



OPEN

DATA DESCRIPTOR

United States Precinct Boundaries and Statewide Partisan Election Results

Brian Amos^{1,3}, Steven Gerontakis^{2,3} & Michael McDonald^{1,3} ✉

We describe the creation and verification of databases of all precinct boundaries used in the United States 2016, 2018, and 2020 November general elections, enhanced with election results for all partisan statewide offices. United States election officials report election results in the smallest geographic reporting known as the precinct. Scholars and practitioners find these election results valuable for numerous use cases. However, these data cannot be augmented with other geographically-bound data, such as U.S. Census data, without precinct boundaries. Here we describe the collection of precinct boundary data from state and local election officials, sometimes provided in GIS formats, images, text descriptions, and – in rare cases – verbally. We describe how we verify boundaries with other election data, such as geocoded voter registration files. Our open-source data has appeared in redistricting litigation argued before the United States Supreme Court; and has been used by state and local redistricting authorities, media organizations, advocacy groups, scholars, and a vibrant community of mapping enthusiasts.

Background & Summary

A requirement for democratic accountability is that governments report election data. In the United States there is no national entity responsible for administering elections; responsibility devolves to sub-national state and local election administrators. Without centralized national administration, election officials report election data in non-standard formats that pose significant barriers to creating a unified national database¹. To preserve the secret ballot, election administrators report aggregate election results in geographic units commonly known as states, districts, counties, townships, and precincts (the names of sub-state entities vary across the country). The smallest of these geographies are precincts, which are on the scale of neighborhoods. Election officials use precincts to identify the polling location at which a voter will cast an in-person vote and the offices that will appear on the ballot². To augment election results with other contextual data one requires their geographic boundaries³.

State, district, county, and township boundaries – and contextual data related to them – are readily available from the U.S. Census Bureau and other sources. However, there is no nationwide database of accurate precinct boundaries, which poses challenges to their collection and standardization. To fill this gap, we created national precinct boundary databases enhanced with vote counts for candidates of all partisan statewide offices for the 170,098 precincts used in the 2020 U.S. November general election; the 152,217 used in 2018 (CA is in-process at the time of this writing), and the 177,202 used in 2016.

There are notable efforts to collect precinct boundaries from the roughly 3,000 local election officials across the country who are the primary data curators. The U.S. Census Bureau requests states provide precinct boundaries collected from their localities so they may be included in the Census Bureau's geographies for reporting demographic statistics. The principal use case of these data are redistricting of legislative districts following the decennial census. Phase Two of the Census Bureau's Redistricting Data Program – when the Bureau collects precinct boundaries – takes place in a year ending in '7' preceding a decennial census (<https://www.census.gov/programs-surveys/decennial-census/about/rdo.html>). Unfortunately, these boundaries can be inaccurate geographic representations of precincts used in most elections³. Precinct boundaries are time-bound, in that election officials frequently modify precincts from one election to the next for administrative reasons. Additionally, our quality control routinely identifies incorrect boundaries even when obtained directly from local election

¹Wichita State University, Department of Political Science, Wichita, Kansas, 67260, USA. ²University of Florida, Department of Political Science, Gainesville, Florida, 32611, USA. ³These authors contributed equally: Brian Amos, Steven Gerontakis, Michael McDonald. ✉e-mail: michael.mcdonald@ufl.edu

officials. It is for these reasons a labor-intensive effort is required to collect and verify precinct boundaries used in each election. Following the 2010 census, a team from Harvard and Stanford universities attempted to collect 2008 general election precinct boundaries, but discontinued their efforts (<https://projects.iq.harvard.edu/eda/home>). Authors of this essay contributed to that effort and seek to continue it.

Stakeholders find geographically-bound precinct election results important to understanding democratic processes, representation, and governance. A primary use is redistricting, where stakeholders desire district partisanship measures derived from statewide office elections⁴. Our databases are incorporated into online redistricting mapping and evaluation tools, such as Dave's Redistricting App (<https://davesredistricting.org/>), DistrictBuilder (<https://www.districtbuilder.org/>), and PlanScore (<https://planscore.org/>). Numerous media organizations use our database in their election coverage, including CNN, *New York Times*, *Washington Post*, and *Wall Street Journal*. State governments used our databases for their redistricting, including New York, Ohio, and Virginia. A related use case is voting rights litigation, where experts analyze precinct data to estimate racial voting patterns using a technique known as ecological inference⁵. Among the court cases using our databases was a successful challenge to Alabama's congressional districts as a racial gerrymander decided by the U.S. Supreme Court in *Allen v. Milligan* 599 U.S. 1 (2023). Scholars use our databases to evaluate redistricting outcomes^{6,7}, to develop new partisan gerrymandering metrics and solutions^{8–11}, to analyze effects of the U.S. Census Bureau's differential privacy policies¹², to estimate racial voting patterns among congressional districts¹³, to analyze polarization among state legislators¹⁴, to analyze Latino voting patterns¹⁵, to analyze suburban voting patterns¹⁶, to analyze U.S. COVID policies^{17–19}, to analyze local crime policies²⁰, to map campaign donation patterns²¹, and to augment other geographically bound databases with partisan voting data^{8,22}.

Statewide partisan offices on the ballot vary among the states depending on the election. In a presidential election year, all states have the presidential election on their November ballot. Appearances of other offices vary with timing of U.S. Senate elections and state laws regarding when and what state offices are elected. As depicted in Table 1, we report precinct election results for thirty-six offices spanning the 2016, 2018, and 2020 November general elections, with a combined 1,736,409 cells.

Methods

We produce nationwide electronic geographic information system (GIS) databases of precincts used in the 2016, 2018, and 2020 United States November general elections, augmented with precinct-level election results for statewide partisan offices. We on occasion produce and publicly release databases for selected elections other than November federal elections, such as primary elections, when we have capacity to create these databases. There are two important data production phases: the creation of a precinct boundary database for each state and the creation of precinct-level election results that can be joined to these boundaries.

States and the federal government use various names for what are generally known as precincts. States may also call these “wards” or “election districts.” The U.S. Census Bureau acknowledges this naming variety by calling these geographies “voting tabulation districts” or VTDs. We use the common name “precincts” to refer to any such small geographic boundaries used by election officials for managing elections and reporting election results. Our naming convention includes smaller geographies created when election officials occasionally report election results for precincts split by legislative district boundaries. Likewise, the local government that manages federal and state elections is generally what is known as a county. States may have other names for county-equivalents, such as “parishes” and sub-county governments such as “townships” or “districts” may administer elections. We use the common name “county” to refer to any local government responsible for maintaining precinct boundaries and reporting election results.

Precinct boundaries. There is no standard format for precinct boundary data. Frequently, governments publish precinct boundaries in an electronic format known as shapefiles – a proprietary GIS format developed by software vendor ESRI, and since cracked for general use. We prefer election officials to provide a shapefile (or any similar electronic GIS formats, hereafter we simply call all GIS formats “shapefile”) since this allows us to edit boundaries if we detect and verify errors. We obtain shapefiles from the U.S. Census Bureau, state election officials, and local governments. We ultimately convert *all* boundary data we ingest into the shapefile format for re-dissemination. This approach allows us to attach precincts' election results to the boundary data as attributes of the precinct shapes.

The second-most frequent map format election officials provide us is a map image. Sometimes a map image is itself created with GIS software, but election officials may be unable to provide us with the shapefile. Usually this is because election officials did not create the map; another local government agency or an outside vendor created it. When possible we navigate local government contacts or submit open records requests to obtain the canonical shapefile. These steps are not always fruitful, as the outside agent may charge a fee or be otherwise unable to provide the shapefile. Not all map images are created by GIS software. Sometimes election officials will provide a hand-drawn map. Sometimes county officials cannot provide an electronic image file, in which case election officials may take a cell-phone picture of their map and forward it to us.

The third-most frequent map format election officials provide us with is a written or verbal description. We encounter this situation in rural counties or small townships that have few precincts and little GIS capacity. While this may seem antiquated, in reality election officials do not commonly manage which precincts their registered voters are assigned to using electronic maps. Instead, they manage their voter registration databases using master street address files that identify which precinct each street address range is located in (e.g., even numbered 100–198 Main Street is associated with precinct 1). When we geocode voter registration files during our quality assurance protocols we may detect errors in these master street address files when we observe a street range assigned to one precinct while surrounding neighbors are assigned to another²³. We have consulted with the Colorado and Virginia state governments to identify and rectify these errors, and alerted numerous

Office	2016	2018	2020
Attorney General	34108	100486	34551
Auditor	24490	36862	24743
Chief Financial Officer	143	6152	144
City Council Member		143	
Clerk of the Supreme Court		672	
Commissioner of Agriculture	4425	24086	4333
Commissioner of Insurance	10763	8660	10982
Commissioner of Labor	2704	4609	2662
Commissioner of Public Lands	7197	13045	7464
Commissioner of School and Public Lands		737	
Comptroller	10090	39439	
Corporation Commissioner		1951	1948
Council Chairman		143	
Delegate to the U.S. House	143	143	144
Governor	26124	114083	25247
Lieutenant Governor	13939	21865	14577
President	177202		170198
Mayor		143	
Public Service Commissioner	5129	5074	5073
Public Utilities Commissioner	747	737	737
Railroad Commissioner	8832	8936	9014
Secretary of Commonwealth		2173	
Secretary of State	18759	72828	17811
State Appeals Court	4196	6190	6551
State Board of Education	4812	4801	4756
State Controller		3004	
State Court of Criminal Appeals	8832	10928	10986
State Mine Inspector		1489	
State Supreme Court	15028	15126	15565
State University Regent	3010	3136	
Superintendent of Public Instruction	8771	9761	3328
Tax Commissioner		424	
Treasurer	28473	61680	29141
U.S. House	5688	3618	3609
U.S. Senate	115565	112084	78044
University Board of Regents/Trustees/Governors	4812	4801	4756
Total	177202	152717	170198

Table 1. Statewide Office Precinct Count *Note:* CA 2018 is in production at the time of this writing.

localities elsewhere of potential issues. These errors pose a dilemma as to which boundaries to report. Since our use case involves the geographic location of voters, we generally opt to generate modified precinct boundaries that include these verified assignment errors.

Lastly, rarely local officials do not respond to our multiple requests for their precinct boundaries (in extreme cases we have made repeated contact attempts spanning more than a year). In these cases we rely upon other information that provide clues as to the precinct boundaries, such as geocoded voter registration addresses with precinct identifiers, or other local districts and governmental units that local election officials align precinct boundaries with (depending on state and local practices). A detail we frequently wrestle with is local municipality annexations in localities that require municipal and precinct boundaries to coincide. We often rely on published annexation reports to verify the correctness of the precinct boundaries. When we detect conflicts, we contact local election officials to resolve them.

We create our own GIS precinct maps or modify shapefiles that we verify have errors. We nearly always have a reference precinct map to start from that we obtained from the U.S. Census Bureau or a shapefile created by us from a prior election. When we do not have any initial resource, we created a machine learning algorithm to fabricate a starter precinct map from census geography as a basemap using geocoded voter registration addresses and their associated precincts. Automation rarely yields usable precinct boundaries without extensive additional editing, since geocodes are imprecise, particularly in rural areas²³. Once we have a base map, we examine ancillary data to create the most accurate representation that we can.

In Tables 2, 3 we present statistics drawn from our extensive documentation to provide a sense of the scope of our work. Table 2 presents the states from Alabama to Missouri and Table 3 presents Montana to Wyoming,

State	Total Units	2020 Units Adjusted	2018 Units Adjusted	2016 Units Adjusted	2020 Non-Precinct Vote Allocated	2018 Non-Precinct Vote Allocated	2016 Non-Precinct Vote Allocated
Alabama	67	67	67	67	Y	Y	Y
Alaska	20	0	0	0	Y	Y	Y
Arizona	15	3	3	2	N	N	N
Arkansas	75	9	26	36	N	Y	Y
California	58	6	–	0	N	–	N
Colorado	64	7	6	7	N	N	N
Connecticut	169	55	56	57	N	N	N
Delaware	21	4	3	0	Y	Y	Y
District of Columbia	1	0	1	1	N	N	N
Florida	67	29	28	34	Y	Y	N
Georgia	159	20	9	5	N	N	N
Hawaii	5	0	0	0	N	N	N
Idaho	44	16	18	14	Y	Y	Y
Illinois	102	2	4	11	Y	N	Y
Indiana	92	43	56	56	N	Y	Y
Iowa	99	12	7	6	N	N	N
Kansas	105	21	20	11	N	N	N
Kentucky	120	0	–	10	Y	–	Y
Louisiana	64	7	5	5	Y	Y	Y
Maine	425	3	3	0	Y	Y	Y
Maryland	24	0	0	0	N	Y	Y
Massachusetts	351	16	19	15	N	N	N
Michigan	83	8	4	2	Y	Y	Y
Minnesota	87	0	0	0	N	N	N
Mississippi	82	82	82	82	N	N	N
Missouri	115	60	90	90	Y	Y	Y

Table 2. Counties with Precinct Changes and States Requiring Vote Reallocation (Alabama to Missouri). *Notes:* “Units” for all states are counties except AK and DE (state legislative districts) and CT, ME, MA, NH, RI, and VT (cities and townships). KY 2018 had no statewide partisan election. CA 2018 is in production at the time of this writing.

with totals for the entire United States. The first column presents the number of geographical “units” within each state, which are counties for all states except those that use state legislative districts to group precincts (Alaska and Delaware) and those that use cities or towns (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont). In the next three columns we present the number of geographical units where, for the 2020, 2018, and 2016 November elections, we edited at least one precinct boundary by splitting a precinct into two or more parts, merging it with another precinct, or drawing new precincts. In a given election, we edit at least one precinct – and often more – in about one-sixth of the geographical units. These are the geographic units where we make changes. Our quality assurance protocols include checking all precincts whether they require further editing or not.

In six states we did not modify any precincts across all three election years (Alaska, Hawaii, Maryland, Nevada, Oklahoma, Utah, and Wisconsin). State election officials in all of these states – except Nevada – provide a statewide precinct shapefile. Indeed, nearly all states have a statewide map either disseminated by the state or available from the Census Bureau’s Phase 2 redistricting data program production. We welcome a statewide precinct map, but its existence does not guarantee our mapping work is done. We often find issues – errors or out-of-date boundaries – that we resolve by collecting precinct maps from localities.

Tables 2, 3 under-represent the breadth of our mapping work. We do not count certain tasks that we perform regularly. We do not count adjusting precincts for city and town annexations in states where precincts do not cross local government boundaries. When we collect maps from local governments, electronic representations of local governments’ boundaries may not nicely conform with one another due to technical GIS details, such as differing map projections. We do not count ensuring precinct boundaries from different sources do not result in overlapping precincts – and in a rare cases do not extend outside state boundaries. We do not count where we created precinct maps from sources other than those obtained from election officials such as local land parcel maps, which require extra effort to identify as a source for precinct boundaries and to collect. We do not count when we create electronic representations of local governments’ precincts that localities adopt as their official precinct map.

Election results. Precinct boundary data are most valuable when augmented with election results. After constructing a statewide precinct map for an election of interest, we merge election results for partisan statewide

State	Total Units	2020 Units Adjusted	2018 Units Adjusted	2016 Units Adjusted	2020 Non-Precinct Vote Allocated	2018 Non-Precinct Vote Allocated	2016 Non-Precinct Vote Allocated
Montana	56	1	1	1	N	N	N
Nebraska	93	21	0	0	Y	Y	Y
Nevada	17	0	0	0	Y	Y	Y
New Hampshire	221	2	2	2	N	N	N
New Jersey	21	0	1	4	Y	Y	Y
New Mexico	33	6	7	7	N	N	N
New York	62	14	23	15	Y	Y	Y
North Carolina	100	3	0	0	Y	Y	Y
North Dakota	53	8	24	24	N	N	N
Ohio	88	18	41	45	N	N	N
Oklahoma	77	0	0	0	Y	Y	Y
Oregon	36	4	4	4	N	N	N
Pennsylvania	67	64	65	64	N	N	N
South Carolina	46	1	1	0	Y	Y	Y
South Dakota	66	39	40	40	Y	Y	Y
Tennessee	95	8	3	1	Y	Y	Y
Texas	254	11	11	12	Y	Y	Y
Utah	29	0	0	0	Y	Y	Y
Vermont	247	25	25	25	Y	Y	Y
Virginia	133	36	32	37	Y	Y	Y
Washington	39	0	0	0	N	N	N
West Virginia	55	55	55	55	N	N	N
Wisconsin	72	0	0	0	Y	Y	Y
Wyoming	23	10	10	10	N	N	N
U.S. Total	4,528	791	846	852	25	26	26

Table 3. Counties with Precinct Changes and States Requiring Vote Reallocation (Montana to Wyoming and U.S. Total). *Notes:* “Units” for all states are counties except AK and DE (state legislative districts) and CT, ME, MA, NH, RI, and VT (cities and townships). KY 2018 had no statewide partisan election. CA 2018 is in production at the time of this writing.

offices. This step serves as an additional validation check, as described below. Our focus on statewide partisan offices is primarily for purposes to measure the “normal vote,” the partisanship of a precinct, district, locality, or other geography⁴. The scope of offices includes U.S. President, U.S. Senator, Governor, and other statewide offices. U.S. President is the only office appearing on all ballots in the presidential election years of 2016 and 2020. Other offices vary with the timing of U.S. Senate elections – a third of the seats are scheduled for election every two years – and the varied state offices whose terms expire in a given election. In a midterm election year, such as 2018, the president does not appear on the ballot, but states tend to elect governors and other state offices. Not all states do, however, raising the occasional possibility a state has no partisan statewide office on the ballot. When no statewide partisan office is on the ballot, we try to collect precinct boundaries augmented with U.S. House results. We also include U.S. House races for single-district states since these qualify as a statewide partisan office.

