

## APPENDIX:

### Defining the database

The 2016 Primaries Project Database of Candidates (the “database”) provides demographic, issue position, party category, and election return data for every candidate who appeared in a major party primary during the 2016 election cycle. Candidates include all major party House and Senate candidates who were on the ballot in 2016. The database also included the candidates of special elections that took place in 2016, including, for example, the Ohio 8<sup>th</sup> District special election held to fill the seat of former Speaker John Boehner (R-OH). The dataset includes over 1,400 candidates and includes nearly 100,000 cells of data collected from February to September of 2016.

The database was created in three steps. First, for each state, a list of candidates was made available by state secretary of state’s offices. These lists, often posted after filing deadlines passed, included information about the number of candidates running in each district-party primary. All of the major party candidates appearing on these lists were included in our dataset and were only deleted if they withdrew from the primary election or were disqualified/suspended from being featured on the primary ballot.

In Washington and California, minor party candidates were also included in the database since both states use a top-two primary system. Additionally, Libertarians and Alaskan Independence Party candidates in Alaska were included in the dataset since these candidates appear on the same ballot as Democrats. Further, as stated in footnote 3, Louisiana candidates were *not* included in our dataset because Louisiana does not have a traditional primary process.

In Virginia, Connecticut, and Colorado, partisans either nominate or filter candidates through conventions or assemblies. To capture the full primary process, the database includes candidates who competed at the convention and assembly level but did not move on to the primary ballot.

Second, after the list of candidates was determined, candidate websites were coded. To find the candidate websites (when they were not available by the state secretary of state), our primaries team earnestly searched to make sure every website was included. In the cases of candidates without websites, two members of our coding team confirmed that the website was not readily available (or, at the very least, not immediately available to the curious citizen). In terms of websites, one website per candidate was coded with priority given to campaign websites, then to incumbent .gov websites, and lastly to campaign Facebook pages. Personal Facebook pages, Twitter accounts, YouTube pages, and other social media accounts were not included in the database coding. If an incumbent’s campaign website was bare and redirected to her .gov website, the coding team coded the .gov page.

Third, described above in our methodology section, candidate websites were coded within four weeks of the primary election and then second-checked within two weeks of the election according to the rules stated above. Candidates were coded according to party categories, demographic information, and issue positions. Information about election returns were then added following each primary election.

## **Classifying party categories**

Following the party categorization from our 2014 Primaries Project, each Republican in the dataset was assigned to one of four categories: Business/Establishment Republican, Tea Party Republican, Conservative Republican, or Libertarian Republican. Democrats were assigned to one of three categories: Establishment, Progressive, or Moderate. In the cases of candidates without websites or with very little information (or unclear information), candidates were coded as fitting the Other category.

To assign the major party candidates to these categories, there was a four-step logic (i.e. the process would only move to step two if step one did not render a party category classification). First, the candidate websites were examined for any personal identification to one of these categories. Examples of self-identification were sometimes explicit (e.g. “conservative” might have been in the candidate’s slogan) but also sometimes implicit (e.g. a candidate mentioning how her business experience led her to politics). Whenever there were competing party categories (e.g. one can imagine a candidate who might be placed in either the Conservative Republican or Tea Party Republican categories), the following party category ranking system was in place: Tea Party, Libertarian, Conservative, and then Business/Establishment; Progressive, Moderate, and then Establishment. This ranking system allowed the coding scheme to prioritize assignment to more narrow groups.

Second, after personal identification, incumbent candidates were placed in the category according to caucus membership. For Democrats, memberships in the Progressive Caucus,<sup>1</sup> the (former) Populist Caucus,<sup>2</sup> and the (former) Moderate Democrats Working Group<sup>3</sup> placed candidates in the corresponding party categories. For Republicans, membership in the (former) Tea Party Caucus<sup>4</sup> placed incumbents in the Tea Party Republican category. Additionally, candidates in New York were also coded as Conservative Republican, Libertarian Republican, or Progressive Democrat if they were running on the Conservative Party, Libertarian Party, or Working Family Party lines, respectively, according to the Candidate Petition List by the New York Secretary of State.

Third, endorsements listed on the campaign website also determined party categories. If a candidate listed an endorsement from Tea Party Express, that candidate was counted as a Tea Party Republican. If a different candidate listed an endorsement by the U.S. Chamber of Commerce, then that candidate was categorized as a Business/Establishment Republican. In the case of multiple endorsements across party categories, endorsement placement and other contextual factors were taken into consideration.

Fourth, in the cases of candidates failing to meet any of the above party category criteria, candidates were coded according to the candidate issue positions. If, however, there was still confusion (or not enough information conveyed by a candidate’s issue positions), the candidate was simply coded as Other.

## **Candidate demographic data notes**

The following demographic characteristics of the candidates were collected by our team: candidate gender, marital status, education, and whether he or she serves or served in the military. When available by state secretaries of state, the demographic variables were taken from the states’ “List of Primary Candidates.” More often, however, these variables were taken from the candidates’ websites. In introducing themselves to voters, most candidates offer

biographical webpages (e.g., “Meet Cindy” or “About Cindy”). For the candidates who did not provide information about these demographic characteristics, we noted them as “No Information.”

In addition to the above-stated demographic variables, our team also coded dichotomous measures of whether the candidate has experience in elected office or whether the candidate has previously run for Congress. Following our coding practices, these pieces of data were only marked positive if that information appeared on a candidate’s website. (E.g. If a candidate was both a veteran and a state legislator but she did not list that information on her candidate website, then the codes for military service and experience in elected office would be listed as “No Information.”) The dataset’s demographic information is not meant to serve as a comprehensive biographical dataset of the candidates but rather a dataset evidencing how these candidates were presenting their biographical and demographic information in their campaigns.

### **Candidate issue positions data notes**

Nineteen issue positions are included in the dataset, with candidates’ issue positions being interpreted from their websites. In most cases, issues were found on webpages labeled “Platform,” “Issues,” “Press Releases,” or “Endorsements.” A candidate’s entire public website was scanned by our team to investigate that candidate’s issue positions including any positions conveyed by text, pictures, or embedded videos located anywhere on the website. The only website content on candidate websites that was not included in our coding was social media widgets due to their dynamic nature.

For each issue position, candidates were coded into one of four categories (labelled below as 1’s, 2’s, 3’s, and 4’s). In all of the issues below, the fourth category (4’s) was reserved for candidates who provided “No information” on that particular issue.

**Affordable Care Act.** 1’s indicated candidates explicitly supported the Affordable Care Act, included those who lauded the bill’s effects. 2’s were candidates who opposed the Act (often referring to it as Obamacare) and advocated for the Act’s repeal or defunding. 3’s included candidates who forward complex, complicated, or unclear positions. Also, 3’s included those candidates who argued the Act needs serious improvement (that is, to be “fixed” or expanded).

**Minimum Wage.** 1’s were candidates who explicitly supported raising the federal minimum wage. 2’s indicated a candidate expressly argued against raising the federal minimum wage. 3’s in this category again represented candidates who held complex, complicated, or unclear positions. 3’s also included candidates who wanted to leave the issue to the states. Candidates who mentioned they were in favor of a “living wage” but *did not* specifically mention anything about raising/not raising the minimum wage were coded as 4’s.

**Taxes.** 1’s in this category stood for raising taxes on corporations and/or wealthy individuals. 2’s in this category opposed raising taxes (or sometimes even mentioned their advocacy for lowering taxes for all). 3’s were complex/complicated/unclear positions as well as those who favored Tax Reform (but said nothing about tax rates), favored a FairTax or Flat Tax, or held one position about corporate taxes and an opposite position about tax rates on wealthy individuals.

**Business Regulations.** 1's indicated the candidate was in favor of government regulations on business (e.g. "smart regulations on business"). 2's indicated the candidate was against business regulation (e.g. "oppose burdensome regulations and red tape"). 3's were again reserved for candidates with complicated, complex, and unclear positions.

**National Debt/Deficit.** 1's were candidates who explicitly made comments in favor of increasing the national debt or the federal deficit. 2's indicated the candidate called for either lowering the national debt or deficit reduction. 3's were candidates with complicated, complex, and unclear positions. This issue code did not code comments surrounding the debt ceiling.

**Social Security.** 1's were candidates who simply stated they wanted to protect the status quo Social Security system. 2's included candidates who proposed reforms to reshape Social Security, such as increasing the retirement age, means-testing beneficiaries, or privatization. (Candidates who proposed reforms to raise the cap on Social Security Payroll taxes were counted in our **Taxes** code.) 3's were candidates with complicated, complex, and unclear positions.

**Immigration Reform.** 1's indicated candidates were in favor of comprehensive immigration reform, one that explicitly includes a path to citizenship for illegal immigrants. 2's were candidates who oppose comprehensive immigration reform, included those who oppose amnesty, favor "building a wall" or further border security, favor "*enforcing* our current immigration laws," or argued reform should only happen after the border is secured. 3's included candidates with complicated, complex, and unclear positions.

