VRDI 2019 Your State Guide

Daryl DeFord

June 25, 2019

1 Introduction

One of the really interesting features of VRDI is getting to take a deep dive into the specific redistricting situation of a single state. This is a great opportunity to see how the process works end-to-end and get some broad data analysis experience. This is the cutting edge of redistricting research, as for most of these states no one has ever constructed an ensemble of plans. That means that these aren't just homework assignments with predetermined correct answers and there is plenty of room for creativity and new approaches. You will undoubtedly run in to unique situations that require creativity and outside-the-box thinking to solve, good luck!

2 Data

As we saw in Week 1, data processing for redistricting analysis is a challenging task, even when using data that comes directly from the census. We will provide dual graphs and shapefiles for census geometries (Counties, County Subunits, Tracts, and Block Groups)[1] together with some population and demographic information. Most of these graphs do not yet have the current districting plans attached - when we get to that point we will have you use the `maup` software to round off the current plan.

For the initial your state presentations, we suggest that you use the block group data and analyze the data columns we have prepared. Even with the census geographies, there are a few common things that may need some additional attention.

- Islands may need to be attached to the mainland (even for some surprising states like GA)
- There may be missing data in some of the columns (ACS data doesn't cover all the block groups in AZ)
- There may be topology problems that make it difficult to evaluate geographic compactness measures (lots of places - pass the options reproject = False and ignoreerrors=True to Graph.fromfile)
- Donuts may need to be merged if you want to try flip runs (KS is disconnected because of donuts)
- I'm sure you will all discover exciting new errors that no one has ever seen before! We'll get through it together.

For some states it will be possible to get precinct data with demographics and elections. We have prepared 14 states worth of jsons here. For some other states, it may be possible to get reasonable shapefiles from the Secretary of State or other online resources. If this seems appealing to you, feel free to reach out to one of the VRDI faulty to get some suggestions about how to get started.

[1] and maybe blocks if my computer doesn't explode ☺

3 Initial Questions

Before we get to ensemble related questions there are some details about the redistricting rules and practices in your state. A great place to get started is reading the corresponding Ballotopedia article `https://ballotpedia.org/Redistricting_in_[insert-name-of-your-state-here]_after_the_2010_census` which provides details about the passage process, lawsuits, who drew the lines, etc. Here are some initial types of questions you should think about:

- Starting points:
  - How many districts (Congressional, state house, state senate)?
  - Current partisan balance (both legislators and executive branch)
  - How much public input is allowed in the process?
  - Has any previous ensemble analysis been carried out in your state?
- Line drawing:
  - Who actually draws them (in principle)?
  - Who actually drew them (in practice)?
  - How do they get passed through the legislature?
  - What partisan interests impacted the most recent process?
  - Do district lines have to nest?
- What are the redistricting rules around:
  - Population balance
  - Compactness
  - Communities of interest
  - Splitting rules
  - etc.
  - What are the current values of these metrics for your state?
- Partisan metrics (for recent elections)
  - What is the seats–votes curve?
  - Mean median
  - Efficiency gap
  - Partisan bias
  - etc.
- Litigation
  - Were there lawsuits in the last drawing cycle?
  - Was the state covered by preclearance?
  - Were there historical examples of important Supreme Court cases in the state?
  - Have there ever been VRA cases in your state?
- Apportionment:
  - How did the census population changes impact where the districts were placed?
  - Which substantial minority groups are represented by the current districts?
  - Which major cities/counties have their own districts?

Not all of these questions are equally interesting for each state, so be sure to note which ones are important for you. Additionally, this list only provides a starting point - there are plenty of other properties you could[2] investigate.

[2] and should

4 Ensembles

The big goal is to be able to perform (at least preliminary) ensemble analyses on your state. Using the `GerryChain` software you should build several ensembles incorporating different collections of constraints, proposals, starting plans, and everything else you can come up with. Early on, we will not have access to partisan data for most states, so the focus will be on demographic measurements and getting prepared for when that data exists.

For week 2 you should focus on block groups[3]. You should vary the required population balance and compactness constraints and make lots of colorful maps from your examples. It takes a lot of experience to generate good ensembles and tune the values to make the outputs useful - take this opportunity to experiment and find out what works well for your state. In addition to the demographic statistics, you should measure some chain heuristics to evaluate mixing. Things like how often pairs of adjacent nodes are placed in the same district, how often each edge is a cut edge, which boundary edges are cut, etc.

The values and measurements that will make sense depend heavily on the properties of your specific state. That said, there are a few questions that should be answerable on all states to start with. These include things like:

- What is the most population balanced plan found by your ensembles (for each type of plan, with each type of units)
- What is the maximum number of majority minority districts found by your ensembles (for each type of plan, with each type of units)?
- What is the minimum number of counties split by a plan in your ensembles (for each type of plan, with each type of units)?
- What is the most compact plan found by your ensembles (for each type of plan, with each type of units)?
- etc.

Remember that you can address these questions for state legislative districts as well. Eventually, we want to establish baselines for the partisan metrics in each state. As with the demographic measurements we would like to know how these metrics interact with the chain constraints and the choices you make in constructing the dual graph and proposal. For this week a sample workflow might look like:

1. Build a starting plan with the recursive tree method
2. Round off the starting plan using `maup`
3. Build four ensembles using ReCom using two different population constraints and two different compactness constraints
4. Build a Flip ensemble on the counties[4]
5. For each plan in each ensemble, measure every metric that you have written a a function for on every column you have access to.
6. Try out some acceptance functions to optimize county splits, population balance, or compactness
7. Make lots of histograms ☺
8. Look at the histograms[5]
9. Show us the interesting ones on Friday

There is a working template for AR in the your state repo that follows most of these steps.

[3] or counties. The question of how close you can get to a reasonable districting plan built out of counties is an interesting one.

[4] It might get stuck - that is interesting too.

[5] This might seem obvious but I have GB of histograms on my hard drive that no one has ever looked at.

5 What errors?

Here is an undoubtedly incomplete list of issues that you might encounter while trying to evaluate your state:

- Islands
- Missing Data
-