

Twitter accounts and demographic data of U.S. candidates in the 2022 midterm election

Jeremy Gelman¹, Valeriya Mechkova², Daniel Pemstein², Brigitte Seim², and Steven Wilson²

¹ University of Nevada, Reno

² Digital Society Project

United States of America Dataset 2022, Version 1.0

1 February 2022

Background

This data includes the Twitter accounts for candidates who ran for the United States Congress, governor, and state legislative office in 2022. For state legislative candidates, it only includes the Democratic and Republican candidates who ran in the general election. For federal and governor races, it includes primary candidates as well. Additionally, the demographics of federal and governor candidates were collected, while only a limited number of variables were collected for state legislative candidates (e.g., gender, incumbency status).

The tweets associated with the identified accounts were collected by the Digital Society Project using Twitter's Academic Research Product. If you are interested in using these posts for academic research, contact the Digital Society Project for more details on access.

The authors thank Olivia GrafMank, Talia Guzman, Solanna Schuck, Madeleine Stuart, and Amanda Vaskov for their excellent research assistance.

Verification of Twitter Accounts

Two research assistants were tasked with collecting every Twitter account associated with a candidate. This includes official government accounts as well as campaign, personal, and press accounts. Account type was distinguished using the following rules:

1. Official accounts – Used by incumbents or current officeholders in their official capacity. These are generally linked to their official webpage (e.g., their House of Representatives website).
 - @SenKlobuchar; RepAOC
2. Election accounts – Used by candidates as part of their campaigning. If linked to a website, it is to their campaign one.
 - @amyklobuchar; @AOC

3. Personal accounts – These are personal accounts that are not obviously linked to their campaigning.
 - @BenSasse
4. Media/office accounts – Explicitly maintained by communications staff.
 - @HawleyPress

To identify accounts, the RAs used Google and Twitter searches as well as links from official governmental websites, campaign websites, other social media accounts, and third-party candidate trackers like Ballotpedia. Both RAs searched for every candidate after a state's filing deadline, usually between 2 weeks and 1 month apart.

Demographic Information

Three research assistants collected candidate demographic information for federal and governor candidates. Completeness varies by what is available online and for many first-time or amateur candidates, little information was available. Variables like veteran status and education discipline are based on candidate self-reports. Election results were collected from State Secretary of State websites and the Associated Press.

Codebook

BIOID (string): Unique ID for each candidate.

- Each candidate's ID is a concatenation of the state they are running in (FIPS ID), the office they are running for (House = 01, Senate = 02, Governor = 03, State Lower House = 04, State Upper House = 05), the district they are running in (Senate = 0, governor = 99), and the candidate order they were collected in (this order is arbitrary for a given seat).
- This can be represented as: FIPS + office (House = 01, Senate = 02, Governor = 03, State Lower House = 04, State Upper House = 05) + district number (Senate = 0, governor = 99) + candidate order number

COUNTRY NAME (string): United States of America

TIME SPECIFIC COUNTRY NAME (string): United States of America

COUNTRY ID (numeric): 20 (V-Dem Country ID)

COUNTRY TEXT ID: United States of America

Correlates of War (COW) COUNTRY ID: 2

FNAME_ORIG: First name of candidate

LNAME_ORIG: Last name of candidate

FNAME: First name of candidate and identical to FNAME_ORIG

LNAME: Last name of candidate and identical to LNAME_ORIG

Twitter User Name Variables:

GOVTWIT: Government Twitter account

GOVTWIT_ID: Government Twitter account user ID

PTWIT: Personal Twitter account

PTWIT_ID: Personal Twitter account user ID

OTWIT: Other Twitter account

OTWIT_ID: Other Twitter account user ID

OTWIT_TYPE: Other account type: political in-office account (1), a political campaign account (2), a political press account (3), or a personal account (4)

Note: Candidates may have multiple other Twitter accounts. This is denoted in the data with multiple columns of the OTWIT variables (e.g., OTWIT_1, OTWIT_2).

Electoral System and Seat Variables:

OFFICE1: Office candidate is running for (e.g., U.S. House of Representatives, State Assembly)

OFFICE1_LEVEL: 0: Supranational (e.g., EU), 1: National, 2: Subnational (U.S. State)

OFFICE1_CONSTITUENCYLEVEL1: United States of America

OFFICE1_CONSTITUENCYLEVEL2: U.S. state that the seat is in

OFFICE1_CONSTITUENCYLEVEL3: Geographic district of seat (e.g., Congressional District 1). If the office is elected statewide, then variable is left blank.

OFFICE1_TERRITORIAL: 0 Non-territorial constituency; 1 Territorial constituency: delimited by geographically delimited boundaries

- Always coded as 1

OFFICE1_RESERVED: 0: Standard constituency, 1: constituency reserved for citizens with certain ethnic, religious, linguistic, racial or other

- Always coded as 0, including VRA districts.

OFFICE1_ELECTORALSYSTEM:

- 0: First-past-the-post (FPP, aka plurality) in single-member constituencies. The candidate with the most votes wins the seat.
- 1: Two-round system in single-member constituencies. Like FPP except that a threshold — usually 50% + 1 — is required to avoid a runoff between the two top vote-getters.
- 2: Alternative vote in single-member districts. Voters rank-order their preferences for the candidates who compete for a single seat. If any candidate receives an absolute majority of first preferences, s/he is elected.
- If not, then the least successful candidates (based on first-preferences) are eliminated and their votes reallocated to the second-preferences. This process is repeated until a candidate reaches 50% +1 of the votes.
- 3: Block vote in multi-member districts. Electors have as many votes as there are seats within that district and can rank-order them (within or across parties) as they please.
- 4: Party block vote in multi-member districts. Voters cast a vote for a single party (but not for individual candidates within the party's list). The party with the most votes (i.e., a plurality) wins all the seats in that district.
- 5: Parallel (SMD/PR). Some seats are in single-member districts (allocated by FPP or two-round electoral rules) and other seats are in multimember districts (allocated by some form of PR). These districts are overlapping, meaning that each elector votes twice: once in the single-member district race and once in the multi-member district race. Results are independent.
- 6: Mixed-member proportional (SMD with PR compensatory seats). Some seats are in single-member districts (allocated by FPP or two-round electoral rules) and other seats are in multi-member districts (allocated by some form of PR). These districts are overlapping, meaning that each elector votes twice: once in the single-member district race and once in the multi-member district race. Results are not independent. Specifically, the multimember seats are used to rectify disproportionalities achieved in the single-member district election — by adding seats, as necessary. This means that the representation of parties in the legislature is determined entirely by the PR ballot. It also means that the result of an MMP election is similar to the result of a PR election: parties achieve representation according to their nationwide vote share (on the PR ballot).
- 7: List PR with small multi-member districts (mean district size < 7). Each party presents a list of candidates for election within a district. Electors vote for a party, and parties receive seats in (rough) proportion to their overall share of the vote. Mean district size is less than seven.

- 8: List PR with large multi-member districts (mean district size > 7). Each party presents a list of candidates for election within a district. Electors vote for a party, and parties receive seats in (rough) proportion to their overall share of the vote. Mean district size is greater than seven.
- 9: Single-transferable vote (STV) in multi-member districts. Electors rank-order candidates nominated for a district. Candidates that surpass a specified quota of first-preference votes are elected. The remaining seats are chosen by reallocating the votes of the least successful candidates to elector's second- (or third-) preferences until the specified quota is reached. This process is repeated until all seats for that district are filled.
- 10: Single non-transferable vote (SNTV) in multi-member districts. Each elector chooses a single candidate. The candidates with the most votes (a plurality) win. (The number of winners is of course determined by the size of the district.)
- 11: Limited vote in multi-member districts.
Electors have more than one vote but fewer votes than the number of seats in the district. The candidates with the most votes (a plurality) win. (The number of winners is of course determined by the size of the district.)
- 12: Borda Count in single- or multi-member districts. Electors use numbers to mark preferences among candidates and each preference is assigned a value. For example, in a ten-candidate field a first preference is worth one, a second preference is worth .9, and so forth. These are summed and the candidate(s) with the highest total(s) is/are elected.

OFFICE1_MAGNITUDE (numeric): Number of seats elected in the constituency in which the candidate is running.

OFFICE1_CLOSEDLIST: 0 The candidate is running on an open or flexible party list

1 The candidate is running on a closed party list; NA There is no party list

- Always blank

OFFICE1_LISTNUM: Position on the list in OFFICE1 in ordered list systems. NA for other systems.

- Always blank

OFFICE1_ROUND: For multi-round (e.g., runoff) systems. NA for single-round systems.

- States with majority run-off requirements are listed as 2 if run-off occurred. States with ranked choice, instant run-off voting are listed as 1. Multiple rounds of voting are not recorded as separate observations.

OFFICE1_MAXROUNDS:

- 1 for single-round systems
- 2 for majority run-off systems.

OFFICE1_TERMLENGTH: Maximum term length of elected officials for OFFICE1, in years

OFFICE1_WITHDRAW PRIMARY:

- 0: Ran in primary election

- 1: Withdrew from primary
- 2: Disqualified from primary
- Blank: Did not run in primary

OFFICE1_PRIMARY ELECTION RESULT:

- 0: not elected
- 1: elected
- Text: Description of how general election candidates advanced if they did not win a primary election
- Blank: Did not run in a primary election

Note: For Louisiana's jungle primary, no run-off elections were necessary, so this variable was left blank.

