

## Current Progress

**Updated Idea:** In order to maximize democracy I wish to redistrict the United States in a couple ways. First, to combat gerrymandering the entire united states will be partitioned into 435 (the same as before) congressional districts, each differing in population by at most 1. Next, in order to improve democracy in a federal election, each of these congressional districts will be combined into states such that the margin between democratic and republican voters is minimized, i.e. making every state as "swing-statey" as possible

### Current Progress:

I have figured out how to perform the first task using [this](#) paper, which also provides a github repo of rather unhelpful, undocumented, yet functional code to preform to do exactly the first task.

I have not made any tangible progress on the second task (see roadblock) but given in order to group the 435 datapoints into 50 (more likely 48) states, I plan to use a spatially aware kmeans algorithm

### Roadblock

The major roadblock I have been working to find solutions to is unfortunately data related. I have the population data by census tract which is good, but only have the dem% and rep% data by county. For a few states, voter records (for all voters) are available online, but most have either a paywall or require in person requests at a records office.

I have been able to find a [this](#) collection of Harvard surveys with a pretty high sample size, representative of the US, that if I combined the zip-level data from 2006-2020, dropped duplicates, and mapped to census tracts I could build a rough estimate of %dem and %rep for each census tract. This method has the issue though that averaging across all years would mute out any areas that have swung politically since 2006, and since the model is designed to be applicable to current day, not looking at the data cross-sectionally has its drawbacks.

If this method does not work I use the county data as is (there are only 3006, meaning will be less interesting). An alternate method would be to assume that the dem/rep is uniformly distributed across the zipcodes (or potentially using some correlated variable) and mapping the dem/rep of the county to zipcodes that way.