

Abstract

Life course racism refers to the systemic discrimination that individuals from marginalized racial or ethnic groups experience throughout various stages of their lives. This perspective suggests that each life stage influences the next, and together the social, economic and physical environments in which we live profoundly impact our health. For our project, we endeavored to quantify life course structural racism. We analyzed it via two pathways (structural violence and limited/restricted resource access) and three domains (government representation, environmental pollution and media & marketing) using county-level data. We found that across time periods, a county was more likely to experience significant racism in a time period—regardless if they did or did not experience significant racism in the period before. While further research will not be done until after this course, Dr. AJ will continue to examine the impact of life course structural racism on dementia in older Black adults.

Background

The Multifaceted Life Course for Structural Racism (MLCSR) measure is a compilation of publicly available datasets that capture aspects of structural racism. It looks at structural racism through two pathways (structural violence and limited/restricted resource access), ten domains (residential segregation and gentrification, property ownership, government representation, policing and incarceration, income and poverty, occupational segregation and employment, healthcare and healthy food access, environmental pollution, and media and marketing) and three historic periods: before the Civil Rights Act of 1968, Desegregation/Integration (1969-1999) and Modern Times (2000-present). For this task, we will use a life course perspective to examine how government representation, media & marketing and environmental pollution interact with structural racism to affect dementia in older adults racialized as Black. Thus, this raises the question: **Does structural racism have an influence on dementia?**

We hypothesize that greater exposure to life course structural racism increases the risk of incident dementia and cognitive decline in late life. The nature of this project lends itself to multiple phases and we are solely contributing to the first. In this data collection phase, we hope to tune our variables of interest and collect data that begins to capture something as inherently abstract as structural racism. The following phases, which will not take place until well after this course is over, mainly involve analyzing the impact of structural racism on dementia for people racialized as Black.

Analytic Approach/Statistical Model

Participants were a community-dwelling sample of older adults racialized as Black over 50 years, in the study. We linked participant data by county at birth, at adolescence (ages 10-18), adulthood (ages 19-50), and midlife (ages 51 and older) to the aforementioned datasets.

Due to the nature of our project, we first need to develop our own binary criteria (0,1) for structural violence in each item under different domains and pathways, and then we collect or scrape data from public sources in order to actually label it. Instead of working with direct data due to a lack of sources, most of the time we have to work with proxy data. For example, we may not be able to get the data on where Black residents lived before the Civil Rights Movement, so data on redlined areas may help determine where they lived because redlined areas are highly correlated with people of color.

Regarding statistical models, time series analysis is a key component of our research, since our project spans across three different time periods. Time analysis enables us to, for example, explore how structural violence has evolved and impacted the targeted population over time.

Literature Review

For the Environmental Pollution domain, some previous studies (Christopher, et al., 2021) have been conducted to reveal the fact that regions where people of color live were facing more severe fine particulate matter (PM2.5) pollution, which can cause lung or heart problems. However, few studies have focused on the impact of Hazardous Air Pollutants, or HAP, on these regions. HAPs, such as Benzene, are air pollutants emitted by industries or facilities that can cause cancer or other health effects. PM2.5 can sometimes be the carrier of HAPs, but often these two things are discussed separately. Therefore, we decided to work with the county-level HAP data of industries and facilities to determine if counties with a higher proportion of people racialized as Black than the national average are facing more HAP emissions from local facilities. Regarding Superfund Sites, on the other hand, some research suggested that people of color live disproportionately close to superfund sites (Johnson, 2020), which, distinct from conventional landfills, are locations that are polluted by hazardous wastes. The research first shows that most superfund sites are located in the proximity of government-assisted housing, and then states the fact that a considerable portion of residents living in this kind of housing are people of color, thus proving the point that people of color are facing environmental racism. Our approach, in comparison, is to find if counties with a greater proportion of Black residents than the national average in the desegregation period are more likely to have superfund sites.

Black media plays a pivotal role in amplifying the narratives, experiences and stories of the Black community. In 1827, the first African American newspaper in the United States was founded: *Freedom's Journal*. Unlike other publications of the time, it provided a crucial platform

for Black voices to address important issues and challenge harmful stereotypes. During its run, *Freedom's Journal* was adored by many in the community because it humanized individuals that were often dehumanized by society and mainstream media. Since its debut, more than 3000 African American newspapers have appeared and disappeared (Library of Congress). Between 1827 and 1865, there were as many as 40 African American newspapers and magazines. Post-Reconstruction years also gave rise to the greatest increase in African American newspapers (Library of Congress). Black folks had more educational opportunities, a growing urban middle class and the support of religious groups. Now, many of these publications cease to exist due to the lack of monetary support and racism. While there is no county-level data for Black newspapers, we aim to explore the limited access to Black media during time period 1.

The Voting Act of 1965 was a milestone marking a major shift for Black people in politics. This important piece of legislation prohibited racial discrimination in voting practices. Upon its ratification, the Black community was finally able to *legally* participate in the nation's system of democracy. In fact, this amendment did not just allow for a significant increase in the number of registered Black voters; it opened the door for more Black legislators across all levels of government. Within the following decade, from 1970 to 1980, "the total number of Black elected officials in the United States tripled, from just 1,469 to 4,912." Ideally, this act would have allowed Black voters the right "to elect representatives of their choice"; however, implicit bias and racist policies mitigated the potential for diverse representation throughout the branches of government. This was especially true for the judicial branch.

The United States has two court systems—the federal court system and the state court system. Federal courts consist of three main types: District Courts, Courts of Appeals, and the Supreme Court. District Courts (i.e., "small case courts") hold trials for cases, in which the district judge applies legal principles to make a decision for or against the defendant. The Court of Appeals hears appeals and challenges to decisions made in trials from the District Courts. The Supreme Court is the highest court in the United States. Supreme Court judges defend the Constitution by holding the final decision on constitutional questions and other appeals regarding constitutionality. Unlike in state courts, decisions made in the federal court system do not just affect the plaintiff and defendant, they hold the power to affect the entire nation. A prime example is *Plessy v. Ferguson*—whose ruling not only required Black train passengers to ride in partitioned cars, but also legally upheld racial segregation across the nation for almost seven decades. Court rulings, such as this, highlighted the dire need for such an influential and authoritative system to have better Black representation.

From 1970 to 2000, there have been over 1500 federal judges appointed in the United States; only 125 of them were Black. In an era of civil and political unrest amongst the Black community, researchers have questioned if the lack of Black judges within the federal court system has had an impact on significant trials. In 1988, a study by Welch et al. demonstrated that while it is difficult to definitively state what impact Black judges have on cases, "in the crucial

decision to incarcerate, having more black judges increases equality of treatment.” In fact, a more recent Harvard study conducted by Dr. Sen demonstrated patterns amongst the evaluation of rulings made by Black federal judges versus that of their white counterparts. The study indicated that case decisions made by Black federal district judges were more likely to be overruled than that of white judges, by up to ten percentage points. It also clarified that this increased likelihood “cannot be explained away by factors such as qualification disparities or types of cases.” The presence of Black judges within the federal court system has created—and continues to create—a significant impact; however, the potential of this impact is diminished because of their disproportionately low representation and systemic biases of the environment around them.

Our work also focuses on the lack of Black representation in our state government. In order to understand why this is an issue, one must be familiar with the structure of our state government. The state government is modeled after the federal government and follows the same three branch structure. The executive branch is headed by the state governor, with six other positions that are held in all fifty states: attorney general, superintendent of schools, insurance commissioner, agriculture commissioner, labor commissioner and public service commissioner. Beyond this, the structure varies after this from state to state because of the freedom that individual states are given in designing their government. In every state except Nebraska, the legislative branch, Congress, consists of the Senate and the House of Representatives. The Senate is the smaller upper chamber where officials serve longer four year terms and the larger lower chamber where officials typically serve two year terms is the House of Representatives. Finally, the judicial branch is led by the State supreme court. The court’s structure and judicial appointments are determined by legislation or the State constitution (“State and Local Government.” *The White House*).

Next, it is crucial to think about the power that the state government has. State governments play an important role in our daily lives because they have specific responsibilities that are *solely* theirs. These include but are not limited to; education, running welfare and social benefits programs/general distribution of aid, regulation of industry, maintaining state highways and local roads, and protecting people from local threats and maintaining a justice system. This is a wide range of responsibilities that might seem intangible at first, however we quickly start to see how race intersects with the state’s powers when we start to get more specific. For example, house members heavily contribute to the policy work for things like criminal justice reform and voting rights, two issues which have historically impacted Black folks disproportionately (Maloy, and Trust. “Functions of State and National Government.”).

Many proponents of equal Black representation in government are well aware of the fact that more Black politicians does not necessarily mean an emphasis on Black rights and centering Black issues—yet it is still an indicator of a healthy and equitable democracy. Descriptive representation is a measure of how well an elected representative resembles their constituents.

Descriptive representation is necessary for many reasons beyond the push for equal representation in government as a benchmark of equality. For example, there is research supporting the claim that Black politicians do a better job attending to the needs/interests of Black Americans than any other demographic (*Black Politicians Are More Intrinsically ... - Wiley Online Library*). Furthermore, Black state representatives are necessary because their election can inspire more Black folks to participate in politics and they are more likely to initiate policies that the mass Black public supports, among many other things (*When and Why Minority Legislators Matter - Annual Reviews*).

Now that we have established the need for Black representation at the state level, it proposes the following question: How well are Black folks currently represented in the state government? There is a plethora of research available to support the claim that it is not good enough as is. For starters, 89% of officeholders nationwide are white. That leaves only 11% of these seats for various minority groups with many differing interests. Representation is severely disappointing in the US Senate and state governor's offices, and even local government positions and the US presidency when you look beyond the state level. Building on the history of Black participation in politics outlined above, voter suppression efforts and gerrymandering are two of the biggest modern restraints on the political power of Black voters. Despite these glaring issues, state legislatures are the sole branch of state government that accurately mirror the Black population in the US. For example, more than half of state legislatures had either nearly proportional or overrepresentation of Black folks in 2020. On the surface this seems like a step in the right direction until you break it down to the house and senate levels. There are currently only three Black senators— this is even more disheartening when you recognize the authority that the Senate has relative to members of the house. At the conclusion of our literature review, one thing is clear above all else: Black folks are underrepresented in our government at the state level, which compels us to dig deeper and analyze this statistically (Noe-Payne, Mallory. "Black Representation in Politics Is Poor Nationwide, except in State Houses.", "The State of Black Representation in the US Today." *Public Wise*).

Items

In the environmental pollution domain, EP.SV3 captures if counties with more Black populations suffer from more severe air pollution in the year 2020. This item was configured using data from Census.gov and United States Environmental Protection Agency (EPA)'s National Emissions Inventory (NEI). To compute this item, counties with Black population percentages greater than or equal to the national average in the year 2020 were marked first as target counties. Then, target counties with annual hazardous air pollutant (HAP) emissions in the year 2020 greater or equal to the non-target counties' mean annual emissions were found. For this item, 1 signifies a target county with HAP emissions in the year 2020 equal to or exceeding the mean emissions of non-target counties, and 0 signifies a non-target county or a target county with HAP emissions in the year 2020 less than the mean emissions of non-target counties.

EP.SV2 captures if counties with more Black populations contain a landfill site. This item was configured using data from the Integrated Public Use Microdata Series (IPUMS) and EPA's Landfill Methane Outreach Program (LMOP). To compute this item, firstly, target counties (same definition as target counties in EP.SV3) in 1990 are marked using 1990 IPUMS data. Next, landfill sites were filtered so that only sites that were active before and in the year 1990 were kept. Then, all counties that have such a site within their boundary were found with the help of Google Geocoding API. For this item, 1 signifies a target county in 1990 that has a landfill site that is active before or in the year 1990, and 0 signifies a non-target county or target counties that do not have an active landfill site before or in the year 1990.

EP.LRA2 captures if counties with more Black populations contain a Superfund site. This item was configured using data from IPUMS and EPA. To compute this item, firstly, target counties (same definition as target counties in EP.SV3) in 1990 are marked using 1990 IPUMS data. Next, counties with a Superfund site that were in the National Priority List were listed. For this item, 1 signifies a target county in 1990 that has a Superfund Site in the National Priority List, and 0 signifies a non-target county or a target county that does not have a Superfund Site in the list.

EP.SV1 captures the severity of natural disasters in once redline areas. This item was configured using data from the United States Federal Emergency Management Agency (FEMA) and Riley Champine and Justin Madron's HOLC research. To compute this item, first, data from the HOLC research were cleaned to find the redlined neighborhoods in terms of 2020 census tracts. Then, the disaster dataset was used to find counties that have the number of natural disasters not below the national average level. For this item, 1 signifies a county that contains a redlined neighborhood (HOLC rating below "B") and have equal or more occurrences of natural disasters than the national average, and 0 signifies a county that does not contain a redlined neighborhood or counties that contain a redlined neighborhood but do not have more occurrence of natural disasters than the national average.

EP.LRA1 captures the accessibility of public parks in once redlined areas. This item was configured using data from Trust For Public Land, Riley Champine and Justin Madron's HOLC research, and Esri Data and Maps. To compute this item, firstly, the shape files of the once redlined areas (HOLC rating below "B") and the 10-min-access range of all the parks in the United States are imported into ArcGIS and overlaid to the US-county layer. Then, the geographical intersection between redlined areas' shapes and park access shapes are taken. Finally, counties that have at least one redlined area and are not completely covered by park access range are marked. For this item, 1 signifies a county that contains at least one redlined area and is not completely covered by the 10-min park access range, and 0 signifies a county that does not contain redlined areas or contains a redlined area but completely covered by the 10-min park access range.

The MM.LRA1 variable measures limited and/or restricted resources within the media & marketing domain during time period 1 (before 1968). This item was configured using the Historical African American Newspapers Available Online archive at the James A. Cannavino Library. To compute this item, years before 1968 were filtered from the dataset. Then, each county that had a Black newspaper was identified. Counties that did not have a Black newspaper prior to 1968 received a 1, signifying that racism was present. Whereas counties that had a Black newspaper received a 0, signifying that racism was not present in this domain.

The GR.SV2 variable measures structural violence within government representation during time period 2 (i.e. 1969-1999). This item was configured using data from Wikipedia's "List of African American federal judges" and the "Jurisdiction" section of all 94 District Courts' websites. To compute this item, the count of Black judges that were active in service for each District Court for every year starting from 1969 to 1999 was aggregated. Following, this count was used in tandem with information on each District Court's jurisdiction to calculate the count of active Black judges per county. For each county, a GR.SV2 input of 1 signifies that there were no Black judges in the District Court presiding over a specified county from 1969 to 1999. Opposingly, 0 signifies that in this time there was at least one Black judge in the presiding District Court for the county.

The GR.LRA2 variable measures the Black community's limited resource access within government representation during time period 2 (i.e. 1969-1999). This item was configured through a transformation of the information from the annual breakdown of GR.SV2. To compute this item, each county that had one or more increases in the number of Black federal judges serving from 1969 to 1999 was identified. The counties who had at least one increase received an input of 0, signifying that there was no racism experienced in that county within the scope of this variable. For the rest of the counties—all of which either had the number of Black federal judges decrease over time or remain stagnant—received an input of 1, there was racism experienced in that county within the scope of this variable.

The final variable of interest that we focused on in government representation is limited resource access in the modern times (GR.LRA3). We chose to quantify this by investigating the ratio of Black representation in state offices to the Black population in that state, for any decade between 2000-2020. This item was configured using data from the various sources listed in the table below. To compute this item, we computed the percentage of elected Black state representatives in each state from 2010 to 2020, then compared this to the percentage of the Black population in the same state. In this instance, adequate representation means the percentage of Black elected officials is greater than or equal to the percentage of the Black population in the state. 1 signifies inadequate representation in government and 0 signifies adequate representation in government. Each county within the same state will have the same value for this variable since we collected data at the state level.

Process

Environmental Pollution domain, Time Period 1

For the first time period, or before the Civil Rights Movement, a large portion of the datasets are in a shapefile format, which can be read using ArcGIS. The workflow is as follows.

The first step was to import every shapefile into ArcGIS. The base layer is a map of the United States having boundaries and details of each county in the year 2020. Next, HOLC (redlined areas) shapefiles, in terms of 2020 census tracts, were overlaid on the base layer. Then, some spatial operations, such as spatial join and spatial intersection, were done in ArcGIS to mark counties with at least one neighborhood rated by HOLC below “B”, which formed the target counties in this time period. Next, environmental datasets or shapefiles were imported into ArcGIS. The new datasets or shapefiles were then joined to the combined shapefile in step 1 either spatially or using shared features like FIPS codes. For example, for EP.LRA1, the new shapefile is the 10-minute access range of all the parks in the U.S. A spatial intersection was done between the new park access shapefile and the existing county-level map to show what percentage of the county was covered by the park access range. Finally, the shapefiles with the new attributes added were exported to allow for the final “and” operation. The “and” operation here mainly involves two attributes: 1) is the county a target county; 2) is the county under environmental racism or limited access.

Environmental Pollution domain, Time Period 2

Regarding the Desegregation period, or period 2, the data analysis was divided into two main parts.

The first step was to gather demographic data from IPUMS. Notice that the demographic data source has switched from Census.gov to IPUMS. This is because IPUMS has more comprehensive and accurate census data in earlier times. Same as before, target counties were marked using the same criteria, except that here the time changed from 2020 to 1990, which is the year that most datasets were available. The second step involves the introduction of a series of landfill datasets from the EPA. For each dataset used, the location of each landfill site was matched to the corresponding county, which will be discussed later in this section. Then, a county is marked as 1 if that county has a certain type of landfill site, based on the item we are considering. The final step is again taking the “and” between the labels of the first and second step, and outputting the final label of the items.

When working with the data from this time period, one major difficulty was to match city names to their corresponding county names. The cause of this problem is that some dataset only includes the state name and city name, but no county names. This problem was solved using the

Google Geocoding API, which takes the input of city names and returns the county they belong to.

Environmental Pollution domain, Time Period 3

Regarding the modern time period, or time period 3, the data analysis process follows a general pattern. Firstly, using 2020 nationwide demographics census data from Census.gov, the percentage of Black populations of each county is calculated, and each of these percentages are then compared to the percentage of Black population in the whole nation in the year 2020. If it is higher or equal to the national average, then it is marked as a target county. Otherwise, it is a non-target county. Then, environmental pollution-related datasets are introduced into analysis. EP.SV3, for example, that dataset is the 2020 county-level Hazardous Air Pollutants (HAP) emission dataset. On this dataset, a statistics is generated from the existing features to determine if a county is experiencing excessive environmental pollution. Then, this new dataset with features engineered was joined into our main table for the next step. The final step was to do an “and” operation on the two criteria: 1) is the county a target county? 1 stands for yes, and 0 stands for no; 2) is the county experiencing excessive environmental pollution? Again, 1 stands for yes, and 0 stands for no. Then the final label for the item for this county is the output of this “and” operation.

Although there were very few missing values in this time period, one difficulty we encountered was to join different tables based on geographical identification. FIPS code formats differ from dataset to dataset, so finding the pattern of differences and writing code to unify the formats was one of the key challenges of the analysis.

Media & Marketing domain, Time Period 1

For the media and marketing domain, the item originally looked different. The variable was formerly defined as the number of times a Black newspaper was cited in the *Green Book*, a guidebook that deemed safe locations for Black travelers. The goal was to list each Black newspaper in the *Green Book* and find the county it corresponds to. If the county was listed in the *Green Book*, then it would receive a 0 and 1 vice versa. Upon reviewing all available editions of the *Green Book*, only two years mentioned a Black newspaper (1946 and 1947). Given that two years would not be a sufficient amount of data, the variable was redefined and a new dataset was used.

In our second search for data, we came across an archive of African American newspapers from the James A. Cannavino Library. The library had Black newspapers from various time periods, such as the Antebellum and Civil War Era, Reconstruction, Gilded Age, Pre-World War I, World War I - World War II and the 20th and 21st Centuries. The archive also included useful information such as the newspaper name, city, state, available editions and the

years the newspaper ran. Given we had a new dataset, we decided to redefine the MM.LRA1 variable to be the presence of a Black newspaper in a county. If a county had a Black newspaper, it was coded as 0 and 1 otherwise.

One key thing to note about the data cleaning process for this dataset was that the years were structured as a character value instead of numeric values. For example, the years, 1883, 1884, 1885 and 1886, were structured as 1883-1886. We manually had to change the years in the text file before writing any code. Once that was resolved, the county names had to be assigned to each city in the dataset. One problem that was persistent throughout the project were the county names. Initially, we wanted to merge an existing dataset that contained the city names and county names with the media and marketing dataset, however we realized that this method would not render the results we needed. The dataset with the county names did not contain all possible county names, only present-day counties. For context, there are currently 3143 counties, however there have been a total 3258 county names over time. As aforementioned, county names and boundaries have changed; counties that were present in time period 1 were not present in time period 2 or time period 3. In the city of Miami, for instance, the counties were changed from “Dade County” to “Miami-Dade County. For this item, we decided that it would be best to manually assign the county names given that they were listed in each original newspaper.

Lastly, while the Black newspaper archive that we found did not contain any “missing” values, it is important to note that not all Black newspapers were listed in the archive. Given that everything from this time period is paper, it would have been difficult to retrieve these historical documents because printed newspapers were not saved or stored onto an online database in the early 1800s. Only the digitization of select historical newspapers has been supported in recent times through the National Digital Newspaper Program.

Government Representation domain, Time Period 2

Initially, the variable for government representation in period 2 (i.e., 1969-1999) looked slightly different. The original variables were supposed to identify counties with Black representation in local offices. However, dealing with local offices in this time period presented several challenges. The main challenge stemmed from the fact that local governments are heavily saturated. There is no end of offices, agencies, and positions in local government. This, in turn, makes it hard to confirm that all Black people who have ever worked in local government are properly accounted for across 50 states over the course of 30 years. This problem, coupled with the fact that there is already a scarce amount of data for this time period, caused us to reassess the definitions of our variables. As to not compromise the thoroughness and accuracy of the variables, we decided to explore other levels of government to quantify this metric.

When looking into the data available on state government positions for this time period, we ran into the same issues that we had with the local government. As a result, we decided that national offices were the best direction for these variables, since there are ample verified records

about positions held on the national level. The only issue is that we are only interested in state-level or county-level data, and most national offices pertain to serving the nation as a whole. The only viable options were to look into representation in the House of Congress or the federal court system (specifically the District Courts and the Courts of Appeal). Upon reading the 1988 paper “Do Black Judges Make a Difference?”, we realized that there was an already-proven relationship between Black representation of federal judges and the livelihood of the Black community. As a result, we redefined our variables to focus on representation of Black federal judges in the District Courts.

Originally, the variables GR.SV2 and GR.LRA2 were supposed to include Black judges who actively served in the Court of Appeals and the District Courts from 1969 to 1999, however there were two reasons why this was not viable for the variables. First, the Court of Appeals has jurisdiction over entire states, not counties. Because most of the other data in the set is on the county level (or generalized to be on the county level), using information about the judges on the Court of Appeals would be inconsistent. Opposingly, data on the District Courts existed on the county level. Additionally, by 1999, there had been at least one judge actively serving in every circuit within the Court of Appeals. Thus, for variable GR.SV2, all counties for all states would have had representation and been coded with a 0. That did not convey much information about the *true* representation of Black judges nationwide throughout time period 2; thus, the data collected for the Court of Appeals was no longer used.

In order to encode the variables GR.SV2 and GR.LRA2, information listing all the Black judges who have ever served in the United States District Courts was scraped, cleaned, and wrangled. Ideally, this data would have been joined with a publicly-available dataset detailing the jurisdiction of each federal district and circuit. Unfortunately, such a dataset did not exist already. Moreover, there was no local place that listed the county-level jurisdictions of all the federal districts in the United States; all 94 districts had that information detailed in varying places on their individual websites. As a result, two datasets were created by manually searching for and inputting the needed information in an excel sheet until all states and counties of the United States (including Puerto Rico) were accounted for. This was especially challenging because of changes made to counties and federal districts in time period 2. For example, in 1997, Dade County was renamed Miami-Dade County. Additionally, on October 1, 1981, the Eleventh Federal Circuit was created as a subset of the Fifth Federal Circuit, including federal districts from Alabama, Florida, and Georgia. To ensure the accuracy of the datasets, research was done about each individual federal circuit, district, and county to see if any changes went into effect during the time period. The resulting datasets detailed each county in the United States, the state it is in, the federal district the county belongs to, and the federal circuit the district is in.

Government Representation domain, Time Period 3

It is extremely difficult (likely impossible) to find an all encompassing dataset of state elected officials from 2010-2020 that includes demographic information. There are many roles at the state level that only exist in certain states, within the executive branch alone. The lack of uniform offices and centralized data can be attributed to the freedom that states are given in governing themselves. As a result, we chose to look at four specific offices and use them as a case study of the remaining positions at the state level. These positions are Congress (both the house and the senate), state governor, state attorney general, and state supreme court justice. For all four, it was more feasible to look for an exhaustive list of every Black person to hold each office ever. Then, essentially working backwards, we used the details of their terms to narrow it down to those that served between 2010 and 2020. Finally, to find the percentage of Black state representatives for each role, we divided by the total number of people who served during that same time period. Again, due to the lack of centralized information regarding state officeholders, it was impossible to find an *exact* number for the total number of people who served in each position from 2010 to 2020. We had the option to either estimate this number using an LLM or other information online, but our entire team decided we should use the most conservative estimate possible. For example, there are 52 house members in the state of California, so we assumed that all 52 people held their seats for the entire decade from 2010 to 2020. This is a very naive assumption, and almost certainly an underestimate in most offices/states during this time. However, Black people are still significantly underrepresented in our state government when we aggregate the data by summing across positions and finding the percentage of all state representatives who are Black, regardless of the office they held.

A few other important notes from the data collection and cleaning process include the handling of different types of “missing” data. We collected data spanning over more than a century at the county level, but county names and boundaries have changed over that time span. For example, there are numerous counties in Virginia that were renamed over time. This means we must include them in our final merged dataset, but there is no data for them in some time periods. We chose to impute missing values of counties with the same value from other counties in the same state for the GR.LRA3 variable. We imputed the percentage of the Black population with the statewide mean.

We also had to be meticulous in computing the total number of eligible seats per position because of the criterion that these people must be elected. For example, the Colorado State Supreme Court is one of many that has its justices appointed by the state governor. For this reason, the total number of elected state representatives in Colorado does not include the seven seats from its state supreme court.

This process was more difficult than expected and we had to be creative about the way we searched for the data that was relevant to our analyses. Once we settled on this “backwards” approach, it made things much easier. It was disappointing to realize that there was no widely available answer to a question as simple as, “What percentage of state attorney generals in the U.S. were Black, from 2010 to 2020?” Furthermore, there is no widely accessible data that one can use to easily find the answer to this question. Scraping the data from different sources and being intentional about the different assumptions we made before merging the data is extremely

important work— this is why the majority of our labor went into compiling extensive datasets that can only begin to answer deceptively complex questions, including “Which states in the U.S. had sufficient Black representation between 2010 to 2020?”

Results and Exploratory Data Analysis

For figures 4-7 in Appendix B, each figure shows the extent of racism in its respective time period. Notice here the color of each county is determined by a score we coined to create a weighted sum of all the items in that time period for a better visualization. The higher the score is, the darker the color is, and the more severe racism that county encounters. The score is defined as follows:

$$\frac{1}{|I|} \sum_{i \in I} \log \left(\frac{n}{\sum_{j \in L_i} 1_{\{j=1\}}} \right)$$

In this formula, I is the set of items in a time period; L_i is the set of labels of the item i (i.e. set of 1/0 labels of all counties under that item); n is the number of counties.

The intuition of this score is that it overweighs items that are very sparse (meaning that the item is 1 for only a few counties) and downweighs items that are not sparse. After the score of each county is calculated, standardization is done.

From the plots, one can observe that across all three time periods, large cities are subject to more severe structural racism than other areas. For instance, New York City, NY, Miami, FL, and Houston, TX, are indicative of this pattern. The cause of this pattern might be complicated, but one hypothesis is that redlining activities are mainly conducted in populated cities, and that redlining activities may also be more well documented in large cities, thus the pattern.

Another observation is that the extent of structural racism is diminishing for most parts of the country, as the major color of the maps becomes lighter as time period progresses from period 1 to period 3. This reflects the historical trend that more efforts were made to address structural racism after the Civil Rights Movement. Notice that the score is not related to the number of items in that time period, since the raw score is divided by the number of items during that time period. Therefore, the number of items will not influence the result.

A comprehensive map (figure 7), which includes all items we worked on across all the time periods, is included here to better visualize the geographical patterns of structural racism. Notice here the score is just an unweighted sum of all the items. Apart from the patterns we discussed before, speaking on a larger scale, the eastern and southeastern U.S. are subjected to more severe structural racism activities than the western or northern part. This may be caused by some complicated historical reasons.

Figure 3 in Appendix B is an alluvial diagram representing the flows of counties and their experiences with racism throughout each time period. Specifically, it illustrates the number of counties experiencing significant racism across the time periods. It is important to note that this alluvial only considers eligible variables—those developed during the scope of this capstone, not the project overall. That considered, a county is identified as experiencing significant racism in a given time period if it experienced racism in more than 50% of the eligible variables set in the time period. For example, since time period 1 has three eligible variables (i.e., EP.LRA1, EP.SV1, and MM.LRA1), a county would experience significant racism in this time period if it had coded inputs of 1 for at least 2 of the variables.

According to the alluvial diagram, a majority of counties were not experiencing significant racism in time period 1. This result is particularly surprising considering what is commonly known to be true about the experiences of the national Black community during this time (e.g., slavery and segregation). This specific result in the diagram is affected by two of the three eligible variables that are from the Environmental Pollution domain. As we see in the bar chart (Figure 1 in Appendix B), a majority of counties did not experience racism in all variables in the Environmental Pollution domain. This considered, the results of the alluvial may be more on par with what we know to be true when the rest of the variables for time period 1 are coded and taken into account. Nevertheless, the diagram shows that a majority of counties not experiencing significant racism in time period 1 did experience it in time period 2. In fact, across time periods, a county was more likely to experience significant racism in a time period—regardless if they did or did not experience significant racism in the period before. It is also worth mentioning that from time period 2 to time period 3, the number of counties that went from experiencing racism to not experiencing racism is almost equal to the amount of counties that went from not experiencing racism to experiencing racism; the count is off by only 7 counties. This suggests that with only eligible variables considered, once a county experiences significant racism in a time period, it is likely that it will continue to experience significant racism going into modern times.

Future Works

In the future, we plan to run some analyses on the data regarding the Black judges who have been appointed to serve on a federal court from 1970 to 2000. We also plan on looking into the breakdown of federal judicial districts throughout those years. We know what they are in current times, but we are unsure if they have changed in the past; thus, we are proactively taking the necessary measures to confirm the accuracy of each state's count. When all the data has been gathered and processed, we plan to visualize our findings through a temporal map—a map with user controls to explore data over time. Ideally, this will allow viewers to see how Black representation in the federal courts changed throughout the 30 years of this period.

We will also finish collecting and merging all of the state representation data so the result is one clean dataset that we can aggregate to find the answer to the original question: What is the ratio of Black representation in state offices to the Black population in that state, for any decade between 2000-2020? From there, we hope to analyze our results across time periods and domains to investigate how these things interact with one another under the large umbrella of structural racism and violence against people racialized as Black.

In the future, Dr. AJ hopes to use the MLCSR to further research and investigate the role of multifaceted structural racism on incident dementia and mild cognitive impairment in older adults racialized as Black. Previous research by the Alzheimer's Association and Pohl et al. (2021) suggest that interpersonal and structural racism are contributors to racial and ethnic disparities in cognitive aging ("Experiences of Racism Associated with Poor Memory, Increased Cognitive Decline"; "Relationship between Residential Segregation, Later-Life Cognition, and Incident Dementia across Race/Ethnicity"). Additionally, we hope to develop other items in the MLCSR, such as policing/incarceration. By defining other items in the MLCSR, we can explore how these variables interact with each other and how they influence cognitive aging.

Conclusion

The life course perspective emphasizes the significance of specific life stages in shaping health outcomes and emphasizes the link between individuals and social institutions. Thus, defining each item within the MLCSR was crucial. As marginalized individuals, we prioritized data collection over results, aiming to ensure data quality before training any machine learning model. Data isn't solely numerical; it represents the stories of individuals and their communities. Compiling a dataset like the MLCSR holds promise for enabling future ML models to understand abstract concepts such as racism. Historically, ML algorithms tend to perform poorly on data pertaining to Black folks. Per Gianattasio et al. (2020), current dementia predictive algorithms are ill-suited for racial/ethnic disparities work given differences in sensitivity and specificity by race/ethnicity. While we did not complete all 60 variables in the MLCSR, the nine we completed paint a much larger picture of structural racism and violence in the United States over a long period of time.

Ethical Considerations

Regarding dementia research, it is important to consider how cognitive impairment may affect one's decision making; participants may have no recollection of being in the study. It is also imperative that no participants' personal information is released after the research is published.

Appendix A: Contributions

1. Background
 - a. Abstract - Woomy
 - b. Research question, problem statement, motivation - Woomy
 - c. Hypothesis - Imani & Dillon
2. Analytic Approach/Statistical Model - Longyuan
3. Literature review
 - a. Pertaining to environmental pollution - Longyuan
 - b. Pertaining to government representation in Period 2 & federal court system - Imani
 - c. Pertaining to government representation in modern times & state offices - Dillon
 - d. Pertaining to media and marketing - Woomy
4. Items, Data Sources, Process
 - a. EP.LRA1, EP.SV1, EP.LRA2, EP.SV2, EP.SV3 - Longyuan
 - b. GR.LRA3 - Dillon
 - c. GR.SV2 & GR.LRA2 - Imani
 - d. MM.LRA1 - Woomy
5. Results and Exploratory Data Analysis - Longyuan, Imani, Dillon
6. Future Works - Longyuan, Imani, Dillon, Woomy
7. Conclusion - Woomy
8. Ethical considerations - Woomy

Appendix B: Data Sources

Environmental Pollution domain — EP.SV3 and EP.LRA2

	format	Available features	Scraping Approach	Data Quality
2020 Census National Demographics data	csv	Percentage of Black or African American alone or in combination with one or more other races in 2020	Downloaded from https://data.census.gov/table/DECENNIALDP2020.DP1	No missing data on key features
2020 Census County Level Demographics data	csv	-County GEO ID (FIPS) -County name -Percentage of people racialized as black -Total population	Downloaded from https://data.census.gov/table/DECENNIALDP2020.DP1?g=010XX00US\$0500000&d=DEC+Demographic+Profile	No missing data on key features
2020 County-Level Facility HAP Emission data	xlsx	-State and county name -County FIPS code -Annual HAP emission amount	Downloaded from https://awsedap.epa.gov	No missing data on key features
All National Priority List (NPL) Superfund Sites data	xlsx	-State and city name - Site epa id - Action start and end time	Downloaded from https://semspub.epa.gov/work/HQ/100000013.pdf	No missing data on key features
1990 5% (1-in-20) census data	csv	-Public Use Microdata Area(PUMA) code -Ethnicity	Extracted from https://usa.ipums.org/	No missing data on key features

Disaster Declaration Summaries	csv	-Location of the disaster -Time of the declared disaster -disaster type	Downloaded from https://www.fema.gov/openfema-data-page/disaster-declarations-summaries-v2	No missing data on key features
Home Owner's Loan Corporation (HOLC) Redlined Areas Dataset	shp	-Geometry of the redlined areas -Ratings of each area	https://github.com/americanpanorama/Census_HOLC_Research/blob/main/2020_Census_Tracts/Tracts_2020_HOLC.geojson	No missing data on key features
10-min Access Range of US Parks	shp	-Geometry of the ranges	https://www.tpl.org/park-data-downloads	No missing data on key features
Esri Data and Maps - 2020	shp	-Geometry of counties	https://hub.arcgis.com/datasets/esri::usa-counties/explore	No missing data on key features

Figure 1

Government Representation domain — GR.SV2 and GR.LRA2

	Format	Variables of interest	Scraping Approach	Data quality	Purpose
"List of African American federal judges"	Csv	Judge, District, Began Active Service, Ended Active Service	Scrape from https://en.wikipedia.org/wiki/List_of_African_American_federal_judges	No missing values in variables of interest	Finding the number of Black federal District Court judges from 1969-1999, and the counties their district served

the “Jurisdiction” section of all 94 District Courts’ websites	Text	Counties of Jurisdiction	Manually wrote	No missing values in the variables of interest	Identifying what counties are under the jurisdiction of each District Court
---	------	-----------------------------	-------------------	---	--

Figure 2

Government Representation domain — GR.LRA3

	Format	Variables of interest	Scraping Approach	Data quality	Purpose
“CCEST2020-ALLDATA6”	Csv	“BA_MALE”, “BA_FEMALE”, “HBA_MALE”, “HBA_FEMALE”, “TOT_POP”	Downloaded from the official Census website; CCEST2020-ALLDATA	No missing values in variables of interest	Finding the percentage of Black population by state between the years 2010 and 2020
“List of minority governors and lieutenant governors in the United States	Tabular	“Minority ethnicity”, “State”, “Term start”, “Term end”	Downloaded the table from Wikipedia; Minority Governors	No missing values in variables of interest	Finding the total number of Black state governors
“Black- American Members by State and Territory”	Tabular	“State or Territory”, “First Took Office”	Scrape from the website Black American Congress Members	No missing values in variables of interest	Finding the total number of Black house members (Senate & House)
“African American State	Text/Tabular	Term	Scrape from the website Black State	No missing values in variables of	Finding the total number of Black state

Supreme Justices since 1870”			Supreme Court Justices	interest	supreme court judges
“List of Minority Attorney Generals in the United States	Tabular	Ethnicity, State, Term Start, Term End	Scrape from the website Minority Attorney Generals	No missing values in variables of interest	Finding the total number of Black state attorney generals
“State Supreme Courts”	Tabular	Court, Method of selection, Number of justices	Scrape the data from State Supreme Courts	No missing values in variables of interest	Finding the total number of elected supreme court justices per state
“United States House of Representatives Seats by State”	Tabular	State, Representatives	Scrape the data from US House of Representatives Seats	No missing values in variables of interest	Finding the total number of house members by state

Figure 3

Media & Marketing domain — MM.LRA1

	Format	Variables of interest	Scraping Approach	Data quality	Purpose
Historical African American Newspapers Available Online: Chronology	Text/Tabular	Publication Place, Publication Dates Available Online	Downloaded from https://libguides.marist.edu/c.php?g=87271&p=562076	No missing values, however the years are structured oddly	Finding the number of Black newspapers in each county by year

Figure 4

Appendix C: Visualizations

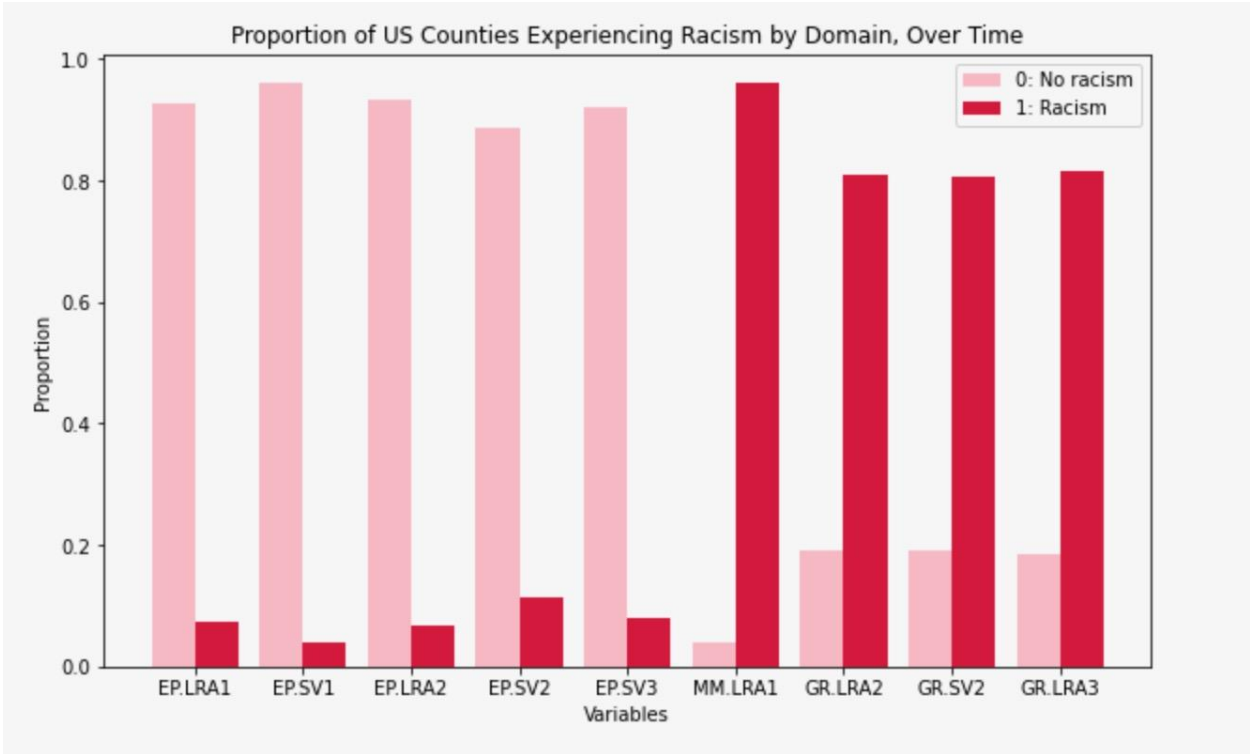


Figure 1

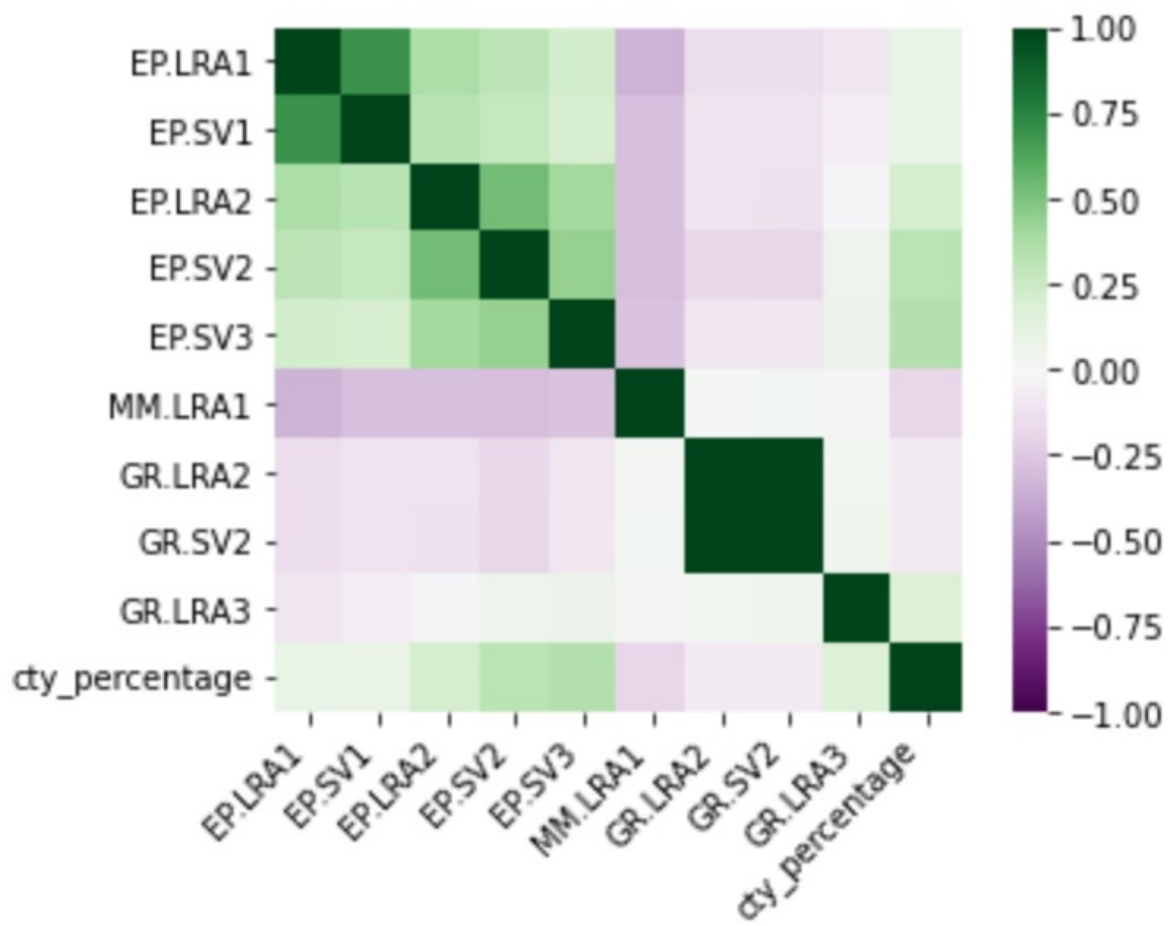


Figure 2

The Number of Counties Experiencing Racism Across Time Period

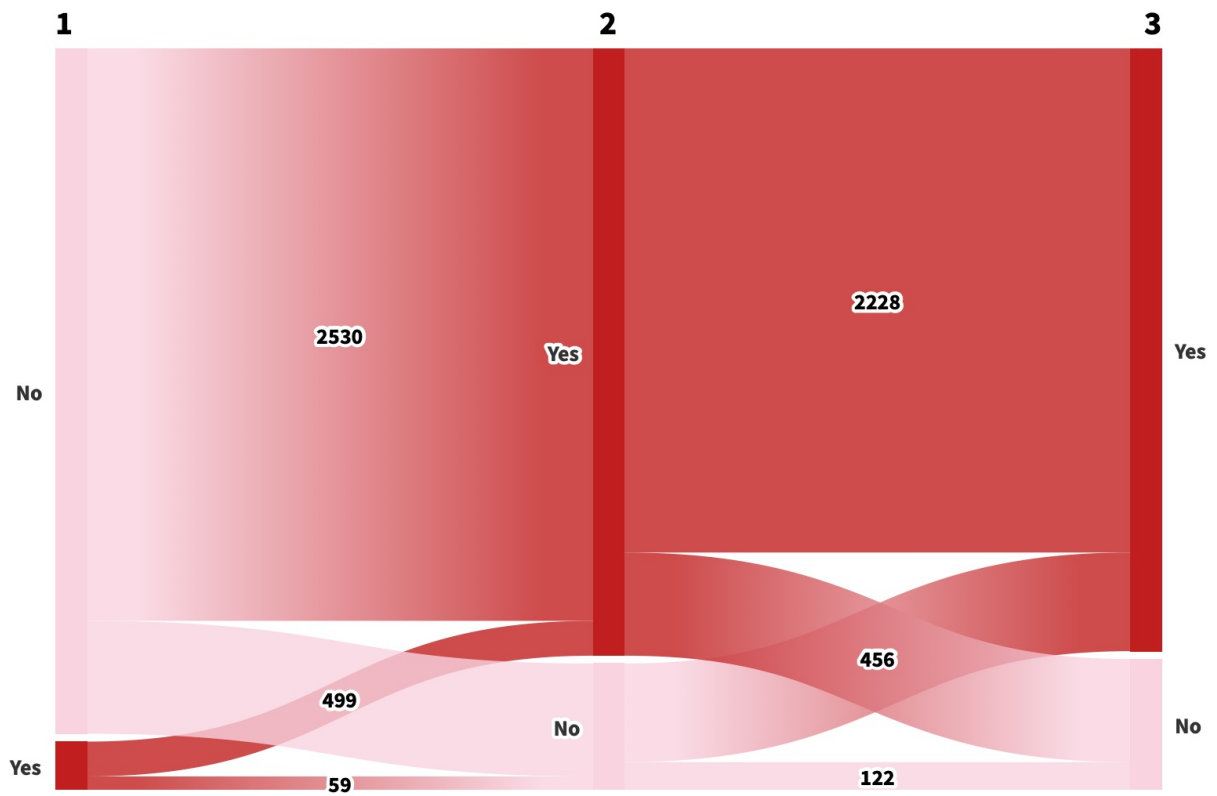


Figure 3

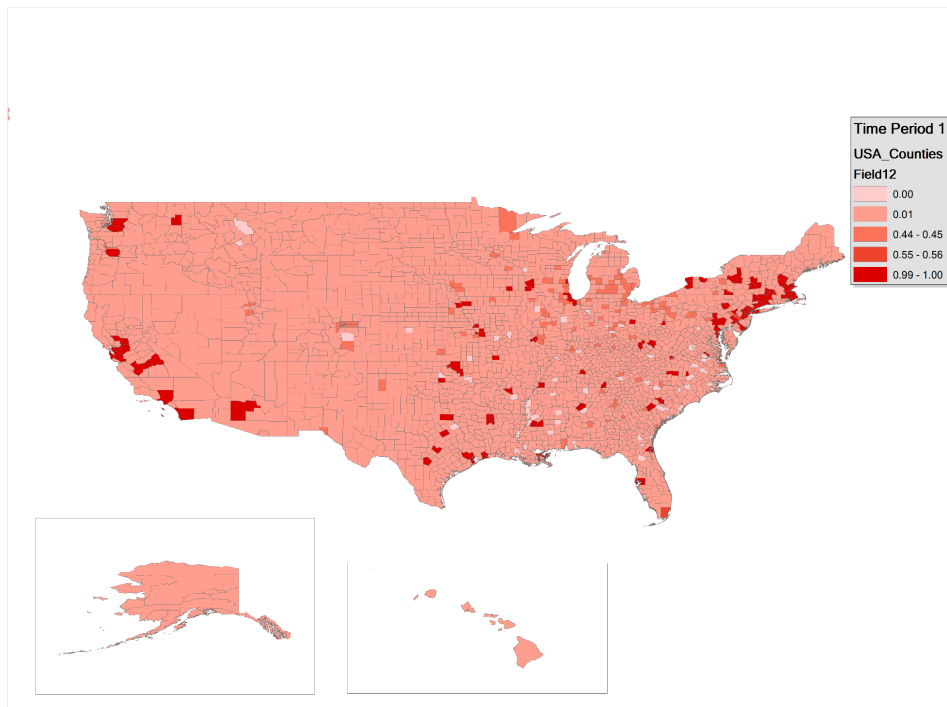


Figure 4

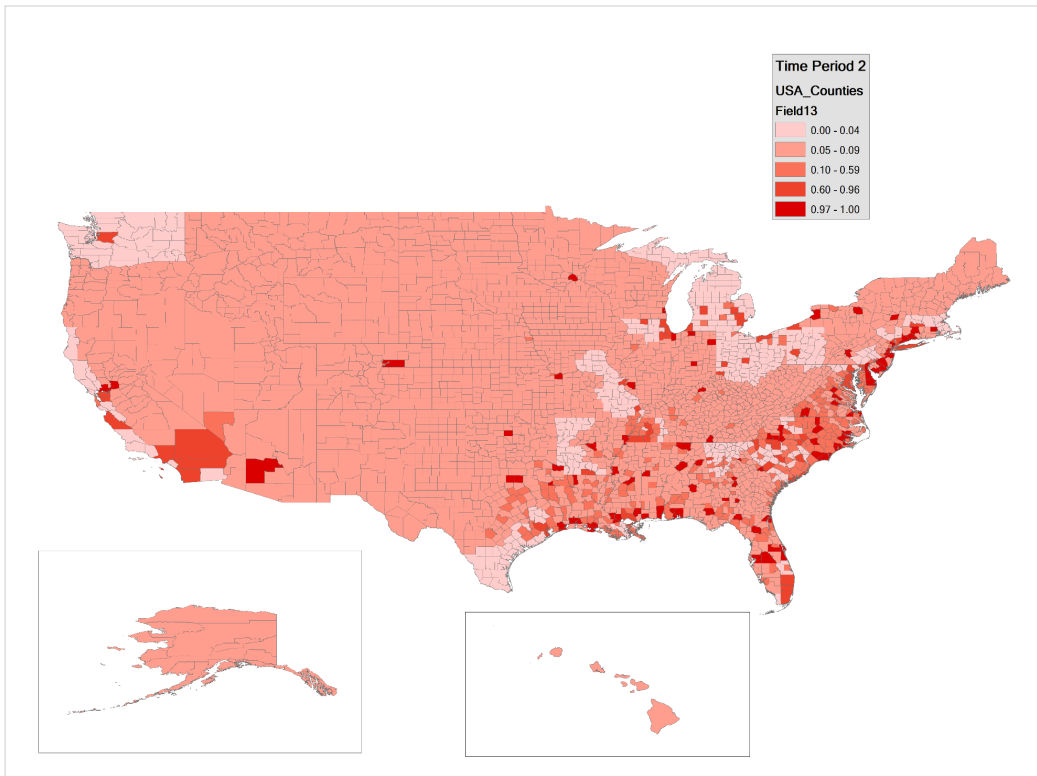


Figure 5

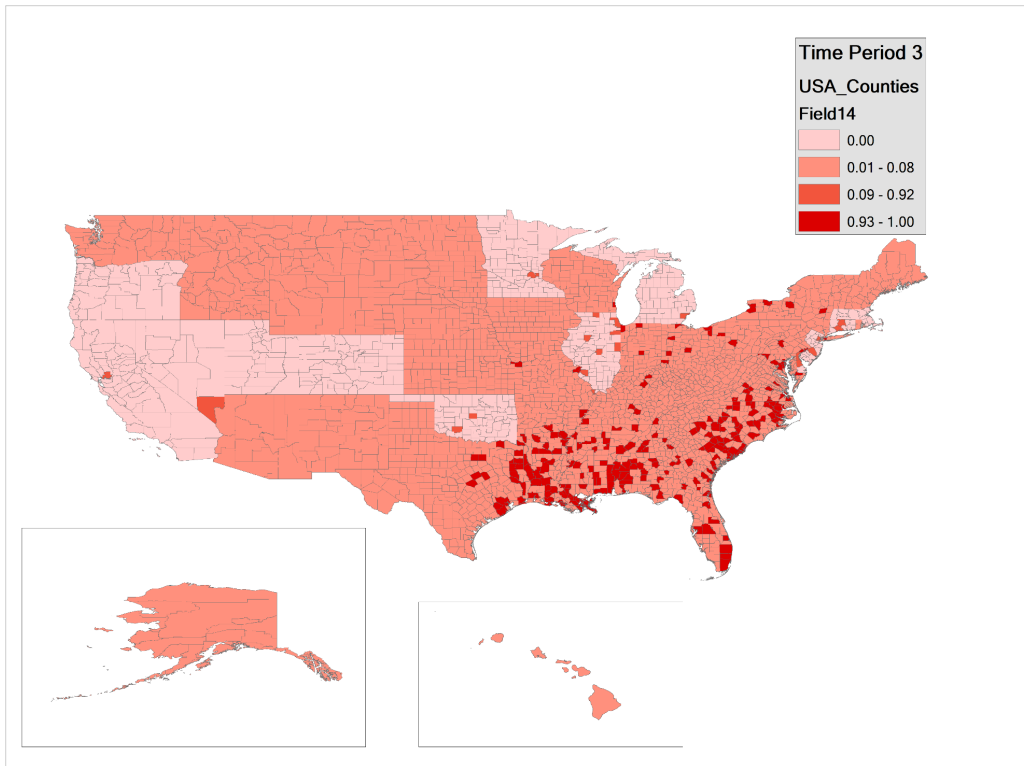


Figure 6

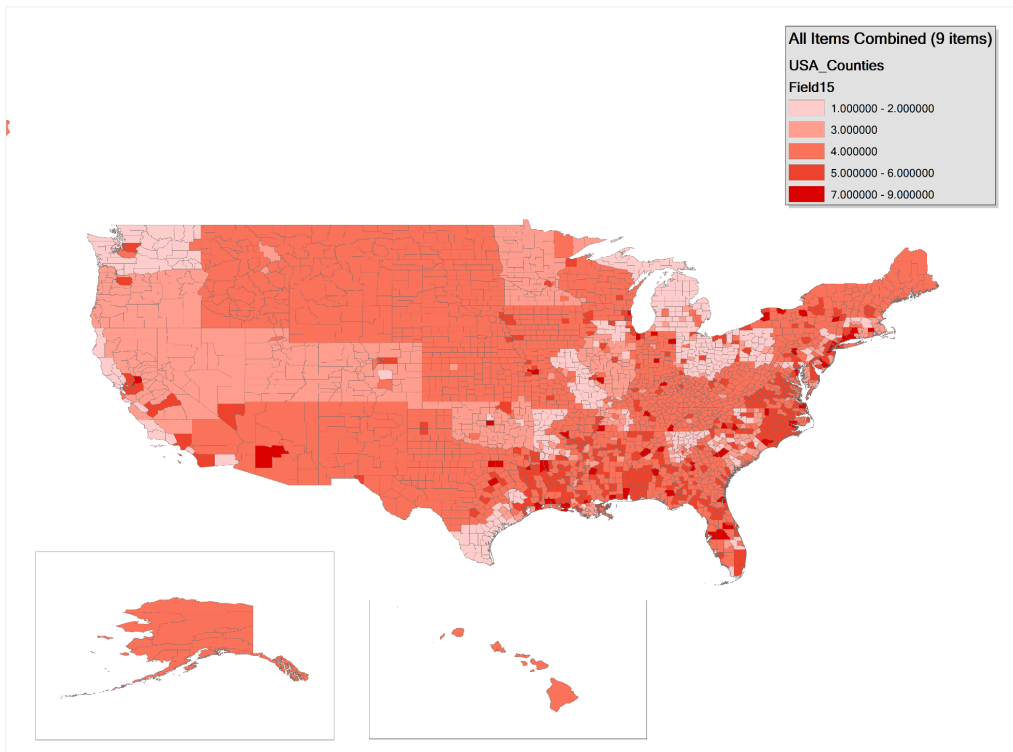


Figure 7

Appendix D: Sources Cited

“African American Newspapers.” Research Guides at Library of Congress, Library of Congress, guides.loc.gov/african-american-newspapers.

Black Politicians Are More Intrinsically ... - Wiley Online Library, onlinelibrary.wiley.com/doi/full/10.1111/ajps.12018. Accessed 28 Oct. 2023.

Brown, Anna. “Black Americans Have Made Gains in U.S. Political Leadership, but Gaps Remain.” Pew Research Center, Pew Research Center, 22 Jan. 2021, www.pewresearch.org/short-reads/2021/01/22/black-americans-have-made-gains-in-u-s-political-leadership-but-gaps-remain/.

Christopher W. Tessum et al. PM_{2.5} pollutants disproportionately and systemically affect people of color in the United States. *Sci. Adv.* 7, eabf4491 (2021). DOI: [10.1126/sciadv.abf4491](https://doi.org/10.1126/sciadv.abf4491)

“Court Role and Structure.” United States Courts, www.uscourts.gov/about-federal-courts/court-role-and-structure. Accessed 28 Oct. 2023.

“Experiences of Racism Associated with Poor Memory, Increased Cognitive Decline.” AAIC, 1 Aug. 2022, [aaic.alz.org/releases_2022/racism-cognitive-decline.asp](https://www.aaic.org/releases_2022/racism-cognitive-decline.asp).

Fine, Amy, and Milton Kotelchuck . Rethinking MCH: The Life Course Model as an Organizing Framework Concept Paper , Nov. 2010, unnaturalcauses.org/assets/uploads/file/Rethinking%20MCH%20-Fine.pdf. Accessed 13 Dec. 2023.

Gianattasio, Kan Z, et al. “Development of Algorithmic Dementia Ascertainment for Racial/Ethnic Disparities Research in the US Health and Retirement Study.” National Center for Biotechnology Information, U.S. National Library of Medicine, Jan. 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC6888863/.

Johnson, M. (2020, July 1). New report underscores racial prejudices in Superfund sites. The Hill. <https://thehill.com/policy/energy-environment/505526-new-report-underscores-racial-prejudices-in-superfund-sites/>

- Maloy, and Trust. "Functions of State and National Government." Functions of State and National Government, EdTech Books, 1 Jan. 1970, edtechbooks.org/democracy/functionsofstate.
- Noe-Payne, Mallory. "Black Representation in Politics Is Poor Nationwide, except in State Houses." NPR, NPR, 15 Aug. 2020, www.npr.org/2020/08/15/902811982/black-representation-in-politics-is-poor-nationwide-except-in-state-houses.
- Pohl, Daniel J, et al. "Relationship between Residential Segregation, Later-Life Cognition, and Incident Dementia across Race/Ethnicity." International Journal of Environmental Research and Public Health, U.S. National Library of Medicine, 26 Oct. 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC8583156/#B1-ijerph-18-11233.
- Sen, Maya. "Is Justice Really Blind? Race and Reversal in U.S. Courts." The Journal of Legal Studies, Volume 44, Number S1, January 2015, DOI: <https://doi.org/10.1086/682691>.
- Solomon, Danyelle, et al. "Systematic Inequality and American Democracy." Center for American Progress, 14 Feb. 2023, www.americanprogress.org/article/systematic-inequality-american-democracy/.
- "State and Local Government." The White House, The United States Government, 12 July 2022, www.whitehouse.gov/about-the-white-house/our-government/state-local-government/.
- "The State of Black Representation in the US Today." Public Wise, 13 Feb. 2023, publicwise.org/publication/the-state-of-black-representation-in-the-us-today/.
- "Understanding the Federal Courts." NAACP, 3 Dec. 2021, naacp.org/find-resources/know-your-rights/understanding-federal-courts#:~:text=Within%20the%20federal%20system%2C%20there,the%20court%20of%20final%20review.
- Welch, Susan, et al. "Do Black Judges Make a Difference?" American Journal of Political Science, vol. 32, no. 1, 1988, pp. 126–36. JSTOR, <https://doi.org/10.2307/2111313>. Accessed 29 Oct. 2023.
- When and Why Minority Legislators Matter - Annual Reviews, www.annualreviews.org/doi/10.1146/annurev-polisci-033011-205028. Accessed 28 Oct. 2023.