Bristol, Damariscotta
Dresden, Edgecomb, Newcastle,
South Bristol
- Oxford County
- 59 Rumford
- 60 Mexico
- 61 Buckfield, Canton, Dixfield, Hartford,
Peru, Sumner
- 62 Hebron, Paris, West Paris, Woodstock
- 63 Norway, Oxford, Waterford
- 64 Brownfield, Denmark, Fryeburg, Hiram,
Lovell, Porter, Stow, Sweden
- 65 Andover, Bethel, Byron, Gilead,
Greenwood, Hanover, Newry, Roxbury,
Stoneham, Upton, Lincoln Pl.,
Magalloway Pl.
- Penobscot County
- 66 Bangor
- 67 Brewer
- 68 Old Town
- 69 Millinocket
- 70 Corinna, Exeter, Newport, Stetson
- 71 Hampden, Newburgh
- 72 Carmel, Dixmont, Etna, Hermon,
Levant, Plymouth
- 73 Dexter, Garland
- 74 Alton, Bradford, Charleston, Corinth,
Edinburg, Glenburn, Hudson,
Kenduskeag, Veazie
- 75 Clifton, Eddington, Holden, Orrington
- 76 Orono
- 77 Bradley, Enfield, Howland, Lagrange
Maxfield, Milford, Seboeis Pl.
- 78 Lincoln
- 79 Burlington, Chester, Greenbush,
Greenfield, Indian Island Voting
District, Lee, Lowell, Mattawamkeag,
Passadumkeag, Springfield, Winn
Woodville, Carroll Pl., Drew Pl.,
Grand Falls Pl., Lakeville Pl.,
Prentiss Pl., Webster Pl.
- 80 East Milinocket, Medway, Patten,
Stacyville, Mount Chase Pl.
- Piscataquis County
- 81 Abbot, Greenville, Guilford, Monson,
Shirley, Wellington, Willimantic,
Blanchard Pl., Elliottsville Pl., Kingsbury Pl.
- 82 Atkinson, Bowerbank, Brownville, Medford,
Milo, Sebec, Barnard Pl., Lake View Pl.
- 83 Dover-Foxcroft, Parkman, Sangerville
Sagadahoc County
- 84 Bath
- 85 Bowdoin, Bowdoinham, Georgetown,
Richmond, West Bath
- 86 Arrowsic, Phippsburg, Topsham, Woolwich
- Somerset County
- 87 Skowhegan
- 88 Fairfield
- 89 Canaan, Detroit, Palmyra, Pittsfield
- 90 Athens, Bingham, Cambridge, Harmony,
Hartland, Moscow, Ripley,
St. Albans, Brighton Pl.
- 91 Cornville, Madison, Mercer,
Norridgewock, Smithfield
- 92 Anson, Embden, Jackman, Moose River,
New Portland, Solon, Starks,
Caratunk Pl., Dennistown Pl.,
Highland Pl., Pleasant Ridge Pl.,
The Forks Pl., West Forks Pl.
- Waldo County
- 93 Belfast
- 94 Belmont, Freedom, Islesboro, Knox,
Liberty, Lincolnville, Montville,
Morrill, Northport, Palermo, Searsmont
- 95 Brooks, Burnham, Jackson, Monroe,
Swanville, Thorndike, Troy, Unity, Waldo

