

Exploring Maine's Historical Records and Ranked Choice Voting

Cindy Hong, 2027

This summer, my research focused on the analysis and conversion of Maine electoral records and census data into easily digestible form for future work, later shifting to understanding the various models used to analyze elections using the system of ranked choice voting.

The process began with collecting all recent records of Maine's elections ranging from 2010 to 2024. This included but was not limited to: presidential elections, state congressional races, and referendums. All of these elections were already digitized by the government previously and were meant to be converted to a .csv format for better, more efficient handling in future projects.

The following phase of my research involved working with data from the American Community Survey (ACS). Similar to the US Census, the ACS provides data collected state by state on over one thousand demographic variables such as age, race, career, and living conditions. It provides data from both a 1-year and 5-year span. I reviewed these variables with the goal of finding potential proxies for economic variables. In order to reduce the number of economic variables with collinearity, I used RStudio to calculate the variance inflation factors (VIFs) among the larger set of variables through linear regression. Those with VIFs of five or higher were discarded. Out of the initial set, I narrowed it to a broad twenty variable set and a final subset of five variables. The remaining variables were chosen as they represented distinct categories that contributed to economic prosperity.

After working on the ACS, my next task was to select the best tools to digitize historical voter records of Maine from the 1870s onwards through testing the accuracy of the tool on the handwritten documents. In the current landscape of OCR (optical character recognition) software, I found the most useful tools to be ABBYY FineReader, and two other companies that specialize in processing documents into manageable formats. An ongoing project, I look forward to seeing a full digitization of these records, as it is rare to encounter electronically available records from this time period.

The final section of my research explored the Plackett-Luce and Bradley-Terry models as ways to evaluate and analyze ranked choice elections. I collaborated with fellow student researcher, Lieren Peterson, to observe the structure of Minneapolis' individual district elections, building the Plackett-Luce model in RStudio. We calculated individual ballot likelihoods and simulated the elections with a great degree of accuracy.

I am extremely grateful to have done this work this summer and find that I have strengthened my technical skills in both coding and statistical analysis, as well as efficiency. Through Professor O'Brien, I learned the importance of ensuring my code in RStudio was streamlined and systematic, especially given the large datasets I was working with. Overall, this experience deepened both my understanding and passion for statistical analysis, and I look forward to my future endeavors in the field of mathematical research.

Faculty Mentor: Professor Jack O'Brien

Funded by the Bowdoin Research Award Grant