OFFICE1_PRIMARY ELECTION PERCENTAGE: Percentage vote received in primary election

OFFICE1_WITHDRAW GENERAL

- 0: Ran in general election
- 1: Withdrew from general election
- 2: Disqualified from general election
- Blank: Did not run in general election

OFFICE1_GENERAL ELECTION RESULT:

- 0: not elected
- 1: elected
- Blank: Did not run in a general election

OFFICE1_GENERAL ELECTION PERCENTAGE: Percentage vote received in primary election

OFFICE1_ASSUMED OFFICE:

- 0: did not assume office
- 1: assumed office
- Blank for candidates who did not advance to general election

OFFICE1_RESULT DISCREPANCY: Missing if previous two variables are 0/0 or 1/1.

Note: Office2 variables are identical to those above but are used in cases where a candidate runs for two seats (e.g., a state legislator runs as an incumbent and for the U.S. Congress).

Party Variables:

*Note: Comparative party datasets generally only include Democrats and Republicans in their coding. If researchers want third-party candidates, the **Party Name, English** variable is where that information is included. Additionally, independent candidates are coded using different*

*terms, which varies by state (e.g., Independent, No Party Affiliation). The **Independent** variable captures all these variations.*

INDEPENDENT:

- 0: The candidate is running under a party/coalition label
- 1: The candidate is officially running as an independent, with no party/coalition label

PARTY NAME, ENGLISH: Same as V-Party v2paennname. The most common English name for this party in political science datasets as recorded in Party Facts' core parties (Döring and Regel 2019).

- This includes party names as listed on ballots, which includes many more parties than in the V-Party dataset.

PARTY NAME, ORIGINAL: Same as V-Party v2paornname. The most common original name for this party in political science datasets as recorded in Party Facts' core parties (Döring and Regel 2019).

PARTY SHORT NAME: Same as V-Party v2pashname. Harmonized short name based on Döring and Regel 2019

NUMERIC PARTY IDENTIFIER: Same as V-Party v2paid. Unique, numeric identifier for each party as recorded in Party Facts' core parties dataset (Döring and Regel 2019). In case a party changed names of scholars, sources, etc. used (slightly) different names when referring to this entity.)

PARTY FACTS ID: Same as V-Party pf_party_id. From Party Facts (Döring and Regel 2019).

CMPCODE: CMP Party Code--Comparative Manifestos party code, if available

CHES PARTY CODE: Unique Party ID for each party from Chapel Hill Expert Survey (CHES), if available.

GPS PARTY CODE: Unique numeric code (ID_GPS) for each party in the Global Party Survey (GPS) dataset, if available.

LIST NAME, ORIGINAL: Official name of the list for which the candidate is running. Code as "Independent" if no party/coalition affiliation.

- Always NA

OFFICE1_INCUMBENT: Incumbency status in constituency 1.

- 0: Does not currently hold a seat in the contested body
- 1: Currently holds the seat they are contesting (in list systems this means elected from this list, or replaced someone elected from this list last time)
- 2: Currently holds a seat in the contested body, but not in the constituency/district in which they are running (e.g., incumbent in a different FPTP constituency)

Note: Since the 2022 elections were held after a redistricting year, incumbents were coded as 1 if they retained seats that were geographically similar to those they held previously. Incumbents were coded as 2 if their seats were dramatically changed and/or were placed in a district with another incumbent.

OFFICE2_INCUMBENT: Same coding scheme as OFFICE1_INCUMBENT

YEAR-START: 2022

MONTH-START: Month that in-person voting began

DAY-START: Day that in-person voting began

YEAR-END: 2022

MONTH-END: 11

DAY-END: 8

Note: Election start dates are the earliest date in-person voting was available to voters anywhere in the state.

Demographic Variables:

GENDER (numeric): 0: male; 1: female; 2: third gender

BIRTHDATE (date): YYYY/MM/DD

EDUCATION (numeric): Highest Education Level:

- 1: No Post Secondary School
- 2: Technical School
- 3: Bachelors or equivalent
- 4: Professional degree (e.g. Law or MBA)
- 5: Masters or PHD finished

Note: Advanced degrees that are required to practice a profession (medical doctor, nursing, dentistry, etc.) and Masters in Public Administration and Public Policy are coded as professional degrees.

EDUCATION DISCIPLINE (numeric): Principal course of study in undergraduate degree, where applicable

- 1: Agronomy, agriculture
- 2: Engineering
- 3: Math, Computer Science, Statistics, or other applied math
- 4: Biology, Chemistry, Physics, or other natural science
- 5: Medicine
- 6: Economics, Business, Management
- 7: Social Sciences
- 8: Law
- 9: Humanities
- 10: Education
- 11: Military
- 12: Other (a residual category)

ETHNICITY/RACE (string):

White
White – Hispanic
Black
Asian
Latino/Hispanic
Native Hawaiian or Other Pacific Islander
Native American
Biracial/Multiracial

Note: The Ethnicity variable includes any additional information a candidate provided about their race/ethnicity. This can be used to create additional categories for individuals whose race is not well-captured by this schema.

IMMIGRANT (numeric):

0: Held citizenship in country of election at birth
1: Obtained citizenship in country of election after birth

COUNTRY OF BIRTH (string): Country where candidate was born

CURRENT PROFESSION (numeric): Job held at time of candidacy, or just prior to candidacy, including political office. Use ISCO88 codes:

PRE-POLITICAL PROFESSION (numeric): Last non-elected official job held prior to this election. Can be the same as current profession for non-incumbents

Note: Professions are coded using ISCO88 codes

(<https://www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf> and include the following additions:

- *Retirees = 9703*
- *Business owner, unspecified 1121*

OFFICE1_LAST CANDIDATE:

0: Was not a candidate in the most recent election for this office (either in this constituency or elsewhere)
1: Was a candidate in the most recent election for this office (either in this constituency or elsewhere)

OFFICE2_LAST CANDIDATE: Same coding as OFFICE1_LAST CANDIDATE

LAST NATIONAL LEGISLATIVE CANDIDATE:

0: Was not a candidate in the most recent prior national legislative election
1: Was a candidate in the most recent prior national legislative election

PRIOR NATIONAL LEGISLATIVE CANDIDATE COUNT: Integer count of the number of times, before this election, that the candidate has run for national legislative office (in any house) in the last 20 years.

PRIOR NATIONAL LEGISLATIVE CANDIDATE COUNT COMPLETE:

- 0: Insufficient official records to check candidate going back 20 years
- 1: Candidate checked against official records covering the last 20 years

LAST NATIONAL EXECUTIVE CANDIDATE:

- 0: Was not a candidate in the most recent prior national executive election
- 1: Was a candidate in the most recent prior national executive election

PRIOR NATIONAL EXECUTIVE CANDIDATE COUNT: Integer count of the number of times, before this election, that the candidate has run for national executive office, in systems with direct election to national office, in the last 20 years.

PRIOR NATIONAL EXECUTIVE CANDIDATE COUNT COMPLETE:

- 0: Insufficient official records to check candidate going back 20 years
- 1: Candidate checked against official records covering the last 20 years

Previous Public Office Variables

PRIOR HIGH PUBLIC OFFICE NAME 1: Original language office name

PRIOR HIGH PUBLIC OFFICE NAME ENGLISH 1: English translation of office name

PRIOR HIGH PUBLIC OFFICE TYPE 1:

- 0: Legislator, lower or only chamber
- 1: Legislator, upper chamber
- 2: Legislator-executive (i.e., prime minister)
- 3: Legislator-minister (cabinet level minister in parliamentary system)
- 4: Elected executive (e.g., president, mayor)
- 5: Appointed executive (cabinet minister level; don't code at local level, but would include Euro Commissioners at Supranational)
- 6: Judge

PRIOR HIGH PUBLIC OFFICE LEVEL 1:

- 0: Local (any office held below the state level)
- 1: Regional (state-level office in the US)
- 2: National
- 3: Supranational (include name of supranational organization in Comments)

PRIOR HIGH PUBLIC OFFICE START 1: YYYY-MM-DD

PRIOR HIGH PUBLIC OFFICE END 1: YYYY-MM-DD

Notes:

- *These variables are repeated as many times as necessary to include a politician's entire career.*
- *If a politician held a seat in a body and the seat changed, both seats are coded (e.g., US House District 1 and US House District 2)*
- *Dates an office is held and birthdate are coded to the most specific level available.*

- *In almost all cases, the most recent office is coded first. However, researchers should verify this is the case before assuming as such. In a few cases, these variables were coded as the oldest office first.*

OTHER POLITICAL EXPERIENCE: Has the candidate ever held an official position in one of the following organizations?

- 1: Trade union
- 2: Employers' organization
- 3: Interest group
- 4: Nongovernmental organization (NGO) or international non-governmental organization (INGO)
- 5: Social movement
- 6: Party youth organization
- 7: Local party organization
- 8: Regional party organization
- 9: National party organization
- 10: Party executive
- 11: Celebrity
- 12: Sports person

EDUCATION DISCIPLINE GRADUATE: Principal course of study in graduate degree, where applicable.

- 1: Agronomy, agriculture
- 2: Engineering
- 3: Math, Computer Science, Statistics, or other applied math
- 4: Biology, Chemistry, Physics, or other natural science
- 5: Medicine
- 6: Economics, Business, Management
- 7: Social Sciences
- 8: Law
- 9: Humanities
- 10: Education
- 11: Military
- 12: Other (a residual category)

ETHNICITY: Any ethnic/nationality information the candidate provided.

VETERAN: 1 if candidate identified is a veteran; blank otherwise.