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. These trends appear in data from some states as early as 2008 but are present in all six states from 2016 forward.

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.
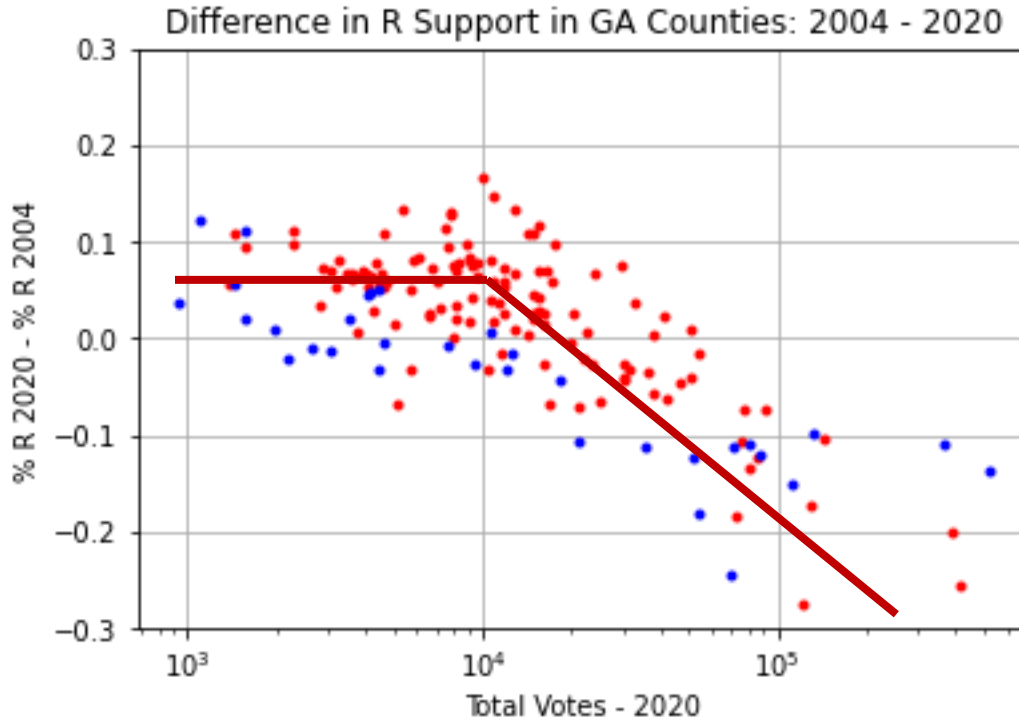
Figure 1. Differences in Republican presidential voting in GA counties, 2004-2020

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 22). **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.

This trend, in particular the piecewise nature of it, is highly indicative of an artificial process rather that natural demographic effects. In order to test the plausibility of it being associated with a vote switching algorithm (as is claimed to exist by some parties), the author proposes 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).
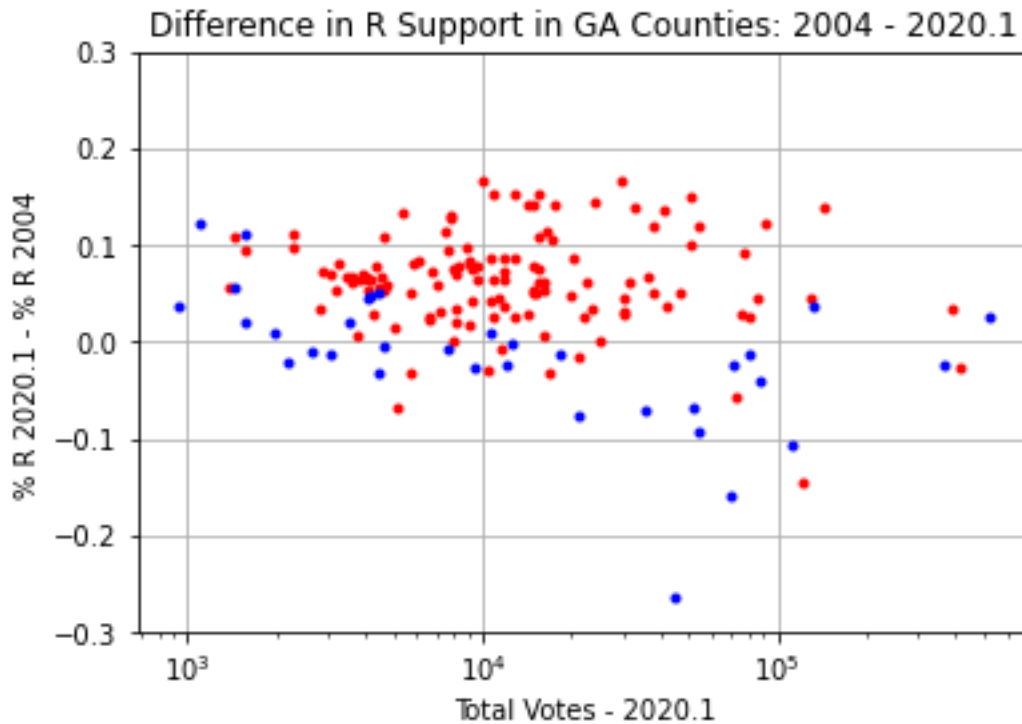

Figure 2. Adjusted ($k = 0.225$, $L_0 = 4$) differences in Republican voting in GA counties, 2004-2020

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 (moved from R to D).

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.
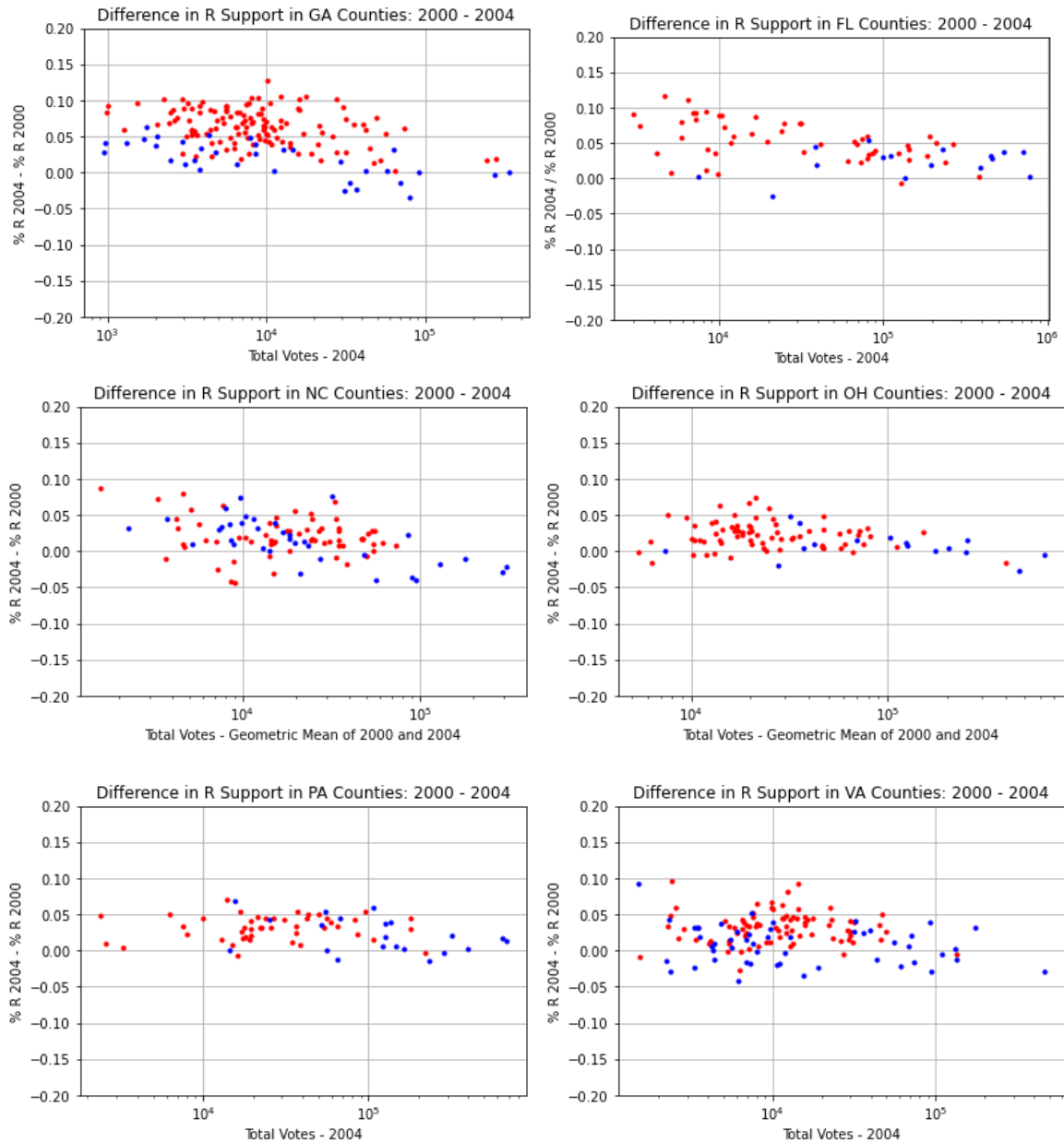
Figure 3. Voting Pattern Differences between 2000 and 2004

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 a constant slope in family with the other anomalies.

In the 2008 election, we also see several examples of patterns within the realm of expectation. Figure 4 shows the differences between 2004 and 2008 that are not indicative of an anomalous slope. 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.
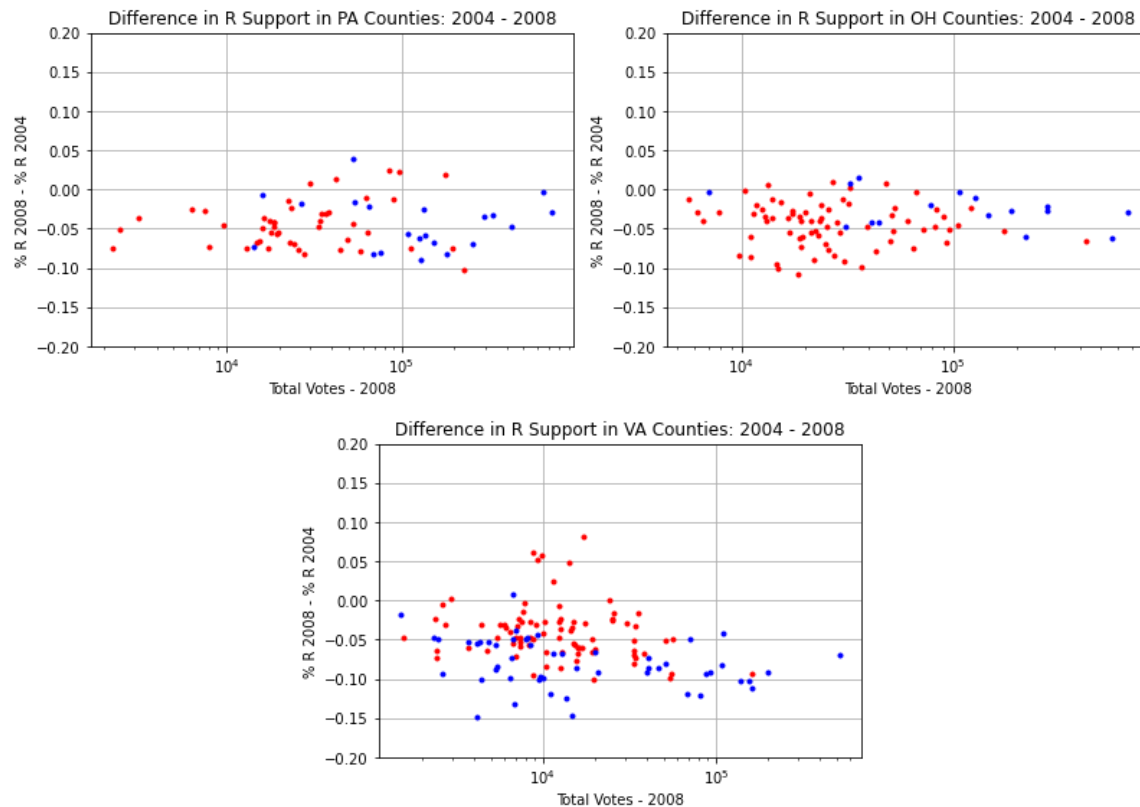

Figure 4. Unremarkable Voting Pattern Differences between 2004 and 2008

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 piecewise behavior in the differential percentage points with a transition point 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.
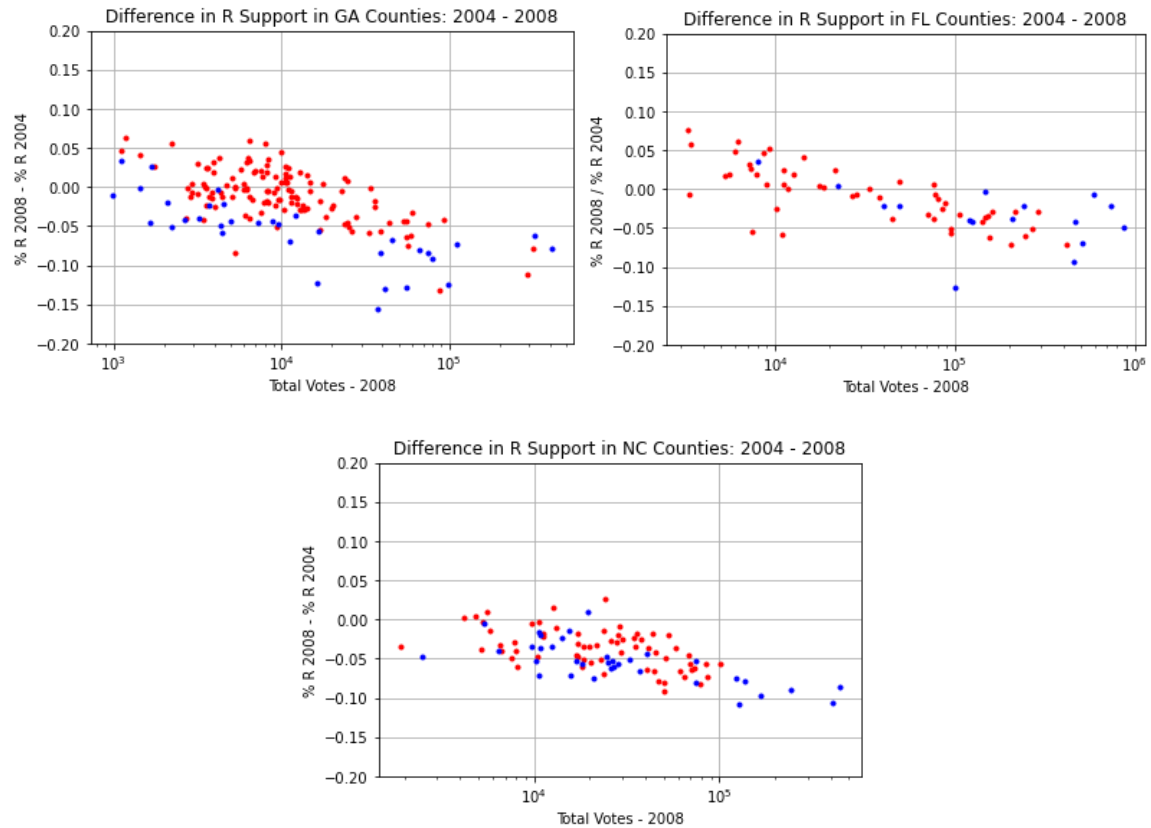
Figure 5. Anomalous Voting Pattern Differences between 2004 and 2008

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.
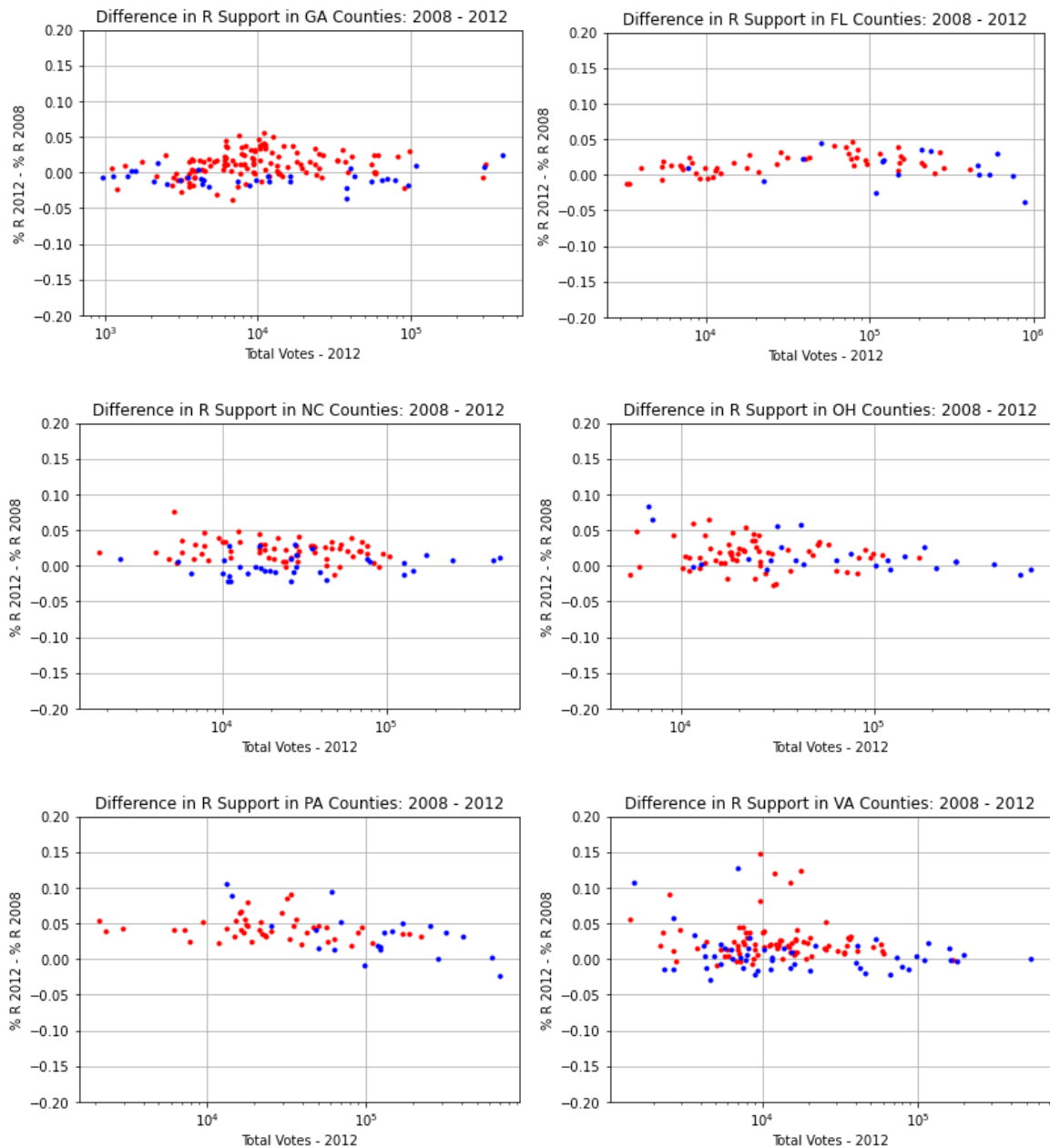
2012

Figure 6. Unremarkable Voting Pattern Differences between 2008 and 2012 (except FL)

Figure 6 shows the differential percentage statistics between the 2008 and 2012 elections. These data are within the bounds of expected distributions for an incumbent election with tight variance, except for FL, which shows a "hump" in the middle of the data. This "hump" will be further analyzed later in the report. (The data in PA could indicate a downward slope at the high end of the graph, but this effect is slight and has not been investigated in this report).
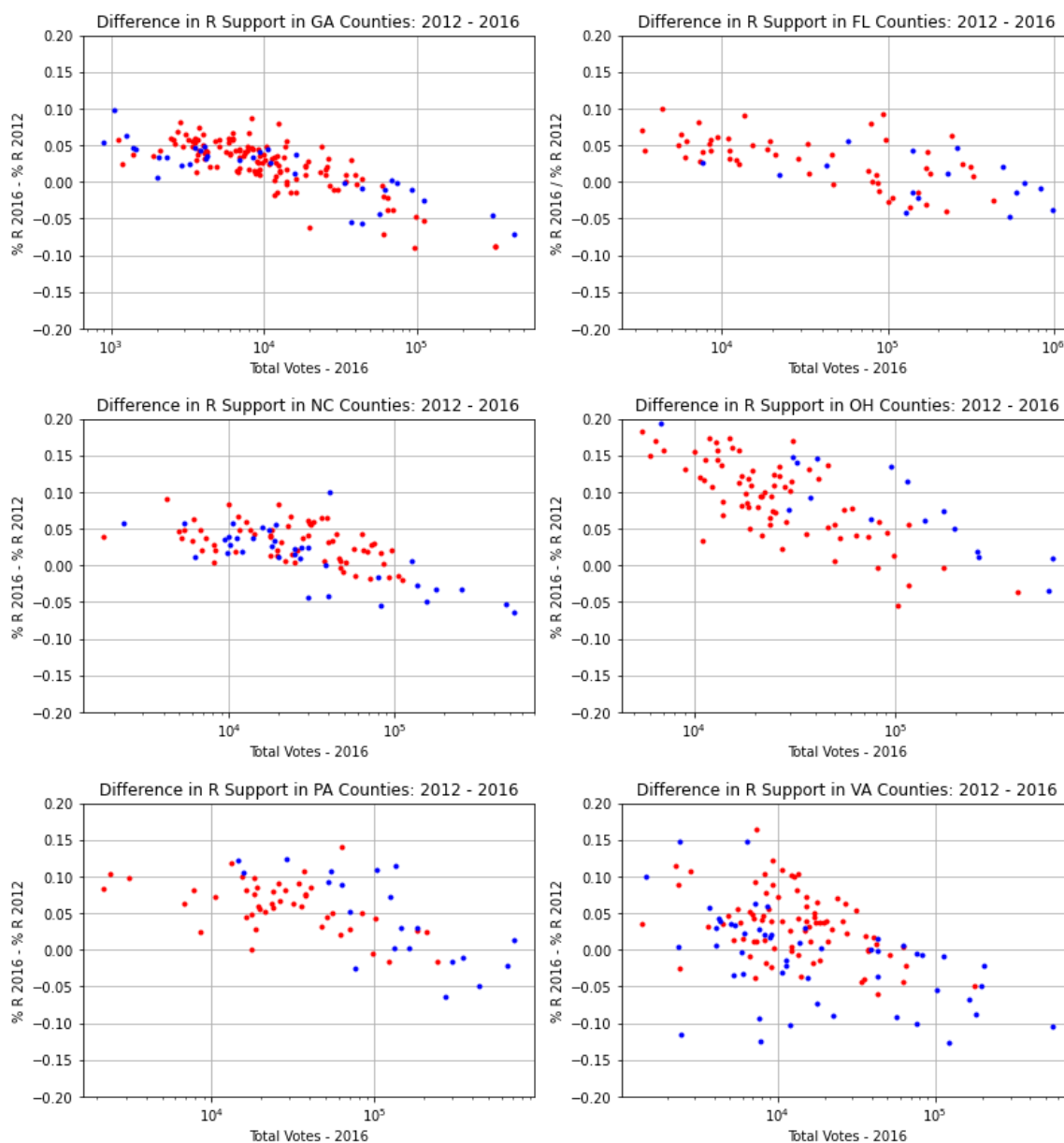
Figure 7. Anomalous Voting Pattern Differences between 2012 and 2016

Figure 7 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.
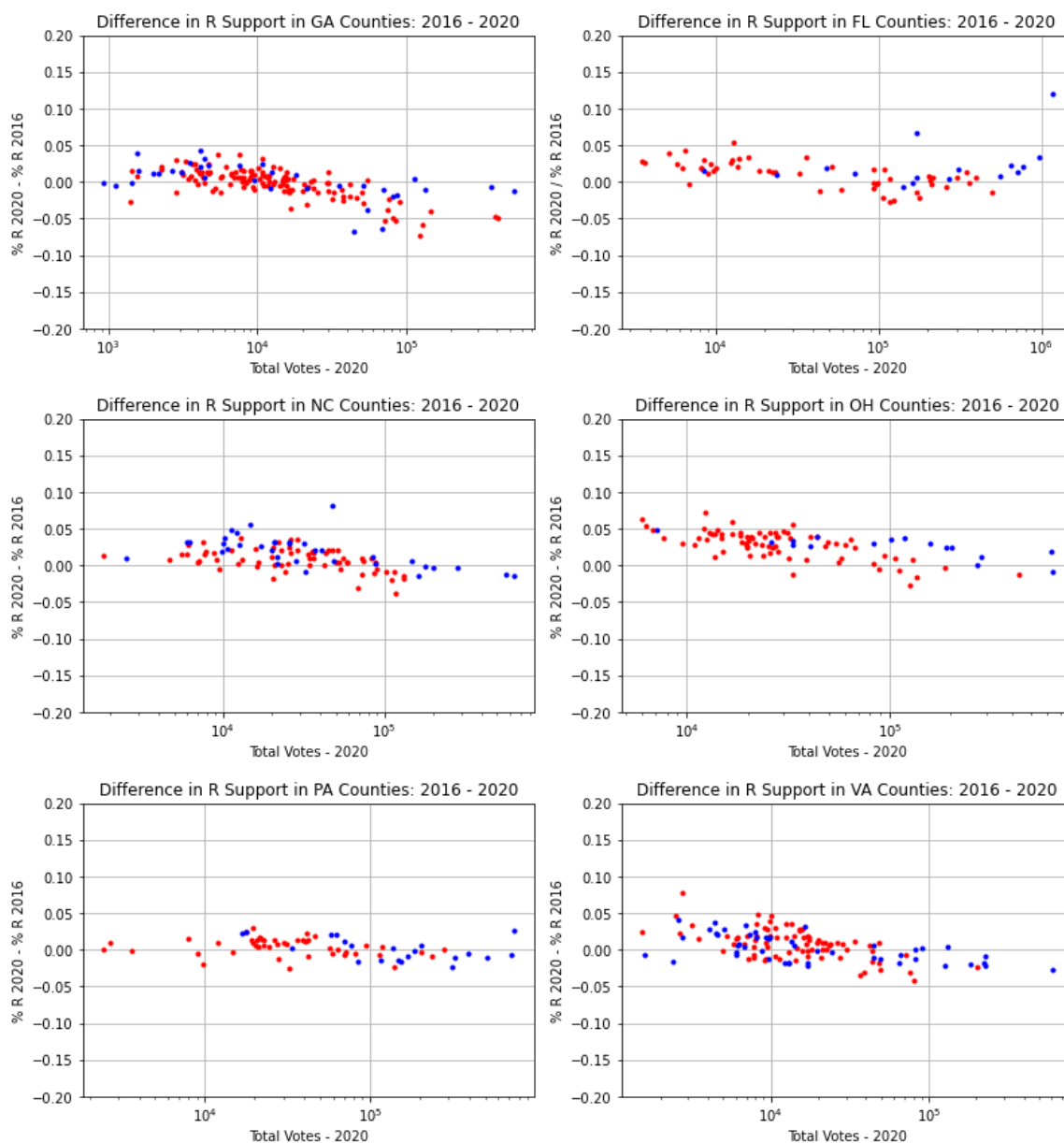
2020

Figure 8. Anomalous Voting Pattern Differences between 2016 and 2020

Figure 8 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).

Cumulative Effect (2004-2020)

In the previous subsections, we saw that the sloping trend in Republican presidential voting percentage points referred to as the "anomaly" begins in various states in either 2008 or 2016 in comparison to the corresponding previous presidential election. Figure 9 shows the overall effect of this sloping trend over time in each state. (The y-axis limits in the Figure 9 plots have been doubled compared to previous plots, and the x-axis limits have been fixed to be common from 1k to 1M votes per county. This allows an accurate side-by-side comparison.)
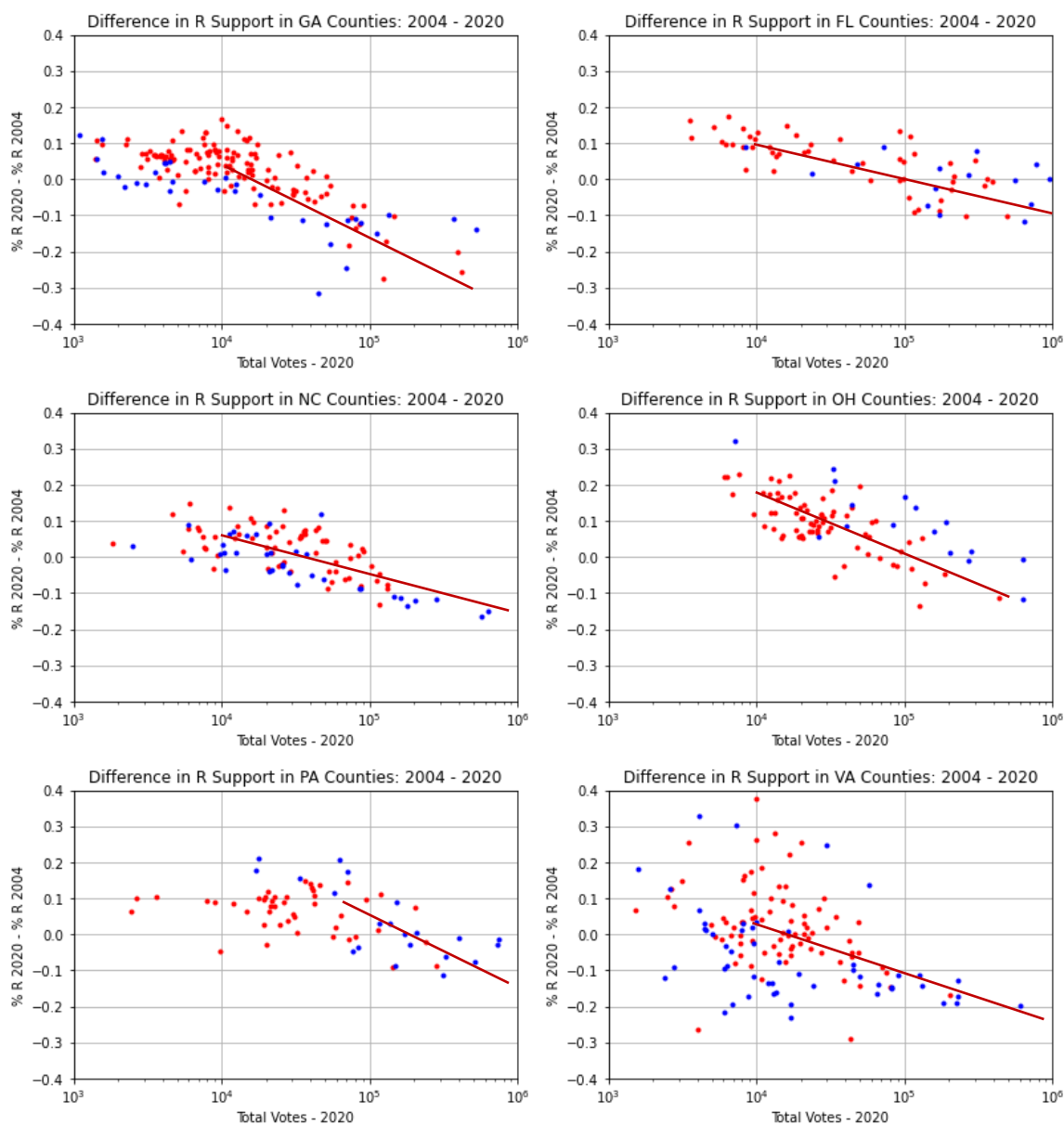

Figure 9. Anomalous Voting Pattern Differences between 2004 and 2020

Figure 9 shows that each state exhibits a downward sloping trend on the log scale. The slope of the trend is at least 0.1/dec and up to 0.2/dec. The exact slope is difficult to determine from these graphs. (A more precise analysis technique is developed later in the report.) The author has drawn lines on the

plots – these do not represent any formal numerical result (such as linear regression analysis) but merely a visual aid to the reader. In most states, this trend is stronger than the overall variability of the data (VA has an unusually large data variance for smaller counties). In GA, the trend clearly exhibits a piecewise nature beginning at 10k votes per county. In other states, the relative lack of counties with total votes less than 10k makes this piecewise nature less clear, but in NC and VA, this appears to be the case as well.

As these states are quite diverse demographically, including states with distinct African American, Hispanic, and White blue-collar subpopulations, it seems highly unlikely that such similar trends could be rooted in common demographic shifts. Additionally, such shifts tend to manifest in clusters for counties where those subpopulations congregate, rather than a trend with consistent logarithmic slope over all counties. These facts, along with the strong piecewise nature of the trend in GA suggest a potential artificial cause.

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 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 more technical as it presents the derivation of the parameter fitting method. Less technical readers may wish to review the proposed equation for differences in votes and skip to the next section for a detailed description of the analysis method.

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 (slope discontinuity) 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 in county i.) Therefore (for counties larger than T_0)

$$\Delta R_i = \rho_i R_i$$

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 significant 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 discontinuity point of the anomaly should be clearly visible where the slope changes.** The slope should ideally be measured in within the first decade from the onset point, as otherwise the (neglected) higher order terms in the Taylor expansion may begin to change the slope.

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

GA - Detailed Demonstration of Analysis Methodology

This section demonstrates in detail the analysis methodology that is used to estimate the parameters k and T_0 for the vote swapping model

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

for each election and state. In the previous section, we found that the log of the ratio of R percentage points between two elections gives a function which is approximately linear in the log county size, i.e.

$$\ln \hat{r}_{12} = \ln \frac{\tilde{R}_2/T_1}{\tilde{R}_1/T_2} \approx \ln \frac{R_2}{R_1} \frac{T_1}{T_2} + (k_1 - k_2)\log_{10}\left(\frac{T_2}{T_0}\right) = (k_1 - k_2)\log_{10}(T_2) + C$$

The slope of the log ratio plot is the difference of the parameter k in each election.

The steps for estimating k and T_0 for a new election (and producing a corrected result) are as follows.

1. Begin with an election for which $k = 0$ is assumed. This may be a "baseline" election (previous to the assumed operation of the algorithm), or data from a previous election for which the effect of the algorithm has already been removed.
2. Form the log (reported) percentage point ratio between the new election (for which k is unknown) and the previous election ($k = 0$) for each county and plot this against the log10 total votes for that county. Then the slope of this plot gives an estimate of k, since

$$\ln \hat{r}_{12} \approx \ln \hat{r}_{12,true} + k \log_{10}\left(\frac{T_2}{T_0}\right)$$

3. Estimate T_0 by determining where the slope of the graph changes discontinuously.
4. Remove the effect of the algorithm in each county by subtracting ΔR_i D votes and adding ΔR_i R votes to/from the original data set and label this as a new (modified) data set.
5. Repeat the log ratio plot between the previous election and the modified version of the new election.
6. If the plot appears (nearly) flat and/or resembles an expected demographic pattern, then stop. Otherwise, adjust k and T_0 as necessary, and repeat 4-6.
7. Examine the differences in percentage points between the two elections for both original and adjusted data to ensure that the corrections are reasonable. If not, adjust k and T_0 as necessary, and repeat 4-7.
8. The modified data set may now be used as the comparison data set for the next election (since we assume $k \approx 0$ with the modification).
9. Finally, at the end of the sequence of elections, compare the **final** corrected election data to the **initial** "baseline" election as a sanity check (in both log ratio and differential percentage). This ensures that the final results are not corrupted by accumulated errors in the previous analyses (effectively "closing the loop").

This method will now be illustrated for the 2008, 2012, 2016, and 2020 elections in GA.

2008

Since there was no anomaly in the comparison between the 2000 and 2004 election, we take the 2004 election as a "baseline" ($k = 0$). We form the log ratio plot between 2004 and 2008. This plot is shown in Figure 10.

We now notice the advantage of the log ratio metric – red and blue counties are both brought into "family" with each other. Although they are offset by a nearly fixed amount on the y axis, they have a similar slope (and discontinuity in the slope). This represents a clearer picture for the purposes of estimating our parameters than the percentage point difference plots presented earlier.

Using Figure 10, we can see that there is an apparent discontinuity in the slope around 10k votes per county, so we set $T_0 = 10^4$. The slope of the trend on the graph beyond 10k votes appears to be 0.1/dec. This is true for both red and blue counties (although the blue data are more sparse and there are some significant outliers). Therefore, we estimate $k = 0.1$ for the 2008 election in GA.

Moving on to step 4, we now correct the data according to the proposed algorithm with these estimated parameters in each county. We label this data set "2008.1".
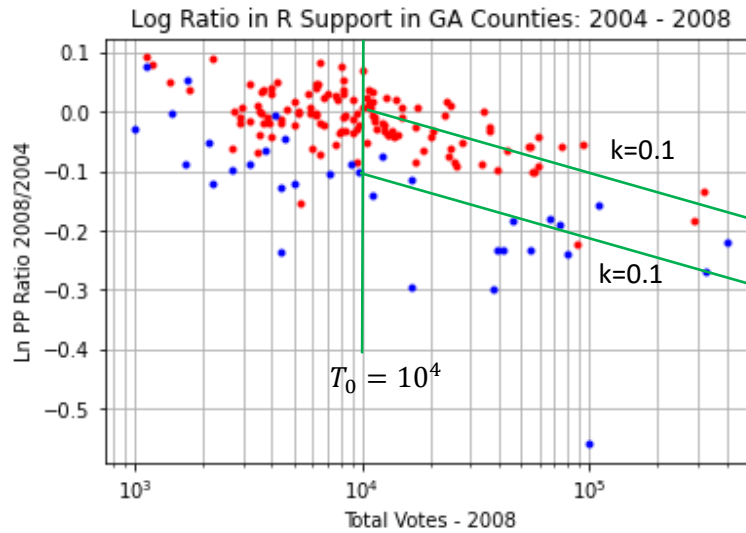
Figure 10. Log Ratio of Republican PP between 2004 and 2008 in GA

Now, performing Step 5, we replot the log ratio between 2004 and 2008.1 (the modified data). This is shown in Figure 11. We see that the data appear flat, so that the anomalous trend has been effectively removed.
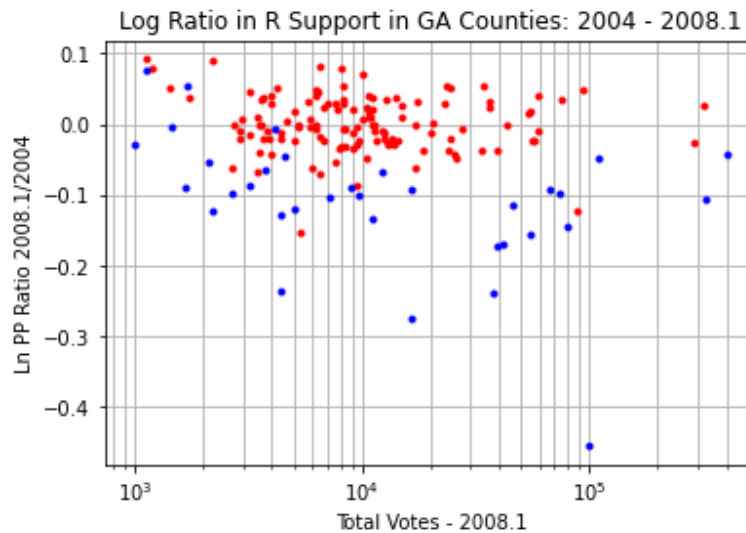
Figure 11. Adjusted Log Ratio of Republican PP between 2004 and 2008 in GA

We can now view the results of this adjustment in the original data space of percentage point differences. These are shown in Figure 12. The left plot is the original data, and the right plot is the adjusted data (2008.1). We see that the right-hand plot resembles more strongly an expected voting pattern for this election (compare to e.g. Figure 4 in other states). Blue counties moved 5 pp toward Obama compared to Kerry almost uniformly, and red counties retained a similar level of support for McCain as they did for Bush in 2004.
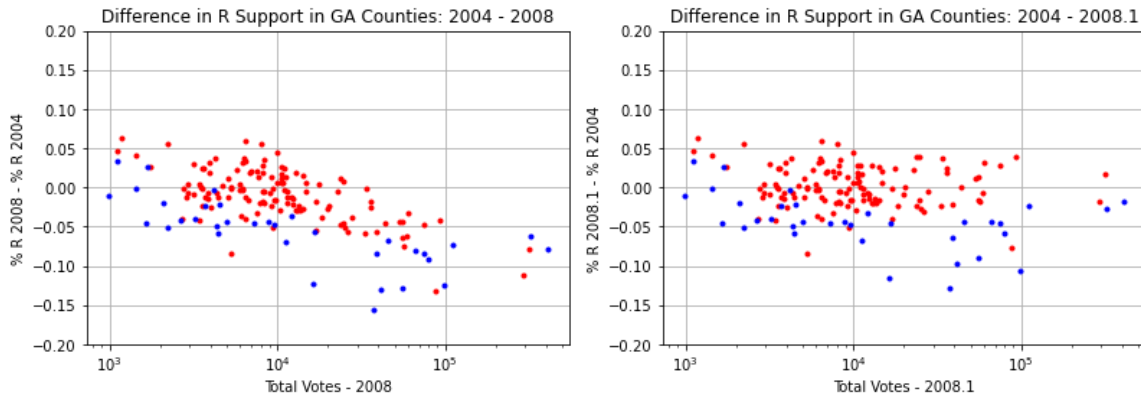
Figure 12. Original (left) and Adjusted (right) Differences in 2004 and 2008 Elections in GA

2012

We can now move forward in a couple of ways to analyze 2012. We could continue using 2004 as the baseline comparison election or use the corrected 2008.1 data. The latter approach will be illustrated here.
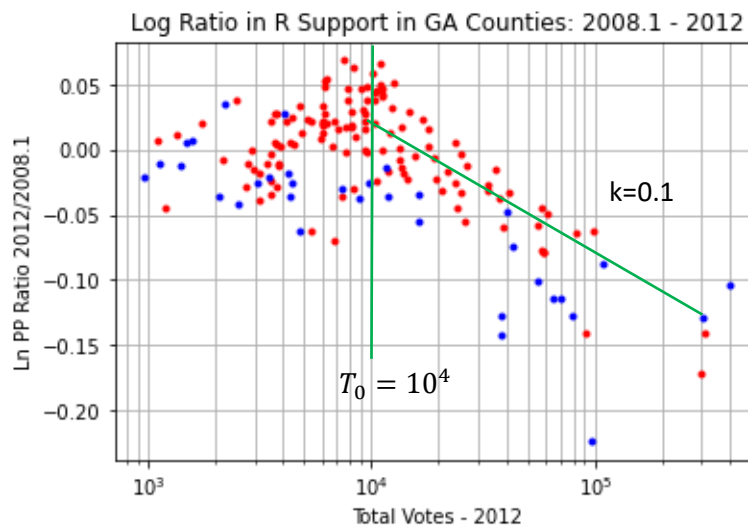
Figure 13. Log Ratio of Republican PP between 2008 (adjusted) and 2012 in GA

Figure 13 shows the log ratio between the adjusted 2008 data and 2012. The discontinuity of slope is again evident at 10k votes, so we keep $T_0 = 10^4$. The slope of the graph appears to be at least k=0.1. By referencing Figure 6, we see that the anomaly does not appear in the original differential percentage

data between 2008 and 2012, so we do not expect any additional action of the algorithm in this election beyond $k = 0.1$.

