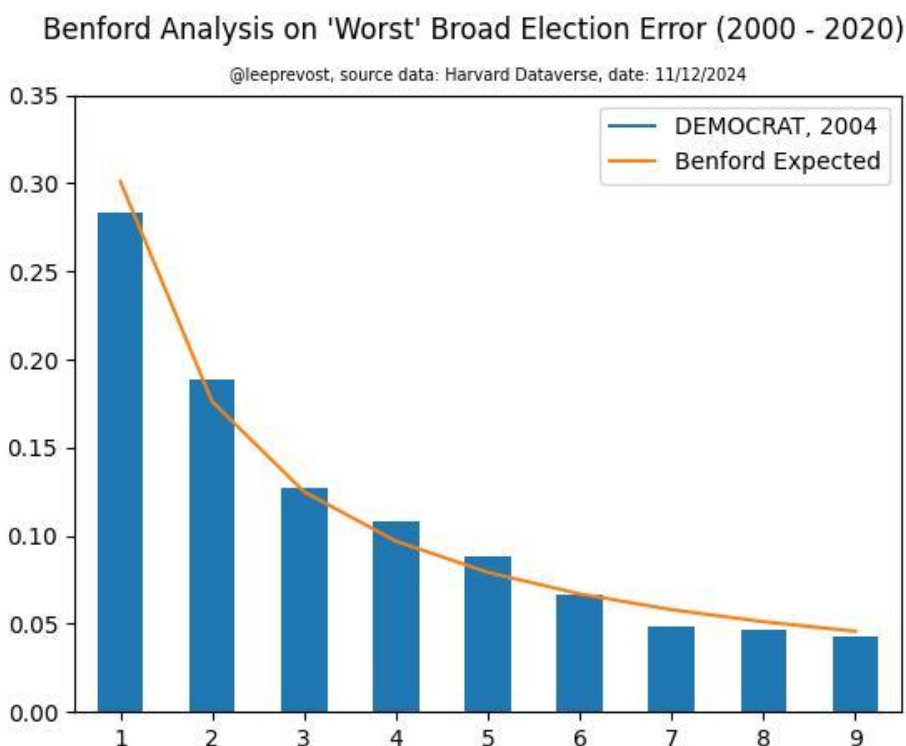


Benford Oddities in 2020 County Level US Presidential Election Data

Lee Prevost, 11/20/24

I'm aware of discussions that Benford's Law is misused in the study of elections, particularly in its use on precinct level data.ⁱ

