

Arizona voter roll ID numbers

Preliminary Report

Andrew Paquette, PhD

10/26/2024

Introduction

This investigation of Arizona's voter rolls was prompted by a request from the Election Fairness Institute Inc. 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 and Texas: Hidden attributes in voter records enable covert tracking in populous counties.

Hawaii¹: 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 Arizona'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:
 - o Algorithmic segregation of number ranges
 - o Systematic categorization
 - o 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 a minimum of about 590,529 cloned records in Arizona's current database, a number sufficient to justify ID-tracking algorithms. All of Arizona's 15 counties employ a complex algorithm mapping VRAZ numbers to Registration ID (RID) numbers. Notably, Maricopa and Pima counties, which together account for 65.24% of all registrations, have variations not found in other counties.

Data Sources and Processing

Database Files

Data was obtained from the Arizona Secretary of State website (azsos.nextrequest.com) in 14 files:

- 2 PDF documents containing data documentation
- 12 voter registration record files, organized by:
 - Voter status (Active, Canceled, Inactive, Suspicious)
 - Party affiliation (Democrat/Republican/Other) for Maricopa County only
 - Geographic location: Maricopa County, Pima County, and two combined files for smaller counties (Apache-La Paz and Mohave-Yuma)

Initial Processing

All classifications embedded in file names were preserved in fields when consolidated into a single database. Approximately 500,000 records were found to have exact duplicates, resulting in over 1 million total records in this group. An SQL script identified and removed these duplicates, leaving only one copy of each unique record. After importing and processing for duplicate removal, 6,851,732 records remained for analysis.

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

Arizona'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
3. First Name + Last Name + Phone Number (rare but highly reliable)

Statistical Validation

Based on actual name distribution data from Arizona's full voter roll (population 6,851,732), statistical analysis predicts approximately 680 false positives using Method 1, 25 false positives using Method 2, and less than 1 false positive using Method 3. These estimates are derived from the observed frequency of 459,773 unique last names and 185,011 unique first names in the population, accounting for an 80-year span of birth years. Method 3, while most accurate, is limited by phone number availability in voter records.

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. Method 3, with less than 1 expected false positive, found 67,464 matches. The total unique matches across all methods (1,181,058) yields a minimum estimate of 590,529 clone registrations (8.62% of registered voters), after accounting for the original records needed to generate each match. 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

Match field	Total	Exclusive	Active	Pct Active
First and Last Name & Year of Birth	1,176,645	481,861	555,632	47.22%
First and Last Name, Middle Initial & Year of Birth	691,109	66	241,393	34.93%
First and Last Name & Phone Number	67,464	4,347	22,727	33.69%
Any 1 of 3	1,181,058	486,274	558,052	47.25%
Estimate excess ID numbers (half of total)	590,529			
Total records	6,851,732			

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 71,984 in 2008 and 93,356 in 2016. The data reveals two distinct patterns. A steady increase in both total volume and percentage of clone registrations from 1990 (14.75%) to 2012 (23.27%) A sharp decline after 2020, with clone percentages dropping from 11.17% to 7.62% in 2024

Pre-1990 registrations show 79,832 clones (15.20%). During 1990-2004, clone percentages remained stable at 15-18%. After Arizona implemented HAVA (Help America Vote Act) in 2004, clone percentages climbed steadily, peaking at 24.17% in 2010. This suggests HAVA's ID requirements paradoxically made it easier to create clone registrations (Table 2).

Table 2 Clones by year of registration, Arizona

	<1990	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Clones	79,832	12,106	5,827	22,550	4,483	9,819	7,386	19,477	6,515	11,070	8,565	25,191
Total	525,243	82,054	38,744	156,407	30,185	58,471	47,511	114,646	39,773	62,945	53,179	150,353
Pct Clones	15.20%	14.75%	15.04%	14.42%	14.85%	16.79%	15.55%	16.99%	16.38%	17.59%	16.11%	16.75%
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Clones	6,765	21,805	15,291	58,728	14,634	42,477	33,736	71,984	32,161	50,853	36,734	66,715
Total	38,788	120,029	88,193	320,610	72,418	197,201	154,979	331,433	145,435	210,434	161,002	286,695
Pct Clones	17.44%	18.17%	17.34%	18.32%	20.21%	21.54%	21.77%	21.72%	22.11%	24.17%	22.82%	23.27%
	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Clones	35,389	46,632	44,519	93,356	37,606	70,368	48,969	58,204	19,712	24,251	13,839	23,509
Total	161,175	202,146	210,354	445,829	202,027	379,345	292,122	521,277	197,476	265,014	179,795	308,440
Pct Clones	21.96%	23.07%	21.16%	20.94%	18.61%	18.55%	16.76%	11.17%	9.98%	9.15%	7.70%	7.62%

While only 47.25% of identified clone records remain active, their continued presence in voter rolls is problematic regardless of status. Even if Arizona authorities identified and deactivated these duplicates, their mere creation violates election law. The presence of over half a million active clone records (558,052), combined with evidence that many deactivated clones were previously active for multiple years, suggests systematic registration issues that cancellation alone doesn't address. Maintaining these records in the system, even when deactivated, creates unnecessary vulnerability since status changes require only simple database updates.

Algorithms

Arizona uses two voter ID formats: 8-digit Registration IDs (RID) and 10-digit VRAZ Voter IDs. While all records have RIDs, VRAZ IDs are only present in 3,611,351 of 4,868,993 total records.

VRAZ IDs begin with a two-digit county code, assigned alphabetically (01=Apache through 15=Yuma). However, Maricopa and Pima counties use "M" and "P" prefixes respectively, instead of their alphabetical codes (08 and 11), despite having sufficient number ranges available in the standard format. The number ranges unused by Maricopa and Pima, have not been assigned to any other counties (Table 3).

Table 3 Arizona RID and VRAZ ID county ranges, sorted by RID

County	County ID	RID MIN	RID MAX	Gap to next	Range	Alpha VRAZ	VRAZ at RID MIN	VRAZ at RID MAX	Gap to next	VRAZ Range	VRAZ Assigned
Maricopa	8	20,000,000	20,119,041		119,041	M	3,570,350	3,824,576		254,226	67,912
Pima	11	20,119,042	21,393,299	1	1,274,257	P	922,243	2,441,312	-2,902,333	1,519,069	795,089
Maricopa	8	21,393,300	24,983,426	1	3,590,126	M	3,463,164	5,111,834	1,021,852	1,648,670	2,547,141
Graham	5	24,983,427	25,023,352	1	39,925		500,016,919	500,040,007	494,905,085	23,088	37,824
Greenlee	6	25,023,353	25,034,252	1	10,899		600,001,849	600,010,799	99,961,842	8,950	10,267
Apache	1	25,034,255	25,129,781	3	95,526		100,071,637	100,095,207	-499,939,162	23,570	91,735
Gila	4	25,129,782	25,200,650	1	70,868		400,057,769	400,081,304	299,962,562	23,535	66,472
Cochise	2	25,200,652	25,350,563	2	149,911		200,104,575	200,155,336	-199,976,729	50,761	145,852
Santa Cruz	13	25,350,569	25,404,324	6	53,755		1,300,041,762	1,300,052,260	1,099,886,426	10,498	29,552
La Paz	7	25,404,325	25,427,127	1	22,802		700,000,313	700,021,634	-600,051,947	21,321	21,781
Yuma	15	25,427,128	25,600,824	1	173,696		1,500,183,470	1,500,243,574	800,161,836	60,104	95,541
Pinal	12	25,600,826	26,020,121	2	419,295		1,200,258,888	1,200,416,918	-299,984,686	158,030	251,051
Navajo	10	26,020,122	26,216,365	1	196,243		1,000,156,246	1,000,193,460	-200,260,672	37,214	142,828
Mohave	9	26,216,366	26,507,194	1	290,828		900,214,465	900,287,806	-99,978,995	73,341	194,174
Coconino	3	26,507,195	26,713,779	1	206,584		300,001,025	300,204,116	-600,286,781	203,091	187,813
Yavapai	14	26,713,781	27,040,677	2	326,896		1,400,227,087	1,400,325,446	1,100,022,971	98,359	183,961
Maricopa	8	27,040,839	28,315,219	162	1,274,380	NA	NONE	NONE	NONE	NONE	NONE
Mixed		28,315,221	28,560,797	2	245,576	NA	NONE	NONE	NONE	NONE	NONE

Counties are assigned sequential RID ranges with minimal gaps between them. Each county also has a dedicated VRAZ number range, identified by either numeric prefixes or 'M'/'P' for Maricopa and Pima. While both numbering systems continue to be maintained in parallel, records created after 2018 (RIDs above 27,040,839) are assigned only RID numbers, not VRAZ numbers.