**Gun Control.** 1's were candidates who were in favor of any gun control, including background checks. 1's also included candidates who listed endorsements from various gun control organizations. 2's included candidates who opposed gun control or supported "no exceptions to the second amendment." 2's also included candidates listing endorsements from the National Rifle Association or other pro-second amendment organizations. 3's were candidates with complicated, complex, or unclear positions.

**Criminal Justice Reform.** 1's included candidates who support criminal justice reform, including candidates seeking solutions to fix incarceration rates and even those who explicitly support "Black Lives Matter." 2's were candidates who opposed criminal justice reform, including those who explicitly issue positions in opposition to "Black Lives Matter" or in support of "Blue Lives Matter." 3's were candidates with complicated, complex, or unclear positions.

**Abortion.** 1's indicated candidates hold a pro-choice position, including candidates who explicitly support reproductive rights, Planned Parenthood funding by the U.S. government, and "a woman's right to choose." 1's also included candidates listing endorsements from Planned Parenthood and other explicitly pro-choice organizations. 2's were candidates holding the pro-life position, including those candidates who call for the government to stop providing funding to Planned Parenthood. 2's also included candidates listing endorsements from various Right to Life organizations. 3's were those candidates with unclear, complex, and complicated positions (such as the candidate who argued he was "both pro-life and pro-choice.").

**Same-Sex Marriage.** 1's included candidates who explicitly support marriage equality (i.e. same-sex marriage) or "LGBT rights." 2's were candidates who argued in favor of "traditional family" or that marriage is "between one man and one woman." 3's included those with unclear, complicated, or complex issue positions.

**Climate Change Regulations.** 1's were candidates who support regulations and other measures to combat climate change/global warming. 2's were candidates who explicitly opposed climate change regulations and candidates who denied the existence or effects of climate change. 3's were candidates with unclear, complicated, or complex issue positions.

**K-12 Education Reform.** 1's included candidates who supported federal proposals for major education reform (including supporting increased federal funding towards education and supporting Common Core.) 2's were candidates who called for local solutions to education reform, including candidates who opposed common core, support cutting the Department of Education, support local charter school programs, or support "returning education decisions to parents and teachers." 3's included candidates with complicated, unclear, or complex issue positions.

**Campaign Finance Reform.** 1's included candidates who favor campaign finance reform, including overturning Citizens United, banning SuperPAC's, or implementing public campaign funding. 2's included candidates who were in favor of the status quo campaign finance system. 3's again included those with unclear, complicated, or complex issue stances.

**Handling ISIS Terrorism Abroad.** 1's indicated a candidate's support of status quo efforts to combat terrorism abroad. 2's were candidates who called for increased American intervention to combat terrorism abroad, including boots on the ground to fight ISIS. 3's included those with unclear, complicated, or complex issue stances.

**National Security Agency Surveillance.** 1's included candidates who supported data collection by the National Security Agency (NSA). 2's were candidates who opposed NSA data collection (or even made broader arguments against "government surveillance" or "big brother"). 3's were candidates with complicated, complex, or unclear issue positions.

**Benghazi Investigation.** 1's included candidates who oppose further investigation into the 2012 attack on U.S. facilities in Benghazi, Libya. 2's were candidates who supported further investigation into the attack. 3's were candidates with complicated, complex, and unclear positions.

**Defense Budget.** 1's were candidates who support reducing military spending (i.e. support cuts to the military budget). 2's were candidates who oppose reducing military spending, including those candidates arguing for a fully funded and strong military. 3's were candidates with complicated, complex, and unclear positions.

**Syrian Refugees.** In the final issue code there was an alternative coding scheme. Candidates were coded 1 if they *mentioned* Syrian Refugees or a "Muslim immigration ban." Candidates who did not mention this issue at all were coded 0.

---

<sup>1</sup> <https://cpc-grijalva.house.gov/caucus-members/>

<sup>2</sup> <https://web.archive.org/web/20141228120124/http://braley.house.gov/press-release/23-members-congress-across-america-join-establish-populist-caucus/>

<sup>3</sup> <http://politicalticker.blogs.cnn.com/2009/03/18/moderate-dems-working-group-organized-in-senate/>

<sup>4</sup> <https://web.archive.org/web/20121212040121/https://teapartycaucus-bachmann.house.gov/membership>