Analysis of Voting Anomalies in Several States 2000-2020

Full disclosure: The author of this report voted for President Trump in 2020. He holds a STEM PhD and has over 10 years of experience in data analysis for defense systems.

Note: if you are viewing the MS Word version of the report through Chrome (or anything other than actual MS Word), the equations may not show up correctly. In this case, you may want to view the PDF version of the report.

Summary

The 2020 US presidential election continues to be mired in controversy amid accusations of voter fraud. This report presents an examination of the voter data in six states from 2000-2020 with the goal of making an independent determination that these allegations are plausible. The conclusion of the analysis is that massive, systemic voter fraud is not ruled out due to strong, structured trends in differential voting statistics repeated across all states examined that do not appear (collectively) to have a natural demographic explanation.

The clearest example is in the state of GA. Figure 1 shows differences in Republican presidential percentage point (pp) scores for each GA county between the 2004 and 2020 elections. (The label of blue or red is determined based on the county's preference in 2008). The data are plotted against a log scale on the X axis by total number of votes in each county. The data for counties below 10k votes appear normal, but the data from counties with 10k-100k votes show a consistent decrease in Republican voting percentage with an almost perfect slope of 25pp/dec (logarithmic decade, not years). This means that (on a linear scale) the decrease in Republican votes follows the trend of a logarithmic curve with population size.
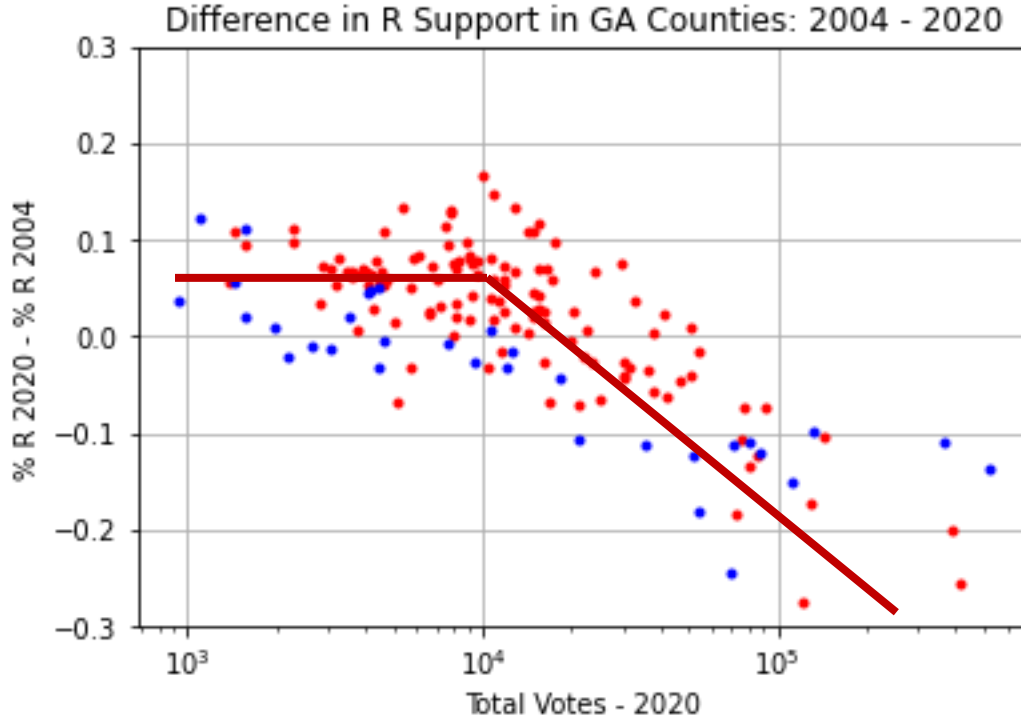
Figure 1. Differences in Republican presidential voting in GA counties, 2004-2020

The remainder of this report shows that this large trend is due to separate trends of the same form but with smaller magnitude occurring in the differential data of the 2008, 2016, and 2020 elections. From Figure 1, it appears that the metro Atlanta counties do not fit the stated trend, but it will be shown that in fact they are in family with the other counties when the difference in relative Republican leaning is accounted for (See Figure 14). **Trends of the same form and similar magnitude occur in every other state examined in this report for the 2016 and 2020 election, and some for the 2008 election**. The states chosen were FL, GA, NC, OH, PA, and VA. **These states were chosen randomly by the author**, although their choice was influenced by the fact that these are generally considered "swing" states politically and/or were reported to have other anomalies in the 2020 election.

In fact, the regularity of the trend and results leads the author to propose the following functional form for an algorithm that is consistent with the trends in the data. **This algorithm is conjecture**, however, performing its inverse on the data appears to yield results consistent with normal voting and demographic patterns. Therefore, it represents a reasonable "educated guess" as to a systemic process that could generate the observed anomalies.

If T_i is the number of total votes in a given county and R_i are the Republican votes, then the "missing" Republican votes in that county (ΔR_i) appear to follow the following equation

$$\Delta R_i = \begin{cases} k \log_{10}(T_i/T_0) R_i & \text{if } T_i > T_0 \\ 0 & \text{otherwise} \end{cases}$$

where k is a slope parameter and T_0 the intercept (onset) parameter. For the data in Figure 1, $k = 0.225$ and $T_0 = 10^4$. If the effect of this equation is applied to the data in reverse, it yields the data in

Figure 2. (In this report, the label XXXX.1 for a year is used to denoted "adjusted" data according to this conjectured trend).
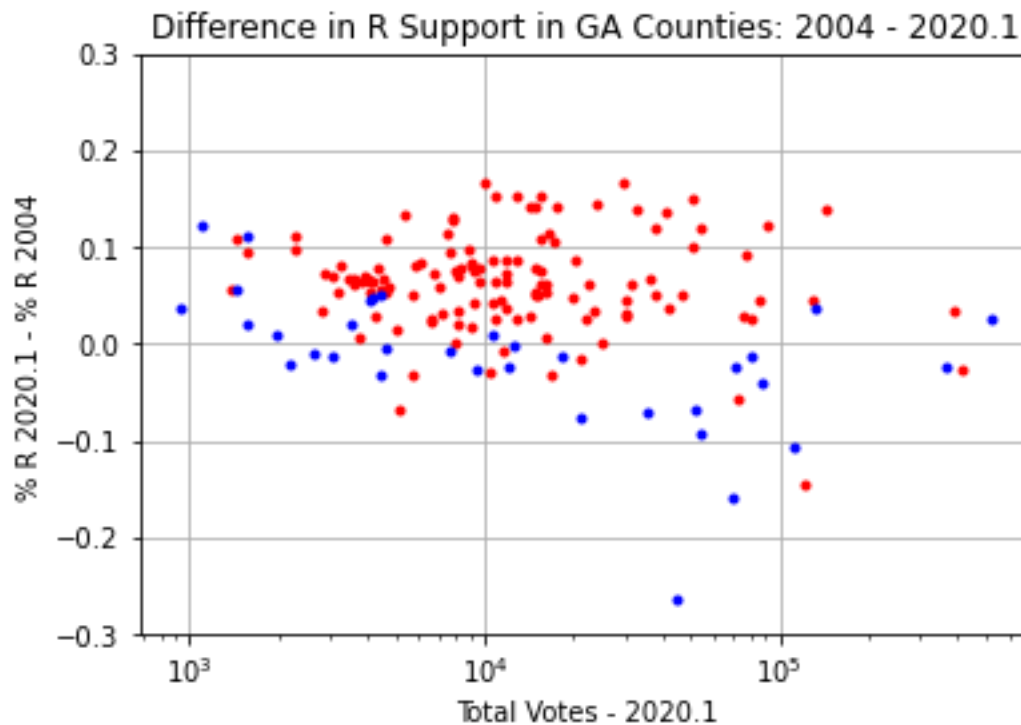

Figure 2. Adjusted (k = 0.25, L_0 = 4) differences in Republican voting in GA counties, 2004-2020

The data in Figure 2 are much more in line with what we would expect to see in a standard demographic analysis over time. There are clear clusters of Republican and Democratic voting patterns. The total number of votes in the adjustment is 610k.

Tables at the end of each analysis section summarize the estimated impact of the anomalies if they are in fact due to artificial manipulation by the proposed algorithm. In multiple states, this analysis shows the potential for 100s of thousands to millions of switched votes based on this assumption.

Furthermore, a cursory analysis of the adjusted data sets shows that the proposed algorithm leads to corrections that appear resilient under leading digit analysis (Benford's law).

Control Cases

First, let us establish a baseline by looking at voting patterns that do not show anomalous results. For example, in the 2004 election, electronic voting was in its infancy, so systemic intervention of the kind alleged in 2020 should be conclusively absent.
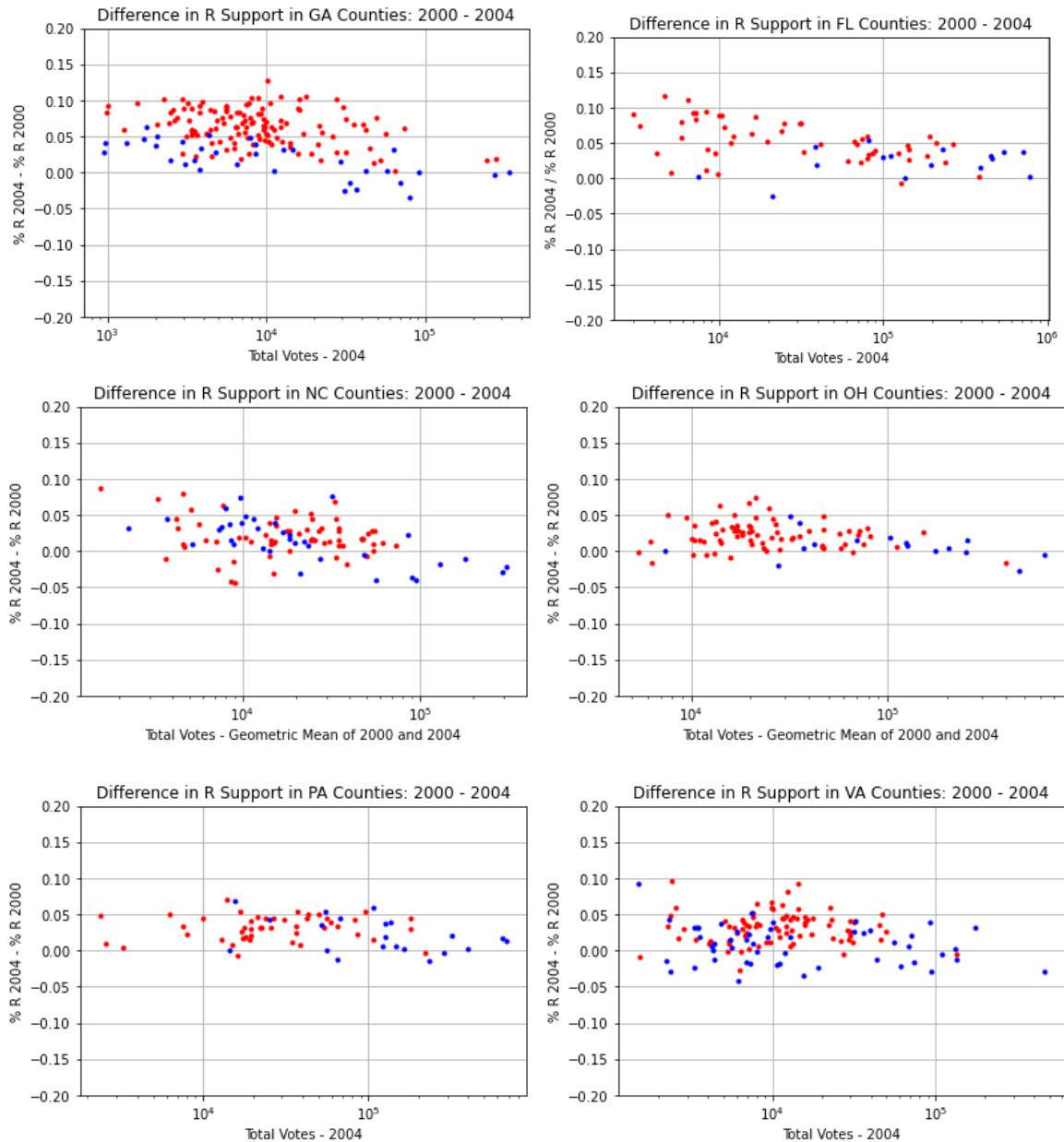
Figure 3. Voting Pattern Differences between 2000 and 2004

Examining Figure 3, we see some clustering behavior typical of smaller vs larger and blue vs red counties. This election proceeded highly along party lines. We see that John Kerry was not a compelling candidate to the red counties, which tended to favor Bush by about 5 pp more than in the 2000 election. Kerry was able to gain support compared to Gore in some larger blue counties (and highly blue smaller counties), but overall support for Bush did not decline by more than a few pp in these places.

The FL differential data appears to exhibit a slight downward trend like that discussed in the intro summary. However, application of the log ratio test developed in later in the report indicates that this data results from distinct shifts in two separate clusters rather than an constant slope in family with the other anomalies.

In the 2008 election, we also see several examples of patterns within the realm of expectation. In the swing states of PA, OH, and VA, Obama improved his support by an average of 5 pp compared to Kerry in in a trend visible across all counties.
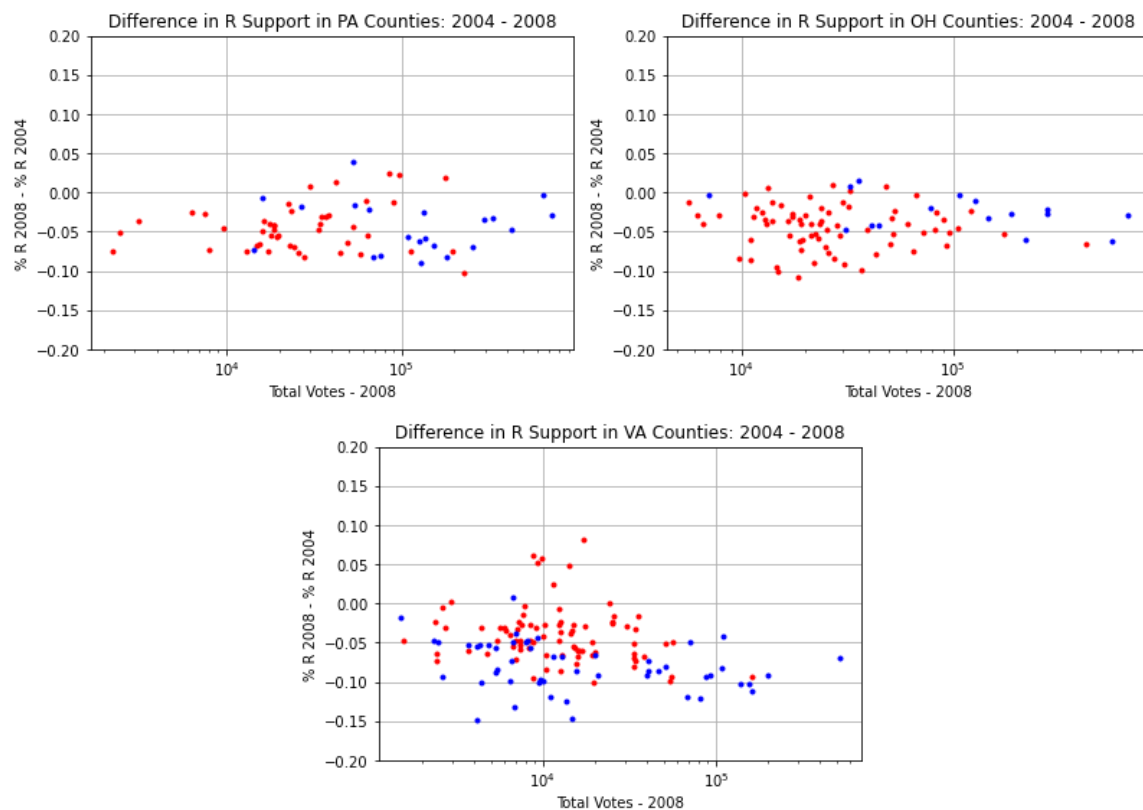

*Figure 4. Unremarkable Voting Pattern Differences between 2004 and 2008*

Comparing the 2004 to the 2008 data reveals another expected result – the variance of the data in the 2008 election among the counties is larger than in the 2004 election. This is expected since 2004 was an incumbent election and 2008 was not. In general, people tend to have a strongly formed opinion about an incumbent president which is difficult to sway barring significant externalities.

The comparisons between 2008 and 2012 data are not shown here, but are in most cases unremarkable, except for FL, which is a case that will be discussed later. Overall, they show a slight decrease in support for Obama, but are otherwise flat with quite a small variance.

Anomalous Cases

2008

While the 2008 election appeared normal in PA, OH, and VA, the data from GA, FL, and NC reveal the beginning of the logarithmic anomaly that is the focus of this report. In particular, the data from GA in Figure 5 reveal a clear bimodal behavior in the differential percentage points beginning with counties of 10k voters or more. The data in FL and NC show a similar onset of the behavior at 10k votes, with a similar slope. The apparent slope of the anomaly in each case is 10 pp/dec.
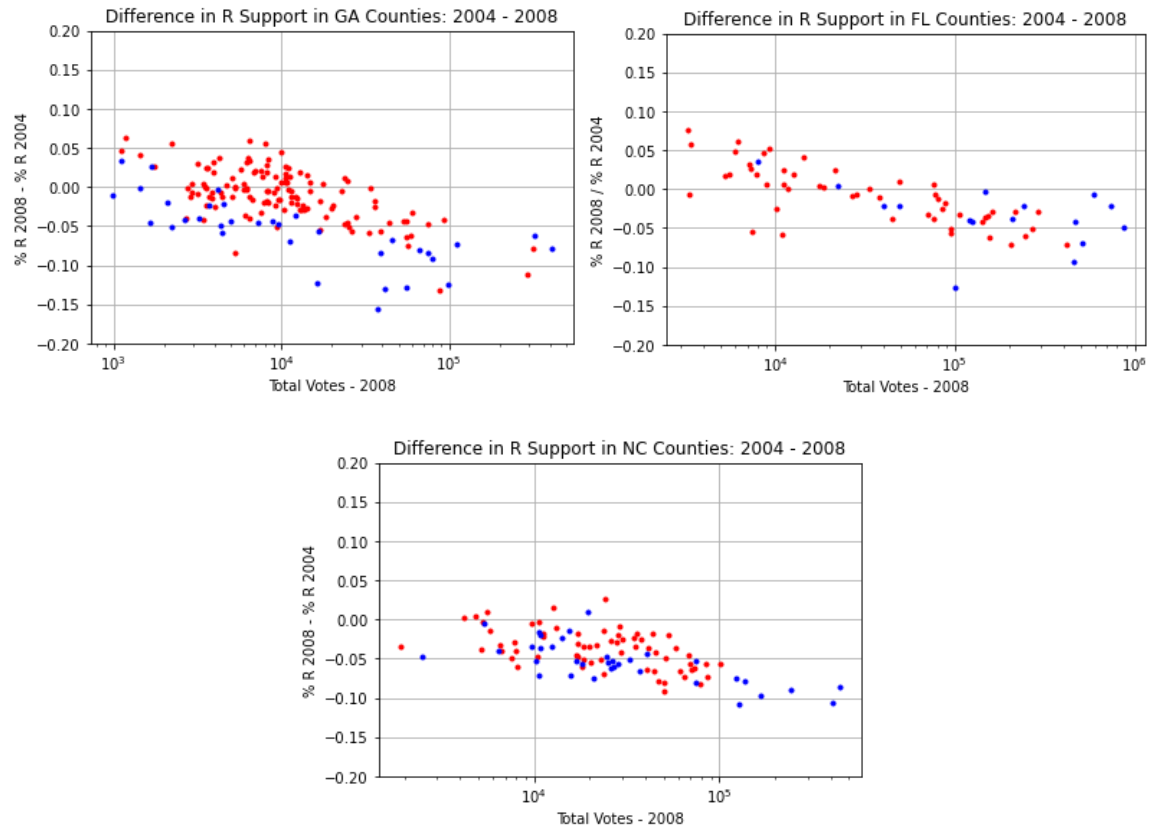
Figure 5. Anomalous Voting Pattern Differences between 2004 and 2008

By themselves, these data might be discounted due to demographic or statistical anomalies associated with the historic 2008 election. (In particular, the pattern is less remarkable in FL and NC, which have larger and fewer counties than GA.) However, by the 2016 election, these anomalies had spread to every state examined in this report.
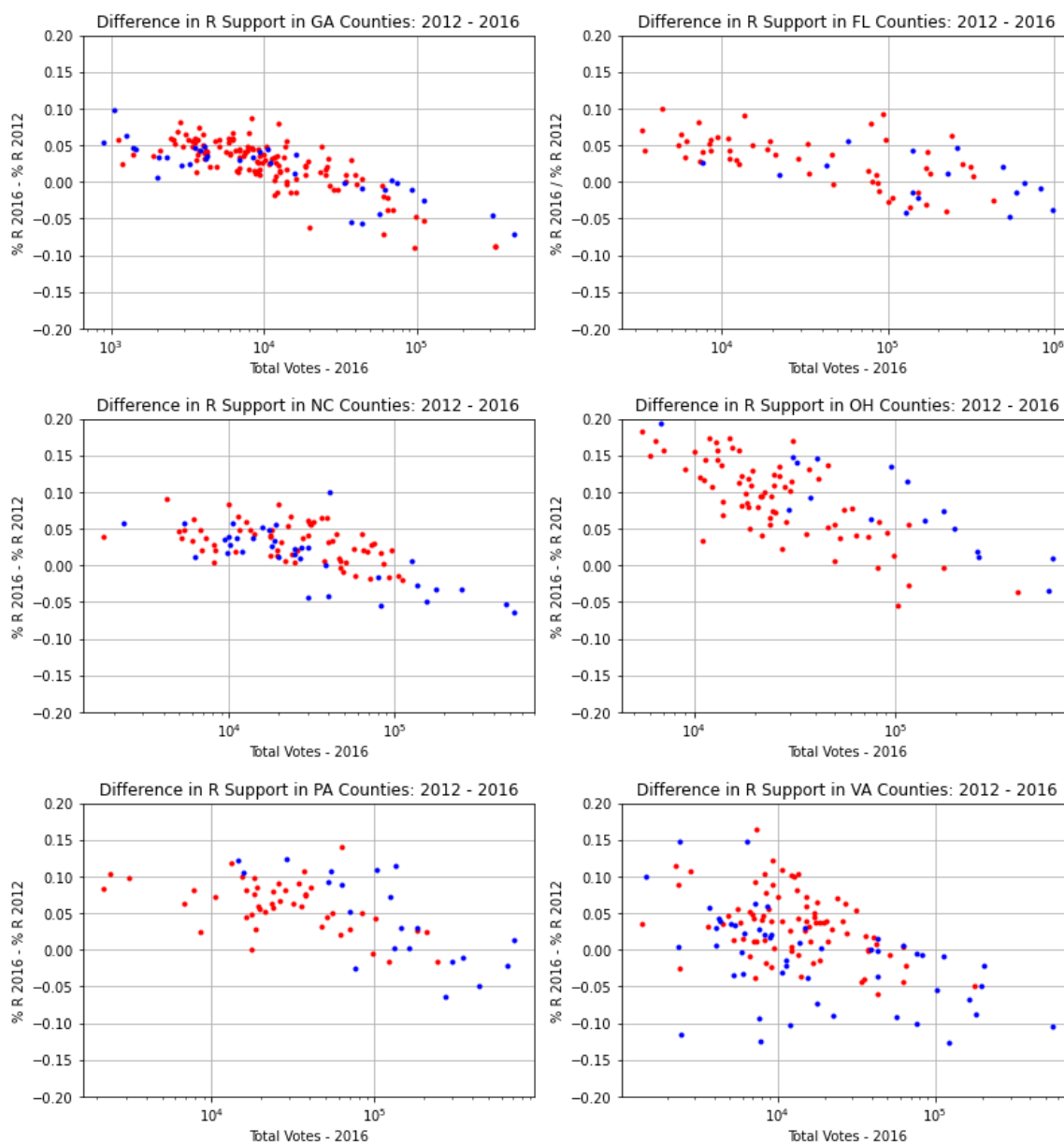
Figure 6. Anomalous Voting Pattern Differences between 2012 and 2016

Figure 6 shows that in 2016, every state examined in this study shows evidence of the logarithmic voting anomaly. In GA, FL, and NC, the anomaly appears to begin at the same county size as in 2008, 10k voters. The apparent slope of the anomaly in all 6 states is close to 10pp/dec. (However, we will soon see that FL is a special case with regard to this trend). The specific slope and onset of the trends will be further analyzed in later sections.

2020
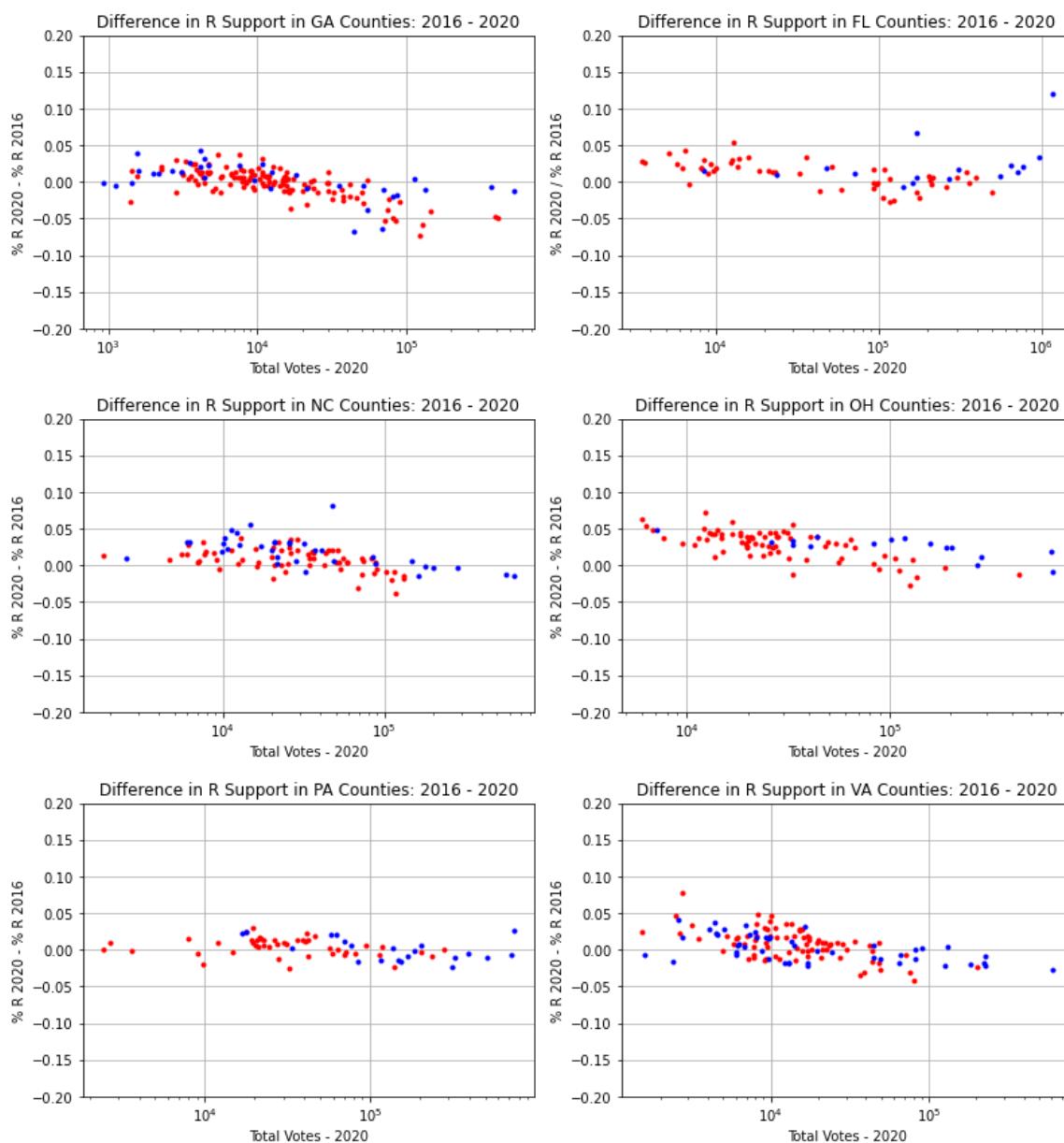
Figure 7. Anomalous Voting Pattern Differences between 2016 and 2020

Figure 7 shows that the anomaly is present again in every state examined in this report, albeit with much smaller magnitude. In general, it is only detectable in some of the states due to the extremely low variance of the data. In general, this low variance indicates that the voters in this data set are extremely firm in their opinion of Trump, with a handful of outliers. Indeed, the standard deviation of the data from the trend is not more than a few percentage points.

In this data set, the slope of the logarithmic anomaly appears close to 2.5-5pp/dec. The onset points of the anomaly in each state appear identical to 2016 and 2008 (if applicable), (except for FL, which has exceptional structure discussed in detail later).

Modeling the Anomaly

The existence of a consistent logarithmic data anomaly in the differential percentage point data of counties across multiple elections and states lends significant circumstantial evidentiary support to the notion that an artificial process may have influenced those votes. The next logical step is to attempt to fit a model to these anomalies.

It should be noted that the results of this section and following sections are conjecture on the author's part (albeit an "educated guess" based on the data). Readers are invited to examine the results with a critical mind and consult external sources of evidence to verify the plausibility of the model proposed.

Note: This section is highly technical. Less technical readers may wish to review the proposed equation for differences in votes and skip to the final bold print describing the methodology for determining parameters.

In particular, we propose the following model. If T_i is the number of total votes in a given county and R_i are the total Republican votes, then the "missing" Republican votes in that county (ΔR_i) appear to follow the equation

$$\Delta R_i = \begin{cases} k \log_{10}(T_i/T_0)\, R_i & \text{if } T_i > T_0 \\ 0 & \text{otherwise} \end{cases}$$

where k is a slope parameter and T_0 the intercept (onset) parameter. For the purposes of this analysis we will assume the votes are changed from R to D (rather than simply deleted).

Two things should be noted about this model. Firstly, the "reported" R vote total is not the true total (based on the working assumption). Secondly, the difference in R votes will have a small effect on the percentage point score in counties that lean heavily D, as the relative proportion of votes changed will be small compared to heavily R counties.

To address the first concern, let us define the proportionality constant $\rho_i = k \log_{10}(T_i/T_0)$ to simplify the analysis. (This represents the proportion of R votes changed.) Therefore (for counties after the onset)

$$\Delta R_i = \rho_i R_i$$

Let $\tilde{R}_i$ be the reported Republican vote count. Then

$$R_i = \tilde{R}_i + \Delta R_i = \tilde{R}_i + \rho_i R_i$$

Solving for R_i, we have

$$R_i = \frac{\tilde{R}_i}{1 - \rho_i}$$

Now, to address the disproportionate number of R voters in various counties, we propose to examine the ratio of the vote percentages between the two elections. Let $\tilde{r}_{12}$ be the ratio formed used the reported percentage point data.

$$\tilde{r}_{12} = \frac{\tilde{R}_2/T_1}{\tilde{R}_1/T_2} = \frac{R_2}{R_1}\frac{(1 - \rho_2)T_1}{(1 - \rho_1)T_2}$$

The logarithm of this ratio can be decomposed into three terms, one containing the log of the true ratio, one related to the existing action of the proposed model on election 1, and the last relating to the action of the proposed model on election 2.

$$\ln \hat{r}_{12} = \ln \frac{R_2}{R_1} \frac{T_1}{T_2} - \ln(1 - \rho_1) + \ln(1 - \rho_2)$$

Using the Taylor expansion for logarithms around 1,

$$\ln(1 - \rho_2) = -\rho_2 + O(\rho_2^2)$$

$$\ln \hat{r}_{12} = \ln \frac{R_2}{R_1} \frac{T_1}{T_2} + (\rho_1 - \rho_2) + O(\rho_2^2, \rho_1^2) = \ln \frac{R_2}{R_1} \frac{T_1}{T_2} + \left(k_1 \log_{10} \left(\frac{T_1}{T_0} \right) - k_2 \log_{10} \left(\frac{T_2}{T_0} \right) \right) + O(\rho_2^2, \rho_1^2)$$

We can further approximate since, due to the fact that $T_1 \approx T_2$, their logarithms are similar compared to the onset point. (If $k_1 = 0$ is assumed, this approximation need not be made).

$$\ln \hat{r}_{12} \approx \ln \frac{R_2}{R_1} \frac{T_1}{T_2} + (k_1 - k_2) \log_{10} \left(\frac{T_2}{T_0} \right)$$

Assuming the first term (the log "true" ratio) produces predominantly random/clustering effects (similar to Figure 3) without direct slope, this equation suggests that if we plot the natural log of the percentage point ratios between elections, the slope of this graph will be $k_1 - k_2$. The sloping effect should be common to the entire set of counties, including D-leaning counties. The onset point of the anomaly should be clearly visible where the slope changes. The slope should be measured in within the first decade from the onset point, as otherwise the (neglected) higher order terms in the Taylor expansion will begin to change the slope.

We now proceed to apply this analysis methodology to the states in consideration.
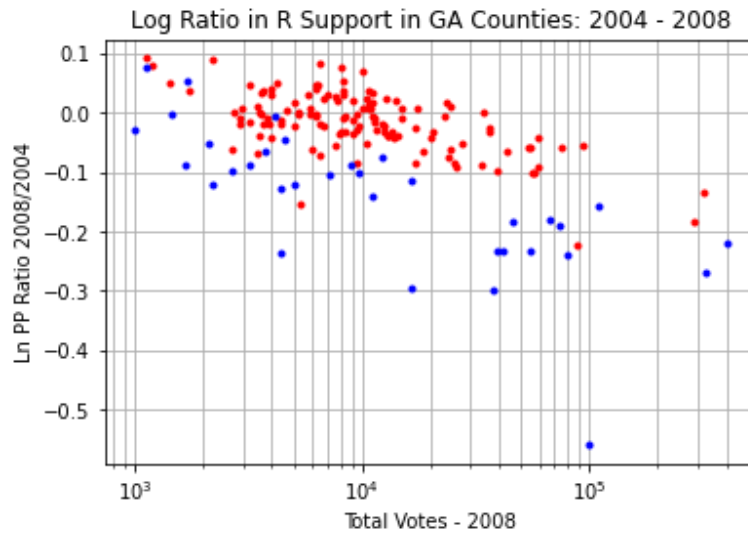
GA Anomaly Analysis

Figure 8. Log Ratio of Republican PP between 2004 and 2008 in GA

Figure 8 supports the earlier assessment of $k = 0.1$ for slope parameter and $T_0 = 10^4$ for the onset parameter for the 2008 election in GA (assuming k=0 for 2004). Plotting the data in this way reveals that the same trend exists in the D-leaning counties as the R-leaning counties, albeit offset by approximately 0.1 on the graph.

As the total of number of votes grows, the data become increasingly non-linear (quadratic). This is due to the neglected terms in the Taylor series approximation of $\ln(1 - \rho_2)$ in the above derivation, since $\rho_2 \ll 1$ no longer holds. Therefore, for the purposes of this model, the approximation of k should be made in the first decade of the graph subsequent to the onset point.

We can now "adjust" for this trend in the 2008 data. We will label this adjusted data "2008.1". Figure 9 shows this adjusted data, which resembles much more strongly an expected voting pattern for this election (compare to e.g. Figure 4 in other states)
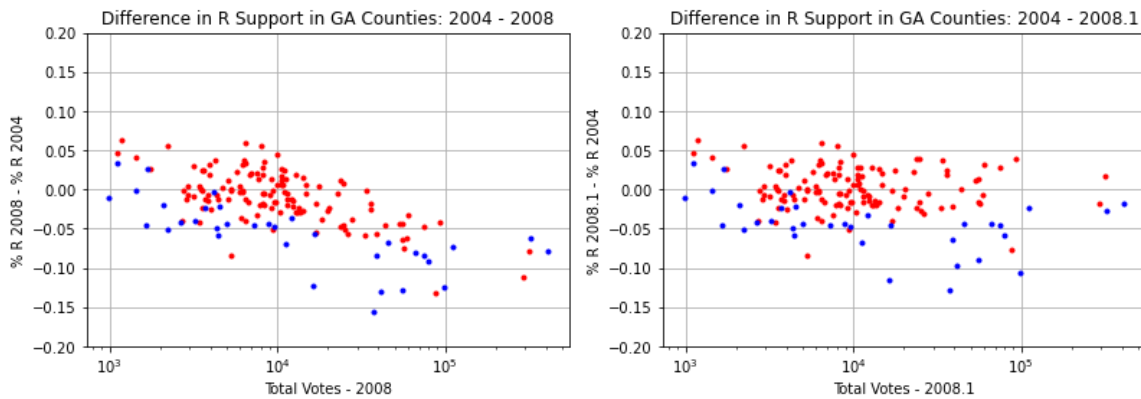
Figure 9. Original and Adjusted Differences in 2004 and 2008 Elections in GA

As the 2012 to 2008 data showed no obvious anomaly, we will assume the model parameters remain $k = 0.1, T_0 = 10^4$. (Note k must remain constant or else the graph would slope up!)
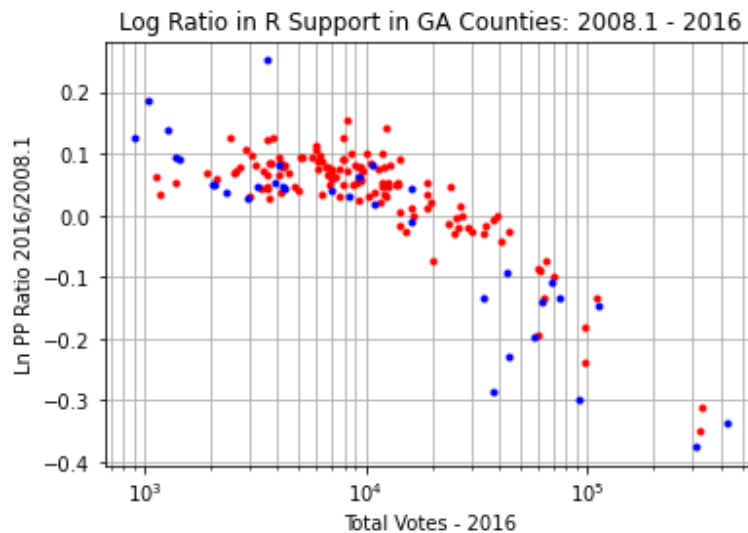
Figure 10. Log Ratio of Republican PP between 2008 (adjusted) and 2016 in GA

In order to determine k for 2016, we apply the methodology to the ratio the adjusted 2008.1 data (for which k=0 is assumed) to 2016. We see a slope of 0.2/dec emanating from the same onset point, which indicates $k = 0.2, T_0 = 10^4$. Again, the trend for "blue" counties overlaps the red trend (this time with no appreciable offset). (Note that, as k increases, the initial linear slope needs to be measured closer to the onset point, due to increasing Taylor series truncation error).
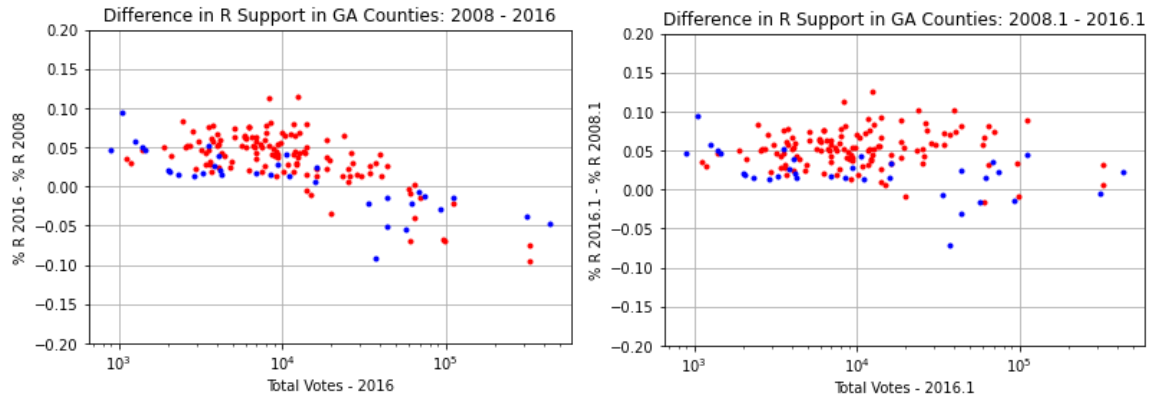

Figure 11. Original and Adjusted Differences in 2008 and 2016 Elections in GA

Adjusting the 2016 data according to the proposed model yields the results in Figure 11, which again, are much closer to expected voting patterns.
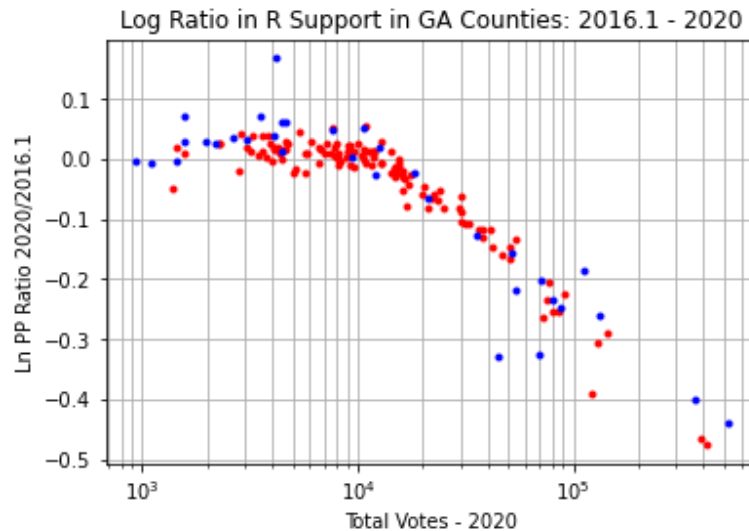
Figure 12. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in GA

Figure 12 compares the adjusted 2016 results to 2020. The first decade suggests parameters of $k = 0.225, T_0 = 10^4$. (This result is unusually easy to read off the chart, because the variance is low, and the log ratio below 10k votes is virtually zero.)
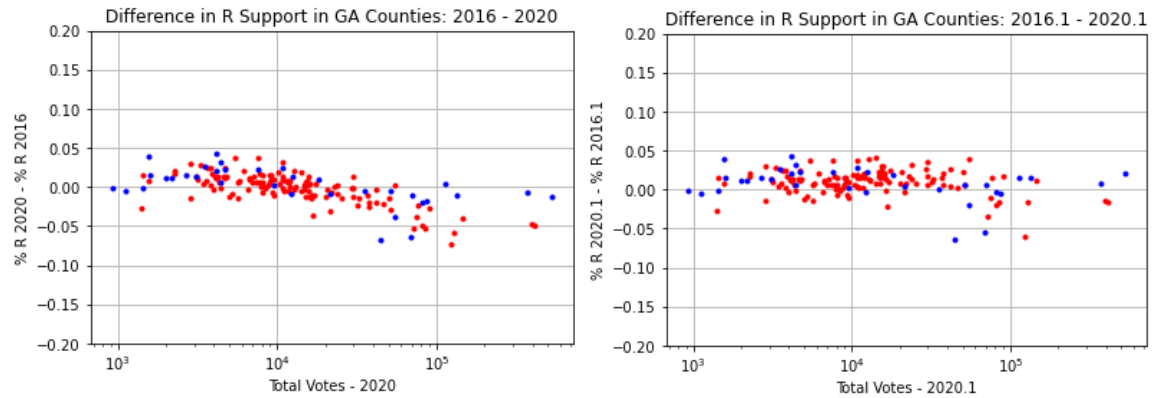
Figure 13. Original and Adjusted Difference in 2016 and 2020 Elections in GA

Figure 13 shows the result of adjusting the data to account for the model. The data now strongly resemble the expected pattern for an incumbent election in which virtually no one has changed their mind about the candidates in question.
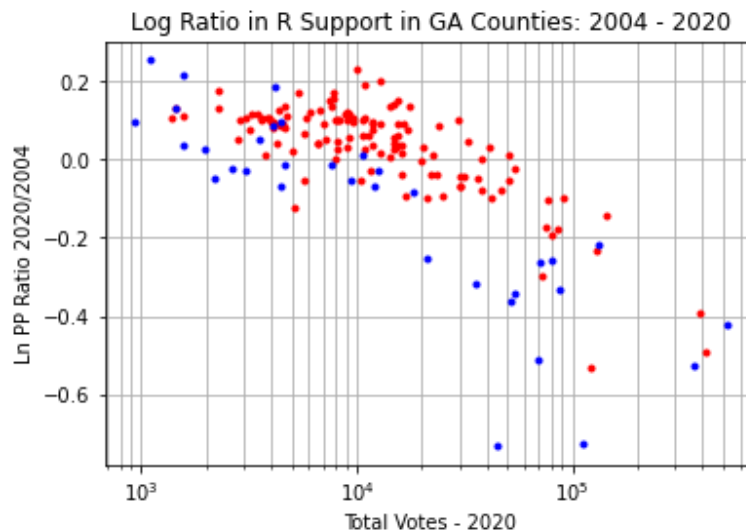
Figure 14. Log Ratio of Republican PP between 2004 and 2020 in GA

If the reader objects to the large slope of Figure 12 being a result of the compounding of adjustment over adjustment, the 2020 data can be compared to the "baseline" election of 2004. Figure 14 shows that the hypothesized value of $k = 0.225$ for the 2020 election is consistent with the trend in these unmodified data sets.
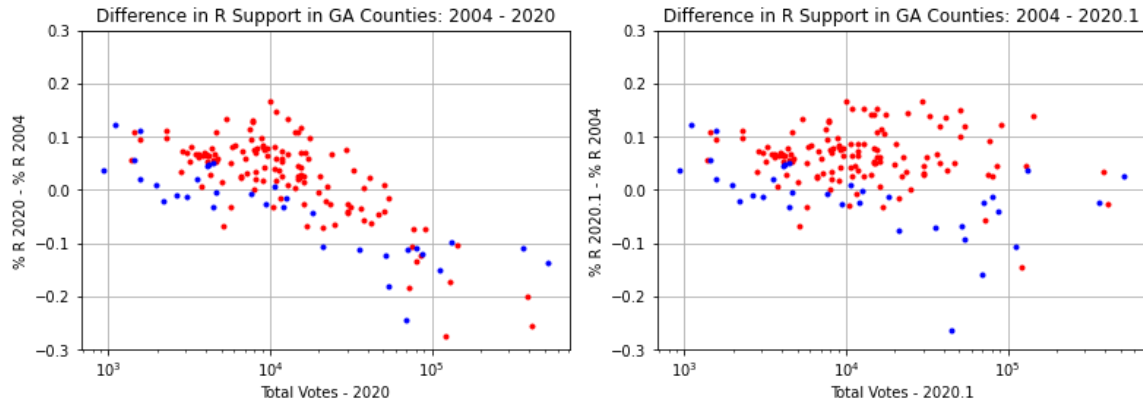
Figure 15. Original and Adjusted Difference in 2004 and 2020 Elections in GA

Continuing to use 2004 as a baseline and comparing the "adjusted" 2020 data, we see that the proposed model results in much more plausible voter statistics between these years than the original data set (left).