- 96 Frankfort, Prospect, Searsport,
Stockton Springs, Winterport
- Washington County
- 97 Addison, Beals, Beddington,
Centerville, Cherryfield, Columbia,
Columbia Falls, Deblois, Harrington,
Milbridge, Steuben, Whitneyville
- 98 East Machias, Jonesboro, Jonesport,
Machias, Machiasport, Marshfield,
Northfield, Roque Bluffs, Wesley
- 99 Cooper, Crawford, Cutler, Dennysville,
Lubec, Pembroke, Whiting,
Plantation No. 14
- 100 Alexander, Charlotte, Danforth,
Eastport, Indian Township Voting
District, Meddybemps, Perry,
Pleasant Point Voting District,
Robbinston, Talmadge, Topsfield,
Vanceboro, Waite, Codyville Pl.,
Grand Lake Stream Pl., Plantation No. 21
- 101 Baileyville, Calais, Princeton, Baring Pl.
- York County
- 102 Biddeford
- 103 Sanford
- 104 Saco
- 105 Kittery
- 106 Kennebunk
- 107 Old Orchard Beach
- 108 York
- 109 Eliot, South Berwick
- 110 North Berwick, Wells
- 111 Acton, Berwick, Lebanon
- 112 Alfred, Arundel, Kennebunkport, Lyman
- 113 Buxton, Dayton, Hollis, Limington
- 114 Cornish, Limerick, Newfield,
Parsonsfield, Shapleigh, Waterboro

Appendix F: Data Set Utilization Check List

You may want to utilize the following checklist of questions before conducting analyses with this data set.

Keep in mind that for a small proportion of cases, the contest-candidate-partyline is not the unit of analysis, as they are further broken down by county. I constantly forget this.

You should use the district designations dname, dist, geopost, and mmdpost for analysis, such as lagging variables. The variable “ddez” merely displays how districts were presented in returns, and are sometimes not standardized across time within the same district and redistricting period. The variable “ddez” is best suited for display in files that workers are inputting data about legislators or candidates into, as there is less visual difference between the contents of “ddez” and how returns look than the variables dname, dist, geopost and mmdpost. The properties of “ddez” are not such that it would be appropriate for making comparisons in the same district over time, especially because district designations are sometimes changed in state sources with returns.

When lagging variables, do you want to use the last time a seat in a district was up, or the last time the district was up? If a district has an alternating component to its seats (i.e., “dtype”=4, for example), these are not the same. If you do lag on the basis of when the last time a seat was up, you should use “termz” in conjunction with “year” to facilitate lagging.

You generally shouldn’t use lagged values when redistricting has occurred since the last time a district was up.

Do you want to include all types of districts in your analysis, or just single-member districts? In other words, what values of “dtype” should be included in the analysis? How do you want to deal with free-for-all multi-member districts?

Should you include Louisiana, given its unique election system?

How do you want to deal with California and Washington’s top-two primary systems? California adopted this with the 2012 elections, while Washington adopted it with the 2008 election.

What type of elections do you want to include in your analysis? In other words, what values of “etype” should be included in the analysis?

Generally, you want to exclude elections that were not at a normal time. Even if an election has a value of deter=1, it may be a special election designed to fill a seat when a candidate died, and the regularly scheduled election did not result in a winner that would go to the legislature.

Do you want to include elections where “etype”=“gs” in your analysis? (These are usually state senate elections in a state with four-year senate terms that are held in November of an even numbered year after only two years of a state senate term has been served.) I would generally say “yes.”

How do you want to deal with uncontested elections in your analysis? How do you want to deal with elections that are contested because of non-major party candidates, but do not have any Democrats or alternatively, do not have any Republicans?

Keep in mind “mixed uncontested” elections. These are contests for two or more seats, and there is at least one Democrat running, and at least one Republican running, but if you add together the number of Democrats and Republicans running, it equals the number of seats to be won. This means that the outcome of the contest is pre-determined, and is probably inappropriate for your analysis.

How do you want to deal with under-contested FFA-MMD elections in your analysis? In other words, elections where there are either fewer Democrats and/or fewer Republicans running than there are seats to be won in that election in that district, although there is at least one Democrat, and one Republican running.

Do you want to include non-partisan elections in your analysis (Minnesota before 1974 and Nebraska)? These are cases for which partyz=“nonpart.”

How do you want to deal with elections that occurred in the “solid South,” elections in the past that were almost inevitably won by Democrats?

Do you want to include elections that have a Democrat “fused” with a third party candidate, or a Republican “fused” with a third party candidate? In other words, these major party candidates are also picking up the votes of third party supporters, and what they mean about Democratic or Republican support may be different than elections in which there is no such fusion. Should the votes of non-major parties be added to the votes of the major party candidate they are fused with? Fusion mostly occurs in CT, NH and NY. Does a Democratic (Republican) candidate receive more votes in years they are fused with non-major party candidates than in years where they aren’t?