The placement of unrestricted IDs after range-limited ones, and the out-of-sequence Maricopa and Pima ranges, obscures the county range structure. While database users may know about county-encoded VRAZ prefixes, this information isn't public. Since the algorithms operate on original ID ranges rather than current county listings, moved voters' records further mask these patterns.

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).

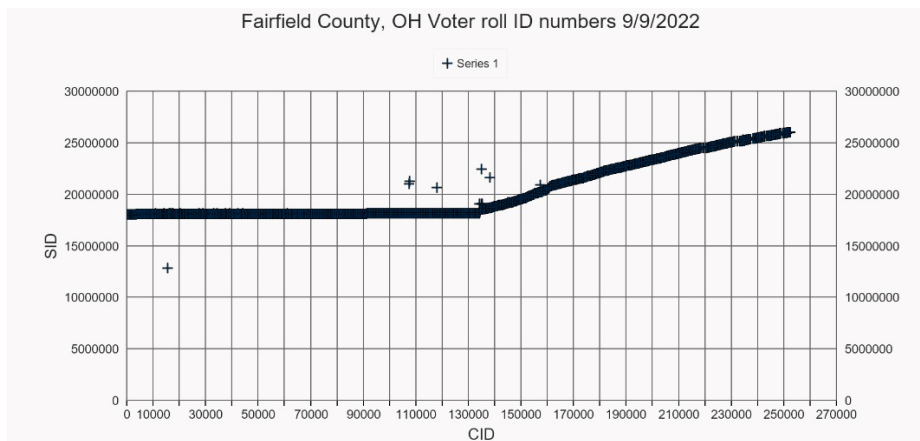


Figure 1 Fairfield County, Ohio, scatterplot CID and SID numbers show correlation over time

Scatterplots

Scatterplots of Arizona's 15 counties reveal non-standard ID assignment patterns. While Apache County (Figure 2) shows a general ascending trend similar to Fairfield County, OH, it contains several anomalies:

- A main diagonal line with regular progression
- A distinct vertical column of scattered points midway
- An isolated horizontal block of records
- An abrupt pattern change in the final quarter of records

These features suggest deliberate complexity rather than natural registration progression or system updates.

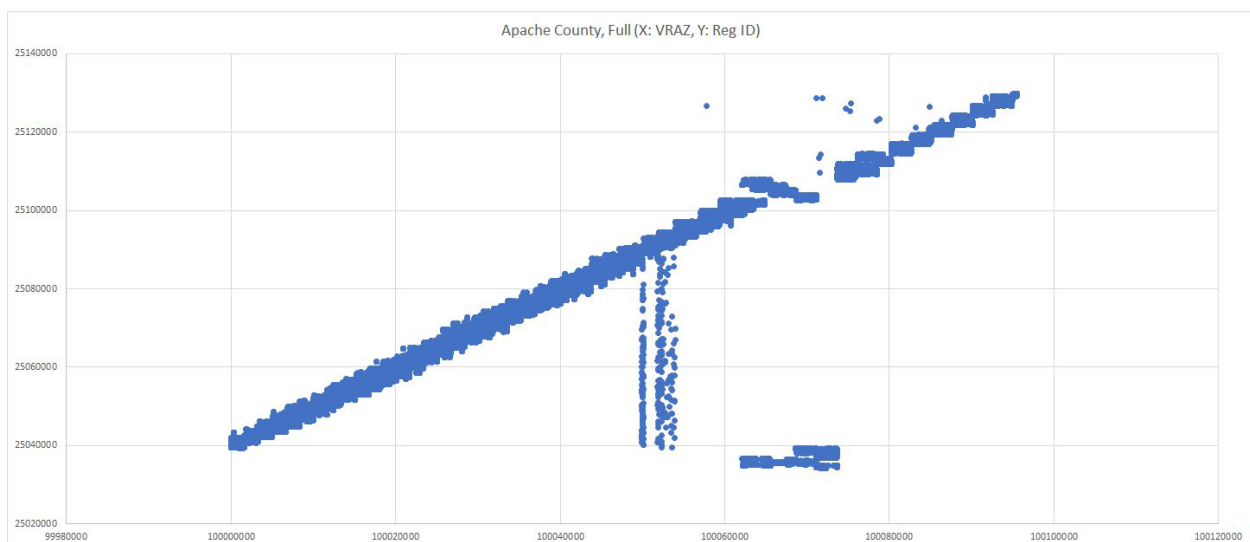


Figure 2 Apache County, AZ scatterplot (X: VRAZ, Y: RID)

A closeup of Apache County's scatterplot (Figure 3) reveals complex patterns hidden in the apparently linear progression. Rather than sequential assignment, the ID numbers form distinct clustered segments. Both VRAZ and Registration IDs show overlapping ranges, with multiple VRAZ numbers corresponding to similar RID ranges and vice versa. This structured dispersal appears intentional rather than random.

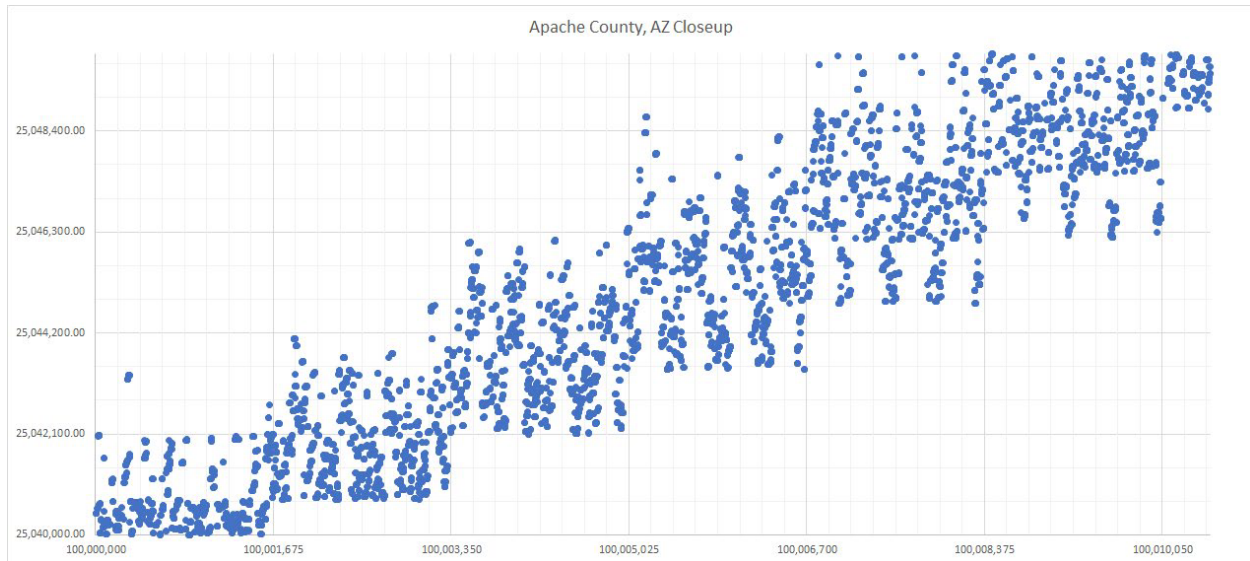


Figure 3 Apache County, AZ scatterplot closeup view, (X: VRAZ, Y: RID)

Random database artifacts or independent ID assignments would show irregular scattering or simple linear relationships. Instead, we see mathematically precise clustering with regular spacing and consistent segment sizes. These patterns repeat across different ID ranges and maintain their structure despite overlapping assignments. Such geometric regularity requires deliberate coordination between the two ID systems.

Unlike other counties with single ranges, Maricopa has three distinct RID ranges - appearing first, third, and seventeenth in the sequence of seventeen total ranges. Each range shows unique characteristics, warranting separate analysis.

