Georgia voter roll ID numbers

Preliminary Report

Andrew Paquette, PhD

11/2/2024

Introduction

This investigation of Georgia's voter rolls was prompted by research I have conducted in other states that has found:

New York: An estimated 2 million illegal "clone" records, along with four unusually complex and well-hidden algorithms used in ID assignment. These algorithms can predict voter status, identify clones, reveal deleted SIDs, and add hidden attributes to records (Paquette 2023).

New Jersey: An encoded identification system that transforms and reverses ID numbers, potentially allowing covert record identification (Paquette, in press).

Arizona: Approximately 500,000 clone records, and two ID number assignment algorithms. The first is similar to an algorithm found in NY, and the second appears to be identical to one of NY's 4 algorithms.

Pennsylvania: ID numbers grouped by last digit prior to mapping to state ID creates added data channels for potentially hidden attributes and record tracking.

Ohio and Texas: Hidden attributes in voter records enable covert tracking in populous counties.

Hawaii[i]: A tagging mechanism on UUID numbers segregated ~10% of records, which have since been deleted.

These findings suggest the possibility of hidden attributes in voter roll data fields, particularly in unique identifiers like State ID (SID), County ID (CID), and Legacy ID (LID) numbers.

A fundamental rule of database management is that all data should be transparent, traceable, and used only for its intended purpose. The algorithms found in various state databases violate this rule by introducing what amounts to undocumented attributes into the database. This makes it untraceable by normal means and can enable manipulations that violate the intended purpose of the databases.

This analysis is based on a version of Georgia's voter rolls dated October 9, 2024.

This preliminary report seeks to identify:

1. Patterns in ID number assignments that could encode additional information through:
    - Algorithmic segregation of number ranges
    - Systematic categorization
    - Predictable sequences
2. Whether such patterns, if found, go beyond standard ID assignment methods
3. Irregular records in sufficient quantities to justify covert tracking

Note: While all ID systems use algorithms, this analysis focuses on detecting unusually complex methods that could be used to embed or organize information within the ID structure itself.

While time constraints prevent a full solution of any algorithms found (unlike in NY), their presence and capabilities can be demonstrated without complete reversal.

Initial results reveal enough potential cloned records in Georgia's current database to justify the use of an ID number indexing algorithm. All of Georgia's 159 counties employ a complex algorithm to assign Registration ID (RID) numbers.

Data Sources and Processing

Database Files

Data was obtained from a copy of the Georgia State Board of Elections voter registration database, containing 7,219,904 records as of 10/8/2024. Of Georgia's 159 counties, detailed analysis focused on the four most populous: Cobb, DeKalb, Fulton, and Gwinnett Counties. Approximately 60 additional counties were examined in less detail, while statewide patterns were analyzed across all counties regardless of individual county examination level.

Clone Records

Clone/Duplicate distinction

Duplicates are records identical in all fields. The "Original" is the first record in any matching group, while "Duplicates" are additional identical records to be deleted.

Cloned records, like biological clones, can differ from their original yet share core identifying traits. While clones may vary in many fields, they share enough personal identifying information (PII) to strongly indicate they represent the same person. Each clone has its own voter ID number, allowing it to function independently in the voting system. Under HAVA Section 303(a)(1)(A), each voter should have only one "unique identifier" in the state system. Having multiple voter IDs for the same person creates illegal multiple registrations that can be used independently, unlike harmless duplicate records.

Legal Context

New York law establishes a specific method to prevent the creation of duplicate records: registration applications must be checked against existing records using first name, last name, and date of birth. When these match, further verification using driver's license or last four SSN digits is required. If one of these also match, processing a new registration with a different voter ID would violate federal and state law. While this matching protocol is designed to prevent duplicate records, it would also prevent clones. The presence of numerous clones in state databases indicates non-compliance with these requirements.

Clone Detection Methodology

Georgia's voter database provides birth year (not full birth date) for matching voter records. Three matching methods were used to identify clone registrations:

1. First Name + Last Name + Birth Year
2. First Name + Last Name + Middle Initial + Birth Year

Statistical Validation

Based on name distribution analysis from a comparable state database (Arizona, population 6,851,732), and given Georgia's similar voter roll size of 7,219,904, we can estimate approximately 715-720 false positives using Method 1 (Last+First+Birth Year), and 26-27 false positives using Method 2 (adding middle initial). These projections are scaled from Arizona's observed frequency of 459,773 unique last names and 185,011 unique first names across an 80-year span of birth years, with Georgia's slightly larger population suggesting marginally higher false positive rates.

Findings (Clone records)

The number of clone registrations found far exceeds statistical expectations. Against an expected 680 false positives for Method 1, we found 1,176,645 matches. Method 2, with an expected 25 false positives, found 691,109 matches. Of particular concern, 856,292 of the LF-BY matches (87.5%) and 153,077 of the LFI-BY matches (86.7%) are in currently active registration status, meaning both the original record and its associated clone(s) are simultaneously active. Even with conservative estimates of false positives (under 1,000 total across all methods), the impact on these findings is statistically negligible (Table 1).

Table 1 Clone counts by match method and status

GA	Active	Inactive	LF-BY Match	LF - Est Clone	LFI-BY Match	LFI - Est Clone	Duplicate	total
	6,272,100	947,804	979,119	489,560	176,518	88,259	0	7,219,904
Active			856,292	428,146	153,077	76,539	0	
Inactive			122,827	61,414	23,441	11,721	0	
Exclusive			802,865	401,433	264	132	0	

Analysis of registration dates shows expected spikes in presidential election years (shaded: 1992, 1996, 2000, 2004, 2008, 2012, 2016, 2020), with peak clone registrations reaching 44,785 in 2020 and 30,959 in 2016. The data reveals two distinct patterns: remarkable stability from pre-1990 through 2008 (maintaining 7.60-8.53%), followed by a steady decline from 2009 onwards, reaching 5.86% in 2024. While Arizona showed an increase in clone percentages from 1990 (14.75%) to a 2010 peak (24.17%) followed by sharp decline, Georgia's pattern reveals a more gradual, consistent decrease. Note: Clone counts and percentages represent half of the total matches found, accounting for the original record needed to generate each clone, thus avoiding double-counting of match pairs (Table 2).

Table 2 Clones by year of registration, Georgia

	<1990	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Clones	31,607	3,270	1,651	8,959	1,501	5,386	5,892	8,934	3,992	5,168	3,876	8,735
Total	394,287	38,822	19,735	107,032	17,587	64,044	71,563	109,418	50,672	65,010	51,806	112,583
Pct Clones	8.02%	8.42%	8.36%	8.37%	8.53%	8.41%	8.23%	8.16%	7.88%	7.95%	7.48%	7.76%
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Clones	3,875	5,645	5,726	13,482	6,415	8,391	8,515	18,631	5,082	7,533	7,482	16,815
Total	51,480	74,848	77,656	174,821	85,535	116,584	117,284	245,209	70,705	106,053	106,818	239,711
Pct Clones	7.53%	7.54%	7.37%	7.71%	7.50%	7.20%	7.26%	7.60%	7.19%	7.10%	7.00%	7.01%
	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Clones	7,796	12,395	10,538	30,959	23,834	35,005	30,246	44,785	17,490	32,159	25,682	22,116
Total	115,661	181,888	164,142	461,959	360,224	527,839	470,504	691,020	281,128	552,869	465,976	377,385
Pct Clones	6.74%	6.81%	6.42%	6.70%	6.62%	6.63%	6.43%	6.48%	6.22%	5.82%	5.51%	5.86%

A critical finding is that 87.5% of identified clone pairs (856,292 LF-BY matches) remain simultaneously active in voter rolls. Unlike typical duplicate registrations where one record is usually inactive, these pairs maintain concurrent active status. While deactivation could address some risk, the creation of these clones violates election law, and their maintenance in the system - even if deactivated - leaves unnecessary vulnerabilities due to the ease of status changes.

