A Method to Audit the Assignment of Registered Voters to Districts and Precincts

Brian Amos, Visiting Assistant Professor, University of North Florida
Michael P. McDonald, Associate Professor, University of Florida

Abstract

Electoral boundaries are an integral part of election administration. District boundaries delineate which legislative elections voters are eligible to participate in and precinct boundaries identify, in many localities, where voters cast in-person ballots on Election Day. Election officials are tasked with resolving a tremendously large number of intersections of registered voters with overlapping electoral boundaries. Any large scale data project is susceptible to errors, and this task is no exception. In two recent close elections these errors were consequential to the election outcome. To address this problem, we describe a method to audit the assignment of registered voters to districts and precincts. We apply the methodology to Florida’s voter registration file to identify thousands of registered voters assigned to the wrong state house district, many of which local election officials have verified and rectified. We discuss how election officials can best use this technique to detect registered voters assigned to the wrong electoral boundary.

Prepared for the 2018 Election Science, Reform, and Administration Conference held at the University of Wisconsin; Madison, Wisconsin. July 27-28, 2018
Introduction

Nearly all United States legislators, at all levels of government, are elected to office by the people residing in a defined geographic area known as a district. Periodically, district boundaries are redrawn through a political process known as redistricting. Precinct boundaries, which determine which polling place many voters will cast an Election Day ballot, are drawn through an administrative process known as reprecincting, which can sometimes have a political component, too. Redistricting politics are intense, and we do not wish to rehash the extensive scholarship. Suffice to say that epic political battles are waged to determine exactly where district boundaries should lie. We are interested here in the nuts and bolts election administration task of assigning voters to electoral boundaries - districts and precincts - so that they can participate in elections for the offices they are intended to vote in. Making sure voters are given the correct ballot seems like it should be easy, but two recent problematic elections demonstrate administrative errors can happen.

In June 2018, dozens of voters in Habersham County, Georgia received a letter from their county’s Office of Elections and Registration informing them that they had been assigned to the incorrect State House district due to “a past voting precinct redistricting issue.” This mistake was not an inconsequential clerical error. In May, the House district in question had held its primary election for the 2018 cycle, and the Republican race was decided by just 67 votes. The losing candidate has challenged the results, and a judge may overturn the election and force a re-vote based on the mistake.

---

1 We use the term “precincts” to refer to what the Census Bureau called “Voting Tabulating Districts” or VTDs, 2 https://nowhabersham.com/some-habersham-county-voters-were-assigned-to-wrong-state-house-districts/
In the November 2017 elections in Virginia, at least 384 registered voters in the northeastern part of the state were assigned to incorrect State House of Delegates districts, of whom at least 147 cast a ballot. 3 125 of these voters who turned out were incorrectly assigned to Senate District 28, a number larger than the Republican candidate’s 82-vote margin of victory. Despite the error, a federal judge declined to overturn the results, and an appeal to the ruling failed. 4

From a naive point of view, these errors should not happen. Election officials could feasibly visit registered voters’ addresses, determine which side of an electoral boundary they reside, and assign the voter to the corresponding district. In theory, this approach would produce an error-free assignment of voters to districts. In practice, election officials work with data-driven representations of a jurisdiction’s geography to manage the scale of the task. This introduces two types of error. First, a particular map may be inconsistent with the real world: for instance, the map could have errors, like placing an address at the wrong location, or it could be outdated, like not including a new residential subdivision. Second, two different maps used in the process may be inconsistent with each other, which could lead to the same point of interest falling into different districts depending on which map is being referenced. In addition, there is always the possibility of clerical error.

We refer to this phenomenon as administrative redistricting, whereby election officials effectively move registered voters into a legislative district different than the one they are assigned to by law. The issue tends to receive attention only in close elections where losing candidates have the resources to detect its presence. To shed light on the scope of the problem,

---

3 http://townhall.virginia.gov/L/GetFile.cfm?File=C:\TownHall\docroot\meeting\151\26912\Agenda_SBE_new_v1.pdf
we develop a methodology to detect administrative redistricting by geocoding voter registration addresses and overlaying district boundaries onto them. The methodology can be used to identify voters assigned to any wrong electoral administrative unit, including precincts. When we apply the methodology in Florida, we find thousands of registered voters assigned to the wrong state legislative district. We have communicated these instances to local election officials, who have verified and corrected the assignment errors. We conclude with a discussion of how we can improve the integrity of elections by auditing the assignment of registered voters to electoral boundaries.

