Codebook for House primary elections (1980-2018):

District level variables:

raceid: [form year(XXXX)jacobson_district(XXXX)party(X)]
For example, if primary elections are being observed in 1990, for the
first district in Alabama the id would be as follows:
199001010 (Republican)
199001011 (Democrat)
199001019 (Top-two / jungle primary – no partisan primary)

year: Year of the election

stcd: First two numbers are the state code (i.e. 01: Alabama, 02: Alaska, 50:
Wyoming), second two numbers are the district number

state: Name of the state

incub_party: 0 = Republican incumbent
1 = Democratic incumbent
2 = Open seat held by Republicans
3 = Open seat held by Democrats
4 = Open seat, new (redistricting)
5 = Two incumbents (redistricting)

cand_party: 0 = Republican
1 = Democrat
9 = not one of the two major parties

redist: 1 = Redistricted boundaries
0 = Not redistricted

fr: 0 = Incumbent is not a freshman
1 = Freshman elected in previous general election
2 = Freshman elected in special election more than 1 year earlier
3 = Freshman elected in a special election less than one year ago
9 = Seat not defended by majority party incumbent
99 = New seat / two incumbents

For redistricting years (years ending with a '2'), fr variable is coded
"99" for districts with two incumbents or an open seat.

primary_type: 0 = Race used an open primary system
1 = Race used a closed primary system
2 = Race used a semi-closed primary system
3 = Top-Two Primary
4 = Louisiana-style primary (used in LA and special cases)
5 = Non-partisan blanket primary

candnumber: Total number of two-party candidates running in each party's primary. For Louisiana-style primaries and top-two primaries total candidates are listed because primaries have no partisan distinctions.

Write-in candidates are omitted for the candnumber because they are not reported consistently in FEC election returns or state election archives.

dem_pres: Democratic presidential two-party vote in the district for the previous election. Code "." indicates third-party or redistricting.

dem_vote: Democratic congressional two-party vote in the district for this general election. Codes 100 & 0 indicate previous primary was unopposed (100 – incumbent party unopposed or top-two single party; 0 – no out-party challenger). Code "." indicates third-party, redistricting, or Louisiana.

prev_vote: Democratic congressional two-party vote in the district for the previous general election. Codes 100 & 0 indicate previous primary was unopposed (100 – incumbent party unopposed or top-two single party; 0 – no out-party challenger). Code "." indicates third-party, redistricting, or Louisiana.

incname: Name of the seat's incumbent at the time of the primary

contest_type: 0 = Primary
1 = Caucus/convention, no sufficient opposition
2 = Caucus/convention, with sufficient opposition
3 = Caucus/convention/committee, no oppositional primary
4 = Other Primary Type

Convention distinction made because in certain states (Virginia and Connecticut) there is a pre-vetting process before the primary. Primaries are not fully "open" for all candidates to enter. Candidates must receive a specific percentage of the vote for a primary to occur.

contested: 0 = Unopposed primary
1 = Opposed primary
4 = No primary, convention

Candidate-level variables:

candidate: Name of the candidate in the election

candvotes: Actual number of votes for the candidate in the primary. "1" indicates uncontested primary; certain state election laws do not require votes to be counted in uncontested elections.

tvotes: Total number of votes cast in this party's primary. "1" indicates uncontested primary; certain state election laws do not require votes to be counted in uncontested elections.

candpct: Percentage of votes received by the primary candidate (candvotes/tvotes). "1" indicates uncontested primary; certain state election laws do not require votes to be counted in uncontested elections.

runoff: 0 = Candidate did not garner enough votes to participate in runoff
1 = Candidate participated in runoff election
. = No runoff election

runoff_winner: 0 = Candidate lost runoff election
1 = Candidate won runoff election
. = Candidate did not participate in runoff election

winner: 0 = Candidate lost election
1 = Candidate won election

Winner variable does not include the runoff winner; winner is the winner(s) of the initial party primary.

winwin: 0 = Candidate lost election
1 = Candidate won election

Winwin variable includes the runoff winner; winwin is the candidate who proceeded to the general election.

inc: 0 = Candidate is a challenger
1 = Candidate is the incumbent
Note: if there is another current representative running in a race that has not been designed a "two incumbent" race through redistricting, this "incumbent" is not coded 1. Their quality is coded "30".

gender: 0 = Candidate is male
1 = Candidate is female

Important Notes for Odd Primary States:

Louisiana: In most instances, only one election is held – there is not separate primary and general election. If one candidate garners 50% of the vote the candidate is declared the winner; if not candidate garners 50% of the vote a runoff election is held. All candidates are listed on the same ballot, including third party candidates. Louisiana did not use its typical primary system in 2008 and 2010.

Because there are no partisan primaries, we have taken several steps to accurate represent Louisiana's unique primary system. The party component of the raceid variable is coded as "9" because there is no separate democratic or republican primary. Louisiana's primary type if coded as "5" because it does not accurately fall into any other category. Unlike other states, with the exception of states with top-two primaries, third party candidates are included. This code is used in the dataset in one other instance for some Texas primaries in 1996 where court-mandated district changes required a Louisiana-style election to be held.

Washington: Until a 2002 Supreme Court decision, Washington used a non-partisan blanket primary system where all candidates were listed on the same ballot without party affiliation. The top Democrat and Republican advanced to the general election along with third party candidates who garnered a specific percentage of the vote. No major coding changes; third party candidates are not included in the dataset.

Hawaii: For several elections Hawaii utilizes a non-partisan blanket primary. No major coding changes; third party candidates are not included in the dataset.

Alaska: For certain elections Alaska utilizes a non-partisan blanket primary. In the early 1990's challenges from the Republican party fractured Alaska's election laws instituting separate systems for Democrats and Republicans – these disparities are included in the dataset. In non-partisan blanket primaries there are no major coding changes; third party candidates are not included in the dataset.

Connecticut & Utah: For all election years Connecticut and Utah use a modified convention system where party conventions nominate an official party candidate; however, if a non-winning convention participant garners a specific percentage of the vote a primary is held between the non-winner and official party candidate. The necessary percentage varies between states and across years. See the FEC for specific information by year. Information on Connecticut and Utah's use of a convention and/or primary is found in the contest_type variable.

Virginia: Local parties in each district determine if a convention or a primary is held. These determinations vary by year and by party. If a party chose to utilize a convention-only the contest_type if coded "4". If a party chose to utilize a primary the contest_type if coded "0" or "1". Virginia does not have oppositional primaries like Connecticut and Utah; there is no mix of convention and primary types.

Resource & Coding Information:

The current data, 2000-2016, was collected in combination with Stephen Pettigrew (Harvard University), Dr. Karen Owen (Reinhardt University), Dr. Emily Wanless (Augustana College), Joel Sievert (Texas Tech), Ryan D. Williamson (University of Georgia), and Jason S. Byers (University of Georgia). Stephen Pettigrew, along with help from Dr. Karen Owen and Dr. Emily Wanless, coded the years 2000-2010 as part of his Master's thesis at the University of Georgia. Joel Sievert, Ryan D. Williamson, and Jason S. Byers coded the years 2012-2016 as research assistance for Dr. Jamie L. Carson.

All election returns were collected from the America Votes book series and FEC official reported election results. Data were supplemented with district-level variables from Gary Jacobson's House general election dataset. Gary's data is often a useful starting for coding a specific year since you already have the general election outcome coded for both sides (which reduces the amount of data that needs to be coded at the primary stage).

Information about candidate quality and experience (for the 2000-2010 elections) was collected by searching for news records about the candidates. Specifically, it started with a Google News search for '[candidate name] [state name] Congress' and restricted the dates of the results to be only from the calendar year of the election in question. Then a similar search was conducted using Lexis Nexus Academic. Finally, simple Google searches were utilized. The later dates (2012-2016) were generally conducted using a simple Google search, being that most campaign information for candidates were readily available on the Internet.

There might be times when you are unable to find any information relating to a candidate's quality. If, after you have exhausted the searches from above you are unable to find any information on a particular candidate, code them as a 0 for quality. Our reasoning follows along with that of Gary Jacobson, meaning that if there is no information relating to a candidate then it is probably safe to assume that they held no previously elected position.

Given the extensive amount of hand-coding in this data, there are likely some coding errors. If you find any, please let us know.

Data Issues, Fixes, & Rationale:

- For 1980-1996 Louisiana was coded to be a partisan primary state, meaning instead of having being coded as a jungle primary Democrats and Republicans were divided into different categories. Third-party candidates were omitted. Louisiana from 1980-1996 was recoded.
- Louisiana, Utah, Connecticut, and Virginia were missing in 2000-2010; coding previously was incorrect or certain convention races were left off. These states were reviewed and recoded.
- Bernie Sanders is coded as a Democrat for simplicity, winner and incumbent primary codes reflect this change.
- Raceid variable was recoded so the party component was coded "9" for top-two primaries and Louisiana primaries. Non-partisan blanket primaries were kept the same because vote totals were provided by party and the top candidate from each party was chosen keeping enough elements of the traditional partisan primary to allow for the distinction to be included.
- All runoff elections from 1980 to 2016 were coded; any extra data points added to capture runoff elections (extra lines in Excel) were deleted.
- Primary type was recoded from 1980-1994 using the America Votes books only for those elections that had "special" primaries (non-partisan blanket, top-two, etc.)
- For primaries that had to be re-held due to illegal voting or court-challenged district boundaries, the official (second) primary was included in the dataset. The first (unofficial) primary was not included. In the original data, these unofficial results were reported and not the official results.
- For certain runoff elections included, there was no winner coded for the initial primary. The vote total winner was coded and the runoff winner was coded separately.
- The freshman variable has been coded in our dataset as a district-level variable, in the Jacobson data it is coded as a district level variable as well but in the context of the general election.
- For certain years when a top-two primary was used, the third-party candidates were not included. These were added and vote totals were modified to reflect this change.
- For redistricting years there was some confusion / inconsistency in who the "incumbent" was – using America Votes, Congress.gov, and Wikipedia these entries were updated.
- The seat variable pulled from Jacobson's data was incorrect for some years. Jacobson's variable captures the state of the seat as it enters the general election – this is specifically problematic for years when the incumbent loses in the primary. For these instances the seat is marked "open, held by x party" – however, in the primary election the seat is not open as the incumbent has not yet lost. This variable was recoded and updated in the current dataset. Further, some of Jacobson's coding options were consolidated for ease and simplicity.
- For certain specialty primaries, new codes were added to encapsulate more information into the dataset. Most notably, several new primary_type options were added and options for contest_type were added to ensure unique primaries

could be excluded from the data if a user desired. The additions to the contest_type variable were collapsed to create a dichotomous variable called contested.

- Several additional variables were created not included in the original dataset for analysis
- If a current member of Congress runs in another district in a given year that is not a redistricting year where the race is designated to have "two incumbents" that member's incumbency is not coded "1", their quality is coded "30".
- If an incumbent is listed twice (was nominated in both primaries) the incumbent is deleted from their non-party primary
- For certain elections that have a top-two primary, only one winner is coded (California 2014, Washington 2010).
- For some years, the total of "wingen" does not equal 435 (there are less winners listed than there should be). There are two potential explanations here: (1) the general election winner was a third party candidate that was not listed (ran outside of a non-partisan blanket, Louisiana, or top-two primary) (2) there was no candidate that ran in the primary and one was chosen by a local caucus.