Table 1. Summary of Modeled Adjustments to GA Election Data

Year	T_0	k	Total Changed Votes	Official R Vote Margin	Official R %	Predicted R %
2008	10k	0.1	171k	205k	52.2	56.6
2012	10k	0.1	174k	308k	53.4	57.9
2016	10k	0.2	397k	211k	50.4	60.0
2020	10k	0.225	610k	-12.7k	49.3	61.5

Lest we lose sight of the meaning of what is presented here amongst the technicalities, let us take a minute to interpret this final result. This model (if true), predicts that, for a county of 100k voters, 22.5 percent of R votes were switched to D in the 2020 election. Table 1 shows that the total predicted effect added over all counties for 2020 is approximately 610k votes switched, or nearly 13% of all votes cast in the election. If true, this would turn GA from a 49.3% loss to a 61.5% win for Trump.

FL Anomaly Analysis

The trends involved in the FL vote statistics are easily among the most complex seen in this data set. In several years, it appears two applications of the proposed model may have been made to the data in several years, starting at distinct onsets.
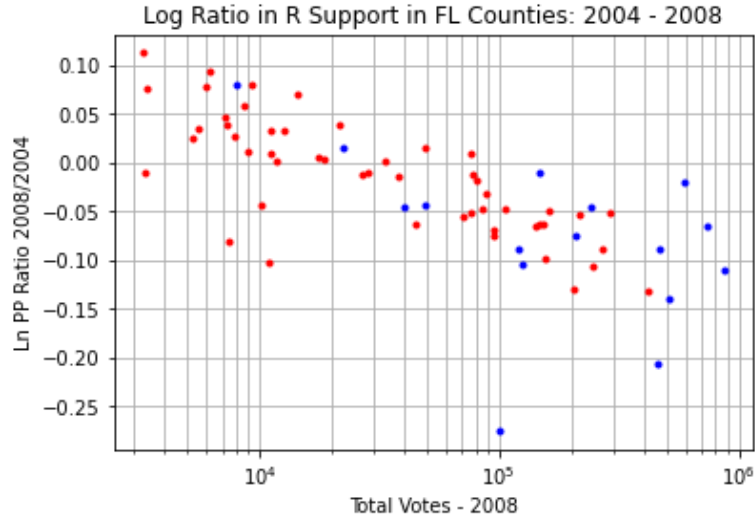
Figure 16. Log Ratio of Republican PP between 2004 and 2008 in FL

Figure 16 shows the anomaly is present with a slope of $k = 0.1$. The onset point is difficult to estimate due to the lower limits of the data and the continuation of the apparent slope. Conservatively, we estimate an onset point of $T_0 = 10^4$ rather than $T_0 = 5000$. The results of this adjustment are shown in Figure 17. The adjusted data are free of the anomaly and appear reasonable.
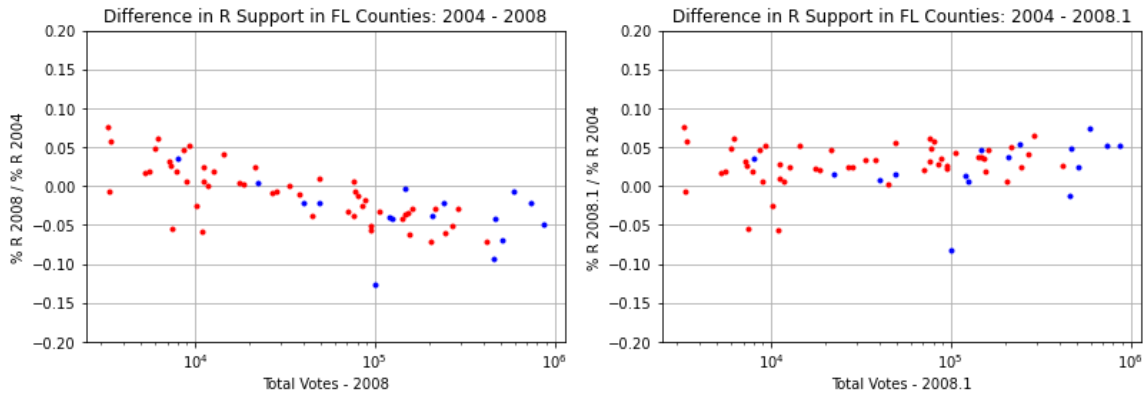

Figure 17. Original and Adjusted Difference in 2004 and 2008 Elections in FL

Unlike in other states, in the 2012 election, the parameters of the anomaly appear to have changed in FL. Two distinct onset points are visible at $T_0 = 10^4$ and $T_0 = 10^5$. Therefore, the adjustment of this data was performed in two steps. First, the estimated slope of $k = 0.05$ at $T_0 = 10^4$ evident in the left plot of Figure 18 was removed. The adjusted data (2012.1) was plotted again against the adjusted 2008 data (2008.1). This revealed an additional slope of $k = 0.15$ with an onset of $T_0 = 10^5$, which was adjusted to produce the data set 2012.2.
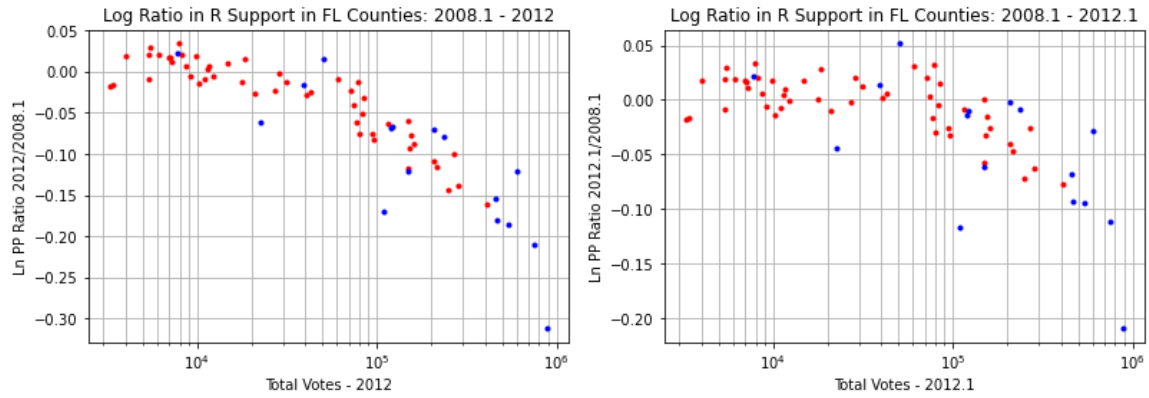
Figure 18. Log Ratio of Republican PP between 2008 (adjusted) and 2012 in FL

Figure 19 shows the results of applying these adjustments (right) in relation to the original data (left). One sees that the unsightly "hump" around 80k votes has been removed, and the new data are consistent with the low variance distribution expected in an incumbent election. **This adjustment is notable, as it represents the only occurrence in this analysis where the slope parameter k was seen to (locally) decrease. In all other cases, k increases over time or remains the same.**
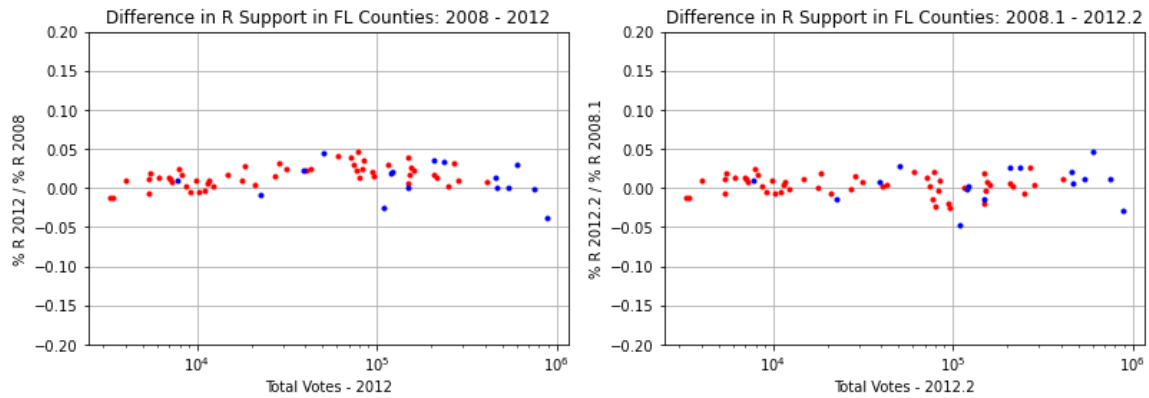
Figure 19. Original and Adjusted Difference in 2008 and 2012 Elections in FL

Figure 20 shows the difference between the adjusted 2012 data and the 2016 election data. Again, we see that there appear to be two onset points at $T_0 = 10^4$ and $T_0 = 10^5$. The first anomaly has slope $k = 0.15$ and is adjusted in the data set 2016.1. The second anomaly is more difficult to estimate due to the apparent presence of a distinct cluster of increased support for Trump in high population counties. Following the lower envelope of the data suggests the slope could be around 0.1. It was found that $k = 0.08$ produces data that is overall consistent with the 2020 data. However, this slope number in particular should be treated with lower confidence.
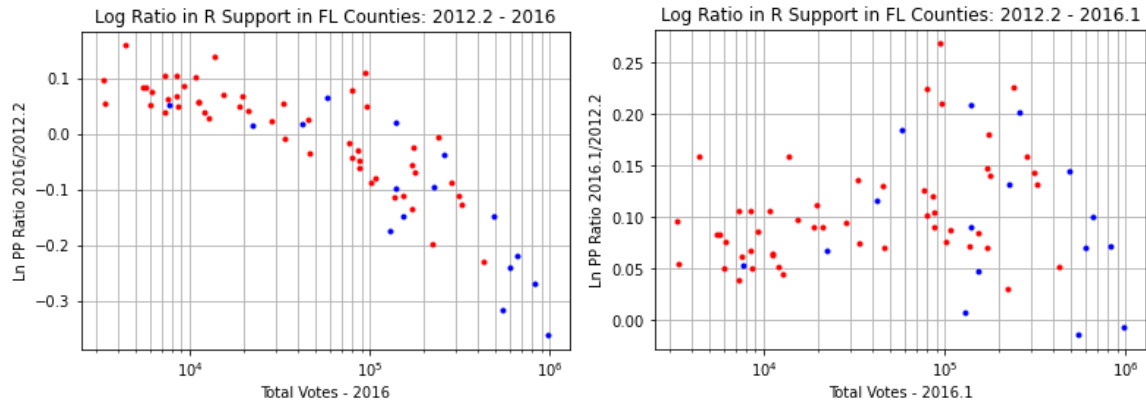
Figure 20. Log Ratio of Republican PP between 2012 (adjusted) and 2016 in FL

Examining the overall adjusted result in Figure 20 shows a very different picture than the original results. In particular, there is a cluster of 20 pp increase in support for Trump vs Obama in counties with greater than 100k voters. The author has not investigated this in detail but believes it to be linked to strong Hispanic voter support for Trump. This cluster is effectively masked in the original data by the effect of the conjectured adjustments. The reader is invited to draw their own conclusions as to the potential origin of this cluster.
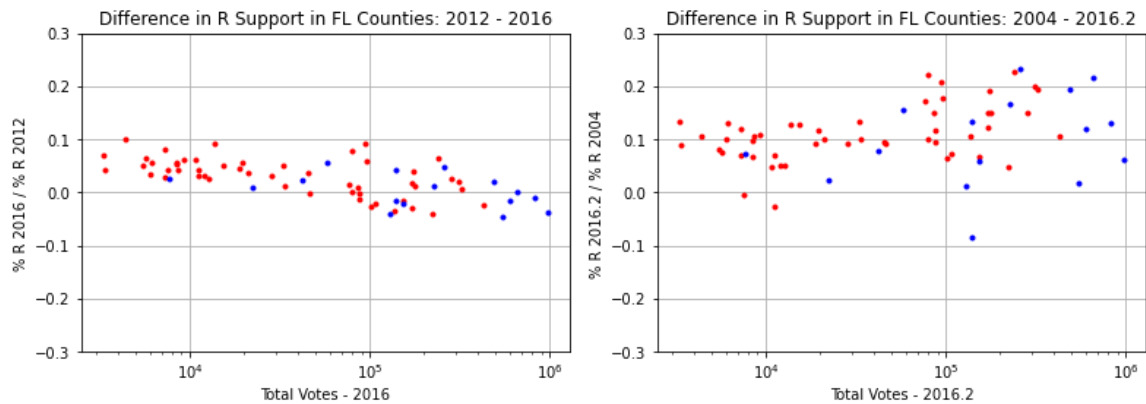
Figure 21. Original and Adjusted Difference in 2008 and 2012 Elections in FL

For 2020, Figure 22 shows that the anomaly in FL appears to once again follow a fixed slope with a single onset. (Full disclosure: the 0.08 slope adjustment for 2016.2 was fine tuned by the author to better match the single slope assumption for this graph. The charitable reader may choose to view this as backward recursion rather than a vulgar fudge.) The onset parameter is $T_0 = 10^4$ and the slope is approximately $k = 0.18$.
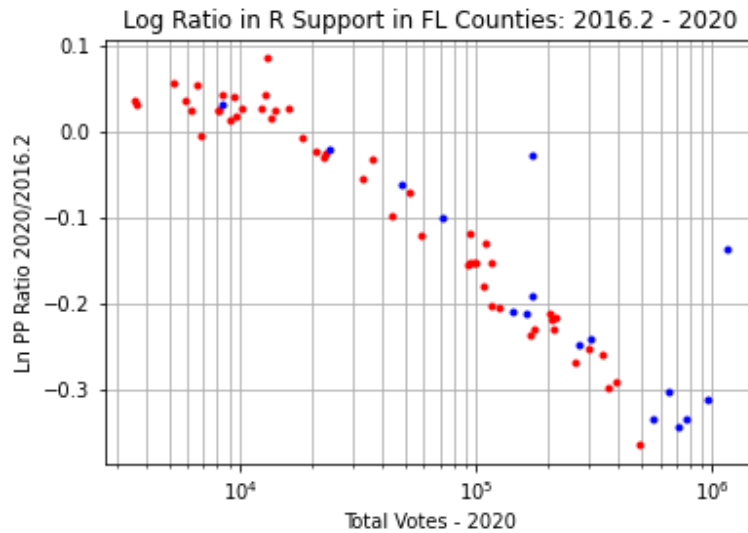
Figure 22. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in FL

Applying these parameters yields the adjusted data set 2020.1 shown in Figure 23. The adjusted data removes the trend visible between 10k and 100k votes, yielding a mostly flat graph consistent with an incumbent election (compare to Figure 19 2012 data).

