

**Racial Gerrymandering in North Carolina: Analyzing Voter Distribution and  
Representation from 2016-2020**

Kyle Guanzon, Amanda Li, Hannah Wen

Department of Statistics, University of California, Davis

STA 141B: Data & Web Technologies for Data Analysis

Dr. Peter Kramlinger

December 11, 2024

## **Abstract**

Gerrymandering, or the act of manipulating electoral district boundaries to give a particular party or group unfair advantage, is a common tactic used to suppress minority voters in the United States. North Carolina is one of many historically gerrymandered states, but in recent years, efforts have been made to draw fairer district lines. For this reason, we compared North Carolina's 2016 district map to the 2020 district map by analyzing the spatial distributions of various demographic groups to assess whether or not these redistricting efforts indeed resulted in fairer boundaries. All data was collected by web-scraping and PDF-parsing North Carolina's voter statistics, demographic groups, and population numbers. Our results show that, after the district map was redrawn for the 2020 election, many districts exhibited less characteristics of cracking and packing voters. This analysis is important because it helps to determine whether recent redistricting efforts have succeeded in fairly representing all voters, particularly minority groups, and provides a look into the ongoing fight for electoral fairness.

All code can be accessed in this [GitHub repository](#). Interactive visualizations can be found on [ArcGIS](#).

## **Introduction and Background**

On December 5th, 2016, citizens David Harris and Christine Bowser stood before the United States Supreme Court to challenge a system that was deliberately designed to undermine their votes. The two resided in District 1 and District 12 of North Carolina—districts that contained 33.4% of the state's entire Black population (Jenning, 2021). Following months of legal battles with the North Carolina district court, Harris and Bowser took to the Supreme Court and argued that the state's use of the Voting Rights Act in their redistricting plan was merely a pretext to pack Black voters into two districts to weaken their voting power across the state. One

year later, the Court ruled in favor of Harris, declaring that North Carolina's plan constituted an unconstitutional racial gerrymander that violated the Equal Protection Clause. This resulted in the landmark case *Cooper v. Harris*.

*Cooper v. Harris* is just one example of the widespread practice of gerrymandering in the United States, where electoral district boundaries are manipulated to give a particular party or group an unfair political advantage. The two most common techniques used in gerrymandering are "packing" and "cracking." Packing involves concentrating voters of a particular group into a single district to reduce their influence in other districts—a tactic central to the issue in *Cooper v. Harris*. Cracking, on the other hand, involves spreading voters from a particular group across multiple districts, diluting their voting power so they cannot form a majority in any district. An example of cracking occurred when North Carolina A&T State University—the largest historically Black college and university in the U.S.—was split across two districts, both of which were reliably Republican (Meyers, 2019). Redistricting, or the act of redrawing district lines, may determine where residents can vote, whom they can vote for, and even how responsive elected officials are to constituents' requests (Fernandez, 2021). These practices are often employed to suppress the political power of communities of color and manipulate electoral outcomes.

Following the 2017 Supreme Court case *Cooper v. Harris*, subsequent legal proceedings continued to unfold. In June 2019, the Supreme Court reversed course in *Rucho v. Common Clause*, declaring that federal courts could not intervene in cases of partisan gerrymandering, as they presented a "political question beyond the reach of federal courts." Justice Kagan, dissenting in the case, argued that by refusing intervention in political gerrymanders, the Court was effectively encouraging a "politics of polarization and dysfunction" that could "irreparably

damage our system of government” (Rucho v. Common Clause, 2019). Despite this setback, North Carolina saw a shift towards improved districting—in alignment with Kagan’s argument—when the General Assembly passed House Bill 1029 in October 2019. This bill established fairer district boundaries for the 2020 presidential election.

This back-and-forth of ongoing political battles in North Carolina reflect the state's persistent struggle with racial bias in its redistricting practices. For this reason, we have selected North Carolina as the focal point of our project, positing it as a unique case study for analyzing the impact of redistricting on minority voting power and political representation. In light of these legal proceedings, this paper seeks to address the following research questions:

1. How have the spatial distributions of demographic groups within North Carolina's congressional districts shifted following the redrawing of district boundaries by the General Assembly in 2020?
2. To what extent have redistricting efforts in North Carolina affected the voting power of historically suppressed voter groups?

To address these questions, we will examine the racial implications of North Carolina's district lines between 2016 and 2020 and present data that highlights the disproportionate impact on minority voters. We hypothesize that the 2020 redistricting plan has led to more equitable representation for Democratic and Black registered voters. Ultimately, this paper aims to provide a data-centric understanding of voting rights, fair representation, and the power of minority voices amidst North Carolina's legal tug-of-war over redistricting efforts.

### **Methodology: Data Acquisition**

We chose to collect data on racial demographics, voter registration, and population statistics to better understand the effects of racial gerrymandering in North Carolina. Our data

was acquired through two primary processes: web-scraping and PDF extraction. All the data we collected was analyzed using Python and ArcGIS.

We began by web scraping the [Voter Registration Statistics Search](#) website using *BeautifulSoup*, which provides demographic statistics on voter registration for each county in North Carolina based on specific dates. To ensure consistency in the data, we extracted the list of counties for each election year and cross-checked them between 2016 and 2020 to confirm that the county list remained unchanged. For our analysis, we selected November 8, 2016, and November 3, 2020, as the key election dates. Then, we scraped the voter registration demographics for each county by parsing the HTML of the search page. For each county, we iterated through the relevant data blocks, creating a dictionary to store the demographic statistics such as party affiliation (e.g., "Republican") and gender (e.g., "Female"). After acquiring all the county-level demographic data, we compiled it into a single large dictionary, which was then converted into a *Pandas* DataFrame and exported as a CSV file. We repeated this process for both 2016 and 2020, creating separate files for each year, as well as a combined dataset with an additional column indicating the collection date.

Many challenges arose when we web-scraped the data. We noticed that each county's information was only accessible through a dynamically-loaded dropdown menu powered by *Java*. However, upon further inspection, we discovered that all the data for each county was embedded in the HTML of the original search page, which we accessed by simply using the date search feature. Another obstacle we faced was missing data. When a new search query was entered, the county demographics would sometimes disappear from the HTML despite being displayed on the page. The changing layout made it more difficult to identify the specific elements we needed to extract. It is also important to note that the data did not include

information on Asian or Multiracial racial groups. However, since these groups made up less than 10% of North Carolina's population combined, and given that historical gerrymandering in the state has primarily impacted Black voters, we determined that these specific demographics were not critical to our analysis (Census Bureau, 2023).

While searching for relevant websites that contained North Carolina's population data, we discovered that all county and district population data were available in PDF tables, which we downloaded from the [North Carolina General Assembly](#) website. Using the *PdfReader* class from the *PyPDF2* Python library, we extracted text data from both the 2016 and 2020 files. The text data was unstructured, with random line breaks and spaces. To organize the text, we applied various regex patterns to match sequences of letters, integers that may include commas, and decimal values followed by the percent symbol. After we structured the text, we transformed the text into DataFrames and exported the data into CSV files.

We opted to use the *PyPDF2* library because more modern libraries like *PyMuPDF* were incompatible with the formatting of the congressional files. Additionally, existing libraries that were created for tabular extraction, such as *Tabula* and *Camelot*, either had deprecated features or required an older version of *Java*. Although *PyPDF2* was the most viable option, we encountered some challenges during the extraction process. In particular, the use of color in the 2016 PDF led to data loss, so we had to manually retrieve a few lines of data from that year. Fortunately, this was only required for a small portion of the data. Another challenge we faced was variation in text formatting in the 2016 and 2020 PDFs, meaning we could not use a single script to automate the extraction for both years. As a result, we wrote separate scripts for the 2016 and 2020 PDFs.

## Methodology: Data Processing

Using web-scraped data on county-registered voters and district distribution data extracted from PDFs, we merged the two DataFrames to create a comprehensive table representing the demographics of each county. An additional column was created to store a list of the counties within each district, ensuring a relationship between counties and districts. The data was then aggregated by district to calculate the total population of registered voters for various demographic groups at the district level, rather than by county. During this process, we encountered challenges due to inconsistent data types, so all columns were standardized to the integer type after applying regex manipulation. We also created a function to compute columns representing the percentage of each demographic group relative to both the total number of registered voters and the district population. Finally, the data was validated by ensuring the aggregated county populations aligned with the original district population figures.