Defining Electoral Boundaries

As mentioned, we do not wish to dwell on the political process of redistricting and reprecincting; we wish instead to focus now on election administration procedures of assigning registered voters to the districts and precincts they are eligible to vote within. Electoral boundaries are defined in law by two common methods. Historically, boundaries are defined by what are known as *metes and bounds*. Metes and bounds describe electoral boundaries by geographic and political features, such as roads, water, city boundaries, and so on. The 2012 New York state Assembly districts are described in typical metes and bounds language:

1. First district. In the county of Suffolk, that part of the town of Brookhaven bounded by a line described as follows: Beginning at the intersection of Edwards Avenue South and the Brookhaven/South Hampton town line, thence southwesterly along said line to a line extended easterly from the East Moriches census designated place to the Brookhaven/South Hampton town line, thence westerly along said line...

The transition from metes and bounds to maps started in 1890, when the Census Bureau began delineating and reporting decennial population statistics for selected small geographic

---

areas in New York City. Over time, the Census Bureau extended coverage by mapping smaller geographies across a greater range of the country. By 1990, the Census Bureau had mapped the entire country down to what are commonly known as census blocks, which are geographic units similar to city street blocks in urban areas, but may follow other geographic features, such as road medians, streams, railroads, and so on. Census cartographers may also use political boundaries for census block boundaries, working with states and localities to collect and verify the location of these political boundaries. Census blocks can be geographically small, and a typical medium-sized state can have hundreds of thousands of census blocks.

As the Census Bureau canvassed the nation to create localized cartography, some states abandoned metes and bounds in favor of defining their electoral boundaries by census geography. Doing so ensures that legislative districts meet legal equal population requirements, since the Census Bureau reports decennial census total population statistics within their defined geographic units. An electoral boundary defined in census geography will list the census cartographic units assigned to each district, such as counties, census places, tracts, block groups, blocks, and so on. These legal definitions typically take the form of a list of unique identifiers for each piece of census geography assigned to a district.

Metes and bounds may coincide with census geography, even if census geography is not explicitly mentioned. The New York Assembly metes and bounds uses some census geographic units, such as the “East Moriches census designated place” (census places are a part of the census geographic hierarchy). Indeed, it is likely that the roads, town boundaries, and other features used in New York’s metes and bounds coincide with census geography since Census Bureau cartographers use these features to define census geography.

---

Moving from metes and bounds to census geography can cause ambiguity, however. Census blocks are one geographic type among many that the Census Bureau releases, and these different geographies are interconnected: although a census block will not be explicitly defined by the streets that make up its boundary as would occur in a metes-and-bounds system, its boundaries may perfectly coincide with streets found in the Census Bureau’s road network geography release. This raises a question: what happens when the Census Bureau’s road network map is inaccurate? This question is not a hypothetical one. The Census periodically releases updates to its street network geography; most of the differences from year to year will be the addition of newly built roadways, but existing streets may see corrections in their locations, and these corrections occasionally have substantive implications for election administration by virtually moving the location of resident addresses.

The Census Bureau releases updated census blocks along with its updated road networks each year, and blocks whose boundaries follow roads that have seen corrections will be adjusted accordingly. Likewise, the Census Bureau releases updated political district boundaries when corrections are made for states that define theirs based on census geography. Perhaps in the Census’s eyes, these changes do not represent an actual shift in boundary lines or the movement of voters from one district to another, as their underlying definition of a particular census block is based abstractly on geographic features in a metes-and-bounds sort of manner. It is unclear, though, whether a state that defines its districts based on census geography is doing so based solely on the current census maps at the time of redistricting, or if it is instead defining the districts based on the implicit ties these maps have to the real world that the Census Bureau is using internally. The latter interpretation perhaps eases the need for perfect maps and allows for corrections, but the lack of explicit, public definitions of census block boundaries, and therefore
district boundaries, is troubling. Furthermore, the interpretation makes the use of non-census-based mapping tools problematic, as the system becomes unmoored from a basic geographic coordinate system; an update to census geography could move a particular latitude and longitude reading to a new district, even if the claim is that no voter has actually been moved.