The outliers in Figure 23 are noteworthy, in particular the one for Miami-Dade county, which shows that Trump increased his support by 20 pp vs 2016. The author finds it unlikely that this single large county would exhibit such an outsized increase in an incumbent election. A more likely hypothesis is that there is something wrong with the number in 2016. The reader is left to draw their own conclusions.
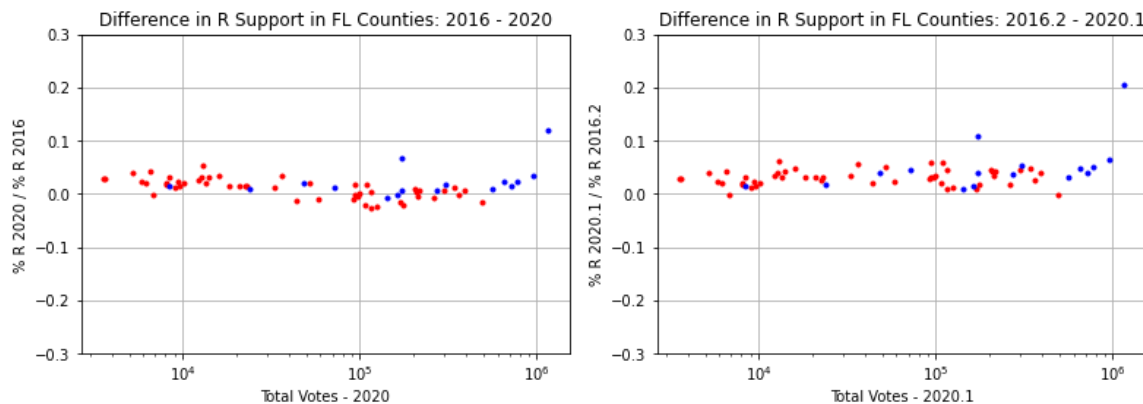
Figure 23. Original and Adjusted Difference in 2016 and 2020 Elections in FL

Again, to protect against the danger of piling adjustment upon adjustment in this analysis, we compare the final 2020 results to the "baseline" 2004 election. Figure 24 shows the log ratio analysis for these elections. The data are quite noisy due to the length of time between these elections, but the initial decade of the data appears to support the predicted slope of $k = 0.18$ (considering the increasing onset of quadratic terms in the log Taylor series)
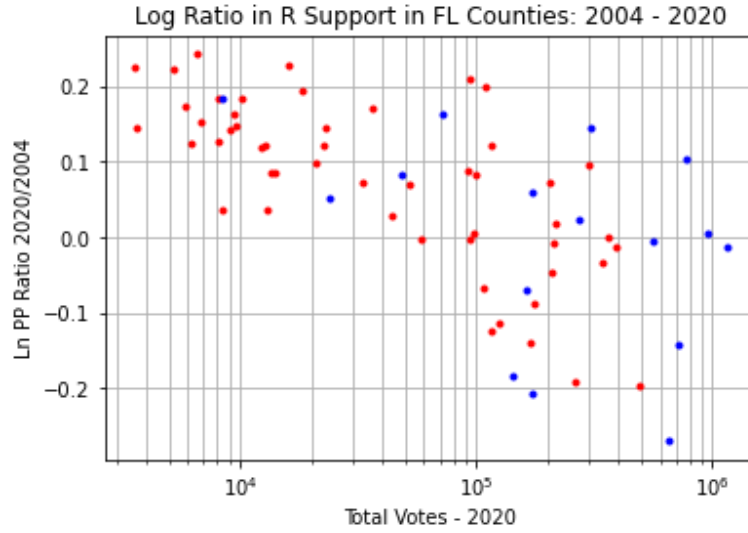
Figure 24. Log Ratio of Republican PP between 2004 and 2020 in FL

Figure 25 compares the original difference between the data sets and the difference from 2004 to the adjusted 2020.1 data set. The data appear reasonable, assuming the reader accepts the likelihood of the 20 pp cluster of Trump support increase in large counties.
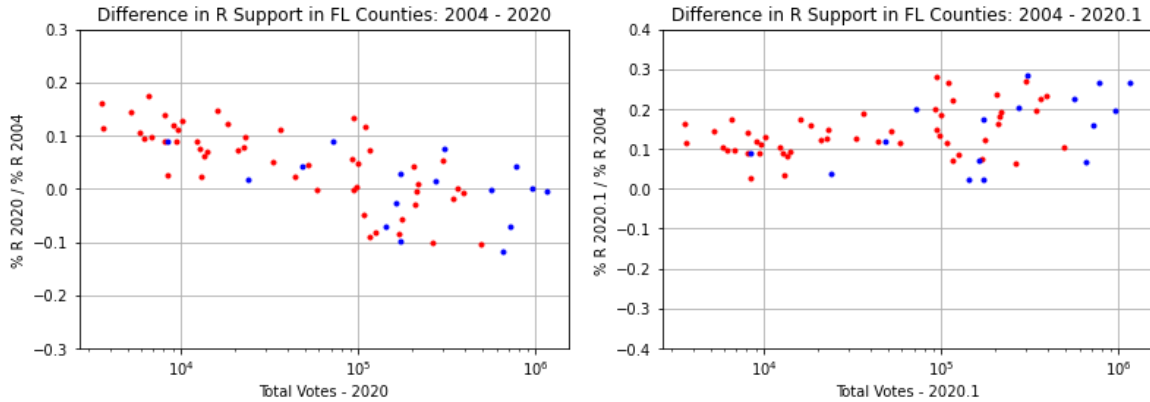

Figure 25. Original and Adjusted Difference in 2004 and 2020 Elections in FL

Table 2. Summary of Modeled Adjustments to FL Election Data

Year	T_0	k	T_0'	k'	Total Changed Votes	Official R Vote Margin	Official R %	Predicted R %
2008	10k	0.1			648k	-205k	48.4	56.3
2012	10k	0.05	100k	0.15	625k	-73.2k	49.1	56.5
2016	10k	0.15	100k	0.08	1492k	113k	48.6	64.3
2020	10k	0.18			2146k	372k	51.2	70.6

The summary of FL modeling results is shown in Table 2. In the years where a dual-onset behavior was observed, the T_0' and k' columns capture the second correction parameters. Removing the anomaly

from the 2008 and 2012 vote date turns FL from a 50% win for Obama to a 56% loss in both elections. Furthermore, the model suggests that for 2020 the parameters $T_0 = 10^4$ and $k = 0.18$ result in an estimated total of 2.15 million votes switched from R to D. If true, this would represent a historic 71% landslide for Trump in FL for the 2020 election.

NC Anomaly Analysis

Like FL and GA, the anomalies in NC voting data begin in 2008. Using the log ratio plot, in Figure 26 we see data that support parameters of $T_0 = 10^4$ and $k = 0.1$, as in GA and FL for 2008.
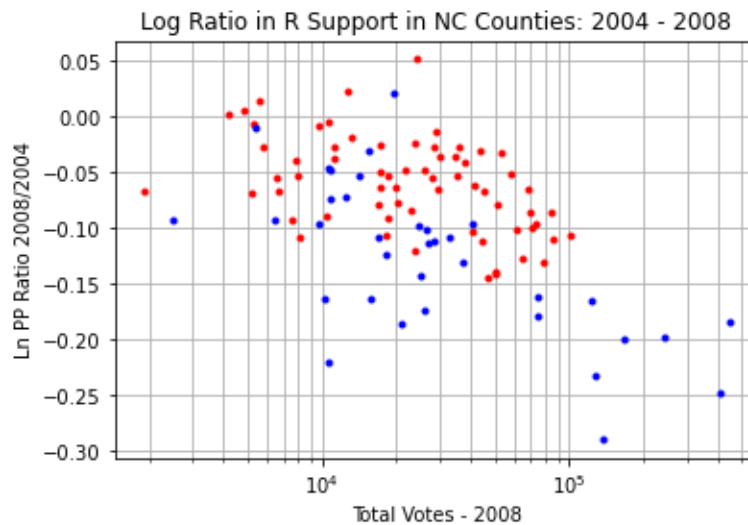
Figure 26. Log Ratio of Republican PP between 2004 and 2008 in NC

Figure 27 shows the adjusted data vs the original. The adjusted data show that Obama enjoyed a boost of 2-3 pp over Kerry in NC and up to 5 pp in blue counties.
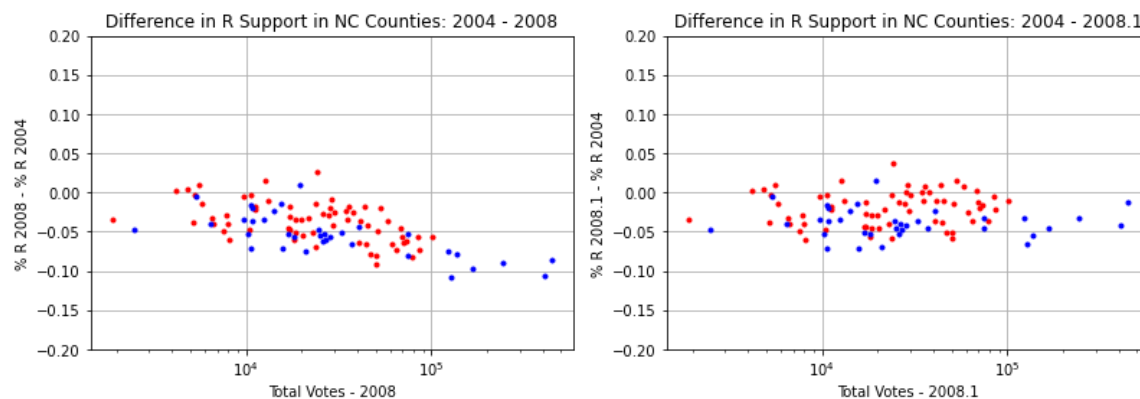
Figure 27. Original and Adjusted Difference in 2004 and 2008 Elections in NC

As in GA, the ratios between the 2008 and 2012 election were unremarkable. In 2016, the anomaly resurfaces. Figure 28 supports parameters of $T_0 = 10^4$ and k around 0.2. $k = 0.2$ seemed to supply too much correction to the slope, so an more modest $k = 0.175$ is assumed.
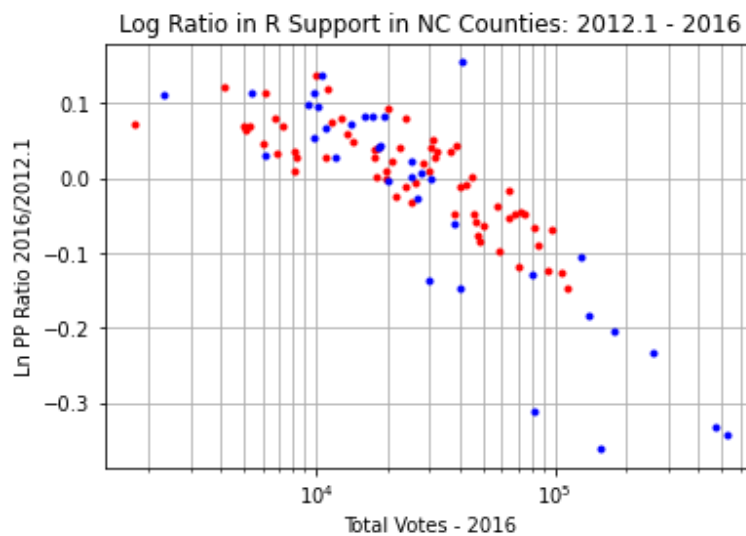

Figure 28. Log Ratio of Republican PP between 2012 (adjusted) and 2016 in NC

The original and adjusted data between these years are shown in Figure 29. The adjusted data show a 5 pp boost for Trump on average, except in several outlier blue areas.
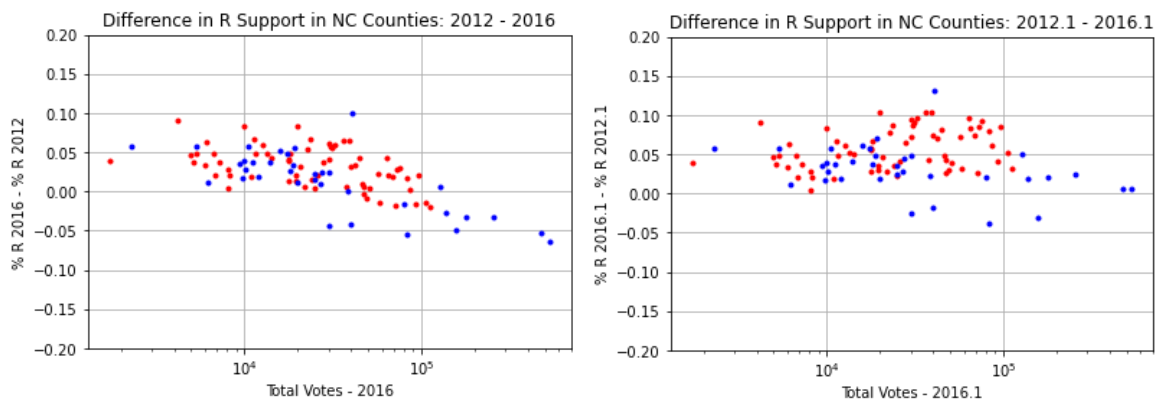
Figure 29. Original and Adjusted Difference in 2012 and 2016 Elections in NC

Figure 30 shows the log ratio plot between the adjust 2016 data and raw 2020 data. Since the anomaly is downward sloping from 2016 to 2020 in the original data, this support a slight increase in k. Figure 30 support an estimate of $k = 0.20$ and $T_0 = 10^4$.
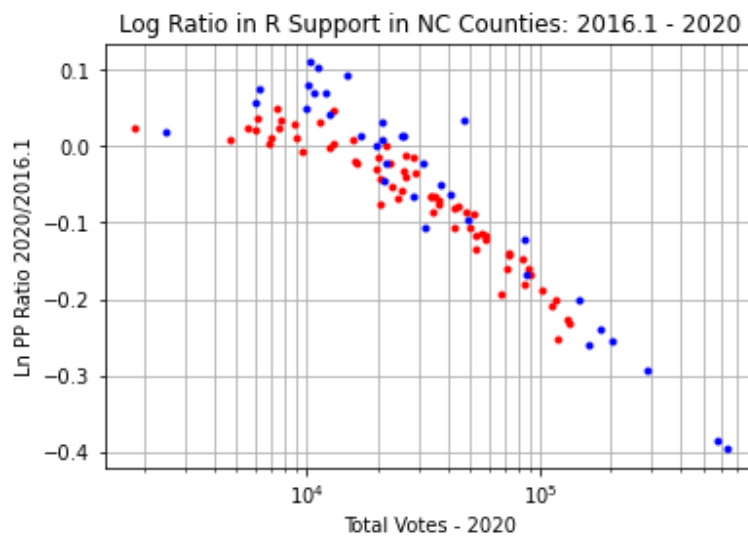

Figure 30. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in NC

Again, the data seem reasonable when corrected, as shown in Figure 31. This data suggests that Trump improved his lead in NC by a handful of pp in the 2020 election, which is consistent with the adjusted data in other states.
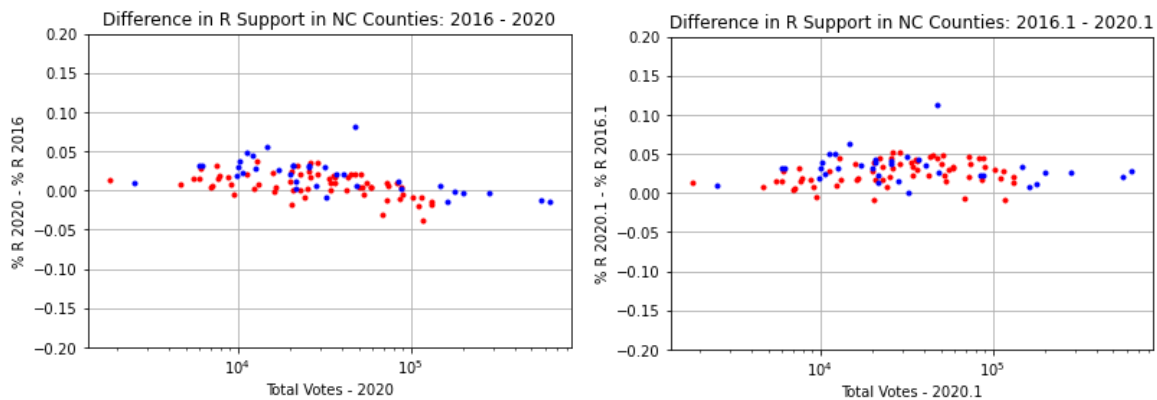
Figure 31. Original and Adjusted Difference in 2016 and 2020 Elections in NC

