California voter roll ID numbers

Preliminary Report

Andrew Paquette, PhD

11/24/2024

Introduction

This investigation of California's voter rolls was prompted by a request from the LAGOP. Their request 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).

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, Arizona, Georgia, 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 two versions of California's voter rolls dated November 10 and November 15, 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.

Analysis found that ID numbers in California's 28th district were assigned non-sequentially through an algorithmic process, an unnecessary complexity given the public nature of this database. In addition, 2,285 (out of 4,927) clone-candidate records were identified. Based on patterns from other states like New York, where in-county clones represented roughly 0.002% of cross-county clones, the actual number could be significantly higher when comparing across districts.

Data Sources and Processing

Database Files

Data source: LAGOP (Los Angeles Republican Party) provided 10 Excel files of voter data by city. Because it is limited to California's 28th District, it isn't possible to:

- Confirm if all clone pair members were found
- Apply methods used in other states to identify and analyze algorithmic patterns

This limitation exists because complete clone detection requires comparing records across all districts.

Initial Processing

Two sets of voter files were analyzed - one from 11/10/2024 with age data only (344,004 records), and an updated set from 11/15/2024 that included birth dates and registration dates (397,828 records). In addition to the large influx of new records (53,824), comparison revealed an anomaly: records sharing the same name consistently had the same age value but different birth dates. The statistical persistence of this pattern suggests deliberate record generation rather than random occurrence. This anomaly raised the importance of the clone analysis portion of this research above that in other reports, where ID number algorithms were more important than any detail about the clone records themselves.

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. In the District 28 data, there are no duplicates, identified as any 2 or more records with identical Voter ID numbers.

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 many state databases indicates non-compliance with these requirements.

Clone Detection Methodology

The following matching methods were used to identify clone registrations:

1. First Name + Last Name + Age (11/10/2024 DB)
2. First Name + Last Name + DOB (11/15/2024 DB)
3. First Name + Last Name + Phone Number (rare but highly reliable)

Statistical Validation

With 350,193 unique names and 47,635 duplicate full names in a total database of 397,828 records, duplicate names represent only 12% of records. The distribution pattern is revealing: from 43,633 unique first names and 80,346 unique last names, only 47,635 full name duplicates emerge. For a false positive clone match to occur, individuals must share the same full name (12% probability) plus identical age or birth date. Given the 93 birth years represented in both databases, the probability of matching on both name and age is approximately 0.13% (12% × 1/93). While a more precise calculation that takes into account the actual age distribution in the data—where certain age ranges are more common than others—yields a slightly higher probability of about 0.18% (12% × 0.0146), this difference does not materially affect the statistical conclusions. In a congressional district-sized population of ~398K records, it is therefore statistically improbable that a significant number of identified clone candidates are false positive matches.

Findings (Clone records)

Analysis identified the following clone patterns:

- 11/10/2024 Database (Name + Age matches): 3,010 records (1,478 unique, 1,532 clones)
- 11/15/2024 Database (Name + Age matches): 4,927 records (2,285 unique, 2,642 clones)
- 11/15/2024 Database (Name + DOB matches): 10 records (5 unique, 5 clones)

While these in-district clones represent less than 1% of total records, comparative data from other states suggests the actual number is likely much higher. For example, New York's Nassau County has 332 in-county clones but 169,054 total clone candidates when compared against the full state database. Even in Texas, where the ratio is lower, Harris County's 529 in-county clones expand to 14,466 candidates in cross-county comparisons. This suggests our district-only analysis in California significantly understates the total number of clones.

Notably, clone identification in other states used name plus birth date matching, not age. Applied to California's District 28, this method finds only 5 clones in 10 records, seemingly suggesting the higher numbers found using age matching (1,532 and 2,510 clones) are false positives. However, this interpretation is problematic for two reasons: first, it implies a volume of false positives far exceeding statistical probability; second, it ignores the systematic pattern of records sharing identical ages but having different birth dates within the same or adjacent years, indicating deliberate manipulation rather than random occurrence.

Anomalous birth dates

Nassau County, NY Comparison

To validate the statistical significance of age matching patterns observed in District 28, a control analysis was performed using the Nassau County, NY voter database. Despite being more than twice the size (979,518 records vs 397,828), Nassau County shows a pattern consistent with natural demographic duplication: 167 name/age match combinations involving 334 records, with nearly all of these (166 combinations, 332 records) sharing birthdates (Table 1). This represents the expected pattern - when records share the same name and age, they almost always share a birthdate because they likely represent the same individual.

