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 assumption: 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:

- Plot 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:

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

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:

$$y = \rho W y + X\beta + \epsilon$$

where ρ is the spatial autoregressive coefficient, Wy represents the spatially lagged dependent variable, and $X\beta$ 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?

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

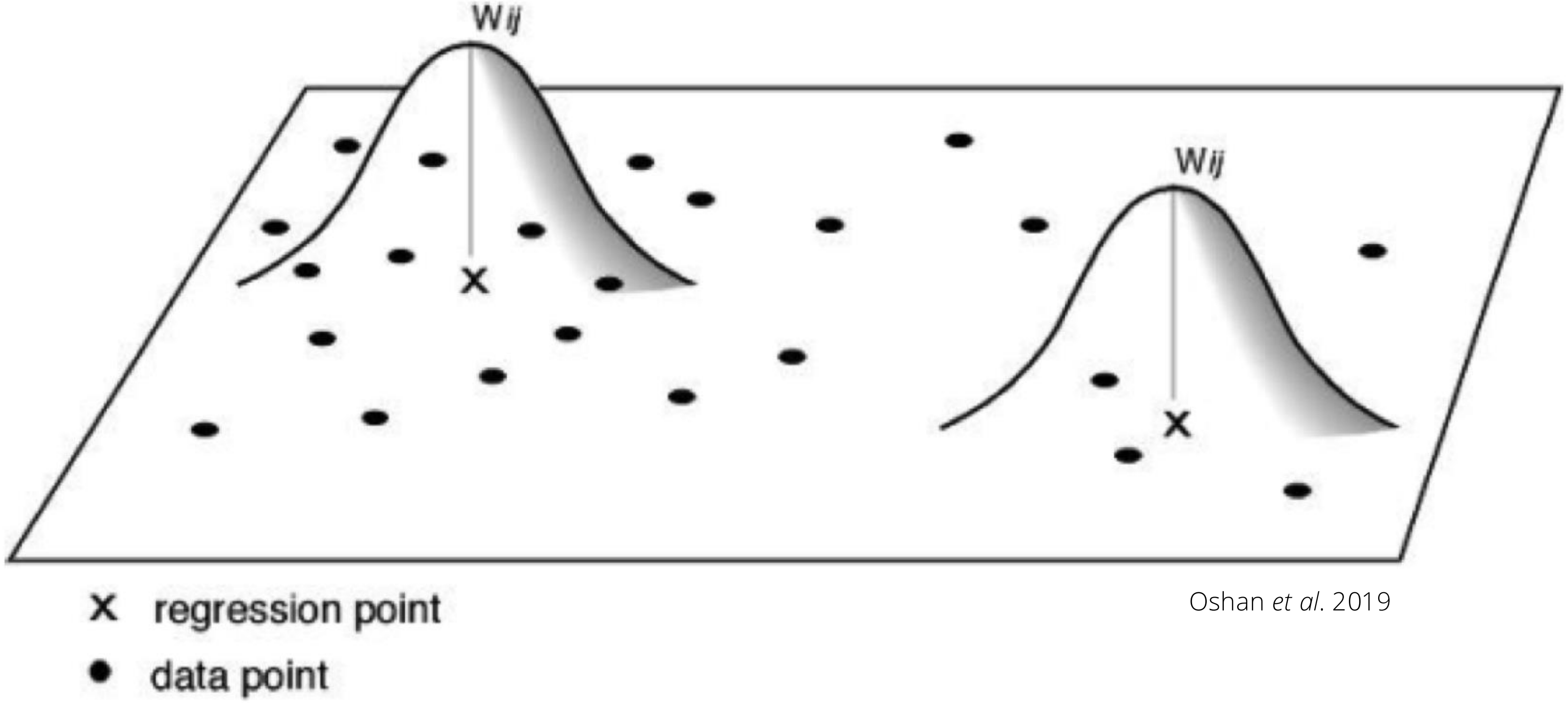
- Unlike traditional global models, which estimate a single set of parameters for the entire study area, geographically weighted statistics allow for parameter estimates that vary across different locations.
- Local means, local standard deviations, local variances.
- Typically uses some kernel function 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."

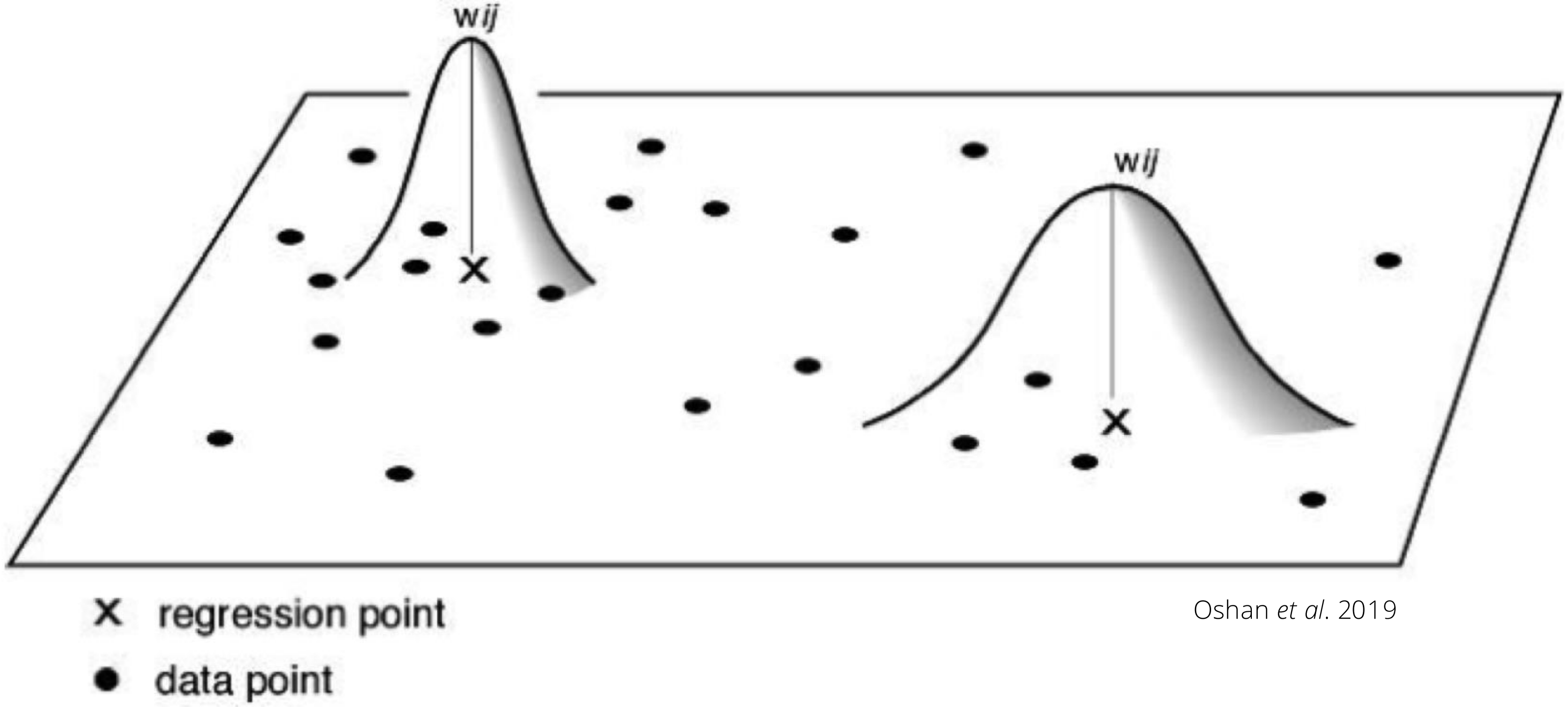
Walter Tobler 1970

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.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk

SA-TIED Geospatial Analysis Workshop

More spatial analysis

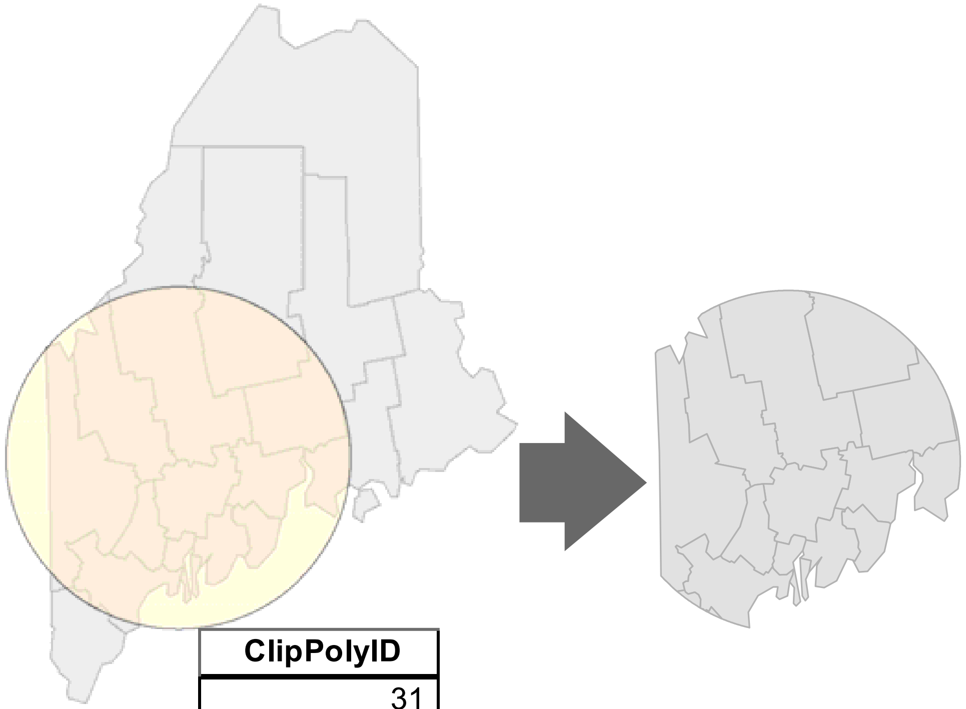
UCL

More spatial analysis

- Spatial operations.
- Accessibility analysis.
- Geodemographic classification.
- Raster data.

