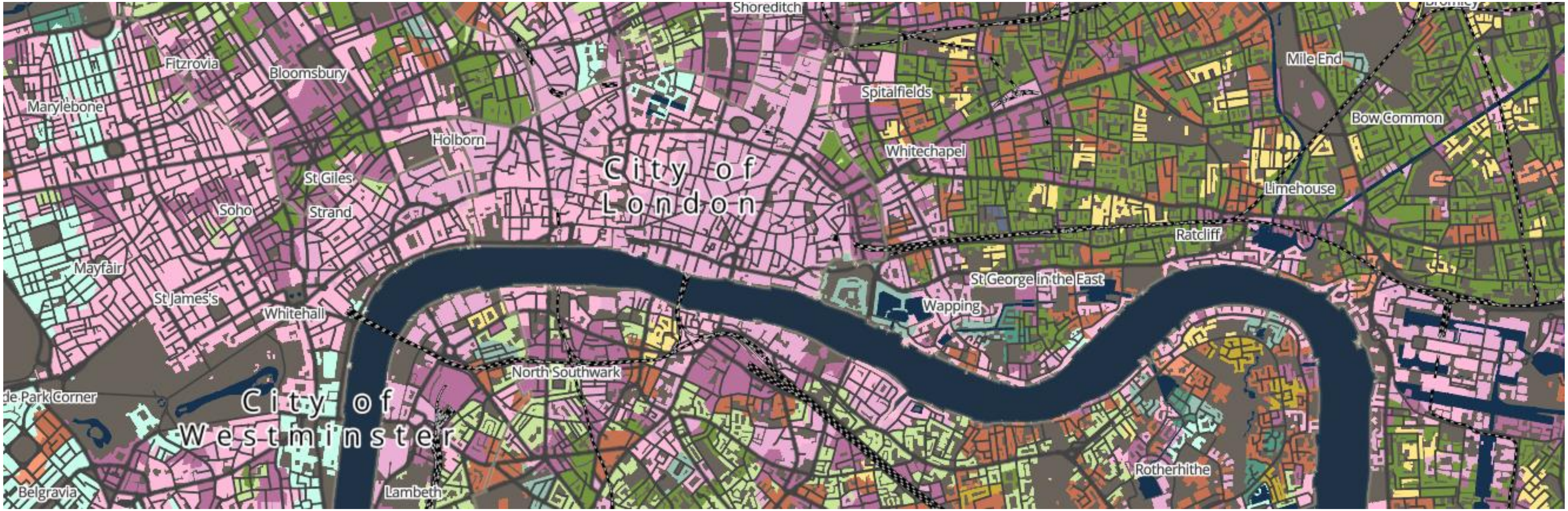
Geocomputation

Spatial Autocorrelation



Module outline

W1	Reproducible Spatial Analysis	Core Spatial Analysis
W2	Spatial Queries and Geometric Operations	
W3	Point Pattern Analysis	
W4	Spatial Autocorrelation	
W5	Spatial Models	
W6	Raster Data Analysis	Applied Spatial Analysis
W7	Geodemographic Classification	
W8	Accessibility Analysis	
W9	Beyond the Choropleth	Data Visualisation
W10	Complex Visualisations	

This week

- Spatial dependence
- Measuring spatial autocorrelation
- Spatial weights matrix

Spatial dependence

Spatial dependence

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

Walter Tobler 1970

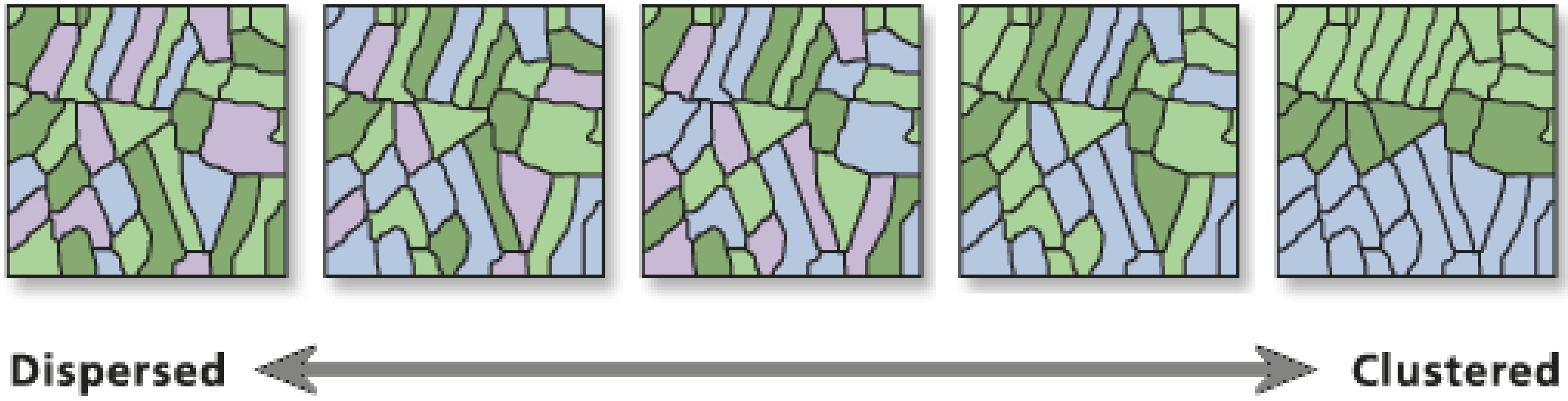
Spatial dependence

- Spatial dependence refers to the concept that the value of a variable at one location is influenced, to some extent, by the value of the same variable at nearby locations.
- This is often understood through the concept of distance decay, where the influence decreases as distance increases.