Philosophical and legal thorniness aside, the previous paragraphs illustrate one potential source of error due to inconsistent maps. In a real example we identified, a county used an outdated Census Bureau map as the reference for its district boundaries, but an up-to-date map to geocode voter addresses and place them within those districts. As a result, a portion of a neighborhood whose streets had seen corrections in later Census releases were assigned to the incorrect districts in the voter registration file.

Assigning Voters to Districts and Precincts

When a registered voter requests a ballot, election officials need to provide a correct ballot listing all of the elections the voter is eligible to participate in. Creating all the possible permutations of ballots, what are known as ballot styles, and verifying their accuracy can be a labor-intensive task where the myriad of congressional, state legislative, judicial, and local government districts overlap. Large localities may have hundreds of ballot styles. Ensuring each voter is given the correct ballot style for where they live is subsequently another labor-intensive task.

To manage the assignment of registered voters to districts, most jurisdictions append to each individual voter registration record identifiers for each district a voter resides in. There are two procedures by which election officials make these district assignments. The first procedure involves election officials creating street address ranges associated with each district. For
example, 100 to 199 Main Street would be associated to district one and any voter registration address within the range is similarly assigned to district one. Typically, a master address database consists of all known street segments falling within city blocks, with separate records for the even and odd numbered sides of streets. Identifying even and odd sides of streets is necessary since election boundaries often use streets as borders. For example, the even 100 to 198 Main Street might be assigned to district one, while the odd numbered 101 to 199 Main Street might be assigned to district two. The even and odd addresses in the next 200 to 299 Main Street block would be similarly assigned to the appropriate districts, and so on.

The advent of the use of detailed census geography to define electoral boundaries can pose problems for some addresses that election officials assign using street segments. Buildings often do not directly abut streets, but instead have a “setback gap”: for example, in a suburban neighborhood, residents usually have a front lawn separating their house from the street their address is located on. In rural areas, the actual location of a voter’s home may be quite far from the road defining their voting address. In a real example we identify, an apartment complex was set back far enough from its road that a district boundary ran within the gap, causing voters to be assigned to the wrong districts when the election official used the street segment to identify the complex’s district rather than its true physical location.

Another method to assign registered voters to districts involves geocoding. Election officials may use geocoding procedures to pinpoint the latitude and longitude of addresses, and then overlay district boundaries onto these pinpoints to assign registered voters to districts. Geocoding algorithms generally work by corresponding an address to be geocoded with a database that contains latitude and longitude coordinates for addresses, roads, and political and administrative units. The Census Bureau creates an extensive system of geographic data in the
course of their cartography work, some of which is specifically designed for use in geocoding, and commercial vendors may supplement Census Bureau data with other databases, such as local assessor data. When a geocoder encounters an unrecognized address, it will use heuristics to produce an approximate latitude and longitude. A typical first pass heuristic is to infer coordinates from a street address. For example, if a geocoder does not recognize 150 Main Street as a valid address in its database, it will interpolate a location situated halfway between the even side of the end point coordinates of 100 and 198 Main Street. A first attempt may fail if a geocoder does not recognize a street name, or a street number does not fail within an existing street range. A geocoder will then attempt to infer what it can from higher geographic levels, such as zip codes and city names, and locate an address at the center of these geographic entities.

Obviously, as the geocoder filters through the heuristic process, the coordinates produced become less precise. The highest level of confidence is “rooftop” accuracy, when the address is contained in the geocoder’s database and the confirmed latitude and longitude of the building at that address is available. If a geocoder must infer latitude and longitude by interpolating between end coordinates on a street, however, two sources of error may arise. First, 150 Main Street may not fall directly in the middle of 100 and 198 Main Street, despite the suggestion of the numbers. Second, the geocoder must guess at the setback gap. Most geocoders by default will apply a standard, non-zero setback to all addresses so that points will not fall directly on roads, but underestimating this gap can lead to situations like the apartment complex mentioned above. If a road does not run perpendicular to an address’ plot of land, a setback gap can produce coordinates that are located in a neighbor’s yard. If the geocoder cannot match the street address at all and must rely solely on city or zip code geography, the potential error is quite high. These
issues are more prevalent in rural areas, which tend to have more long winding roads and wide variation in setbacks.

