Mapping and analyzing racial geography in the United States using National Racial Geography Dataset

Anna Dmowska[a,1], Tomasz F. Stepinski[b]

[a]Institute of Geoecology and Geoinformation, Adam Mickiewicz University, Poznan, Poland
[b]Space Informatics Lab, Department of Geography and GIS, University of Cincinnati, Cincinnati, OH 45221-0131, USA

Abstract

This paper aims to demonstrate how the newly developed National Racial Geography Dataset (NRDG2020) and the R package *raceland* can be utilized to address racial geography problems similar to those discussed in the demographic literature. An example of such investigations is mapping racial distribution, assessing how segregation varies at different scales, comparing the segregation and diversity between different parts of the MSA, and calculating bi-racial segregation measures. The NRGD2020 dataset provides precalculated US-wide GIS layers that allow for the visualization and quantitative analysis of racial distribution, segregation, and diversity for any place in the conterminous United States. The NRGD2020 was created based on 2020 US Decennial census block data and utilizes the Racial Landscape (RL) methodology implemented in R package *raceland*. The NRGD2020 dataset with the detailed, reproducible tutorials illustrating how to perform calculations using R package *raceland* are available at http://socscape.edu.pl

1. Introduction

Analyzing the racial composition and spatial distribution of race-specific sub-populations has a long tradition among demographers. In demographic studies, the spatio-racial distribution is commonly summarized using the concepts of racial residential segregation or racial diversity. Racial diversity describes the level of racial inhomogeneity of the population (White, 1986), whereas racial residential segregation is a spatial concept that identifies how people of different ethnoracial groups are spatially distributed within the analyzed area (Massey and Denton, 1988). The two concepts together describe racial geography (Huiping et al., 2005). The levels of racial segregation and diversity are usually summarized by single-number indices (Massey and Denton, 1988; Reardon and Firebaugh, 2002). The indices are typically calculated for the largest metropolitan areas and are mainly used to rank cities based on their segregation level. Therefore, they do not provide broader information about the spatio-racial pattern of racial segregation and diversity in the United States.

Fewer studies focus on incorporating maps into demographic research and connecting numerical assessments of segregation with visualizations of where racial groups reside within an analyzed area. Logan (2015) points out that maps illustrating the spatial pattern of different ethnoracial groups are the most powerful tool for spatial analysis of racial segregation and diversity. Recently, researchers have turned more towards spatial analysis, but mapping racial distribution is still not widely used. Maps can be difficult to integrate into demographic studies due to limited access to free, ready-to-use resources. The Census Bureau provides data for aggregated areas (i.e., census tracts, census blocks) in census tables. The Census table stores the census subdivision's (an aggregated area) ID and the corresponding racial composition. Data of this type are better suited for calculating single-number indices. Making maps based on census data requires obtaining spatially-referenced data (census boundaries) and access to GIS software. Free, ready-to-use map resources are still limited to a few projects. Mixed Metro (https://mixedmetro.com/) provides the racial diversity maps that are the result of classifying census tracts into diversity-dominant race types (Holloway et al., 2012). Mixed Metro maps are available to download for 1990-2020 (Chipman et al., 2009-2023). Racial diversity maps show the racial character of the neighborhood in a simple and understandable manner. However, such visualization cannot be directly

*Corresponding author. *Email address:* dmowska@amu.edu.pl

connected with the numerical assessment of segregation given by the indices (Dmowska and Stepinski, 2023). There are also a few resources for racial dot maps (https://www.nytimes.com/interactive/2015/07/08/us/census-race-map.html , Race and Ethnicity in the US by Dot Density (2020 Census) provided by ESRI). A racial dot map conveys the spatial distribution of a population's density and the full racial composition on a single map (Dmowska and Stepinski, 2019; Roth, 2010). Other resources include, for example, Social Explorer (https://www.socialexplorer.com/) or Census Demographic Data Map Viewer that provides census-based maps for 2020. In 2013, we launched the SocScape project (Dmowska et al., 2017; Dmowska and Stepinski, 2017a). This project provides free, ready-to-use maps of racial distribution and racial diversity in the conterminous United States and US metropolitan areas in 1990, 2000, 2010, and 2020. The resources include:

- High-resolution population, race-specific, and racial diversity grids for 1990, 2000, 2010, 2020 (available as GeoTiffs) for the conterminous US, counties, and metropolitan statistical areas (Dmowska et al., 2017; Dmowska and Stepinski, 2017a).

- Racial dot maps calculated based on 2010 high resolution demographic grids for each county in the conterminous US (available as ESRI Shapefiles) (Dmowska and Stepinski, 2019, 2017b).

- Historical census tracts showing racial composition and diversity/dominant race classification in US cities between 1910-2010 (available as ESRI Shapefiles).

Recently, we introduced the National Racial Geography Dataset for 2020 (NRGD2020). The NRGD2020 is a collection of precalculated GIS layers that allow for the visualization and quantitative analysis of racial distribution for any place in the conterminous United States at a resolution of 30-meters. In addition to 30-m resolution layers, it also includes 10 racial diversity and 10 segregation grids depicting the level of diversity or segregation at the scale of 72km, 36km, 24km, 18km, 12km, 9km, 6km, 3km, 1.5km, 0.75km. All layers are available as GeoTiffs at http://socscape.edu.pl, and can be used in GIS software to prepare detailed maps of racial geography in the analyzed area. The National Racial Geography Dataset (NRGD2020) was created based on 2020 US Decennial census block data and utilizes the Racial Landscape (RL) methodology (Dmowska et al., 2020).

The Racial Landscape method (RL) introduces a consistent framework for visualization and quantification of the spatial distribution of racial sub-populations in arbitrary, user-defined regions using high-resolution race-specific grids instead of census subdivisions. The race-specific grid means that the region is divided into small (smaller than the smallest census aggregation area) racially homogeneous cells. A cell has two attributes: its race category and its population density. Such a grid is called a racial landscape (RL grid). RL grid visualizes racial distribution but at the same time also provides geospatial data to calculate two metrics to assess the level of racial segregation and diversity in the analyzed area. In the RL method, indices evaluating the level of segregation and diversity are calculated directly from the grid. RL method provides two types of results: (1) a racial map (similar to the dot map) that shows the racial distribution within an analyzed area; (2) two indices that quantify the level of racial segregation and diversity. By using the same data in the RL method, the numerical assessment of segregation given by the indices is connected with the visualization of how racial groups are spatially distributed within an analyzed area. Additionally, the analyzed area can be divided into smaller tiles, and the indices for each tile can be calculated and mapped to show how racial segregation and diversity change within the area on the scale of a tile. By changing the size of a tile, we can map the spatial variation of segregation and diversity on a different scale.

This paper presents examples of how NRGD2020 can be employed to address racial geography problems similar to those examined in the demographic literature. An example of such investigations is mapping racial distribution, assessing how segregation varies at different scales, comparing the segregation and diversity between different parts of the MSA, and calculating bi-racial segregation measures.

The paper is organized as follows. Section 2 describes the NRGD2020 product and concisely explains the Racial Landscape method and its implementation in the R package *raceland*. Section 3 focuses on using NRGD2020 for mapping racial distribution. Section 4 presents three use cases showing how racial geography problems can be addressed using NRGD2020. Section 5 presents conclusions.

2. Data and Methods

2.1. National Racial Geography Dataset

The National Racial Geography Dataset (NRGD2020) provides a high-resolution racial

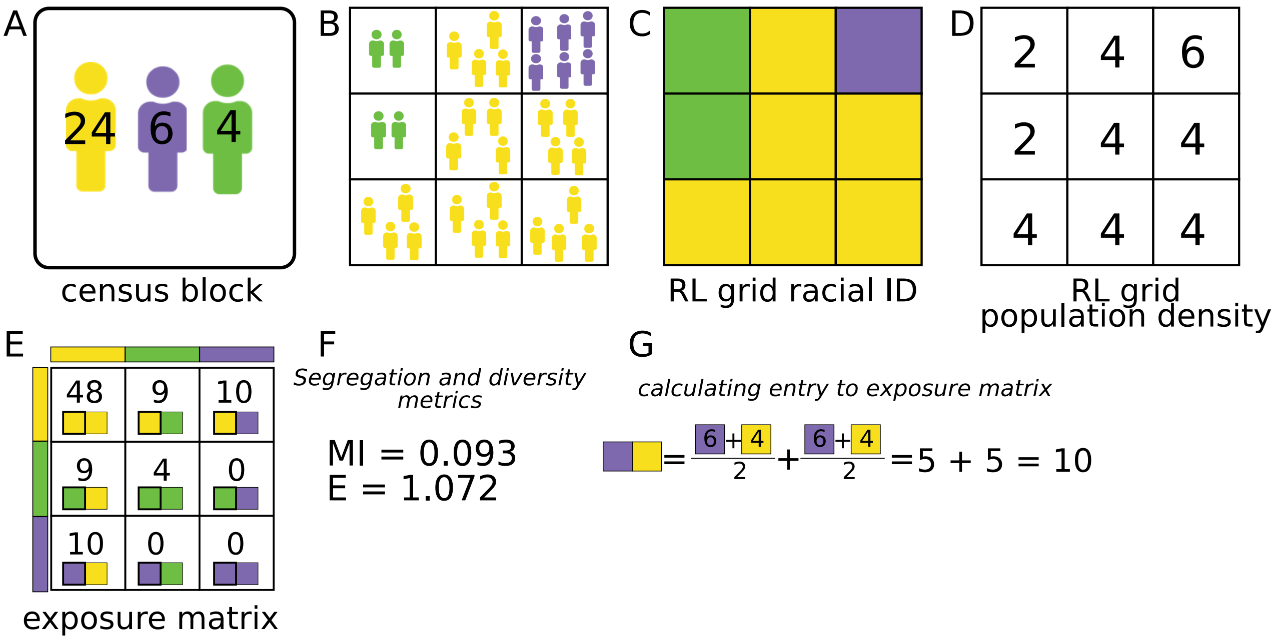

Figure 1: Construction of the racial landscape grid and calculation of the racial diversity and segregation indices using RL method: (A) census block, (B) people randomly redistributed to monoracial cells, (C) RL grid racial ID, (D) RL grid population density, (E) exposure matrix, (F) Segregation and diversity metrics calculated from the exposure matrix, (G) an example of calculating entry to the exposure matrix.

database for the conterminous US in 2020 that includes 4 types of GIS layers:

- *RL image* – a high-resolution RGB image that provides a US-wide visualization of racial geography at the resolution of 30m.
- *RL grid racial ID* and *RL grid population density* – a 30m resolution grid that provides an input to the *raceland* package in R computational environment (R Core Team, 2023) for calculating segregation and diversity metrics for an arbitrary area. *RL grid racial ID* is a categorical raster in which each cell has a label corresponding to one of six races (1- American Indians, 2 - Asians, 3 - Blacks, 4 - Hispanics/Latino, 5 - others (people who declared two or more races), 6 - Whites). *RL grid population density* is a raster in which each cell has a value of local population density.
- *Diversity grids* are available at ten different scales. These layers allow for visualizing and quantifying the racial diversity for any arbitrary areas. Each cell has assigned a Hill's number (see section 2.2 for explanation).
- *Segregation grids* are available at ten different scales. These layers allow for visualizing and quantifying the racial segregation for any arbitrary areas. Each cell has assigned the value of mutual information (see section 2.2 for explanation).

2.2. Racial Landscape method and its implementation in R package raceland

The technical description of the RL method can be found in the original RL paper (Dmowska et al., 2020). Here, we only provide a concise description of the RL methodology to highlight how the NRGD2020 dataset and the RL method can be used to address racial geography inquiries similar to those examined in demographic literature. Therefore, we focused on constructing the RL grid and calculating segregation and diversity indices directly from high-resolution grids.

Figure 1 shows the construction of the racial landscape grid (*RL grid racial ID*). Firstly, we start from the US census data aggregated at the block-level (Fig.1A). Such data provide information about the racial composition within a block, but we don't know the actual spatial distribution of people within a block. We use block-level data as they are the smallest available census subdivisions. Secondly, the RL method randomly redistributes people into monoracial cells (Fig.1B).

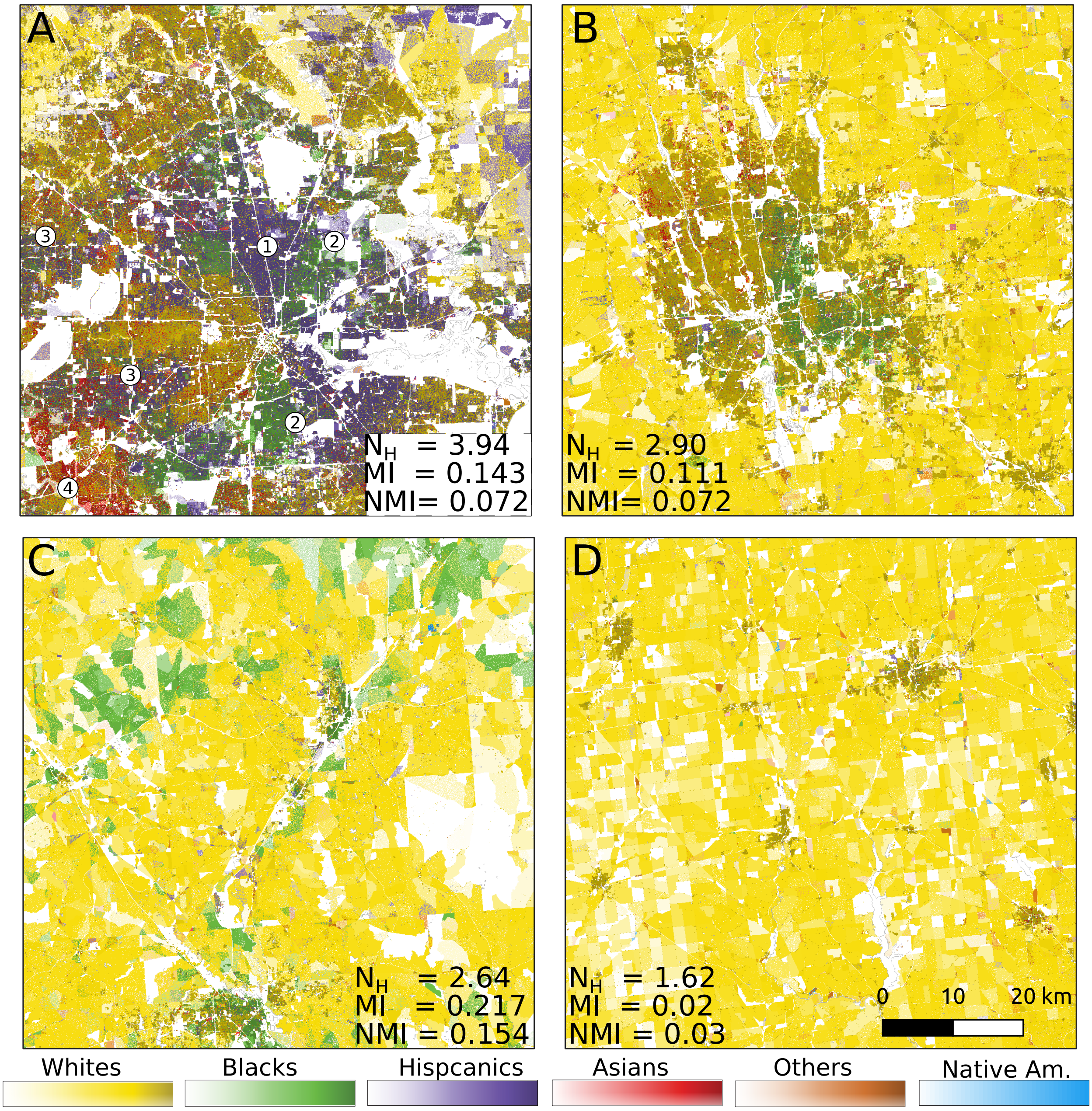

Figure 2: Mapping racial distribution using RL image layer in four areas of the size 72x72km. (A) Houston, TX. (B) Columbus, OH with the surroundings rural areas. (C) The rural area in Mississippi with two small cities. (D) The rural area northwest from Cincinnati, OH.

In the next step, we create two layers: *RL grid racial ID* and *RL grid population density*. A *RL grid racial ID* is a categorical raster generated by assigning to each cell racial ID (Fig 1C). In addition, we also create an *RL grid population density* in which each cell has assigned population densities based on the number of people of a given race within a block (Fig. 1D). A RL grid can visualize racial patterns while also providing geospatial information for calculating racial diversity and segregation metrics. In the RL method, the spatio-racial pattern is quantified using an exposure matrix. An exposure matrix is a modification of a co-occurrence matrix, widely used in landscape ecology to summarize land cover patterns. The co-occurrence matrix is a two-way table of the size K x K summarizing cells adjacencies (K is a number of categories). In the exposure matrix, each pair contributes the value of their average population density instead of 1. Fig. 1G shows an example of calculating the entry to the exposure matrix for the pair of yellow-purple cells. Further, an exposure matrix is summarized using two metrics derived from Information Theory (Shannon, 1948) – entropy and mutual information (Fig.1E). Entropy measures racial diversity. It can be translated into standardized entropy or Hill's number (Hill, 1973) for better interpretation. Hill's number is a straightforward transformation $N_H = a^E$, where a is the base of a logarithm used in calculation of entropy (RL method uses log_2), and E is a value of entropy. Hill's number depicts the significant number of racial groups present in an area. In the RL method, segregation is visually depicted as same-color cell clumps. The larger the clumps of same-color cells, the larger the segregation. Mutual information is a measure of clumping. In Information Theory, mutual information MI is interpreted as a reduction of uncertainty of the race of adjacent cells if the race of the focus cell is known. From a demographic point of view, the large MI value indicates higher segregation – the spatio-racial pattern creates larger clumps. The small values of MI indicate a mixed pattern (in the mixed pattern, knowing the race of the focus cell does not help to guess the race of the adjacent cells, therefore there is a small reduction of uncertainty).

The RL method was implemented in the R package *raceland* (Dmowska et al., 2020, 2021). This package provides a complete computational framework and allows one to perform calculations for the area of a single city. The package usage has been described in three vignettes that show how to calculate segregation metrics for any user-defined region and how to summarize local racial patterns of racial landscapes at different spatial scales. These examples are available in the package documentation at https://cran.r-project.org/web/packages/raceland/.

3. Mapping racial distribution using NRGD2020

One of the tasks in the demographic studies relates to mapping racial distribution within an analyzed area. NRGD2020 provides ready-to-use resources for mapping racial distribution, segregation, and diversity levels at different scales. The precalculated layers can be opened in GIS software (QGIS, ArcGIS). By cropping the US-wide layer to the extent of the analysis, a map can be easily generated for any part of the United States. Examples of such maps are presented in Figure 2 and Figure 3.

Figure 2 utilizes the *RL image* layer to illustrate the spatio-racial pattern in four areas: the area centered in Houston, TX (Fig. 2A); the city of Columbus, OH, with the surrounding areas (Fig. 2B); the rural areas with small towns in Mississippi (Fig. 2C), and the rural area located northwest of Cincinnati, OH (Fig. 2D). Each panel also shows three indices calculated directly from RL grids. Hill's number (N_H) corresponds to the diversity level. It indicates the number of racial groups with a sizable population present within the mapped area. The level of racial segregation is expressed by the mutual information (MI) or by the normalized mutual information ($NMI = MI/E$).

An RL image is best observed when zoomed in (i.e., 4X) since it reveals many details related to racial patterns. The spatio-racial pattern in Houston, TX (Fig. 2A) is composed of four racial groups with sizable population. In Houston, TX we can see the areas mostly populated by Hispanics/Latinos (shades of purple), Blacks (shades of green), Whites (shades of yellow), and Asians (shades of red). The distinct areas inhabited by each racial group indicate a higher level of segregation (MI = 0.143). Zooming in, we can also see the area marked as (1) is primarily Hispanic. Hispanics also have a significant share of the population in areas dominated by Blacks (marker (2)). The areas marked (3) are highly diverse that is depicted by mixed patterns. The Asian population is mixed with Whites (marker (4)). Figure 2B depicts Columbus, OH, and its surroundings. The yellow hues indicate the boundary between urban and rural areas surrounding Columbus. In Columbus, OH we can identify areas dominated by three ethnoracial groups located in different parts of the city. Blacks mostly inhabit the east part of Columbus, OH Whites mostly occupy the west part, and Asians mainly reside in the northwest part. Such a pattern corresponds to a similar level of segregation as seen in Houston in Fig. 2A. However, Columbus lacks highly diverse areas exhibiting mixed patterns. Panels C and D on Figure 2 show two examples from rural areas and small cities. Fig.2C shows the area in Mississippi with two small cities – Harrisburg (47 000 population) in the south and Laurel (18 000 population). Whites occupy the majority of the area with significant Black populations. After zooming in, we can also see the Asian population in Harrisburg (red cells on Fig.2C). The highest value of MI among the presented examples visually manifests itself by large clusters of green (Black population) and yellow cells (White population). Fig. 2D shows the area northwest of Cincinnati, OH. There

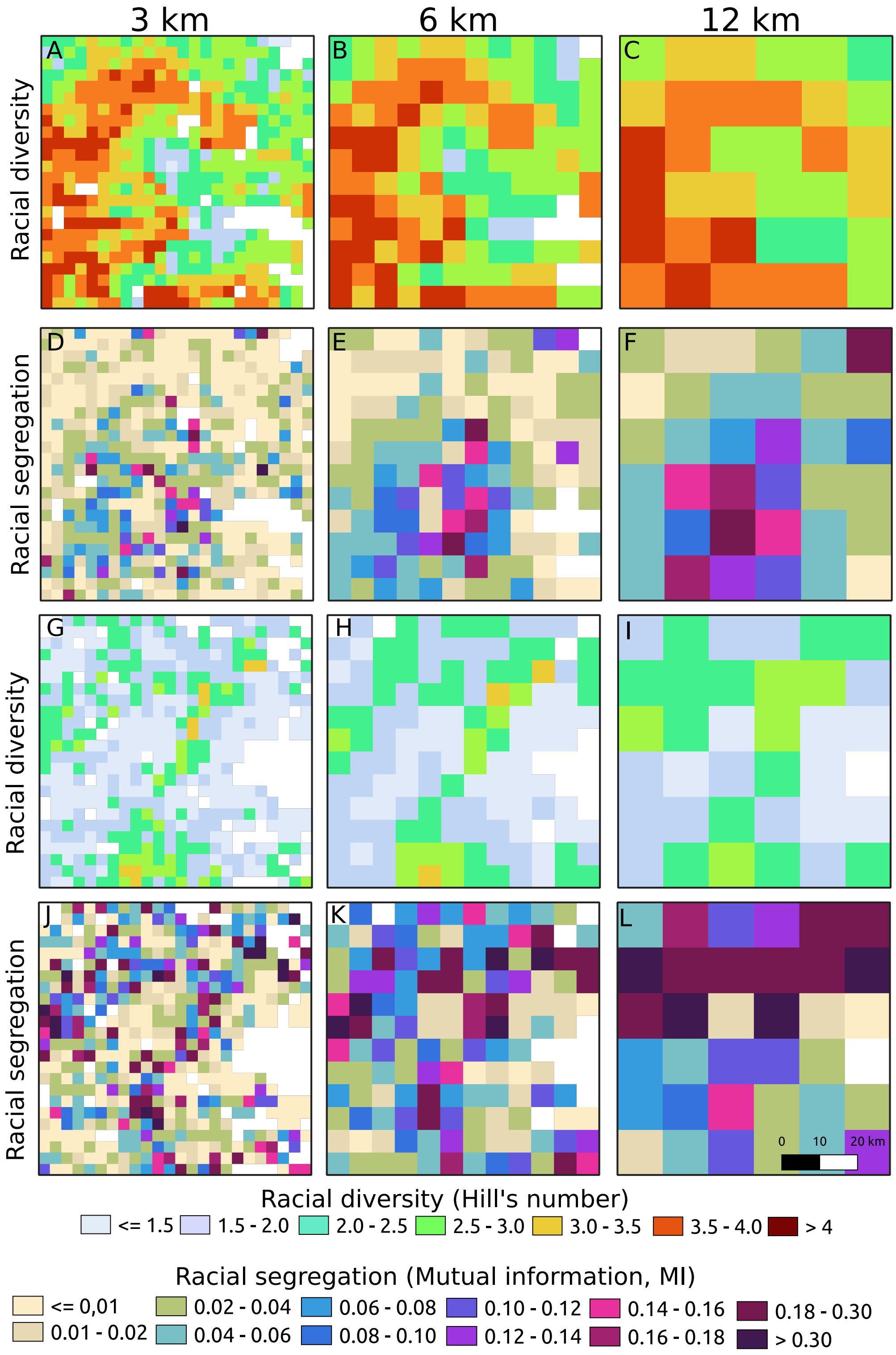

Figure 3: Examples of racial diversity and segregation maps at the scale of 3km, 6km, 12km. (A-C) Racial diversity maps for Houston, TX. (D-F) Racial segregation maps for Houston, TX. (G-I) Racial diversity maps for rural areas in Mississippi. (J-L) Racial segregation maps for rural areas in Mississippi. For the reference to the maps presented in the panels A-F see also RL image in the Fig. 2A. For the reference to the maps presented in the panels G-L see also RL image in the Fig. 2B.

is a dominance of Whites in the area, as evidenced by the majority of different shades of yellow. This is visually manifested by the small values of MI and N_H. The NRGD2020 dataset provides *RL image* layers that cover the conterminous US, allowing us to explore the spatio-racial pattern everywhere in the US, including small cities.

Figure 3 shows racial diversity and segregation maps at the scales of 3km, 6km, and 12km for the two areas presented in Figure 2: Houston, TX, and the rural area in Mississippi. In the RL method, the scale is defined by the size of the tile on which the area is divided. A scale of 3km, for instance, means that the area is divided into tiles of 3x3km, and for each tile, the metrics (MI, N_H) are calculated and mapped.

Figure 3 is composed of 12 panels arranged in four rows. The first and second rows show levels of diversity and segregation in Houston, TX. The third and fourth rows show the diversity and segregation levels in rural Mississippi.

The analysis of Figure 3 yields the following observations:

- As the scale increases, so does the level of diversity or segregation. The larger the area, the more likely it will be populated by a variety of racial groups.
- Rural areas in Mississippi are less diverse and more segregated than Houston's urban areas. It manifests visually as clusters of green and yellow cells located in different parts of the mapped area.
- On the 3km scale in Houston areas, blue tiles represent Hispanic-dominated areas, whereas orange and red tiles (N_H between 3 and 4) represent racially mixed areas.
- At 12km, the spatio-racial pattern includes at least three racial groups occupying each tile.
- At the scale of 3km and 6km in the Mississippi area (panels I and J), we can clearly distinguish the location of small cities.

4. Use cases

The examples below illustrate how NRGD2020, together with *raceland* package, can be used to address racial geography problems similar to those discussed in demographic literature.

The examples include:

- comparing metrics of racial diversity and segregation between different parts of the metropolitan statistical area (MSA);
- constructing segregation profiles for the metropolitan statistical areas;
- calculating bi-racial segregation indices.

The R code to reproduce the examples is available from the section NRGD2020 at http://www.socscape.edu.pl. In addition, the Appendix A in this paper provides a computational framework to calculate entropy E and the mutual information MI for any user-defined region at any scale using the precalculated GIS layers (*RL grid racial ID* and *RL grid population density*) as an input data to the R *raceland* package. Box 1 shows the R code to perform the calculation.

4.1. Comparing racial diversity and segregation metrics between different parts of metropolitan statistical area (MSA)

A common inquiry in demographic research is the comparison of racial diversity and segregation measures within metropolitan areas. Such zones include, for example, core city and suburban areas.

Such analysis can also be performed using the NRGD2020 dataset and RL method implemented in the *raceland* package. In order to obtain the results presented in table in Fig. 4A two steps must be taken. A data preparation step involves cropping RL grid layers (*RL grid racial ID* and *RL grid population density*) to the boundaries of MSA, principal city, and suburban areas. The data should be prepared using GIS software rather than R due to the size of RL grids. As a second step, racial diversity and segregation metrics are calculated for each zone using the *raceland* package in R. The R code enabling calculations is presented in Box 1 in the Appendix A. Detailed, reproducible tutorial is presented in the section NRGD2020 at http://www.socscape.edu.pl.

Figure 4A shows the results for the Atlanta Metropolitan Statistical Area (MSA) in 2020. Based on typology proposed by Lichter et al. (2023), the counties within Atlanta MSA were classified into three suburban zones (inner-ring suburbs, outlying suburbs, fringe suburbs) and a principal city (established using US Census place subdivisions). There is a significant difference in the value of segregation and diversity within the MSA. The inner ring and outlying suburbs have one more racial group than the principal city. Rural fringe

A

zone	***MI***	N_H	***E***
MSA	0.193	3.68	1.88
principial city	0.258	3.23	1.69
inner ring suburbs (1)	0.316	3.66	1.87
outlying suburbs (2)	0.135	3.71	1.89
fringe suburbs (3)	0.123	2.13	1.09

B

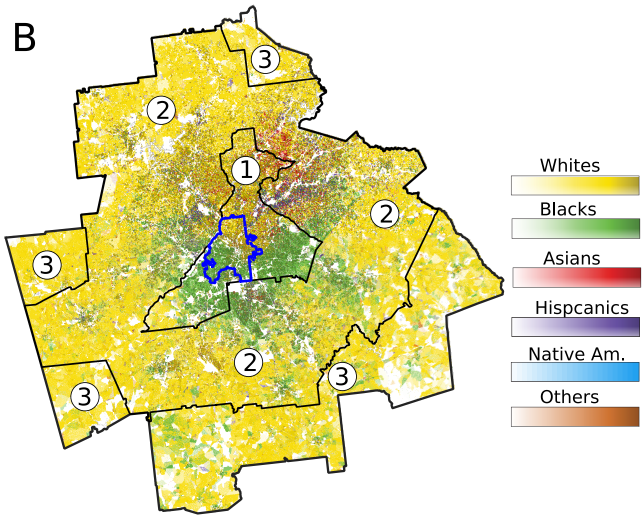

Figure 4: Comparing the segregation and diversity metrics between different parts of the Atlanta, MSA. (A) RL image showing the spatio-racial pattern in the Atlanta, MSA. Imposed are the boundaries of principal city (blue), and suburban areas. (B) The value of segregation and diversity metrics for zones within Atlanta, MSA.

suburbs on the outskirts of the MSA are primarily populated by two racial groups with a sizable population ($N_H = 2.13$).

The use of NRGD2020 and the *raceland* package significantly reduces the time and effort necessary to obtain such results. Furthermore, the RL method provides a visual representation of the results along with the numerical results. Such visualization obtained from the RL image can serve as a visual confirmation of the results. Fig. 4B illustrates how the spatio-racial pattern changes within the Atlanta MSA based on the RL image overlaid with four zones.

4.2. Segregation profiles

Another topic relevant to demographic studies is how segregation changes with scale (Reardon et al., 2008, 2009; Hennerdal and Nielsen, 2017; Clark et al., 2015; Jones et al., 2015; Olteanu et al., 2019; Owen et al., 2021). Reardon et al. (2008) introduced the concept of the segregation profile, which represents a function describing the segregation level across different scales. The x-axis shows the scales on which segregation is calculated; the y-axis shows the level of segregation. To compute the profiles, they multiple times calculated the information theory index H (Theil and Finizza, 1971; Theil, 1972) for local circular subdivisions with varying radius.

The example below shows how to construct a segregation profile using the precalculated segregation grids available in the NRGD2020 dataset. The segregation profile constructed with the NRGD2020 dataset differs conceptually from that introduced by Reardon et al. (2008). The difference arises from the different methods of calculating local values of segregation. The RL method doesn't calculate H values for local circular subdivisions. Instead, the scale is determined by dividing the analyzed area into square tiles of a particular size. Next, the value of segregation is calculated for each tile (expressed as mutual information MI). The value of segregation in a particular scale is calculated as the average of all MI values in this scale.

The NRGD2020 dataset contains 10 US-wide pre-calculated layers showing segregation metrics at scales of 72km, 36km, 24km, 18km, 12km, 9km, 6km, 3km, 1.5km, and 0.75km. In segregation grids, the scale is defined by cell size. For example, the scale 3km means that the area was divided into tiles 3x3km, and the metrics were calculated for each tile. The resulting grid has a resolution of 3x3km.

The example depicted in Figure 5 compares the segregation profiles calculated for five of the largest metropolitan statistical areas: New York NY, Los Angeles CA, Chicago IL, Dallas TX, and Houston TX. Los Angeles, CA and Dallas, TX show similar segregation profiles. On a scale of 750 meters, the segregation level in New York, NY is the lowest, while it is the highest in Chicago, IL. On the other hand, at the scale of 3km, the least segregated city is Los Angeles CA, whereas Houston, TX and Chicago, IL had similar segregation levels and were the most segregated among the analyzed cities.

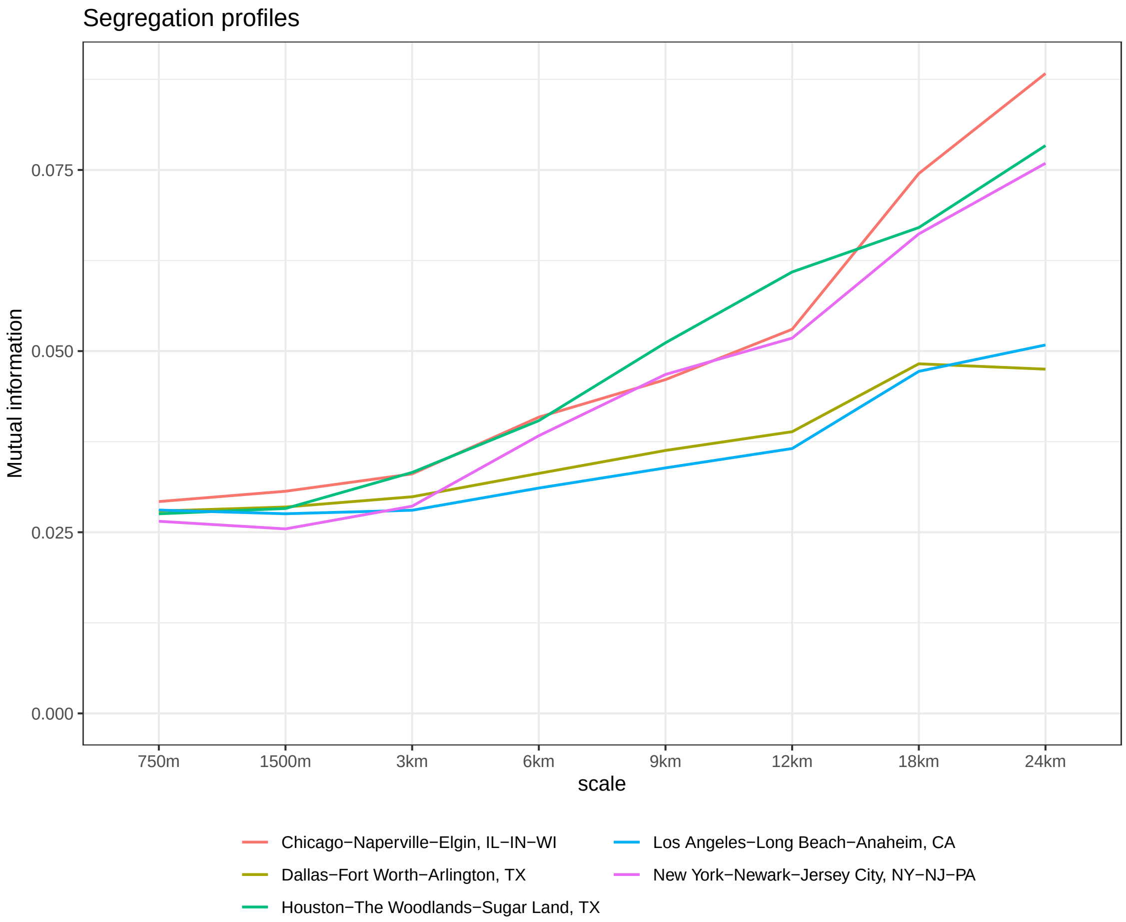

Figure 5: Segregation profiles for five the largest metropolitan areas.

The R code to create segregation profiles is available in the section NRGD2020 at http://socscape.edu.pl

Segregation profiles can be constructed for precalculated scales or for any user-defined scale. The NRGD2020 together with the *raceland* package allows to calculate segregation at user-defined scale (see the code in Box 1), and next use this layers to compute segregation profiles.

4.3. *Calculating racial segregation metrics for particular racial group*

Demographic studies also often examine the level of segregation between two racial groups. For example, how segregated is the Hispanic population? Traditionally, the dissimilarity index *D* assesses segregation between two groups (i.e., Logan et al. (2022)). In other studies, authors used multiracial information theory index *H* to assess the segregation level between one racial group and the remaining groups (i.e., Elbers (2021)).

Our example illustrates how to use the *raceland* package and the NRGD2020 to assess the level of segregation between one racial group and the remaining groups. Such calculations require one preprocessing step – reclassification of the *RL grid racial ID* into two categories: 1 – corresponding to one group (i.e. Hispanic), and 0 – corresponding to the remaining groups. The rest of the calculations are performed using the code presented in Box 1 in Appendix A. The complete tutorial showing how to perform calculations is available at http://socscape.edu.pl.

Figure 6 shows the disparity in segregation levels of Black and Hispanic/Latino populations in Atlanta, MSA in 2020. The results show that the Black population is much more segregated than the Hispanic/Latinos.

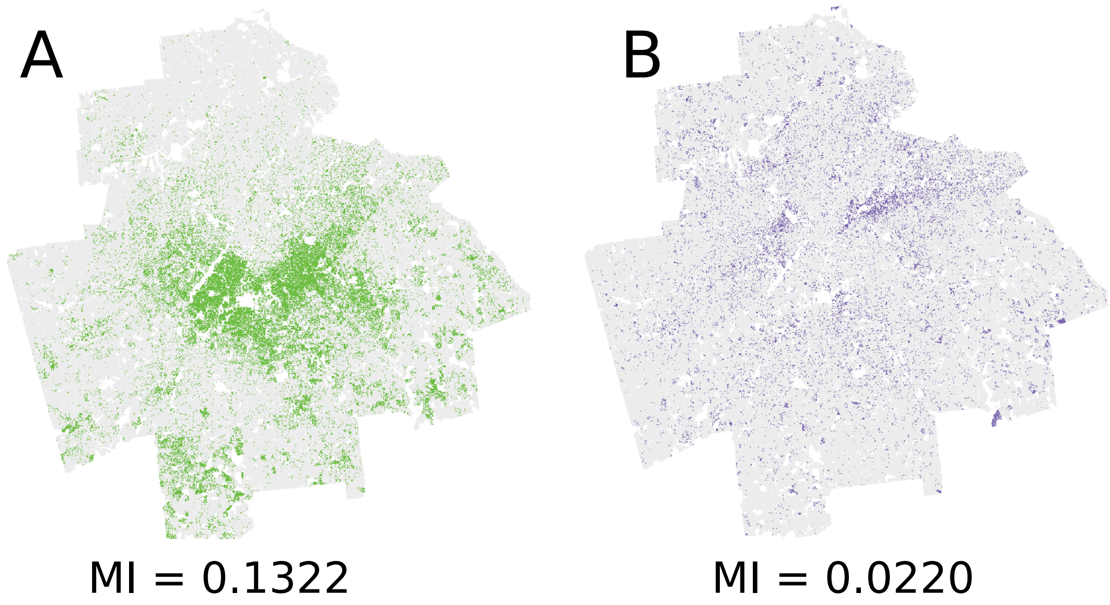

Figure 6: The distribution of Blacks (A) and Hispanic/Latino sub-populations in the Atlanta, MSA.

5. Conclusion and Discussion

One of the main challenges in demographic studies is acquiring the necessary census data. Generally, census data are available in tabular form at various levels of aggregation. Subdivision boundaries are stored separately in ESRI shapefile format. For spatial analysis, census data and boundary files had to be separately acquired and linked together. Such a process for larger areas is time-consuming, and requires computational effort. Moreover, creating maps from census data requires a few preprocessing steps, such as calculating racial percentages or categorizing data by dominant race and/or diversity types (Holloway et al., 2012; Wright et al., 2014; Fasenfest et al., 2004; Bader and Warkentien, 2016). Creating a segregation map at the census tract level would be even more difficult. Traditionally used single-number indices require the division of the analyzed area into subdivisions. Consequently, to map segregation at the census tract level, one needs to calculate the H index for each census tract using block-level data.

Our long-term goal is to provide the demographer's community with easy-to-use resources for mapping racial geography in the United States. In 2013, we started the SocScape project (http://socscape.edu.pl) – a research project that aims to offer free, ready-to-use resources of maps for visualizing and analyzing residential segregation and racial diversity in the conterminous United States and US metropolitan areas in 1990, 2000, 2010, 2020. A newly developed National Racial Geography Dataset 2020 provides another contribution to the SocScape project.

This paper demonstrates how to address important demographic research issues using the NRGD2020 dataset and the *raceland* package.

The advantage of using the NRGD2020 dataset along with the *raceland* package can be summarized as follows:

- NRGD2020 offers easy-to-use resources for mapping US racial geography, including racial distribution, segregation, and diversity maps at different scales.

- US-wide NRGD2020 dataset is available in GeoTiff format and is very easy to use within GIS software (QGIS, ArcGIS). A map of racial distribution, segregation, or diversity can be prepared by cropping the dataset based on the region of the analysis.

- NRGD2020 provides input for the *raceland* package in R for quantifying segregation and diversity in regions defined by the user. Racial segregation and diversity can be calculated on various scales. As demonstrated by Dmowska and Stepinski (2023) on the example of 51 MSA, the RL-based segregation metric provides similar ranking results as the traditionally used information theory index H (Theil and Finizza, 1971; Theil, 1972).

- Utilizing NRGD2020 and the *raceland* package allows connecting the numerical assessment of segregation given by the indices with the visualization of how racial groups are spatially distributed within an analyzed area.

In summary, NRGD2020 and the *raceland* package provide computational frameworks to visualize and quantify US racial geography. The next step will be to calculate similar datasets for the years 1990, 2000, and 2010 to allow for a temporal analysis.

References

Bader, M. D., Warkentien, S., 2016. The fragmented evolution of racial integration since the civil rights movements. Sociological Science 3, 133–166.

Chipman, J., Wright, R., Holloway, S., Ellis, M., 2009-2023. Mixed-Metro.US: Mapping Diversity and Segregation in the USA. URL `https://mixedmetro.com/`

Clark, W. A., Anderson, E., Östh, J., Malmberg, B., 2015. A multiscalar analysis of neighborhood composition in Los Angeles, 2000–2010: A location-based approach to segregation and diversity. Annals of the Association of American Geographers 105(6), 1260–1284.

Dmowska, A., Stepinski, T., 2023. Improving the understanding of the us urban racial geography and its temporal change using the racial landscape method.

Dmowska, A., Stepinski, T., Nowosad, J., 2021. A pattern-based approach to analysis and visualization of spatio-racial distribution.

Dmowska, A., Stepinski, T. F., 2017a. Mapping changes of racial composition in the united states: 1990–2010. In: Population Association of America 2017, Chicago, April 26 - 29 2017. https://paa.confex.com/paa/2017/meetingapp.cgi/Paper/10564.

Dmowska, A., Stepinski, T. F., 2017b. Mapping racial diversity using grid-based racial dot maps and racial diversity maps. In: Population Association of America 2019, Austin, April 10 - 13 2017. https://paa2019.populationassociation.org/uploads/191266.

Dmowska, A., Stepinski, T. F., 2019. Racial Dot Maps Based on Dasymetrically Modeled Gridded Population Data. Social Sciences 8 (5), 157.

Dmowska, A., Stepinski, T. F., Netzel, P., 2017. Comprehensive framework for visualizing and analyzing spatio-temporal dynamics of racial diversity in the entire united states. PLoS One 12 (3), e0174993.

Dmowska, A., Stepinski, T. F., Nowosad, J., 2020. Racial Landscapes–a pattern-based, zoneless method for analysis and visualization of racial topography. Applied Geography 122, 102239.

Elbers, B., 2021. Trends in US Residential Racial Segregation, 1990 to 2020. Socius 7, 1–3.

Fasenfest, D., Booza, J., Metzger, K., 2004. Living Together : A New Look at Racial and Ethnic Integration in Metropolitan Neighborhoods, 1990-2000. Tech. rep., Brookings Institution.

Hennerdal, P., Nielsen, M. M., 2017. A multiscalar approach for identifying clusters and segregation patterns that avoids the modifiable areal unit problem. Annals of the American Association of Geographers 107 (3), 555–574.

Hill, M. O., 1973. Diversity and evenness: a unifying notation and its consequences. Ecology 54, 427–432.

Holloway, S. R., Wright, R., Ellis, M., 2012. The racially fragmented city? Neighborhood racial segregation and diversity jointly considered. The Professional Geographer 64, 63–82.

Huiping, L., Fernandez, S., Ganguly, A., 2005. Racial geography, economic growth and natural disaster resilience. House of Representative. Committee on Science and Technology, Washington, DC.

Jones, K., Johnston, R., Manley, D., Owen, D., Charlton, C., 2015. Ethnic residential segregation: A multilevel, multigroup, multiscale approach exemplified by London in 2011. Demography 52 (6), 1995–2019.

Lichter, D. T., Thiede, B. C., Brooks, M. M., 2023. Racial Diversity and Segregation: Comparing Principal Cities, Inner-Ring Suburbs, Outlying Suburbs, and the Suburban Fringe. RSF: The Russell Sage Foundation Journal of the Social Sciences 9 (1), 26–51.

Logan, J. R., 2015. Chalanges of spatial thinking. In: Howell, F. M., Porter, J. R., Matthews, S. A. (Eds.), Recapturing Space: New Middle-Range Theory in Spatial Demography. Springer, Ch. 2, pp. 11–36.

Logan, J. R., Stults, B. J., et al., 2022. Metropolitan Segregation: No Breakthrough in Sight. Tech. rep., Working Papers 22-14, Center for Economic Studies, U.S. Census Bureau.

Massey, D. S., Denton, N. A., 1988. The dimensions of residential segregation. Social forces 67(2), 281–315.

Olteanu, M., Randon-Furling, J., Clark, W. A., 2019. Segregation through the multiscalar lens. Proceedings of the National Academy of Sciences 116(25), 12250–12254.

Owen, G., Manley, D., Johnston, R., Birabi, T., Song, H., Wang, B., 2021. Multi-scale inequality and segregation: Theory and estimation. Urban Inequality and Segregation in Europe and China: Towards a New Dialogue, 207–231.

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`

Reardon, S. F., Farrell, C. R., Matthews, S. a., O'Sullivan, D., Bischoff, K., Firebaugh, G., 2009. Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990–2000. Social Science Research 38 (1), 55–70.

Reardon, S. F., Firebaugh, G., 2002. Measures of Multigroup Segregation. Sociological Methodology 32(1), 33–67.

Reardon, S. F., Matthews, S. A., Sullivan, D. O., Lee, B. A., Firebaugh, G., Farrell, C. R., 2008. The geographic scale of metropolitan racial segregation. Demography 45 (3), 489–514.

Roth, R. E., 2010. Dot density maps. The Encyclopedia of Geography, 787–790.

Shannon, C. E., 1948. A mathematical theory of communication. Bell system technical journal 27(3), 379–423.

Theil, H., 1972. Statistical Decomposition Analysis. Amsterdam: North.

Theil, H., Finizza, A. J., 1971. A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts. The Journal of Mathematical Sociology 1(2), 187–193.

White, M. J., 1986. Segregation and diversity measures in population distribution. Population index, 198–221.

Wright, R., Ellis, M., Holloway, S. R., Wong, S., 2014. Patterns of racial diversity and segregation in the United States: 1990–2010. The Professional Geographer 66(2), 173–182.

Appendix A – Using NRGD and *raceland* package

In Figure A1, we provide a computational framework for calculating entropy E and mutual information MI for any user-defined region at any scale using GIS layers (*RL grid racial ID* and *RL grid population density*) as inputs to the R *raceland* package. Figure A1 illustrates an example of the input grids (Fig. A1B and A1C) and the resultant grid (Fig. A1D) for the principal city of Atlanta, GA.

Box 1 shows the R code to perform the computations, including the calculation of segregation and diversity metrics for user-defined regions and selected scales.

In the first step, we read RL grids to R using *terra* package. Secondly, we calculate segregation and diversity metrics for the analyzed area using the function *calculate_metrics()* from the *raceland* package.

We can also divide Atlanta into tiles, i.e., 600 x 600 m, and calculate metrics for each tile. It would allow the creation of a map showing how diversity and segregation varies within a city. In the *raceland* package, scale is defined by the size parameter expressed as a number of cells. The actual size of the tile is $sizeparameter \times cellresolution$. In our example, we define size as 20, which gives us $20 \times 30m = 600m$. The last part of the code uses the *create_grid()* function from *raceland* package to create the spatial object required to prepare a final map.

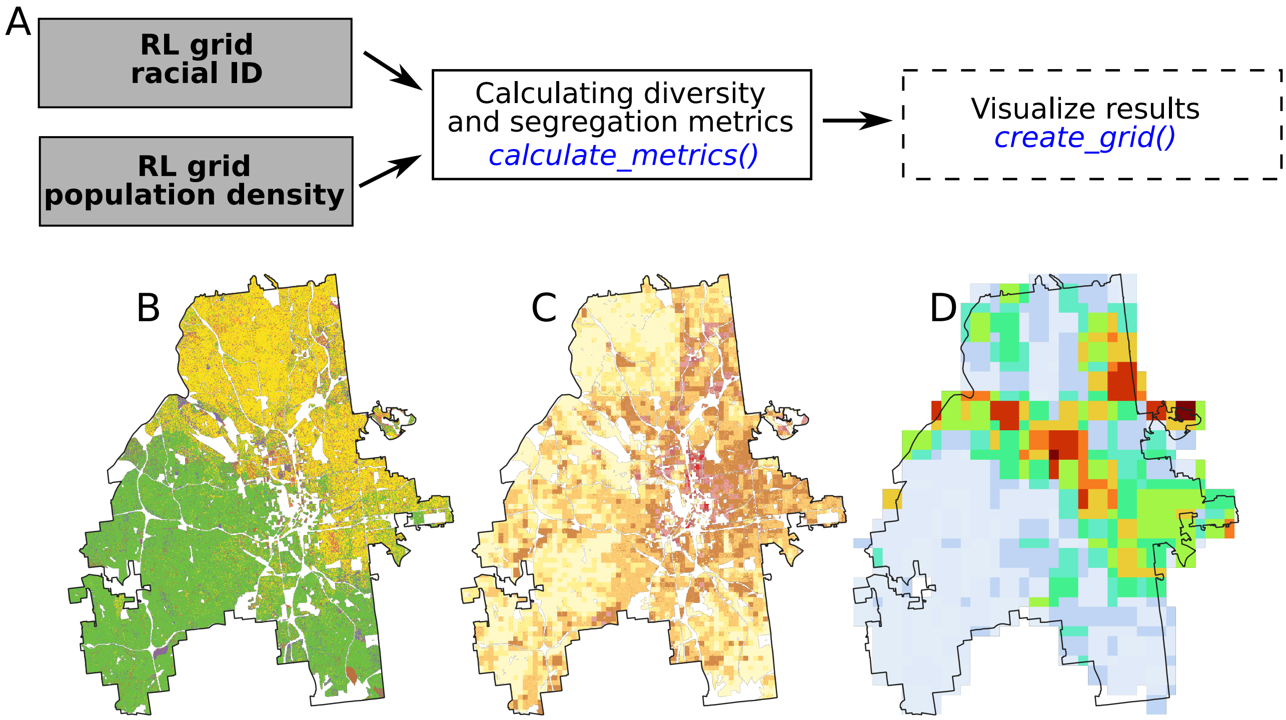

Figure A1: (A) A computational framework for analyzing and visualizing racial distribution using NRGD2020 dataset and *raceland* package. (B) Input layer *RL grid racial ID* for the principal city of Atlanta, GA. (C) Input layer *RL grid population density* for the principal city of Atlanta, GA. (D) Racial diversity map at the scale of 600m for the principal city of Atlanta, GA.

Box 1. Using NRGD2020 and *raceland* package

```
library(raceland)
library(terra)
library(sf)
#set working directory to the folder containing RL grids.
setwd("")

#read RL grid racial ID layer to R
rl = rast("rl_city.tif")
#read RL grid population density layer to R
rd = rast("rd_city.tif")

#calculate metrics using raceland package
metr = calculate_metrics(rl, rd, fun = "mean", threshold = 1)
#return metrics in a user-friendly format
res = c(ENTROPY = metr$ent, HILL = 2^metr$ent,
MI = metr$mutinf, NMI = metr$mutinf/metr$ent)
round(res, 4)

#calculate metrics for the scale of 600m (20x30m).
metr_scale = calculate_metrics(rl, rd, fun = "mean", threshold = 1, size = 20)

#transformation of entropy to Hil's number
metr_scale$Hill = 2^metr_scale$ent

#metrics can be joined with the spatial object to prepare visualization.
grid_sf = create_grid(rl, size = 20)
attr_grid = dplyr::left_join(grid_sf, metr_scale, by = c("row", "col"))

#save grid to prepare map in GIS software
st_write(attr_grid, "attr_grid.shp")
```