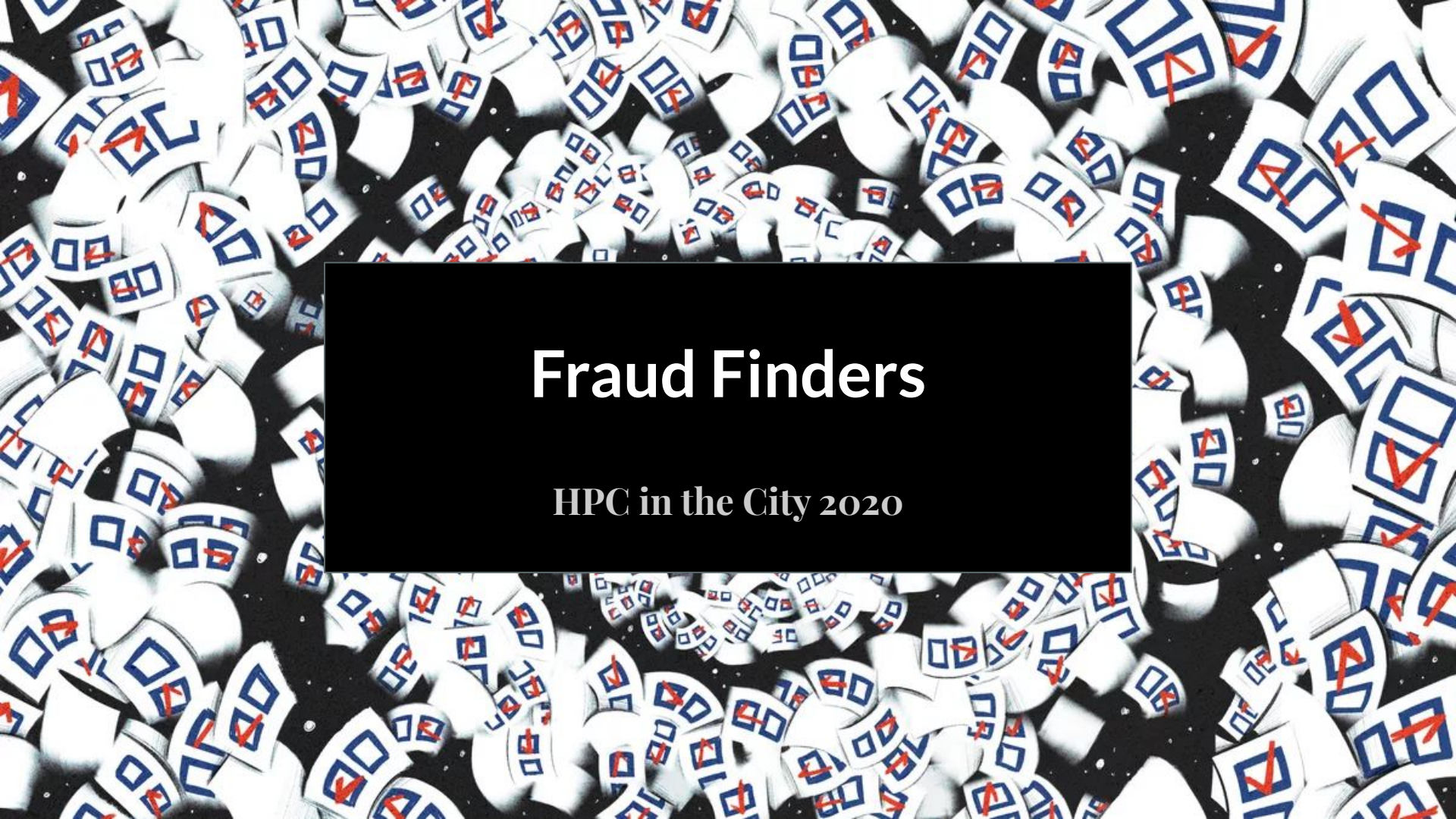


Fraud Finders

HPC in the City 2020

Index

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

BALLOT
BOX
VOTER
FRAUD

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.

