Spatial Autocorrelation Part A

First Session

Orlando Sabogal-Cardona
PhD researcher
University College London UCL

First law of geography

"Everything is related to everything else, but near things are more related than distant things"
(Tobler, 1970)

Photo by Conor Luddy on Unsplash

Key concepts

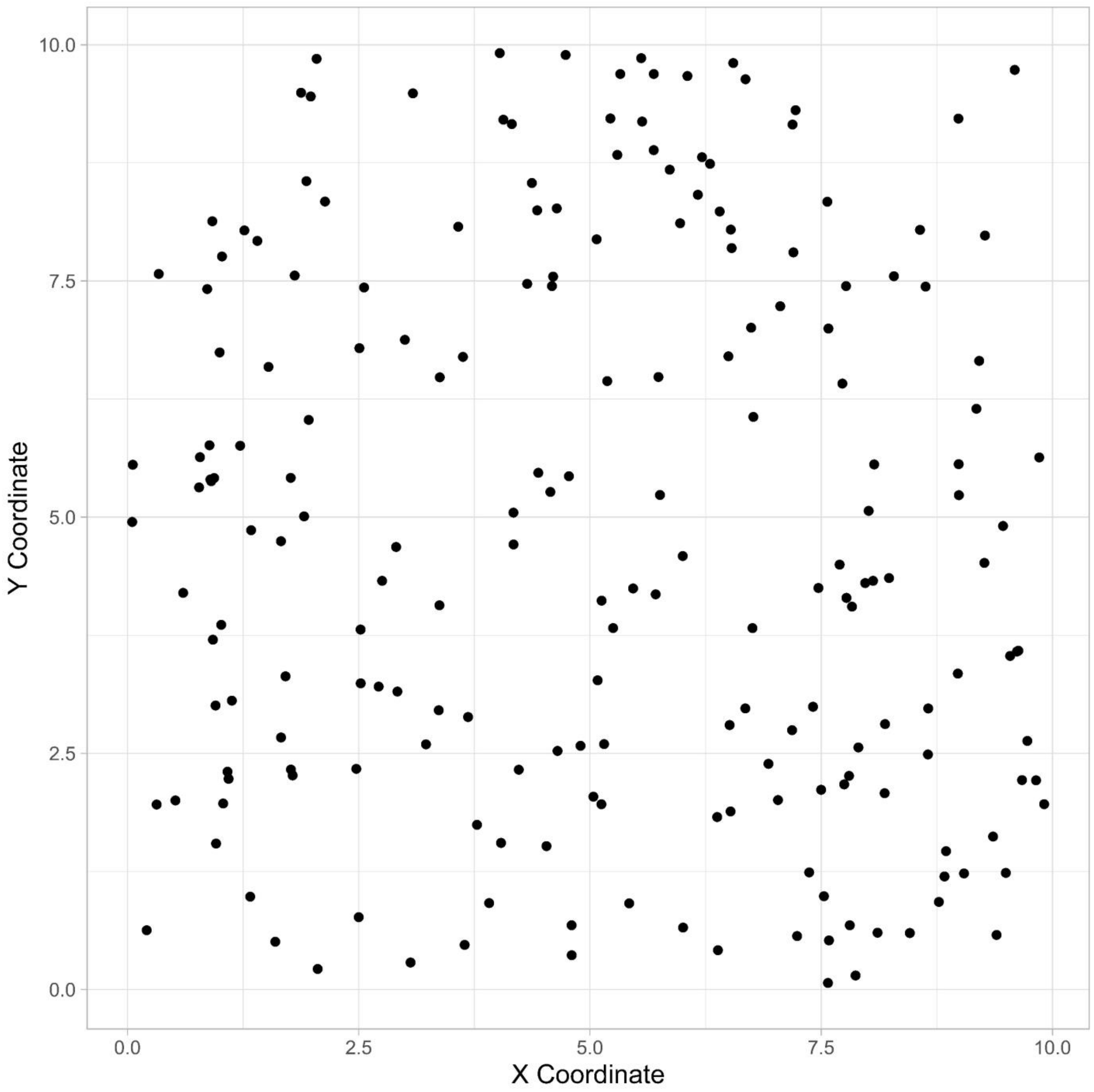
Spatial randomness
Spatial homogeneity
Spatial heterogeneity
Spatial dependence
Spatial autocorrelation

Do not worry, all these ideas will come to your mind as we move forward in the lectures