We create databases of precinct-level vote totals for every partisan statewide office. We attempt to tally results for every candidate, including write-in candidates, when available. When election officials recognize “official” write-in candidates through a formal ballot qualification process, election officials may report their vote totals alongside all major and minor party candidates, which allows us to report votes for each of these candidates. Sometimes election officials may choose to report all write-in candidates in a single category, in which case we report only the aggregate write-in tallies.

We collect election results in a suitable electronic format to merge with precinct boundaries. Most often election officials report candidates’ votes in an electronic spreadsheet of some sort. Some states and counties still report election results in a scanned portable document format (pdf), either generated from software or a scanned image, requiring conversion to an electronic spreadsheet. Our experience with pdf to spreadsheet conversion software is we nearly always need to perform additional cleaning and reformatting.

Merging election results with precinct boundaries requires a common and unique identifier in both data sources. Sometimes precinct identifiers in these databases are different. Often these are minor inconsistencies resolved through visual inspection of names. Infrequently, we may contact local election officials to resolve inconsistencies. Our preference is to produce shapefiles with precinct identifiers as they appear in election results so that others may merge more election results data, if desired.

“At-large” precincts may cover an entire state, county, or a sub-unit within a county, such as districts or localities. Some special-purpose precincts have very small boundaries, typically consisting of the city block where an election office is located. We typically treat these government office precincts the same as at-large precincts in our data processing. Depending on the locality, election officials may create these precincts to report election

results for mail ballots, in-person early votes, overseas votes, provisional ballots, disability-assistance votes, or some combination of these. Not all localities report election results by voting methods in special at-large precincts, as some may tabulate and report these votes within the voters' home geographically-bound precincts. When possible, we collect election results that allocate these votes to voters' home precincts and often these reports are available from county – not state – election officials, entailing additional data collection.

The last three columns of Tables 2, 3 provide a sense of the extent where we allocate at-large votes for geographic units within a state during the 2020, 2018 and 2016 November elections. In about half the states we apportion at least one – usually more – geographic unit's votes. Typically, we apportion a candidate's at-large votes to geographically-bound precincts proportional to the candidate's votes within the geographically-bound precincts. For example, if Joe Biden receives 100 votes in an at-large precinct and two geographically-bound precincts within it have 600 and 400 votes for Biden, we apportion 60 of the at-large Biden votes to the first precinct and 40 to the second. We apportion fractions such that the largest remainders are awarded first so that the resultant precinct counts are whole numbers and tally correctly to the county-level results.

An important distinction between our database and other precinct databases, such as the database produced by the MIT Election Data and Science Lab¹, is that these other data providers report at-large precincts as separate rows and do not disaggregate to geographically-bound precincts as we do. These differing approaches primarily involve our respective use cases. We are most interested in measuring candidate votes cast within the geographic bounds of a precinct, even if we must estimate votes by disaggregating votes election officials report in at-large precincts. Our approach permits analyses that account for the political character of a geographic unit that splits counties, such as a legislative district. Their use case is primarily to measure candidate votes within precincts of all types, which enables analyses of voting by different methods.

When we complete our allocations, we verify our votes tallies with official county-level election reports. We investigate discrepancies when precinct election results do not match exactly. In most cases vote tally discrepancies reveal errors in our data production, but we encounter very rare circumstances where either precinct-level or county-level election results are in error or incomplete. Sometimes discrepancies are by design. States may censor small vote tallies to protect voters' confidentiality and the secret ballot. The North Carolina State Board of Elections adds a small amount of noise to their state's precinct results per state law whenever a candidate receives one hundred percent of the vote within a reporting unit and voters' choices would be revealed.

Replication. As described, there are two components to our data collection, precinct boundaries and precinct election results. It is possible to replicate collection of precinct election results. Indeed, a team of MIT scholars created databases of precinct election results, and our teams have shared information on our data collection efforts¹. Retrospective replication of precinct boundaries following the procedures we describe is technically possible, but not practical. An independent team or scholar would encounter logistical difficulties reproducing retrospective versions of our precinct boundary maps. Election officials rarely archive precinct boundary data or maps, even when they are produced in electronic formats that would facilitate archiving. As time passes and turnover occurs within election offices institutional memory about precinct boundaries fades. Prospective continuation of our data production is feasible, but is labor intensive, more so without our knowledge of where hurdles exist and how to navigate over them.