Neither street segments nor geocoding is a fool-proof method of assigning registered voters to districts. When election officials encounter a problematic voter registration application, they place voter registration applications into a suspense status while they conduct additional research. Most commonly with respect to addresses, voter registration applications are placed into a suspense status when a street address is unrecognized, as may occur with new construction, an error by an applicant, or illegible handwriting. Some addresses might not commonly be recognized as a valid street address, such as a rural address that is merely a description of the property, or an urban address that identifies a park or alley where a homeless person resides. Indian reservations can be particularly problematic where residents’ street addresses are associated with their mailbox, located miles from their physical home. These vagaries illustrate a virtue of localized election administration, as local residents know their neighborhoods the best.

Auditing Methodology

Our methodology to detect when registered voters are assigned to the wrong district is similar to the geocoding methodology used by some election officials to assign their registered voters to districts, described above. We geocode voter registration addresses, overlay district boundaries, and identify instances where a voter registration district identifier is not the same as the overlaid district. An innovation that distinguishes our approach from election officials is that we use four geocoding databases in sequence, two based on direct address matching, two based on interpolating street addresses. We automate our detection methodology using ERSI’s
ArcGIS software, which makes it possible to implement the procedure on any voter file with relatively minimal effort.

Our approach works well as a check on the street segment approach since we can independently verify district assignments using an alternative methodology. Our methodology has detected voters assigned to incorrect districts even in localities that use a geocoding system to assigned voters to districts. We attribute the latter to the use of different geocoding databases. Indeed, in one instance we conferred with a local election official who verified their geocoder had misplaced an apartment complex.

We expect different performance from different companies’ geocoding databases because these are generally proprietary collections that companies have spent considerable effort to develop. We infer from documentation that a typical geocoding database starts with the Census Bureau’s Topologically Integrated Geographic Encoding and Referencing (TIGER) products.\(^7\) The Census Bureau has invested extensive resources into mapping the country to aid in conducting its decennial census and its sampled surveys, such as the American Community Survey. Keeping these data current is a tremendous challenge, given that new houses are continually being built and some are torn down. Companies supplement the TIGER databases with local tax assessor and surveyor maps, the Post Office’s National Change of Address files, information from consumer credit monitoring agencies, and so on. Companies may use different databases, different algorithms to merge records, and have different timing on when they collect data. This leads to companies producing different geocoding databases.

We successfully geocode more voter registration addresses by using multiple geocoding databases layered on top of one another. We further observe that we geocode more records by

layering legacy geocoding databases from the same company. We attribute this phenomenon to how street addresses occasionally change, such as when a street is renamed or when street numbers are renumbered. Voter registration addresses become out of sync with commercial vendors’ geocoding databases when voter registration databases are not updated as quickly as commercial vendors’ geocoding databases. A legacy geocoding database may thus geocode addresses the most current database will not. It is also possible to observe the reverse, where a voter registration database has recent addresses that do not exist in a geocoding database.

Election officials occasionally receive voter registration application from a new address, which they duly investigate, approve, and enter into their databases as appropriate. Election officials have told us of driving to unrecognized addresses to verify a house exists. Commercial vendors may detect the new address at a later time when new data become available to them.

At the end of the geocoding process, matched points are graded based on which database was able to locate the address and the confidence the database had in the match. As described above, rooftop accuracy represents the highest level of confidence, and serves as our first cut into exploring errors. We supplement these with street address interpolation matches, especially in areas that have been identified as having problems using the rooftop matches. Layering these databases is valuable. By itself, ESRI’s rooftop-accuracy-level database has delivered coordinates for between 70% and 85% of addresses in the states we have tested. The address interpolation database released publicly by the Census Bureau (the “Address Range Feature” shapefile) has identified about 90% of addresses, but with the caveat of lower confidence in the
level of accuracy. In combination, though, our rooftop and street address interpolation databases regularly provide us matches for about 98% of registrants in our databases.⁸

Since these geocoded addresses are now represented as points on a map defined by latitude and longitude, we are able to match them with the districts that we believe they should be assigned to. We do so by applying a spatial join on the geocodes to a district map, which is simply an operation that reports where two different geographies coincide. Which map represents the most accurate version of the districts may vary by jurisdiction, but with congressional and state legislative districts, we generally use the most up-to-date boundaries provided by the Census Bureau. After the join is completed, we can compare the district assignment found in the voter registration file to the assignment that we find through our geocoding process, and can further investigate cases where these assignments do not align.