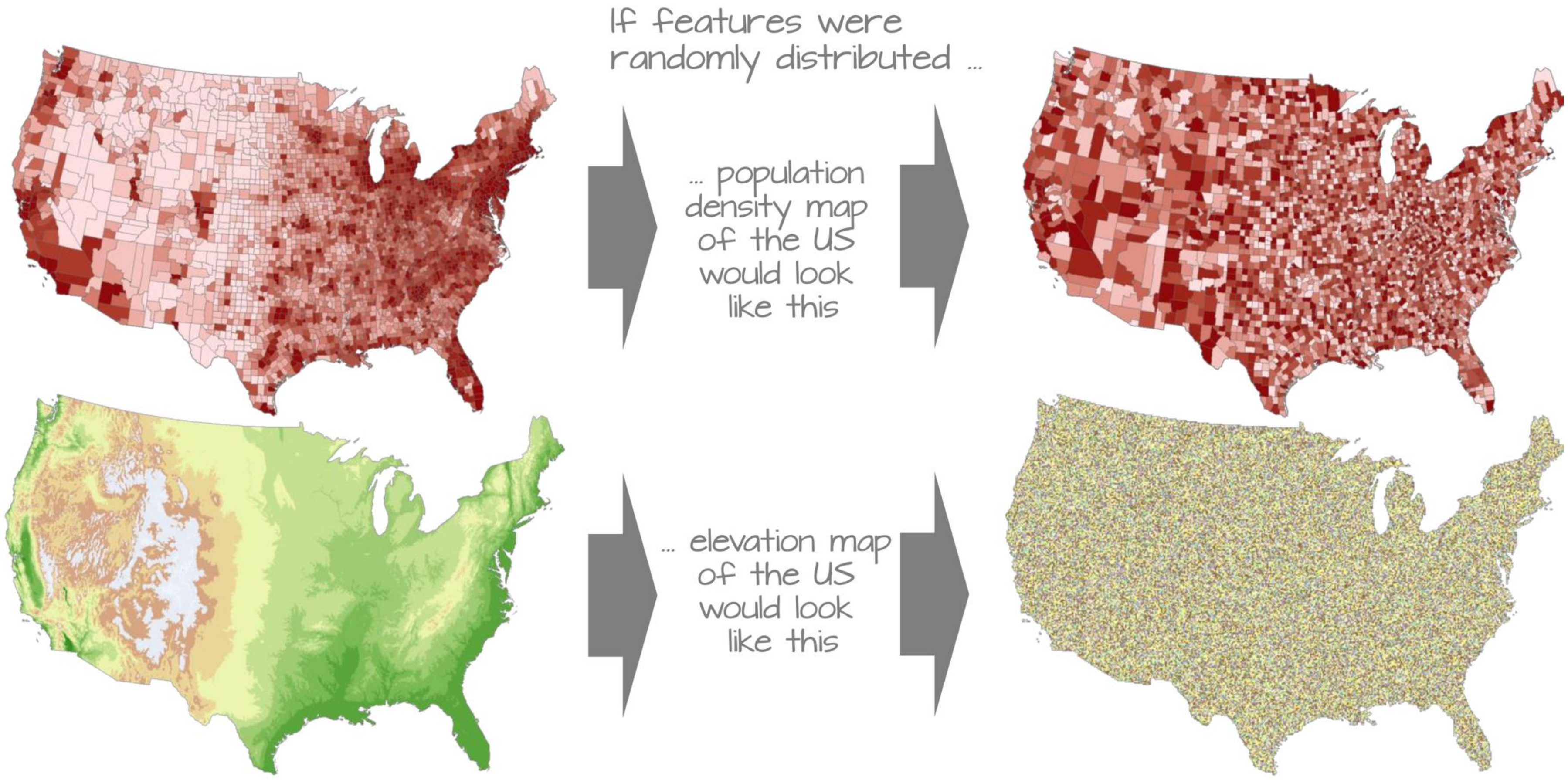
Spatial autocorrelation

- Measurement of spatial autocorrelation is the idea of formalising spatial dependency: measuring the degree to which similar values cluster together in space.
- By measuring spatial autocorrelation, we try to identify hotspots where high values are concentrated versus areas where low values are concentrated.
- Spatial Autocorrelation indicates the absence of Complete Spatial Randomness (CSR).
- CSR suggests that a pattern is entirely the result of random chance, with no underlying spatial structure.

Spatial autocorrelation



Spatial autocorrelation



Gimond, M. 2025. Intro to GIS and Spatial Analysis. [online]
https://mgimond.github.io/Spatial/index.html

Spatial autocorrelation

Spatial Autocorrelation can be measured in two ways:

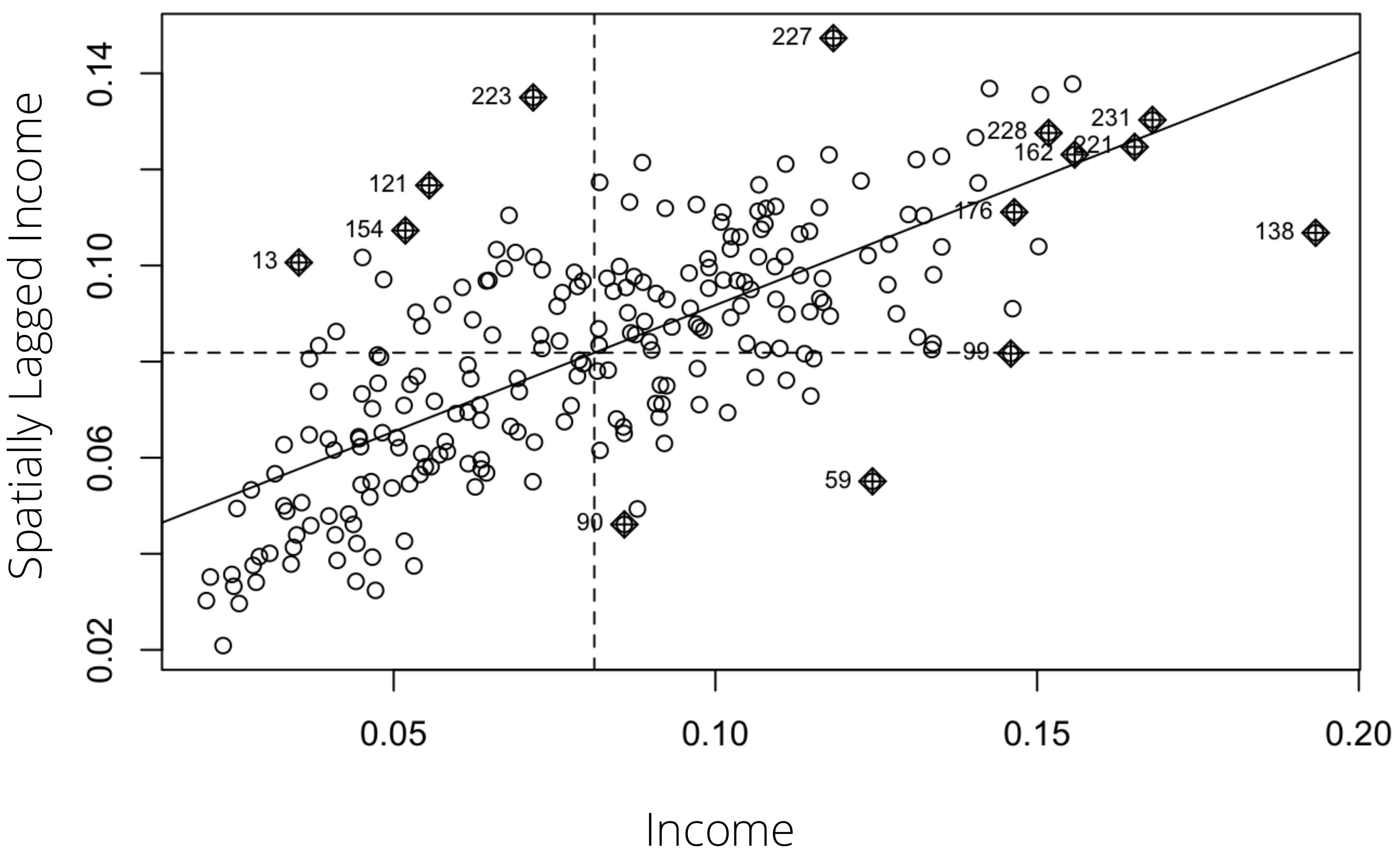
1) Global Spatial Autocorrelation: This assesses the overall spatial dependence across the entire dataset.
2) Local Spatial Autocorrelation: This focuses on the differences between each unit of analysis and its neighbors.

Global spatial autocorrelation

Moran’s I

- Moran's I: The most commonly used indicator of global spatial autocorrelation.
- Identifies neighbours for each target feature (e.g. polygon) and summarises their values by computing their means to create a spatially lagged variable value.
- Plots the target feature’s value against its spatially lagged mean value and fits a linear model to the points.
- The slope of the fitted line (β estimate) is the Moran’s I statistic.

Moran’s I



Spatially Lagged Income
Income
0.02
0.06
0.10
0.14
0.05
0.10
0.15
0.20
227
223
231
228
162
221
121
176
154
138
13
99
59
90

