An Electoral System in Crisis

lulu Fries'dat & Anselmo Sampietro

in collaboration with Fritz Scheuren

July 18, 2016

Having confidence in our elections is central to our faith in our government and all of the decisions that we make collectively as a nation. But are the candidates who win the ones we actually vote for?

A large and growing body of research provides convincing evidence that U.S. electronic voting equipment in many areas throughout the country is not counting the votes accurately. This could be due to malfunctions in computer equipment that in 43 states is over a decade old, and long past its natural life. However, in many cases, the evidence strongly suggests that fraud is the likely explanation. These problems have been occurring since at least 2004, and are certainly present in the current 2016 presidential primaries.

The documentation consists of statistical graphs analyzing data from five presidential cycles, as well as off-year races from across the country. The data illustrates that there are unusually large discrepancies between small precinct and large precinct election returns, and noticeable differences between hand-counted and machine-counted precinct results. Even in isolation, the data gives cause for concern. The statistical evidence is reinforced by physical evidence and congressional hearings: manual recounts that do not match the totals of the machines being audited; and testimony under oath about direct knowledge of tampering with electronic voting equipment.

We examined the election results of the 2016 presidential primaries, and found irregularities in the overwhelming majority of the twenty-one states that we analyzed. The data indicates, in particular, that the totals reported in the Democratic race between Hillary Clinton and Bernie Sanders may not be correct. In state after state, independent examination by two separate analysts found suspect statistical patterns giving Clinton inflated percentages, that in all likelihood are not fully based on actual votes; and leaving Sanders with what appear to be artificially depressed totals.

The difference between the reported totals, and our best estimate of the actual vote totals, varies considerably from state to state. However, these differences are significant—sometimes more than 10%—and could change the outcome of the 2016 Democratic presidential primary. We found irregularities in the 2016 Republican presidential primary as well, and while concerning, we do not believe they are large enough to change the outcome of that race.

Fritz Scheuren, a member of the statistics faculty at George Washington University, and a former president of the American Statistical Association, has been a collaborator in this research. Examining the data from the study, Scheuren said, "As a statistician, I find the results of the 2016 primary voting

unusual. In fact, I found the patterns unexpected [and possibly even] suspicious. There is a greater degree of smoothness in the outcomes than the roughness that is typical in raw/real data."

It is important to note that the fact that a candidate benefits from irregularities does not imply that a candidate is responsible for them.

In January 2014, The Presidential Commission on Election Administration published a report stating, "Perhaps the most dire warning the Commission heard in its investigation … concerned the impending crisis in voting technology. Well-known to election administrators, if not the public at large, this impending crisis arises from the widespread wearing out of voting machines purchased a decade ago (p.62.)" This report was issued over two years ago, but unfortunately very little has been done since then to rectify the problem. So the issues we are reporting here, of security problems on old and failing machines, are not surprising. However we did find security issues with even newer electronic voting equipment, such as the machines in New York State.

At a congressional briefing on voter suppression, held on April 21, 2016, Rep. Hank Johnson (D-Georgia) expressed grave concern about the security of the voting equipment: "There is a very insidious, treacherous and deceitful method of voter suppression, and it has to do with the integrity of the voting process itself… one possibility, and I think it's a very good one, is that someone's manipulating the counting of the votes. Someone is hacking into these computers that tabulate the votes."

An Environment of Corruption

The portrait of an electoral system in crisis is further supported by reports from election integrity organizations, media outlets, and individuals on social media that voting is increasingly taking place in a corrupt environment. This contextual evidence of voters purged from the rolls, registrations lost in the mail, party registrations being changed without a voters' knowledge or intent, voters being sent incorrect ballots, a shortage of ballots, polling places being closed, discouragingly long lines in targeted precincts and states, and disturbingly large disparities between initial exit polls and official results, lends credence to the argument that if one form of fraud is already in play, another form of fraud is more plausible. This information is being aggregated by election integrity groups such as Election Justice USA, through voter testimonials and lawsuits that are in progress around the country.

Figures 1 and 1A are examples of disenfranchised voters from the 2016 presidential primaries. Stories like these have been ubiquitous in many states, including Arizona, New York and California. More of these instances are documented in this article on Heavy.com.
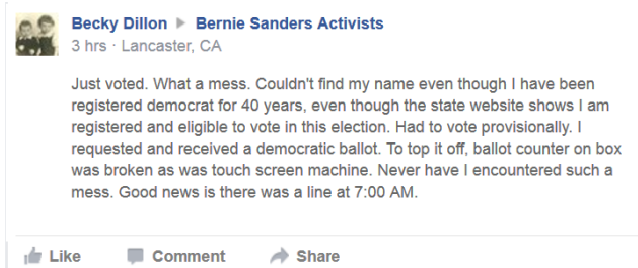
Becky Dillon ▶ Bernie Sanders Activists
3 hrs · Lancaster, CA

Just voted. What a mess. Couldn't find my name even though I have been registered democrat for 40 years, even though the state website shows I am registered and eligible to vote in this election. Had to vote provisionally. I requested and received a democratic ballot. To top it off, ballot counter on box was broken as was touch screen machine. Never have I encountered such a mess. Good news is there was a line at 7:00 AM.

👍 Like   💬 Comment   ➤ Share

Fig. 1 — Facebook post: Becky Dillon, a California voter forced to vote via provisional ballot, June 7, 2016

Line 388 A (REV)-CI-10

BOARD OF ELECTIONS IN THE CITY OF NEW YORK

CERTIFICATE TO REQUEST A COURT ORDER
TO CAST A BALLOT ON THE VOTING SYSTEM

4/19/16
Date

TO THE SUPREME COURT OF THE STATE OF NEW YORK:

We, the Inspectors of Elections of ___66___ Election District ___53___ Assembly District

Name and Address of Poll Site

In the County of ___Kings___ certify that:

___CHLOE PECORINO___
Name

has not been permitted to cast a ballot on the voting system because:

☐ Voter's name does not appear in the Poll Site Voter Registration List.

☐ Poll Site Voter Registration List shows Voter enrolled in _____ party.

☐ Poll Site Voter Registration List indicates "ID REQ" and voter does not have ID/refuses to show it, and declines to vote by Affidavit Ballot.

☐ Voter claims s/he registered with Department of Motor Vehicles or_____(insert applicable agency) and Voter's name does not appear on Poll Site Voter Registration List.

☑ Other: _Registration was lost in mail._

This applicant may go to the Supreme Court Justice presiding in the Borough after going to the Board's borough office for determination of his or her right to cast a ballot on the voting system for this Election District.

_____
Republican Inspector

_____
Democratic Inspector

Fig. 1A — Court order request of Chloe Pecorino, a first-time New York voter whose registration was lost in the mail. She was unable to cast a regular ballot in the Democratic presidential primary despite requesting a court order on April 19, 2016

Documents courtesy of Election Justice USA

This is part of the text of the affidavit that Ms. Pecorino filed with Election Justice USA:

> "I, Chloe Pecorino, remain unregistered in the state of New York as a Democrat. I have made multiple attempts to confirm my voter registration with the DMV and the Brooklyn Board of Elections … I registered through a change of address form with the DMV in early March. I registered as a Democrat. I know the DMV received the form because my new address is on file as of March 18th, 2016. I never received confirmation about my registration which is why I've been calling the Brooklyn Board of Elections for the past three weeks."

Some of these tampering issues may be related to a December 2015 massive data leak, reported by researcher Chris Vickery, that included "personal, public, and some non-public information on 191 million registered voters." According to a *Forbes* article by Thomas Fox-Brewster, Vickery found "300GB of voter data, which includes names, home addresses, phone numbers, dates of birth, party affiliations, and logs of whether or not they had voted in primary or general elections. The data appears to date back to 2000." *Forbes* stated that the information was openly available online, and that "It would appear every registered US voter is included in the leak."

How Would We Know if the Voting Machines Were Not Counting the Votes Correctly?

The best way to check would be to count the ballots by hand, or examine any paper or electronic trail available in a thorough and public audit. This is not happening. According to a database compiled by Citizens for Election Integrity, only 12 states require a post-election audit of "every contest and ballot issue voted on the ballot." Even that estimate is generous. For example, New York is listed as one of those 12 states, but in 2015, its post-election audit law was changed from requiring a three-percent hand count audit to simply running those ballots through the machine again. Alan Goldston, a New York election law consultant, said "this is not a recount at all."

So the short answer to this question is: We wouldn't know if the totals were wrong. Or would we? Would there be other indications that the machine-count is not accurate?

Evidence

If voting-machine results were inaccurate on a regular basis, there would be some evidence of it. One indicator would be that votes counted by machines would give different results than votes counted by hand. In fact, this is now being seen in elections all over the country.

In the 2016 Democratic primary in Kings County, New York (Brooklyn,) a group of affidavit ballots were hand-counted by a group of volunteers. Comparing the hand-counts with the machine-counts, there is a noticeable difference (Figure 2). In every single assembly district we examined, except one, Hillary Clinton performed better when the votes were counted by machine. This is a small sample of the overall ballots cast, but the consistency of the results makes a convincing case that something is amiss.
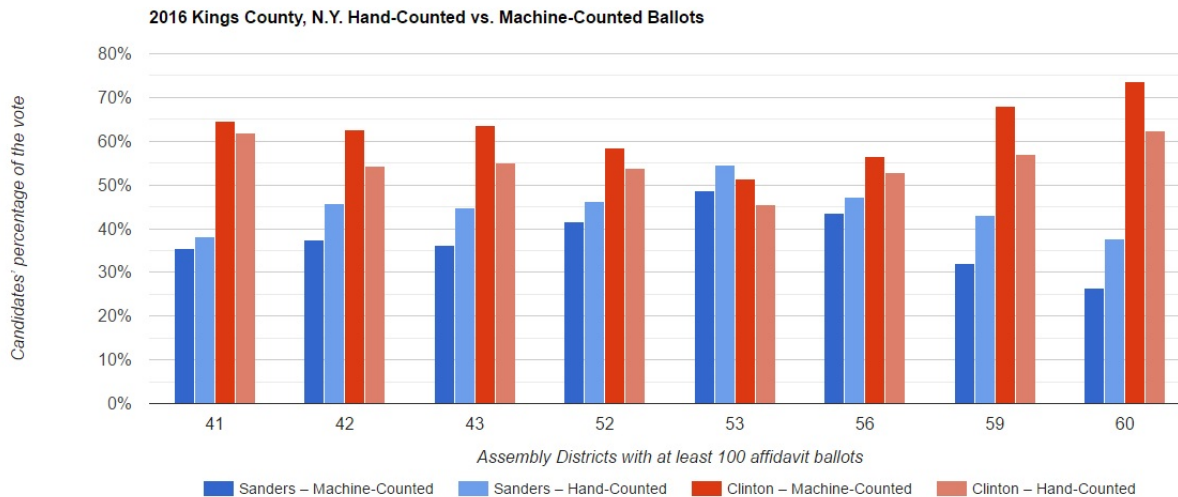
2016 Kings County, N.Y. Hand-Counted vs. Machine-Counted Ballots

Fig. 2 — Hand-counted ballots show
a consistently higher return for Sanders in the 2016 New York presidential primary
Graph by Anselmo Sampietro

Comparisons have been made previously between voting results in hand-counted precincts and machine-counted precincts. When there have been discrepancies, they have been passed off as the result of demographics. This is a reasonable concern, since it is possible that voters of a particular political perspective could tend to choose a particular type of voting equipment. However, in this instance, because the two sample sets (hand-counted and machine-counted) are from identical precincts, with voters participating in the same election on the same day—there is no demographic variable to take into account.

These affidavit ballots are from voters who were not able to vote by regular ballot. In the New York 2016 primary, over 120,000 voters were purged from the rolls in Brooklyn alone, and a large number of voters also had their voter registration changed without their knowledge or intent.

Sanders voters tend to be younger and more independent, so one might think that they would be less likely to register ahead of time, and more likely to show up in the affidavit sample. However of the over 120,000 affidavit ballots cast, only about 30,000 were actually certified and counted. It is that final "approved" subset being counted in our study. Those votes would have only included officially registered Democrats, not independents or late registrants. Those officially approved affidavit votes, when counted by hand, are showing a consistently higher percentage for Sanders than when the votes are counted by machine. There are two possible explanations for this. One is that the machines are counting the votes differently. The other is that the voters who were forced to use affidavit ballots were targeted Sanders voters. Possibly, both of these factors are at work. Either way, the data indicates the footprint of manipulation in the election, and calls into question the validity of the reported results.

Figure 2A shows the results of a recount in Hillsborough County in the 2008 New Hampshire Democratic primary. There were differences in almost every precinct between the original machine count and the manual recount.

NH PRIMARY RECOUNT
PRESIDENT OF THE UNITED STATES · DEMOCRATIC

Hillsborough County	Clinton, d	Recount		Edwards, d	Recount		Obama, d	Recount	
Amherst	970	963	-7	424	427	+3	1,309	1,308	-1
Antrim	195	194	-1	106	111	+5	272	272	
Bedford	1,630	1,630		529	530	+1	1,614	1,616	+2
Manchester Ward 3	551	558	+7	165	165		458	459	+1
Manchester Ward 4	696	702	+6	237	237		519	524	+5
Manchester Ward 5	683	619	-64	255	217	-38	404	365	-39
Manchester Ward 6	875	880	+5	348	350	+2	587	590	+3
Manchester Ward 7	753	755	+2	270	272	+2	421	421	
Nashua Ward 1	1,021	1,026	+5	439	439		744	744	
Nashua Ward 2	773	775	+2	290	290		666	666	
Nashua Ward 3	788	790	+2	338	341	+3	637	637	
Nashua Ward 4	683	685	+2	164	164		416	417	+1
Nashua Ward 5	1,030	959	-71	405	377	-28	673	678	+5
Nashua Ward 6	967	976	+9	333	334	+1	520	523	+3
Nashua Ward 7	859	863	+4	295	298	+3	480	481	+1
Nashua Ward 8	789	791	+2	291	291		681	683	+2
Nashua Ward 9	881	888	+7	390	394	+4	786	791	+5
New Boston	366	366		160	159	-1	482	479	-3
Sharon	33	33		19	19		80	80	
Temple	97	97		73	73		199	197	-2
Weare	630			255			580		
Wilton	284	285	+1	180	179	-1	386	286	-100
Windsor	13	13		15	15		15	15	

* results from NH SoS recount website as of 6:30 PM EST 1-20-08
www.sos.nh.gov/recountresults.htm

Fig. 2A — 2008 NH Dem. presidential primary
The manual re-count shows large discrepancies with the original totals
Source: The Bradblog

In the 2016 Wisconsin and Massachusetts presidential primaries, there have also been stark differences between the candidates' percentages in hand count and machine count precincts.

Hacking

The examples provided above, showing differences between hand counts and machine counts, cannot be explained by demographics. If the discrepancies are not due to demographics, there is either some issue with the voting machines or the hand counts. Information surrounding those two protocols suggests that the problem would be with the machines and not the hand counts.

In his well-researched post on the odd results of the 2016 Massachusetts Democratic primary, Theodore de Macedo Soares points out that Canada, Australia, Denmark, France, Ireland, Italy, Sweden, and Spain are among the 59 countries that rely on hand-counted paper ballots to determine their results. In contrast, concerns about security and accuracy have plagued electronic voting machines wherever they have been implemented.

In 2009, Germany's highest court banned the use of computers in the voting process amidst concerns that the process was not transparent. Jonathan Simon, a Harvard-educated attorney who is the co-director of the Election Defense Alliance, says on his website, "There's virtual unanimity among the experts who have studied electronic voting machines that insiders or hackers can change the results of elections without leaving a trace." He cites studies from Johns Hopkins, Princeton, University of Michigan, The Brennan Center For Social Justice at NYU, the states of California and Ohio, and even the U.S. Government Accountability Office to back up his claim.

J. Alex Halderman teaches computer and network security at the University of Michigan and has successfully compromised numerous voting systems. He paints a vivid and unnerving description of one hack, "Within 36 hours of the system going live, our team had … almost total control of the server software, including the ability to change votes and reveal voters' secret ballots." Halderman points out that the threat to our elections could be coming from political players inside our country—or even from abroad. Testifying before the D.C. Board of Ethics and Elections about one of his many voting-machine hacks, Halderman makes it clear that the risks are not theoretical. "While we were in control of these systems we observed other attack attempts originating from computers in Iran and China. These attackers were attempting to guess the same master password that we did. And since it was only four letters long, they would likely have soon succeeded."

You can view Dr. Halderman performing a successful hack on a voting machine in this clip from the documentary "Holler Back — [not] Voting in an American Town."

Who is Responsible?

At this point, we are unable to say who might be responsible for any data breaches to the voting equipment. There could be any number of independent players who would benefit from the victory of a

particular candidate and would be willing to take action to influence the results. Our research also indicates that in some elections the footprint of more than one unofficial player is evident.

What is that Pattern?

We are now going to utilize a different technique in our search for evidence of election-result irregularities. We will focus primarily on statistical irregularities, and by that we mean results that defy statistical laws. The technique we are using is called the "Cumulative Precinct Vote Tally Chart," also known as a CVT (cumulative vote tally) graph, or CVS (cumulative vote study). The CVT graph has a number of advantages in examining election results. Exit polls and the discrepancies between them and the official results have received a lot of attention in the 2016 presidential cycle. The CVT graph uses actual votes, and not post-vote surveys, so the results are more conclusive than exit poll comparisons. Secondly, although it is based on solid statistical protocols, it does not require statistical training to understand, and is therefore suited to helping both the statistical and the non-statistical community fully grasp the large distance between the vote totals currently being reported and the statistical norm. Third, statistician Beth Clarkson explains that election data tends to have a lot of "noise." She likes the CVT analysis because, "It allows you to see a trend that is difficult to spot in a noisy data set. "

The technique is based on the Law of Large Numbers (Figure 3.) Investopedia provides a straight-forward explanation, "A principle of probability and statistics which states that as a sample size grows, its mean will get closer and closer to the average of the whole population."
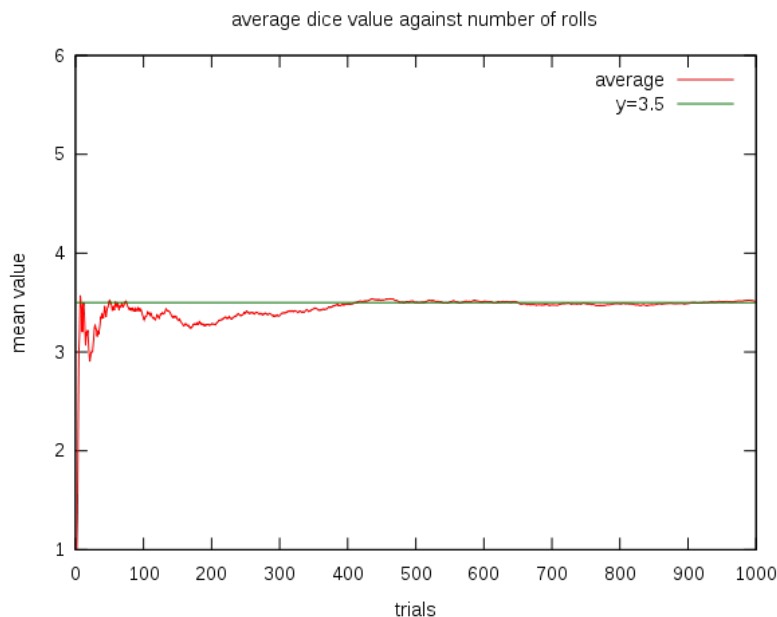

Fig. 3 — Illustration of the Law of Large Numbers using rolls of a single die
Wiki graph by NYKevin

Interpreting this law for elections, the sample size is the number of votes, and the mean is the candidate's percentage. In practice what happens is that the larger a sample of votes that you collect, the closer you should get to the candidate's average percentage of support in that locale. This is easy to see in action. If you and your friends support a candidate, it does not mean the candidate has that level of support overall. But a broader sample of voters in your community will generate a more accurate picture of the candidate's actual level of support. This is the basic concept behind all polling; and this is the principle that is the foundation for the CVT graphs.

Using CVT graphs to demonstrate irregularities in election results has been controversial. As a result, we're going to relay the methodology and backstory of the technique, confirm that it accurately demonstrates a statistical pattern that exists; investigate whether there is a demographic explanation for that pattern and explore what, if anything, the pattern signifies.

The CVT graph shows the precincts added together cumulatively from the smallest to the largest along the X-axis. On the Y-axis it shows the two candidates' percentages (Figures 4 and 4A). In these instances from 2000 and 2004, the CVT graph resembles the graph illustrating the Law of Large Numbers. Because the precincts are added together cumulatively as you move further right on the graph, it becomes harder and harder for any individual precinct to overcome the average percentage of all the votes that have been added up so far, and the data tends to chart as a flat line, at least it did until 2004. Sometime around 2004, or possibly a little earlier, other patterns emerge that we will discuss shortly.
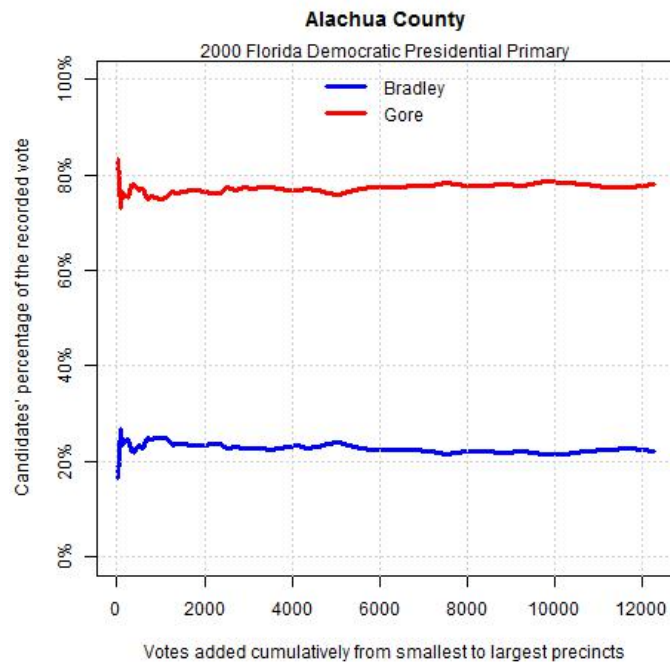

Fig. 4— 2000 Alachua Florida Democratic presidential primary
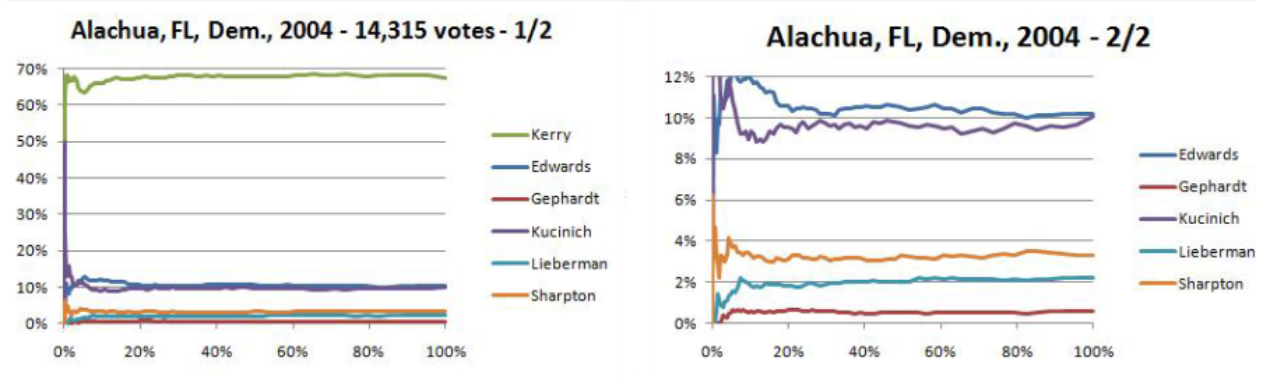
Graph by Anselmo Sampietro

Fig. 4A— 2004 Alachua Florida Democratic presidential primary
(The second graph is a zoomed-in view of the bottom half of the first graph.)
from Ron Paul Forum by "Liberty1789"

You may be surprised to see some of the above graphs credited to "Liberty 1789." One of the reasons for the controversy surrounding the CVT graph is that it was developed on the Internet by non-professionals outside of academic statistical circles by forum users posting under pseudonyms. You couldn't really ask for a worse start for a statistical method to be taken seriously.

The graph was first used in 2012 by a group of Ron Paul supporters who had strong analytical and engineering skills. The first formal presentation of the technique was made by two of those Ron Paul supporters, Choquette and Johnson, in two online papers. But according to Choquette, the idea of charting the precincts from the smallest to the largest was conceived by an engineer named Phil Evans, who used the online handle "The Man."

Evans remembers the night he first started to notice an unusual pattern in the election returns. "In 2012 I was watching CNN report on the GOP primary results in New Hampshire and what struck me was that [Ron] Paul received double the percent in small precincts as in large. I wondered what that could be." Evans designs and builds industrial machinery, and his work involves complex data analysis. He became fascinated with the question: Why would one candidate get such a larger percentage of the votes in the large precincts?

After studying the data intensively for six weeks, Evans came to a conclusion that stunned him — but also made sense. He became convinced that in the large precincts, some of the candidates' votes were being shifted to another candidate. Why only in the large precincts? It would be easier to disguise the differences, he thought. In the small precincts with only a few voters, the shift would be much more noticeable. There were at least two ways it could be done—through software in the machines; or through the software used when the totals were centrally tabulated. He wanted to illustrate the vote-switching he believed was occurring. He says, "Six weeks later I had figured out a method for expressing this using Excel and released a paper that is still online today."

Evans says his initial graphs from that paper were modified by another forum user, "Liberty 1789," into the "Cumulative Precinct Vote Tally Chart." Evans and his fellow Ron Paul supporters began using it to graph many of the election results of the 2012 Republican primary. Here is the pattern they saw in state after state: a candidate receives a higher percentage of votes in large precincts than he/she receives in small precincts. This increase occurs in a mathematically proportionate pattern, in other words, as the precincts get larger the candidate's support gets larger at the expense of other candidates. Often this increase is enough to change the outcome or the dynamic of the election. Which candidate receives increased support in the larger precincts depends on the particular race. In 2012, the candidate that benefited from the pattern in almost every race was Mitt Romney (see Figure 5).
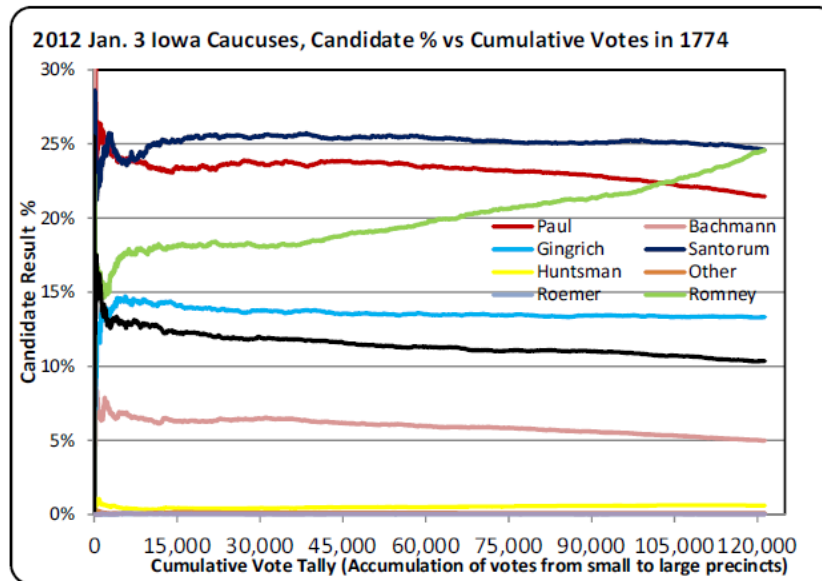
Fig. 5 — 2012 Iowa Republican presidential caucus
source: "Republican Primary Election 2012 Results: Amazing Statistical Anomalies"
graph by Choquette and Johnson

After Ron Paul lost the election, Evans suspected that Paul had been cheated. Evans says, "It was frustrating because he was giving speeches in large venues with thousands lined up outside beyond capacity, while the other candidate[s] were somewhat lonely." It didn't make sense to Evans, but he had no way to prove that Paul's votes had been stolen.

In the fall of 2012, Choquette and Johnson wrote up the findings of the forum and circulated their papers widely on the Internet and via email. In those papers they occasionally use the word "alleged;" but for the most part they made bold claims like, "When candidate Mitt Romney is on the ballot he always gains votes through Vote Flipping" (v1.6 p. 4) and, "This document exposes what may very well be the greatest case of election fraud ever to occur in US history (v 1.6 p. 20.)" They sparked considerable discussion within the election reform community. However, their study was received in the statistical community with (understandable) skepticism.

The most obvious flaw in Choquette and Johnson's paper is their claim that Democratic party elections "don't show this problem." It turns out there are many Democratic party elections that exhibit this pattern too. However, despite this weakness, their statistical graphs have been confirmed to be accurate in three separate studies (Clarkson, Plattsmouth Nebraska High School Scientific Logic Class, and Lindeman). Each of these found more elections where the pattern appears.

We asked Kellie Ottoboni, a graduate student at UC Berkeley, to confirm the accuracy of the graphs in all three of these studies and she replicated and confirmed the accuracy of one graph each of Clarkson; the Nebraska High School Scientific Logic Class; and two of Mark Lindeman's graphs.

Beth Clarkson, who conducted one of the studies, is a quality control engineer with a doctorate in statistics. She read Choquette and Johnson's paper and tested their technique herself on a number of elections. In the elections Clarkson examined in Kansas, Ohio, and Wisconsin, she found the same unusual increase for one candidate in the large precincts. As a statistician she found the results "terribly surprising."

Clarkson published an article in *Significance*, affirming both the studies' analysis, and conclusions, saying, "The data I've analyzed supports their hypothesis that we have a serious, pervasive, and systematic problem with electronic voting machines." She is currently suing Kansas election officials for permission to audit the paper trail of one of the elections she analyzed in order to compare the machine's paper records with the recorded results. So far her audit has not been permitted, and she remains concerned: "If fraud were occurring, these are the kind of patterns we would expect to see."

Clarkson discovered that statistical patterns and a candidate's percentage of the vote share, vary between different models of electronic voting equipment (Figure 6). She also found that there were statistical irregularities favoring more than one candidate, leading her to surmise, "the manipulation is not limited to a single powerful operator. My assessment is that the data reveals multiple (as least two) agents working independently to successfully alter voting results."
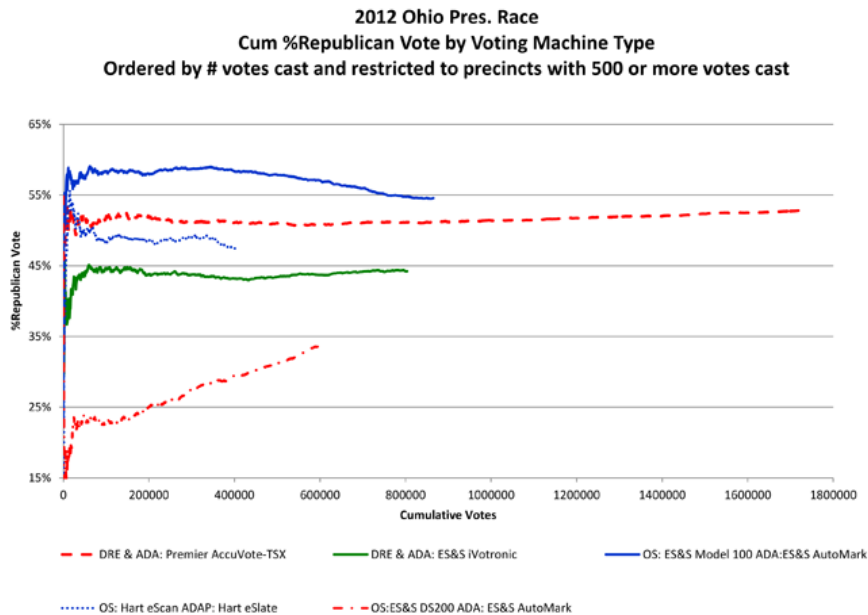
2012 Ohio Pres. Race
Cum %Republican Vote by Voting Machine Type
Ordered by # votes cast and restricted to precincts with 500 or more votes cast

Fig. 6 — 2012 Ohio presidential general election (Mitt Romney vs. Barack Obama)
Statistical patterns and candidates' percentages vary with different models of voting equipment
Graph by Beth Clarkson

In Figure 6, the ES&S DS200 vote-scanner shows an irregular statistical pattern that favors Romney (red dots.) But the ES&S DS100 and the Hart Escan show an irregular pattern that favors Obama (blue lines.) Clarkson is graphing the percentages as they impact the Republican vote, so when a line is going up, it is favoring the Republican (Romney) and when a line is coming down, it is favoring the Democrat (Obama). Kellie Ottoboni of UC Berkeley replicated Clarkson's research and confirmed its accuracy (Figure 6A).
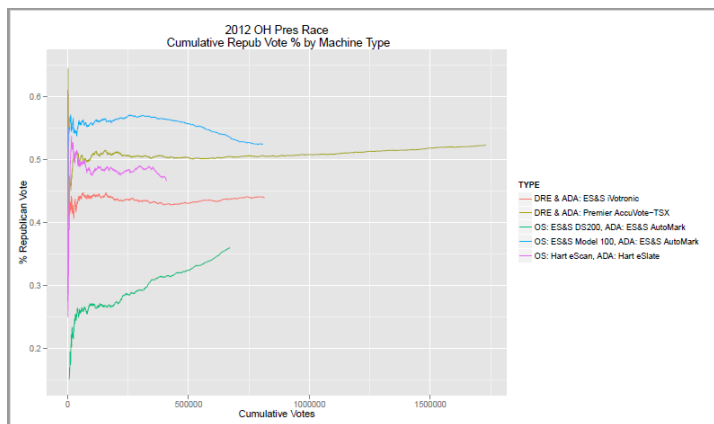
Fig. 6A – 2012 Ohio presidential general election (Mitt Romney vs. Barack Obama)
Clarkson's research was replicated and confirmed accurate by Kellie Ottoboni of UC Berkeley
Graph by Kellie Ottoboni

There are some DREs in this election demonstrating a normal (relatively flat) statistical pattern. But it would be ill-advised to conclude that those machines are secure. Based on conversations with security experts like Matt Bishop at UC Davis and Halderman at the University of Michigan, it is almost always possible to breach the security of these machines. After hacking the Washington, D.C. Internet voting pilot program, Halderman said, "If this particular problem had not existed, I'm confident that we would have found another way to attack the system."

Columbia University political scientist Mark Lindeman and data scientist Levi Bowles have both published work confirming the existence of the pattern, but arguing that it is not indicative of fraud. We found their research flawed and their logic unconvincing, and have provided a detailed breakdown of these issues later in this paper.

One fact that is clear from all these studies is that in many U.S. elections, certain candidates are receiving an increased share of the vote as the precincts get larger. The crucial question is, Why? Is there an innocuous demographic explanation for the increase? Or is it something that is indicative of error or fraud?

When Did the Pattern Begin?

We are unable to pin down exactly when the pattern originated. In a fascinating trip down election fraud memory lane, writer Victoria Collier describes numerous troubled U.S. elections. It would be instructive to do a statistical analysis on one of the races that she cites as an "up-set" like Chuck Hagel's 1996 Nebraska Senate victory. "Three days before the election … a poll conducted by the *Omaha World-Herald* showed a dead heat [but] Hagel trounced Nelson by fifteen points," Collier says. "This divergence from pre-election polling was enough to raise eyebrows across the nation."