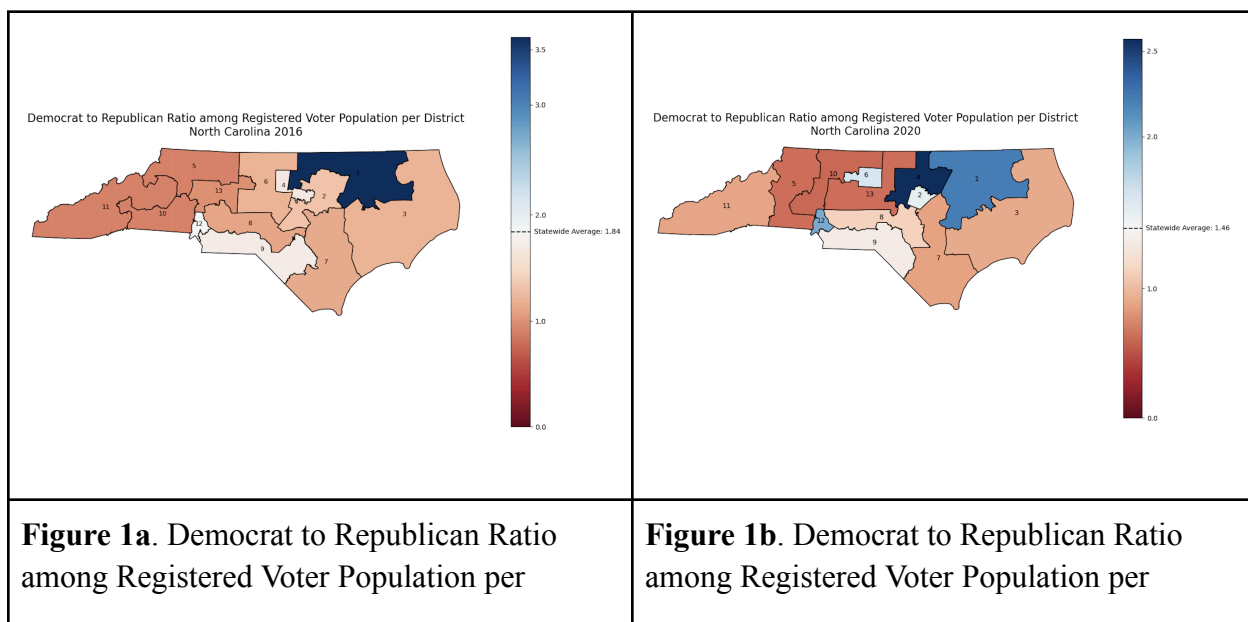
In addition to using Python, we also used ArcGIS Pro to evaluate racial gerrymandering in North Carolina. District and county shapefiles were obtained from the [North Carolina General Assembly](#) website, which we then merged with district demographic and voter attributes derived from the earlier PDF extraction. We imported shapefiles and layers and assigned features to be drawn on the map. The layers included the 2016 and 2020 Congressional Districts, North Carolina county maps, and 2016 and 2020 Election Day Polling Places. After merging the district-level demographic data, we conducted spatial statistical analyses, specifically using Moran's I and standard distance computations.

None of our team members had prior experience with Geographic Information Systems (GIS), so this data processing stage posed several challenges. Due to compatibility issues, we had to manually reformat and reshape the CSV and shapefiles in Excel, adjusting rows and

columns and adding attributes to align the data with the shapefiles. After making these manual corrections, we were able to successfully merge and process the files. Despite this learning curve, we selected ArcGIS for its ability to analyze spatial and demographic data, which would not have been achievable through traditional statistical analysis methods alone.

## Results

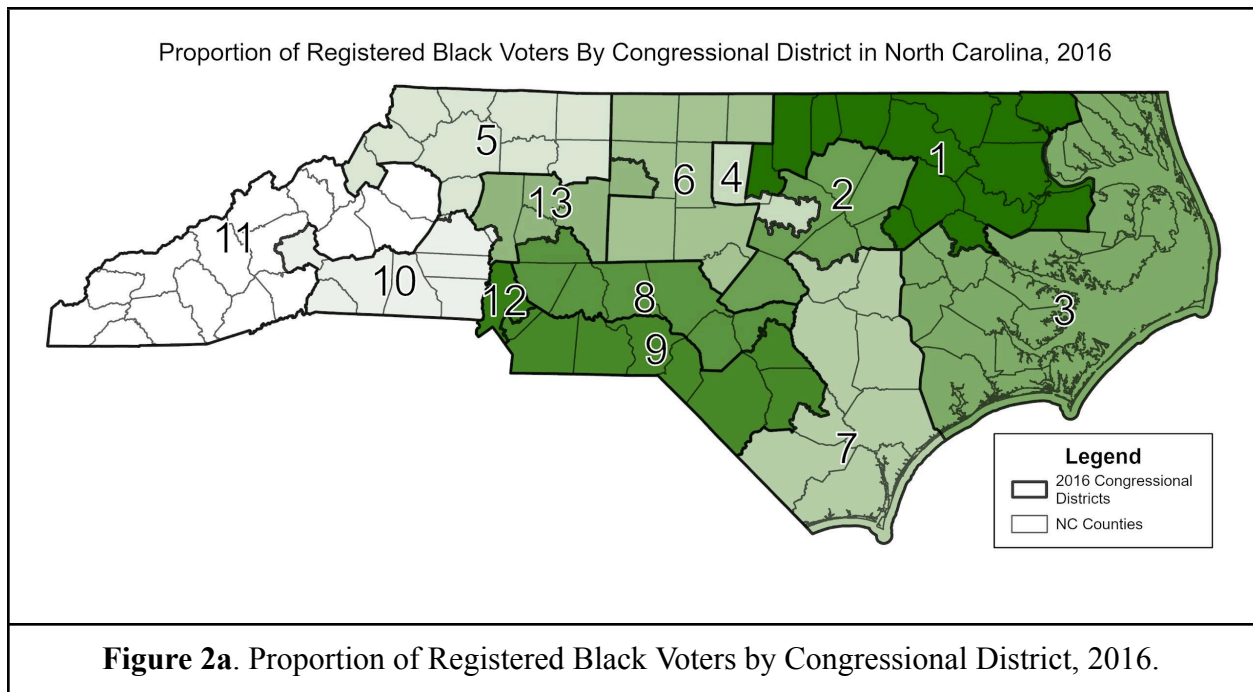
Our analysis of North Carolina's voter distribution from 2016 to 2020 revealed shifts in the concentration of Democratic and Black voters across congressional districts. In 2016, the statewide average ratio of Democratic registered voters to Republican registered voters was 1.84. This ratio decreased to 1.46 in 2020. The 2016 distribution, as shown in Figure 1a, illustrates a pattern of Democratic voter concentration, where District 1 exhibited extreme packing of Democratic voters. The majority of other districts either aligned with the statewide average ratio of 1.84 or leaned predominantly Republican. In contrast, the 2020 distribution, depicted in Figure 1b, demonstrates a more dispersed pattern of Democratic voter concentration, where five districts (1, 2, 4, 6, and 12) emerged as majority Democratic. This represents a significant increase from the 2016 map, where only District 1 showed such a strong Democratic majority.

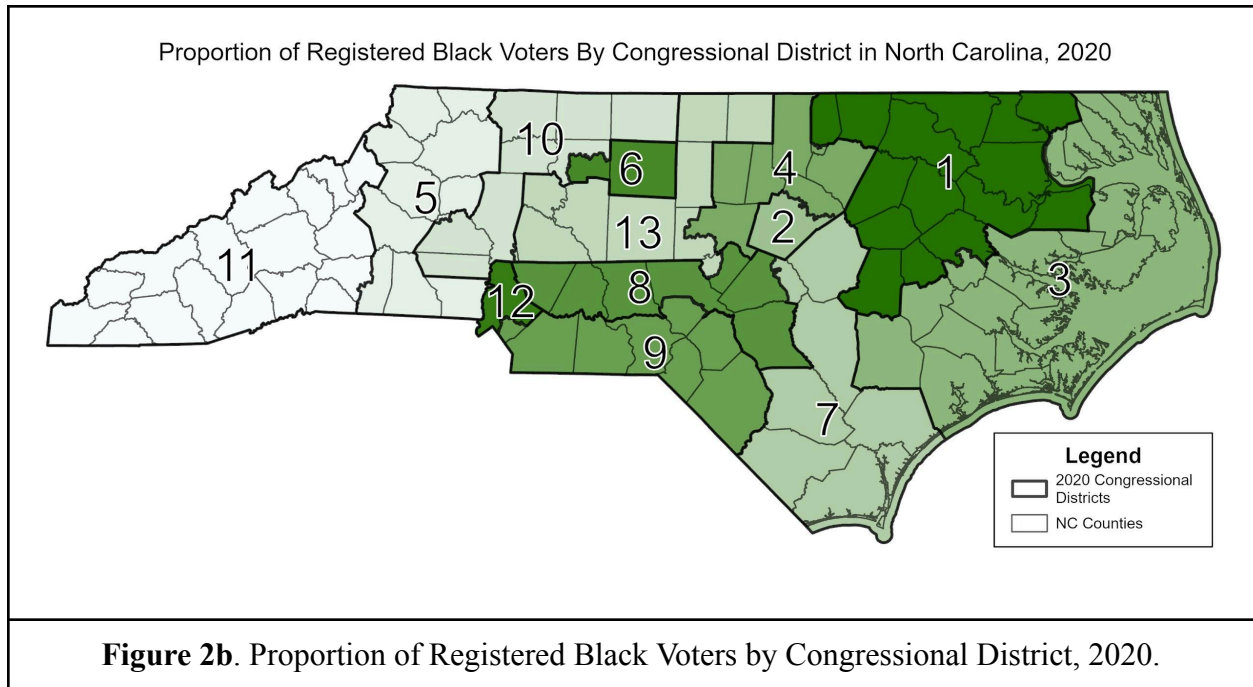




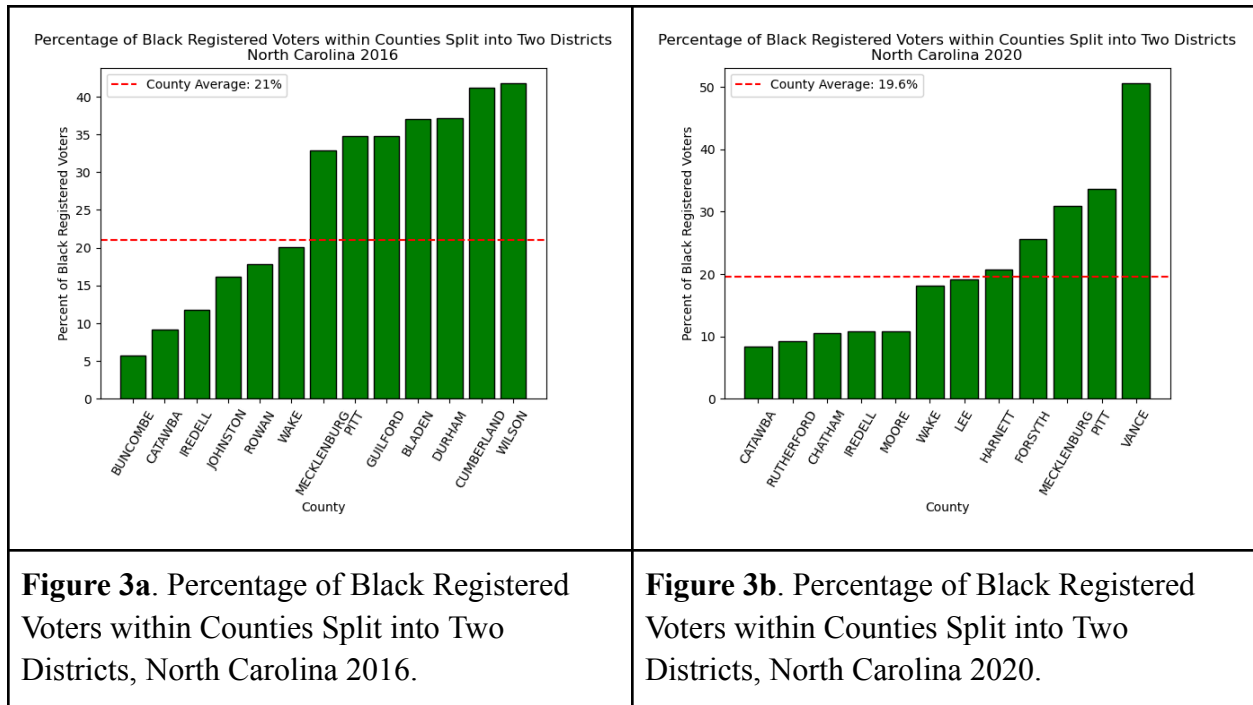
District, North Carolina 2016.	District, North Carolina 2020.
--------------------------------	--------------------------------

We were also interested in exploring changes in the spatial distribution of racial demographics, with a particular focus on North Carolina's Black population. In 2016, a large proportion of Black registered voters were concentrated in Districts 1 and 12. By 2020, as shown in Figure 2b, this concentration extended to District 4 and District 6. The spatial distribution of registered Black voters across the districts reveal significant shifts in voter concentration from 2016 to 2020. Notably, District 6 was redrawn in 2016, resulting in the separation of North Carolina A&T State University, the largest HBCU in the United States. By 2020, the district was redrawn to include the university in its entirety, resulting in a significant rise in Black voter concentration, as seen in Figure 2b.





Furthermore, we investigated shifts in the practice of cracking. To do this, we focused on counties that were divided into two districts. While county splitting is typically used to ensure balanced populations across districts, we aimed to understand how the percentage of Black registered voters in these counties changed after the redistricting. As shown in Figure 3a, counties that were split in 2016 accounted for 21% of the Black registered voter population, on average. By 2020, this percentage had decreased to 19.6%, on average. In 2016, seven counties that were split had a higher percentage of Black voters than the county average. However, by 2020, this number had dropped to five counties with above-average Black voter populations, as displayed in Figure 3b. It is also worth noting that the total number of counties split into two districts decreased by one between 2016 and 2020.



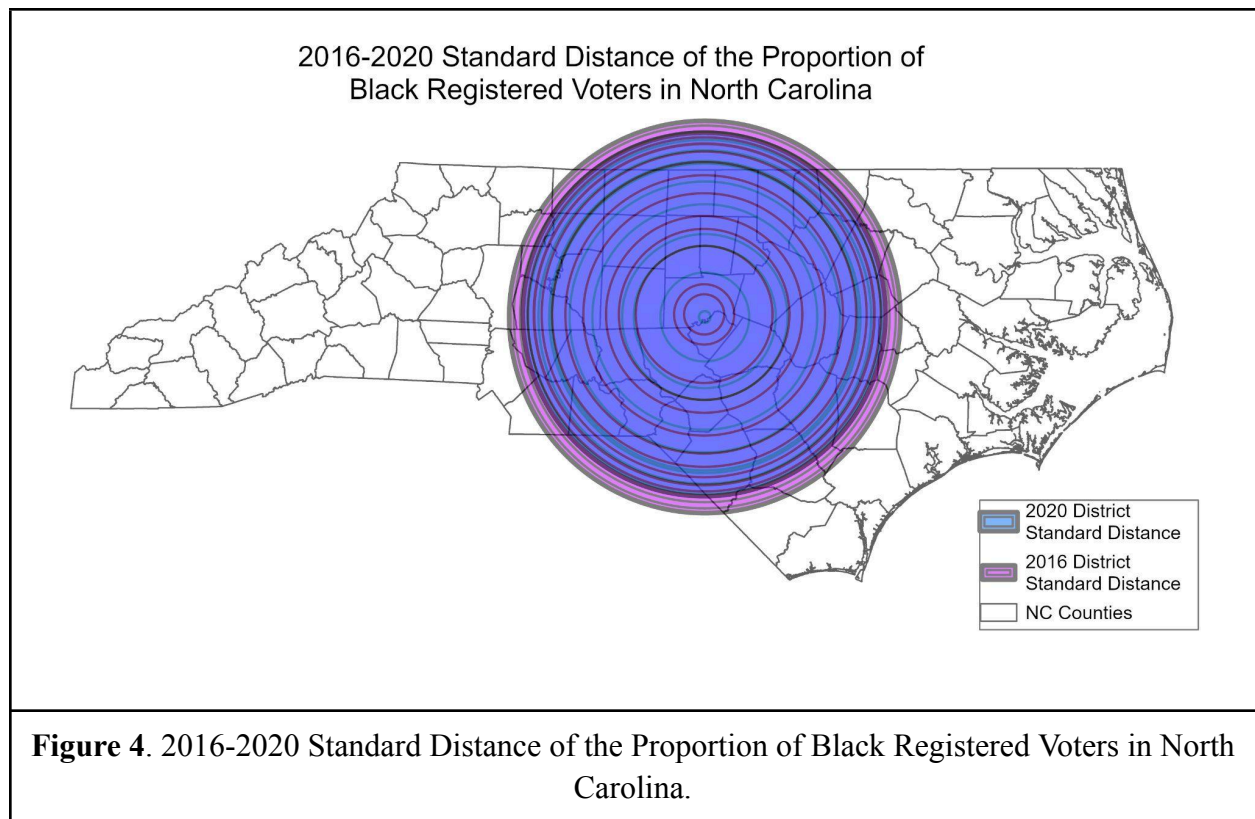
Lastly, we calculated spatial autocorrelation (Moran's I) and standard distance to better understand the clustering of registered Black voters over time. In Table 1, the Moran's Index dropped by 0.1 from 2016 to 2020, indicating that registered Black voters were distributed across North Carolina with fewer dispersion patterns. Comparing the results from 2016 and 2020, Standard distance decreased from 175210.38 units in 2016 to 162398.17 units in 2020, suggesting a trend that Black registered voters became more geographically concentrated over time, rather than being widely dispersed. Additionally, the mean center coordinates shifted slightly northeast, from (-8819056.84, 4243999.38) in 2016 to (-8819638.62, 4246323.38) in 2020. The center shift could reflect demographic and social dynamic changes, and/or variations in voter registration patterns over the four years.

**Table 1.** Spatial Autocorrelation (Global Moran's I)

<b>Global Moran's I Summary</b>	<b>2016</b>	<b>2020</b>
Moran's Index	0.191410	0.082530
Expected Index	-0.083333	-0.083333
Variance	0.045686	0.033190
z-score	1.285396	0.910432
p-value	0.198654	0.362595

**Table 2.** Spatial Statistics: Standard Distance

Year	Standard Distance	Mean Center X	Mean Center Y
2016	175210.38	-8819056.84	4243999.38
2020	162398.17	-8816938.62	4246323.38



## Discussion

The results demonstrate a clear shift in the geographic concentration of Democratic and Black registered voters across North Carolina's Congressional districts between 2016 and 2020. In 2016, the concentration of Democratic voters was highly localized, particularly in District 1, where we observed the packing of Democratic voters into a single district to reduce their influence in surrounding districts. The extreme concentration of Democratic voters in District 1, as shown in Figure 1a, ensured Republican dominance in neighboring districts. However, the 2020 distribution, as illustrated in Figure 1b, reflects a substantial shift toward a more balanced and competitive electoral map. The emergence of five majority-Democratic districts (Districts 1, 2, 4, 6, and 12) indicates a significant increase from 2016, when the Democratic party only dominated District 1. The increase in Democratic representation suggests that the 2020 redistricting plan reduced the instances of previous packing.

We specifically focused on Black voters in our analysis of voter racial composition between 2016 and 2020, given their historical oppression in North Carolina dating back to the Jim Crow era (Democracy NC, 2021). In addition, we compared White and Black registered voters as these two groups represented the largest racial demographics in North Carolina during both years. The spatial distribution of registered Black voters across the districts from 2016 to 2020 displays notable shifts in voter concentration. Specifically, District 6 saw a marked increase in Black voter concentration, likely due to the redrawing of district lines to incorporate North Carolina A&T State University. Additionally, the number of split counties decreased from 2016 to 2020. This aligns with our hypothesis that the 2020 redistricting would result in a higher concentration of Black voters in areas that were previously cracked.

Moran's I results show decreased spatial clustering, indicating that redistricting efforts redistributed Black registered voters across congressional districts. This aligns with our Standard Distance findings, which show a smaller radius in 2020 compared to 2016. The reduced Standard Distance suggests a more compact geographic distribution of Black registered voters. However, an increase of physical concentration does not inherently indicate improved representation. Figure 2a and Figure 2b reveal that, although clustering across districts decreased, packing persisted or even intensified in certain districts. Districts 1, 4, and 6 exhibit darker shading in 2020, indicating higher proportions of Black registered voters, which may suggest localized packing. While the reduction in Moran's I implies progress in mitigating cracking, the persistence of packing in specific districts shows continued inequities.

## **Conclusion**

Our results suggest that the 2020 redistricting plan reduced instances of both cracking and packing, but only in specific districts. Districts 1 and 12 saw a decrease in the packing of Democratic voters, likely due to their involvement in legal proceedings. Similarly, District 6, which had previously split North Carolina A&T State University, was redrawn to include the entire university, a move that suggests a reduction in cracking. However, our spatial analysis indicates that, overall, these changes did not result in a complete overhaul of the system. The 2020 redistricting efforts may have reduced the overall dispersion of Black registered voters, but at the cost of disproportionately concentrating their voting power, thus limiting their influence in surrounding areas. While the 2020 redistricting efforts in North Carolina represent a step forward in addressing the concerns raised by legal battles such as *Cooper v. Harris*, the ongoing patterns of packing minority populations indicate that further reforms are necessary to ensure that these groups have equal power in the political process.

**Group Contribution**

Amanda Li conducted the literature search, wrote most of the paper, and scraped text from PDFs. Kyle Guanzon worked on ArcGIS data and shapefiles, conducted spatial statistical analyses, and created an interactive ArcGIS Web Map for exploration into the findings. Hannah Wen web scraped the North Carolina registered voter demographics by county, joined the county and district demographic data, and conducted part of the data analysis.

## References

*Cooper v. Harris*. (n.d.). Oyez. Retrieved December 2, 2024, from

<https://www.oyez.org/cases/2016/15-1262>

*“Election Integrity” Act is Jim Crow 2.0 Voter Suppression*. (2021, March 30). Democracy NC.

<https://democracync.org/research/election-integrity-act-is-jim-crow-era-voter-suppression/>

Fernandez, M. (2021, May 11). Power on the line(s): Making redistricting work for us. *Legal*

*Defense Fund*. <https://www.naacpldf.org/redistricting-report/>

Jenning, L. K. (2021, September 29). As communities of color grow, racial gerrymandering takes center stage. *YES! Magazine*.

<https://www.yesmagazine.org/democracy/2021/09/29/census-gerrymandering-racial-redistricting>

Meyers, D. (2019, November 7). Experts identify the worst examples of gerrymandering. *The*

*Fulcrum*. <https://thefulcrum.us/worst-gerrymandering-districts-example>

*North Carolina General Assembly*. (n.d.). Retrieved December 2, 2024, from

<https://www.ncleg.gov/>

*Rucho v. Common Cause*. (n.d.). Oyez. Retrieved December 2, 2024, from

<https://www.oyez.org/cases/2018/18-422>

*U.S. Census Bureau quickfacts: North Carolina*. (n.d.). United States Census Bureau QuickFacts.

Retrieved December 4, 2024, from

<https://www.census.gov/quickfacts/fact/table/NC/PST045223>