Critical Computational Geographies – Measures of Segregation

Technical Appendix

Nathan Alexander, PhD

Introduction

I outline the technical documentation for the Fall 2025 Quantitative Histories Workshop series on *Critical Computational Geographies*. In this particular section, we focus on questions about indicators used to quantify segregation. This section is part of a long-term project of the Quantitative Histories Workshop focused on exploring the dynamic features of context in probability and high-dimensional data.

Conceptual Model

High-dimensional data are characterized by the relationship between the data's dimensions (the number of features) and the data sample (number of observations). In an ideal interdisciplinary model that is informed by the various fields of human development, there is a potential to understand how the number of data features relate to the sample, and what methodological selections are used to characterize the set of indicators used in a mathematical model.

We will use U.S. census data to consider a measure of dissimilarity and spatial maps to: (1) observe land and coverings of different variables that quantify race, and (2) engage in assessing the various spatial conditions that inform racial isolation, or a series of dividing walls that separate one group from another group (Short, 2011).

Example: A Theory of Dividing Walls

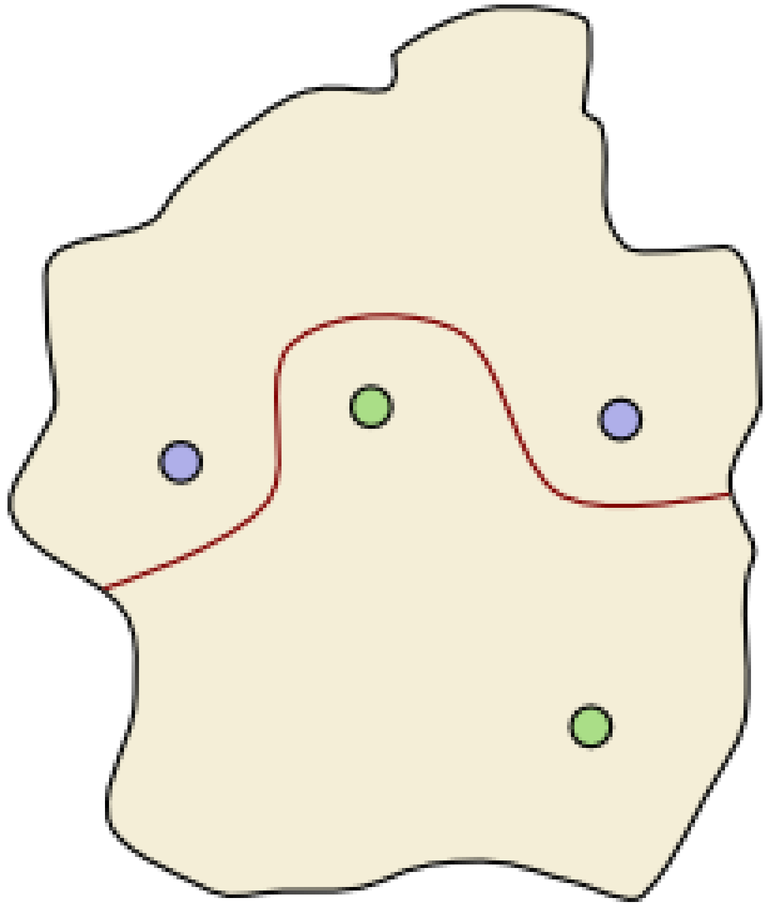


Figure 1: A sample dividing wall separating one group from another group

Short's (2011) *Dividing Wall's Theorem* presents a simplified topological equivalence to consider the conditions of segregation and isolation when a population of individuals are split into two groups. In the current instance, we will examine dissimilarity in a measure of Black and non-Black populations over some geographical area, G.

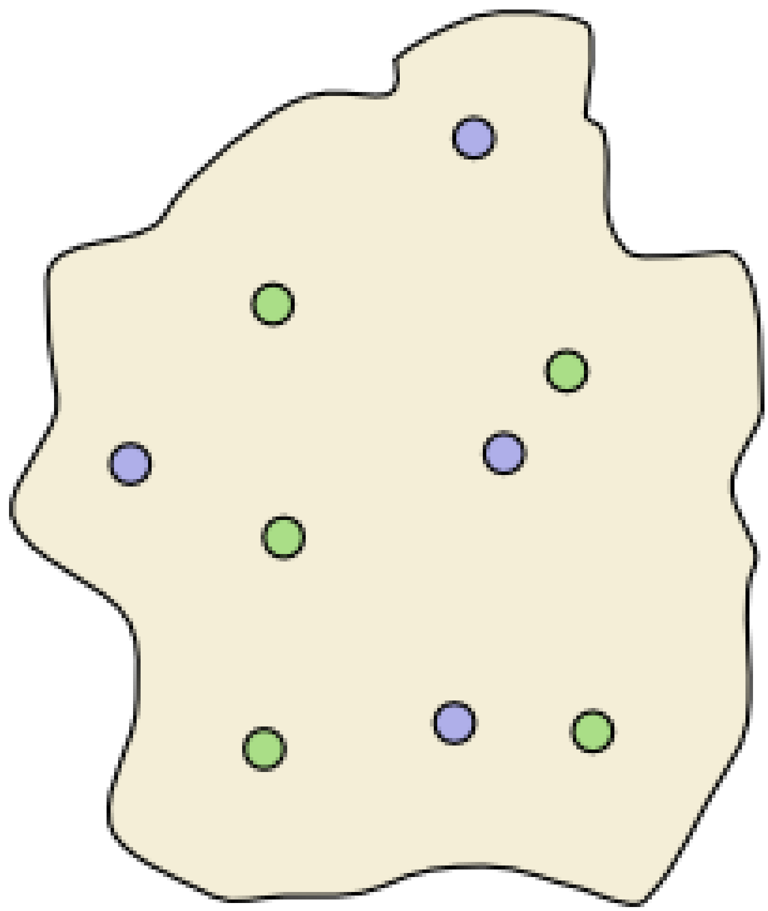


Figure 2: Is there a dividing wall?

Theorem 1. Given any configuration of blue and green towns, there is a dividing wall that separates blue towns from green towns.

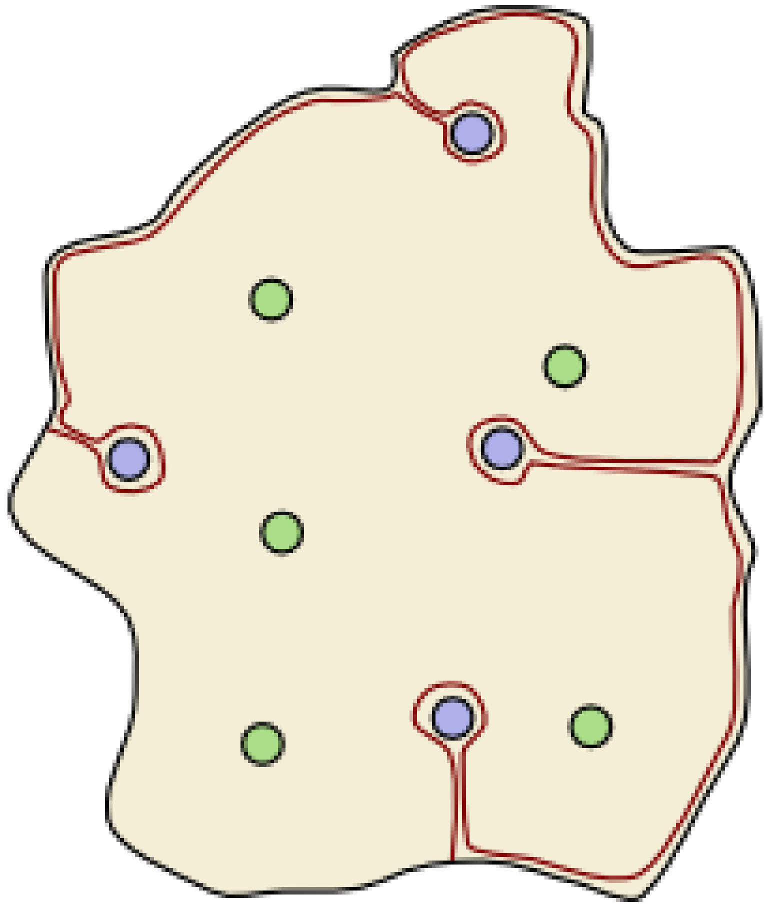


Figure 3: Visual proof of Short's (2001) Dividing Wall's Theorem

In the conceptual analysis above, there were only two tracts present and they completely segregated the two groups by the dividing wall. This conceptual model is a base example: two groups, two tracks, complete segregation.

Computational Model

The index of dissimilarity will be used to quantify the evenness with which Black residents are distributed relative to other non-Black racial groups across census tracts. Our current analysis builds on the perspectives that these base measures may be foundational in understanding and quantifying the dimensions of antiblackness in high-dimensional data sources.

The index of dissimilarity, D, is a demographic measure of the evenness with which two groups are distributed across geographic units within a larger geographical area. The measure quantifies the percentage of one group that would need to relocate to achieve an even distribution across all units in the geographical area. A value of $D = 0$ corresponds to complete integration, while $D = 1$ indicates complete segregation.

$$D = \frac{1}{2} \sum_{i=1}^{N} \left| \frac{a_i}{A} - \frac{b_i}{B} \right|$$

where

- N is the number of geographic units (e.g., census tracts),
- a_i is the population of group A (e.g., Black residents) in unit i,
- A is the total population of group A,
- b_i is the population of group B (e.g., non-Black residents) in unit i,
- B is the total population of group B.

D measures the unevenness or lack of even distribution between Black and non-Black residents across the geographical units of a region G. The index D takes on values from 0 (complete integration) to 1 (complete segregation) and represents the fraction of a group's population that would need to relocate to achieve an even spatial distribution. For example, if $D = 0.60$, 60% of one group would need to move to different areas to achieve integration.

Example: Hypothetical City Census Tracts

We can build on the hypothetical example provided by Dr. Rodney Green (n.d.) where he offers a tutorial of the dissimilarity index. In the example, Green (n.d.) models five tracts containing between 10 and 200 residents. I recreate Green's example below:

Consider the following hypothetical city with five census tracts.

Table 1: Hypothetical distribution of Black (B) and White (W) households across five census tracts with intermediate calculations toward the Index of Dissimilarity.

Tract	b_i	w_i	$\frac{b_i}{B=300}$	$\frac{w_i}{W=500}$	absolute difference
1	$b_1 = 50$	$w_1 = 10$	0.1667	0.0200	0.1467
2	$b_2 = 200$	$w_2 = 40$	0.6667	0.0800	0.5867
3	$b_3 = 10$	$w_3 = 100$	0.0333	0.2000	0.1667
4	$b_4 = 30$	$w_4 = 200$	0.1000	0.4000	0.3000
5	$b_5 = 10$	$w_5 = 150$	0.0333	0.3000	0.2667
					$\sum = 1.47$

where,

- $B = \sum b_i = 300$ is the total number of Black households,
- $W = \sum w_i = 500$ is the total number of White households.

In Green's example, the index of dissimilarity D is computed as follows:

$$D = \frac{1}{2} \sum_{i=1}^{N} \left| \frac{b_i}{B} - \frac{w_i}{W} \right|$$

with $N = 5$ we update our index:

$$D = \frac{1}{2} \sum_{i=1}^{5} \left| \frac{b_i}{B} - \frac{w_i}{W} \right|$$

we then update our population, where $B = \sum b_i = 300$ and $W = \sum w_i = 500$.

$$D = \frac{1}{2} \sum_{i=1}^{5} \left| \frac{b_i}{B = 300} - \frac{w_i}{W = 500} \right|$$

So we now have:

$$D = \frac{1}{2} \sum_{i=1}^{5} \left| \frac{b_i}{300} - \frac{w_i}{500} \right|$$

We then subsitute our values in the expansion, starting with total populations:

$$D = \frac{1}{2} \left(\left| \frac{b_1}{300} - \frac{w_1}{500} \right| + \left| \frac{b_2}{300} - \frac{w_2}{500} \right| + \left| \frac{b_3}{300} - \frac{w_3}{500} \right| + \left| \frac{b_4}{300} - \frac{w_4}{500} \right| + \left| \frac{b_5}{300} - \frac{w_5}{500} \right| \right)$$

and then values from each neighborhood in the numerators:

$$D = \frac{1}{2} \left(\left| \frac{50}{300} - \frac{10}{500} \right| + \left| \frac{200}{300} - \frac{40}{500} \right| + \left| \frac{10}{300} - \frac{100}{500} \right| + \left| \frac{30}{300} - \frac{200}{500} \right| + \left| \frac{10}{300} - \frac{150}{500} \right| \right)$$

or, more succinctly:

$$D = \frac{1}{2} \sum_{i=1}^{5} \left| \frac{b_i}{B} - \frac{w_i}{W} \right| = \frac{1}{2}(0.1467 + 0.5867 + 0.1667 + 0.3000 + 0.2667) = 0.7334$$

Green notes that 73.3 percent of either Black households would need to relocate to another tract to achieve an even distribution. In his discussion, Green first holds the White population constant in each tract and points to Title VII of the Civil Rights Acts when "White neighborhoods became available to Black households that previously had been constrained, by law and extra-legal practices, to live in densely populated inner cities" (Green, n.d.). He then presents the example when there is a swap to achieve racial parity across the census tracts. We modify the model and diverge from Green's example toward another end, based on our theoretical framework centered on segregation as one measure of antiblackness.

Research Question

How segregated are Black residents from other racial groups across census tracts, as quantified by the index of dissimilarity, D?

Data and Methods

For this analysis, we will use data from the U.S. Census Bureau, which includes information from the decennial census and the American Community Survey (ACS) 5-year estimates.

There are some essential items needed to generate our maps and begin our investigations. First, please make sure you have requested and stored your Census API key for easy access. Next, you will need to get set up in R and the R and the RStudio IDE (or Posit Cloud) and load the necessary packages and libraries.

Finally, you will need to select a geographical area that you would like to explore.

```r
# Install packages as needed
# install.packages(c("tidycensus", "tidyverse", "mapview", "mapgl", "quarto"))

# Load your Census
#CENSUS_API_KEY='your_api_key'

# Load necessary libraries
library(tidycensus)
library(dplyr)
library(ggplot2)
library(sf)
library(viridis)
library(scales)
```

Model Assumptions

We will suppose that a geographical area, G, consists of N tracts such that

$$G = \{g_1, g_2, g_3, ...g_N\} = \{tract_1, tract_2, tract_3, ..., tract_N\}$$

where,

- G is the set of census tracts that fully cover a geographical area,
- g_i is the i-th tract such that g_1 = tract 1, g_2 = tract 2, g_3 = tract 3, ...,
- N is the number of geographic units (e.g., census tracts)

We assume that G can be modeled by discrete data over a minimum population of n individuals, where there is at least one individual in each unit, i.e., $n \geq N$ (so that no unit in G is empty, i.e., all geographical units contain at least one individual).

We also modify the group meanings in the model to attend to the theoretical framework centered on measures of segregation that support our continued understanding of antiblackness. Specifically, we have:

$$\hat{D} = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{b_i}{b} - \frac{o_i}{O}\right|$$

where

- N is the number of geographic units (e.g., census tracts),
- b_i is the population of Black residents in unit i,
- B is the total population of Black residents,
- o_i is the population of non-Black (other) residents in unit i,
- O is the total population of non-Black (other) residents.

Findings

DC

Given the structure of DC, we use `geography = "tract"` on the variable `B02001_003` for Black alone.

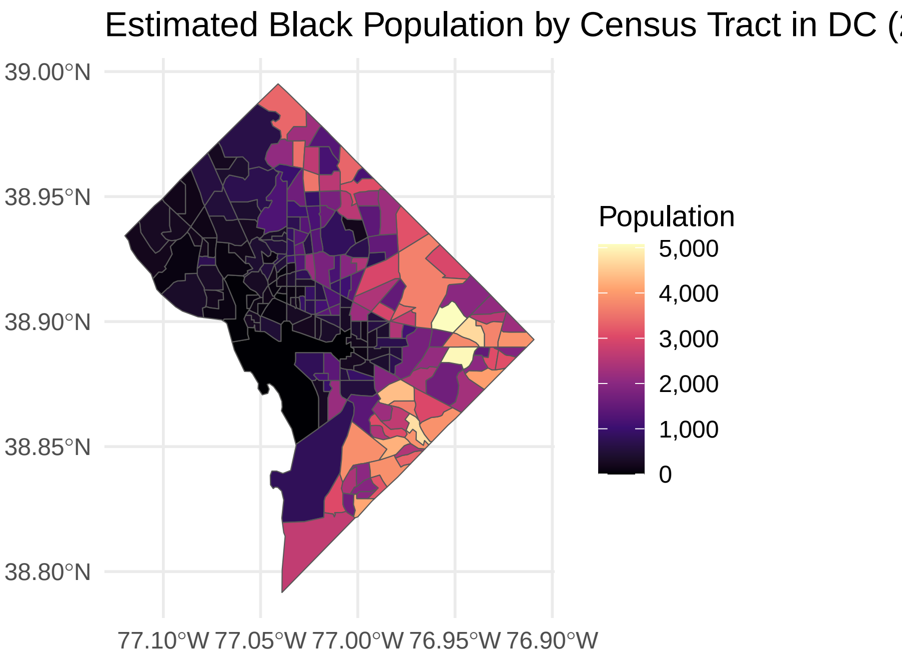
```r
dc_tracts <- get_acs(
  geography = "tract",
  variables = c(black = "B02001_003", # Black/African American population alone
                total = "B01001_001" # Total population
  ),
  state = "DC",
  year = 2023,
  geometry = T,
  output = "wide"
)
```

We can then look at the first few and last few rows of our estimates.

```r
dc_tracts <- dc_tracts %>%
  mutate(nonblackE = totalE - blackE) %>%
  select(GEOID, blackE, nonblackE, totalE)
```

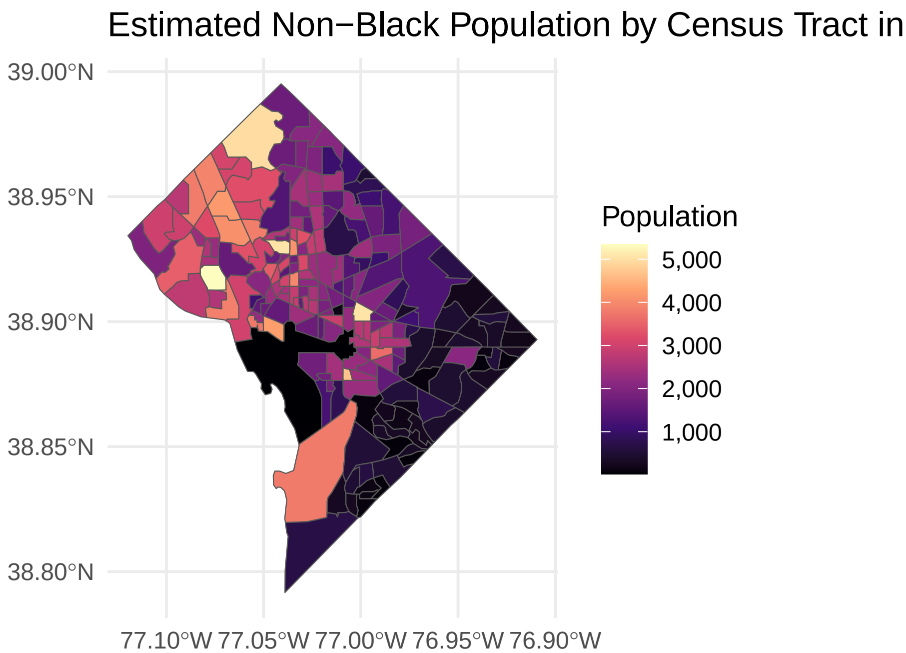
We'll then plot our data to get an initial visual of the variable.

```r
ggplot(dc_tracts) +
  geom_sf(aes(fill = blackE)) +
  scale_fill_viridis_c(option = "magma",
                       na.value = "grey50",
                       labels = comma) +
  labs(title = "Estimated Black Population by Census Tract in DC (2023)",
       fill = "Population") +
  theme_minimal()
```