For now, we can state that races that we examined from 2004 and earlier did not show the pattern of increased candidates' percentages in large precincts. Looking again at Figure 4, we see that in the 2000 and 2004 races in Alachua, Florida, each candidate's share of the votes is roughly the same in small and large precincts. However, by 2008, this is not the case in many races around the country (see Figure 7).
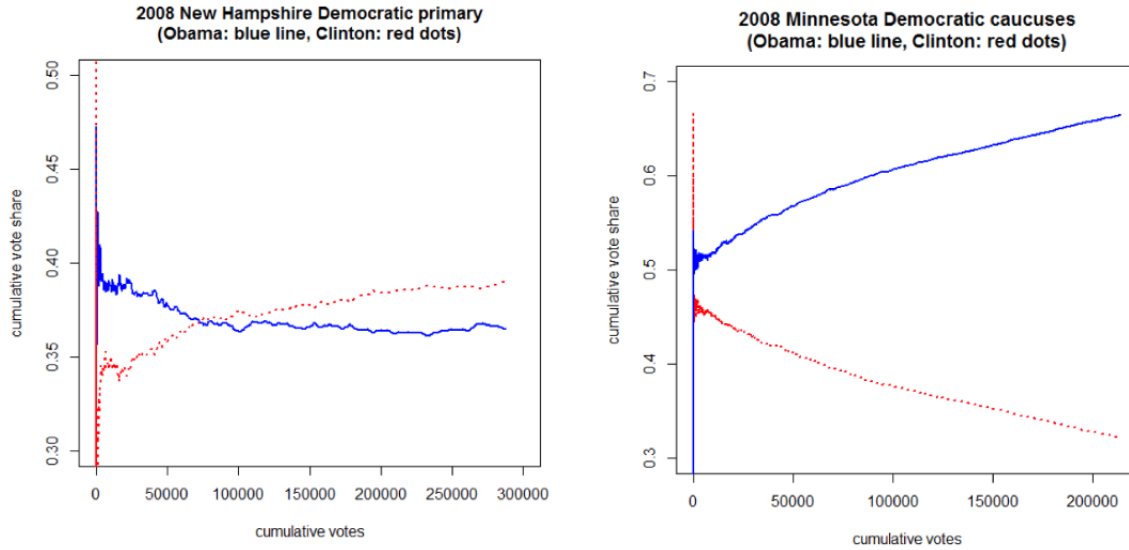
Fig. 7 — 2008 New Hampshire and Minnesota Democratic presidential primaries
In races around the country candidates' begin receiving larger percentages of the vote in large precincts
Graphs by Mark Lindeman

In the New Hampshire and Minnesota Democratic primaries, as well as in other races in 2008, candidates receive a larger percentage of the votes as the precincts get larger. In New Hampshire, the pattern benefits Clinton. In Minnesota, Obama is the one who gains vote shares in the large precincts. The 2008 New Hampshire Democratic presidential primary was also the race we looked at initially where the manual recount did not match the original machine totals (Figure 2A).

Figure 8 shows the Wisconsin Republican primary in two different election cycles. In 2000, no candidate has much of an increase in the large precincts. But in 2016, Ted Cruz's percentage noticeably increases in the large precincts, while Donald Trump's percentage of the vote goes down.

These two comparisons also demonstrate that the pattern is happening in both Democratic and Republican races.
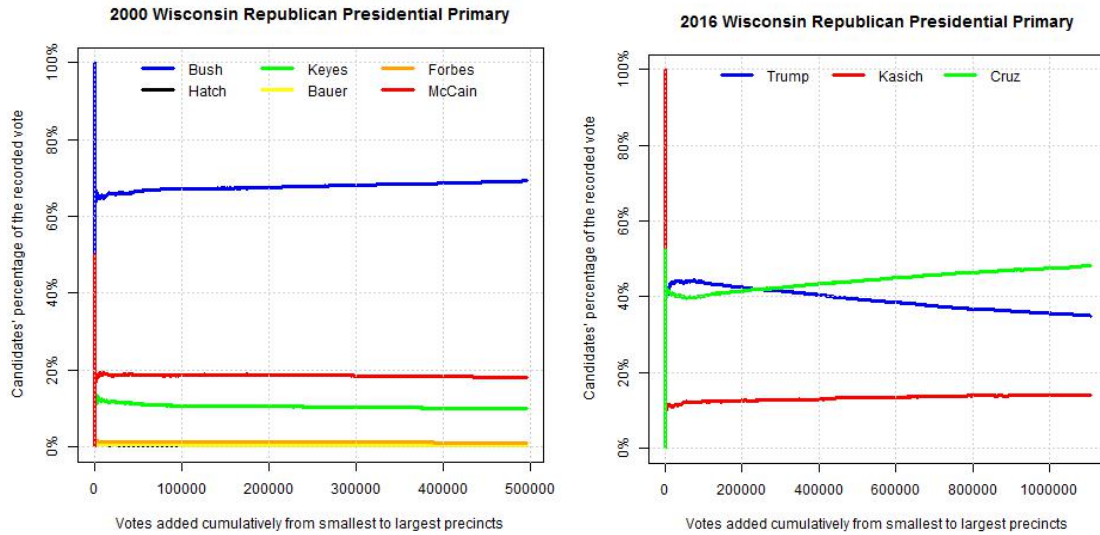
Fig. 8 – 2000 and 2016 Wisconsin Republican presidential primaries show different statistical patterns
Graphs by Anselmo Sampietro

Doing further research on historical races will help identify possible early appearances of the pattern. Collier says, "Throughout the 1980s and 1990s, the use of optical scanners to process paper ballots became widespread." But probably the most seminal year for electronic voting equipment was 2002, when states across the country experienced a large influx of computer-based voting systems, with the passage of the (perhaps ironically named) Help America Vote Act.

Smooth, Unidirectional, and Mathematically Predictable.

Figure 9 shows a graph of the 2016 Louisiana Democratic Primary. The analysis is by Beth Clarkson and Anselmo Sampietro confirmed its accuracy.
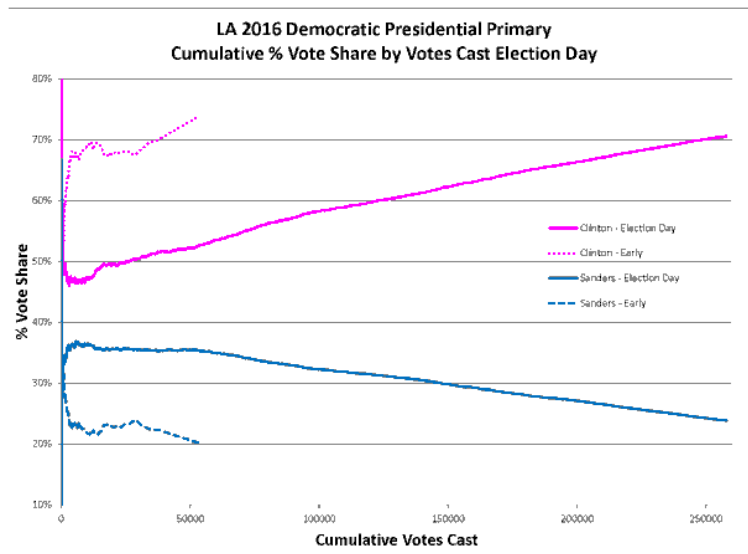
Fig. 9 — 2016 Louisiana Democratic presidential primary
Candidates' vote shares vary by as much as 37% between small and large precincts
Graph by Beth Clarkson

This graph is in complete violation of the Law of Large Numbers. For a candidate to receive this level of increased support in the large precincts, each new precinct must be so heavily weighted that it defies the average of all the other precincts that have already been added together. This is a major statistical irregularity.

In the small precincts, the difference between Clinton and Sanders is approximately 10% (Clinton 48%–Sanders 38%). However, in the largest precincts the difference between the candidates is 47% (Clinton 70%–Sanders 23%.) That is a difference of 37% support between the smallest precincts and the largest precincts.

To see how heavily weighted the large precincts are, we graphed them separately, county by county, dividing the largest 25% from the remaining 75% (Figure 10). Within almost every county, Clinton receives a higher percentage of the vote in large precincts by unusually high margins, sometimes by close to 40%. In Washington County you can literally see the moment that the data starts to change around 600 votes.
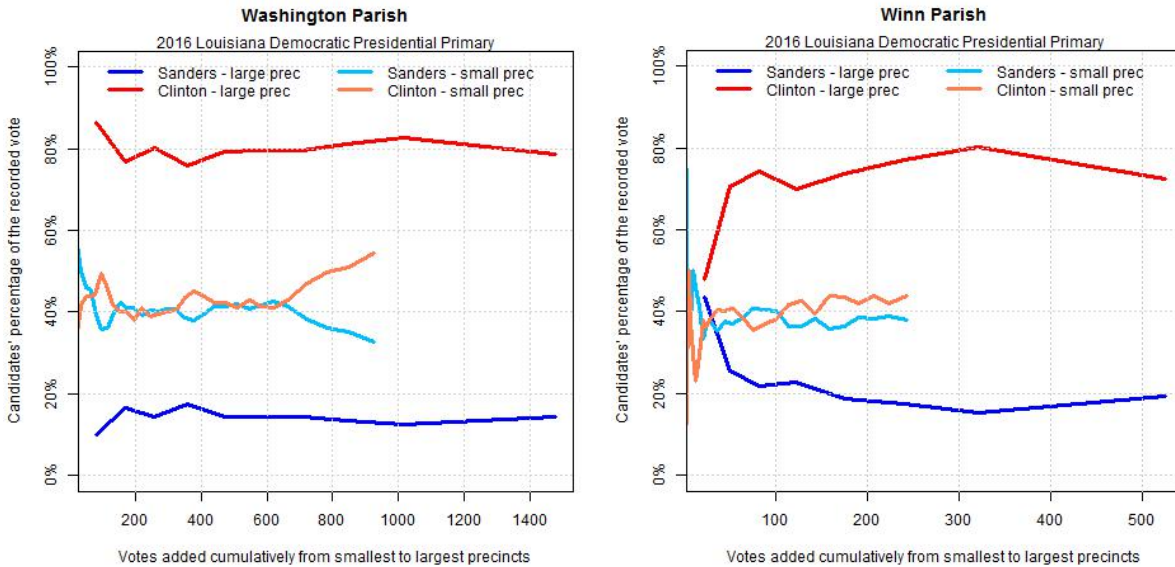
Fig. 10 – 2016 Louisiana Dem. presidential primary, county level data - the difference in the candidates' percentages between small and large precincts is unusually high - in some cases close to 40%

Graphs by Anselmo Sampietro

There are three other characteristics of this data that are suspect:

1) **The data is smooth**. The lines in the overall state chart go straight up and straight down; and lines of data in the large precincts are also quite straight. This is what Dr. Scheuren is referring to in the opening of the paper when he says, "There is a greater degree of smoothness in the outcomes than the roughness that is typical in raw/real data."

2) **The data is unidirectional.** In the statewide results, the data only moves in one direction: Clinton goes up; Sanders goes down. The percentages never demonstrate the kind of ups and downs caused by organic voting behavior.

3) **The data follows a mathematically predictable pattern.** Clinton's support is increasing in a mathematically predictable way. In each progressively larger precinct she gets a slightly larger level of support. This is a possible indication that a mathematical algorithm has been applied to the results.

Demographic Factors

The data we've examined so far shows that the pattern emerged in various locations sometime after 2000. This argues against a demographic explanation, since the demographics are consistent, and it is the time period that is changing.

Digging deeper into the Louisiana data further undermines the theory that demographic factors are responsible for the increased vote shares in large precincts. Here, the sheer size of the difference between the small and large precincts has no plausible demographic explanation.

Nate Cohn at the *New York Times* has insisted that the candidates' percentages being correlated with precinct size in the South is due to a large concentration of black voters in the precincts that support Clinton.
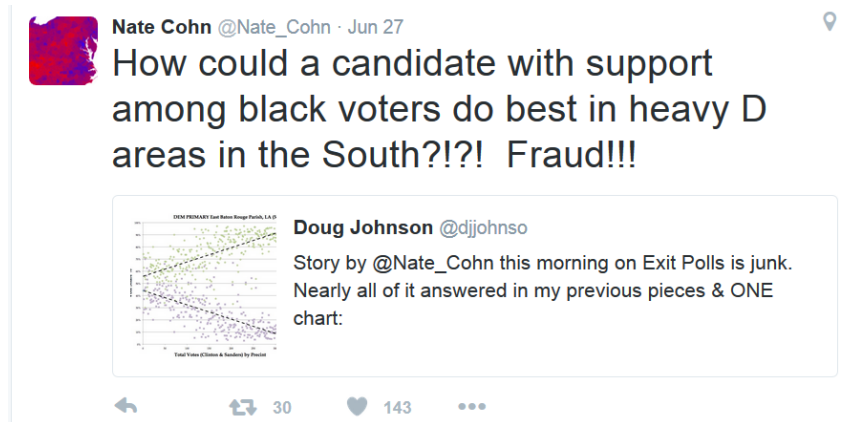

Fig. 11 – Nate Cohn's tweet makes a weak, sarcastic argument, implying that all of the statistical irregularities in the South can be explained by Clinton's lead among black voters

In a tweet posted on June 27, 2016 (Figure 11), Cohn takes an unfortunately flip attitude toward the evidence—which allows him to pass on analyzing the data, but does not help anyone else sort out whether this is a pattern that we need to be concerned about. The underlying question of Cohn's tweet is this: Is there a demographic explanation that can account for a 37% shift in support for Clinton from the smallest precincts to the largest precincts? For instance, is this correlation happening because Clinton's support is so strong among blacks and there are so many additional blacks in the large precincts that it can account for a 37% increase?

Just a brief look at the demographics in Louisiana, contradicts this theory. The 2010 census summary file (search Washington Parish, Louisiana – and then click on "Race and Hispanic or Latino origin, for example) indicates that both Washington and Winn parishes have a black population of 31% (Figure 11A). So even if 100% of black voters were supporting Clinton, (which they are not) and even if they all lived in the large precincts, it could not explain a 37% increase in her support in the large precincts.

Geography: Washington Parish, Louisiana				Geography: Winn Parish, Louisiana		
Subject	**Number**	**Percent**		**Subject**	**Number**	**Percent**
RACE				RACE		
Total population	47,168	100.0		Total population	15,313	100.0
One race	46,649	98.9		One race	15,142	98.9
White	31,482	66.7		White	10,223	66.8
Black or African American	14,625	31.0		Black or African American	4,702	30.7

Fig. 11A – Census data shows there is not a large enough black population in Winn or Washington Parish to explain the percentage of support Clinton is receiving in the large precincts

Even if you could explain the increase in Clinton's support in Louisiana by saying that there are more blacks in those large precincts, it does not explain Cruz's increased support in large precincts in Wisconsin, or Romney's increased support in large precincts in almost every state in 2012. While one might be inclined to look to demographics to explain discrepancies in one state or another, the fact that the *same* pattern is present across multiple states with very *different* demographics argues strongly against it being caused by demographic factors alone.

Despite the lack of hard data, the idea that these correlations can be explained demographically remains popular. Eitan Hersh, a member of the political science faculty at Yale, told us that he would bet money on a demographic explanation. We appreciate his gambling spirit, let's see if his confidence is well-placed.

First, let's take a close look at the prevailing narrative that the most predictive demographic factor of whether a voter will support Clinton or Sanders is race. This is actually not the case. Articles in *The Atlantic*, the *Los Angeles Times,* and *Vox* have all reported that the single most salient factor in determining whether a voter will support Clinton or Sanders in the 2016 primary is age. *Vox* quoted political scientist Alan Abramowitz as saying, "It was age, and beyond that nothing mattered," after he "ran a multivariate analysis to help figure out this question."

In the same *Vox* article, Jeff Stein refines the narrative that Clinton is beating Sanders among black voters by pointing out that "several polls have put Sanders ahead of Clinton among *young* African-Americans; in the Reuters polling data, for instance, Sanders beats Clinton by 25 points among black voters aged 18 to 29." According to Lee Miringoff, director of the Marist College Institute for Public Opinion, in the *Los Angeles Times* piece, "The age factor seems to trump everything."

How much advantage does Clinton have among older voters? We took the average of three polls to determine Clinton's approximate statistical lead with older voters (which, based on these polls, we're defining as 45 and older):

- NBC News/*Wall Street Journal* poll from April 18, 2016 (quoted in the *Los Angeles Times*) gives Clinton a 27% lead with voters over 50
- Iowa entrance poll gives Clinton a 23% lead with voters 45–64 (*The Atlantic*)
- Iowa entrance poll gives Clinton a 43% lead with voters over 65 (*The Atlantic*)

Averaging these three polls show Clinton with a 31% lead among older voters. So even if the largest precincts in Louisiana consisted of only voters over the age of 45 (which clearly they do not), you still could not explain a shift of 37%, using the most important demographic factor in the contest.

Is it possible that a combination of factors such as race and age could explain the differences in these percentages? In theory, yes. But the precincts that would have to exist to justify these percentages have no bearing on reality. They would need to be unusual neighborhoods where older black and white voters live together in mixed-race, large precincts, with very few young people. In actuality, black neighborhoods tend to consist of families and people of all ages; and mixed-race neighborhoods are not typically made up primarily of people over forty-five – not in Louisiana – and not in other parts of the country where we see this pattern occurring over and over again.

In Louisiana, the demographic argument seems to evaporate when the data is carefully investigated. What explanation *is* supported by the data? Let's look at the 2016 Wisconsin primary to answer that.

Phil Evans, the engineer who originated the "Cumulative Precinct Vote Tally Chart," did an analysis of the 2016 Wisconsin Republican Primary (Figure 12). The graph compares hand-counted counties to the machine-counted counties in the rest of the state. The pattern of larger vote shares in larger precincts is not present in the hand-counted counties. All candidates receive approximately the same percentage of the vote in both small and large precincts. However, in the machine-counted counties, Cruz does better as the precincts get larger; Trump fares worse than he did in the hand-counted counties and 17% worse in the largest precincts. Sampietro confirmed these results and also charted them as scatterplots (Figure 12A).
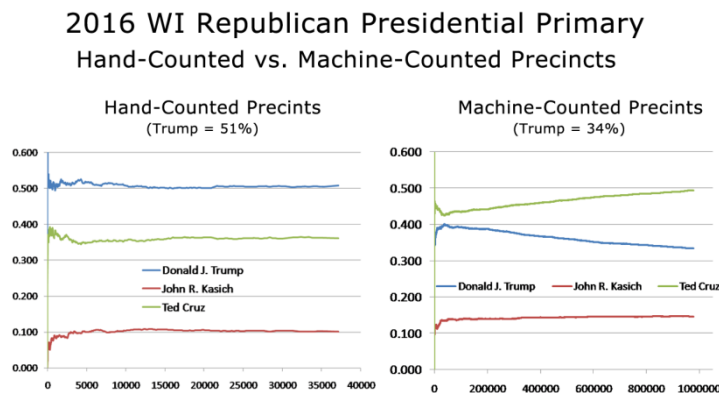

Fig. 12 – 2016 Wisconsin Republican presidential primary
Hand-counted counties and machine-counted counties show different results
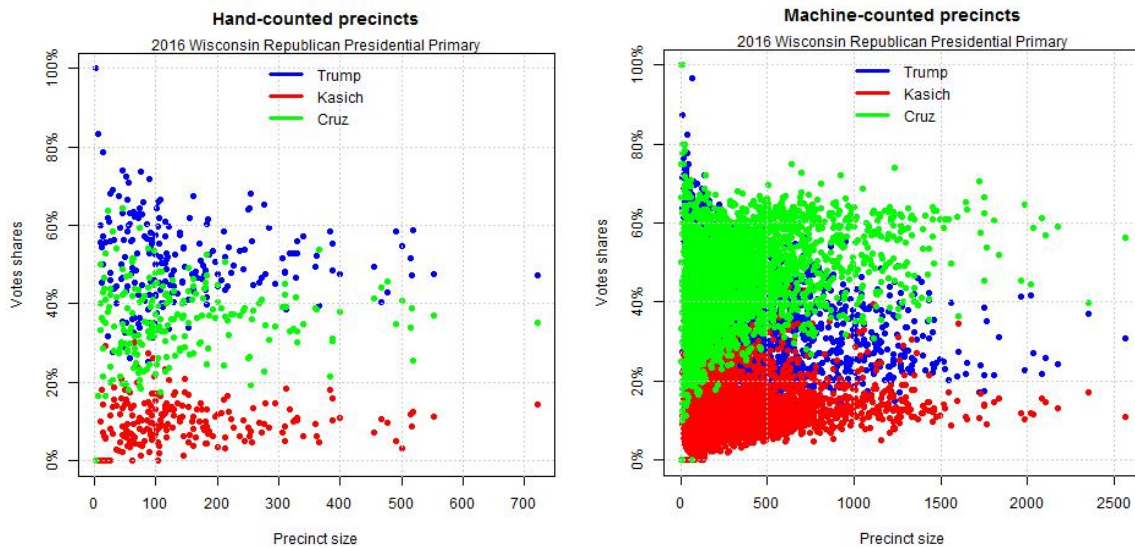*graphs by Phil Evans*

Fig. 12A – 2016 Wisconsin Republican presidential primary
Hand-counted counties vs. machine-counted counties (scatterplots)
graphs by Anselmo Sampietro

We have created scatterplots of our research, but in general are demonstrating the line chart, because the pattern is more visible there. Each dot on the scatterplot represents the percentage of votes a candidate receives in a single precinct: the higher the dot, the larger the percentage of votes for the candidate in that precinct. The further to the right a dot is, the larger the precinct. In the hand-counted Wisconsin results, the dots are level: precinct size doesn't influence the percentage of votes for each candidate, and the percentages are relatively consistent. In the machine-counted results, Cruz's percentage of votes systematically increases as the precincts get larger. The descending share of votes for Trump is also visible.

Once again, one candidate does inexplicably better in the large precincts, but this time there is also an association with the voting equipment—because unlike Louisiana, Wisconsin counts some votes by hand—allowing for an added layer of comparison. We have an association, but is it a cause and effect association? Are the machines causing this increase in a candidate's totals, or is there some other explanation?

The choice of voting equipment is not random. For example, it could be that the precincts that choose to count their votes by hand are also precincts that include more "independent thinkers" and are therefore more attracted to the maverick campaign of Trump.

If that is the case, then we would expect the same independent thinkers in the same precincts on the Democratic side to support the maverick campaign of Sanders. Is that what we find? Do the Wisconsin hand-counted precincts give Sanders a greater percentage of support than the Wisconsin machine-

counted precincts? No, they do not (Figure 13). Sanders receives approximately 57% of the vote in both hand-counted and machine-counted precincts in that race.
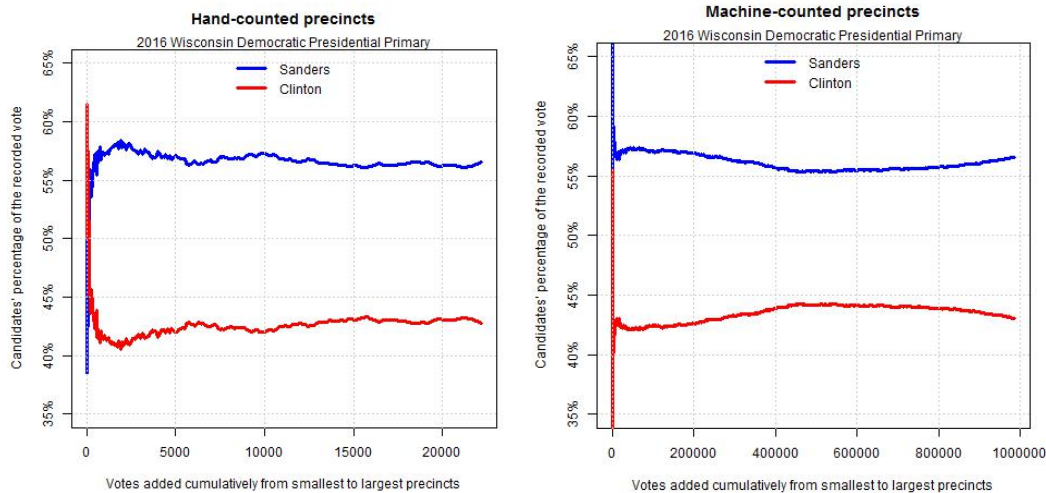

Fig. 13 – 2016 Wisconsin Democratic presidential primary - unlike the Republican presidential primary this race shows very little correlation between precinct size and any candidates' percentage of the vote
graphs by Anselmo Sampietro

This seems strange. If it's a demographic pattern, why isn't it evident on both sides? The media has commented on the [striking similarities](#) between the Trump and Sanders campaigns. Both candidates are political outsiders, delivering scathing criticisms of the status quo and opposing trade deals. Both campaigns have done well with white male voters. It would seem likely that a demographic pattern large enough to affect the Republican race by 17% would also have some impact on voters in the Democratic race.

Key Arguments Part 1

We have now established six key arguments that the pattern we are witnessing is irregular; is not due to demographic factors; and is in some way connected with the use of electronic voting equipment:

1) Data shows that elections in the same location developed the pattern after 2000, contradicting a demographic explanation.*
2) The pattern violates the Law of Large Numbers.
3) The pattern is smooth, unidirectional and mathematically predictable.
4) The large discrepancies in the percentages between small and large precincts defy a plausible demographic explanation.

5) The pattern is evident in a wide variety of geographies, with very different demographics, further contradicting a demographic explanation. We have already shown the pattern in Iowa, Ohio, Wisconsin, Florida, and Louisiana. The pattern has been confirmed by two analysts in the 2016 presidential primary in Connecticut, Delaware, Florida, Illinois, Kentucky, Louisiana, Massachusetts, New York, North Carolina, Ohio, South Carolina, Tennessee, West Virginia, and Wisconsin. It has additionally been confirmed by at least one analyst in the 2016 presidential primary in Alabama, California, Georgia, Oklahoma, and Texas.

6) The pattern appears in machine-counted races, but not hand-counted races. Public hand counts with a secure chain of custody are considered internationally the most secure form of counting votes, so this points to the voting machines as a likely source of the problem.

*Phil Evans told us that the only races they found the pattern in prior to 2000, were in elections where an individual or officials had been convicted of fraud. We have not yet had a chance to verify that information.

New York State — Multiple Issues Raise Serious Concerns

In the 2016 New York Democratic presidential primary, the CVT graphs of certain counties are irregular (Figure 15). Both Richmond County (Staten Island) and Kings County (Brooklyn) reveal a strong correlation between precinct size and candidates' percentages.
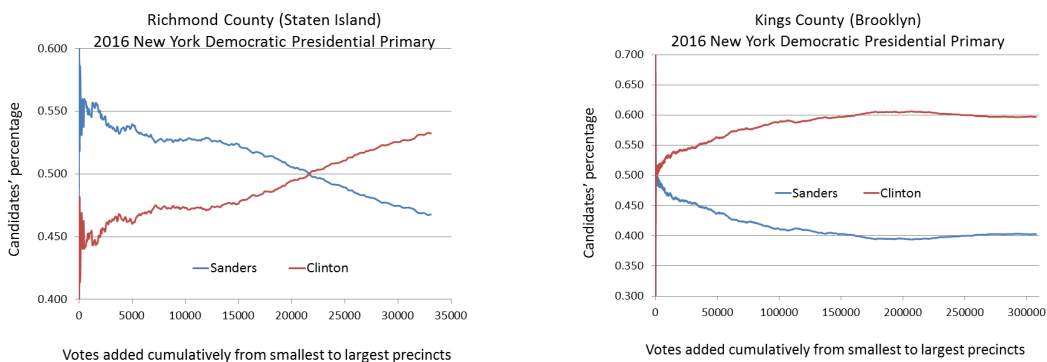

Fig. 15 — 2016 New York Democratic presidential primary
Richmond and Kings County, both in New York City, show irregular CVT graphs
graphs by Phil Evans

Earlier in the paper we demonstrated that hand-counted and machine-counted ballots in Kings County give different percentages for the candidates. The suspect CVT graph in Kings County reinforces the

findings of irregular election results presented earlier, and paints an increasingly troubled portrait of potentially compromised vote totals. But there is another odd factor in this state's results.

Precise Percentages

We are showing the New York City graphs (Figure 15) to illustrate a very specific point. The data supports the idea that the overall state vote totals are being massaged to achieve a predetermined percentage.

The final reported totals in New York state were almost exactly 58% Clinton to 42% Sanders. It is necessary to go to the third decimal digit to see a difference: 57.995 versus 42.005.

[Doug Johnson Hatlem](#) reported on this issue, pointing out that, "The overall results in New York, as announced on election night, deviated from a perfect 58–42 split by 0.005345. That's 97 votes out of over 1.8 million."

In Kings County, the reported totals were almost exactly 60% Clinton, to 40% Sanders: (59.72% Clinton, 40.27% Sanders.) The .27 difference is caused by about 800 votes out of 300,000. In the Bronx, the percentages were almost precisely 70% to 30% (69.59% Clinton to 30.41% Sanders), the difference being just 616 out of 151,908 total votes. The Bronx County CVT graph is also irregular (Figure 16).
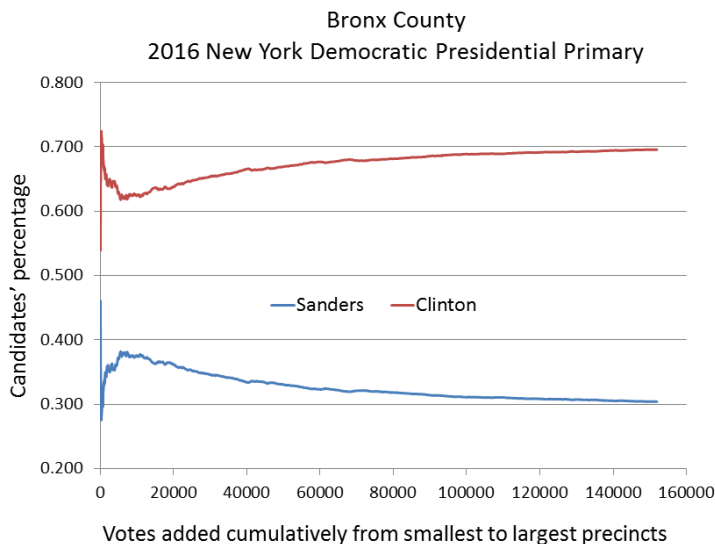

Fig. 16 — 2016 New York Democratic presidential primary - the Bronx County CVT graph shows a strong correlation between precinct size and candidate percentage

graph by Phil Evans

In an interview, J. Alex Halderman was confident that it would be possible to assign a percentage to a particular candidate, "If you're manipulating the central counting systems, then you can make an overall adjustment." Asked whether it's possible to get access to the central counting software, he replied, "Probably. It's been our experience in the last decade with the Diebold systems, for instance, that it was certainly possible in the machines and counting systems we examined."

New York City already had two unusual pieces of data giving cause for concern:

- The difference between hand-count and machine-count results
- Irregular CVT graphs in multiple counties

Add to that:

- Precise percentages in the total election results could be an indication that the results have been manipulated to achieve a specific percentage for the candidates.

These are three separate but reinforcing facts, illustrating why the totals for this state are suspect.

A Normal Graph

Columbia County, New York — a hand-counted county — offers a good point of comparison to the above graphs (Figure 17).

Columbia County (Hand-Counted)
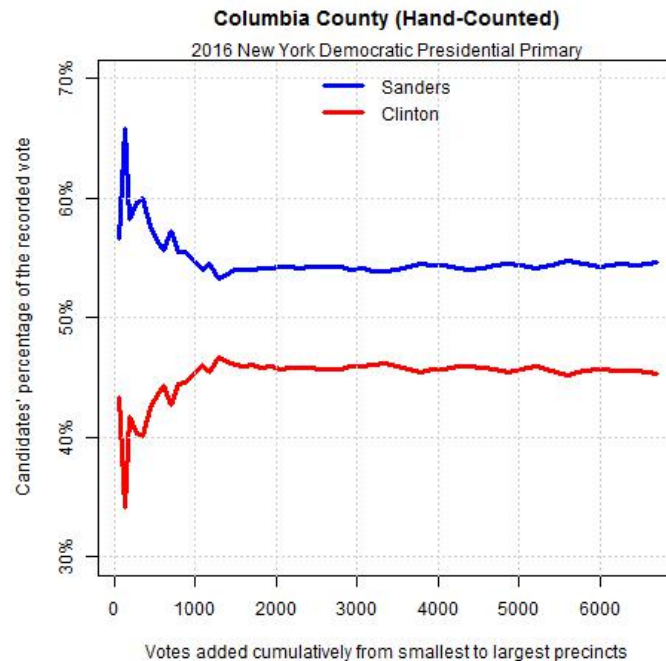2016 New York Democratic Presidential Primary

Fig. 17 — 2016 New York Democratic presidential primary, Columbia County
This hand-counted county has a normal CVT graph
graph by Anselmo Sampietro

In this graph, there is a large degree of fluctuation both up and down on the left side of the graph. Then, by about 1,300 votes, the graph settles into a fairly even straight line, with small, random fluctuations. It maintains a basically flat line through all of the largest precincts. This is a very normal-looking CVT graph.

Columbia County Democratic Election Commissioner Virginia Martin says they look at every single ballot in competitive races. In an interview with the *Register-Star,* she describes their process: "When a voter scans a ballot, it drops down into a ballot bag at the bottom of the optical scanning machine. At the end of the day, two inspectors, one Republican and one Democrat, open the machine, pull out the bag and zip it shut." Then they proceed to count 100% of the ballots by hand. In an interview on public radio, she said that Columbia County's vote-counting process is focused on security, accuracy, and transparency: "We are very, very careful about the chain of custody … Everything gets hand counted, and the results

that we certify are based on that hand count." She concluded by pointing out that the hand count is open to the public.

Opportunity, Incentive, and Rationalization

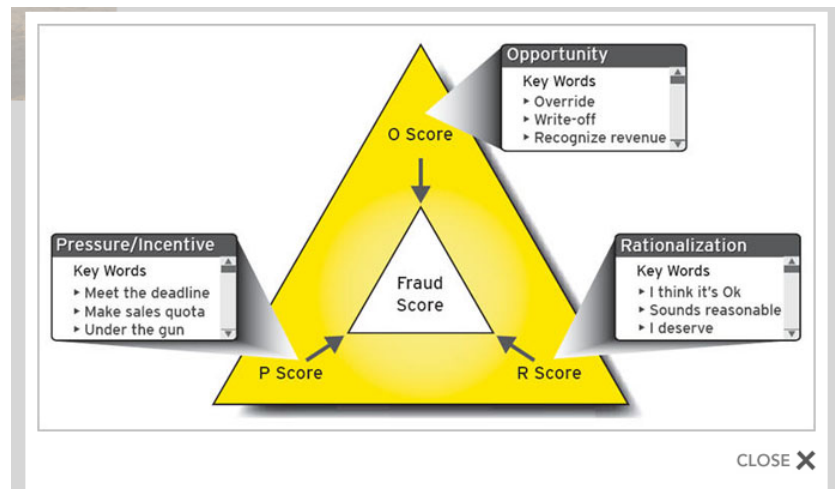
Fig. 18 – The Fraud Triangle: opportunity, incentive, and rationalization

Fraud investigations in other industries look for three factors: opportunity, incentive, and rationalization. The ease with which Dr. Halderman and other security experts have been able to penetrate and control the software of the voting equipment, makes it apparent that the opportunity exists.

The other factors, incentive and rationalization, are also present. "There's a lot at stake," says Jonathan Simon of the Election Defense Alliance. "Control of the American government, control of the economy, control of the military. When there's a lot at stake there's an incentive to fraud."

Rationalization? It's unfortunate, but there are individuals and groups active in the political landscape who feel justified taking actions that cross ethical boundaries. On the Democratic side, leaked documents revealed that the Democratic National Convention supported Clinton as the eventual nominee, although their public position was one of neutrality. The Guccifer document, "2016 GOP presidential candidates," says, "Our goals in the coming months will be to frame … the eventual nominee early and to provide a contrast between the GOP field and HRC [Hillary Rodham Clinton.]"

On the Republican side, former Florida Republican Party chairman Jim Greer testified in 2012 that Republicans had meetings about "keeping blacks from voting."

These are just two examples. The reality is that political campaigns have operatives who work to discredit and target the other side, and that these campaigns sometimes cross ethical boundaries.

Rising Threat of Data Attacks

The advanced age of many of the voting machines being used in the areas we studied makes it quite possible that at least some of the strange patterns we are witnessing are due to hardware or software failures. But the unidirectional results of the statistical patterns make it unlikely that error is the central source of these issues.

In answering the question of whether or not the data in our voting machines is being breached, it is important to be aware of how many major government, financial institutions, and high-profile industries in the country have been successfully attacked. In May 2016, New York Attorney General Eric Schneiderman released a statement saying, "his office has received an over 40% increase in data breach notifications involving New Yorkers so far this year." IRS Commissioner John Koskinen told *Fortune* magazine, "We are basically attacked or at least probed over a million times a day." One of those attacks on the IRS resulted in a spectacularly successful May 2015 hack in which, "hackers had used Get Transcript [an IRS tool] to steal the personal information of 724,000 people." The U.S. Army and Navy, the Pentagon, and NASA were all successfully penetrated by a hacker named Gary McKinnon. The *Guardian* quoted McKinnon as saying he could scan "65,000 machines in less than nine minutes." This was between 1999 and 2002; hackers have developed considerably more sophisticated protocols since then.

Banks and corporations are also defending against constant attacks on their data and financial resources. In February 2016, CNN reported that hackers stole cash from 100 banks and rigged ATMs to spew cash in "one of the largest bank heists ever," totaling approximately $1 billion in stolen funds. Other financial institutions that have been successfully attacked include JPMorgan Chase, Citigroup, the Federal Reserve Bank of New York, and security analysts working for Bank of America. The hacker collective Anonymous has, by themselves, hacked the Church of Scientology, Hidden Wiki, San Francisco BART, the Department of Justice, and the World Trade Organization. Major corporate attacks include the Target data breach that exposed the financial information of 40 million customers, the Sony email scandal, and Adrian Lamo's attack on the *New York Times*, which was apparently so easy that according to geek.com, Lamo "created an entry in the [New York Times] Op-Ed database for himself, complete with cellphone number, real name, and e-mail address. In the description field of the database, he simply entered 'Computer hacking, national security, communications intelligence.'"

Evaluating the Opposition

Two reports have been filed that, while confirming the pattern of increased candidate percentages in large precincts, argue that it is not an indication of error or fraud.

In November 2015, Mark Lindeman, a political scientist at Columbia University, confirmed the existence of the pattern. But he took issue with the conclusions of previous studies, dismissing their analysis as "unsupported" and having "no foundation." Lindeman's analysis does not hold up to close scrutiny. He attacks Choquette and Johnson without presenting any real data to back his assertions. He often misinterprets evidence and selectively ignores facts that don't match with his theory. His addendum critiquing Clarkson is equally weak. Clarkson agrees that his analysis is not statistically supported, writing to us, "My own work including share of registered Republicans shows that even when that data is included, the number of votes cast remains a significant factor, which contradicts his analysis."

A second data analyst, Levi Bowles, covered Clarkson's work in a series of five blog posts titled, *Kansas Election Fraud*. Bowles is concerned, as are we, that the patterns we are witnessing could be due to demographic issues, and he makes his point emphatically in his comments, "…there isn't good statistical evidence that the machines are working incorrectly. The evidence, is that there is an underlying correlation [of increased percentages with larger precincts], though after we recognize the world is complex and creation/existence of precincts is not a random, stochastic process, we see that correlation completely disappear."

Bowles' critique does not provide an explanation for the appearance of the pattern since the year 2000. Precincts have never been randomly created districts. So why wasn't this pattern present in earlier elections?

Furthermore, the creation of precincts is impacted by a variety of factors, but those factors vary from state to state. Yet we are witnessing the pattern across a broad selection of states, with strikingly different demographics such as Wisconsin and Louisiana. Each of those states has its own process for creating precincts, so the existence of the pattern across such a broad array of states argues against it being due to factors that affect precinct formation.

To summarize our critique of this effort: Bowles goes beyond rhetorical arguments and utilizes a statistical analysis to compare the correlation of demographic factors with the correlation of precinct size; but he fails to apply the right statistical model, delivers weak or non-relevant evidence, and eventually does validate Clarkson's work, against his own intent.

Large Republican Precinct Theory Not Valid

Bowles does come up with one intriguing theory we found worth pursuing. He speculates that the reason for Republican candidates gaining larger percentages in the larger precincts is that the largest precincts are actually conservative-leaning suburban precincts. He creates a map that shows that at least some large precincts are conservative in their political outlook. We asked Clarkson to separate out the suburban "conservative" precincts from the "democratic-leaning" urban precincts and graph them separately. If Bowles' theory is correct, that the correlation of higher Republican percentages is due to large conservative-leaning suburbs, then the correlation will show up even more strongly in those conservative suburbs. Is that what happens?

No. In the Republican-leaning suburbs, there is no increase in the Republican candidate's percentage in the large precincts (Figure 19). There are fewer Republicans inside the city of Wichita (37% vs. 48%), but the increase in the share of the Republican vote in the larger precincts is plainly evident there. Bowles theory, that this statistical pattern is due to the existence of large conservative precincts in the suburbs, is not correct—at least in this instance. Whatever is causing the pattern, here it is visibly impacting only the inner-city Democratic-leaning precincts. That raises the question of whether these inner-city precincts have been targeted in some way.
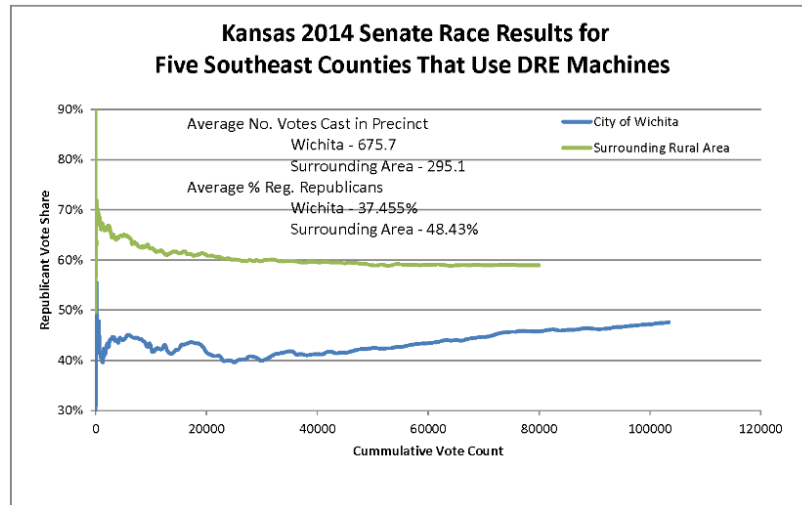

Fig. 19 — 2014 Kansas Senate race - the increase of a candidate's percentage in the large precincts is only seen in the inner city precincts, not the suburban precincts

Graph by Beth Clarkson

Key Arguments Part 2

Continuing with our list of the key points that we have established:

7) Election results with precise percentage points, in races that are already showing statistical irregularities, raise further concern that the results may not be based on actual votes.
8) The consistency with which one candidate benefits from the pattern argues against it being generated by a random computer error.
9) In the Kansas 2014 Senate race the irregular statistical pattern is evident inside the city limits of Democratic-leaning Wichita, but not in the more Republican suburban areas. This argues against the theory that the pattern is due to large suburban precincts being made up of a greater percentage of Republican voters.

Left Side / Right Side Fluctuation

There tends to be a lot of fluctuation on the left side of the CVT graph. This is to be expected, because in small precincts, before a lot of votes are accumulated, the graph will swing from the influence of just a few votes. Occasionally we have seen graphs with such steep unidirectional movement on the left side of the graph that it is concerning. The 2016 Delaware Democratic presidential primary is an example of this (Figure 20).
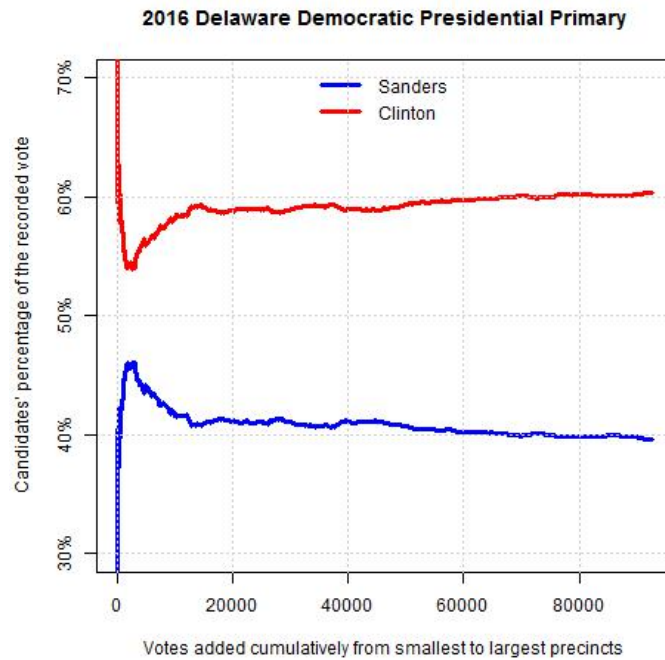

Fig. 20 — 2016 Delaware Democratic presidential primary
The slope on the left is unusually steep and unidirectional
graph by Anselmo Sampietro

We found screenshots online confirming that the actual vote count for Sanders had decreased sharply after early reporting (Figure 21). These two factors together lead us to say that the results in Delaware bear further investigation and may not be accurate.
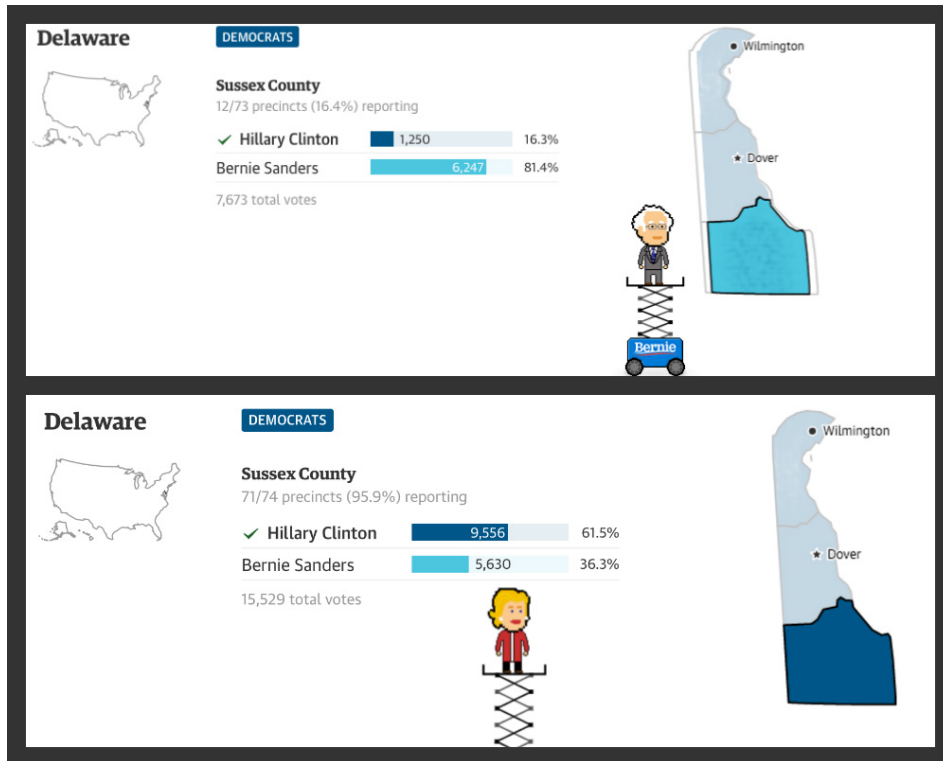
Fig. 21 — 2016 Delaware Democratic presidential primary, Sussex County
Sanders votes go down from 6,247 to 5,630
screenshot courtesy of Reno Berkeley

Reversed Effect

In the 2016 primary election, on two occasions we saw irregular charts that tilted in Sanders's favor (Figure 22). But this was rare. Our theory is that vote totals are being manipulated to achieve a pre-determined percentage, and that in this process votes might be shifted at various times between multiple candidates.
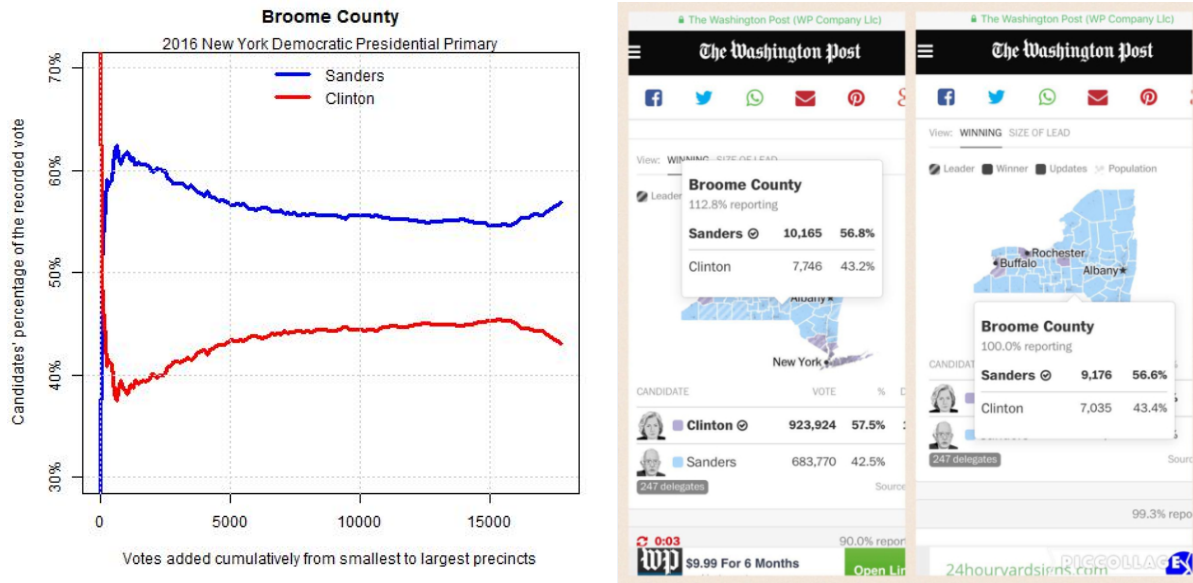
Fig. 22 — 2016 New York Democratic presidential primary, Broome County
The CVT graph shows Sanders with an unusual gain in votes in the largest counties
A screenshot shows 112.8% reporting, followed by both candidates experiencing a loss of votes
graph by Anselmo Sampietro & screenshot courtesy of Aimee Rox Coleman

Voter Fraud vs. Election Fraud

There is one other critique that is worth taking the time to examine. Professor J. Celeste Lay's comment in an interview with Doug Johnson Hatlem is a good example of this:

"Most of this discussion is driven by Sanders supporters who are disappointed he is not winning and want to claim he has more support in the Democratic Party than he actually does … Until proven otherwise, I'll go with the numerous studies demonstrating the infinitesimal amount of voter fraud in U.S. elections."

Lay is conflating voter fraud with election fraud. They are opposite behaviors. Voter fraud is when a single voter is trying to scam the system by voting twice, or voting when he or she is not authorized to do so. It has been demonstrated to be rare in the U.S. Election fraud refers to the illegal interference with the process of an election. In election fraud, the voters are the victims of an organized effort to rob them of their rightful influence. How common is it?

According to Walter R. Mebane, Jr. and Allen Hicken, both on the political science faculty at the University of Michigan, election fraud occurs in a majority of democracies. In their 2015 *Guide to Election Forensics*, they write, "During the first half of 2015 alone, for example, allegations of election fraud occurred in Bangladesh, India, Israel, Macedonia, Nigeria, Pakistan, Russia, Togo, the United Kingdom, the United States, and Zambia … data compiled by Kelley and Kolev (2010) on national

elections conducted in more than 170 countries from 1978–2004 indicate … 61% of countries experienced some degree of (known) cheating."

In correspondence, Mebane was firm that, "the problem of detecting election fraud really is extremely difficult." This raises the larger issue: Is it incumbent on the voting public to police elections and prove that fraud is taking place? Or is it more appropriately incumbent on election officials to provide convincing evidence that the election results are accurate and secure?

Testimony Under Oath

It would be helpful if someone would come forward and testify under oath that they had tampered with or been asked to tamper with the election results of voting machines. In his article on Medium, Spencer Gundert reminded us that Clint Curtis did exactly that.

In 2004, computer programmer Clint Curtis testified under oath before the U.S. House Judiciary Committee that he "wrote a prototype for [at that time] present Congressman Tom Feeney," of a program that "would flip the vote 51–49. Whoever you wanted it to go to and whichever race you wanted to win." Curtis testified that he was told by his company that, "we needed to control the vote in South Florida (5:23.)"

In this clip from the documentary, *Uncounted*, Curtis demonstrates the "flip." At 4:17 he says, "Twenty-four lines of code. You never see it."

Future Elections

Even as doubts rise about the accuracy and security of electronic voting machines, counties continue to invest millions of dollars purchasing them. In March Marion County Indiana spent 1.4 million dollars purchasing electronic voting equipment that included the ES&S DS200 in-precinct tabulators; the same machines that produced irregular election results in the 2016 New York primary. In Sedgwick County Kansas, the county commissioners approved a measure this week to spend 7.8 million over ten years on new ES&S machines as well. Beth Clarkson was at the meeting, and expressed her concerns, but the vote to purchase the machines still passed. That vote was counted by hand.

The 2016 Primary

The majority of the data we examined suggests that the two candidates currently slated to accept their party's nomination in the 2016 presidential primary races, received a different number of votes than what has been officially reported.

On the Republican side, statistical analysis indicates that Donald Trump probably received more votes than what has been reported and certified. Because he was able to overcome his opposition, even with the irregularities, his selection as the presumptive Republican nominee is supported by the data.

As we stated in the opening, this is not the case on the Democratic side. The overwhelming majority of the almost two dozen states that we analyzed, demonstrated irregularities. In almost every instance the discrepancies favored Hillary Clinton. In all likelihood the current results have assigned her a greater percentage of the vote than she may have actually received, while simultaneously under-reporting Bernie Sanders' legitimate vote share.

The difference between the reported totals, and our best estimate of the actual vote, varies considerably from state to state. However these differences are significant – sometimes more than 10% – and could change the outcome of the election. We intend to report on the percentage that each state may be off, based on a statistical analysis of as many states as possible.

It is hard to conceive of a legitimate transfer of power following an election that has been this flawed. We recommend that many of these elections be examined, and if found to be inaccurate, decertified. Where paper ballots are available, it would be informative to count them by hand. Where paper ballots are not available, it would be more equitable to hold a second vote, one that does not utilize any electronic voting equipment. It would be best if the counts are videotaped, and for observers to be able to watch in close enough proximity to verify the accurate count of every ballot. Possibly the ballots themselves need to be individually photographed and itemized. Hand counts, historically, have also been subject to election fraud, so the protocols must focus on transparency, chain of custody, and secure and verifiable results.

We understand that this is unprecedented. While the U.S. does have a long history of election fraud, we do not believe it has ever been this well-documented prior to the end of the election cycle.

Philip B. Stark, a statistics professor at UC Berkeley, who has been instrumental in designing new auditing techniques for elections agrees that, "Closer scrutiny of elections is necessary and welcome … that requires a paper trail, convincing evidence that the paper trail is complete and accurate, and a risk-limiting audit of the paper trail or a full manual tally to provide convincing evidence that the paper trail matches the announced result."

Beth Clarkson adds, "It's possible that to do a good audit would be more expensive and less transparent than a hand count because to do a good audit requires expertise."

Securing Our Democracy

The lack of accuracy in our elections is truly a betrayal of our ancestors who fought and died for the democratic process. It is a betrayal of the soldiers who lost their feet to amputation in the Valley Forge winter. It is a betrayal of the women who went to prison and starved themselves to join the franchise. It is a betrayal of the civil rights workers who died for the right to register to vote. There can be no debate

about whether or not the vote is accurate. We must know that it is accurate the way we know that the Earth revolves around the sun and not vice versa.

The people who win these elections will decide whether or not we go to war, how many people get what jobs, where our children attend what quality schools, the quality of the air we breathe and the water we drink, and so much more. The entire identity of our nation rests on our self-image as a self-correcting democracy whose leaders are accountable to the voters that elect them. To not know with 100% confidence that those leaders are the leaders that we actually voted for is the very essence of an existential crisis. This is a state of emergency. We must move rapidly to secure the integrity of the vote.

We need to immediately implement robust audit procedures. Then, as rapidly as possible, we must join the other legitimate democracies of the world and implement a system of paper ballots, hand-counted in a secure process that is open to the public, invites media scrutiny, and has strong chain of custody protocols. In this way, we can achieve accurate, verifiable results. Each citizen of the United States, and indeed the world, deserves this from us.

About the Team

lulu Fries'dat is an Edward R. Murrow award-winning journalist and documentary filmmaker. She received a Best Documentary award for her first feature-length documentary (producer/director) *Holler Back — [not] Voting in an American Town*; a film that explores systemic issues in our elections that discourage voter participation. Clips are available for viewing here. Her network news experience includes assignments as an editor for *CBS Evening News*, *Nightline*, *Sunday Morning*, *The Today Show*, and *Good Morning America*; She produced and edited profiles of Democratic candidates for MSNBC and has done long-format documentary work with NBC News and CNBC. She was on the editing team of *Gideon's Army*, an Emmy-nominated documentary that follows the personal stories of public defenders in the Deep South. For a change of pace, she created and directed voter registration PSAs for VH1 that featured Triumph the Insult Comic Dog. She has produced and directed frequently with her independent production company Shugah Works, including three short films for New York City schools. Her full bio is available here. Follow her on twitter @shugahworks.

Anselmo Sampietro holds a Master of Statistics degree from the University of Bologna, Italy and has also studied at the University of Technology Sydney, in Australia, and the University of Warsaw, in Poland. He currently leads a team of data analysts for a company based in London. He collaborated (through InnoCentive) with General Fusion, a Canadian start-up that is developing clean and reliable nuclear fusion reactors, to build a statistical model predicting plasma performance. He has specialized in the use of statistical analysis software R, which is the analytic tool used throughout this article.

Fritz Scheuren is a Senior Fellow and Vice President at NORC in the Center for Excellence in Survey Research. Scheuren has an unparalleled record of work on complex substantive tasks related to

sampling and to the analysis of data from government agencies and private sector institutions. Notable work at NORC includes a landmark project resolving issues with Indian Trust Fund accounts at the U.S. Department of the Interior as well as work with the Federal Reserve, particularly the Survey of Consumer Finances and at the Census Bureau. Scheuren is currently leading impact evaluations of Millennium Challenge Corporation (MCC) programs in Georgia and Lesotho, and formerly advised the MCC on impact evaluations in Armenia and Vanuatu. He brings a proven ability to solve hard, non-standard problems and the theoretical background to back these solutions up with sound statistical arguments.

Scheuren also serves on the Statistics Faculty at The George Washington University, where he had been a Visiting Professor of Statistics. At GWU, he created a successful survey sampling certificate program which he still teaches. Most of his over 450 applied and theoretical papers, presentations, monographs, and books are on sampling aspects of data collection primarily in a survey context, with an emphasis on sample design and estimation, process quality, and the handling of missing data. He has written over half a dozen books.

Scheuren served at the 100th President of the American Statistical Association and chaired the ASA Sections on Survey Research Methods and, later, the Social Statistics Section. He has been their Scientific Secretary at the International Association of Survey Statisticians. He is a Fellow of the American Statistical Association and the America Society for the Advancement of Science.

Election Justice USA – the paper was assisted in its research and development by Election Justice USA, a national non-partisan coalition of seasoned election integrity experts, statisticians, attorneys, journalists and activists, whose mission is to make sure each American's right to vote can be exercised without issue in accurate and honest elections.

Links to Election Results

[2000 WI Republican presidential primary](#)

[2000 FL Alachua County Democratic presidential primary](#)

[2004 FL Alachua County Democratic presidential primary](#)

[2016 DE Democratic presidential primary](#)

[2016 LA Democratic presidential primary](#)

2016 New York Democratic presidential primary:

[http://vote.nyc.ny.us/downloads/csv/election_results/2016/20160419Presidential%20Primary/0100020 0000Citywide%20Democratic%20President%20Citywide%20EDLevel.csv](#)

[2016 NY Broome County Democratic presidential primary](#)

[2016 NY Columbia County Democratic presidential primary](#)

[2016 WI Republican and Democratic presidential primary](#)

Links to Choquette and Johnson's Election Data

[Version 1.5](#)

[Version 2.1](#)

[Links to Clarkson's Election Data](#)

[Links to Lindeman's Election Data](#)

[Voting Equipment by Type](#)

[Wisconsin voting equipment](#)