Geographic Data Science

Exploratory spatial data analysis

Carmen Cabrera and Elisabetta Pietrostefani

ESDA

Exploratory spatial data analysis

E Exploratory

S Spatial

D Data

A Analysis

What ESDA involves

Exploratory

Discovery and investigation without fixed assumptions.

Spatial

Patterns and processes with space and geography at their core.

Data analysis

Statistical techniques.

Questions for ESDA

Questions it helps answer

- Is the variable concentrated over space?
- Do similar values tend to occur close together?
- Where are particular values clustered?

Questions it helps us ask

- What is behind the pattern? What process could generate it?
- Why do we observe particular spatial clusters?

Geographic Data Science

Spatial autocorrelation

Spatial autocorrelation

Tobler’s first law of geography

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

Waldo Tobler (1970)

Spatial autocorrelation

- A statistical representation of Tobler's law
- The spatial counterpart of traditional correlation

The degree to which similar values occur in similar locations.

Positive

Similar values occur close together.

Negative

Similar values occur further apart.

Examples

Positive spatial autocorrelation

Income, poverty, vegetation, temperature…

Negative spatial autocorrelation

Supermarkets, police stations, fire stations, hospitals…

Scales of spatial autocorrelation

Global: clustering

Do values tend to be close to other similar or dissimilar values?

Local: clusters

Are there specific areas with an unusually high concentration of similar or dissimilar values?

Geographic Data Science

Global spatial autocorrelation

Global spatial autocorrelation

Global spatial autocorrelation

Clustering: an overall spatial pattern in the distribution of values.

Positive

Similar values close together: high-high or low-low.

Negative

Similar values further apart, with high-low neighbours.

How can we measure it?

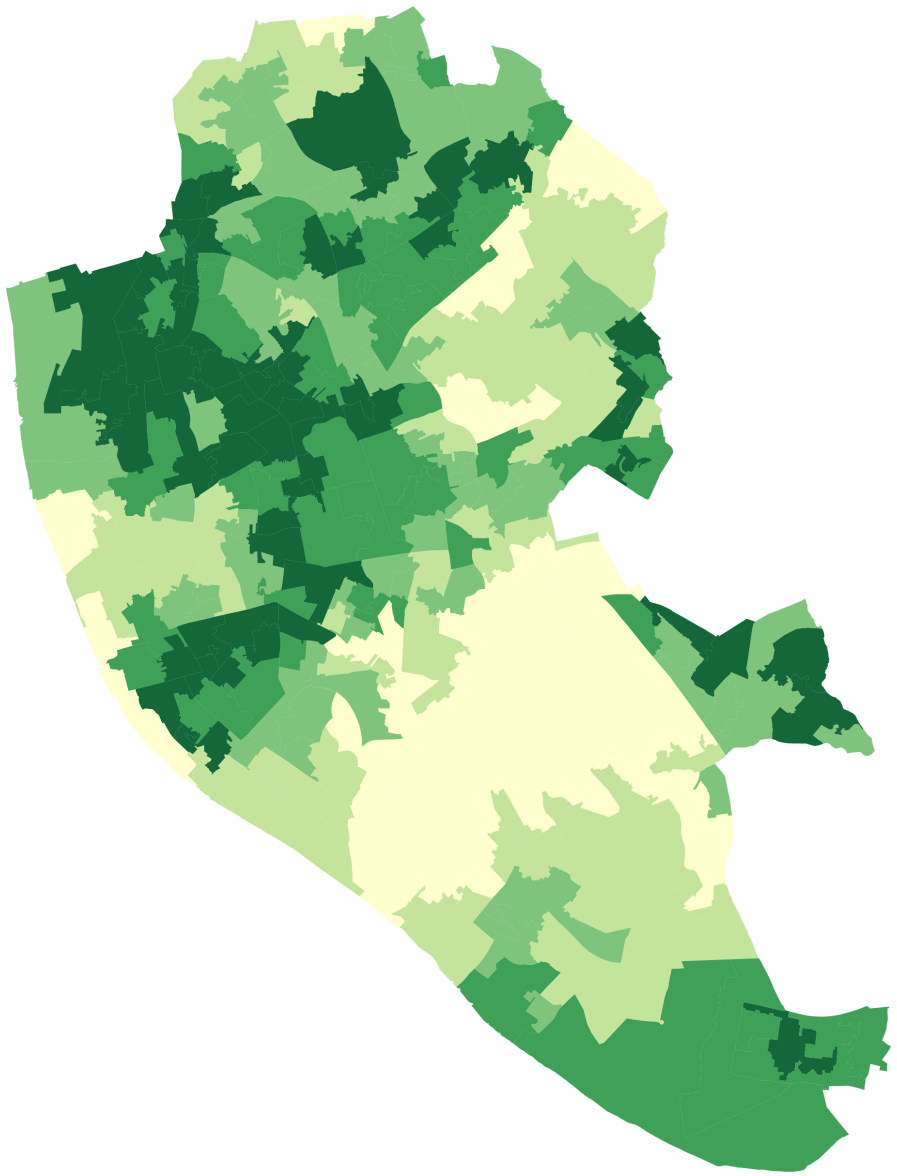
The Moran plot

- Horizontal axis: a variable
- Vertical axis: its spatial lag
- Preferably use a standardised variable and spatial weights matrix

