Geographic Data Science - Lecture V

Space, formally

Dani Arribas-Bel

Today

- The need to represent space formally
- Spatial weights matrices
  - What
  - Why
  - Types
- The spatial lag
- The Moran Plot

Space, formally

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

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

W

*N x N positive matrix that contains **spatial relations***

W

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

$$W$$

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

$$w_{ij} = \begin{cases} x > 0 & \text{if i and j are neighbors} \\ 0 & \text{otherwise} \end{cases}$$

$w_{ii} = 0$ by convention

W

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

$$w_{ij} = \begin{cases} x > 0 & \text{if i and j are neighbors} \\ 0 & \text{otherwise} \end{cases}$$

w_{ii} = 0 by convention

...What is a ***neighbor****???*

Types of W

A neighbor is "somebody" who is:

- Next door
- Close
- In the same "place" as us
- ...

Types of W

A neighbor is "somebody" who is:

- Next door $\rightarrow$ **Contiguity**-based Ws
- Close
- In the same "place" as us
- ...

Types of W

A neighbor is "somebody" who is:

- Next door $\rightarrow$ **Contiguity**-based Ws
- Close $\rightarrow$ **Distance**-based Ws
- In the same "place" as us
- ...

Types of *W*

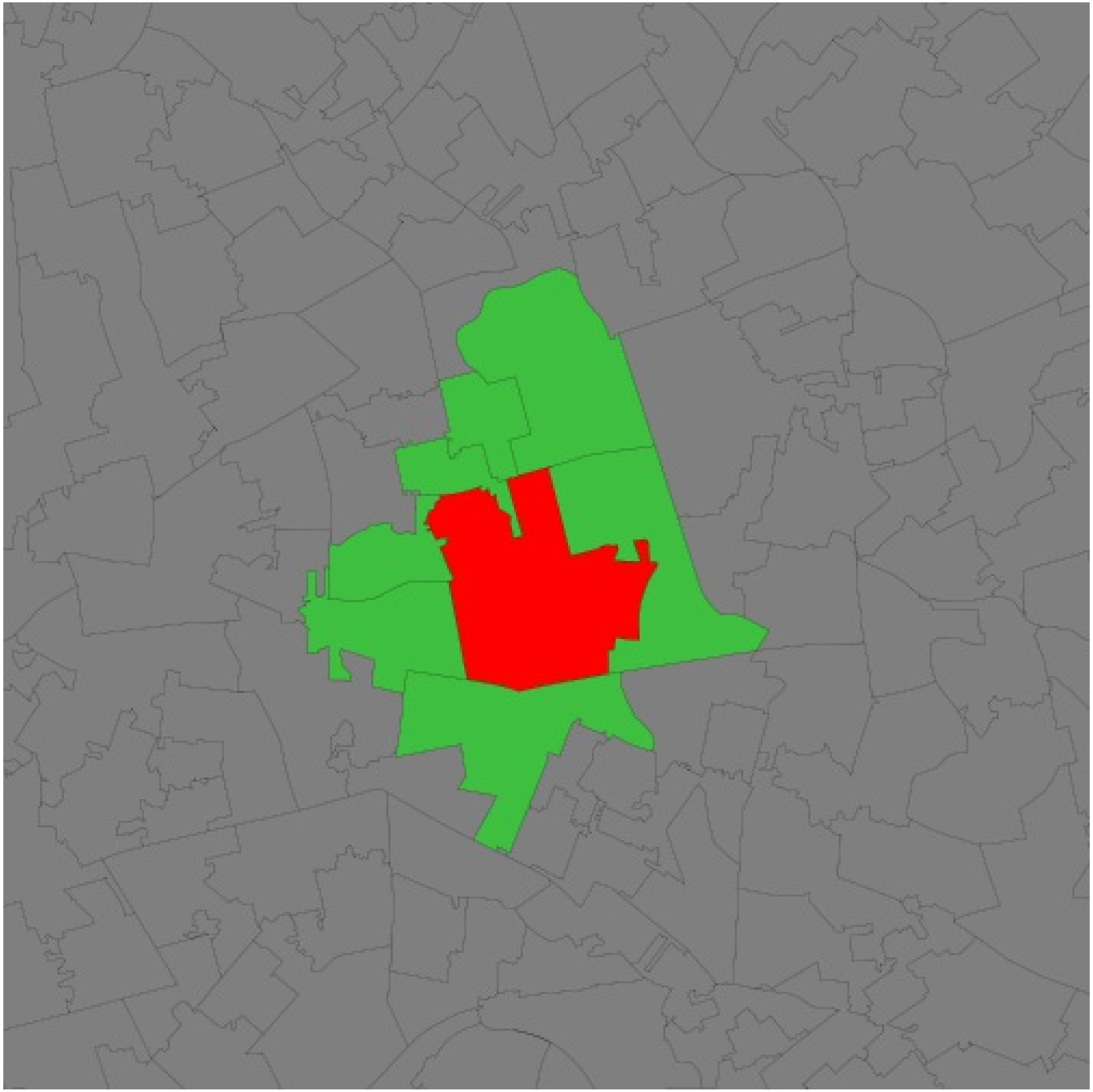
A neighbor is "somebody" who is:

- Next door $\rightarrow$ **Contiguity**-based *W*s
- Close $\rightarrow$ **Distance**-based *W*s
- In the same "place" as us $\rightarrow$ **Block** weights
- ...

Contiguity-based weights

Sharing boundaries to any extent

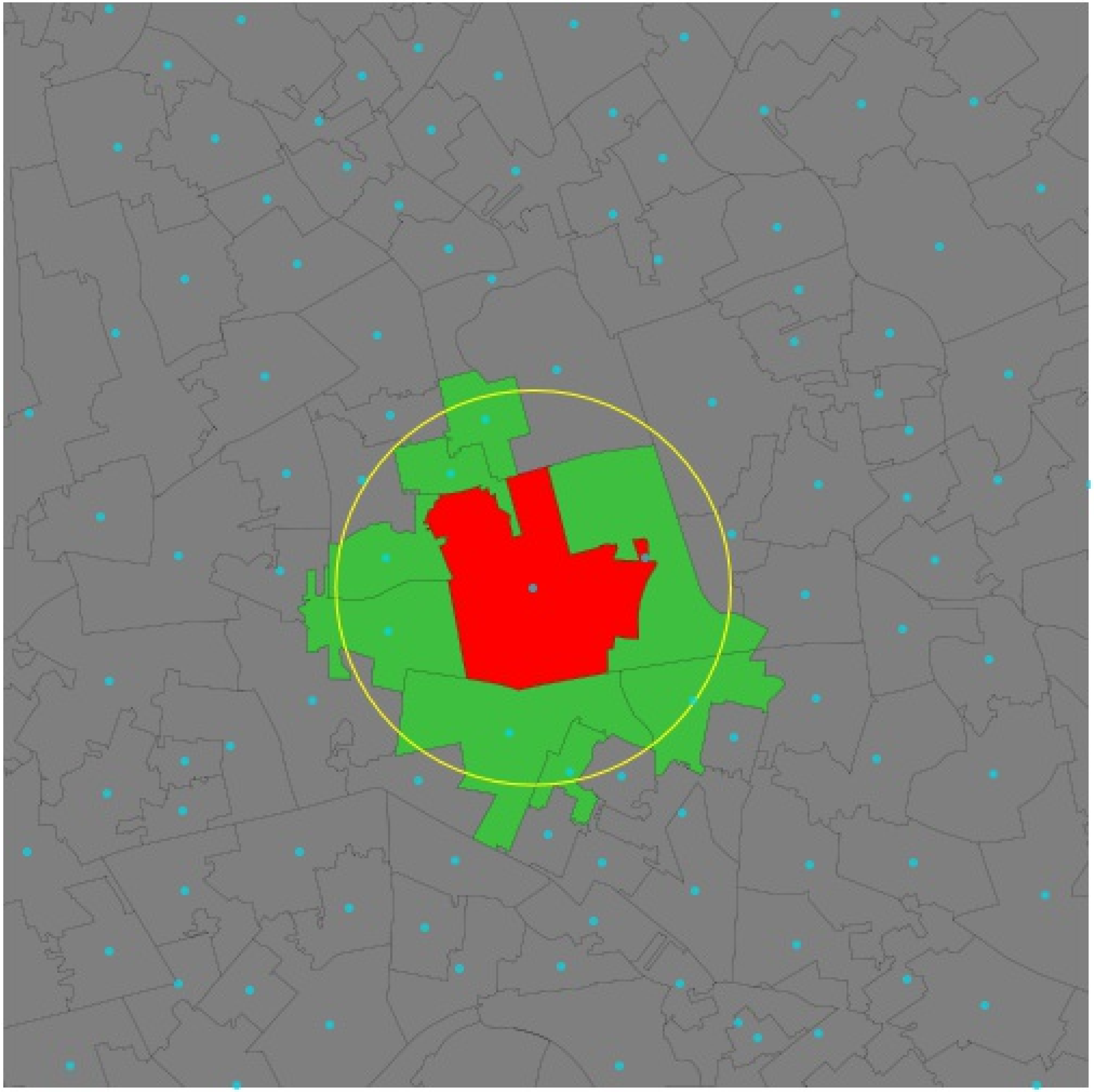
- Rook
- Queen
- ...

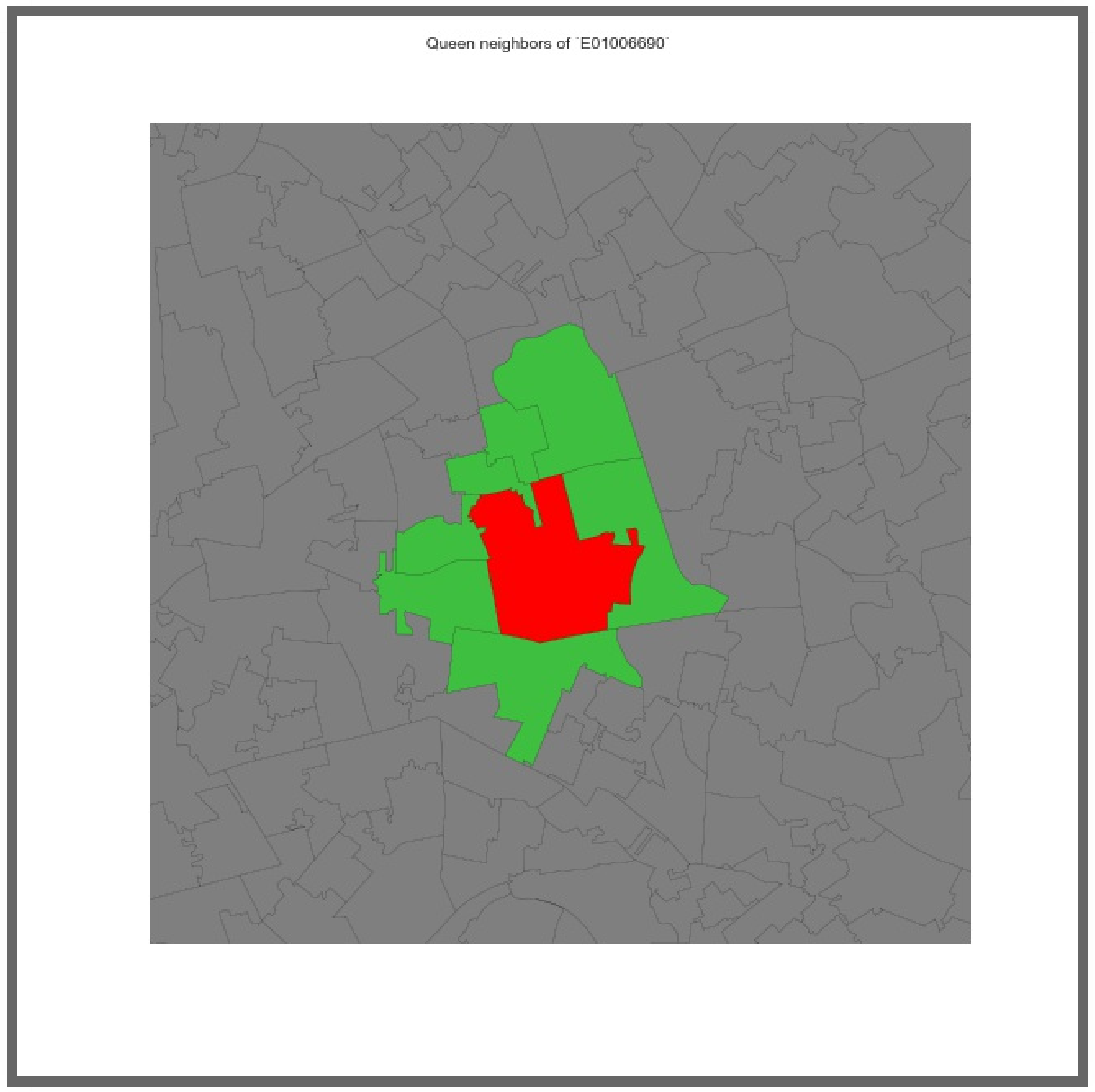
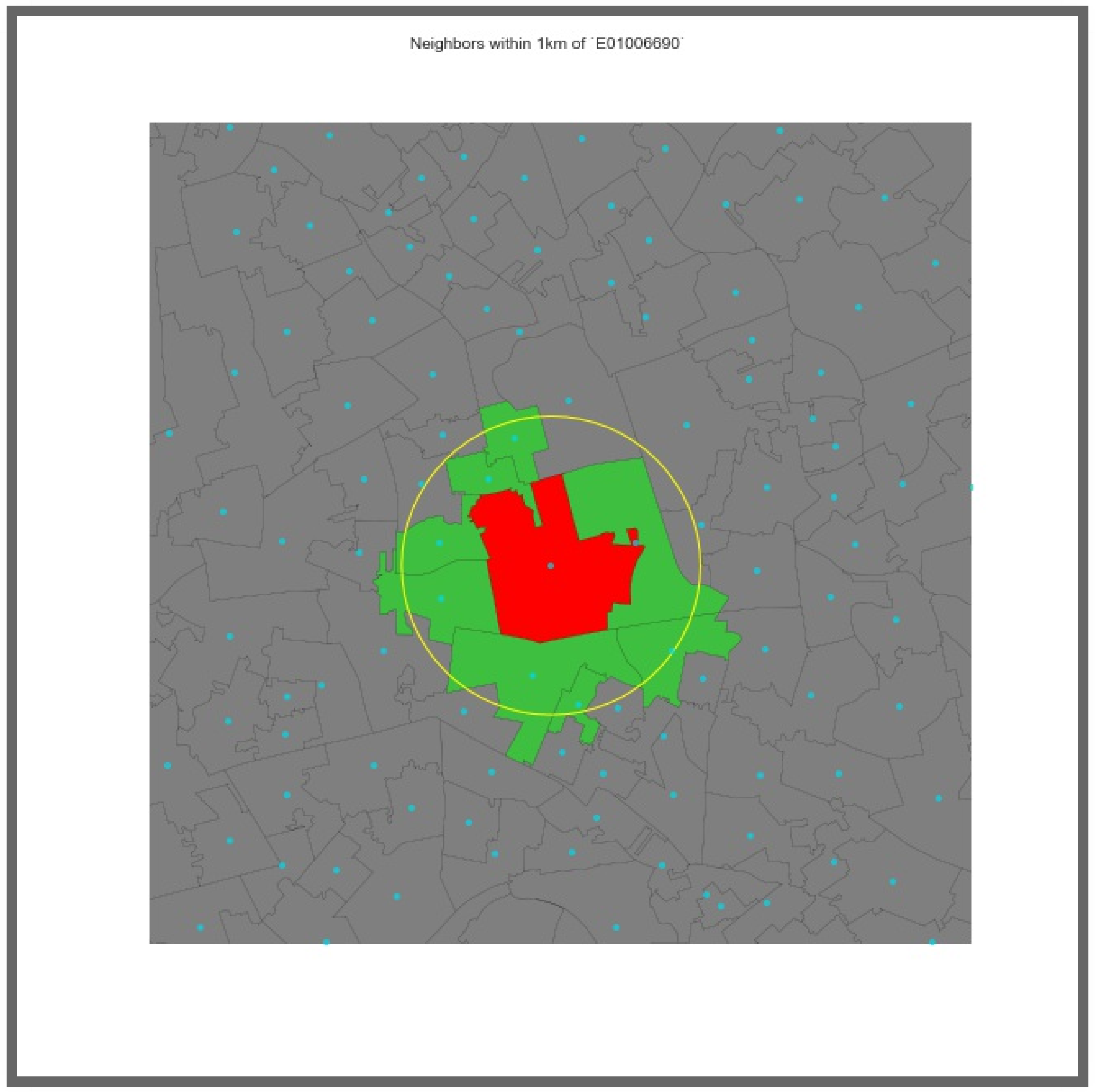


Distance-based weights

Weight is (inversely) proportional to distance between observations

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



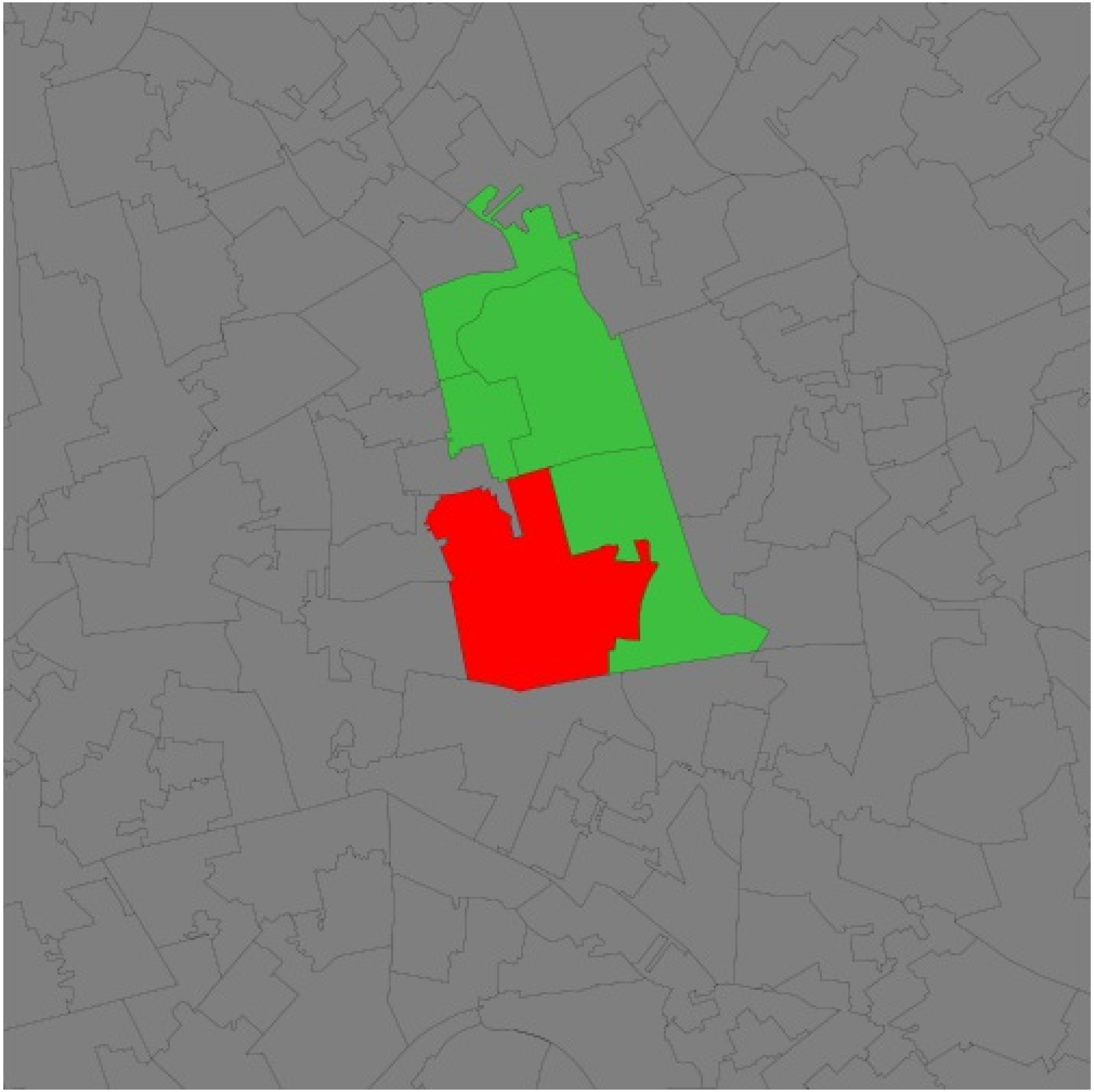

Queen neighbors of `E01006690`

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
- ...



Other types of weights

- Combinations of the above
- Kernel
- Statistically-derived
- ...

See Anselin & Rey (2014) for an in-detail discussion.

How much of a neighbor?

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

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

- One $w_{ij} = 1 \rightarrow$ **Binary**
- Some proportion ($0 < w_{ij} < 1$, **continuous**) which can be a function of:
    - Distance
    - Strength of interaction (e.g. commuting flows, trade, etc.)
    - ...

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) $\rightarrow$ Contiguity weights
- Accessibility $\rightarrow$ Distance weights
- Effects of county differences in laws $\rightarrow$ Block weights

Do your own (contiguity) weights time!

1	2	3
4	5	6
7	8	9

1	2	3
4	5	6
7	8	9

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

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:

$$\bar{w}_{ij} = \frac{w_{ij}}{w_{i\cdot}}$$

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

The spatial lag

The spatial lag

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

The spatial lag

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

$$Y_{sl} = WY$$

$$y_{sl-i} = \sum_j w_{ij} y_j$$

- 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*

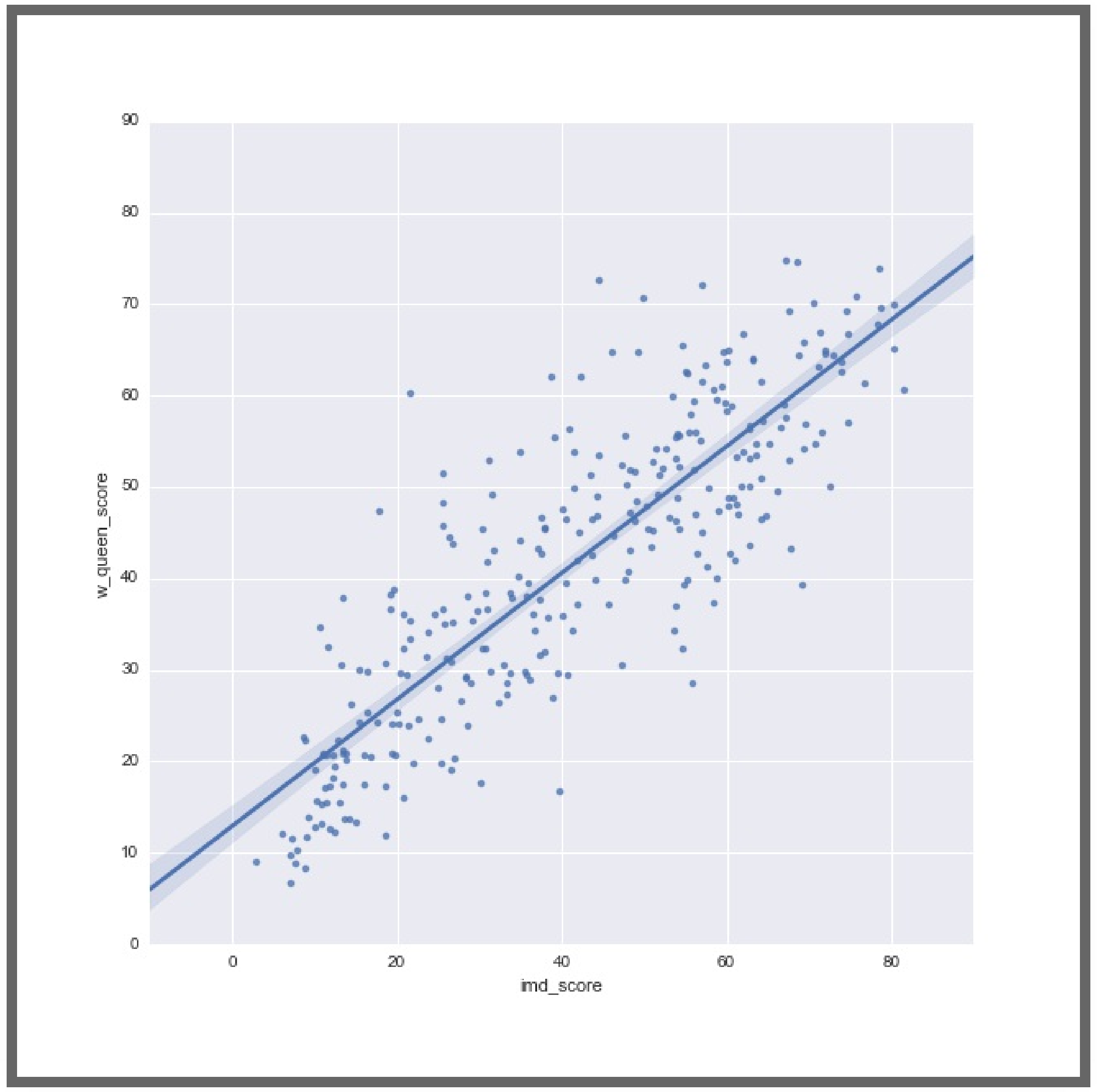
- Common way to introduce space formally in a statistical framework
- Heavily used in both ESDA and spatial regression to delineate neighborhoods. Examples:
    - Moran's I
    - LISAs
    - Spatial models (lag, error…)

Moran Plot

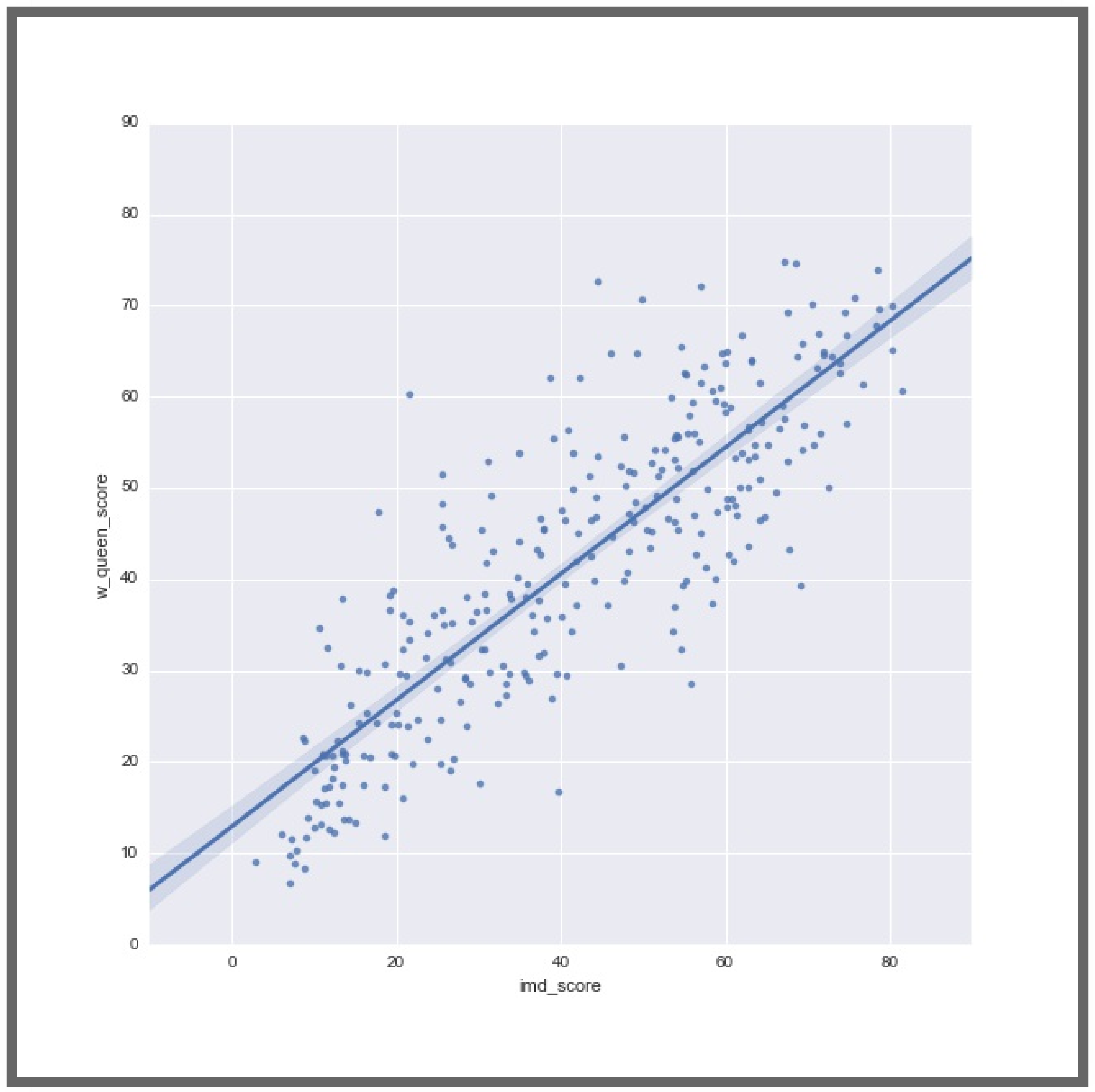
Moran Plot

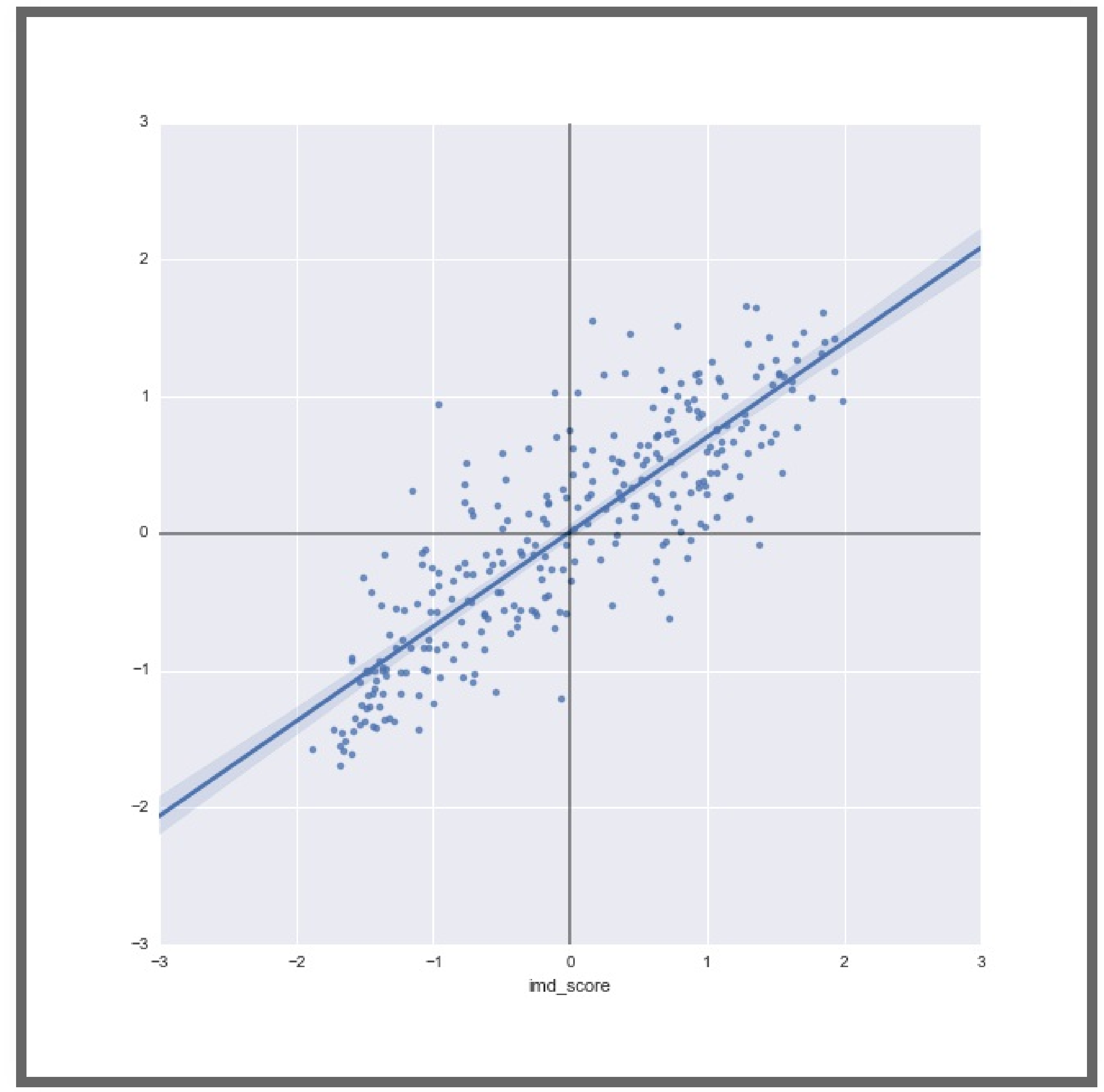
- Graphical device that displays **a variable** on the horizontal axis against **its spatial lag** on the vertical one
- Usually, variables are standardized ($\frac{y - \text{mean}(y)}{\text{std}(y)}$), which divides the space into **quadrants**
- Tool to start exploring **spatial autocorrelation**

Moran Plot



Moran Plot


w_queen_score
imd_score
0
10
20
30
40
50
60
70
80
90


imd_score
−3
−2
−1
0
1
2
3

Recapitulation

- Spatial Weights matrices: matrix encapsulation of space
- Different types for different cases
- Useful in many contexts, like the spatial lag and Moran plot, but also many other things!