How do you want to deal with elections that have one candidate running as both a Democrat and a Republican? This is either when partyz=“b,” or when one candidate has both partyz=“d” and partyz=“r” on different “lines.” If the candidate in question is deemed to “really” be a Democrat, should you count their Republican votes as Democratic or Republican votes? Do you wish to utilize “trueparty” to assign one party to such candidates (see Appendix H).

Do you want to include “over-contested” elections in your analysis? These are elections that have more Democrats and/or more Republicans running than there are seats to be won in that election in that district. Aside from cases in CA, NV and WA, not many elections are over-contested, and many of these are probably instances of write-in candidates who are also Democrats or Republicans. Another way of dealing with these elections is by limiting the number of Democrats, and also limiting the number of Republicans, to the number of seats to be won in the election.

Which values of the variable “uncert” should result in excluding a case from analysis? One alternative is to exclude a case if the following values appear: incompleteelection, dist, party, etype, esets, generalproblem, outcome, year, writeinstatus.

You should not include elections in which a candidate is running that does not have a value for the variable “vote,” and who is running in a contested election.

For most elections, you should not include write-in votes in vote totals. The amount of votes obtained by write-ins is not reported in official sources for many states and year, and was not collected for many of the elections in states and years that do report them. This makes the measure of the number of write-ins cast in elections inconsistent across states.

How do you want to deal with elections where write-in candidates (i.e., partyz=“writein”) obtained a substantial portion of the vote?

Beware of candidates who are not coded as write-ins who probably are write-ins. The code provided in Appendix J provides guidance on identifying these candidates.

How do you want to code write-in Democrats and write-in Republicans? For an analysis of the determinants of election outcomes, elections with such candidates should be excluded if they received a “substantial” proportion of the vote. It is unclear how these cases should be dealt with in an analysis of the relationship between seats and votes.

When measuring the propensity for incumbents to run, take care not to count elections in which they have merely been written onto the ballot by a few voters. The incumbent in question may have actually decided to not run in that case.

How do you want to deal with elections where there is a “substantial” third party (i.e., non-major party, partyz=“nonmaj”) presence? A “substantial third party presence” might include elections where a non-major party candidate is an incumbent, or where the non-major party candidate receives more than a given percent of the vote. Should the cutoff be different for elections that are under- or uncontested for major party candidates? In other words, if an election for one seat has one Democrat, one Republican, and a third-party candidate who obtains 30 percent of the vote, we may well want to exclude that election in an analysis of the determinants of Democratic vote share. On the other hand, if an election has one Democrat and one third party candidate who receives 30 percent of

the vote, we may still classify that election as “uncontested-no Republican candidate” in the analysis of the determinants of Democratic vote share (say, as a lagged variable coded as an uncontested election).

How do you want to deal with incumbency in the beginning of the dataset? (Some of the candidates identified as incumbents in the beginning of the data set may have been appointed incumbents, and not elected incumbents. See description of the variable “exper.”)

How do you want to deal with elections where a candidate recently ran on a different party’s label? This is especially consequential when that switch is from Democrat to Republican or vice versa, and when the party switching candidate won the last election they ran in.

Appendix G: How to Best Lag Variables

An alternative method of lagging variables compared to the typical method of doing so with time series commands is offered here. The general idea is to create a dataset that will have the lagged variables in it, and then merge it with the main dataset after replacing the variable “year” with “year” + “termz.”

The specific steps to do this are as follows.

First, collapse the data to the election as the unit of analysis. Compute the quantities of interest that will be lagged in that dataset (i.e., Democratic vote share for the election, number of Democratic incumbents, etc.).

In the main dataset, create a new battery of variables representing dnamez, distz, geopostz and mmdpostz, with new names (say, with a “2” suffix). Replace their values with redist1, redist2 and redist3 (as appropriate) when the variable “redist” is equal to “2” or “4.”

Save the main dataset with a new name (say, “templag”), and rename dnamez, distz, geopostz and mmdpostz to have “2” suffixes. Then replace the variable “mmdpostz2” with the values of “specpost” when “specpost” is non-missing.

Then replace the values in “year” with the values of “year” plus “termz.”

Rename all variables you wish to lag with the suffix “lag.”

Then drop all variables except for the variables you wish to lag, and the variables necessary for the merge (see below).

Then merge these variables into the main file on sid, sen, dnamez2, distz2, geopostz2, and mmdpostz2. Drop cases that represent cases that only existed in file “templag.”

Next, when “redist” is equal to “1” the lagged values that were just brought in should either be recoded as system missing, or they should be treated with extreme care and designated with a special status if they are going to be used for further analysis. Keep in mind that just because a district has the same designation as a district before redistricting need not mean they have anything to do with each other (i.e., as in Arkansas and Montana).

Now that the entire process has been explained, how to deal with two related matters that only effect a small number of cases will be more understandable. When the collapse of the main file is done, the variable “termz” must be retained, but it is sometimes not constant at the election level, and so its minimum and maximum within an election must be recorded when the collapse is conducted. Two copies of one election must then be made, one with the minimum value of “termz,” one with the maximum value of “termz.” In this way, the same election will go on to be the lagged value of two different elections in the future. This is necessary for appropriate comparisons over time because its seats have different term lengths.

The next problem is that when year is replaced with year plus termz, there will be a small number of cases within groups of cases defined by sid, sen, dnamez, distz, geopostz, and mmdpostz that are not just one case (i.e., are not “identified” by those cases). However, you only want one observation for each such group, since there is only one election a lagged value is required for. To do this appropriately, the data must be collapsed on sid, sen, dnamez, distz, geopostz, and mmdpostz. The quantities you wish to lag must be averaged in this collapse, but the computed average should be weighted by the number of seats in the elections in question (the variable “eseats”).

Appendix H: “True” Party

There are three ways to define a fused Democratic-Republican’s “true” party. 1) When running as just a Democrat or Republican, which one has it been or will it be in the future? 2) When both Democratic and Republican votes for the same candidate are reported as separate lines, which line gets the most votes? 3) How are they listed in the Directory I (Council of State Governments various years), as “D,” “R,” “D-R” or “R-D”? I take the party that is listed first as indicating which party the CSG took the legislator to “really belong to” / caucus with. Or, alternatively, one could look up who this individual caucuses with in the legislative record, which I only infrequently did.

Unfortunately, one doesn’t always have #1, because sometimes candidates are D-R fused in all observed elections they’ve run in. One doesn’t always have #2, because votes for the separate lines aren’t always reported, like for the New Hampshire House. One does always have #3, but it unfortunately relies on my interpretation of what the order of the D and the R mean, and assumes the CSG is always correct.

What is the correlation between the two indicators of “true party” within each of the three pairs of indicators? The analyses below were conducted on October 27, 2010 and are based off an analysis of 1968 to 2008 elections. Overall, 1,773 candidates ran as fused Democrat-Republicans in this time period. 1,528 can be coded as Democrat or Republican on the basis of running as just one of those parties in other years. Using separate lines allows coding 610 of those, and the two methods together fail to code 142 candidates. There is 84% agreement between those two methods when both are observed (N=507).

The results of these cross-tabs indicate that utilizing 1) prior or future parties as party designations for fused d-r candidates, 2) votes for separate lines of d and r for fused d-r candidates to assign parties is valid. The first method listed agreed with the roster 97.3% of the time. The second method agreed with the roster 86.3% of the time.

The results also indicate that the roster considering legislators d versus d-r, or r versus r-d doesn’t matter that much. If the roster gave a hyphenated party, it agreed (agreement determined by the first party listed when the roster gave a hyphenated party) with the first method listed 95.9% of the time, while non-hyphenated parties agreed 98.1% of the time (n was large in both categories of hyphenation in this cross-tab). If the roster gave a hyphenated party, it agreed with the second method (i.e., using whether there were more Democratic or Republican votes) 60.0% of the time, while non-hyphenated parties agreed 88.2% of the time (N=40 among hyphenated cases in this last cross-tab).

Appendix I: How to Create a List of Sitting Legislators from SLERs for Each Session Year

The variable termz is instrumental in making the following work. The following Stata code creates a list of sitting legislators in each session year.

- *produces a list of sitting legislators, along with
- *the districts they were elected from,
- *the district they currently represent (different from above sometimes if a holdover state senator (or holdover ND state house member)
- *the year in which they were elected (electyear)

- *each legislator is represented once per "session year"
- *"session year" is still observed if a legislature wasn't in session in that calendar year.
- *"sessyear" starts in the year after a November general election, and ends with the end of the term the legislator served. If a legislature adjourns in the early part of the year after an election year, the election year is taken as the end year.
- *For example, if electyear=1988, sessyear=1989 and 1990 for a two year term.

- *The variable "trueparty" is used for the legislators' party.

*REGIME

- *Is the redistricting regime from which they were elected.

*REGIMEHOLD

- *Is the redistricting regime of the map they were brought into by their holdover status.

clear

version 15.1

set varabbrev off, permanently

cd C:\Users\User\Dropbox\04_SLERs_Additions\CODE\CreatesUOALegislatorSessYear

use 001_193slers1967to2016_20180813

keep if deter==1 & outcome=="w"

collapse (sum) deter, by(year sid sen ddez dname dno geopost mmdpost etype cand candid trueparty hold hold1 hold2 hold3 termz tenure1 tenure2 regime redist redist1 redist2 redist3 esests dseats dtype)
gen mod=mod(termz,1)

sum termz

local ccc=r(max)

forvalues aaa=1/`ccc' {

local bbb=`aaa'-1

gen sessyear`aaa'=year+`aaa' if (termz>=`aaa')&mod==0

replace sessyear`aaa'=year+`bbb' if (termz>=`bbb')&mod==.5

gen tenure1`aaa'=tenure1+`bbb' if (termz>=`aaa')&mod==0

replace tenure1`aaa'=tenure1+`bbb' if (termz>=`bbb')&mod==.5

gen tenure2`aaa'=tenure2+`bbb' if (termz>=`aaa')&mod==0

replace tenure2`aaa'=tenure2+`bbb' if (termz>=`bbb')&mod==.5

}

rename year electyear

drop tenure1 tenure2 deter mod

reshape long sessyear tenure1 tenure2, i(electyear sid sen ddez dname dno geopost mmdpost candid) j(yearofterm)

drop if sessyear==.

- *All holdover districts are observed to be applied in the third year of a holdover legislator's term, with one exception. The KY Senators elected in 1981 who ended up as holdover senators (one left office after two years) had five year terms, and there was redistricting in 1986. So they are holdover in their fourth and fifth years in office, not third, fourth and fifth years in office.

```

gen sessdname=hold1
gen sessdno=hold2
gen sessgeopost=hold3
replace sessdname=dname if hold==.|yearofterm<3
*only dno has to be adjusted for KY, as dname and geopost don't exist for that state senate for the relevant years.
replace sessdno=dno if hold==.|(yearofterm<3&[!(sid==17&sen==1&electyear==1981)])
replace sessdno=dno if yearofterm<4&(sid==17&sen==1&electyear==1981)
replace sessgeopost=geopost if hold==.|yearofterm<3

*IDENTITY
*bysort sid sessyear candid: assert _N==1
*26 contradictions
*bysort sid sen sessyear candid: assert _N==1
*8 contradictions
bysort sid sen sessyear candid: gen sum1=_N
list sid sen sessyear if sum1!=1
*the ca ones look familiar. There are ky cases, but they aren't when the switch to even years was. They all look to
be redistricting related, as expected.
bysort sid sen sessyear candid dname dno geopost mmdpost: assert _N==1
*That passes the test, as expected.
*bysort sid candid electyear sessyear: assert _N==1
*assertion is false
bysort sid sen candid electyear sessyear: assert _N==1
*assertion is true.
bysort sid sen candid electyear yearofterm: assert _N==1
*assertion is true.

*REGIMEHOLD
gen regimehold=regime
*The following changes the values of regimehold for holdover legislators.
*The regime value must be the value of sessyear-1 when yearofterm=3 if hold=1, with the exception of the KY
replace regimehold=sessyear-1 if yearofterm==3&hold==1&[!(sid==17&sen==1&electyear==1981)]
replace regimehold=sessyear-2 if yearofterm==4&hold==1&[!(sid==17&sen==1&electyear==1981)]
replace regimehold=sessyear-1 if yearofterm==4&hold==1&(sid==17&sen==1&electyear==1981)
replace regimehold=sessyear-2 if yearofterm==5&hold==1&(sid==17&sen==1&electyear==1981)
gen these=(yearofterm==3|yearofterm==4)&hold==1&[!(sid==17&sen==1&electyear==1981)]
replace these=1 if (yearofterm==4|yearofterm==5)&hold==1&(sid==17&sen==1&electyear==1981)

*KY
*I didn't do the following correctly, so the following corrects the mistake above.
replace regimehold=1983 if sen==1&sid==17&sessyear>1983&sessyear<1993

save 102slersUOALegislatorSessYear20180813, replace

*CHECK
*The following block doesn't alter the data, just checks.
*See how often regimehold is followed by an identical value in that district. If it isn't followed by an identical
value, it might be because of 1) a mistake in the assumptions behind the code or 2) districts that "never see the light
of day."
*The district-sessyear has to become the unit of analysis for the test.
clear
use 102slersUOALegislatorSessYear20180813
bysort sid sen dname dno geopost mmdpost sessyear: gen sum2=_N
bysort sid sen dname dno geopost mmdpost sessyear regimehold: gen sum3=_N
assert sum2==sum3
*no difs, great.
bysort sid sen dname dno geopost mmdpost sessyear these: gen sum4=_N
*assert sum2==sum4

```

```

*The above assert was false, but I can see how that would be. No problem.
drop sum2 sum3 sum4
collapse (max) these, by(sid sen dname dno geopost sessyear regimehold)
bysort sid sen dname dno geopost sessyear: assert _N==1
*All 1, good
bysort sid sen dname dno geopost (sessyear): gen id=1 if _n==1
replace id=sum(id)
gen tempyear=sessyear
tsset id sessyear
by id: gen theselead=these[_n+1]
by id: gen regimeholdlead=regimehold[_n+1]
by id: gen yearlead=tempyear[_n+1]
gen yeardif=yearlead-sessyear
egen difregime=diff(regimehold regimeholdlead)
*If regimehold is constant within a chamber-sessyear, then difregime being=1 shouldn't matter under the below
conditions (i.e., these==1&theselead==0&yeardif==1).
bysort sid sen sessyear: gen sum1=_N
bysort sid sen sessyear regimehold: gen sum2=_N
*how often is regimehold dif in the same chamber-sessyear?
tab sid if (sum1!=sum2)&sen==1
*4830 times out of 93187 cases in state senates. Whenever there is re-redistricting, this often happens, so it doesn't
mean errors. The amount there is about right I'd guess.
tab sid sen if (sum1!=sum2)
*Happens a bunch in state houses also, about three times more often.
assert difregime==0 if these==1&theselead==0&yeardif==1&(sum1!=sum2)
*all cool, good.
clear

```

Appendix J: Program to Restructure SLERs for Analysis of Elections

The following code makes the unit of analysis the contest, i.e., a year-month-day-sid-sen-dname-dno-geopost-mmdpost-etype / district-post-election year / district-post-contest.

To be added.

Appendix K: Program to Merge SLERs Lists of Sitting Legislators with Other Lists

To be added.

Appendix L: Merging SLERs with Census Data

To be added.