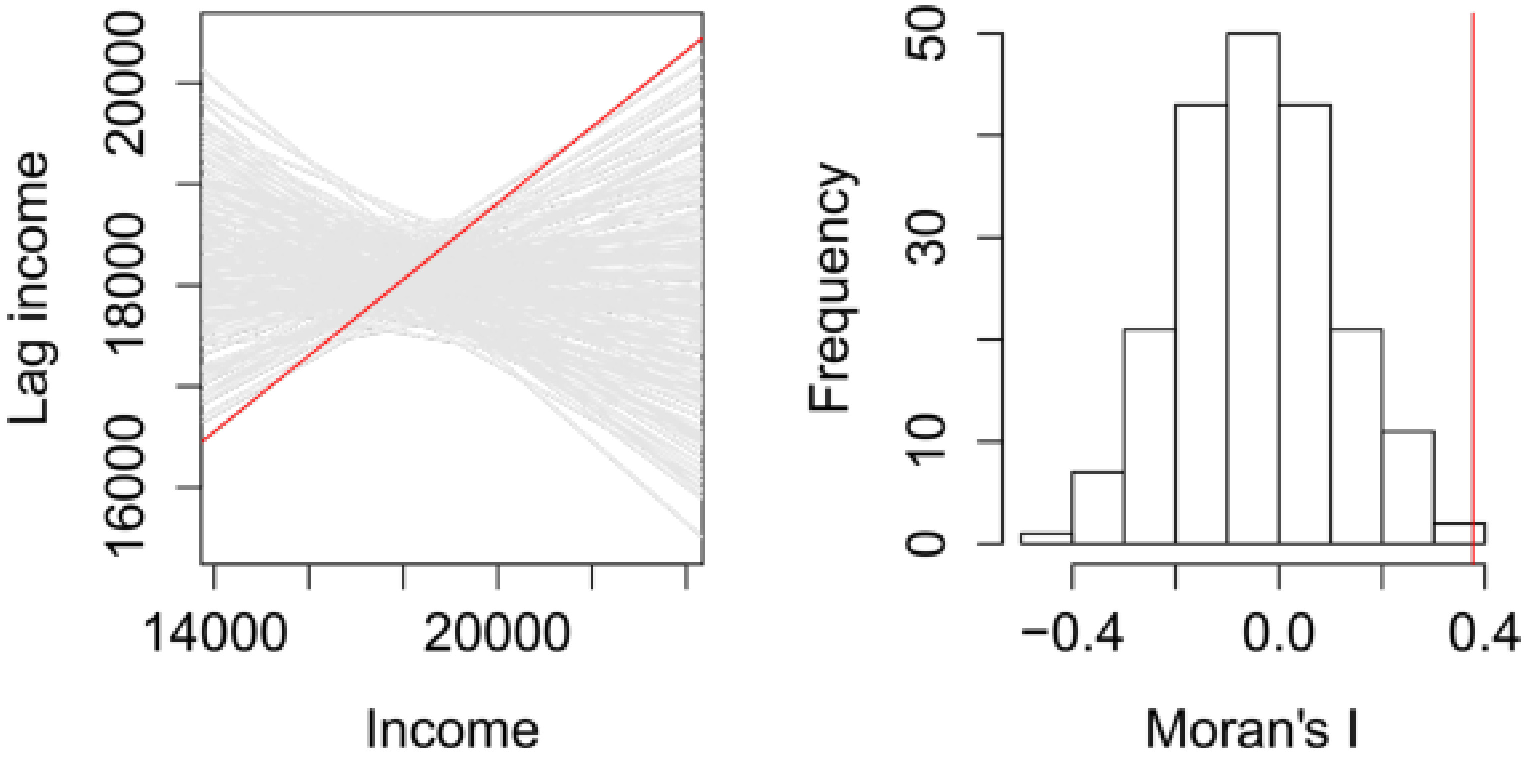
Moran’s I

- The Moran’s I statistic typically ranges from -1 to +1:
  - +1 Indicates perfect clustering (positive spatial autocorrelation).
  - 0 Suggests a random pattern (no spatial autocorrelation).
  - -1 Indicates perfect dispersion (negative spatial autocorrelation).
- How to assess the significance of the relationship?

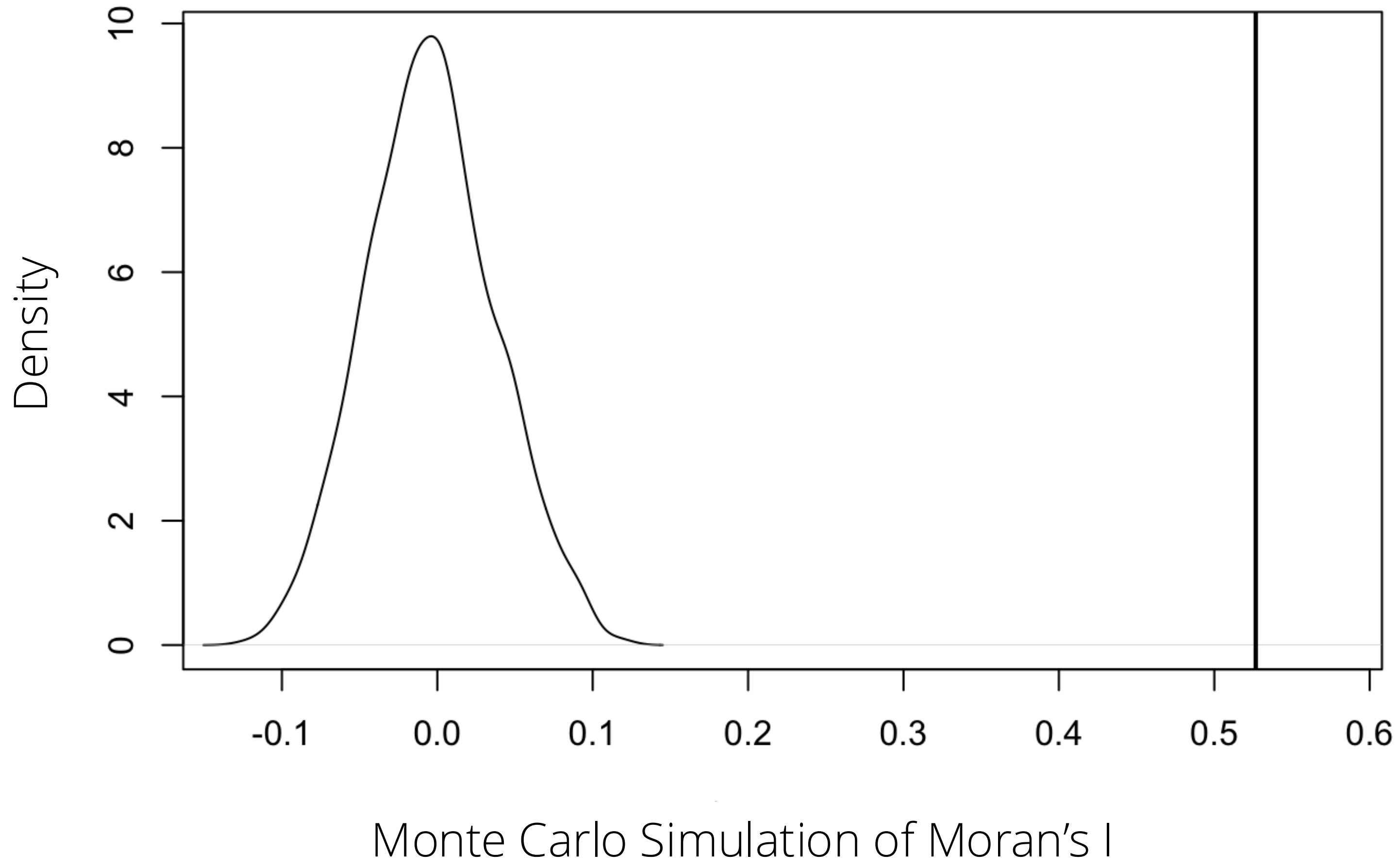
Moran's I

- To understand whether our relationship is significant, we can use either an analytical approach or a computational approach. The latter is the preferred option as it does not require making any assumption about the shape and layout of our data set – for this we can use a Monte Carlo test.
- This approach randomly and repeatedly assigns values to polygons in the data set.
- The output is a sampling distribution of Moran's I values under the (null) hypothesis that attribute values are randomly distributed across the study area.

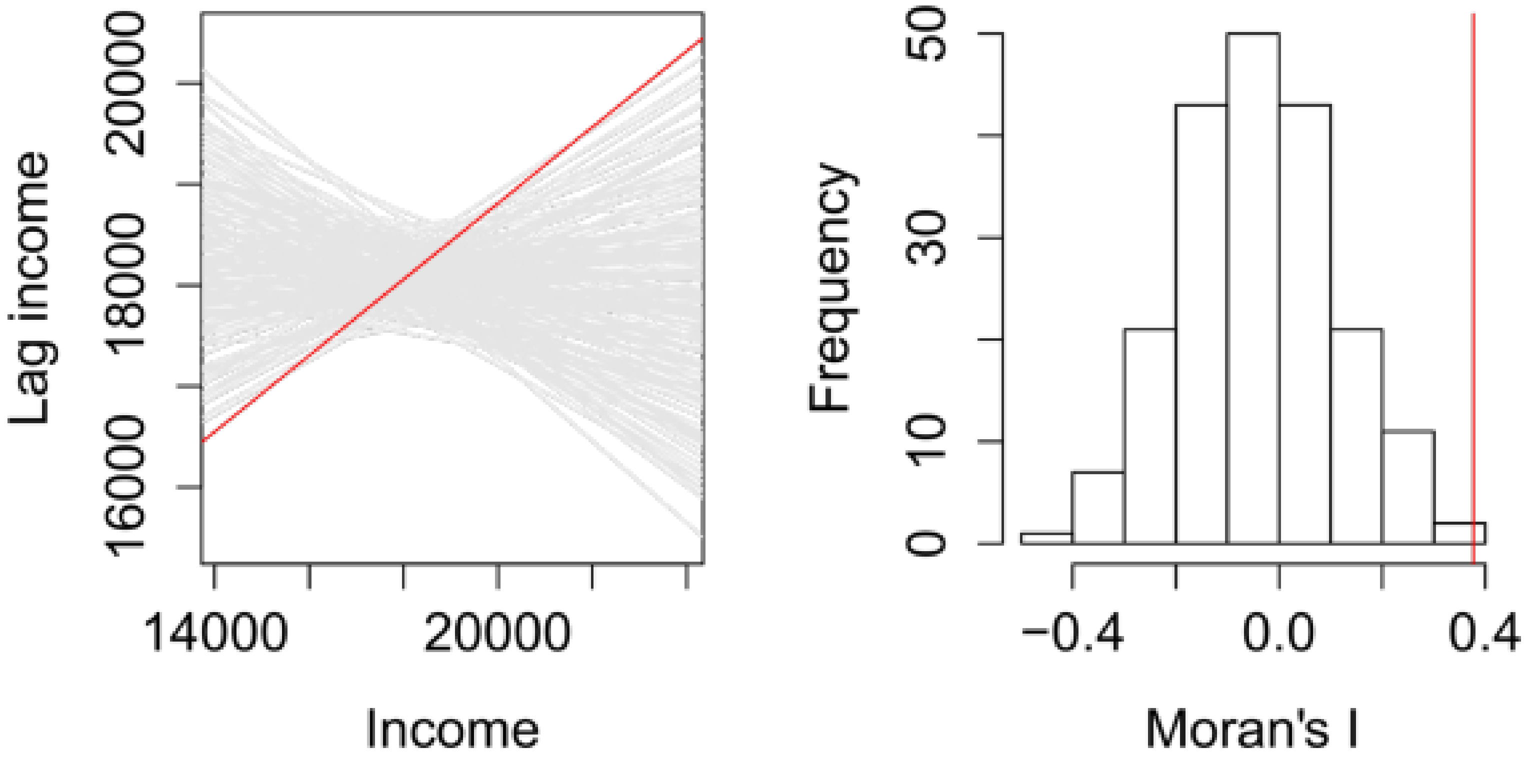
Moran’s I


Lag income
16000
18000
20000
14000
20000
Income
Frequency
0
10
30
50
−0.4
0.0
0.4
Moran's I

Moran’s I


Density
0
2
4
6
8
10
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Monte Carlo Simulation of Moran’s I

Moran’s I

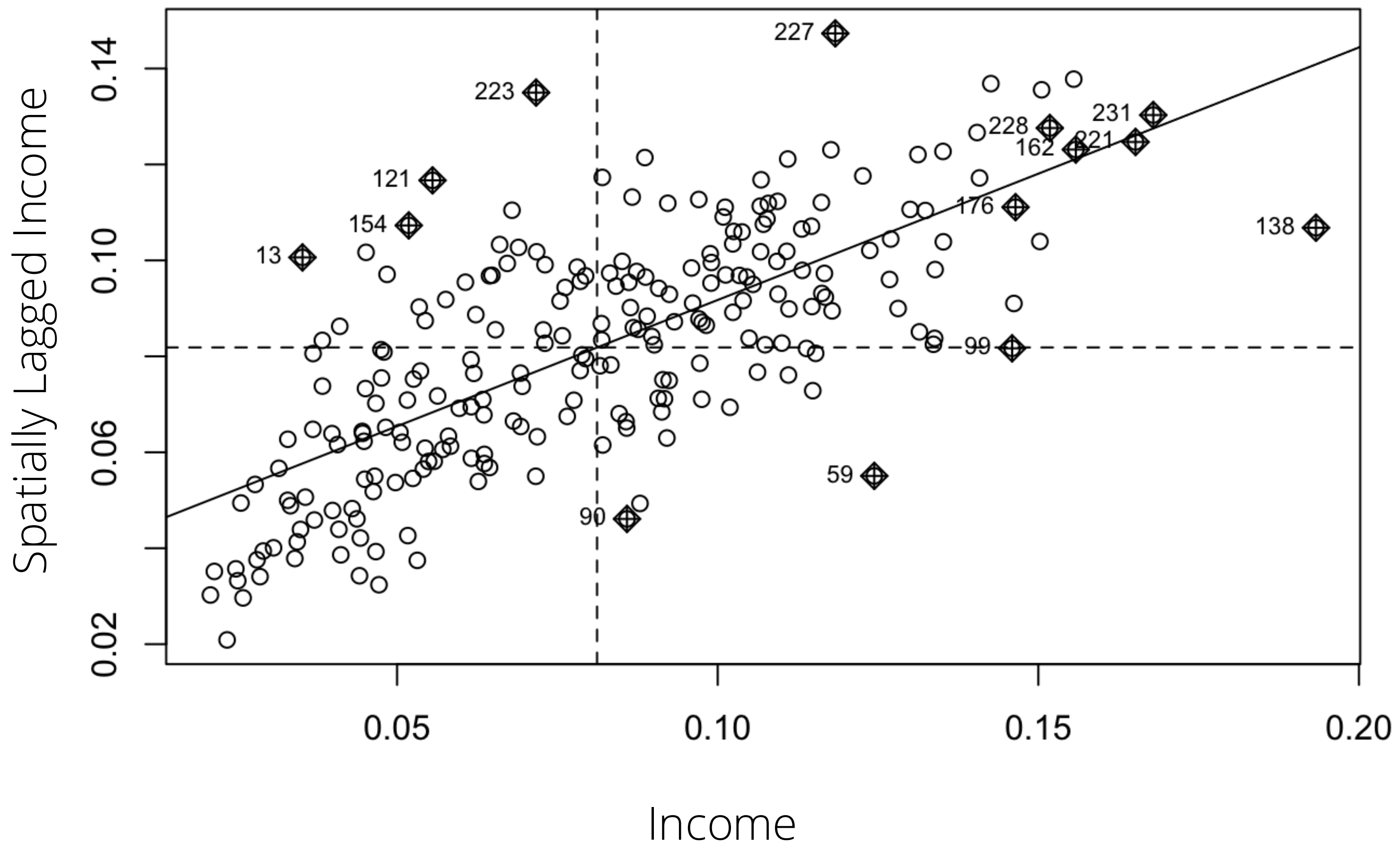

Lag income
16000
18000
20000
14000
20000
Income
Frequency
0
10
30
50
−0.4
0.0
0.4
Moran's I

Moran’s I

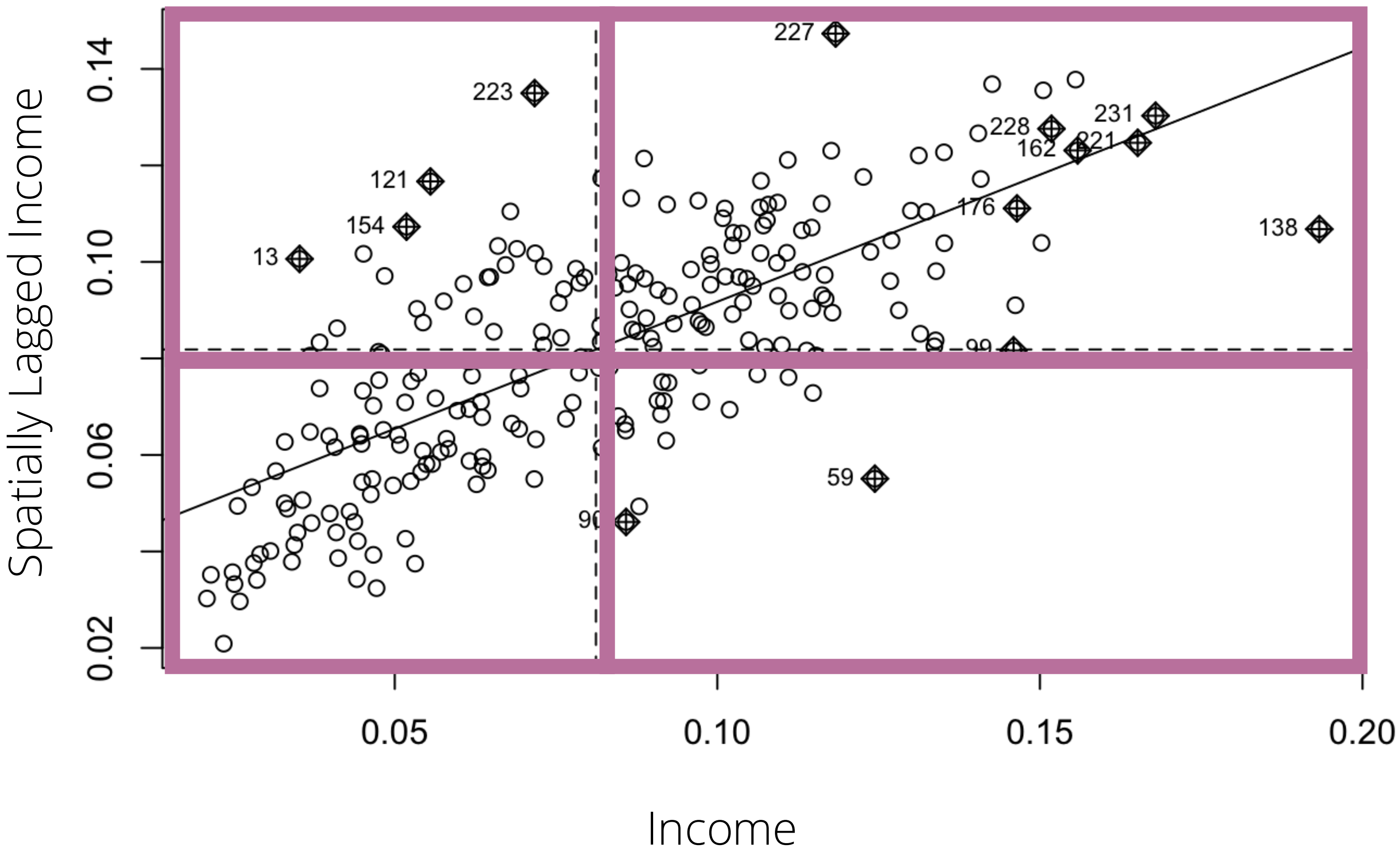
- A pseudo *p*-value is generated from the simulation results.
- For instance: if out of 199 simulations, just one simulation result is more extreme than our observed statistic , *p* is equal to (1 + 1) / (199 + 1) = 0.01. This is interpreted as “there is a 1% probability that we would be wrong in rejecting the null hypothesis.”
- Be aware, that the pseudo *p*-value is only a summary of the results from the reference distribution and should not be interpreted as an analytical *p*-value (assumption of normality and normal distribution).

Local spatial autocorrelation

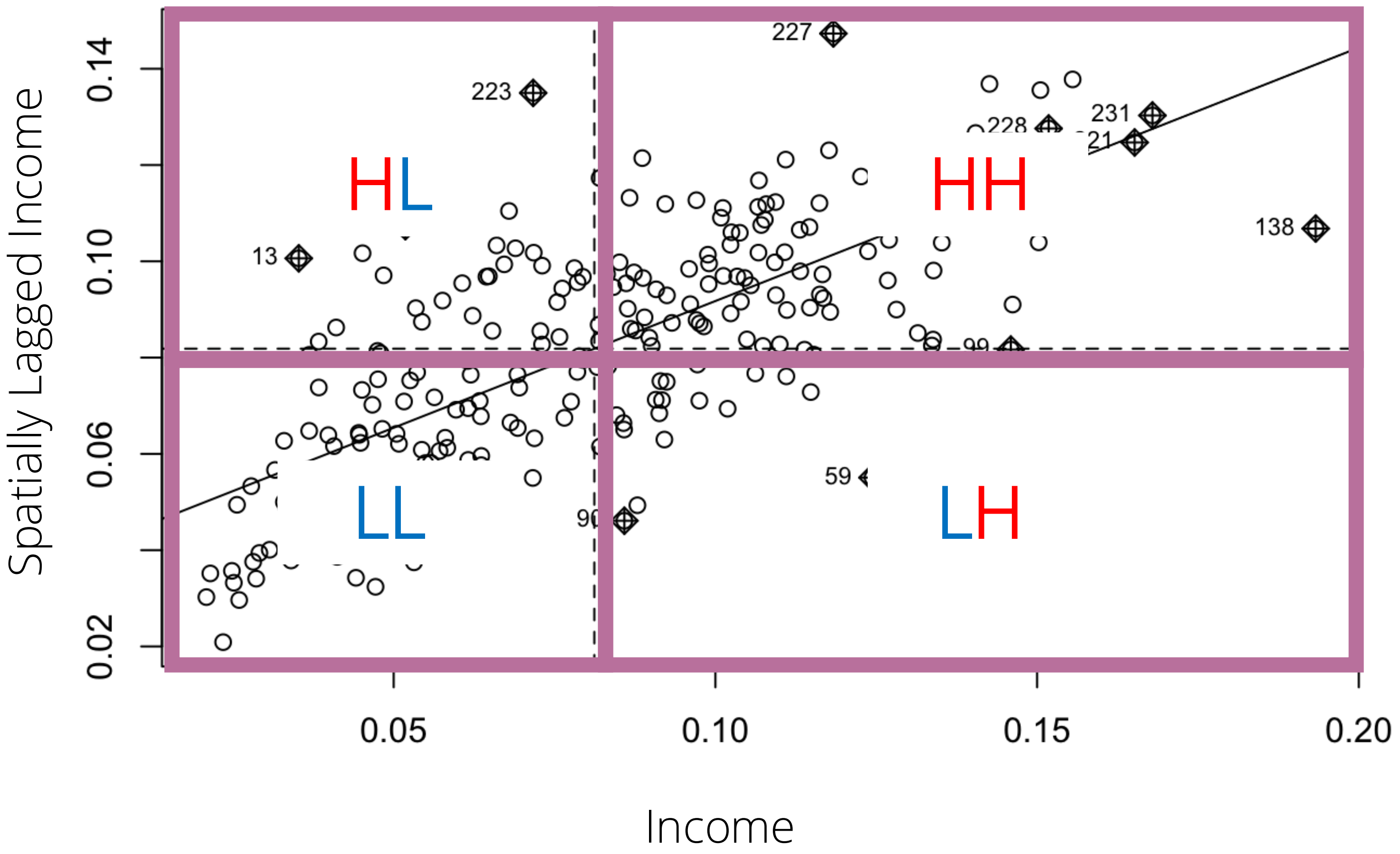
Local Moran’s I



Local Moran’s I



Local Moran’s I



Local Moran’s I

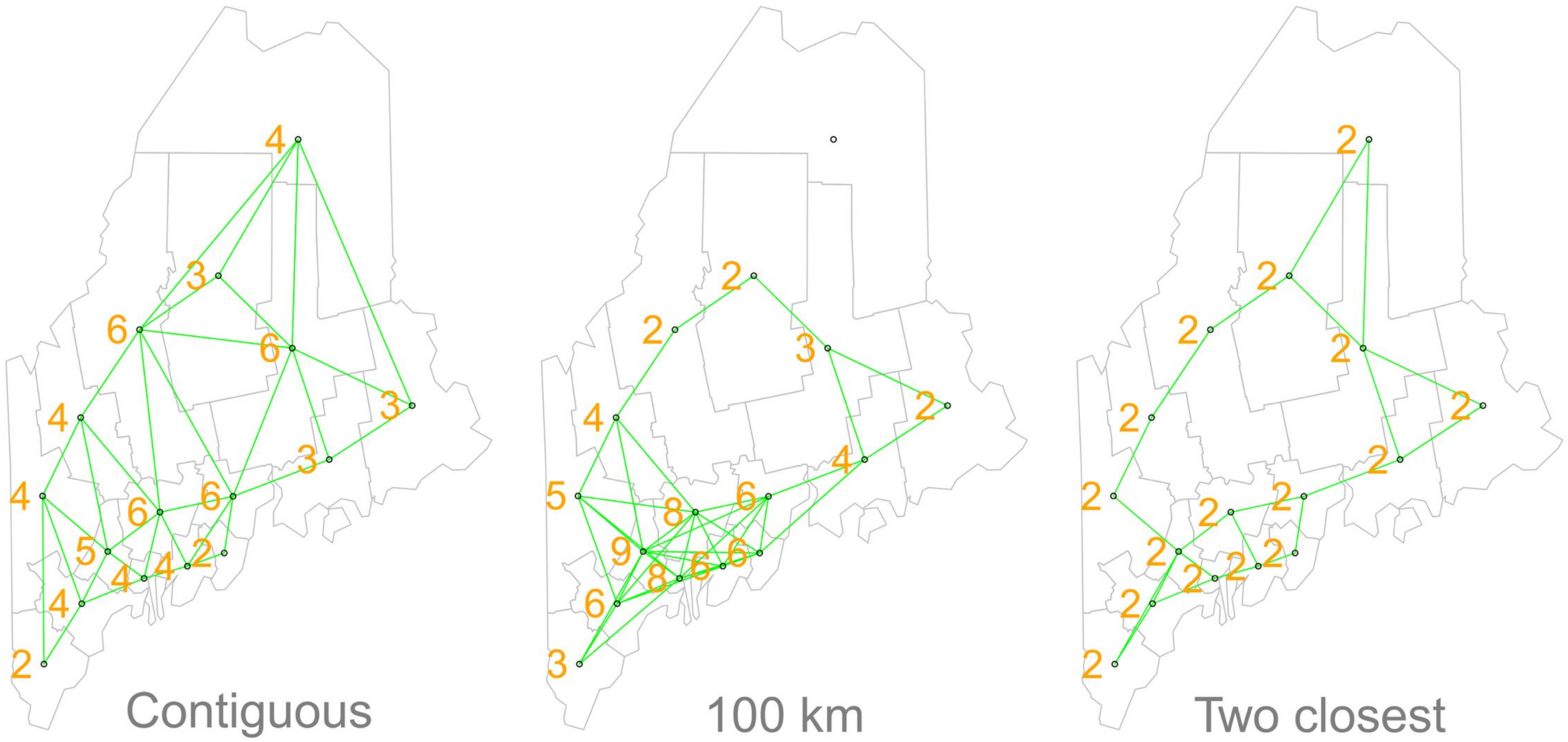
- Decomposing the Moran’s I statistic: Local Moran’s I
- Assesses spatial autocorrelation at the local level by evaluating each feature and its surrounding neighborhood.
- Four cluster types: high-high, low-low, but also outliers: high-low, low-high.
- Commonly known as cluster and outlier analysis.
- Monte Carlo simulation can be used to assess significance of these clusters.

Spatial weight matrix

Who is your neighbour?

- To calculate any measure of spatial autocorrelation, it's essential to understand how spatial units relate to each other as neighbours, specifically how we define their spatial relationships.
- There are two primary approaches to defining neighbours:
  - Proximity: Neighbours are determined based on the distance between spatial units. A unit is considered a neighbour if it is within a certain distance from another unit.
  - Contiguity: Neighbours are defined by shared boundaries or edges. Spatial units are considered neighbours if they touch each other directly, either along edges (rook contiguity) or at corners (queen contiguity).

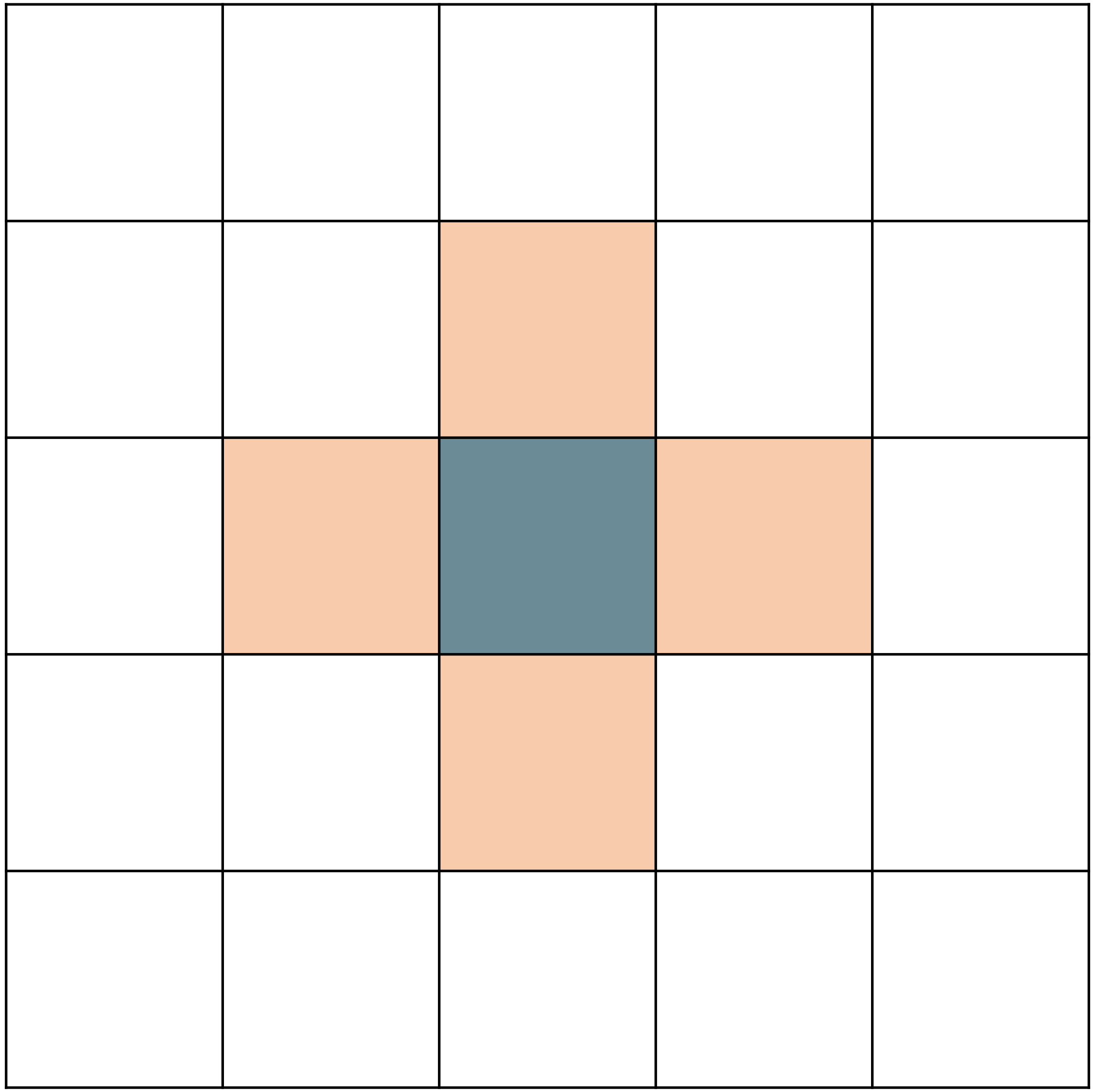
Who is your neighbour?



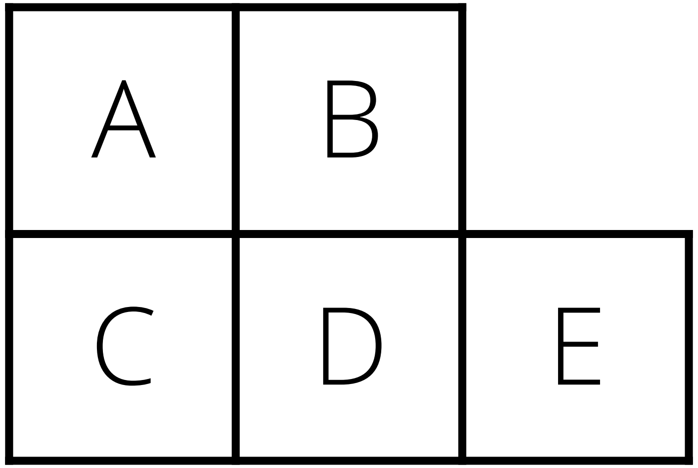
Spatial weights matrix

- How to use the relationships between neighbouring features?
- Spatial weight matrix: an `N x N` positive matrix (`W`) that summarises all spatial relationships between the features in a dataset.
- It provides a formal expression of spatial dependency by quantifying the influence of each feature on every other feature based on their spatial proximity or other relevant criteria.

