Oklahoma voter roll ID numbers

Preliminary Report

Andrew Paquette

9/5/2024

Introduction

The state of New York's voter rolls uses multiple complex algorithms to map County ID (CID) to State ID (SID) numbers. At least two of the algorithms conceal attributes that can be used to identify certain categories of suspicious records based on ID number alone. New Jersey uses a different algorithm to assign ID numbers that can be used to clandestinely track records of interest. In Ohio, at least 3 of 88 counties used complex algorithms to assign ID numbers. These allow segregation of records and thus the assignment of attributes not contained in data fields.

It is not proven that any of the algorithms were designed for a nefarious purpose or have been used this way. However, the fact that they can be used to conceal information from the public and regular users of the systems represents a valid security risk. The presence of the algorithms is unusual enough to warrant further investigation in other states, to determine how widespread the use of these types of algorithms is nationwide.

Oklahoma

I have spent literally one day looking at Oklahoma's voter rolls. Much less if you subtract the time it took to import each county's database into a master database for study. In comparison, it took weeks before the first hints of voter roll algorithms were found in Ohio and New Jersey, and even longer in New York With the caveat that this is not enough time to yield a definitive response either way, here are a few preliminary observations:

Voter Status

The rolls I viewed, downloaded two days ago on 9/2/2024, do not appear to have any purged records. The only status conditions are "A" and "I", presumably for Active and Inactive. This suggests that instead of purging records, Oklahoma deletes invalid registrations. Deleting instead of purging may be safer for election integrity, but it also makes it more difficult to find any algorithms if present.

Voter ID

I looked at the county and state versions of the voter rolls. At the county level, there is a single voter ID number, labeled "Voter ID". In the state rolls, a similar field is found, and no other. A comparison shows that the Voter ID in the state records matches the Voter ID number in county records. This indicates there is no separate county and state ID number as in NY, OH, and some other states. If true, this rules out the possibility of a mapping algorithm such as the one found in NY. It does not rule out the type of algorithm found in New Jersey, which does not use a county ID.

The Help America Vote Act (HAVA) requires that states provide a unique identifier to voters within the state. It does not require an additional county ID.

Clones

"Clone" records are records with unique ID numbers that appear to have been assigned to voters who already have a different number assigned to them. Because Oklahoma does not maintain purged or deleted records in their rolls, all clone records could potentially be used to double vote.

To locate clones, a match field was created by concatenating last name, first name, and date of birth. These search items are what the state of New York is supposed to use, under the law, to identify attempts at multiple registration (whether intentional or unintentional.) Using this method, 7,057 potential clones were identified statewide in Oklahoma.

Due to lack of time, the number of unique ID numbers involved was not checked, but this translates to approximately 3,500 problematic registrations, or possible excess votes. Some of these are likely false positives, but in other states this is usually a small percentage of the total.

The numbers involved are small, but in a state with a relatively small population as Oklahoma, it is enough to affect close races where the margin of victory was less than 3,500 in state races, or less than 332 in a county like Tulsa, which has 664 possible clones, or Osage, which has 128, or about 64 potential excess votes.

Impossible registration dates

Statewide, 179,598 out of 2,381,508 (7.54%) records have a registration date that is earlier than the date of birth listed in the record. Either the registration or birth dates are false. False data, though common in many state voter rolls, violates HAVA:

HELP AMERICA VOTE ACT (HAVA) OF 2002: SECTION 303(A)(4) OF HAVA (52 U.S.C. § 21083(A)(4)) MANDATES: "THE STATE ELECTION SYSTEM SHALL INCLUDE PROVISIONS TO ENSURE THAT VOTER REGISTRATION RECORDS IN THE STATE ARE ACCURATE AND ARE UPDATED REGULARLY…" FURTHERMORE, SECTION 303(A)(2)(A)(II) REQUIRES THAT THE STATEWIDE VOTER REGISTRATION LIST BE: "ACCURATE, COMPLETE, AND CURRENT."

To be fully compliant, these records must be either removed or corrected.

Scatterplots

Scatterplots can be used to quickly determine via visual examination whether numbers follow or deviate from expected values. To make a scatterplot, 2 variables must be identified. In NY and OH, SID and CID were used, to understand the relationship between the two numbers, which should be linear. OK lacks CID numbers. For this reason, a scatterplot of the Voter ID against Registration Date was made for every county, to see if there is a linear relationship.

If so, then ID numbers would increase as registration dates become more recent. A good example of a normal scatterplot comes from Fairfield County, OH (Figure 1).
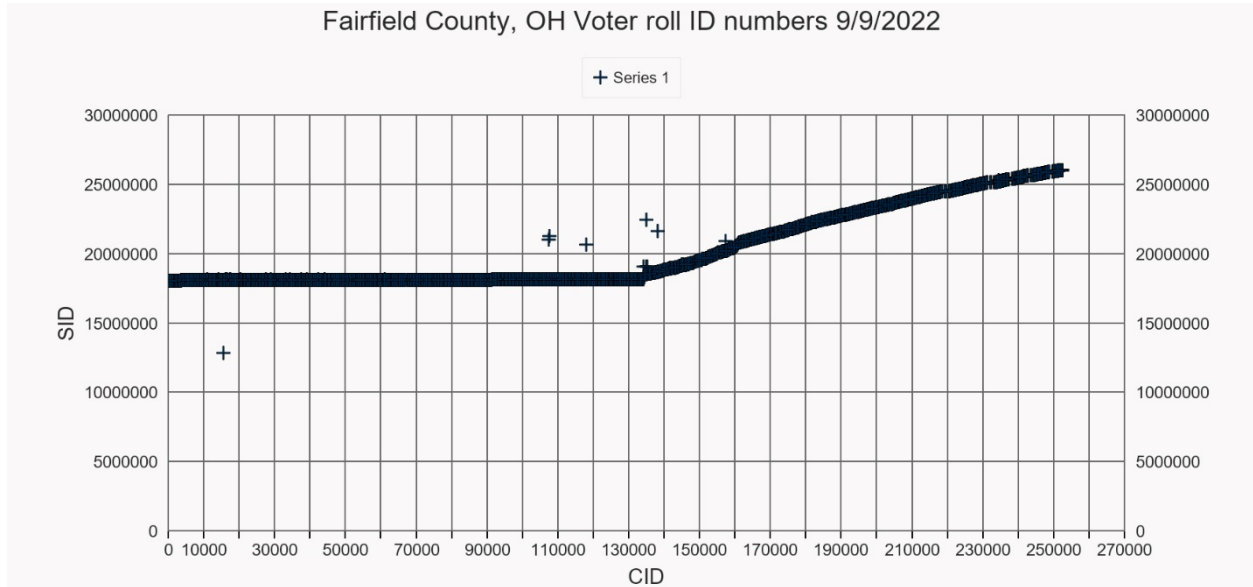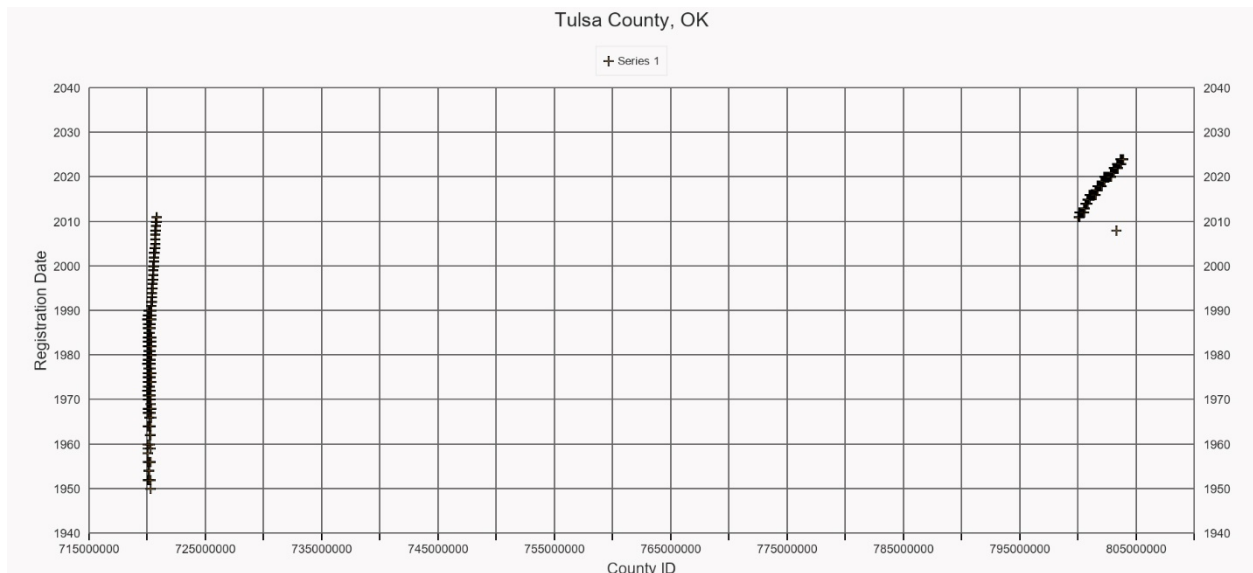
Figure 1 Fairfield County, OH scatterplot showing normal distribution of CID and SID numbers

Oklahoma's scatterplots tend to be more similar to Ohio's Fairfield County than the more unusual plots found in all NY counties and some Ohio counties. The following plot from Tulsa County resembles all other counties when fully zoomed out to view the entire county.

What this plot tells us is that ID numbers generally ascend as registration dates become more recent. There is a large break in CID numbers between about 720,000,000 through 800,000,000 that occurs in 2012. This break is found in all other OK counties. The reason for this is unclear, particularly for a state with a population size that is unlikely to ever exceed the available unused numbers.

A close-up of numbers on the left of the plot reveals another break in the numbers in the year 1990, after which they ascend normally. Earlier numbers do not follow a normal ascending pattern, but are found in any year from 1950-1990, regardless of number size. That is, a high number from the series is

just as likely to be from 1950 as 1990, but later numbers always ascend with the year. This is different from some counties and bears further investigation (Figure 2).
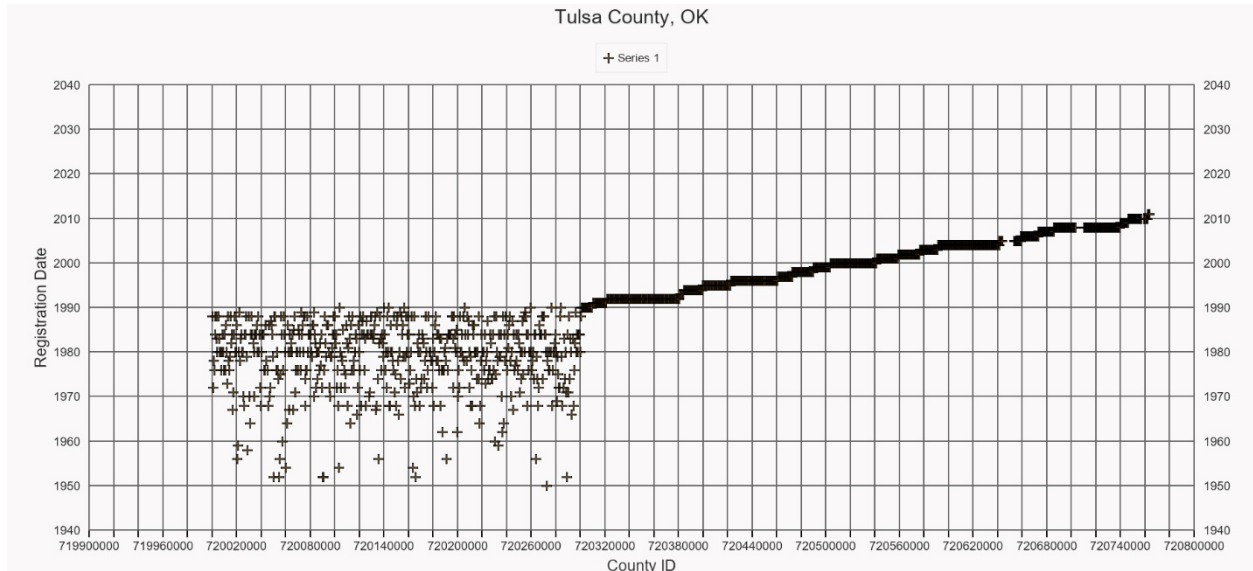

Figure 2 Tulsa ID numbers and Registration dates, closeup

Registrations

Annual registrations were checked based on the Registration Date field. A chart of these values in Tulsa for the years 2000-2024 ascends, with the exception of presidential election years, when it spikes. This is expected behavior. However, the size of the registration spikes in 2016 and 2020 are unusual. Each represents about 10.0% of all registrations at the time. This is a high value, though higher values have been recorded elsewhere (Figure 3).
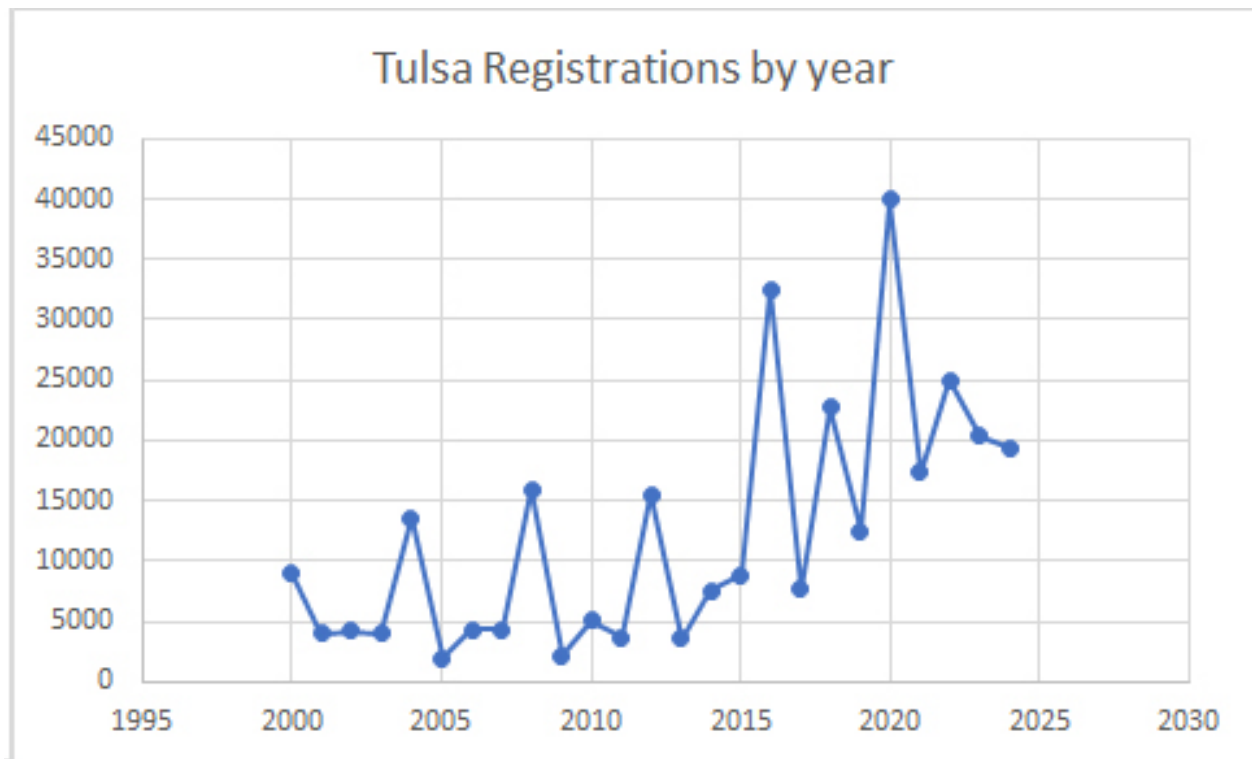
Figure 3 Tulsa registrations by year 2020-2024

Algorithms

There are no obvious signs of the use of any overly complex or intentionally hidden algorithms in Oklahoma at this stage of analysis. However, this was not expected, as it took longer to find examples in other states. It is not possible to say with any confidence or justification that they exist or don't exist at this time.

Conclusion

The lack of purged and deleted records, as well as the use of a single voter ID number, reduce the amount of data that can be explored to determine whether any method of tagging or tracking records was implemented in Oklahoma. Lack of time also prevented a proper examination of records for this property.

The presence of cloned records and false data recommend corrective measures be taken, and a more thorough investigation for other anomalies initiated. Election law is designed as a zero-trust system, so that all possible contingencies are dealt with. If any of these provisions are ignored, or remain uncorrected when faults are found, the electoral franchise is corrupt and cannot be trusted.