Data Records

We post our 2016, 2018, and 2020 general election shapefiles for each of the fifty U.S. states plus the District of Columbia on public archives. Our original repository is the Harvard Dataverse^{24–26}. We mirror the Harvard Dataverse archive at the Election Lab at the University of Florida data archive (<https://election.lab.ufl.edu/data-archive/>) We post updates to our databases at both locations. Some databases in addition to the 2016, 2018, and 2020 general election shapefiles can be found only on the Election Lab archive. These databases include primary elections and state general elections – such as those taking place in odd-numbered years – that are held outside federal general election years.

Shapefiles are used for GIS applications and are in practice a collection of files, one of which is a file that includes the attributes stored in a dBase format accessible by most statistical and spreadsheet software. Statewide precinct boundary shapefiles tend to be large, so we produce a separate shapefile for each state. This assists us and our user-base two-fold. From our end, our production workflow is to release each state as it is completed rather than delaying release until a nationwide file is complete. This enables our users to obtain data for their states of interest as soon as we complete our work.

The attributes for each precinct record include information to uniquely identify precincts within a state. Precinct names are not always unique among counties within a state. Precincts are named or numbered by local election administrators and it is thus possible two localities within the same state can use the same precinct identifier, particularly when they sequentially number precincts. We standardize geographic identifiers within a state, but do not adopt a standardized schema across states. States have as few as one geographic field (Delaware – which embeds state legislative district identifiers in its precinct codes) and up to fourteen fields (Georgia – which identifies parts of precincts split by legislative districts) needed to uniquely identify precincts. In some cases, these identifiers are duplicates to a degree, with separate fields identifying a geography with a code and a long text name. We adopt the full schema used by a state for their election results, which facilitates the merging of election results to precinct boundaries beyond the statewide partisan offices we provide. Further complicating data schema is that they may change from one election to the next.

We identify individual candidates by a ten-character code, for example, G20PRERTRU. The first character denotes the election type, which can be 'G' for a general election, 'C' for recount results, 'P' for a primary, 'S' for a special election, and 'R' for a runoff election. The second and third characters denote the last two digits of the year of the election, '16,' '18,' or '20.' The fourth through sixth characters reference the office code, a list of

Code	Office
AGR	Agriculture Commissioner
ATG	Attorney General
AUD	Auditor
COC	Corporation Commissioner
COU	City Council Member
DEL	Delegate to the U.S. House
GOV	Governor
H##	U.S. House, where ## is the district number ('AL' denotes at large)
INS	Insurance Commissioner
LAB	Labor Commissioner
LAN	Commissioner of Public Lands
LTG	Lieutenant Governor
PRE	President
PSC	Public Service Commissioner
RRC	Railroad Commissioner
SAC	State Appeals Court (in Alabama, Civil Appeals Court)
SCC	State Court of Criminal Appeals
SOS	Secretary of State
SSC	State Supreme Court
SPI	Superintendent of Public Instruction
TRE	Treasurer
USS	U.S. Senate

Table 4. Statewide Office Codes.

which is provided in Table 4. The seventh character is a political party code. Major political parties are identified as 'D' for Democrat and 'R' for Republican; codes for various minor state political parties are identified in our documentation. The eighth through tenth characters represent the first three characters of a candidates name. Unusual exceptions to our candidate schema are described in our documentation.

Technical Validation

We utilize processes to verify precinct boundary correctness similar to those we use to draw maps from scratch – geocoding and comparing boundaries to existing local boundaries. In addition, we compare precincts to prior versions available to us through our work, either drawn ourselves or collected from other sources. Boundaries that do not change when we expect they should are suspect. Election officials in rapidly growing urban areas often create new precincts with new polling places to better meet voting demand. Election officials often conform precinct boundaries with other local political boundaries, so we may expect new precinct boundaries following legislative redistricting at any level of government, especially when precincts define districts for local governments such as city or county legislatures. In the extreme, we observe precinct boundaries that appear to be at least a decade out of date in that they are the same as those submitted to the Census Bureau as part of their 2010 Phase 2 Redistricting Data collection.

