Assessing Adherence to Benford's Law in Presidential Elections

Brooks Groharing, William Jewell College
Co-Author/Faculty Advisor: David McCune, William Jewell College

MAA MathFest, 8/6/21

The Claim

In the 2020 presidential election, Biden's county vote totals in Pennsylvania fail to follow a relationship described by Benford's law; this is mathematical proof of substantial voter fraud and/or election tampering in Biden's favor.

Our goal: Place Biden's Pennsylvania results in the wider context of county-level Presidential election results, and assess the above claim using statistical methods.

Benford's Law

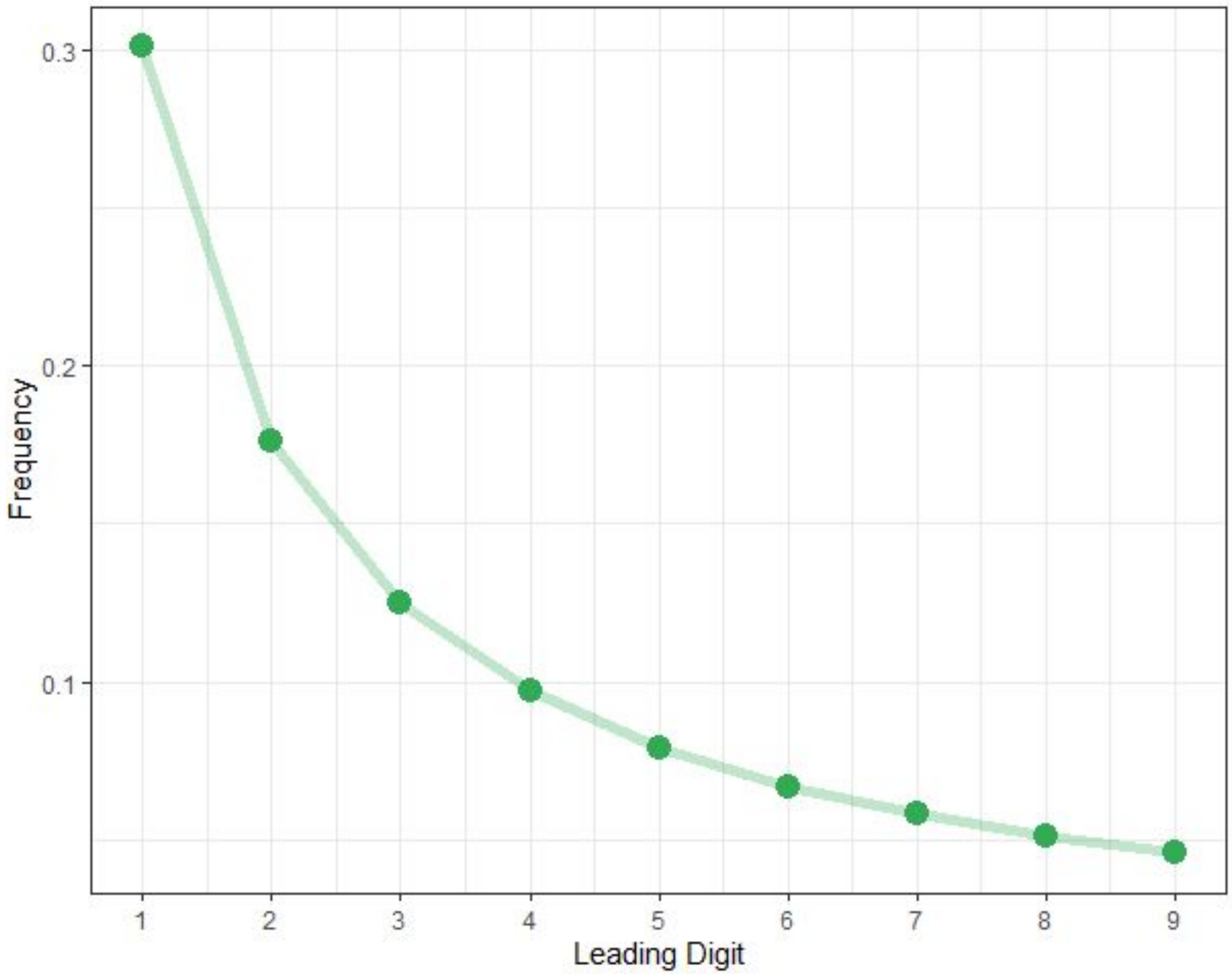
Observation: in many naturally occurring datasets, the possible leading digits 1-9 are distributed non-uniformly; low digits occur most frequently, with each higher digit occurring with a lower probability.

Benford's law predicts the frequency of each digit *i*, in such datasets, as $\log_{10}(1+1/i)$.

Ex: molecular weights, town populations, lengths of rivers

Leading Digit	Expected Frequency
1	30.1%
2	17.6%
3	12.4%
4	9.69%
5	7.92%
6	6.69%
7	5.80%
8	5.12%
9	4.58%

Predicted First Digit Frequencies, Benford



0.3
0.2
0.1
Frequency
1
2
3
4
5
6
7
8
9
Leading Digit

Benford’s Law

Previously, Benford’s Law has been shown to model digit frequencies in normal accounting data; a sizeable body of research supports using this as a litmus test for tax fraud.

The same relationship has not been shown for U.S. election data; nonetheless, Benford’s law was used to make various claims of fraud following the 2020 Presidential election.

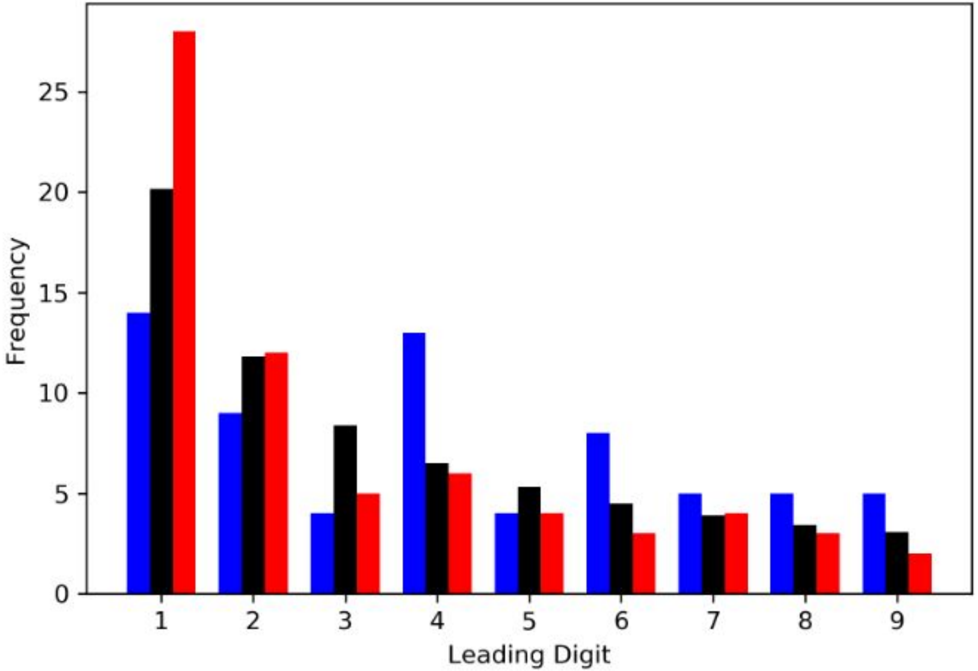
Leading Digit	Expected Frequency
1	30.1%
2	17.6%
3	12.4%
4	9.69%
5	7.92%
6	6.69%
7	5.80%
8	5.12%
9	4.58%

Predicted First Digit Frequencies, Benford

Pennsylvania, 2020

Why *might* one expect Benford's law to apply?

- County-level vote totals can span a range of magnitudes
  - Compare precinct*
- PA has a decent number of counties. (67)



Breaking down the claim

In the 2020 presidential election, Biden's county vote totals in Pennsylvania fail to follow a relationship described by Benford's law; this is indicative of substantial voter fraud and/or election tampering.

Three interrelated assumptions:

1) Under normal circumstances, county-level vote totals should follow Benford's law.
2) Biden's results from Pennsylvania in 2020 are unusually un-Benford-like.
3) Deviations from Benford's law indicate foul play.

Assumption 1

Under normal circumstances, county-level vote totals should follow Benford's law.

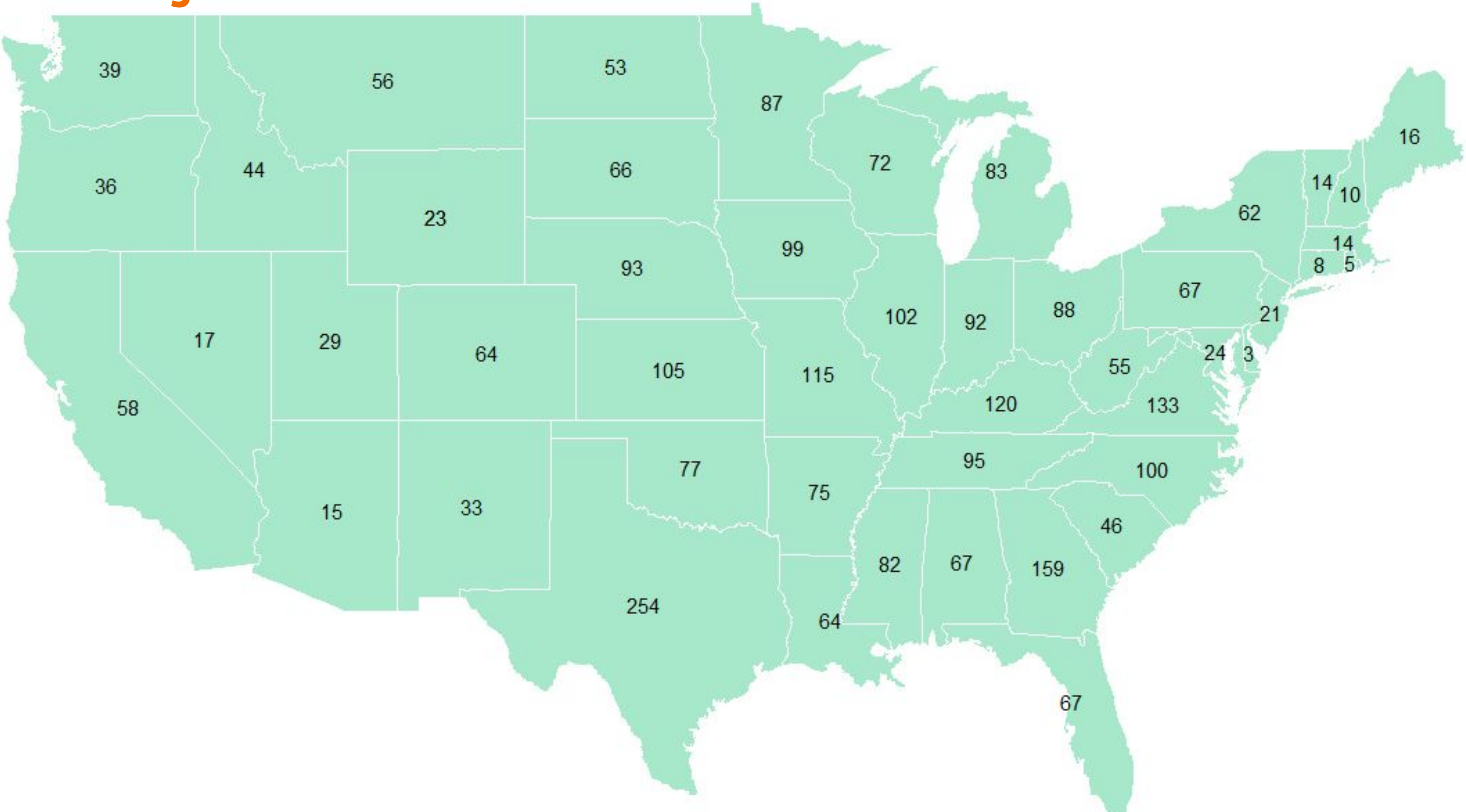
Assumption 1: These data should follow Benford's Law

To assess this, we evaluated the "Benfordness" of county-level presidential results, as a whole.

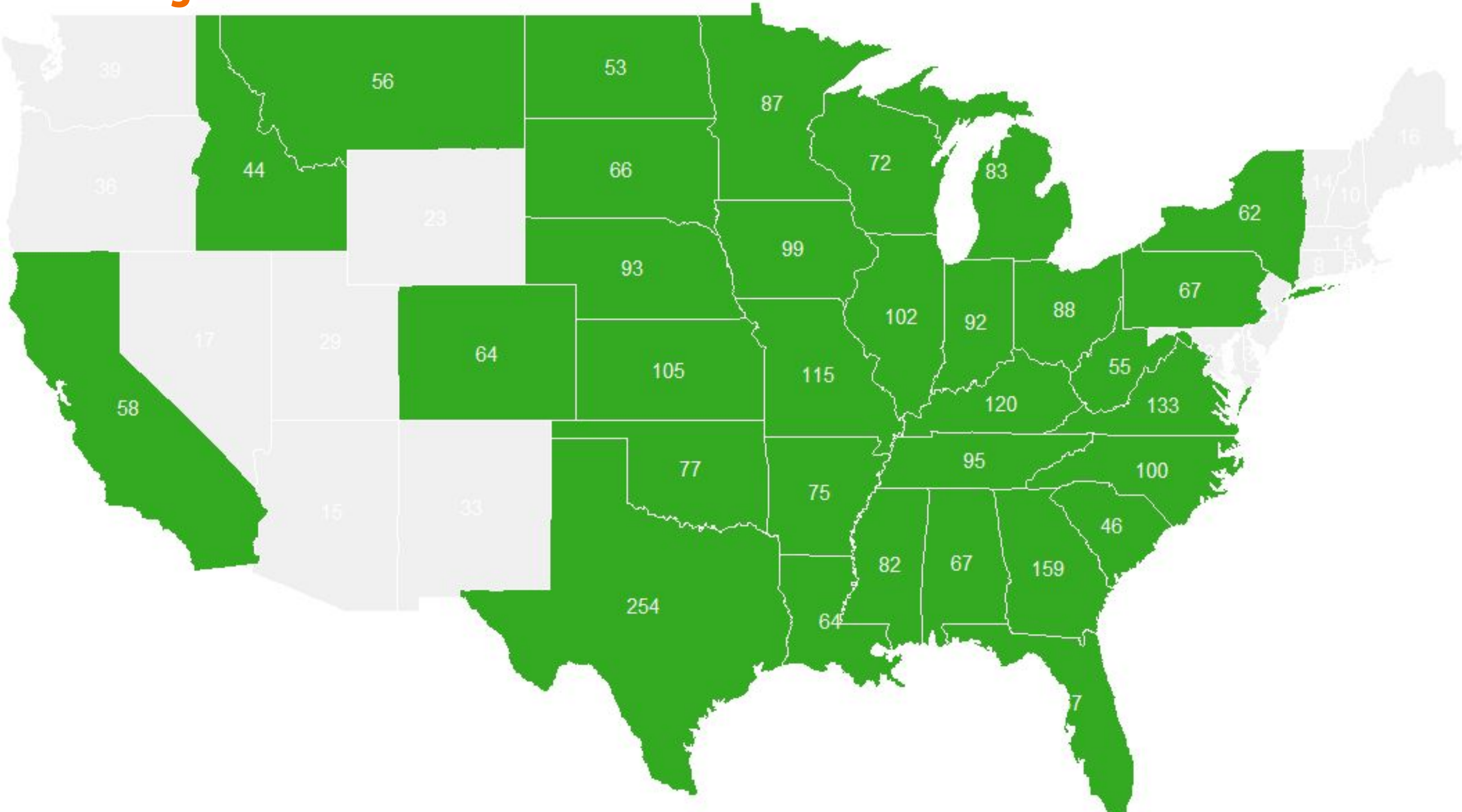
Three sub-steps:

a) Compile a dataset of relevant elections
b) Compare each set of state/year/party results to Benford Predictions
c) Assess Benford deviations at the single-digit level

1A) Selecting Data: Which states?



1A) Selecting Data: Which states?



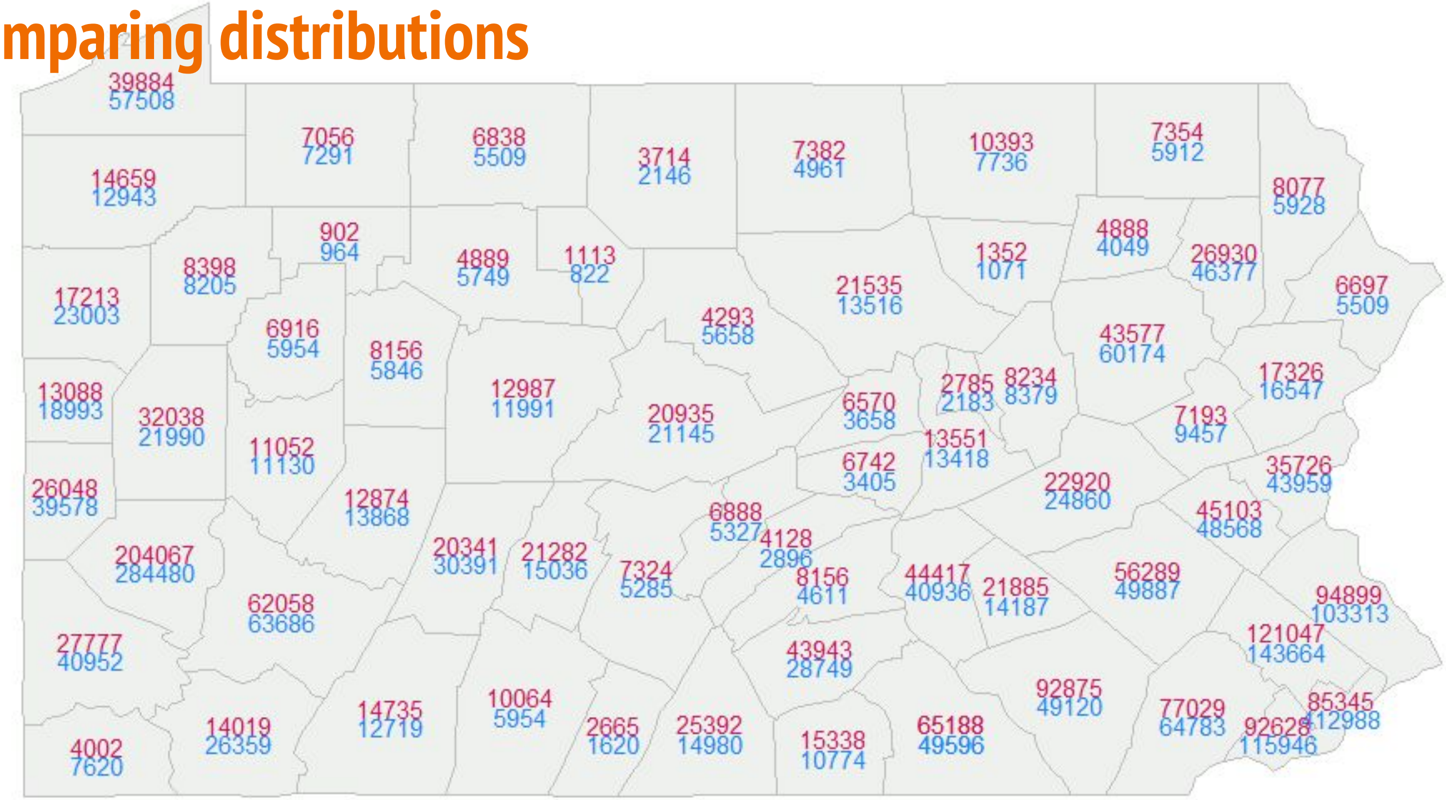
1A) Selecting Data: What years?

- Our focus is on PA, 2020; the farther back you go, the less relevant the data will be.
  - **Murky factors:** shifting party ideology, demographic changes, etc.
- 1992: Ross Perot earned 18.9% of popular vote. (1996- only 8.4%)
  - Most recent strong, Independent showing.
  - Again, unclear what effect this would have on “Benfordness.”

Our dataset:

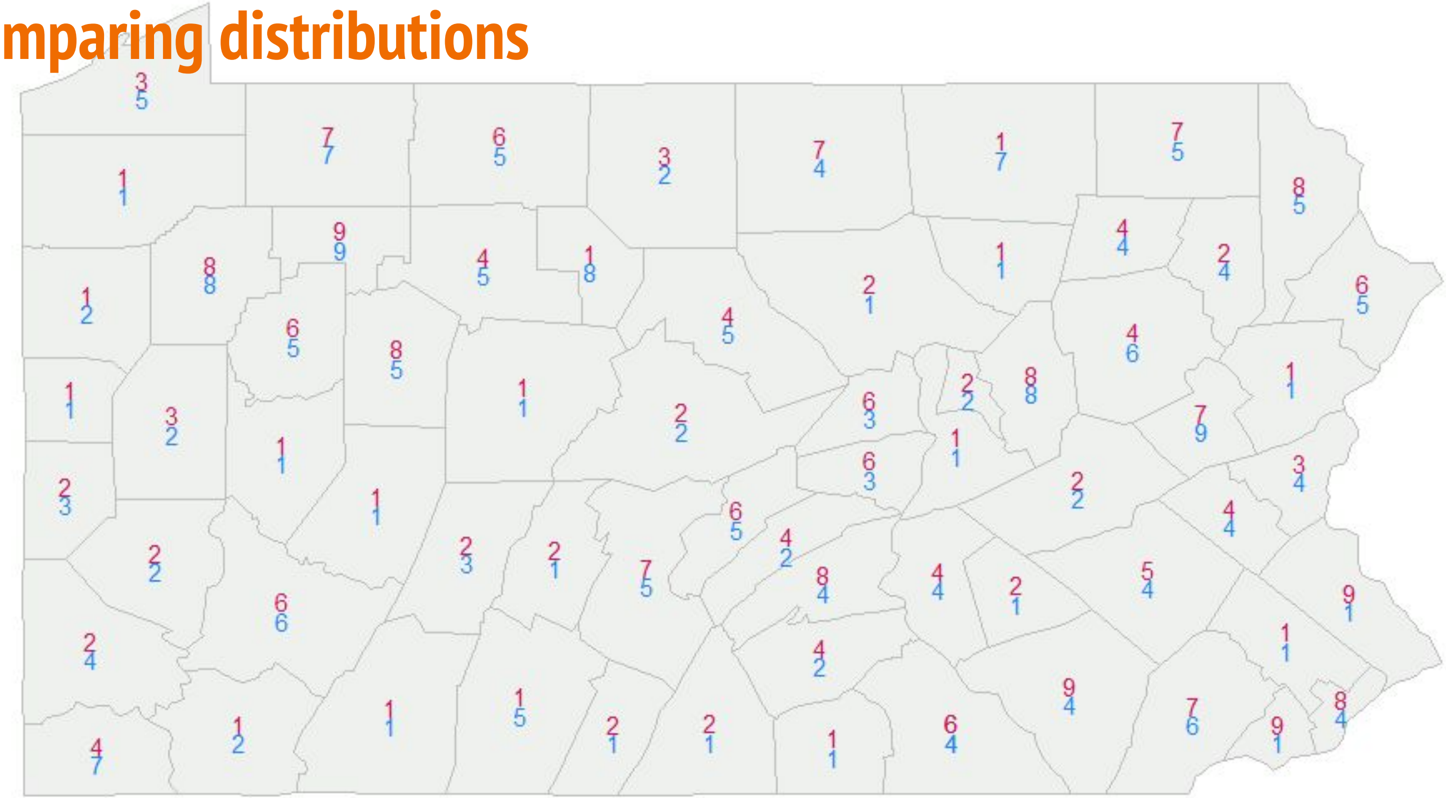
1996, 1998, 2000, 2004, 2008, 2012, 2016, 2020

1B) Comparing distributions



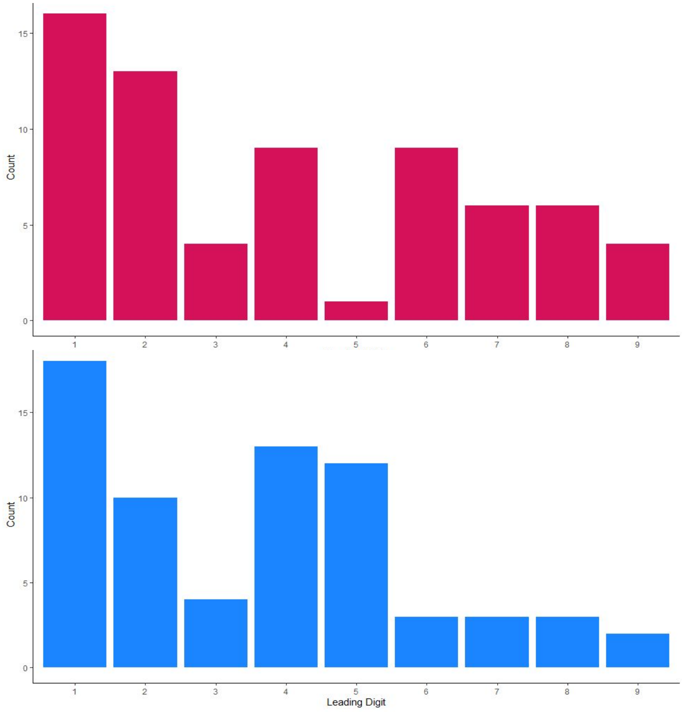
Pennsylvania, 1996

1B) Comparing distributions



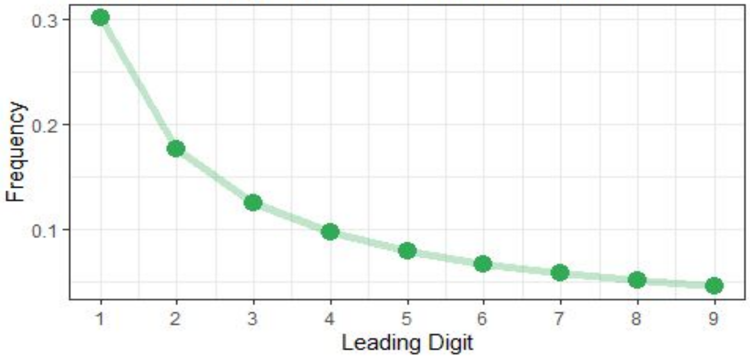
Pennsylvania, 1996

1B) Comparing distributions



vs

vs



Observed Counts

Predicted Distribution

1B) Comparing distributions

χ^2 test: Assesses "goodness of fit"

Tests the null hypothesis that the observed distribution follows the expected distribution.

If $\chi^2 > \chi^2_{crit}$, we can reject this hypothesis. The two distributions differ at the 5% confidence level.

We chose α 0.05 for our acceptable Type I error rate. Re-testing at lower α levels didn't change our conclusions.