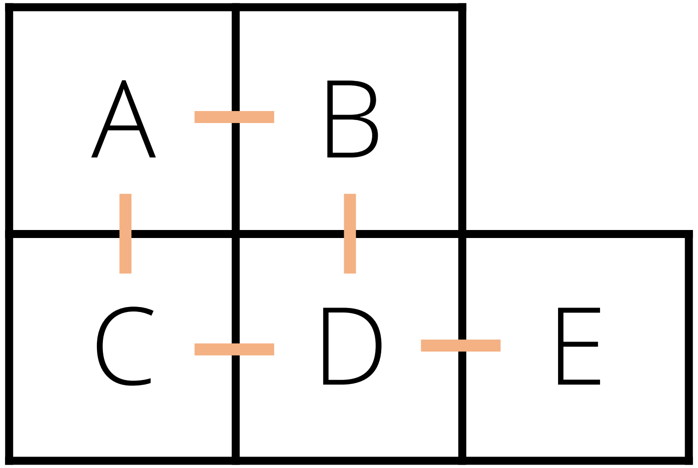
Spatial weights matrix



Spatial weights matrix

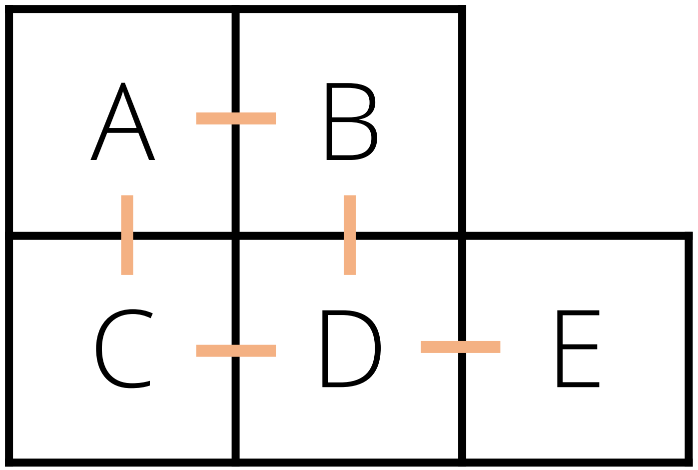


Spatial weights matrix



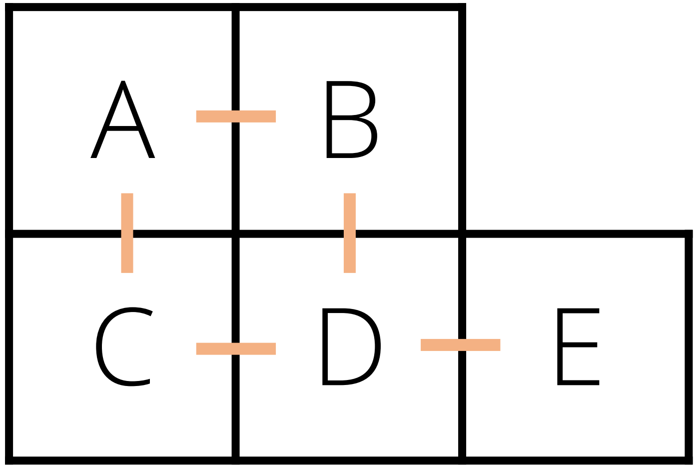
Spatial weights matrix

	a	b	c	d	e
a	0	1	1	0	0
b	1	0	0	1	0
c	1	0	0	1	0
d	0	1	1	0	1
e	0	0	0	1	0



Spatial weights matrix

	a	b	c	d	e		w
a	0	1	1	0	0		2
b	1	0	0	1	0		2
c	1	0	0	1	0	=	2
d	0	1	1	0	1		3
e	0	0	0	1	0		1

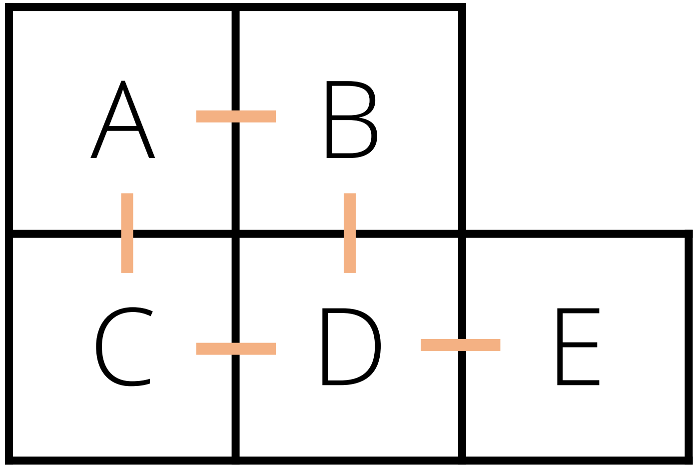


Spatial weights matrix

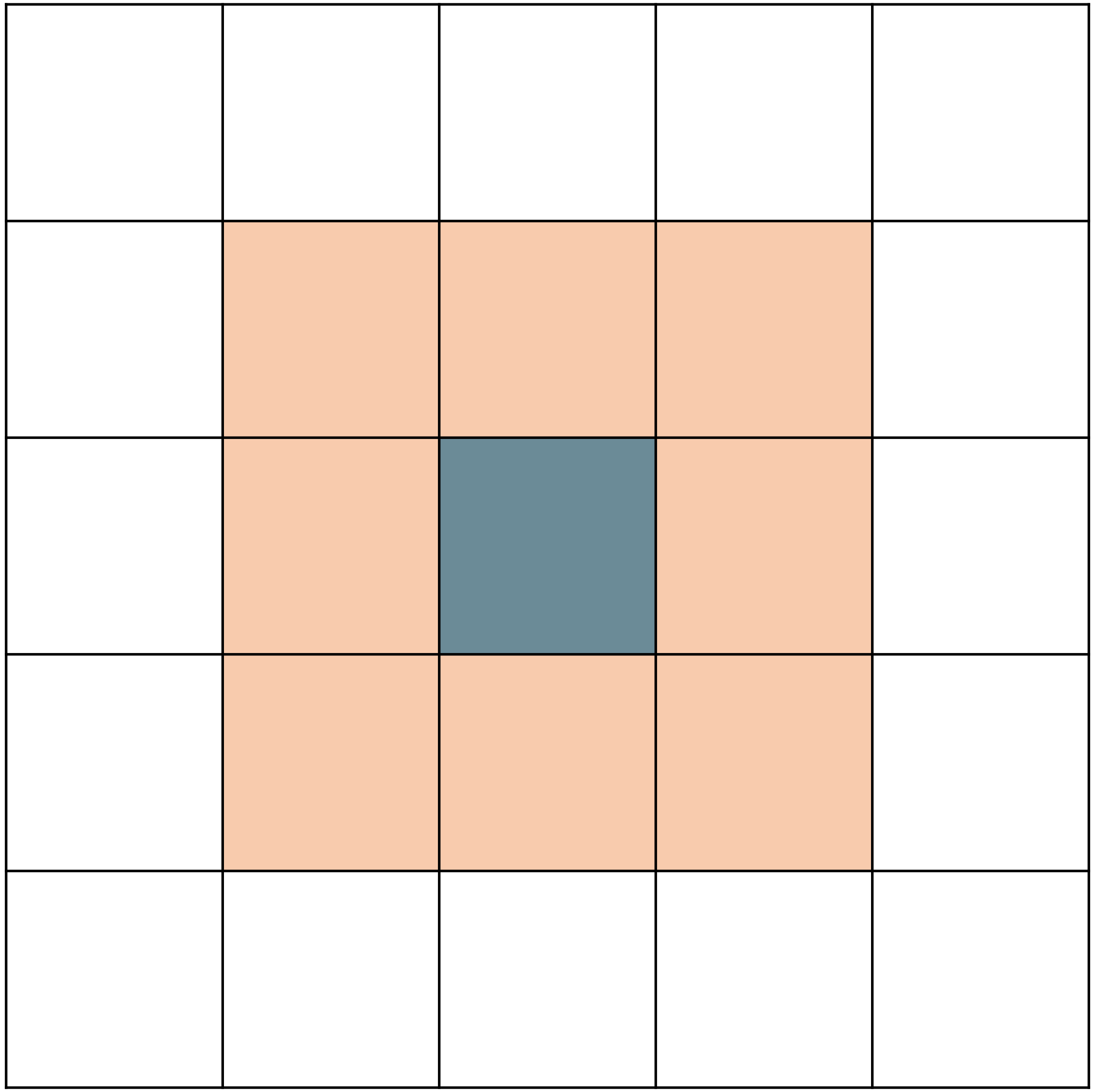
	a	b	c	d	e
a	0	.5	.5	0	0
b	.5	0	0	.5	0
c	.5	0	0	.5	0
d	0	.33	.33	0	.33
e	0	0	0	1	0

