Data and Democracy

The University of Chicago Data Science Institute identified a major gap in election research: decades of U.S. election data existed only as scanned PDFs, making large-scale analysis virtually impossible. The Data & Democracy team built a custom pipeline powered by AI-based optical character recognition to convert those 51 PDF files (over 3,000 pages) into machine-readable data for analysis and visualization.

The team partnered with OpenElections—an open-source repository of U.S. election returns—submitting processing scripts and formatted data to extend their archive. The team's contributions will ensure that every state's federal results from 1920-1996 integrate seamlessly into the standardized framework they provide.

To make this archive widely accessible, the team published an independent website hosted via GitHub Pages, where visitors can browse/download machine-readable election data from 1920-1996, view interactive maps and visualizations, read about the team's OCR approach and validation methods, and meet the team to learn about the project goals.

By transforming paper records into ready-to-use data, the team empowers historians, political scientists, journalists, and civic technologists to uncover long-term voting trends and deepen understanding of American democracy.
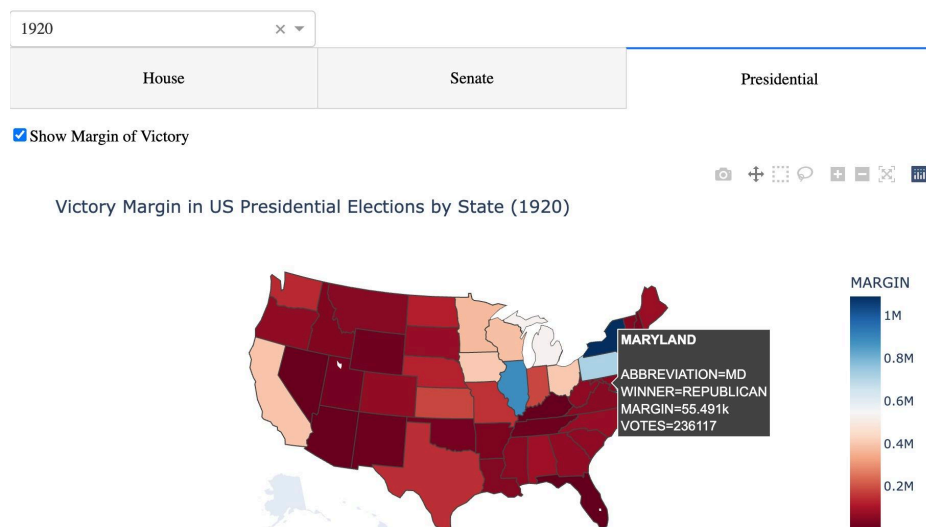

Figure 1: Sample Visualization for 1920 Data