Administrative Redistricting

When governments create district boundary lines through redistricting, they do so with the intent of having residents vote in races based on location of their home addresses. The actual process of assigning registered voters to districts is more complicated, and when election officials assign voters to the wrong district, we call this administrative redistricting. Election officials may not intentionally make errors, but their effect is that some voters are given the wrong ballot, which may have dire consequences for the integrity of a particularly close election.

We have presented election officials with some of the suspected errors identified by our methodology, and they have validated and rectified them. In the course of our communications, we identify a number of ways by which administrative redistricting happens. Some involve

---

⁸ Due to the large uncertainty, we choose not to include databases that work in lower levels of accuracy in our geocoding hierarchy, such as zip code matching.
human error, some involve over-reliance on geocoding algorithms, some involve obscure
updates to census geography, and some involve district lines. Our list may not be exhaustive,
since there may be other causal factors we have yet to encounter. We obscure the source of the
following examples, since election officials have been generously responsive to our
communications, and we mutually seek to improve election integrity.

Human operator error occurs in various guises. Human error may occur when an election
official assigns a street segment to the wrong district. These errors are easily detected by our
methodology, appearing as a line of geocoded addresses along a street block flagged as assigned
to the incorrect district. An example of a street segment error is presented in Figure 1. The red
line represents a district boundary. The dots represent geocoded addresses that are colored by the
district they are assigned to in the voter registration file. The black line represents a single road
that crosses the two districts. In the western portion of the road, houses with even street numbers
(yellow) are correctly assigned, but houses with odd street numbers (blue) are not.

Figure 1. Example of Street Segment Error
In another case of human operator error, during the course of a local redistricting, registered voters within a single precinct split by the new local districts somehow had their state House district changed to match the new local district lines. Figure 2 presents this precinct before the city commission redistricting (left) and after (right). The red line represents a state legislative boundary, and dots represent geocoded addresses that are colored by the state legislative district they are assigned to in the voter file. Though the state district line did not change, the assignments did; the changes follow those made in the city commission districts. The assignment error persisted across several elections from the time of the local redistricting until we alerted local election officials of the error.

Realignment of roads or local political boundary lines may create ambiguities in district boundaries. If a district border is described in metes and bounds, a diversion of that road can
technically cause a change in district assignment if the move is around a residential area. Not as obvious, such a division may also create ambiguities for census geography since roads often form census block boundaries. In a personal communication with a state’s geographer, we learned that the particular state’s policy is to shift the census block boundary, thereby affecting any individuals residing in the affected area. Similar issues may arise from political annexations or other realignments of political boundaries.

Figure 3. Example of Census Geography Changes Resulting in Assignment Error

Census Bureau cartographers continually update census geography to account for changes to physical and political boundaries, such as the realignment of roads and political boundaries. They may also address issues that are endemic to cartography, such as how well the
virtual maps and physical terrain correspond. Geographic updates can result in administrative redistricting if election boundaries are not updated to reflect changes to the cartography, while geocoding databases are updated. These out-of-sync geographies can inadvertently shift registered voters near a district border out of their assigned district. We present such a scenario in Figure 3. We represent a state legislative boundary as defined by census block geography in the 2013 TIGER data (red) and 2017 TIGER data (black). The changes to the district boundary lines are due to corrections to census geography, rather than actual changes to the district boundaries. The county in question uses a geocoder that accurately places voters on a map, but is still using the 2013 district boundaries for district assignment purposes; dots on the map represent registered voters assigned to the wrong district.