The merging of election results serves as another verification check. The number of geographically-bound precincts with reported election results should align with the number found on a map, setting aside at-large precincts. Precinct names may provide clues that precinct boundaries changed. Sometimes election officials split precincts into two or more precincts because a precinct's number of registered voters has grown to a point where voters are better served with the creation of a new polling location. Election officials will often signify these child precincts with a suffix of 'A' and 'B' or '1' and '2,' which serve as indicators of areas needing attention. In the reverse, local election officials may also consolidate two or more precincts into one precinct, usually resulting in the disappearance of suffixes or a precinct name. We may also detect a boundary realignment when one precinct unexpectedly gains votes over the last election and a neighbor loses votes. Some localities name precincts after their polling place, and name changes may – but do not always – signal new boundaries. Precinct changes due to local annexations are not always obvious from elections data, since these relatively small adjustments do not often result in name changes or changes in the number of precincts. For these, we collect annexation notices filed by local governments.

In the course of our work we have encountered oddities. A rural election office that burned down along with all its election data. Rural counties that allow voters to decide which precinct they live in, and which polling place they will vote at, creating intermingled precincts that defy boundaries, so we create one precinct for the entire county. We've discovered individuals and even an entire neighborhood assigned to vote in the wrong county, which we verified with election officials. On rare occasions we identify errors in precinct boundaries and in certified vote totals. We work with election officials to correct these issues so that overall election administration may be improved. We've shared election maps we've created with election officials who do not have GIS capacity, so they may have accurate representations. Our work has even been included in a few localities'

2020 Phase 2 Redistricting Data Program transmission of precinct boundaries to the U.S. Census. We strive for perfection but know the reality of working with big data is we will not catch all errors. Our large user-base (our databases have over two hundred thousand downloads) includes thousands of mapping enthusiasts who create election results maps for dissemination on social media. They include tens of thousands of users who create DIY redistricting plans using online mapping applications. Our users act as crowd-sourcing agents, and we welcome and research their error reports.

Usage Notes

Our databases are released under a Creative Commons Attribution 4.0 International license (<https://creativecommons.org/licenses/by/4.0/deed.en>). Users are welcome to share and adapt our work as long as they provide appropriate credit. Unfortunately, attribution has at times been challenging, perhaps due to the success of our work. We have observed peer-reviewed published research attribute our work to other organizations that re-disseminate our databases. We hope this essay will provide future users a viable and persistent citation to our work.

Code availability

We used no customized software for our databases.

Received: 11 June 2024; Accepted: 22 October 2024;

Published online: 29 October 2024

References

- Baltz, S. *et al.* American election results at the precinct level. *Sci. Data* **9**, 1–12, <https://doi.org/10.1038/s41597-022-01745-0> (2022).
- Herrnson, P., Hanmer, M. & Niemi, R. The impact of ballot type on voter errors. *Am. J. Polit. Sci.* **56**, 716–730, <https://doi.org/10.1111/j.1540-5907.2011.00579.x> (2012).
- Amos, B., McDonald, M. & Watkins, R. When boundaries collide: constructing a national database of demographic and voting statistics. *Public Opin. Q.* **81**, 385–400, <https://doi.org/10.1093/poq/nfx001> (2017).
- McDonald, M. P. Presidential vote in legislative districts. *State Polit. Policy Q.* **14**, 196–204, <https://doi.org/10.1177/1532440014529291> (2012).
- King, G. *A Solution to the Ecological Inference Problem* (Princeton University Press, Princeton, NJ, 1997).
- Warshaw, C., McGhee, E. & Migurski, M. Districts for a new decade – partisan outcomes and racial representation in the 2021–2022 redistricting cycle. *Publius: The J. Fed.* **52**, 428–451, <https://doi.org/10.1093/publius/pjac020> (2022).
- Artes, J., Kaufman, A. R., Richter, B. K. & Timmons, J. F. Are firms gerrymandered? *Am. Polit. Sci. Rev.* 1–21, <https://doi.org/10.1017/S0003055424000558> (2024).
- de Benedictis-Kessner, J., Lee, D. D. I., Velez, Y. R. & Warshaw, C. American local government elections database. *Sci. Data* **10**, 912, <https://doi.org/10.1038/s41597-023-02792-x> (2023).
- Dobbs, K. W., King, D. M. & Jacobson, S. H. Redistricting optimization with recombination: A local search case study. *Comput. & Oper. Res.* **160**, 106369, <https://doi.org/10.1016/j.cor.2023.106369> (2023).
- Dobbs, K. W., Swamy, R., King, D. M., Ludden, I. G. & Jacobson, S. H. An optimization case study in analyzing missouri redistricting. *INFORMS J. on Appl. Anal.* **54**, 162–187, <https://doi.org/10.1287/inte.2022.0037> (2024).
- Palmer, M., Schneer, B. & DeLuca, K. A partisan solution to partisan gerrymandering: The define–combine procedure. *Polit. Analysis* 1–16, <https://doi.org/10.1017/pan.2023.39> (2023).
- Kenny, C. T. *et al.* The use of differential privacy for census data and its impact on redistricting: The case of the 2020 u.s. census. *Sci. Adv.* **7**, <https://doi.org/10.1126/sciadv.abk3283> (2021).
- Kuriwaki, S., Ansolabehere, S., Dagonel, A. & Yamauchi, S. The geography of racially polarized voting: Calibrating surveys at the district level. *Am. Polit. Sci. Rev.* **118**, 922–939, <https://doi.org/10.1017/S0003055423000436> (2024).
- Hunt, C. & Rouse, S. M. Polarization and place-based representation in us state legislatures. *Legislative Stud. Q.* <https://doi.org/10.1111/lsq.12441> (2023).
- Fraga, B. L., Velez, R., Yamil & West, E. A. Reversion to the mean, or their version of the dream? latino voting in an age of populism. *Am. Polit. Sci. Rev.* 1–9, <https://doi.org/10.33774/apsa-2023-764r1> (2024).
- Rastogi, A. & Jones-Correa, M. Not just white soccer moms: Voting in suburbia in the 2016 and 2020 elections. *RSF: The Russell Sage Foundation J. Soc. Sci.* **9**, 184–203, <https://doi.org/10.7758/RSF.2023.9.2.08> (2023).
- Grossman, G., Kim, S., Rexer, J. M. & Thirumurthy, H. Political partisanship influences behavioral responses to governors’ recommendations for covid-19 prevention in the united states. *Proc. Natl. Acad. Sci.* **117**, 24144–24153, <https://doi.org/10.1073/pnas.2007835117> (2020).
- Kitchens, K., Harris, S. & Miller, K. What matters in school reopening plans: an analysis of the impact of school board demographics. *Polit. Groups, Identities* **12**, 186–216, <https://doi.org/10.1080/21565503.2023.2224765> (2024).
- Wang, B. S., Rodnyansky, S., Boarnet, M. G. & Comandon, A. Measuring the impact of covid-19 policies on local commute traffic: Evidence from mobile data in northern california. *Travel. Behav. Soc.* **34**, 100660, <https://doi.org/10.1016/j.tbs.2023.100660> (2024).
- Beck, B., Antonelli, J. & LaScala-Gruenewald, A. Neck-restraint bans, law enforcement officer unions, and police killings. *Criminol. & Public Policy* <https://doi.org/10.1111/1745-9133.12658> (2024).
- Denes, M., Scanlon, M. & Schulz, F. Disclosure in democracy. *SSRN* <https://doi.org/10.2139/ssrn.4154777> (2022).
- Hughes, S., Kirchhoff, C. J., Conedera, K. & Friedman, M. The municipal drinking water database. *PLOS Water* **2**, e0000081, <https://doi.org/10.1371/journal.pwat.0000081> (2023).
- Amos, B. & McDonald, M. P. A method to audit the assignment of registered voters to districts and precincts. *Polit. Analysis* **28**, 356–371, <https://doi.org/10.1017/pan.2019.44> (2020).
- Amos, B., Gerontakis, S. & McDonald, M. *Voting and election science team: 2016 precinct-level election results* <https://doi.org/10.7910/DVN/NH5S2I> (2024).
- Amos, B., Gerontakis, S. & McDonald, M. *Voting and election science team: 2018 precinct-level election results* <https://doi.org/10.7910/DVN/UBKYRU> (2024).
- Amos, B., Gerontakis, S. & McDonald, M. *Voting and election science team: 2020 precinct-level election results* <https://doi.org/10.7910/DVN/K7760H> (2024).

Acknowledgements

We thank our funding supporters: the Alfred P. Sloan Foundation, the Houston Endowment, Resilient Democracy, and individual donors to the University of Florida Foundation's Election Science Group account. Research assistants who assisted with data collection include Maxwell Clarke, Robert Della Salle, Karl Klarner, Sara Loving, Evan Smith, and Mario Villegas. Michal Migurski independently provided some data assistance. We would like to thank numerous state and local officials who kindly responded to our requests.

Author contributions

S.G. was primarily responsible for data collection and processing. B.A. was primarily responsible for some data collection and processing, and voter file geocoding. M.M. was primarily responsible for project management and fundraising. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com