This is confirmed by "closing the loop" by applying the $k = 0.1$ and $T_0 = 10^4$ correction to the 2012 data to form the 2012.1 data set (Step 4) and replotting the log ratio (Step 5). The updated log ratio is shown in Figure 14. The data do not appear to have an anomalous slope, so we may assume these choices of parameters are approximately correct.
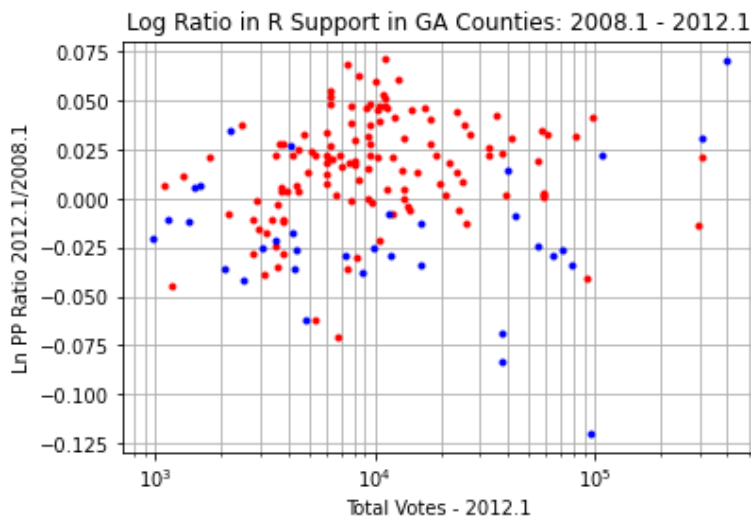

Figure 14. Adjusted Log Ratio of Republican PP between 2008 and 2012 in GA

Finally, the adjusted differences in percentage points can be viewed in comparison to the original data. This is shown in Figure 15. The charts are virtually identical and show no visible sloping anomaly. This was an incumbent election for Obama, and the data exhibit a characteristic low variance centered nearly around zero. It appears Obama lost some support in mid-sized red counties, but otherwise performed similarly to 2008.
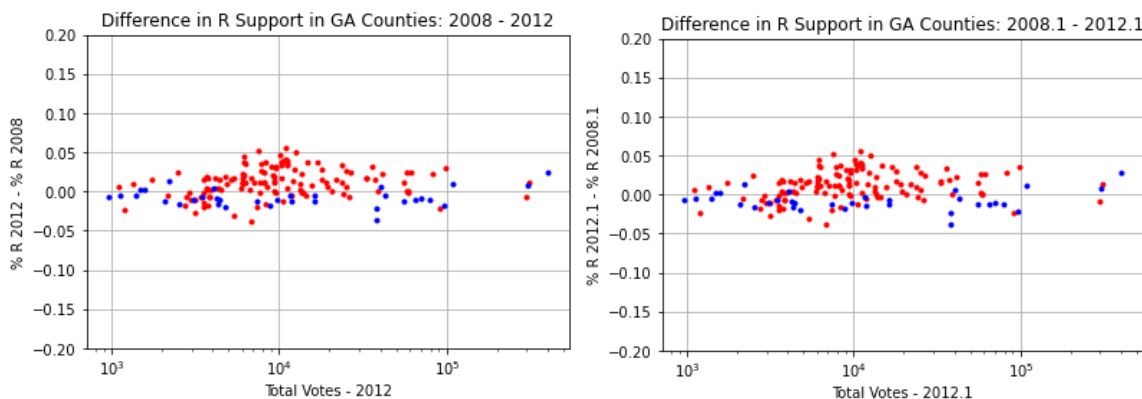
Figure 15. Original (left) and Adjusted (right) Differences in 2008 and 2012 Elections in GA

2016

We repeat the same steps again for 2016, using 2012 (adjusted data) as the baseline election. Figure 16 shows the log ratio plot between these elections. Here we see the slope of the graph has increases significantly to around k=0.2. The discontinuity in the slope remains at $T_0 = 10^4$.
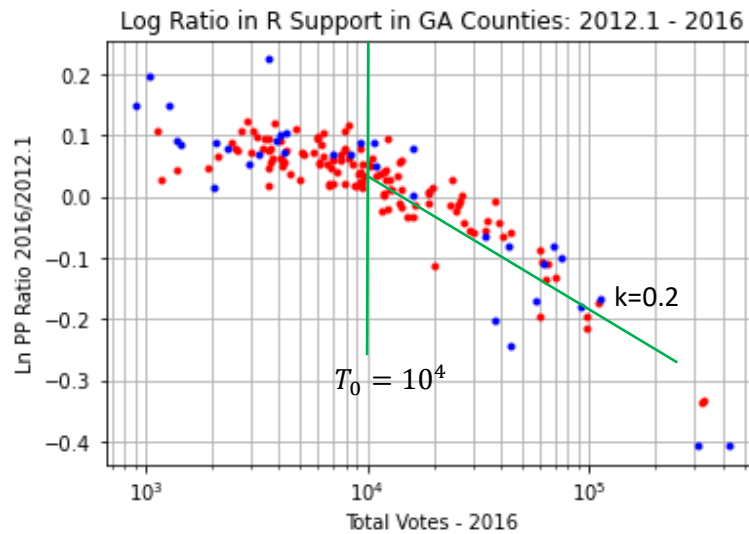
Figure 16. Log Ratio of Republican PP between 2012 (adjusted) and 2016 in GA

Again, we apply these parameters to correct the data, forming data set 2016.1, and replot the log ratio. The adjusted log ratio is shown in Figure 17. The data appear flat and without discontinuity in slope, indicating a good parameter match.
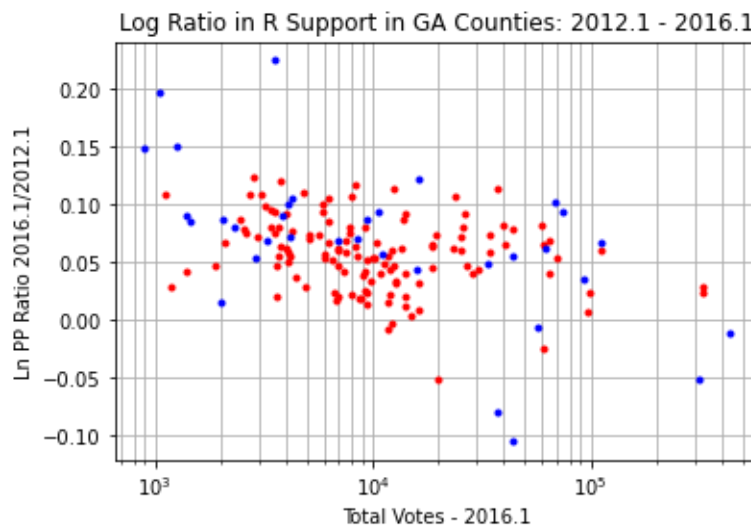
Figure 17. Adjusted Log Ratio of Republican PP between 2012 and 2016 in GA

Figure 18 shows the original and adjusted differential data between these elections based on this parameter fit. The adjusted data on the right resemble an expected voting pattern in which Trump received on average 5pp more support throughout GA than Romney, except in Metro Atlanta where his support was similar. There is no apparent anomaly in the data.
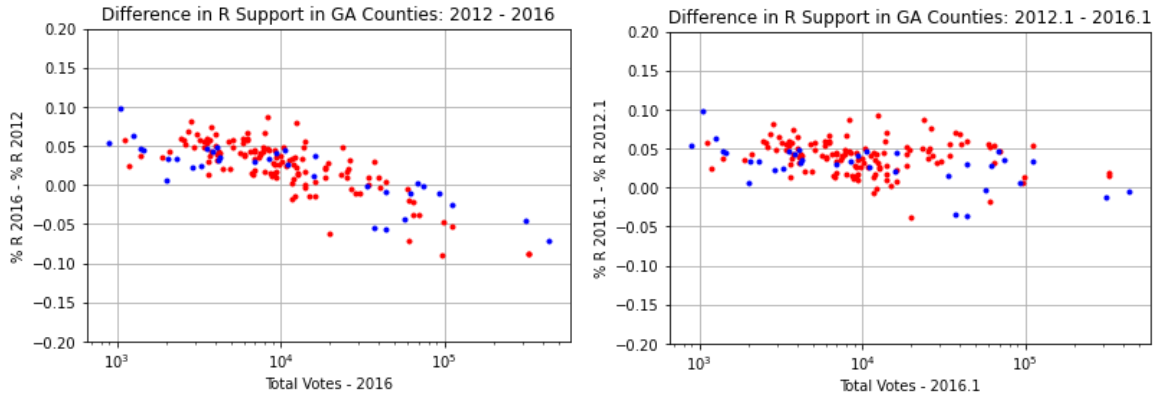
Figure 18. Original (left) and Adjusted (right) Differences in 2012 and 2016 Elections in GA

2020

Finally, the 2020 election is compared to the adjusted 2016 data to determine the new parameters. Figure 19 shows the log ratio for these two data sets. A slope of at least k=0.225 is apparent, perhaps even as high as k=0.25. We use the former figure for a more conservative analysis. The slope discontinuity point remains at $T_0 = 10^4$.
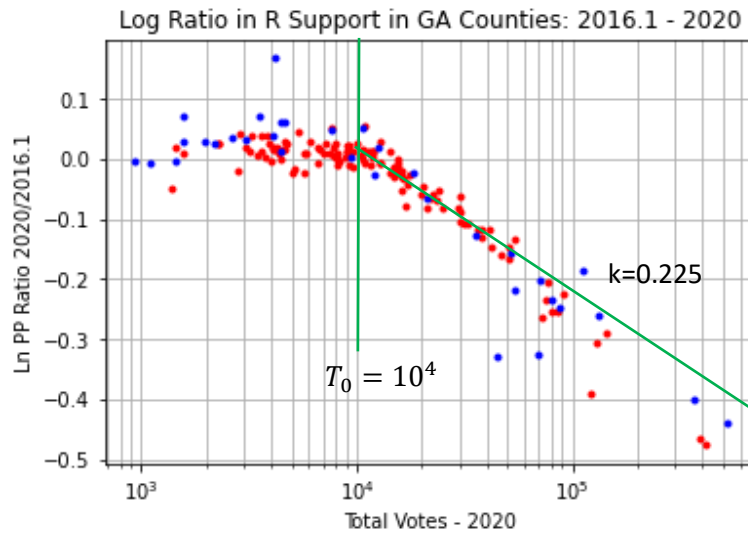

Figure 19. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in GA

Again closing the loop, we correct the 2020 data according to these parameters to obtain data set 2020.1. The log ratio of this adjusted data set is plotted against the adjusted 2016 data in Figure 20. The data (especially near the slope discontinuity point) appear nearly flat indicating a good match. (It could be argued that there is some residual slope to the graph, since the red dots appear to descent for larger counties. However, it is not unexpected that Trump might have lost some support in the red Metro Atlanta counties compared to 2016, so we leave the parameters as they are.)
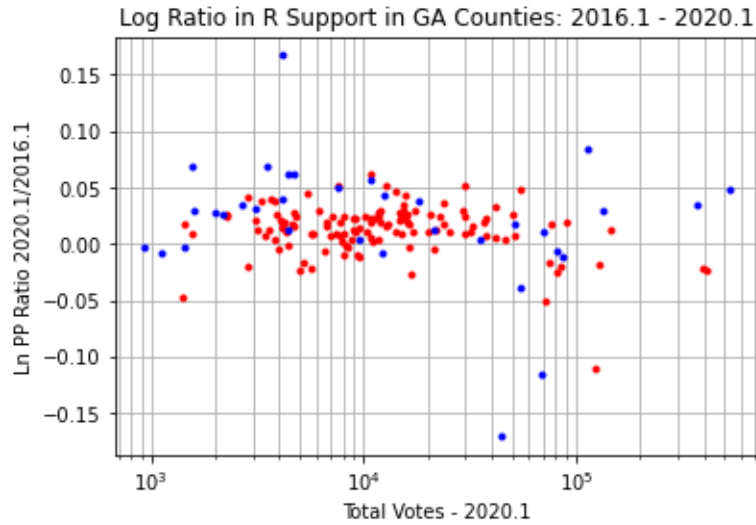
Figure 20. Adjusted Log Ratio of Republican PP between 2016 and 2020 in GA
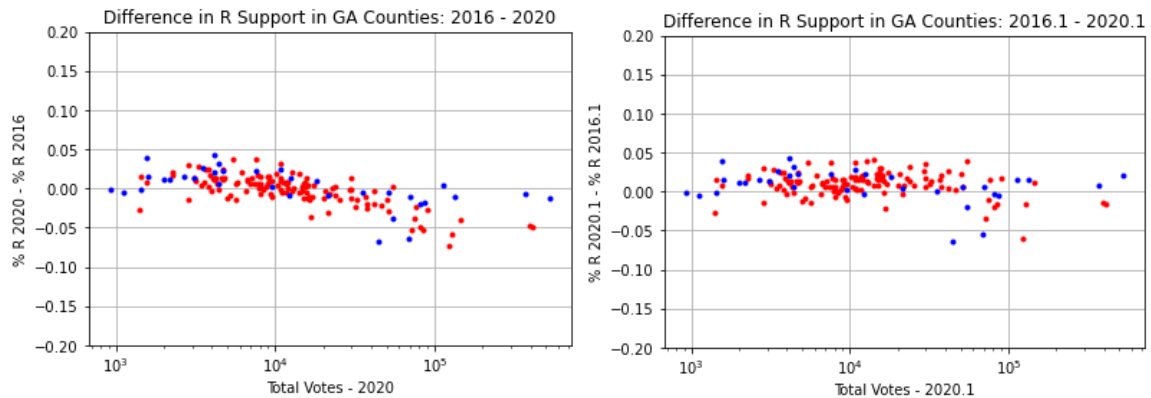

Figure 21. Original (left) and Adjusted (right) Difference in 2016 and 2020 Elections in GA

Figure 21 shows the original and adjusted difference in percentage points between the 2016 and 2020 elections. The adjusted 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. No anomalous trend is present.

Sanity Check

The reader may have noticed that the previous analysis walked through adjustments in the data of four separate elections, with the adjustment in each subsequent election being calculated based on the adjusted data from the last election. Because of this, it is a good idea to repeat the analysis for 2020 (the final adjustment) against an **unadjusted election**. The closest unadjusted data set for which no anomaly was observed is 2004. Therefore, we repeat the 2020 analysis to double check the parameters.

Figure 22 shows the log ratio between the original data from both the 2004 and 2020 elections. Due to the period of time between these elections, the variance of the data is much higher than the 2016 comparison in Figure 19. However, a slope discontinuity point of $T_0 = 10^4$ is still indicated, along with

an slope of (at least) $k = 0.225$. (This slope would have been difficult to determine directly with any precision if we had begun with this comparison and not walked through the 2016 data).
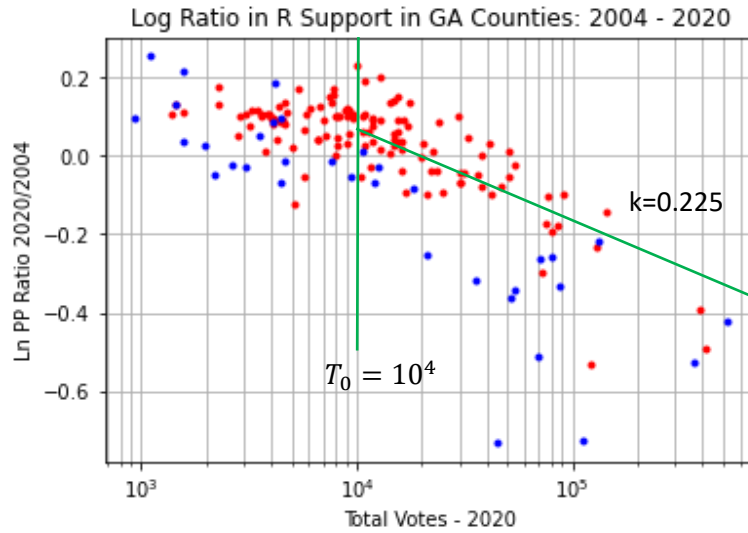

Figure 22. Log Ratio of Republican PP between 2004 and 2020 in GA

Figure 23 shows the log ratio between the original 2004 election data and the adjusted 2020 data. The data appear flat (within natural variability), indicating that the parameters fitted for 2020 through the course of the earlier analysis are indeed reasonable.
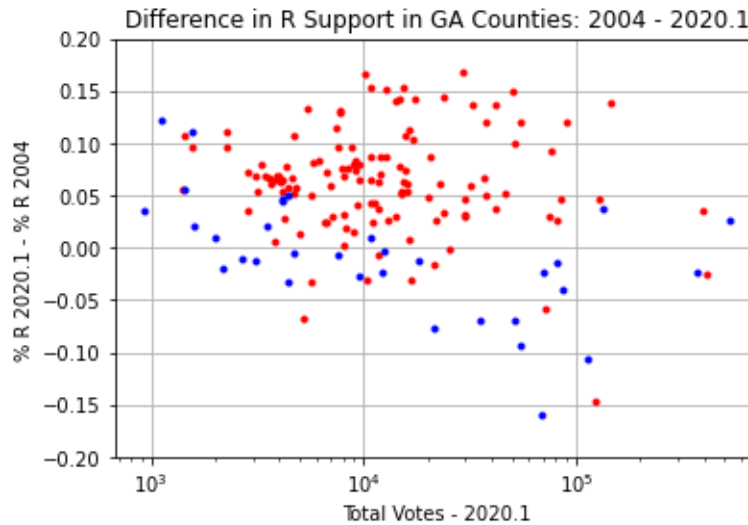
Figure 23. Adjusted Log Ratio of Republican PP between 2004 and 2020 in GA

Finally, Figure 24 shows the effect of the 2020 correction on the differential percentage data between 2020 and 2004. The right-hand plot shows the typical clusters we might expect to see in the evolution of a state over two decades. Trump appears to have gained about 5 pp on average compared to Bush, however, there are clear clusters of increased polarization in the mid-sized counties. Overall, the relative level of support for Trump and Bush is nearly unchanged in Metro Atlanta (although the size of that area

relative to the rest of the state has surely grown in population). The reader may judge for themselves which plot they find more reasonable.
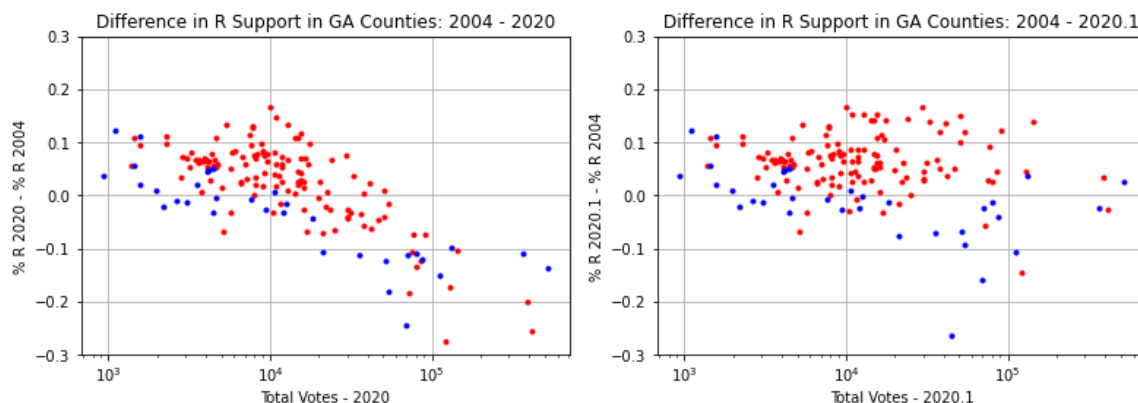

Figure 24. Original (left) and Adjusted (right) Difference in 2004 and 2020 Elections in GA

Summary Results

Table 1. Summary of Modeled Adjustments to GA Election Data

Year	T_0	k	Switched Votes	Official R Vote Margin	Official R %	Predicted R %
2008	10k	0.1	171k	205k	52.2	56.6
2012	10k	0.1	174k	305k	53.3	57.8
2016	10k	0.2	397k	211k	51.0	60.7
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 the results of this analysis in terms of actual vote numbers, rather than differential percentages and log ratios. 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.

The methods of this section will now be repeated for the other states under consideration, however, the discussion of the steps followed will be abbreviated, except in cases where deviations are made or exceptional circumstances occur. If the reader becomes confused as to how the analysis in the rest of the report was performed, he should refer back to this section and review the process.

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 algorithm may have been made to the data in several years, with slope discontinuity points at county sizes of both 10k and 100k voters.
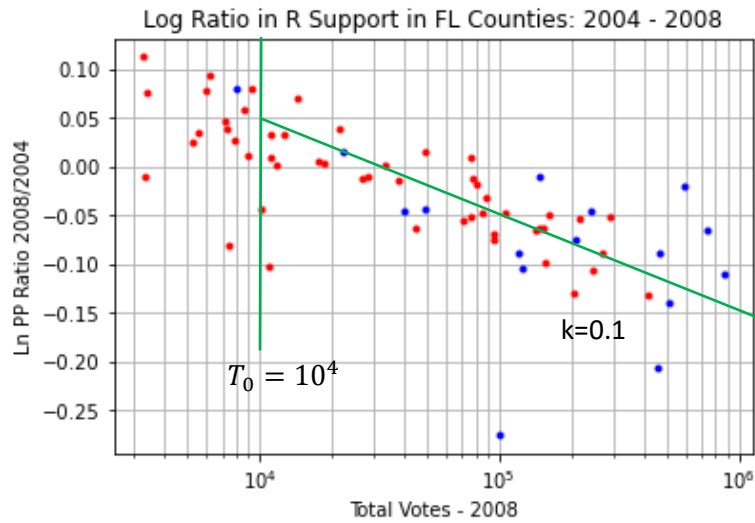
2008

Figure 25. Log Ratio of Republican PP between 2004 and 2008 in FL
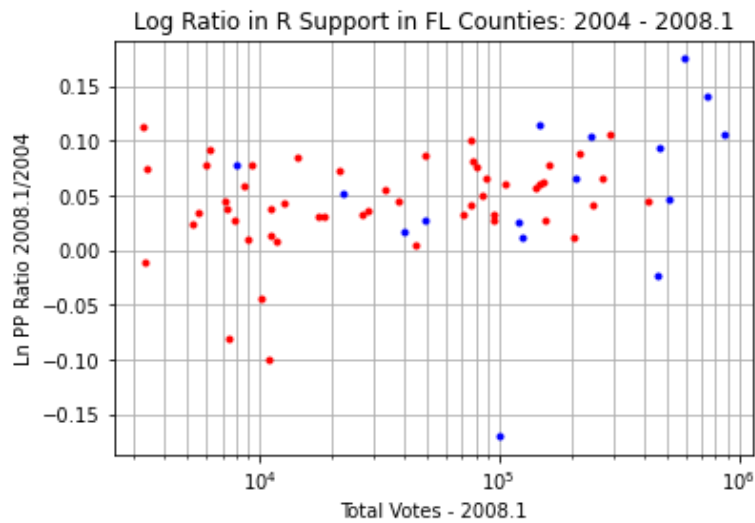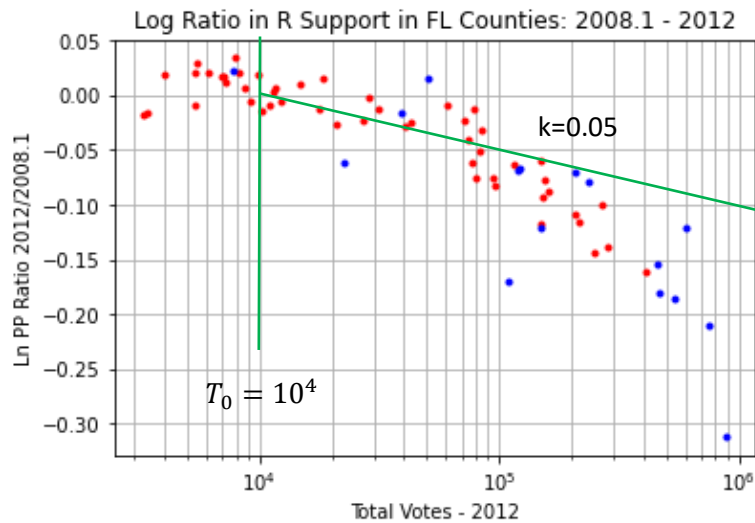

Figure 26. Adjusted Log Ratio of Republican PP between 2004 and 2008 in FL

Figure 25 shows the log ratio between the 2004 and 2008 elections. The graph has a clear slope of $k = 0.1$. The slope discontinuity point is difficult to estimate due to the fact that FL has much fewer small counties than GA. Conservatively, we estimate $T_0 = 10^4$. Figure 26 shows the log ratio of the data set adjusted according to these parameters (2008.1) compared to 2004, which appears flat. (The three largest counties in FL appear as interesting outliers in this data set. We will see this pattern repeated later as well.)

The results of this adjustment to the differential percentage point data are shown in Figure 27. The adjusted data are free of the anomaly and appear reasonable. The adjusted data indicate that FL supported Obama a handful of pp less than Kerry, a somewhat surprising result. (Officially, Obama won FL in 2008). In particular, there is a curious lack of support for him in the three largest counties. This may indicate that they do not follow the same algorithmic trend as the other counties and would be

"overcorrected" by this analysis. Therefore, final vote summary numbers should be taken with a moderate grain of salt.
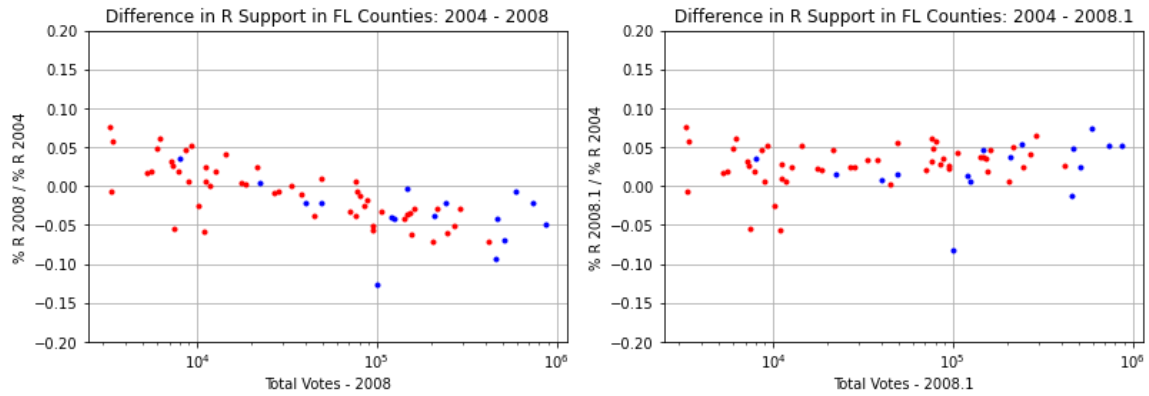
Figure 27. Original (left) and Adjusted (right) Difference in 2004 and 2008 Elections in FL

2012

Unlike in other states, in the 2012 election, the parameters of the anomaly appear to have changed in FL. Figure 28 shows the log ratio between the adjusted 2008 data and the 2012 data. Two distinct slope discontinuity 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 Figure 28 was removed to create the data set 2012.1. The log ratio technique was then repeated by comparing this adjusted data (2012.1) against against the adjusted 2008 data (2008.1). These results are shown in Figure 29.

Figure 28. Log Ratio of Republican PP between 2008 (adjusted) and 2012 in FL
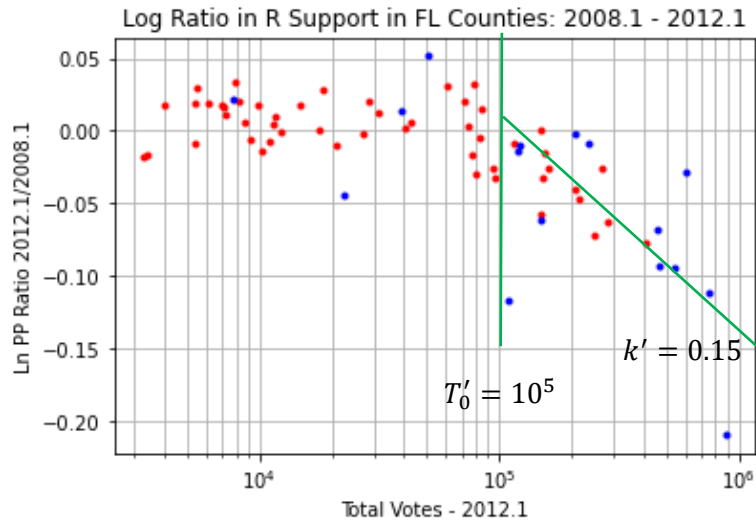
Figure 29. First Adjusted Log Ratio of Republican PP between 2008 (adjusted) and 2012 in FL

Figure 29 shows that the anomaly between 10k and 100k votes has effectively been removed by the first correction. There is an additional slope discontinuity point at $T_0' = 10^5$ with slope $k' = 0.15$. (This discontinuity could begin as far back as 80k votes, but we choose 100k as a conservative estimate) Applying this second correction yields the data set 2012.2.
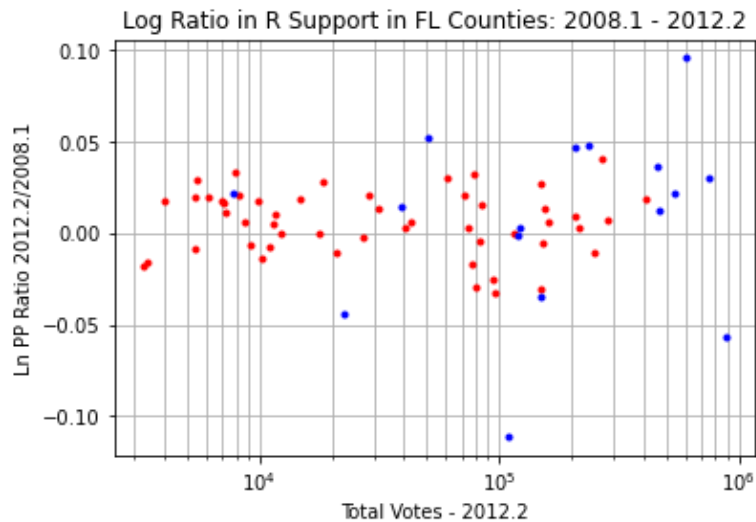

Figure 30. Second Adjusted Log Ratio of Republican PP between 2008 (adjusted) and 2012 in FL

Figure 30 shows the final log ratio between the 2012 data with both adjustments and the adjusted 2008 data. The graph appears flat and both anomalies have been removed. The data for two of the largest three counties again appear to be outliers, further supporting the notion that these data are not in family for 2008.

Figure 31 shows the differential voting percentage statistics for both the original and adjusted data sets. One sees that the unsightly "hump" around 80k votes in the original data set 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.

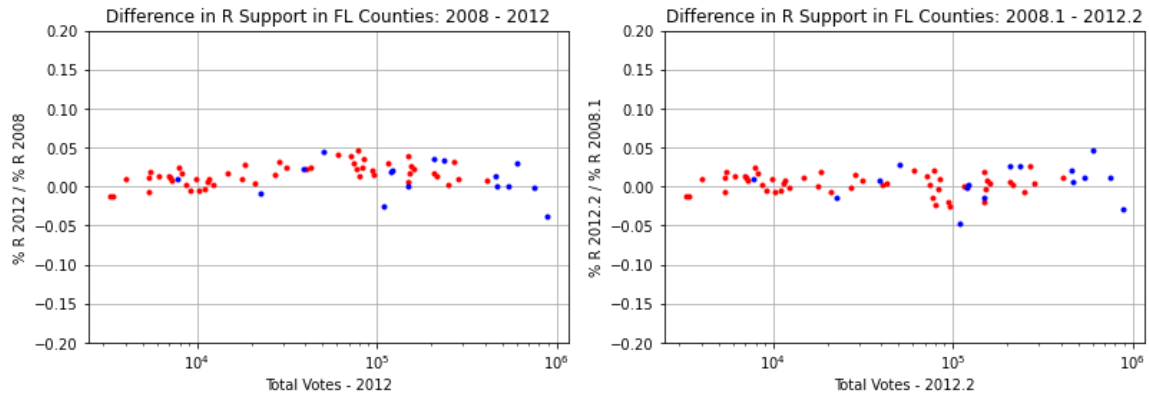
Figure 31. Original (left) and Adjusted (right) Difference in 2008 and 2012 Elections in FL

2016

Figure 32 shows the log ratio between the adjusted 2012 data and the 2016 election data. Again, we see that there appear to be two slope discontinuity 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 log ratio of 2016.1 is then plotted against 2012.2 in Figure 33.
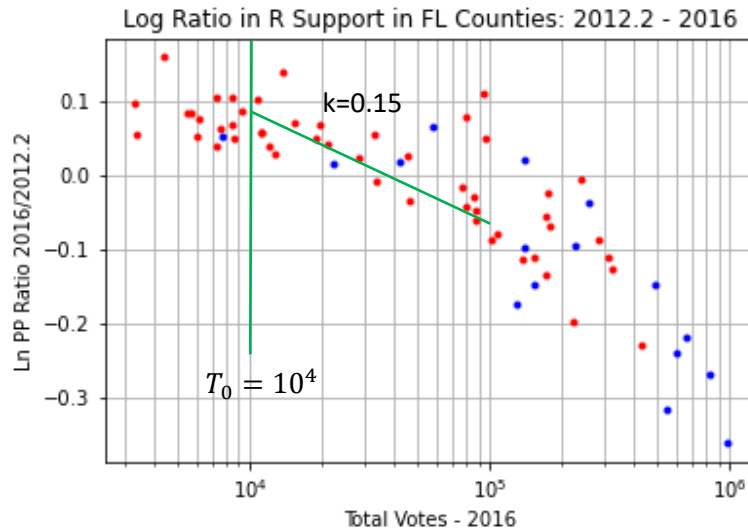
Figure 32. Log Ratio of Republican PP between 2012 (adjusted) and 2016 in FL
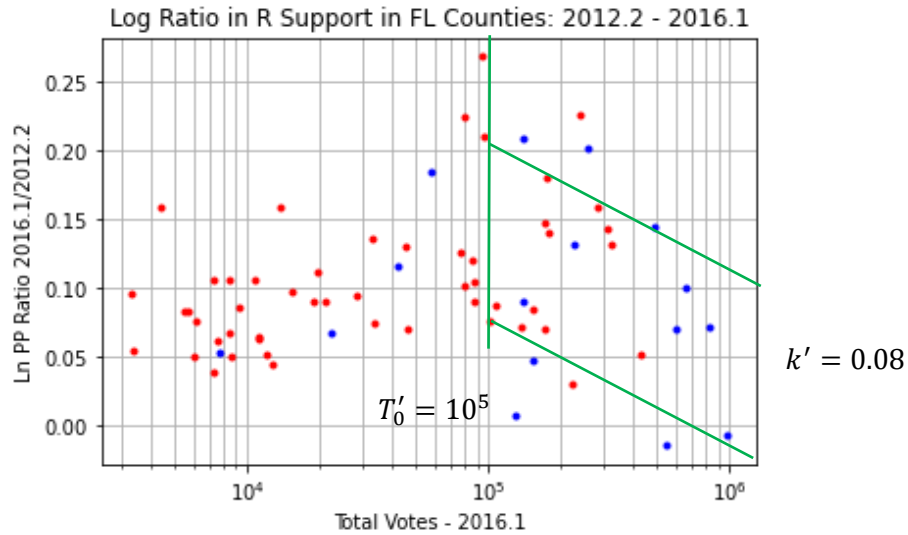
Figure 33. First Adjusted Log Ratio of Republican PP between 2012 (adjusted) and 2016 in FL

The slope of the second anomaly (from 10k to 100k votes) in Figure 33 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 (which appears to follow a single-slope correction model). This estimate is not inconsistent with Figure 33.
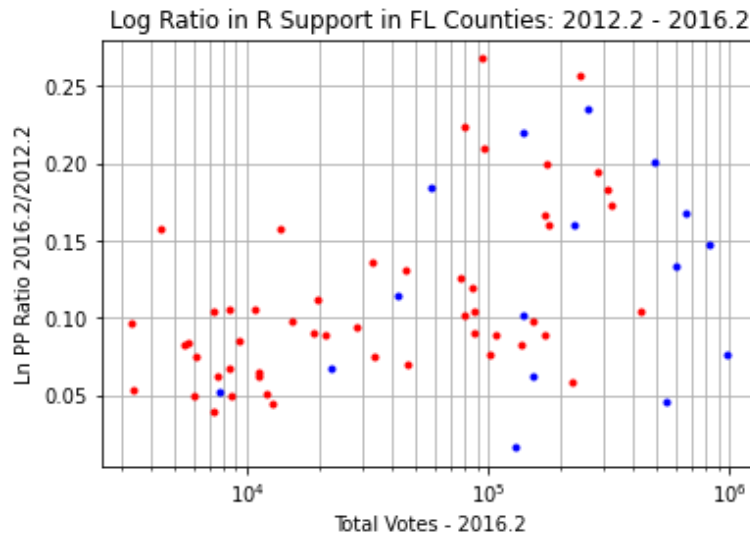
Figure 34. Second Adjusted Log Ratio of Republican PP between 2012 (adjusted) and 2016 in FL

Finally, the log ratio of the 2016 data with both slope adjustments (2016.2) vs the adjusted 2012 data is plotted in Figure 34. It is hard to say the graph appears flat above 100k votes, because there seems to have been a significant shift in voting patterns in this election for these counties. However, it looks much better than Figure 32 and Figure 33. The data below 100k votes appear to have no trend. (If anything, the k slope above 100k votes could be larger, so this is a conservative analysis).

Figure 35 shows the original and adjusted differences in percentage points between 2012 and 2016. The adjusted data shows a very different picture than the original results. In particular, there is a cluster of nearly 15 pp increase in support for Trump vs Obama that is almost totally invisible in the original results. This cluster warrants some further investigation.
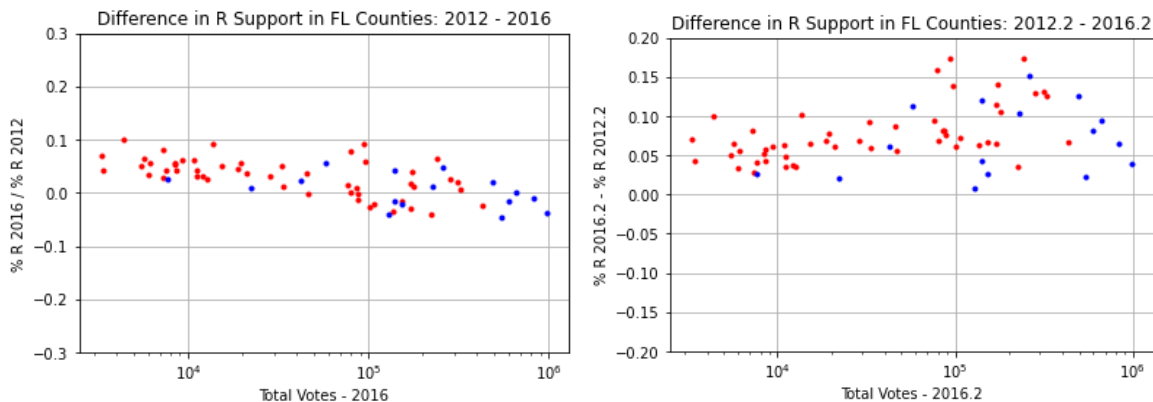

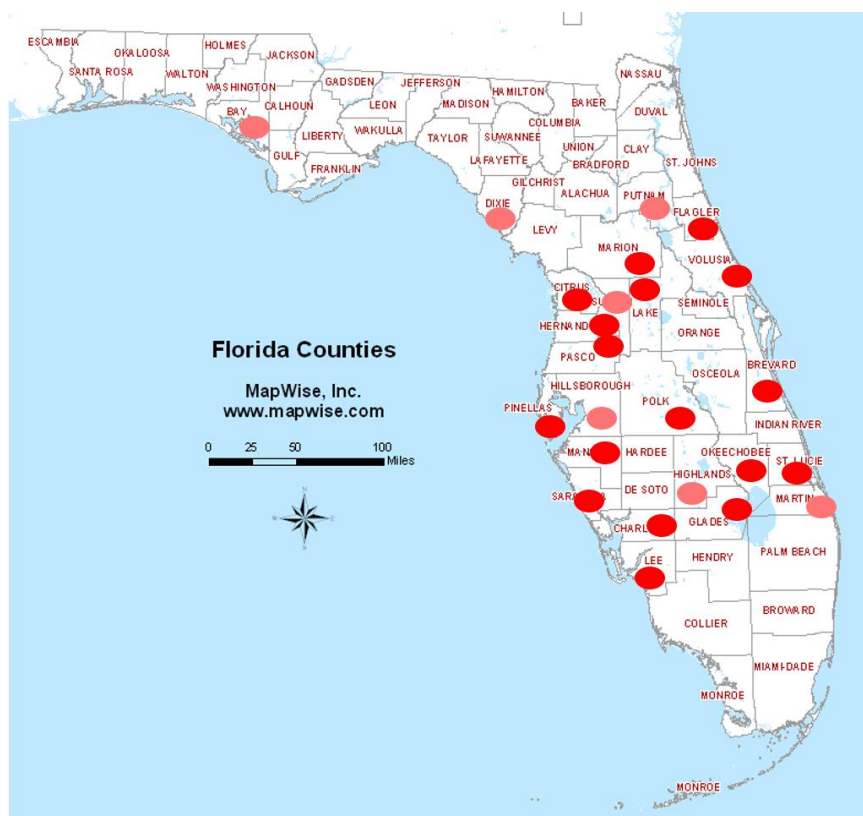Figure 35. Original (left) and Adjusted (right) Difference in 2012 and 2016 Elections in FL

Figure 36. FL Counties with Greater than 8pp (pink) or 10pp (red) R Increase in Adjusted 2016 Data Relative to 2012

Figure 36 shows the counties that had an increase (in the adjusted data) of at least 8 pp in R vote from 2012 to 2016 (pink) and those with over 10pp increase (red). All of these counties (with the exception of Bay in the panhandle) are in central FL. In fact, the list fills almost the entirety of central FL except for Orlando. These counties are demographically distinct from the rest of FL, being highly composed of

retirees. Therefore, the author finds this a feasible cluster, and far more feasible than the original data which show these counties decreased in R support logarithmically with size.

2020

For 2020, Figure 37 the log ratio plot in comparison to 2016 (adjusted). This shows that the anomaly in FL appears to once again follow a fixed slope of approximately $k = 0.18$ with a single slope discontinuity at $T_0 = 10^4$. (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. Occam's razor applies here as well.)
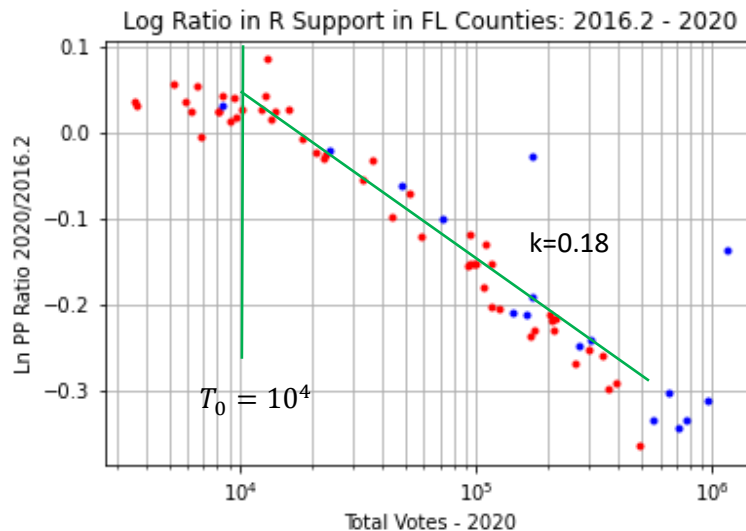
Figure 37. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in FL

Figure 38 shows the log ratio of the 2020 data adjusted according to these parameters against the 2016 data. There are several enormous outliers in this data set. The counties with over 0.1 adjusted log ratio are:

- Hendry – 12k voters, 40% Hispanic (third highest Hispanic % in state)
- Osceola – 172k voters, 48% Hispanic (second highest Hispanic % in state)
- Broward – 957k voters, 13% Hispanic, 30% Black (largest black population in the state, higher percent only in several small N FL counties)
- Miami-Dade – 1156k voters, 57% Hispanic (highest Hispanic % in state)

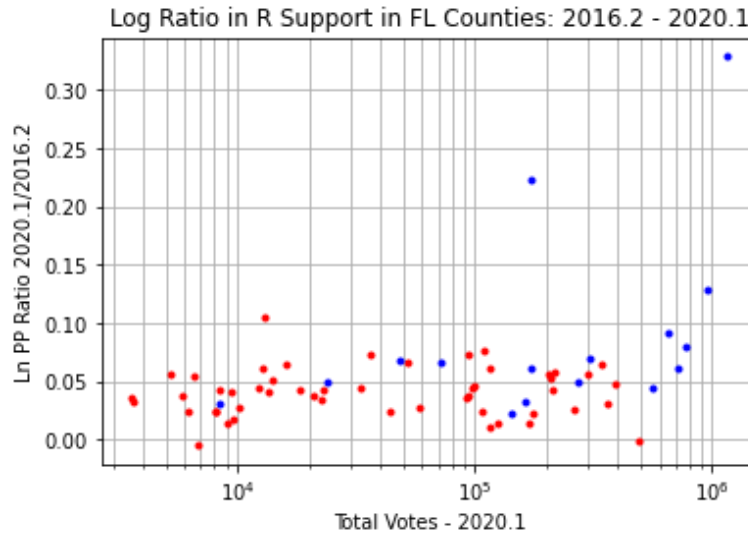These data indicate an enormous surge of support for Trump among minority voters in FL in 2020.

Figure 38. Adjusted Log Ratio of Republican PP between 2016 (adjusted) and 2020 in FL

Figure 39 shows the comparison of the original and adjusted differential percentage data for 2020 compared to 2016. The data have the tight variance associated with an incumbent election, except for the two counties Miami-Dade and Osceola, which are highly Hispanic. The adjusted data show a steady increase in support of a handful of pp across the state, except for the noted outliers. There is no anomalous slope trend remaining.
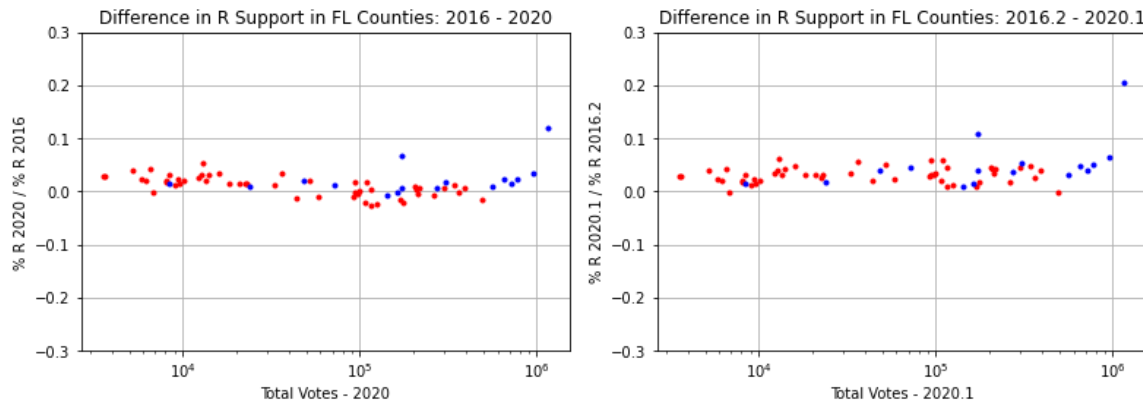

Figure 39. Original (left) and Adjusted (right) Difference in 2016 and 2020 Elections in FL

Sanity Check

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 40 shows the log ratio analysis for these elections. The data are quite noisy due to the length of time between these elections, but indicate that a slope of $k = 0.18$ is plausible.
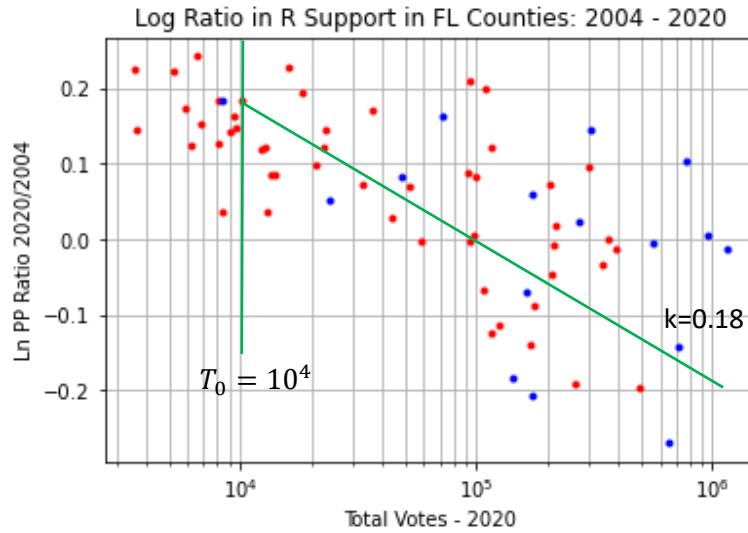
Figure 40. Log Ratio of Republican PP between 2004 and 2020 in FL

Figure 41 shows the log ratio between the 2004 election and the adjusted 2020 data. The data appear flat over 10k to 100k voters, and the data above 100k voters are consistent with the central FL and minority voter clusters previously discussed.
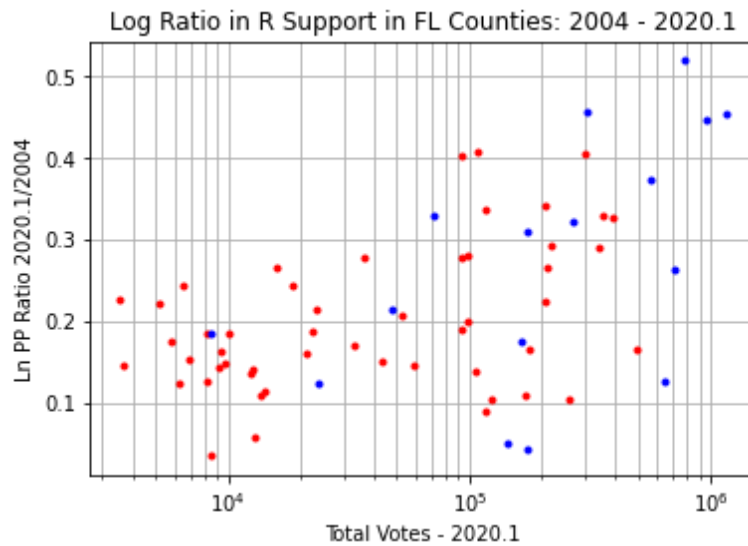

Figure 41. Adjusted Log Ratio of Republican PP between 2004 and 2020 in FL

Figure 42 compares the original difference between the data sets and the difference from 2004 to the adjusted 2020.1 data set. The anomalous trend has been removed, and we see the central FL and minority voter clusters clearly.
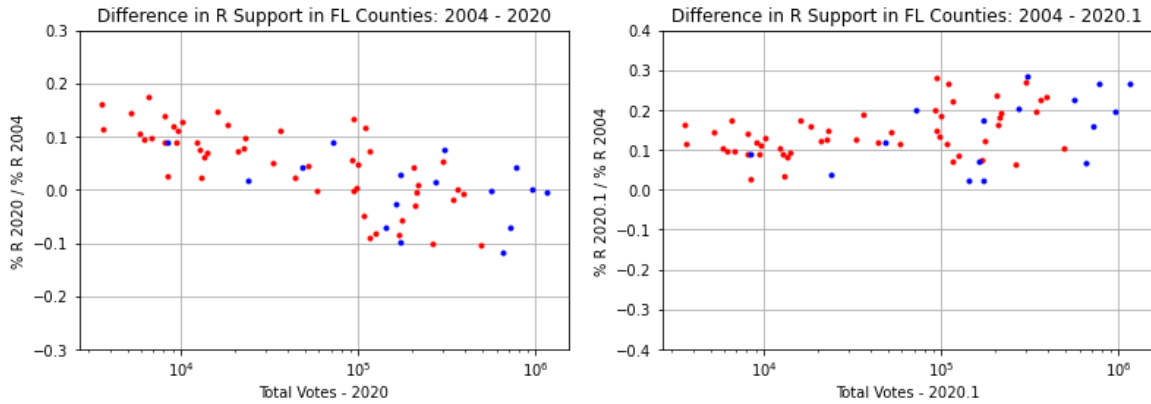
Figure 42. Original (left) and Adjusted (right) Difference in 2004 and 2020 Elections in FL

Summary Results

Table 2. Summary of Modeled Adjustments to FL Election Data

Year	T_0	k	T_0'	k'	Switched Votes	Official R Vote Margin	Official R %	Predicted R %
2008	10k	0.1			648k	-236k	48.2	56.0
2012	10k	0.05	100k	0.15	625k	-74.3k	49.1	56.5
2016	10k	0.15	100k	0.08	1492k	113k	49.0	64.9
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-correction 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

The vote data for NC is more difficult to analyze, because the vast majority of counties are clustered between 10k and 100k voters (a "hump" on the graph), with only a handful of (heavily blue) counties exceeding this. Referring to the "control" data (Figure 3), it seems that these large, blue counties tend to line up with the bottom of the "hump" in terms of voting pattern. This general principle is used in the data analysis in this section. However, the reader may wish to treat the slope numbers from this section with lower confidence.

2008

Like FL and GA, the anomalies in NC voting data begin in 2008. Figure 43 shows the log ratio between the 2004 and 2008 Republican percentage point totals in NC. The slope for the vote switching model can be read as approximately $k = 0.1$ with a discontinuity point of $T_0 = 10^4$.
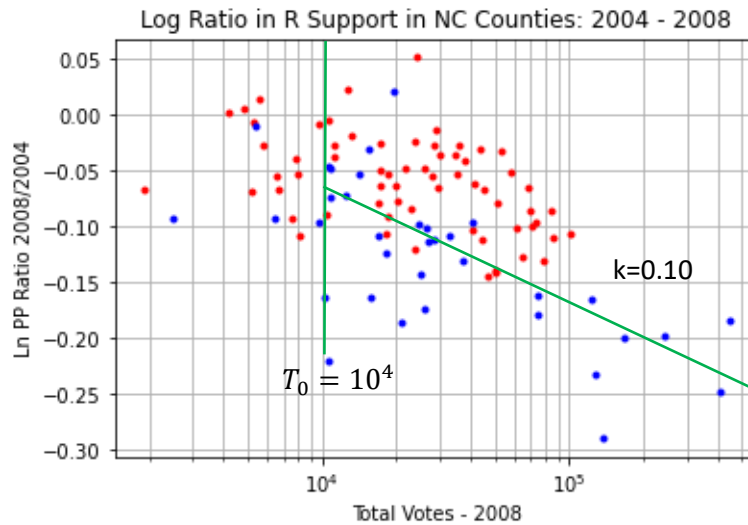
Figure 43. Log Ratio of Republican PP between 2004 and 2008 in NC

Figure 44 shows the log ratio of the 2008 voting data with the proposed corrections applied (2008.1) to the 2004 election. The sloped trend appears to have been removed.
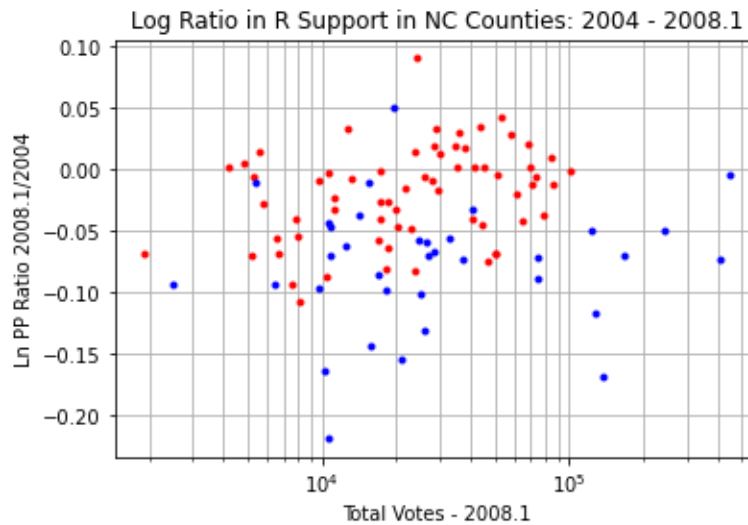

Figure 44. Adjusted Log Ratio of Republican PP between 2004 and 2008 in NC

Figure 45 shows the differential percentages for the adjusted data (2008.1) vs the original. The adjusted data are without an anomalous trend and show that Obama enjoyed a boost of 2-3 pp over Kerry in NC and up to 5 pp in blue counties.
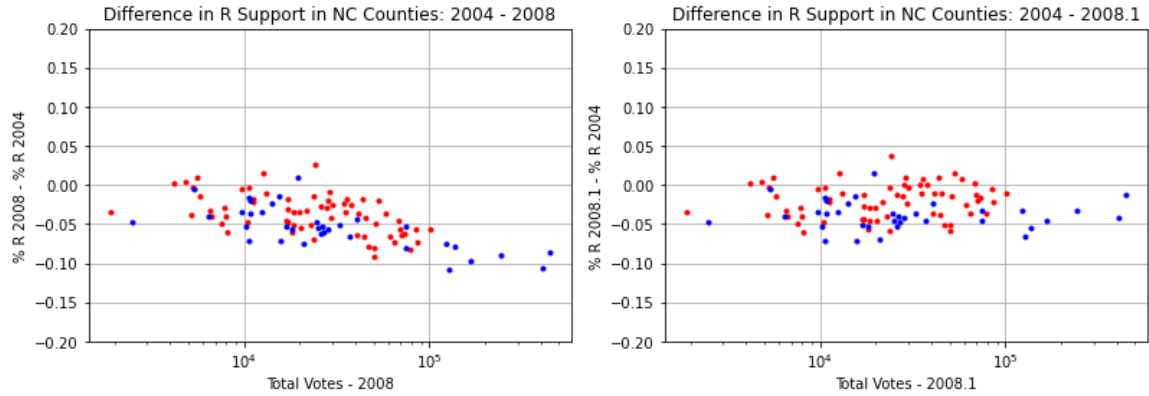
Figure 45. Original (left) and Adjusted (right) Difference in 2004 and 2008 Elections in NC

2012

As in GA, the differential percentages between the 2008 and 2012 election were unremarkable. This suggests no change in the model parameters. For brevity, the author will simply show that the log ratio plot between the adjusted 2008 data and the 2012 data adjusted using the parameters $k = 0.1$ and $T_0 = 10^4$ appears flat (Figure 46). The differential percentages between the adjusted elections are virtually identical to those in the original data (Figure 6).
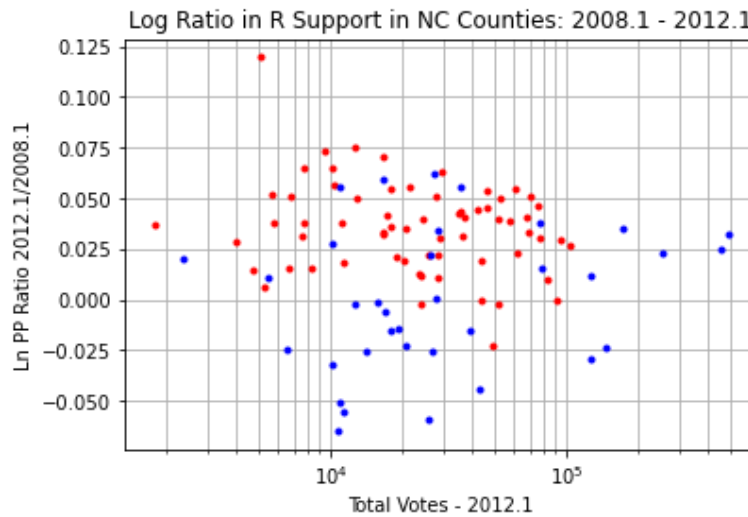
Figure 46. Adjusted Log Ratio of Republican PP between 2008 (adjusted) and 2012 in NC

2016

In 2016, the anomaly in the differential percentage data resurfaces. Figure 47 shows the log ratio between the adjusted 2012 data and 2016. The graph indicates parameters of $T_0 = 10^4$ and k around

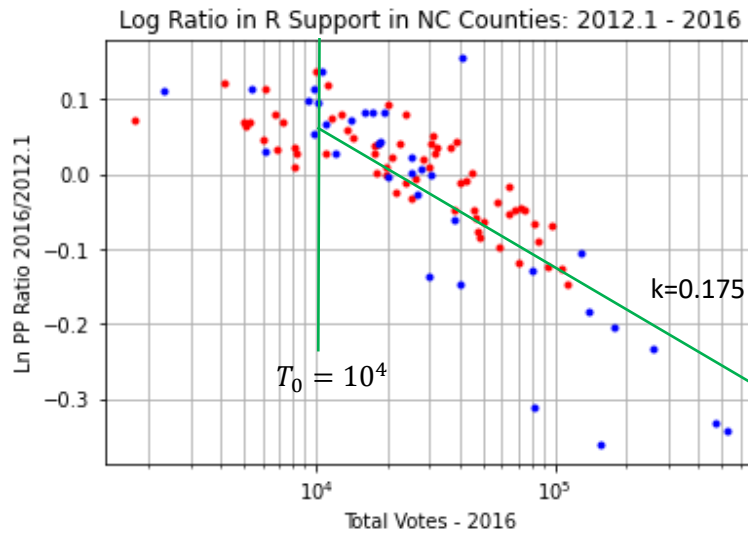0.2. $k = 0.2$ seemed to supply too much correction to the slope, so a more modest $k = 0.175$ is assumed.

Figure 47. Log Ratio of Republican PP between 2012 (adjusted) and 2016 in NC

Figure 48 shows the adjusted log ratio after applying the correction suggested by these parameters to 2016 data (2016.1). The slope of the graph appears flat without discontinuity.
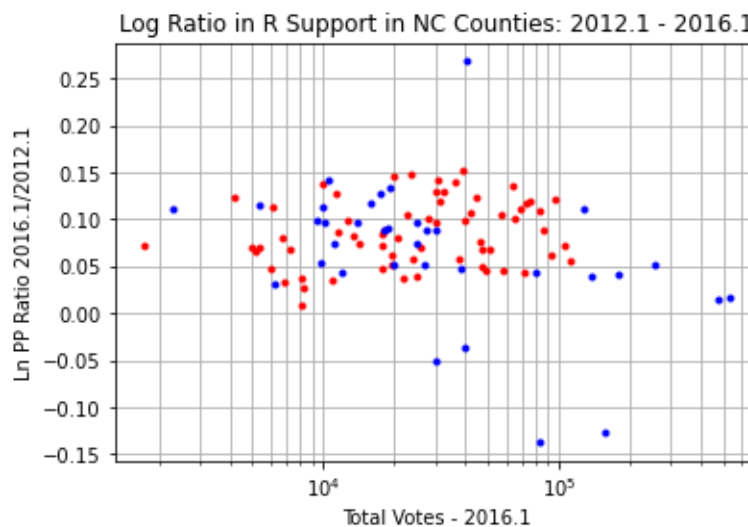

Figure 48. Adjusted Log Ratio of Republican PP between 2012 (adjusted) and 2016 in NC

The original and adjusted differential percentages between these years are shown in Figure 49. The adjusted data show a 5 pp boost for Trump on average, except in several outlier blue areas.
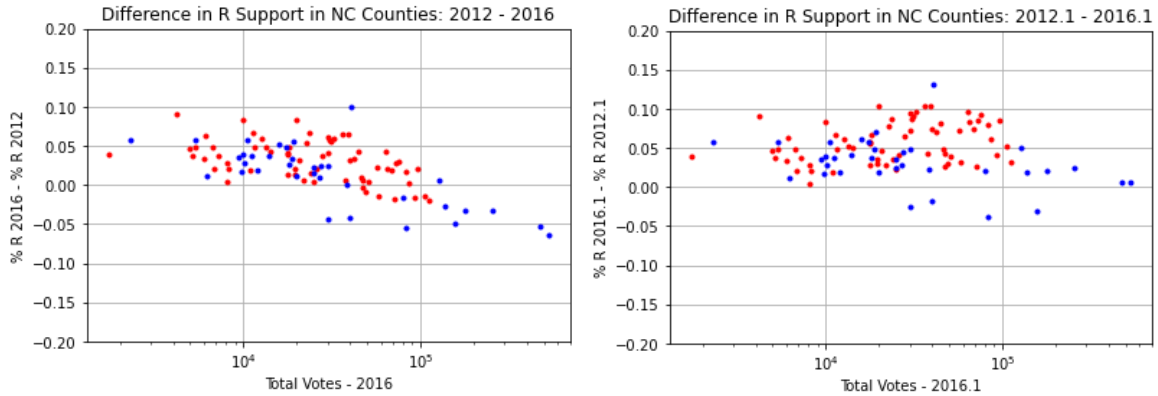
Figure 49. Original (left) and Adjusted (right) Difference in 2012 and 2016 Elections in NC

2020

Figure 50 shows the log ratio plot between the adjusted 2016 and raw 2020 data. Since the anomaly is downward sloping from 2016 to 2020 in the original data (Figure 8), this suggests a slight increase in k. Figure 50 (conservatively) support an estimate of $k = 0.20$ and $T_0 = 10^4$.
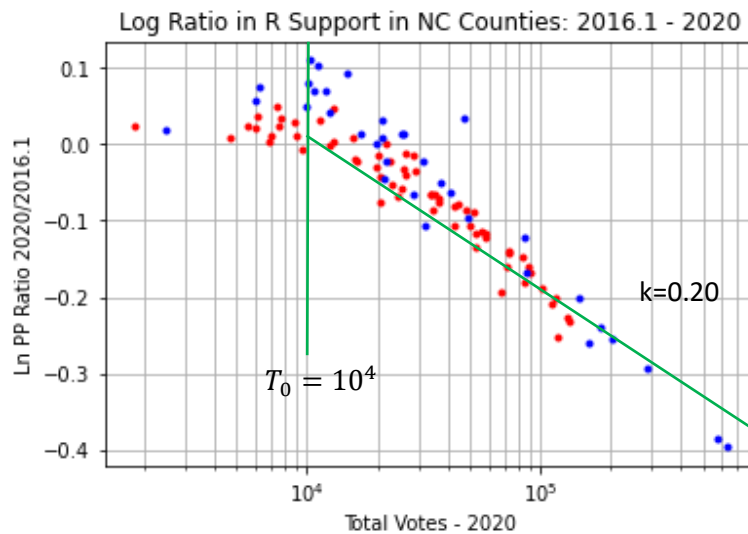

Figure 50. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in NC

Figure 51 shows the change in the log ratio between 2016 and 2020 after the adjustments suggested by these parameters is made. The trend line now appears flat with no discontinuity. The reader may debate the degree to which the large county data have been straightened, however, the overall difference in these counties appears similar to the average increase in support in other blue counties. (Note that the y-axis scale of this plot is very tight!) For the purposes of this (admittedly approximate) analysis, we will accept the estimated parameters and examine their implications.
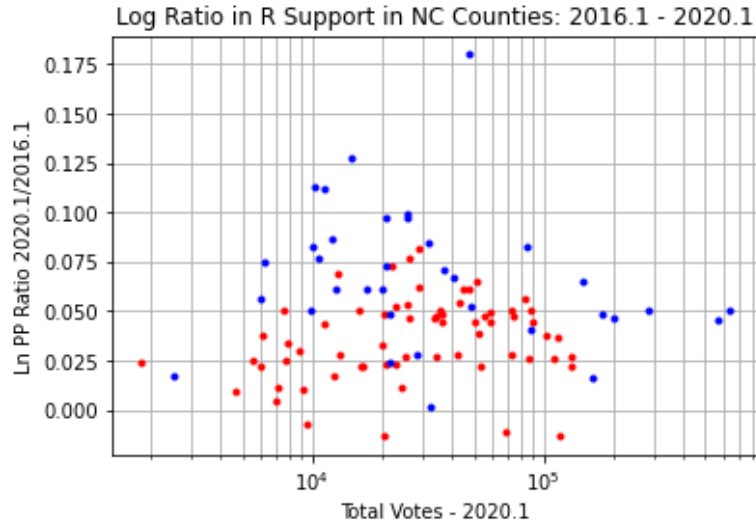
Figure 51. Adjusted Log Ratio of Republican PP between 2016 (adjusted) and 2020 in NC

Again, the differential percentage data between these two elections seem reasonable when corrected, as shown in Figure 52. 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.
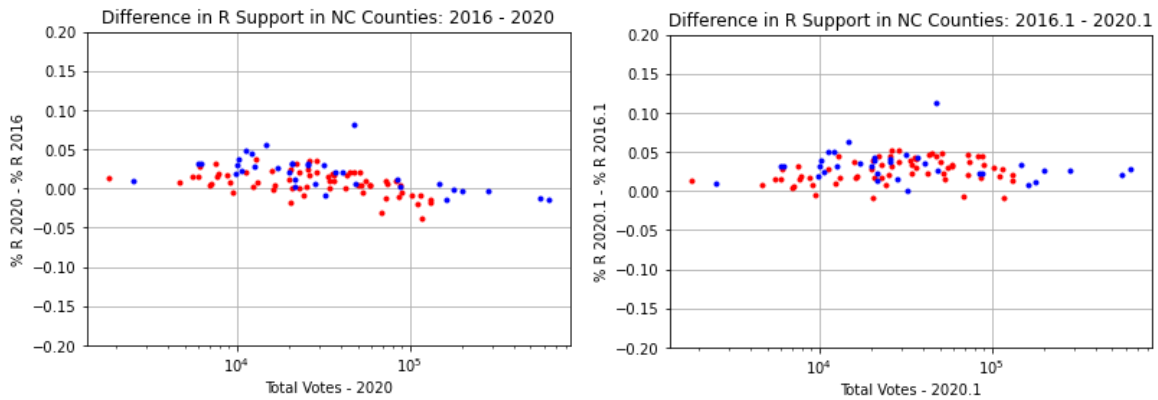

Figure 52. Original (left) and Adjusted (right) Difference in 2016 and 2020 Elections in NC

Sanity Check

Once again, we will use the 2004 election as a "baseline" to sanity check the results.   Figure 53 shows the log ratio metric plotted between these elections. This data confirms the parameters $T_0 = 10^4$ and $k = 0.2$ are reasonable.
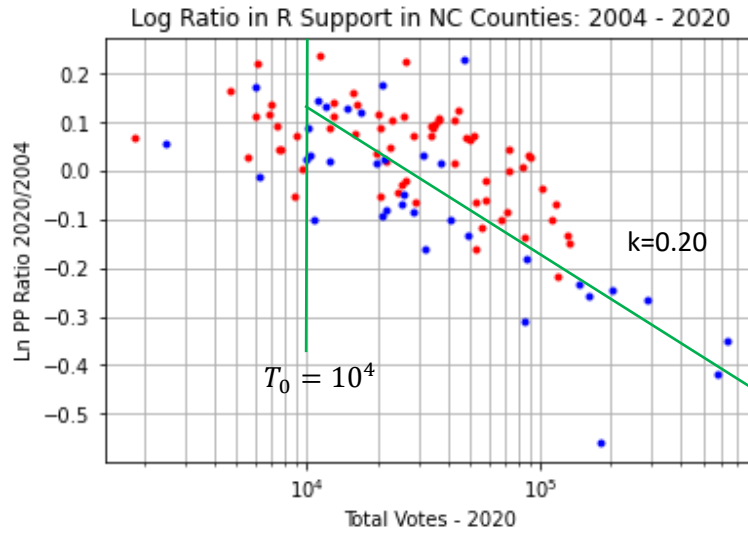
Figure 53. Log Ratio of Republican PP between 2004 and 2020 in NC

Similarly, Figure 54 shows that the plot of the log ratio data between 2004 and the adjusted 2020 data appears flat and without slope discontinuity, further confirming the parameter estimates.
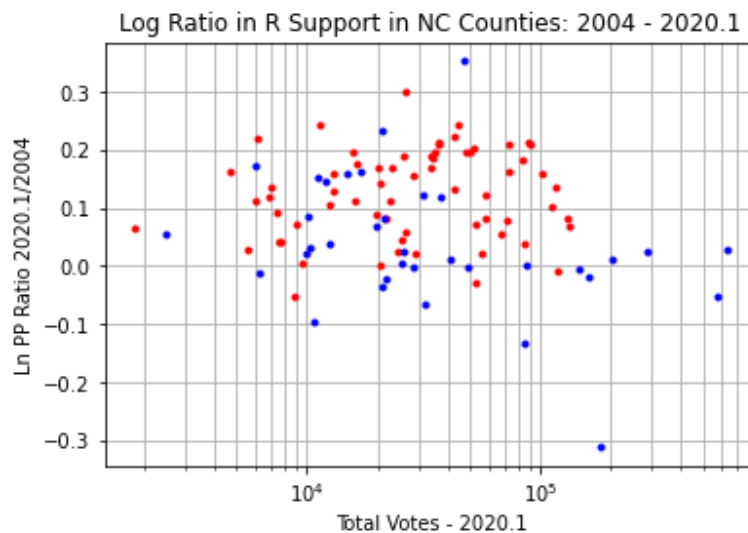

Figure 54. Adjusted Log Ratio of Republican PP between 2004 and 2020 in NC

The adjusted differences in percentages between 2004 and 2020 are shown in Figure 55. The adjusted data appear much more reasonable than the original, showing clear clusters of behavior of red and blue counties. The adjusted data suggests that mid-sized red counties in NC supported Trump significantly more than Bush (10pp) and blue counties only moderately so.
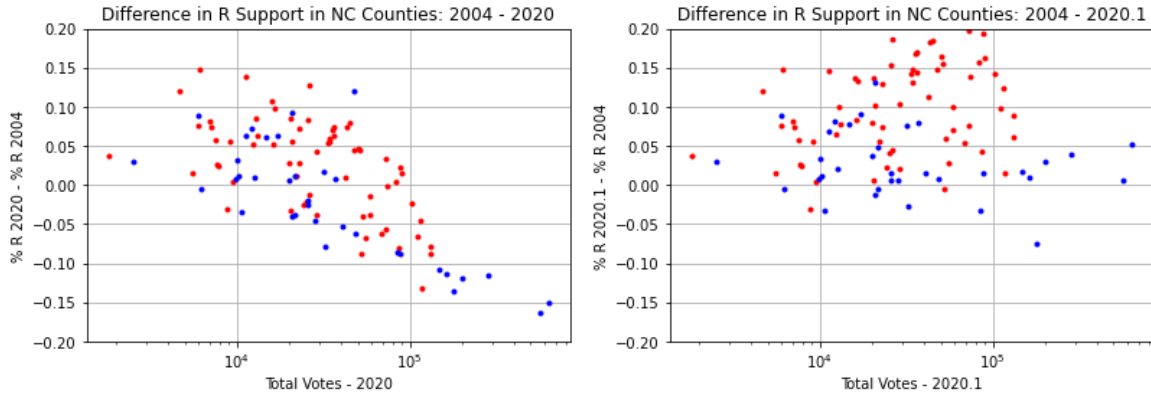
Figure 55. Original (left) and Adjusted (right) Difference in 2004 and 2020 Elections in NC

Summary Results

The overall summary of estimated parameters and their resulting impact is shown in Table 3. In particular, these corrections overturn 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	Switched Votes	Official R Vote Margin	Official R %	Predicted R %
2008	10k	0.1	205k	-14.2k	49.4	54.1
2012	10k	0.1	227k	92.0k	50.4	55.4
2016	10k	0.175	455k	173k	49.8	59.4
2020	10k	0.20	687k	74.5k	50.0	62.4

OH Anomaly Analysis

Unlike GA, FL, and NC, there is no apparent anomaly in 2008 and 2012 data for OH (See Figure 4 and Figure 6). In this data set, the anomaly begins in 2016.

2016

Figure 56 shows the log ration between the 2012 and 2016 elections. The graph suggests a discontinuity parameter of $T_0 = 10^4$ and slope of $k = 0.15$ for the anomaly model. The discontinuity parameter is difficult to confirm, since there are only a handful of counties below 10k voters. However, there is no evidence of a discontinuity higher than that number, so we estimate this conservatively. Likewise, the slope parameter is less clear due to the variance of the data. However, $k = 0.15$ appears to fit well the offset trends in both the red and blue counties.
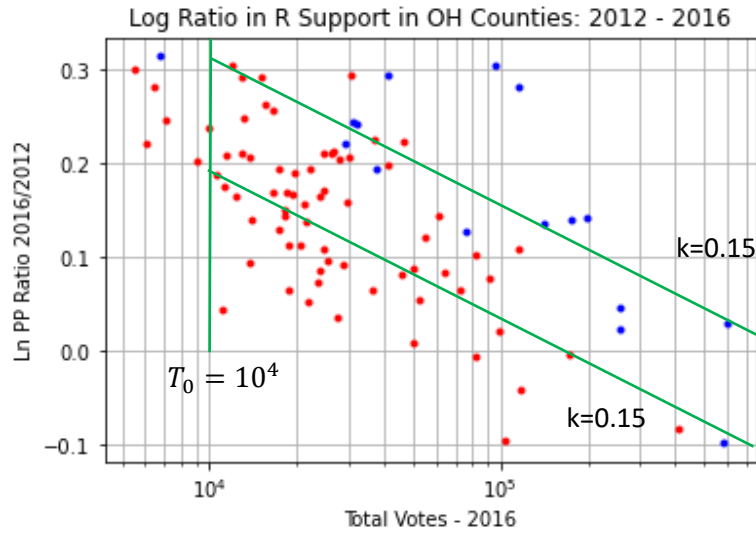
Figure 56. Log Ratio of Republican PP between 2012 and 2016 in OH

Figure 57 shows the log ratio between the 2012 and 2016 elections after applying the adjustments consistent with the estimated parameters. This graph shows clear trends of red and blue county behavior without any apparent slope or discontinuity.
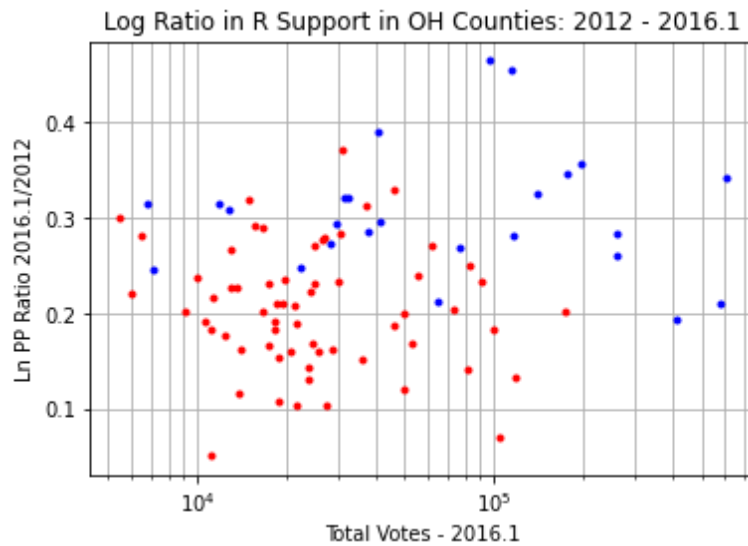

Figure 57. Adjusted Log Ratio of Republican PP between 2012 and 2016 in OH

The original and adjusted differential percentage data are shown in Figure 58. The adjusted data show a 15 pp swing toward Trump in OH vs Romney over almost all counties. This trend may be compared with the 2004-2008 swing towards Obama in Figure 4. (In this report, the author has deliberately fixed the axes of the differential percentage point plots between adjacent elections in order to provide a fair comparison, but the data here begin to run off the plot.)
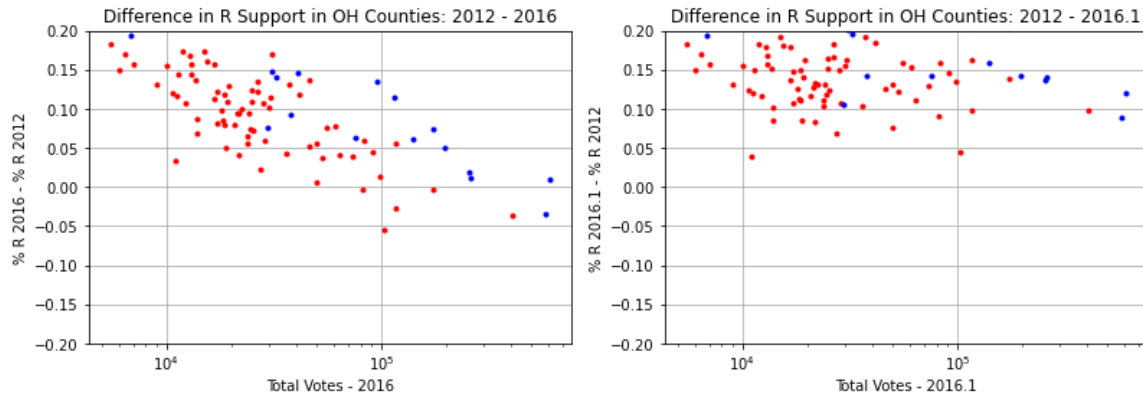
Figure 58. Original (left) and Adjusted (right) Difference in 2012 and 2016 Elections in OH

2020

In 2020, there is an additional downward trend to the anomaly in the differential percentage data (Figure 8), supporting an increase in k. Figure 59 is consistent with $k = 0.175$ for this election, with $T_0 = 10^4$. (In fact, it appears that a lesser value for T_0 might even be justified due to the lack of apparent slope discontinuity in this plot. We cannot say with any reasonable certainty due to the small number of counties less than that size. For a conservative analysis, we will leave T_0 as it is.)
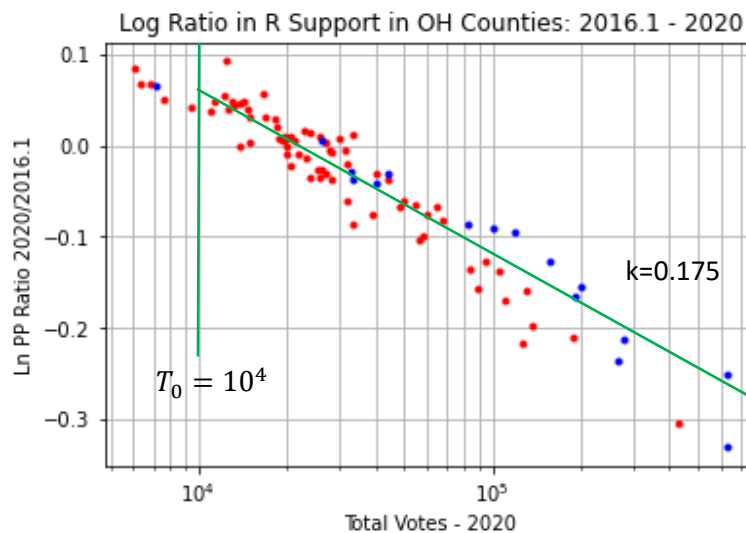

Figure 59. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in OH

Figure 60 shows the log ratio between 2016 and 2020 after adjusted according to the estimated parameters. The data appear flat with separate, offset trends for red and blue counties.
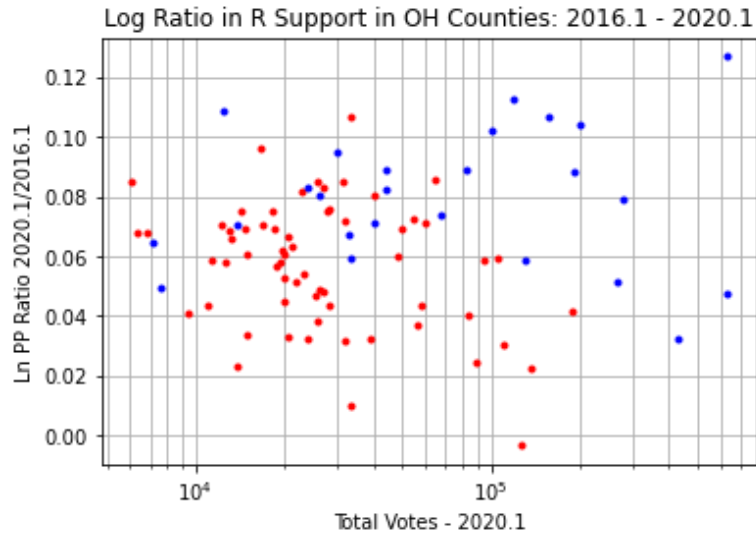
Figure 60. Adjusted Log Ratio of Republican PP between 2016 (adjusted) and 2020 in OH

Figure 61 shows the original and corrected differential percentages between 2016 and 2020. The anomalous trend line in the original data has been effectively removed.
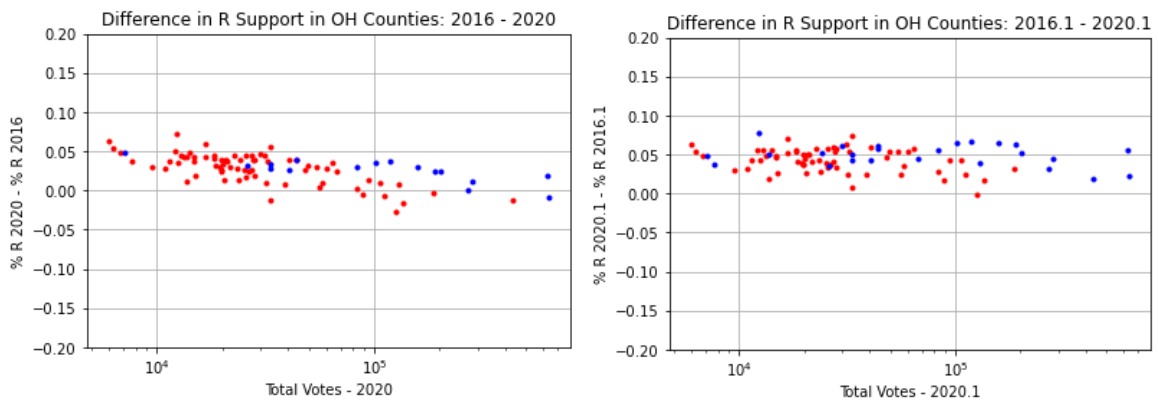

Figure 61. Original (left) and Adjusted (right) Difference in 2016 and 2020 Elections in OH

Sanity Check

As in the other states, we compare the final adjustments to an unadjusted election as a sanity check. This time, since we assumed no anomaly until 2016, 2012 is the nearest unadjusted data set. Figure 62

shows the log ratio between 2012 and 2020. The graph indicates that the slope estimate of $k = 0.175$ for the proposed 2020 adjustment is reasonable.
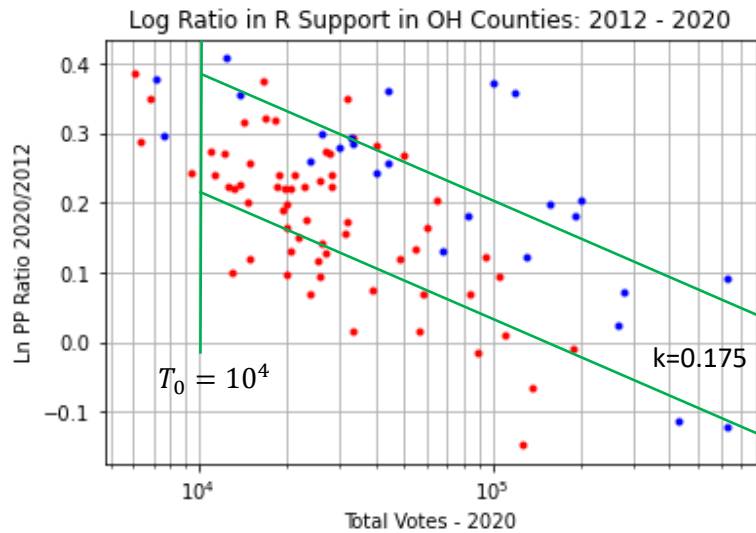

Figure 62. Log Ratio of Republican PP between 2012 and 2020 in OH

Figure 63 shows the log ratio between 2012 and the adjusted 2020 data. The data appear flat and without obvious discontinuity in slope, confirming the feasibility of these parameters.
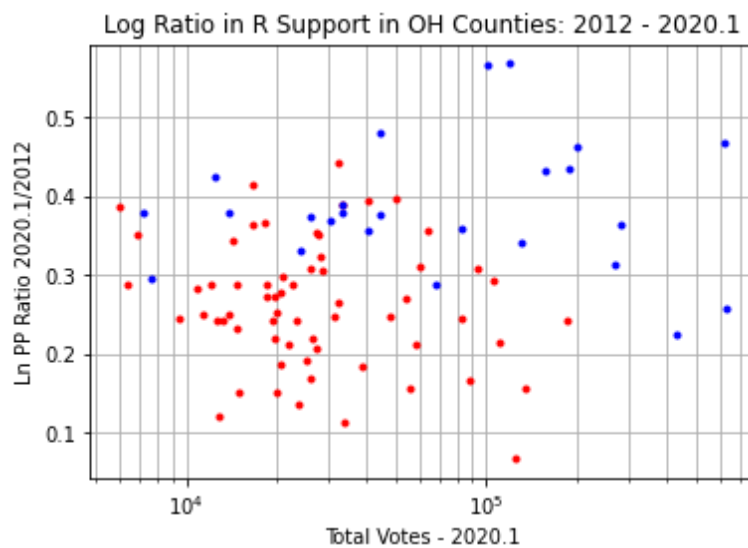
Figure 63. Adjusted Log Ratio of Republican PP between 2012 and 2020 in OH

Figure 64 shows the original and corrected data for these two elections. The adjusted data appear free of the trend and quite reasonable. Overall, OH appears to have shifted 20 pp to Trump from Romney. (Again, this may be compared with the 2004-2008 swing towards Obama in Figure 4). This is consistent with the overall demographic uniformity of OH as a state.
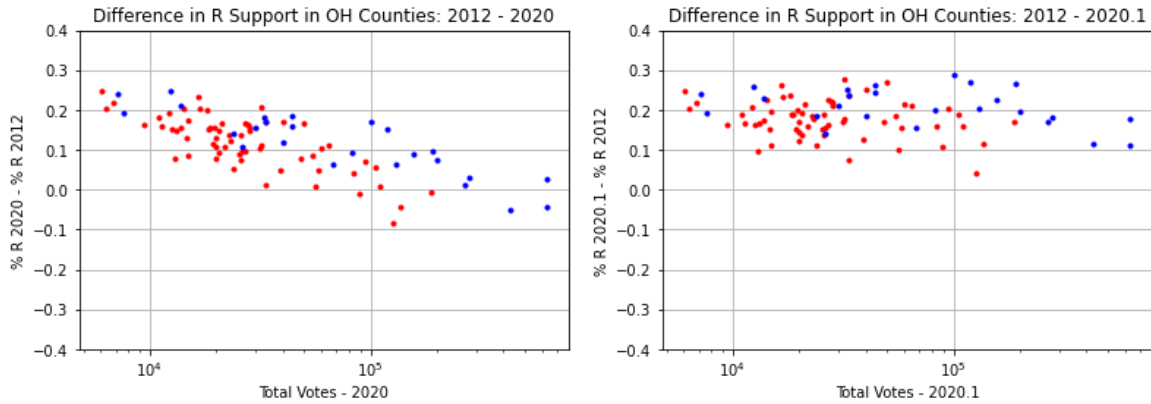
Figure 64. Original (left) and Adjusted (right) Difference in 2012 and 2020 Elections in OH

Summary Results

Finally, Table 4 summarizes the overall effect of the corrections on the data. The adjustments do not cause any changes in the outcome, but strengthen Trump's "marginal" win to landslide territory in 2020.

Table 4. Summary of Modeled Adjustments to OH Election Data

Year	T_0	k	Switched Votes	Official R Vote Margin	Official R %	Predicted R %
2016	10k	0.15	521k	447k	51.3	61.1
2020	10k	0.175	722k	476k	53.3	65.5

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.

2016

As in OH, PA does not show any anomaly before the 2016 election. The log ratio analysis for 2016 compared to 2012 is shown in Figure 65. However, the discontinuity point in PA appears to be different than from other states. There is no apparent slope in the log ratio before $T_0 = 5 \times 10^4$. After this point, there is an estimated slope of $k = 0.175$.
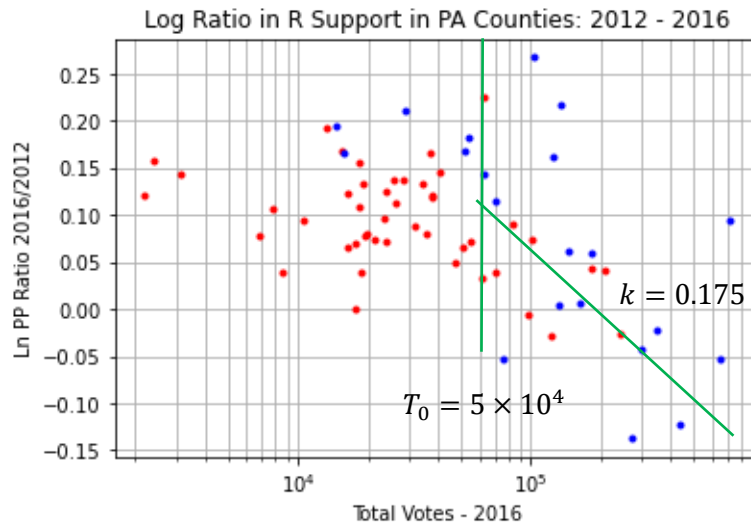
Figure 65. Log Ratio of Republican PP between 2012 and 2016 in PA

Figure 66 shows the log ratio of the data corrected using these estimated parameters (2016.1) with the 2012 election. The data do not appear to have an obvious trend or slope discontinuity, indicating a reasonable estimate.
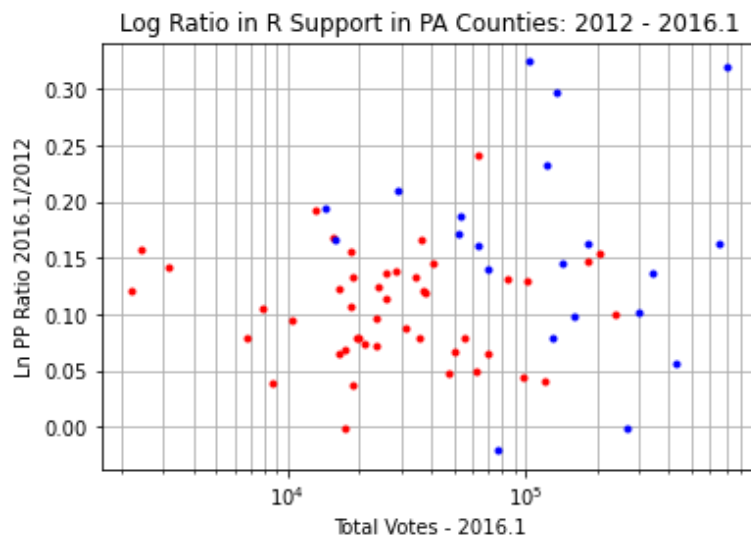

Figure 66. Adjusted Log Ratio of Republican PP between 2012 and 2016 in PA

Figure 67 shows the original and adjusted differential percentage data between 2012 and 2016. The adjusted data appear reasonable with clear clusters of support and show a shift of around 5-7 pp toward Trump from Romney.
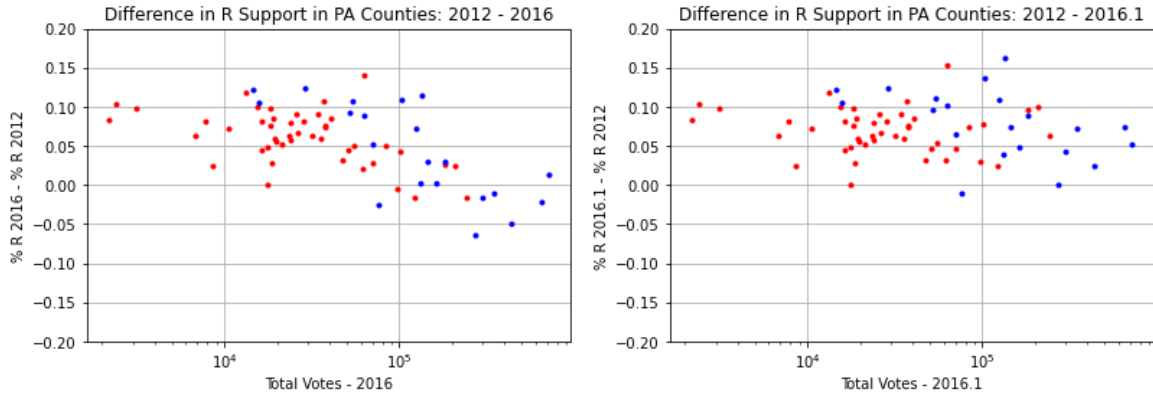
Figure 67. Original (left) and Adjusted (right) Difference in 2012 and 2016 Elections in PA

2020

For 2020, the anomaly appears as a very slight downward slope in Figure 8. This suggests a small incremental increase in k. Figure 68 shows the log ratio between the 2016 and 2020 elections. The parameters $T_0 = 5 \times 10^4$ and $k = 0.2$ are consistent with this graph.
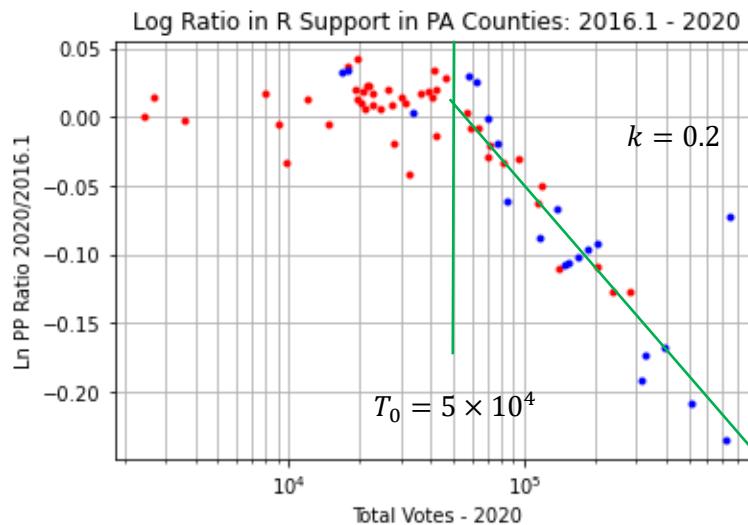
Figure 68. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in PA

Figure 69 shows the log ratio between 2020 and 2016 after these proposed corrections have been applied. The data are nearly flat and show the low variance associated with an incumbent election. Philadelphia county (43% African American) shows up as a significant outlier. This is consistent with the increased minority support for Trump in 2020 indicated by the FL data.
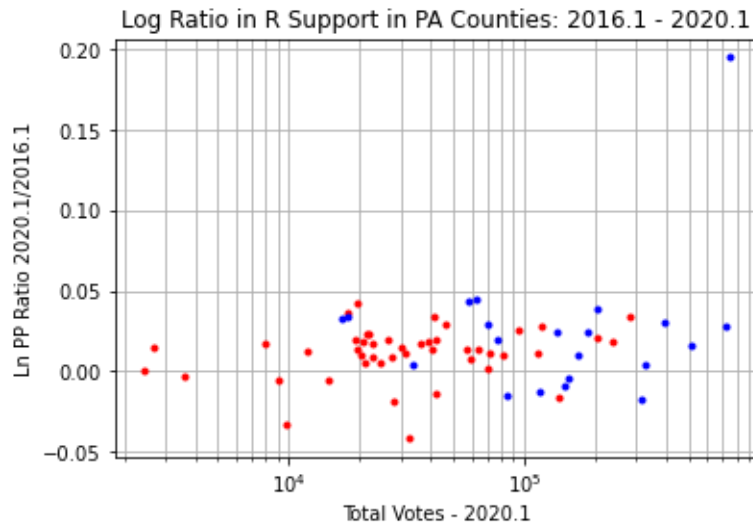
Figure 69. Adjusted Log Ratio of Republican PP between 2016 (adjusted) and 2020 in PA

Figure 70 shows the original and adjusted differential percentage points between 2016 and 2020. The trend in the data has been effectively removed by the adjustment, and the data appear reasonable.
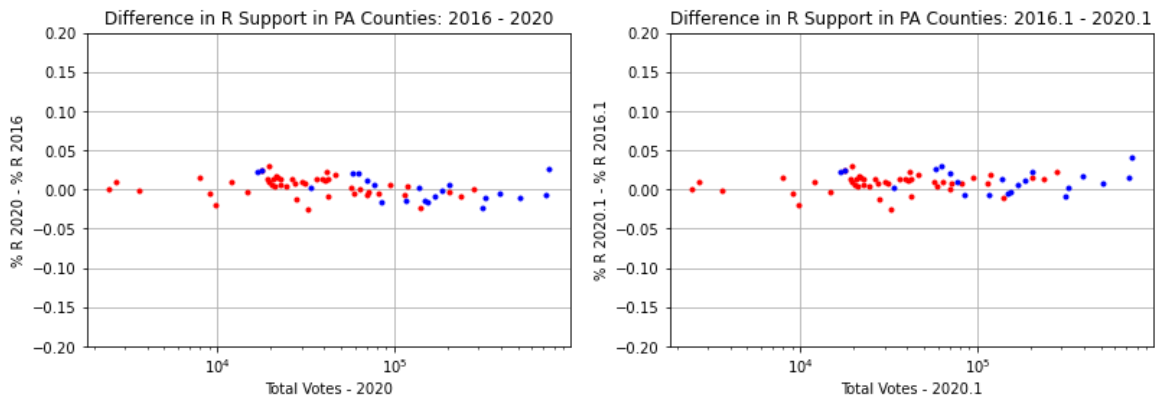
Figure 70. Original (left) and Adjusted (right) Difference in 2016 and 2020 Elections in PA

Sanity Check

Again, we perform a sanity check by comparing to the unadjusted 2012 election as a baseline. Figure 71 shows that a slope of $k = 0.125$ is supported by the log ratio data between 2012 and 2020.

Figure 71. Log Ratio of Republican PP between 2012 and 2020 in PA

Additionally, Figure 72 shows that the log ratio between 2012 and the 2020 data after adjustment by these parameters appears without trend or slope discontinuity.

Figure 72. Adjusted Log Ratio of Republican PP between 2012 and 2020 in PA

Figure 73 shows the differences in percentage points between 2012 and 2020 after adjustment. The adjusted data have not strong trend and indicate an overall shift of nearly 10 pp toward Trump compared to Romney.
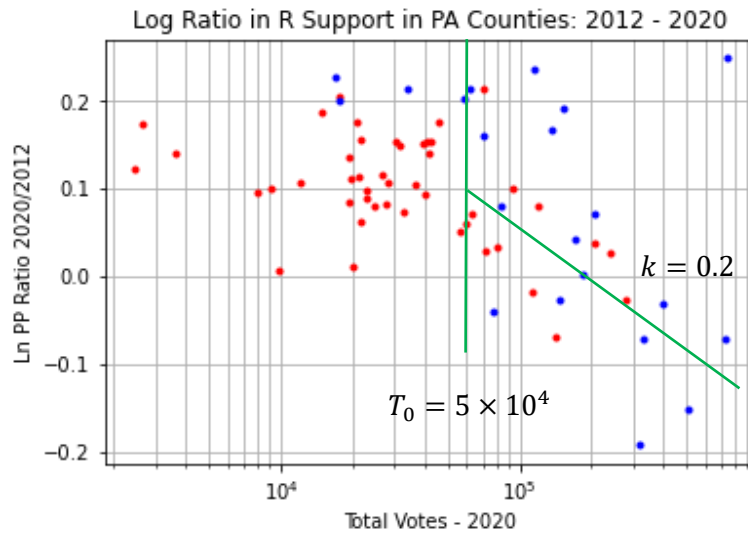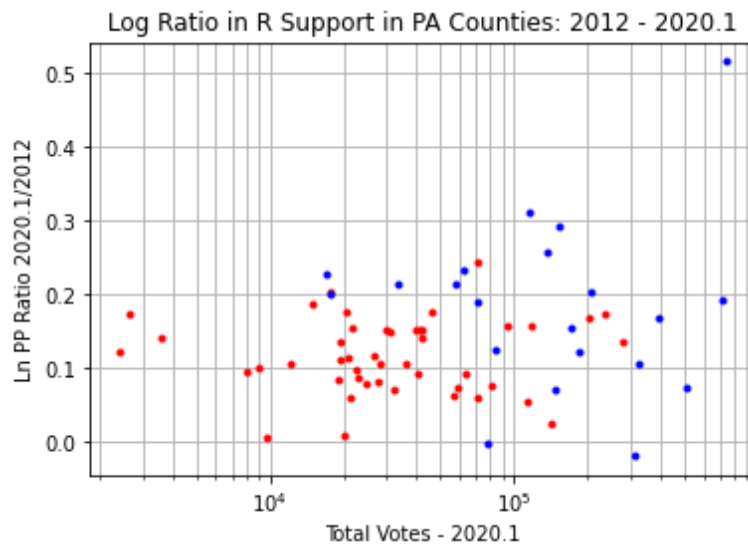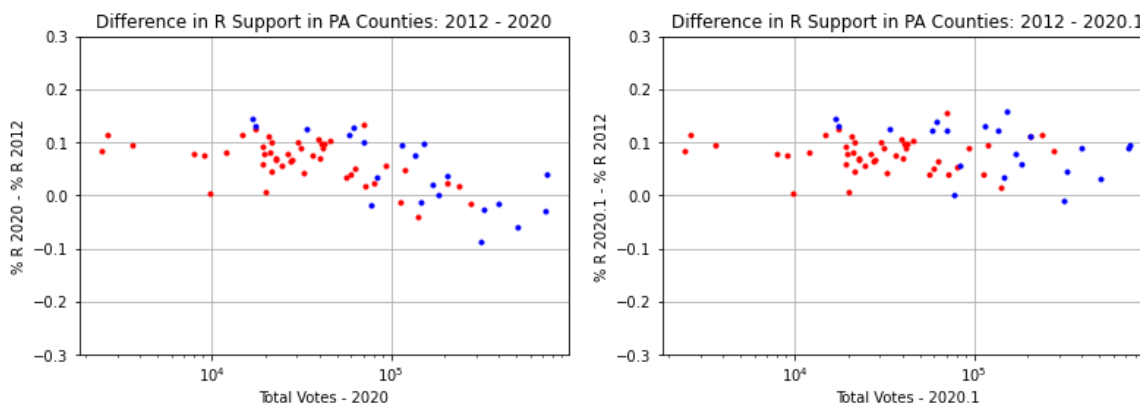

Figure 73. Original (left) and Adjusted (right) Difference in 2012 and 2020 Elections in PA

Summary Results

Table 5 summarizes the overall effect of these corrections. The proposed corrections turn the 2020 election into a clear win for Trump with 55% 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	50k	0.175	294k	44.3k	48.6	53.4
2020	50k	0.20	430k	-81.7k	48.8	55.0

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.

2016

As in OH and PA, VA did not exhibit anomalous data previous to 2016. Figure 74 shows the log ratio analysis for VA 2012 vs 2016. The graph suggests a slope discontinuity point of $T_0 = 10^4$ and a slope around $k = 0.1$.
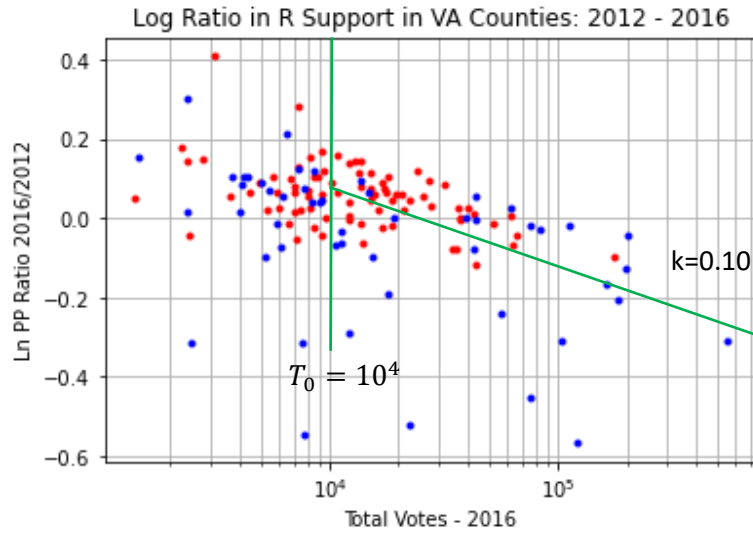
Figure 74. Log Ratio of Republican PP between 2012 and 2020 in VA

Figure 75 shows the log ratio between the 2012 and adjusted 2016 data. The data appears nearly flat and has no apparent slope discontinuity. There appears to be a slight residual negative slope, but its magnitude is well within the variance of the data. Therefore, we move forward with these estimates.

Figure 75. Adjusted Log Ratio of Republican PP between 2012 and 2020 in VA

Figure 76 shows the original and adjusted differential percentage points between the 2012 and 2016 elections. The trend in the R and D clusters has been removed. The data show VA as a highly divided state with widely varying opinions of Trump.
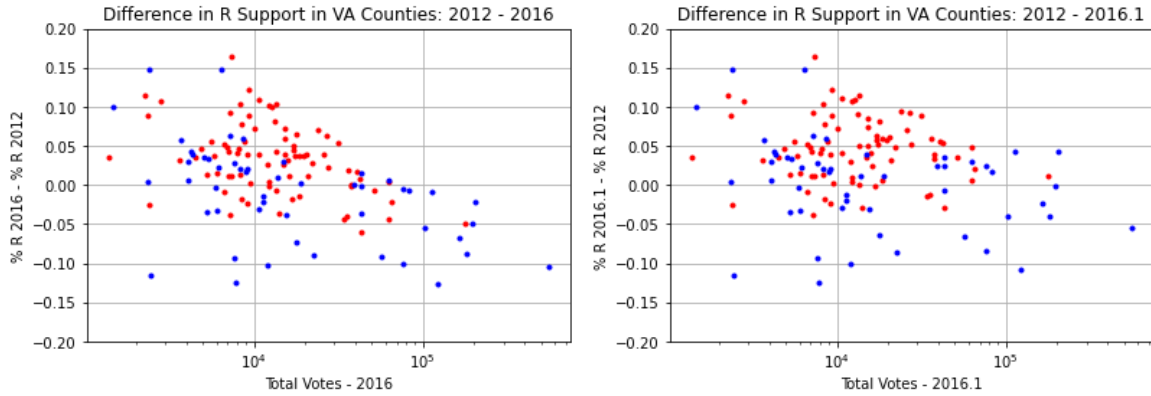
Figure 76. Original (left) and Adjusted (right) Difference in 2016 and 2020 Elections in VA

2020

In 2020, the increase in the anomaly in differential data compared to 2016 (Figure 8) appears slightly stronger than in other states. The log ratio comparison of 2020 to the adjusted 2016 data supports an increase of the slope to $k = 0.15$.
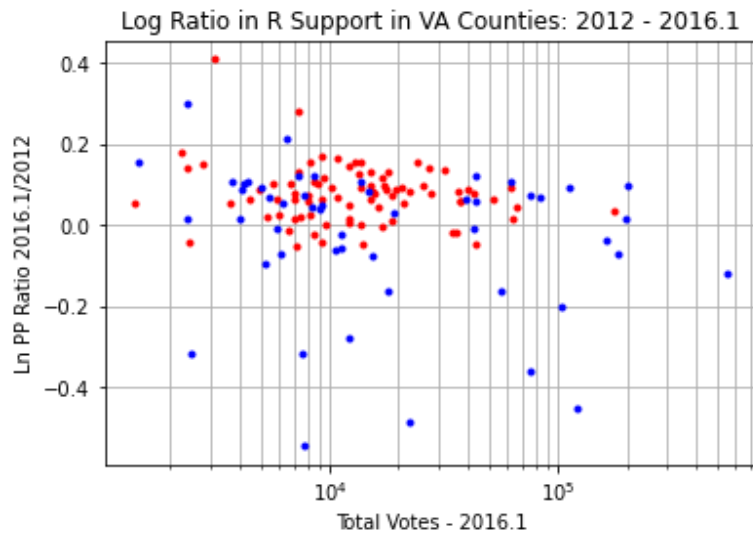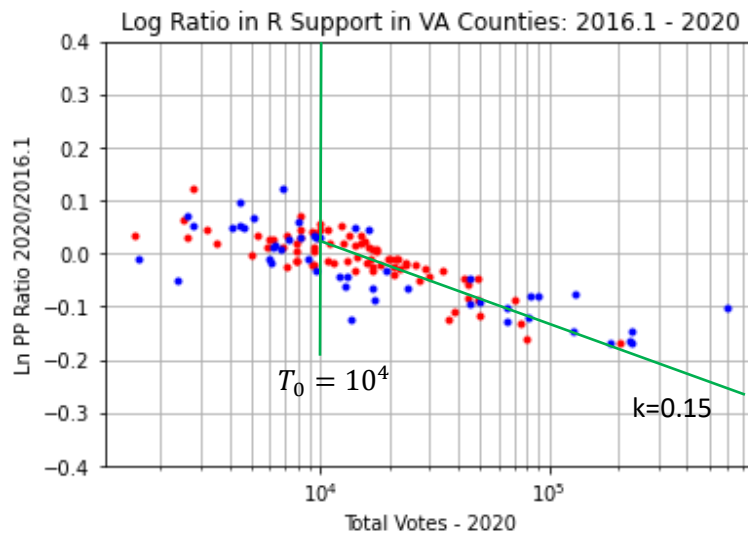

Figure 77. Log Ratio of Republican PP between 2016 (adjusted) and 2020 in VA

Figure 78 shows the log ratio between 2016 and the adjusted 2020 data after the estimated parameters are applied. The overall trend is flat, however Fairfax county is a significant outlier. Fairfax county does not have a large AA or Hispanic minority population, so the nature of this outlier is unknown.
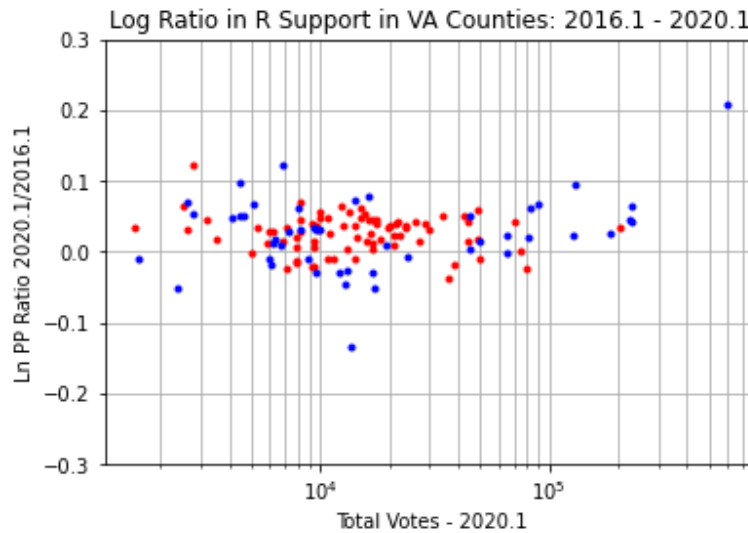
Figure 78. Adjusted Log Ratio of Republican PP between 2016 (adjusted) and 2020 in VA

Figure 79 shows the original and adjusted differential percentage point data between the 2016 and 2020 elections. The adjusted data has removed the sloping 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.
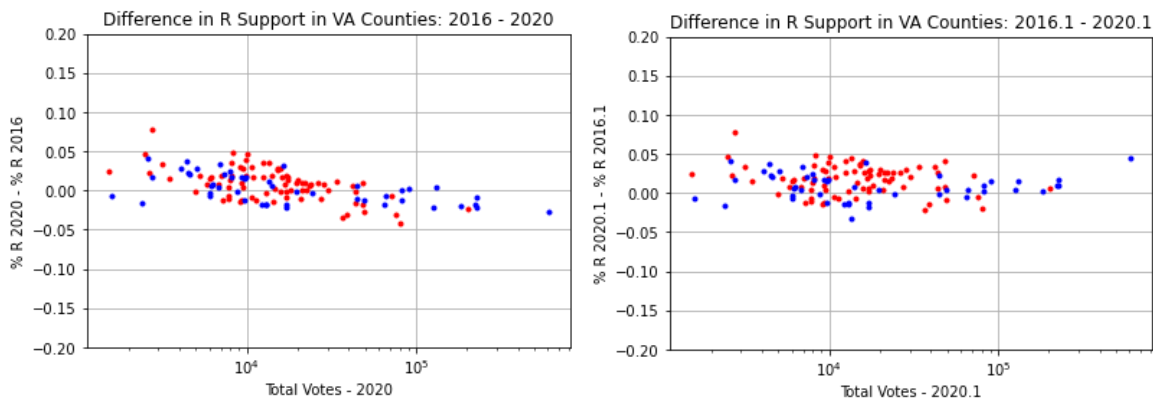

Figure 79. Original (left) and Adjusted (right) Difference in 2016 and 2020 Elections in VA

Sanity Check

As in the other states, we compare the final results to an unadjusted election data set. Again, 2012 was used as the baseline election. Figure 80 shows that the estimated slope parameter of $k = 0.15$ is

supported by the data. (In fact, the slope could be higher, but we leave it here for a more conservative analysis).
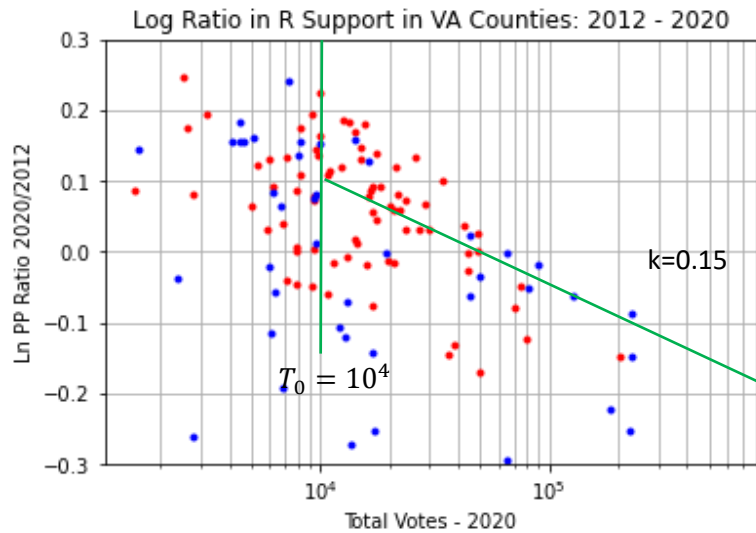
Figure 80. Log Ratio of Republican PP between 2012 and 2020 in VA

Figure 81 shows the log ratio between 2012 and 2020 after applying the adjustments to 2020. The graph appears nearly flat. (There might be a small negative slope remaining).
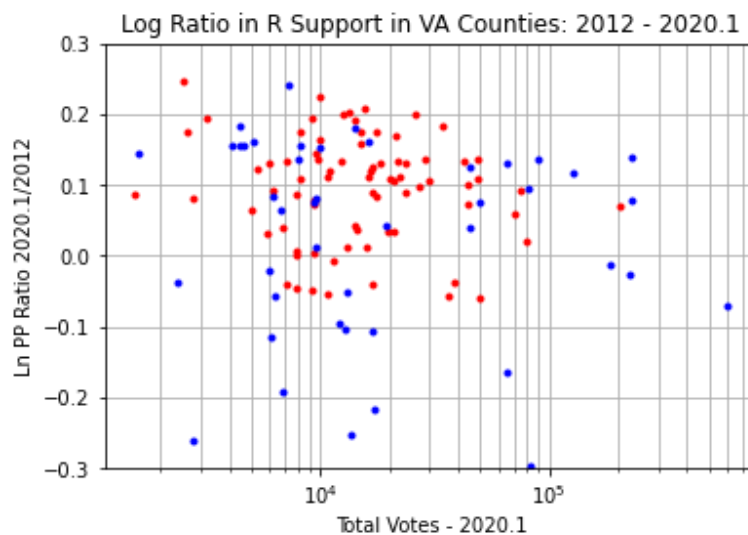
Figure 81. Adjusted Log Ratio of Republican PP between 2012 and 2020 in VA

Figure 82 shows the differential percentage points between 2012 and 2020 for the original and adjusted data. The trend in the original data has been effectively removed. Notably, Fairfax county does not appear as an outlier in this data set, which tends to suggest that the outlier in the comparison between 2016 and 2020 is due to an anomaly in the 2016 data (which is not involved in Figure 82).
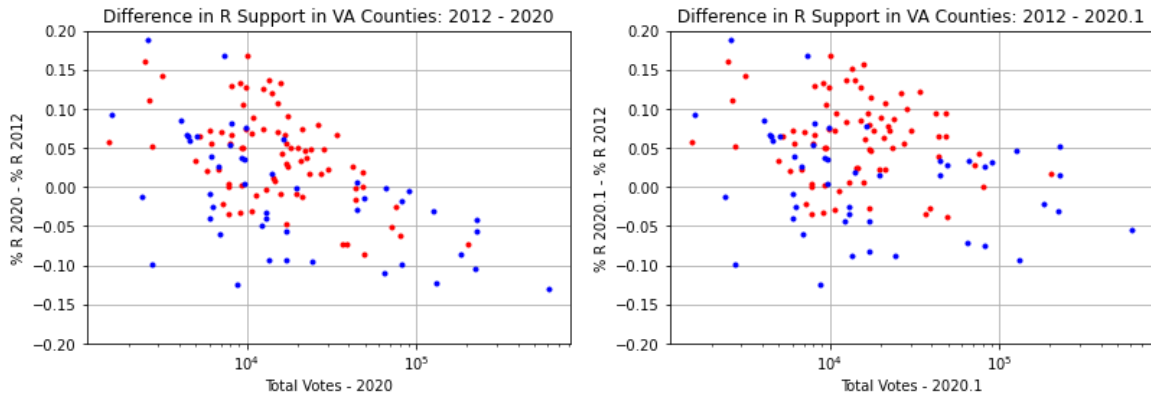
Figure 82. Original (left) and Adjusted (right) Difference in 2012 and 2020 Elections in VA

Summary Results

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. Since the parameter estimates are determined by approximate graphical techniques, they 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	Switched Votes	Official R Vote Margin	Official R %	Predicted R %
2016	10k	0.10	148k	-212k	44.4	48.2
2020	10k	0.15	279k	-451k	44.0	50.3

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 83 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 moot.
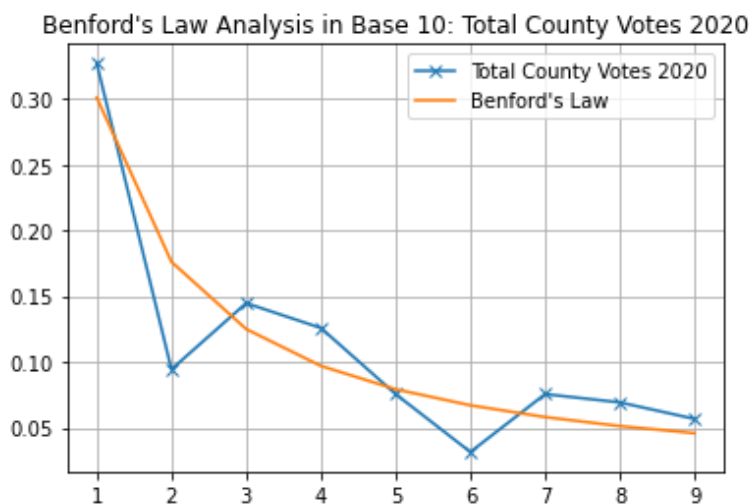
Figure 83. Leading Digit Distribution of Total County Votes in GA 2020

Figure 84 shows 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 83. 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 84. Leading Digit Distribution of Unadjusted and Adjusted R Votes in GA 2020

In contrast, Figure 85 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.
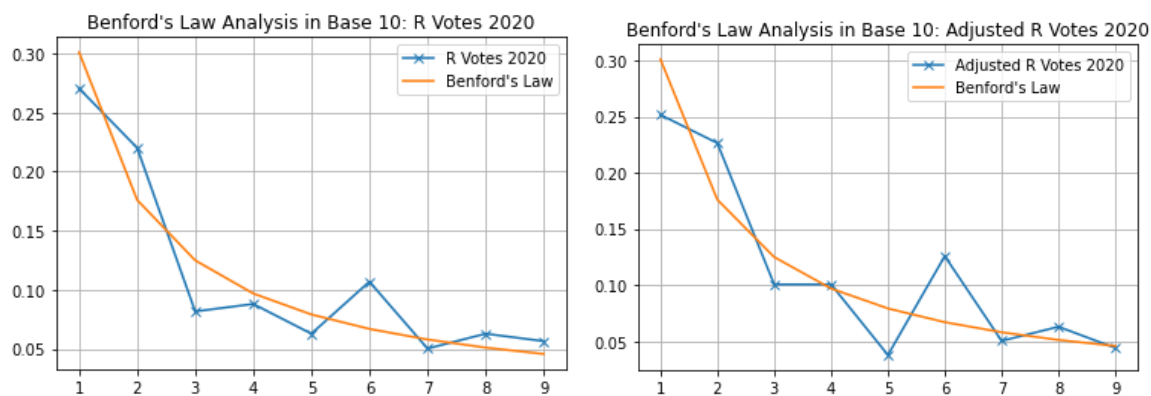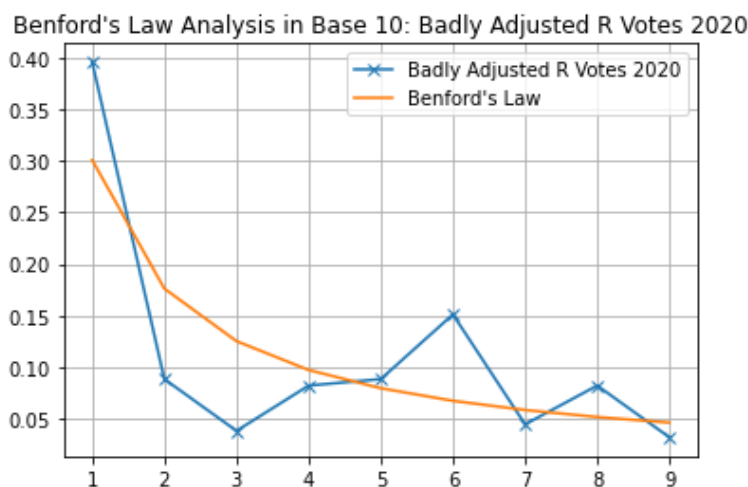

Figure 85. Leading Digit Distribution of Contrived Adjustment to R Votes in GA 2020
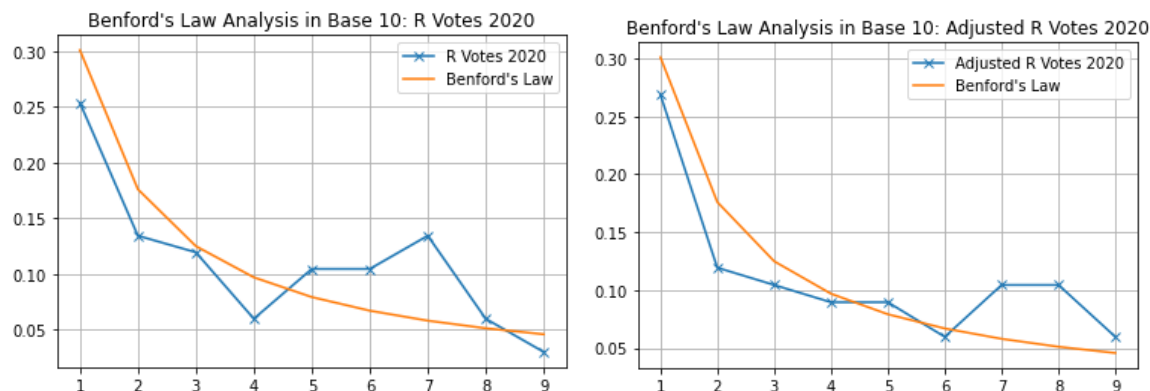
Figure 86. Leading Digit Distribution of Unadjusted and Adjusted R Votes in FL 2020

In order to check our assumption further, Figure 86 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 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 anomalies exhibit a common slope discontinuity point of 10,000 voters, except in FL where two points of discontinuous slope were observed in 2012 and 2016, and in PA, where the discontinuous slope parameter is 50,000 voters. (Moderate confidence, high for GA)
- 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.

Cross-Examination

In this section, the author has attempted to address some anticipated concerns regarding this report. The reader may judge for themselves the strength of the straw men erected here.

You voted for Trump and are a naked partisan. Why are you spreading disinformation propaganda on behalf of a racist tyrant intent on discrediting American democracy for his own personal gain?

I have tried my utmost to make this report as transparent as possible to avoid any hint of disinformation. All of the raw data and analysis code I used to create the report have been made freely available for you to review, tweak, and revise as you see fit. I encourage every reader to do this and confirm my results. I have done most of the hard work for you, spending many hours of my own, uncompensated time to do so in a holiday season. (I thank my wife for her patience in not turning me out of the house for neglect.)

The allegations being made by various parties about voting in America are highly disturbing. If true, they represent a stunning failure of both major political parties to secure our elections and a clear threat to American democracy which is well beyond partisan politics or the welfare of any particular candidate. They should be seriously considered, **but by no means taken at face value**. This is the purpose of this report, to provide a (as transparent as possible) "gut check" as to the feasibility of what other parties are presenting as evidence. **By itself, it is not proof of anything**. However, it is circumstantial evidence

which may be combined with other forms of direct evidence to build a case, either in a court of law, or in the minds of our fellow countrymen.

The results of this report indicate that it is plausible that an algorithm which changes votes from R to D has been in operation since 2008 in some states and since 2016 in all states examined. This is well before the current election, and well before Donald Trump (or related parties) would have any opportunity to plant evidence and attempt to discredit election results for their own gain.

Why are the trends considered anomalous? Haven't certain states and areas become more progressive over time? Why shouldn't larger counties vote more Democratic?

There are a few major effects that can shift support for political parties within a given state. One is turnout. If one assumes an increased enthusiasm for voters of one party over another, this will result in a proportional increase in support for that party over the other. This is precisely the kind of effect which produces nearly constant offset in the log ratio plots across all counties**.** A prime example of this is in the 2004-2008 OH data, which is indicative of increased Democratic turnout for Obama. (Note that overall turnout increases or decreases common between the parties should produce almost no effect in the log ratio, as the proportional effect is the same.) A related effect is increased voter registration % among a given political party (assuming similar widespread application across the state in question).

Another major effect is the realignment of certain demographic groups. These effects are evident in the 2016 and 2020 FL data. In these cases, retirees and (AA and Hispanic) minorities appear to have both shifted greatly for Trump compared to Romney. This produced clusters of counties that shifted with a similar number of pp in the first case, and in the second case, isolated counties that were outliers due to concentrated presence of the minorities in question.

A third effect is the migration of progressive population to urban areas. This would indeed change the voting patterns of these areas. However, in the states under examination, the urban counties are a small fraction of those examined. (Such patterns might have a significant effect on the overall vote, but not on the log ratios or differential percentages).

Furthermore, the states in question are very different demographically. GA has significant AA population throughout the state (with 20/159 counties above 50% AA population). FL has diverse demographics that are geographically corralled. OH is nearly 80% white. PA has significant AA population only in Philadelphia. One would not expect common, systemic trends to exist between these states over multiple elections.

Although I am not a demographer or political scientist, I am unaware of any demographic effect that produces trends in political alignment that are linearly correlated to the log of the total number of voters in the county involved and/or should produce a linear slope on a log ratio plot. These kinds of trends, especially given their replication across multiple states and election cycles, are highly indicative of an artificial process. In particular, the piecewise nature of the trend (clearly in GA and somewhat less clearly in other states) is to me the "smoking gun" which suggests artificiality. This behavior is extremely difficult to explain through a natural demographic process.

You just adjusted slopes on a graph till you thought they were flat. How do you know the numbers are right? How does this prove anything?

I have never claimed in this analysis that my parameter estimates were "correct", or that even the algorithm proposed is "correct". (Indeed, if such an algorithm were in operation, it would likely contain random elements in addition to the deterministic adjustment.) The point of the analysis is to show that the data are not inconsistent with the operation of an algorithm similar to the one proposed. This is demonstrated by estimating "reasonable" parameters for such an algorithm and showing that the data resulting removal of the effect is not nonsensical. Furthermore, the summary data illustrate that the effect of such an algorithm would not be slight, but in the 100s of thousands to millions of votes. Certainly, the specific numbers presented are based on speculative analysis.

Again, this report does not prove the existence of anything, it merely suggests that a vote-changing algorithm that operates in a similar way is not inconsistent with the facts. Additional, direct evidence is required to make a proper case for the existence of such an algorithm.

The reader may wish to view the results in the spirit of the popular show "Mythbusters". The analysis in this report indicates the notion of a vote changing algorithm is "plausible" rather than "busted/confirmed".