OCTOBER 16, 2020 | JUDICIAL WATCH

New Judicial Watch Study Finds 353 U.S. Counties in 29 States with Voter Registration Rates Exceeding 100%

Pinned Tweet

Donald J. Trump ✓
@realDonaldTrump

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???

8:46 AM · Jul 30, 2020 · Twitter for iPhone

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.

69.2K Retweets

US elections · LIVE

Voter fraud of any kind is exceedingly rare in the US, election experts confirm

With ballot counting continuing and the presidential race being called for Joe Biden, experts and officials say there has been no evidence of widespread voter fraud in the 2020 US election. The election process is secure and voter fraud of any type is incredibly rare according to The Associated Press and Reuters. Officials and experts warn that the most interference in US elections, whether from foreign or domestic players, comes in the form of misinformation campaigns, many of which are intended to create distrust in the US's electoral process.

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!

NEWS EXCLUSIVE

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

By Jon Levine

August 29, 2020 | 5:24pm | Updated

Sidney Powell: "We've Identified 450,000 Ballots that Miraculously ONLY have a Vote for Joe Biden"

814,083 views • Nov 8, 2020

18K 922 SHARE SAVE

Our Team
© 2020 MARVEL

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

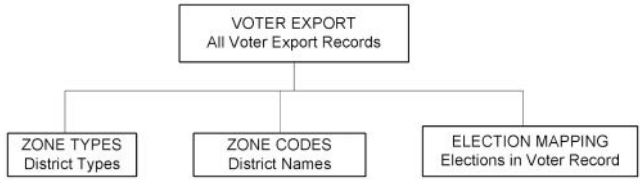
Dataset

Dataset: Pennsylvania Full Voter Export - 5.8GB

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

Voter Export Overview

These Voter data files contain the complete files of all Voter data including District Zone types within the district fields, Zone Codes within the district fields that link to district names and up to 40 elections found in the Voter Export record.



The attached tables for each file type indicate Field Description, Field Type and Comments.

```
df = pandas.read_csv("/content/drive/My Drive/Colab Notebooks/HPC in the City/ADAMS FVE 20201102.txt", sep='\t', lineterminator='\r', names=result)
df
```

```
/usr/local/lib/python3.6/dist-packages/IPython/core/interactiveshell.py:2718: DtypeWarning: Columns (12,21,24,29,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,
  interactivity=interactivity, compiler=compiler, result=result)
```

	ID Number	Title	Last Name	First Name	Middle Name	Suffix	Gender	DOB	Registration Date	Voter Status	Status Change Date	Party Code	House Number	House Number Suffix	Street Name	Apartment Number	Address Line 2	
0	001003958-01	NaN	WEIKERT	MARK	K	NaN	M	11/20/1932	01/01/1955	A	01/11/2017	R	13	NaN	FULTON DR	NaN	NaN	NEW OXF
1	\n"001006201-01"	NaN	STOUFFER	JOHN	C	NaN	M	02/03/1927	01/01/1959	A	09/06/2017	D	6375	NaN	CHAMBERSBURG RD	#105	NaN	FAYETTEVI
2	\n"001006466-01"	NaN	ADKINS	ELLEN	E	NaN	F	09/06/1920	01/01/1960	I	07/10/2018	R	2990	NaN	CARLISLE PIKE	NaN	NaN	NEW OXF
3	\n"001008846-01"	NaN	SMITH	JACK	A	NaN	M	04/08/1934	01/01/1964	I	07/31/2020	R	1639	NaN	COON RD	NaN	NaN	ASPI
4	\n"001010053-01"	NaN	RISS	JEAN	A	NaN	F	10/06/1940	01/01/1965	A	09/01/2015	R	2304	NaN	CHAMBERSBURG RD	NaN	NaN	BIGLERVI
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
72334	\n"111157206-01"	NaN	WARREN	FREDERICK	S	III	M	05/04/1972	10/28/2020	A	10/28/2020	NF	191	B	GREENBRIAR RD	NaN	NaN	Y SPRI
72335	\n"111158380-01"	NaN	DECAVAIGNAC	LINDSEY	PAYTON	NaN	F	10/12/2001	10/29/2020	A	10/29/2020	D	300	NaN	N WASHINGTON ST	0567	NaN	GETTYSBU
72336	\n"111158407-01"	NaN	DIPRETE	CHARLOTTE	CALAT	NaN	F	03/09/1999	10/29/2020	A	10/29/2020	NO	42	NaN	E LINCOLN AVE	NaN	NaN	GETTYSBU
72337	\n"111160304-01"	NaN	WASSUTA	MICHAEL	NaN	NaN	M	02/22/1952	10/30/2020	A	10/30/2020	NF	2490	NaN	CHAMBERSBURG RD	NaN	PO BOX 4865	GETTYSBU
72338	\n	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

72339 rows × 153 columns

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

DEPARTMENT OF STATE

English Español HOME OTHER LINKS

Election Ballot Status

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

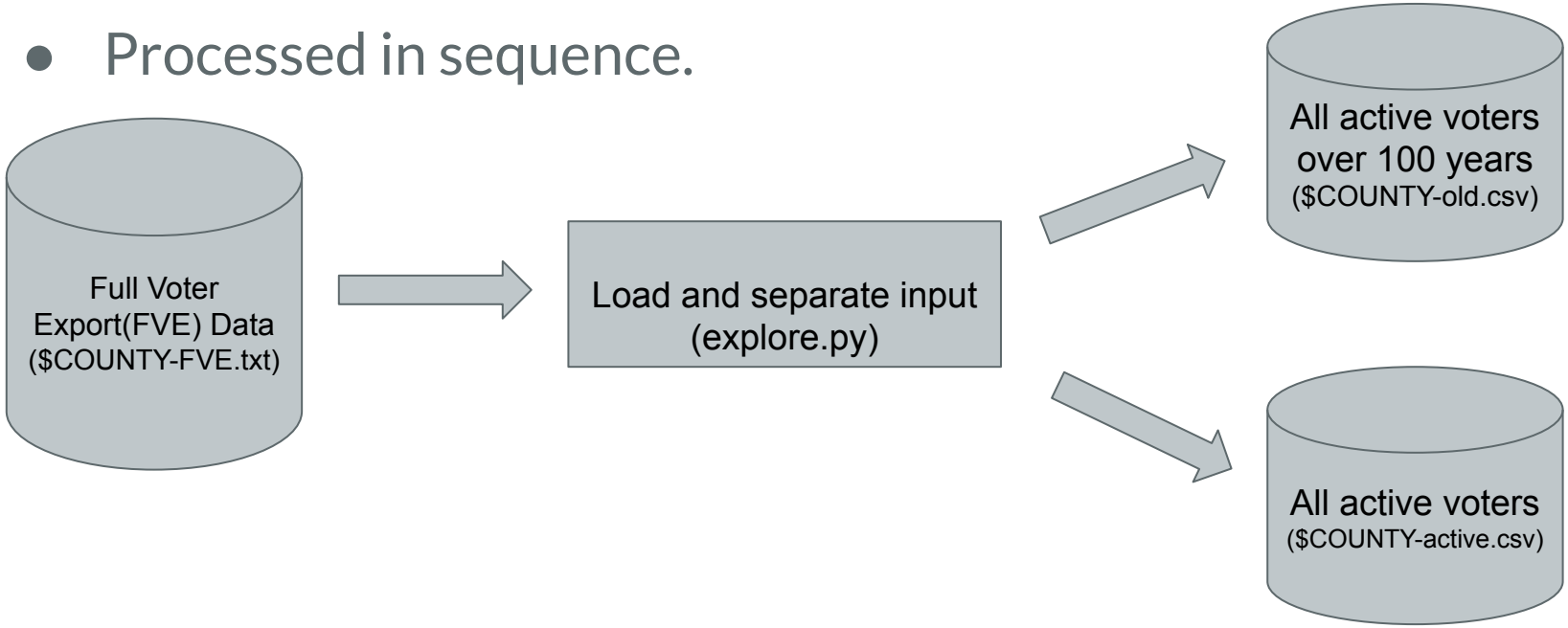
Your Ballot Status Result(s)

Ballot Type	Election	Application Received	Application Processed	Ballot Mailed On	Ballot Received	Status
MAIL-IN	2020 GENERAL ELECTION	10/19/2020	10/19/2020	10/23/2020		Pending

Workflow - Input

For each county's FVE file - 67 jobs

- Processed in sequence.

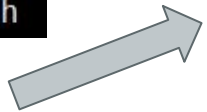


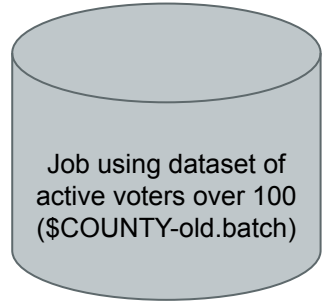
Workflow - Query Batch Job Creation

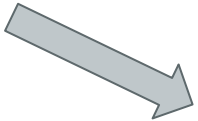
For each county's "old" and "active" datasets - 134 jobs

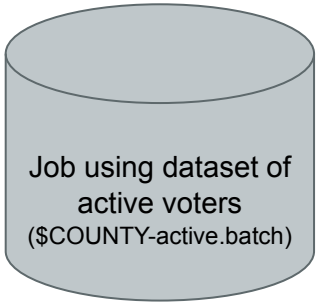
```
while read cnty; do
  echo "$cnty"


  /bin/su -c "cat <<'EOF'> $cnty"${1}".batch
```

Batch Template Script (create-batch.sh)





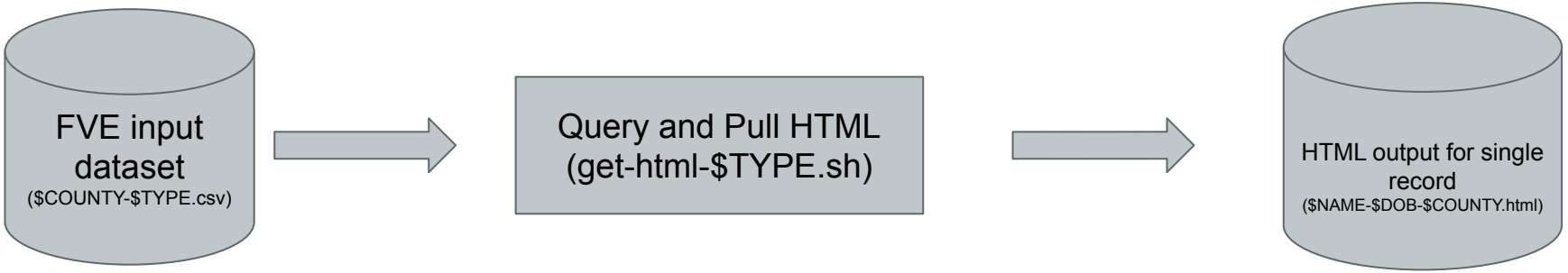





```
#!/bin/bash
#SBATCH -N 1
#SBATCH --mem=4G
#SBATCH --ntasks-per-node=1
#SBATCH -e /mnt/orangefs/team3/jobs/work/slurm-%j.err
#SBATCH -o /mnt/orangefs/team3/jobs/work/slurm-%j.out
#SBATCH -D /mnt/orangefs/team3/jobs/work/jobs
```

Workflow - Query Batch Job

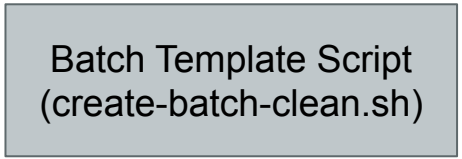
- Query PA ballot tracker with registration data
- Pull HTML output of ballot data

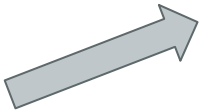


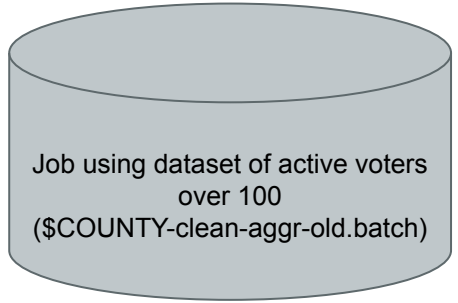
Clean/Aggregate Batch Job Creation

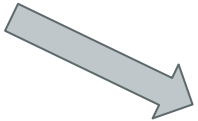
For each county's "old" /"active" HTML output datasets - 134 jobs

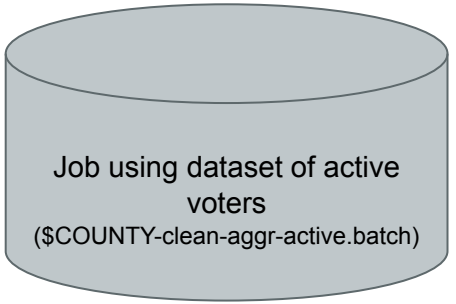
```
while read cnty; do              # for each county
  echo "$cnty"


  /bin/su -c "cat <<'EOF'> "${cnty}"-clean-aggr-"${1}".batch    # create batch script
```









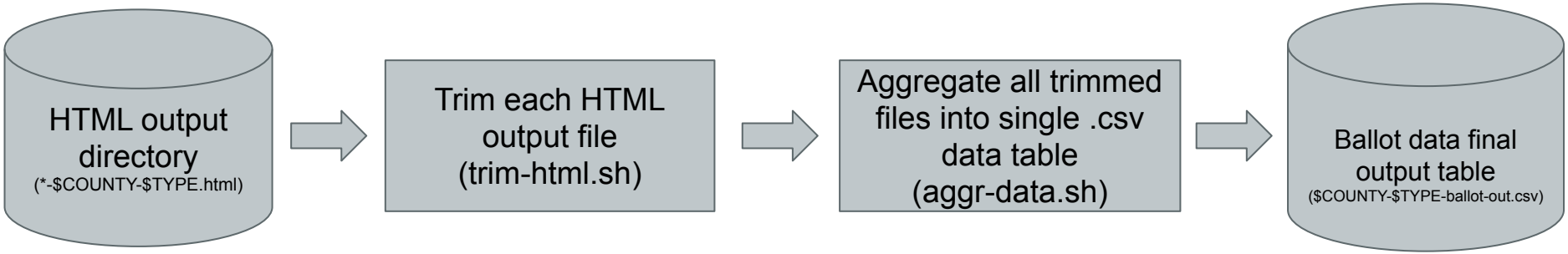



```
#!/bin/bash
#SBATCH -N 1
#SBATCH --mem=4G
#SBATCH --ntasks-per-node=1
#SBATCH -e /mnt/orangefs/team3/jobs/work/slurm-%j.err
#SBATCH -o /mnt/orangefs/team3/jobs/work/slurm-%j.out
#SBATCH -D /mnt/orangefs/team3/jobs/work/jobs
```

Workflow - Clean/Aggregate Batch Job

For each record:

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



Challenges

Challenges:

- Retrieving/formatting valid curl query
- Bypassing reCaptcha
- Underestimating run times of larger county datasets(ex. 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

© 2019 MARVEL

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

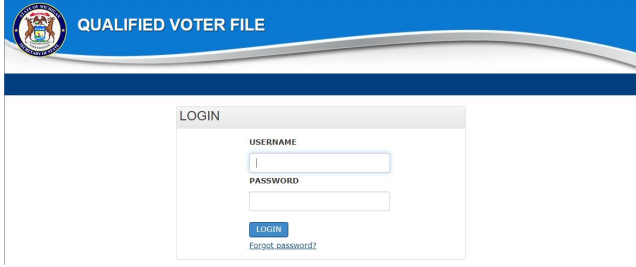
We found no evidence supporting major voter fraud in PA.

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.



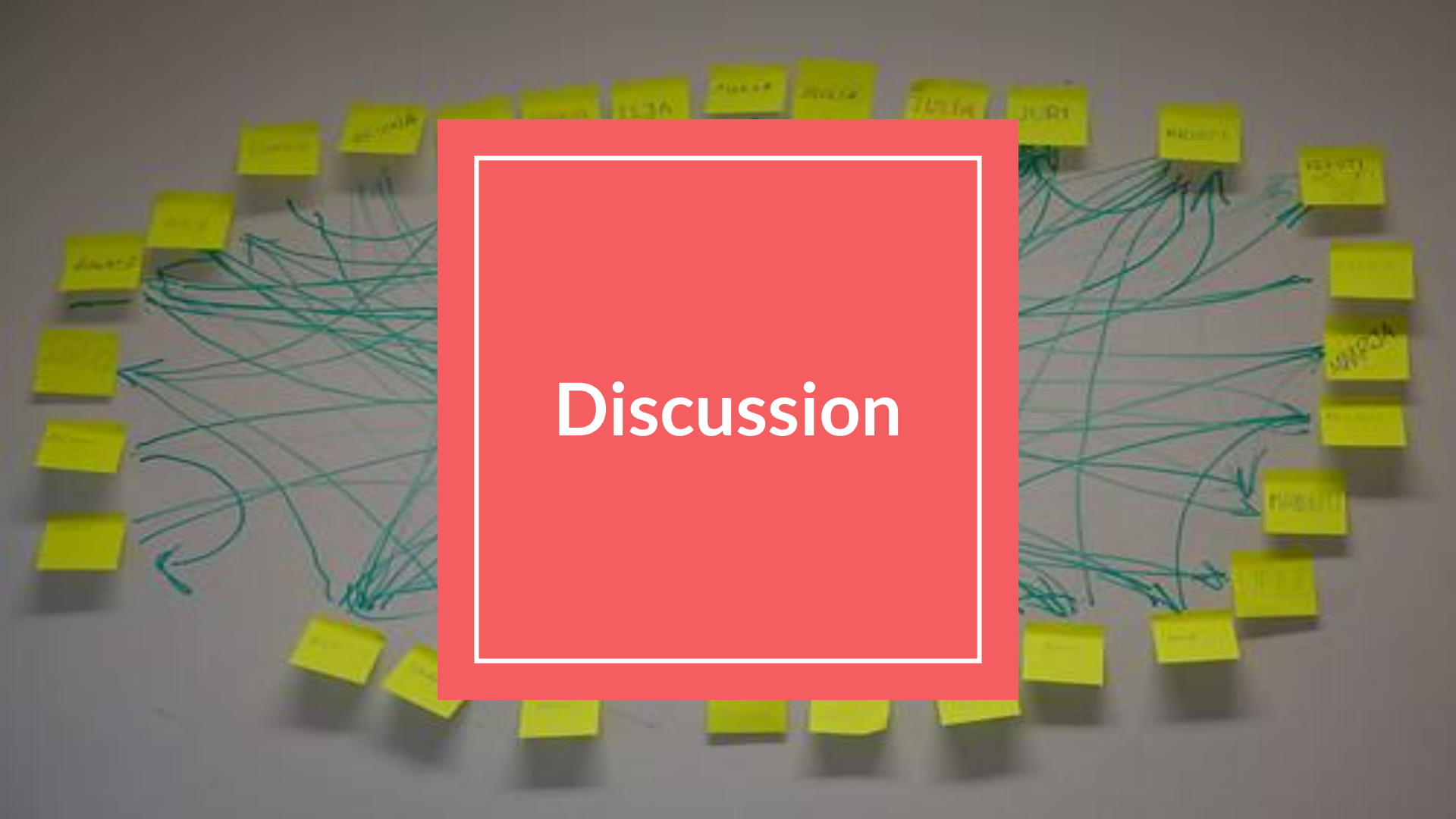
Michigan M.C.L.A. 168.522 168.509gg	Any person may request the voter file.

PRICING

Voter Registration List Pricing Policy (Adopted December 23, 2016)

Statewide Voter List	$250.00

Georgia Ga. Code Ann., § 21-2-225 § 21-2-225.1	Available to the public.


Discussion

Sources

FVE dataset:

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

PA Voting/Election stats:

https://www.dos.pa.gov/VotingElections/OtherServicesEvents/VotingElectionStatistics/Pages/VotingElectionStatistics.aspx

PA Census Data:

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

Voter List Access Laws:

https://www.ncsl.org/research/elections-and-campaigns/access-to-and-use-of-voter-registration-lists.aspx