=

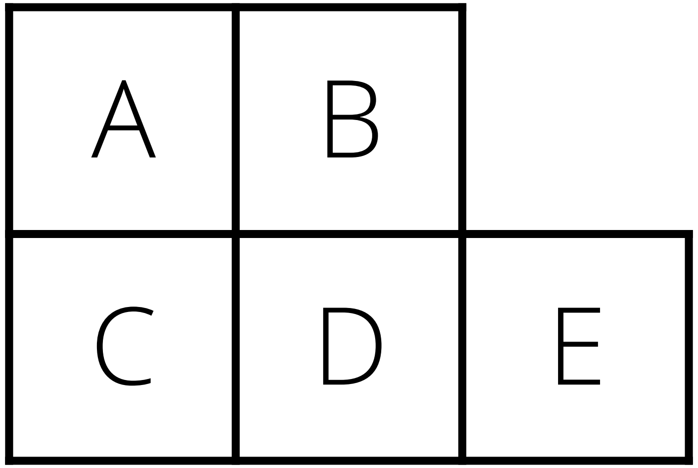
w
2
2
2
3
1



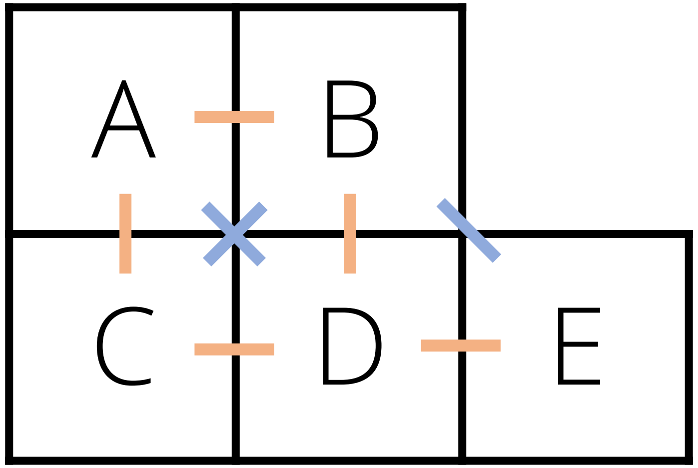
Queen



Spatial weights matrix

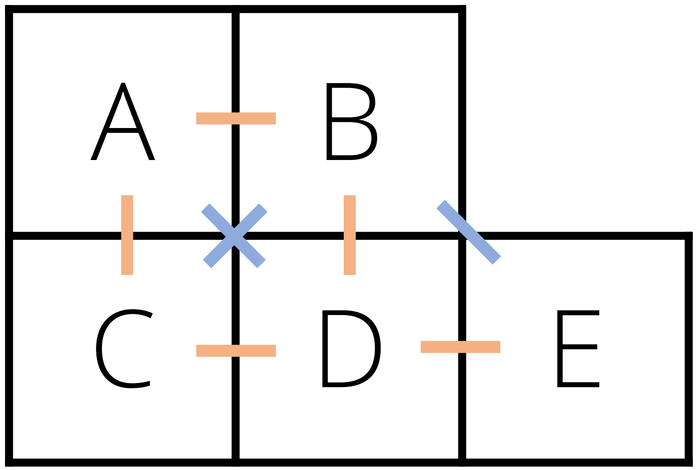


Spatial weights matrix

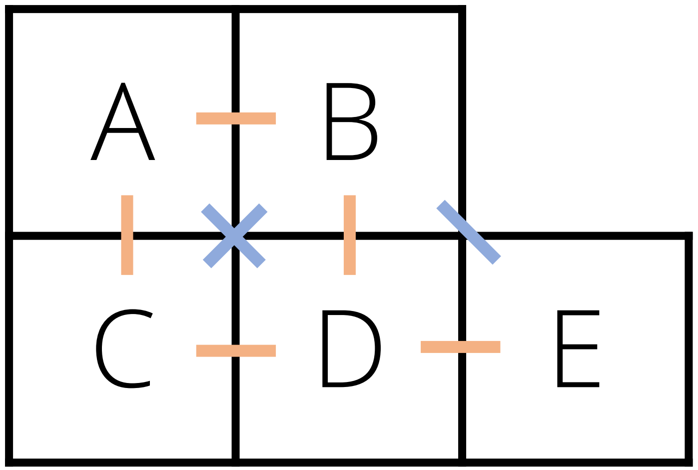


Spatial weights matrix

	a	b	c	d	e
a	0	1	1	1	0
b	1	0	1	1	1
c	1	1	0	1	0
d	1	1	1	0	1
e	0	1	0	1	0



Spatial weights matrix



	a	b	c	d	e
a	0	1	1	1	0
b	1	0	1	1	1
c	1	1	0	1	0
d	1	1	1	0	1
e	0	1	0	1	0

=

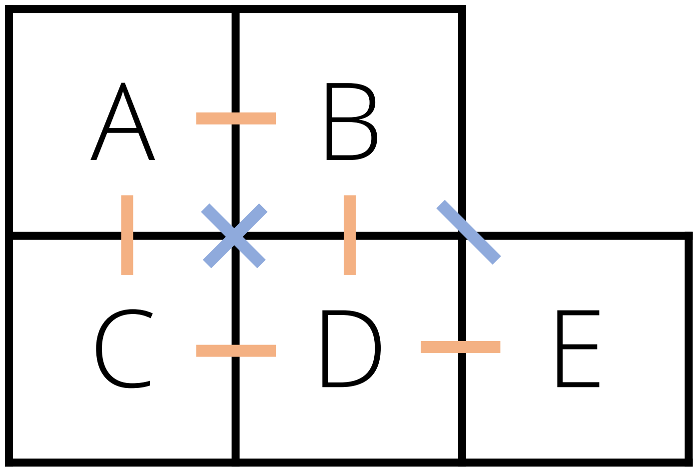
w
3
4
3
4
2

Spatial weights matrix

	a	b	c	d	e
a	0	.33	.33	.33	0
b	.25	0	.25	.25	.25
c	.33	.33	0	.33	0
d	.25	.25	.25	0	.25
e	0	.5	0	.5	0

=

w
3
4
3
4
2



Spatial weights matrix

Standardisation can be done in different ways. In `spdep` package:

- "B" coding scheme	no standardisation (heterogeneity between zones)
- “W” coding scheme	row standardisation
- “C” coding scheme	global standardisation; weights are standardised so that the sum of all weights is equal to the total number of entities
- “U” coding scheme	weights are standardised so that the sum of all weights equals 1

Topology

Conclusion

- Measuring spatial autocorrelation is essential for understanding spatial relationships and patterns in data.
- We discussed two common measures of spatial autocorrelation, though other methods also exist.
- A spatial weights matrix is required to calculate the spatially lagged variable used in these measures.
- The way neighbours are defined (proximity or contiguity) directly affects the results of spatial autocorrelation tests.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk