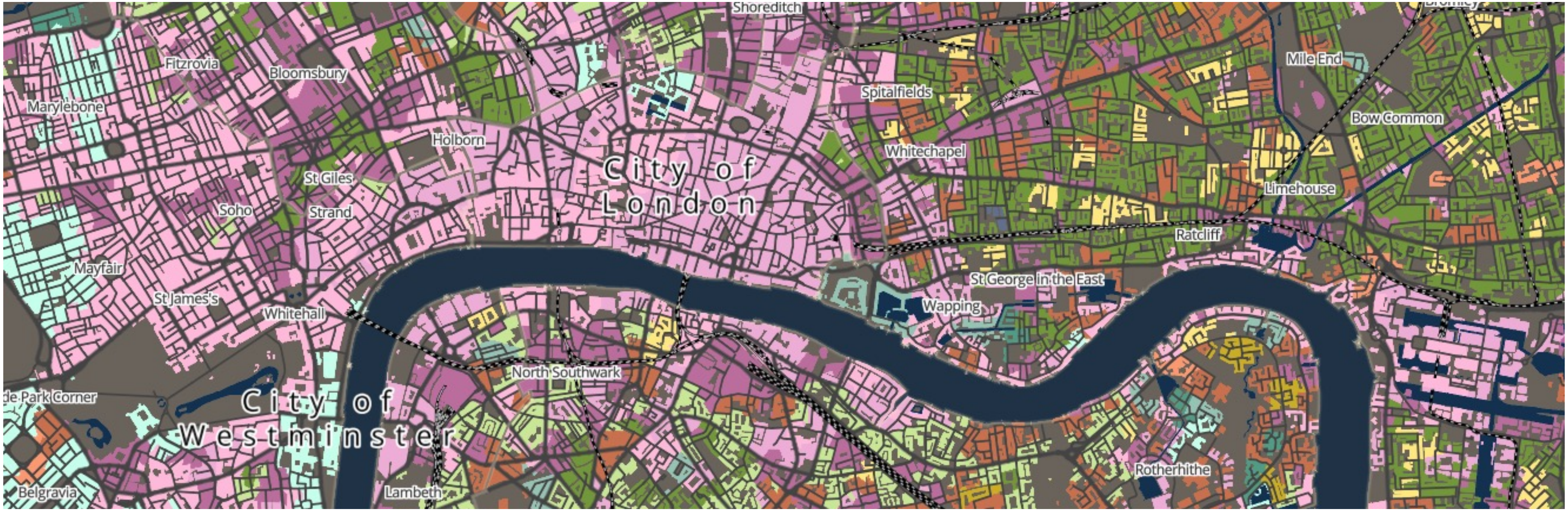
Geocomputation

Spatial Autocorrelation



Where are we at?

Part I: Foundational Concepts

W1	Geocomputation: An Introduction	
W2	GIScience and GIS software	QGIS
W3	Cartography and Visualisation	QGIS
W4	Programming for Data Analysis	R
W5	Programming for Spatial Analysis	R

Where are we at?

Part II: Core Spatial Analysis

W6	Geometric Operations and Spatial Queries	R
W7	Point Pattern Analysis	
W8	**Spatial Autocorrelation**	

Part III: Advanced Spatial Analysis

W9	Rasters, Zonal Statistics and Interpolation	R
W10	Transport Network Analysis	

This week

- Spatial dependence and spatial autocorrelation
- Measuring spatial autocorrelation: Global Moran’s I, Local Moran’s I, Getis-Ord Gi*
- Spatial weight matrix

Spatial dependence

Spatial dependence

- Spatial dependence is the idea that the observed value of a variable in one location is dependent (to some degree) on the observed value of the same variable in a nearby location.
- Often understood as distance decay – an idea which is used in many geographical applications (e.g. spatial interpolation, spatial interaction).

Spatial dependence

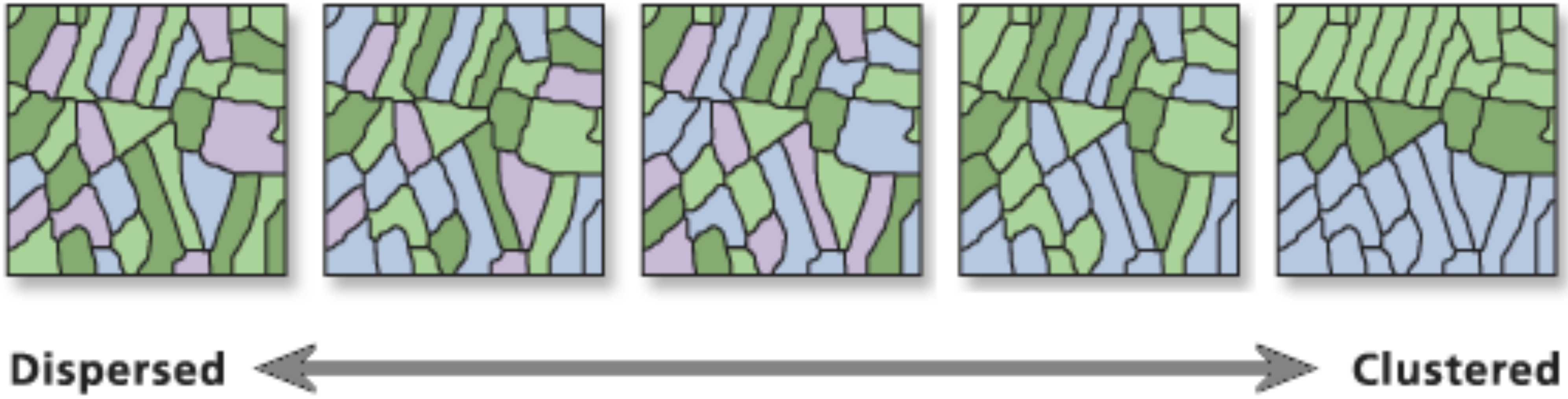
“Everything is related to everything else, but near things are more related than distant things.”

Walter Tobler 1970

Spatial autocorrelation

- Statistical way of formalising spatial dependency.
- Spatial version of “normal” correlation.
- In short: the degree to which similar values cluster in space.
- Questions: Is our data clustered, randomly distributed ,or dispersed? Can we find hotspots of high values versus low values? Can we find areas where high values exist directly next to low values?

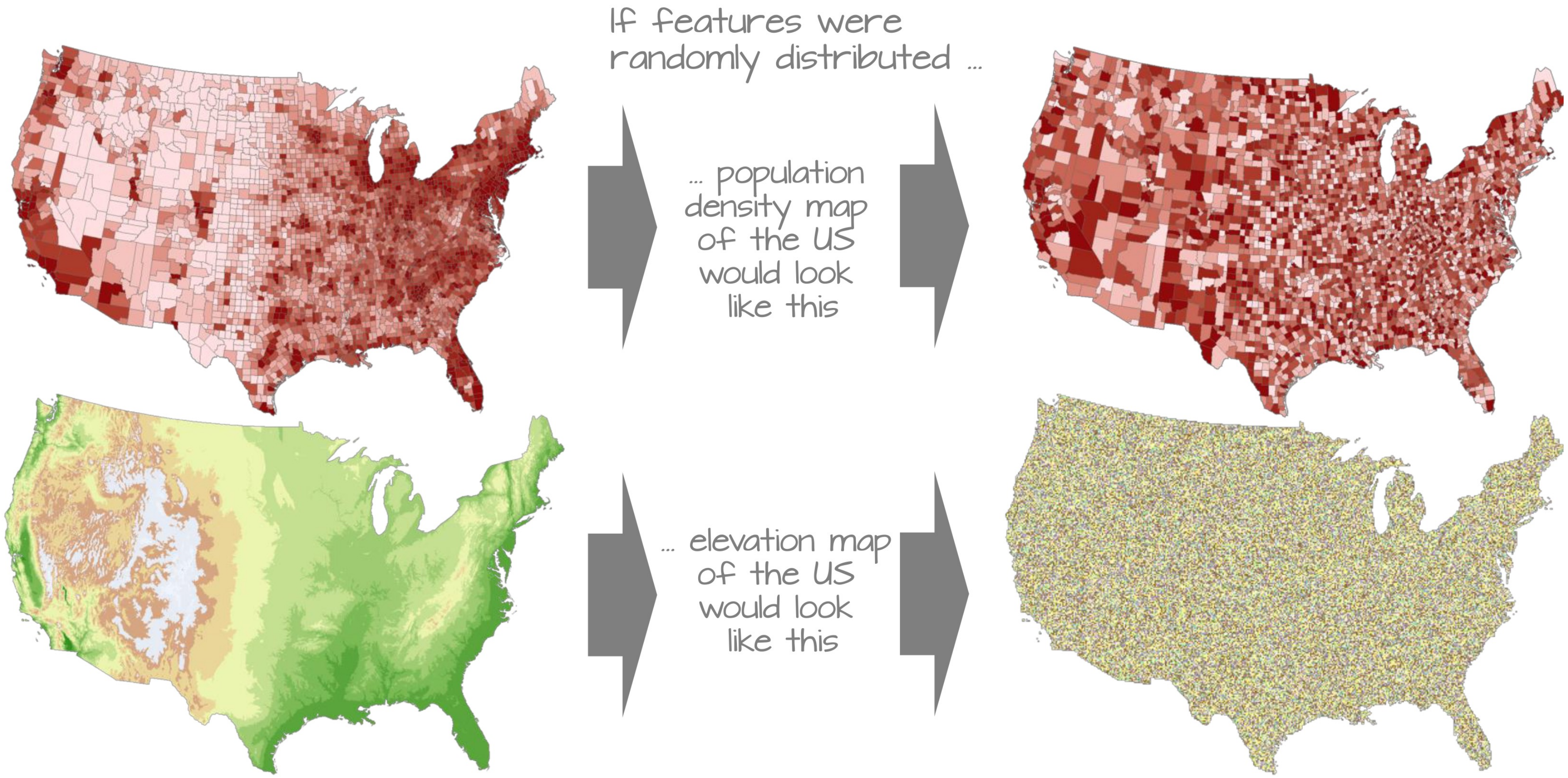
Spatial autocorrelation



Spatial autocorrelation

- Spatial autocorrelation – either positive or negative – implies the absence of Complete Spatial Randomness (CSR).
- CSR means that a pattern is completely made up by chance.
- In most cases, the distribution of attribute values will seldom show evidence of CSR.

Spatial autocorrelation



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

Spatial autocorrelation

Two ways:

1) Global: What is the overall spatial dependence across the entire data set? Studying at a global level will tell you how clustered, dispersed or random the data is distributed over the entire area studied.
2) Local: What is the difference between each unit of analysis and its neighbours? Studying at the local level, you can find areas of greater contrast by seeing if places are quantifiably more like or dislike their neighbours than the average other place.

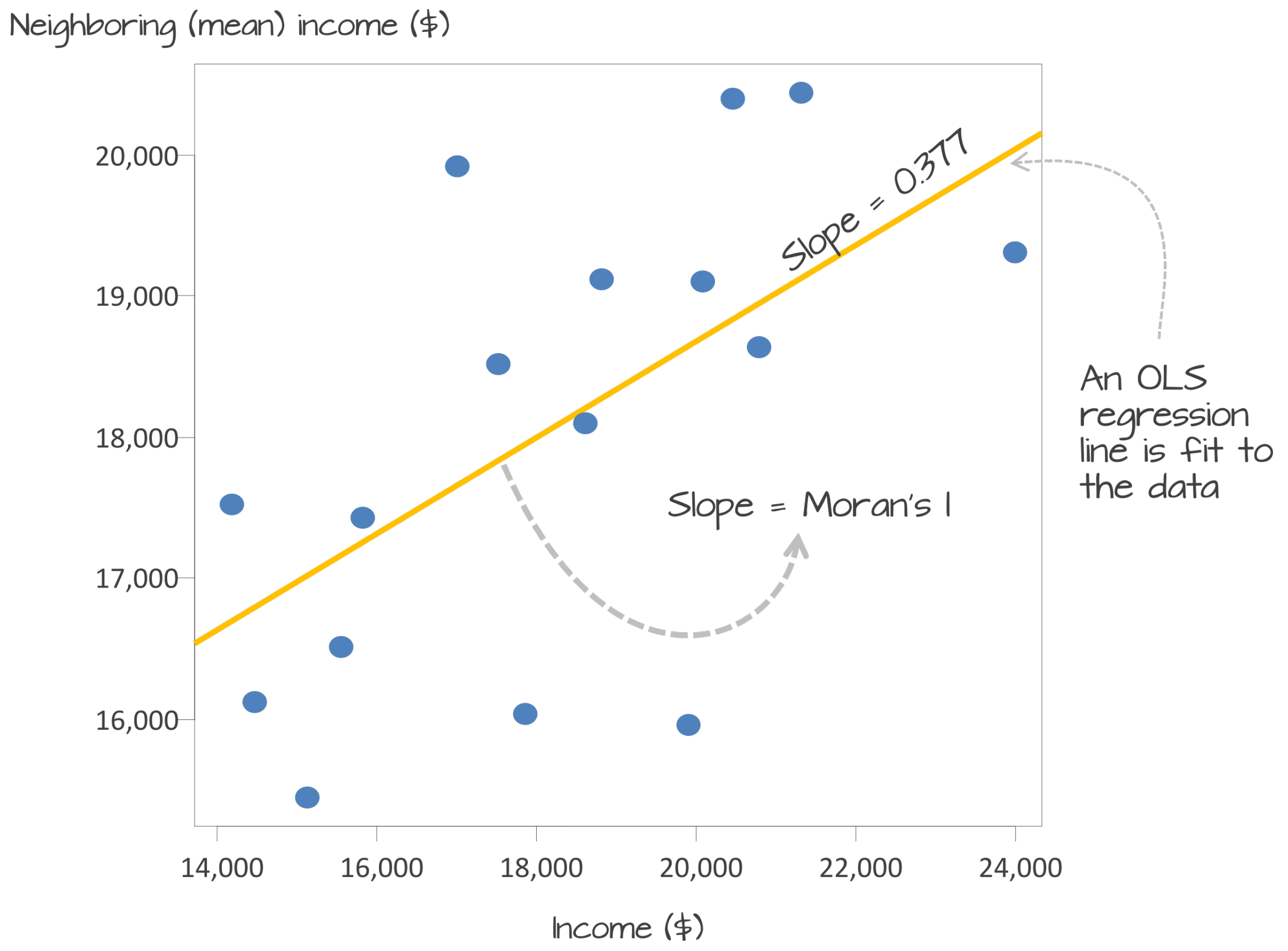
Global spatial autocorrelation

Global Moran’s I

- The most commonly used indicator of global spatial autocorrelation.
- Works through identifying neighbours for each target feature (e.g. polygon) and summarising their values by computing their means (spatially lagged variable value).
- We can then plot the target feature’s value against the spatially lagged mean value, repeat this for all features in your data set, and fit a linear model.
- The resulting β estimate is what is called your Moran’s I statistic.

Global Moran’s I

Global Moran's I



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

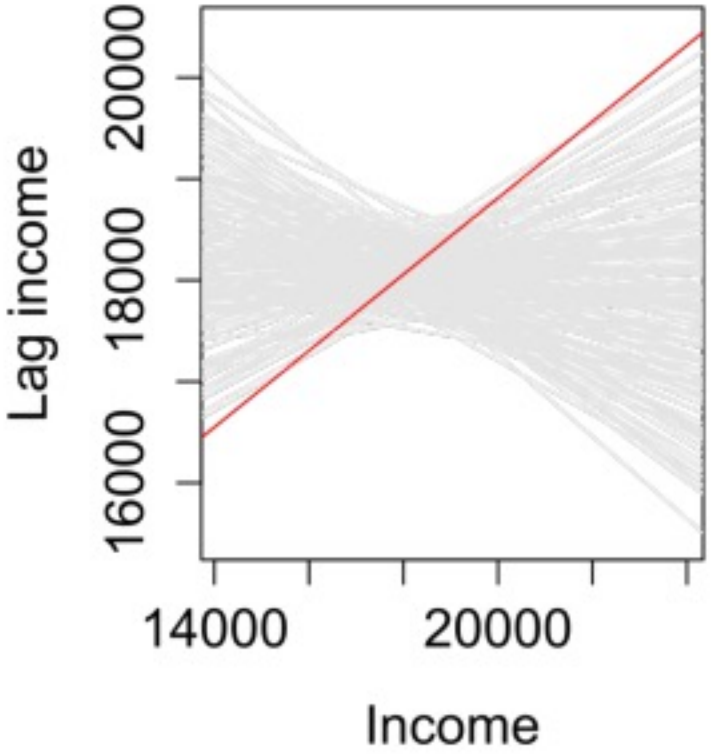
Global Moran’s I

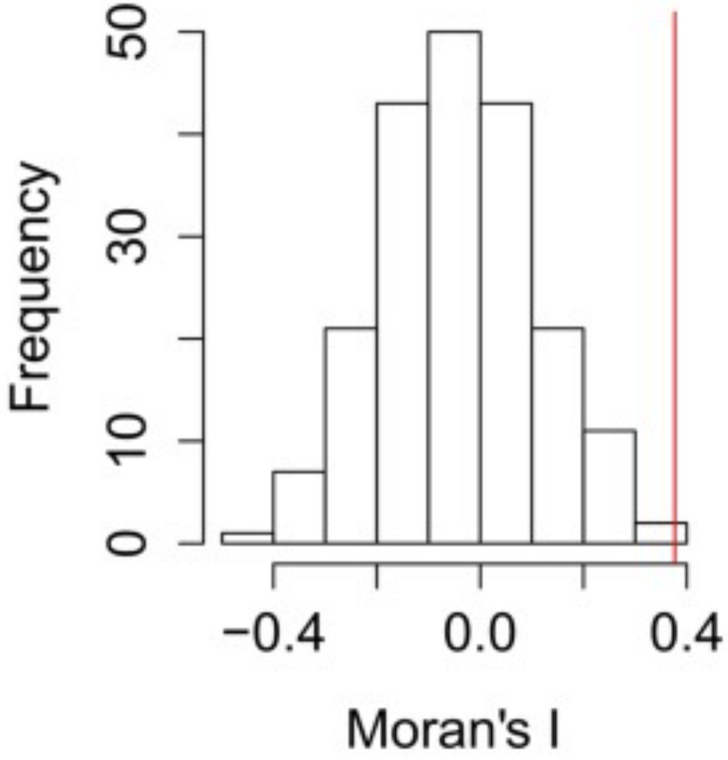
- Output is the global Moran’s I statistic, which is the slope of the linear fit of a variable (attribute) and its surrounding values.
- The statistic evaluates whether the pattern is clustered, dispersed or random.
- How to generate the significance of the relationship?

Global 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.

Global Moran's I





Global 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 spatial autocorrelation

- Looking for areas that are significantly different and deviate from the overall pattern: portions of a map where values are correlated in a particularly strong and specific way.
- Local Indicators of Spatial Association (LISA)
- Compares the observed map with many randomly generated ones to see how likely it is to obtain the observed associations for each location (inference).
- Several measures: Local Moran’s I, Getis-Ord Gi*

Local Moran's I

- A localised measure of autocorrelation. More commonly known as cluster and outlier analysis.
- The Local Moran's I tries to determine for each feature whether:
    - if the neighborhood is significantly different from the study area; *and*
    - if each feature is significantly different from its neighborhood.
- Four types: high-high, low-low, but also outliers: high-low, low-high.

Getis-Ord Gi*

- Where do “high” and “low” values cluster in space.
- More commonly known as hot-spot analysis – finding statistically significant clusters of high and low attribute values.
- The Getis-Ord Gi* tries to determine for each feature whether:
  - each feature's neighborhood is significantly different from the study area
  - If yes, then the feature is categorised as being part of a hot spot (if significantly higher) or cold spot (if significantly lower).

To keep in mind

- Spatial autocorrelation is all about the values of some feature and whether there is spatial patterning (clustering) for these values – and whether they show some form of spatial patterning.
- For point or event data this means that you will need some form of aggregation (e.g. counts of the event within a grid or within an administrative boundary) to run these measures.

Spatial weight matrix

Spatial weight matrix

- For a statistical method to be explicitly spatial, it needs to contain some representation of the spatial context (with spatial context being: neighbours!).
- How do we capture space so that we can conduct statistical analysis? Spatial Weight Matrix.
- `N x N` positive matrix (`W`) that summarises all spatial relations between all features: formal expression of spatial dependency between observations.

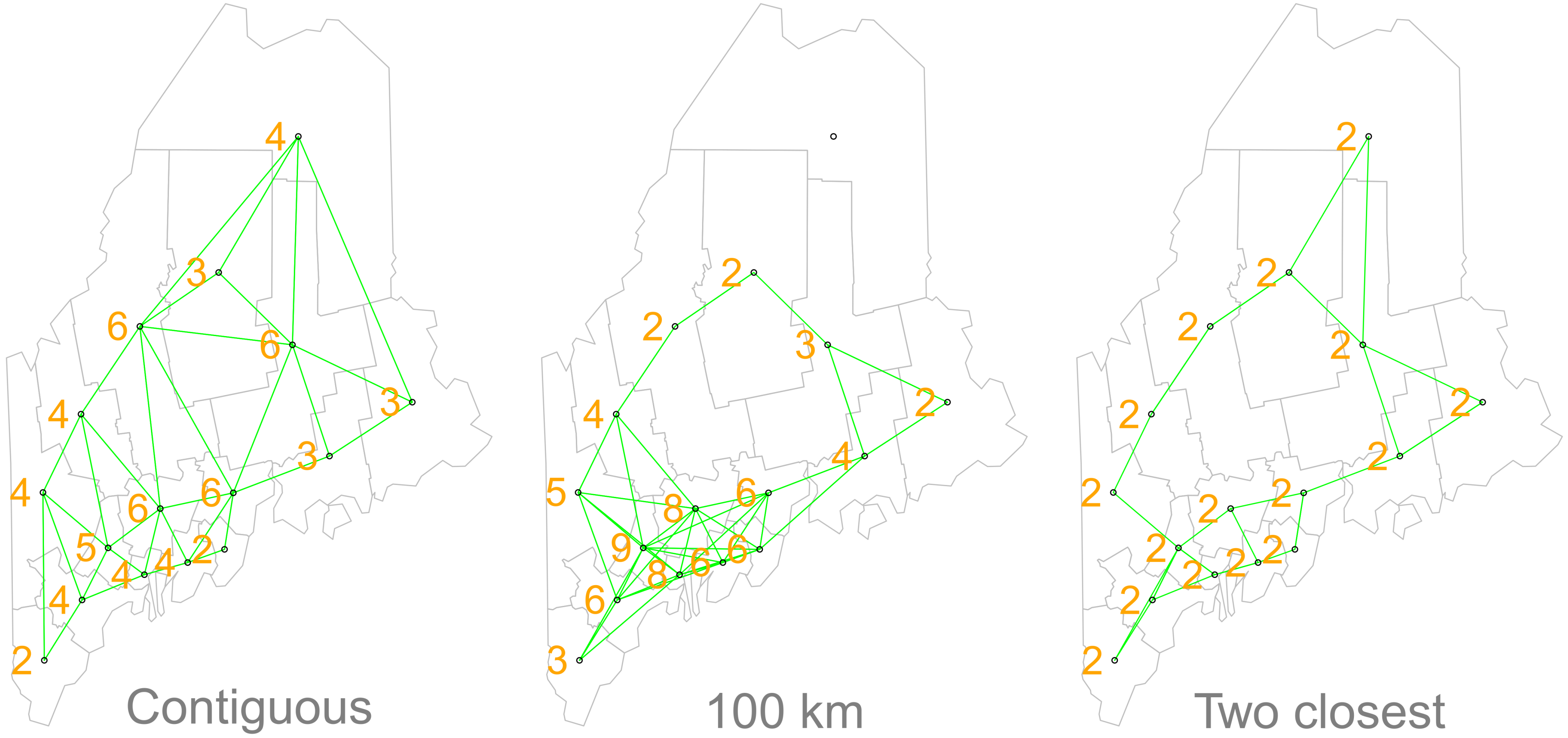
Defining neighbours

- To calculate any measure of spatial autocorrelation, we need to understand how our spatial units relate to each other as neighbours, i.e. how do we conceive their spatial relationship with one another.
- There are two approaches to defining neighbours: through proximity or through contiguity.

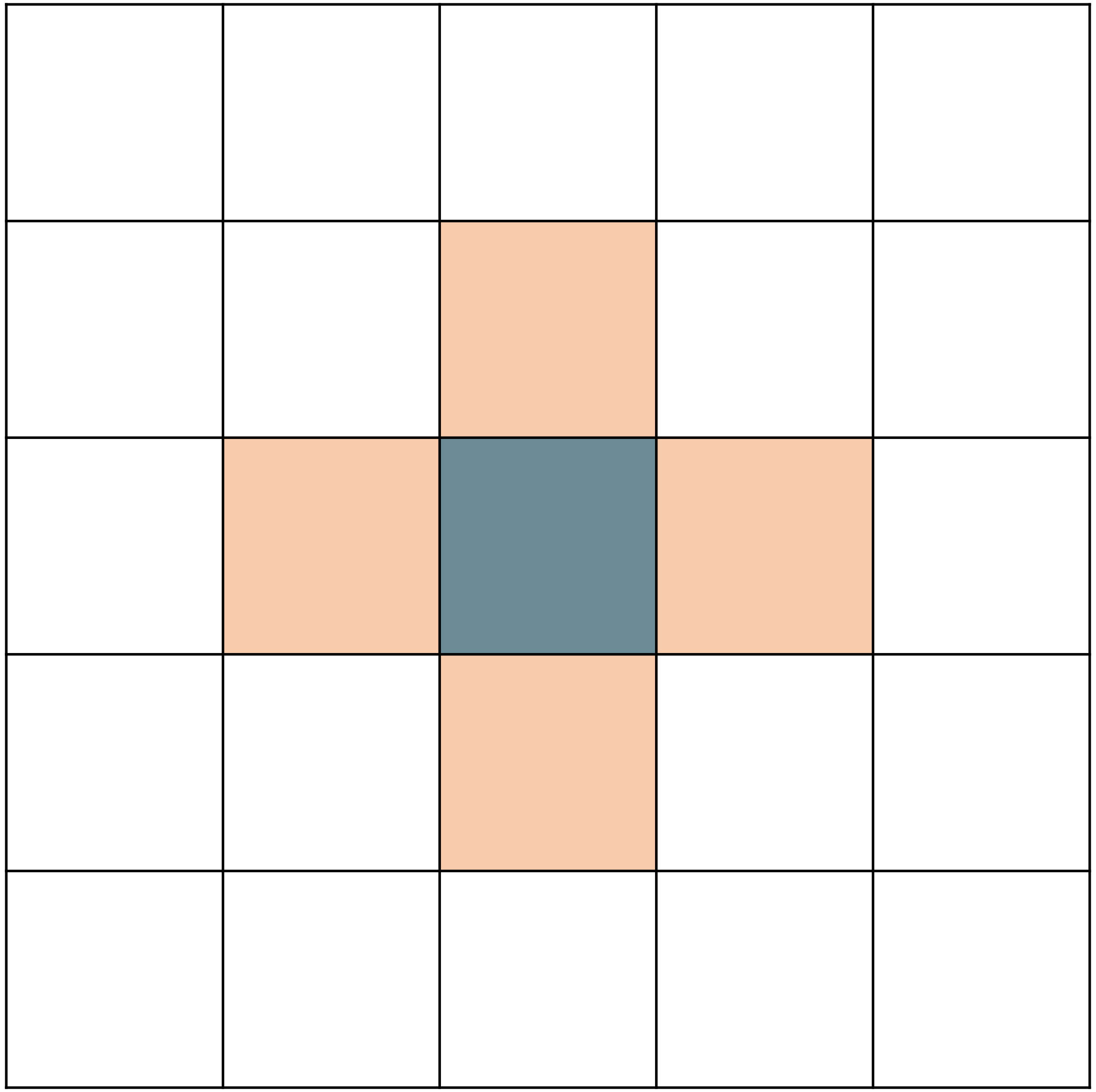
Defining neighbours

- Distance (proximity): units that are within a specific distance of one another, which itself can be defined in numerous ways: Fixed Distance; Inverse Distance, Travel Time, K Nearest Neighbours.
- Contiguity: units that are spatially next to one another and connect to one another via the same side (edge), vertex or both.

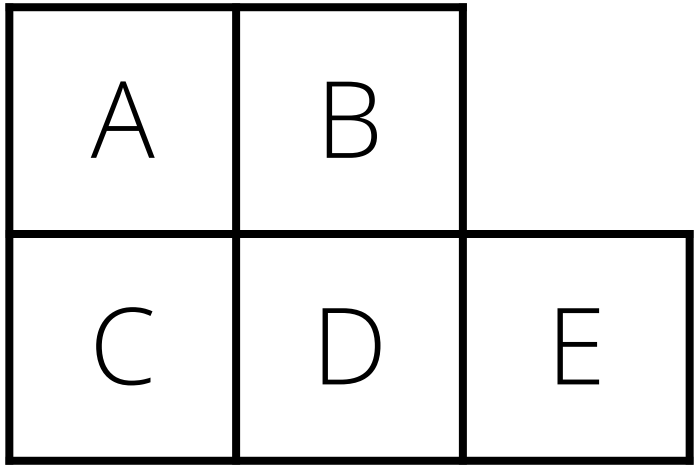
Different conceptualisations, different neighbours



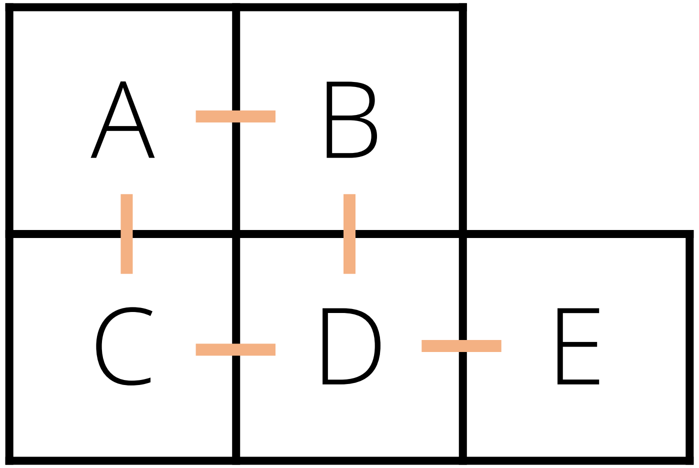
Rook



Spatial weights matrix

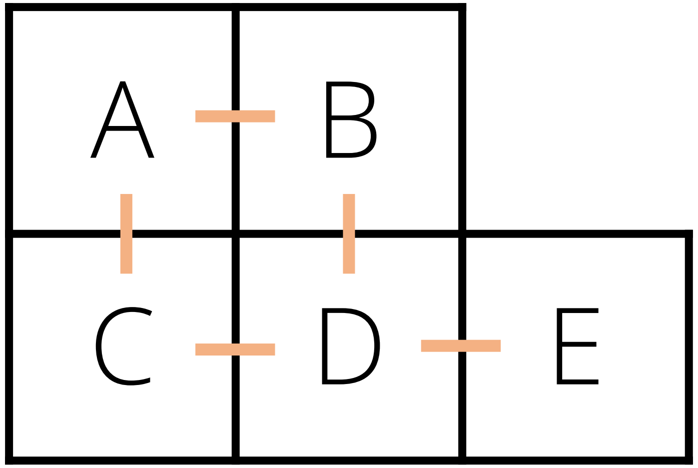


Spatial weights matrix

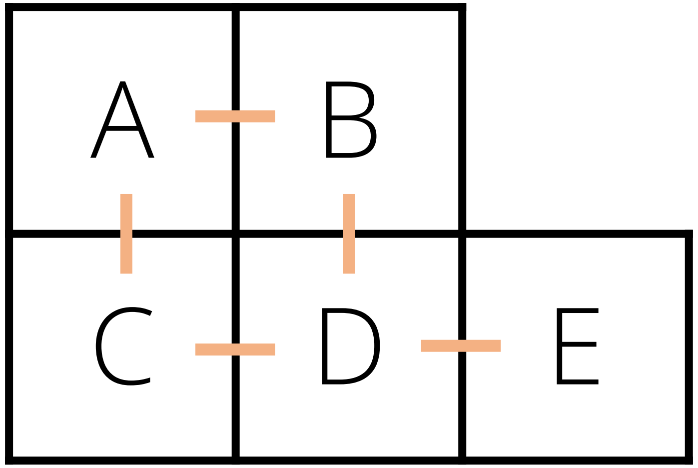


Spatial weights matrix

	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

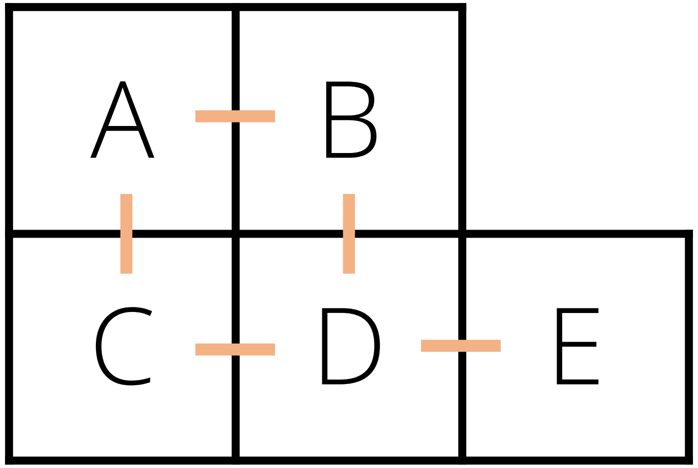


	a	b	c	d	e
a	0	1	1	0	0
b	1	0	0	1	0
c	1	0	0	1	0
d	0	1	1	0	1
e	0	0	0	1	0

=

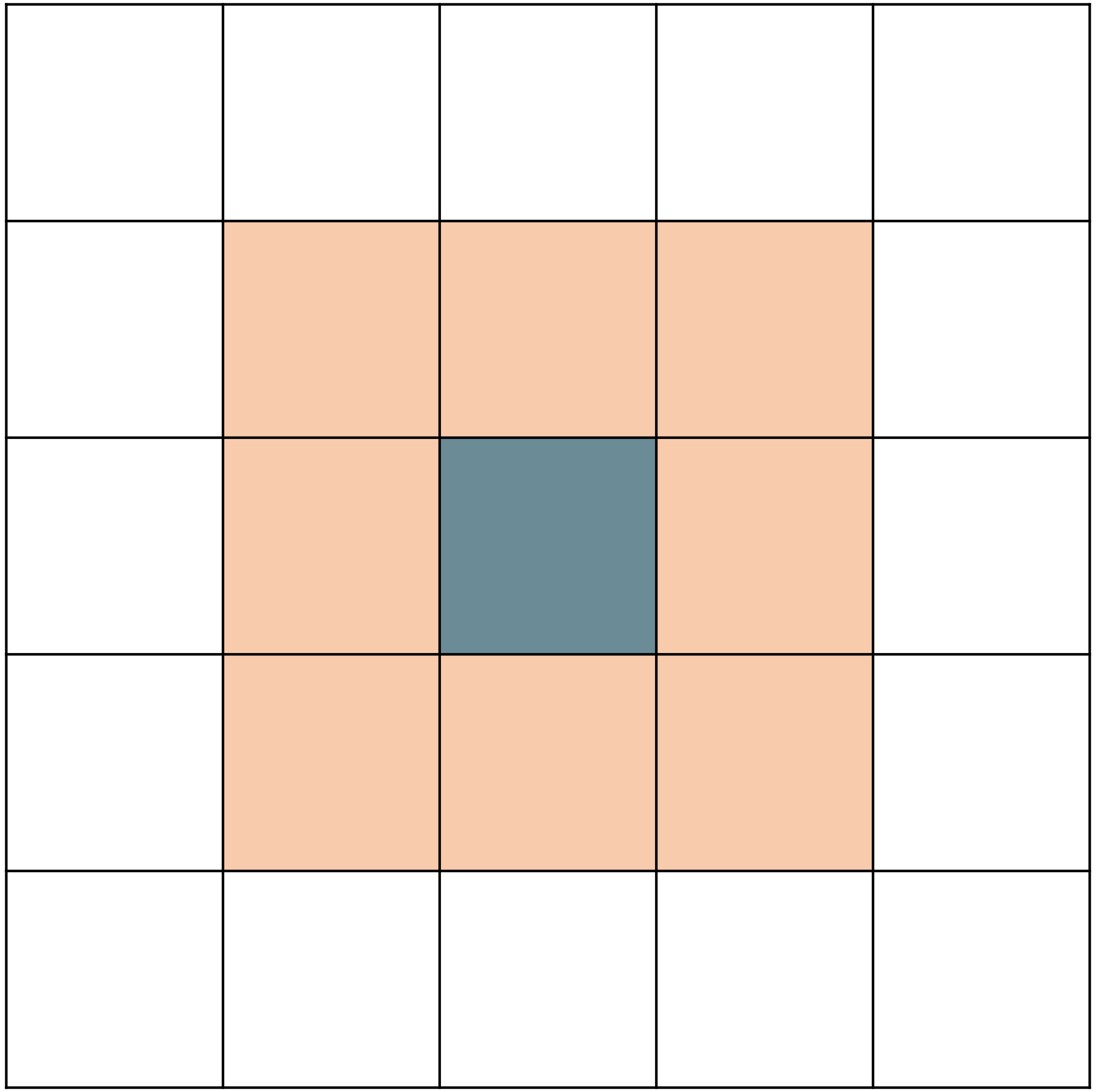
w
2
2
2
3
1

Spatial weights matrix

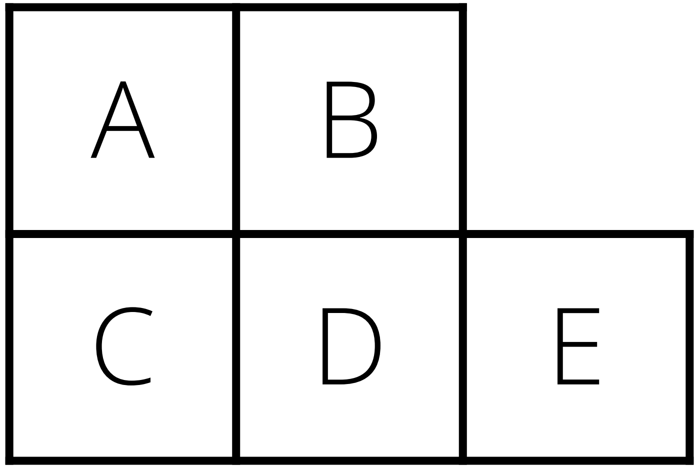


$$
\begin{array}{c|ccccc}
 & a & b & c & d & e \\
\hline
a & 0 & .5 & .5 & 0 & 0 \\
b & .5 & 0 & 0 & .5 & 0 \\
c & .5 & 0 & 0 & .5 & 0 \\
d & 0 & .33 & .33 & 0 & .33 \\
e & 0 & 0 & 0 & 1 & 0
\end{array}
=
\begin{array}{c}
w \\
\hline
2 \\
2 \\
2 \\
3 \\
1
\end{array}
$$

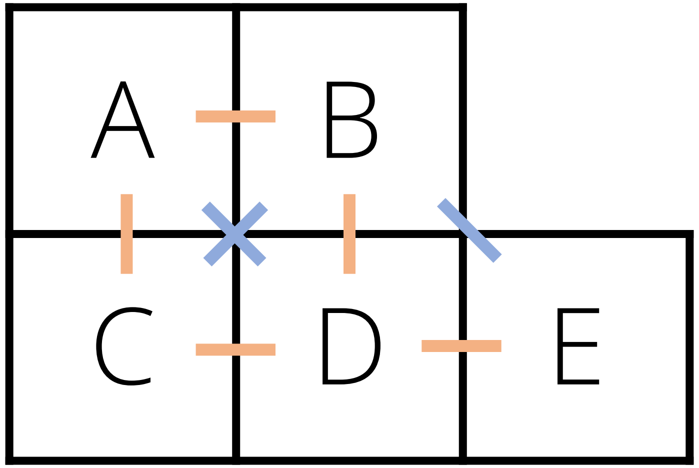
Queen



Spatial weights matrix

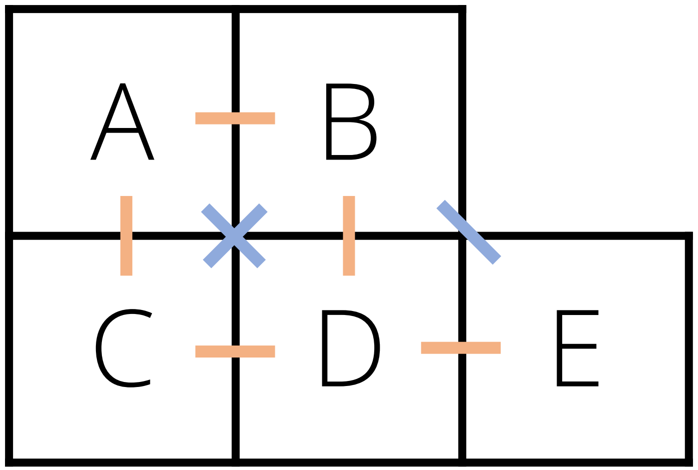


Spatial weights matrix

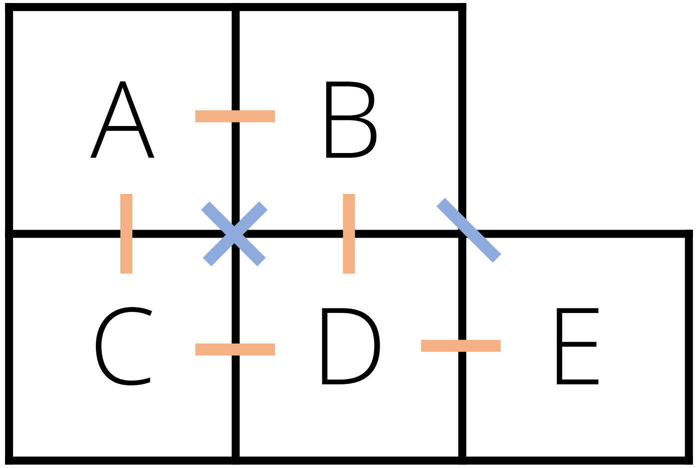


Spatial weights matrix

	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



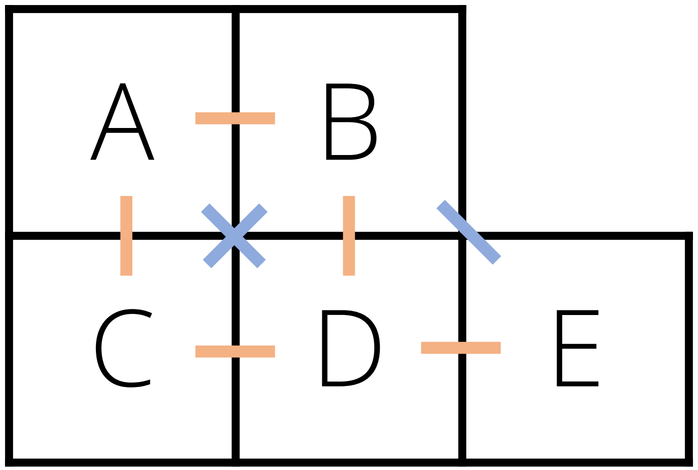
	a	b	c	d	e		w
a	0	1	1	1	0		3
b	1	0	1	1	1		4
c	1	1	0	1	0	=	3
d	1	1	1	0	1		4
e	0	1	0	1	0		2

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 important for understanding spatial relationships.
- We discussed three measures – but there are others.
- We need a spatial weights matrix to calculate the spatially lagged variable that is used in these measures.
- The specification of the neighbourhood can impact the results of these tests.
- Presence of spatial autocorrelation has implications for analysis: spatial error model, spatial lagged model, Geographically Weighted Regression.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk