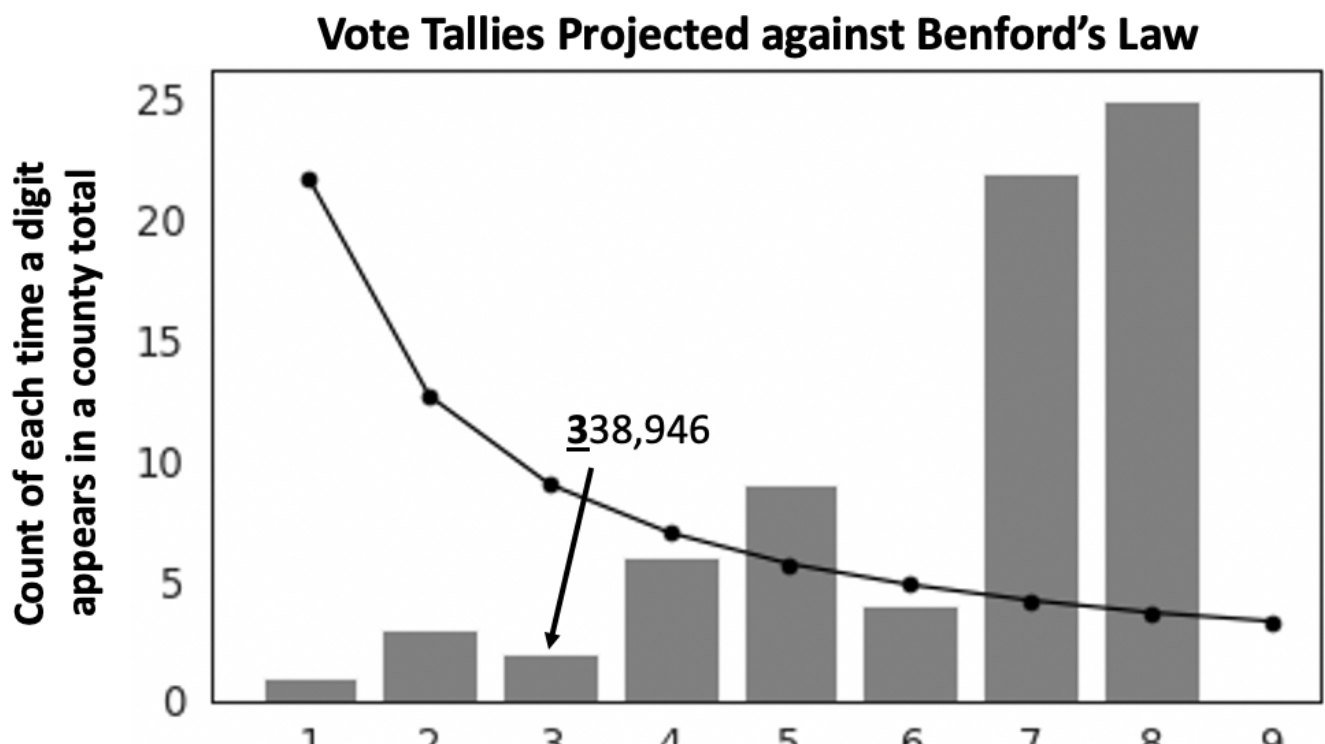Home   News   Policy Analysis   Rapid Response   Events

Nov 6

Vote Data Patterns used to Delegitimize the Election Results

Authors: Joe Bak-Coleman, Morgan Wack, Joey Schafer, Emma Spiro, Jevin West, University of Washington Center for an Informed Public

Vote Tallies Projected against Benford's Law

1   2   3   4   5   6   7   8   9

Leading Digit

The figure above shows the leading digit of reported vote tallies across select counties. For instance, the final tally in Dane County, Wisconsin was **3**38,946. This would count for one county in the **3** column. But why would anyone care to look at this kind of frequency distribution? Data forensic experts use these distributions to investigate fraud. They look at whether empirical distributions of leading digits deviate from a special distribution described by Benford's Law. The law posits that leading digits of numbers are more likely to be smaller numbers (e.g., **1**) than larger numbers (e.g., **9**).

Armchair investigators during the election have already begun to argue that too many of the submitted vote totals begin with larger single digit numbers (7 or 8 for example), which is being spun as evidence of voter fraud. We caution against this conclusion. Having the distribution of leading digits stray from the expected percentages predicted by Benford's Law can happen by chance, though it is more common when the law's assumptions are violated, as they often are with vote tallies. Benford's Law, and other math-based inquiries, can be used to detect voter fraud, but the vast majority of these violations are not conclusive evidence of fraud.

Simon Newcomb first recognized this law of anomalous numbers in the 19th century, observing that the most worn pages in a logarithmic table book are in the pages with 1s versus 8s and 9s. The law was later formalized by Frank Benford, who looked at various empirical data--from astronomical object sizes to populations of cities to house address numbers--and found miracuously that they followed this same pattern across disparate data sets. However, the law holds consistently when certain assumptions are met: all numbers must be equally likely to appear (i.e., you can't only tally batches of 6 votes and expect the totals to start with 7, 8, or 9) and the numbers must span multiple orders of magnitude, such as ranging from 100 to 10,000,000 [†]. Violations of these assumptions lead to violations of the law. For vote tallies, all numbers are equally likely, but not all states meet the second assumption. In the state of Nevada, Esmeralda County has around 900 people while Clark County has over 2,250,000 people. In the state of Vermont, the bounds are much narrower.
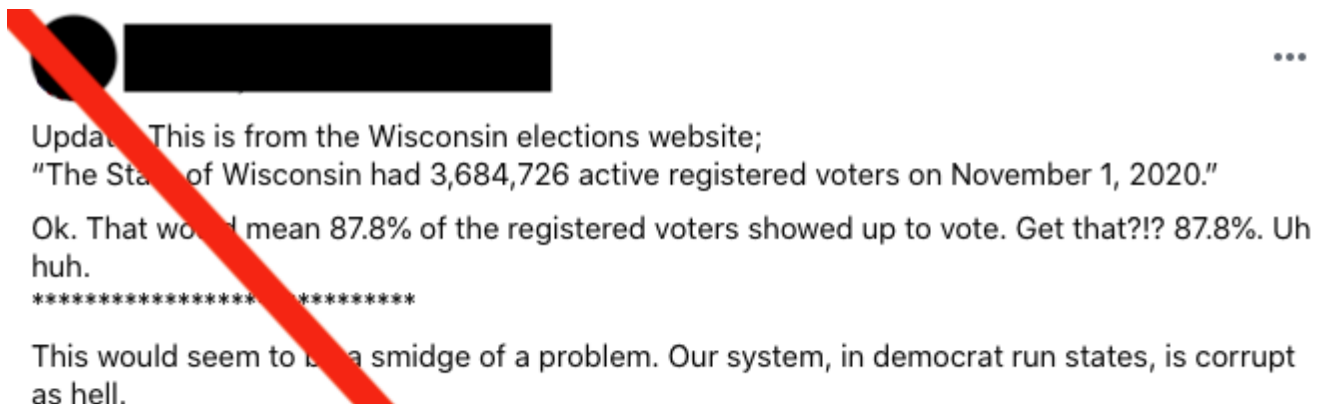
Returning to our voting tally in Figure 1, you will see that the tallies deviate from the line of expectation. So, does this mean fraud? Does it mean that vote counters were up to something nefarious? In this case, absolutely not. First, the example above is a simulation based on a computer script, rather than one based on real voter data. If we consider the final output of this 72 county simulation, it ends up looking like Figure 2:

Final Vote Tallies Projected against Benford's Law

Final Vote Tallies Projected against Benford's Law
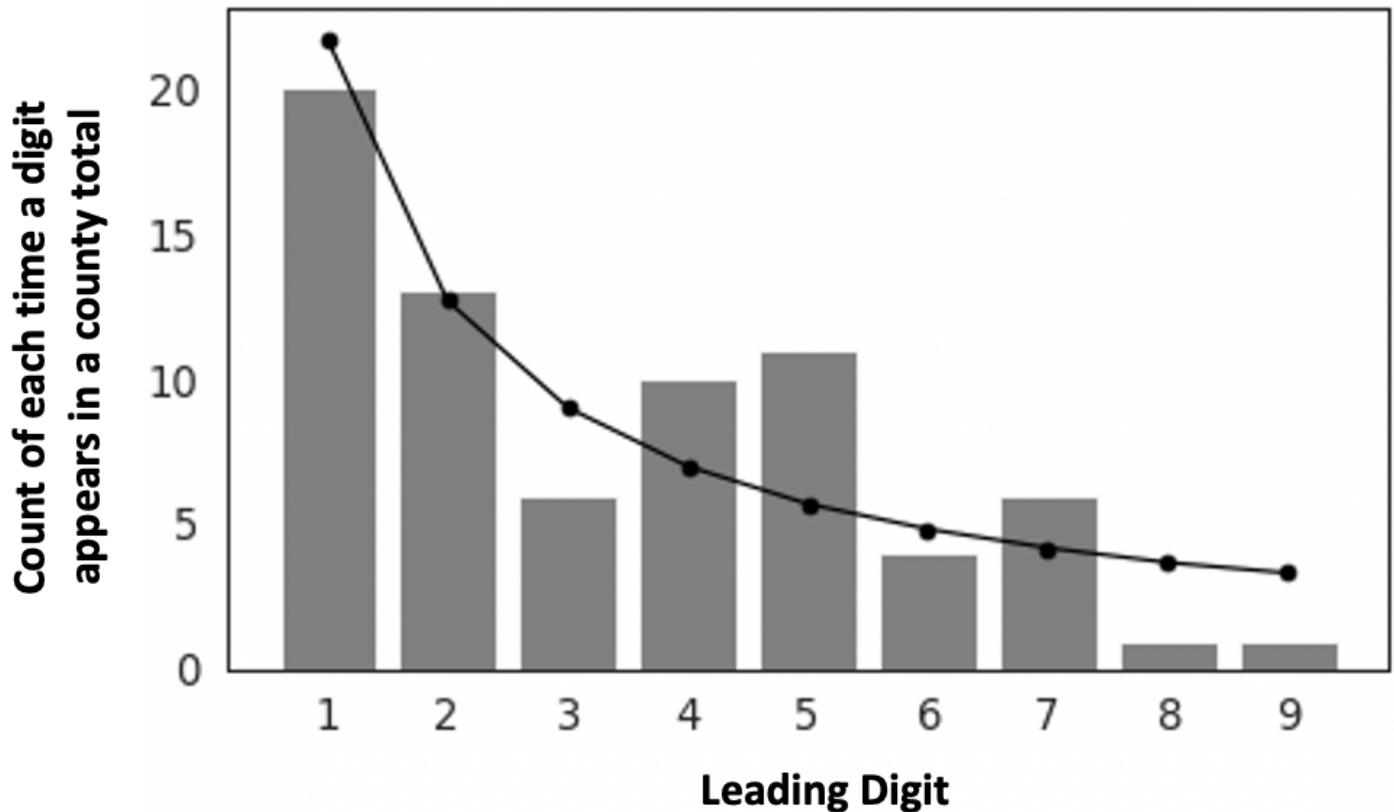
These final results are more predictable and follow the expected counts more closely, but still exhibit expected deviations. These same deviations are occurring in the voting counts currently being reported in the 2020 election. Our aim in this post is to prepare the public and journalists for these misleading arguments and to provide context for the claims already being made online.

Use of Benford's Law to "prove" voter fraud is but one of many strategies trying to use numbers to prop up spurious conclusions. More commonly, data that don't rely on underlying mathematical theories, but are similarly flawed, are cited as evidence for misleading claims. This use of "mysterious" numbers is a common tactic of Qanon supporters attempting to disprove the existence of COVID-19. Moreover, in the last forty-eight hours alone, we have seen this strategy utilized in attempts to delegitimize turnout rates in addition to vote counts. For example, consider the widely cited and since debunked tweet by a verified user (pictured below):

Upda__ This is from the Wisconsin elections website;
"The Sta__ of Wisconsin had 3,684,726 active registered voters on November 1, 2020."

Ok. That wo__ mean 87.8% of the registered voters showed up to vote. Get that?!? 87.8%. Uh huh.
*****************___********

This would seem to __ __ smidge of a problem. Our system, in democrat run states, is corrupt as hell.

Mike Coudrey ✓
@MichaelCoudrey

BREAKING: Wisconsin has more votes than people who are registered to vote.

Total number of registered voters: 3,129,000

Total number of votes cast: 3,239,920

This is direct evidence of fraud.

This is misleading for several reasons. Wisconsin has around 3,600,000 registered voters on Election Day, which is greater than the number of votes cast. Additionally, since Wisconsin allows same-day voter registration, this number could be even larger. Provisional ballots could be assigned to voters registering the day of the election.

Elsewhere on social media, users have used projections of exactly 270 votes for Biden in the electoral college to suggest that fraud has occurred:
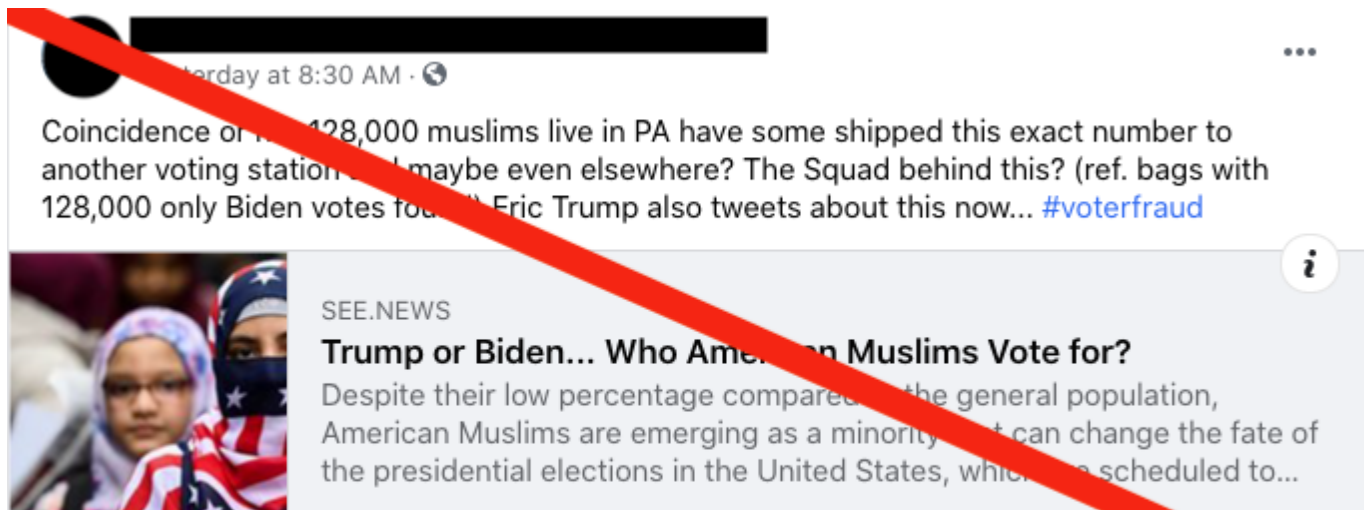

Nov 5, 2020 at 8:59 AM

I don't believe its just a coincidence that this election has to be as absolute stressful as possible by random event. s also overly fishy that Biden will get exactly the number he needs to win. No more, no less. Seems like such hokey sham. Everything riding on Nevada sitting around taking days to count ballots for six electoral votes & biden needs 6 electoral votes to win. Its like 1/2 a % difference between Biden & Trump. The

Its like 1/2 a %... iherence between Biden & Trump. The media twisting ...d propaganda isn't helping the situation either. T...ey should not even be calling the election at the leve...f what's been going on.

See less

Joe Biden

264

270...win

Donald Trump

214

Won   L...ds

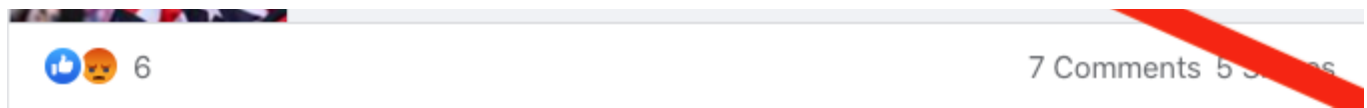Candidates   Electoral votes

71   71

Likewise, skeptics are already employing the numbers associated with a narrow Biden win as evidence that it was pre-planned, for example with Democrats "producing" exactly enough votes to win (notably only in the presidential race but not in the Senate or other contentious races down the ballot). More outlandish claims have used this format to spread completely unfounded narratives tying specific groups for voter fraud based on their population size:

...erday at 8:30 AM ·

Coincidence or...   ...28,000 muslims live in PA have some shipped this exact number to another voting station... maybe even elsewhere? The Squad behind this? (ref. bags with 128,000 only Biden votes fou...) Eric Trump also tweets about this now... #voterfraud

SEE.NEWS

Trump or Biden... Who Ame... ...n Muslims Vote for?

Despite their low percentage compared... ...the general population, American Muslims are emerging as a minority... ...t can change the fate of the presidential elections in the United States, whic... ...e scheduled to...

Amidst the deluge of data, specifically numerical quantities, that is bound to accumulate in an election that impacts over 328,000,000 people, with over 142,000,000 votes already tallied across 435 districts, an absence of any irregularities would in and of itself be irregular. Despite the obviousness of this conclusion, arguments that rely on "evidence" of violations of mathematical laws are not considered amidst this context. As a result, we caution everyone to be on the lookout for these erroneous numerical statements during uncertain periods of vote counting and in the weeks, months, and years to come.

Modeling observations: Our simulation model, that produced the data used in this post, is a simplification of what is inherently a complex process that varies state to state. We simulated counties casting a number of votes based on what was reported by NYT for Wisconsin in 2020. At each time step, we assumed that a poisson-distributed number of votes were counted with an expected rate of 1000 votes/step. In reality, the rate is likely somewhat higher for larger counties, but we note that this assumption nonetheless captures the phenomenon of larger counties taking longer to count. The rate limits, in combination with a large number of smaller counties, leads to early counties being approximately the same order of magnitude. At this stage, the assumptions that lead to Benford's law are violated leading to the patterns generated in the Figure 1 above. Only once all counties have been counted does the distribution approach something consistent with Benford's law, seen in Figure 2. Even at this stage, the distribution of county sizes still makes it unlikely to exactly match expectations. A more complete model might include non-random voting patterns whereby rural counties lean a different direction than urban ones. This, compared with the relationship between the rate of vote counting and county or precinct size would probably cause more drastic violations of assumptions. As this is a rapid response, incorporating this complexity was impractical. For a more rigorous discussion of Benford's law and its limited utility for detecting electoral fraud see Deckert, Myagkov, Ordeshook 2011.

Updated 11/18/2020: Footnote added regarding processes that can generate Benford's distribution with fewer orders of magnitude.

[†] Of course other processes may also generate similar distributions, even over fewer orders of magnitude (see example 10 of this reference). But without reason to suspect such a process is operating, we cannot *expect* to see the distribution predicted by Benford's law if we are looking at data that cover on a small range.

⟨   **Election Delegitimization: Coming to you Live**   **Media Largely Frames Trump's Victory Declaration as False in Headlines**   ⟩