Gerrymandering Case Study Rubric

DS 4002 – Fall 2023 – Ravza Aykan
Due: Dec 11
Submission format: Upload link to GitHub repository on UVA Canvas

Individual Assignment

Why am I doing this?
This case study allows you to leverage your data science knowledge by using image analysis techniques to detect and measure gerrymandering in various states in the US. As you work through this assignment, you will be exposed to the ways that data analysis can be used in a real-world context with potential implications for public policy and governance.

What am I going to do?
The GitHub repository for this case study can be found at https://github.com/ravza22/DS4002_CS2/. You will obtain maps of current and hypothetical, non-gerrymandered district boundaries for various states of your choosing from FiveThirtyEight's "The Gerrymandering Project". The GitHub repository contains maps for several states as examples. After obtaining the maps of district boundaries, you will use Python to determine the counts of RGB colors from the images. After determining proportions of color composition (where each color corresponds with a chance of being represented by each party) for both the current and hypothetical states, you will create a frequency table for each state and conduct a chi-squared test on the table in R to determine whether the shares of votes in each state are significantly different between their current district boundaries and their theoretical, non-gerrymandered district boundaries. You will then use the declination angle values for each state (which can be found at https://observablehq.com/@sahilchinoy/gerrymandering-the-declination-function) to determine whether your model (i.e., the chi-squared test) yielded accurate results regarding whether states were gerrymandered.

Your final deliverables should include:
- The maps for the states you chose to analyze
- A data dictionary
- Well documented, commented source code
- A GitHub repository containing all materials used

Tips for success:

- Do not approach the analysis with a partisan mindset – instead, select at least one state with a current Republican majority and one state with a current Democrat majority.
    - A higher-level analysis would consist of more than two states in your dataset, so choose as many states as is feasible to analyze
- Make your file names (i.e., the maps) and your variable names easily interpretable.
    - map1, map2, proportion1, proportion2, etc., are not easily interpretable and can cause issues down the road as you analyze more states
- You will be working in both Python and R. Familiarize yourself with both languages to the best of your ability before starting this case study in order to complete it more efficiently

How will I know I have succeeded?

You will meet expectations on this case study when you successfully follow and complete the criteria in the rubric below:

Spec Category	Spec Details
Formatting	• One GitHub repository (submitted via link on Canvas) ▪ Create a new GitHub repository for this assignment titled 'CS2_Gerrymandering' that contains ▪ README.md ▪ LICENSE.md ▪ Source Code File ▪ Your data (i.e., the images you chose) ▪ REFERENCES.md
README.md	• Brief summary of what you've produced for the case study, this does not have to be very detailed but should provide enough information to orient people to your repository

Source Code File	Well documented Jupyter Notebook file and R Script that contains the code used to execute your image and statistical analyses. In the source code, you must include: • The states you chose • Color composition for each state • Frequency table for each state • Chi-sq test for each state • Comments throughout, and especially in the R Script when interpreting the results of the chi-sq test
REFERENCES.md	Markdown File titled “REFERENCES.md” with citing any resources (journal articles, websites, etc.) referenced in helping you create your model in IEEE Documentation style. Also include brief annotations under each citation on how each reference informed/helped you for this case study.

Acknowledgements: Thank you to Professor Alonzi for providing the rubric structure!