ENSEMBLE ANALYSIS OF FLORIDA STATE LEGISLATIVE DISTRICTS

RYAN BAUSBACK AND KEVIN KNUDSON

Abstract. In an effort to determine whether or not the Florida state legislative district maps form a partisan gerrymander we generate several hundred thousand valid plans and simulate elections using actual voter tabulation data from various 2016 and 2018 contests. Our results show that the current maps are an extreme outlier in the distribution of results.

1. Introduction

The advent of increased computing power has ushered in a new era of precision in drawing legislative maps, often to partisan ends. The most egregious examples, such as the current North Carolina Congressional district map, have led to lawsuits contesting their constitutionality. The July 2019 ruling by the U.S. Supreme Court, decided by a 5-4 margin, renders such claims non-justiciable in federal courts, effectively kicking the problem back to the states.

The North Carolina case is notable for its use of ensemble analyses of valid districting plans. The idea is simple enough: to determine if a particular plan is reasonable one generates a large number of valid plans, tallies the wins for each party using prior election data, and then compares the number of wins for the proposed plan to the estimated distribution. If the plan is an extreme outlier then it may be a partisan gerrymander. The current North Carolina plan, with its 10-3 advantage for the Republicans, is indeed an outlier; the expected distribution is 7-6 Democrat [1].

In this note we perform an ensemble analysis on the Florida state legislative districts. Florida is a rather "purple" state–most statewide elections are closely contested and are often decided by a percentage point or two. The state has 27 Congressional districts; in 2016 the election outcome was 16-11 R/D and in 2018 it was 14-13 R/D. Given the statewide votes in those years these distributions are fairly reasonable.

The state legislature is a different story, however, especially when it comes to the Florida State House of Representatives. There are 120 seats in the body and after the 2018 election the distribution of seats is 71 Republican, 47 Democrat, 2 vacant. On the other hand, the 2018 governor and U.S. Senate elections were extremely close (the latter required an automatic recount), which suggests that the current State House districts might be gerrymandered. Our calculations provide evidence for this–a typical random districting plan could be expected to yield 56 Democratic representatives. So it seems that the Republicans may have a slight built-in geographic advantage, but that a close statewide election should yield a more equitable distribution of seats in the State House.

Date: August 20, 2019.

The results presented here are not without some issues. As we discuss in Section 2, the precinct shapefile we used as a starting point had some serious problems. This forced us to begin our Markov chains with a random districting plan instead of the current one. Given the enormous size of the space of possible plans, we could very well be generating partitions that are nowhere near the current state of affairs. However, we ran the simulations several times, on two different computers, with different starting points each time, generating 100,000 plans in each run. All of them satisfy compactness and contiguity constraints and are within 2% of the ideal population distribution, just like the current plan. In all cases we obtained similar outcomes; typical distributions are shown in Figures 2-5 below.

We also studied the State Senate. The current distribution of seats is 23-17 R/D while our simulations suggest that 21-19 or even 20-20 is the more likely outcome.

This work was carried out in the summer of 2019 while the first author was a member of the University of Florida's SURF program. We thank the university for its support.

2. Data cleaning

Data cleaning was the most tedious part of performing an ensemble analysis of the Florida House of Representatives, given the nature of the precinct shapefile we used. This shapefile, which we acquired from the Florida Science Election Team, suffered from a number of crippling issues including innumerable overlaps and islands, polygon segments with no area, and non-contiguous precincts [2]. All of these issues had to be addressed before any analysis could begin.

The zero area segments and disconnected points from the precinct polygons were eliminated using QGIS software. By performing a cycle of using the "Check Validity" tool to determine the current number of errors followed by the "Fix Geometries" tool, these issues were whittled down to a more manageable amount. However, even after all of the errors caught by "Check Validity" such as duplicated points were fixed, many of the polygons were still labelled as invalid. A scan of all of the nodes of the problem precincts revealed that QGIS had failed to identify numerous floating points, which had to be dealt with by hand. This only partially fixed the issues, since many of the precincts such as those in the Florida keys had large pieces that were completely disconnected from each other.

In order to avoid redrawing the entire map of the Florida keys, we enlisted the help of Daryl Deford from the Massachusetts Institute of Technology (MIT), who informed us of a much more efficient solution. We could import the shapefile into Python using Pysal. This package allowed us to create a listing of "Rook-style" contiguity weights based on the distance of each of the precinct polygon's centroids from each other [8]. The list could in turn be used to construct a NetworkX graph whose edges could be more easily manipulated to get the shapefile into working order.