The first range (Figure 4), shows structured patterns similar to Apache County, but truncated. The overlay grid reveals:

- Uniform block widths in VRAZ numbers (X-axis)
- Regular subdivisions: each block quartered in both dimensions
- Consistent RID spacing (Y-axis), despite a narrower range than VRAZ

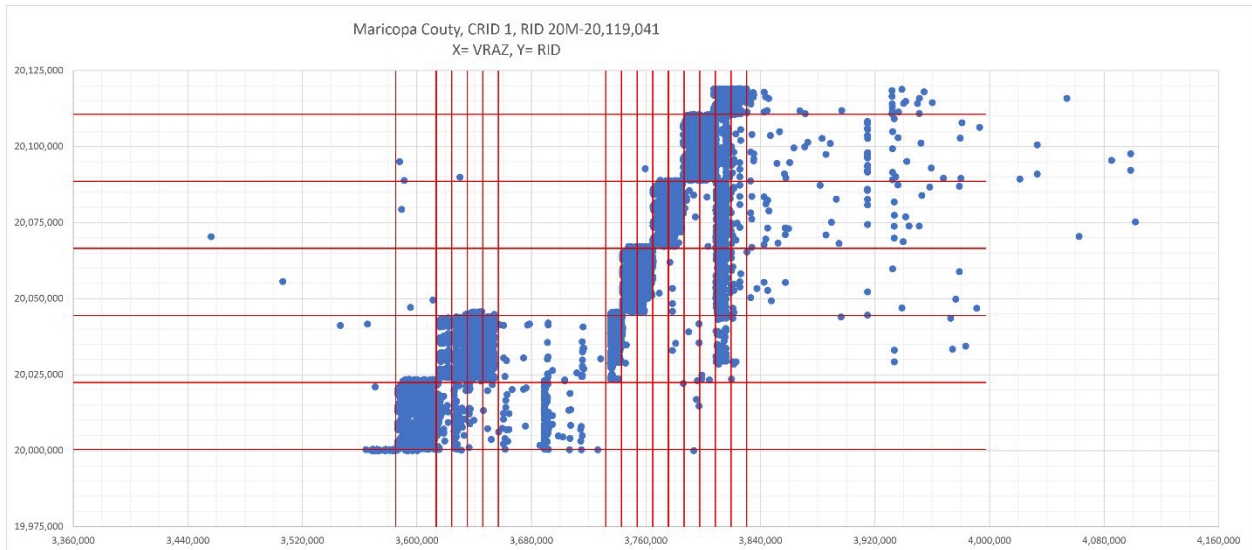


Figure 4 Maricopa County, AZ, Tranche 1, with overlay. (X: VRAZ, Y: RID)

A close-up of Maricopa's first block reveals an intricate, interwoven structure of ID assignments (Figure 5), The pattern shows:

- Multiple parallel diagonal strands that cross and interleave
- Regular spacing between strands, creating a braided appearance
- Consistent angles and distances in the crossings
- Precise coordination between VRAZ and RID progressions

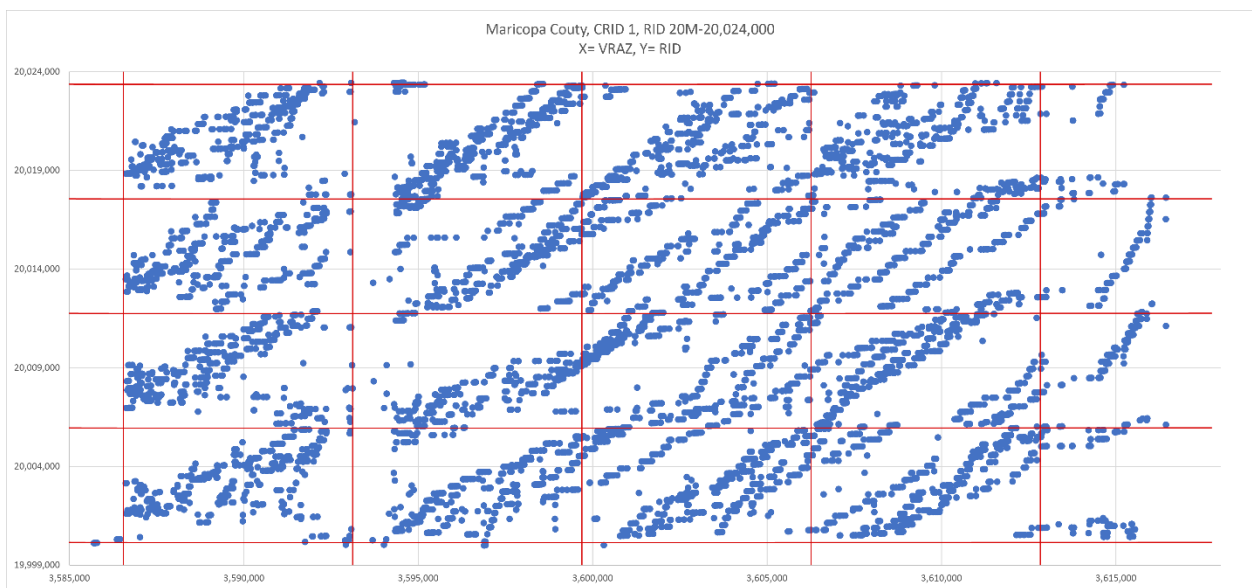


Figure 5 Maricopa County, AZ, "braided" pattern in ID number close-up

A rank analysis of these numbers reveals a deliberate pattern in how VRAZ and RID numbers correspond. When sorted by RID, the relationship between VRAZ and RID ranks follows a precise mathematical sequence:

The block is divided into sections labeled a-d (Figure 6), each containing 8 groups of 1,000 numbers. While RID ranks progress sequentially upward (Y-axis), VRAZ ranks (X-axis) follow a cyclic pattern:

First rank pair: 4X/1Y

Second rank pair: 1X/2Y

Third rank pair: 2X/3Y

Fourth rank pair: 3X/4Y

This sequence repeats vertically through each block, creating the characteristic braided pattern. A standard ID assignment would show matching or nearly matching ranks; this pattern shows intentional mathematical redistribution.

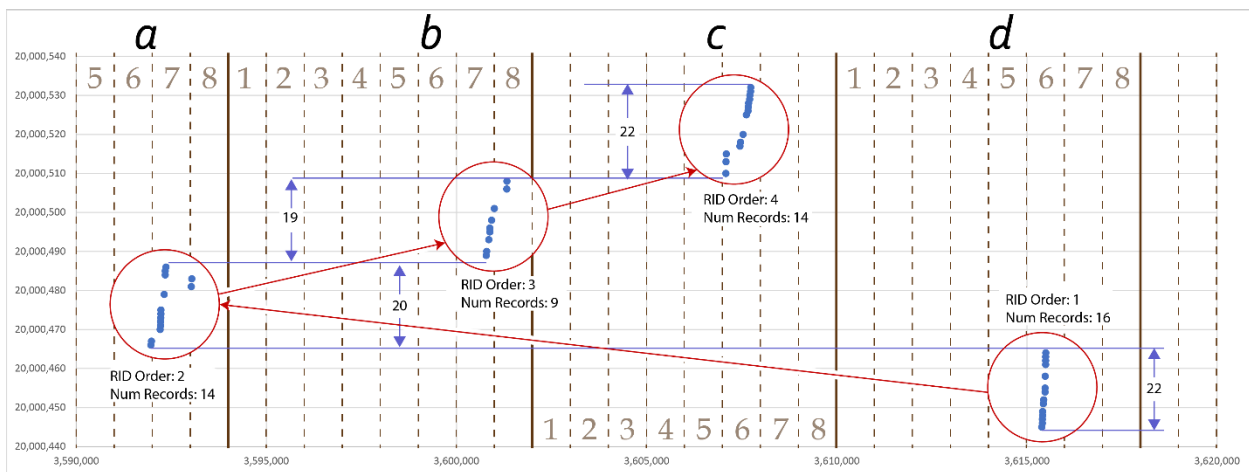


Figure 6 Close-up Maricopa County, rank value analysis (X: VRAZ, Y: RID)

The 'Braid' algorithm shares key features with New York's 'Shingle' algorithm (Figure 7) though they differ in execution. Both use cyclic permutation and slope-aligned clustering, but the Shingle (found in ~700,000 of New York's 21.5M records) shows a looser, more stepped pattern compared to the Braid's tight interweaving. Their structural similarities suggest related design principles despite distinct implementations.

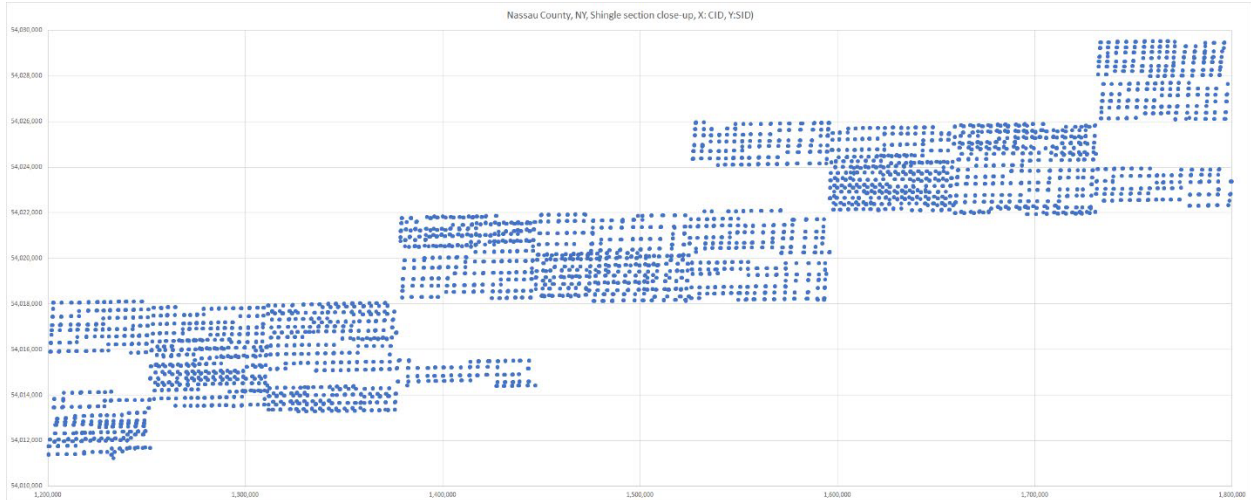


Figure 7 Nassau County, NY, Shingle algorithm section close-up (X: County ID, Y: State ID)

Maricopa's second tranche (Figure 8) combines two distinct patterns: the Braid algorithm seen in the scattered diagonal sections, and a dense block pattern similar to New York's 'Metronome' algorithm. The block pattern fills the coordinate space by distributing ID numbers across the full range of both axes, like a graphics program's flood fill. This approach requires disconnecting ID assignments from registration dates and risks number collisions by using the entire range of one axis for each step of the other.