And the data for non-Black individuals.

```
ggplot(dc_tracts) +
  geom_sf(aes(fill = nonblackE)) +
  scale_fill_viridis_c(option = "magma",
                       na.value = "grey50",
                       labels = comma) +
  labs(title = "Estimated Non-Black Population by Census Tract in DC (2023)",
       fill = "Population") +
  theme_minimal()
```



Now that we know our maps feature is working, we begin our investigation.

We first make a `dc_black_pop` data frame.

```
# Get Black population by tract in DC
dc_black_pop <- get_acs(
  geography = "tract",
  variables = "B02001_003",  # Black alone
  state = "DC",
  year = 2023,
  geometry = T
) %>%
  mutate(estimate_black = estimate)
```

We then grab the total population by tract in DC, we will call it `dc_total_pop`.

```
# Get total population by tract in DC
dc_total_pop <- get_acs(
  geography = "tract",
  variables = "B01003_001",  # total population
  state = "DC",
```

```r
  year = 2023,
  geometry = F # note that we have geometry turned off here
) %>%
  mutate(estimate_total = estimate)
```

We then combine the black population with the total population.

```r
dc_combined <- left_join(dc_black_pop, dc_total_pop, by = "GEOID") %>%
  mutate(estimate_nonblack = estimate_total - estimate_black) %>%
  st_as_sf()
```

We then look at our combined data with the other estimates.

```r
dc_combined %>%
  relocate(GEOID, estimate_total) %>%
  arrange(desc(estimate_black)) %>%
  head()
```

```
Simple feature collection with 6 features and 12 fields
Geometry type: POLYGON
Dimension:     XY
Bounding box:  xmin: -76.99229 ymin: 38.84459 xmax: -76.93487 ymax: 38.90822
Geodetic CRS:  NAD83
        GEOID estimate_total
1 11001009602           5527
2 11001007703           7140
3 11001007502           4999
4 11001007803           4968
5 11001007601           4949
6 11001007404           4304
                                                           NAME.x variable.x
1 Census Tract 96.02; District of Columbia; District of Columbia B02001_003
2 Census Tract 77.03; District of Columbia; District of Columbia B02001_003
3 Census Tract 75.02; District of Columbia; District of Columbia B02001_003
4 Census Tract 78.03; District of Columbia; District of Columbia B02001_003
5 Census Tract 76.01; District of Columbia; District of Columbia B02001_003
6 Census Tract 74.04; District of Columbia; District of Columbia B02001_003
  estimate.x moe.x estimate_black
1       5072   629           5072
2       5016  1421           5016
3       4734   915           4734
```

```
4       4676   834           4676
5       4384   860           4384
6       4245   659           4245
                                                         NAME.y variable.y
1 Census Tract 96.02; District of Columbia; District of Columbia B01003_001
2 Census Tract 77.03; District of Columbia; District of Columbia B01003_001
3 Census Tract 75.02; District of Columbia; District of Columbia B01003_001
4 Census Tract 78.03; District of Columbia; District of Columbia B01003_001
5 Census Tract 76.01; District of Columbia; District of Columbia B01003_001
6 Census Tract 74.04; District of Columbia; District of Columbia B01003_001
  estimate.y moe.y estimate_nonblack                       geometry
1       5527   590               455 POLYGON ((-76.96222 38.8995...
2       7140  1051              2124 POLYGON ((-76.9575 38.88363...
3       4999   922               265 POLYGON ((-76.97574 38.8608...
4       4968   857               292 POLYGON ((-76.95101 38.8955...
5       4949   901               565 POLYGON ((-76.9901 38.87135...
6       4304   648                59 POLYGON ((-76.99199 38.8537...
```

