

SA-TIED Geospatial Analysis Workshop

Overview



Dr Justin van Dijk



j.t.vandijk@ucl.ac.uk



Objectives

We will explore the following topics:

- Fundamentals of using R for data analysis.
- Creating thematic maps using R.
- Quantifying the degree of spatial dependence in a dataset.
- Incorporating space into statistical models.

Schedule

Day 1 – Morning	R for Data Analysis
Day 1 – Afternoon	R for Spatial Analysis
Day 2 – Morning	Spatial Autocorrelation
Day 2 – Afternoon	Spatial Models

SA-TIED Geospatial Analysis Workshop

S04 – Spatial Models



This session

- Spatial models.
- Geographically weighted statistics.
- Geographically weighted associations.
- More spatial analysis.

Linear models

- Used to determine the relationship or association between a dependent with one or more independent variables.
- Important assumptions: homoscedasticity and independence of residuals.
- Violating this assumption can lead to inefficient estimates and unreliable hypothesis tests.

Linear models

When building a model based on spatial data:

- Map the residuals of the linear model to visually inspect for spatial patterns.
- Calculate Moran's I statistic on the residuals to assess spatial autocorrelation.
- If spatial autocorrelation is present, fit a **spatial linear model** to account for it.
- Recalculate Moran's I statistic on the residuals of the spatial model to confirm that the autocorrelation has been addressed.

Spatial models

A **spatial error model** adjusts for spatial autocorrelation by adding a spatially lagged error term to the regression equation:

$$y = X\beta + v, v = \lambda Wv + \epsilon$$

where $X\beta$ represents the standard regression components, λ is a spatial autoregressive parameter, W represents the spatially weights matrix, and u is a vector of spatially autocorrelated errors.

Spatial models

A **spatial lag model** incorporates a spatially lagged dependent variable, which is the weighted sum of the dependent variable values in neighboring locations, into the regression equation:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where ρ is the spatial autoregressive coefficient, $\mathbf{W}\mathbf{y}$ represents the spatially lagged dependent variable, and $\mathbf{X}\boldsymbol{\beta}$ represents the standard regression components.

Spatial models

- Both the spatial error and spatial lag models assume that the relationships between variables are the same across the study area, with adjustments made only for spatial dependencies.
- A Lagrange Multiplier Test can be used to make a decision as to which of these two models is most appropriate.
- What about non-stationarity?

Geographically weighted statistics

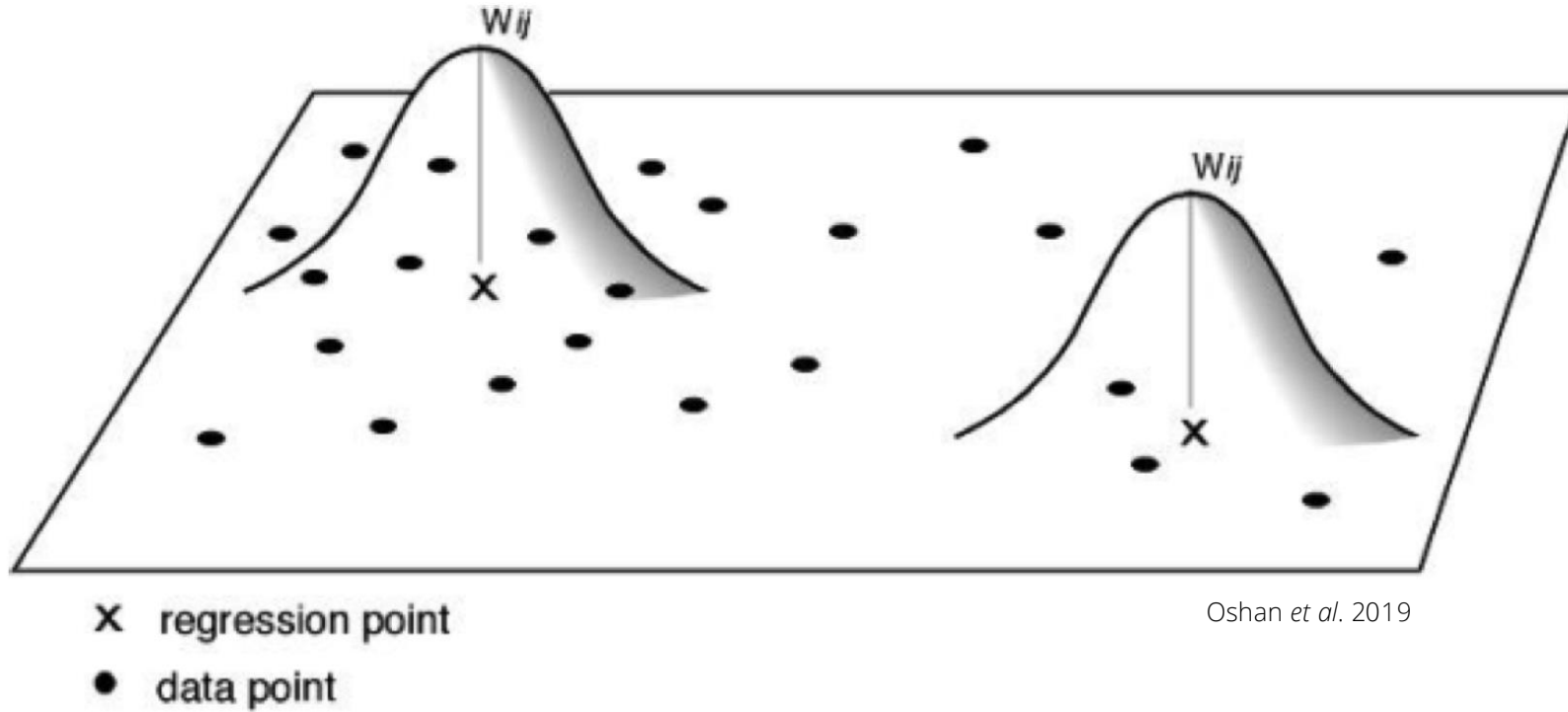
- Unlike traditional global models, which estimate a single set of parameters for the entire study area, geographically weighted statistics allow for parameter estimates that vary across different locations.
- Local means, local standard deviations, local variances.
- Typically uses some **kernel function** to weigh observations based on their distance from the location of interest.

Geographically weighted statistics

“Everything is related to everything else, but near things are more related than distant things.”

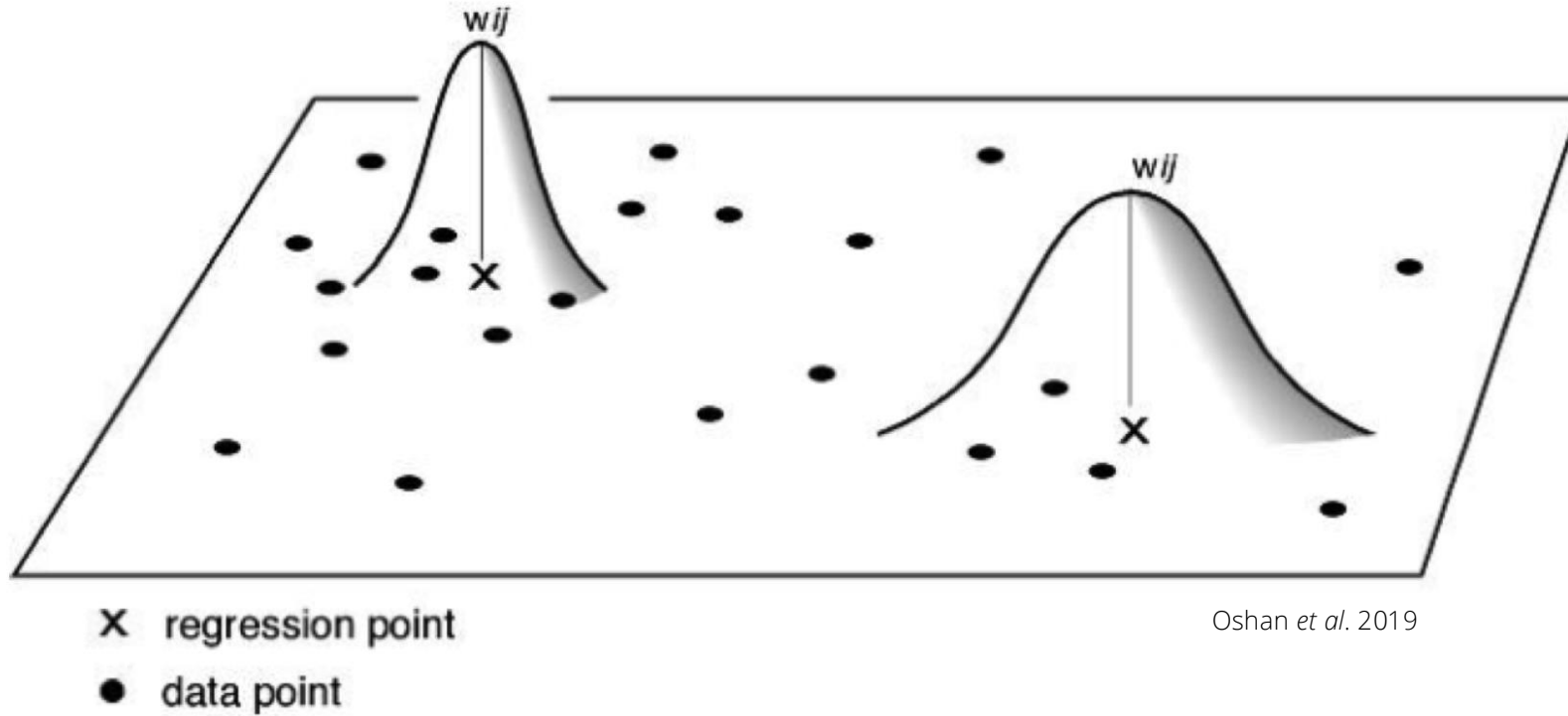
Walter Tobler 1970

Geographically weighted statistics



Oshan *et al.* 2019

Geographically weighted statistics



Oshan *et al.* 2019

Geographically weighted associations

- These ideas can be extended to correlation and regression:
 - Geographically weighted correlation (GWC)
 - Geographically weighted regression (GWR)
- The basic GWR equation is:

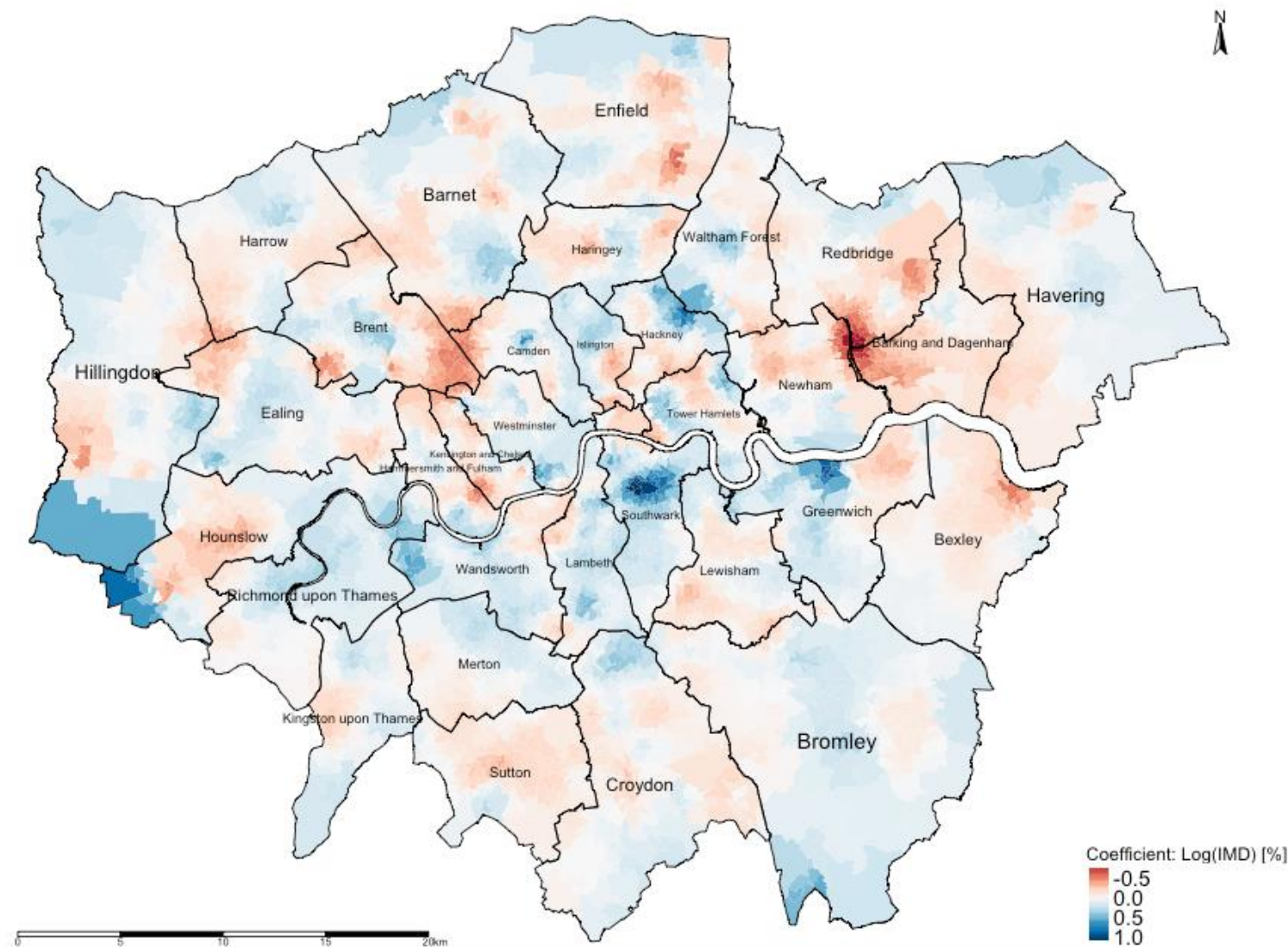
$$y_i = \beta_0(v_i, v_i) + \sum_{k=1}^p \beta_k(v_i, v_i) x_{ik} + \epsilon_i$$

where (v_i, v_i) are the coordinates of location i and $\beta_k(v_i, v_i)$ are the location-specific coefficients.

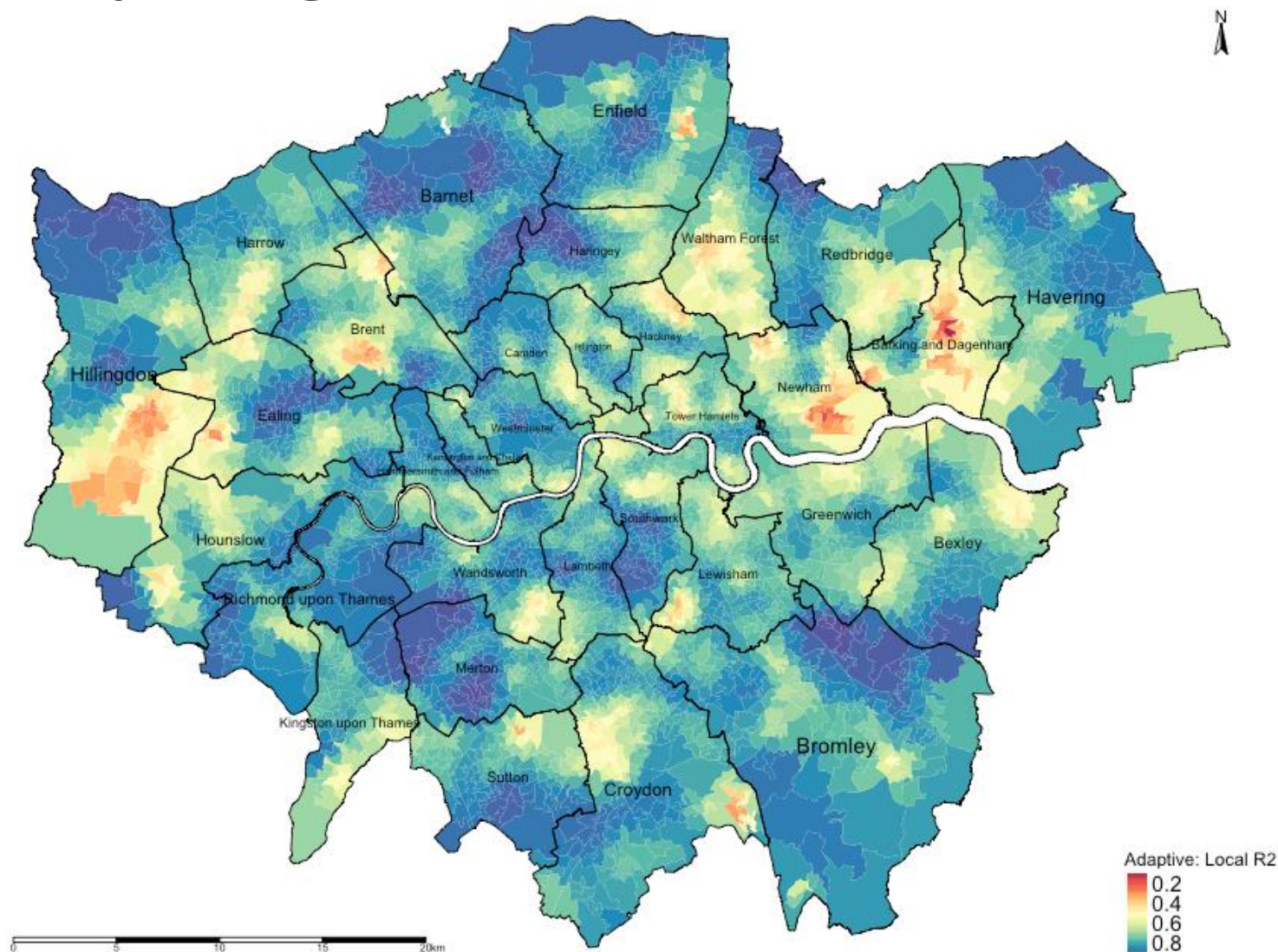
Geographically weighted associations

- Each area has its own set of regression coefficients.
- Each location has its own R^2 value.
- Each area has its own standard errors for the coefficient.
- More recently: bandwidths can vary between different variables.

Geographically weighted associations



Geographically weighted associations



Questions

Justin van Dijk
j.t.vandijk@ucl.ac.uk



SA-TIED Geospatial Analysis Workshop

More spatial analysis



More spatial analysis

- File formats.
- Map projections.
- Digitisation and geocoding.
- Spatial operations.
- Accessibility analysis.
- Geodemographic classification.
- Raster data.

File formats: Geopackage

- A GeoPackage is an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial data.
- It stores spatial data layer as a single file, based upon an SQLite database.
- How to spot in the wild: `.gpkg`

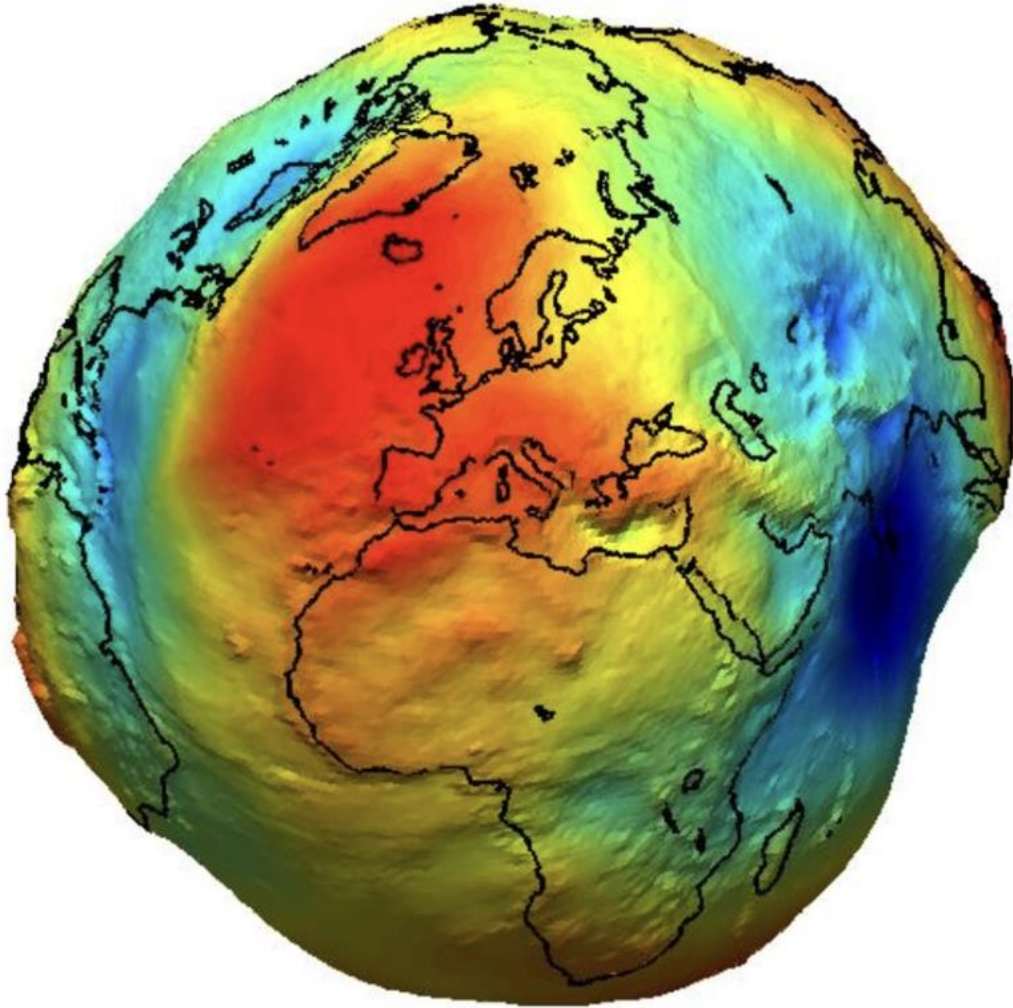
File formats: Shapefile

- `.shp` contains the feature geometry. *Mandatory.*
- `.shx` index file which stores the position of the feature's ID in the `.shp` file. *Mandatory.*
- `.dbf` stores all attribute information associated with the records. *Mandatory.*
- `.prj` contains the coordinate system information and projection. *Optional but not really.*
- `.xml` general metadata. *Optional.*
- `.cpg` encoding information. *Optional.*
- `.sbn` optimisation file for spatial queries. *Optional.*

