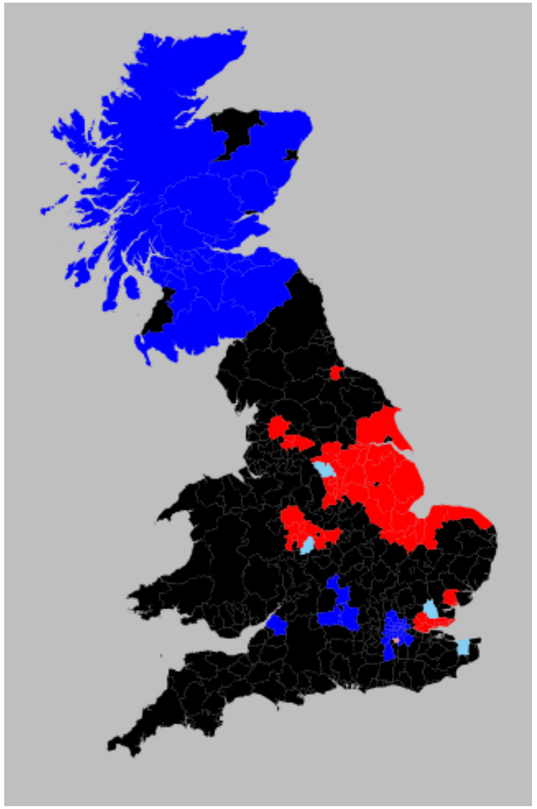
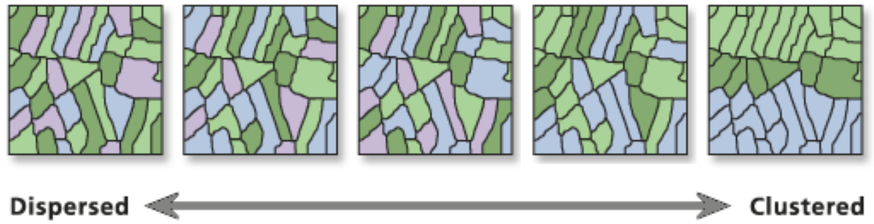
Lecture 5: Spatial autocorrelation

Instructor: Michael Szell
Mar 3, 2022

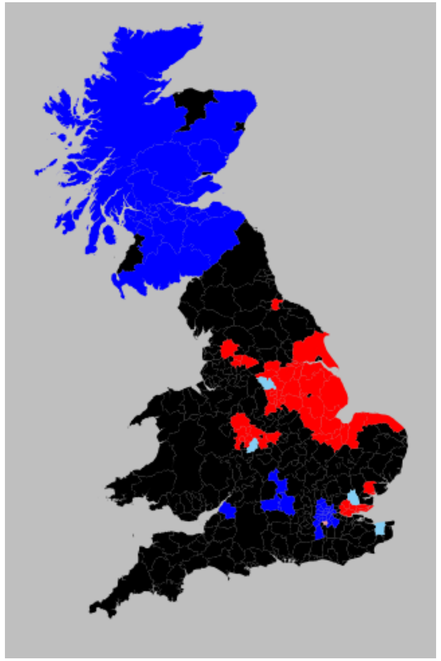


Today you will learn about spatial correlation

Questions to ask in Exploratory spatial data analysis



Use in Python with example



Global and local measures

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}$$

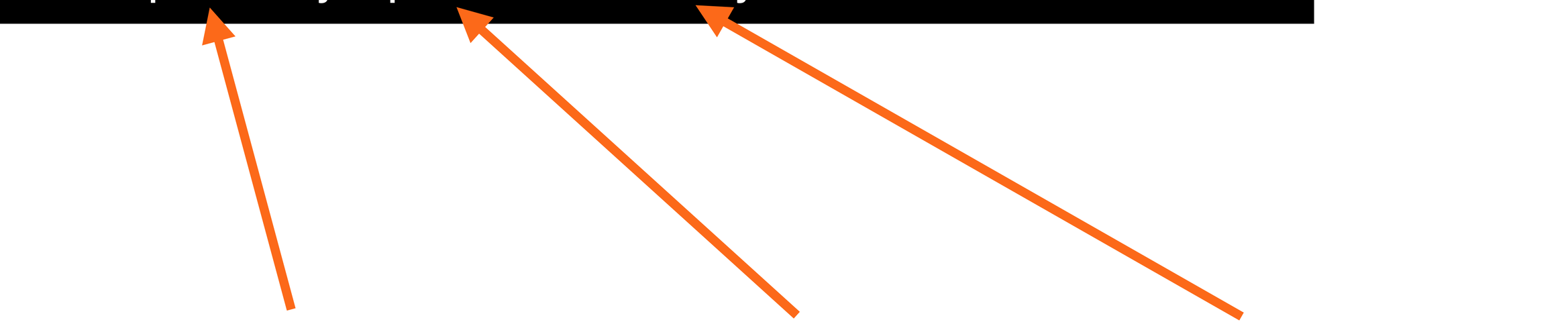
$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j \; ; \; m_2 = \frac{\sum_i z_i^2}{n}$$

Exam project info

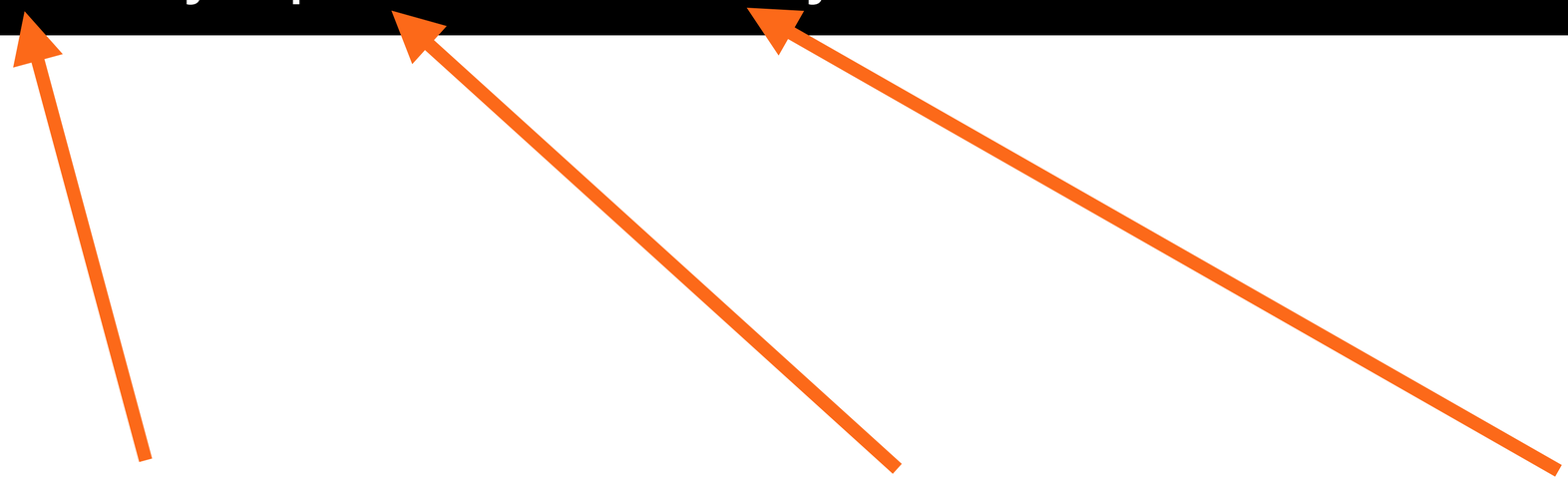
GDS asks: How do things relate in space?

> Everything is related to everything else, but near things are more related than distant things.
>
> Tobler's 1st law of geography

ESDA: Exploratory Spatial Data Analysis



ESDA: Exploratory Spatial Data Analysis



Assumption-free investigation of spatial patterns with statistical techniques

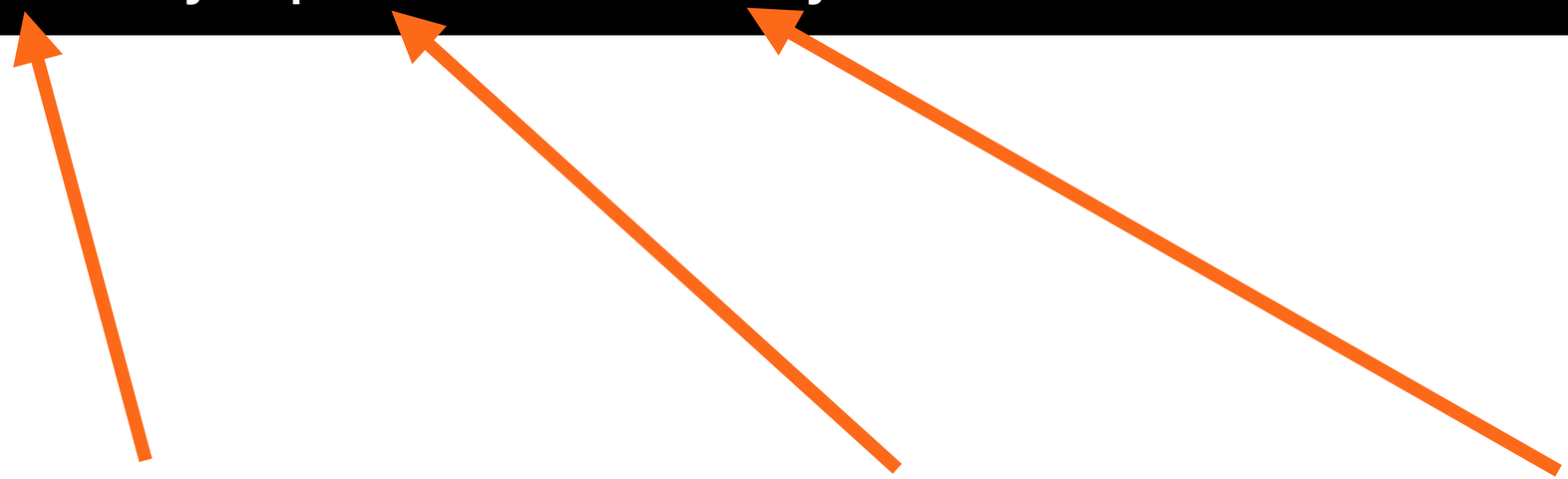
Answer:

Is a variable concentrated over space?

Do similar values locate closeby? If so where?

Given a choropleth map, is there a significant statistical pattern?

ESDA: Exploratory Spatial Data Analysis



Assumption-free investigation of spatial patterns with statistical techniques

Answer:

Is a variable concentrated over space?

Do similar values locate closeby? If so where?

Given a choropleth map, is there a significant statistical pattern?

Ask:

What process is behind a pattern?

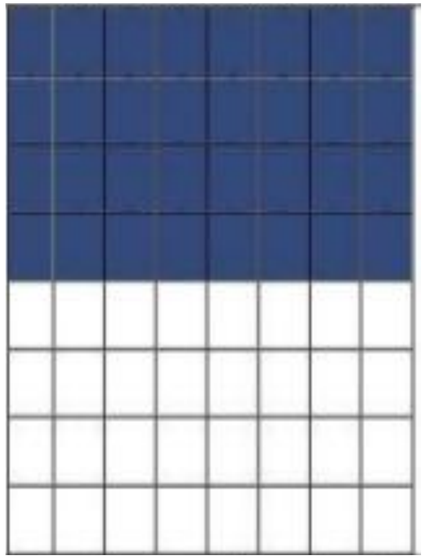
Why are there certain clusters?

Spatial autocorrelation: Positive vs. Negative

Is the spatial counterpart of traditional correlation

Positive

similar values are closeby



Negative

similar values are further apart

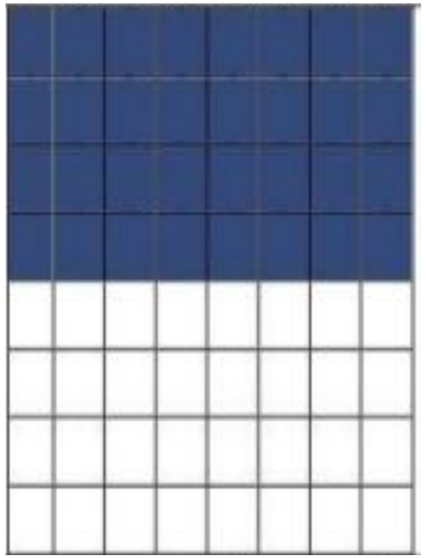
?

Spatial autocorrelation: Positive vs. Negative

Is the spatial counterpart of traditional correlation

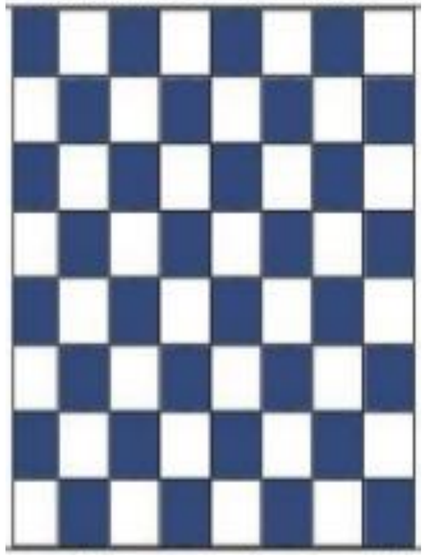
Positive

similar values are closeby



Negative

similar values are further apart

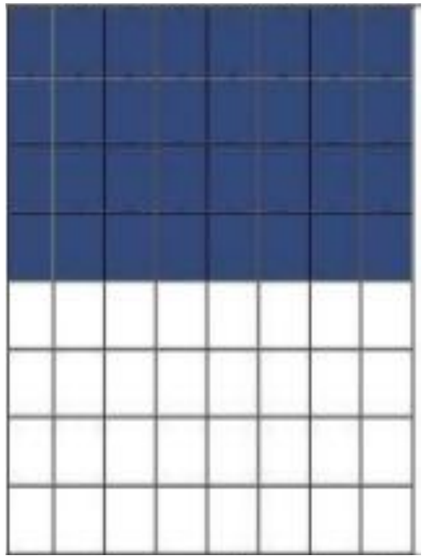


Spatial autocorrelation: Positive vs. Negative

Is the spatial counterpart of traditional correlation

Positive

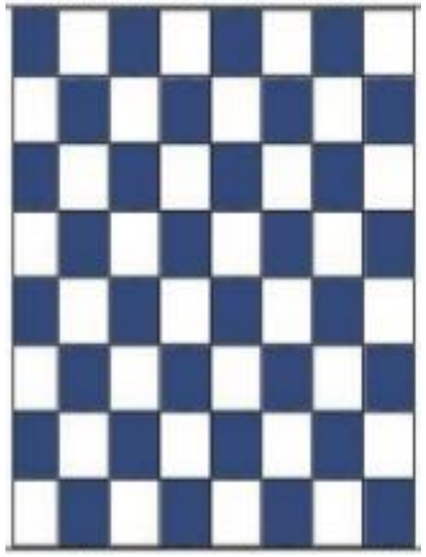
similar values are closeby



income, poverty, covid cases, vegetation, temperature,...

Negative

similar values are further apart

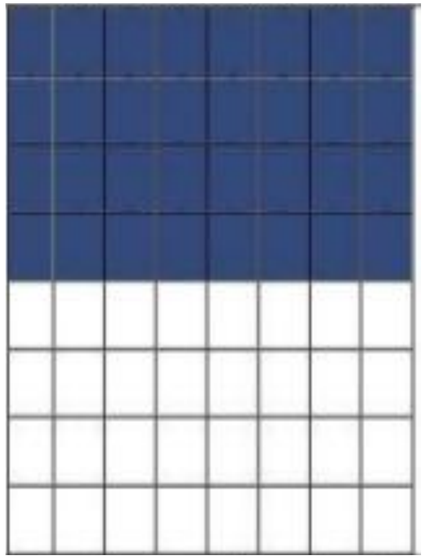


Spatial autocorrelation: Positive vs. Negative

Is the spatial counterpart of traditional correlation

Positive

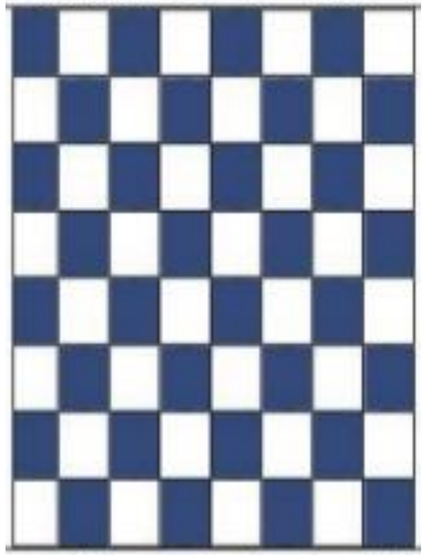
similar values are closeby



income, poverty, covid cases, vegetation, temperature,...

Negative

similar values are further apart



tigers, pharmacies, fire/police/ metro stations, hospitals,...

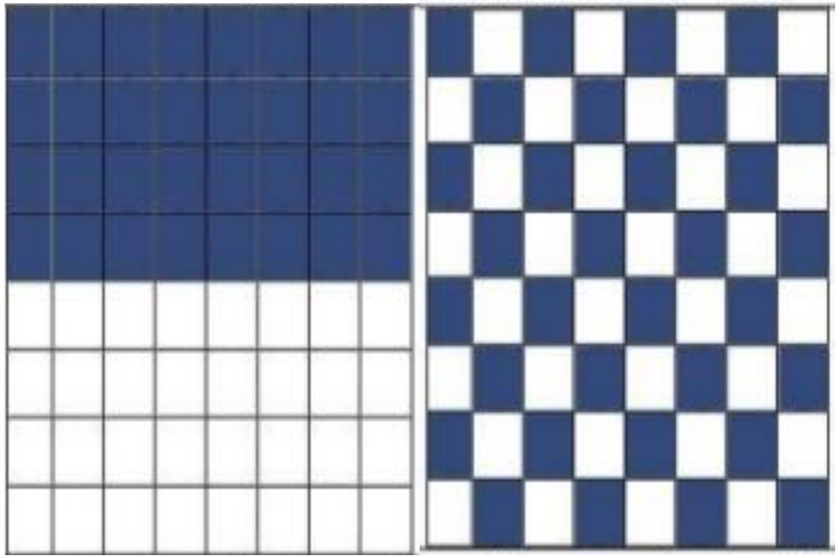
WHY DO COMPETITORS OPEN THEIR STORES NEXT TO ONE ANOTHER?

Spatial autocorrelation: Global vs. Local

Is the spatial counterpart of traditional correlation

Global

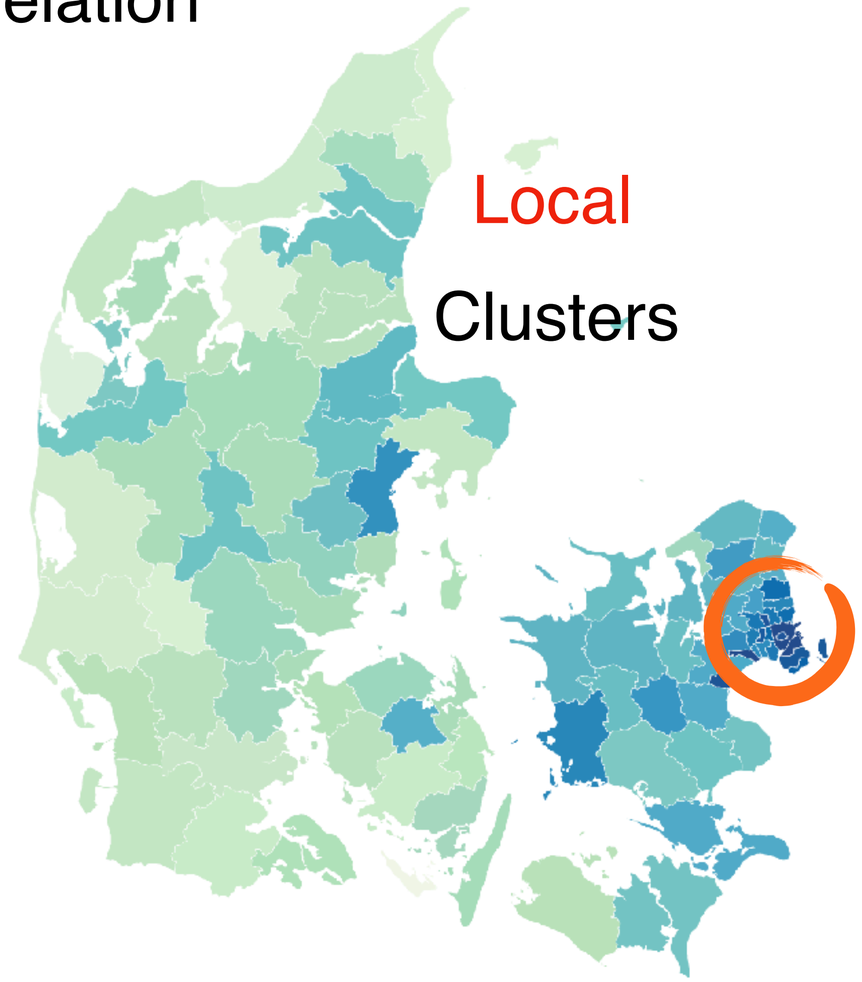
Clustering



Do values tend to be close to (dis)similar values?

Local

Clusters

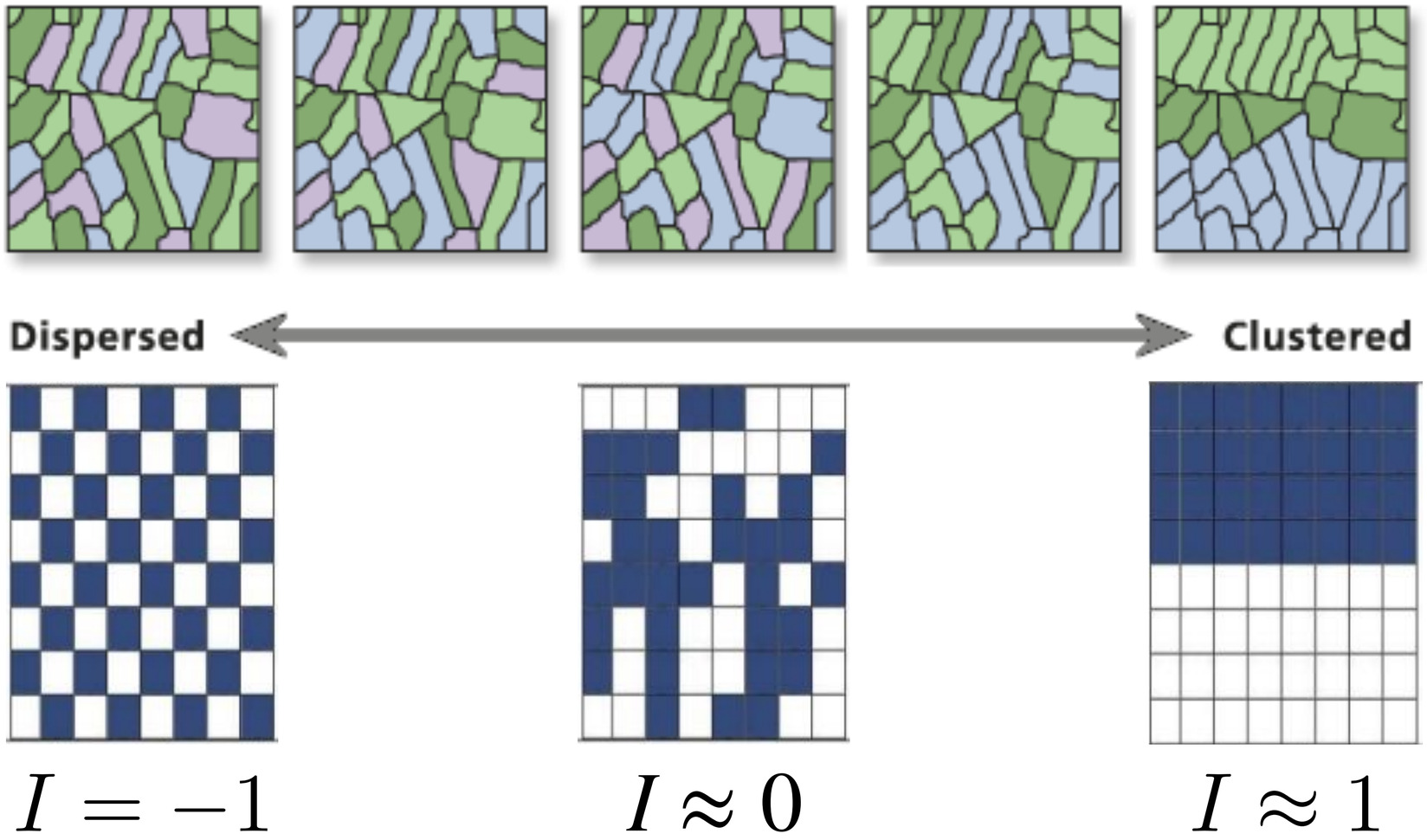


Are there areas with an extraordinary concentration of (dis)similar values?

Global Spatial Autocorrelation

Global spatial autocorrelation: Moran's I

Moran's I measures the average correlation between the value of a variable at one location and the value at nearby locations.

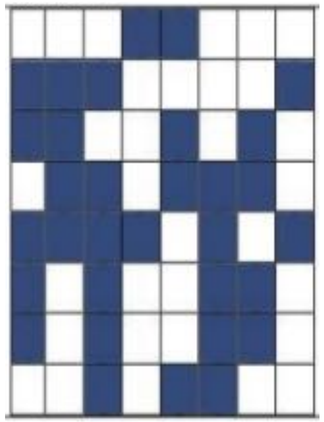


Global spatial autocorrelation: Moran's I

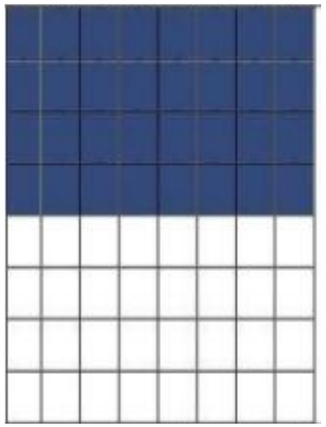
Moran's I measures the average correlation between the value of a variable at one location and the value at nearby locations.

Statistical inference with a significance value:
How likely is it to obtain a map like observed from a purely random pattern?

$I = -1$



$I \approx 0$



$I \approx 1$

Global spatial autocorrelation: Moran's I

Moran's I measures the average correlation between the value of a variable at one location and the value at nearby locations.

$$y_i$$

$$y_{\text{lag},i} = \sum_{j=1}^{n} w_{ij} y_j$$

$$\rho = \frac{1}{n-1} \sum \left(\frac{y_i - \bar{y}}{s_y} \right) \left(\frac{x_i - \bar{x}}{s_x} \right)$$

Global spatial autocorrelation: Moran's I

Moran's I measures the average correlation between the value of a variable at one location and the value at nearby locations.

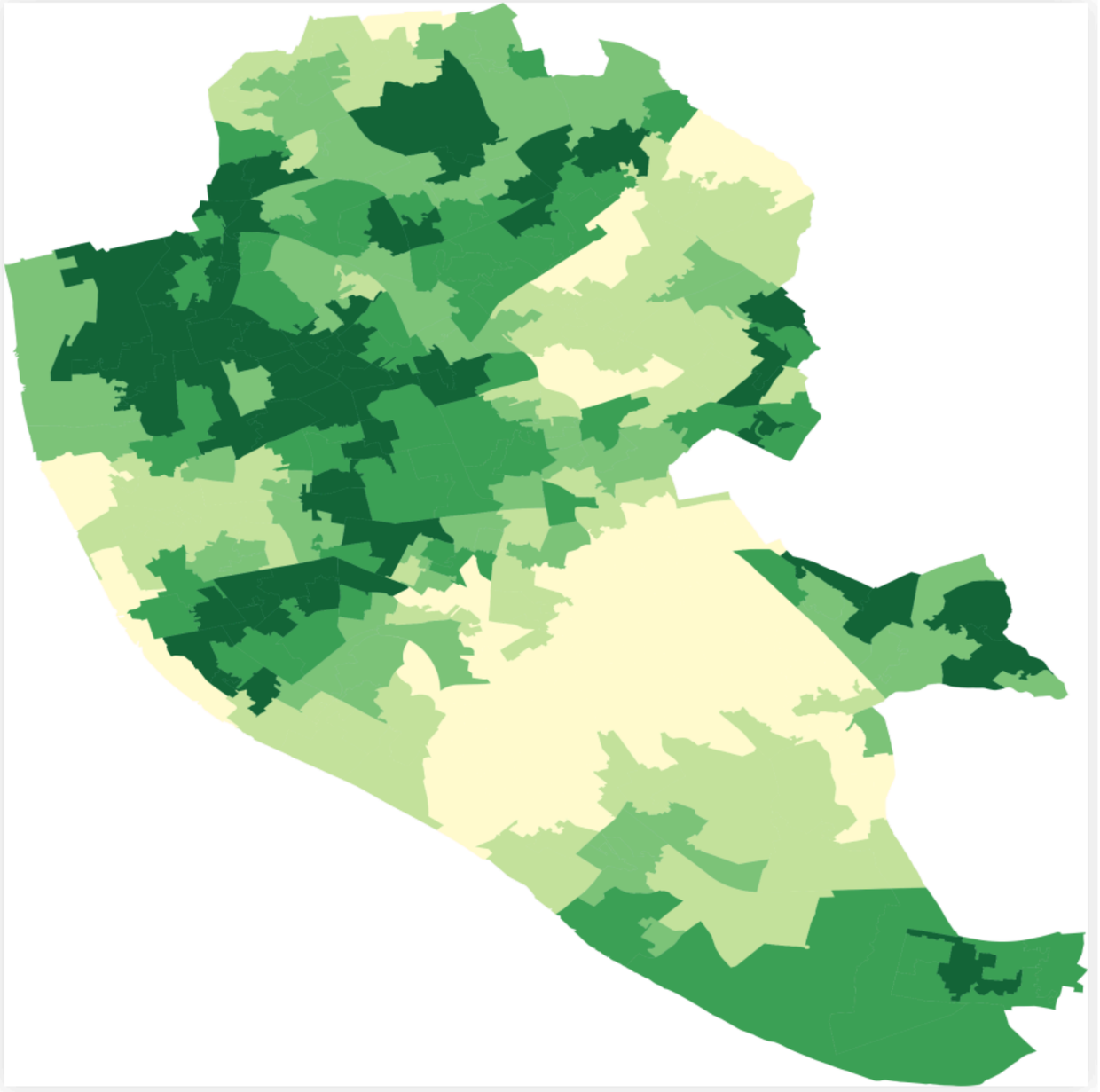
$$y_i$$

$$y_{\text{lag},i} = \sum_{j=1}^{n} w_{ij} y_j$$

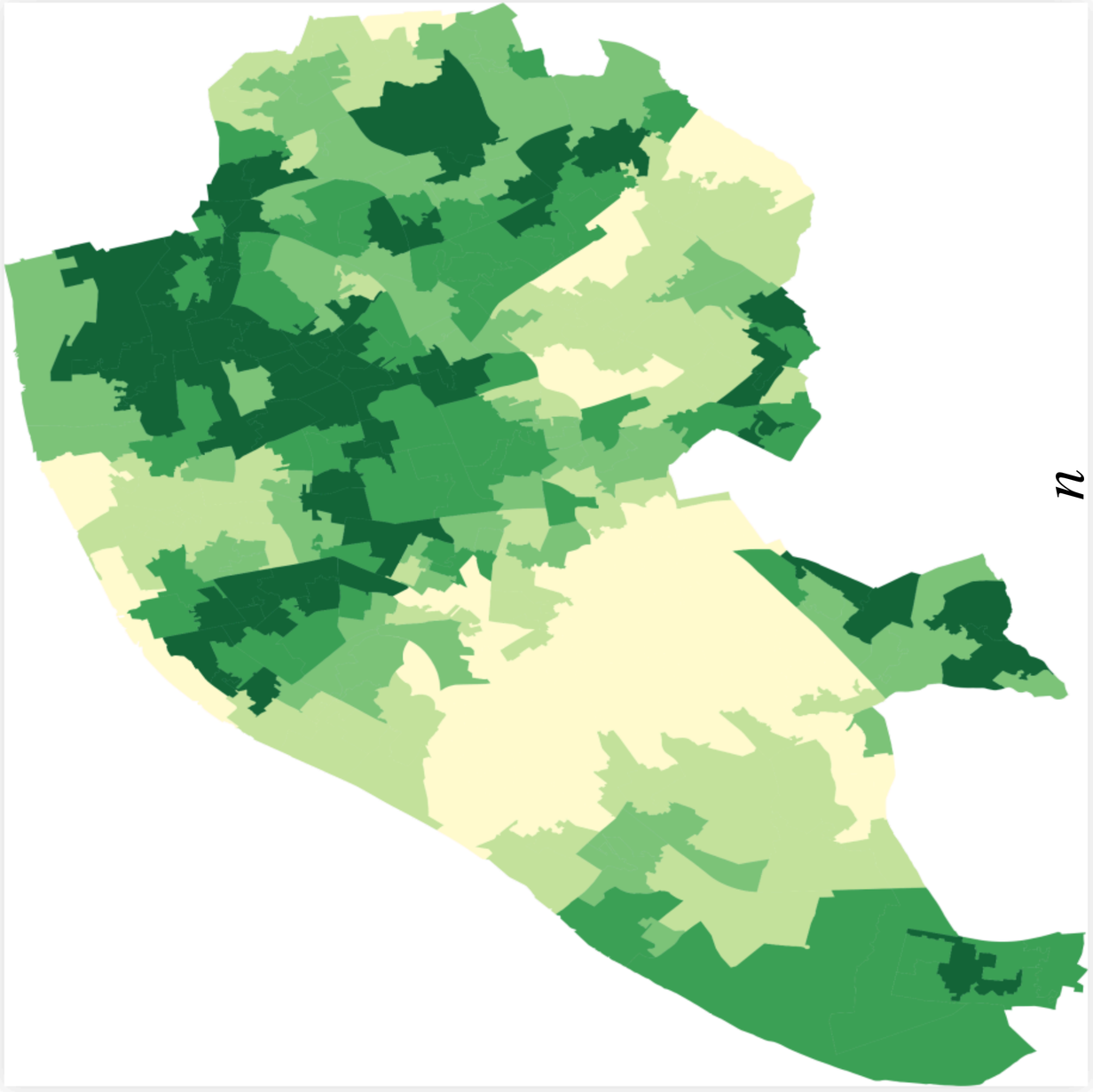
$$\rho = \frac{1}{n-1} \sum \left(\frac{y_i - \bar{y}}{s_y} \right) \left(\frac{x_i - \bar{x}}{s_x} \right)$$

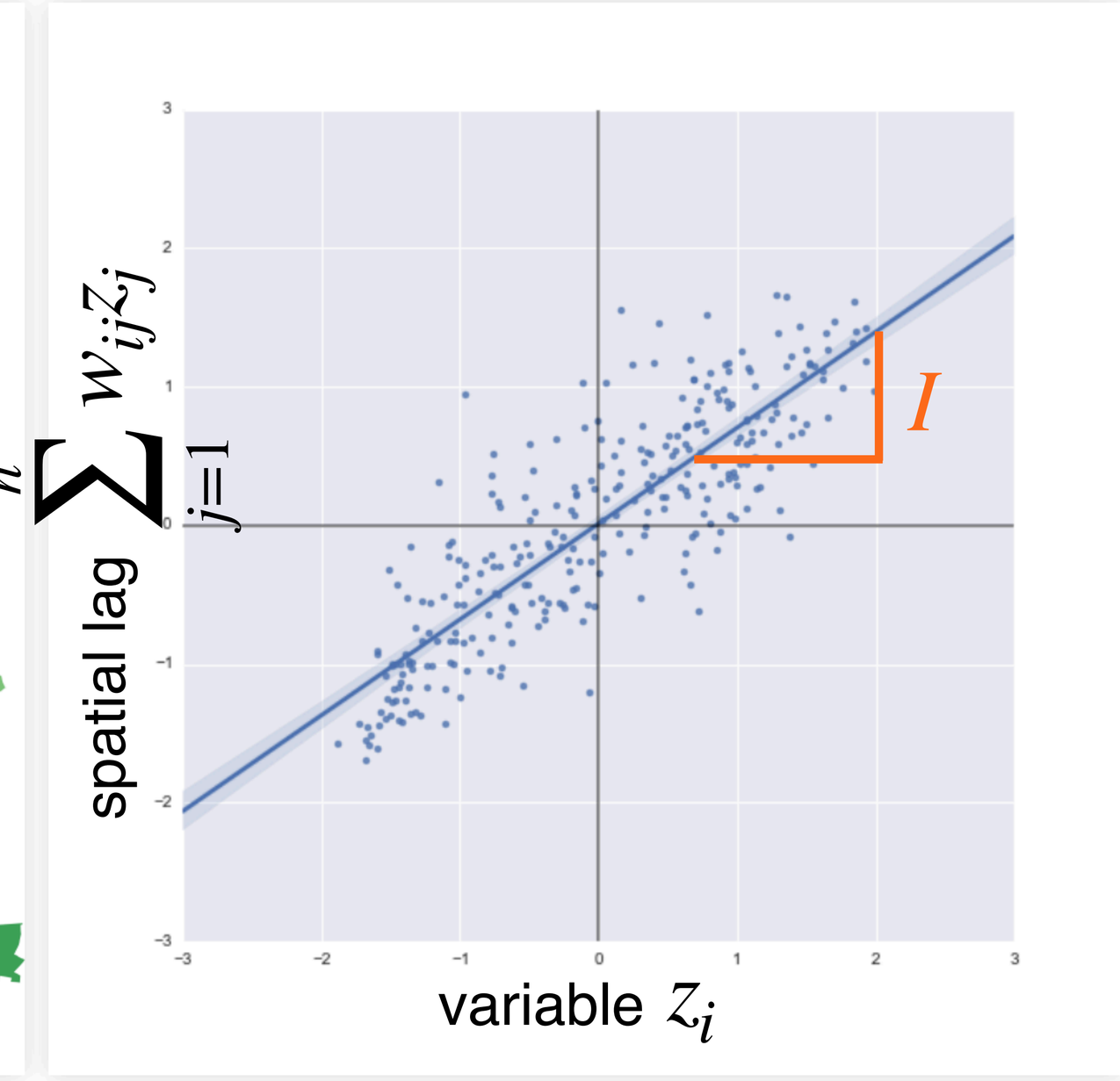
If we call $z_i = \left(\frac{y_i - \bar{y}}{s_y} \right)$, then:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}$$



Moran's I is the regression slope in the Moran plot





Local Spatial Autocorrelation

Local spatial autocorrelation is about local clusters

Cluster = Portion of a map where values are correlated in a particularly strong or specific way

Local spatial autocorrelation is about local clusters

Cluster = Portion of a map where values are correlated in a particularly strong or specific way

Positive	High-High: Hotspot
	Low-Low: Coldspot
Negative	High-Low: Spatial outlier
	Low-High: Spatial outlier (donut)

LISA: Local Indicators of Spatial Association

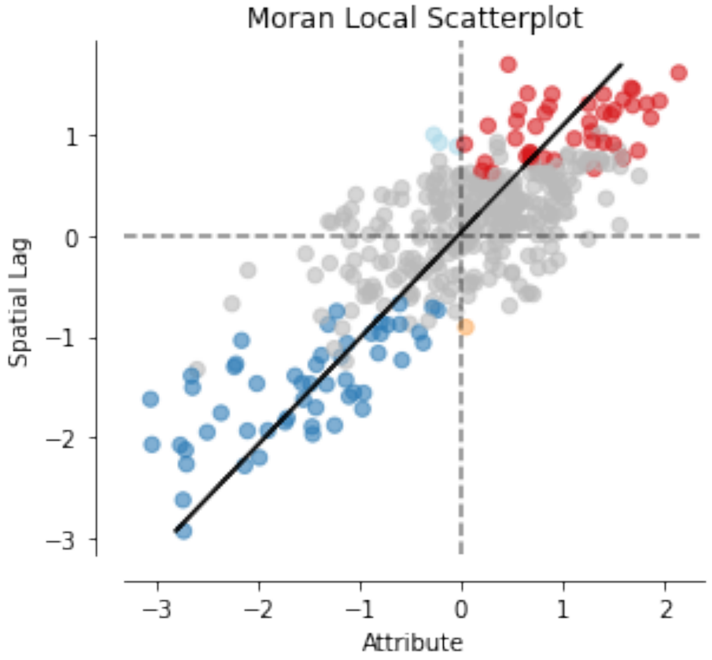
Like a "local Moran's I"

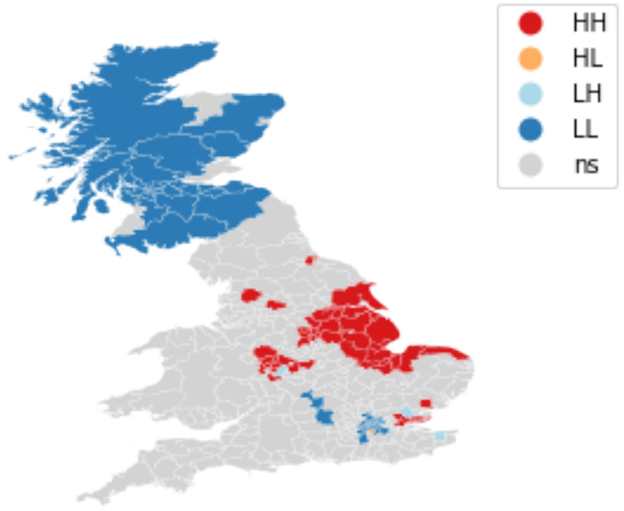
How much is each object's relation with its neighbors different from the relations of other objects and their neighbors on the map?

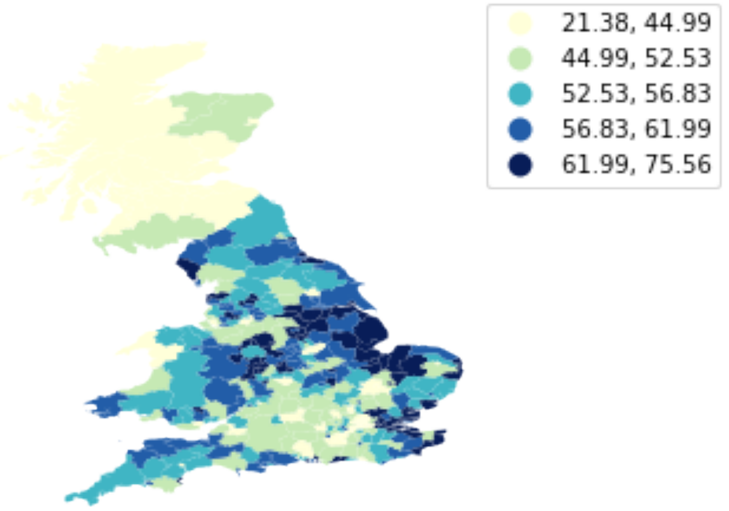
LISA: Local Indicators of Spatial Association

Like a "local Moran's I"

How much is each object's relation with its neighbors different from the relations of other objects and their neighbors on the map?







LISA: Local Indicators of Spatial Association

Like a "local Moran's I"

How much is each object's relation with its neighbors different from the relations of other objects and their neighbors on the map?

$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j \; ; \; m_2 = \frac{\sum_i z_i^2}{n}$$

$$\sum_i I_i = \gamma I$$

If W is standardized, then: $\gamma = \sum_i \sum_j w_{ij}$

Anselin, Geographical analysis 27.2 (1995)

Jupyter

https://learnit.itu.dk/mod/resource/view.php?id=155919

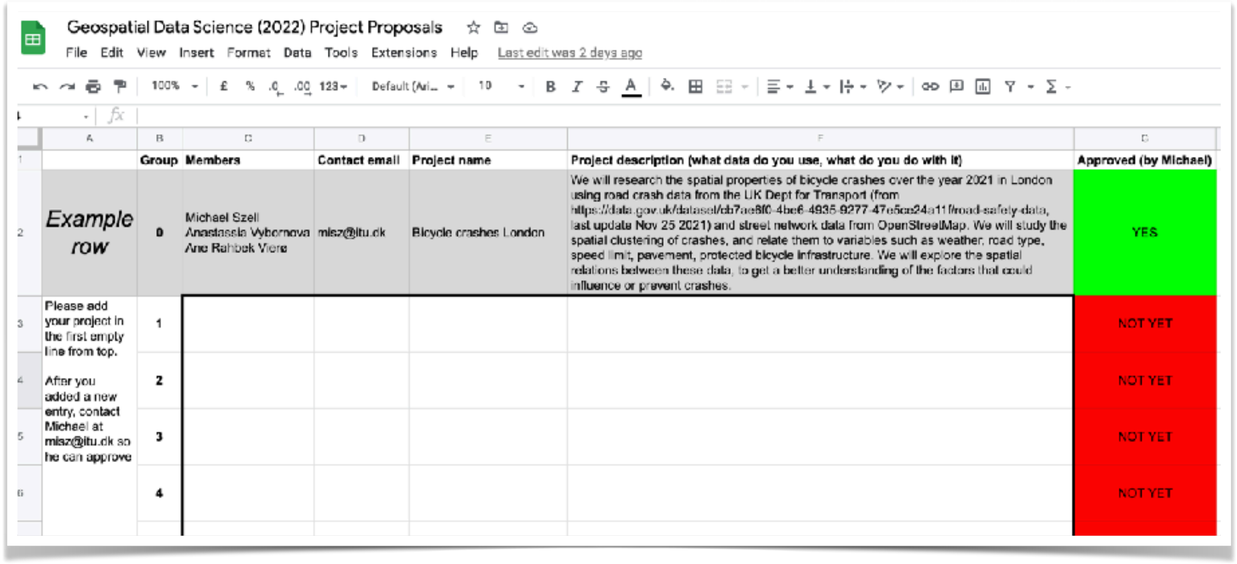
Geospatial Data Science Exam Project Instructions, Spring 2022

The submission is a written project report about the application of geospatial data science either to **answer a research question** or to **create a prototype of a digital product**. It may range from a technical workflow proof of concept to research data exploration. The project should explore or solve a problem with a geospatial dimension and may focus on any aspect of spatial data collection, visualization, analysis, or statistical evaluation. The submission has two parts: (a) commented code deposited on Github (or similar code repository) and (b) the associated report that describes the project and links to the code repository.

If anything is left unclear after reading this document, please ask in the forum or Michael in class.

Project approval

Before starting to work on your project, you must form a group of 2 or 3 people, and submit a very short project proposal at the latest **by March 31st** at: [tinyurl link]



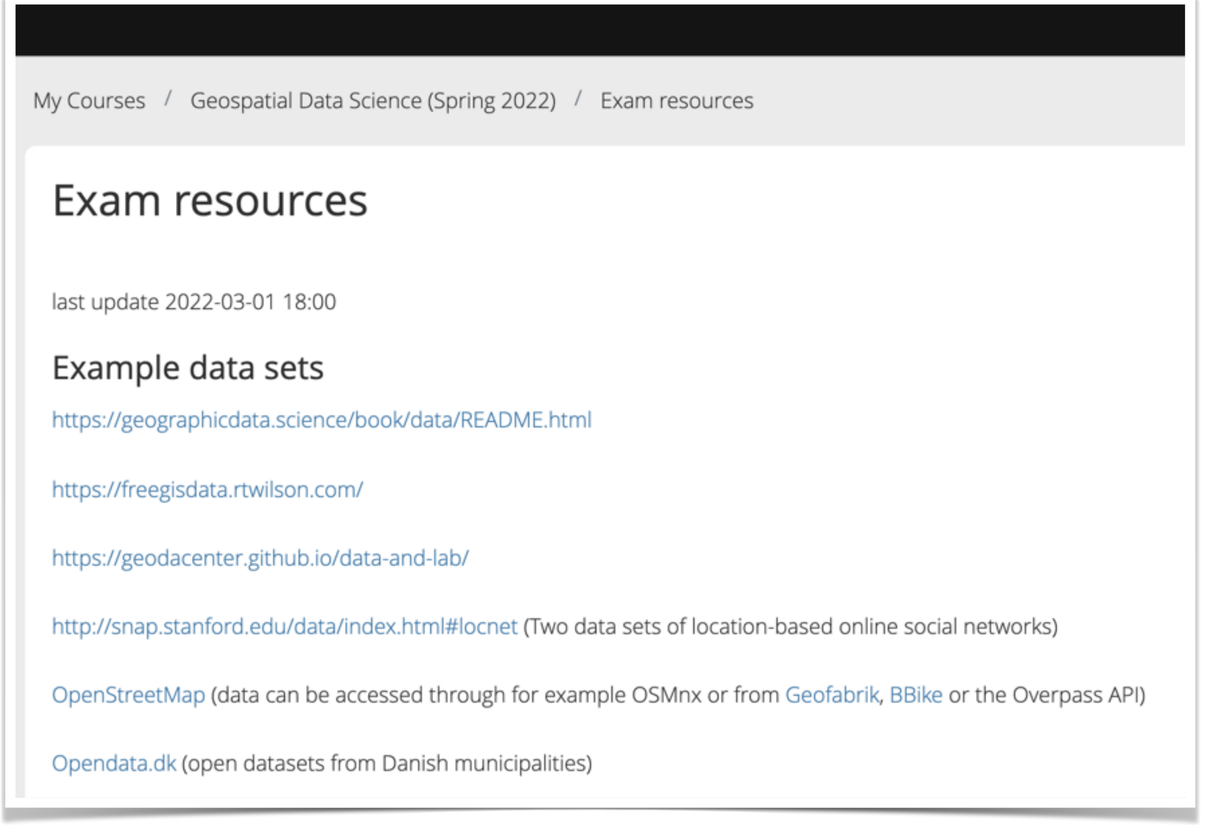
Geospatial Data Science (2022) Project Proposals

File Edit View Insert Format Data Tools Extensions Help Last edit was 2 days ago

	Group	Members	Contact email	Project name	Project description (what data do you use, what do you do with it)	Approved (by Michael)
Example row	0	Michael Szell Anastassia Vybornova Ane Rahbek Vierø	mlsz@itu.dk	Bicycle crashes London	We will research the spatial properties of bicycle crashes over the year 2021 in London using road crash data from the UK Dept for Transport (from https://data.gov.uk/dataset/cb7ae8f0-4be6-4935-9277-47e5ca24a11f/road-safety-data, last update Nov 25 2021) and street network data from OpenStreetMap. We will study the spatial clustering of crashes, and relate them to variables such as weather, road type, speed limit, pavement, protected bicycle infrastructure. We will explore the spatial relations between these data, to get a better understanding of the factors that could influence or prevent crashes.	YES
Please add your project in the first empty line from top. After you added a new entry, contact Michael at mlsz@itu.dk so he can approve	1					NOT YET
	2					NOT YET
	3					NOT YET
	4					NOT YET

[tinyurl link]

https://learnit.itu.dk/mod/page/view.php?id=155887



My Courses / Geospatial Data Science (Spring 2022) / Exam resources

Exam resources

last update 2022-03-01 18:00

Example data sets

https://geographicdata.science/book/data/README.html

https://freegisdata.rtwilson.com/

https://geodacenter.github.io/data-and-lab/

http://snap.stanford.edu/data/index.html#locnet (Two data sets of location-based online social networks)

OpenStreetMap (data can be accessed through for example OSMnx or from Geofabrik, BBike or the Overpass API)

Opendata.dk (open datasets from Danish municipalities)

Sources and further materials for today's class

Geographic Data Science with Python

https://geographicdata.science/book/notebooks/06_spatial_autocorrelation.html

https://geographicdata.science/book/notebooks/07_local_autocorrelation.html

https://darribas.org/gds_course/content/bF/concepts_F.html