Algorithms

Georgia assigns unique voter Registration IDs (RIDs) consisting of 8 digits. Analysis reveals a complex system where counties are allocated multiple, overlapping ranges of RID numbers. Unlike states such as NY, PA, and AZ, which assign discrete ranges to counties, Georgia's system shows sophisticated patterns of range allocation and usage across its 159 counties. This complexity is particularly evident in the simultaneous use of multiple ranges within each county, and the way new ranges are systematically activated over time while maintaining activity in earlier ranges (Table 3).

Table 3 Georgia ID number overlap, by county (note that these are all 8 digit numbers with leading zeroes omitted)

CLARKE	HABERSHA	FULTON	BARROW	COBB	GWINNETT	FORSYTH	CHEROKEE	GILMER	DOUGHERT	GLYNN	CRISP
1,199	1,223	1,338	1,389	1,391	1,528	2,180	2,288	3,520	4,934	4,710	5,469
1,200	1,275	1,993	2,454	2,530	2,277	2,591	4,776	5,967	4,943	4,711	5,551
2,257	1,328	2,362	2,458	3,430	3,644	2,603	4,777	7,051	4,947	4,989	6,576
2,995	1,358	2,694	2,465	3,535	4,412	2,739	6,133	9,148	4,960	7,004	7,192
3,136	1,359	3,166	2,557	3,781	4,487	2,933	9,325	9,164	4,970	7,433	7,256
3,715	1,366	3,390	3,165	5,141	4,770	3,131	10,373	14,256	4,975	7,488	7,257
3,745	1,368	3,667	3,404	6,293	4,905	3,194	10,420	14,889	4,979	7,491	7,657
3,746	1,374	3,786	3,533	7,645	5,609	3,379	11,342	20,191	4,982	7,764	7,679
4,192	1,458	6,618	3,855	8,734	5,753	4,590	11,978	21,477	4,983	8,376	8,183
4,442	1,460	7,098	4,465	8,873	7,056	4,680	12,090	23,098	4,984	8,657	8,246
4,471	1,635	7,752	4,694	9,021	7,544	4,681	15,591	23,405	4,985	9,304	8,247

Preliminary results

Voter registration records typically show correlated progression of ID numbers and registration dates. For example, Fairfield County, OH maintained constant SID numbers until a system change (around CID 170,000), after which SID numbers increased steadily with CIDs (Figure 1).

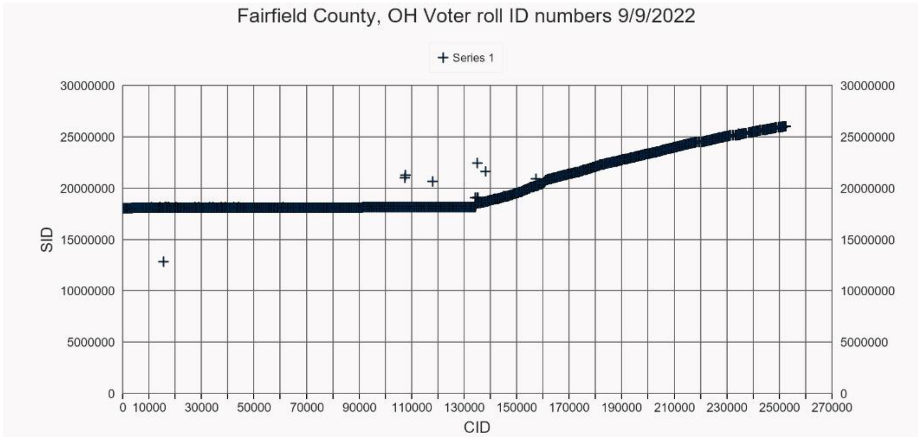

Figure 1 Fairfield County, Ohio, scatterplot CID and SID numbers show correlation over time

Scatterplots

Scatterplots of Georgia's most populous 4 counties reveal non-standard ID assignment patterns. Gwinnett County (Figure 2), representative of patterns found across other examined counties, shows a general ascending trend but contains several notable anomalies:

- Multiple concurrent ranges being actively used (evident in vertical stacking of points)
- Clear 'stairstep' progression into new ID ranges while maintaining activity in lower ranges
- Distinct boundary gaps (notably around 9-10M and 14-15M ranges)
- Sharp vertical jumps to new ranges (particularly visible in 15-18M range)
- Continuous activity in lower ranges (2-8M) even as higher ranges become active

These patterns, particularly the simultaneous use of multiple ranges and systematic progression to new ranges while maintaining old ones, suggest engineered complexity rather than natural registration progression or routine system updates.

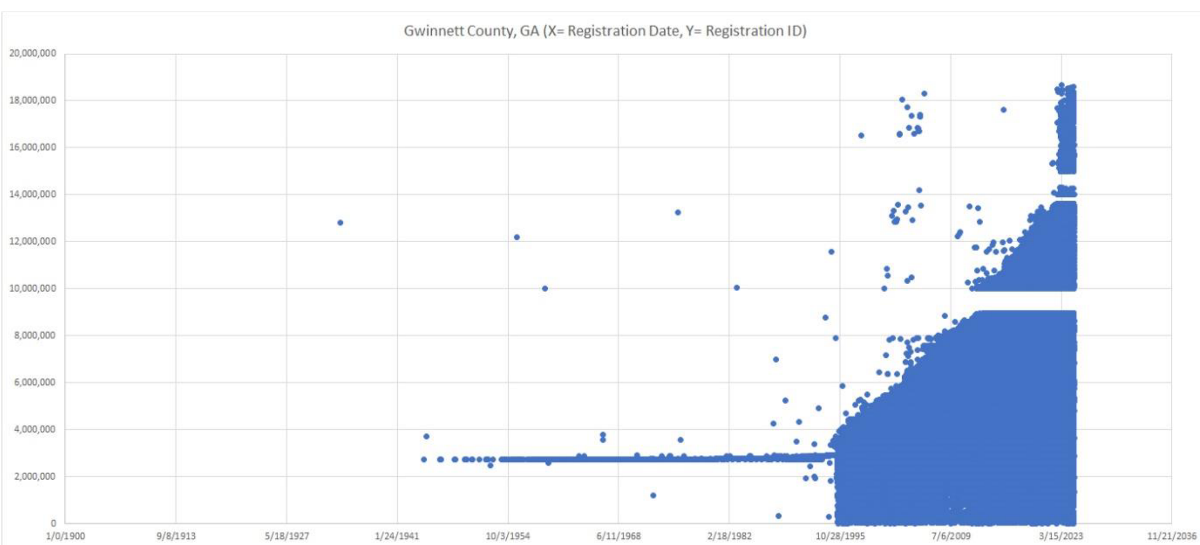

Figure 2 Gwinnett County, GA scatterplot (X: Reg Date, Y: RID)

Analysis of RID assignments by range and year (Figure 3) reveals a systematic pattern where the same number ranges are utilized for new registrations across all years. For example, 2016 registrations span ranges from 1-12M, similar to other years, rather than using a sequential block of numbers. The distribution of records within each range shows remarkable consistency - the 1-1M range varies only between 141-311 registrations per year, 2-3M between 753-2,109, and similar bounded variations in other ranges.

ID Range list	2016	2017	2018	2019	2020	2021	2022	2023	2024	Average
1-1000000	311	206	270	193	255	162	214	141	207	218
1000001-2000000	621	343	533	460	557	310	394	273	425	435
2000001-3000000	2,109	1,483	1,828	1,406	1,982	927	1,189	753	1,231	1,434
3000001-4000000	1,919	1,350	1,707	1,289	1,750	859	999	659	1,030	1,285
4000001-5000000	2,334	1,711	2,075	1,580	2,194	989	1,191	789	1,185	1,561
5000001-6000000	3,498	2,668	3,056	2,448	3,362	1,459	1,740	1,194	1,674	2,344
6000001-7000000	4,061	2,790	3,907	3,047	4,190	1,727	2,139	1,318	2,060	2,804
7000001-8000000	4,072	3,343	4,524	3,446	4,917	1,950	2,351	1,649	2,410	3,185
8000001-9000000	4,604	3,175	4,453	3,989	6,277	2,375	3,217	2,410	3,251	3,750
9000001-10000000	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10000001-11000000	30,796	2,894	4,553	4,128	6,396	2,653	3,699	2,773	4,155	6,894
11000001-12000000	13,425	27,419	34,674	3,589	5,904	1,978	3,167	2,664	3,769	10,732
12000001-13000000	5	1	5,544	33,298	40,823	1,771	2,883	2,227	3,459	10,001
13000001-14000000	1	0	8	0	3,862	18,189	41,059	2,843	3,064	7,670
14000001-15000000	0	0	0	0	1	0	9	438	41	54
15000001-16000000	1	0	0	0	1	0	14	13,499	744	1,584
16000001-17000000	0	0	0	0	0	0	0	7,002	301	811
17000001-18000000	0	0	0	0	0	0	2	6,715	7,105	1,536
18000001-19000000	0	0	0	0	0	0	0	39	4,507	505
19000001-20000000	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Total registrations/Year	67,757	47,383	67,132	58,873	82,471	35,349	64,267	47,386	40,618	

Figure 3 Overlapping ID ranges compared by date

Notable features include the complete absence of assignments in the 9-10M and 19-20M ranges, clear 'stairstep' progression through higher ranges (visible in the diagonal pattern from 11-14M across 2016-2022), and consistent use of lower ranges throughout. This constrained variation within ranges, combined with the dispersal of contemporaneous registrations across multiple ranges and precise avoidance of duplicate assignments, suggests a sophisticated algorithmic distribution controlling both range utilization and volume allocation.

A scatter plot of all Georgia registrations from 10/8/2024 (Figure 4) demonstrates how registration IDs are distributed across multiple ranges within a single day. The plot reveals systematic clustering at several distinct levels (notably around 12-13M, 15-17M, and 18M), with consistent spacing between clusters. These records, spread across 80 counties, show clear overlapping range usage - multiple counties receiving IDs from the same ranges simultaneously. The vertical stacking of points at similar ID ranges but different record counts illustrates how the same ranges are shared across different counties during concurrent registration activity.