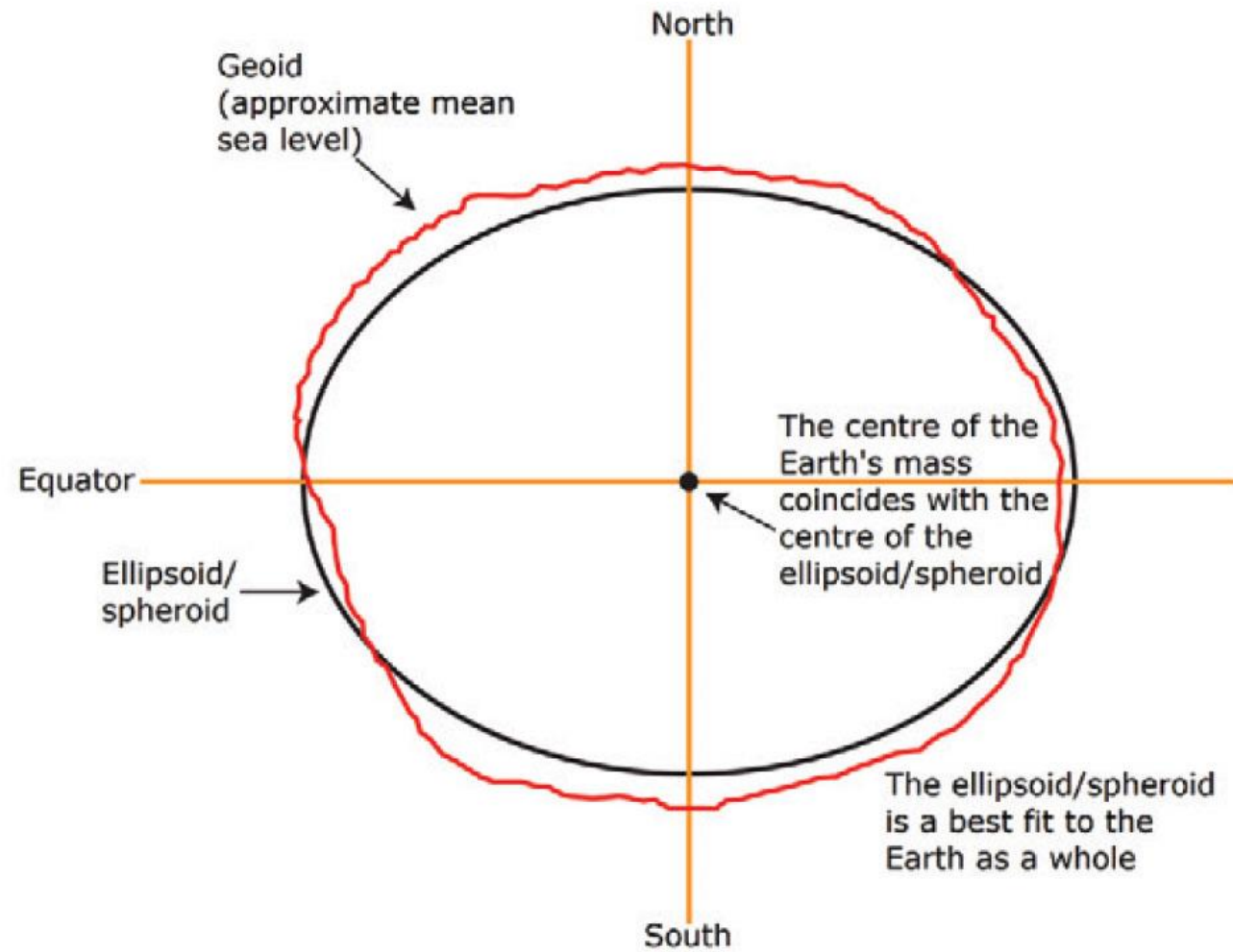
File formats: Shapefile



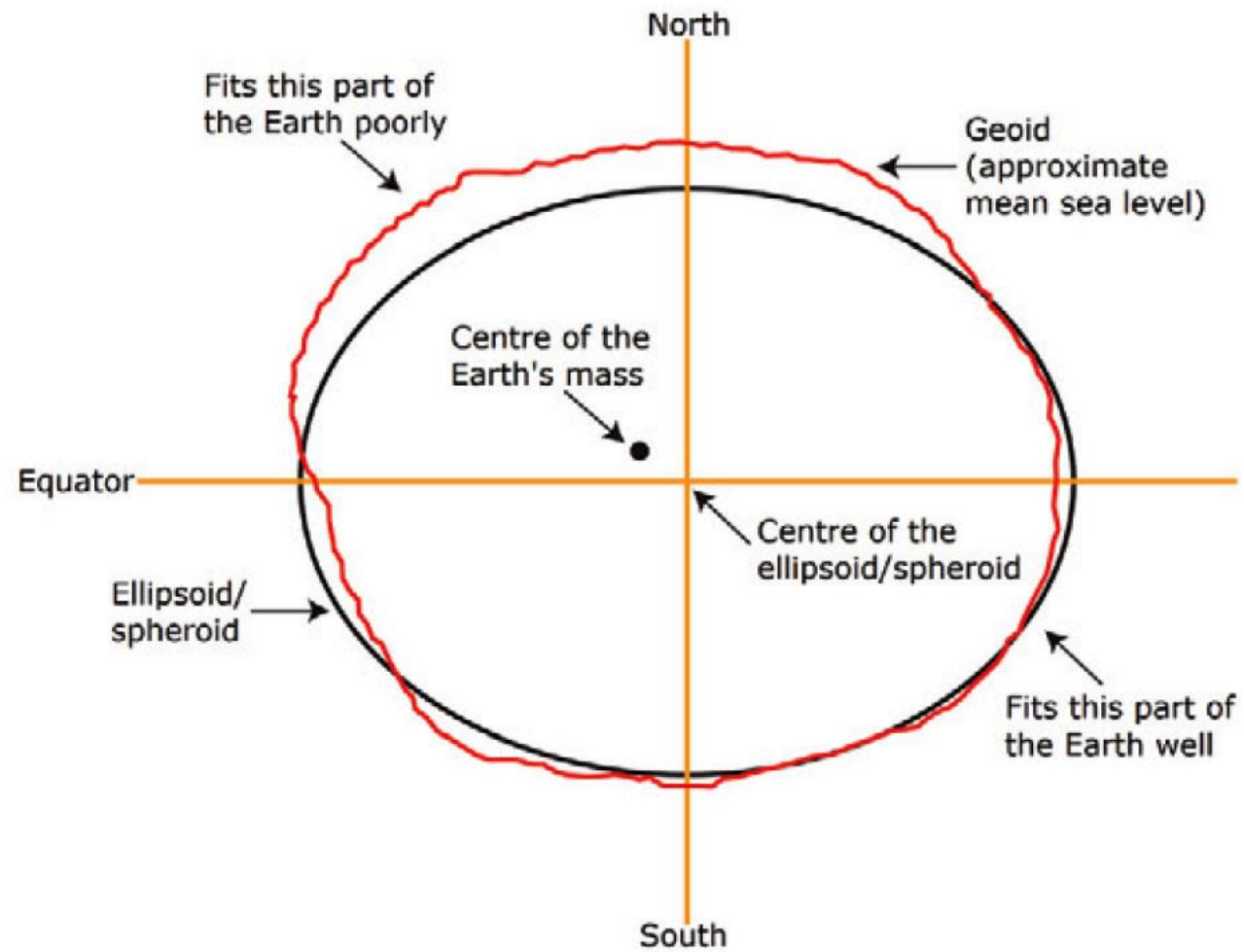
Map projections



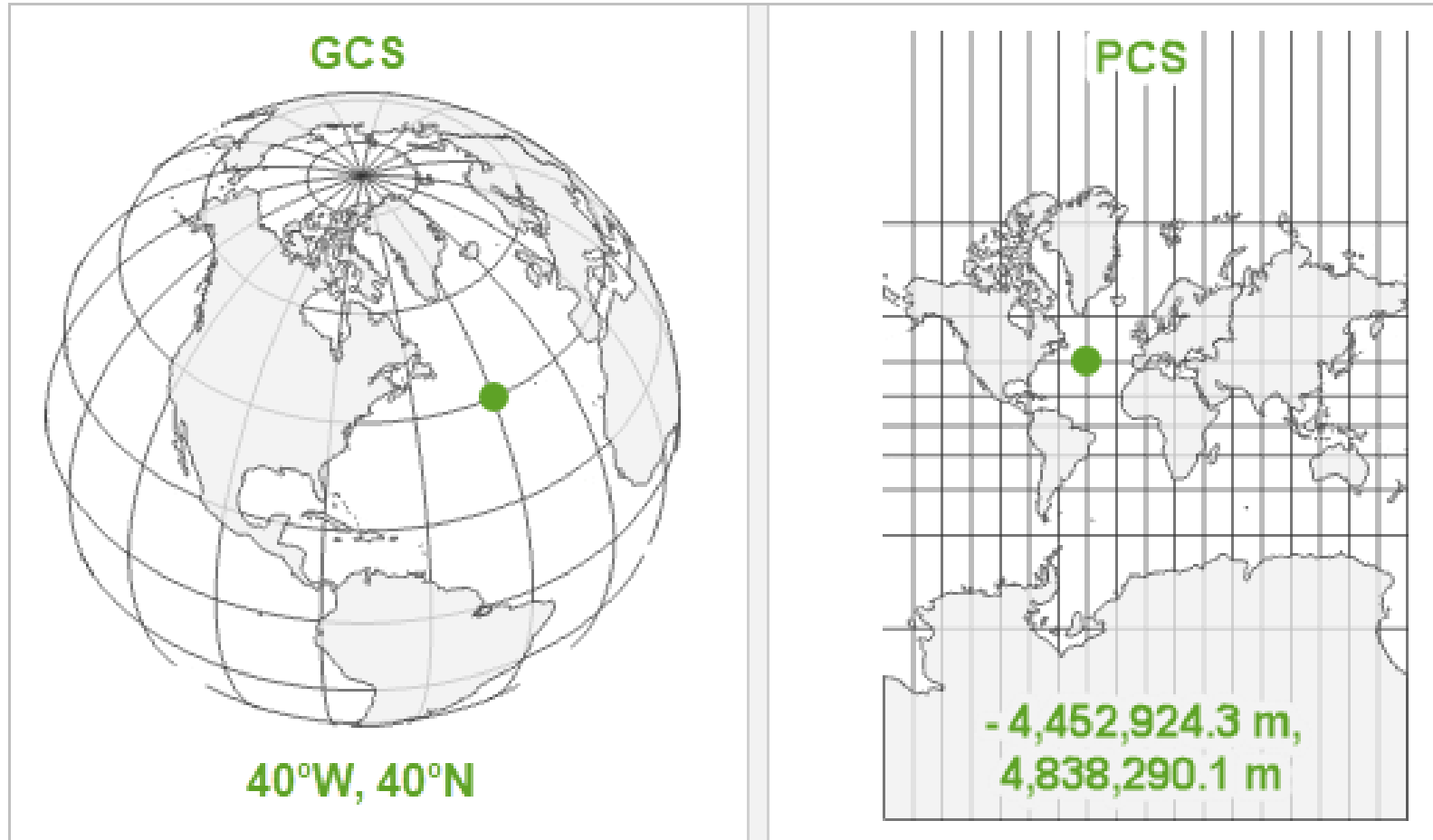
Map projections



Map projections



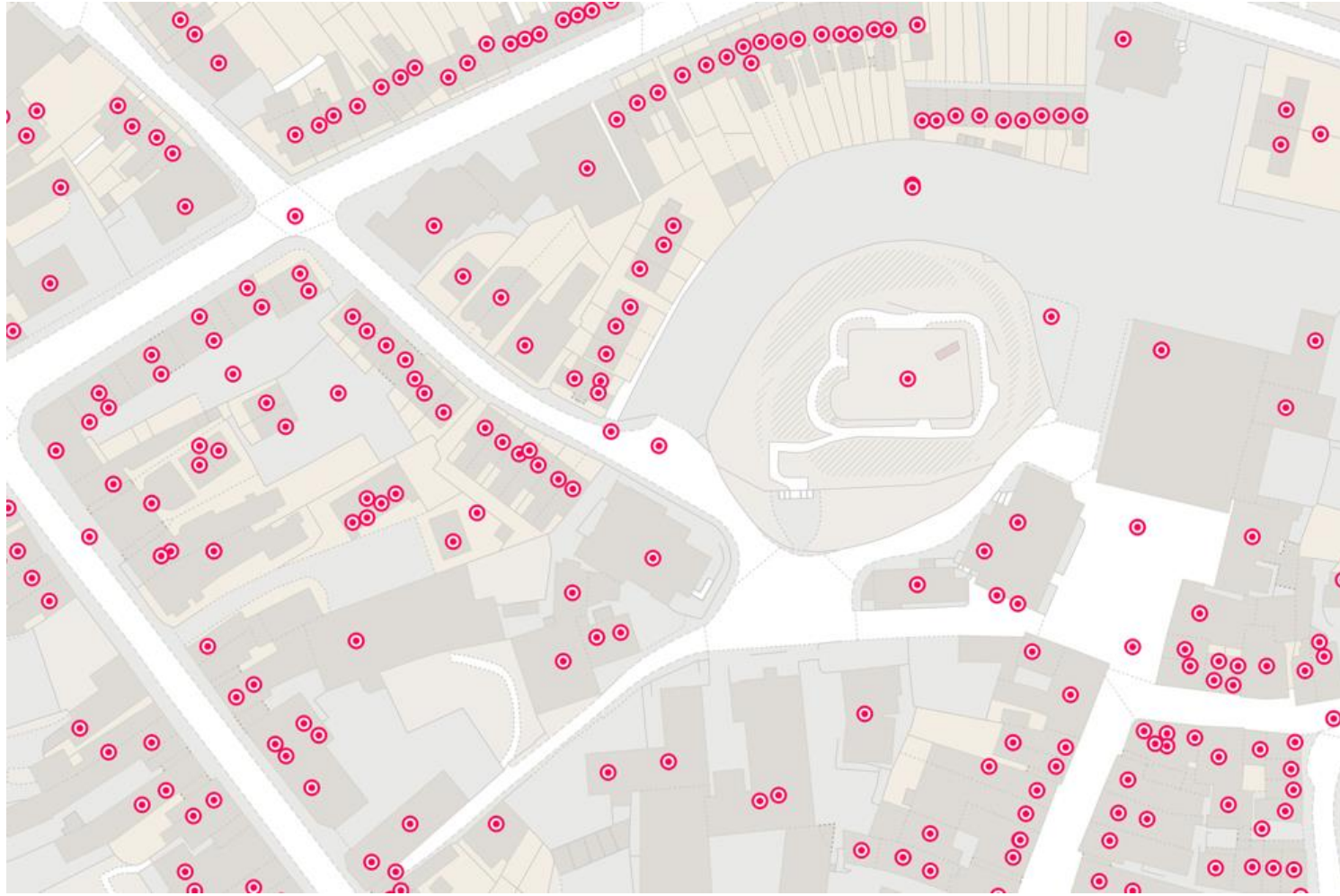
Map projections



Geocoding

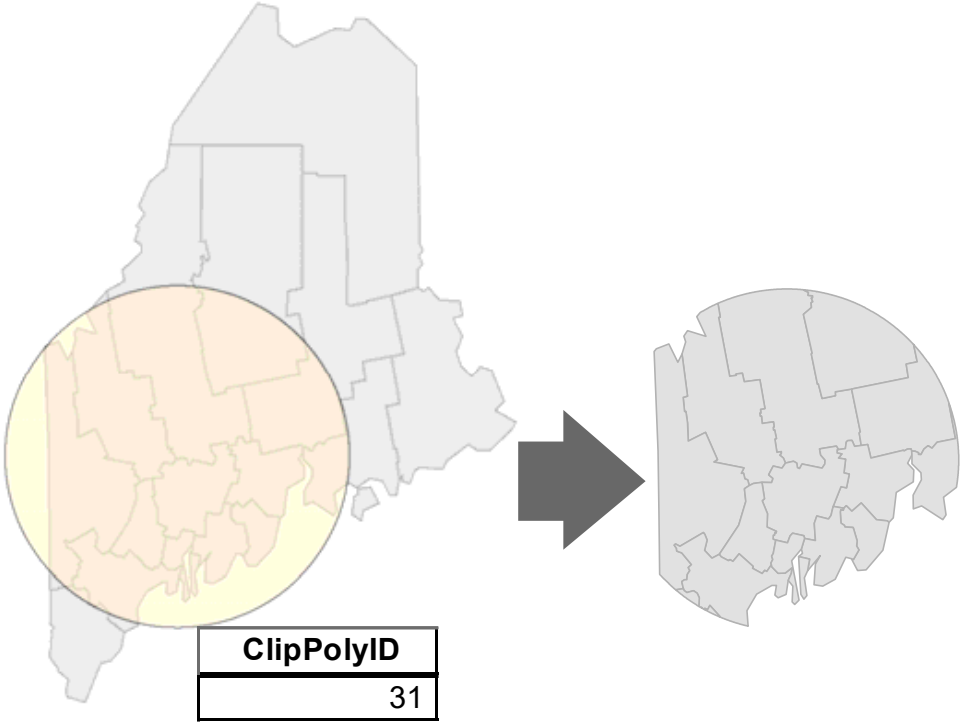
Forename	Surname	Address
Justin	van Dijk	Flat 18 Terry House SW22NT London

Geocoding



Spatial operations

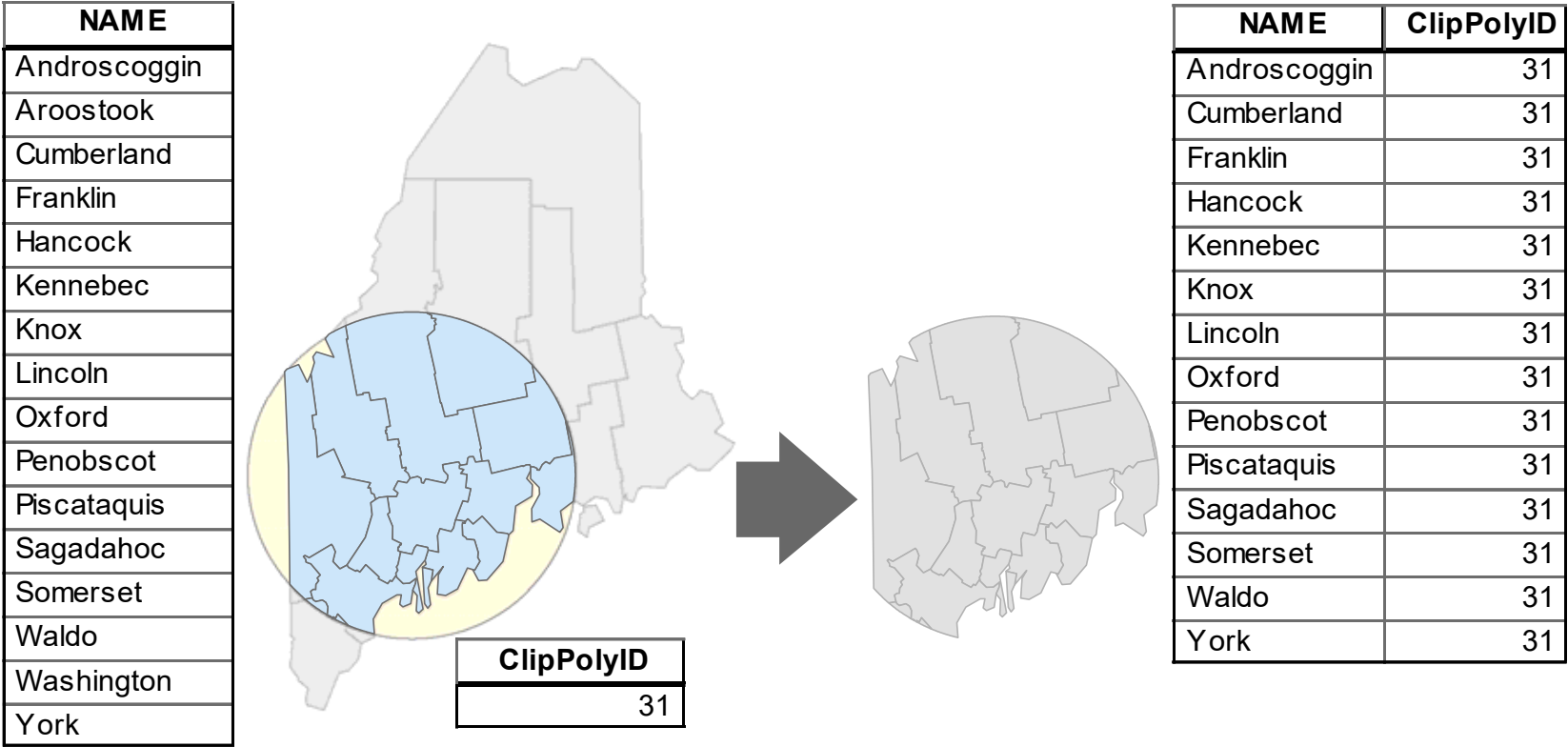
NAME
Androscoggin
Aroostook
Cumberland
Franklin
Hancock
Kennebec
Knox
Lincoln
Oxford
Penobscot
Piscataquis
Sagadahoc
Somerset
Waldo
Washington
York



NAME
Androscoggin
Cumberland
Franklin
Hancock
Kennebec
Knox
Lincoln
Oxford
Penobscot
Piscataquis
Sagadahoc
Somerset
Waldo
York

