Seeking a Spectral Signature for Gerrymandering

Yoni Ackerman

Introduction

In this paper I will be exploring the relationship between state elections and district gerrymandering. Gerrymandering is the practice of carefully drawing state voting districts to dilute the voting power of certain socioeconomic groups. In particular, I would like to know if gerrymandering has a spectral signature in the frequency domain. Because of its political nature, it would be easy to abuse this topic, and let it devolve into a "data-driven" diatribe motivated by some underlying partisan agenda. However, as I have not yet received checks from either George Soros or the Kochs, I will attemp to maintain a reasonable academic distance. I will have at my disposal two main data sources: (1) House and Senate election results for all 50 states between the years 1939-2011, and (2) shape files of every state congressional district for the same period. Both data
cleaning and transformation, which I will spend as little time describing as possible.

With the datasets adequately prepped (an arduous task that I will omit from this discussion), I will apply basic time series methods from (Shumway and Stoffer 2011), in the time domain and a few few from the frequency domain, such as the periodogram and spectrogram. I am not looking for causal relationships, nor am I interested in uncovering a 'smoking gun' of American Election Engineering. It should be obvious (should have been more obvious to me when designing this project) that elections and district drawing are processes far more complex than these data can capture. It would be a mistake on my part to label all gerrymandering as "bad" or some electoral pattern (such as high volatility) as "good". To reiterate, my goal is only to search for frequency-based similarities between state electoral records and relate them, when possible, to the presence of gerrymandering.

State Congress Political Leaning

data source: dataverse.harvard.edu

These are yearly data collected from each state in the union between 1939 and 2015, inclusive. The available data are organized by year, state, and election year, and present 96 other cofactors related to that state's government (house, senate and govenor), in that year. I will almost exclusively use data on the number of Democrats and Republicans in the State Houses during election years. The reason for this is the consistency of House elections. All (around 200) seats in the House are open every two years (as opposed to the Senate, where only a handful of seats reopen each election),thus election results give a much clearer snapshot of a state's political landscape. To get an idea of what these data look like, we can examine how "right" or "left" each state's House. For each state i at year t, define H_{it} = Number of House Republicans − Number of House Democrats. Figure 1 shows H_t for all states.

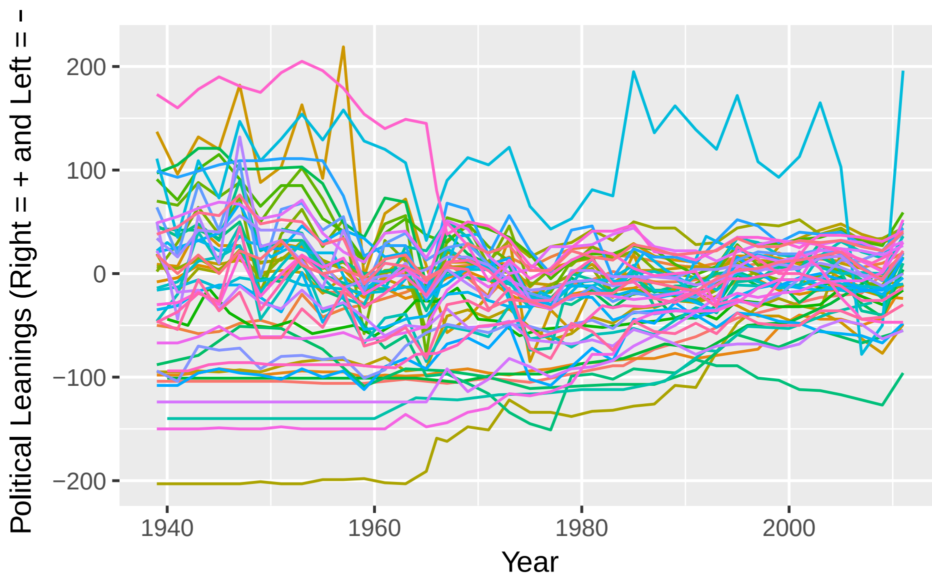

Figure 1: Yearly State House Political Leaning

Given the above plot, and a little thought about how American politics have changed over time, it seems safe to assume that these time series are not stationary. To address this, I deseason the data by taking first differences. Thus, for state i and year t define $X_{it} = Y_{it} - Y_{i(t-1)}$. This variable describes the change in state political leaning per year, and it will be the focus of much of this project. We visualize X_t for a few states in figure 2.

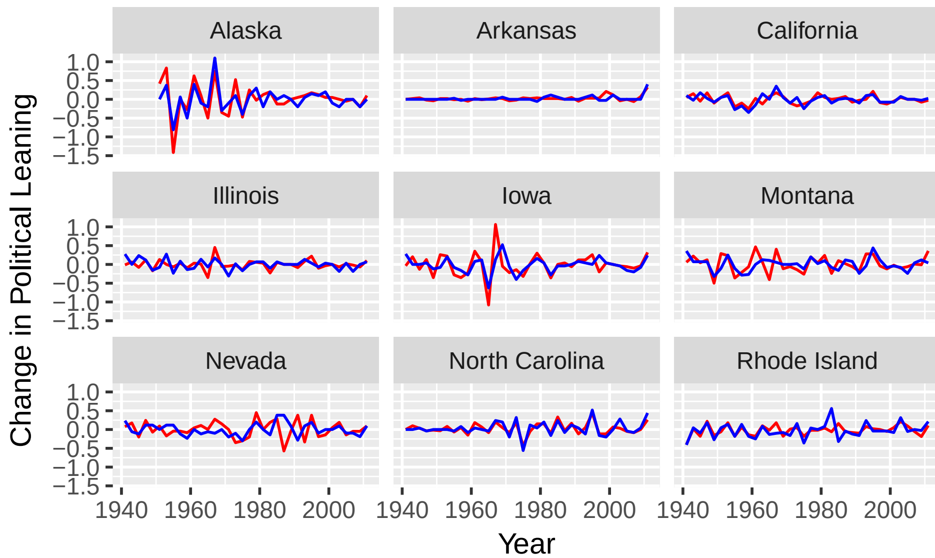

Figure 2: Yearly Change in State House Political Leaning (Blue is Senate, Red is House)

The first thing to do with these data is to examine the linear dependence of the X_t on their lagged values. So I proceed by calculating the autocovariances for each X_t, as well as the equivalent time series for the state senates (just for comparison). The results are summarized in figure 3.

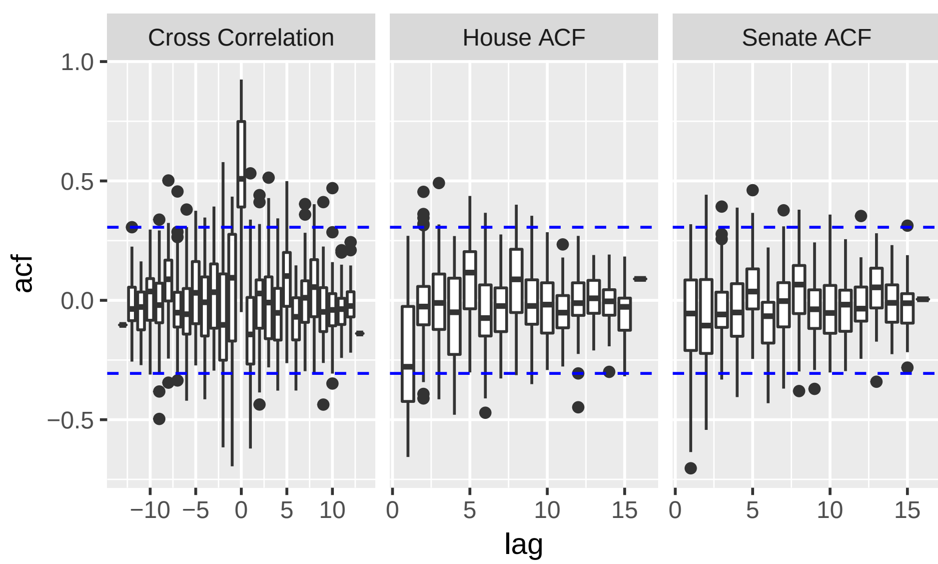

Figure 3: Autocovariance Boxplots

Each box plot represents the interquartile range (plus max and min) of all states' ACF values at the corresponding lag. The blue line represents a 95% confidence interval for significance. There are only two lags which are consistently significant across multiple states (not necessarily all of them): lag 1 in the House plot and lag 0 in the crosscorrelation plot. House elections occur every two years and non-election year data was omitted. This means that many states have a significant, negative correlation in there House voting patterns from election to election. That is, many states have consistently flip-flopped Republican to Democrat and vice versa every election.

Next we look at the State Election data in the frequency domain, the ultimate goal being to find groups of frequencies correlated with the gerrymanding index (to be described later). State periodograms are shown in figure 4.

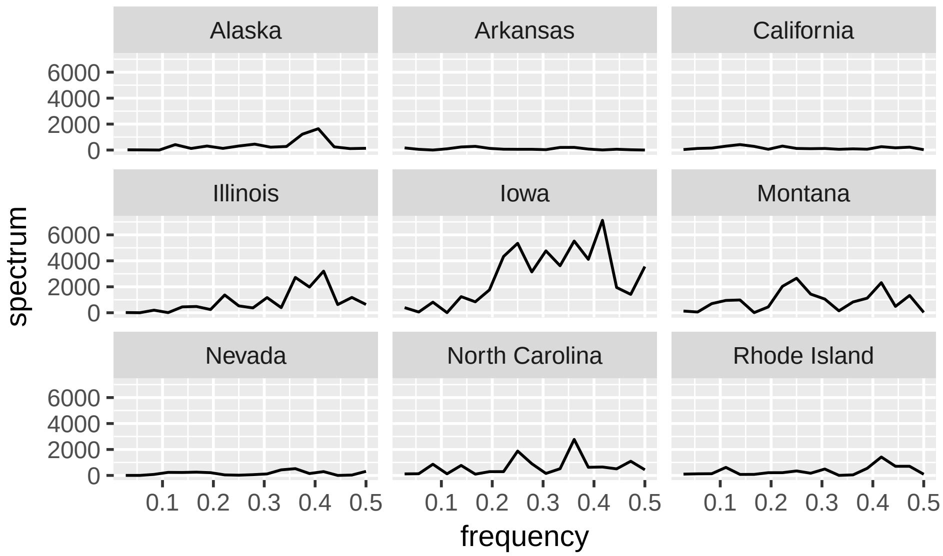

Figure 4: Select State Periodograms

Just like the original first-difference data, state periodograms vary widely. They are determined not just by fickle (if not random) voter behavior, but also by complex historical events. Because of this, it does not seem reasonble to expect to see the effects of redistricting (which changes every 10 years) in state periodograms. Instead, we should be looking at time series on the time-scale that redistricting occurs. The tool that will help us to do so is the spectrogram. The figures 5, 6, and 7 show spectrograms for a few states.

Before moving to the gerrymandering data, there is one last perspective from which I looked at the state election data: potential functions. For this approach, I follow the procedure layed out in (Brillinger 2015). Going back to the original time series H_t of state house data, and adding in its senate equivalet S_t, I look at the trajectory of $(x_t, y_t) = (H_t, S_t)$ for each state. Using these data and their first differences, I estimate the state's potential function, $P : \mathbb{R}^2 \to \mathbb{R}$, by assuming the form:

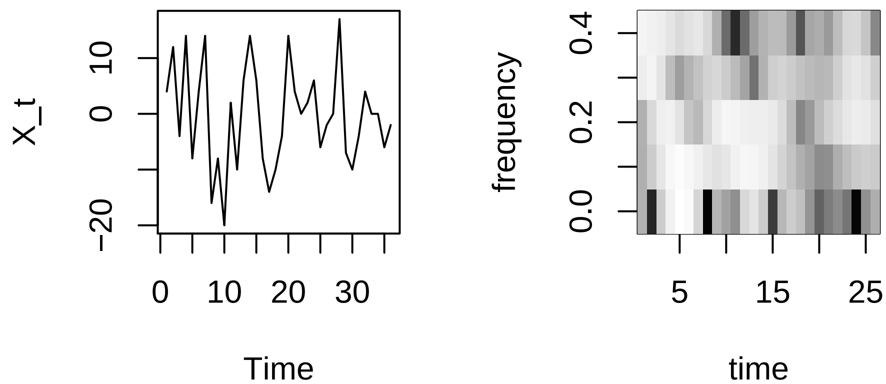

Figure 5: Time Series and Spectrogram of California

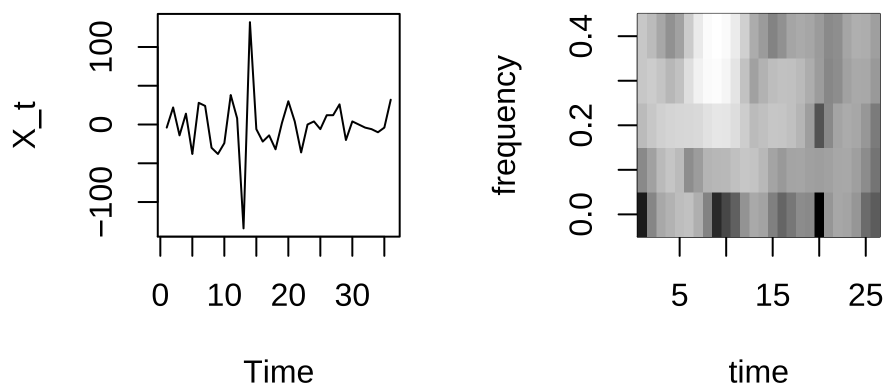

Figure 6: Time Series and Spectrogram of Iowa

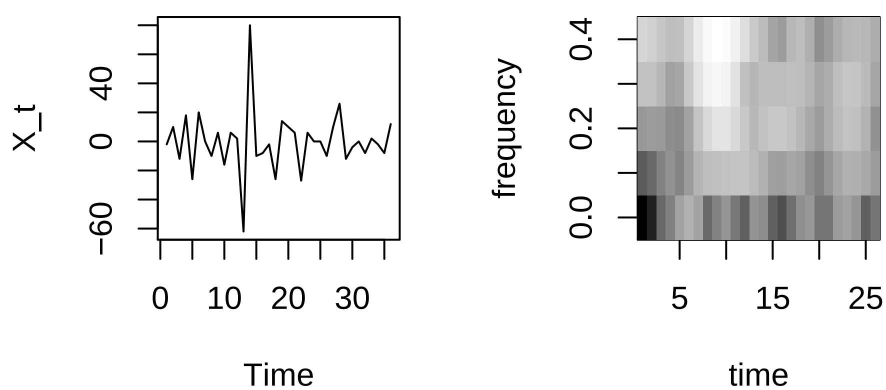

Figure 7: Time Series and Spectrogram of Illinois

$$P(x,y) = \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 xy + \beta_5 y^2 + \beta_6 x^3 + \beta_7 x^2 y + \beta_8 xy^2 + \beta_9 y^3$$

which as has gradient:

$$\begin{bmatrix} 1 & 0 & 2x & y & 0 & 3x^2 & 2xy & y^2 & 0 \\ 0 & 1 & 0 & x & 2y & 0 & x^2 & 2xy & 3y^2 \end{bmatrix}$$

multiplied by:

$$\begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 & \beta_7 & \beta_8 & \beta_9 \end{bmatrix}^T$$

Then using the model

$$r(t_{i+1}) - r(t_i) \approx \nabla P(r(t_i), t_i)(t_{i+1} - t_i) + \Sigma(r(t_i), t_i)\mathbf{Z}_i \sqrt{t_{i+1} - t_i}$$

I performed a linear regresson on the first differences of the (x_t, y_t) trajectories to estimate the β vector.

Plots 8-11 are approximations to the “Political Landscape” of each state. It is interesting to see how different these potentials are. In particular, California’s trajectory appears to exist on a relatively flat potential surface, while the other three state trajectories are on more undulating potential surfaces. If I had more time, I would like to incorporate the gerrymandering data into the potential function’s regression as a dampening effect and examine the differences between the two results. (Note, the banding on the images is an artifact from the plotting tool that was used.)

The next step is to associate the spectra at each time interval with the degree of gerrymandering in that interval, and by insodoing, find some signals. To do this, we should first have a look at the gerrymandering data.

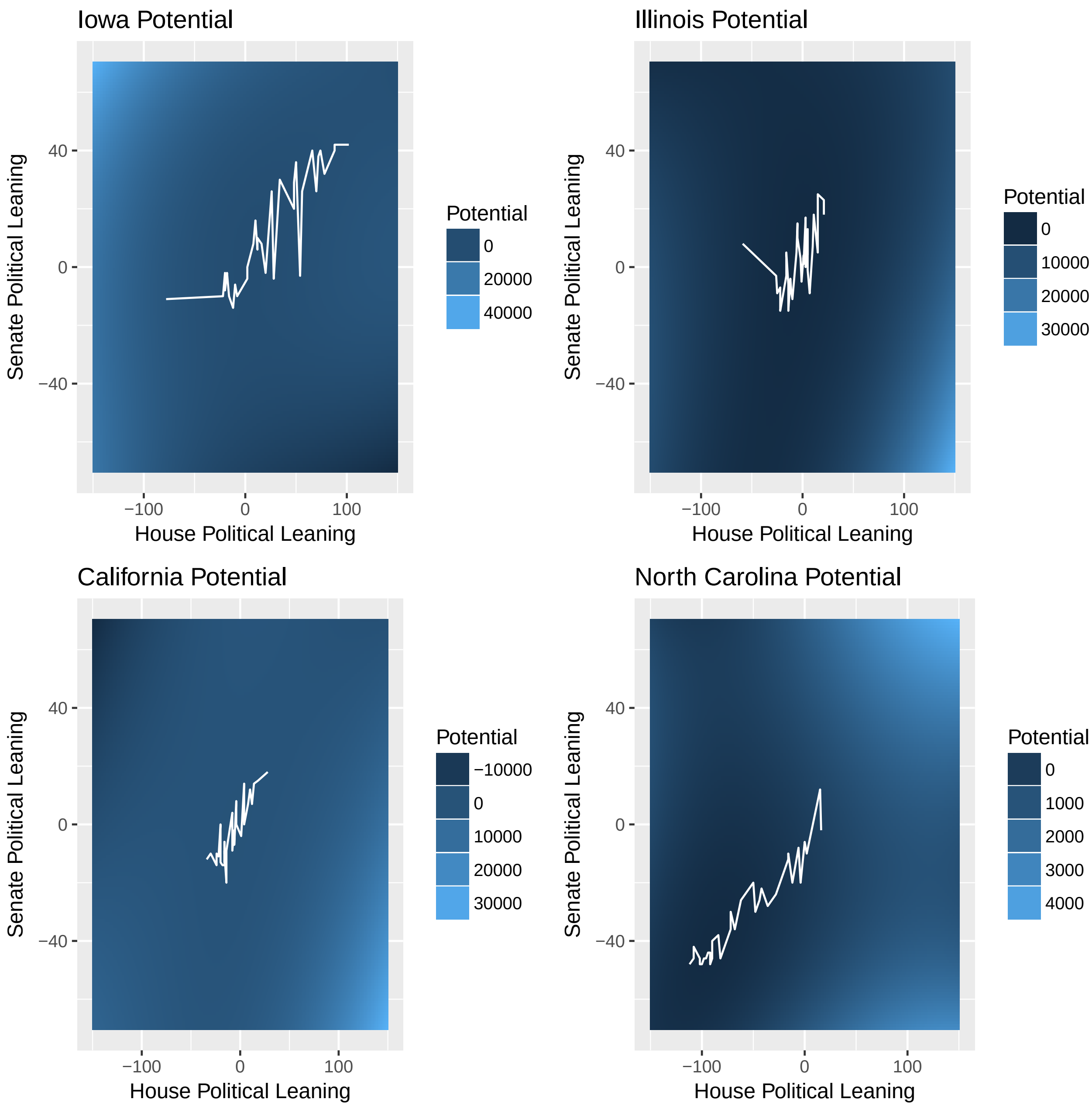

Figure 8: Potential functions and trajectories for selected states

Gerrymandering Data

data source: cdmaps.polisci.ucla.edu

This dataset began as a set of shape files describing the spatial boundaries for every congressional district in the United States since the 1800's. I calculated each state's gerrymandering index in a given year t as follows:

$$g_{s,t} = \frac{\sum_{d \in \text{Districts}} h_{d,t}}{\sum_{d \in \text{Districts}} a_{d,t}}$$

Where g_s is the gerrymandering index for state s, h_d is the area of the convex hull of district d, and a_d is the area of district d. Figure 9 shows time series of this variable.

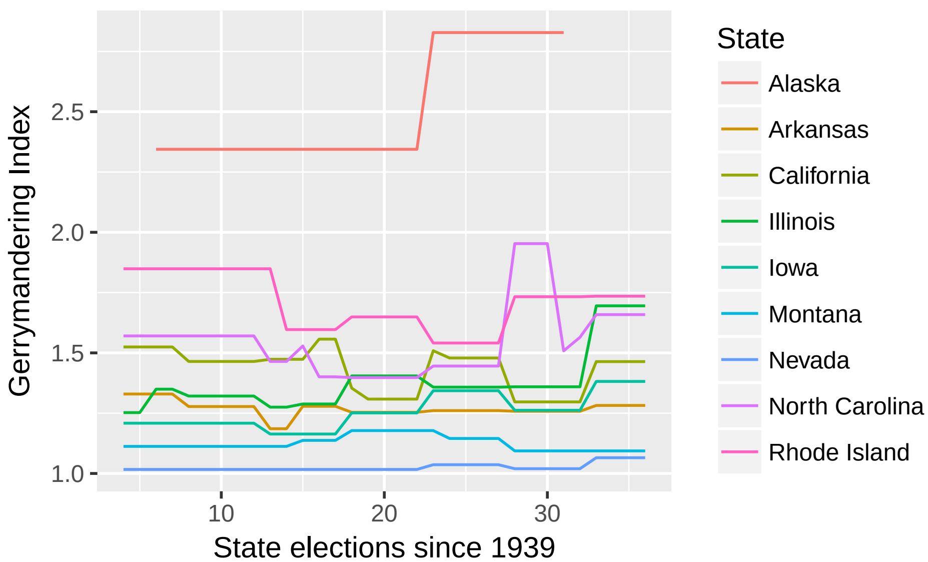

Figure 9: Time Series of Gerrymandering Index for select States

Technically, these data should be step functions that change every 5 elections (districts are redrawn every 10 years). A couple of things could be behind the discrepancies: 1) special redistricting that I am not aware of or, more likely, 2) something changed with the data-collection, measurement, or input processes behind these data. Either way, I will proceed with the data as they are - there simply wasn't time to go digging to deep.

Figure 10 shows the empirical distribution of gerrymandering indicies. This is what was to be expected: the distribution of gerrymandering index has shifted to the rightover time.

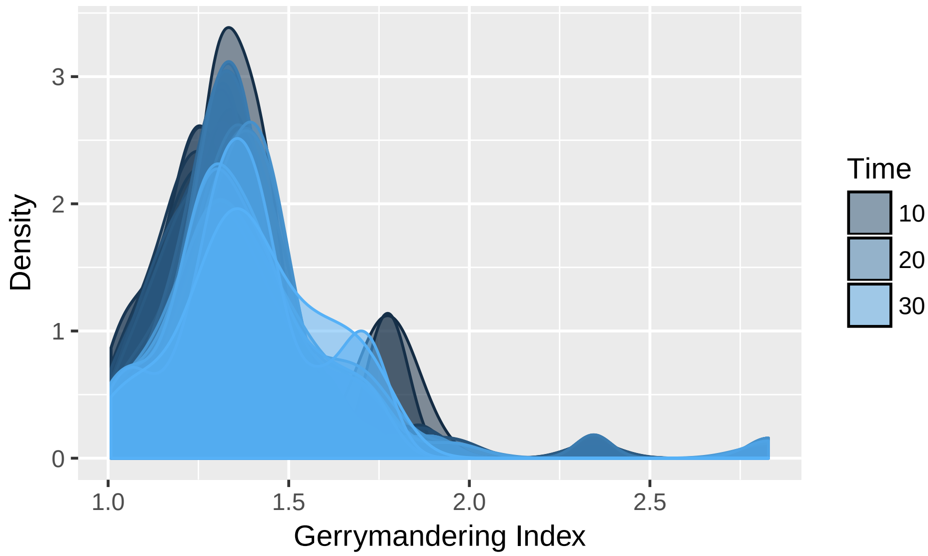

Figure 10: Empirical Distribution of Gerrymandering Index

Combining The Gerrymandering Index with Election Data

I first plotted gerrymandering index against the first-differenced election data, as shown in figure 11. I am tempted to see a slight trend in this data: as gerrymandering increases, the variance of the first differences decreases. This would make sense, as the point of gerrymandering is often to create "safe seats". Thus, the more (nefarious) gerrymandering a state experiences, the less we would expect to see political change. To explore, I binned the data by .125-width bins, and found the standard deviation of each bin. The resulting plot is shown in figure 12.

Given the figure 12 results, the trend appears rather anomalous - the high variance bins do not appear indicative of an overall trend. If I had more time and space, I would want to explore the states that exhibit high variance even in the presense of high gerrymandering. My theory is that one party managed to gain enough political control (not by overwhelming physical presence, but by influence and wiles) to allow them to re-write the entire district map and thereby allow a sea change of sets won to their party at the time of the next election. Should this be the case, we'd expect a single election of high turnover, followed by an extended period of little exchange between parties.

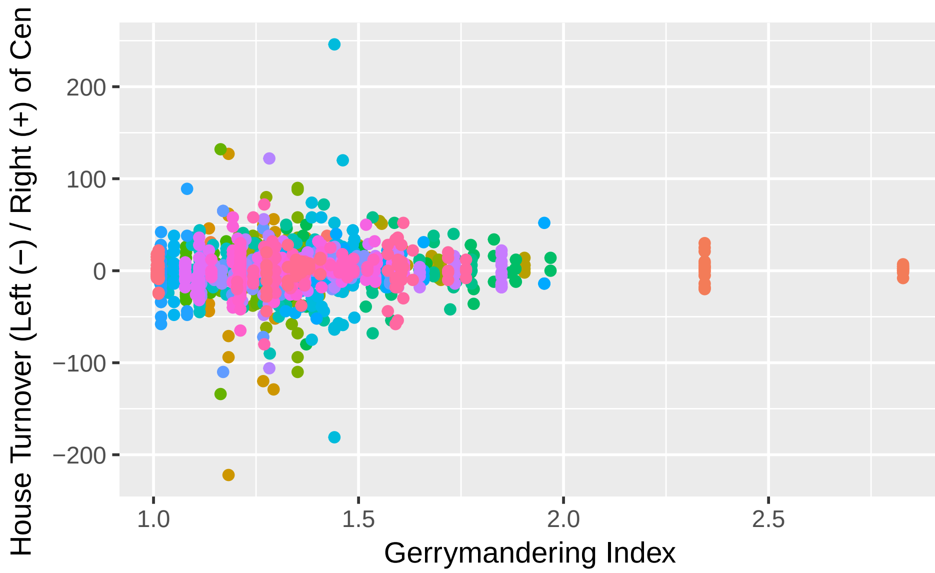

Figure 11: Gerrymandering vs First Differences of House Elections (Colors indicate individual states)

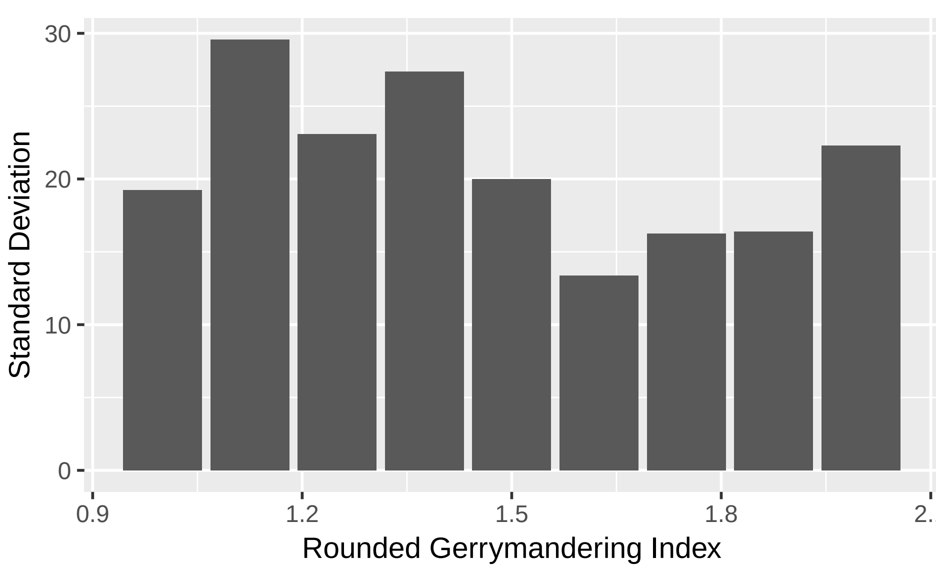

Figure 12: Gerrymandering index vs Standard Deviation of Change in Political leaning.

Is there a spectral signature of gerrymandering?

The following final analysis is my attempt to explore my overall question: does gerrymandering have a particular frequency signature? By that I mean, do certain frequencies appear/disappear as the gerrymandering index of a state changes? To explore this question statistically, I followed the following steps:

1) For each state, I calculated the spectrogram of X_t using a 10-election window with an overlap of 3.
2) I then record the periodogram of each window along with that window's left-most time point, as well as the gerrymandering index (or average of gerrymandering indices, if the window extends over two re-districting periods) and the state associated with that window. I keep the recorded data in an R data.frame.
3) I then calculate the euclidean distances between all the rows of the data.frame (excluding the columns indicating state and gerrymandering index).
4) I then visualize these distances by performing multi-dimensional scaling on the distance matrix to find a set of points in $\mathbb{R}^2$ that faithfully captures the distances between the rows. In this visualization, each point represents the periodogram of a window from some state's X_t series. Thus, I can also associate each point with a gerrymandering index and state.

The output of this procedure is shown in figure 13. Unfortunately, it doesn't say much. I was hoping to see clusters of points with similar sizes (gerrymandering indices), however there is no evidence of such a clear pattern. Instead, it looks like a host of varying gerrymandering indices can be associated with quite similar frequency patterns. Should I have more time/space, I would repeat this analysis with a variety of window sizes and overlaps, and further try to cluster the results both by gerrymandering index and by spectral density.

Conclusion

It is very unlikely that a universal effect of gerrymandering can be seen in the spectral density of X_t - certainly, my analysis didn't discover one. One major obstacle that I witnessed in my analysis was

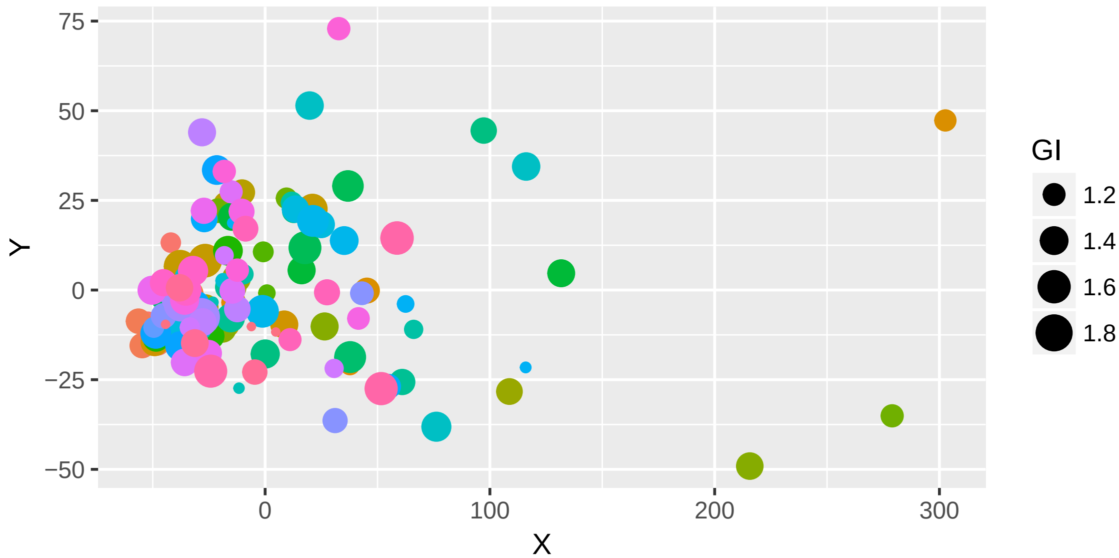

Figure 13: 2D projection of frequency distance

the vagueness of the gerrymandering index. Each state possess its own complex political landscape based on the location, culture, and economy of that state. To say that the convexity of a district is representative of the same political intent across all states is simplistic. That is, a state's decision to redraw a district to have a non-convex shape does not necessarily imply malicious political intent. Without this direct association, there was just too much noise in the gerrymandering data for my analysis to find a signal.

References

Brillinger, David R. 2015. "Marine Mammal Movements: Data Analysis and Theory."

Shumway, R H, and D S Stoffer. 2011. *Time Series Analysis and Its Applications*. Vol. 97. doi:10.1007/978-1-4419-7865-3. http://www.springerlink.com/index/10.1007/978-1-4419-7865-3.