Once again, we will use the 2004 election as a "baseline" to sanity check the results. Figure 32 shows the log ratio metric plotted between these elections. The data show clear support for the parameters $T_0 = 10^4$ and $k = 0.2$.
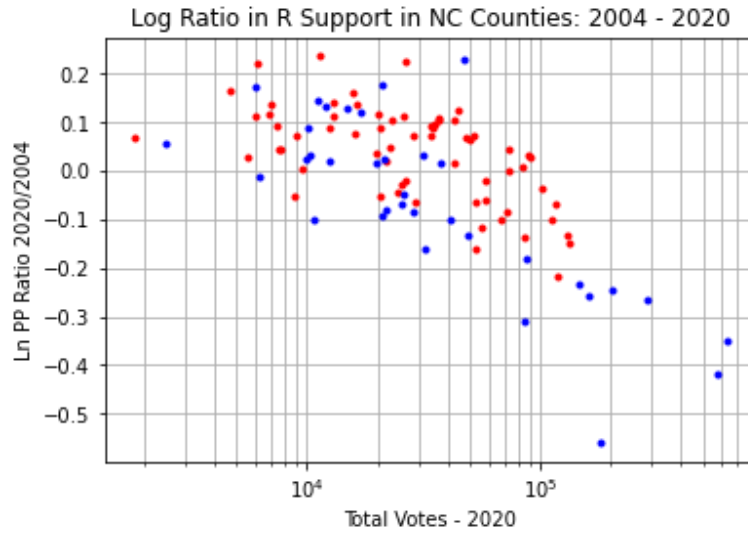
Figure 32. Log Ratio of Republican PP between 2004 and 2020 in NC

Similarly, the adjusted voting patterns in Figure 33 appear much more reasonable than the original. This data suggests that mid-sized counties in NC supported Trump significantly more than Bush.
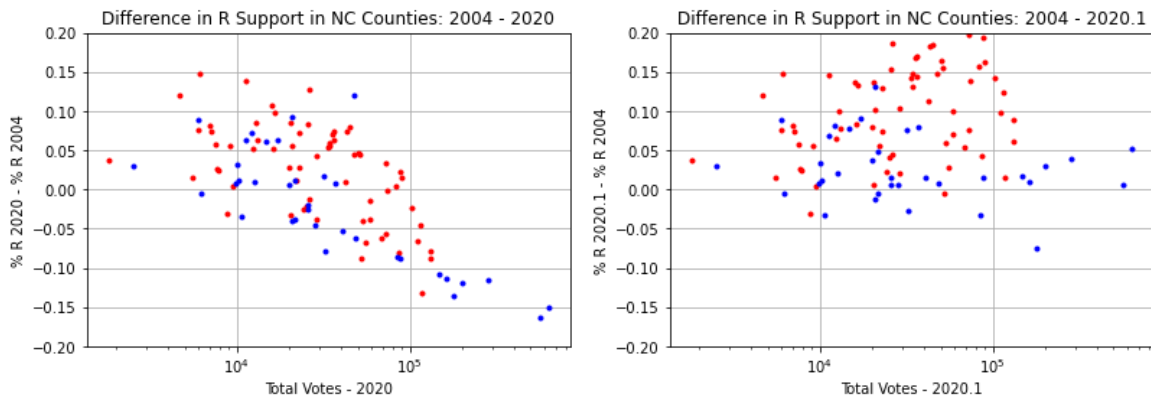

Figure 33. Original and Adjusted Difference in 2004 and 2020 Elections in NC

The overall summary of estimated parameters and their resulting impact is shown in Table 3. In particular, these corrections turn the 2008 win for Obama into a loss and show 60% support for Trump in his elections, rather than a close race.

Table 3. Summary of Modeled Adjustments to NC Election Data

Year	T_0	k	Total Changed Votes	Official R Vote Margin	Official R %	Predicted R %
2008	10k	0.1	205k	-13.7k	49.9	54.4
2012	10k	0.1	227k	97.5k	50.6	55.6
2016	10k	0.175	454k	173k	49.8	59.4
2020	10k	0.20	687k	74.5k	50.1	62.5

OH Anomaly Analysis

Unlike GA, FL, and NC, there is no apparent anomaly in 2008 and 2012 data for OH (See Figure 4). In this data set, the anomaly begins in 2016. Figure 34 supports parameters of $T_0 = 10^4$ and $k = 0.15$ for the anomaly model. The onset parameter is somewhat difficult to determine due to the small number of counties with less than 10k votes, so we will move forward conservatively with 10k.
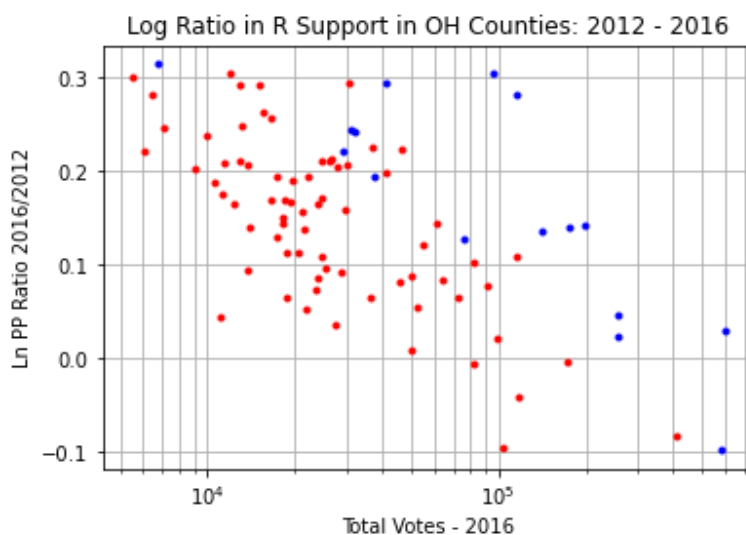

Figure 34. Log Ratio of Republican PP between 2012 and 2016 in OH

The corrected data in Figure 35 show a 15 pp swing toward Trump in OH vs Romney. In this report, the author has deliberately fixed the axes of these plots in order to provide a fair comparison, but the data here begin to run off the plot.
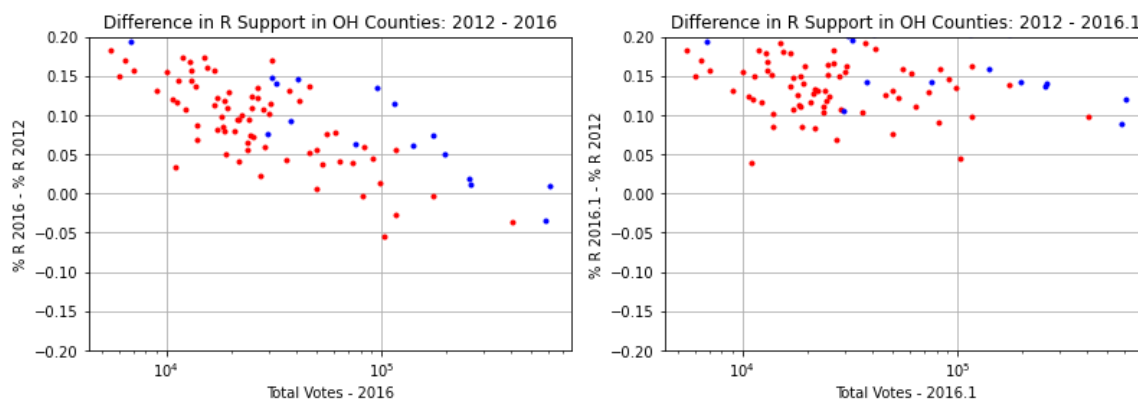

Figure 35. Original and Adjusted Difference in 2012 and 2016 Elections in OH

In 2016, there is an additional downward trend to the anomaly, supporting an increase in k. Figure 36 appears to support $k = 0.20$ for this election, with $T_0 = 10^4$ remaining the same.
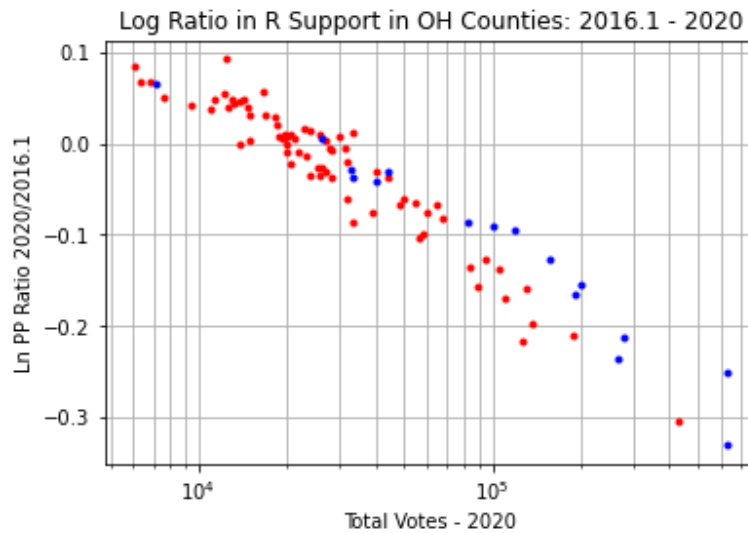

Figure 36. Log Ratio of Republican PP between 2016 and 2020 in OH

Figure 37 shows the original and corrected vote data. It could be argued that the corrected data now slope up, so that k is too high. Referring back to Figure 36, there appear to be a separate cluster of blue counties offset from the red ones which show extra gain. This would support the appearance of the graph as a separate cluster of increased support rather than an upwards slope. Conservatively, $k = 0.175$ could be assumed, but we will leave it at 0.2 to move forward with the analysis.
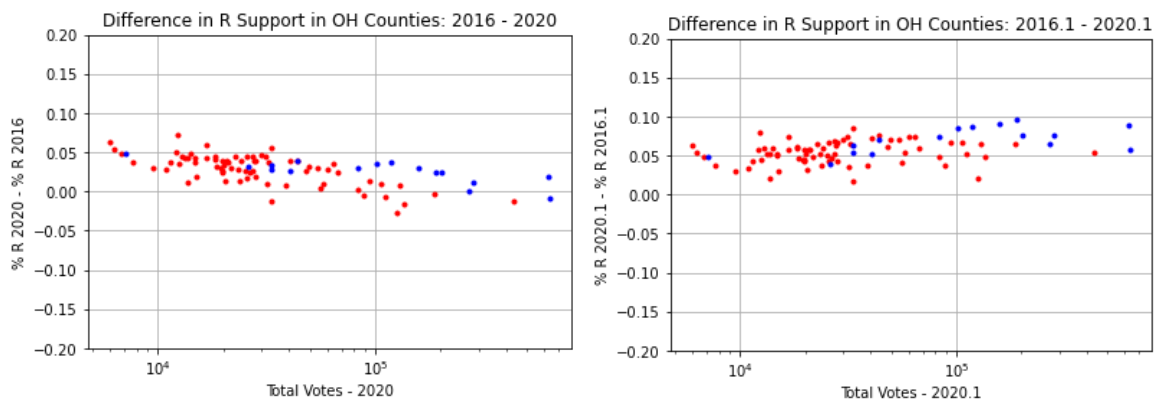

Figure 37. Original and Adjusted Difference in 2016 and 2020 Elections in OH

As in the other states, we compare the final adjustments to an unadjusted election as a sanity check. Figure 38 shows that, in comparison to 2004, the slope of $k = 0.2$ for the proposed 2020 adjustment is reasonable. It also supports the existence of the offset blue cluster in the larger counties.
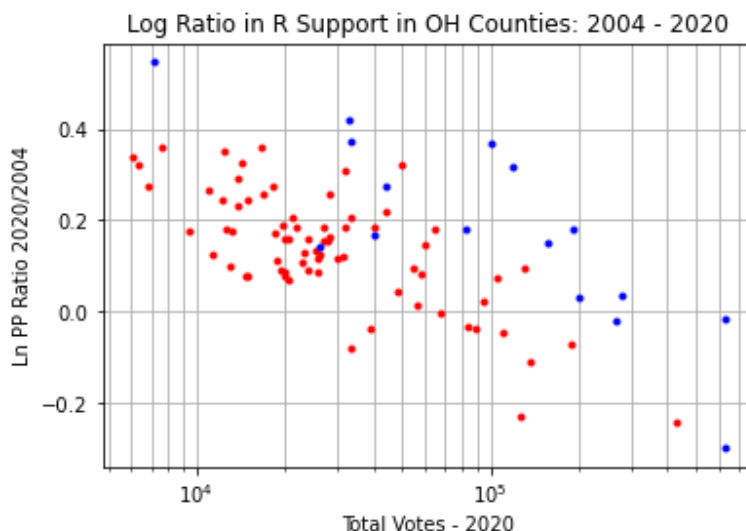

Figure 38. Log Ratio of Republican PP between 2004 and 2020 in OH

Figure 39 shows the original and corrected data for these two elections. The chosen parameters are well supported by the resulting data. Overall, OH appears to have shifted 20 pp to Trump from Bush.
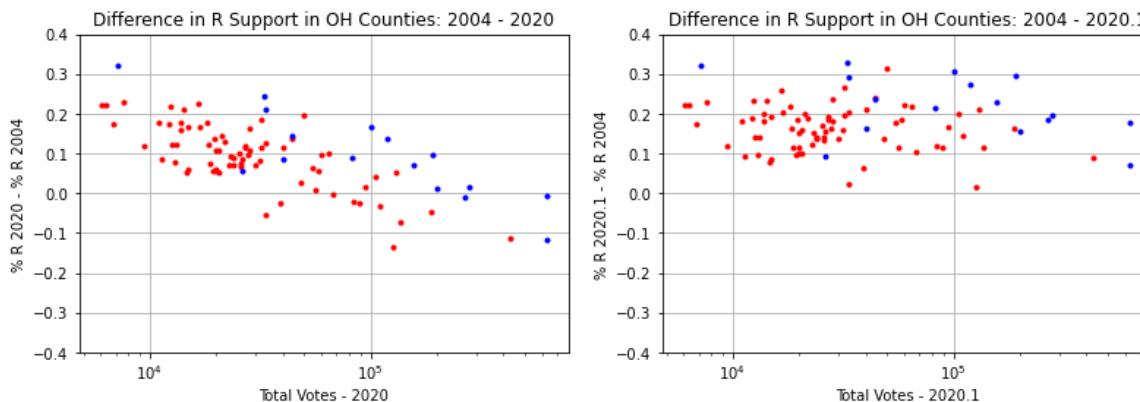

Figure 39. Original and Adjusted Difference in 2016 and 2020 Elections in OH

Finally, Table 4 summarizes the overall effect of the corrections on the data. One sees that even if a more conservative k was assumed in 2020, the result would still be a Trump landslide.

Table 4. Summary of Modeled Adjustments to OH Election Data

Year	T_0	k	Total Changed Votes	Official R Vote Margin	Official R %	Predicted R %
2016	10k	0.15	521k	447k	51.3	60.7
2020	10k	0.20	864k	476k	53.3	67.9

PA Anomaly Analysis

Caveat: The 2020 data analyzed for PA do not include the mail in votes proscribed by the US Supreme Court, per the SOS website. The author used the SOS-published data as of approximately 11/25/2020.

As in OH, PA does not show any anomaly before the 2016 election. The ratio analysis for 2016 compared to 2012 is shown in Figure 40. As in other states, the onset parameter appears to be $T_0 = 10^4$. However, the relative lack of data clusters crossing this threshold make determining the slope more difficult. The other data appear to suggest a slope of around $k = 0.20$.
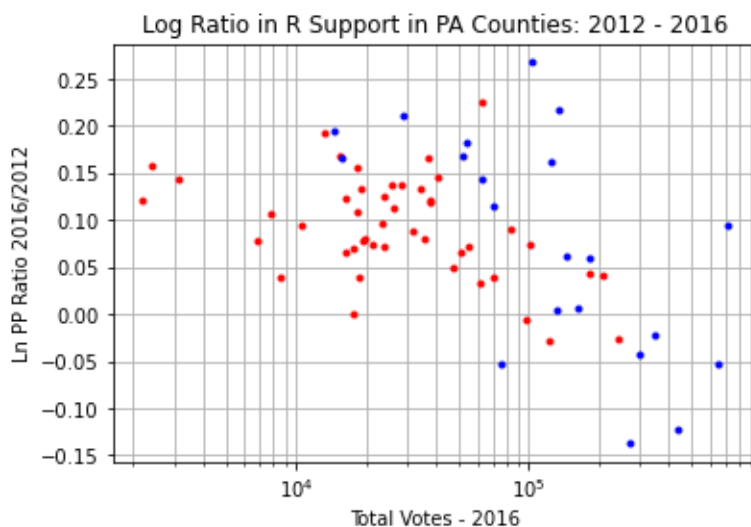

Figure 40. Log Ratio of Republican PP between 2012 and 2016 in PA

Figure 41 shows the result of applying these corrections compared to the original data. The chosen parameters appear reasonable and show a shift of around 5-7 pp toward Trump from Romney.
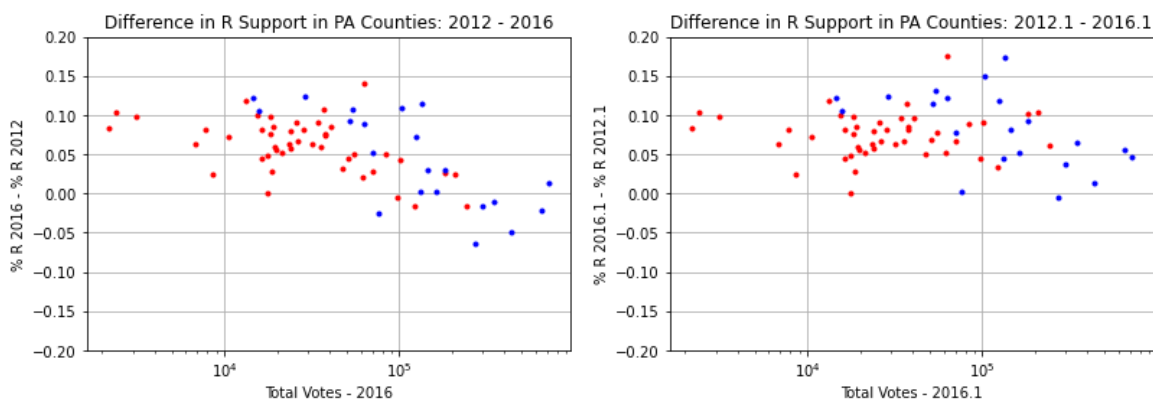

Figure 41. Original and Adjusted Difference in 2012 and 2016 Elections in PA

For 2020, the anomaly appears as a very slight downward slope. This suggests an incremental correction to k. Figure 42 shows the log ratio, but the parameters are difficult to determine from this. The parameters $T_0 = 10^4$ and $k = 0.225$ are assumed.
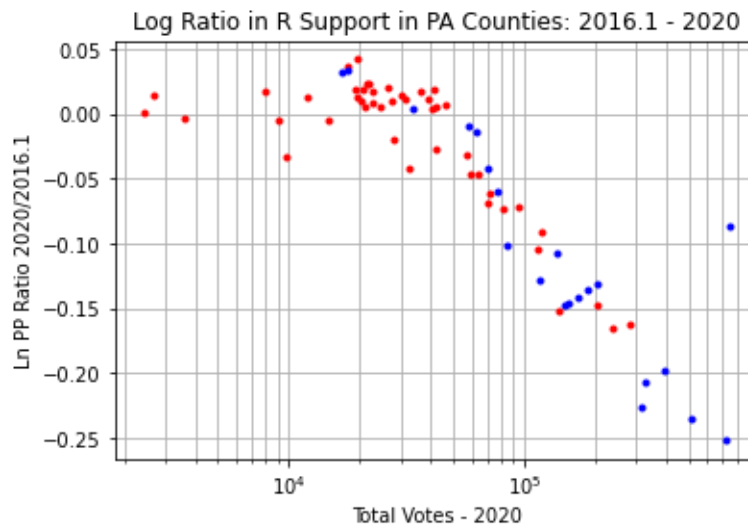

Figure 42. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in PA

Figure 43 shows that these corrections produce a trendless data set with low variance, as might be expected in an incumbent election.
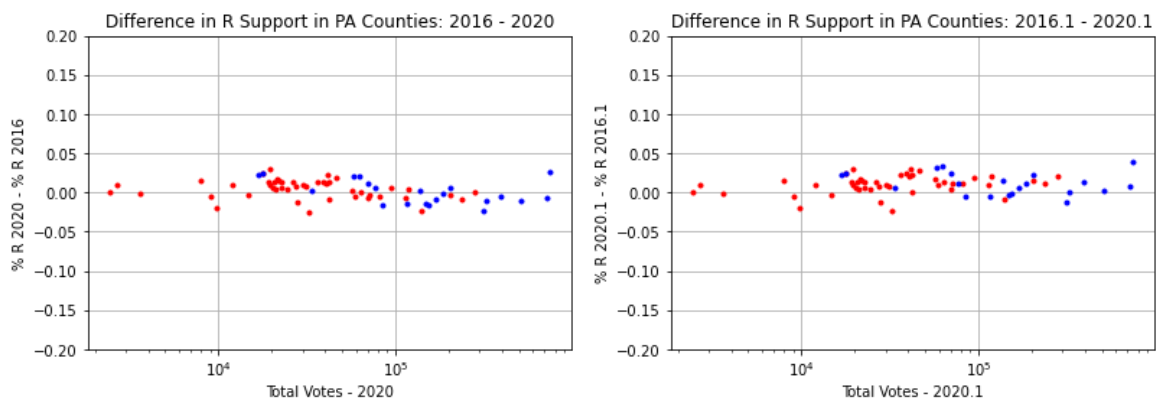

Figure 43. Original and Adjusted Difference in 2016 and 2020 Elections in PA

Again, we perform an sanity check by comparing to the unadjusted 2004 election as a baseline. Figure 44 shows that a slope of $k = 0.225$ is supported by the data (the lower envelope is useful for this purpose).
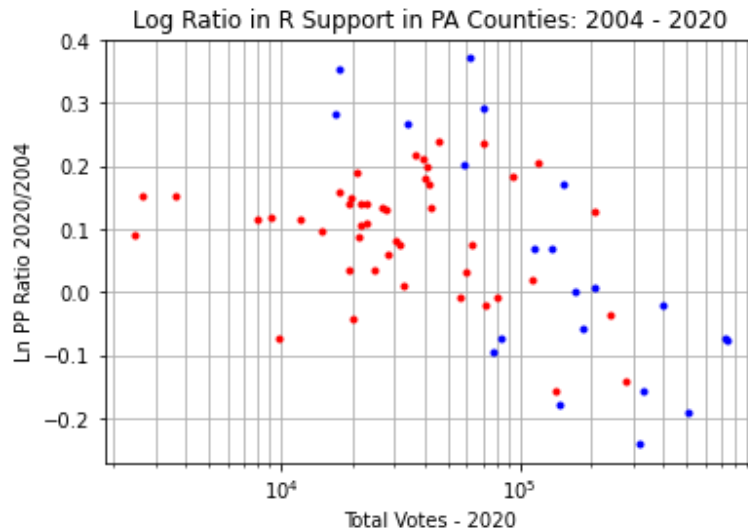

Figure 44. Log Ratio of Republican PP between 2004 and 2020 in PA

Additionally, the adjusted data set shown in Figure 45 appears much more reasonable than the original.
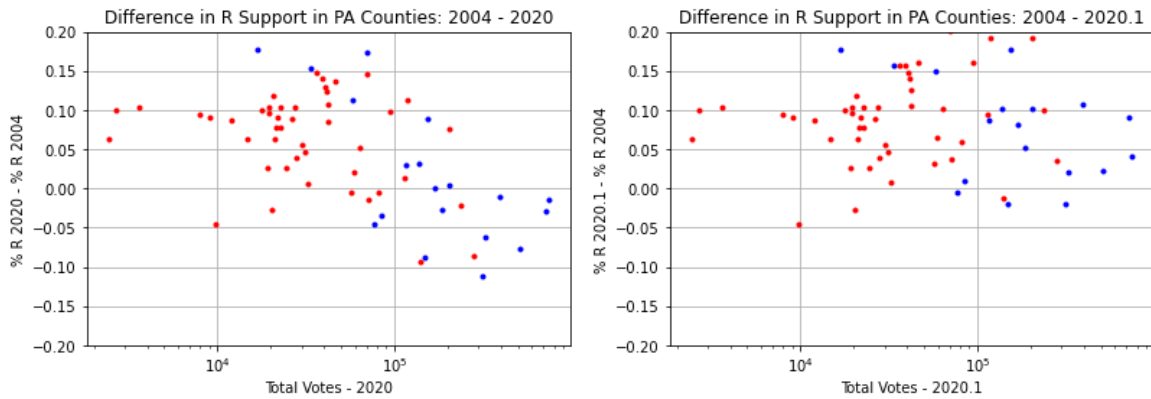

Figure 45. Original and Adjusted Difference in 2004 and 2020 Elections in PA

Table 5 summarizes the overall effect of these corrections. The proposed corrections turn the 2020 election into a clear win for Trump with 56% of the vote.

Table 5. Summary of Modeled Adjustments to PA Election Data

Year	T_0	k	Total Changed Votes	Official R Vote Margin	Official R %	Predicted R %
2016	10k	0.20	389k	47.3k	48.2	54.5
2020	10k	0.225	532k	-81.7k	48.8	56.5

VA Anomaly Analysis

Caveat: VA is a somewhat unusual state because some cities are treated like counties for the purposes of election statistics. These cities can pop in and out of existence as legal entities, and there were 2-3 cases of this from 2000-2020. The author's solution was simply to remove these cities from all data sets. Some extreme outliers appear in the VA data in several years (outside plot axes) and may be due to these regroupings of voters.

As in OH and PA, VA did not exhibit anomalous data previous to 2016. Figure 46 shows the log ratio analysis for VA 2012 vs 2016. The data support the parameters $T_0 = 10^4$ and $k = 0.1$ (at least).
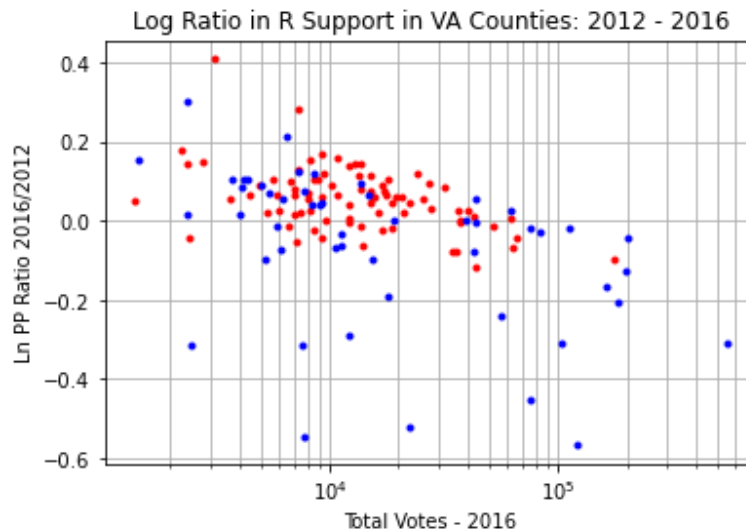

Figure 46. Log Ratio of Republican PP between 2012 and 2020 in VA

Figure 47 shows that these corrections effectively remove the trend line compared to the original data and result in overall reasonable data. The data show VA as a highly divided state with widely varying opinions of Trump.
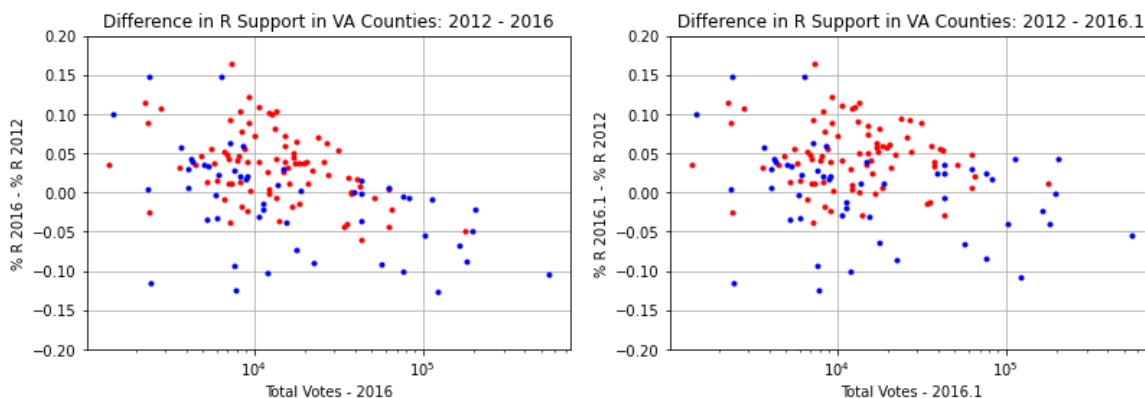

Figure 47. Original and Adjusted Difference in 2016 and 2020 Elections in VA

In 2020, the increase in the anomaly appears slightly stronger than in other states. Figure 48 support an increase of the slope to $k = 0.15$.
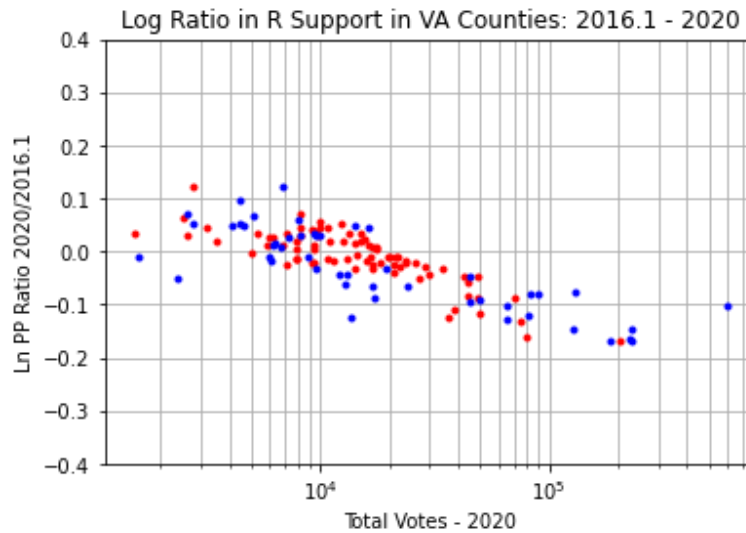

Figure 48. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in VA

Figure 49 shows the results of applying this correction, which effectively removes the trend. Compared to the original data, Fairfax county (far right) has changed its behavior dramatically. The author's conjecture is that, rather than showing an increase for Trump by 5 pp in 2020, Fairfax county does not follow the trend of the overall anomaly in 2016, and thus the applied correction produces this outlier.
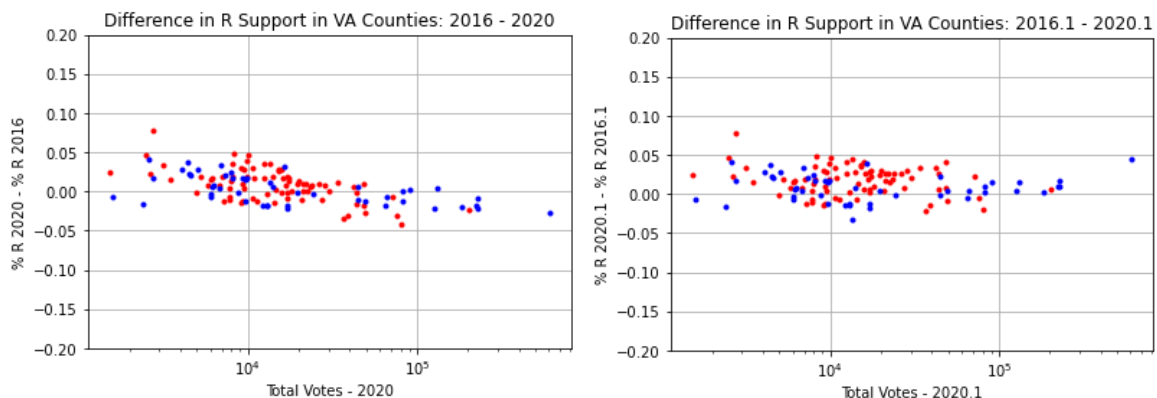

Figure 49. Original and Adjusted Difference in 2016 and 2020 Elections in VA

As in the other states, we compare the final results to an unadjusted election data set. This time, 2012 was used as the baseline election. Figure 50 shows that the estimated slope parameter of $k = 0.15$ is

supported by the data. (In fact, the parameter could be higher, but we leave it here for a more conservative analysis).
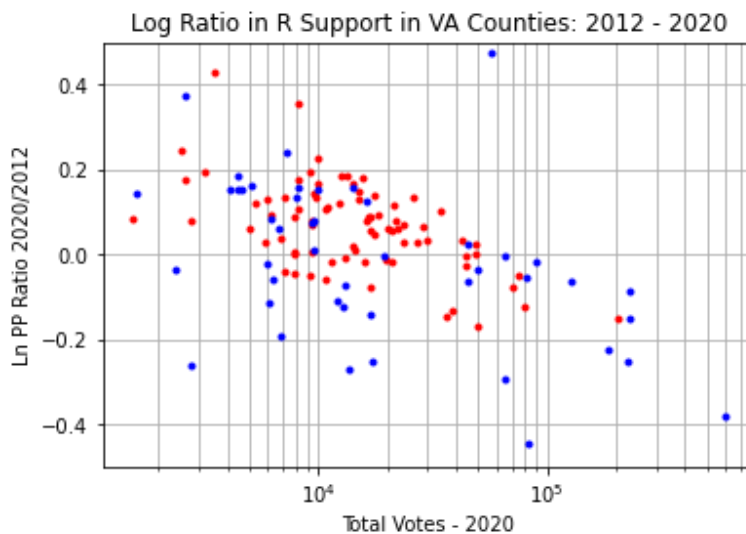

Figure 50. Log Ratio of Republican PP between 2012 and 2020 in VA

Figure 51 shows the results of applying this correction. Notably, Fairfax county does not appear as an extreme outlier in this data set, which tends to support the author's earlier conjecture. Figure 51 also seems to suggest that the slope parameter might be higher that the 0.15 estimated.
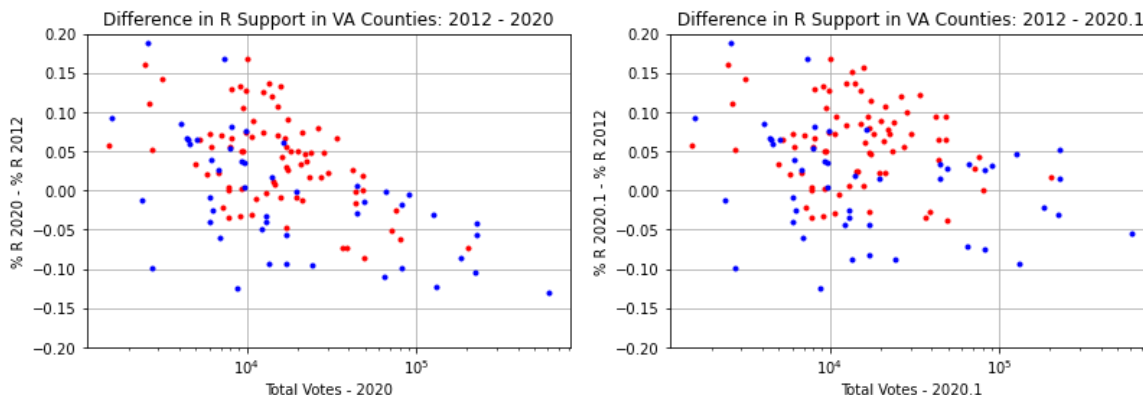

Figure 51. Original and Adjusted Difference in 2012 and 2020 Elections in FL

Table 6 shows the overall result of applying these corrections to the data. The corrected data show tossup elections for Trump in 2016 and 2020, rather than an overwhelming loss. The slope parameter estimates lack the precision to make a clear determination as to whether Trump might win under these corrections.

Table 6. Summary of Modeled Adjustments to VA Election Data

Year	T_0	k	Total Changed Votes	Official R Vote Margin	Official R %	Predicted R %

| 2016 | 10k | 0.10 | 131k | -212k | 44.4 | 47.7 |
| 2020 | 10k | 0.15 | 229k | -451k | 44.2 | 49.4 |

Benford Analysis and the Proposed Algorithm

Benford's law is a standard forensic statistical analysis applied in multiple fields, in particular forensic accounting. It is based on the principle that the leading digits of figures follow a predictable distribution under certain conditions. In particular

- The data span several orders of magnitude (in the chosen base)
- The data are approximately power law distributed

The second factor indicates that the data are approximately uniformly distributed on a log axis. In the many plots shown in the previous sections, it is evident that this is approximately true for the total number of voters by county in certain states and less true for others.

Due to the allegations surrounding the 2020 election, Benford analysis has enjoyed new popularity amongst those looking for anomalies. Opinions are divided as to how usefully the analysis applies to election data (frequently along partisan lines). The author will not attempt to address the body of these concerns here or attempt to highlight Benford anomalies in the preceding data sets.

However, the author makes the following conjecture: **adjustments of the sort described by the proposed algorithm will not have a material effect in a Benford analysis**. This conjecture is based on the fact that the corrections are logarithmically correlated to the original data, and Benford's law is based on a power law assumption. A proper verification of this notion would involve more extensive mathematical analysis which is beyond the scope of this paper. However, we can test this conjecture on a superficial level by simply applying Benford analysis to the original and adjusted data sets from this report and noting any glaring anomalies.

We begin by examining the data from the 2020 election in GA. A useful heuristic is that the data for vote totals over all counties should resemble Benford's law to a similar extent as the party vote totals for major candidates.

Figure 52 shows that the county total vote data resemble the general shape of Benford's law, with some noise at 2 and 6. Therefore, we may conclude that this analysis is not completely inapplicable.
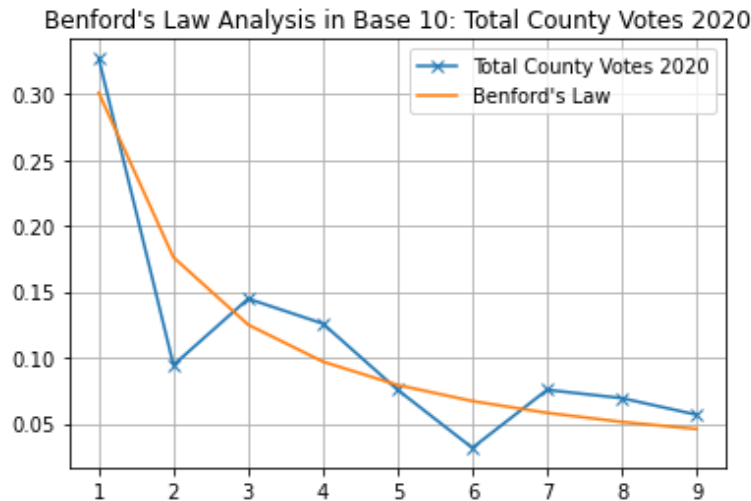
Figure 52. Leading Digit Distribution of Total County Votes in GA 2020

Figure 53 show the leading digit distribution of the R vote totals in this data set for both the original and adjusted data set. Both data sets resemble Benford's law to a similar extent as the totals data in Figure 52. Recall that Table 1 shows an adjustment of 610k votes from the original to the adjusted data set, more than 10% of the total votes in the election. This adjustment appears to have minimal effect on the Benford analysis.
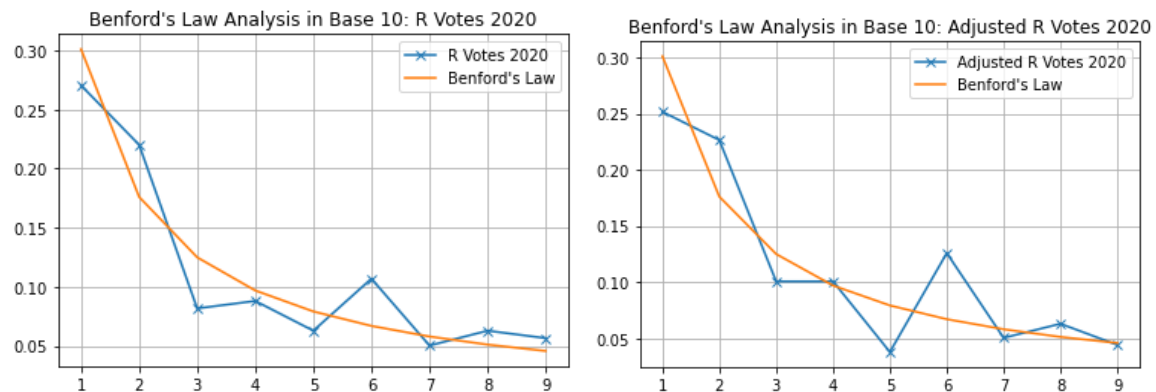

Figure 53. Leading Digit Distribution of Unadjusted and Adjusted R Votes in GA 2020

In contrast, Figure 54 shows an adjusted 2020 data set for which 4000 votes have simply been added (ham-handedly) to each of the 159 counties in GA. This data set shows a clear relative corruption of the distribution at digits 1,2 and 3 compared to Benford's law.
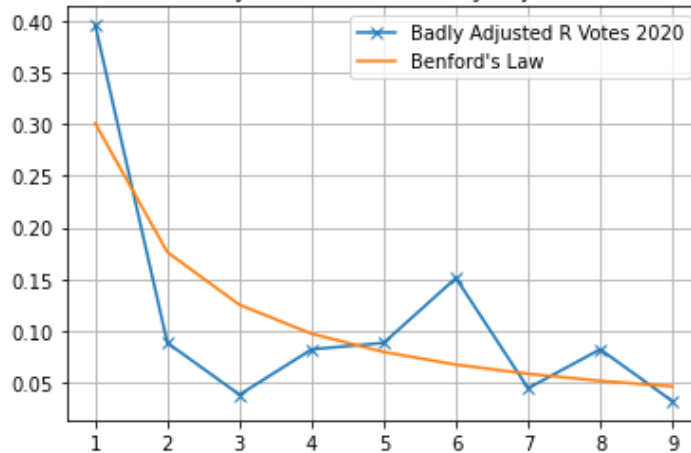
Figure 54. Leading Digit Distribution of Contrived Adjustment to R Votes in GA 2020
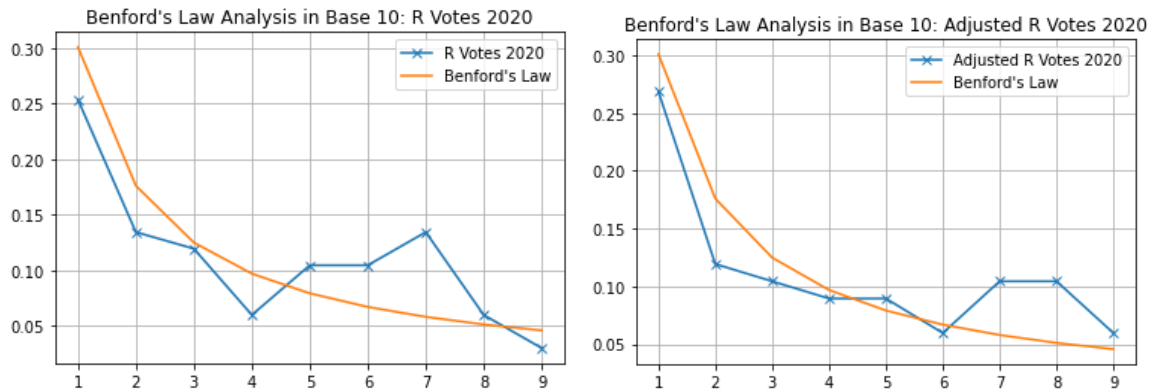

Figure 55. Leading Digit Distribution of Unadjusted and Adjusted R Votes in FL 2020

In order to check our assumption further, Figure 55 shows the leading digit distribution analysis for R votes in FL 2020 for both the original and unadjusted data. (FL has significantly fewer counties than GA, so more deviations from the law are expected as a result of the smaller sample size). It is seen that the adjustment of 2.1 million votes between the data sets, the largest in this report, does not produce highly anomalous results in the Benford analysis.

These data tend to support the author's conjecture that adjustments to the data according to the proposed algorithm are resilient to Benford's law analysis. To further establish this fact would require a separate study beyond the scope of this document.

Conclusions

Based on the analysis in this document, the author makes the following assessments

- The data in all 6 states under study exhibit a clear trend in differential voting patterns over multiple elections which is logarithmically correlated to the number of voters in a given county. For 3 states, the anomalies appear in 2008 and in 2016 for the remainder. (High confidence)
- The anomalies increase in (cumulative) slope over time, except for one partial data set in FL 2012 (High confidence)

- The anomalies exhibit a common onset point of 10,000 voters, except in FL where two onset points of distinct slope were observed in 2012 and 2016. (High confidence)
- The consistency of the structure, magnitude, and presence of the anomalies relative to expected voter data are consistent with an external adjustment of the data. (Moderate confidence)
- The anomalies are consistent with a simple piecewise algorithm of switched votes from R to D which is affine in the logarithm of total voters (Moderate confidence)
- The parameters for the algorithm can be readily estimated by appropriate data analysis as demonstrated in the report. The resulting corrected data appears to follow normal voting patterns (Moderate-low confidence)
- The adjustments based on the algorithm parameter estimation suggest significant changes in election results, including strong Trump wins in GA and PA in 2020 and a loss for Obama in FL and NC in 2008. (Moderate-low confidence)
- The proposed algorithm appears resilient to leading digit (Benford) analysis (Low confidence)

Based on these assessments, the author concludes that the data in these 6 states provide significant circumstantial evidentiary support to claims of widespread artificial vote manipulation in elections since 2008. The author encourages readers to compare this evidence to that available from a wide variety of other sources and methods to further assess these claims, conducting their own independent analysis as they are able.