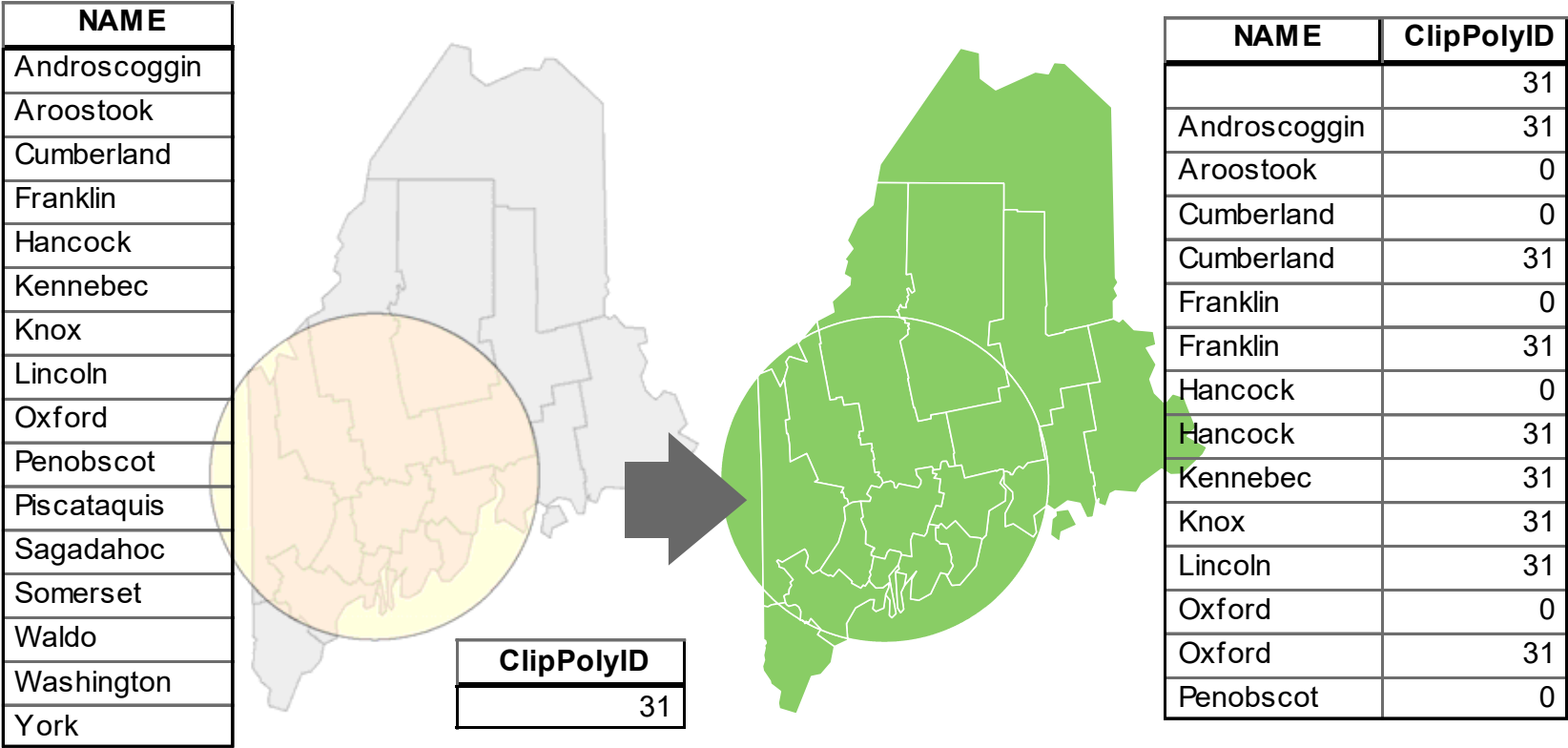
Gimond, M. 2021. Intro to GIS and Spatial Analysis. [online]
<https://mgimond.github.io/Spatial/introGIS.html>

Spatial operations



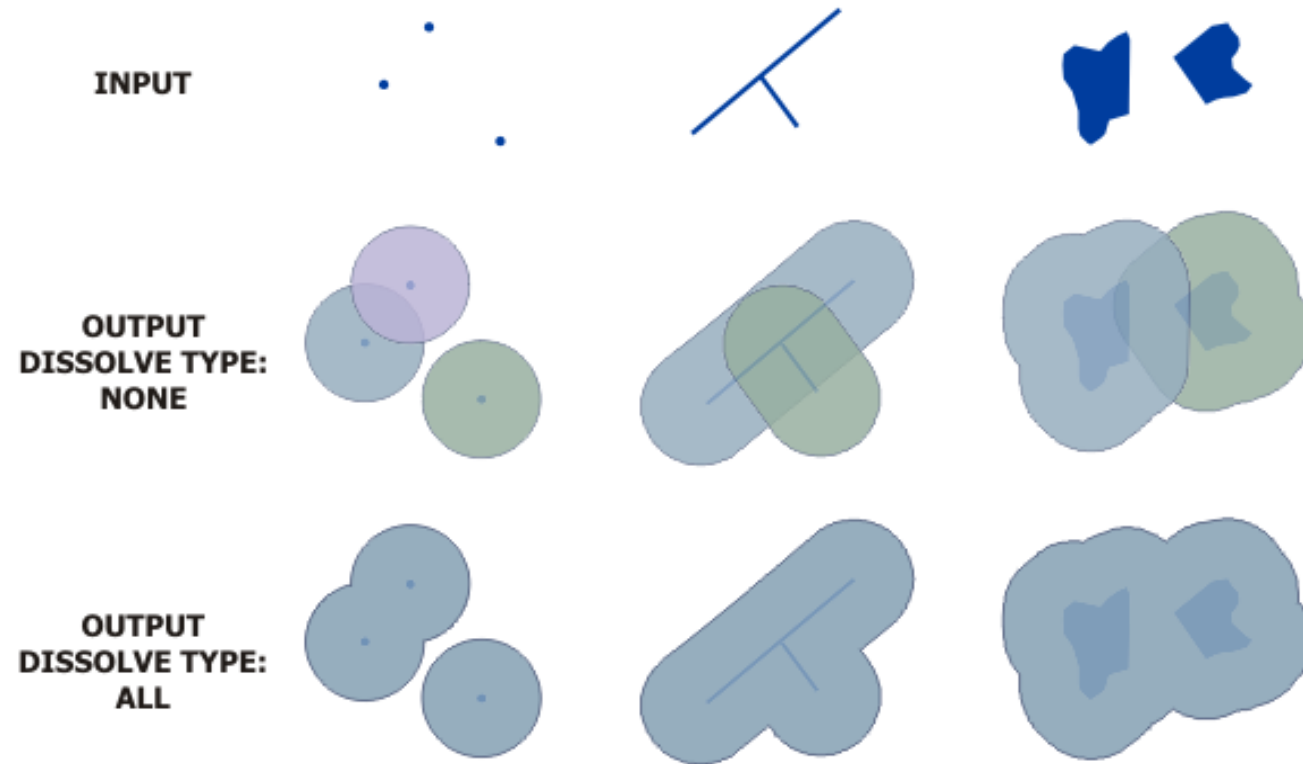
Gimond, M. 2021. Intro to GIS and Spatial Analysis. [online]
<https://mgimond.github.io/Spatial/introGIS.html>

Spatial operations



Gimdond, M. 2021. Intro to GIS and Spatial Analysis. [online]
<https://mgimond.github.io/Spatial/introGIS.html>

Spatial operations



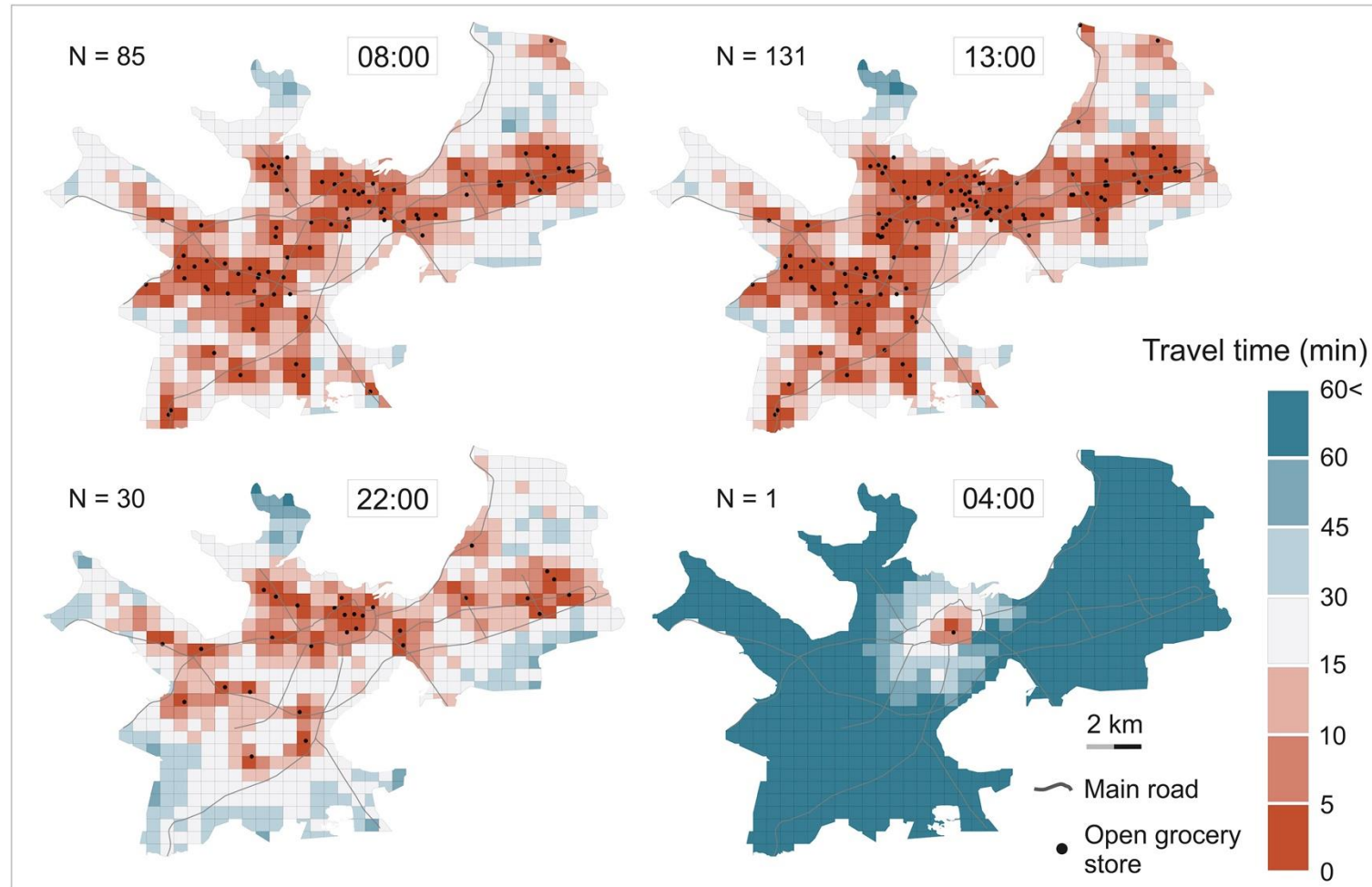
ESRI. 2021. Buffer. [online]