Another roadblock to performing the analysis was adding population data to the shapefile. Census data can only be acquired at the block level and therefore has to be aggregated up to the precinct level. This is done by comparing the overlapping areas between precinct-level and block-level shapefiles. We downloaded the decennial census data for 2010 for each county one at a time (since there is no option for full state download) [4]. We then organized it by comparing the GeoIds to the TIGER/Line block level shapefile ones in Excel using the "vlookup" function, before combining them in QGIS [3].

In order to actually perform the aggregation, the GeoPandas function sjoin was first used to label each block based on which precinct contained it. Because of the major problems with the geometry of the precinct shapefile, the normal assignment of blocks to precincts using maup.assign(blocks, precincts) could not be performed [5]. Instead, we devised a workaround using the same principles with the help of Ruth Buck and colleagues at the Metric Geometry and Gerrymandering Group (MGGG) [6]. The solution involves using the "maup.intersections()" function to compute the overlaps between the output of gpd.sjoin and the precinct shapefile. Because the blocks also do not nest nicely, the coverage ratio for each element was then calculated by dividing the area of the intersections by the area of the blocks. This was then used to create the Pandas series assigning each block to a precinct, which was then aggregated together using the normal Pandas "groupby(assignment).sum()" operation. Further discrepancies occurred during this process since a number of the precincts were smaller than the blocks they overlapped, meaning the blocks were assigned to larger precincts. After further research, it was determined that those precincts consisted of election offices that handled absentee ballots, meaning they were non-residential. The population of those precincts was therefore set to zero.

Even after all of this, however, there are still some issues with the maps. There are three noticeable holes in the plans (see Figure 1). Two of these, in the southern part of the state, correspond to Lake Okechobee and to the Everglades National Park. Given that there are no voters in these areas this is no problem. There is an additional hole in the panhandle of the state, between Tallahassee and Pensacola. This is a low-density area with few people, so its omission does not affect the results of our simulations significantly. Still, our estimates should be seen as just that–an approximation to what is really going on in the state.

3. 2016 ELECTIONS

To perform the ensemble analysis, we determined that a random-walk approach to running the Markov chain would be most effective due to time constraints. A random initial plan was generated using the Gerrychain function recursive_tree_part(). Then, 100,000 plans per election were generated using a technique known as "recom" (An example plan is in Figure 1). Recom involves combining two existing districts together and then randomly splitting them according to the precinct boundaries. However, any new districts created, as well as the initial plan's districts, must satisfy a number of pre-defined constraints. For our runs, we used 1) the contiguity constraint, 2) a compactness constraint based on the compactness of the initial districts, and 3) a population constraint within two percent of the ideal population (total population divided by 120 districts). These are the most basic criteria for a districting plan under Florida law and will be known as the *normal constraints.*

Two different 2016 elections were used in the ensemble analysis: 1) the presidential race between Donald Trump and Hillary Clinton, and 2) the United States Senate race between Marco Rubio and Patrick Murphy. Florida House of Representatives races were not used since many incumbents ran unopposed, resulting in there being no election data for more than a few races. For each election, 100,000

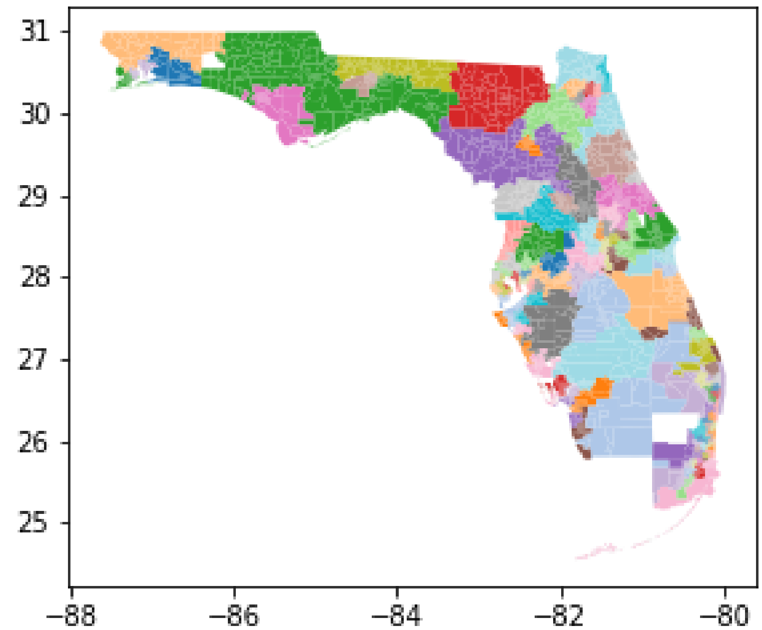

FIGURE 1. A random districting plan for the Florida State House of Representatives

plans were generated under normal constraints and the number of districts won by Democrats tallied for each plan.

A typical 100,000-step run of plans required approximately 4 hours of computing time on a MacBook Pro with a 2.3 GHz Intel Core i5 processor and 16 GB RAM. Interestingly, the corresponding run for State Senate plans took nearly twice as long, even though such a plan has only 40 districts.

As we can see from Figure 2, the 2016 Presidential election data predicts that the Democrats should win somewhere between 54 and 60 seats in the Florida State House of Representatives out of 120 given a random districting plan. The 2016 United States Senate election (Figure 3), where Rubio (R) won by a landslide, predicts between 42 and 49 seats. In reality, the Democrats only won 41 seats in the Florida State House in 2016 [9]. This evidence suggests that the Florida State House districting plan may be gerrymandered in favor of the Republicans, given that the actual number of Democrat seats won is an outlier in both cases.

4. 2018 ELECTIONS

Ensemble analysis was also performed using two different 2018 elections: 1) the Florida governor's race between Andrew Gillum and Ron DeSantis, and 2) the United States Senate race between Rick Scott and Bill Nelson. Both races were extremely close and went to recounts. For each election, 100,000 districting plans were generated under normal constraints and the number of districts won by Democrats tallied for each plan.

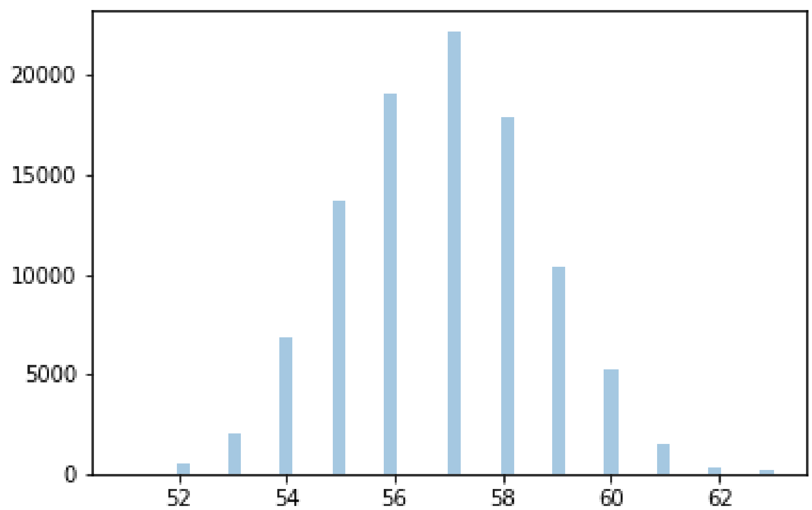

FIGURE 2. Histogram showing the predicted number of Florida State House of Representatives districts won by the Democrats using the 2016 Presidential election data

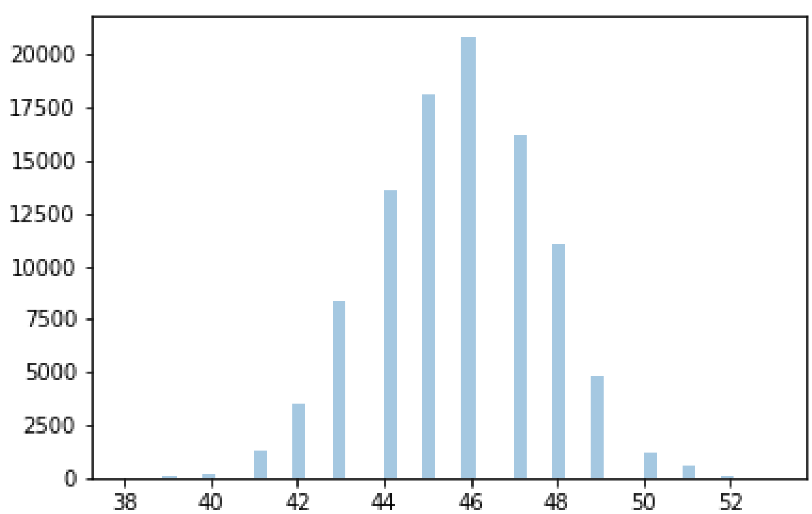

FIGURE 3. Histogram showing the predicted number of Florida State House of Representatives districts won by the Democrats using the 2016 US Senate election data

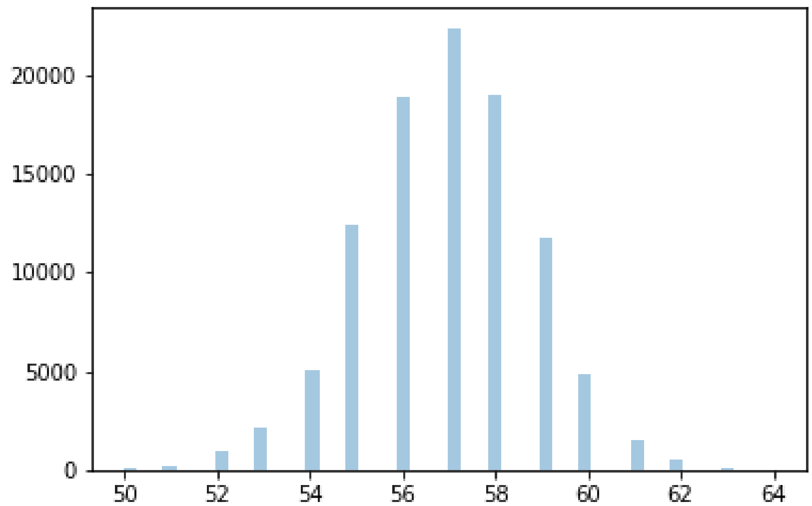

Figure 4. Histogram showing the predicted number of Florida State House of Representatives districts won by the Democrats using the 2018 US Senate election data

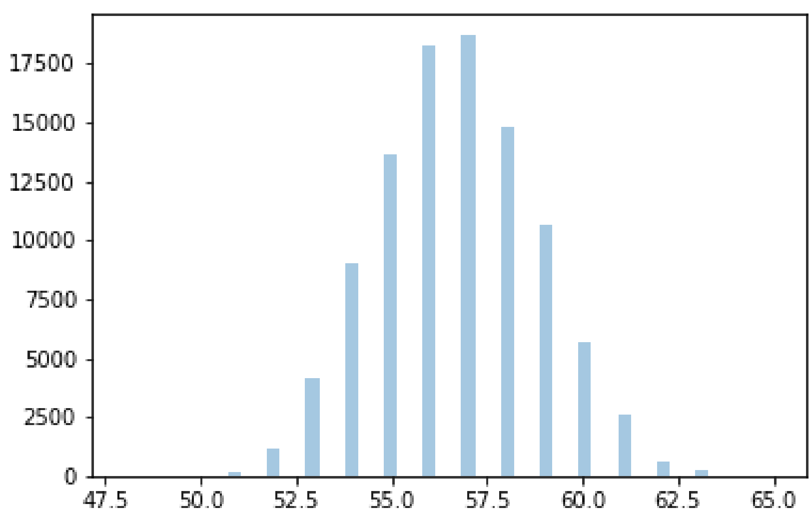

Figure 5. Histogram showing the predicted number of Florida State House of Representatives districts won by the Democrats using the 2018 Florida Governor election data

As we can see from Figures 4 and 5, both elections predict that the Democrats should hold between 53 and 60 seats in the Florida State House given a random districting plan. The Democrats only won 47 seats out of 120 in 2018 [10]. Since this result is an extreme outlier in both cases, it suggests that the Florida State House of Representatives may be gerrymandered in favor of the Republicans. This evidence only becomes stronger when combined with the 2016 election data conclusion and the fact that the districing plan for the Florida State House has not been redrawn since 2015.

5. Ensemble Analysis of the Florida Senate Districting Plan

Additionally, we also performed ensemble analysis on the Florida State Senate using the same "random-walk" method as the one above, with a few modifications. Total population was instead divided by 40 to calculate ideal population, and "recursive_tree_part" was seeded with a range of 40 instead of 120 for a different initial plan. Interestingly, it took almost double the time (8 hours) to complete each 100,000 step run for the Senate compared to the House.

The same 2016 and 2018 elections were also used. Analysis using the 2016 United States presidential election data (Figure 6) suggests that the Democrats should have approximately 19 or 20 seats in the Florida State Senate, while analysis using the 2016 United States Senate election data (Figure 7) suggests that they should have approximately 15 or 16 seats. Given that the Democrats currently hold 17 seats in the Florida State Senate, the presidential election data suggests it might be slightly gerrymandered in favor of the Republicans, while the 2016 US Senate election data suggests that this is not the case.

However, if we also perform ensemble analysis using the 2018 election data, this contradiction is cleared up. Both the 2018 Governor election data (Figure 8) and the 2018 US Senate election data (Figure 9) suggest that the Democrats should have approximately 19 or 20 seats in the Florida State Senate. This indicates that the 2016 US Senate election may be once again underestimating how gerrymandered the Florida State Senate is, given that Rubio won by such a large margin. Overall, the total evidence suggests that the Florida State Senate may be slightly gerrymandered in favor of the Republicans, although not as severely as with the Florida State House of Representatives.

6. Future work

The results presented here are an encouraging first step toward understanding the political landscape of state elections in Florida. As we approach what is certain to be a viciously contested redistricting fight after the 2020 Census it will be important to have solid data. To this end, a better precinct shapefile needs to be built so that a more accurate analysis can be performed. The Florida Legislature, to its credit, has precinct-level election results available on its website going back several years; merging these with a more robust shapefile will then enable a more thorough investigation.

References

[1] J. Mattingly and C. Vaughn, *Redistricting and the will of the people*, Arxiv:1410.8796 (2014), preprint.

[2] Florida precinct shapefile https://bit.ly/30R5lh0

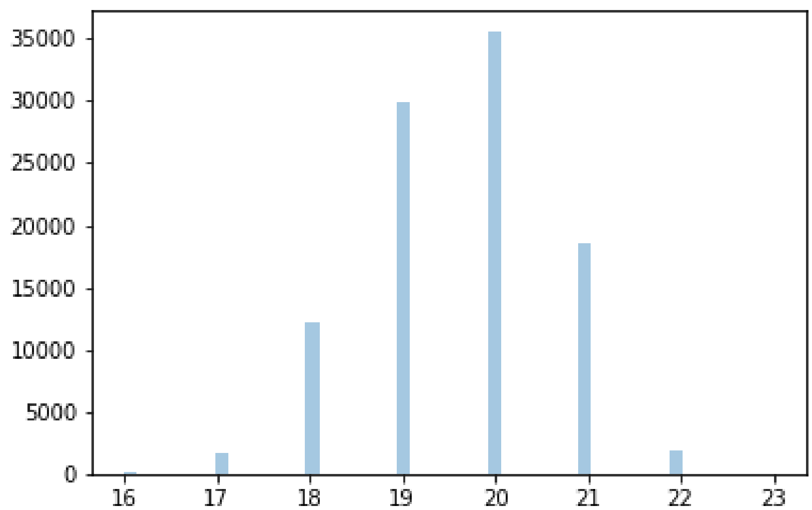

Figure 6. Histogram showing the predicted number of Florida State Senate districts won by the Democrats using the 2016 Presidential election data

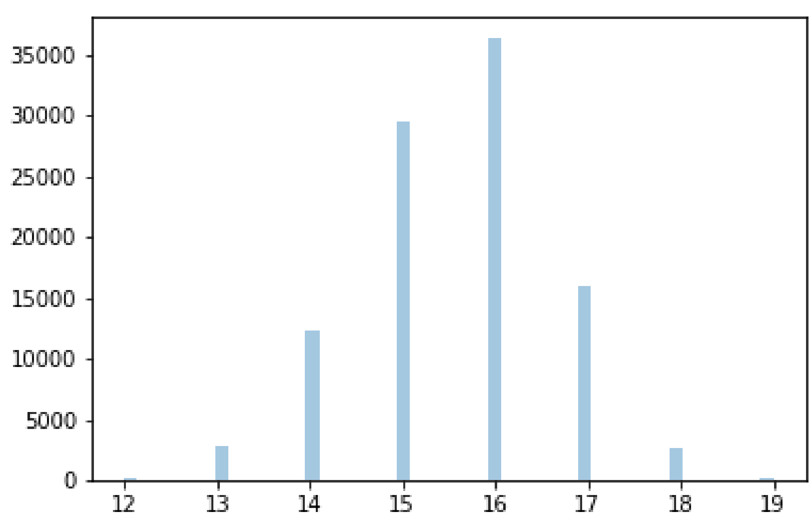

Figure 7. Histogram showing the predicted number of Florida State Senate districts won by the Democrats using the 2016 United States Senate election data

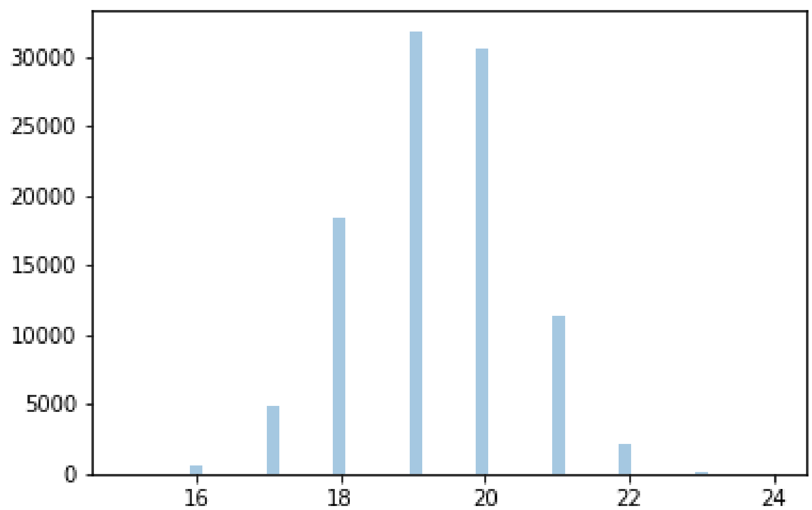

FIGURE 8. Histogram showing the predicted number of Florida State Senate districts won by the Democrats using the 2018 Florida Governor election data

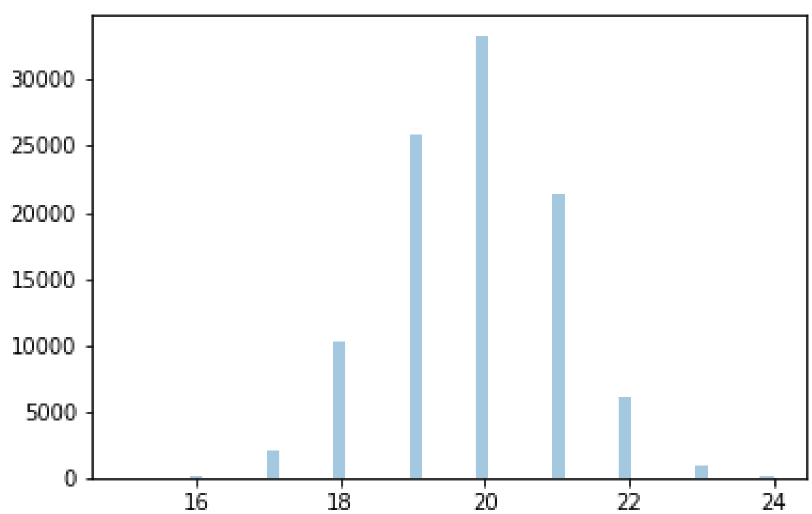

FIGURE 9. Histogram showing the predicted number of Florida State Senate districts won by the Democrats using the 2018 United States Senate election data

[3] Census shapefile https://www.census.gov/cgi-bin/geo/shapefiles/index.php
[4] Census data https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
[5] Maup package https://github.com/mggg/maup
[6] Metric Geometry and Gerrymandering Group https://mggg.org/
[7] GerryChain https://github.com/mggg/GerryChain
[8] Pysal https://bit.ly/2Mjtqsz
[9] Florida 2016 election data https://bit.ly/32TyklX
[10] Florida 2018 election data https://bit.ly/2YoeFXQ

E-mail address: xaalthe71@gmail.com

Florida State University, Tallahassee, FL

E-mail address: kknudson@ufl.edu

Department of Mathematics, University of Florida, Gainesville, FL 32611