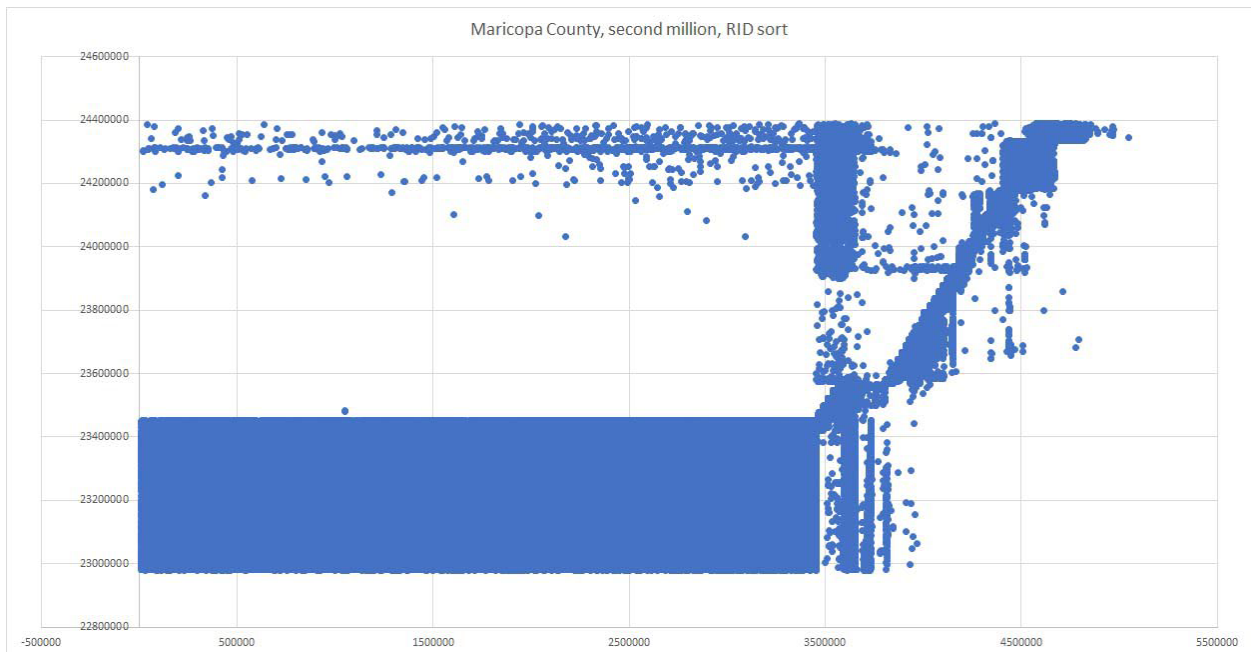


Figure 8 Maricopa County, Second Tranche, medium zoom (X: VRAZ, Y: RID)

A close-up view of the seemingly solid block reveals the Metronome's true structure (Figure 9) Rather than complete coverage, it creates a precise scatter pattern with deliberate spacing between points. The distribution suggests a mathematical formula using rotated coordinate sets, creating what appears to be pseudo-random but geometrically controlled gaps throughout the number space.

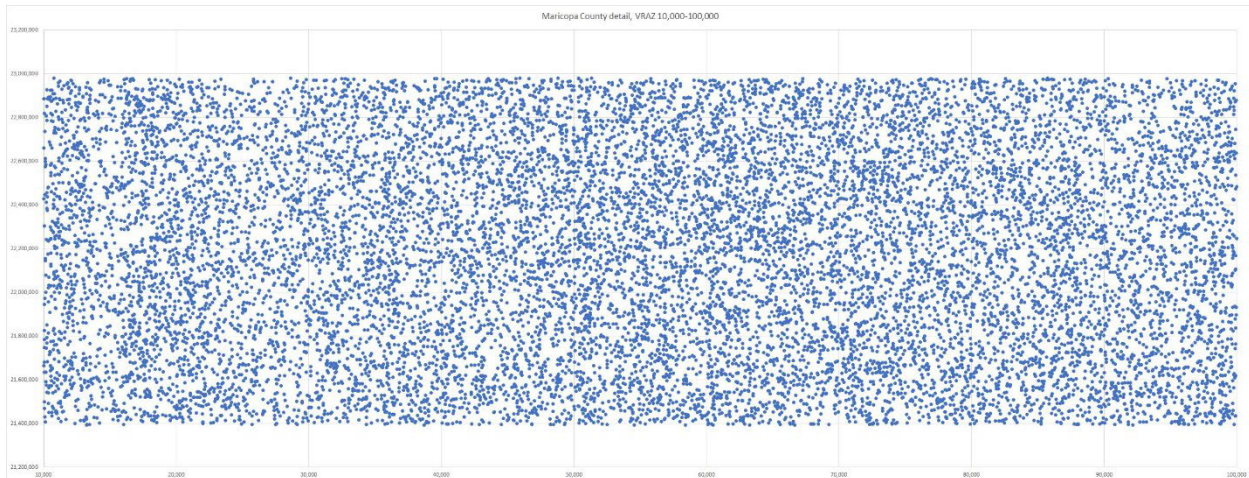


Figure 9 Maricopa County, Second Tranche, closeup (X: VRAZ, Y: RID)

The Maricopa system employs two sophisticated mathematical patterns that share a concerning characteristic: both add unnecessary complexity to what should be simple sequential ID assignments. The 'Braid' pattern uses cyclic permutation to interweave ID sequences, while the Metronome-like pattern creates precisely controlled pseudo-random distributions similar to those used in computer graphics algorithms.

Neither pattern serves any legitimate administrative purpose. Voter registration IDs should be:

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

Instead, both patterns appear designed to obscure relationships between records while maintaining plausible deniability through mathematical precision. This level of algorithmic sophistication in ID assignment actively works against database transparency and auditability.

These algorithmic transforms between VRAZ and RID numbers effectively create a covert data channel. When one set of numbers is mapped to another through complex patterns (like the "Braid" or "Metronome" algorithms), the mapping itself can encode additional information. For example:

1. The position of an ID within these patterns (which strand of the braid, which quadrant of the block) becomes an implicit attribute of that record.
2. Since these positions follow precise mathematical sequences (like the 4-1-2-3 cyclic permutation seen in Maricopa County), they can be used to systematically tag or categorize records without adding visible database fields.
3. This hidden categorization persists even when records are moved or modified, since it's embedded in the relationship between the two ID numbers rather than stored as explicit data.

This capability for covert record tagging becomes particularly concerning given the presence of over 590,000 clone records and 476,000 questionably active records - as these could be systematically tracked and managed through their positions within these algorithmic patterns.

The appearance of similar pattern elements across all Arizona counties - including vertical breaks, dense blocks, and coordinated pattern shifts - indicates these algorithms were implemented at the state level rather than developed independently by counties. Such consistency in unnecessary complexity across jurisdictions suggests centralized control of ID assignment methods.

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. Arizona's estimated 590,529 clones (8.62% of 6.85M records) and its sophisticated ID algorithms suggest the same pattern - complex ID systems correlating with illegal excess registrations.

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 Arizona's 4.3 million active records, 1.1 million show no voting activity in any election from 2020-2024, including both federal general elections (November 2020 and November 2022). While 633K of these were registered after 2019, the remaining 476K 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, yet remain listed as active. 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, Arizona'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 VRAZ and RID numbers
- Deliberate cyclic permutation (4,1,2,3) and pseudo-random distribution
- Consistent pattern maintenance across all counties
- Violation of database best practices

The precision and complexity of these patterns, particularly the Braid and Metronome algorithms, suggest deliberate design rather than administrative convenience.

This study of Arizona's voter rolls reveals evidence of multiple ID number assignment algorithms. These appear overly complex, potentially enabling data segregation and hidden attribute assignment. The presence of 475,946 apparently inactive records for 5 or more years exceeds normal error rates or acceptable administrative standards. Such a large number of problematic records could potentially impact election outcomes if manipulated. The presence of an estimated 590,529 cloned records represents an additional and unnecessary risk to election integrity in Arizona.

These findings suggest potentially problematic management of Arizona'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.

Arizona 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, Arizona 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.

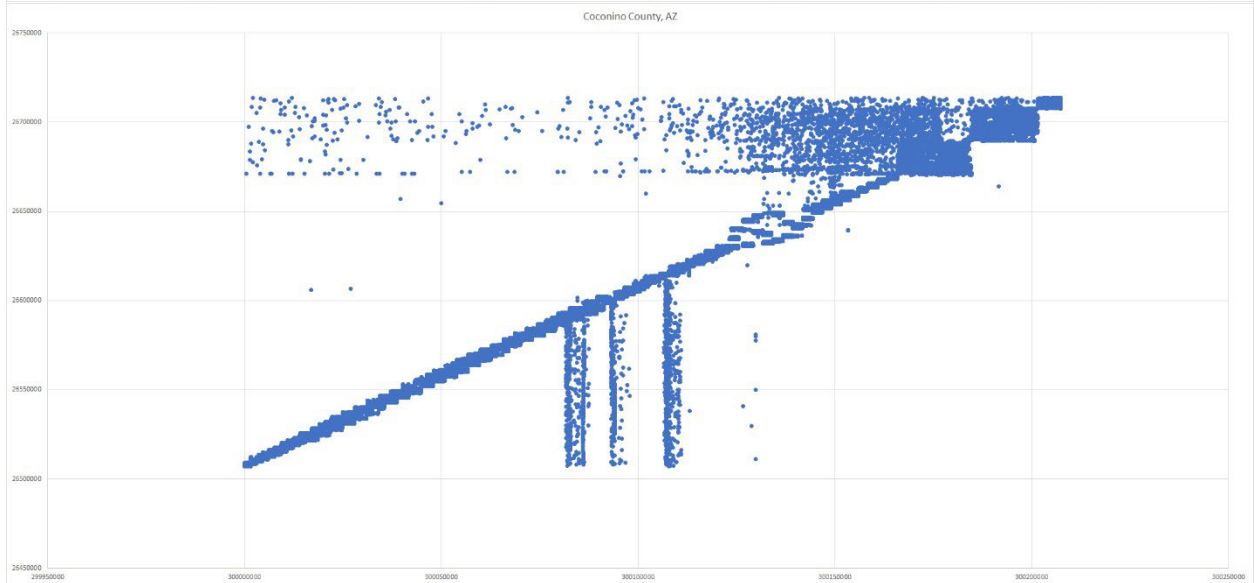
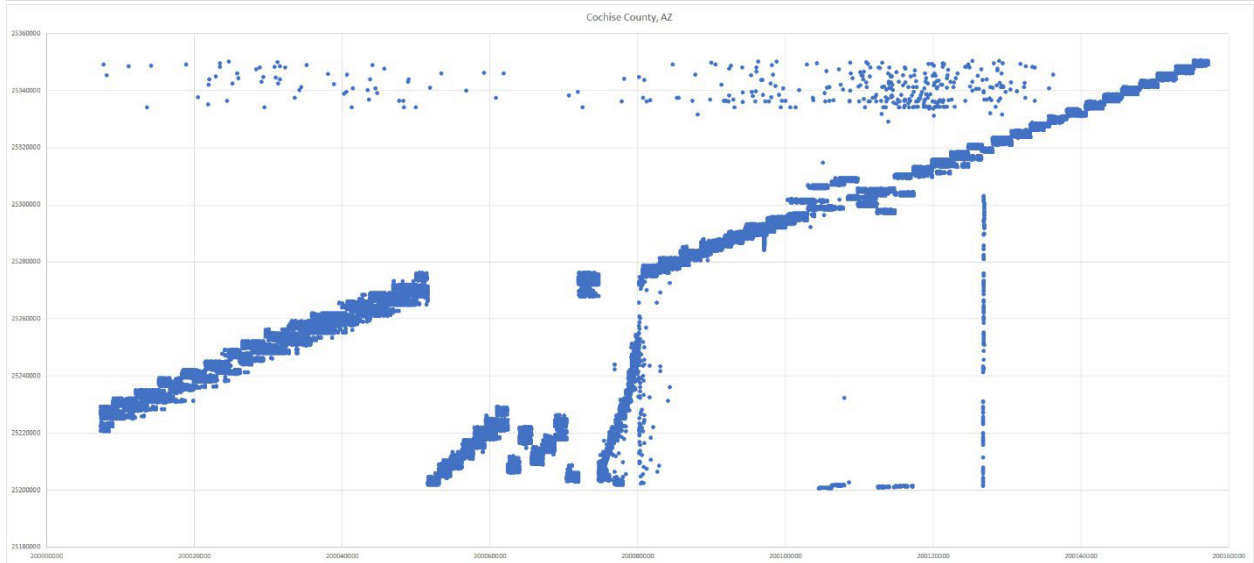
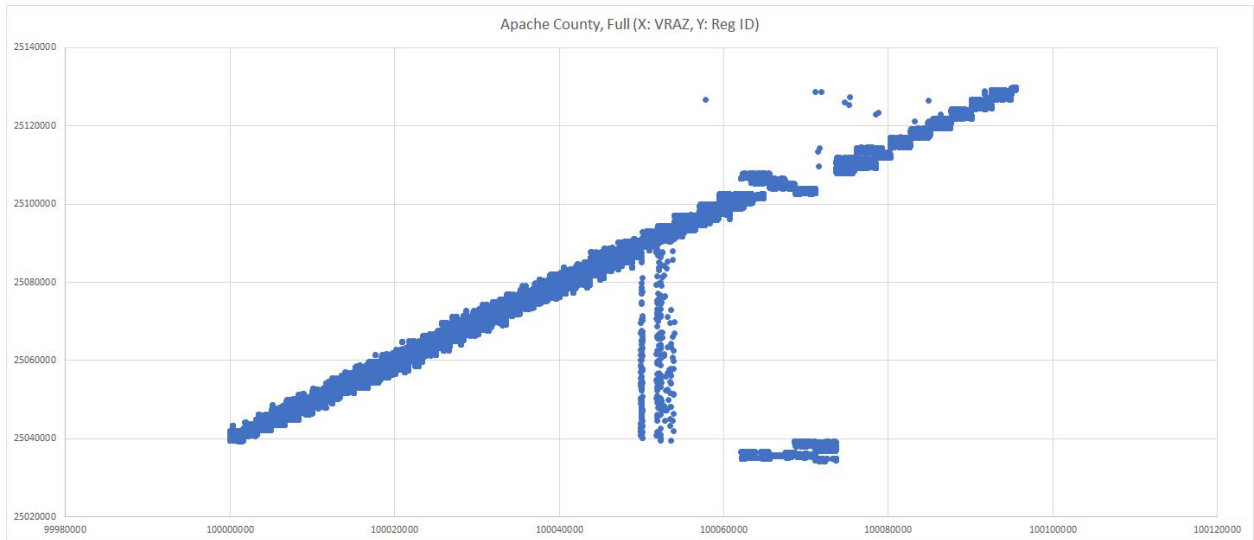
References

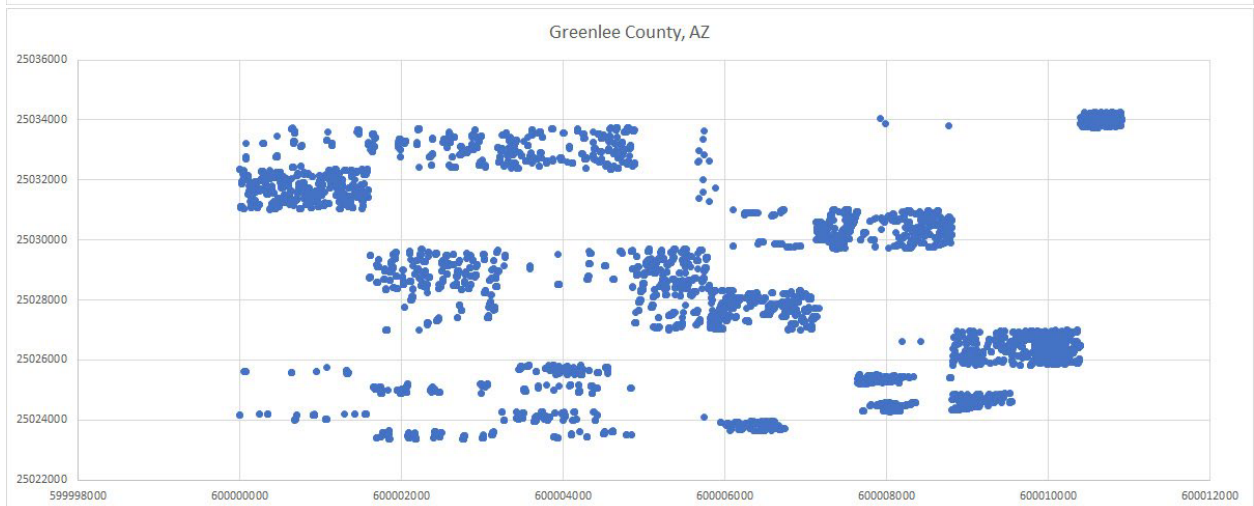
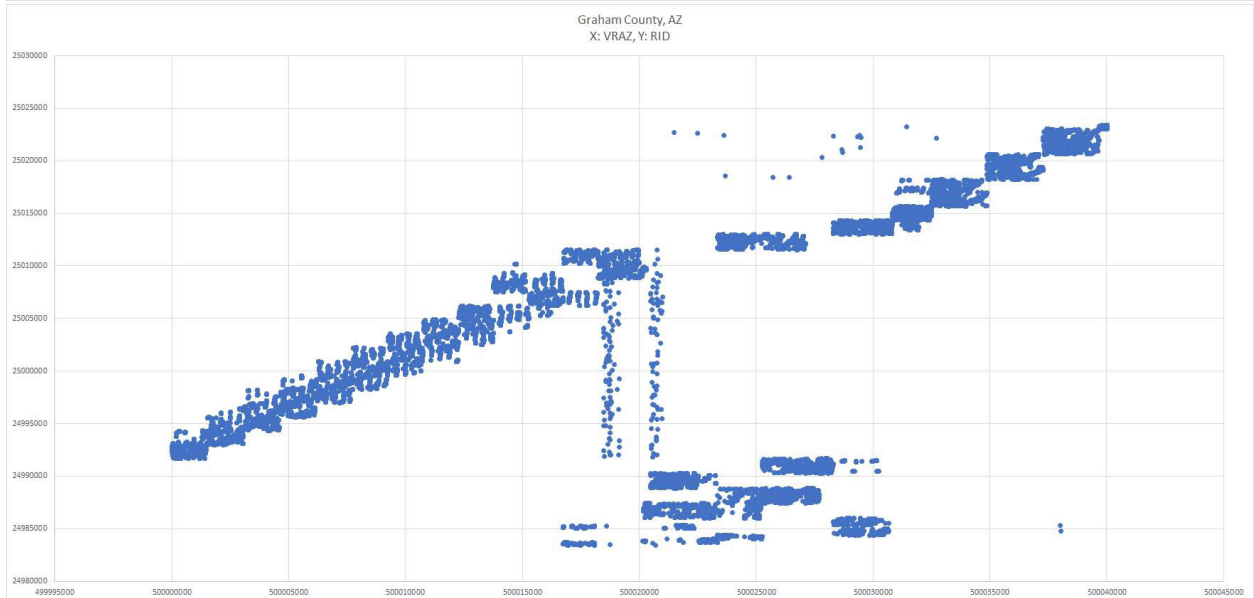
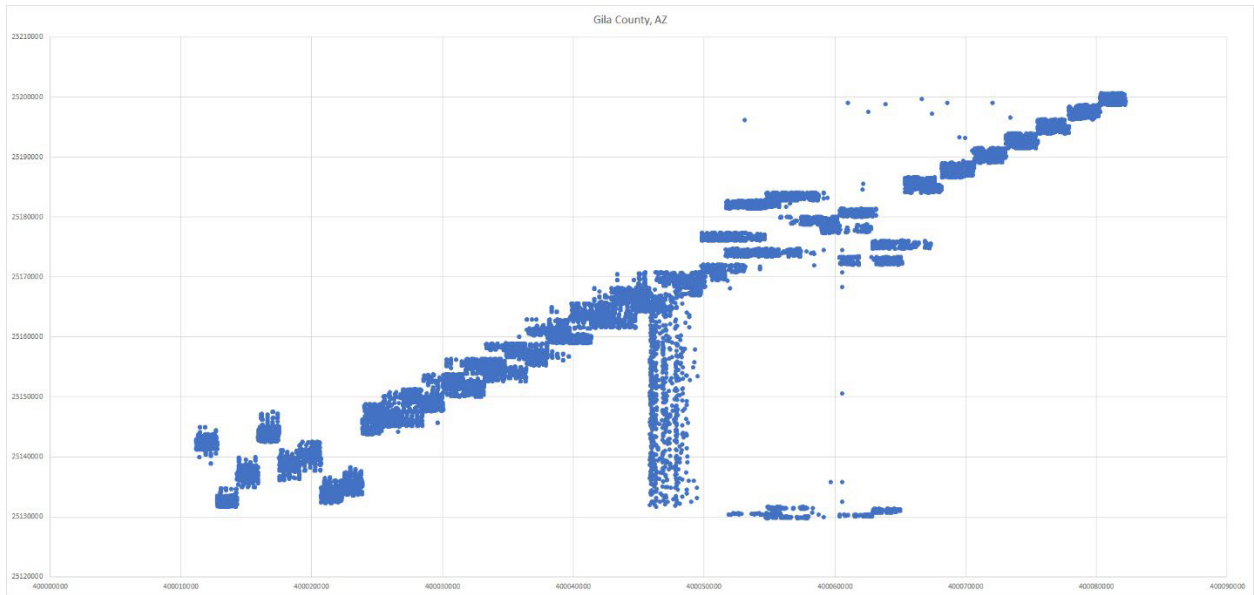
Paquette, A. (2023). "The Caesar cipher and stacking the deck in New York State voter rolls " [Journal of Information Warfare](#) **22**(2): 86-105.

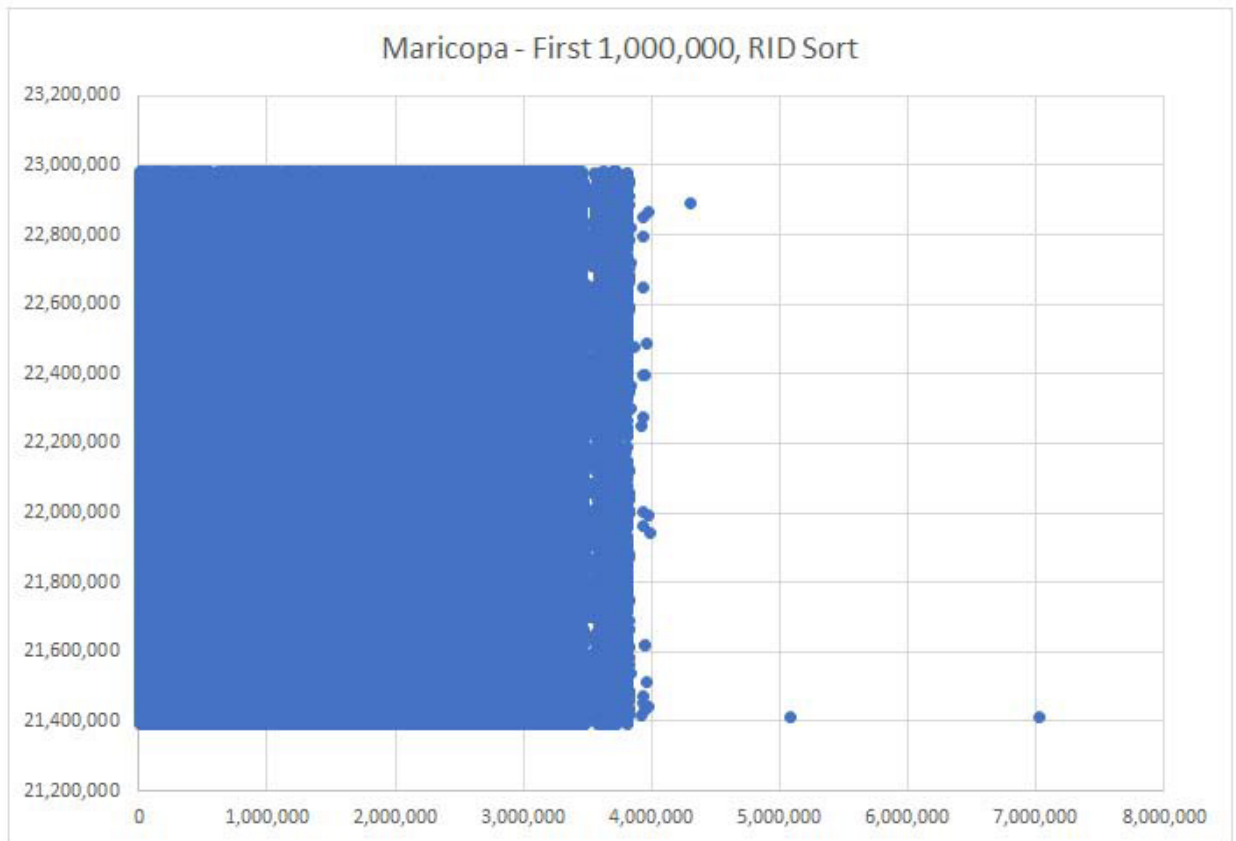
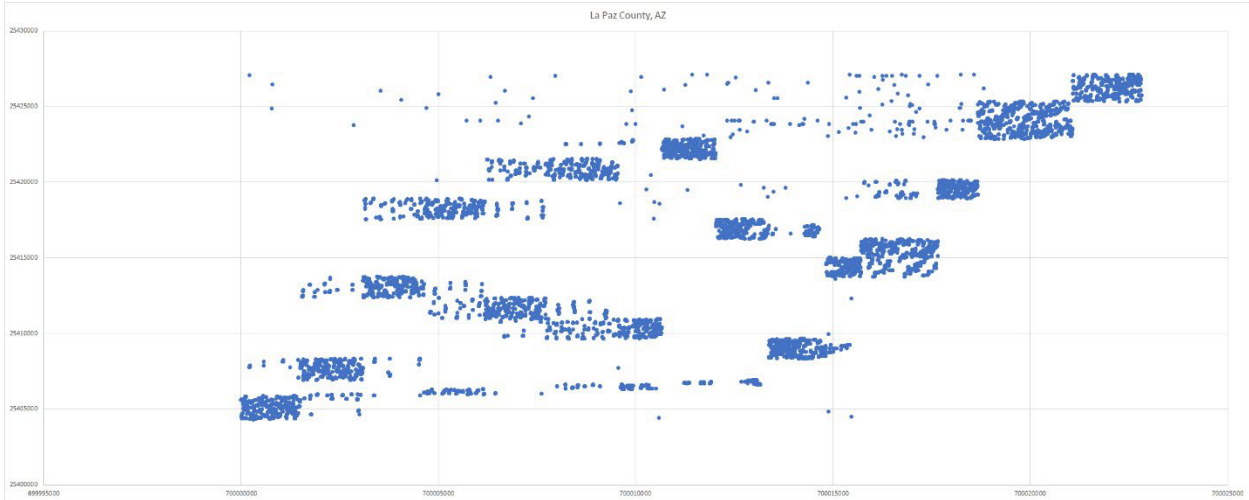
Paquette, A. (2024). "New Jersey voter ID numbers reconfigured with shift cipher." [\(In-Press\)](#).

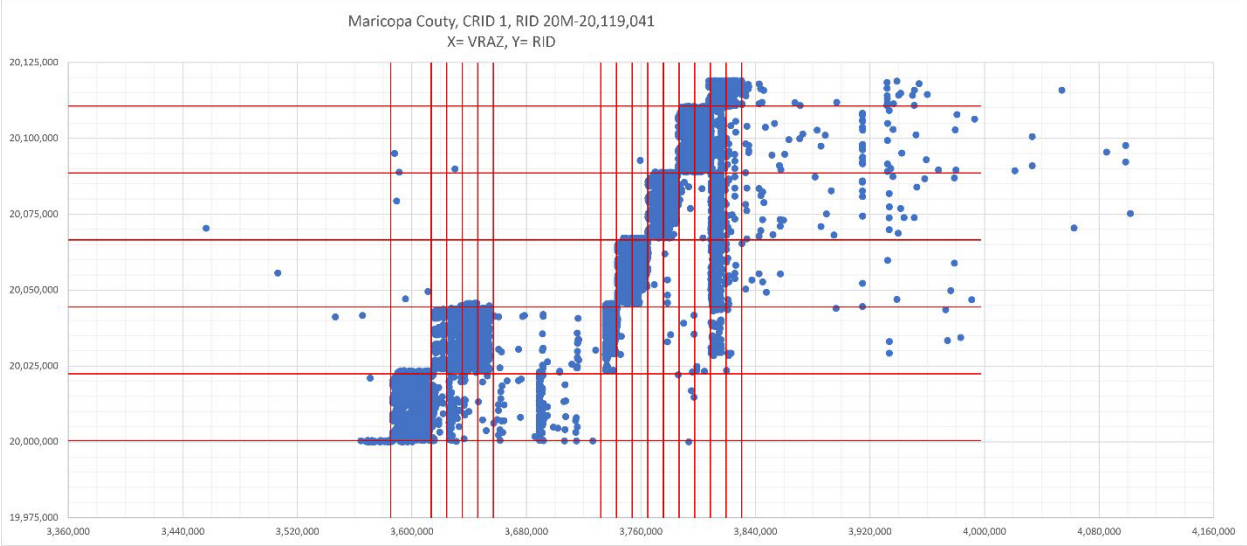
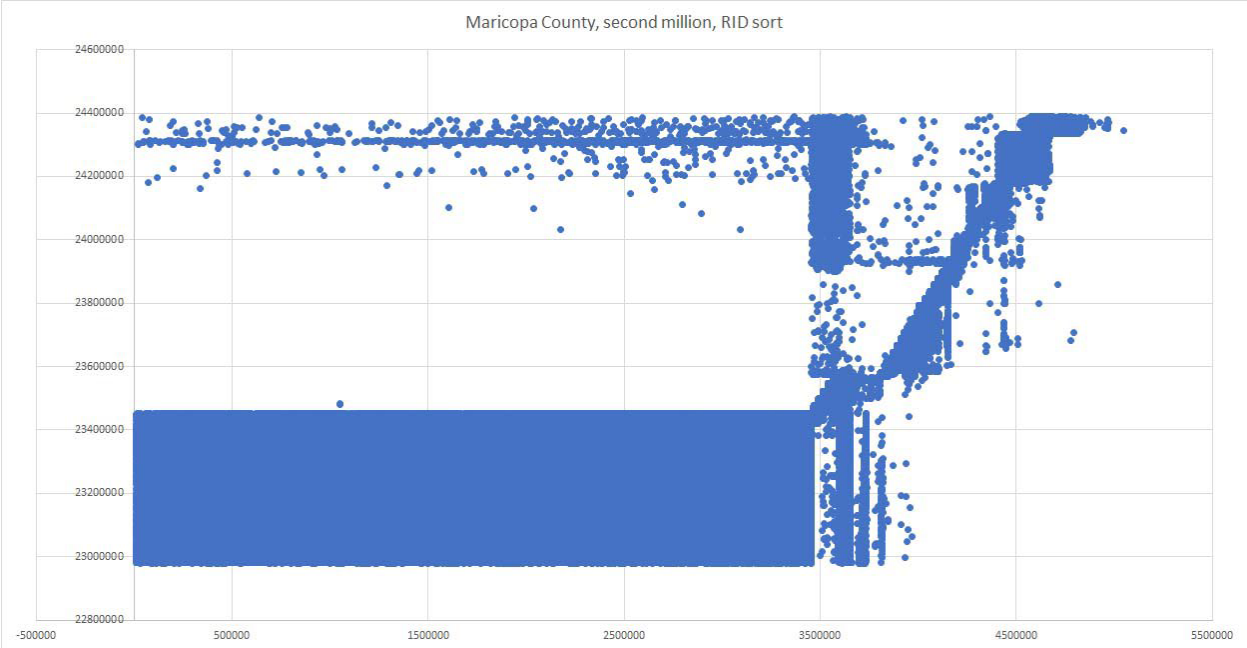
Appendix

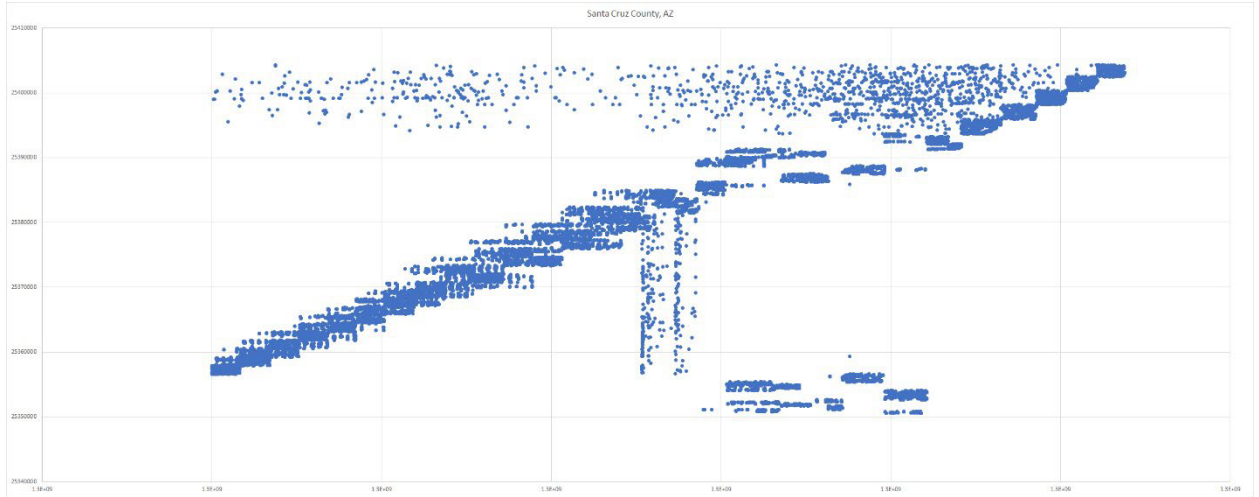
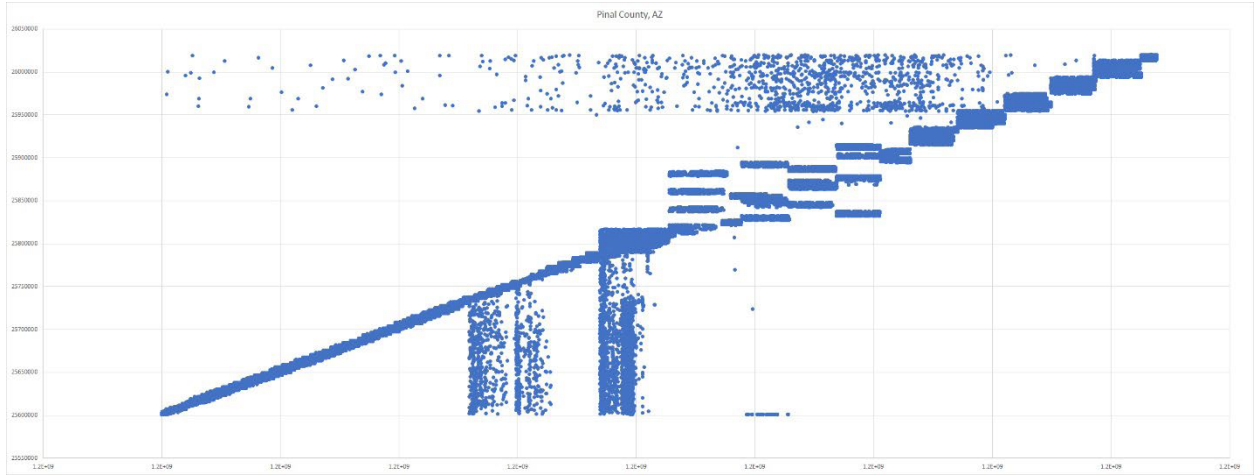
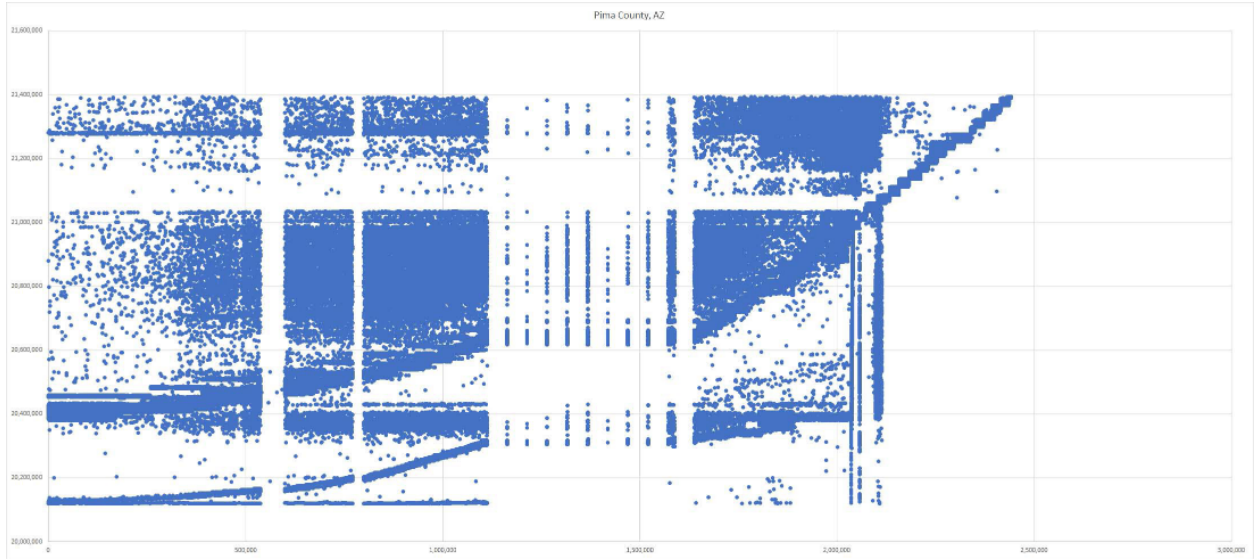
Scatterplots for all 15 Arizona counties, in alphabetical order:

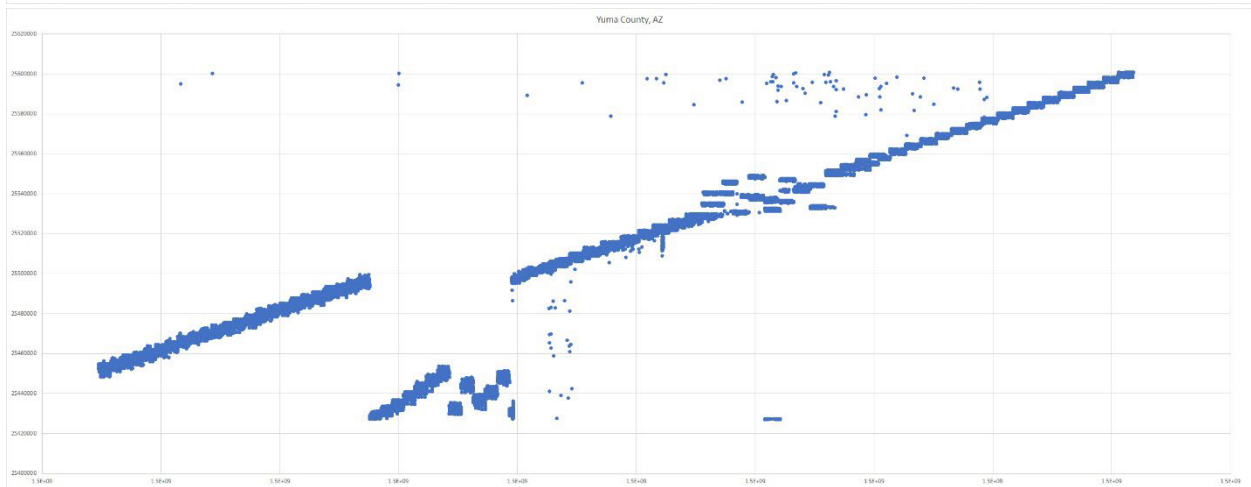
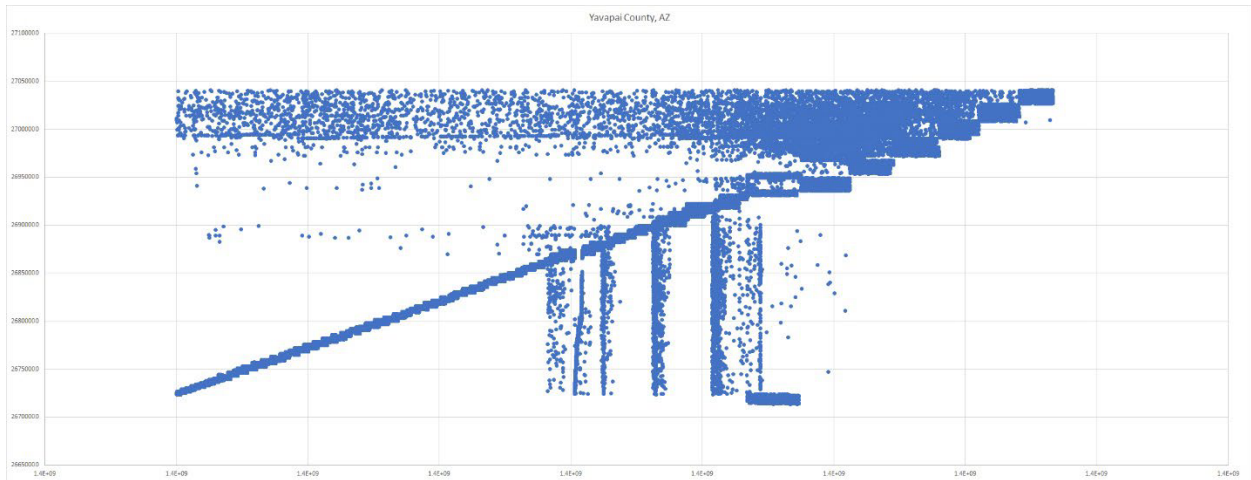












ⁱ This was found by researcher Vico Bertogli, of Pennsylvania