$$\chi^2 = \sum_{i=1}^{9} \frac{(O_i - E_i)^2}{E_i}$$

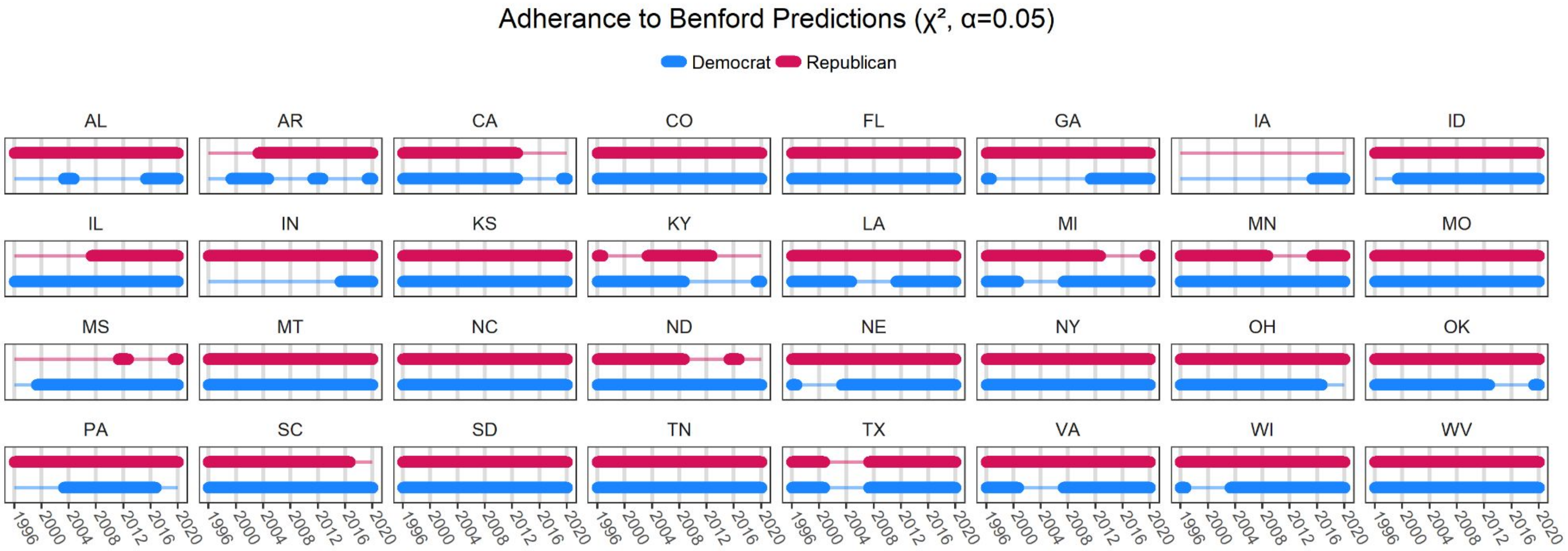
Chi-Square equation

$$\chi^2_{crit} = 15.507$$

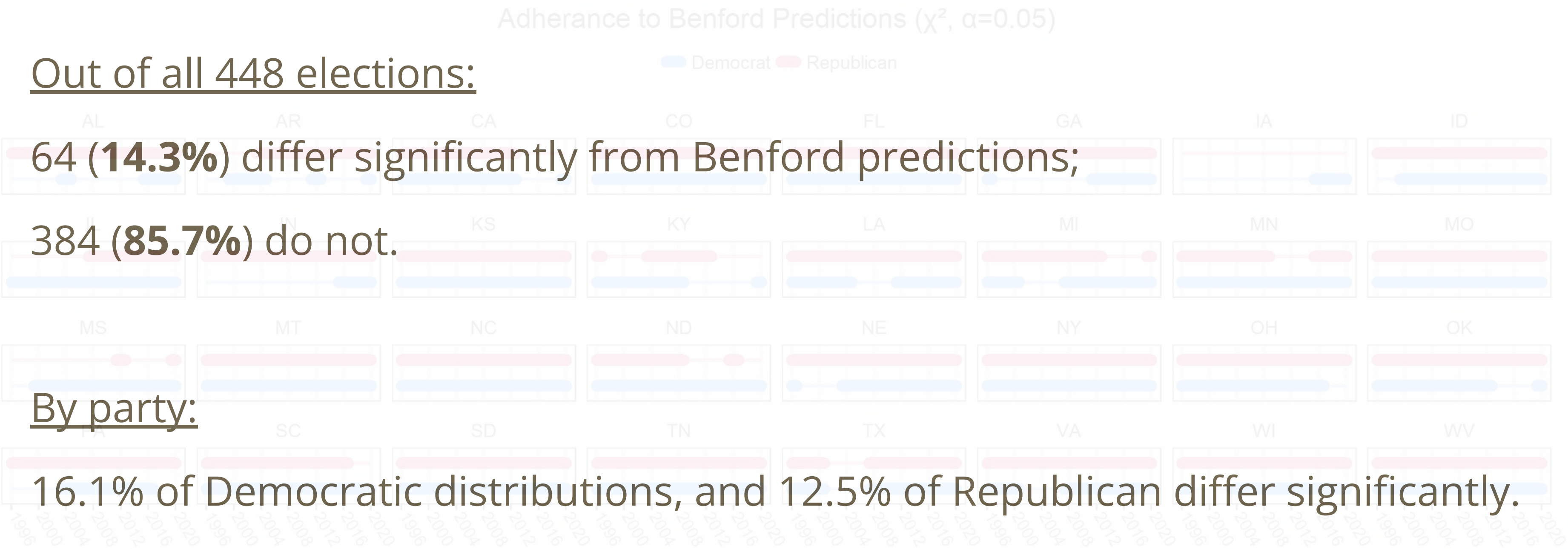
(α 0.05 , 8 d.f.)

Critical Threshold

1B) Comparing distributions



1B) Comparing distributions



<u>Out of all 448 elections:</u>

64 (**14.3%**) differ significantly from Benford predictions;

384 (**85.7%**) do not.

<u>By party:</u>

16.1% of Democratic distributions, and 12.5% of Republican differ significantly.

1C) Per-digit deviations

Z-test: Assesses per-digit deviations

Tests the null hypothesis that an observed count of leading digit *i* in a given state/year/party distribution is statistically the same as the count Benford's law would predict.

If $Z_i > Z_{crit}$, we can reject this hypothesis. The given digit deviates from Benford's law at the 5% confidence level.

Note: A single deviating digit, per Z-test, may not lead to the overall distribution differing per the χ^2 test.

$$Z_i = \frac{|Po - Pe|}{\sqrt{e * (1-e) / n}}$$

$$\text{or } Z_i = \frac{|Po - Pe| - 1/(2n)}{\sqrt{e * (1-e) / n}}, \text{ if } 1/(2n) < Z_i$$

Z-score Equation

$$Z_{crit} = 1.96$$

(α 0.05 , 8 d.f.)

Critical Threshold

1C) Per-digit deviations

Out of all 448 elections:

200 (**44.6%**) had at least one deviating digit.

46 (**10.2%**) had two or more deviating digits.

10 (**2.2%**) differed at three digits.

The number of non-Benford digits did not differ significantly by party.

Assumption 1 Verdict

Assumption 1 (these data *should* follow Benford's law) **is partially supported.**

Benford adherence appears to be the historical norm, but deviations have occurred frequently for candidates from both major parties. This is true whether you compare at the level of overall distributions, or individual digits.

<u>Assumption 2</u>

Biden's results from Pennsylvania in 2020 are *unusually* un-Benford-like.

Assumption 2: Biden’s PA distribution is an outlier

Using the recent election data, we can contextualize Benford’s 2020 PA results.

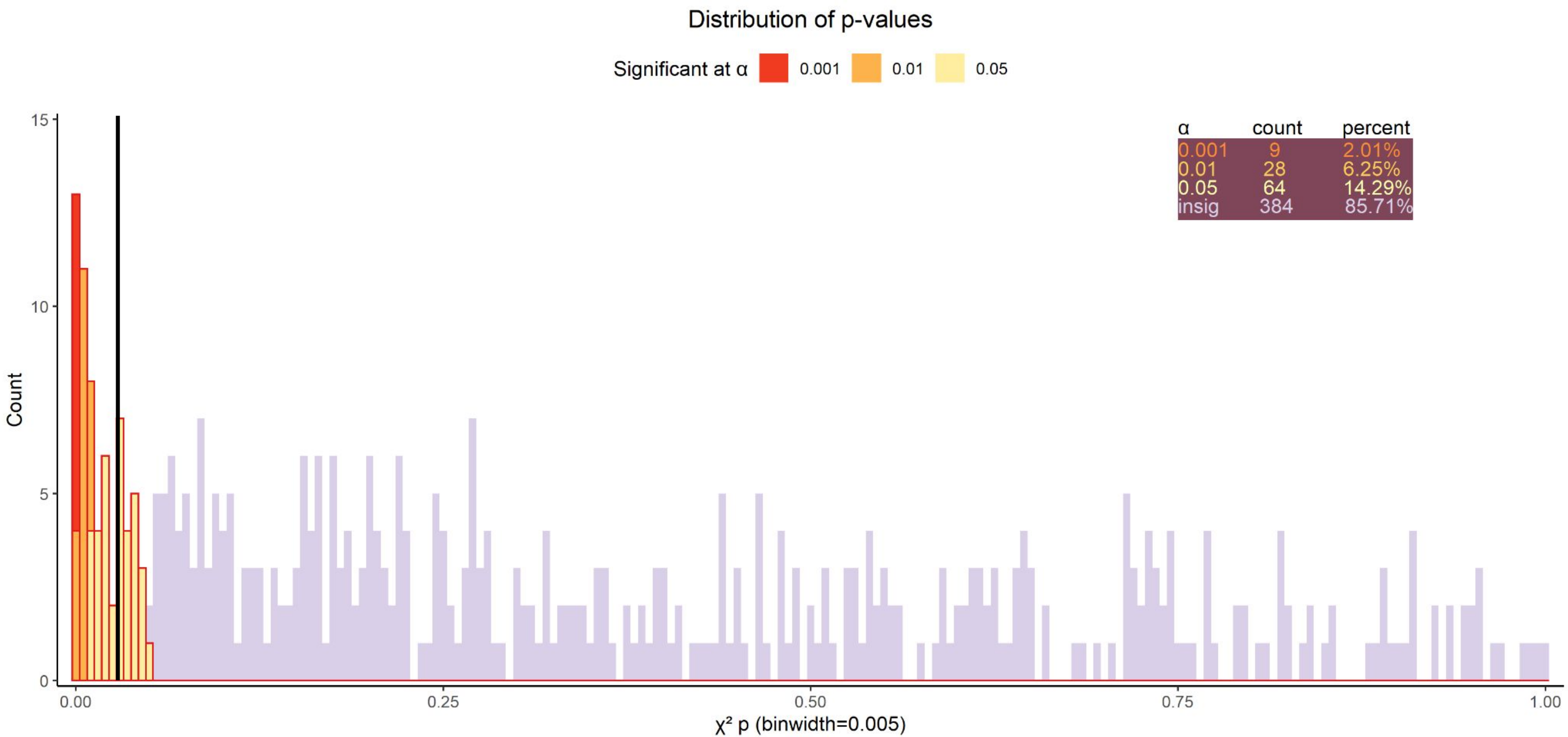
a) Consider degree of deviation, per $\chi 2$ test
b) Consider deviating digits, per Z-test

Assumption 2: PA's $\chi 2$ value

With a $\chi 2$ value of 16.69, Biden's PA results do differ overall from predictions under Benford's law.

However, of the 64 distributions found to be non-Benford overall, 52 (29 D, 23 R) yielded even higher $\chi 2$ values .

Assumption 2: Biden's PA distribution is an outlier



Biden's PA χ2 value deviates from Benford, but is not a significant outlier.

Assumption 2: PA's differing digits

Per the Z-test, Biden's 2020 PA results differ significantly only at digit 4.

This makes it one of 200 party/state/year distributions that differed at one digit. 46 (10.2%) of all distributions differed at more than one.

Assumption 2 Verdict

Assumption 2 is not supported.

Biden's "non-Benfordness" in Pennsylvania do not make the results an outlier in the dataset.

<u>Assumption 3</u>

Deviations from Benford’s law indicate foul play.

Assumption 3: Deviations indicate foul play.

Practical Considerations

Any number of other variables might impact how votes are distributed in a state, from year to year; even if the assumptions 1 and 2 held true, that “at best” might invite some additional research.

Assumption 3: Deviations indicate foul play.

Anomalies do not inherently indicate blame.

If one party benefited substantially from fraudulent votes, that would presumably (again, holding assumptions 1 and 2 to be true) cause their distribution to differ from Benford. The same would happen if the other party decided to systematically throw out their votes.

Assumption 3 Verdict

There is no logical justification to assume that a truly anomalous set of election results must be a consequence of fraud by a particular party.

Assumption 3 is unsupported.

Conclusions

Conclusions

Of three assumptions crucial to the original claim, only the first has some tenuous evidence. Biden’s Pennsylvania results do not stand out as suspicious in the context of recent elections, nor do we observe significant differences in how well the individual parties adhere to Benford’s law.

More broadly, Benford’s law does not model recent county-level Presidential results well enough to be applied as a litmus test.

Thanks!

Link Github repo (somewhat WIP):
https://github.com/brgrhrng/Benford_Project

Further reading:
Mebane (2020): Discusses precinct-level claims
http://www-personal.umich.edu/~wmebane/inapB.pdf

Nigrini (1996): Seminal paper on Benford's law and tax fraud
https://www.proquest.com/docview/211023799

Addendum: Do we have a multiple comparison problem?

We conducted 448 χ^2 tests. Every test has an independent chance of yielding a Type I error, equal to α.

At α 0.05, this basically guarantees some elections are wrongly found to be non-Benford. Maybe we should have used a lower α rate?

α level	p(0 type I errors)	Expected type I errors	Observed deviations
0.05	0	22.4	64
0.01	0.01108	4.48	28
0.001	0.63876	0.448	9

p(0 type 1 errors) = $(1-\alpha)^n$ Expected Errors = $\alpha * n$

However, even at α 0.05, the actual # of deviations far exceeds the expected rate of Type I errors. This is consistent when lower α. So, chance alone can't account for the frequency of non-Benfordness.