Finally, there are the odd corner cases that are not systematic, but make for interesting stories. In one case, a family won a court decision to affiliate themselves with a neighboring county even though their home is plainly not located in that county. In another case, we discovered district boundaries in a state were wrong, but not in the voter registration file: the state’s localities had successfully sued the state to unsplit all precincts divided by state legislative districts, but this information was never transmitted to the state officials responsible for generating electronic district maps.

Disparate Effects

There appears to be a bias in who is affected by administrative redistricting, in that urban voters are disproportionately misassigned in the states we have examined. This is perhaps to be expected, for at least a couple reasons. First, the density of district lines scales with the density of population. A rural county may completely reside in a single district at the congressional and the
state legislative levels, while a more urban county may deal with a dozen or more districts across levels of government; the latter county has many more opportunities for errors. In states with no compactness requirements on the redistricting process, the complexity can multiply. Second, the greater population density of urban areas means that small errors can affect more voters than in rural areas, such as the misassignment of a single street segment.

This disparity in likelihood of urban and rural voters to be misassigned can lead to disparate effects by party registration and race, since these two groups of voters have different general demographics. In our search for misassignments in Florida State House, our baseline was a statewide registration rate of 38% Democrat and 36% Republican, and 16% Hispanic and 13% Non-Hispanic Black. Among the 1,208 voters we identified with rooftop accuracy as being assigned to the wrong House district, the distribution was 46% Democrat and 30% Republican, and 16% Hispanic and 23% Non-Hispanic Black.

Recommendations

It is clear that voters are occasionally assigned to incorrect districts, and that these errors have led to real consequences in close elections, most recently in Georgia and Virginia. The problem is much more extensive than those two examples, though; we and the Washington Post have detected thousands more registered voters assigned to the wrong districts in just Florida and Virginia. Fortunately, none of these additional assignment errors beyond the highly publicized ones have occurred elections close enough to changed election outcomes elsewhere, but there is no reason to believe this luck will continue to hold.

---

We have conducted our district assignment audits from an academic perspective, and have engaged with election officials as white-hat outsiders who wish to improve election integrity. Election officials have been graciously responsive to our communications, and have rectified the errors they have verified. However, we have carried out these efforts with no funding, which is an unsustainable model for continuous auditing. With this in mind, we make the following recommendations.

We recommend our methodology be integrated into existing election management systems. Indeed, we spoke with a vendor who has developed such a reporting feature. However, such systems may be insufficient to detect all errors, since geocoding databases vary in their accuracy and coverage. We thus recommend district assignment audit systems use multiple geocoding databases to flag potential issues that may be missed by a single system.

We recommend that election officials integrate into their election preparations an audit of district assignments. We have observed election officials with the capability of auditing district assignments fail to use the tools available to them. Routinizing and automating district assignment audits can mitigate avoidable problems.

We have conducted our analyses on congressional and state legislative district assignments because the electronic representations of these district boundaries are readily accessible from the Census Bureau. We cannot as easily verify that registered voters are assigned to the correct election districts for other offices, particularly local elections. Collecting these boundaries can be a monumental task, involving individual contact with local election offices across a state. Most states do not have statewide electronic versions of their precinct boundaries, similarly making it difficult for us to verify registered voters are assigned to the correct precinct. A failure to audit precinct assignments can have detrimental consequences for voters who vote in
the wrong precinct, particularly among states that require voters cast ballots in the correct precinct.

We recommend better data management of electoral boundaries, both for districts at all levels of government and for precinct boundaries. Geographic information systems can be deployed on the internet such that state and local governments can work in the same environment as they draw and redraw election boundaries.\(^{10}\) A centralized system will ensure that all of a state’s election boundaries are housed in a single repository, which will enable district assignment audits for all election boundaries. A byproduct of developing a centralized mapping system is cost savings, since a single application can serve an entire state, whereas vendors often currently sell licenses to each locality.

The good news is that we can audit district and precinct assignments. Technological innovations have progressed such that it is possible to develop and deploy of these recommendations, and some vendors have already deployed such systems. We can develop mapping applications that will centralize election boundary collection for districts and precincts. The result will be better election data integrity that will improve voters’ experiences and reduce the costs of elections.

---

\(^{10}\) We offer DistrictBuilder as one such mapping application, available at [www.districtbuilder.org](http://www.districtbuilder.org).