Table 1 District 28/Nassau County clone comparison

	Records	Names	Name/Age combinations	Ages	Dup Name/Age Combinations	Dup Records	Dup Name/DOB Records	Dup Name/DOB Names
District 28	397,828	349,347	395,186	93	2,285	4,927	10	5
Nassau	979,518	945,799	979,351	106	167	334	166	332

This contrasts sharply with District 28's data. Despite its smaller size, District 28 contains 2,285 name/age match combinations involving 4,927 duplicate records - nearly 15 times more combinations and 15 times more records than the larger control database. Yet only 5 pairs (10 records) have matching DOBs, compared to the 99% DOB match rate in Nassau. This inversion of the expected relationship between ages and birthdates, combined with the systematic variation of birthdates within age-matched groups and the statistically impossible frequency of age matches, provides strong evidence that these patterns do not reflect natural demographic occurrence.

The statistical significance of this difference is overwhelming. Nassau's rate of name/age combinations (0.017% of records) represents an expected baseline for demographic duplication in in-county comparisons. District 28's rate (0.574% of records) is 34 times higher - a deviation representing over 100 standard deviations from the expected rate ($p < 10^{-23}$). The near-perfect correlation between age and DOB matches in Nassau (99.4%) versus District 28 (0.2%) further confirms these patterns cannot be explained by random variation or normal demographic distribution. Notably, this analysis represents only in-county comparisons; in other states, cross-county comparisons typically reveal substantially more matches – The most extreme example is New York's Nassau County, whose 332 in-county matches expand to 169,054 when compared against the full state database.

Methodology

The baseline probability of two random records sharing the same age is 1/93 (1.075%), reflecting the 93 different ages found in the District 28 database. Statistical validation was performed using Monte Carlo simulation, a computational method that tests probability by running repeated random trials. An AI-assisted analysis generated 100,000 simulations for each group size (2-10 records) to determine how often age matches would occur by chance. This number of iterations provides stable probability estimates: at 100,000 trials, events with probabilities as small as 0.001% can be detected with 95% confidence, sufficient for analyzing observed patterns where even the most common matches (pairs) should occur in only 1.075% of cases. The simulation accounts for all possible match patterns - both complete matches where all records share an age and partial matches where only some records align. This approach establishes expected match frequencies for comparison against the patterns found in District 28's data. For example, simulation of 100,000 groups of 4 records yields probabilities for 2, 3, or all 4 records matching, providing a comprehensive baseline for evaluating the actual data.

Numerical First Name Group

The District 28 data contained a significant but small outlier group, identified by the first name "1". There are five surnames with this first name: Levy (59 records), Hanson (92), Mason (104), Harrison (121), and Liang (298). These five names, comprising 674 total records, account for 583 age-matched records in systematically varied group sizes (Table 2). The patterns include multiple instances of large matches - from Liang's four groups of 10 matching records to groups of 8, 7, and 6 across the names - far exceeding the typical 2-3 matches seen elsewhere in the data. Of the total records in this group, only 91 (13.5%) are unmatched singles, with the remaining 583 (86.5%) appearing in age-matched groups. While these records are analyzed separately due to their obviously erroneous first name and unusually high record counts, they display the same fundamental pattern observed in the main dataset: multiple age matches with systematically varied birthdates.

Table 2 "1" first name group

FullName	Count FullName Age	Count FullName	10	9	8	7	6	5	4	3	2	Subtotal (Dups only)	1	Total
1LEVY	35	59						1		3	14	42	17	59
1HANSON	40	92				1	2	1	2	7	12	77	15	92
1MASON	55	104						1	3	10	16	79	25	104
1HARRISON	60	121					1	1	1	13	23	100	21	121
1LIANG	72	298	4	2	3	4	10	4	10	11	11	285	13	298
			4	2	3	5	13	8	16	44	76	583	91	674

No Remainder Match Groups

The strongest statistical evidence comes from groups where all records sharing a name also share the same age - "no remainder" matches (Table 3). Analysis identified 379 Name/Age combinations forming complete pairs (758 records), 7 combinations forming triplets (21 records), and 4 combinations forming quadruplets (16 records), totaling 795 records in complete match groups. Within the population of records sharing names, random chance would predict only 2.7 complete pairs and 0.03 triplets, making the observed numbers statistically impossible (>100σ from expected). Even when compared against the full database of 397,828 records, where natural demographic patterns would predict approximately 297 age matches of any type, finding 795 records in complete match groups represents a significant anomaly. This pattern cannot be explained by chance whether examined within the matched-name population or the broader database context.

Table 3 No remainder group

	Match 2	Match 3	Match 4	Total
Name/Age combinations	379	7	4	
Records Involved	758	21	16	795
Expected by chance	2.7	0.03	~0.00	

Partial Match Group

While partial matches account for more total records (3,405) than complete matches (795), their statistical significance is lower due to the higher probability of partial matches occurring naturally. For example, finding exactly 2 matching ages in a group of 3 records has a 3.2% probability by chance, compared to only 0.0108% for a complete match of 2 records. Despite this higher natural probability, the observed patterns still far exceed random chance - 192 groups of size 3 with 2 matches were found where only 6 would be expected, and 150 groups of size 4 with 2 matches where only 9 would be expected. However, these deviations from expected values (32x and 17x respectively) are less extreme than the complete match patterns which showed deviations of over 140x from expected values. This suggests that while both patterns are statistically impossible by chance, the complete matches provide even stronger evidence of systematic manipulation than the partial matches.

A Monte Carlo simulation was conducted on the full database of 397,828 records to determine how closely random chance could replicate the observed patterns. Given the database size and 93 possible ages, random occurrence would predict approximately 31 complete pairs and less than 1 complete triplet across all records. The actual data shows 379 complete pairs and 7 complete triplets, along with 192 partial match triplets. This represents a deviation of over 140 standard deviations from the expected rate, with a probability so small (approximately 1 in 10^10,000) that it exceeds conventional statistical frameworks. Even accounting for natural demographic patterns that might cause age clustering, the magnitude of this deviation conclusively demonstrates that these match patterns cannot be explained by random occurrence.

Significance of non-Matching DOB

The discovery that only 5 name/age matches (10 records total) share identical dates of birth provides a critical statistical contradiction. In a random dataset, we expect and find variety in birthdates - this is natural. However, when we've already identified records with matching names and ages (an event with 1/93 probability), the birthdates for these matches tell an important story. For a given age (say, 54), the birthdates could span two different years depending on when the list was generated and each person's birth month. For example, if the list was generated in June 2024, people aged 54 could be born between June 1969 and June 1970. This natural variation makes the consistent pattern we see - matched ages with differing birthdates - even more striking. Having found 4,927 records in age-matched groups that have already beaten the 1/93 probability, we would expect a significant proportion of these matches to also share birthdates. Instead, we see systematic variation in birthdates within these already improbable age-matched groups, with only 10 records (0.24%) sharing birthdates. This pattern appears across all frequency groups (2-10) and suggests deliberate assignment of different birthdates within age-matched groups rather than natural occurrence.

Possible explanation

The statistical analysis of this dataset reveals several striking patterns: an extremely high rate of age matches within name groups, systematic variation in birthdates within these age-matched groups, and only 5 pairs of matching DOBs among thousands of records. Let's evaluate three possible explanations for these patterns.

- Natural Occurrence: The hypothesis that this represents natural data can be conclusively rejected based on statistical probability. In the smallest groups (pairs), we observed 379 complete matches where only 31 would be expected by chance - a deviation of over 140 standard deviations. This improbability becomes even more extreme in larger groups, where we see multiple instances of 8, 9, and 10 records sharing ages against probabilities of less than 1 in 10^4487. These numbers transcend what we could reasonably attribute to coincidence or demographic patterns.
- Innocent Error: The possibility that these records represent the same individuals with erroneously recorded birthdates fails to explain the systematic nature of the variations we observe. If birthdates were being incorrectly recorded due to clerical errors, system issues, or data entry problems, we would expect to see patterns typical of such errors: missing values, default dates, repeated incorrect values, or random variations. Instead, we see carefully distributed dates within appropriate year ranges that consistently yield the matching ages.
- Deliberate Manipulation: The hypothesis of intentional manipulation best fits the observed patterns. The combination of statistically impossible age matches (from 379 complete pairs to multiple groups of 10), near-zero DOB matches (only 5 pairs among thousands of records), and systematically varied birthdates within age-matched groups suggests deliberate action. The careful distribution of unique birthdates within year ranges that maintain the matched ages indicates an intentional process to create distinct records while preserving age alignment.

The statistical evidence strongly suggests algorithmic control of age assignments in the District 28 data. Natural demographic patterns, as demonstrated in the Nassau County control data, show age matches only when birthdates match. Random error would produce sporadic matches with inconsistent patterns. Instead, the District 28 data shows precise mathematical regularity: matching ages with systematically varied birthdates, and highly consistent group sizes (predominantly 2-3 matches, with larger groups showing identical patterns at decreasing frequencies).

These patterns - systematic variation within strict parameters, mathematical regularity in match frequencies, and consistency across different name groups - are hallmarks of algorithmic behavior rather than natural occurrence or random error.

Purpose of manipulation

Distinct voter ID numbers serve a necessary administrative purpose: preventing the automatic merger of duplicate records and ensuring each registration can function independently in the system. The systematic variation of birthdates while maintaining age matches suggests a sophisticated understanding of how duplicate-prevention systems operate. Most voter registration systems employ a multi-step matching protocol that begins with name and date of birth comparisons before proceeding to additional identifier checks. By varying birthdates while preserving ages, these records would bypass the initial matching threshold that triggers deeper verification - they would appear to be different individuals who happen to share a name rather than duplicate registrations requiring investigation.

The preservation of matching ages represents a particularly sophisticated choice that reveals additional intent. Creating completely random birthdates would have been simpler and equally effective at avoiding duplicate detection. Instead, we observe careful maintenance of age alignment across related records, requiring more complex data manipulation and introducing an unnecessary constraint. This pattern effectively creates a hidden indexing system within the data: related records can be easily identified by their matching ages while appearing unrelated to standard database operations. The age matches function as a form of steganography - hiding information in plain sight by making it appear to be random noise while actually serving as a systematic marking mechanism.

The scale and consistency of the pattern, combined with unique ID assignments and the systematic grouping of records (predominantly in pairs and triplets with consistent patterns up through groups of 10), suggests an organized system for managing these records. The "1" name groups follow the exact same age-matching pattern despite their obviously artificial nature, confirming these patterns resulted from deliberate algorithmic manipulation rather than either natural occurrence or random error.

Additional Explanations and Impact

While these patterns strongly suggest deliberate manipulation, it's important to consider plausible innocent explanations. Data migration or system conversion issues could theoretically create anomalies, but such problems would produce random variations rather than systematic patterns. Privacy protection measures would be applied uniformly across the database, not to specific groups of records. Test or training data would typically be segregated and clearly marked, not integrated with real records while maintaining sophisticated matching patterns that avoid detection.

The significance of these findings extends beyond the raw numbers. While 4,927 duplicate records might seem modest in a database of nearly 400,000 records, this perspective overlooks several critical factors. First, the Nassau County control data suggests the true scope may be much larger - their 332 in-county matches expanded to 169,054 when compared against the full state database. Even Texas, with lower match rates, showed a similar pattern: Harris County's 529 in-county matches expanded to 14,466 in cross-county comparisons. Second, close elections are often decided by margins of a few thousand or even hundred votes, making even a small number of systematically managed duplicate records potentially significant.

Most importantly, the identification of this pattern in one district raises broader concerns. The careful design represents a sophisticated understanding of database systems and duplicate-prevention protocols. If this pattern exists in one district, similar patterns could exist elsewhere but remain undetected precisely because they were designed to evade standard duplicate-detection methods. The presence of structured patterns at elevated frequencies in the "1" name group suggests these are not isolated anomalies but part of a systematic approach to record manipulation.

Registration Dates

New registrations with old dates

The second significant anomaly identified in the CA-28 voter data concerns the sudden appearance of 60,376 voter records between November 10-15, 2024, during the vote counting period following the November 5 election. While the voter roll showed net growth of 53,824 voters during this period (from 344,004 to 397,828), the pattern within the 60,376 new additions raises particular concern.

These records show registration dates distributed across more than a century, from 1900 to present - a pattern previously identified in Arizona's voter rolls by O'Donnell in preliminary research (personal communication, 2024), who documented 57,146 similar records added in a single month with historically distributed registration dates. The timing of these additions during active vote counting in a still-uncalled race, combined with their historical date distribution, suggests systematic rather than organic voter registration activity.

In typical voter roll updates, new registrations generally show recent dates clustered around voter registration deadlines. The appearance of thousands of records with registration dates spanning decades, added as a single batch during the counting period, deviates significantly from this expected pattern (Figure 1).

Analysis of the registration dates from these 60,376 newly-added records reveals two notable patterns. First, their temporal distribution appears perfectly natural for a voter database: 7 registrations from the 1900s, minimal activity through the 1970s, increasing volume through the 1980s (3,467) and 1990s (7,459), peaking in the 2000s (10,699) and 2010s (21,679), with 16,521 from 2020 forward. However, when registration dates are plotted against voter ages, a more complete picture emerges. The resulting scatterplot shows the expected triangular distribution pattern – older voters' registrations spanning from first eligibility to present day, while younger voters cluster in recent years – creating a mathematically coherent relationship between age and registration date (Figure 2). The sophistication of this distribution, appearing overnight in records added during vote counting, suggests careful attention to creating a statistically plausible voter population rather than the random or clustered patterns typically seen in administrative errors or system migrations.

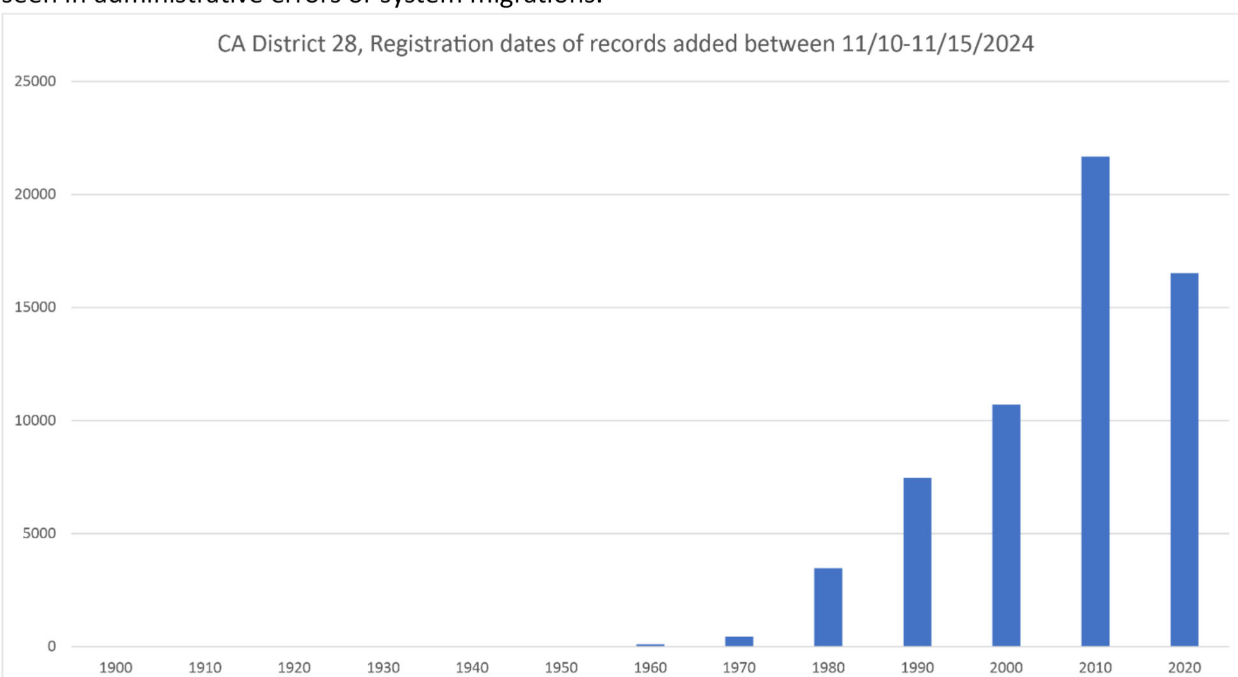


Figure 1 Registrations added between 11/10-11/15/2024 and their registration dates, by decade

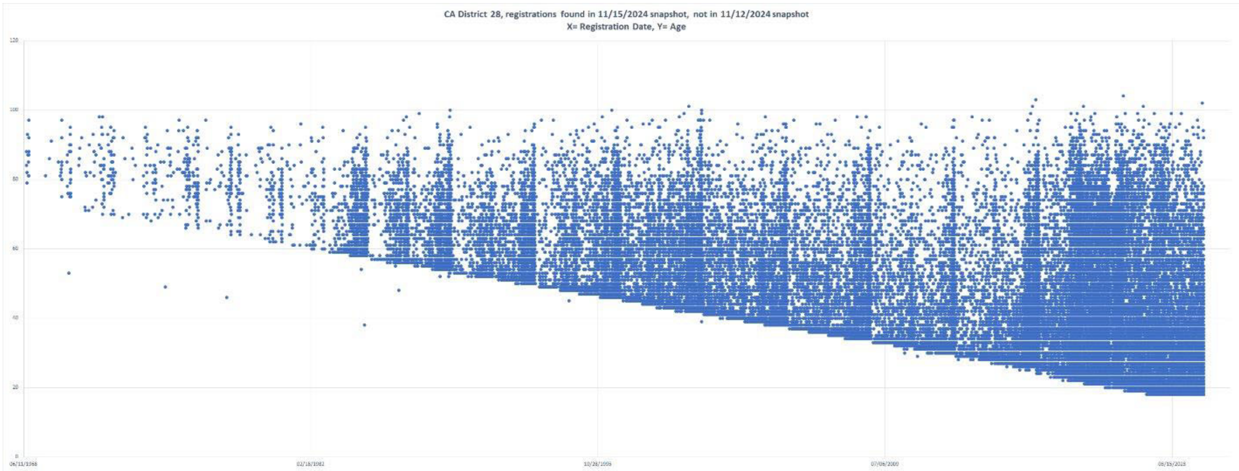


Figure 2 Scatterplot, registration date and age values, recently added records

The algorithmic patterns identified in the clone analysis find a striking parallel in these registration date distributions. Where the clone records showed careful calibration of birthdates to maintain matching ages while avoiding duplicate detection, these 60,376 new records demonstrate similarly sophisticated data engineering. Both patterns reveal attention to maintaining statistical plausibility - the clone records through systematic birthdate variation within age-matched groups, and these registration records through a mathematically precise distribution of ages and registration dates that perfectly mimics natural voter roll patterns. In both cases, the anomaly lies not in the data's structure, which appears statistically normal, but in its sudden appearance: thousands of age-matched records with systematically varied birthdates, and tens of thousands of registration records with a pristine demographic distribution, all materializing in the same five-day period during vote counting. The precision required to generate both patterns suggests not just deliberate manipulation, but sophisticated understanding of both voter database architecture and demographic distributions.

Several innocent explanations were considered in light of California's voter registration systems and processes. The state's VoteCal system, operational since 2016, processes registrations electronically in real-time and maintains daily synchronization across all counties. This modern, continuously-updated infrastructure makes it implausible that legitimate historical records would suddenly materialize during vote counting. While system maintenance and administrative corrections do occur, these processes are designed specifically to prevent the type of large-scale, post-election voter roll changes observed here. Moreover, none of these routine processes explains how legitimate voter records spanning 124 years of registration dates could appear after election day - these registrations, whether from 1900 or 2024, would have been required to exist in the system before votes could be cast. The sophisticated demographic distribution of these records, combined with their timing and the state's own system specifications, further strengthens the case that these additions represent deliberate manipulation rather than any routine administrative process.

Last Voter Activity Date

The LastVoterActivity (LVA) field in California's voter database tracks the most recent interaction between a voter and the election system. Under normal database operations, any update that changes a registration date (RD) would also update the LVA date, making it impossible for a legitimate record to have an LVA earlier than its RD. When LVA equals RD (zero difference), this typically indicates a voter who has never interacted with the system since their initial registration.

Of the 60,376 records that appeared after the November 5, 2024 election, 15,510 (25.7%) show identical registration and last activity dates - many from years ago when they should have been purged under federal and state voter list maintenance requirements. Even more significantly, 34,892 records (57.8%) show last activity dates on or before August 31, 2022, meaning they missed two federal elections before appearing in these rolls. More striking still, 19,650 records (32.5%) show no activity since August 31, 2020, missing three federal elections. Under federal and state voter list maintenance laws, these inactive records should have been removed from the rolls years ago, not appearing as new additions during vote counting. This pattern, combined with 872 records showing the impossible scenario of activity dates preceding registration dates, suggests these records were artificially generated without regard for mandatory voter list maintenance requirements that would have eliminated tens of thousands of these inactive registrations.

Algorithms

The District 28 records use an 8-digit "VoterKey" field for unique identification, allowing up to 99,999,999 values. In standard voter registration systems, ID numbers increase sequentially over time - newer registrations receive higher numbers than older ones. This sequential assignment serves both administrative efficiency and database integrity, providing clear chronological tracking of registration activity.

Scatterplots

When visualizing normal voter databases, ID numbers and registration dates show a distinctive diagonal pattern, as seen in Fairfield County, OH (Figure 3). This represents the expected correlation: newer registrations get higher ID numbers, with only occasional breaks due to documented system changes.

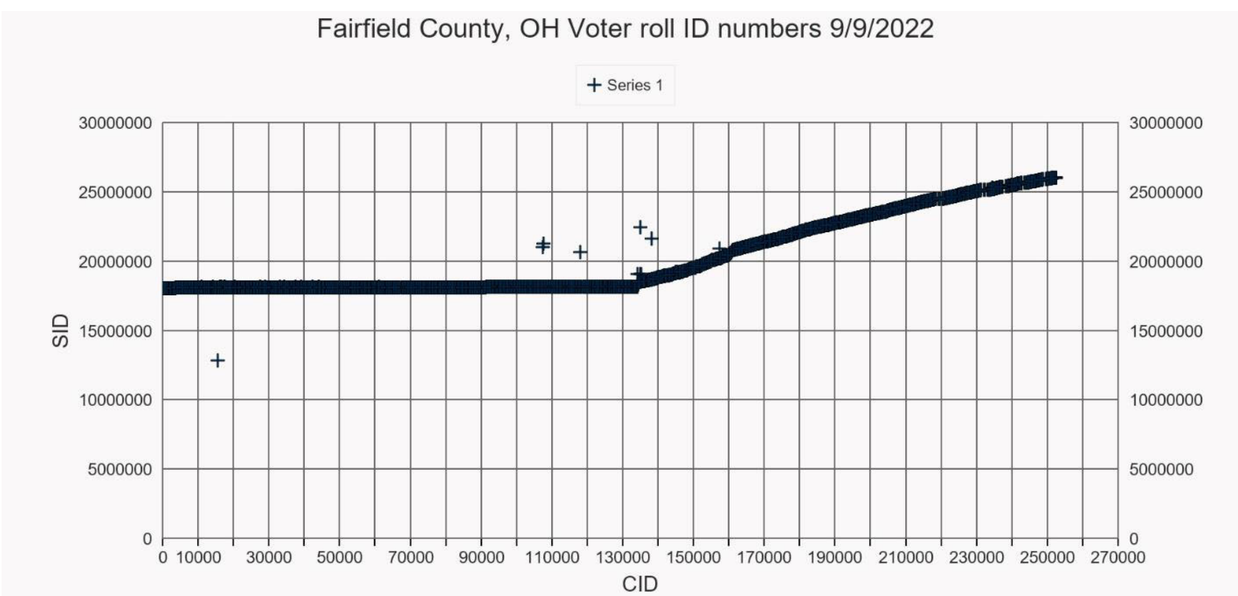


Figure 3 Fairfield County, OH voter ID numbers

District 28's data sharply deviates from this pattern. Analysis reveals that:

1. On any given registration date, voter IDs span the entire range from near zero to around 35 million (Figure 4)

2. This non-sequential assignment serves no legitimate administrative purpose in a public database

3. The data shows clear horizontal stratification suggesting a classification system

4. Vertical striping patterns indicate batch generation of records

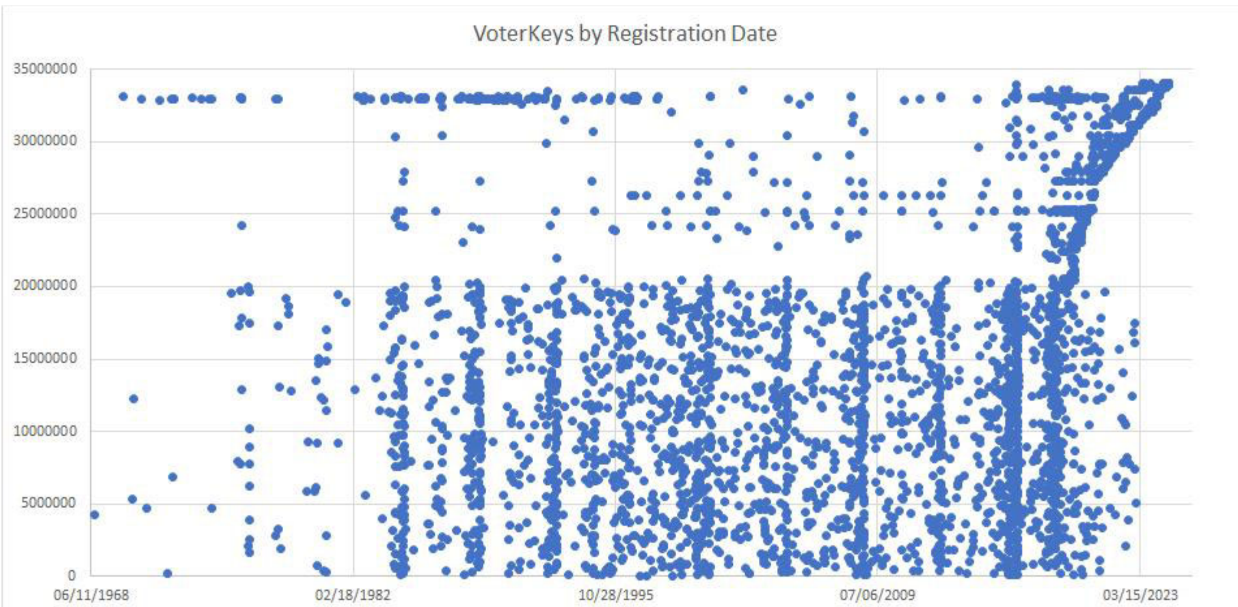
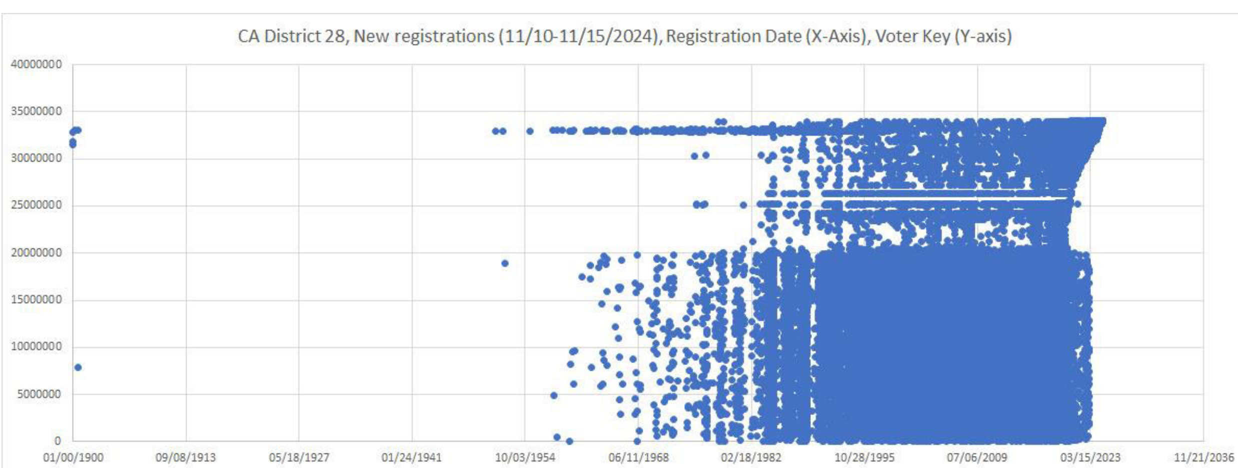


Figure 4 Voter ID numbers and registration dates

Most significantly, the 60,376 records added during vote counting precisely mirror this problematic distribution instead of clustering at the high end of the ID range as new records should. This matching pattern suggests algorithmic generation designed to blend with existing records rather than natural registration activity.



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, making anomalous records easier to detect. The implementation of an unnecessarily complex ID system therefore raises immediate concerns, as it requires additional development effort while reducing transparency - suggesting an intentional design choice rather than administrative convenience. This pattern has been observed in other states where large numbers of duplicate registrations were later discovered: New York and Wisconsin, with their complex ID systems, were found to have approximately 2 million and 500,000 illegal duplicate registrations respectively.

Non-voting active records

If a voter misses both the 2020 and 2022 federal general elections (marking two consecutive federal election cycles with no activity), they should be marked inactive and sent a notice after the 2022 election.

Of District 28's 397,828 records, 126,302 show no voting activity in any election from 2020-2024, including all three federal general elections (November 2020 November 2022, and November 2024). While 4,680 of these were registered after 2019, the remaining 121,622 records show no activity across all listed elections and were registered before 2020. These records should have been marked inactive after the November 2022 election, or purged, yet they remain on the rolls. This represents a clear deviation from statutory requirements for maintaining voter roll accuracy. The persistence of these records through three election cycles raises questions about list maintenance procedures.

Comments

While benign explanations are possible, California'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.

The anomalies identified in District 28's voter rolls display three key characteristics that eliminate innocent explanations:

1. Algorithmic Sophistication: The patterns show deliberate design choices that work against normal database functionality, including complex ID assignment systems and carefully calibrated demographic distributions.

2. Temporal Inconsistency: The sudden appearance of 60,376 records during vote counting, with registration dates spanning 124 years and activity dates that violate federal maintenance requirements, cannot be explained by routine administrative processes.

3. Statistical Improbability: The mathematical precision of demographic distributions and systematic variation of birthdates within age-matched groups exceeds what could occur through natural patterns or random error.

These findings suggest potential systemic issues in voter roll management that warrant further investigation, particularly given similar patterns identified in other states' databases.

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