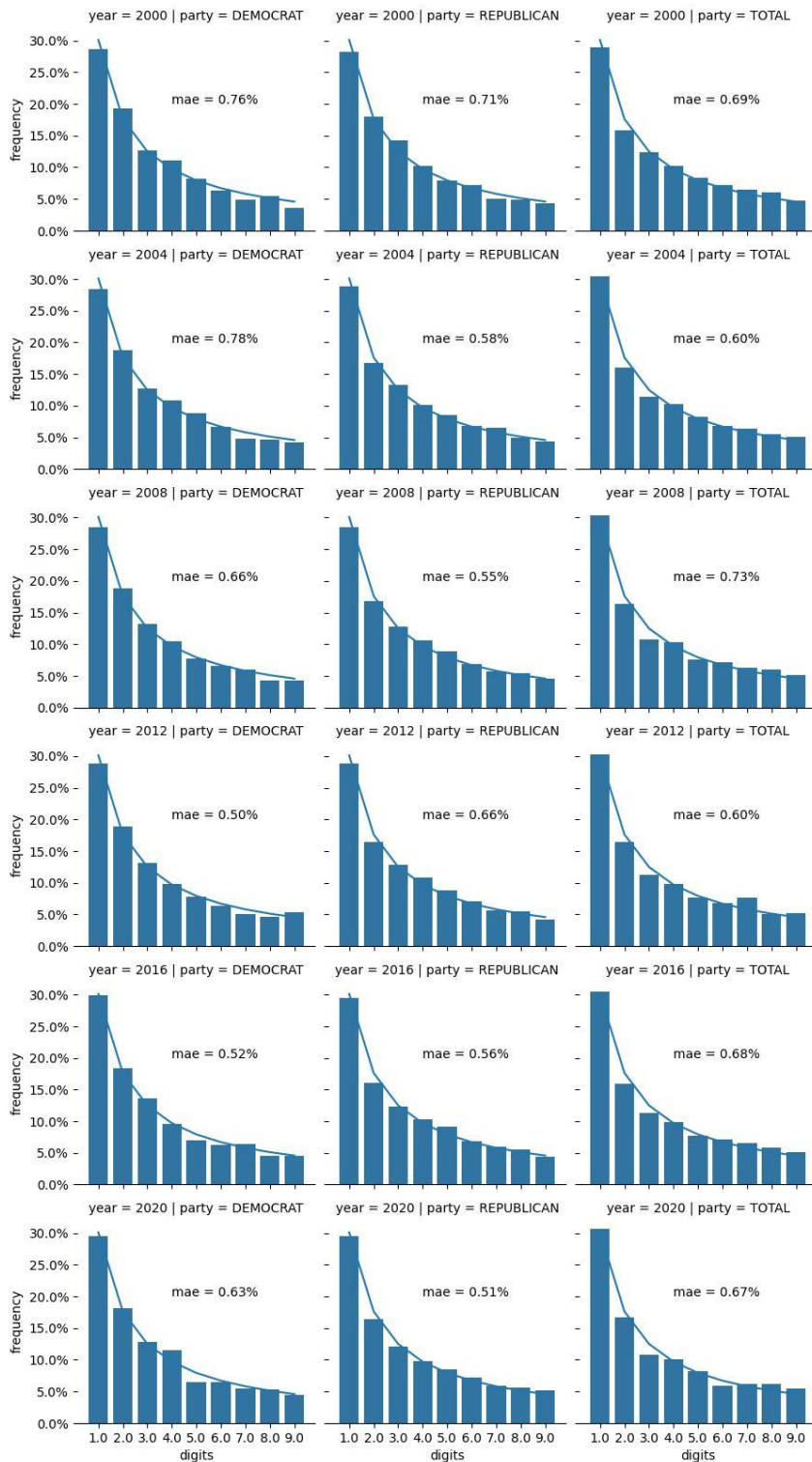
However, in studying the result of the 2020 election using a dataset from Harvard Dataverse/MIT Election Lab on county level election data since 2000, I'm finding Benford's Law to apply very well.



Using MAD or “median absolute deviation” as the error scoring methodology as prescribed by David G. Banksⁱⁱ, I am finding the error rates for all elections since 2000 to be well within the range of conformity outlined by Nigrini.ⁱⁱⁱ All total counts are in Nigrini's “Acceptable Conformity” range.

Guidelines	MAD
Close Conformity	0.000 to 0.004
Acceptable Conformity	0.004 to 0.008
Marginally Acceptable Conformity	0.008 to 0.012
Nonconformity	Greater than 0.012

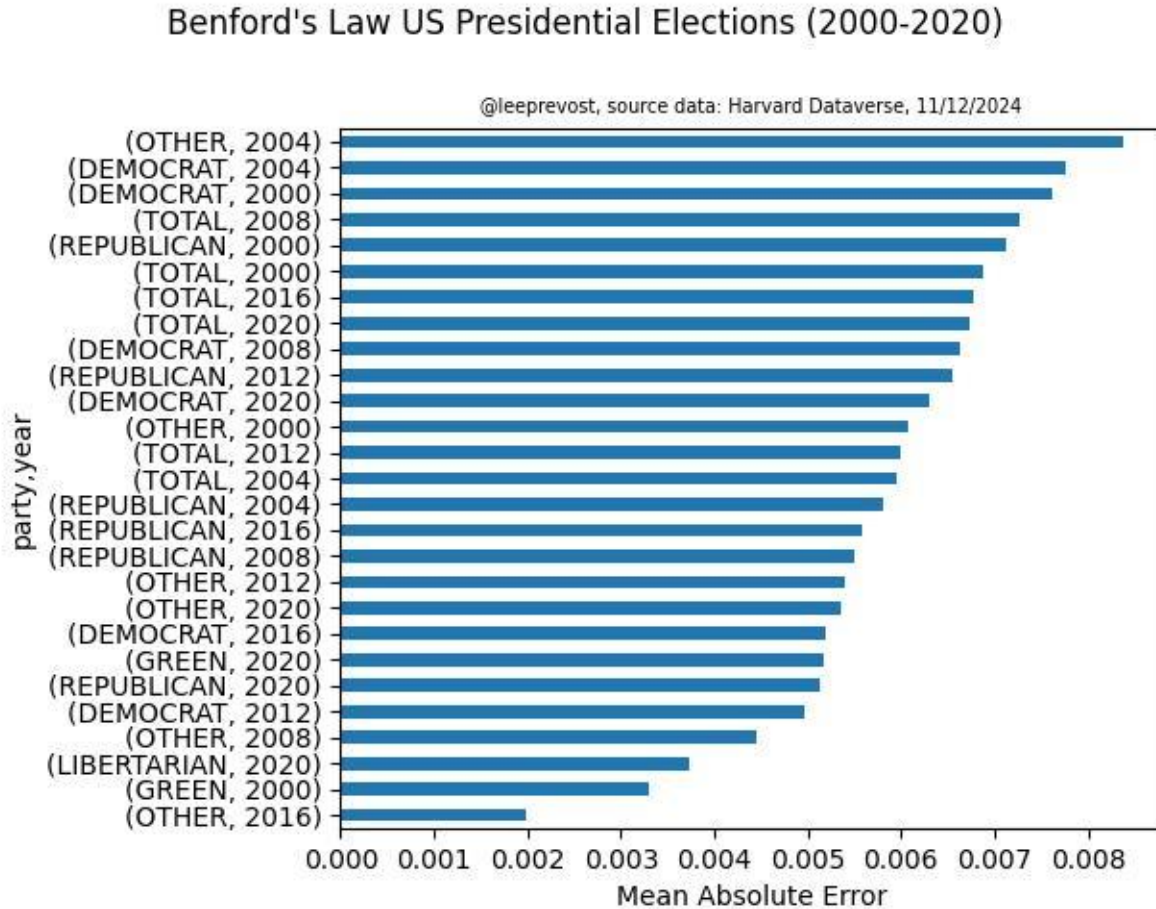
Benford's Law on US Presidential Elections 2000-2020 County Level



@leeprevost, source_data: Harvard Dataverse, 11/14/2024

Note: MAE (Mean Absolute Error as defined in sci-kit learn) is equivalent to “MAD”

The comparison of error rates of elections 2000-2020:



I conclude that the county level data, which is not artificially constrained by the boundaries of a precinct, follows Benford closely.

Even though the party level and total county level data seems to closely conform, I do find that Benford's Law breaks down when used with calculated data such as vote margins.

Closer Inspection of Party Level Data

The seemingly close conformity to Benford's Law encouraged me to look closer at actual digit anomalies over the course of elections.

Raw Benford digit county frequencies are in my [repo](#) along with all sources and methods.

I applied a z-score to all Benford digit count frequencies and then filtered on z-scores whose absolute values were greater than two. This produced an anomalous result list of elections and parties which had vote counts with z-score frequencies that were more than 2 standard deviations from the expected Benford count.

Benford Anomalies 2000-2020

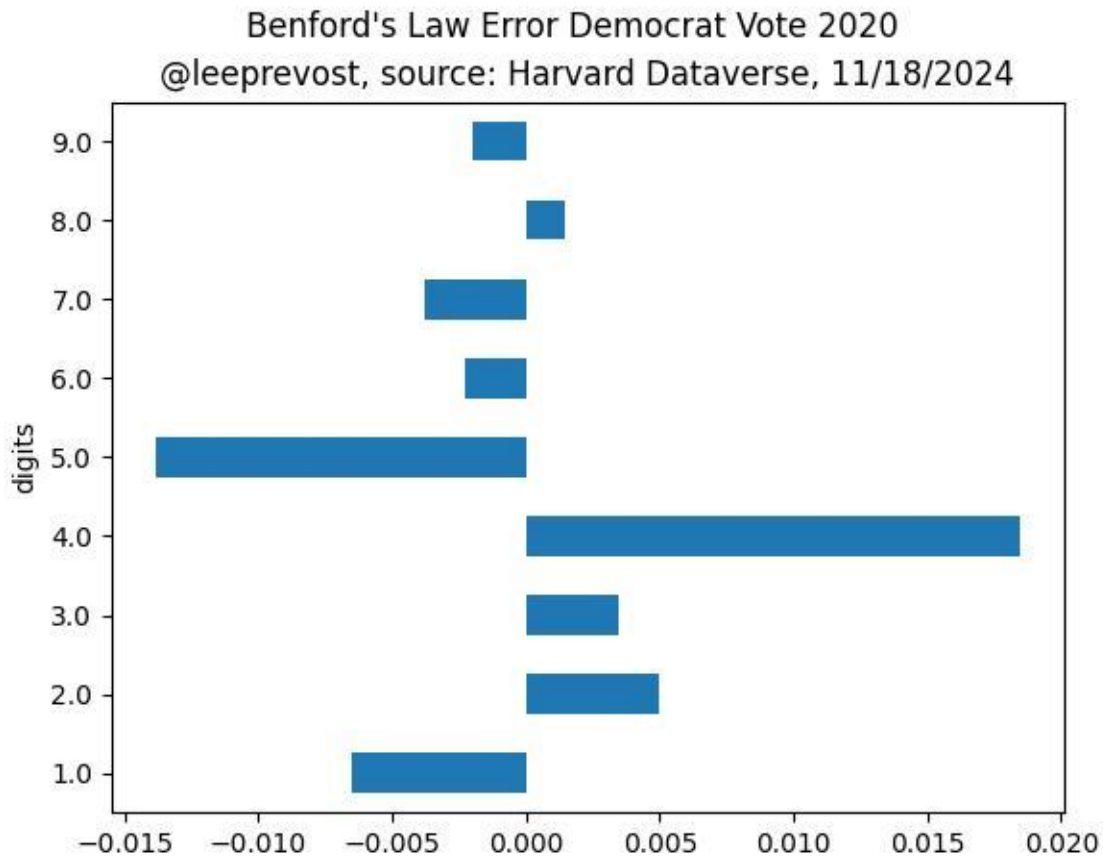
party **DEMOCRAT GREEN OTHER TOTAL**

year **2020 2020 2008 2012 2012**

digits

1.0	0.65%	1.45%	1.94%	1.36%	0.18%
2.0	0.49%	0.35%	0.02%	0.87%	1.25%
3.0	0.35%	0.91%	0.07%	0.21%	1.24%
4.0	1.85%	0.31%	0.75%	0.17%	0.07%
5.0	1.39%	0.25%	0.37%	0.77%	0.21%
6.0	0.23%	0.27%	0.04%	0.62%	0.15%
7.0	0.38%	0.80%	0.14%	0.18%	1.84%
8.0	0.15%	0.37%	0.59%	0.28%	0.11%
9.0	0.20%	0.29%	0.09%	0.40%	0.56%

A closer inspection of the one major party's vote in 2020 yields this error rate at the digit frequency levels:



Conclusion

The significance of these anomalies:

- Of all party level and total presidential election results from 2000-2020, there were 5 counts that were deemed suspicious by the z-score screen.
- Of those 5, one was in 2009, and two each in 2012 and 2020.
- My focus has been on 2020. Of the two in 2020, only one involved a major political party's vote (Democrat).
- It seems that a closer inspection of those counties whose votes shifted outside of the norm to a digit of 4 or from a digit of 5 would be worthwhile.

ⁱ Walter Mebane, <https://websites.umich.edu/~wmebane/inapB.pdf>

ⁱⁱ David G. Banks, [Get M.A.D with the Numbers!](#), Fraud Magazine, September/October 2000

ⁱⁱⁱ Nigrini, "Digital Analysis Using Benford's Law."