Spatial operations

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



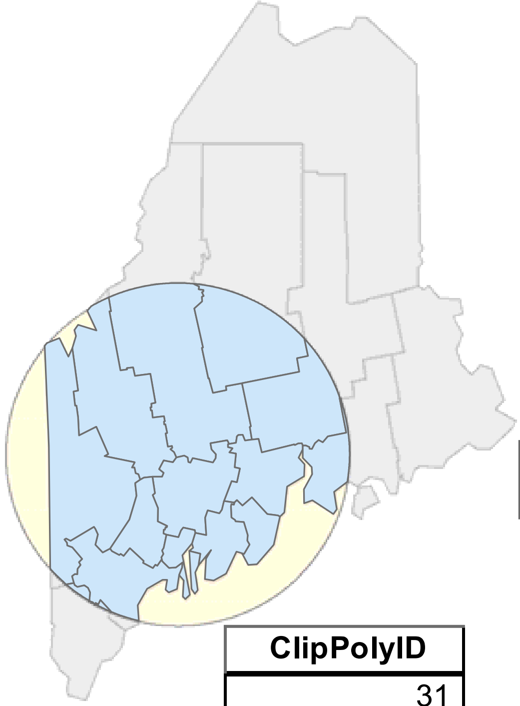
ClipPolyID
31

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

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

Spatial operations

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



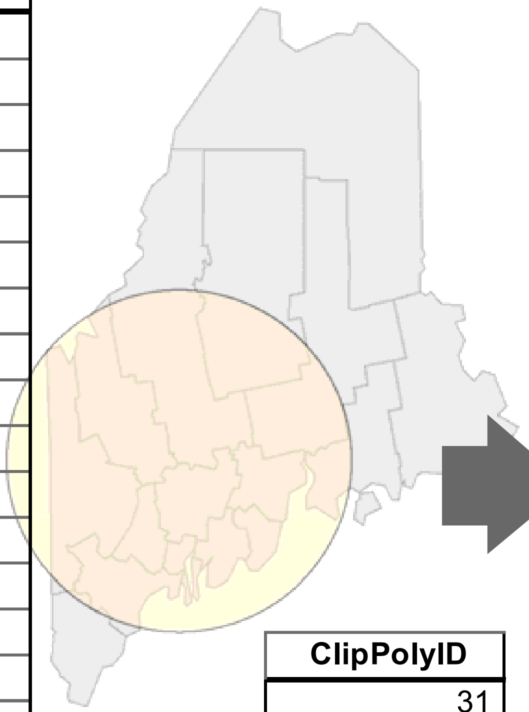
ClipPolyID
31

NAME	ClipPolyID
Androscoggin	31
Cumberland	31
Franklin	31
Hancock	31
Kennebec	31
Knox	31
Lincoln	31
Oxford	31
Penobscot	31
Piscataquis	31
Sagadahoc	31
Somerset	31
Waldo	31
York	31

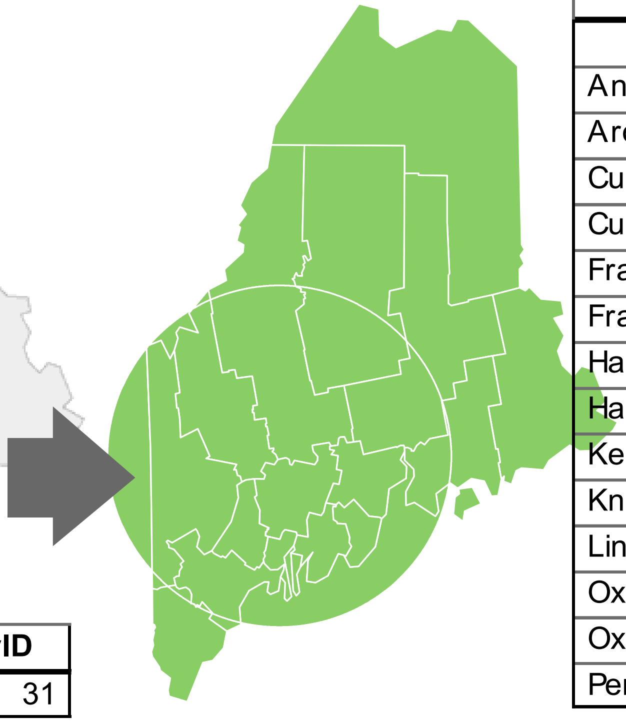
Gimdond, M. 2021. Intro to GIS and Spatial Analysis. [online]
https://mgimond.github.io/Spatial/introGIS.html

Spatial operations

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



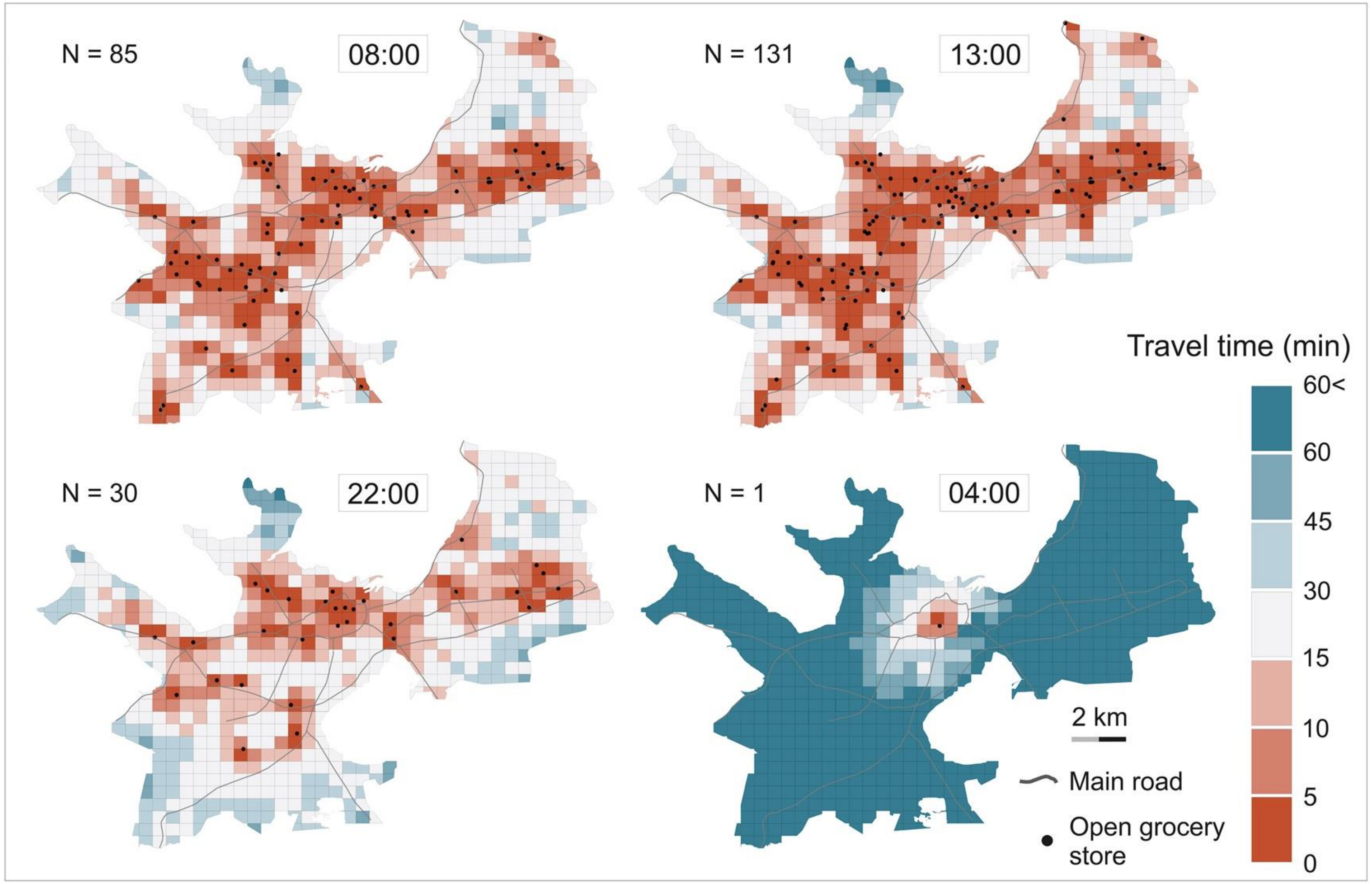
ClipPolyID
31



NAME	ClipPolyID
	31
Androscoggin	31
Aroostook	0
Cumberland	0
Cumberland	31
Franklin	0
Franklin	31
Hancock	0
Hancock	31
Kennebec	31
Knox	31
Lincoln	31
Oxford	0
Oxford	31
Penobscot	0

