Codebook for Datasets
"Senate Countermajoritarianism"
C. Lawrence Evans
April 2024

This document describes the contents of the seven datasets associated with the study. The first six are necessary to replicate results reported in the article and appendix. The last is necessary to construct the state population variable (pre-1901) already included in Dataset 2. For each dataset, the relevant variables are identified below by number based on their order of appearance, and clarifications are provided by accompanying notes.

Dataset 1: votes.dta (see note below)

1) *congress*: ranges from "1" (1789-91) to "117" (2021-23)

2) *rollnumber*: the voteview.com rollcall identifier within the relevant two-year congress

3) *icpsr*: an identification code for individual Senators. When members switch or otherwise leave parties, they receive a different code for roll calls cast with the new partisan affiliation

4) *cast_code*: the standard ICPSR values for categorizing the votes of individual members

0—not a member when vote taken
1—yes
2—paired yes
3—announced yes
4—announced no
5—paired no
6—no
7 or 8—present
9—not voting

5) *state_icpsr*: an identification code for the state represented by a Senator

6) *party_code*: an identification code for a Senator's party. For details, consult https://voteview.com/articles/data_help_parties

7) *majority:* relationship to the party or coalition organizing the chamber

0—member not affiliated with the partisan majority
1—member affiliated with the partisan majority

Note: Observations are by member and roll call, pooled across congresses (n=4,368,550). With one exception, all variables in the "votes.dta" dataset are as downloaded from voteview.com in October 2023, and variable descriptions are directly adopted from that source. The variable "majority" was coded by the author based on

https://www.senate.gov/history/partydiv.htm and is necessary to construct the variable "majfrac" included in independent_vars.dta.

Dataset 2: populations.dta (see note below)

1) *congress*: see description under **votes.dta**

2) *state_icpsr*: see description under **votes.dta**

3) *population*: state population by two-year congress

Note: Observations are by state and congress (n=4,556). For the 57th (1901-03) through the 117th Congress (2021-23), state populations are for the first year of the relevant congress, and the source is annual Census data as compiled by the Federal Reserve Economic Data (FRED), the St. Louis Fed, https://fred.stlouisfed.org/release/tables?rid=118&eid=259194#snid=259234 . The observation date is January of each odd numbered year, and estimates are annual and not seasonally adjusted. The data are as downloaded in October 2023 – minor adjustments occur as the Census and FRED update.

For 1789-1900, population by two-year congress is from decennial Census data as compiled by *Social Explorer* https://www.socialexplorer.com, supplemented by other Census sources (for details, see the notes for Dataset 7 below). Direct annual estimates are not generally available for this period. The lack of annual or biennial data is problematic because populations often increased dramatically in the years immediately preceding and following the admission of a state into the Union. Depending on where statehood falls over a decade, using the most recent territorial figures might significantly underestimate the population, while using the first Census following statehood might lead to overestimations. Even for states already admitted, the rate of population growth across states differed markedly throughout the 19th Century, and such differences are lost if we rely on the previous or the most temporally proximate decennial figures.

As a result, for congresses occurring between census years during 1789-1901, a linear progression is imposed on the decennial data. More concretely, for such congresses, the state populations are a weighted linear combination of the Census estimates immediately preceding and immediately following that congress, with the weights adjusted for the passage of time. For example, for the First Congress (1789-91), the numbers from the 1790 Census are used. For the Second Congress (1791-93), state populations = .8 * (1790 Census figure) + .2 * (1800 Census figure), for the Third Congress (1793-95), state populations = .6 * (1790 figure) + .4 * (1800 figure); for the Fourth Congress (1795-97), state populations = .4 * (1790 figure) + .6 * (1800 figure); for the Fifth Congress (1797-99), state populations = .2 * (1790 figure) + .8 * (1800 figure); for the Sixth Congress (1799-1801), the numbers from the 1800 Census are used; and so on until reliable annual data becomes available following the 1900 Census. The Stata commands used to produce the pre-1901 state populations in this manner are included in the do file labeled "populations_estimates_construction.do," and the dataset necessary to implement the commands (population_decennial_pre1901.dta) is described at the end of this document.

Dataset 3: independent_vars.dta (See notes 1-2 below)

1) *congress*: see description under **votes.dta**

2) *gini*: The Gini coefficient for population dispersion across states within a two-year congress

3) *closevote*: the fraction of roll calls during a congress that were close, where a vote is treated as close if the fraction voting yes is less than .6 and greater than .4

4) *majfrac*: the fraction of the U.S. population represented by members of the majority party in aggregate during a congress

Note 1: Observations are by congress (n=117). The variable "gini is calculated via the adgini package written for Stata and using the population data included in populations.dta. The necessary Stata commands to produce the variable are included in variable_construction.do

Note 2: In constructing majfrac, the linkage of parties to majority status is based on the summary of party divisions by congress included on the Senate website: https://www.senate.gov/history/partydiv.htm. Each Senator is allotted one half of the relevant state population. For cases where more than two Senators represented a state during a congress, the two that participated in the largest number of roll call votes are included, where "participation" entails having cast a yes, no, or paired vote. Totals occasionally differ from the divisions on the Senate website because of membership changes that occurred over the course of a two-year congress. In addition, four Senate delegations are dropped for particular congresses because of minimal participation in Senate roll calls. During the 37th Congress (1861-63), Senators from Arkansas, North Carolina, and Texas cast a small number of roll calls before exiting the chamber due to secession. And during the 57th Congress (1901-03), the Delaware legislature was unable to agree on who would represent the state in the Senate until the waning days of the second session, when one member was able to cast a single roll call on the chamber floor. The necessary Stata commands to produce this variable are included in variable_construction.do.

Dataset 4: motions.dta (see notes 1-3 below)

1) *congress*: see description under **votes.dta**

2) *rollnumber*: see description under **votes.dta**

3) *supermaj* (all congresses)

0—simple majority required
1—treaty ratification
2—cloture with a supermajority requirement
3—super majority budget waiver
4—supermajority required due to special order
5—other supermajority

4) *invoke* (all congresses, includes both supermajority and majority cloture motions)

0—not a cloture roll call
1—cloture not invoked
2—cloture invoked

5) *motiontype* (1945-2022, missing otherwise)

1—passage of entire measures (excluding nominations)

2—amendments
3—cloture (supermajority and majority, but excluding nominations)
4—budget waivers (supermajority and majority)
5—process/other (all other motions, again excluding nominations)
6—nomination related (all votes concerning the confirmation of nominees)

6*) keyvote* (1945-2022, missing otherwise)

0—not a *Congressional Quarterly* key vote
1—a *Congressional Quarterly* key vote

7) *keymeasure*: whether at least one *CQ* key vote occurred on the underlying measure (1945-2022, missing otherwise)

0—no key votes associated with the measure
1—at least one key vote associated with the measure

8) *keyid*: identifier for key measures (1945-2022, missing otherwise)

9) *vpvote:* tie breaking vote cast by the Vice-President (all congresses, missing if no tiebreaker is cast)
1—vice president cast a tie-breaking vote in favor
6—vice president cast a tie-breaking vote in opposition

Note 1: Observations are all motions associated with a roll call, pooled across congresses (n=52,046). Supermajority votes were identified and classified via key word searches of the "Congressional Votes" dataset on voteview.com. A definitive list of motions subject to a supermajority requirement based on Senate rules is provided in https://crsreports.congress.gov/product/pdf/RS/98-779/8. For the "invoke" variable, cloture votes of various types were identified using relevant portions of the Senate website https://www.senate.gov/legislative/cloture/clotureCounts.htm. Roll calls with a supermajority threshold due to a special order (a unanimous consent agreement or UCA) were identified using congress.gov and the Senate website, supplemented by relevant pages of *The Congressional Record*. The "keyvotes" and "keymeasure" variables were coded based on relevant issues of *Congressional Quarterly Weekly Report*. The bill numbers that generally make up "keyid" are from the "Congressional Votes" dataset on voteview.com, supplemented by issues of *CQ Weekly Report*. Where bill numbers or other identifiers are unavailable, a placeholder was entered by the author. For "vpvote," the source is files provided to the author by the Senate Historical Office, supplemented by the Senate website. My thanks to Daniel Holt of the Senate Historical Office for his assistance.

Note 2: Construction of the motion type variable also drew on a range of sources. As mentioned, the identification of cloture motions was based on the Senate website. Nomination related matters were identified by key word searches in the "Congressional Votes" dataset on voteview.com and include all motions related to the consideration or confirmation of nominees, including motions such as cloture that otherwise would fall in the other levels of the variable. For remaining values, construction of the motion type variable relied on the "vote" indicator included in Jason Roberts, David Rohde, and Michael H. Crespin, *Political Institutions and Public Choice Senate Roll-Call Database*, https://ou.edu/carlalbertcenter/research/pipc-votes/, and extended back to 1945 by Anthony Madonna and Michael Lynch as part of the *UGA Congress Project*: https://www.thecongressproject.com/data-and-links. More concretely, included among "entire measures" are the initial and final passage of bills and resolutions, proposed constitutional amendments, conference reports, veto overrides, and the ratification of treaties. The "amendments" category includes amendments in the first

and second degree; substitute amendments; motions to table amendments; motions to agree or disagree, strike, or otherwise delete language pertaining to portions of broader measures; and motions to recede. "Budget waivers" include all motions primarily focused on waiving budget rules, including Gramm-Rudman-Hollings restrictions. "Process/other" is for motions not otherwise placed in categories 1, 2, 3, 4, or 6.

Note 3: Although substantial overlap exists between similarly named values of "supermaj" and "motiontype," the fit is not perfect by design. Value 2 for "supermaj," for example, includes supermajority cloture motions, and thus does not perfectly align with value 3 for "motiontype," which includes all cloture motions not related to nominations. Value 3 of "supermaj," in turn, includes roll calls where the threshold exceeded a simple majority due to a violation of budget rules, while the budget waiver category (value 4) for "motiontype" covers all roll calls categorized as such by the "votes" indicator in the sources referenced in Note 2, including the many waivers that require only a simple majority. A full account of budget points of order in the Senate and the thresholds required to waive them is provided in https://www.budget.senate.gov/budgetpointsoforder. In short, the variables "supermaj" and "motiontype" were constructed for different purposes using different coding criteria and sources. Especially for this dataset, variable construction required occasional judgement calls, and any errors are the responsibility of the author.

Dataset 5: eligible_electorate.dta (see note below)

1) *congress*: see description under **votes.dta**

2) *state_icpsr*: see description under **votes.dta**

3) *potential:* potential electorate (in thousands)

Note: Observations are states by congress (1789-1923), (n=2,092). The source is Walter Dean Burnham, *Voting in American Elections*, Academica Press, Palo Alto, CA, 2010: pages 114-26. My thanks to Andrew Hoffman for carefully entering these data.

Dataset 6: FigA2.dta (see note below)

1) *congress:* see description under **votes.dta**

2) *cm:* the fraction of roll calls that are countermajoritarian

3) *cmnoties*: the fraction of roll calls that are countermajoritarian when ties are excluded

4) *cmvp:* the fraction of roll calls that are countermajoritarian when tie votes are left in the data and vice-presidential tie breakers are integrated into the roll call outcome

Note: Observations are individual congresses (n=117). The value of cm is produced by the Stata commands used to create Figure 2 (included in article_figures.do). The values for cmnoties and cmvp can be produced via commands included in variable_construction.do.

Dataset 7: population_decennial_pre1901.dta (see notes 1-2 below)

1) *state_icpsr*: see description under **votes.dta**

2) *cenyear*: the relevant Census year (e.g., 1790, 1800, and so on)

3) *totalpop*: the population estimate for that state and Census year

4) *nextpop*: the population estimate for that state in the decennial Census that immediately follows (when cenyear=1790, nextpop=1800, and so on). For cenyear=1900, nextpop is missing because of the shift to annual data that become available at that point.

Note 1: Observations are individual states by decennial census, 1900 and before (n=372). This dataset is not necessary to reproduce the results reported in the article and supplementary appendix but was used to estimate the congress-by-congress population entries prior to the 57th Congress (1901-03) included in **populations.dta** (relevant Stata commands are included in population_estimates_construction.do). The main source here is annual Census estimates by state accessed in October 2023 via *Social Explorer*: https://www.socialexplorer.com. The *Social Explorer* tabulations in turn are drawn from Table 4 of https://www.census.gov/library/publications/1909/decennial/century-populaton-growth.html. For the most part, the *Social Explorer* figures for a census year reflect geographic boundaries as they existed at the time, and totals often are not included for territories prior to admittance, which can complicate calculating linear projections. In such cases, I relied on other Census sources (taking steps when necessary to reflect contemporaneous boundaries) to fill in the missing data, especially https://www2.census.gov/library/visualizations/2000/dec/2000-resident-population/ and https://www2.census.gov/library/publications/decennial/1990/population-of-states-and-counties-us-1790-1990/population-of-states-and-counties-of-the-united-states-1790-1990.pdf

Note 2: Particularly during the antebellum and reconstruction years, the Census often was unable to count individuals living in sparsely populated or remote areas, and early territorial and state counts periodically included people more properly assigned elsewhere. A comprehensive summary of these difficulties by state is provided in the "notes" following the county-level data for each state in Part III of https://www2.census.gov/library/publications/decennial/1990/population-of-states-and-counties-us-1790-1990/population-of-states-and-counties-of-the-united-states-1790-1990.pdf. Correcting for such limitations in the evidence is not generally feasible, given standard Census sources, and for the most part the affected numbers fortunately are not large. Still, four adjustments to the dataset were feasible, and each involved substantial population. Where appropriate, the changes are directly entered into the dataset. Otherwise, they are implemented via commands included in population_estimates_construction.do. The four adjustments include:

1) Virginia, 1789-1793. Prior to Kentucky's admission as a state in June of 1792, the area was part of Virginia. The *Social Explorer* population figure for Virginia in 1790 does not include Kentucky. As a result, for the First Congress (1789-1791), the Census figure for Kentucky is added to the Virginia number. For the Second Congress, the estimated figure for Kentucky that Congress (resulting from a linear progression from the 1790 to the 1800 numbers for the new state) is added to the Virginia total. Since Kentucky was represented in the Second Congress for only a handful of the votes cast at the very end of the second session, the area's population is added to the Virginia figure for the entire congress.

2) Virginia, 1861-1865. West Virginia was part of Virginia until June of 1863, when it was admitted into the Union as an independent state. The inclusion of West Virginia within Virginia is reflected in the *Social Explorer* totals until the outbreak of the Civil War. For the purposes of this article, the estimate for Virginia in the 37th Congress (1861-63) reflects a linear progression from the Virginia Census

number for 1860 and what it would have been in 1870 if West Virginia and Virginia had remained united. For the 38th Congress (1863-65), the population estimate reflects that same linear progression but now subtracting the estimate for West Virginia. Virginia did not have voting representation in the Senate during the 39th and 40th Congresses (1865-67 and 1867-69, respectively), so the issue was moot until 1870 Census figures became available at the end of the decade.

3) Massachusetts, 1789-1821. Prior to Maine's admission as a state, it was legally part of Massachusetts. The Massachusetts state legislature, which selected the Commonwealth's delegation in the U.S. Senate, included representatives from Maine, and the apportionment of U.S. House districts to Massachusetts likewise included districts located in geographic Maine. Census tabulations from the time, including the totals in *Social Explorer*, generally provide separate numbers for the two populations. For this reason, I added the population for Maine to the Massachusetts numbers until Maine entered the Union as an independent state in 1820. The Massachusetts linear projection between 1811 and 1820 is anchored on the upper end by what the state's population would have been in 1820 if Maine had not broken away.

4) Texas, 1845-1851. Texas was admitted as a state in 1845. However, no census was conducted for the Republic of Texas in 1840 to enable the calculation of a linear projection leading up to the 1850 Census, which is problematic given the large population changes that occurred over the decade. However, reliable estimates put the population at the time of statehood at about 125,000, so I used that figure when imposing the linear projection. For background, see https://texasourtexas.texaspbs.org/the-eras-of-texas/early-statehood/.