We then calculate the proportion of Black people in all DC tracts.

```r
total_black <- sum(dc_combined$estimate_black, na.rm = TRUE)
total_nonblack <- sum(dc_combined$estimate_nonblack, na.rm = TRUE)

dc_combined <- dc_combined %>%
  mutate(proportion_black = estimate_black / estimate_total)
```

Now we can view the top 20 tracts with the highest proportion of Black individuals.

```r
dc_combined %>%
  mutate(proportion_non_black = 1 - proportion_black) %>%
  arrange(desc(proportion_black)) %>%
  select(GEOID, proportion_black, proportion_non_black) %>%
  head(n = 20)
```

```
Simple feature collection with 20 features and 3 fields
Geometry type: POLYGON
Dimension:     XY
Bounding box:  xmin: -77.01481 ymin: 38.82144 xmax: -76.9094 ymax: 38.90822
Geodetic CRS:  NAD83
First 10 features:
         GEOID proportion_black proportion_non_black
```

```
1  11001007404        0.9862918           0.01370818
2  11001007409        0.9834662           0.01653381
3  11001009700        0.9808168           0.01918317
4  11001009811        0.9789349           0.02106509
5  11001009905        0.9723444           0.02765556
6  11001009907        0.9666508           0.03334921
7  11001007709        0.9645701           0.03542994
8  11001007605        0.9601898           0.03981018
9  11001007808        0.9532278           0.04677223
10 11001007502        0.9469894           0.05301060
                         geometry
1  POLYGON ((-76.99199 38.8537...
2  POLYGON ((-76.98066 38.8454...
3  POLYGON ((-76.99275 38.8309...
4  POLYGON ((-77.00203 38.8309...
5  POLYGON ((-76.92932 38.8807...
6  POLYGON ((-76.94577 38.8808...
7  POLYGON ((-76.9733 38.87839...
8  POLYGON ((-76.98436 38.8666...
9  POLYGON ((-76.92796 38.8918...
10 POLYGON ((-76.97574 38.8608...
```

We can also calculate the proportion of tracts that are above a certain threshold of Black only individuals. Here we set the threshold at tracts with 75 percent or more Black residents.

```
dc_combined %>%
  # Count how many tracts have proportion_black >= 0.75
  summarise(
    total_tracts = n(),
    tracts_above_threshold = sum(proportion_black >= 0.75),
    proportion_above_threshold = mean(proportion_black >= 0.75)
  )
```

```
Simple feature collection with 1 feature and 3 fields
Geometry type: POLYGON
Dimension:     XY
Bounding box:  xmin: -77.11976 ymin: 38.79165 xmax: -76.9094 ymax: 38.99511
Geodetic CRS:  NAD83
  total_tracts tracts_above_threshold proportion_above_threshold
1          206                     51                  0.2475728
                        geometry
1 POLYGON ((-77.05166 38.9870...
```

We see that one-fourth of all DC tracts have a population of more than 75 percent Black.

Finally, we calculate D.

```
dissimilarity_dc = 0.5 * sum(abs(
  (dc_combined$estimate_black / total_black) -
  (dc_combined$estimate_nonblack / total_nonblack)
), na.rm = TRUE)
```

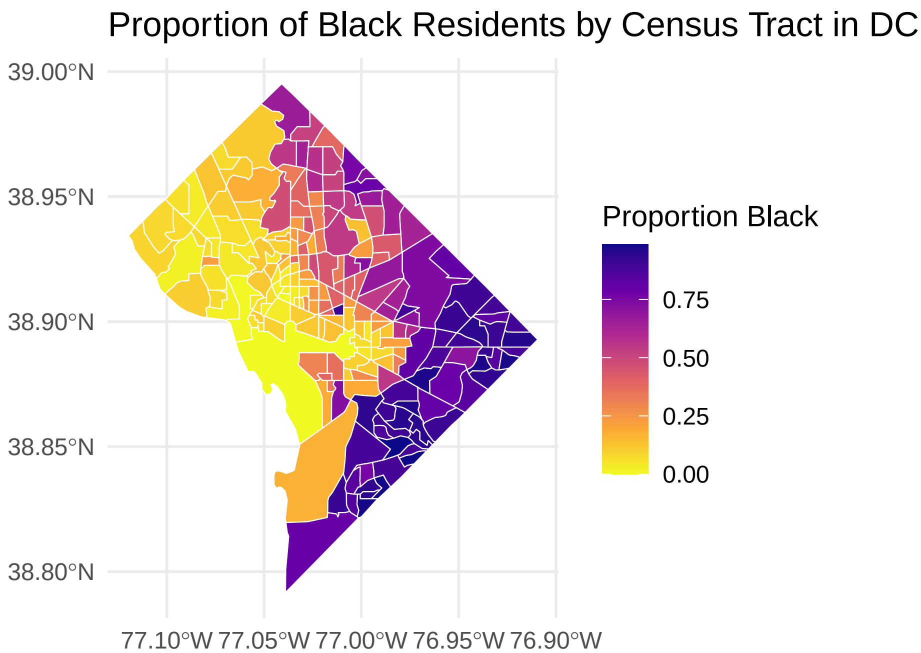
We print our result.

```
dissimilarity_dc
```

```
[1] 0.5916675
```

Values between roughly 0.3 to 0.6 indicate moderate segregation; above 0.6 is high segregation. In this instance, the model result suggests significant residential segregation by race in DC. We close by visualizing our result and the potential dividing wall.

```
# we need to manually convert our combined data frame to include simnple features
dc_combined <- st_as_sf(dc_combined)

ggplot(dc_combined) +
  geom_sf(aes(fill = proportion_black), color = "white") +
  scale_fill_viridis_c(option = "plasma", direction = -1) +
  labs(title = "Proportion of Black Residents by Census Tract in DC",
       fill = "Proportion Black") +
  theme_minimal()
```



Is there a dividing wall?

Given that our model result indicate a significant measure of segregation, we proceed with identifying the dividing wall. The maps we have viewed up to this point give us a clear view of where that wall may be.

```
dc_combined <- dc_combined %>%
  mutate(majority_black = proportion_black > 0.5)
```

We'll join geometries by group.

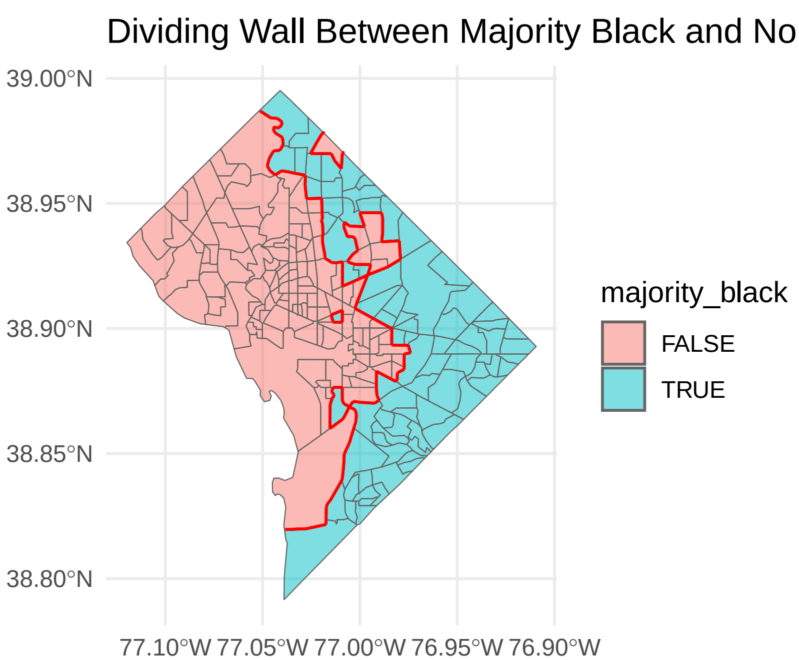
```
# Union geometries by group
union_black <- dc_combined %>%
  filter(majority_black) %>%
  summarise(geometry = st_union(geometry))

union_nonblack <- dc_combined %>%
  filter(!majority_black) %>%
  summarise(geometry = st_union(geometry))
```

We then calculate boundary (difference) between groups (shared border).

```r
boundary_line =
  st_intersection(st_boundary(union_black), st_boundary(union_nonblack))
```

We then plot the base polygons for the dividing wall.

```r
ggplot() +
  geom_sf(data = dc_combined, aes(fill = majority_black),
          color = "grey40",
          alpha = 0.5) +
  geom_sf(data = boundary_line,
          color = "red",
          size = 1) +
  labs(title = "Dividing Wall Between Majority Black and Non-Black Areas") +
  theme_minimal()
```



NC

DC is a speical case since it is a city-state.

For NC, we'll use `geography = "county"` and then search for those counties with higher values of segregation, and then we'll use those ordered values to identify significant segregation and identify potential dividing walls.

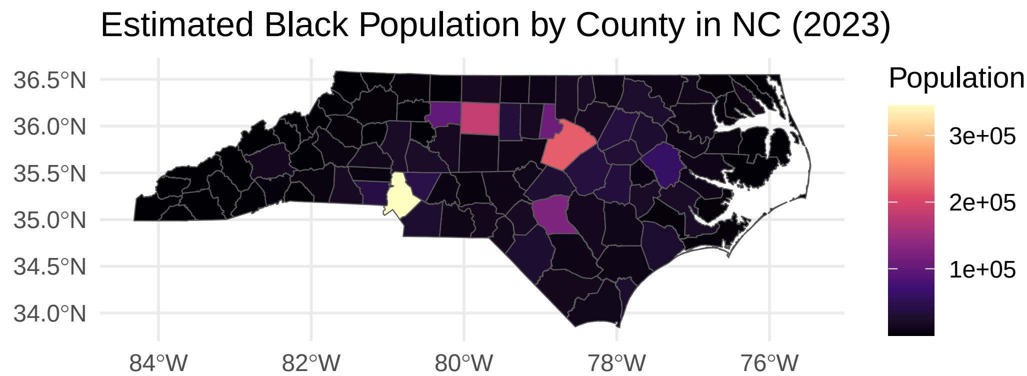
```r
nc_black_pop <- get_acs(
  geography = "county",
  variables = "B02001_003", # Black/African American alone
  state = "NC",
  year = 2023,
  geometry = T
)
```

```r
head(nc_black_pop)
```

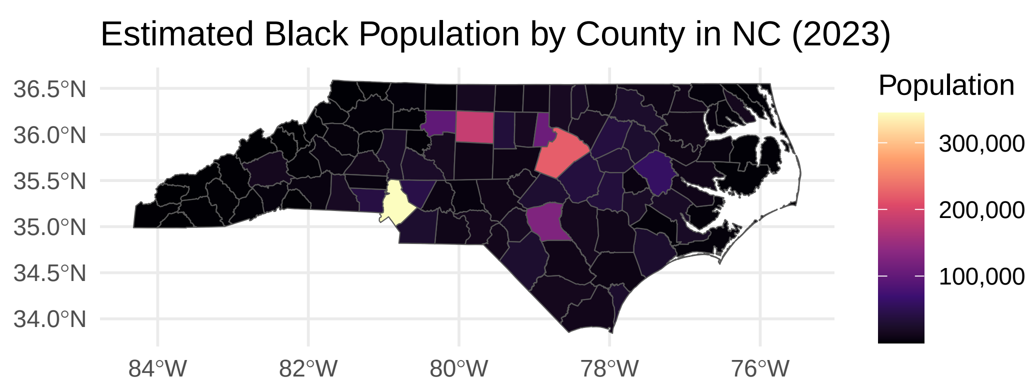
```
Simple feature collection with 6 features and 5 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -83.95288 ymin: 34.44087 xmax: -75.77333 ymax: 36.58812
Geodetic CRS:  NAD83
  GEOID                               NAME   variable estimate  moe
1 37133      Onslow County, North Carolina B02001_003    26157 1219
2 37009        Ashe County, North Carolina B02001_003      289  100
3 37169      Stokes County, North Carolina B02001_003     1543  373
4 37053   Currituck County, North Carolina B02001_003     1509  137
5 37173       Swain County, North Carolina B02001_003      212  112
6 37131 Northampton County, North Carolina B02001_003     9451  187
                        geometry
1 MULTIPOLYGON (((-77.17131 3...
2 MULTIPOLYGON (((-81.74065 3...
3 MULTIPOLYGON (((-80.4502 36...
4 MULTIPOLYGON (((-76.3133 36...
5 MULTIPOLYGON (((-83.94939 3...
6 MULTIPOLYGON (((-77.89977 3...
```

We'll then plot our data to make sure that our map features are working properly.

```r
ggplot(nc_black_pop) +
  geom_sf(aes(fill = estimate)) +
  scale_fill_viridis_c(option = "magma",
                       na.value = "grey50") +
  labs(title = "Estimated Black Population by County in NC (2023)",
       fill = "Population") +
  theme_minimal()
```




```r
ggplot(nc_black_pop) +
  geom_sf(aes(fill = estimate)) +
  scale_fill_viridis_c(option = "magma",
                       na.value = "grey50",
                       labels = comma) +
  labs(title = "Estimated Black Population by County in NC (2023)",
       fill = "Population") +
  theme_minimal()
```



References

Hudson, P. J., & McKittrick, K. (2014). The geographies of blackness and anti-blackness. *The CLR James Journal, 20*(1/2), 233–240.

King, T. L., Navarro, J., & Smith, A. (2020). *Otherwise worlds: Against settler colonialism and anti-blackness*. Duke University Press.

Lu, B., Harris, P., Charlton, M., & Brunsdon, C. (2014). The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spatial Information Science, 17*(2), 85–101.

Sørensen, A. B. (1978). Mathematical models in sociology. *Annual Review of Sociology, 4*, 345–371.

Walker, K. (2023). *Analyzing us census data: Methods, maps, and models in R*. Chapman.

Coda

While the index of dissimilarity offers a clear and interpretable measure of segregation, it is only one facet of complex social dynamics. Future work could extend this analysis by incorporating:

- Isolation and clustering indices to capture different aspects of segregation.
- Temporal dynamics to assess how patterns shift over time.
- Qualitative data integration to contextualize spatial patterns with lived experiences.
- Policy evaluation assessing impacts of urban planning and housing initiatives.

This technical file lays the foundation for quantitative historical geography research by demonstrating reproducible workflows with open Census data and modern R tools. The methodologies presented can be readily adapted to other metropolitan areas and demographic groups for comparative analyses.

Continued interdisciplinary collaboration will deepen our understanding of spatial inequality and support informed efforts to foster equitable communities.

Document No: 20250922-na