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

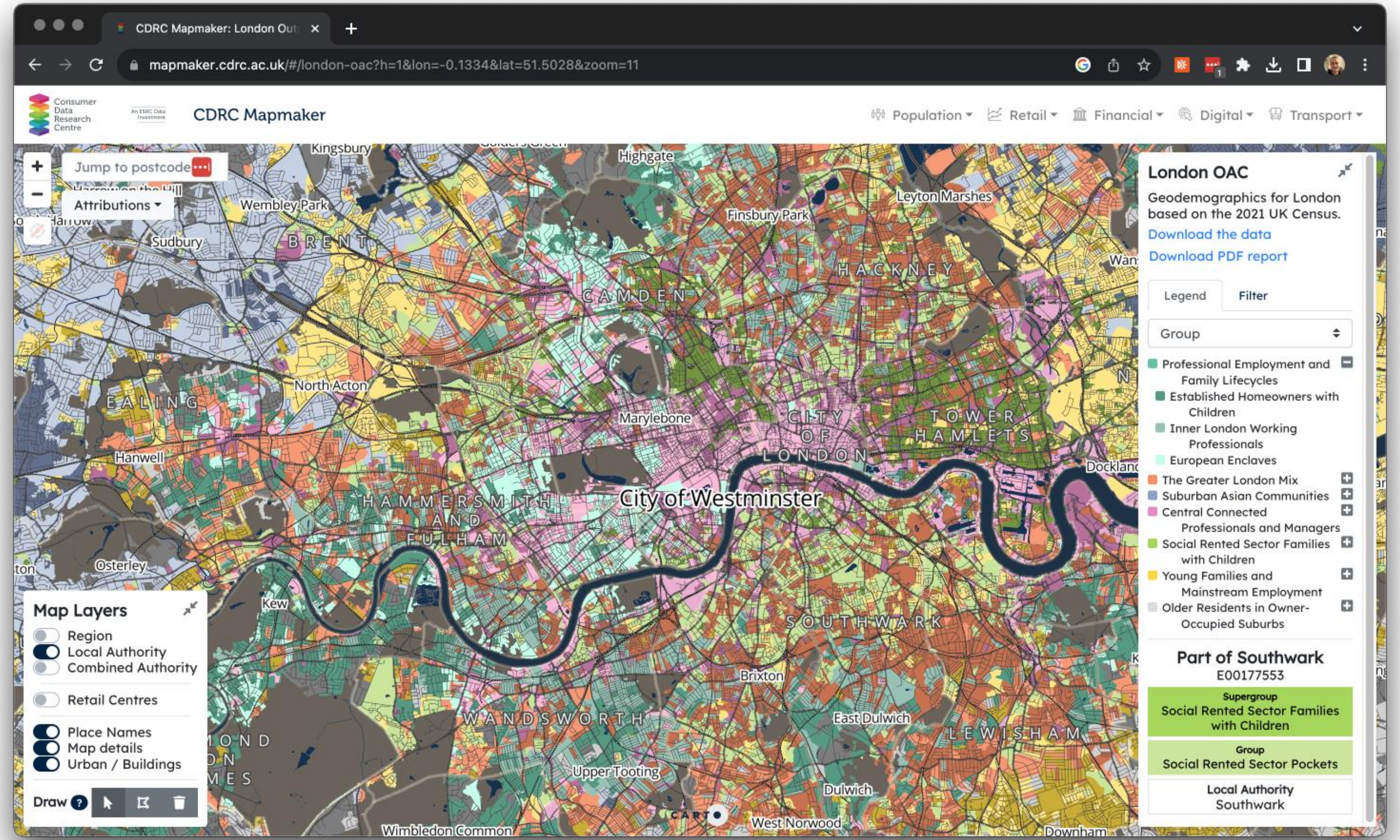
Accessibility analysis



Järv *et al.* 2018

Accessibility analysis

Geodemographic classification



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


8 9 9 10 0 10 10 10 0 0 0 7 5 3 0 0 0 0 1
8 9 9 10 10 0 10 9 9 0 0 5 3 0 0 0 0 0 0
8 8 9 9 10 0 0 9 8 7 5 0 0 0 1 0 0 0 0
5 8 8 9 10 10 0 9 7 5 0 0 5 5 5 0 0 0 1
3 5 8 9 9 10 0 0 3 0 0 0 5 0 0 1 0 0 2
2 5 8 8 9 9 10 0 0 0 1 5 0 0 0 0 0 0 1
2 4 6 8 8 9 0 0 0 1 5 0 0 5 5 5 0 0 1
0 3 6 8 8 0 0 0 0 5 0 5 5 5 5 5 0 0 0
2 2 5 8 0 0 0 0 0 0 5 5 5 5 5 5 3 0 0
0 2 5 0 0 1 2 3 4 4 4 4 4 4 4 5 0 0 0
0 0 0 0 1 1 1 1 4 4 4 4 4 4 4 5 0 0 0
0 0 1 1 2 2 2 2 3 3 3 3 3 3 3 4 0 3 0
1 1 1 1 2 2 3 3 3 3 1 1 1 1 1 2 3 4 3

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk