

# MAPPING AND ANALYZING RACIAL GEOGRAPHY IN THE UNITED STATES USING NATIONAL RACIAL GEOGRAPHY DATASET

Anna Dmowska<sup>1</sup>, Tomasz Stepinski<sup>2</sup>  
(1) Adam Mickiewicz University, Poznan, Poland, (2) University of Cincinnati, Cincinnati, OH, US

## RACIAL LANDSCAPE METHOD

## NATIONAL RACIAL GEOGRAPHY DATASET

The Racial Landscape method introduces a consistent framework for visualization and quantification of spatial distribution of racial groups in arbitrary, user-defined regions. Racial Landscape (RL) is a high-resolution grid divided into small, racially homogenous cells. RL provides a map showing racial distribution, but at the same time serves as geospatial data to calculate racial segregation and diversity indices. By using the same data visual and numerical assessment is connected.

### How to create racial landscape?

**STEP 1:** We start from the US census data aggregated at the census block-level (the smallest and most homogenous census subdivisions). Block-level data provide information about racial composition within a block, but we don't know the actual spatial distribution of people within a block.

**STEP 2:** RL method randomly redistributes people from block into monoracial cells. Each cell has only one racial group. Cells have different population density.

**STEP 3:** We create two layers:  
- **RL racial ID grid** - a categorical raster generated by assigning to each cell racial ID based on the racial group present in this cell.  
- **RL population density grid** - each cell has assigned population density based on the number of people of given race within a block.

RL racial ID and RL population density grids can be transformed into **RL image** - a high-resolution RGB image that provides a visualization of racial pattern within analyzed area.

RL racial ID and RL population density grids serve also as geospatial data used to calculate racial diversity and segregation indices directly from grid.

**STEP 4: Quantification of spatio-racial pattern from grid**  
RL method quantifies spatio-racial pattern using an exposure matrix - a modification of the co-occurrence matrix. The co-occurrence matrix is a two-way table that summarize cell adjacencies (i.e. how many times purple and yellow cell appear next to each other). In the exposure matrix each pair contributes the value of their average population densities instead of 1.

Calculating entry to the exposure matrix

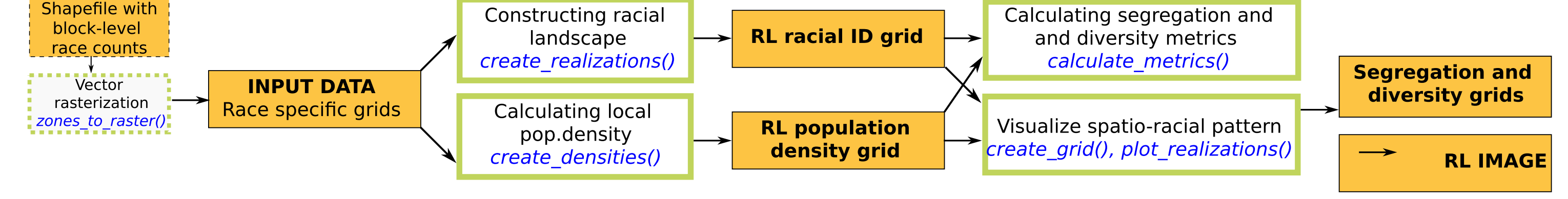
$$\frac{48}{2} + \frac{9}{2} = 5 + 5 = 10$$

**STEP 5: Quantification of racial diversity and segregation metrics**  
The exposure matrix is further summarized using two metrics derived from Information Theory - entropy and mutual information.  
**Entropy (E)** measures racial diversity. It can be translated into standardized entropy or Hill's number ( $N_H = e^E$ ,  $e$  is a logarithm base equal to 2 in RL).  $N_H$  depicts the number of racial groups with a significant shares in the analyzed area.  
**Mutual information (MI)** measures racial segregation. In RL, segregation is visually depicted as clumps created by the same-color cells. MI is a measures of clumping. A large MI value indicates large segregation (racial pattern creates larger clumps). A small MI value indicates mixed racial pattern.

MI = 0.093  
E = 1.072

### R package raceland

R package raceland provides complete computational framework to visualize and quantify racial pattern using RL method. Package is available from the R-cran repository.

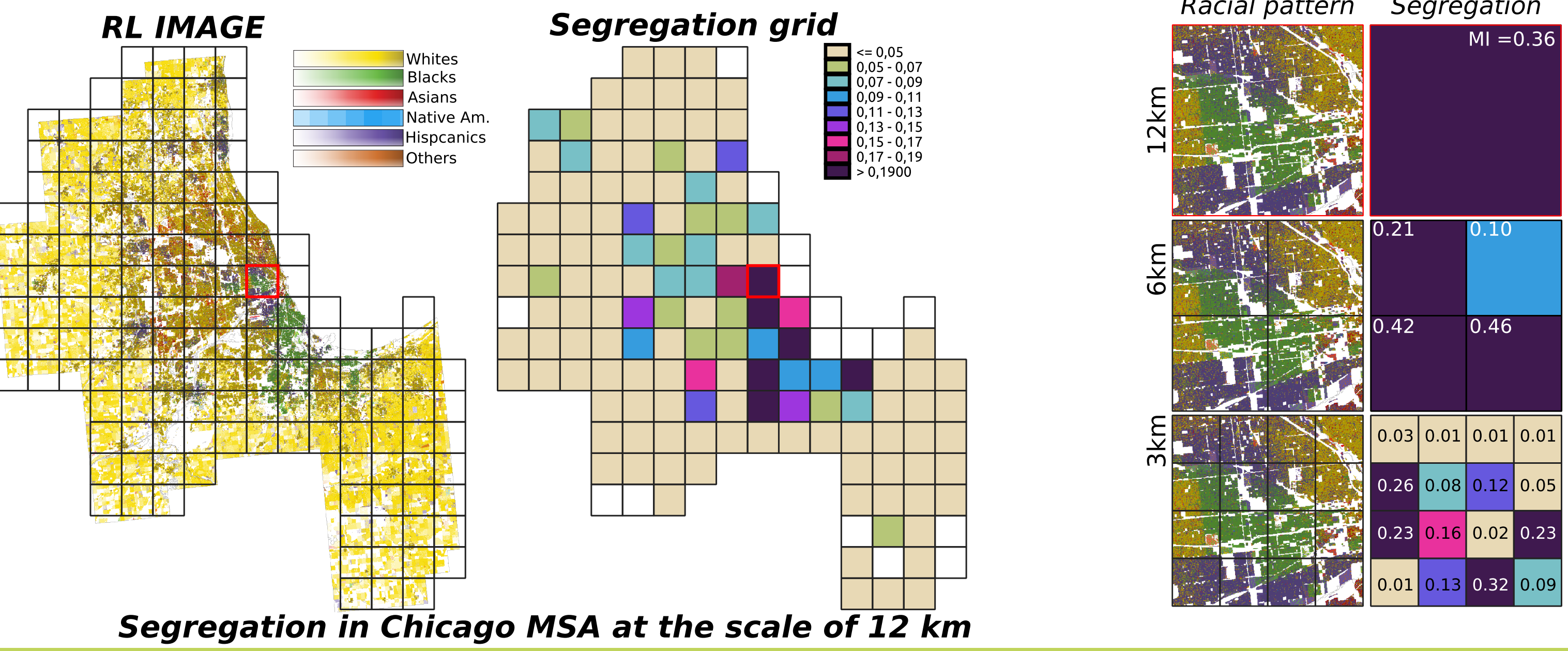


## How segregation changes with a scale?

The NRGD provides 10 precalculated US-wide segregation grids showing the level of segregation at the scale of 72km, 36km, 24km, 18km, 12km, 9km, 6km, 3km, 1.5km, 0.75km.

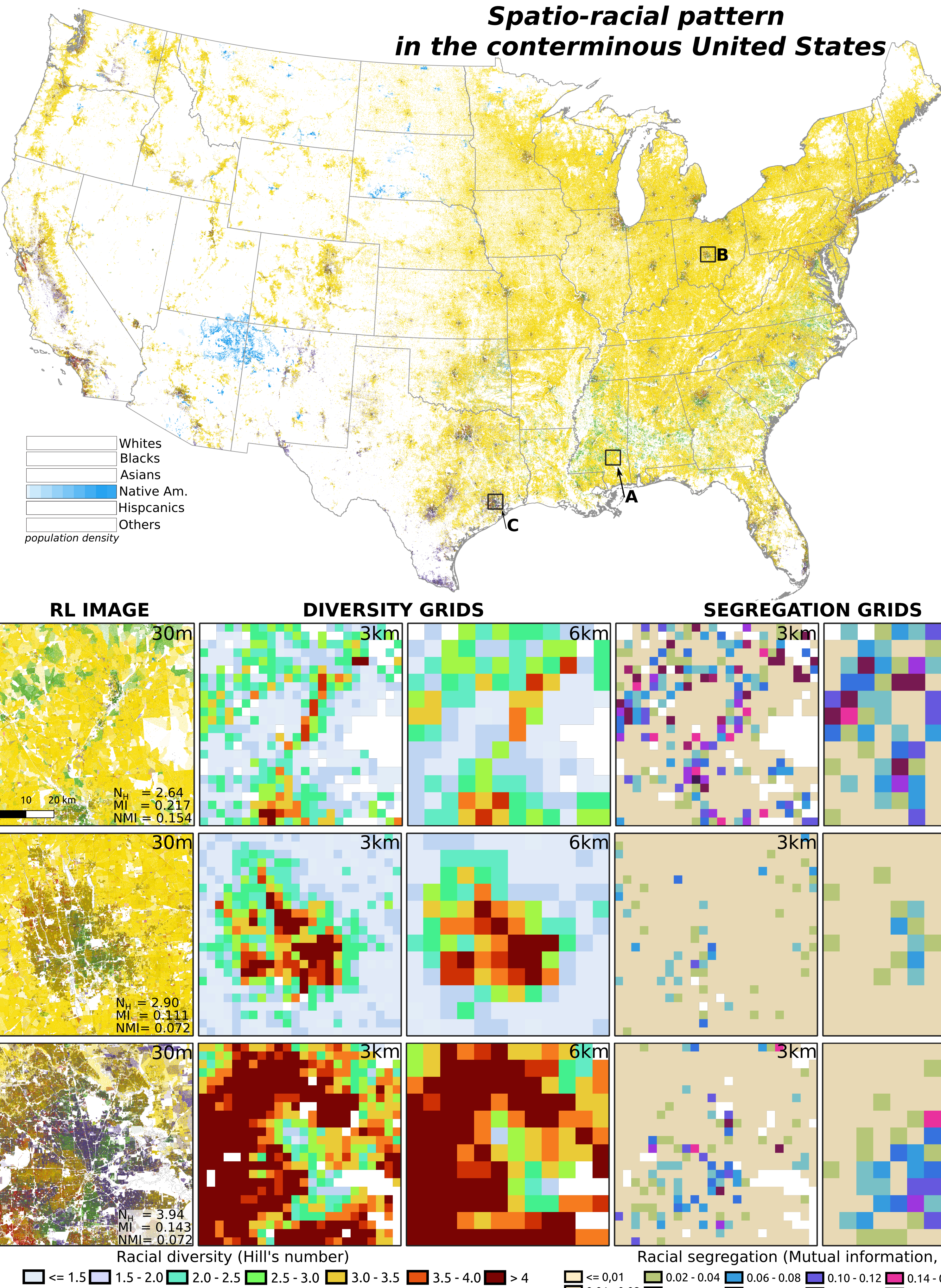
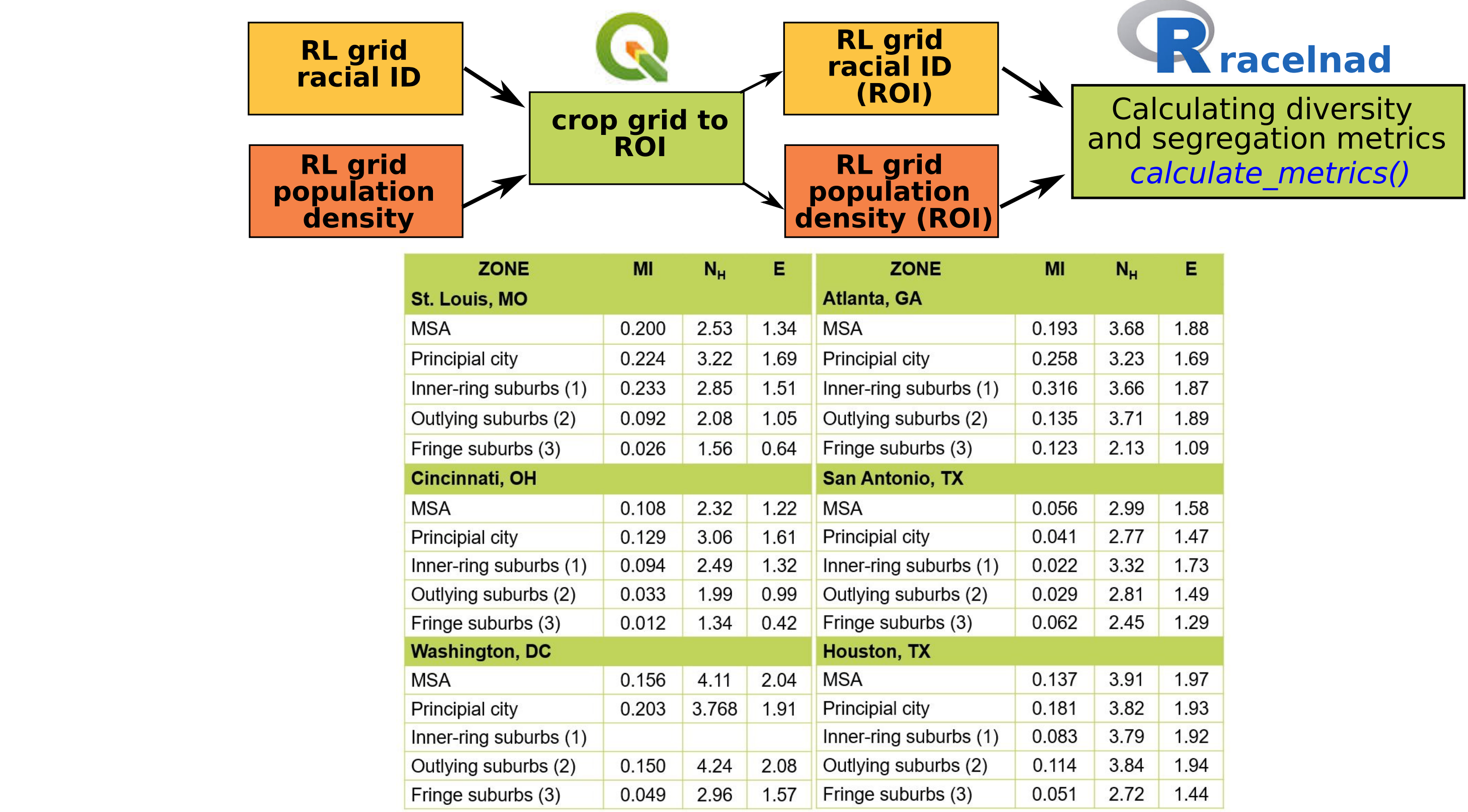
In segregation grids the scale is defined by the cell size. For example, the scale of 6km means that the area was divided into tiles 6x6km and the mutual information metric was calculated for each tile.

Using the *raceland* package with the *RL racial ID* and *RL population density grids* the segregation can be calculated for any user-defined scale.



## Comparing racial diversity and segregation between different parts of the MSA

A common inquiry in demographic studies involves comparing various zones within metropolitan statistical areas (MSAs) regarding racial diversity and segregation. For example, how do racial diversity and segregation change between suburban areas and a principal city? Such analysis can be also performed using NRGD2020 and raceland package.



The National Racial Geography Dataset (NRGD2020) provides a high-resolution racial database for the conterminous US in 2020. The NRGD2020 consists of a collection of precalculated GIS layers that provides ready-to-use resources for visualization and quantitative analysis of racial geography for any place in the conterminous United States. The NRGD2020 has been created using Racial Landscape methodology.

LAYER	RESOLUTION	DESCRIPTION
RL image	30-m	An RGB image that provides a US-wide visualization of racial geography. <i>Use: Visualization</i>
Diversity grids	72-km, 36-km, 24-km, 18-km, 12-km, 9-km, 6-km, 3-km, 1.5-km, 0.75-km	US-wide grids showing spatial variability of diversity (Hill number, N <sub>H</sub> ) at ten different length scales. <i>Use: Visualization and quantification</i>
Segregation grids		US-wide grids showing spatial variability of segregation (MI) at ten different length scales. <i>Use: Visualization and quantification</i>
RL racial ID grid	30-m	A raster in which each cell has a label corresponding to one of six races. <i>Use: An input for the R package raceland which calculates metrics of segregation and diversity for an arbitrary area.</i>
RL population density grid	30-m	A raster in which each cell has a numerical value of local population density. <i>Use: An input for the R package raceland.</i>

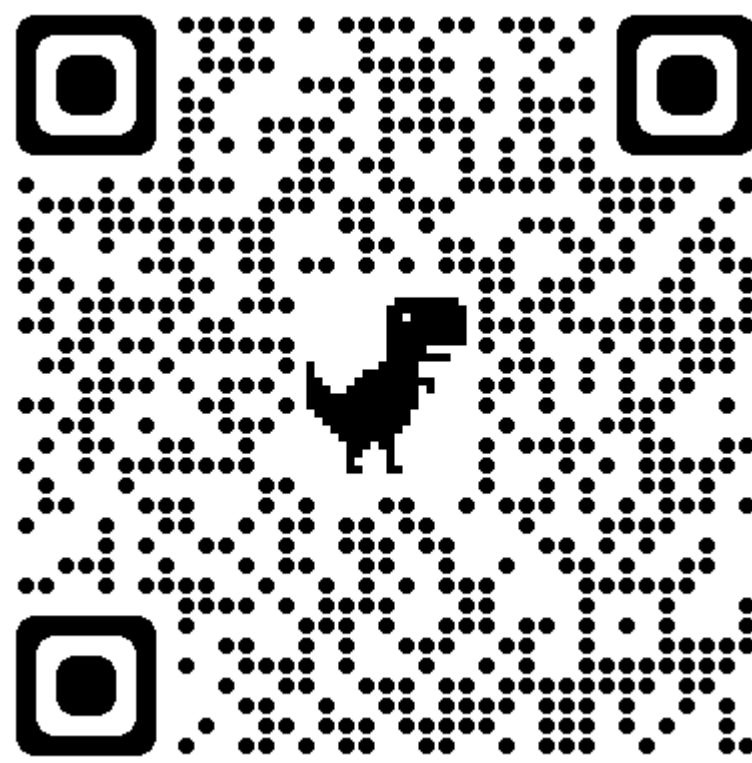
## USE CASES

The National Racial Geography Dataset (NRGD2020) together with the R package raceland can be employed to address racial geography problems similar to those examined in demographic literature. The examples include:

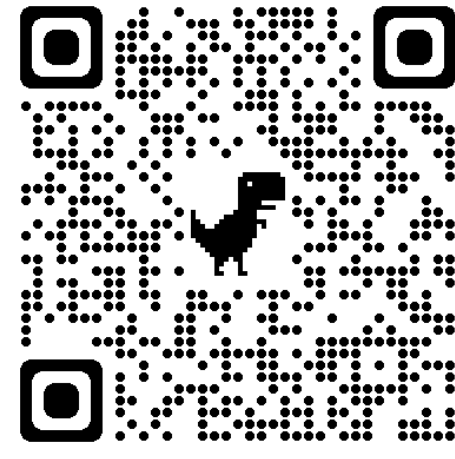
- mapping racial segregation and diversity;
- comparing racial diversity and segregation metrics between different parts of the metropolitan areas;
- constructing segregation profiles that shows how segregation change between different scales;
- calculating segregation at user-defined scales.

## Using NRGD and raceland

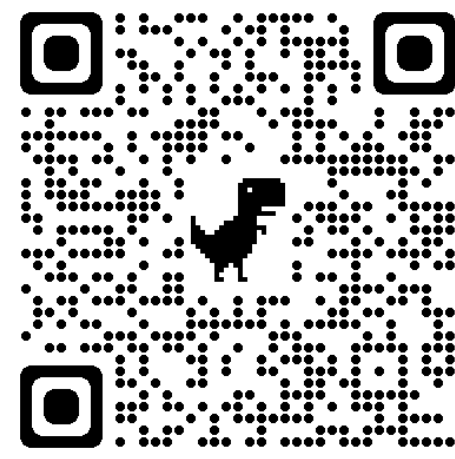
The National Racial Geography Dataset (NRGD2020) together with the R package raceland can be employed to address racial geography problems similar to those examined in demographic literature. The tutorial with the reproducible examples is available at <http://socscape.edu.pl>



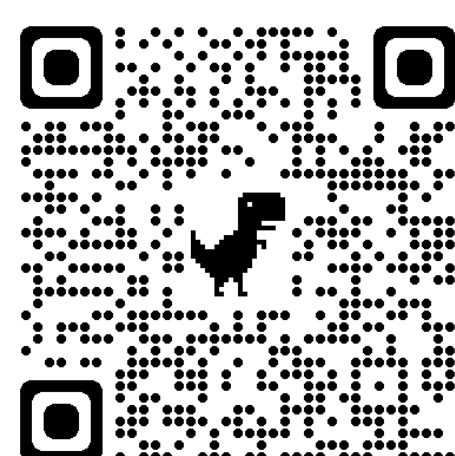
### Reproducible examples of application



Comparing racial diversity and segregation between different parts of the metropolitan area.



Segregation profiles - how segregation change with the scale?



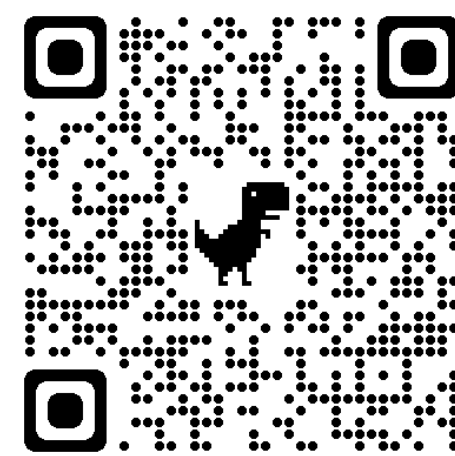
Calculating racial segregation and diversity metrics for the user-defined scale.

### More information

SocScape

Mapping racial segregation and diversity

<https://socscape.edu.pl>



raceland documentation

Contact information:  
Anna Dmowska, [dmowska@amu.edu.pl](mailto:dmowska@amu.edu.pl)