<https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/buffer.htm>

Spatial operations



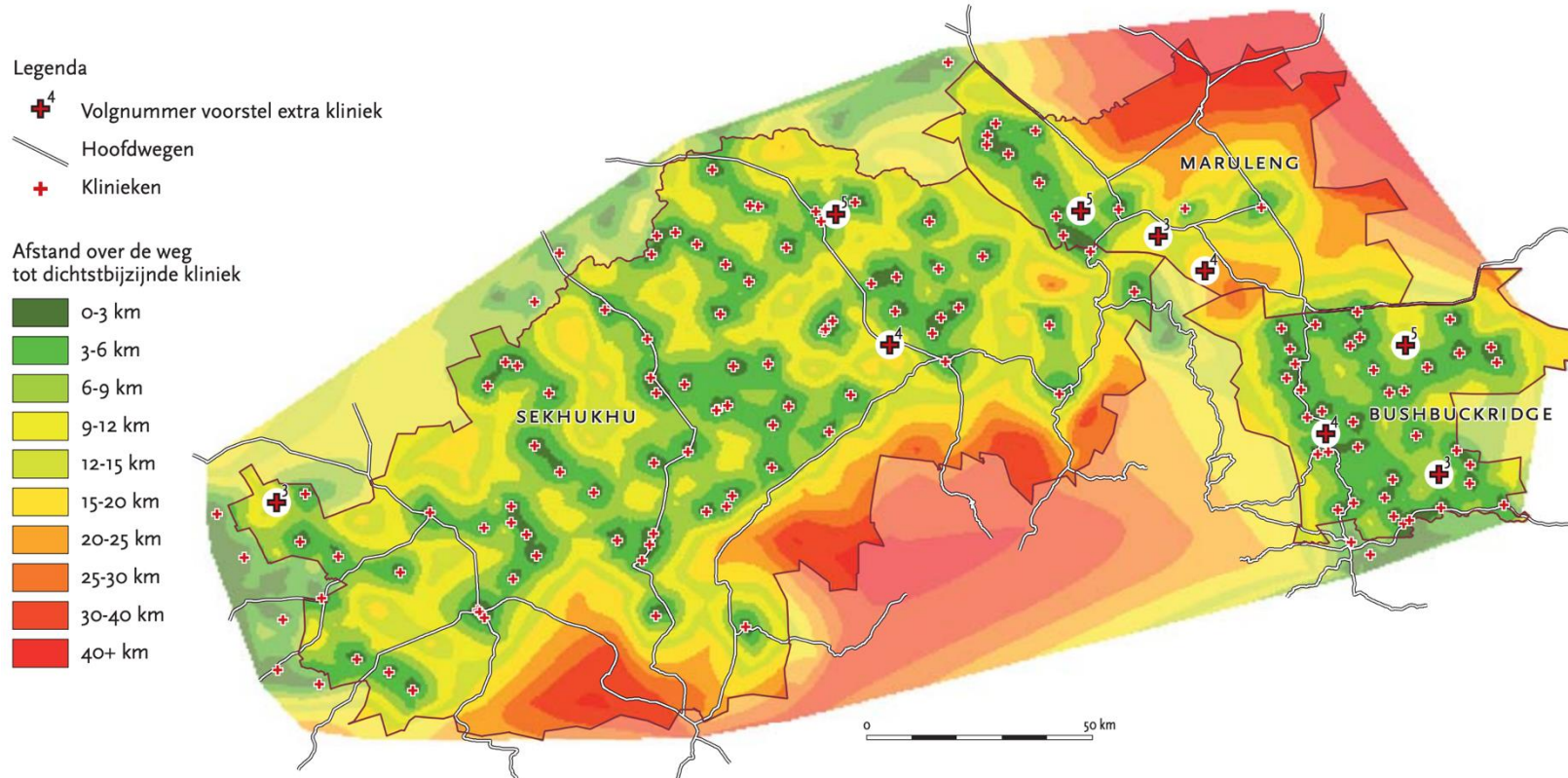
Accessibility analysis



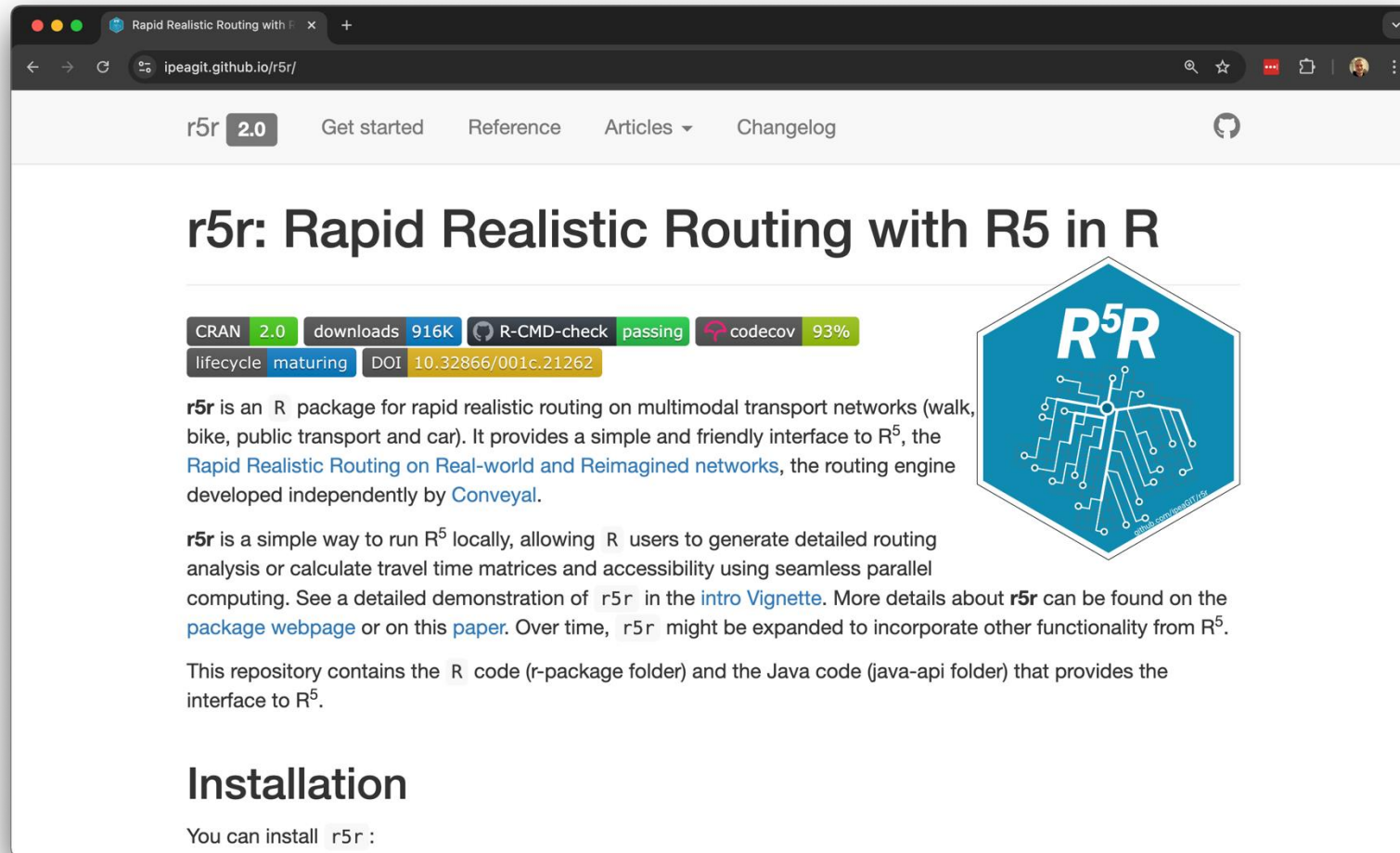
Accessibility analysis



Accessibility analysis



Accessibility analysis

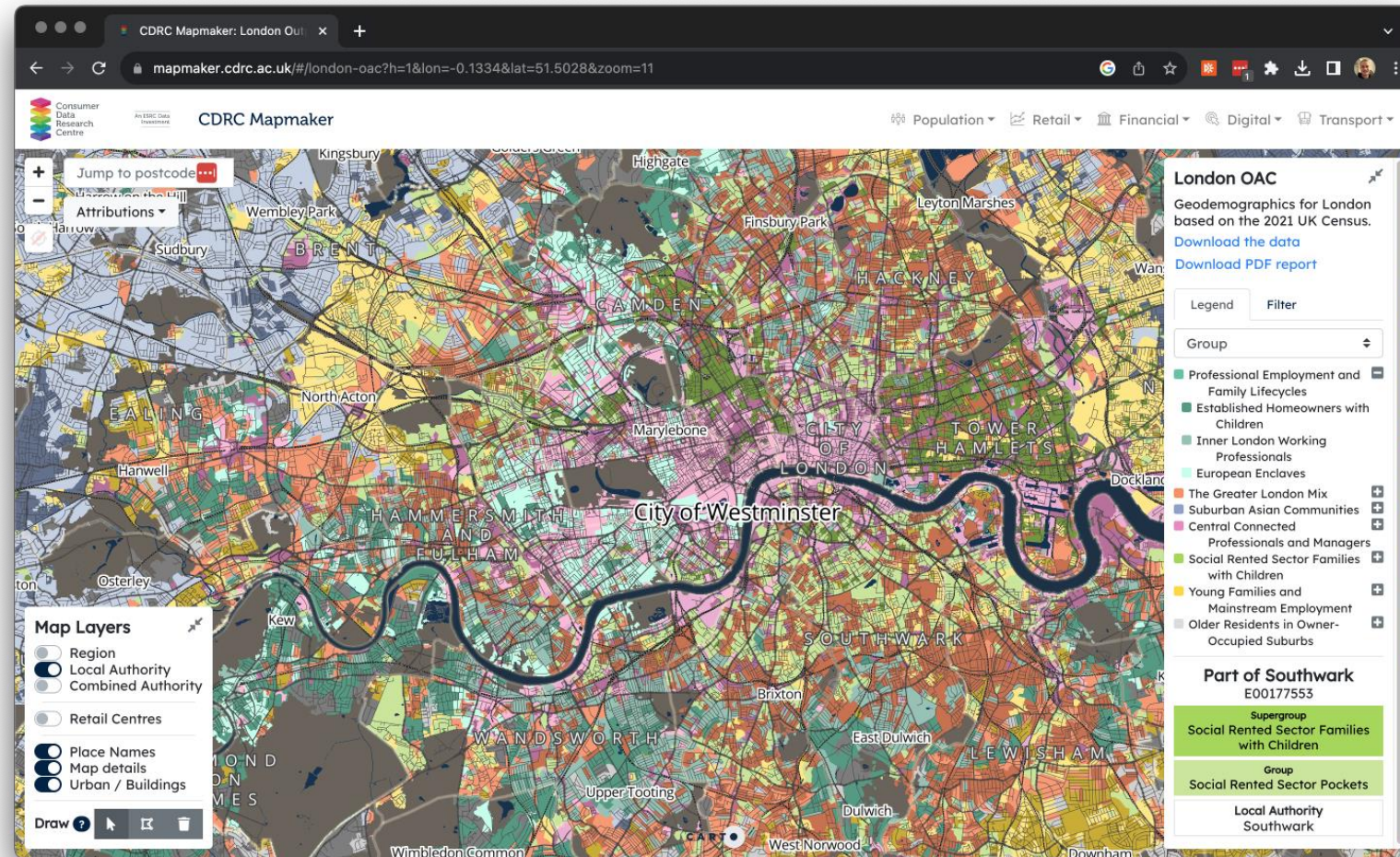


The screenshot shows the GitHub repository page for **r5r** (Rapid Realistic Routing with R5 in R). The page header includes the repository name **r5r 2.0** and navigation links: [Get started](#), [Reference](#), [Articles](#), and [Changelog](#). The main heading is **r5r: Rapid Realistic Routing with R5 in R**. Below the heading, a row of badges displays project statistics: CRAN 2.0, downloads 916K, R-CMD-check passing, codecov 93%, lifecycle maturing, and DOI 10.32866/001c.21262. To the right of the text is a blue hexagonal logo with the text **R⁵R** and a circuit-like pattern. The text describes **r5r** as an R package for rapid realistic routing on multimodal transport networks (walk, bike, public transport and car). It provides a simple and friendly interface to R⁵, the [Rapid Realistic Routing on Real-world and Reimagined networks](#), the routing engine developed independently by [Conveyal](#). It also mentions that **r5r** is a simple way to run R⁵ locally, allowing R users to generate detailed routing analysis or calculate travel time matrices and accessibility using seamless parallel computing. See a detailed demonstration of **r5r** in the [intro Vignette](#). More details about **r5r** can be found on the [package webpage](#) or on this [paper](#). Over time, **r5r** might be expanded to incorporate other functionality from R⁵. This repository contains the R code (r-package folder) and the Java code (java-api folder) that provides the interface to R⁵.

Installation

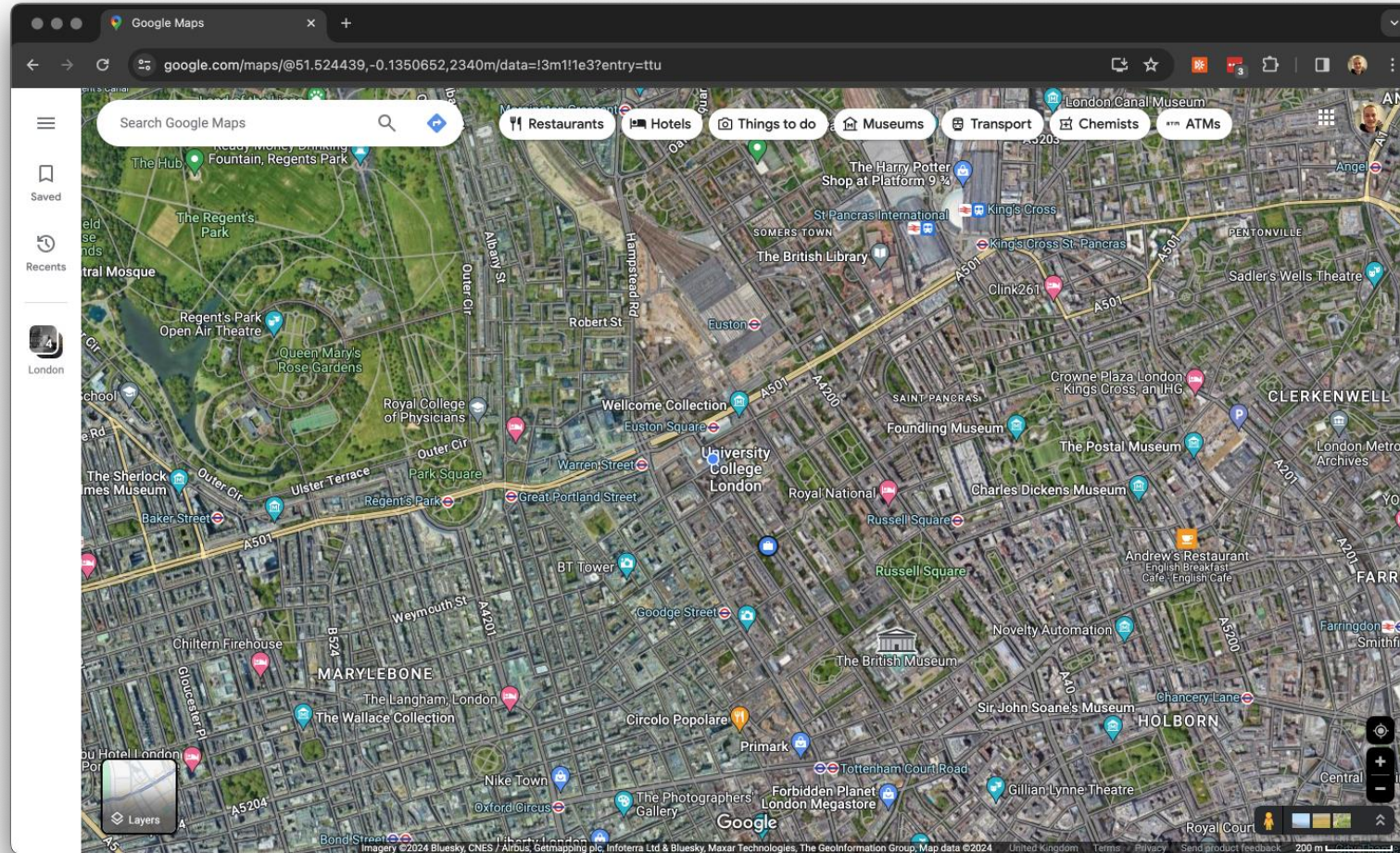
You can install **r5r** :

Geodemographic classification

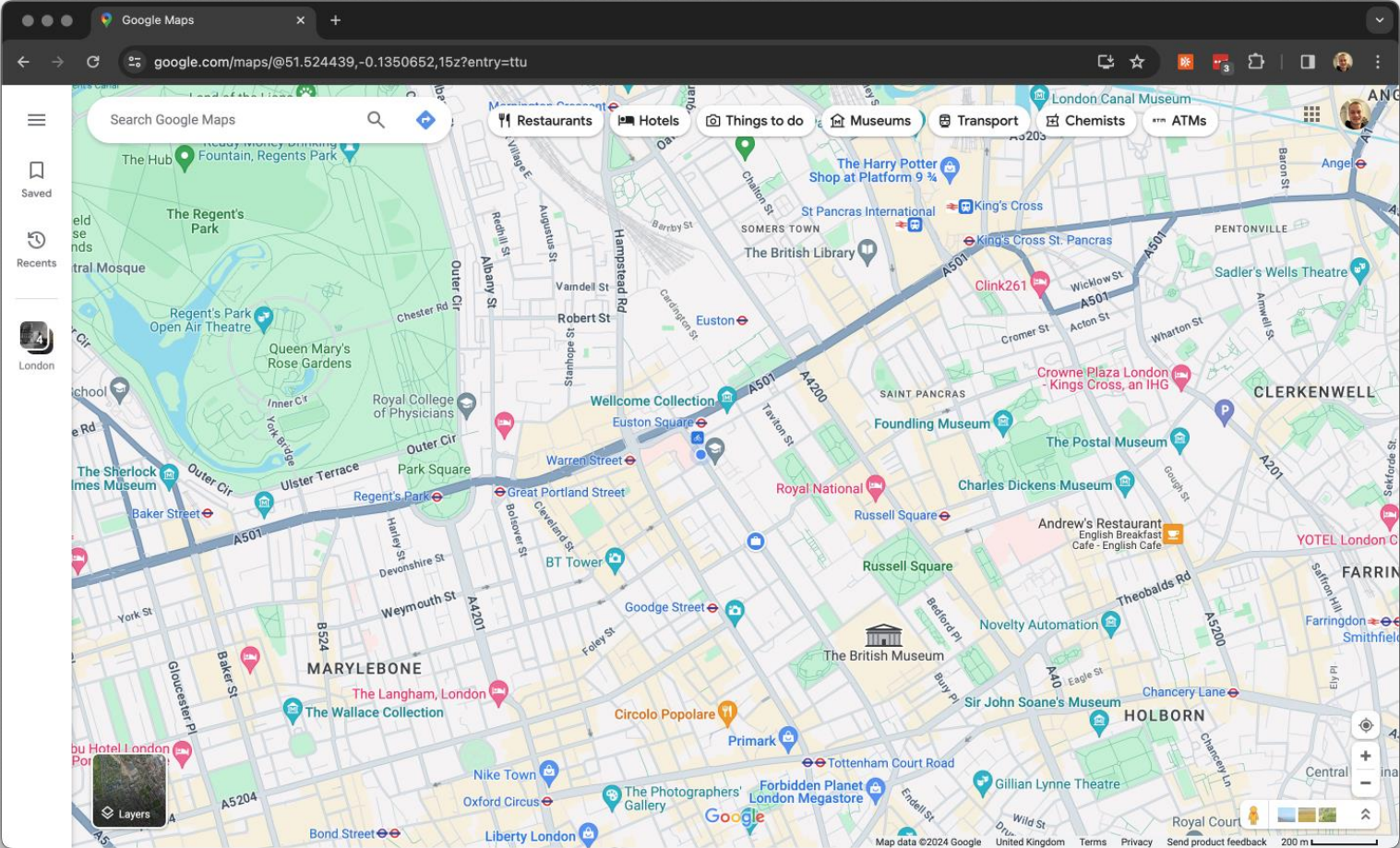


2021 London Output Area Classification on mapmaker.cdrc.ac.uk

GI Science



GIScience



GIScience

- GIScience relies on representing spatial information in a digital format. Traditionally, geographic information is conveyed in two primary ways:.

Vector This method uses a finite set of discrete geometric objects, such as points, lines, and polygons, to represent spatial features.

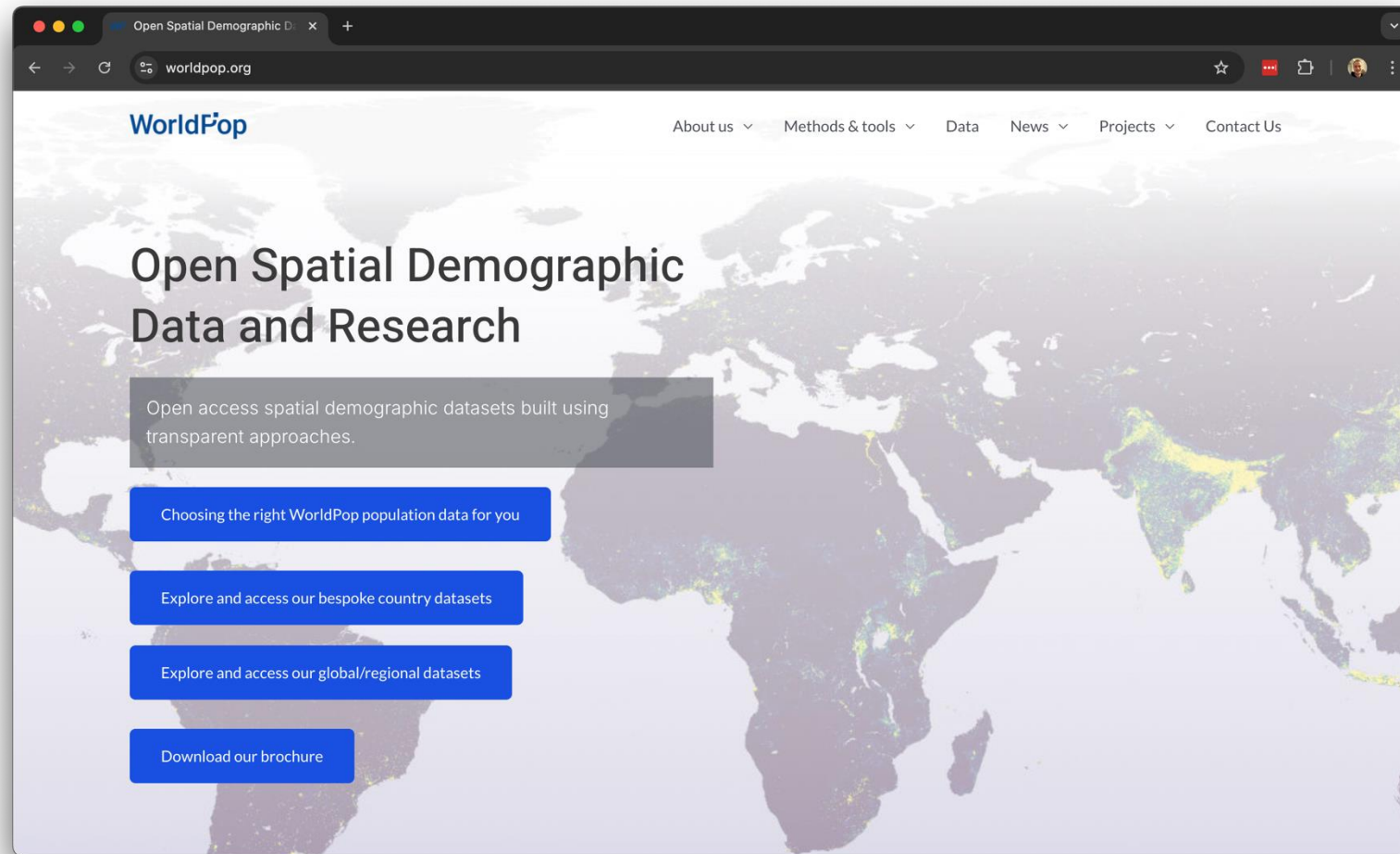
Raster This approach employs images or grids to represent surfaces, with each cell or pixel holding a value, often indicating attributes like colors or measurements.



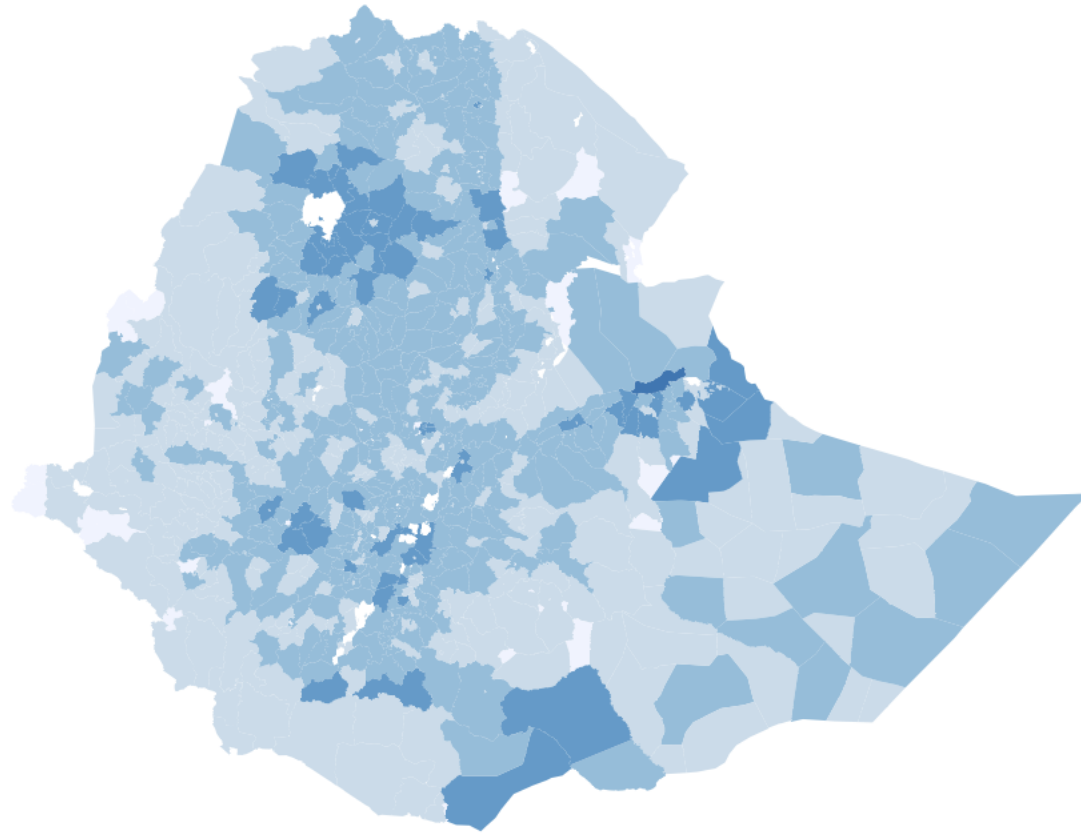




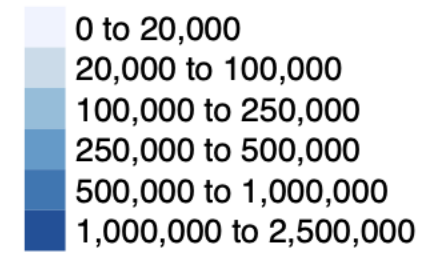
Dasymetric mapping



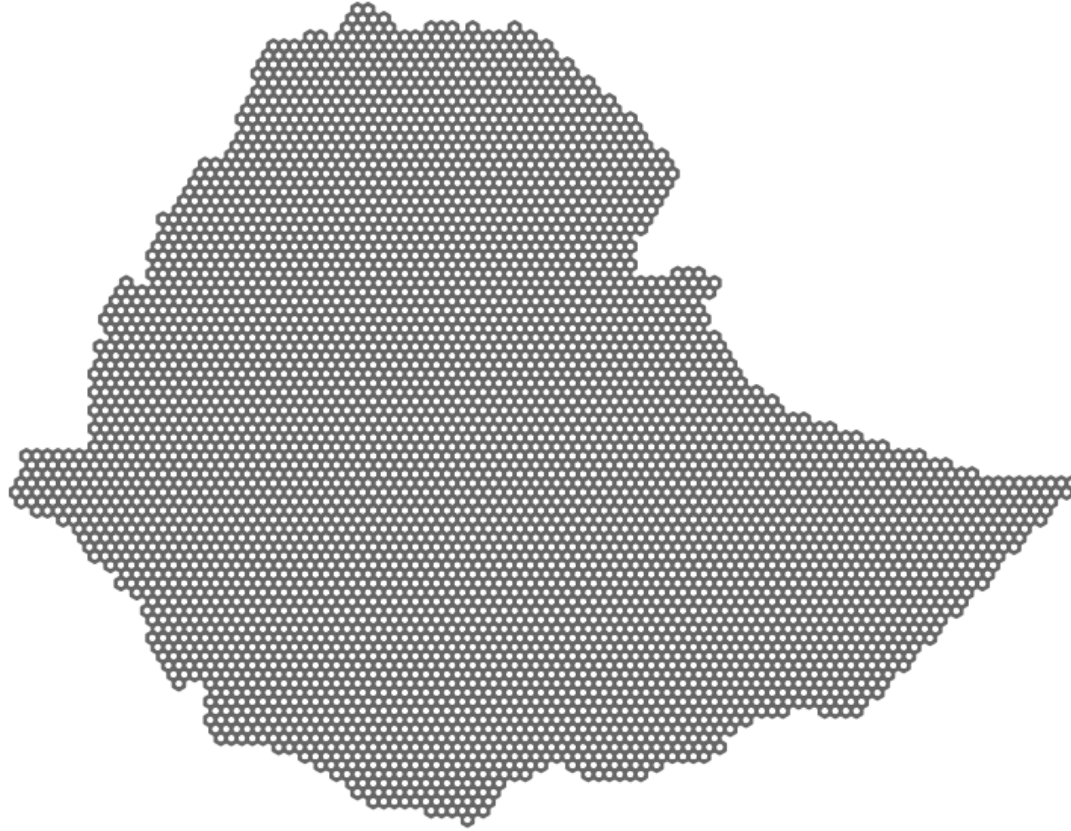
Raster data



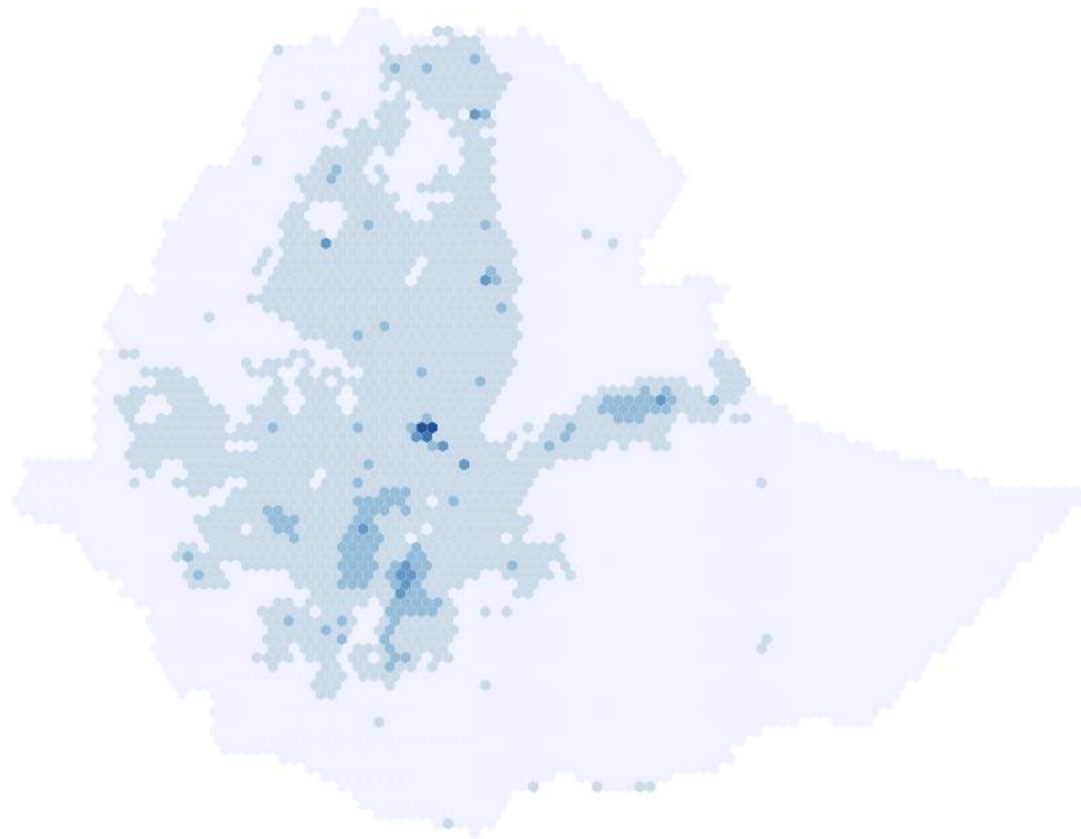
Population count



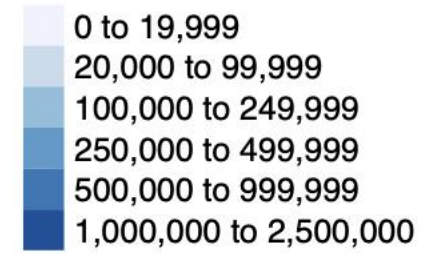
Raster data



Raster data



Population count



Questions

Justin van Dijk
j.t.vandijk@ucl.ac.uk