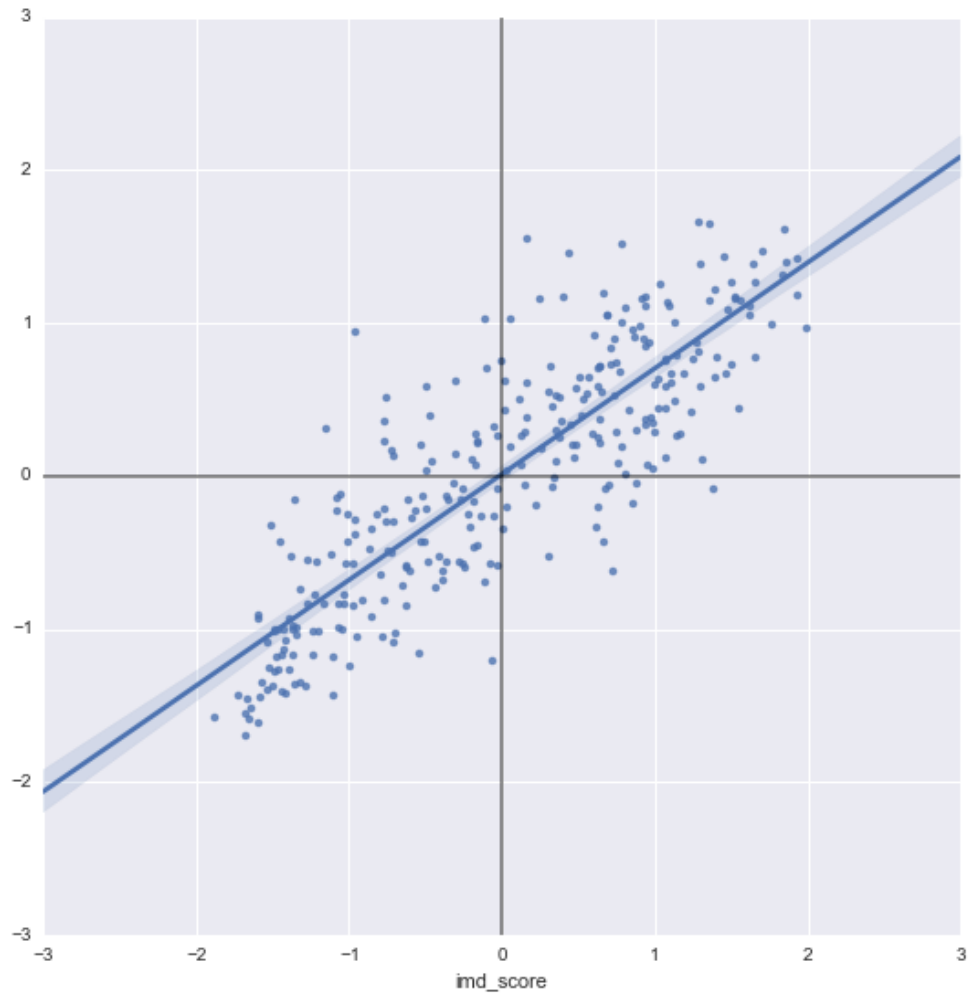
Shows the overall association between values at each location and values in its neighbourhood.

The Moran plot: example

Spatial pattern



Variable and spatial lag



Moran's *I*

- A formal test of global spatial autocorrelation
- Identifies clustering in a variable
- The slope of the Moran plot for standardised data and weights

Inference asks how likely the observed pattern would be under spatial randomness.

Geographic Data Science

Local spatial autocorrelation

Local spatial autocorrelation

Local spatial autocorrelation

Clusters: pockets of spatial instability where values show particularly strong, specific associations.

High-high	Positive association: high values (hotspots)
Low-low	Positive association: low values (coldspots)
High-low	Negative association: spatial outliers
Low-high	Negative association: spatial outliers

LISAs

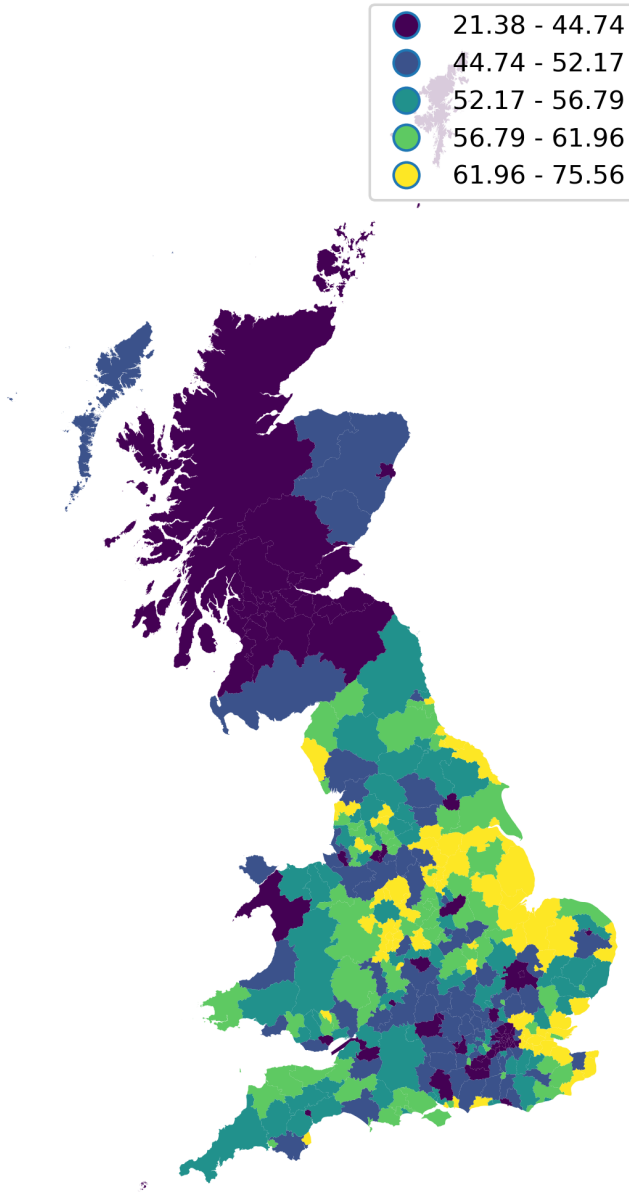
Local Indicators of Spatial Association

- Statistical tests for detecting spatial clusters
- Assess statistical significance at each location

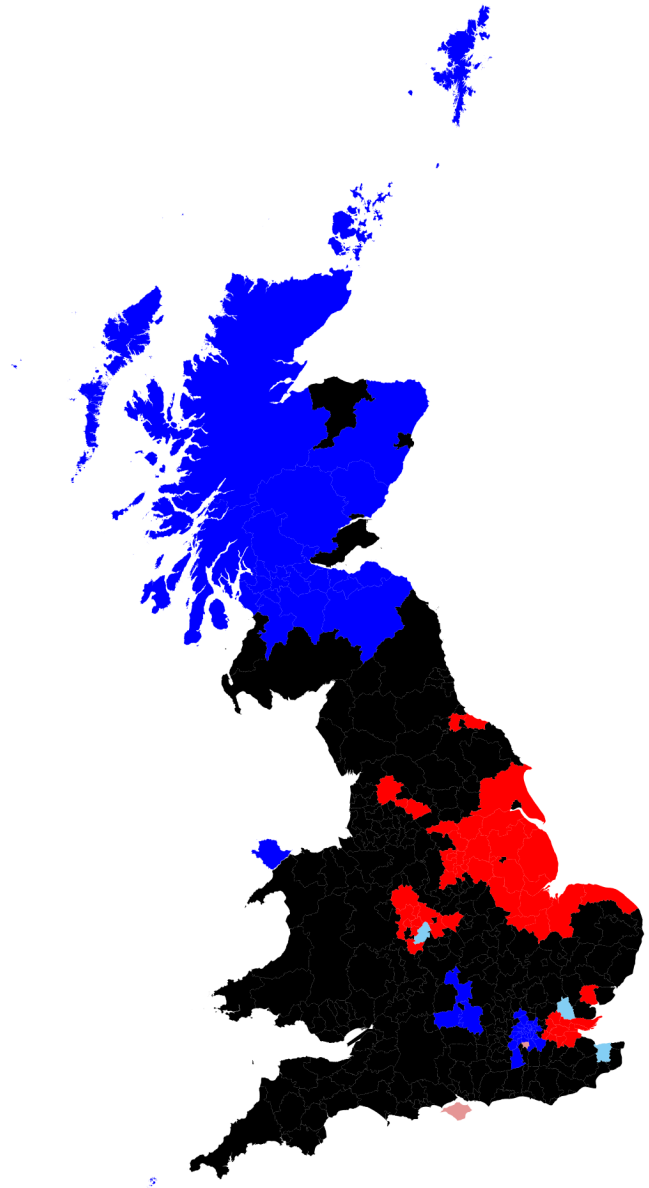
Compare the observed map with many randomly generated maps to assess how likely each local association is.

LISAs: Brexit vote example

Leave vote (%)



Local spatial association



Attribution and licence

These slides were originally created by
Dani Arribas-Bel for a course in Geographic Data Science.

Creative Commons Attribution-ShareAlike 4.0 International Licence

