Fraud Finders

HPC in the City 2020
Index

- Problem
- Team Introduction
- Objectives
- Methods
- Results
- Challenges
- Future Work
- Conclusion
- Sources
Voter Fraud

This election is during a pandemic which resulted in an excessive amount of votes through mail-in ballots.

This particular topic was chosen because of this year’s election, based on questionable evidence of voter fraud:

● The number of voters exceeded 100% in some states/counties.
● Voting fraud is rare but real.
● Blatant misinformation regarding voter fraud in 2020 election from both sides.
● The election was close and is being contested.
New Judicial Watch Study Finds 353 U.S. Counties in 29 States with Voter Registration Rates Exceeding 100%

The Hill @thehill - 55m
GOP Georgia Lt. Gov. Geoff Duncan: “At this point we’ve not seen any sort of credible examples” of voter fraud or voter disenfranchisement.

One Michigan county clerk caught a glitch in tabulation software so they hand counted votes and found the glitch caused 6,000 votes to go to Biden + Democrats that were meant for Trump and Republicans. 47 MI counties used this software. All must check now!

MISLEADING

Confessions of a voter fraud: I was a master at fixing mail-in ballots

Sidney Powell: “We’ve Identified 450,000 Ballots that Miraculously ONLY have a Vote for Joe Biden”

With Universal Mail-In Voting (not Absentee Voting, which is good), 2020 will be the most INACCURATE & FRAUDULENT Election in history. It will be a great embarrassment to the USA. Delay the Election until people can properly, securely and safely vote???
Our Team
Our Team

Mentor

Cole Mcknight

Members

Ronesha Shaw

Frederick Morris

Carlos Miranda
Objectives
Objectives

- Analyze registration and ballot data for 2020 to find evidence supporting or debunking claims of voter fraud.
- Compare results with data from previous elections.
- Create Slurm batch job to query website and get html output.
- Clean html output and organize into dataframe.
Methods
Dataset

Dataset: Pennsylvania Full Voter Export - 5.8GB

- [https://www.pavoterservices.pa.gov/pages/PurchasePAFULLVoterExport.aspx](https://www.pavoterservices.pa.gov/pages/PurchasePAFULLVoterExport.aspx)
- Statewide voter registration data.
- 153 features: Names, Active/Inactive, Voting History, DOB, County, Address, Registration Date, etc
Tools

- PA Ballot Tracker - get current ballot info for a record
- Globus - transfer input data to cluster
- Google Colab - develop python scripts
- CloudyCluster - dynamic resource allocation in GCP
- Slurm/CCQ - HPC schedulers to submit batch jobs
- Formfind - finds form data in HTML output
- jupyter notebook
- Curl - query and pull HTML data from website
- Python/Pandas - load, clean, explore input
- Bash - develop scripts to generate and submit batch jobs
Your Mail-in or Absentee Ballot status can be tracked by completing the fields below. You cannot use the tracker to track the status of a ballot voted in person on Election Day.

First Name (as it appeared on your application)

BEATRICE

Last Name (as it appeared on your application)

HARNESS

Date of Birth (mm/dd/yyyy)

05/31/1916

County

ADAMS

Submit

Your Ballot Status Result(s)

<table>
<thead>
<tr>
<th>Ballot Type</th>
<th>Election</th>
<th>Application Received</th>
<th>Application Processed</th>
<th>Ballot Mailed On</th>
<th>Ballot Received</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIL-IN</td>
<td>2020 GENERAL ELECTION</td>
<td>10/19/2020</td>
<td>10/19/2020</td>
<td>10/23/2020</td>
<td></td>
<td>Pending</td>
</tr>
</tbody>
</table>
Workflow - Input

For each county’s FVE file - 67 jobs

- Processed in sequence.
Workflow - Query Batch Job Creation

For each county’s “old” and “active” datasets - 134 jobs

```
while read cnty; do
echo "$cnty"
/bin/su -c "cat <<'EOF'> $cnty"${1}".batch
```

- List of Counties (counties.txt)
- Batch Template Script (create-batch.sh)
- Job using dataset of active voters over 100 ($COUNTY-old.batch)
- Job using dataset of active voters ($COUNTY-active.batch)
Workflow - Query Batch Job

- Query PA ballot tracker with registration data
- Pull HTML output of ballot data

FVE input dataset ($COUNTY-$TYPE.csv) → Query and Pull HTML (get-html-$TYPE.sh) → HTML output for single record ($NAME-$DOB-$COUNTY.html)
Clean/Aggregate Batch Job Creation

For each county’s “old” / “active” HTML output datasets - 134 jobs

```
while read cnty; do
  # for each county
  echo "$cnty"
  /bin/su -c "cat <<'EOF'> "$cnty"-clean-aggr-"$l".batch # create batch script"

# !/bin/bash
#SBATCH -N 1
#SBATCH --mem=4G
#SBATCH --ntasks-per-node=1
#SBATCH -e /mnt/orangefs/team3/jobs/work/slurm-%j.err
#SBATCH -o /mnt/orangefs/team3/jobs/work/slurm-%j.out
#SBATCH -D /mnt/orangefs/team3/jobs/work/jobs
```

- List of Counties (counties.txt)
- Batch Template Script (create-batch-clean.sh)

Job using dataset of active voters over 100
($COUNTY-clean-aggr-old.batch)

Job using dataset of active voters
($COUNTY-clean-aggr-active.batch)
Workflow - Clean/Aggregate Batch Job

For each record:

- Trim HTML output to get ballot data
- Aggregate ballot data into single file for each county.

HTML output directory
(*-$COUNTY-$TYPE.html)

Trim each HTML output file
(trim-html.sh)

Aggregate all trimmed files into single .csv data table
(aggr-data.sh)

Ballot data final output table
($COUNTY-$TYPE-ballot-out.csv)
Challenges:

- Retrieving/formatting valid curl query
- Bypassing reCaptcha
- Underestimating run times of larger county datasets (e.g., Philadelphia county)
Results
Results

9,117,518 FVE records

67 County FVE datasets

352 batch jobs

8,293,578 queries to PA Ballot Tracker

134 final output ballot data tables
Statistics

Pennsylvania:

- 2020 Eligible voters (citizens >= 18): 9,810,201
- 2020 Active voters: 8,291,014
- 2020 Registration rate: ~85%
- 2020 Active voters > 100 years old: 2564
- 2020 Mail-in ballots > 100 years old: 967
- 2020 Accepted mail-in votes > 100 years old: 814
- 2020 Mail-in ballots: (in-progress)
- 2020 Accepted mail-in votes: (in-progress)
Future Work
Future Work

- Continue analyzing FVE and ballot data.
- Run on other states’ FVE data.
- Develop more test cases to detect fraud.
- Develop classification model using machine learning to find records suspected to be fraudulent.
- Use model to identify significant features in detecting voter fraud.
Conclusion
Conclusions

We found no evidence supporting major voter fraud in PA.

Future investigation is needed in PA and other states.

PA has good policy regarding election security.

Good policy is necessary at a federal level.

Voter data should be free and accessible to the public.
Discussion
Sources

FVE dataset:

https://www.pavoterservices.pa.gov/pages/purchasepafullvoterexport.aspx

PA Voting/Election stats:


PA Census Data:

https://data.census.gov/cedsci/all?q=PA

Voter List Access Laws: