# Geographic Data Science - Lecture VI

## Exploring Space in Data

Dani Arribas-Bel

# Today

- ESDA
- Spatial Autocorrelation
- Measures
  - Global
  - Local

ESDA

**E**xploratory

**S**patial

**D**ata

**A**nalysis

**[Exploratory]**

Focus on discovery and assumption-free investigation

**[Spatial]**

Patterns and processes that put *space* and *geography* at the core

**[Data Analysis]**

Statistical techniques

Questions that **ESDA** helps...

**Answer**

- *Is the variable I'm looking at concentrated over space? Do similar values tend to locate closeby?*
- *Can I identify any particular areas where certain values are clustered?*

Questions that **ESDA** helps...

**Answer**

- *Is the variable I'm looking at concentrated over space? Do similar values tend to locate closeby?*
- *Can I identify any particular areas where certain values are clustered?*

**Ask**

- *What is behind this pattern? What could be generating the process?*
- *Why do we observe certain clusters over space?*

# Spatial Autocorrelation

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

Waldo Tobler (1970)

# Spatial Autocorrelation

-Statistical representation of Tobler's law

-Spatial counterpart of traditional correlation

# Spatial Autocorrelation

-Statistical representation of Tobler's law

-Spatial counterpart of traditional correlation

***Degree to which similar values are located in similar locations***

# Spatial Autocorrelation

-Statistical representation of Tobler's law

-Spatial counterpart of traditional correlation

***Degree to which similar values are located in similar locations***

Two flavors:

- **Positive**: similar values → similar location (*closeby*)
- **Negative**: similar values → disimilar location (*further apart*)

# Examples

**Positive** SA:

**Negative** SA:

# Examples

**Positive** SA: income,

**Negative** SA:

# Examples

**Positive** SA: income, poverty,

**Negative** SA:

# Examples

**Positive** SA: income, poverty, vegetation,

**Negative** SA:

# Examples

**Positive** SA: income, poverty, vegetation, temperature…

**Negative** SA:

# Examples

**Positive** SA: income, poverty, vegetation, temperature…

**Negative** SA: supermarkets,

# Examples

**Positive** SA: income, poverty, vegetation, temperature…

**Negative** SA: supermarkets, police stations,

# Examples

**Positive** SA: income, poverty, vegetation, temperature…

**Negative** SA: supermarkets, police stations, fire stations,

# Examples

**Positive** SA: income, poverty, vegetation, temperature…

**Negative** SA: supermarkets, police stations, fire stations, hospitals…

# Scales

**[Global]**

*Clustering*: do values tend to be close to other (dis)similar values?

# Scales

**[Global]**

*Clustering*: do values tend to be close to other (dis)similar values?

**[Local]**

*Clusters*: are there any specific parts of a map with an extraordinary concentration of (dis)similar values?

# **Global** Spatial Autocorr.

# **Global** Spatial Autocorr.

"***Clustering***"

*Overall trend where the distribution of values follows a particular pattern over space*

# **Global** Spatial Autocorr.

"***Clustering***"

*Overall trend where the distribution of values follows a particular pattern over space*

**[Positive]** Similar values close to each other (high-high, low-low)

# **Global** Spatial Autocorr.

"***Clustering***"

*Overall trend where the distribution of values follows a particular pattern over space*

**[Positive]** Similar values close to each other (high-high, low-low)

**[Negative]** Similar values far from each other (high-low)

# **Global** Spatial Autocorr.

"***Clustering***"

*Overall trend where the distribution of values follows a particular pattern over space*

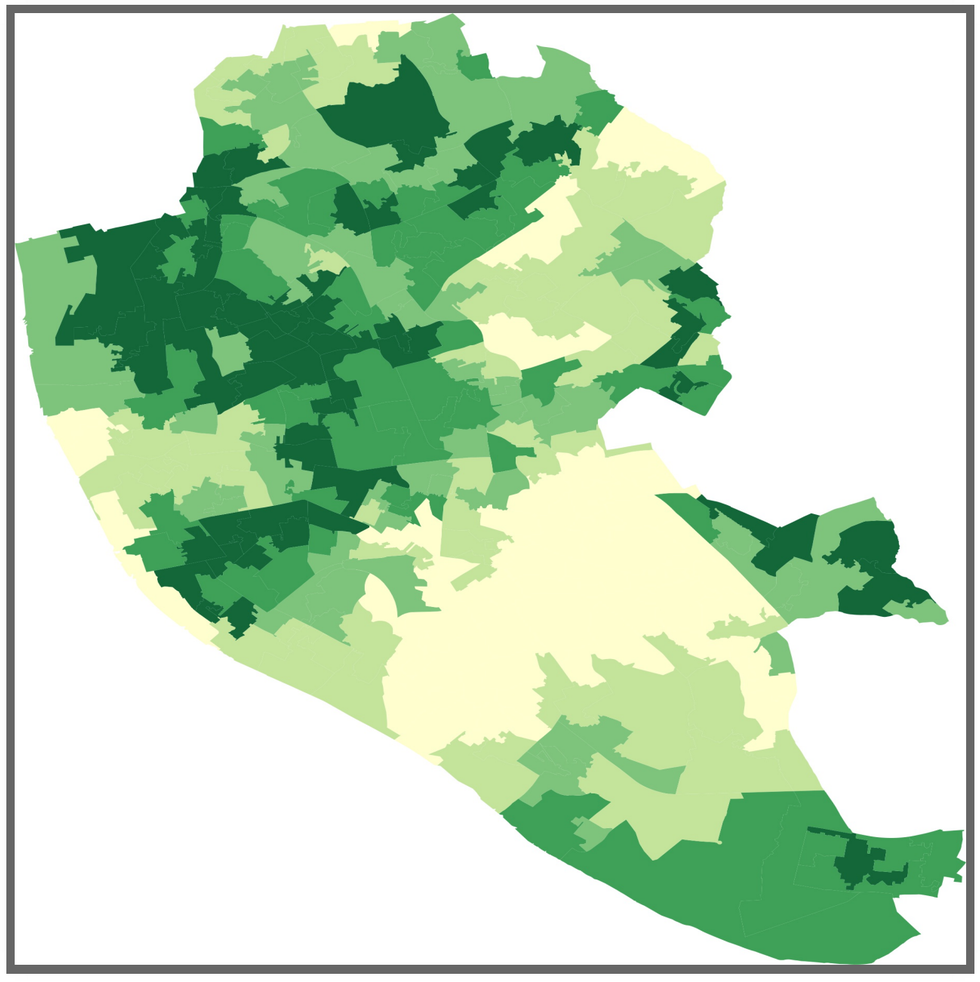
**[Positive]** Similar values close to each other (high-high, low-low)

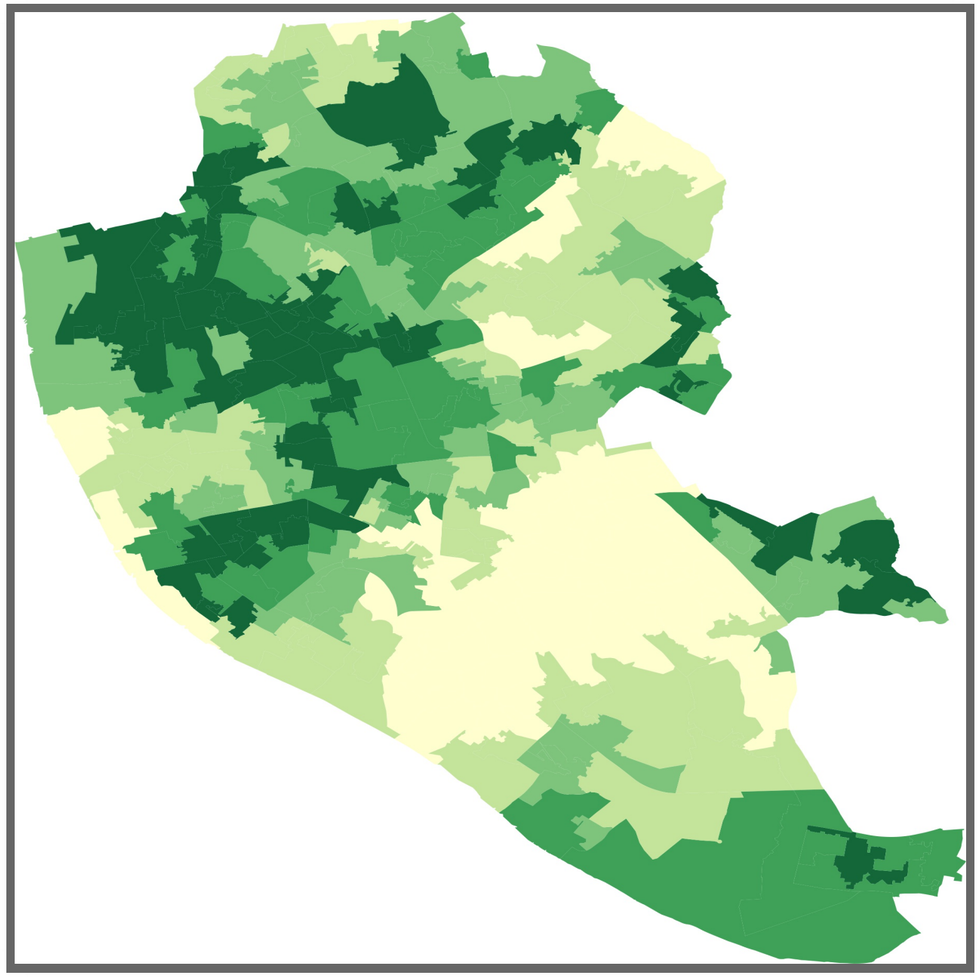
**[Negative]** Similar values far from each other (high-low)

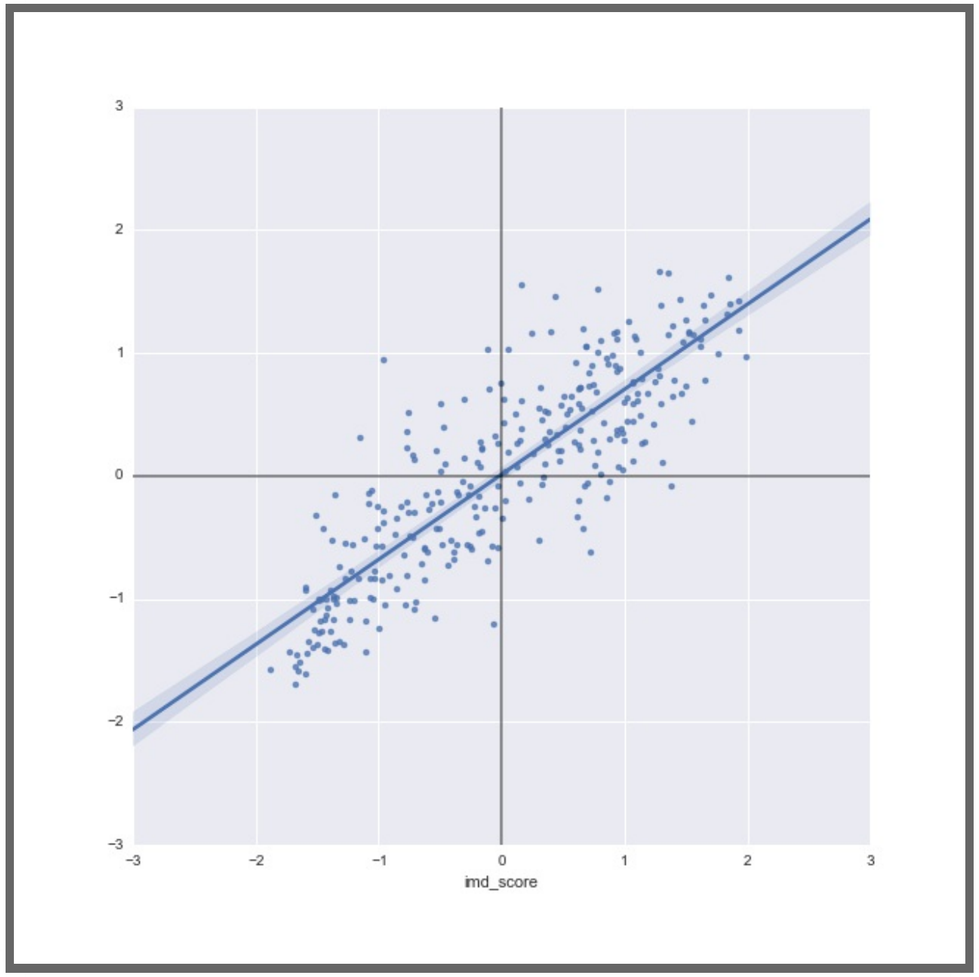
How to measure it???

# Moran Plot

- Graphical device that displays **a variable** on the horizontal axis against **its spatial lag** on the vertical one
- Variable and spatial weights matrix are preferably **standardized**
- Asssessment of the overall association between a variable in a given location and in its *neighborhood*






3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
imd_score

[Interactive Demo]

# Moran's I

**Formal test** of global spatial autocorrelation

**Statistically** identify the presence of clustering in a variable

Slope of the Moran plot

Inference based on how likely it is to obtain a map like observed from a purely random pattern

# Moran's I

**Formal test** of global spatial autocorrelation

**Statistically** identify the presence of clustering in a variable

Slope of the Moran plot

Inference based on how likely it is to obtain a map like observed from a purely random pattern

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(Z_i)(Z_j)}{\sum_i (Z_i)^2}$$

# **Local** Spatial Autocorr.

# **Local** Spatial Autocorr.

*"**Clusters**"*

*Pockets of spatial instability*

Portions of a map where values are correlated in a particularly strong and specific way

# **Local** Spatial Autocorr.

*"**Clusters**"*

*Pockets of spatial instability*

Portions of a map where values are correlated in a particularly strong and specific way

**[High-High]** Positive sp. autocorr. of *high* values (*hotspots*)

# **Local** Spatial Autocorr.

*"**Clusters**"*

*Pockets of spatial instability*

Portions of a map where values are correlated in a particularly strong and specific way

**[High-High]** Positive sp. autocorr. of *high* values (*hotspots*)

**[Low-Low]** Positive sp. autocorr. of *low* values (*coldspots*)

# **Local** Spatial Autocorr.

*"**Clusters**"*

*Pockets of spatial instability*

Portions of a map where values are correlated in a particularly strong and specific way

**[High-High]** Positive sp. autocorr. of *high* values (*hotspots*)

**[Low-Low]** Positive sp. autocorr. of *low* values (*coldspots*)

**[High-Low]** Negative sp. autocorr. (*spatial outliers*)

# **Local** Spatial Autocorr.

*"**Clusters**"*

*Pockets of spatial instability*

Portions of a map where values are correlated in a particularly strong and specific way

**[High-High]** Positive sp. autocorr. of *high* values (*hotspots*)

**[Low-Low]** Positive sp. autocorr. of *low* values (*coldspots*)

**[High-Low]** Negative sp. autocorr. (*spatial outliers*)

**[Low-High]** Negative sp. autocorr. (*spatial outliers*)

# **Local** Spatial Autocorr.

*"**Clusters**"*

*Pockets of spatial instability*

Portions of a map where values are correlated in a particularly strong and specific way

**[High-High]** Positive sp. autocorr. of *high* values (*hotspots*)

**[Low-Low]** Positive sp. autocorr. of *low* values (*coldspots*)

**[High-Low]** Negative sp. autocorr. (*spatial outliers*)

**[Low-High]** Negative sp. autocorr. (*spatial outliers*)

How to measure it???

# LISAs

**L**ocal **I**ndicators of **S**patial **A**ssociation

Statistical tests for **spatial cluster detection** → Statistical significance

**Compares** the **observed** map with many **random**ly generated ones to see how likely it is to obtain the observed associations **for each location**

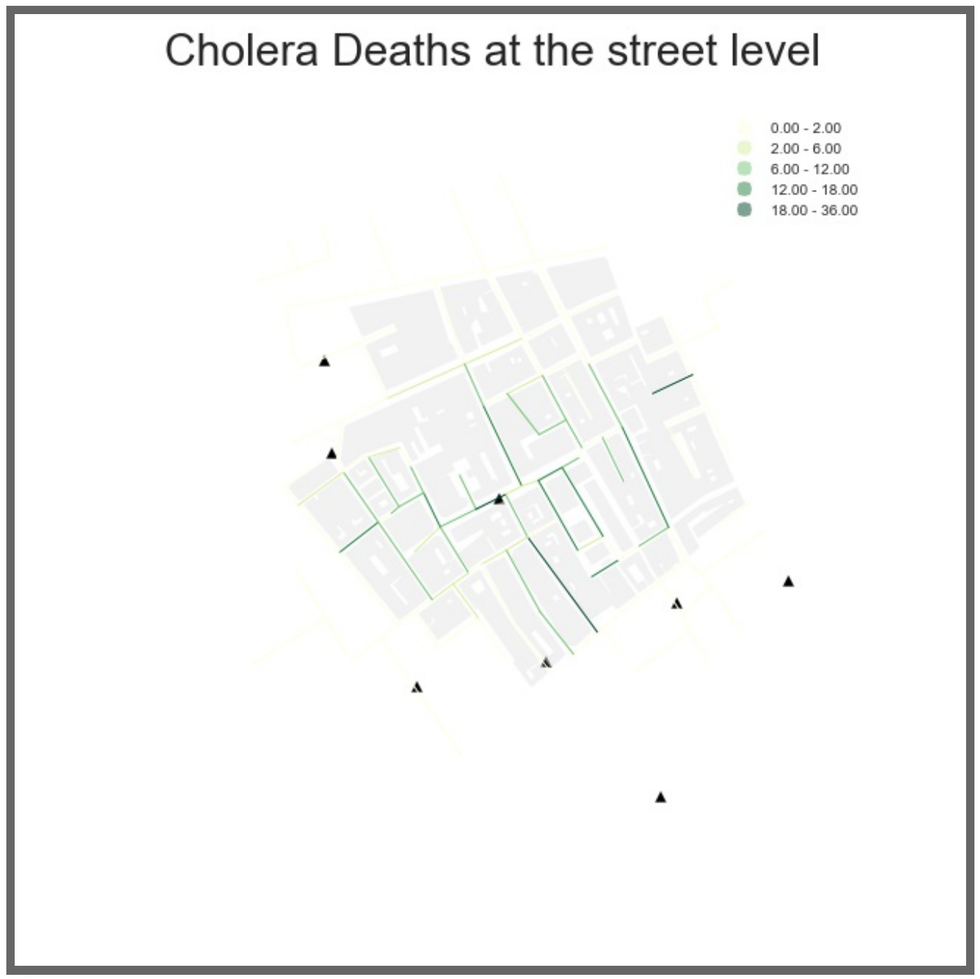
# LISAs

**L**ocal **I**ndicators of **S**patial **A**ssociation

Statistical tests for **spatial cluster detection** → Statistical significance

**Compares** the **observed** map with many **random**ly generated ones to see how likely it is to obtain the observed associations **for each location**

$$I_i = \frac{Z_i}{m_2} \sum_j W_{ij} Z_j \quad ; \quad m_2 = \frac{\sum_i Z_i^2}{N}$$


Cholera Deaths at the street level
0.00 - 2.00
2.00 - 6.00
6.00 - 12.00
12.00 - 18.00
18.00 - 36.00

## Cholera Deaths at the street level


## LISA for Cholera Deaths

# Recapitulation

**ESDA** is a family of techniques to explore and spatially interrogate data

Main function: characterize **spatial autocorrelation**, which can be explored:

- **Global**ly (e.g. Moran Plot, Moran's I)
- **Local**ly (e.g. LISAs)