Project: AI for Redistricting, Spring 2024
Due: 11:59 PM Wednesday, May 8

Description: You will analyze one state's redistricting map. You will write up the results in a project report and present the results in class (either on May 6 or 8). In doing so, you will incorporate all (or nearly all) of the learning outcomes for this course. For your reference, here are the learning outcomes:

- Describe what redistricting is, how it is done, how it can be manipulated, and why it is complicated.
- Calculate various metric values intended to detect gerrymandering. Compare and contrast the ways in which these metrics act, and what they can and cannot detect.
- Describe what a Markov Chain is, what a random walk on a graph is, and how they can be used to create an ensemble of districting maps. Describe the redistricting metagraph and identify some of the complexities and nuances in ensemble creation.
- Explain some of the theory of Markov Chains, stationary distributions, and mixing times.
- Identify the data required to use each of the metrics, and to use GerryChain to create an ensemble of maps. Construct a coherent and accurate dataset that can be fed into GerryChain.
- Apply all of this knowledge to an analyze a map using an ensemble created through GerryChain and using metrics intended to detect gerrymandering. Formulate a coherent discussion of a districting map through this analysis.
- Summarize the legal history as it pertains to racial gerrymandering and partisan gerrymandering. Explain the important factors in the redistricting process, such as commissions, legislatures, communities of interest.
- Understand the legal and social importance of "VRA districts," as well as the three Gingles criteria. Apply tools of Ecological Inference to determine the presence (or absence) of Gingles 2 and Gingles 3.

Additionally, graduate students completing this course will be able to:

- Answer complex theoretical questions pertaining to redistricting and assessing the presence of gerrymandering.
- Assess a plan's responsiveness to partisan swing, and assess a state's capacity for VRA districts.

Required Components:

- Brief description of redistricting and gerrymandering
- Contextual description of state in question
  - Description of when the map was created and by whom
  - Description of legal landscape for that state, if applicable
  - Description of partisan and racial landscape for that state
- Gathering and cleaning of data using MAUP
- Brief description of the method you will be using: a Markov Chain process to implement an ensemble analysis.
- Creating a Markov Chain
- Outlier analysis/histogram for:
  - Efficiency Gap
  - Mean-Median Difference
  - Dem-won (or Rep-won) districts
  - Cut edges
- The "signature of gerrymandering"
- Write a paper reporting your results
- Have a GitHub page for your code
  - Data cleaning code, with excellent documentation
  - Ensemble creation code
  - Any image creation code
- Give a presentation of your results

Grad students only:

- Short Burst analysis
  - Include a brief description of what such an analysis is used for.
  - Include a brief description of what the Short Burst method is.
  - Include appropriate tables and/or visualizations

Supportive Materials

- Go to this link: `https://docs.google.com/spreadsheets/d/1B9jyOztmBl22ntq13yyG-prZKtJbVt` `edit?usp=sharing` to declare your group, and which state you will analyze.

  - I will allow more than one group to look at the same state, but if more than one group considers the same state, they must analyze *different districting maps.*

- Go to this link: `https://www.overleaf.com/read/jrvnhrhkfddv#722213` to see a sample report. You are NOT required to use LaTeX. But you're welcome to do so.

Submission:

- Please use a public shared GitHub repository for your group. Please commit/push frequently so that I can see the progression of your work. It may be helpful for you and your group members to work in branches; if this is something you're not familiar with, I'm happy to walk you through it, just let me know. Include the link to your repository in your project report.

- Upload your project report to Canvas. Your report must contain a link to your GitHub repository for your code. Make sure that everyone in your group selects the correct group members for the Project assignment in Canvas.

- Your implementation must be your work (no copy/paste/edit), although you may look to other sources if you get stumped. If any part of your implementation is derived from another work, **you must cite that work**, include a discussion about what you used, and include a link to the source.

- Make sure to comment your work sufficiently so that you and anyone looking at your code can understand the code easily. This is especially important if any part of your implementation is derived from another work.

- Each group will give a brief, approximately 12-minute presentation of their project and results on either May 6 or May 8.

- If you are concerned that thee workload for your project was unbalanced, please reach out to Prof Veomett to share your concerns at eveomett@usfca.edu

Map, 2xElection data, population data, racial data

Recommended Timeline:

By Monday March 25: Choose group members. Choose state and map. Put all of these on the spreadsheet: `https://docs.google.com/spreadsheets/d/1B9jyOztmBl22ntq13yyG-p edit?usp=sharing`

Week of March 25: Collect and prepare data for your state, being sure to include more than one election. Document data collection/preparation. Research your state/map background. Begin filling that background into the paper.

Week of April 1: Continue data cleaning/collection, finish state/map background research. Begin running ensembles and creating images. Fill in parts of the paper describing methods. This week or next check in with Prof Veomett to make sure you're on track.

Week of April 8 Finish running ensembles and creating images. Fill in parts of the paper describing methods.

Week of April 15 Create marginal box plots. Discuss with group members any findings/conclusions. Run short bursts (if grad group). Discuss with group members any findings/conclusions. This week or next check in with Prof Veomett to make sure you're on track.

Week of April 22: Finish any remaining analyses. Start filling out results and discussion sections. Create any additional/revised/cleaner visualizations. Begin creating slides for presentation.

Week of April 29: Clean paper, clean code, finish presentation slides.

Week of May 6: Give presentation, hand in project report.

Grading:

50% = Implementation Your code must run efficiently, successfully, and accurately. It must do exactly what you claim it does in your project report.

15% = Documentation and Images Your code must be easily readable to another student who has taken this class. The images that you create for your project report must be illuminating, accurate, and complete. They must tell the story of your analysis well.

20% = Project Report Your report must be well-organized. It must be well-written, clear, concise, and with smooth transitions. It must tell the complete story of your analysis.

15% = Presentation Your presentation must be well-organized. All members of your group should participate meaningfully in the presentation. Your presentation must be clear and include the main results of your analyses.