Does the US have a duplicate voter problem?

Yes — but a small one, and it is a data quality problem rather than a fraud problem.

Published January 2024, based on publicly available voter registration data purchased between September and November 2023 under state open records laws. All figures refer to that 2023 snapshot.

49.5M	7	394,396	0.96 v 0.89
Voter profiles	**States analysed**	**Duplicate registrations**	**Democrat v Republican**
Across 7 US states	AR, FL, GA, MI, NC, OH, PA	0.8% of the analysed population	Duplicate rate by party — no meaningful skew

Voter fraud is a politically charged topic in the United States. But strip away the politics and what remains is a data quality problem — one that entity resolution is well suited to diagnose.

We gathered publicly available voter registration data from seven US states where the data is freely obtainable — Arkansas, Florida, Georgia, Michigan, North Carolina, Ohio, and Pennsylvania — representing over 49.5 million registered voters, and ran Tilores entity resolution across the full dataset, comparing every registered voter against every other using name, address, date of birth, and other available fields.

Methodology

The analysis uses three matching techniques in combination:

- **Text similarity matching** — catches minor character differences ("Philip" / "Phillip") and names that differ in spelling but are phonetically the same ("Steven" / "Stephen").
- **Geographic proximity matching** — uses address coordinates, so two addresses within a defined distance can be treated as a candidate match.
- **Temporal range matching** — treats dates falling within a defined window, such as a date of birth out by a few days, as a potential match.

It is the combination of these techniques across the available fields that forms the deduplication rules. Multiple rules cover different duplication scenarios; if any one fires when comparing two records, those records are

treated as duplicates, connected into a single identity graph together with an explanation of which rule was triggered.

KEY LIMITATION

A duplicate registration is not a duplicate vote. Registration files tell you who is on the roll. They only tell you who actually voted if the state publishes that field — and of our seven states, only **Ohio and Pennsylvania** do. Everything below regarding actual double voting is therefore limited to those two states. In the other five the data is not publicly available — the state holds it, but does not publish it, so it was out of reach for an analysis run from public sources.

The headline numbers

Across all seven states we identified **394,396 voters with duplicate registrations**, representing 0.8% of the total voting population in our sample. A duplicate here means an active registered voter holding two or more active registrations — whether in the same county, in different counties within the same state, or in two different states.

That figure demands context. Almost none of it represents fraud. The overwhelming majority represents a genuine data management problem: voters who moved state or county, voters whose names appear in multiple formats, and voters who died but whose registrations were not removed in time. The system is producing duplicates faster than it can clean them up.

The state breakdown

STATE	DUPLICATE REGISTRATIONS	DUPLICATE RATE
Florida	**148,516**	**1.10%**
Pennsylvania	**80,142**	**1.01%**
Georgia	**51,876**	**0.73%**
Ohio	**47,000**	**0.79%**
Michigan	**35,723**	**0.47%**
North Carolina	**20,323**	**0.32%**
Arkansas	**10,816**	**0.92%**
Total	**394,396**	**0.80%**

No state was a dramatic outlier, which is itself a useful finding: the problem is consistent and structural rather than concentrated in a few badly run jurisdictions. The spread between North Carolina at 0.32% and Florida at 1.10% is real, but it is a spread within a common problem rather than evidence that some states have solved it.

County-level variation

Duplicate rates vary significantly below the state level. In Arkansas the statewide average is 0.92%, but Searcy County — with a voting population of 5,805 — reaches 2.06%, or 121 individuals. In Florida, Hardee County has 13,654 registered voters and 314 duplicates (2.3%), while Osceola County has 214,588 registered voters and 5,376 duplicates (1.94%).

That last pair is the instructive one. Hardee and Osceola differ in size by a factor of fifteen, yet both sit at roughly twice the average we measured. Variation of this kind suggests the problem is partly driven by local data management practice rather than by scale or systemic factors alone.

Registration rates: a separate signal

Alongside the duplicate analysis, we saw significant differences between the registration rate of certain counties and their general population. This is a registration rate, not a duplicate rate, and needs reading separately.

Michigan — above 90%

Several counties showed registration levels above 90% of the population — not feasible when roughly 22% of the US population is under 18, and therefore ineligible. The likely explanation is that Michigan is not reconciling its lists against death notifications as effectively as other states.

Arkansas — as low as 26%

At the other end, some Arkansas counties showed unusually low registration — Lincoln County at 26.02%. That points to room for improvement in voter engagement rather than to a data quality fault.

The distinction matters: a roll can be inaccurate because it holds people it should not, or because it is missing people it should hold. Entity resolution addresses the first. The second is a participation question.