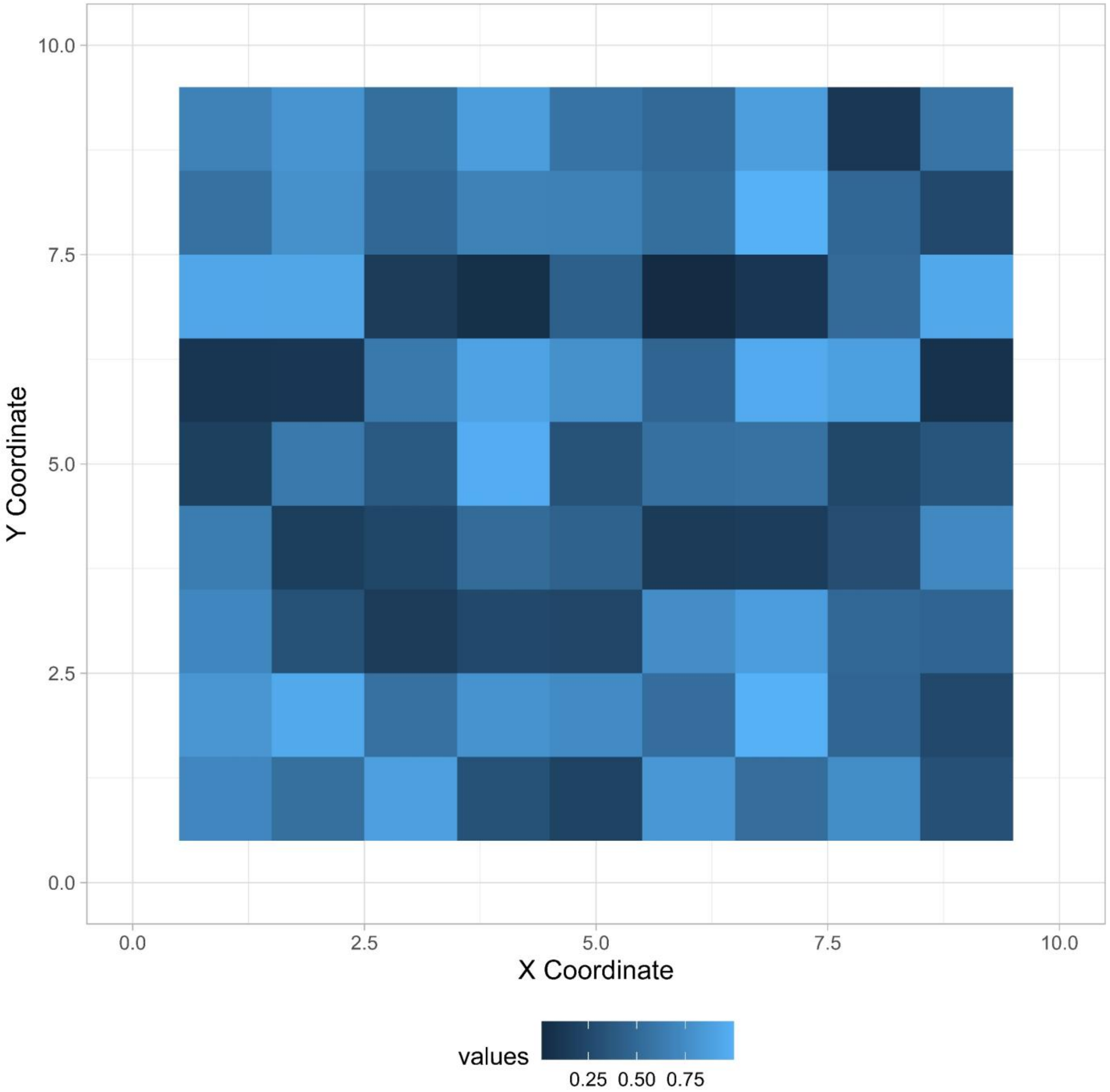
Key concepts

Spatial randomness:

- Lack of structure
- Visually difficult to detect
- When present is not interesting
- We usually want to test if a variable has spatial randomness, and if not, then it is because there is an underlying "spatial structure" we want to understand.
- Used as the Null Hypothesis in a test for spatial autocorrelation (More on this later)



Do you see any pattern?



Do you see any pattern?

What is the opposite to spatial randomness?

Key concepts

Spatial homogeneity (stationarity):

- Mean and variance of an attribute are constant across space

Spatial heterogeneity:

- Other names are: spatial structure, non-stationarity, large-scale global trends
- Different effect across space
- Influence of explanatory variables varies with the location of the observations
- "Structural instability in space“
- We often want to model spatial heterogeneity

Key concepts

Spatial dependence:

- The value of an area is linked to the values of neighborhood areas
- They (the areas) influence each other or are subject to the same unobserved phenomenon
- Observations in close geographic proximity exhibit greater similarity compared to those that are farther apart
- Can you think of an example?

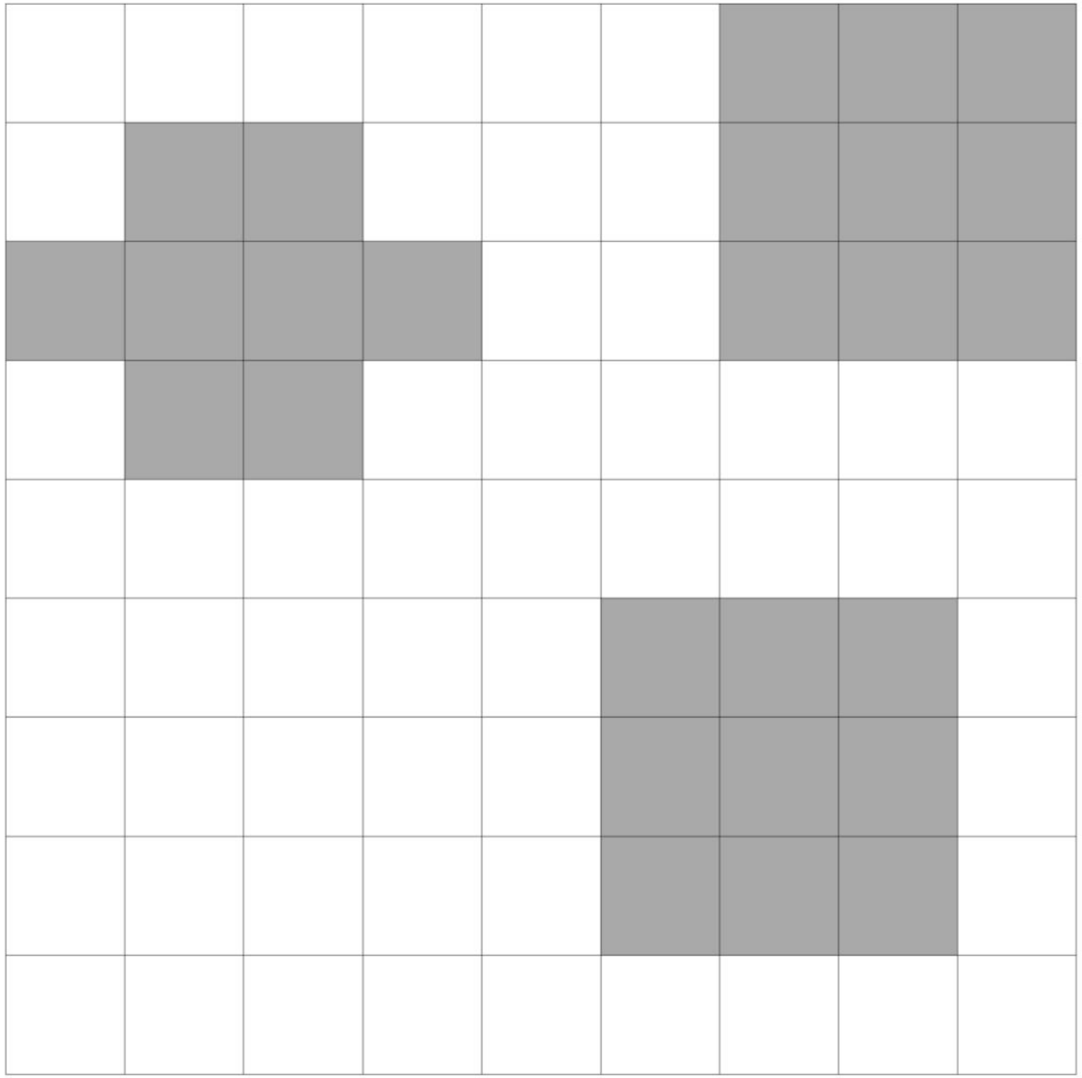
Spatial autocorrelation

- Spatial autocorrelation refers to the statistical relationship between observations in a spatial dataset based on their proximity or spatial arrangement.
- It measures the degree to which similar values or patterns tend to occur near each other in geographic space.
- In other words, spatial autocorrelation assesses whether nearby locations are more likely to have similar values compared to locations that are further apart.

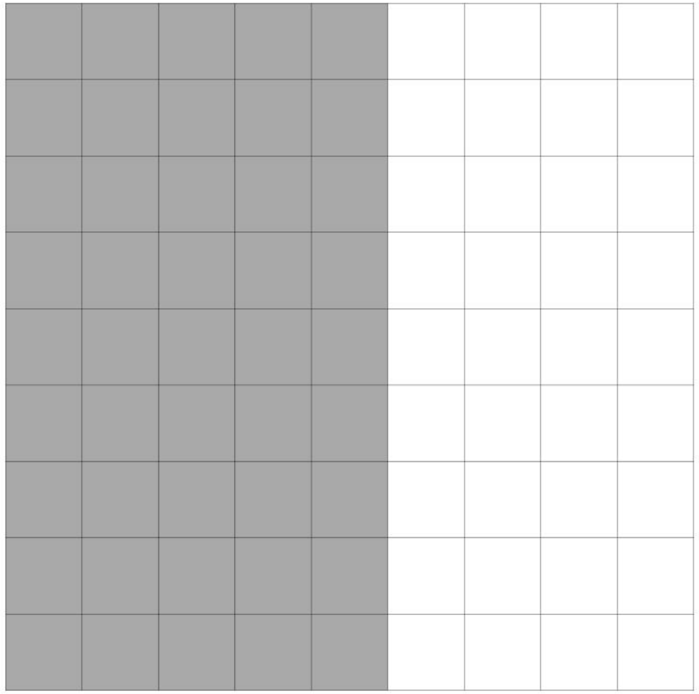
Spatial autocorrelation

"Spatial autocorrelation relates to the degree of dependency between the spatial location and the variable measured at that location (Chainey and Ratcliffe 2005). This spatial dependency could mean that the crime rate in one census area is partly influenced by the crime rate in a neighboring tract; for example, a drug set may sell drugs in one area and their presence may influence the growth of a drug market in the neighboring location.“
Jerry Ratcliffe (Crime Mapping: Spatial and Temporal Challenges) in Handbook of Quantitative Criminology (Alex R. Piquero and David Weisburd, 2010)

