Geographic Data Science

Space, formally

Carmen Cabrera-Arnau and Elisabetta Pietrostefani

Space, formally

For a statistical method to be **explicitly spatial**, it needs to contain some representation of the geography, or **spatial context**

One of the most common ways is through **Spatial Weights Matrices**

- **(Geo)Visualization**: translating numbers into a (visual) language that the human brain *"speaks better"*
- **Spatial Weights Matrices**: translating geography into a (numerical) language that a computer *"speaks better"*.

Core element in several spatial analysis techniques:

- Spatial autocorrelation
- Spatial clustering / geodemographics
- Spatial regression

W as a formal representation of space

W

N x N positive matrix that contains ***spatial relations*** *between all the observations in the sample*

$w_{ii} = 0$ by convention

…What is a ***neighbor****???*

Geographic Data Science

Types of *W*

What is a neighbor?

A neighbor is "somebody" who is:

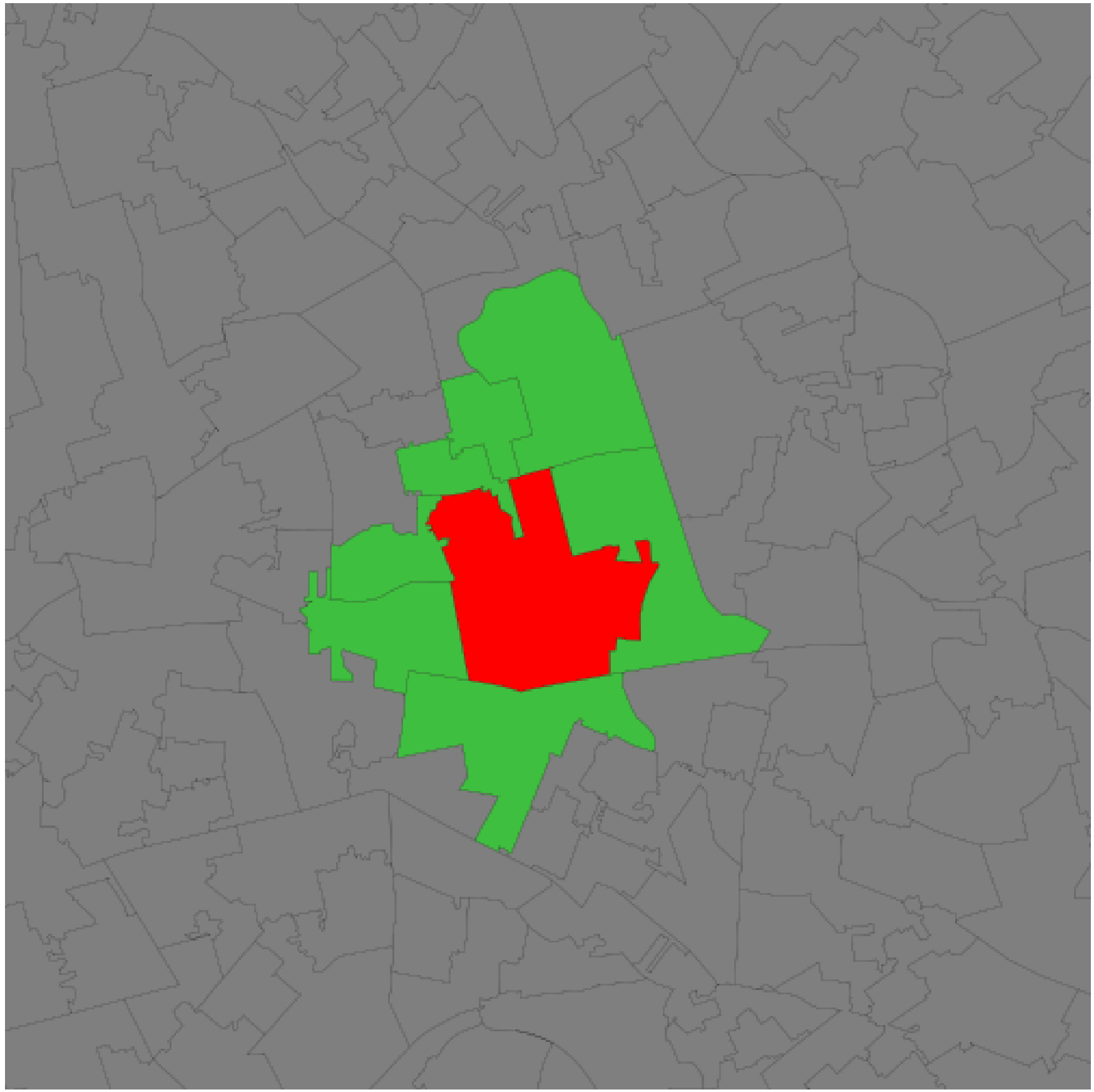
- Next door → **Contiguity**-based *W*s
- Close → **Distance**-based *W*s
- In the same "place" as us → **Block** weights
- …

See Anselin & Rey (2014) for an in-detail discussion and more types of *W*.

Contiguity-based weights

Sharing boundaries to any extent

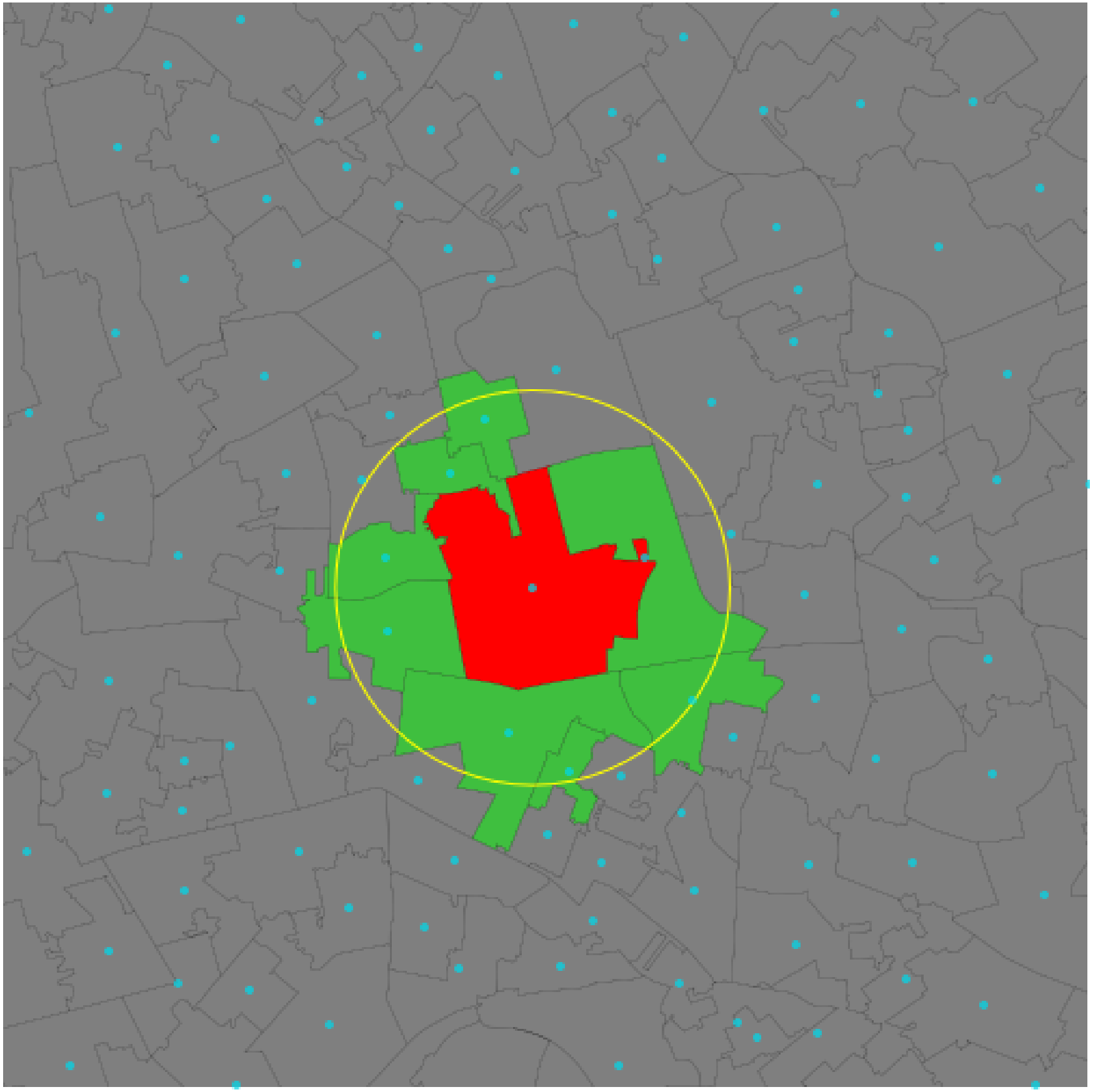
- Rook
- Queen
- …



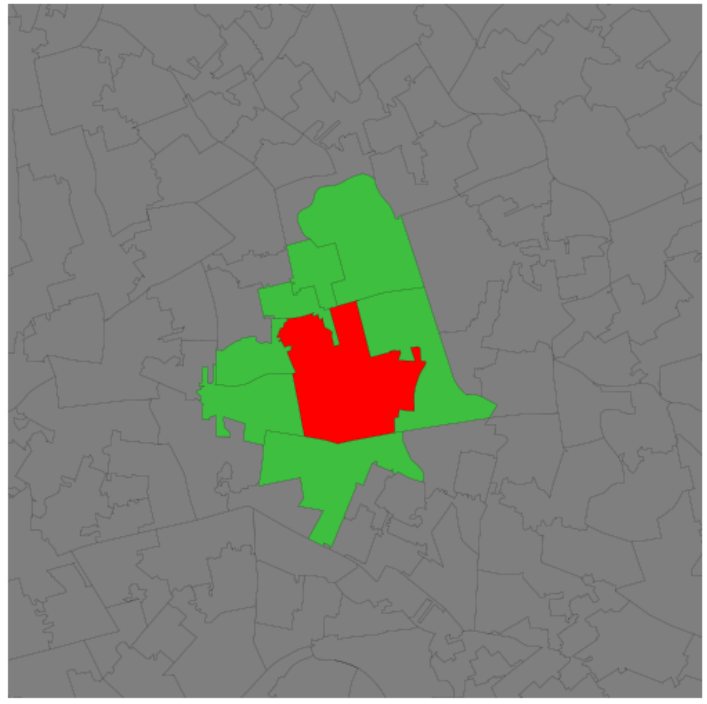
Distance-based weights

Weight is (inversely) proportional to distance between observations

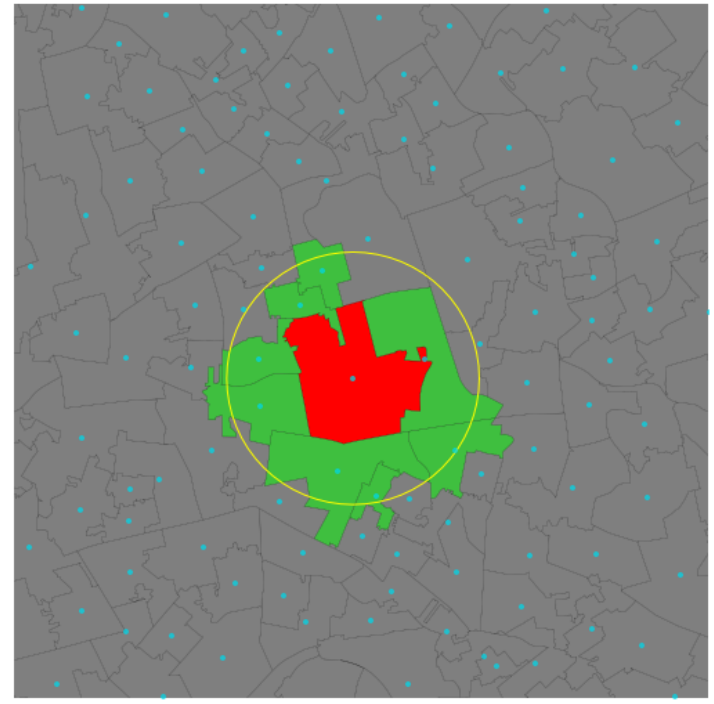
- Inverse distance (threshold)
- KNN (fixed number of neighbors)
- …



Queen neighbors of `E01006690`



Neighbors within 1km of `E01006690`

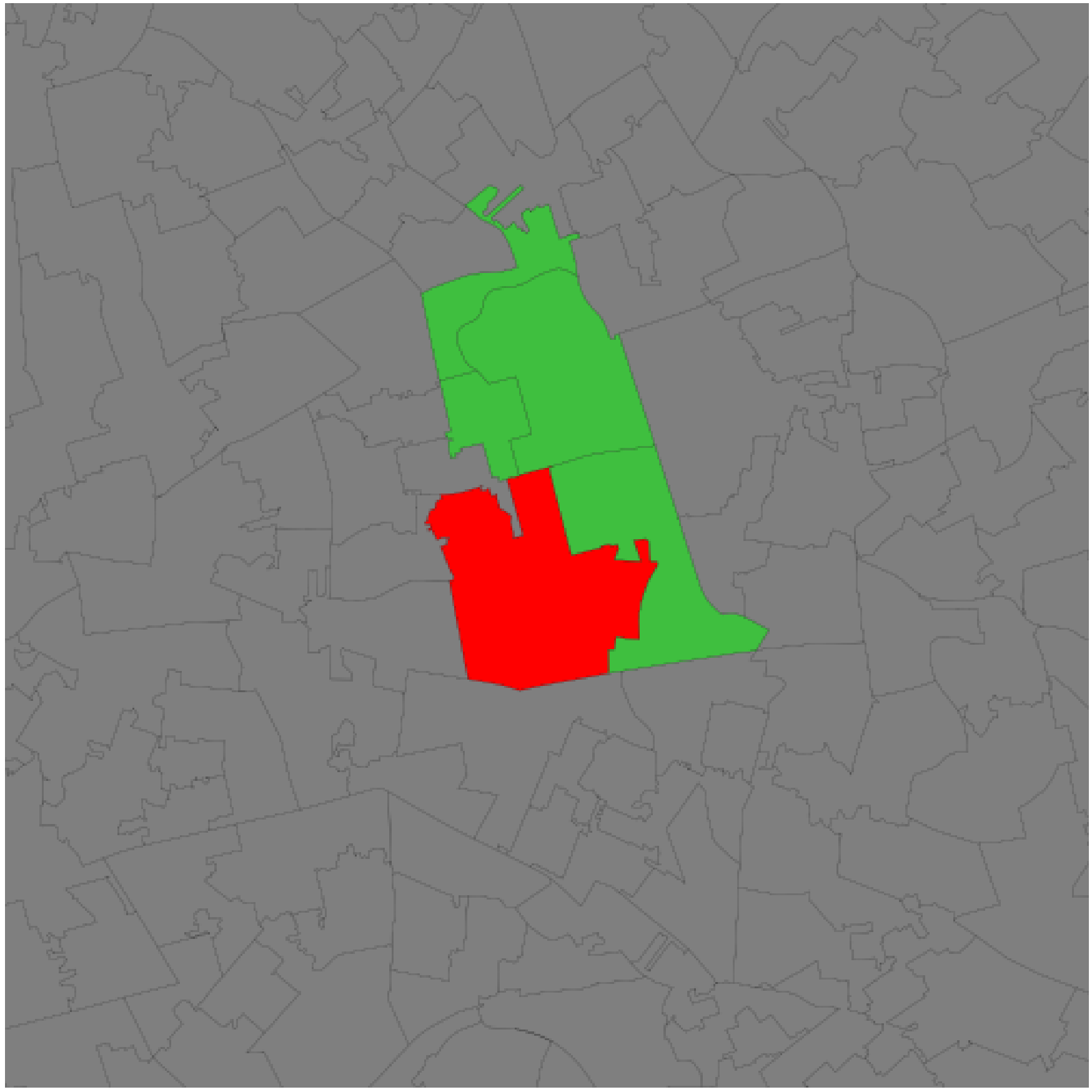


Block weights

Weights are assigned based on discretionary rules loosely related to geography

For example:

- LSOAs into MSOAs
- Post-codes within city boundaries
- Counties within states
- …



How much of a neighbor?

No neighbors receive zero weight: $w_{ij} = 0$

Neighbors, it depends, w_{ij} can be:

Choice of W

Should be based on and reflect the **underlying channels of interaction** for the question at hand.

Examples:

- Processes propagated by inmediate contact (e.g. disease contagion) → Contiguity weights
- Accessibility → Distance weights
- Effects of county differences in laws → Block weights

Standardization

In some applications (e.g. spatial autocorrelation) it is common to *standardize W*

The most widely used standardization is **row-based**: divide every element by the sum of the row:

where $w_{i.}$ is the sum of a row.

Geographic Data Science

The spatial lag

The spatial lag

Product of a spatial weights matrix W and a given variably Y

$$Y_{sl} = WY\,\gamma_{sl-i} = \sum_j w_{ij}\gamma_j$$

- Measure that captures the behaviour of a variable in the neighborhood of a given observation *i*.
- If *W* is **standardized**, the spatial lag is the *average value of the variable in the neighborhood*

- Common way to introduce space formally in a statistical framework
- Heavily used in both ESDA and spatial regression to delineate neighborhoods. Examples:
  - Moran's I
  - LISAs
  - Spatial models (lag, error…)