Party affiliation in duplicate registrations

In every sampled state except Michigan, which does not disclose it, we were able to break duplicates down by registered party. Looking only at voters registered as Democrats or Republicans, Democrats showed a duplicate rate of **0.96%** and Republicans **0.89%** — 121,985 and 120,384 records respectively.

STATE	DEMOCRAT	REPUBLICAN
Florida	**1.30%** 58,455 records	**1.29%** 65,905 records
Pennsylvania	**1.12%** 39,797 records	**0.96%** 30,928 records
Arkansas	**1.10%** 632 records	**1.23%** 1,149 records

STATE	DEMOCRAT	REPUBLICAN
Ohio	**0.68%** 6,686 records	**0.67%** 9,061 records
Georgia	**0.58%** 8,713 records	**0.43%** 7,449 records
North Carolina	**0.37%** 7,702 records	**0.29%** 5,792 records
Overall	**0.96%** 121,985 records	**0.89%** 120,384 records

Michigan is excluded because it does not disclose party affiliation. Voters registered with other parties, or with none, are excluded from this table.

The overall gap is **0.07 percentage points**, and no individual state gap exceeds 0.16 points. The direction is not uniform either: Republican-affiliated registrations show the higher rate in Arkansas, Democrat-affiliated registrations in most other states, and Florida and Ohio are effectively level at a hundredth of a point apart. Duplicate registrations in this dataset are not a partisan phenomenon. They are a function of how long someone has been registered, how often they have moved, and how their name has been recorded — none of which correlates meaningfully with party.

Double voting: what Ohio and Pennsylvania showed

Duplicate registration data is one thing, but on its own it means nothing more sinister than poor processes and untidy records. The larger question is whether anyone actually voted twice.

We only hold actual voting data for Pennsylvania and Ohio, so we examined the duplicates within that subset. There we found approximately 1,000 potential cases. Wanting to be stringent about a claim of this kind, we applied additional deduplication rules and then reviewed the remaining cases manually, one by one.

~1,000
Potential cases from matching

→

61
After stricter rules + manual review

Democratic Party affiliation	**31**	Republican Party affiliation	**21**
Both parties across the two records	**6**	No party affiliation	**3**

Two of the 61 span Pennsylvania and Ohio, both involving absentee votes; the rest sit within a single state. Two examples give a sense of what these look like in practice:

> An individual registered to vote in **West Chester, Pennsylvania** in October 2020, shortly before the presidential election. Nine days later the same individual registered in Philadelphia. There was a minor difference in the name; the date of birth and phone number were identical. Both registrations voted.

> An individual in **Normalville, Pennsylvania** registered twice at the same address, sixteen years apart almost to the day. The only difference between the two records was a small discrepancy in the date of birth. Both registrations voted. This was the most common pattern — a date of birth out by a few days or months, with name and address otherwise unchanged.

We want to be careful about what this is. It is **a candidate set for review by election officials, not a legal finding.** Entity resolution followed by manual review establishes that two records are very probably the same person; establishing that an offence occurred requires investigation by the relevant authorities. The rate is vanishingly small relative to the registrations in those two states — but it was detectable, and only because the vote history was published. Had the other five states published theirs, we would expect to find further cross-state cases.

What already exists: ERIC

We were not the first to run this analysis, and it would be misleading to present it as though we were.

A non-profit organisation called **ERIC** — the Electronic Registration Information Center — already performs cross-state list comparison for its member states using entity resolution, and supplements voter rolls with additional sources such as state DMV records, in a privacy-preserving manner. We compared methodologies with them after publishing our first results. They were finding broadly the same duplicates we were, including the double-vote cases.

Two things follow. First, **detection is not the bottleneck.** The technology to find these records exists, is in use, and works. What happens next — actually reviewing and removing the records — sits with each individual state, and is outside ERIC's control. Second, **coverage is uneven.** A number of states have withdrawn from the consortium in recent years. A state that leaves loses the cross-state view of its own roll, and its departure also removes its records from the pool that remaining member states match against, so the effect is not contained to the state that left.

A state's own Tilores instance

Continuous resolution of your own roll, inside your own AWS account, under your own control. A live view of data quality and something to act on between reporting cycles — plus the same engine for every other identity dataset the state holds.

ERIC membership

The cross-state picture no single state can assemble alone, with data-sharing agreements and supplementary sources already in place. Nothing a state runs internally can replicate it.

A STATE WOULD GET BETTER RESULTS THAN WE DID

Everything in this note was produced from the fields that appear in *public* voter files — name, address, date of birth, registration status, party affiliation. A public showcase can only use public data.

A state election office holds considerably more about each registrant: driver's licence numbers, full or partial Social Security numbers, prior addresses, application and signature history, and links into DMV and vital records systems. Those identifiers improve matching in both directions at once — confirming genuine duplicates that public data cannot resolve, and ruling out false positives where two different people share a name and an approximate date of birth.