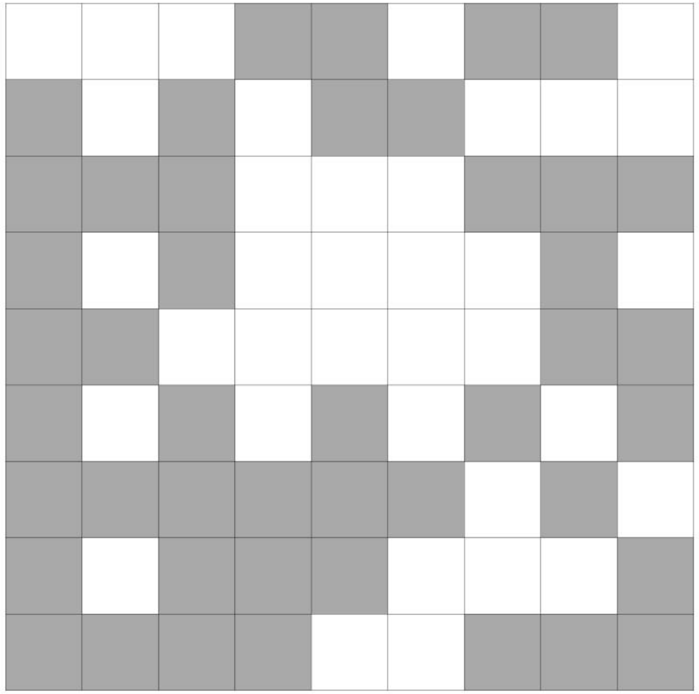
Spatial autocorrelation



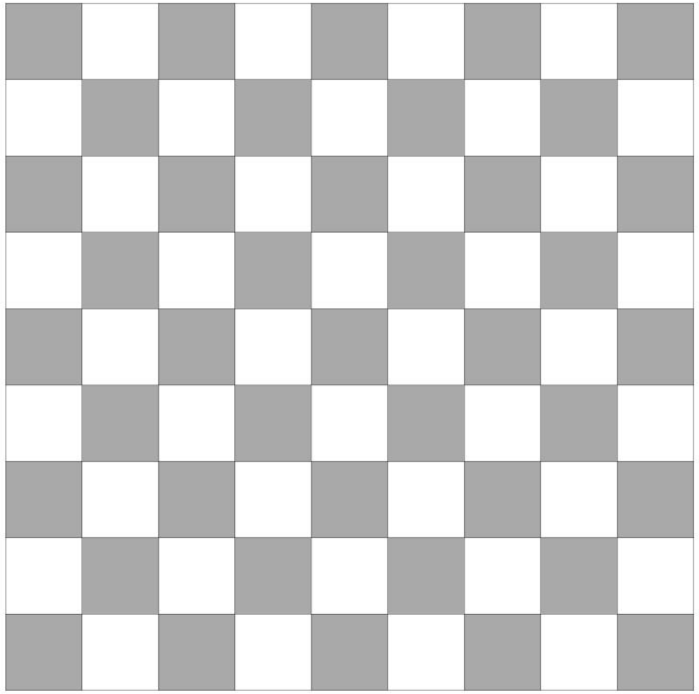
Spatial autocorrelation



Positive spatial autocorrelation



No spatial autocorrelation



Negative spatial autocorrelation

Adapted from George E. Tita & Steven M. Radil (Spatial Regression Models in Criminology: Modeling Social Processes in the Spatial Weights Matrix) in Handbook of Quantitative Criminology (Alex R. Piquero and David Weisburd, 2010)

Spatial autocorrelation vs spatial dependence

- Spatial autocorrelation and spatial dependence are the same but used in different contexts.
- Spatial autocorrelation is used for Exploratory Spatial Data Analysis ESDA
- Spatial dependence for spatial regression modeling

Spatial autocorrelation, so…

- We need to find a way to "measure" spatial autocorrelation
- And we need to find a way to assess if that measurement is correct (do we believe in it?)

Spatial autocorrelation, so…

- We need to find a way to “measure” spatial autocorrelation
- And we need to find a way to assess if that measurement is correct (do we believe in it?)

We do this by using spatial autocorrelation indexes. The most famous is Moran’s I

Spatial autocorrelation, so…

- We need to find a way to "measure" spatial autocorrelation
- And we need to find a way to assess if that measurement is correct (do we believe in it?)

We do this by using spatial autocorrelation indexes. The most famous is Moran's I

Unpacking Moran's I:

- Neighbours
- Mathematical form
- Hypothesis testing (we will take some time here)

Neighbours and the Spatial Weight Matrix

A	B	C
D	E	F
G	H	I

A	B	C
D	E	F
G	H	I

Neighbours for A?

A	B	C
D	E	F
G	H	I

A: {B, D}

A	B	C
D	E	F
G	H	I

A: {B, D}

E: {B, D, F, H}

H: {G, E, I}

Neighbourhood Matrix

	A	B	C	D	E	F	G	H	I
A	0	1	0	1	0	0	0	0	0
B	1	0	1	0	1	0	0	0	0
C	0	1	0	0	1	1	0	0	0
D	1	0	0	0	1	0	1	0	0
E	0	1	0	1	0	1	0	1	0
F	0	0	1	0	1	0	0	0	1
G	0	0	0	1	1	0	0	1	0
H	0	0	0	1	0	1	0	1	0
I	0	0	0	0	0	1	0	1	0

The Weight Matrix

	A	B	C	D	E	F	G	H	I	SUM
A	0	1	0	1	0	0	0	0	0	2
B	1	0	1	0	1	0	0	0	0	3
C	0	1	0	0	1	1	0	0	0	3
D	1	0	0	0	1	0	1	0	0	3
E	0	1	0	1	0	1	0	1	0	4
F	0	0	1	0	1	0	0	0	1	3
G	0	0	0	1	1	0	0	1	0	3
H	0	0	0	1	0	1	0	1	0	3
I	0	0	0	0	0	1	0	1	0	2

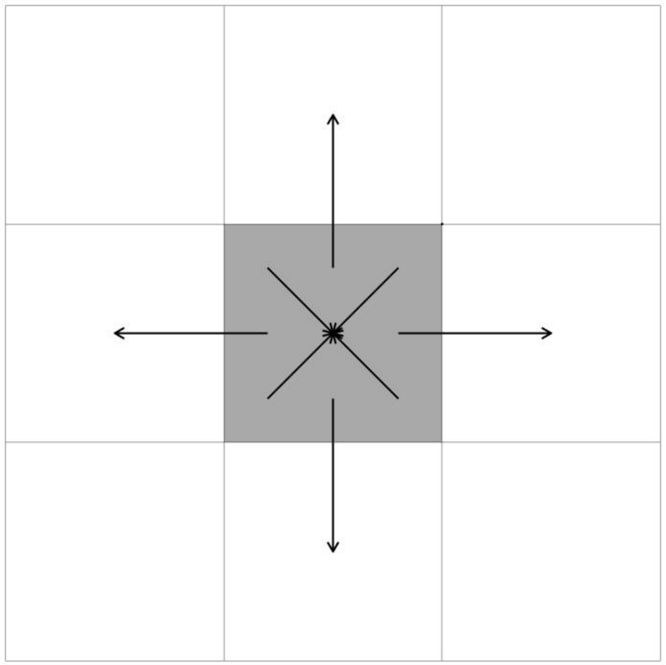
The Weight Matrix

	A	B	C	D	E	F	G	H	I
A	0	0.5	0	0.5	0	0	0	0	0
B	0.33	0	0.33	0	0.33	0	0	0	0
C	0	0.33	0	0	0.33	0.33	0	0	0
D	0.33	0	0	0	0.33	0	0.33	0	0
E	0	0.25	0	0.25	0	0.25	0	0.25	0
F	0	0	0.33	0	0.33	0	0	0	0.33
G	0	0	0	0.33	0.33	0	0	0.33	0
H	0	0	0	0.33	0	0.33	0	0.33	0
I	0	0	0	0	0	0.5	0	0.5	0

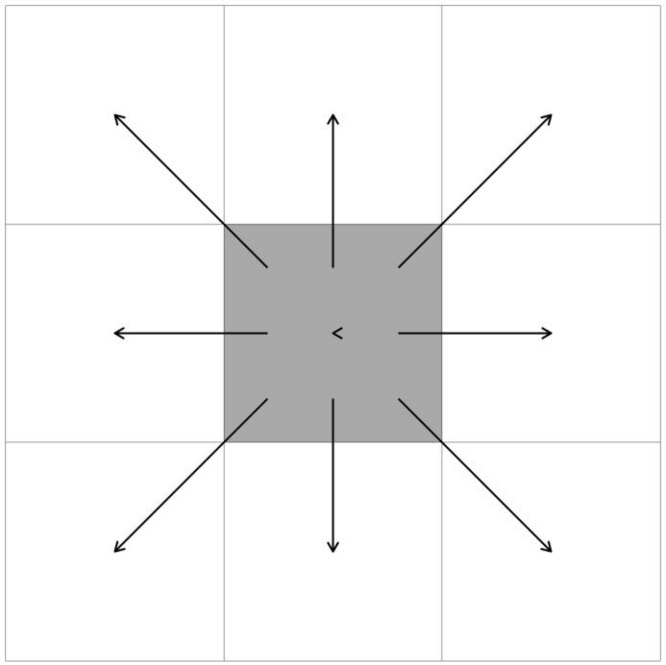
More formally

$if\ there\ are$ "n" zones, then $\mathbf{W}$ is an n$*$n squared matrix with $w_{ij}\ elements$

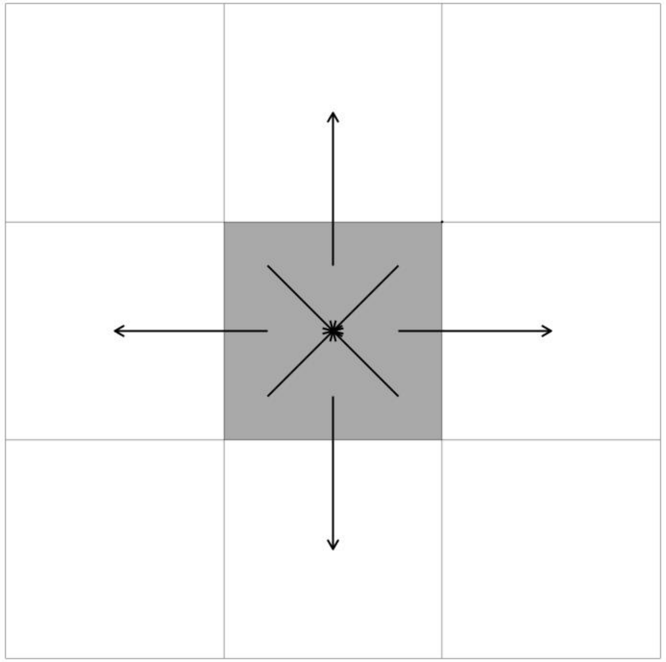
$$w_{ij} = \begin{cases} 1\ if\ \ i\ and\ j\ are\ spatially\ linked \\ 0\ otherwise \end{cases}$$



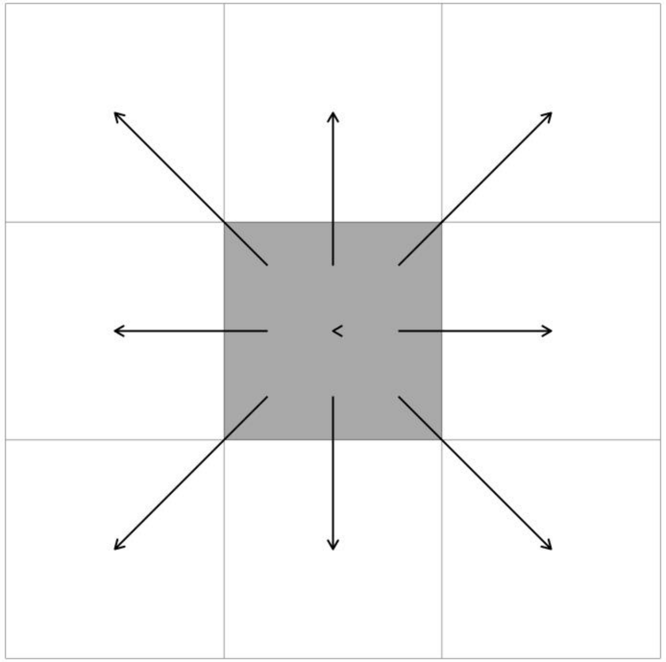
Rook contiguity



Queen contiguity



Rook contiguity

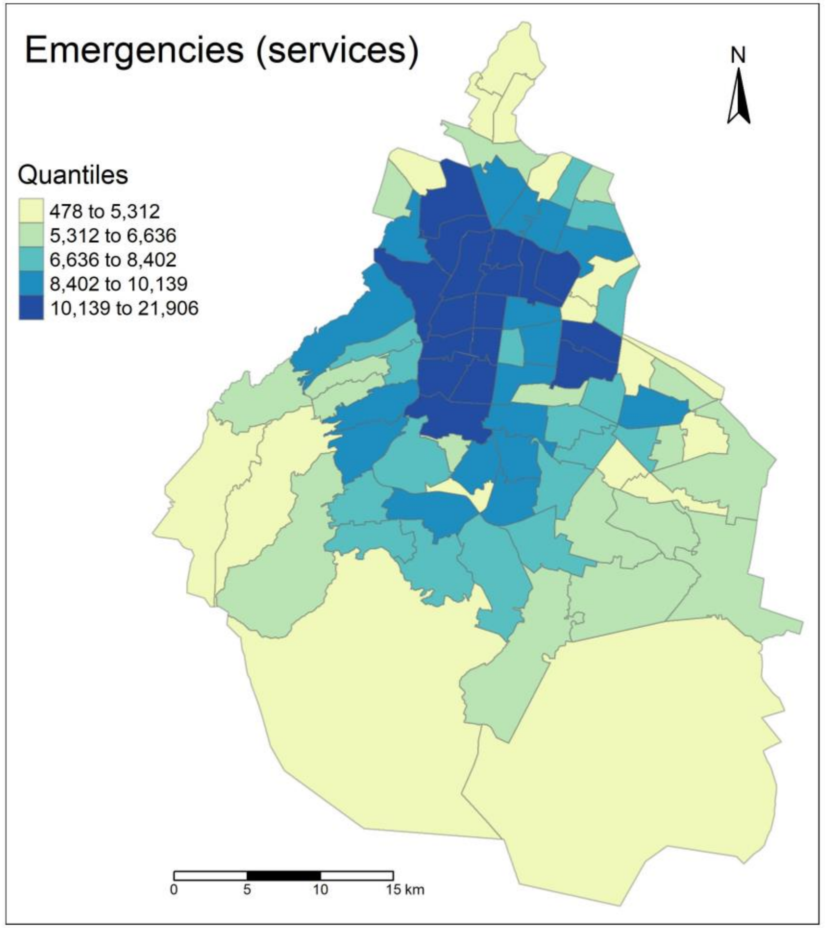


Queen contiguity

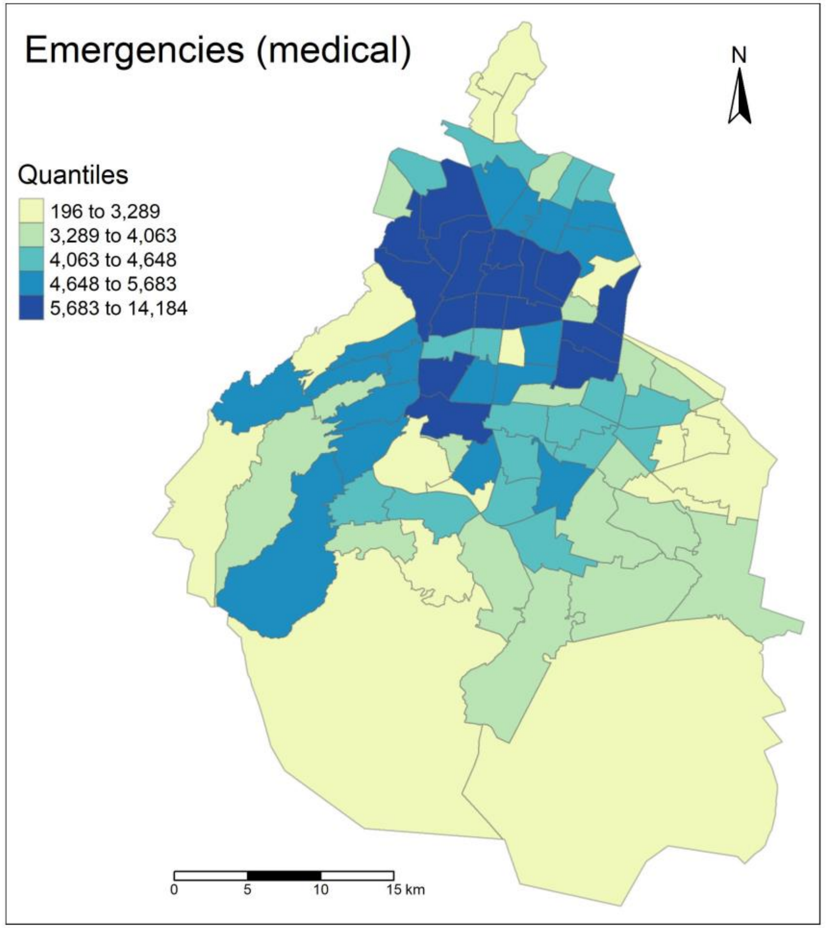
These are not the only alternatives

Distance based neighbourhood structures are also very popular

Morans's I



Is there a systematic pattern?
Is there a spatial structure in the data?
Similar values tend to be clustered together or dispersed?
Is there any spatial clustering?



Morans's I

$$S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

$$\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N}$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Covariance:

The variance measures the spread or dispersion of a single variable. It quantifies the average squared deviation from the mean.
The covariance measures the relationship between two variables. It quantifies how changes in one variable are associated with changes in another variable.

Correlation:

Standardized measure of the linear relationship between two variables. It quantifies the strength and direction of the relationship between the variables. It ranges from -1 to +1.

Morans's I

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

It quantifies the degree of similarity or dissimilarity between attribute values at different locations in a spatial dataset

Instead of two different variables here we are comparing the values in a region to the values in a neighboring region.

- Positive values (close to 1) indicate positive spatial autocorrelation, meaning that similar attribute values tend to be clustered together.
- Negative values (close to -1) indicate negative spatial autocorrelation, suggesting that dissimilar attribute values tend to be clustered together.
- A value of 0 indicates no spatial autocorrelation, implying that there is no systematic spatial pattern in the attribute values.

Spatial autocorrelation, so…

- ~~We need to find a way to "measure" spatial autocorrelation~~
- And we need to find a way to assess if that measurement is correct (do we believe in it?)

~~We do this by using spatial autocorrelation indexes. The most famous is Moran's I~~

Unpacking Moran's I:

- ~~Neighbours~~
- ~~Mathematical form~~
- Hypothesis testing (we will take some time here)

Thank you

Orlando Sabogal-Cardona
PhD researcher
University College London UCL