Computational Analysis of California's 2021 Congressional Districts Using Markov Chain Monte Carlo Recombination

By Kael Adair and Dr. Dustin Potter

COLLIN COLLEGE™

SMU MUM_TX

Background

In 2021, California established new congressional district boundaries, drawn by the independent and non-partisan California Citizens Redistricting Commission (CCRC), formed under the VOTERS FIRST Act. The goal was to reduce partisan bias and provide fair representation for all Californians, addressing the issue of gerrymandering in the previous map.

Gerrymandering is the manipulation of electoral boundaries to favor a specific political party or group, leading to distorted election outcomes and unfair representation.

In order to detect gerrymandering and determine if the CCRC has drawn a more fair and less gerrymandered congressional map, we can utilize a Markov Chain Monte Carlo Recombination algorithm to take the current map and randomly split and combine districts to produce an unbiased ensemble that we can then compare with the current map.

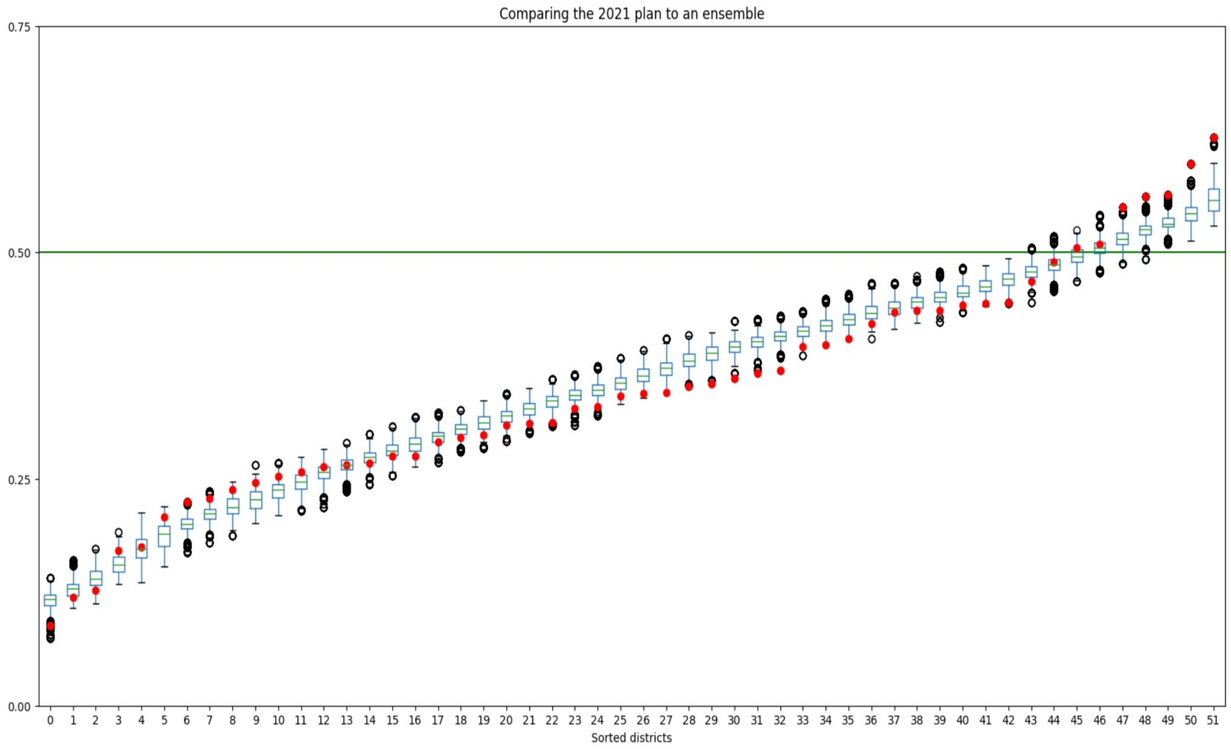
In simpler terms, what we have done is generated over 10,000 new congressional maps and can compare California's new 2021 map to visualize any potential gerrymandering activity.

Methods

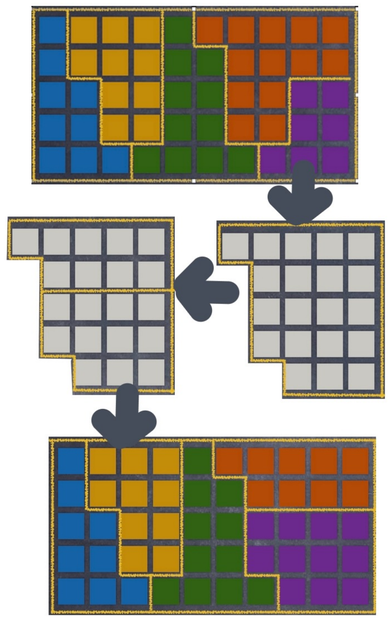
Data Preprocessing and Cleaning:
The initial stage of our analysis involved the collection, cleaning, and integration of relevant data sources. These sources included census blocks, shapefiles, census data, congressional district block assignments, and election data from the ALARM groups dataset for the 2020 election. We employed Python libraries like Pandas, GeoPandas, and Fiona to manage and clean the data, ensuring that the final dataset was free of inconsistencies, missing values, and inaccuracies.

Island Connections:
As part of the data preprocessing, we identified and connected isolated islands to the mainland. This step was crucial to maintain the continuity of the districts and avoid any issues in the subsequent analysis. The connections were established based on the nearest neighboring landmass to ensure minimal impact on the overall district shapes.

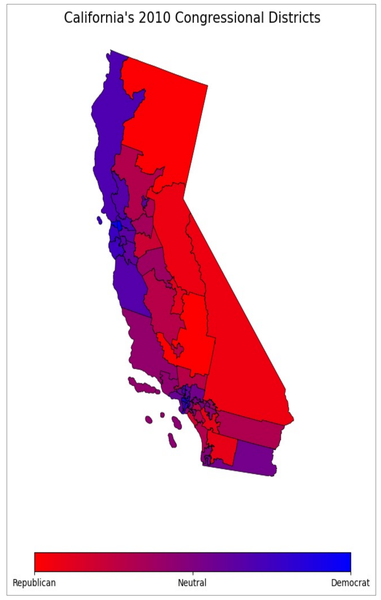
Republican Vote Share: Unbiased Ensemble (Boxes) vs. 2021 Map (Red Dots)

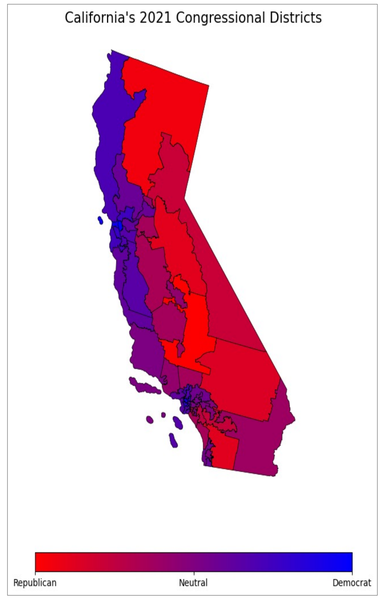


GerryChain



California Congressional Districts (2010 vs. 2021)





Graph Construction:
With the cleaned dataset in hand, we proceeded to construct a graph representation that would facilitate the use of GerryChain for our analysis. Utilizing the NetworkX library, we converted the spatial data into a graph, where each node represented a census block, and edges denoted the shared boundaries between adjacent blocks. This graph representation allowed us to effectively model the relationships between the different geographical units in our dataset.

GerryChain Analysis:
Having prepared the graph, we used the GerryChain Python library to analyze California's 2021 congressional districts. We ran multiple Markov Chain Monte Carlo (MCMC) simulations to explore the space of possible redistricting plans, which allowed us to generate 10,000 new unbiased maps. By comparing the distribution of simulated district plans against the enacted plan, we assessed the fairness and representativeness of California's current congressional districts.

Results

Our analysis revealed that California's new map still exhibits signs of gerrymandering, given that most districts fall outside of the distribution of the unbiased ensemble. However, it appears that this gerrymandering has been intentionally carried out by California to benefit historically underrepresented voters. This approach inadvertently gives the impression of partisan gerrymandering in favor of the Democratic Party.

By comparing the new map to alternative district maps generated using GerryChain and MCMC simulations, we observed that the new districts exhibit some irregular shapes and configurations that seem to benefit the democratic party. However, our findings suggest that these anomalies aim to enhance the representation of underrepresented communities, rather than solely serve partisan interests, and has been done openly.

In conclusion, while California's new map displays gerrymandering characteristics, our hypothesis is that the Citizens Redistricting Commission's efforts to promote fair representation have led to the prioritization of historically underrepresented voters, and more data needs to be analyzed to see exactly what is going on.