Vote history is the clearest illustration. Five of the seven states we analysed record whether a registrant voted but do not publish it — the sole reason we could not test those five for double voting. Each already holds everything needed to run that test on its own data.

The figures here should therefore be read as a floor, not a ceiling. A state running the same engine against its own complete records would find duplicates we could not see from the outside, and would do so with materially higher confidence in every individual match.

What would actually reduce the problem

Wider participation in cross-state matching. The mechanism already exists. The constraint is how many states take part, and whether they act on what comes back.

Automatic death registry synchronisation. Social Security Administration death data should be matched against voter rolls continuously rather than annually or ad hoc — using fuzzy logic, not exact match, because death records and voter records use different name and address formats.

Fuzzy matching as the standard, not exact-match rules. Exact matching misses the large majority of genuine duplicates, because real-world names and addresses are never entered consistently. A rule that only catches perfectly identical records will report a clean roll while the duplicates remain on it.

The same technique protects eligible voters

Entity resolution is not only a tool for removing records. It is equally a tool for not removing the wrong ones — and that matters just as much to the accuracy of a roll.

In 2018 the State of Georgia attempted to implement an "exact-match" requirement, under which a voter registration had to match the individual's driver's licence or Social Security record exactly. Any small variance placed the registration in a "pending" status until corrected. Around 51,000 registrations were pending in Georgia that year, disproportionately affecting voters with names more likely to be misspelled or transliterated inconsistently in official records. Several other states, including Wisconsin and Virginia, had already cancelled or scaled back similar requirements.

The underlying failure is the same one that lets duplicates survive: exact matching cannot cope with the ordinary variance of real-world data. Would it be reasonable for José Rodríguez to be unable to vote because

his driving licence reads Jose Rodrigues, when the date of birth and address match? To a human reviewer these are obviously the same person. Entity resolution lets a computer reach the same conclusion. Any state cross-validating its register against social security, licence, or death records should be doing so with a system that accommodates that variance — the alternative produces two kinds of error at once.

Beyond the electoral roll

We used voter data because it is one of the few large-scale, publicly available identity datasets in the United States — not because electoral rolls are the largest instance of the problem. The same fragmentation appears wherever an agency has been collecting records on people and organisations across separate systems and many years:

Improper payments & duplicate enrolment

The same person enrolled more than once across programmes, agencies, or states under name and address variants that exact-match rules never catch.

Student records across institutions

Students move between schools, districts, and states, re-enrolling each time under a slightly different name or identifier — the same mechanism as a voter who moves. Fragmented learner records break longitudinal outcome tracking and distort per-pupil funding.

Beneficiary & case file consolidation

One person's records spread across systems never designed to share a common key — reassembled into a single view without merging the source systems.

Tax & revenue records

One taxpayer appearing as several across years, filing types, and associated entities — with the links between them never made explicit.

Licensing, permits & DMV records

Multiple licence or permit records for one individual, created over decades of differing data entry standards and formats.

Vendor, grant & procurement integrity

Resolve vendor entities to surface shell companies, related parties, and sanctioned entities across procurement and grant systems.

Where the data lives

Runs inside your environment

Deploy into your own AWS account, including AWS GovCloud. Your records stay inside your own boundary — we do not need a copy.

On-premise and air-gapped

The engine depends only on a key-value store, a queue, and file storage, so it runs on-premise and inside isolated networks.

SOC 2

Tilores is SOC 2 audited. Deployment, access, and retention questions are answered in detail during evaluation.

Resolution without pooling raw data

Cross-agency matching happens at the identity layer rather than by centralising records, so agencies keep custody of their own data.

Talk to us

Tilores is available to US federal, state and local government agencies through our partnership with **Carahsoft**. The quickest next step is a short technical demo against a dataset you care about; we bring Carahsoft in for contract vehicle and procurement questions.

Book a demo: tilores.io/contact · **Carahsoft:** carahsoft.com/tilores · **Government use cases:** tilores.io/solutions/government

Sources and notes. Voter data was purchased between September and November 2023 under state open records laws; costs varied by state (Michigan $23, Pennsylvania $20, Georgia $250). A full list of US voter data accessibility by state is maintained by the National Conference of State Legislatures at ncsl.org/elections-and-campaigns/access-to-and-use-of-voter-registration-lists. Under-18 population share from the US Census Bureau. Georgia exact-match background via Bloomberg, 15 October 2018. This note accompanies the interactive showcase at tilores.io/resources/showcases/us-voter-fraud-analysis. Figures describe duplicate registrations identified by probabilistic matching and are not assertions of unlawful conduct by any named individual.