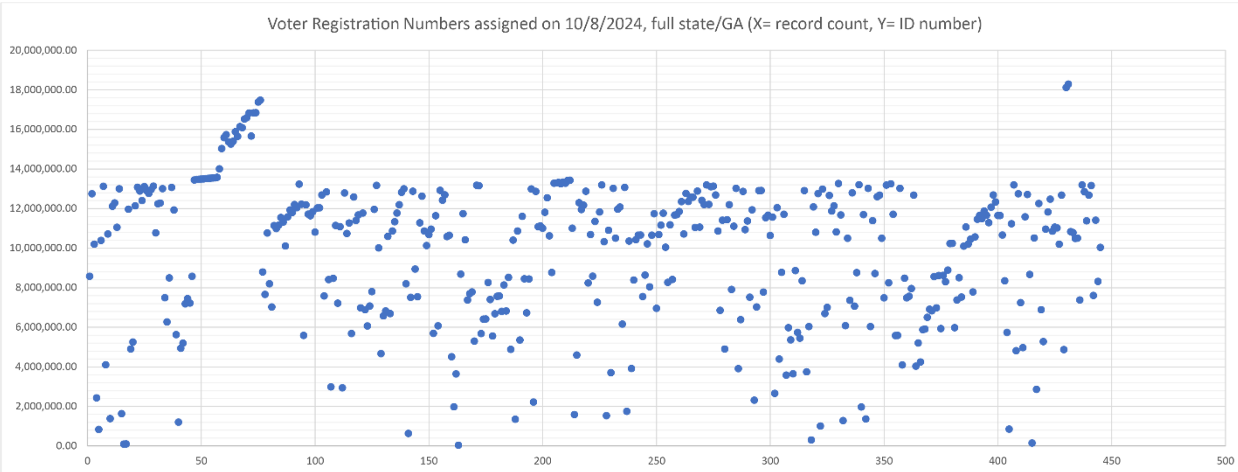

Figure 4 All registrations for the day 10/8/2024

Georgia's voter registration system employs a sophisticated algorithmic pattern that adds unnecessary complexity to what should be simple sequential ID assignments.

Voter registration IDs should be:

- Sequential or near-sequential
- Traceable to registration order
- Simple to audit
- Free from unnecessary complexity

Instead, the pattern maintains mathematical precision while obscuring simple relationships between records. The coordinated range allocation across counties, combined with consistent volume distribution and systematic progression through ranges, suggests centralized control of ID assignment rather than independent county-level implementation. This algorithmic sophistication actively works against database transparency and auditability, while the embedded patterns could enable systematic categorization or tracking through ID number placement within specific ranges.

Complex ID Assignment Systems and Hidden Records

Database ID systems typically use simple sequential numbering unless specific requirements demand more complexity. In voter registration databases, sequential ID numbers provide transparency and easy auditing.

The presence of an unnecessarily complex ID system suggests a need to covertly manage significant numbers of records. This relationship appears in practice: New York and Wisconsin, with an estimated 2 million and 500,000 illegal duplicate registrations respectively, both use complex ID systems. Georgia's data reveals a similar pattern - a sophisticated algorithmic ID distribution system correlating with approximately 489,560 clone registrations (6.78% of 7.22M records). This repeating relationship between complex ID systems and significant numbers of duplicate registrations suggests the systems' complexity serves a specific, if concerning, purpose.

Comments

While benign explanations are possible, Georgia's database practices significantly deviate from industry standards. Privacy and security cannot justify these complex ID systems - the National Voter Registration Act (1993) requires public access to all voter roll data. Any attempt to obscure or protect information through complex ID assignment violates these public disclosure requirements.

Database administration practices also fail to explain the observed patterns. While system evolution, administrative efficiency, backup systems, or multi-office processing might justify some complexity, they cannot account for:

- Sophisticated mathematical relationships between RID number ranges and registration dates
- Consistent pattern maintenance across all counties
- Violation of database best practices

The precision and complexity of the RID assignment algorithm suggests deliberate design rather than administrative convenience.

The presence of an estimated 489,560 cloned records represents an additional and unnecessary risk to election integrity in Georgia.

These findings suggest potentially problematic management of Georgia's voter roll records. The algorithm's use creates a hidden classification system for data segregation, posing a security risk. The high number of questionable records exacerbates this risk, as they could be targets for voter roll misuse - a concern recently realized when Wisconsin mailed absentee ballots to inactive voters.

Georgia should investigate:

- When and by whom the algorithm was introduced
- Its intended purpose
- Associated costs
- Prior awareness among officials
- The presence of clone records

Additionally, Georgia should consider removing all excess (clone) records and those incorrectly marked as active. Retaining unusable voting records serves no legitimate purpose. If preserving voter history is a concern, these records could be archived separately from the active rolls.

These findings suggest potential systemic issues in voter roll management and warrant further investigation.

References

Paquette, A. (2023). "The Caesar cipher and stacking the deck in New York State voter rolls " Journal of Information Warfare **22**(2): 86-105.

Paquette, A. (2024). "New Jersey voter ID numbers reconfigured with shift cipher." (In-Press).

[i] This was found by researcher Vico Bertogli, of Pennsylvania