States are the preeminent geopolitical units of the United States. Americans live under different public institutions and laws in each state. However, state boundaries are counterintuitive, given their importance. For example, in Kansas City, the Missouri-Kansas boundary cuts the metropolitan area in half. In Texas, El Paso is 285 miles from the nearest metro area in Texas (Odessa) but only 45 miles from the closest one in New Mexico (Las Cruces).

What if the state boundaries of the United States matched where Americans live? This project applies machine learning to reimagine the borders of the contiguous United States. The algorithm groups people who live near each other, placing the new state borders in unpopulated expanses of land between groups. The algorithm suggests grouping the contiguous US population into {state\_count} new states. Figure 1 draws boundaries (blue lines) around population centers (brown dots) to illustrate how the new state borders correspond to where people live.

How would the new states differ in typical quality of life and politics? Figures 2a, 3a, and 4a present the Human Development Index (HDI), total population, and the percentage that voted for the 2020 Republican presidential candidate for each new state. HDI is a quality of life score that considers high income per capita, long average lifespan, and high average years of education to be signs of high quality of life. Presidential vote is an indicator of general political inclinations. The new states are ordered the same way in all three bar charts to facilitate side-by-side comparisons. Figures 2b, 3b, and 4b present these statistics as maps.

METHOD: The project applies two-stage agglomerative clustering to formulate new state borders. Agglomerative clustering is a machine learning technique that groups points (places) into clusters. The algorithm initially groups neighboring places and then iteratively adds more sites to each set, starting with places that are very close to all the places already in the cluster. It also merges nearby clusters until the number of clusters is less than a given threshold. The project does this once to split the country into regions, and then again to split each region into new states.