MITRE

MP210086
MITRE PRODUCT

Data Analytics to Enhance Election Transparency

McLean, VA

Authors:
Hyeyon Bastian
Emily Frye
Caroline Gary
Drew Houck
Marc Schneider
Frank Thomason
Brandon Werner

February 2021

Executive Summary

The 2020 United States presidential election occurred during a time when public opinion about election integrity involved levels of concern that were not customary in the nation. To help ensure public trust and confidence by providing some additional transparency, MITRE's National Election Security Lab gathered and analyzed a wide range of relevant data. The team, dubbed "Bald Eagle," operated between October 2020 and early January 2021. Bald Eagle researched several topics and data sources related to the 2020 presidential election in eight swing states. Bald Eagle conducted analysis on data available before Election Day, including data related to early voting, early voting methods, and return and rejection rates of early ballots. The team obtained data from several election and state government websites, organized the data across states into a similar schema, and developed dashboards to visualize the data and related analytical results. After Election Day, the team continued to conduct analysis on election result data and exit poll data. The following section describes the anomaly allegations, data analysis conducted, and associated results.

Throughout the period of analysis and across the variety of techniques used, evidence of fraud or compromise was not found. This report and the conclusions below summarize the team's findings in five key areas:

- Bald Eagle used two statistical analysis methods to research the allegations of ballot harvesting in Georgia. The first compared the number of ballots requested by mail to the number of ballots returned for each county in Georgia. The second examined addresses in Georgia that received multiple absentee ballots. After researching ballot return rates and addresses with multiple accepted ballots in Georgia, the team found no anomalous points through its analysis. There were no suspicious indicators of ballot harvesting. (See Figure 1 and Table 2 for analysis results.)
- A fingerprinting analysis technique derived from Klimek et al. [1] graphs the percentage of votes for the winner vs. the percentage turnout in a district. The resulting graph is a "fingerprint" of the election that can be used to detect certain types of electoral problems, including vote flipping, ballot box stuffing, and ballot harvesting. Using election fingerprinting for the counties in Georgia, the team saw no signs of vote flipping or ballot box stuffing. Statistical evidence derived using a quantitative technique, which is similar to fingerprinting, supports a finding that these analytic techniques do not demonstrate fraud. (See Figure 4 and Figure 5 for analysis results.)
- Widespread vote manipulation by Dominion machines would likely be reflected in the election results data in the form of unexpected spikes in Democratic support that appear in counties using Dominion machines. To determine if any anomalous increases in Democratic support occurred in these counties, the team compared the election results data provided by eight key states in the 2016 and 2020 presidential elections. Bald Eagle split the data into two sets – one set consisting of the 318 counties that used Dominion machines, and one set consisting of the 333 counties that did not, as determined by data from the Verified Voting Foundation [2]. Researching allegations of vote flipping with Dominion voting machines, the team found there was no statistical difference in the results from ballot machines from different vendors. The claim that Dominion machines artificially inflated results to Joseph Biden, Jr.'s advantage is not supported in the election results data. (See Table 7 and Figure 8 for analysis findings.)

- After allegations that more than 100,000 votes were fraudulently dumped for Biden in Michigan, the team downloaded preliminary results from the Michigan Secretary of State's website and compared each county's preliminary results to the 2016 results. Antrim County appeared as an outlier because its support for the Democratic candidate appeared during a brief period of time to have nearly doubled from 2016 to 2020. After investigating Antrim County in Michigan for data anomalies that created temporary unusual increases in votes for Mr. Biden, the team found that the issues were caused by user error, which was discovered and corrected quickly and did not have an impact on the official election results. (See Figure 17 for analysis findings.)
- In October 2020, local news out of Pittsburgh published an article about the Butler County Elections Director stating that the United States Postal Service had lost an unknown number of mail-in ballots [3]. In the days leading up to the statement released by the Butler County Elections Director, the mail-in ballot return rate for Butler County, PA (according to data collected from the Election Returns for Pennsylvania website [4]) was significantly lower than the other counties in Pennsylvania. The team conducted iterative and daily analysis of the return rates, and although many requested mail-in ballots were lost in Butler County, PA, the overall return rate of ballots leading up to Election Day fell in the expected range of all other counties in Pennsylvania. No evidence of nefarious or fraudulent activities was identified. (See Figure 19 for analysis findings.)

In summary, multiple types of analysis found no evidence of fraud, manipulation, or uncorrected error in the eight states included in this research.

Table of Contents

List of Figures

List of Tables

1 Background and Approach

The 2020 United States presidential election occurred during a time when public opinion about election integrity involved levels of concern that have not been customary in the nation [5] [6] [7]. To help ensure public trust and confidence by providing additional transparency, MITRE's National Election Security Lab gathered and analyzed a wide range of relevant data. The team, dubbed "Bald Eagle," operated between October 2020 and early January 2021. Bald Eagle researched several topics and data sources related to the 2020 presidential election in eight swing states. This research focused on the Republican Candidate, President Donald Trump, and the Democratic Candidate, Joseph Biden, Jr. Throughout the period of analysis, our research did not reveal any evidence of fraud or compromise.

The Bald Eagle team focused on understanding voting trend lines (historical and actual), monitoring for anticipated anomalies based on unfolding public perception and new headlines, and performing data analysis on what may have been indicators of unexpected behaviors. The scope of the Bald Eagle research was on eight swing states: Arizona, Florida, Georgia, Michigan, North Carolina, Ohio, Pennsylvania, and Wisconsin. While available data from all of these states were analyzed in this effort, this report focuses on a handful of anomalies. Section 2 describes five areas of analysis related to these swing states and the 2020 election:

1. Alleged ballot harvesting in Georgia.
2. Election fingerprinting in Georgia
3. Alleged irregularities in Dominion voting machines.
4. Voting anomaly in Michigan
5. Lost ballots in Butler County, Pennsylvania

Bald Eagle conducted analysis on data available before Election Day, including data related to early voting, early voting methods, and return and rejection rates of early ballots. The team obtained data from several election and state government websites, organized the data across states into a similar schema, and developed dashboards to visualize the data and related analytical results. After Election Day, the team continued to conduct analysis on election result data and exit poll data. The following section describes the anomaly allegations, data analysis conducted, and associated results.

2 Leveraging Data Analytics to Identify Anomalies

The MITRE team used multiple publicly available data sets for the analysis presented in this section. Data sets included those provided from election websites, state government websites, and the United States Census. Additionally, the team maintained situational awareness of legal and media events occurring between October–December 2020 to monitor any anomalies in the data. This research is summarized in the remainder of this section. Some additional areas of research and analysis are described in Appendix B.

2.1 Alleged Ballot Harvesting in Georgia

2.1.1 What Was Alleged

In November, there were allegations of reported ballot harvesting in multiple states, including Georgia. Ballot harvesting is an illegal form of ballot collection. Ballot collection is the gathering and submitting of completed absentee or mail-in voter ballots [8]. This is an allowable practice in some states in the United States, especially those where voting by mail is more common. Additionally, ballot collection can often provide a facilitated voting option for anyone who may be unable to submit his or her own vote (for example, someone who is incapacitated or in a hospital). However, there are often concerns that ballot collection can turn into the fraudulent activity of ballot harvesting to generate increased votes by improperly completing ballots for a preferred candidate or by destroying ballots presumed to be for a non-preferred candidate. Ballot collecting can take the form of a third party (which could be volunteers or campaign workers) collecting and delivering ballots to a polling location or the offices of election officials.

While some states do not deem ballot collecting illegal, Georgia law states that only the voter, a relative, or an individual residing at the same address can mail or otherwise deliver a ballot to election officials. There are exceptions for individuals who are disabled or confined to a hospital or prison or otherwise detained [8] [9].

2.1.2 Analysis Performed

We used two statistical analysis methods to research the allegations of ballot harvesting in Georgia. The first compared the number of ballots requested by mail to the number of ballots returned for each county in Georgia. The second examined addresses in Georgia that received multiple absentee ballots.

The first analytical approach compared the number of ballots requested by mail to the number of ballots returned for each county in Georgia. If large amounts of absentee ballots were destroyed through ballot harvesting, we would expect the number of ballots returned to be significantly reduced relative to other counties. Alternately, if votes were fraudulently inflated for a candidate, we would expect the number of ballots returned to be higher than the other counties. This finding would indicate that a county had requested many more ballots than it returned or had returned many more ballots than it requested.

Using data collected from the Georgia Secretary of State's website [10], Figure 1 – Log of Georgia Mail-in Ballots Returned vs. Mail-in Ballots Requested plots the number of ballots requested versus returned for all 159 counties in Georgia. Points along the blue line have relatively similar return rates to each other. The logarithm function was used to account for wide variations in population between counties.

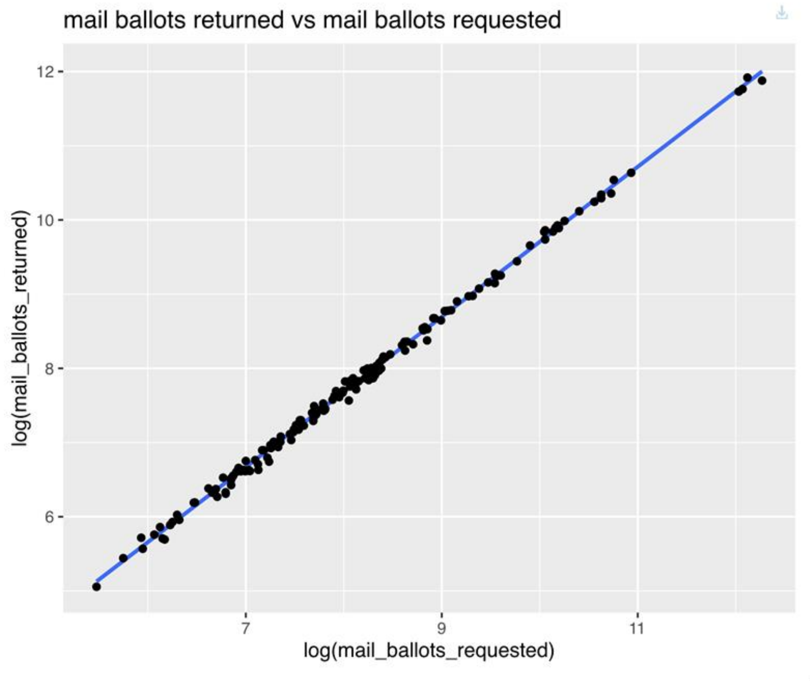

Figure 1 – Log of Georgia Mail-in Ballots Returned vs. Mail-in Ballots Requested

A point located far below the blue line could indicate an anomalous event such as large numbers of ballots being destroyed in a county. Conversely, a point above the blue line could indicate large numbers of ballots being artificially submitted. A point near the blue line indicates the return rate was proportional to the other counties in Georgia.

The lack of any significant deviation from the blue line signifies that there was no indication of widespread ballot harvesting in the Georgia election data.

The second analytical approach related to ballot harvesting in Georgia examined addresses in Georgia that received multiple absentee ballots. Based on analysis of data from the Georgia Secretary of State's website, thirty-eight addresses requested and received more than twenty absentee ballots each [10]. Table 1 – Addresses in Georgia with Multiple Accepted Ballots provides the ten addresses that received the most ballots along with a description of the location.

Table 1 – Addresses in Georgia with Multiple Accepted Ballots

# Accepted Ballots	Mailing Address	County	Description
84	504 COLLEGE DR ALBANY GA	Dougherty	College Dorm (Albany State)
73	2001 S LEE ST AMERICUS GA	Sumter	Senior Center
69	1762 CLIFTON RD NE ATLANTA GA	DeKalb	Emory University
68	342 LOG CABIN RD NE MILLEDGEVILLE GA	Baldwin	Apartments
61	1501 MERCER UNIVERSITY DR MACON GA	Bibb	Mercer University

# Accepted Ballots	Mailing Address	County	Description
61	4255 SMOKECREEK PKWY SNELLVILLE GA	Gwinnett	RV Park
48	1014 CAVE SPRING RD SW ROME GA	Floyd	Darlington Boarding School
47	223 JAMES P BRAWLEY DR ATLANTA GA	Fulton	Clark Atlanta University
43	830 WESTVIEW DR SW ATLANTA GA	Fulton	Morehouse College
40	3219 COLLEGE ST THUNDERBOLT GA	Chatham	Savannah State University

Each of the thirty-eight addresses was analyzed in detail and all were determined to have a plausible reason for receiving multiple ballots. The addresses correlated to facilities such as universities, health care facilities, jails, and senior living communities. The analysis of all thirty-eight addresses is summarized in Table 2 – Address Category of Georgia Addresses with Multiple Accepted Ballots. Because each address had a plausible reason for receiving multiple ballots, the data does not indicate that there was widespread harvesting by mail.

Table 2 – Address Category of Georgia Addresses with Multiple Accepted Ballots

Address Category	Number with 20+ Accepted Ballots
Apartments	3
Boarding School	1
Health Care	6
Homeless Shelter	2
Jail	3
Monastery	1
RV Park	4
Senior Living	8
University	10

2.1.3 Results Identified

The statistical analysis of ballot return rates shows no anomalous points. With the first statistical analysis approach, there were no points far below the blue line that would indicate an anomalous rate of ballots requested to ballots returned (see Figure 1). Therefore, the analysis did not show any suspicious indicators of ballot harvesting. Additionally, the results from both a statewide risk limiting audit and a recount using automated machines were both consistent with the original automated count. The results were certified by the Georgia Secretary of State on November 20, 2020.

With the second analytical approach, ballot harvesting indicators would show that multiple ballots were tied to the same address, but that the address was not affiliated with a facility where multiple eligible voters reside. The analysis of addresses receiving more than twenty absentee ballots revealed that each address had a plausible reason for receiving multiple ballots. There were no potential incidents of ballot harvesting revealed using this analytical method.

2.2 Election Fingerprinting in Georgia

2.2.1 What Was Alleged

In the post-election period, several claims [11] [12] were made that systems made by Dominion Voting Systems and used in Georgia [13] (among other states) were flipping votes from Mr. Trump to Mr. Biden. We investigated these claims using multiple techniques, including the creation of an election "fingerprint" for Georgia, which can be used to detect vote flipping. During the investigation into these claims (see Section 2.3 for additional analysis of Dominion voting machines), six counties in Georgia fell far enough outside of the norm to warrant further detailed investigation. The six specific counties in Georgia that we examined using fingerprinting techniques are Cobb, Douglas, Fayette, Forsyth, Henry, and Rockdale.

A related claim was that ballot counts violated Benford's law [14]. The use of Benford's law was investigated as a method to detect fraud, but we found that using Benford's law to detect election fraud is weak at best, not well-established as an indicator, and not practical to perform correctly at scale. See Appendix C for a detailed discussion of Benford's law.

2.2.2 Analysis Performed

The analysis in this section uses a fingerprinting technique derived from Klimek et al. [1]. This technique graphs the percentage of votes for the winner vs. the percentage turnout in a district. The resulting graph is a "fingerprint" of the election that can be used to detect certain types of electoral problems, including vote flipping, ballot box stuffing, and ballot harvesting. If one of these problems occurs, it will deform the fingerprint. When a large number of precincts are fingerprinted, it is necessary to display the data using a two-dimensional histogram, where the color of the data point represents the number of precincts in the bin represented by that data point. For a discussion of indicators of fair and fraudulent elections in fingerprints, see Appendix A and [1].

Fingerprinting provides insight into fraud because the percentage of votes for the winner as it relates to the percentage of turnout has specific statistical properties, provided some conditions are met. The first condition is the size of the jurisdictions under analysis. The jurisdictions being examined need to have an average population of registered voters ($\bar{n}_{pop}$) between 100 and 5,000 people. If ($\bar{n}_{pop}$) is too small, the behaviors of individuals have too much influence over the point in the fingerprint. If ($\bar{n}_{pop}$) too large, aggregation of jurisdictions can mask problems. In Georgia, the analysis needed to be performed at the precinct level to meet this requirement.

Both Klimek and Borghese/Berchaud [1] [15] show that with proper rescaling of vote rates, the vote rate follows a log-normal distribution. The normalized vote rate for jurisdiction i is v_i, and thus In (v_i) is normally distributed. Visual analysis of an election fingerprint is open to interpretation, but by examining this rescaled distribution we can quantify how well it matches a Gaussian (or normal) distribution. We can do this by examining the skew and kurtosis of the observed distribution.

Skew is a measure of asymmetry around the mean (μ) of a distribution. A skew of zero indicates that the distribution is symmetrical around its mean, and the mean, median, and mode are all equal. A positive skew means that the mean is greater than the median and mode, and a negative skew indicates that the mean is less than the median and mode. A graph of the probability density function (PDF) will "lean" to the right if the skew is negative and will "lean" to the left if the skew is positive. An example of a skewed normal distribution with skew = −0.85 is shown in Figure 2, labeled Skew Normal. The skew of a Gaussian or normal distribution is zero, as the distribution is symmetrical around its mean: the mean, median, and mode are all equal to zero, and the graph of its PDF is symmetrical.

Kurtosis is a measure of the "weight" of the tails. Westfall states in [16] "...[kurtosis's] only unambiguous interpretation is in terms of tail extremity, that is, either existing outliers (for the sample kurtosis) or propensity to produce outliers (for the kurtosis of a probability distribution)." For a Gaussian distribution, kurtosis will always have a value of 3. The logistic distribution shown in Figure 2 has a kurtosis of 4.2. All three distributions displayed have a mean of 0 and a standard deviation of 1.

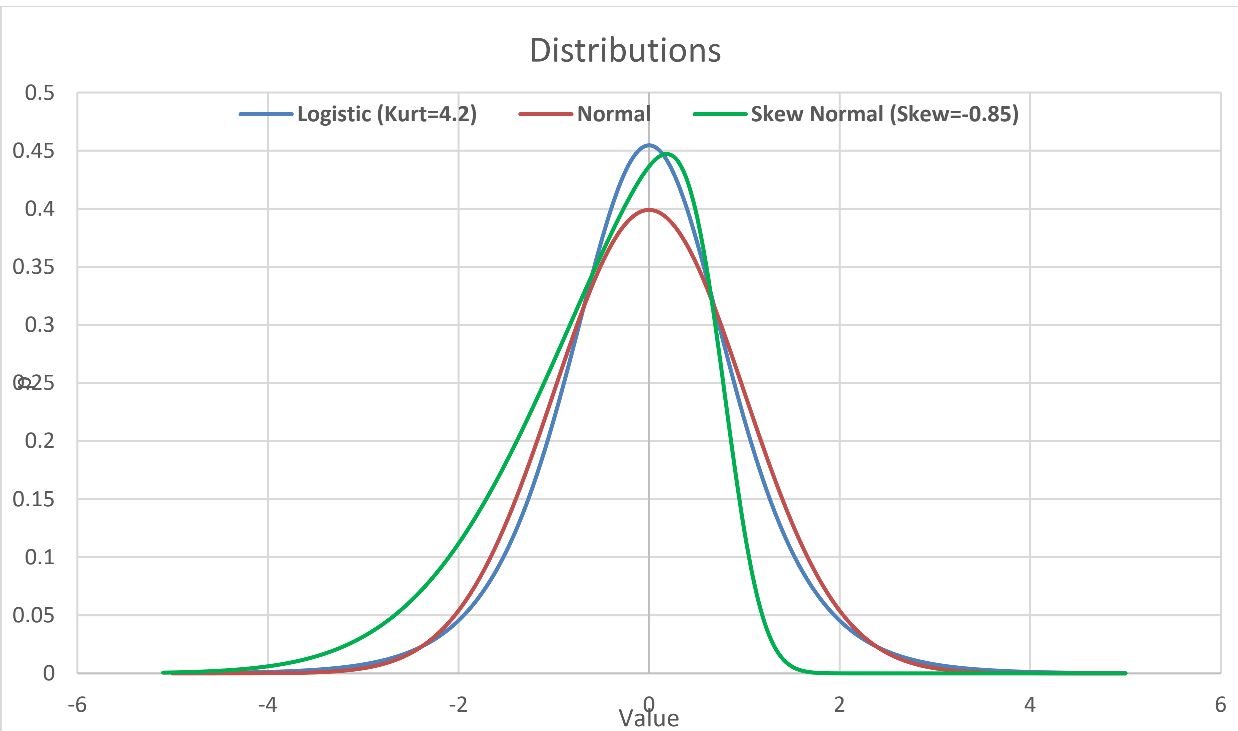

Figure 2 – Probability Distributions

In fair elections, the logarithm of the rescaled vote rate v is approximately normally distributed [1] [15]. The rescaled vote rate (v_i) for a given jurisdiction i, with N_i being the total number of voters in the jurisdiction, and W_i the number of votes for the winner of the election, is given by the following equation:

$$\boldsymbol{v_i = \frac{N_i - W_i}{W_i}}$$

Equation 1 – Rescaled Vote Rate

The rescaled vote rate has a log-normal distribution, thus $\mathbf{ln}(\boldsymbol{v_i})$ is normally distributed. We have calculated the parameters, shown in Table 3, of the observed distribution in Georgia using the rescaled logarithm of the observed values $\mathbf{ln}(\boldsymbol{v_i})$. By examining the skew and kurtosis of the distribution, we can tell how well a Gaussian distribution fits the observed data. A perfect fit would have a skew of 0 and a kurtosis of 3. The Russian presidential election of 2012, which is widely considered to have high levels of fraud [17] [18], had a skew of −2.3 and a kurtosis of 9.7. Canadian parliamentary elections are widely considered to be fair [19] [20], and we used the Canadian parliamentary election of 2011 as a model fair election.[1]

We analyzed all of Georgia's 159 counties and calculated the distribution for the entire state. The statewide data set contained 2,650 precincts with an average precinct size ($\bar{n}_{pop}$) of 2,721. This falls within the values required for fingerprint analysis. The logarithm of the rescaled vote rate v_i (Equation 1) is expected to approximate a Gaussian distribution in a fair election. To understand the observed distribution shown in Figure 3 and the parameters in Table 3 – Statewide Rescaled Normal Distribution, some discussion of Equation 1 is needed.

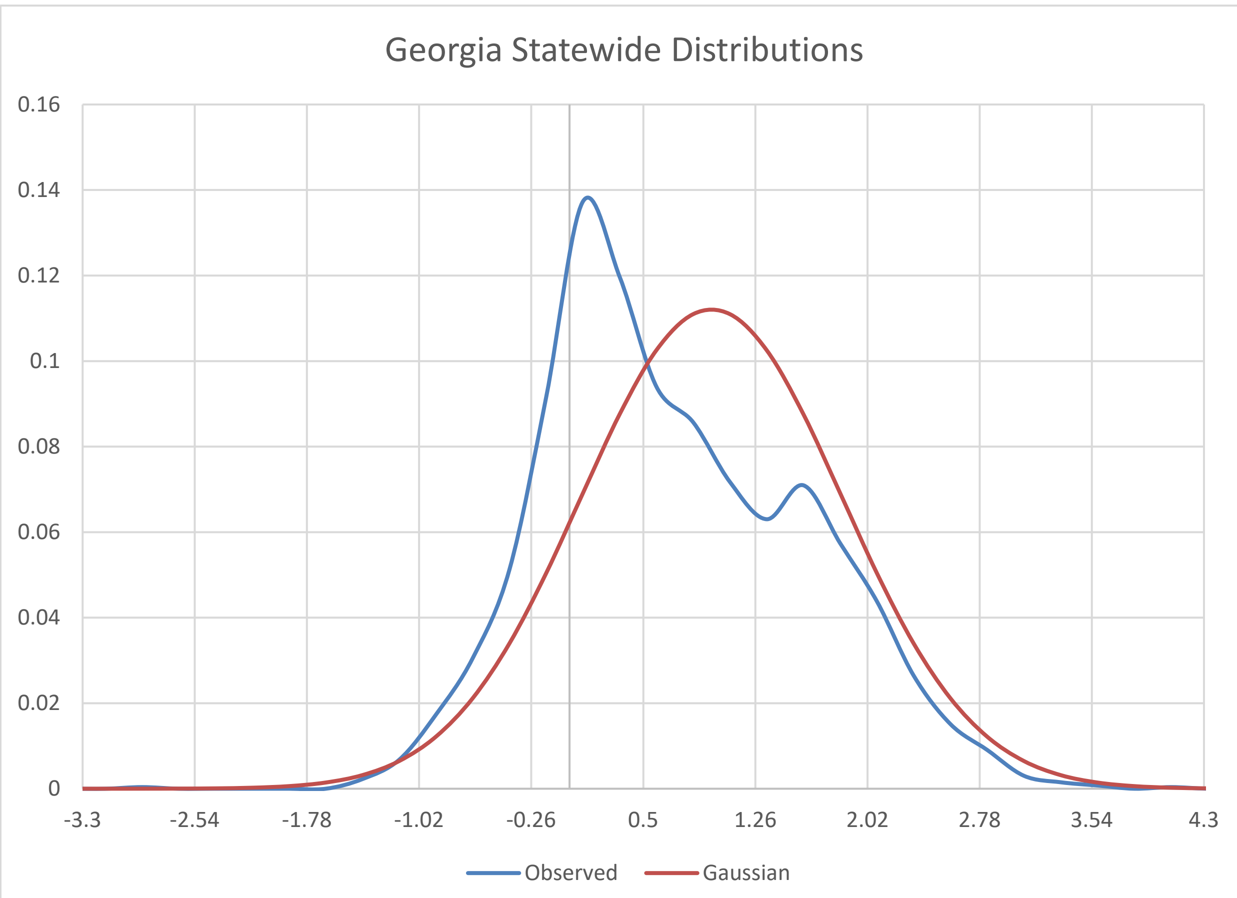

Figure 3 – Georgia Distribution

The value v_i is the ratio of registered voters who did not vote for the winner to the voters who did vote for the winner, in precinct i. A value $v_i > 1$ indicates that there were more voters who either did not vote, or did not vote for the winner, than there were voters who voted for the winner. A value $v_i = 1$ occurs when exactly half of the registered voters in a precinct voted for

1 See Figure 22 for the fingerprint of the Canadian election.

the winner; since the logarithm of 1 is 0 ($\mathbf{ln}(1) = 0$), this is the boundary between precincts where the winner of the election received votes from more than half of all registered voters $\mathbf{ln}(v_i) < 0$, and the precincts where the winner did not receive more than half of all registered voters $\mathbf{ln}(v_i) > 0$.

Table 3 – Statewide Rescaled Normal Distribution

μ	0.84
σ	0.88
Skew	0.36
Kurtosis	2.69

On average, 31% of registered voters did not vote in the presidential election in Georgia, which results in a positive mean, since no candidate received votes from 50% of the registered voters. The skew and kurtosis below 3.0 show that the logarithm of the rescaled vote rate is negatively biased, which can be seen in Figure 3 as the Observed distribution. This indicates that the distribution has a bias towards the winner, which is to be expected since the winner had more votes than their opponents. The low kurtosis indicates that the tails are "lighter" than would be seen in a Gaussian distribution, which is also to be expected since the tails do not go to infinity. As seen in Figure 3-B in [1], the fraudulent elections they examined had high kurtoses (between 6 and 12) for the n_{pop} in the range found in the Georgia precincts.

Looking at the statewide fingerprint shown in Figure 4, we see two clusters: one is centered around 20% of votes for Biden, with 75% turnout; the other cluster is spread along a line at 95% of votes for Biden, with a center of around 60% turnout. We investigated the cluster with the high percentage of votes for Biden and found that all of the precincts were in areas where conventional wisdom indicates there are large numbers of Democrats, thus Biden's percentage of the total vote would be high at any turnout level. These results are consistent with the six-county fingerprint in Figure 5. Because of the large number of precincts, it is necessary to display the statewide fingerprint as a 2-D histogram, but when working with the six-county data, we used a scatter plot where each point represents one precinct. This is equivalent to a 2-D histogram where the background (0 precincts) is white, and the bins are fine enough to have only one precinct per bin. The bins with 1 precinct in them would be colored blue. Because of this difference in display formatting, Figure 4 and Figure 5 are visually very different, but both represent fingerprints.

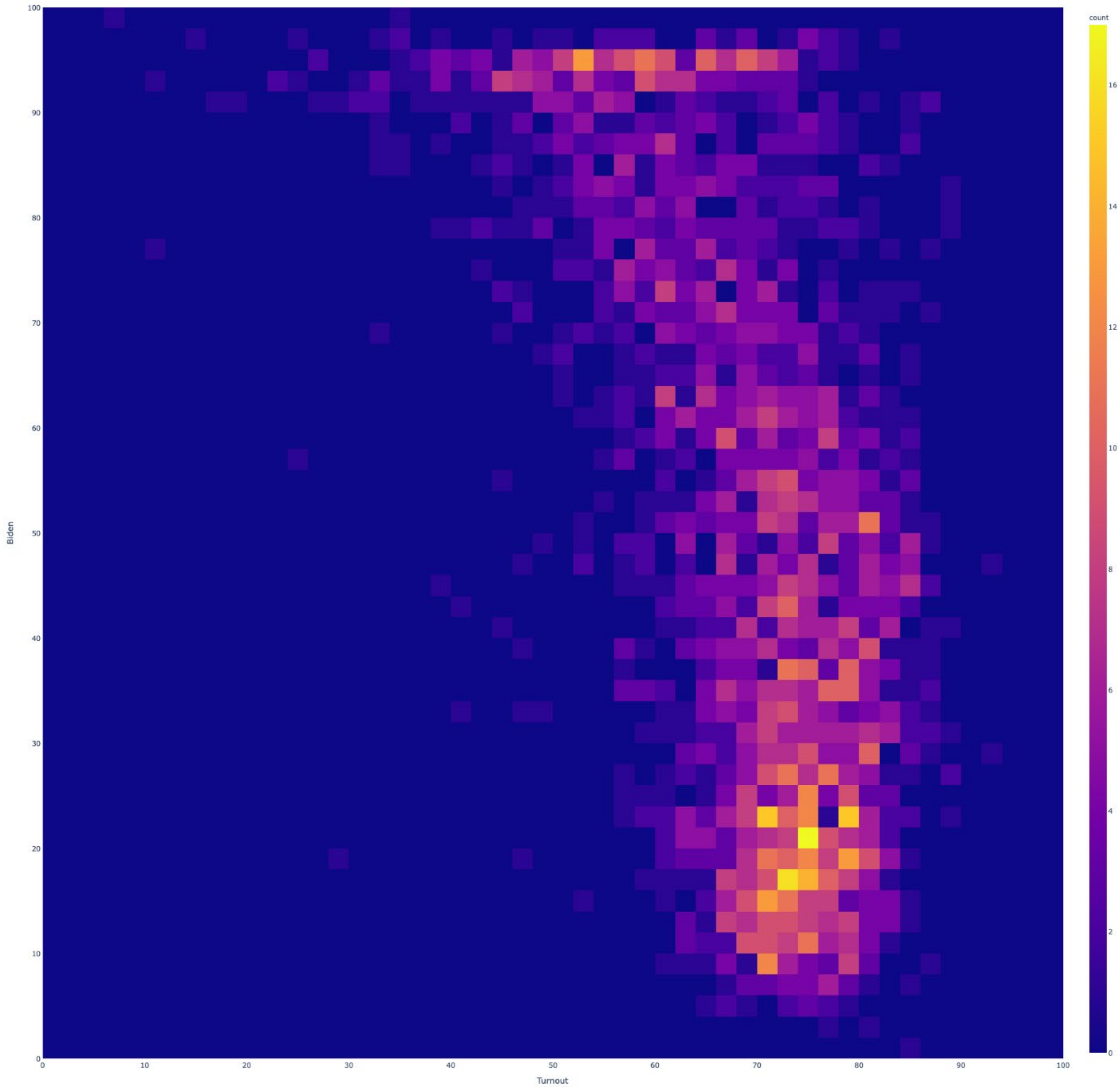

Figure 4 – Statewide Fingerprint of Georgia

2.2.2.1 Further Examination

As mentioned in Section 2.1.1, six counties warranted further examination. For these six counties, there was a total of 279 precincts in the data set, with an average size ($\bar{n}_{pop}$) of 2,971.

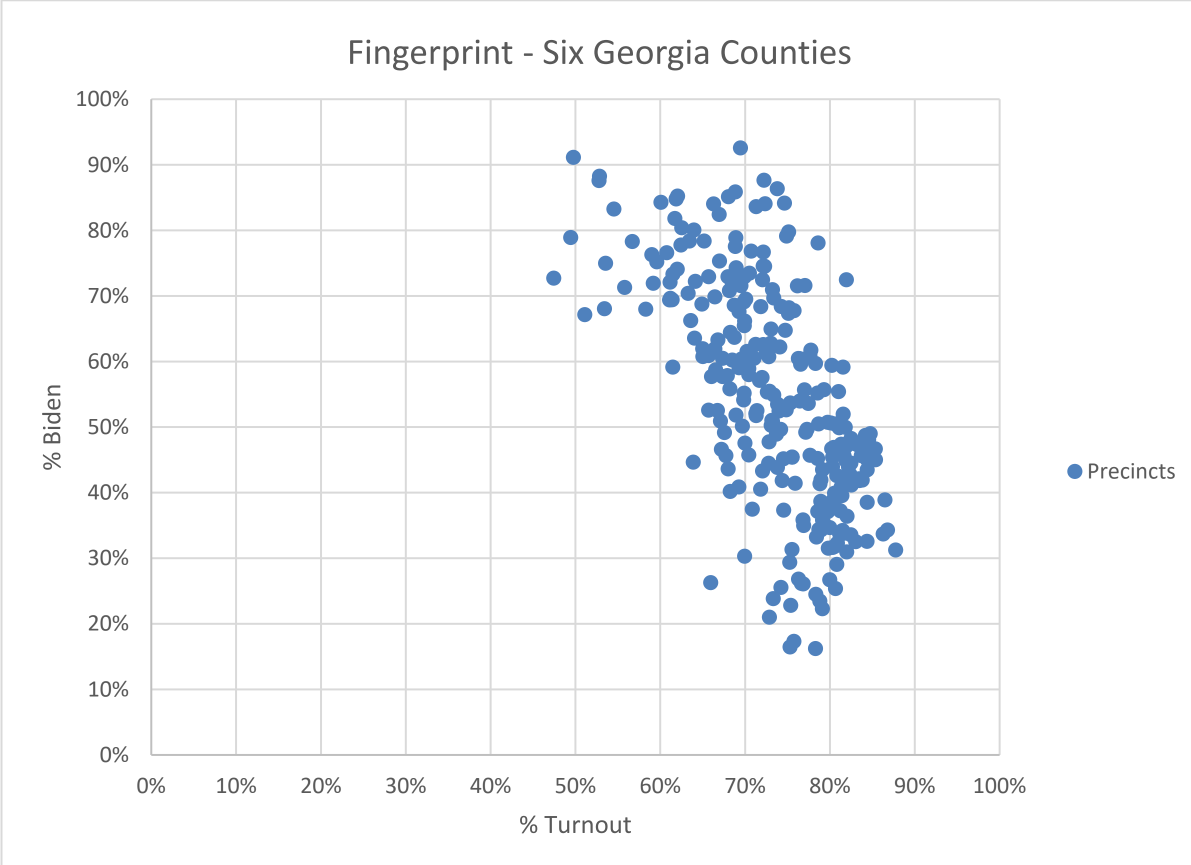

Figure 5 – Fingerprint of Six Georgia Counties

This election had a skew of −0.71 and a kurtosis of 3.4. The values for the six Georgia counties combined, shown in the 3rd and 4th rows of Table 4, are −0.28 and 2.8, respectively.

Table 4 – Parameters of Rescaled Normal Distribution

μ	−0.21
σ	0.77
Skew	−0.28
Kurtosis	2.80

In order to calculate the PDF[2] of the distribution, we first created a histogram using 15 bins, shown in Table 5, with a bin width of 0.35 [21]. This gave us the sample probability of a precinct falling within each bin. All bins not shown in Table 5 had an estimated probability of 0, and the values shown sum to 1. We graphed the estimate of the sample probability of a precinct falling within a particular bin and used interpolation to produce a smooth graph, as seen in Figure 6.

[2] Technically this is a probability mass function (PMF), since we had discrete values for each bin, however we display this as a smooth graph rather than a histogram [21]. In Figure 7, we show a graphical example of how the smooth graph is obtained from a histogram using the data from all precincts in Georgia.

Table 5 – GA Precinct Bin Counts and Sample Probabilities

Bin center value $ln(v)$	Count of precincts in the bin	Sample probability
−2.425	0	0.000
−2.075	2	0.007
−1.725	3	0.011
−1.375	11	0.039
−1.025	14	0.050
−0.675	31	0.111
−0.325	30	0.108
0.025	43	0.154
0.375	59	0.211
0.725	43	0.154
1.075	27	0.097
1.425	12	0.043
1.775	2	0.007
2.125	2	0.007
2.475	0	0.000

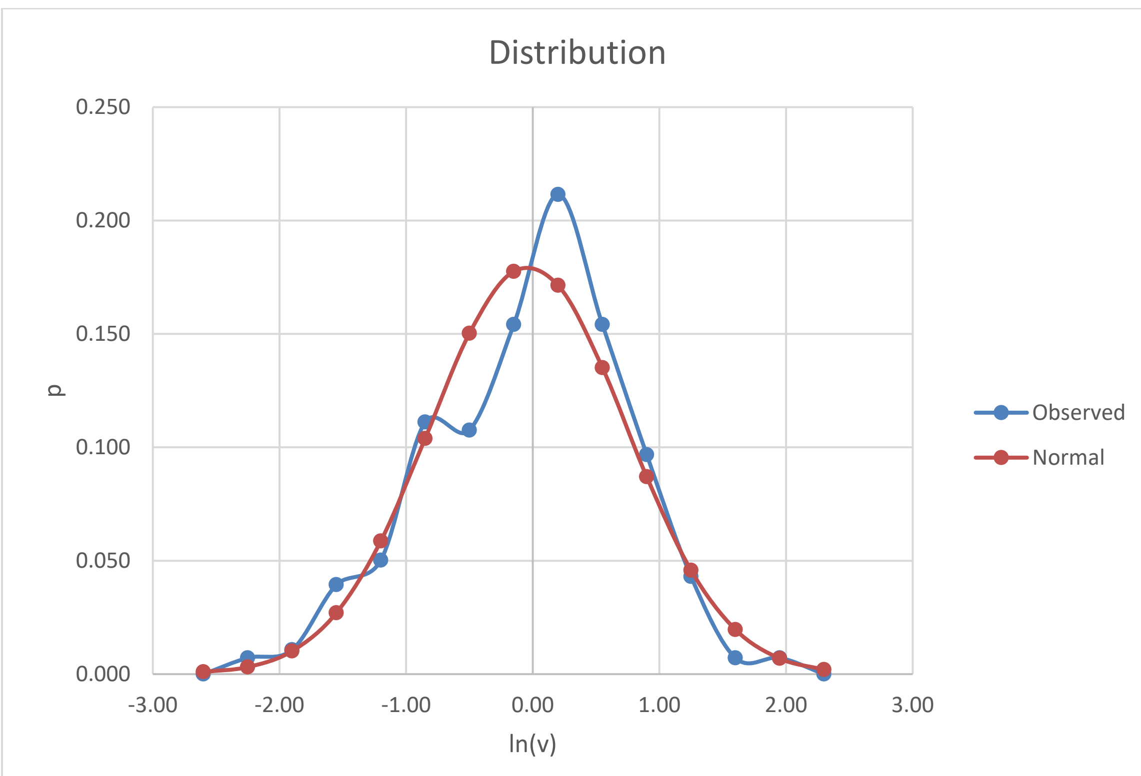

Figure 6 – Observed (GA Six Counties) and Normal Distributions with the Same μ, σ

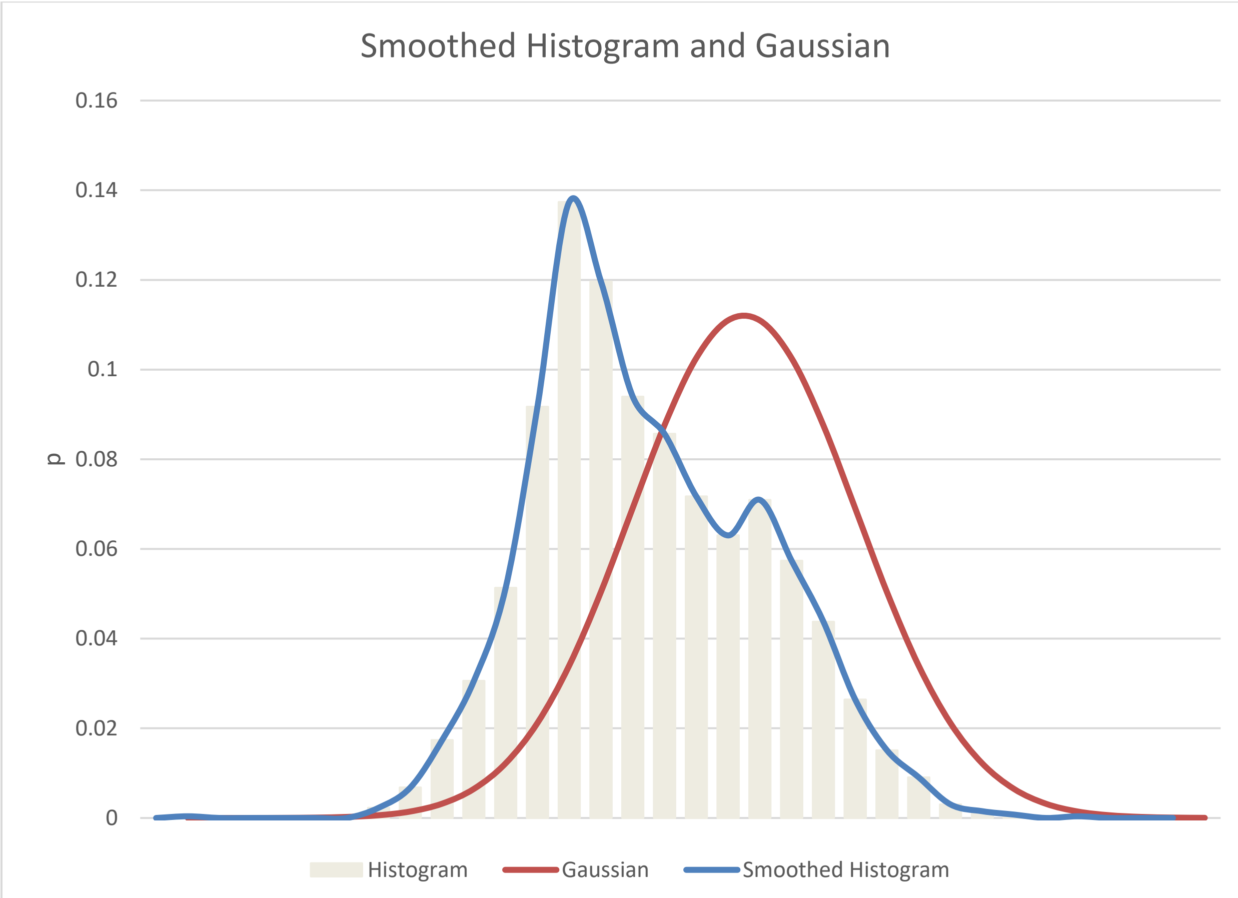

Figure 7 – From Histogram to Smooth Distribution

Since the kurtosis in the observed parameters is less than 3, there is less "weight" in the tails, and the observed (sample) distribution is less likely to produce outliers than a Gaussian distribution. This makes sense, since a Gaussian distribution has tails going to +/- infinity, and the tails of the observed distribution end at −2.6 and 2.3. The fact that the tails are not infinite can be verified by examining the sum of the sample PDF over a specific range. In our case the sum of the sample PDF, which is also referred to as the sample cumulative distribution function (CDF), is 1 for values of $\mathbf{ln}(\boldsymbol{v})$ greater than or equal to 2.3, indicating that there can't be any "additional" probability for values above 2.3; in a graph of this observed PDF, shown in Figure 7, all of the values to the right of 2.3 are zero. Since we start calculating the observed PDF (and sample CDF) at −2.6, which has a probability of 0, and the CDF goes from 0 to 1 over this range, we also know that there is no "additional" probability for values of $\mathbf{ln}(\boldsymbol{v})$ less than −2.6; in a graph of the PDF, all the values to the left of −2.6 are zero.

2.2.3 Results Identified

The statewide analysis and fingerprint of Georgia are consistent with parameters and patterns seen in fair elections.

Neither the six Georgia counties when combined, nor any of the counties examined individually, showed fingerprints that are consistent with the patterns seen in elections where vote flipping, ballot box stuffing, or ballot harvesting have occurred, and all the fingerprints are consistent with fair elections. The patterns seen in the allegedly compromised counties, such as the combined fingerprint in Figure 5, most closely resemble the Swiss parliamentary election of 2008, as shown in [1].

Furthermore, by examining the distribution of logs of rescaled voter rates, we can see, both visually and by examining the skew and kurtosis of the observed distribution, that the distribution is a close fit for a Gaussian distribution, which is to be expected when the election is fair.

The analysis results are consistent with the hypothesis that there was no vote flipping, ballot box stuffing, or ballot harvesting in Georgia, and are not consistent with the widespread claims of fraud that appeared through MITRE's anti-misinformation application SQUINT™ and/or in the media. The voting systems used in Georgia are all the same type, the Dominion Voting ImageCast® X, and this finding (i.e., that analysis shows no evidence of vote flipping or ballot box stuffing) indicates that confidence in these systems is supported by evidence at this time. (See Section 2.3 for additional analysis of Dominion Machines)

This conclusion is consistent with Georgia's risk limit audit ("the hand recount") and machine recount, and with a joint statement from the elections infrastructure government coordinating council and the election infrastructure sector coordinating executive committees.

2.3 Alleged Irregularities in Dominion Machines

2.3.1 What Was Alleged

After the election, there were many allegations made in national media outlets that Dominion Voting Systems machines had switched votes from Donald Trump to Joe Biden. To this date, no evidence of the claims that Dominion Voting machines flipped any votes has been provided to and published by national media outlets [22]. Additionally, social media posts claim that Dominion voting machines deleted large numbers of Trump votes and that states using Dominion Voting Systems machines switched votes from Trump to Biden [23].

2.3.2 Analysis Performed

Widespread vote manipulation by Dominion machines would likely be reflected in the election results data in the form of unexpected spikes in Democratic support. To determine if any anomalous increases in Democratic support occurred, we compared the election results data provided by eight key states in the 2016 and 2020 presidential elections. We split the data into two sets – one set consisting of the 318 counties that used Dominion machines, and one set consisting of the 333 counties that did not, as determined by data from the Verified Voting Foundation [2]. Table 6 shows the number of counties in each state that used Dominion machines.

Table 6 – Number of Counties with Dominion Machines by State

State	Dominion Machines Counties	Non-Dominion Machines Counties
Arizona	1	14
Florida	17	50
Georgia	159	0
Michigan	65	18
North Carolina	0	100
Ohio	18	70
Pennsylvania	14	53
Wisconsin	44	28

The set of counties that did not use Dominion machines was considered a ground-truth data set because these counties would be unaffected by any artificial vote manipulations by Dominion machines. The counties that used Dominion machines were compared to the non-Dominion counties to determine if the data supported any allegations of irregular results.

The metric used for comparison was the change in the percentage of votes for the Democratic candidate from 2020 to 2016. We believed this metric could reflect any widespread manipulation of Dominion voting machines to benefit Biden in 2020 and help identify counties to investigate further.

$$\textit{Change in Democratic Vote Percentage} = (\textit{Biden Vote Pct in } 2020) - (\textit{Clinton Vote Pct in } 2016)$$

As Table 7 shows, the mean change in Democratic support was nearly identical for Dominion and non-Dominion counties. A t-test confirmed that there is no significant difference in the mean of the two sets of data. If Dominion machines were inflating Democratic vote percentages, we would expect the Dominion counties to have a statistically significant greater mean. This finding does not support the allegation of widespread vote manipulation by Dominion machines to benefit Biden.

Table 7 – Change in Democratic Support from 2016 to 2020 for Counties with Dominion Machines and Without Dominion Machines

Change in Democratic Support from 2016 to 2020	Dominion Machine Counties	Non-Dominion Machine Counties
Counties	318	333
Mean	+1.49	+1.51
Median	+1.27	+1.53
Standard Deviation	2.22	2.34

However, there could be some isolated counties with Dominion machines that were manipulating votes. We used two methods to identify anomalous Dominion counties. The first method used a box-and-whisker plot to identify any values that were more than 1.5 times the interquartile range, as seen in Figure 8. This method identified eight counties that were considered outliers because they were above the top "whisker" due to their large increase in Democratic support relative to the other counties. Seven of the counties are in Georgia, and one is in Michigan.

Each dot on the graph represents the difference in vote percentage for the Democratic candidate from 2020 to 2016. A positive value indicates that Biden received a greater percentage of votes in 2020 than Hillary Clinton did in 2016. The color of the dot indicates which candidate won the county in 2020 (Biden is blue, Trump is red), and the size of the dot represents the number of votes recorded in the county.

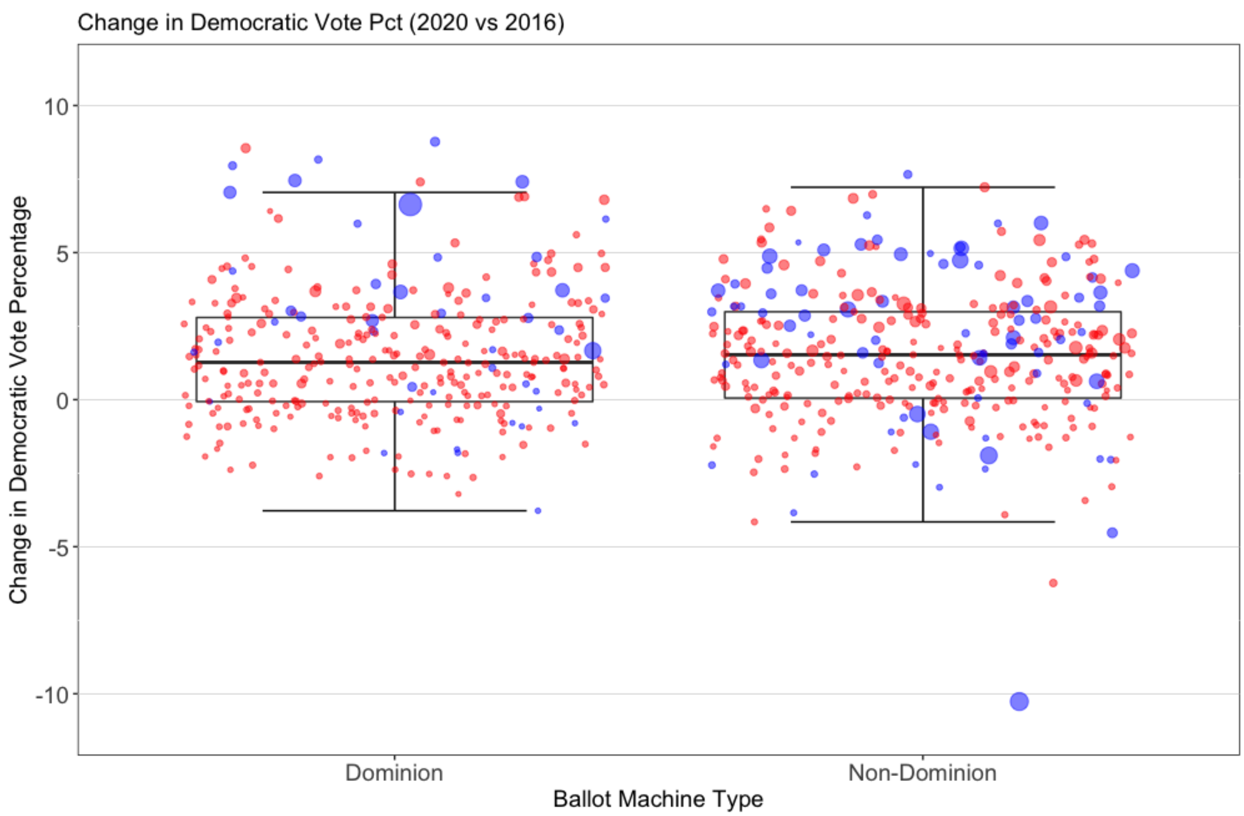

Figure 8 – Box-and-whisker Method to Identify Outlier Counties

The second method looked for counties that were more than two standard deviations away from the mean. While the data sets do not conform to a normal distribution (which is typically required for standard deviations to have any significance), we believed that using the second standard deviation would help identify counties with unexpectedly large increases in Democratic support. This method identified eight additional outlier counties for a total of sixteen counties, as shown in Figure 9. The dots on the graph represent the voting results for each county as described for Figure 8. Additionally, the mean is represented by an X, the second standard deviation is the solid horizontal line, and the third deviation is the dashed line. Counties above the second and third deviations were considered to be potentially anomalous.

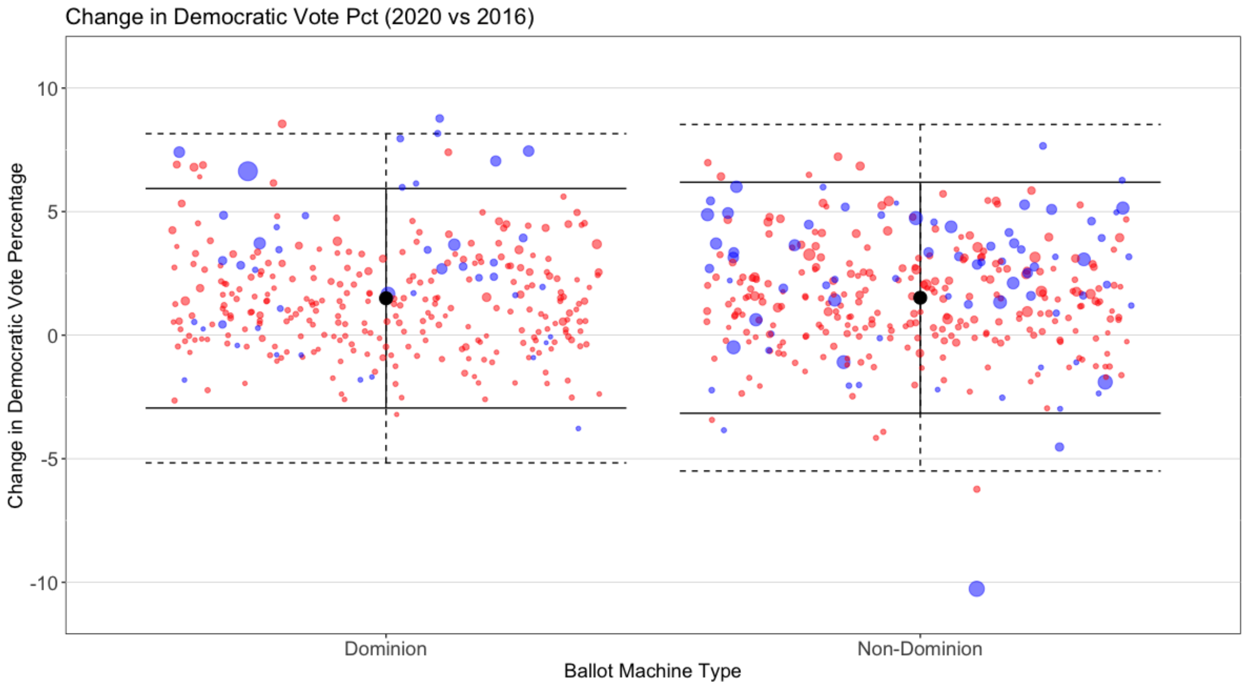

Figure 9 – Standard Deviation Method to Identify Outlier Counties

The sixteen counties that were identified as outliers are shown in Table 8. Ten are from Georgia, four from Michigan, and one each from Arizona and Wisconsin. Seven of the counties were won by Trump, despite large increases in Democratic support. An example is Cherokee County, Georgia, which increased its support for the Democratic candidate by 6.80 percentage points from 2016 but still finished below 30 percent of the total vote.

The outlier counties range in size from 1,557 votes to two million. Fulton County, Georgia, which received a lot of attention after the election, was not considered an outlier in our analysis because the increase in its Democratic support was in line with the rest of the state at 3.6 percentage points.

Table 8 – Counties with Dominion Machines Identified as Outliers

State	County	Change in Dem Vote Pct	Dem Pct 2016	Dem Pct 2020	Winner in 2020	Total Votes in 2020
Georgia	Henry	+8.77	50.93%	59.70%	Biden	122,742
Georgia	Forsyth	+8.55	24.09%	32.64%	Trump	129,305
Georgia	Rockdale	+8.16	61.76%	69.92%	Biden	44,686
Georgia	Douglas	+7.95	54.00%	61.95%	Biden	69,097
Georgia	Cobb	+7.45	48.89%	56.34%	Biden	393,746
Georgia	Gwinnett	+7.41	51.02%	58.43%	Biden	413,865
Georgia	Fayette	+7.40	38.53%	45.93%	Trump	71,993
Michigan	Kent	+7.05	45.00%	52.05%	Biden	361,048
Georgia	Paulding	+6.90	27.88%	34.79%	Trump	85,385
Georgia	Columbia	+6.88	29.40%	36.28%	Trump	80,579

State	County	Change in Dem Vote Pct	Dem Pct 2016	Dem Pct 2020	Winner in 2020	Total Votes in 2020
Georgia	Cherokee	+6.80	22.75%	29.55%	Trump	144,830
Arizona	Maricopa	+6.63	43.69%	50.32%	Biden	2,068,144
Michigan	Keweenaw	+6.41	36.75%	43.16%	Trump	1,557
Wisconsin	Ozaukee	+6.16	36.97%	43.16%	Trump	61,486
Michigan	Leelanau	+6.14	45.90%	52.04%	Biden	16,900
Michigan	Marquette	+5.98	48.65%	54.63%	Biden	37,462

Because Georgia dominates the list of outliers and received most of the media coverage about the alleged Dominion irregularities, we focused our deeper analysis on the ten Georgia counties. Figure 10 shows the percentage of votes the Democratic candidate received in 2020 and 2016 in the ten outlier counties. The '×' indicates the vote percentage in 2016, and the triangle is the vote percentage in 2020. The size of the triangle and the '×' represents the number of votes cast in that county.

In 2020 and 2016, five of the counties selected Trump and five counties selected the Democratic candidate. None of the counties flipped parties in 2020. Four of the counties were strong Trump counties that, despite the large increase in support for the Biden, were still below 40 percent.

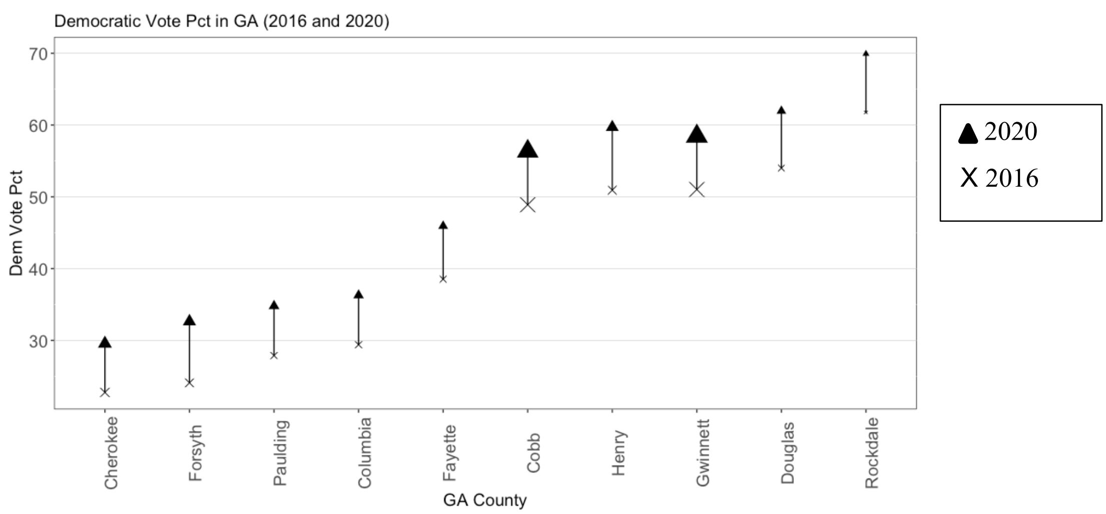

Figure 10 – Change in Democratic Vote Percentage in Georgia Outlier Counties

To put the increased support in context, we analyzed Democratic support in each of Georgia's outlier counties from 2004 to 2020. Figure 11 shows that the increased support in 2020 was not a one-time aberration but is part of a larger increase in Democratic support for the last sixteen years. In 2004, all five red counties gave less than 30 percent of the vote to the Democratic candidate. Except for 2012, all five counties increased their support for the Democratic candidate in each election.

Likewise, the blue counties also increased their support for the Democratic candidate in each election, except for 2012. All five blue counties were below 40 percent in 2004, but by 2016 they had all climbed above 50 percent except for Cobb, which was close at 48.9 percent.

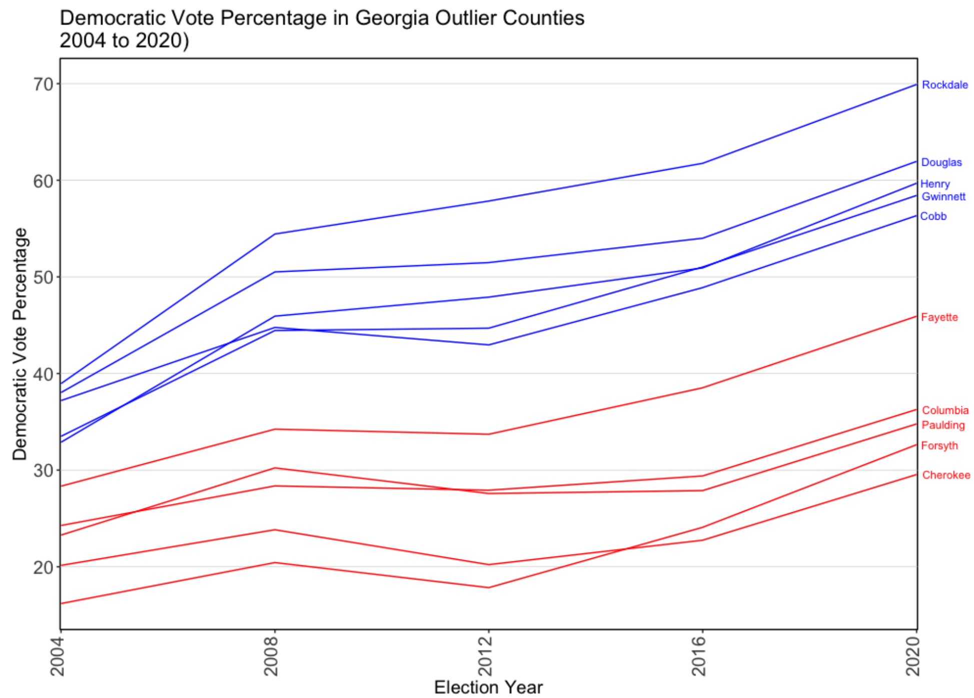

Figure 11 – Democratic Support from 2004 to 2020 in Georgia Outlier Counties

For additional context, we compared the change in Democratic vote percentage in 2020 to the previous three elections in Georgia's outlier counties, when Dominion machines were not used. As Figure 12 shows, while the 2020 election had a greater increase in Democratic support than in 2016 and 2012, it pales in comparison to some of the increases in 2008.

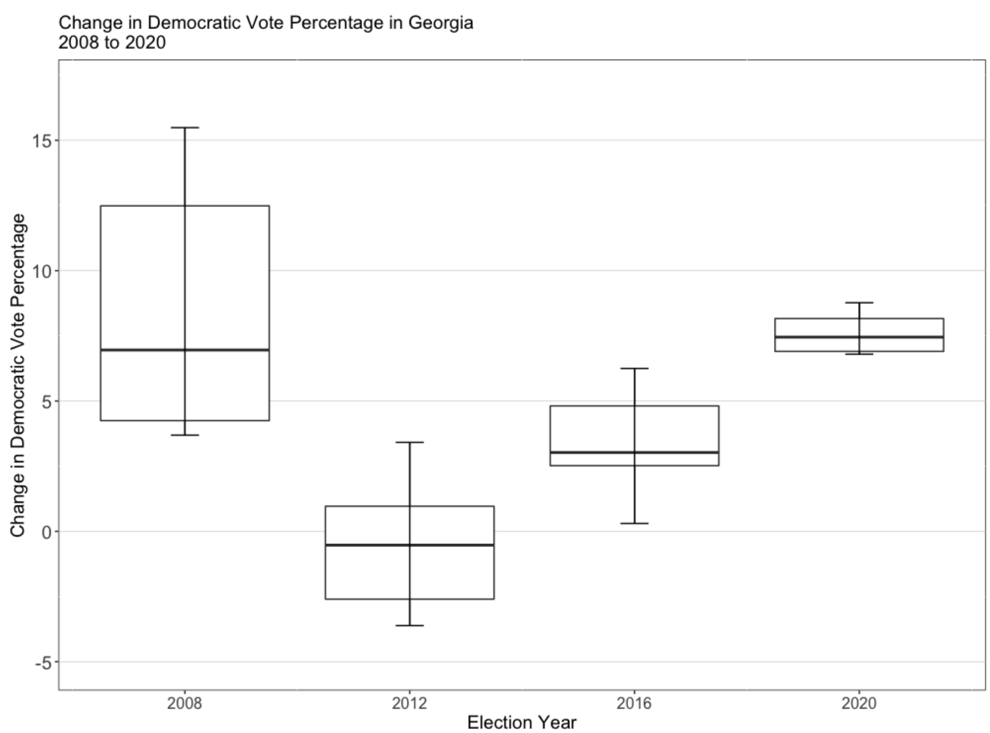

Figure 12 – Change in Democratic Support from 2008 to 2020 in Georgia Outlier Counties

Georgia accounted for ten of the sixteen outlier counties, but Michigan (four outlier counties), Wisconsin (one), and Arizona (one) accounted for the other six. As shown in Figure 13, we graphed the Dominion and non-Dominion counties to see if there was any indication of widespread vote manipulation by Dominion in those states. Michigan and Wisconsin had a statistically significant greater increase in Democratic support in counties that did not use Dominion machines. In Arizona, only one county (Maricopa) used Dominion machines, but its increase in Democratic support was similar to the increases in some non-Dominion counties.

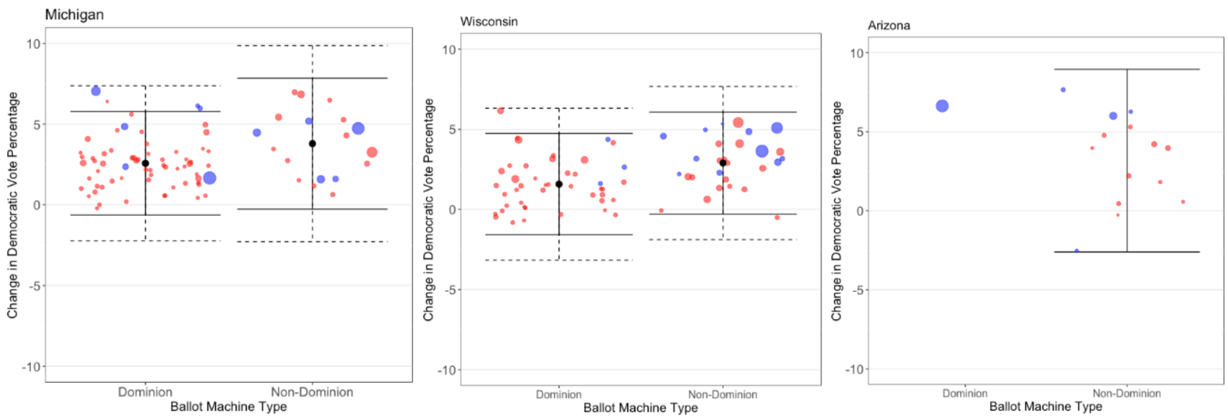

Figure 13 – Change in Democratic Support in Michigan, Wisconsin, and Arizona

2.3.3 Results Identified

We did not find any support from the data of widespread vote manipulations by Dominion machines in the eight key states we examined. Sixteen counties were identified as outliers because they had anomalous increases in their Democratic support from 2016. Ten of the outliers were in Georgia, but we did not find any indication in the data of unexplained vote increases. Instead, the increases seemed to be part of a trend of increased Democratic support since 2004, even in the red counties. Additionally, some of the counties in 2008 experienced Democratic increases that were double the increase in 2020.

The Democratic increases were also analyzed in the other states that had outliers. In Michigan and Wisconsin, the increase in Democratic support was statistically significantly greater in counties that did not use Dominion machines than in the ones that did. In Arizona, only one county (Maricopa) used Dominion machines, but its increase was not unusually different from the non-Dominion counties.

Based on these findings, the data does not support the allegations of vote manipulations by Dominion machines.

2.4 Voting Anomaly in Michigan

2.4.1 What Was Alleged

On the morning after the election, a SQUINT™ report described allegations that more than 100,000 votes had been fraudulently dumped for Biden in Michigan in the middle of the night. Screenshots posted on Twitter allegedly showed that one data dump in Michigan increased Biden's vote total by 138,339 votes while Trump's vote total was unchanged.

2.4.2 Analysis Performed

We downloaded the preliminary results from the Michigan Secretary of State's website on November 4th and compared each county's preliminary results to the 2016 results to identify any anomalous increases in Democratic support.

As seen in Figure 14, Antrim County stood out as a significant outlier because its support for the Democratic candidate nearly doubled from 32.5 percent in 2016 to 62.5 percent in 2020. The graph compares each county's results from 2016 to 2020. A dot above the dash line indicates an increase in Democratic support in 2020.

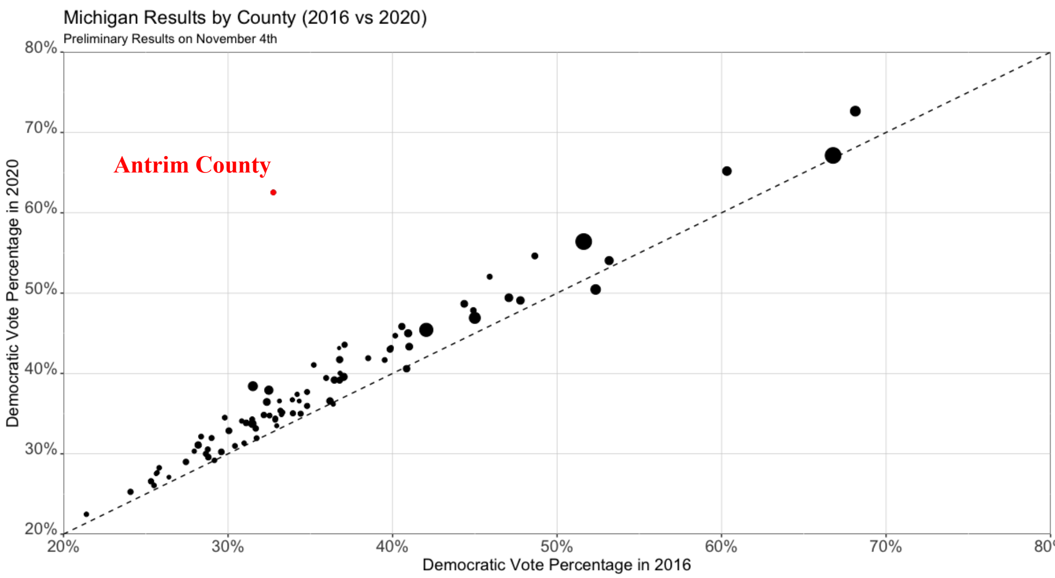

Figure 14 – Michigan Results by County (2016 vs. 2020) on November 4th

In the last 40 years in Michigan, the biggest shift in Democratic support between presidential elections was 13.0 percentage points in St. Clair County in 1996. Antrim County, historically a strong Republican county, supposedly had a shift that was 2.3 times greater and a significant statistical outlier, as shown in Figure **15**.

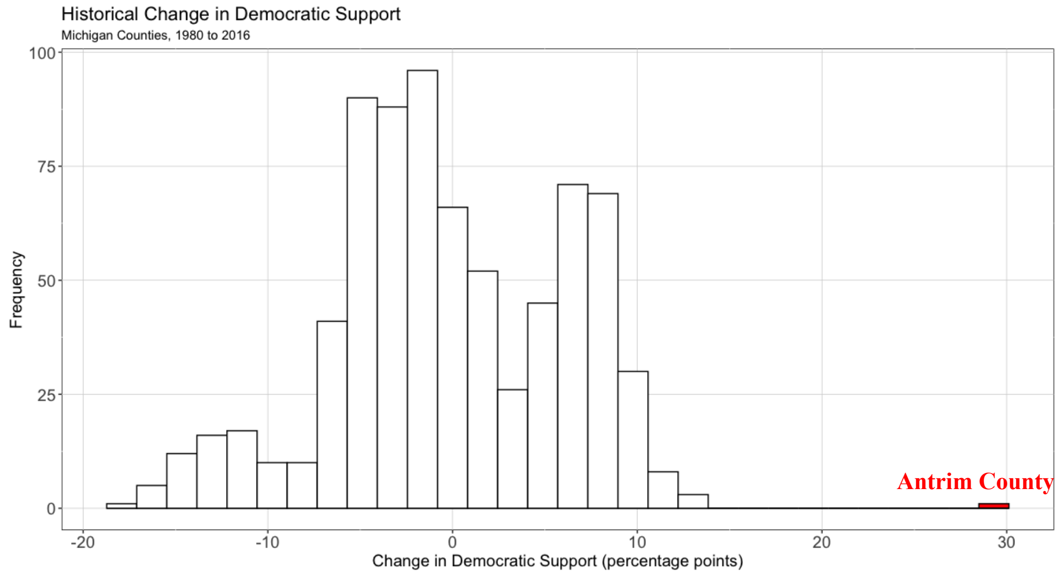

Figure 15 – Historical Change in Democratic Support, Michigan 1980–2016

Figure 16 shows that a Democratic presidential candidate in Antrim County had never exceeded 44 percent of the vote in the last 40 years. The preliminary result of 62.5 percent for 2020 is shown in red.

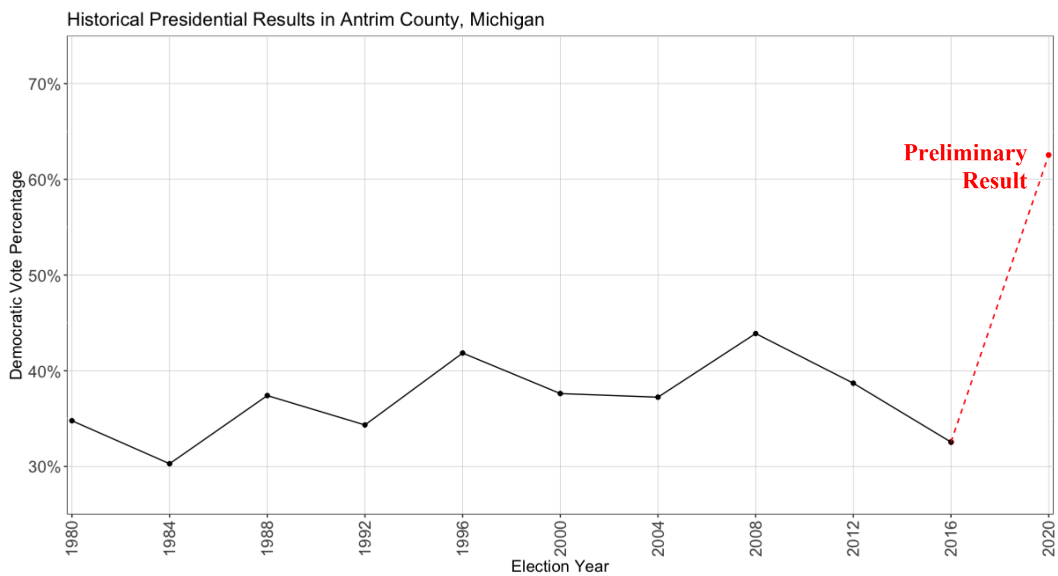

Figure 16 – Historical Presidential Results in Antrim County, Michigan

While Antrim County's anomalous results warranted further investigation for potential fraud, its vote totals were too small to account for all of the 138,339 votes in the allegation. At the time of the analysis, only 12,423 total votes had been reported in Antrim County.

During our analysis, Decision Desk HQ, which produced the election graphics used in the original Michigan allegation, issued a statement that "it was a simple error from a file created by the state that we ingested" [24]. The issue was corrected within 40 minutes of the original report and Biden's vote count was reduced accordingly. The data was corrected before our analysis began, so we were unable to detect the anomalies described in the original allegation.

For Antrim County, the results released by the Michigan Secretary of State later on November 4th showed all zeroes for the county. The following day, new results were released which showed Antrim County in line with its previous historical results, as shown in Figure 17.

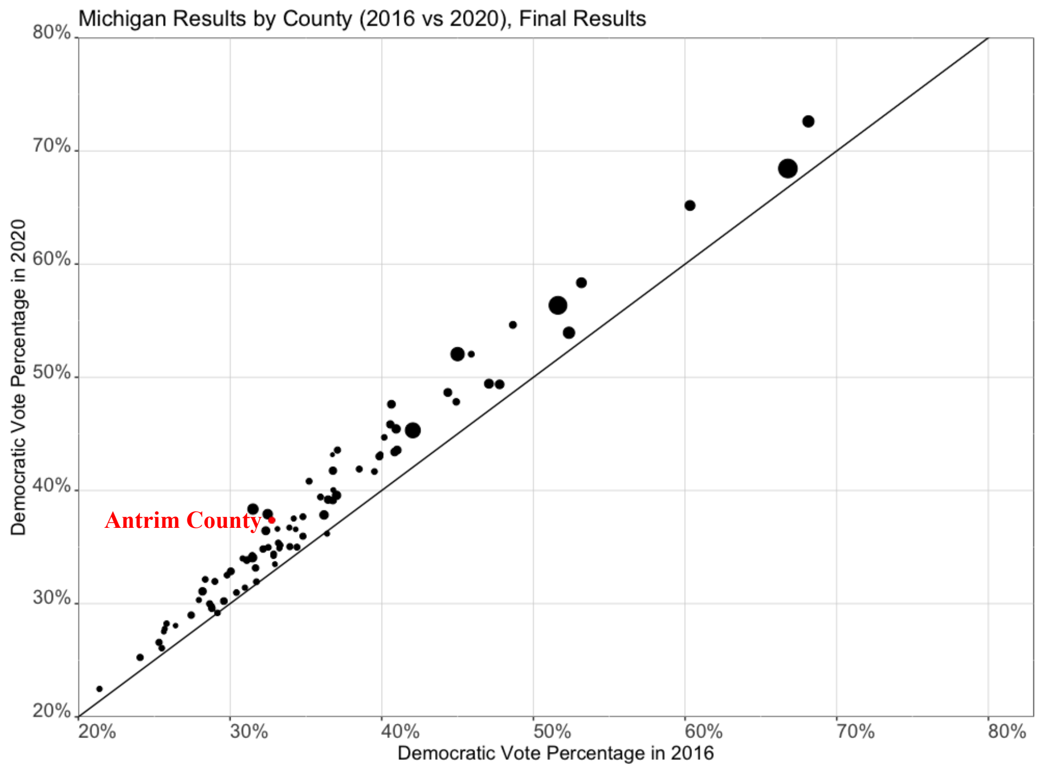

Figure 17 – Percentage of Democratic Votes in 2020 vs. 2016 in Michigan

2.4.3 Results Identified

The original allegation of 138,339 fraudulent votes was explained after an investigation by Politifact and other outlets [25]. A clerk in a Michigan county accidentally added a zero at the end of the results for Biden, changing them from 15,371 to 153,710, which is greater than the population of that county. The data was corrected quickly by the election officials and the subsequent data release showed a decrease in 138,339 votes for Biden. However, the Twitter user switched the order of the screenshots to make it appear as if Biden had gained the votes. The Twitter user subsequently deleted the tweet with the allegation and said, “I have now learned the MI update referenced was a typo in one county” [26].

In response to the unusual voting results in Antrim County, Michigan’s Secretary of State released a statement explaining that the initial results were a mistake and referred to it as an “isolated error” [27]. They explained that this error did not affect the way in which votes were tabulated, and that the issue was a result of “user error,” and “affected only how the results from the tabulators communicated with the election management system for unofficial reporting.” The official results for Antrim County recorded Biden’s percentage as 37.4%, which is consistent with previous elections. A subsequent hand audit on December 17, 2020 by bipartisan election officials verified that the official results were accurate [28].

2.5 Butler County, PA Lost Ballots

2.5.1 What Was Alleged

On October 28, 2020, local news out of Pittsburgh published an article about the Butler County Elections Director stating that the United States Postal Service had lost an unknown number of mail-in ballots [3].

2.5.2 Analysis Performed

In the days leading up to the statement released by the Butler County Elections Director, the mail-in ballot return rate for Butler County, PA (according to data collected from the Election Returns for Pennsylvania website [4]) was significantly lower than the other counties in Pennsylvania (Figure 18).

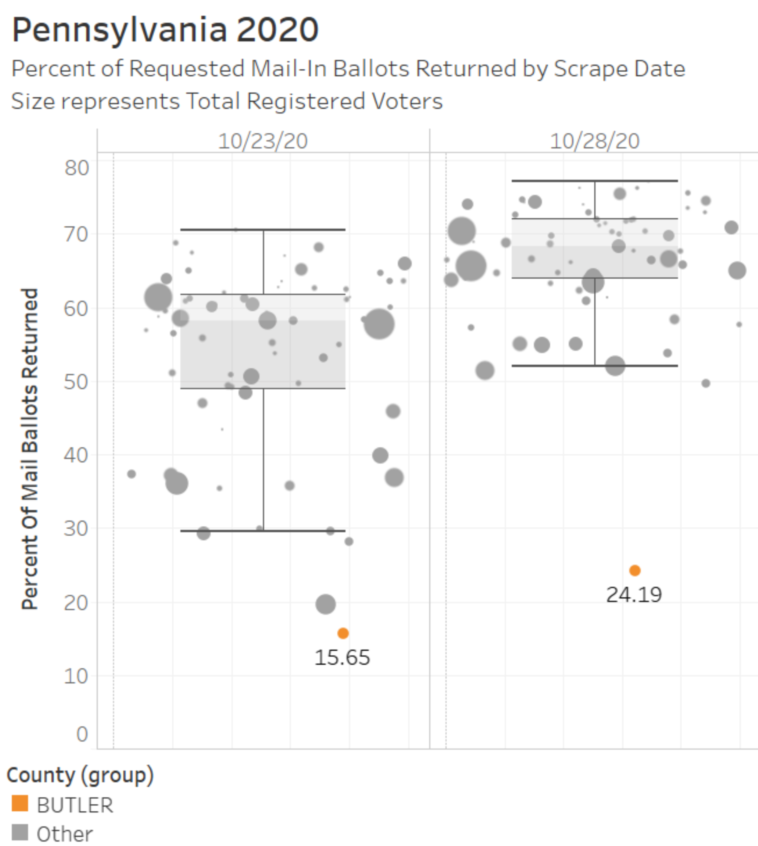

Figure 18 – Mail-In Ballots Returned Leading up to 10/28/20 Alleged Claim

As of October 28, 2020, Butler County showed only a 24.19% return rate for mail-in ballots at the time of data collection. However, the data scrape trends show a shift as of November 1, 2020, when the percent of returned ballots for Butler County jumped from 24.19% to 74.63%, as shown in Figure 19. This can be partially explained by a computer glitch that affected ballots being mailed out in Butler County. The glitch was reported by a local Pittsburgh, PA news outlet (KDKA) on October 14th, which was less than three weeks before Election Day [29] [30].

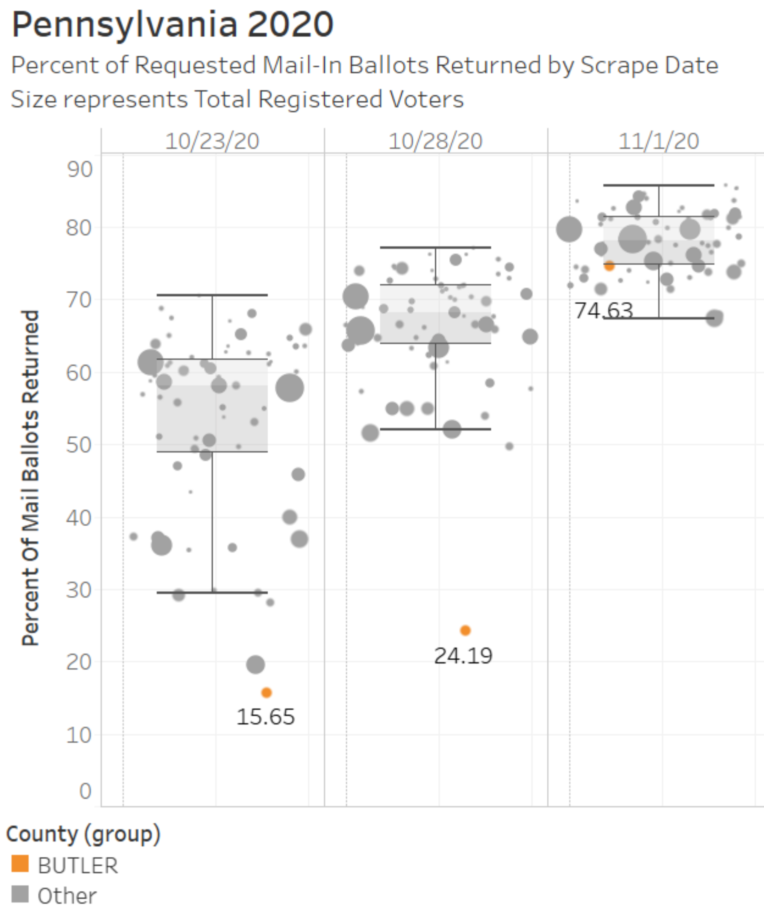

Figure 19 – Shift in Butler County, PA Mail-In Ballots Returned

With the concerns about missing mail-in ballots, the Butler County elections officials provided various options for voters to get their votes in, from keeping the elections office open over the weekend (October 31, 2020 – November 1, 2020), providing provisional ballots at polling places, and even having sheriff's deputies hand-deliver ballots to home-bound voters [31].

2.5.3 Results Identified

Although there were an unknown number of mail-in ballots lost that resulted in the October 28, 2020 announcement by the Butler County Elections Director, the post-announcement mail-in ballots return rate is within the lower quartile and is bordering the second quartile by Election Day 2020 (Figure 20).

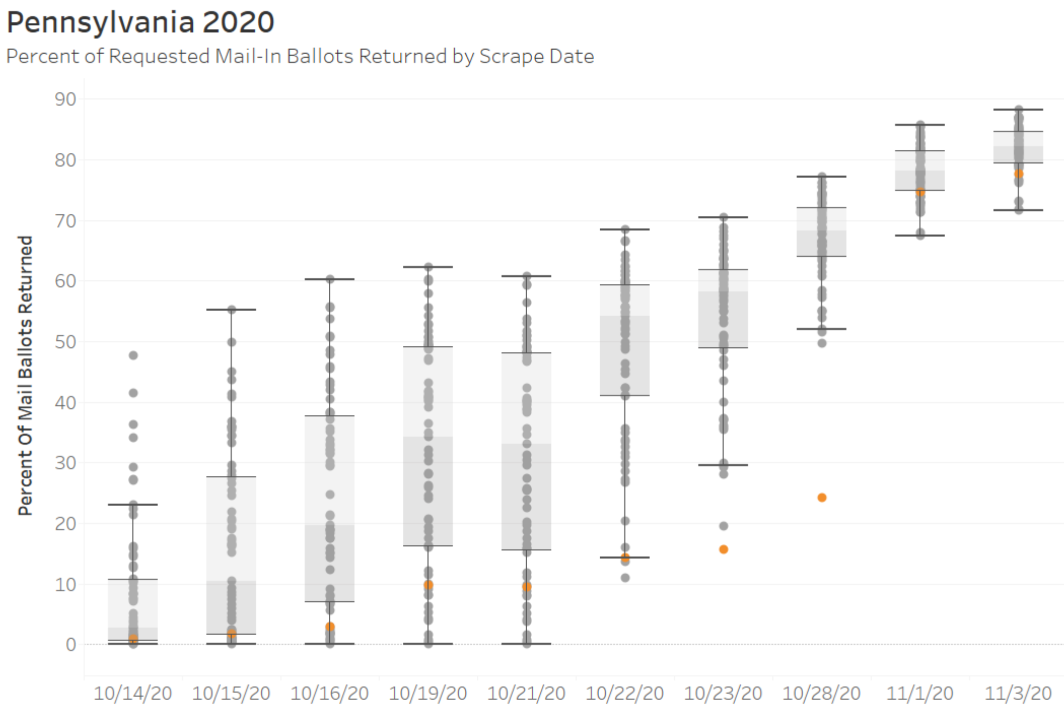

Figure 20 – Butler County, PA Mail-In Ballots Returned Trends

The actions of the Butler County Elections officials following the announcement of the loss of mail-in ballots by the USPS could have led to the significant increase in the mail-in ballot returns.

3 Conclusions

Between October 2020 and early January 2021, the Bald Eagle team researched several topics and data sources related to the 2020 United States presidential election in eight swing states. This report and the conclusions below summarize the team's findings.

- After researching ballot return rates and addresses with multiple accepted ballots in Georgia, the team found no anomalous points through its analysis. There were no suspicious indicators of ballot harvesting.
- Using election fingerprinting for the counties in Georgia, the team saw no signs of vote flipping or ballot box stuffing. Statistical evidence derived using a quantitative technique, which is similar to fingerprinting, supports a finding that these analytic techniques do not demonstrate fraud.
- Researching allegations of vote flipping with Dominion voting machines, the team found there was no statistical difference in the results from ballot machines from different vendors. The claim that Dominion machines artificially inflated results to Joseph Biden, Jr.'s advantage is not supported in the election results data.
- After investigating Antrim County in Michigan for data anomalies that created unusual increases in votes for Mr. Biden, the team found that the issues were caused by user error, which was discovered and corrected quickly and did not have an impact on the official election results.

- Although many requested mail-in ballots were lost in Butler, PA, the overall return rate of ballots leading up to Election Day fell in the expected range of all other counties in Pennsylvania. No evidence of nefarious or fraudulent activities was identified.

In summary, multiple types of analysis found no statistically significant evidence of fraud, manipulation, or error in the eight states included in this research.

4 References

[1] P. Klimek, Y. Yegorov, R. Hanel and S. Thurner, "Statistical detection of systematic election irregularities," *Proceedings of the National Academy of Sciences,* vol. 109, pp. 16469-16473, 2012.

[2] [Online]. Available: https://verifiedvoting.org/equipmentdb/.

[3] [Online]. Available: Source: https://pittsburgh.cbslocal.com/2020/10/28/butler-county-elections-director-says-postal-service-lost-mail-in-ballots/ .

[4] [Online]. Available: https://www.electionreturns.pa.gov/ReportCenter/Reports.

[5] Senate Select Committee on Intelligence, "Report of the Senate Committee on Intelligence on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election Volume 1: Russian Efforts against Election Infrastructure with Additional Views," [Online]. Available: https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume1.pdf. [Accessed 25 August 2019].

[6] R. S. Mueller, "Report On The Investigation Into Russian Interference In The 2016 Presidential Election," U.S. Department of Justice, Washington, D.C., 2019.

[7] T. Romm and I. Stanley-Becker, "Suspicious robocall campaign warning people to ‘stay home’ spooks voters nationwide," The Washington Post, 3 November 2020. [Online]. Available: https://www.washingtonpost.com/technology/2020/11/03/robocall-election-day/. [Accessed 12 January 2021].

[8] [Online]. Available: https://en.wikipedia.org/wiki/Ballot_collection.

[9] [Online]. Available: https://law.justia.com/codes/georgia/2014/title-21/chapter-2/article-10/section-21-2-385.

[10] [Online]. Available: Source: https://elections.sos.ga.gov/Elections/voterabsenteefile.do.

[11] A. Swenson, "Report spreads debunked claims about Dominion machines in Michigan county," Associated Press, 15 December 2020. [Online]. Available: https://apnews.com/article/fact-checking-afs:Content:9847904839. [Accessed 12 January 2021].

[12] Newsmax TV, "Sidney Powell to Newsmax TV: Dominion Designed to ‘Rig Elections’," Newsmax, 2020. [Online]. Available: https://www.wbap.com/news/sidney-powell-to-newsmax-tv-dominion-designed-to-rig-elections/. [Accessed 12 January 2021].

[13] A. Swenson, "Dominion machines didn’t ‘flip’ votes in Ware County, Georgia," Associated Press, 7 December 2020. [Online]. Available: https://apnews.com/article/fact-checking-9773239691. [Accessed 12 January 2021].

[14] F. Benford, "The law of anomalous numbers," in *Proceedings of the American Philosophical Society*, 1938.

[15] C. Borghesi and J.-P. Bouchaud, "Spatial correlations in vote statistics: a diffusive field model for decision-making," *The European Physical Journal B,* vol. 75, pp. 395-404, 2010.

[16] P. H. Westfall, "Kurtosis as Peakedness, 1905–2014. R.I.P.," *The American Statistician,* vol. 68, no. 3, pp. 191-195, 2014.

[17] M. Elder, "Vladimir Putin's critics cry foul over alleged voter fraud in Russian election," The Guardian, 4 March 2012. [Online]. Available: https://www.theguardian.com/world/2012/mar/04/putin-alleged-voter-fraud-russian-election. [Accessed 12 January 2021].

[18] E. Barry and M. Schwirtz, "After Election, Putin Faces Challenges to Legitimacy," The New York Times, 5 March 2012. [Online]. Available: https://www.nytimes.com/2012/03/06/world/europe/observers-detail-flaws-in-russian-election.html. [Accessed 12 January 2021].

[19] OSCE Office for Democratic Institutions and Human Rights, "Canada Parliamentary Elections 23 January 2006 OSCE/ODIHR Election Assessment Mission Report," Organization for Security and Co-operation in Europe, Warsaw, 2006.

[20] OSCE Office for Democratic Institutions and Human Rights, "Canada Parliamentary Elections 19 October 2015 OSCE/ODIHR Needs Assessment Mission Report," Organization for Security and Co-operation in Europe, Warsaw, 2015.

[21] J. L. Devore, "The Normal Distribution and Discrete Populations," in *Probability and Statistics for Engineering and the Sciences*, Belmont, California: Brooks/Cole Pub. Co; 2nd edition, 1987, pp. 146-148.

[22] M. Bartiromo, "Trump legal team to file new ballot lawsuits," Fox News, 8 November 2020. [Online]. Available: https://www.foxnews.com/transcript/trump-legal-team-to-file-new-ballot-lawsuits. [Accessed 26 January 2021].

[23] @realDonaldTrump, "Twitter," 12 November 2020. [Online]. Available: https://web.archive.org/web/20201113000036if_/https://twitter.com/realDonaldTrump/status/1326926226888544256. [Accessed 29 January 2021].

[24] [Online]. Available: https://results.decisiondeskhq.com/.

[25] [Online]. Available: https://www.politifact.com/factchecks/2020/nov/04/tweets/no-biden-did-not-receive-thousands-mysteriously-su/.

[26] [Online]. Available: https://www.politifact.com/factchecks/2020/nov/04/tweets/no-biden-did-not-receive-thousands-mysteriously-su/.

[27] [Online]. Available: https://www.michigan.gov/documents/sos/Antrim_Fact_Check_707197_7.pdf.

[28] [Online]. Available: https://www.michigan.gov/som/0,4669,7-192-47796-547883--,00.html.

[29] R. Jones, "Computer Glitch Leaves Butler County Residents Waiting For Mail-In Ballots Despite Applying Weeks Ago," KDKA, 14 October 2020. [Online]. Available: https://pittsburgh.cbslocal.com/2020/10/14/butler-county-mail-in-ballot-delays-computer-glitch/. [Accessed 25 January 2021].

[30] u/rialucia, "Butler County Mail in Ballots be like...," Reddit, 11 October 2020. [Online]. Available: https://www.reddit.com/r/pittsburgh/comments/j9izbe/butler_county_mail_in_ballots_be_like/. [Accessed 25 January 2021].

[31] A. Wadas, "Butler County And USPS At Odds Over Lost Ballots," KDKA 02 CBS Pittsburgh, 29 October 2020. [Online]. Available: https://pittsburgh.cbslocal.com/2020/10/29/butler-county-lost-mail-in-ballots/. [Accessed 28 January 2021].

[32] Georgia Secretary of State's Office, "GEORGIA EASY VOTER REGISTRATION OPTIONS BREAK NEW RECORDS," Georgia, 06 October 2020. [Online]. Available: https://sos.ga.gov/index.php/elections/georgia_easy_voter_registration_options_break_new_records. [Accessed 20 January 2021].

[33] Georgia Secretary of State's Office, "RESULTS >> PRESIDENT OF THE UNITED STATES," Georgia, 7 December 2020. [Online]. Available: https://results.enr.clarityelections.com/GA/107231/web.264614/#/detail/5000. [Accessed 20 January 2021].

[34] V. Bergengruen, "Weeks Before Critical Runoffs, Georgia Is Becoming Ground Zero for Election Disinformation," *Time,* 4 December 2020.

[35] W. R. Mebane Jr, "Election forensics: Vote counts and Benford's law," in *Summer Meeting of the Political Methodology Society, UC-Davis, July*, 2006.

[36] J. Golbeck, "Op/ed: Breaking the (Benford's) Law?," Maryland Today, University of Maryland, Office of Strategic Communications, 12 November 2020. [Online]. Available: https://today.umd.edu/articles/oped-breaking-benfords-law-1ccdc2ba-c7b0-4515-a1ea-103f16607f5c. [Accessed 13 January 2021].

[37] J. Dacey, "Benford's law and the 2020 US presidential election: nothing out of the ordinary," Physics World, 19 November 2020. [Online]. Available: https://physicsworld.com/a/benfords-law-and-the-2020-us-presidential-election-nothing-out-of-the-ordinary/. [Accessed 13 January 2021].

[38] "Radiolab: Breaking Benford," WNYC Studios, 13 November 2020. [Online]. Available: https://www.wnycstudios.org/podcasts/radiolab/articles/breaking-benford. [Accessed 13 January 2021].

[39] Reuters Staff, "Fact check: Deviation from Benford's Law does not prove election fraud," Thomson Reuters, 10 November 2020. [Online]. Available: https://www.reuters.com/article/uk-factcheck-benford/fact-check-deviation-from-benfords-law-does-not-prove-election-fraud-idUSKBN27Q3AI. [Accessed 13 January 2021].

[40] J. Deckert, M. Myagkov and P. C. Ordeshook, "Benford's Law and the detection of election fraud," *Political Analysis,* vol. 19, pp. 245-268, 2011.

[41] W. R. Mebane, "Comment on "Benford's Law and the detection of election fraud"," *Political Analysis,* vol. 19, pp. 269-272, 2011.

[42] @YoDatElefent, "Biden's Vote Tallies apparently Violate Benford's Law," Twitter, 2020.

[43] W. R. Mebane, "Inappropriate Applications of Benford's Law Regularities to Some Data from the 2020 Presidential Election in the United States," 10 November 2020. [Online]. Available: http://www-personal.umich.edu/~wmebane/inapB.pdf. [Accessed 13 January 2021].

[44] C. Borghesi, J.-C. Raynal and J.-P. Bouchaud, "Election Turnout Statistics in Many Countries: Similarities, Differences, and a Diffusive Field Model for Decision-Making," *PLoS ONE,* vol. 7, no. 5, p. e36289. doi:10.1371/journal.pone.0036289, 2012.

[45] A. E. Kossovsky and S. J. Miller, "Report on Benford's Law Analysis of 2020 Presidential Election Data," [Online]. Available: https://web.williams.edu/Mathematics/sjmiller/public_html/KossoskyMiller_FinalBenfordAnalysis.pdf. [Accessed 13 January 2021].

[46] J. Fernández-Gracia, K. Suchecki, J. J. Ramasco, M. San Miguel and V. M. Eguíluz, "Is the voter model a model for voters?," *Physical review letters,* vol. 112, p. 158701, 2014.

[47] [Online]. Available: https://www.michigan.gov/documents/sos/Antrim_Fact_Check_707197_7.pdf.

[48] A. Rundlett and M. Svolik, "Deliver the vote! micromotives and macrobehavior in electoral fraud," *American Political Science Review,* vol. 110, no. 1, pp. 180-197, 2016.

Appendix A Examples of Election Fingerprints

This appendix contains some examples of election fingerprints showing fair and fraudulent elections. All diagrams are from Klimek, et al. [1]. The axes in these diagrams are % turnout on the horizontal axis (x axis), and % votes for the winner on the vertical axis (y axis).

A.1 Fair Election Fingerprint – Canada 2011

This fingerprint is from [1] and is based upon data from the 2011 Canada federal election.

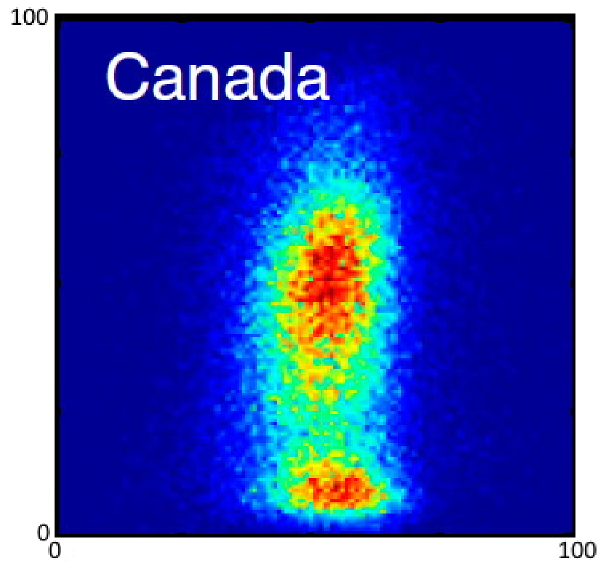

Figure 21 – Example of a Fair Election

A.2 Fraudulent Election Fingerprint – Russia 2012

This fingerprint is from [1] and is based upon data from the 2012 Russia presidential election. It is important to note the "smearing" up and to the right in this fingerprint. This, along with the cluster around the 100%/100% is due to (heavy handed) ballot box stuffing and ballot harvesting. Ballot box stuffing and ballot harvesting result in these patterns because they distort the relationship between turnout and voting for a candidate. If ballots are harvested, and then only the votes for the "right" candidate are counted, this will result in a disproportionate percentage of votes for the winner given the turnout.

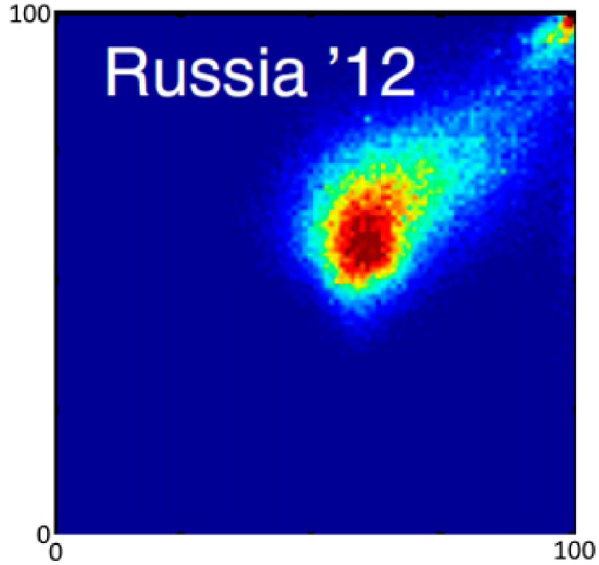

Figure 22 – Example of a Fraudulent Election

A.3 Harvesting of Ballots

A.3.1 Harvesting of Unfilled Ballots

If the winning candidate harvested unfilled ballots, then completed and submitted them, it would result in a fingerprint like that in Figure 22, as this is a form of ballot box stuffing. Turnout will be higher in precincts that have a higher percentage of the votes for the winner.

A.3.2 Harvesting of Completed Ballots

If the winning candidate harvested and discarded completed ballots in precincts likely to lean towards their opponent, then one would expect unusually high votes for the winning candidate, and a lower overall turnout. The turnout would be suppressed because of discarding ballots, yet the winning candidate would have a higher-than-expected percentage of the total vote.

However, if this is done in precincts that lean strongly against the winning candidate, then the winning percentage would be low and turnout would be low, since the number of votes for the winner is low in those precincts. This would result in a smear down and to the left on the fingerprint.

If the losing candidate harvested and discarded completed ballots in precincts that lean strongly towards the winning candidate, then one would expect to see a smearing up and to the left. This would suppress the overall turnout, yet the winning candidate percentage would remain high.

To determine which of the cases is occurring, one would need to examine the precincts to determine which way they are likely to vote to determine whether the evidence supports a claim of ballot harvesting.

Appendix B   Other Issues Examined

Before Election Day, we monitored the data for several metrics to detect any signs of voter suppression or fraud. These metrics include the rejection rates of mail-in ballots, the number of voters who may have voted more than once, and the elapsed time for counties to mail ballots to voters who requested them.

B.1 Duplicate Voters

Because of the high levels of transparency and detailed data available from Georgia, we were able to perform independent analysis to ascertain whether there was merit to various claims about the election process and results in Georgia. In Georgia, more than 7.5 million voters registered to vote in the 2020 general election [32], and almost 5 million votes were cast in the presidential contest [33]. This level of detail and transparency allowed for examination of the claim that voters were casting multiple votes in Georgia [34].

The pre-election data from Georgia contained detailed information on early voters, including their name, voter ID, and the status of their ballot. The data was updated daily and enabled us to monitor for signs of voter fraud, such as a voter submitting multiple ballots.

The standard operating procedure for state voting offices is to cancel any superfluous ballots that are received from the same voter and accept only one ballot. This is not indicative of fraud and occurs most often when a voter requests a mail-in ballot but then decides to vote in person. In that case, the mail-in ballot would be cancelled, and the in-person ballot would be accepted by the voting office. There should be at most only one accepted ballot per voter.

The data from Georgia provided transparent details on which ballots were cancelled or accepted for each voter. We monitored the data for incidents of multiple ballots being accepted from a single voter, which is not a valid procedure. The presence of multiple accepted ballots from a voter in the data may not necessarily indicate fraud – it could be the result of data entry errors. Data entry errors were common in the data and appeared most often in dates, like “9530-02-23" or “0003-11-30” for when a ballot was accepted. There were twenty-nine ballots that were recorded as being accepted after the year 2020. But even with data quality issues, the widespread appearance of multiple ballots being accepted or an upward trend in a single county or precinct could require further investigation to determine the root cause.

In the Georgia data, we detected a total of seventeen voters who appeared to have multiple ballots accepted. These were spread across eight counties, with Gwinnett and Fulton having the most with five each. Table 9 shows the data on multiple ballots accepted from voters.

We were unable to determine from the data whether the seventeen voters had multiple ballots accepted and tabulated, or whether it was a data quality issue. However, because there were no publicly announced investigations, no credible reports of duplicate voting in Georgia, and because the number of duplicate ballots was relatively low (0.0003 percent of the total votes cast in Georgia), we believe the low number of occurrences is not indicative of widespread voter fraud.

Table 9 – Multiple Ballots Accepted for the Same Voter ID

County	Voter ID	Ballot ID	Ballot Status	Ballot Issued Date	Ballot Return Date	Ballot Style
Clayton	8870879	181	Accepted	9/30/20	10/15/20	MAILED
Clayton	8870879	224	Accepted	10/1/20	10/15/20	MAILED
Dekalb	2063623	2681	Accepted	10/15/20	10/15/20	IN PERSON
Dekalb	2063623	2687	Accepted	10/15/20	10/15/20	IN PERSON
Dekalb	2063623	75	Cancelled	9/18/20		MAILED
Dekalb	2150186	3480	Accepted	10/13/20	10/13/20	IN PERSON
Dekalb	2150186	3714	Accepted	10/13/20	10/13/20	IN PERSON
Fulton	2066680	7146	Accepted	10/23/20	10/23/20	IN PERSON
Fulton	2066680	6666	Cancelled	10/23/20	10/23/20	IN PERSON
Fulton	2066680	6677	Accepted	10/23/20	10/23/20	IN PERSON
Fulton	2501632	1348	Accepted	10/15/20	10/15/20	IN PERSON
Fulton	2501632	1346	Accepted	10/15/20	10/15/20	IN PERSON
Fulton	2503470	3483	Accepted	10/16/20	10/16/20	IN PERSON
Fulton	2503470	1018	Accepted	10/8/20	10/26/20	MAILED
Fulton	2527487	5597	Accepted	10/30/20	10/30/20	IN PERSON
Fulton	2527487	5606	Accepted	10/30/20	10/30/20	IN PERSON
Fulton	5977767	3155	Accepted	10/13/20	10/13/20	IN PERSON
Fulton	5977767	2835	Accepted	10/18/20	10/18/20	IN PERSON
Gwinnett	1885063		Accepted		10/7/20	MAILED
Gwinnett	1885063	470	Accepted	9/18/20	10/9/20	MAILED
Gwinnett	4687367	517	Cancelled	10/6/20	10/16/20	MAILED
Gwinnett	4687367	628	Accepted	10/16/20	10/28/20	MAILED
Gwinnett	4687367	235	Accepted	9/18/20	10/28/20	MAILED
Gwinnett	5998434	1123	Accepted	10/13/20	10/21/20	MAILED
Gwinnett	5998434	681	Accepted	9/18/20	10/21/20	MAILED
Gwinnett	8102047	1166	Accepted	10/7/20	10/16/20	MAILED
Gwinnett	8102047		Accepted		10/16/20	MAILED
Gwinnett	8102047					
Gwinnett	8244323	854	Accepted	10/13/20	10/13/20	IN PERSON
Gwinnett	8244323	852	Accepted	10/13/20	10/13/20	IN PERSON
Hall	5756934	4821	Accepted	10/23/20	10/23/20	IN PERSON
Hall	5756934	4666	Accepted	10/23/20	10/23/20	IN PERSON
Henry	12782544	13923	Accepted	10/29/20	10/29/20	IN PERSON
Henry	12782544	341	Accepted	10/12/20	11/1/20	MAILED
Muscogee	5199856	5181	Accepted	10/22/20	10/22/20	IN PERSON
Muscogee	5199856	5178	Accepted	10/22/20	10/22/20	IN PERSON
Richmond	10328335	20745	Accepted	10/29/20	10/29/20	IN PERSON
Richmond	10328335	20743	Accepted	10/29/20	10/29/20	IN PERSON

County	Voter ID	Ballot ID	Ballot Status	Ballot Issued Date	Ballot Return Date	Ballot Style
Richmond	10328335	20741	Cancelled	10/29/20	10/29/20	IN PERSON

B.2 Elapsed Time to Send Ballots

Before the election, there was concern that voting offices would not be able to handle the volume of requests for ballots by mail. To determine if this was an issue in Georgia, we used the Georgia election data to calculate the average number of days for each county to mail a ballot after receiving a request. A county with a very high average could indicate that the county was struggling to meet the demand. We accounted for the fact that ballots could be requested up to 180 days in advance of the election but not mailed until September 15 at the earliest.

We found that the mean number of days to mail a ballot was impressively quick at 1.9 days. Only one county was anomalous – Baldwin County took an average of 6.4 days to mail ballots. Because of the rapid response rate and the presence of only one anomaly, we did not analyze further.

Appendix C Benford's Law and Elections

It has been proposed that Benford's law [14] [35] can be used to detect fraud in elections. Benford's law in its simplest form says that in many cases the leading digits are not uniformly distributed, especially the first digit where the lower numbers are more likely to appear. This theory has appeared in some social media and news reports [36]. In the early 2000s, a supposition was posited [35] that Benford's law [14] applies to the second digit Benford's law (abbreviated 2BL) of voter counts if the election is fair. During the post-election period, there have been claims that voter counts violate "Benford's Law"; sometimes these claims state which digit, such as 2BL, but often these claims do not specify which digit is used [37] [38]. Common claims apply Benford's law to the first, second, or last digit of the vote counts [39].

C.1 Is Benford's Law Applicable to Elections?

Before discussing whether the 2020 voter counts follow 2BL, we need to first examine whether 2BL applies to election vote counts. In [40] it is argued that 2BL doesn't apply to voter counts; in [41] the author of the original paper [35], which suggests that 2BL applies to elections, indicates that there are issues with his original work and that "it may be more precise to refer to second-digit Benford-like tests." In both [40] and [41], all of the authors agree that "a test based on the mean of the second significant digit of vote counts equals 4.187 is useful as a test for the occurrence of election fraud."

Looking at [35], the claim is that relatively large χ^2 values testing for 2BL suggest that maybe there has been fraud. There are several caveats to this statement, including that it does not apply at the voting machine level, but only for precinct-level vote counts, and that there is independence across precincts. Thus, if this test is applied to counts at the county, state, or national level, 2BL is not a valid test.

C.2 Benford's Law Application

The question then becomes "what is being checked when people are claiming that the 2020 presidential election voter counts do not follow Benford's law?", and we need to determine whether it is of any import to examine such a question. At this time, the National Election Security Lab team has not identified any detailed analysis of precinct-level data by anyone claiming that the data violates a 2BL test. We have reviewed claims of fraud detected by Benford's law, but they appear to be using the leading digit [42], which has been shown to not be a good indicator of fraud [39] [43]. It may be that the test described in [35] is simply not being performed for several reasons:

1. The original author proposing the principle later retracted his assertion that the principle applies, so the issue is no longer relevant.
2. Collecting precinct-level data is labor intensive.
3. Many states do not provide precinct-level data.
4. There is no evidence that precincts are independent, and in fact there is evidence that precincts are not independent, but rather are biased along geographic divides [44] (e.g., urban areas tend to vote blue, and rural areas tend to vote red).

One study [45] using a 2BL test might have been performed as described in [35], however the results shown were for tests performed at the state level, and the authors assert that the results were similar at the county and precinct levels. In the study, Kossovsky and Miller show a graph and results using Philadelphia precinct-level data, the conclusion of which is that the data passes the second digit test at the 95% confidence level. Hopefully, Kossovsky and Miller will publish a more detailed paper on the analysis they conducted.