

Studying Gerrymandering with HPC

Sal Saporito (W&M Sociology)

and

Daniel Maliniak (W&M Government and Public Policy)

Social scientists and legal scholars have relied upon measure of geometric shape to determine if legislative districts are racially gerrymandered. Yet, measures of geometric shape are, at best, an imprecise way of detecting gerrymandering since some districts can be irregularly-shaped due to natural and man-made geography—while other districts can be compact but still be racially gerrymandered. The goal of racial gerrymandering is to “pack” a supermajority of racial minorities into one or two legislative districts so that racial minorities represent a numerical minority in the remaining legislative districts in the state. That is, racial packing diminishes the political power of racial minorities by limiting their ability to elect more than one or two candidates to a legislative chamber.

Since measures of shape do accurately detect racially packed districts, Professors Salvatore Saporito and Daniel Maliniak developed a new technique to identify legislative districts that are intentionally packed with racially minorities. An intentionally packed district has much higher shares of racial minorities than is acceptable given the racial composition of the region in which the legislative district is located. Imagine a legislative district in which, say, 70 percent of the people are racial minorities yet only 40 percent of the people in the vicinity of the district are racial minorities. In comparison to an intentionally packed district, an accidentally packed district contains a supermajority of racial minorities if it is located in a geographically expansive region in which the vast majority of voters are also minorities. For example, think of a legislative district located in large, minority-dominated city such as Detroit, MI. In Detroit, it could be the case that, say, 70 percent of the people in a district are racial minorities and 70 percent of the people in the vicinity of the district are also racial minorities. This is an “accidentally” packed district.

To distinguish between naturally packed and accidentally packed legislative districts, Saporito and Maliniak developed a

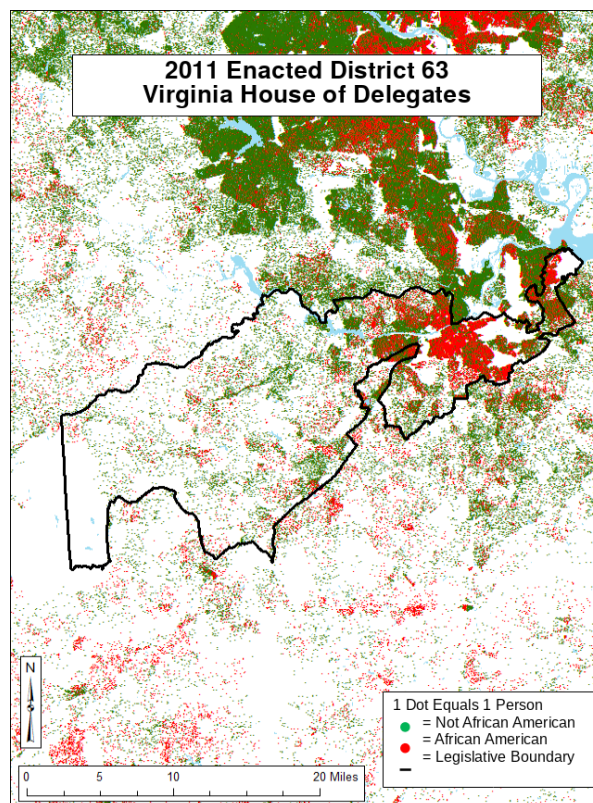


Figure 1: District 63 of the 2011 Virginia House of Delegates was racially packed since 60 percent of the people in the district are African American while, on average, 48 percent of the nearest 700,000 neighbors of district residents are African American.

systematic but computationally intensive way of comparing the racial composition of a legislative district with the racial composition of the region in which a district is embedded. Their approach rests on the notion that racial composition of nearest N neighbors of a district's residents should be, on average, nearly the same as the racial composition of the district. (Here, N is equal to the number of people in a legislative district—or about 700,000 people for a typical U.S. Congressional District.) To determine if the racial composition of a U.S. Congressional District is about the same as the racial composition the nearest neighbors of its residents, the measure makes three basic calculations. First, for each person who lives in district, the measure determines their nearest 700,000 neighbors. Second, the measure calculates the percent of each person's nearest 700,000 neighbors who are racial minorities. Third, the measure calculates the mean percentage of each person's nearest 700,000 neighbors who are racial minorities. The last calculation provides a systematic way of calculating the racial composition of the region in which a legislative district is embedded. The accompanying figure shows an intentionally packed district since 60 percent of the people who live in the district are African American while, on average, 48 percent of the nearest neighbors of district residents are African American.

While the nearest neighbor measure is conceptually straightforward, it is computationally difficult to execute. Calculating the racial characteristics of a single person's nearest 700,000 neighbors takes several seconds on a high-end personal computer. Since there are 330 million people who live in the U.S. and there are thousands of state and federal legislative districts, this project would require about a month just to analyze all of the districts in California. However, a significant reduction in the actual time needed for these calculations can be achieved by running many people at the same time. To achieve this, Jay Kanukurthy (W&M HPC staff) was able to split up all of the work and enable 100-200 calculations to run simultaneously. The result was that within 2 weeks of running, all results were generated for the entire U.S.. These results were generated using the *STATA* statistical software package (<https://www.stata.com>).

Note: This work was highlighted in the July, 5th 2019 issue of the *Daily Press*:

Article: http://enewspaper.dailypress.com/infinity/article_share.aspx?guid=b1bb5481-2ddf-4631-8d94-fb1d50d28d15

Video highlight: <https://www.dailypress.com/visuals/video/117018779-132.html>