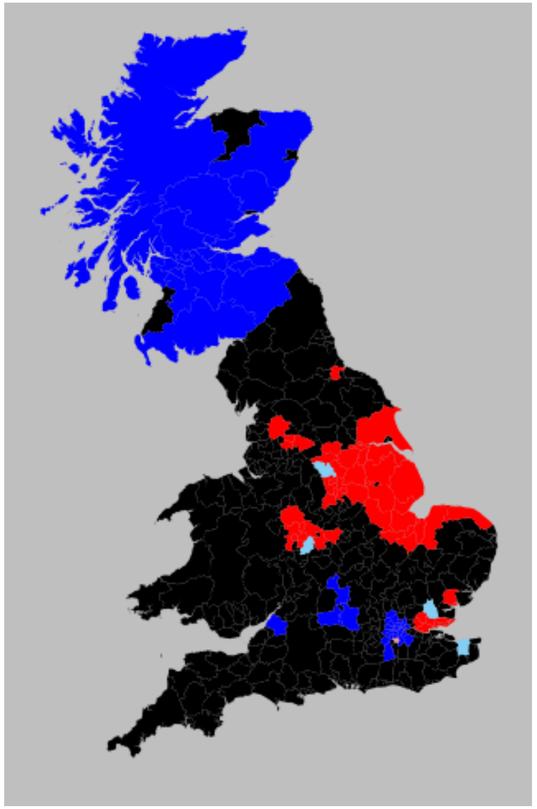
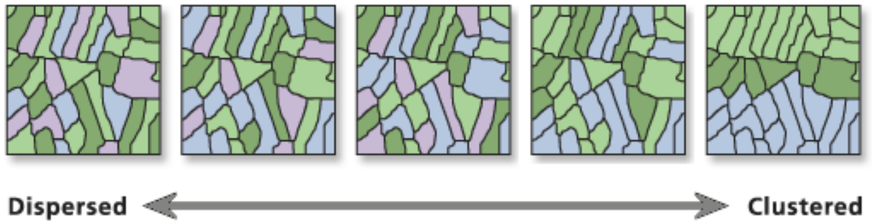
# Lecture 5: Spatial autocorrelation

Instructor: Ane Rahbek Vierø
Feb 27, 2023

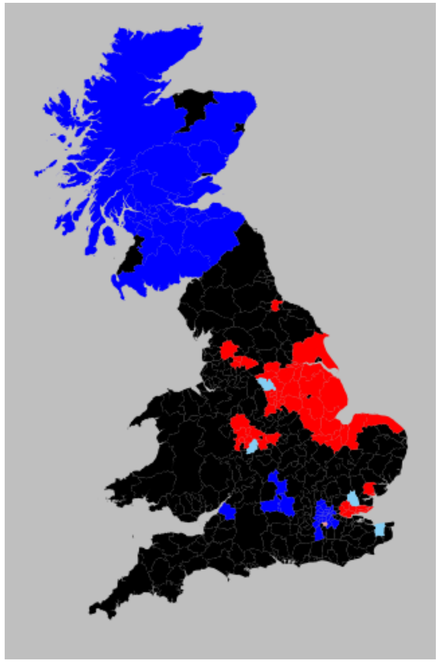


# Today you will learn about spatial correlation

## Questions to ask in Exploratory Spatial Data Analysis




## Use in Python with example

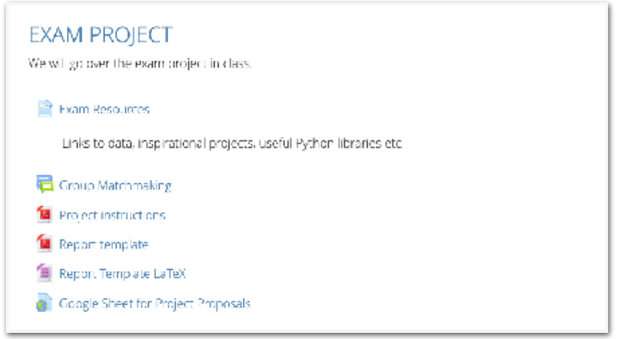


## Global and local measures of spatial autocorrelation

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

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

## Exam project info



# Feedback from early evaluation

More optional readings on LearnIT 

More extensive exercises 

Practical applications of concepts and methods in each lecture 

# Alternatives to Docker

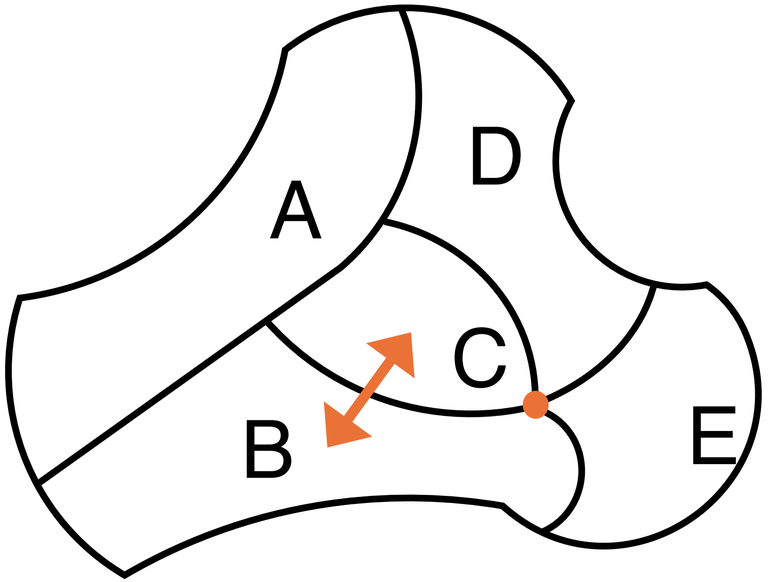
## Install Python Environment

We will use the Python environment gds_env throughout the course.

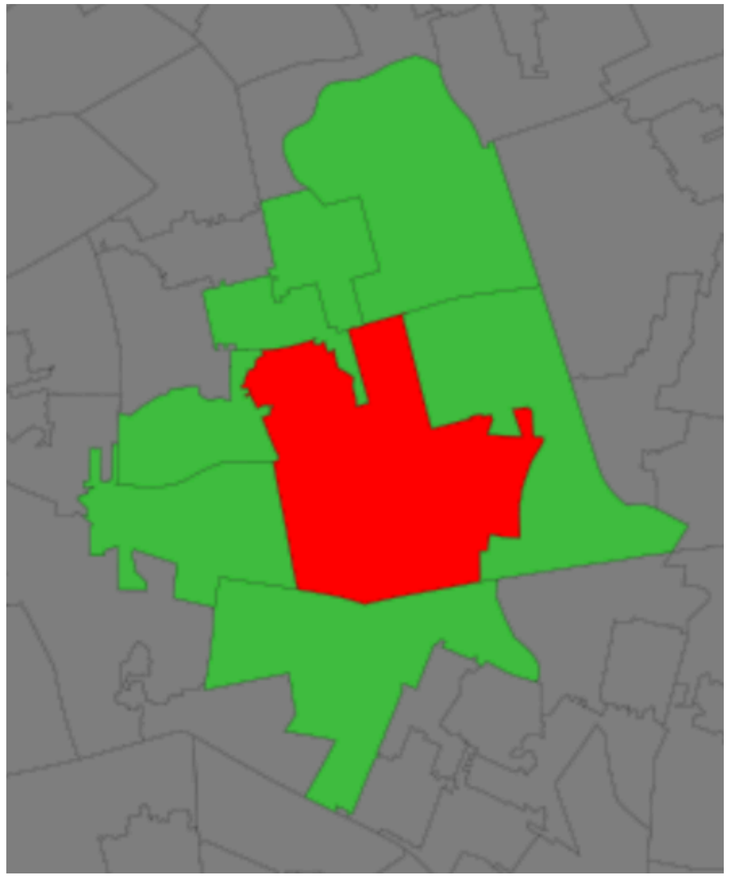
***This is slow and only recommended as a last resort.***

- Navigate to the folder installation_files
- Run *conda env create -f gds_py.yml*
- Activate the environment
- Run *pip install -r pip_requirements.txt*
- Run the notebook *test_gdspy_install.ipynb* to check that everything works

# Spatial weights encodes information about how are objects related to each other in space



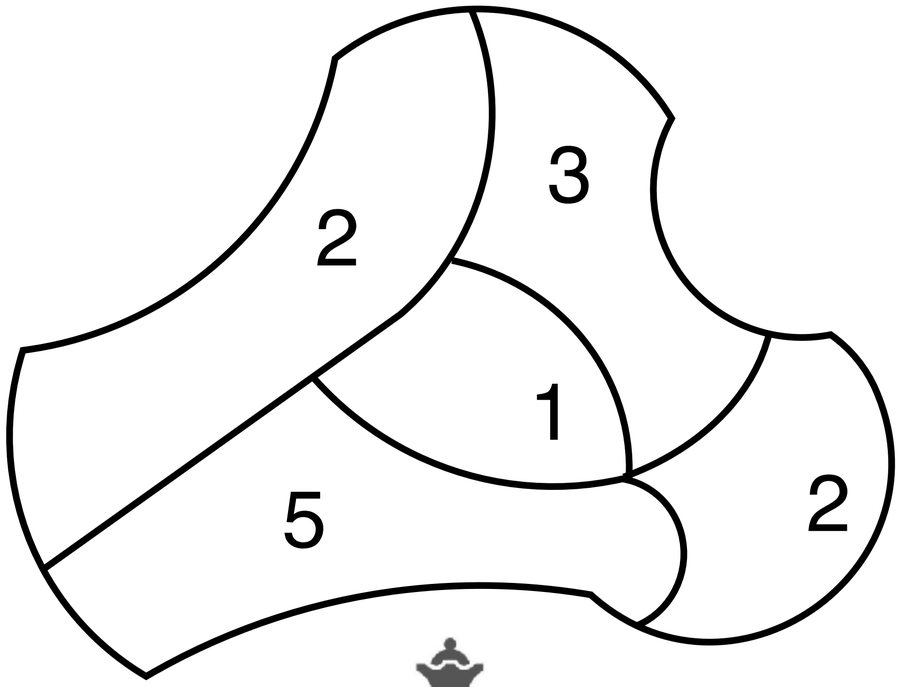

$$W = \begin{pmatrix} 0 & w_{12} & \dots & w_{1N} \\ w_{21} & \ddots & w_{ij} & \vdots \\ \vdots & w_{ji} & 0 & \vdots \\ w_{N1} & \dots & \dots & 0 \end{pmatrix}$$



# The spatial lag is the weighted average value of neighbors

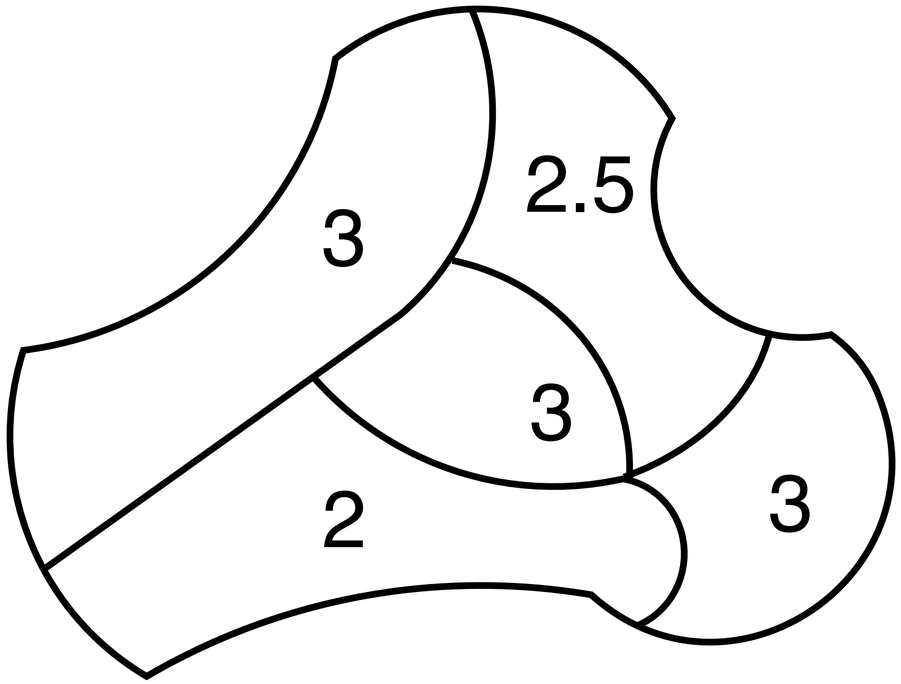
if W is standardized

Values $\mathbf{y}$





Spatial lag $\mathbf{y}_{\text{lag}}$

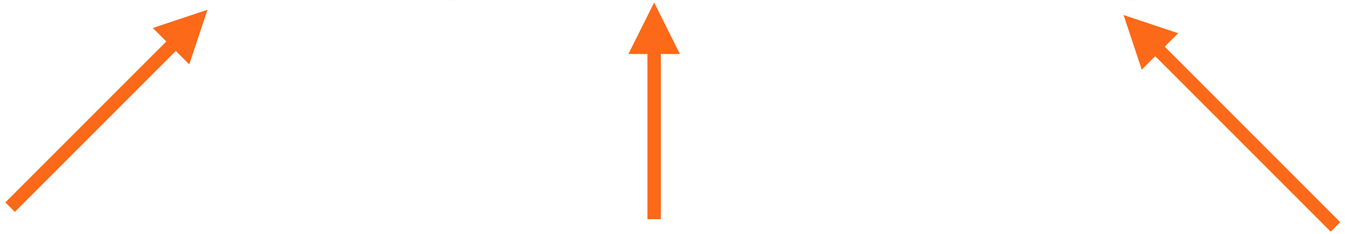



It is a smoother: It brings all values closer to the average

# 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



Assumption-free investigation of spatial patterns with statistical techniques

Is a variable concentrated over space?
Do similar values locate closeby? If so, where?
Given a choropleth map, is there a significant statistical pattern?

Is a variable concentrated over space?
Do similar values locate closeby? If so, where?
Given a choropleth map, is there a significant statistical pattern?

What process is behind a pattern?
Why are there certain clusters?

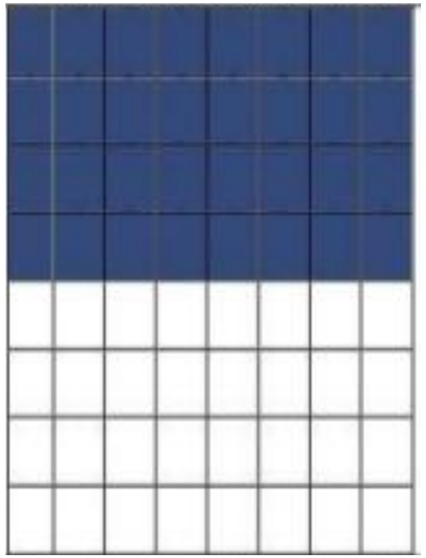
Spatial autocorrelation describes the relationship between values and locations

# Spatial autocorrelation: Positive vs. Negative

Is the spatial counterpart of traditional correlation

## Positive

similar values are closeby



## Negative

similar values are further apart

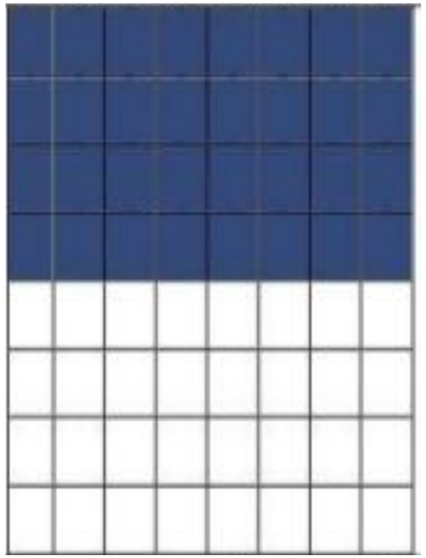
?

# Spatial autocorrelation: Positive vs. Negative

Is the spatial counterpart of traditional correlation

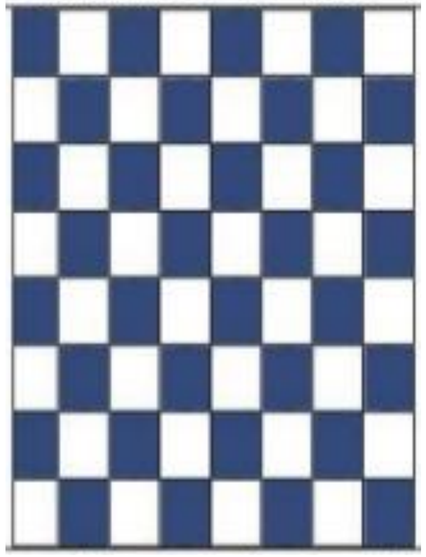
## Positive

similar values are closeby



## Negative

similar values are further apart

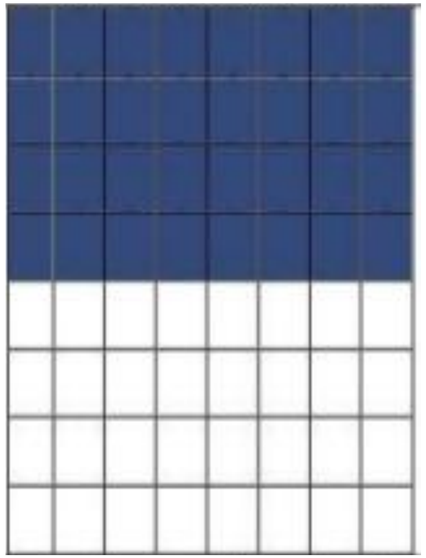


# Spatial autocorrelation: Positive vs. Negative

Is the spatial counterpart of traditional correlation

## Positive

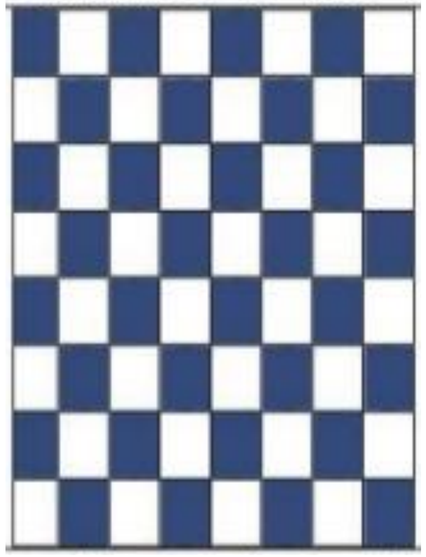
similar values are closeby



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

## Negative

similar values are further apart

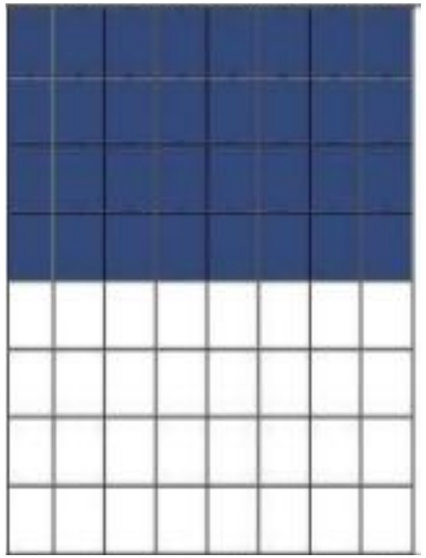


# Spatial autocorrelation: Positive vs. Negative

Is the spatial counterpart of traditional correlation

## Positive

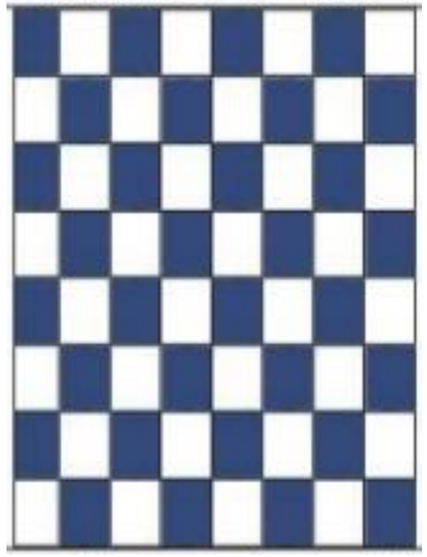
similar values are closeby



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

## Negative

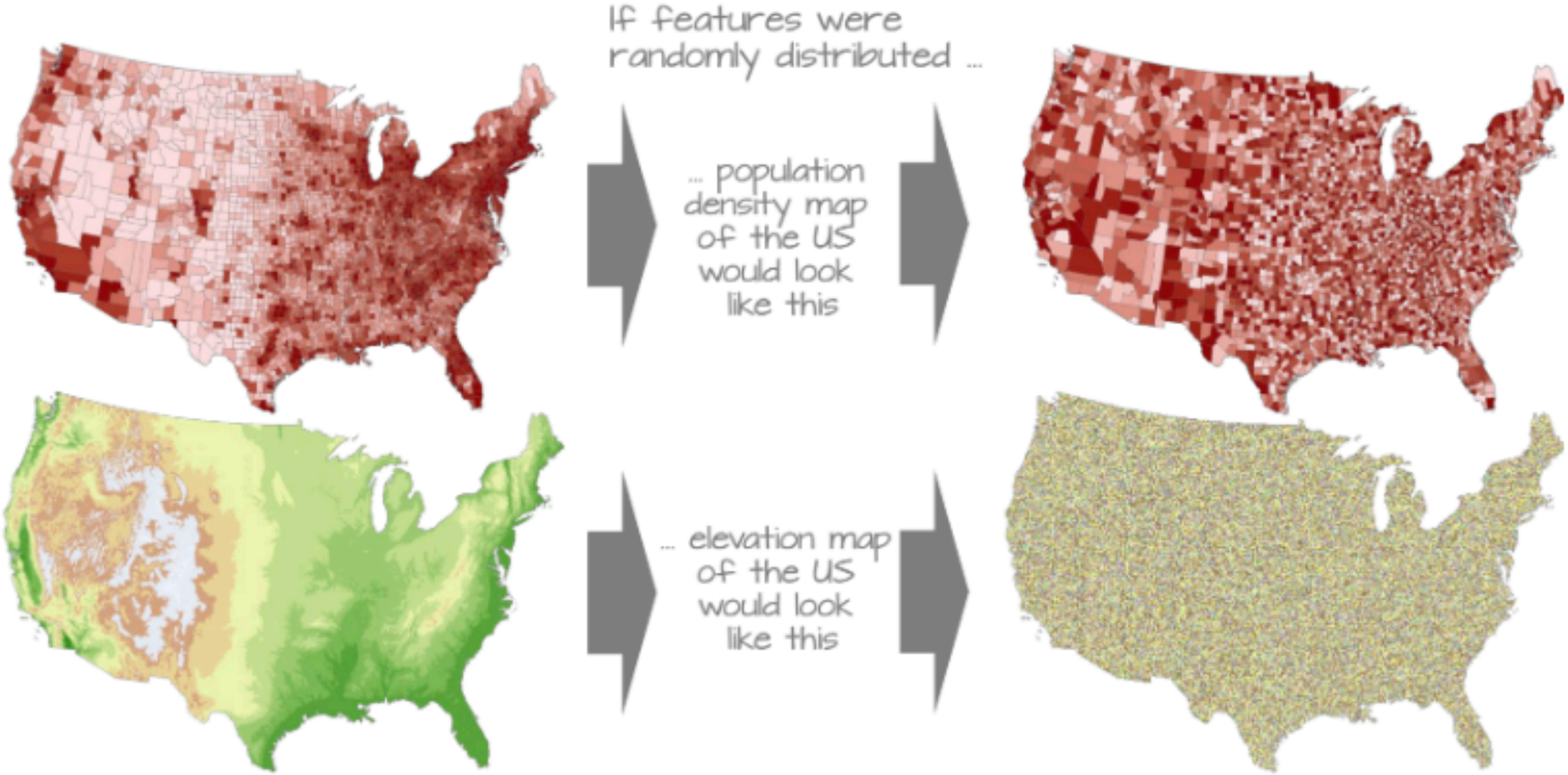
similar values are further apart



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

# WHY DO COMPETITORS OPEN THEIR STORES NEXT TO ONE ANOTHER?

# Random distribution is our null hypothesis

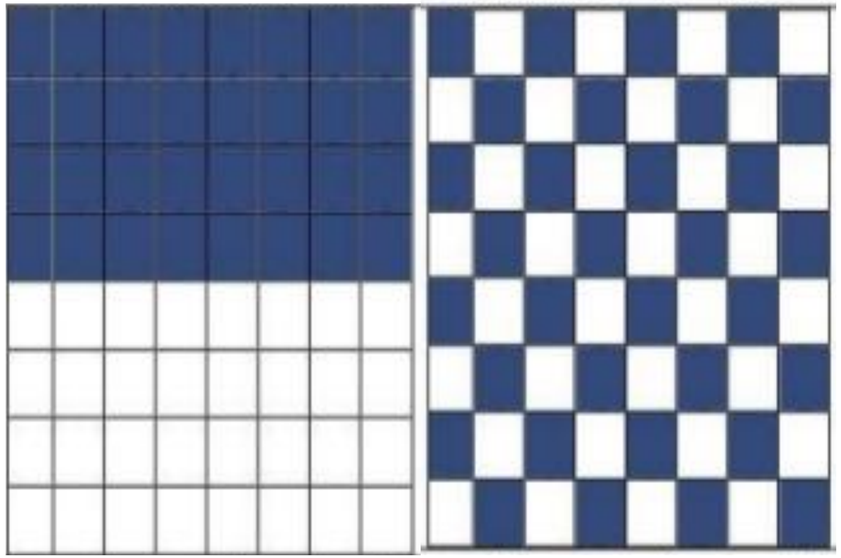



*https://mgimond.github.io/Spatial/spatial-autocorrelation.html*

# Spatial autocorrelation: Global vs. Local

## Global

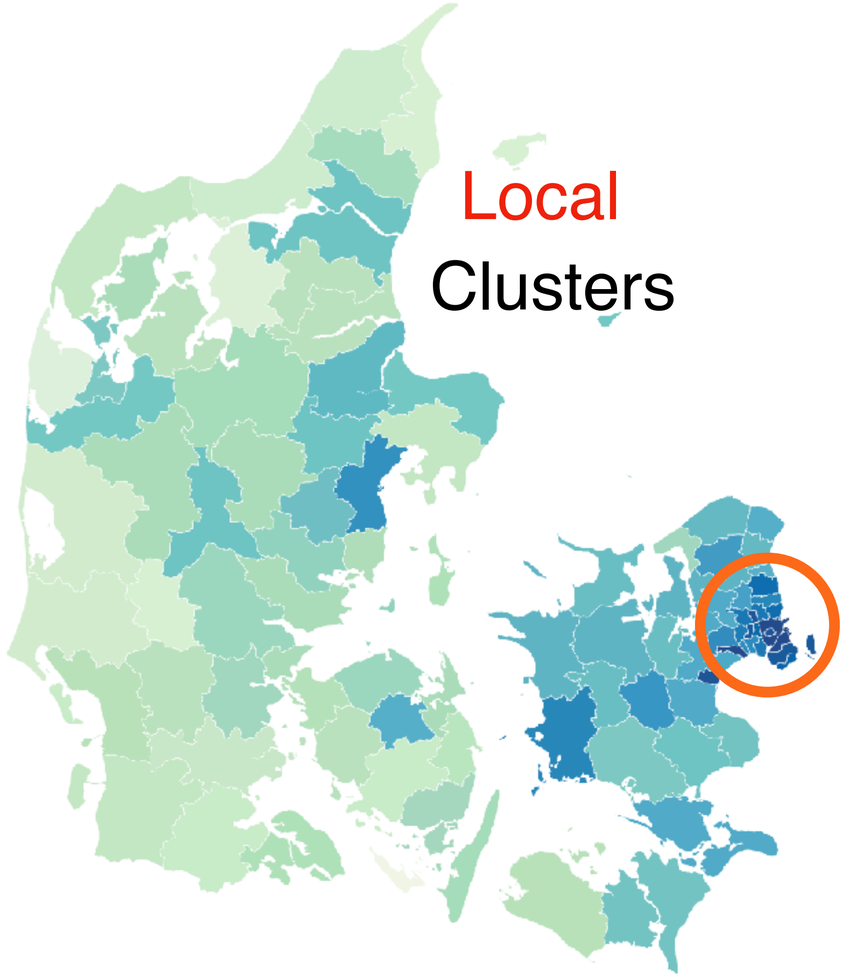
Clustering



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

## Local

Clusters

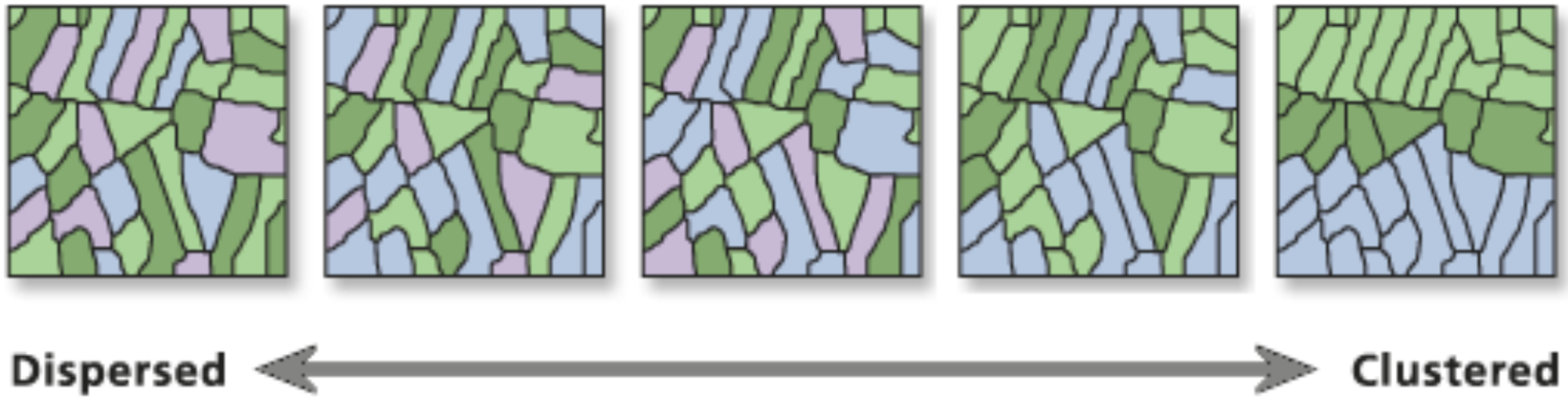


Are there areas with an extraordinary concentration of (dis)similar values?

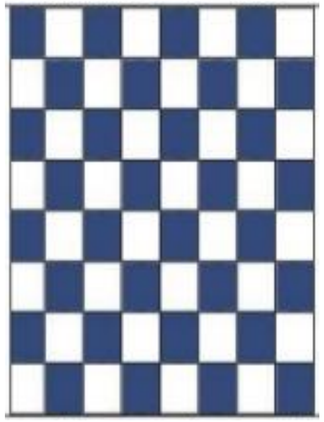
# Global Spatial Autocorrelation

# Global spatial autocorrelation: Moran's I

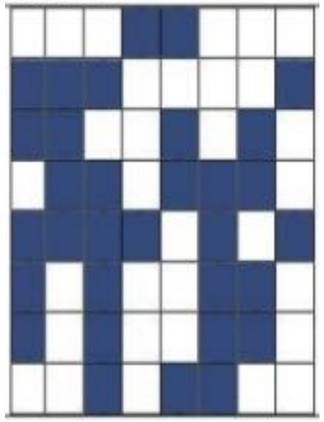
Moran's I measures the average correlation between the value of a variable at one location and the value at nearby locations.



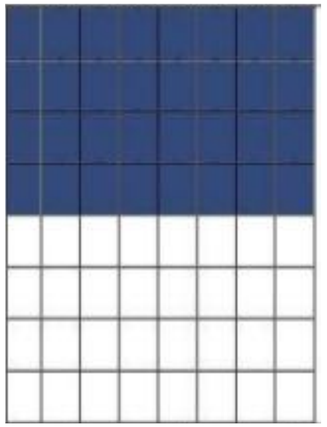
# Global spatial autocorrelation: Moran's I



$I = -1$



$I \approx 0$

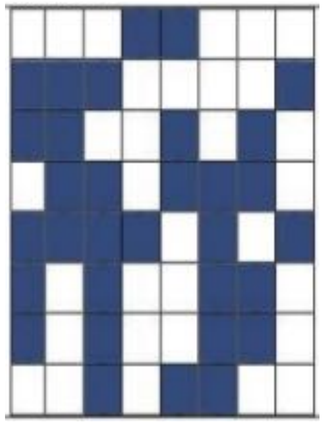


$I \approx 1$

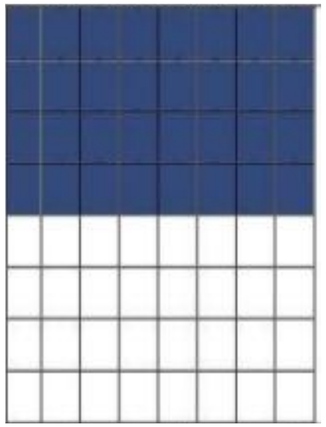
# Global spatial autocorrelation: Moran's I

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.

$y_i$

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

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_i - \bar{Y}) (y_j - \bar{Y})}{\sum_i (y_i - \bar{Y})^2}$$

# Global spatial autocorrelation: Moran's I

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_i - \bar{Y}) (y_j - \bar{Y})}{\sum_i (y_i - \bar{Y})^2}$$

Standardized Moran's I:

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

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

# Global spatial autocorrelation: Moran's I

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{X}) (x_j - \bar{X})}{\sum_i (x_i - \bar{X})^2}$$

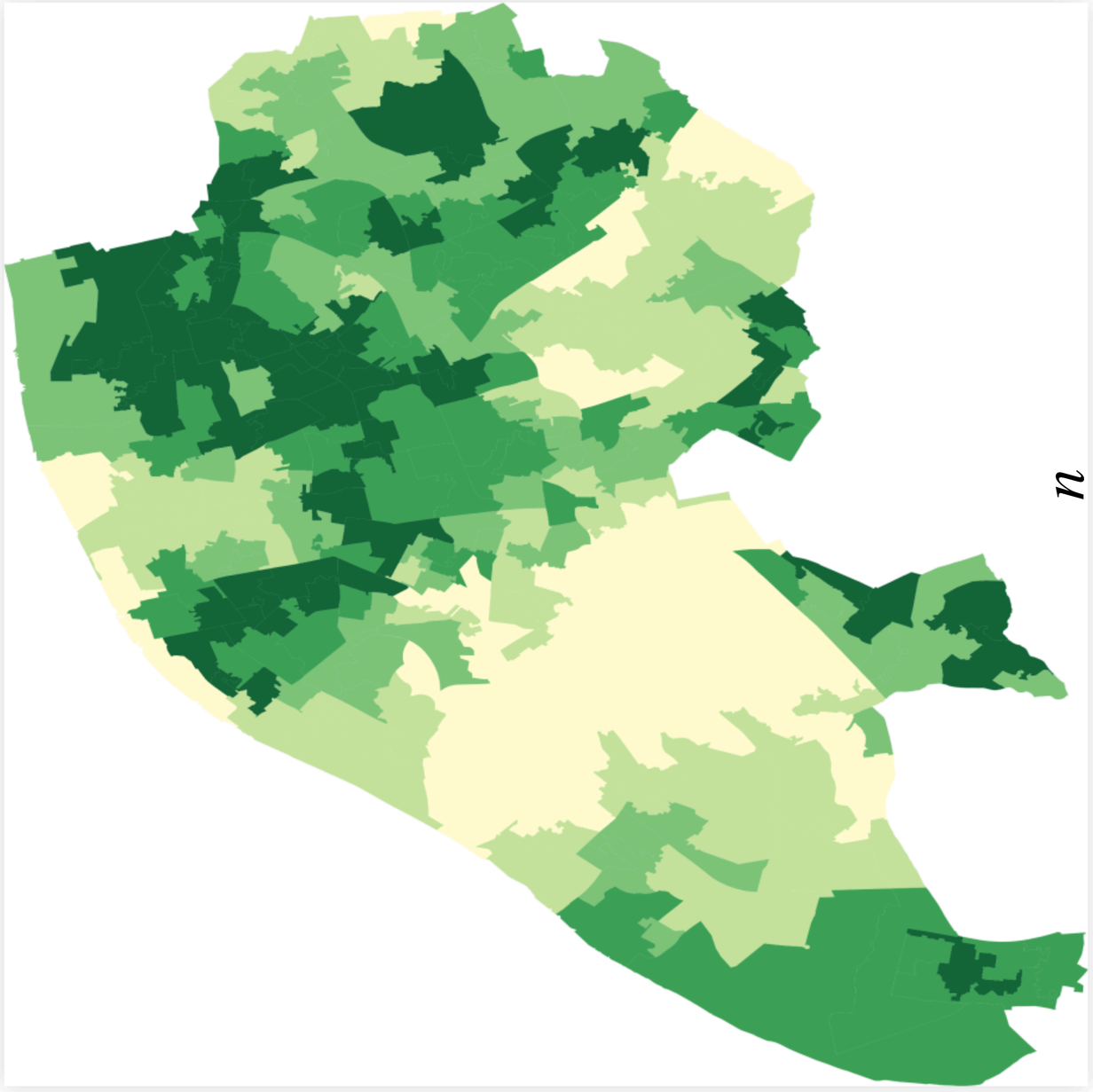
Standardized Moran's I:

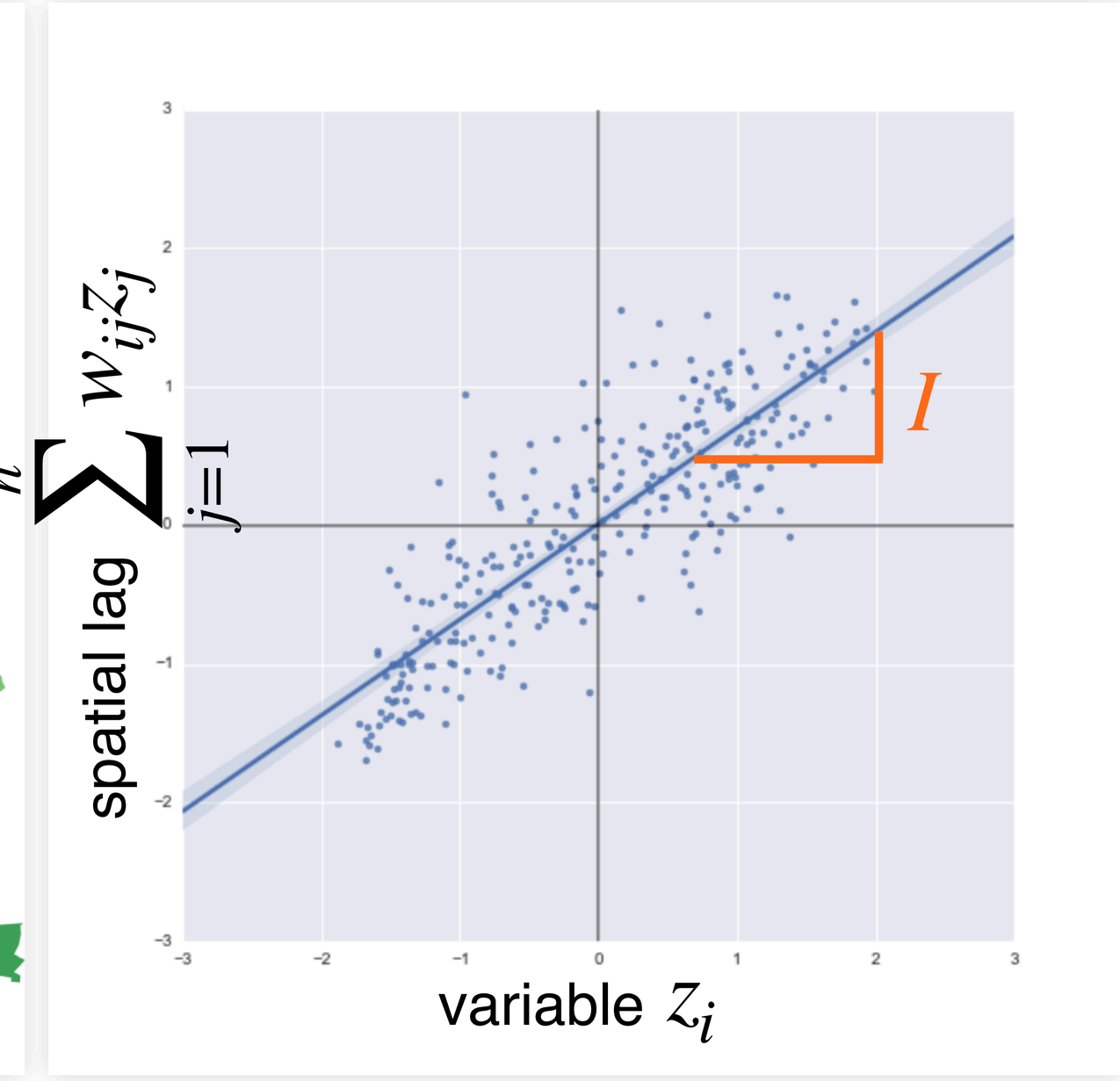
If we call $z_i = \left( \frac{y_i - \bar{y}}{s_y} \right)$, then: $$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}$$

Row standardized Moran's I:

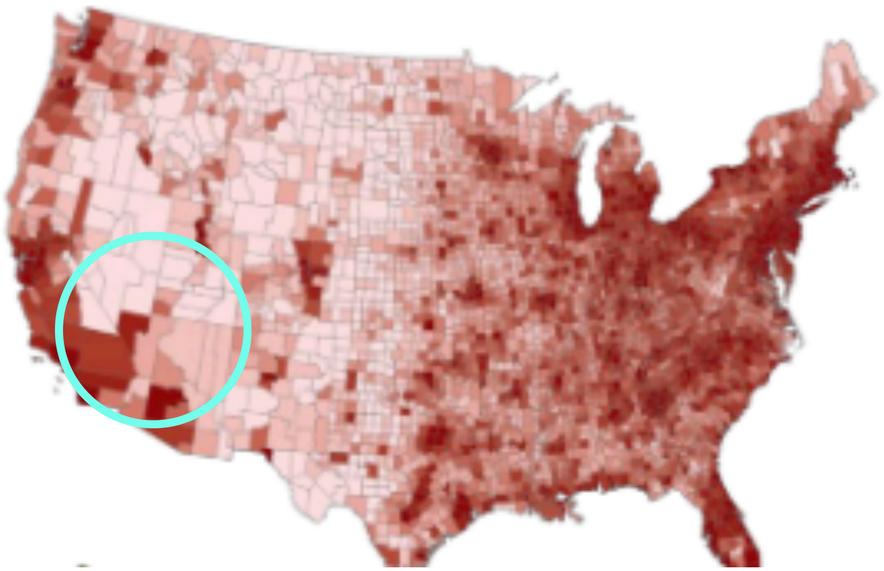
$$I = \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}$$

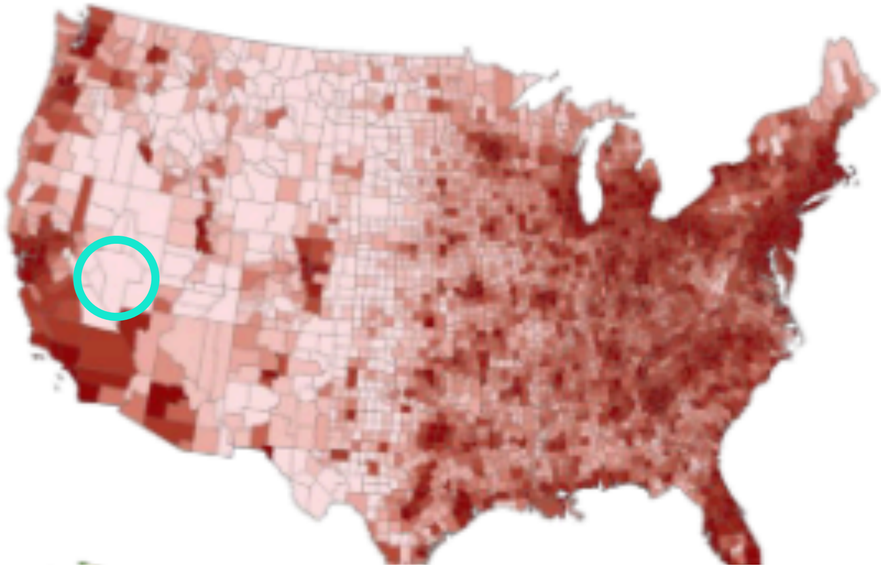
# Moran's I is the regression slope in the Moran plot

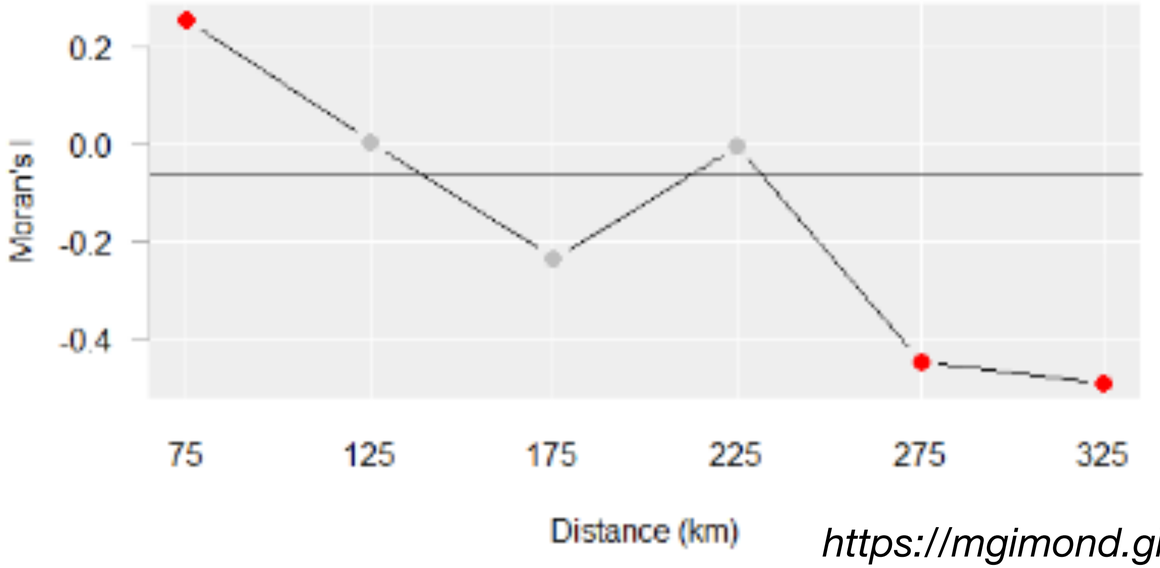




# Scale matters








*https://mgimond.github.io/Spatial/spatial-autocorrelation.html*

# 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

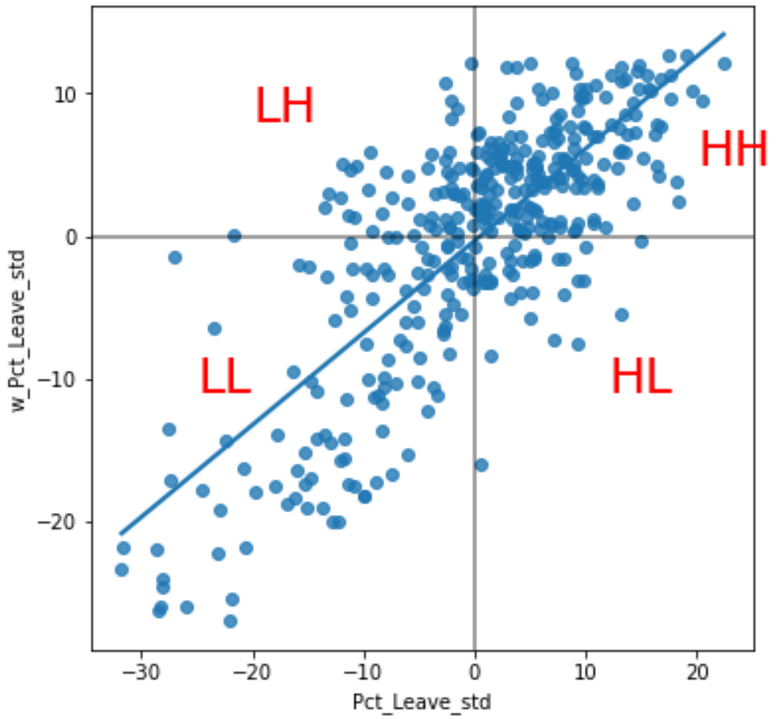
High-High: Hotspot

Low-Low: Coldspot

Negative

High-Low: Spatial outlier

Low-High: Spatial outlier (donut)




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

# LISA: Local Indicators of Spatial Association

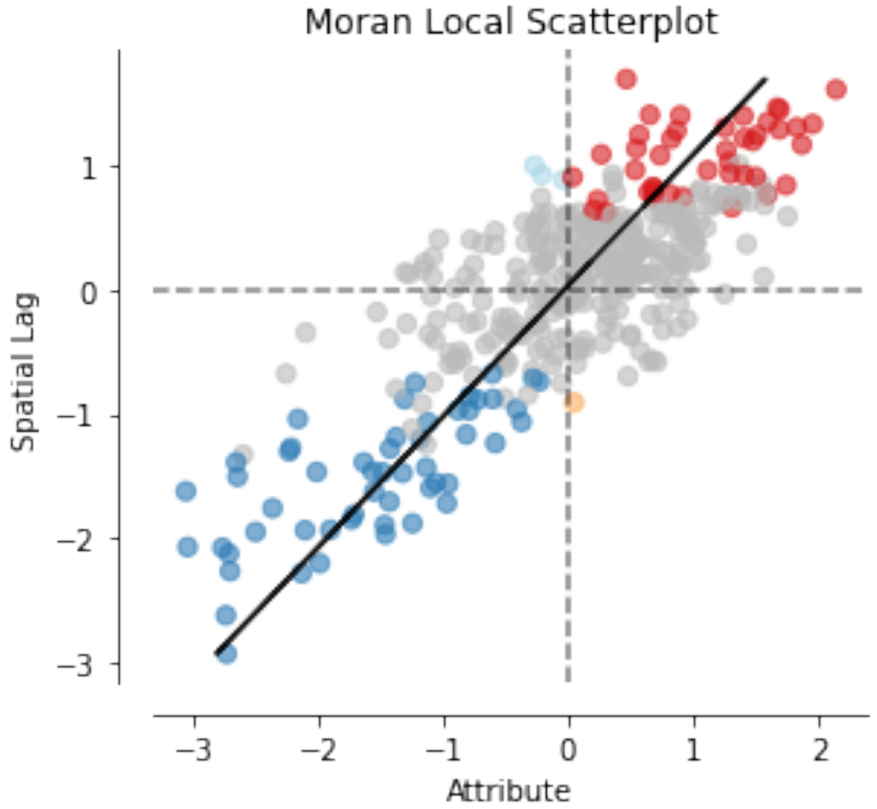
1. Test for spatial cluster

To what extent is each object's relation with its neighbors significantly different from the relations of other objects and their neighbors on the map?

2. Test for significance

# LISA: Local Indicators of Spatial Association

To what extent is each object's relation with its neighbors significantly different from the relations of other objects and their neighbors on the map?



# LISA: Local Moran's I

$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j \qquad m_2 = \frac{\sum_i z_i^2}{n} \qquad z_i = y_i - \bar{y}$$

Anselin, Geographical analysis 27.2 (1995)

# LISA: Local Moran's I

$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j \qquad m_2 = \frac{\sum_i z_i^2}{n} \qquad z_i = y_i - \bar{y}$$

If $W$ is row-standardized, then:

$$\sum_i I_i = \gamma I$$

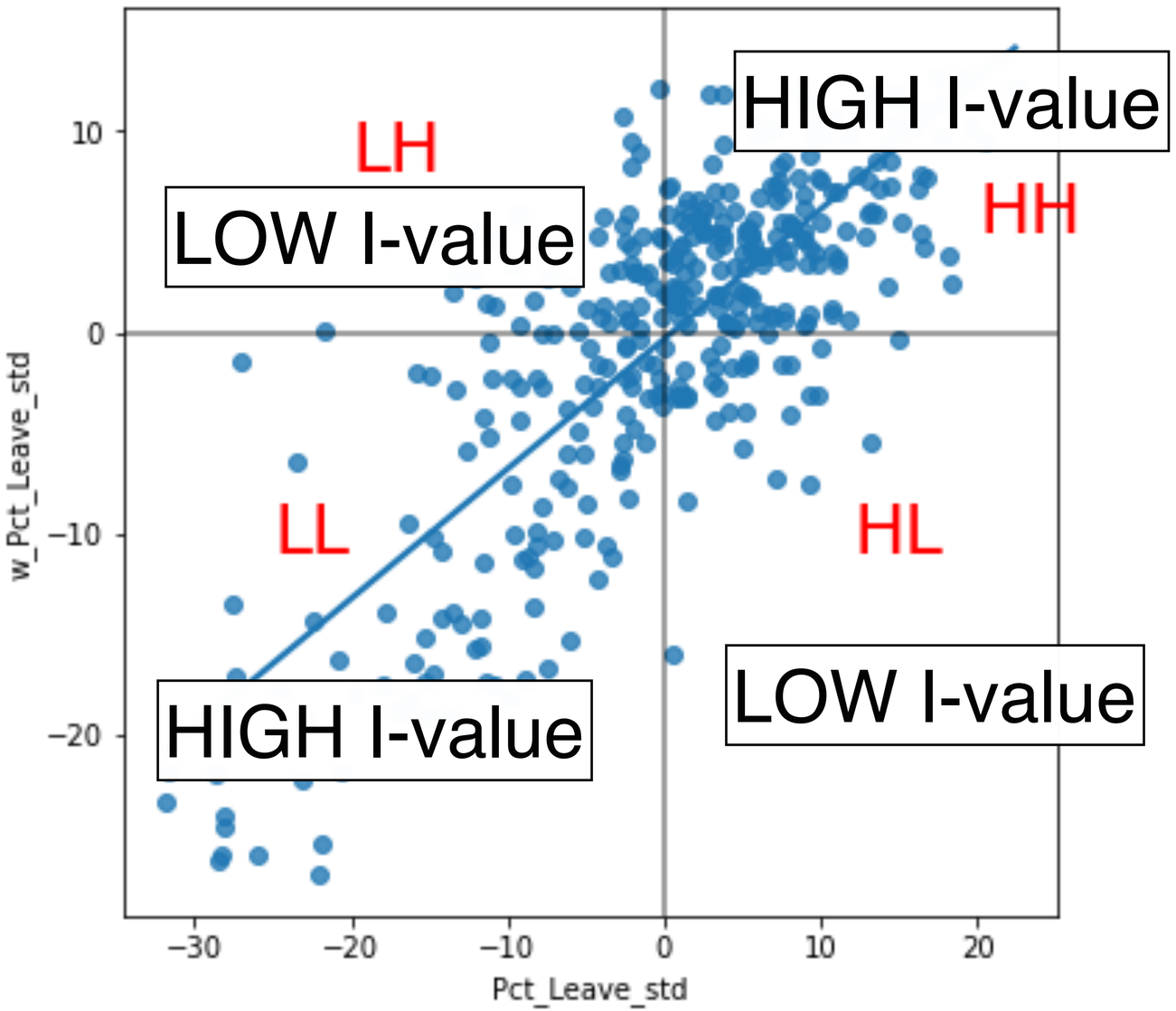
$$\gamma = \sum_i \sum_j w_{ij}$$ = scale factor

$I =$ global indicator of autocorrelation

= Sum of local indicators proportional to global indicator

Anselin, Geographical analysis 27.2 (1995)

# LISA: Local Indicators of Spatial Association



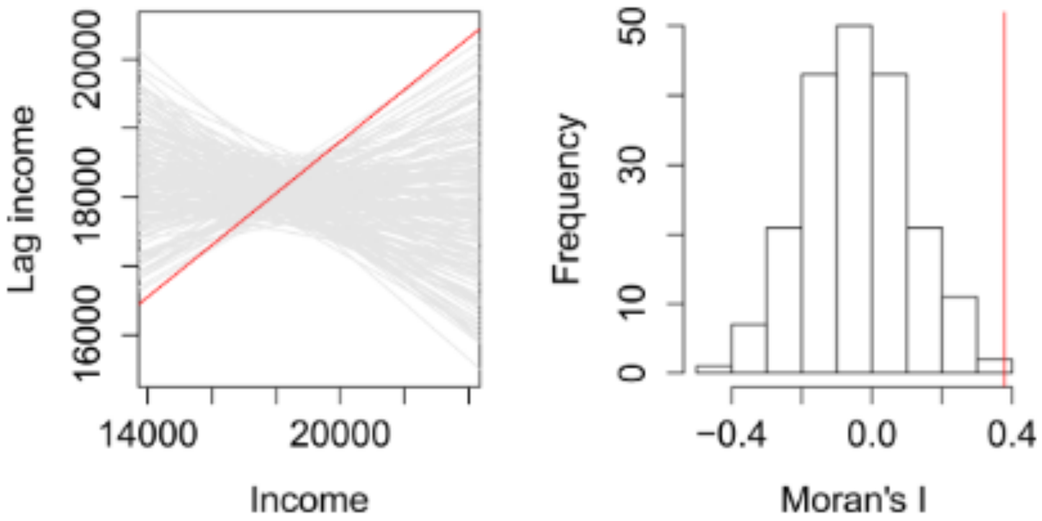
# We need to test whether the pattern is significant

Null hypothesis is a random pattern

*https://mgimond.github.io/Spatial/spatial-autocorrelation.html*

# We need to test whether the pattern is significant

Null hypothesis is a random pattern




$$P_{pseudo} = \frac{N_{extreme} + 1}{N + 1}$$

*https://mgimond.github.io/Spatial/spatial-autocorrelation.html*

# Applications of spatial autocorrelation

Any analysis where we need to understand if a *spatial* process is taking place

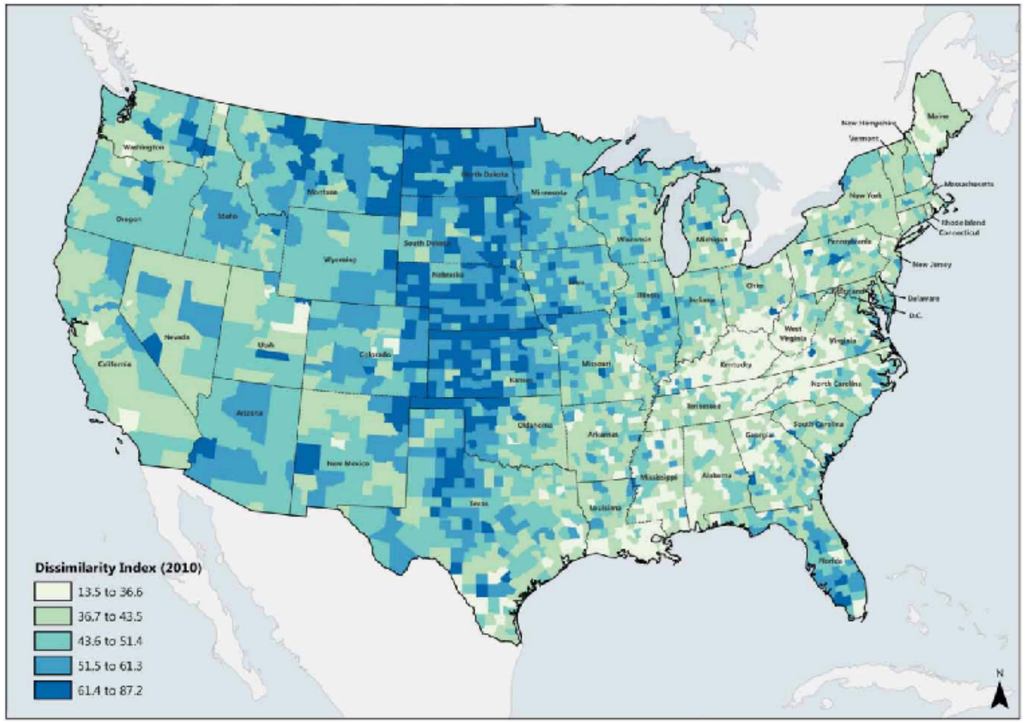



Figure 1. Residential segregation between older and younger adults within counties.

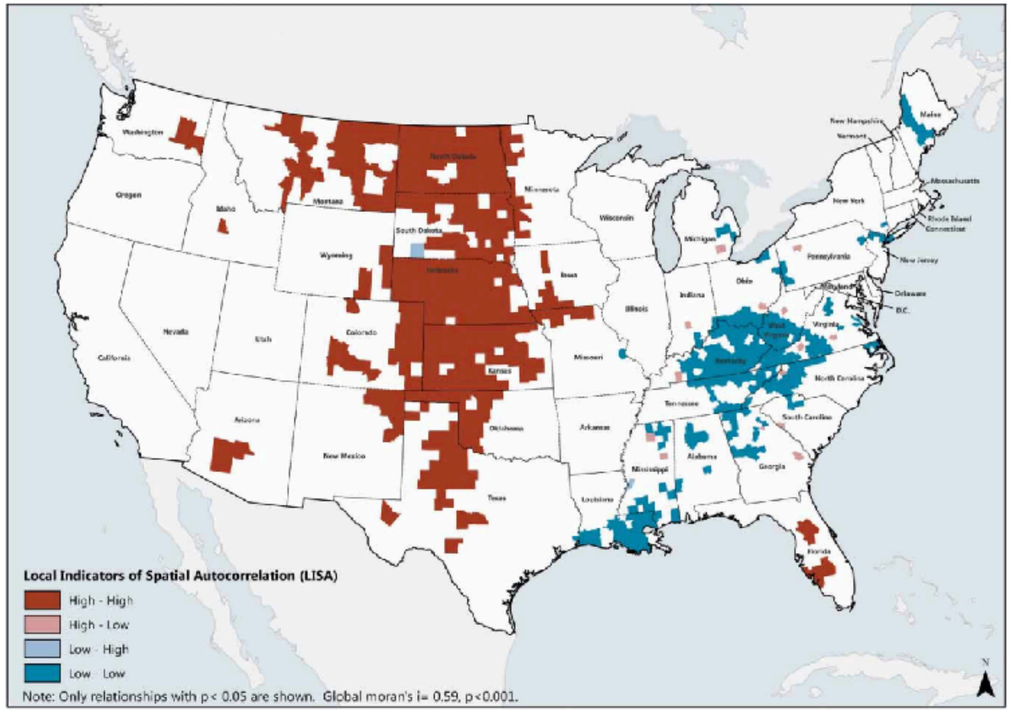



Figure 2. Clusters of age segregation: dissimilarity index of blocks within counties.

# Applications of spatial autocorrelation

Any analysis where we need to understand if places are significantly different

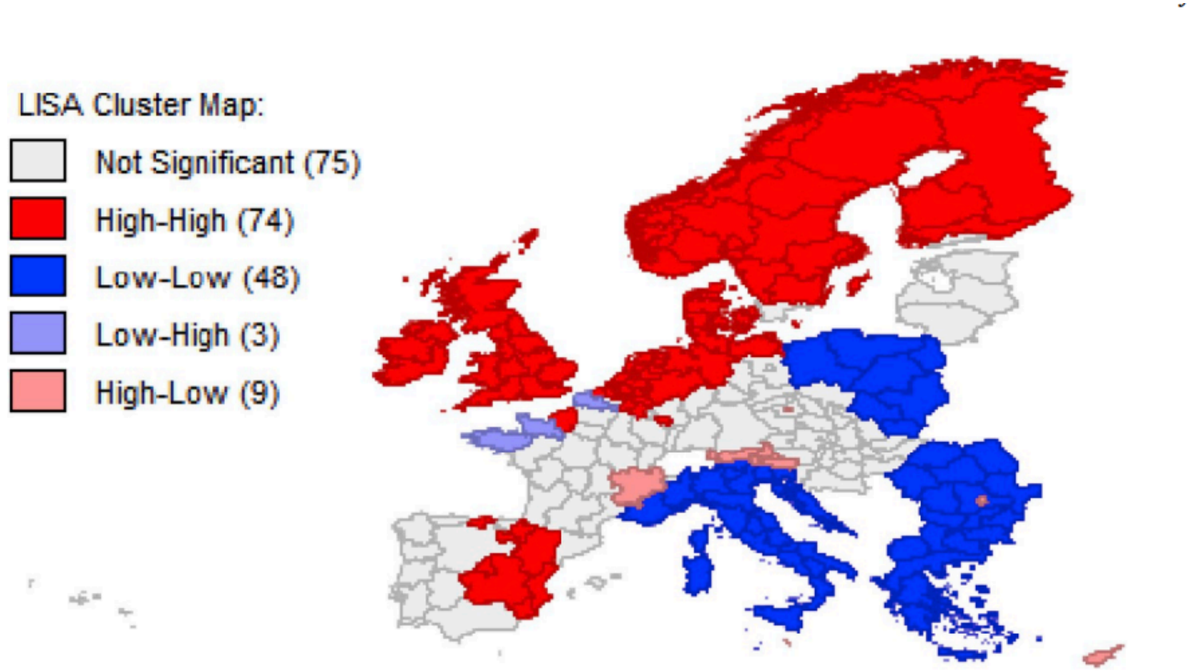



**Fig. 12.** 2016 proportion of population with on-the-go internet access (n = 209, Moran's I = 0.682).

# New library: esda (exploratory spatial data analysis)

Part of PySAL family

Data to examine for spatial autocorrelation (gdf)

Spatial weights

```
mi = esda.Moran(br['Pct_Leave'], w)
```

```
lisa = esda.Moran_Local(br['Pct_Leave'], w)
```

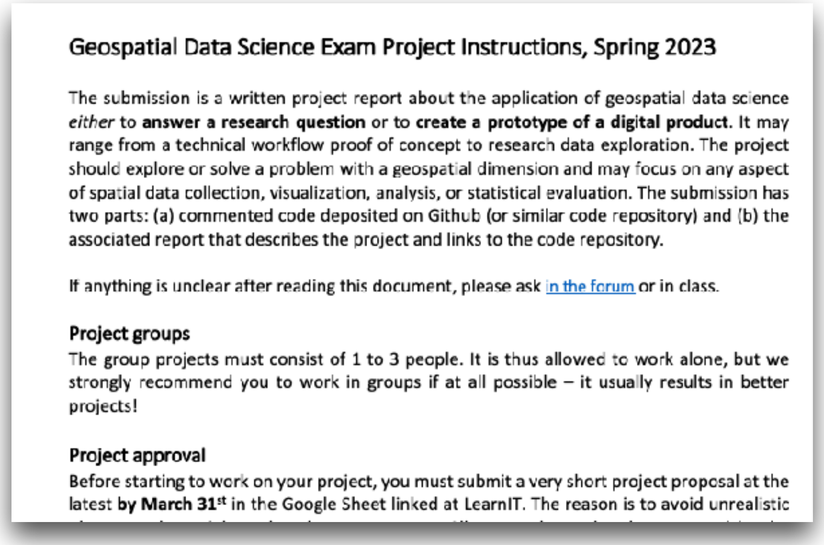
https://pysal.org/esda

# Jupyter

# Project Exam

- Analysis answering a research question *or* prototype of digital product
- Groups (1-3 people) - remember contribution statement
- Hand in report *and* code/product
- LaTeX is recommended (template suggestion on LearnIT)

Detailed instructions on LearnIT



**Geospatial Data Science Exam Project Instructions, Spring 2023**

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 unclear after reading this document, please ask in the forum or in class.

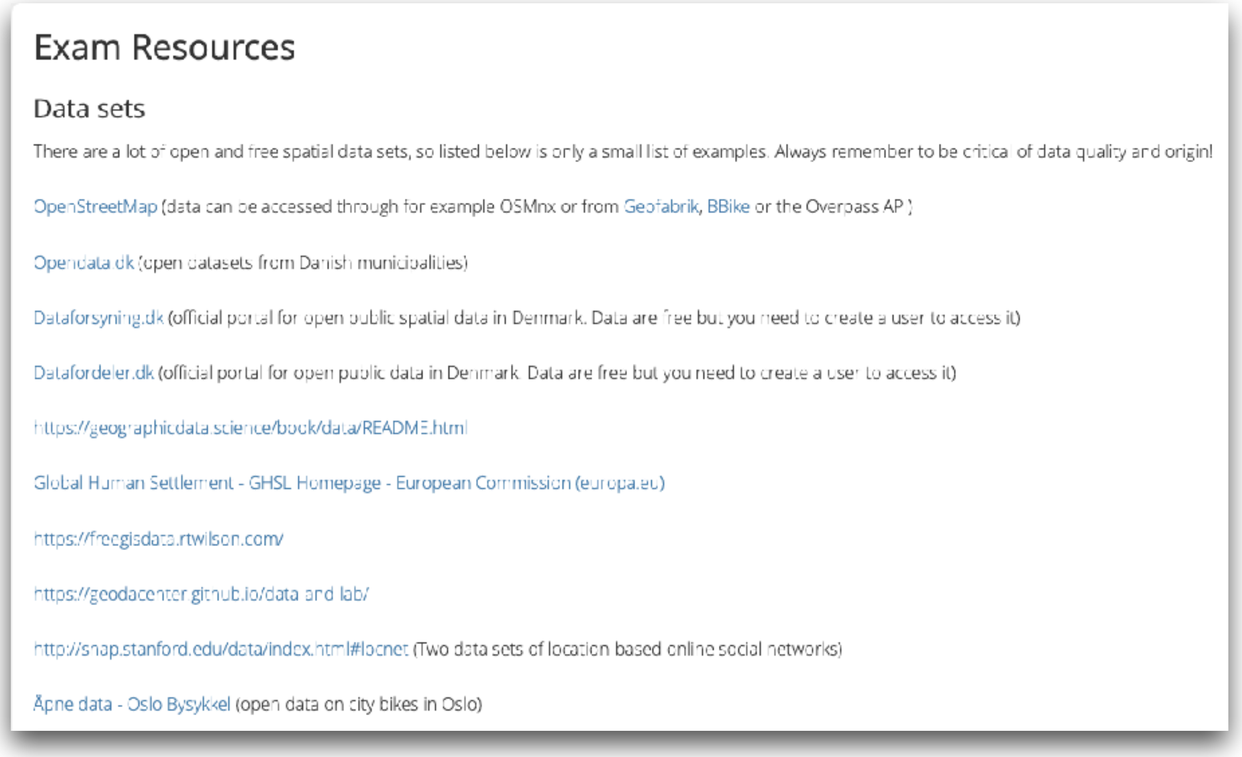
**Project groups**

The group projects must consist of 1 to 3 people. It is thus allowed to work alone, but we strongly recommend you to work in groups if at all possible – it usually results in better projects!

**Project approval**

Before starting to work on your project, you must submit a very short project proposal at the latest **by March 31st** in the Google Sheet linked at LearnIT. The reason is to avoid unrealistic

# Use the exam resources



## Exam Resources

### Data sets

There are a lot of open and free spatial data sets, so listed below is only a small list of examples. Always remember to be critical of data quality and origin!

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

Opendata.dk (open datasets from Danish municipalities)

Dataforsyning.dk (official portal for open public spatial data in Denmark. Data are free but you need to create a user to access it)

Datafordeler.dk (official portal for open public data in Denmark. Data are free but you need to create a user to access it)

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

Global Human Settlement - GHSL Homepage - European Commission (europa.eu)

https://freegisdata.rtwilson.com/

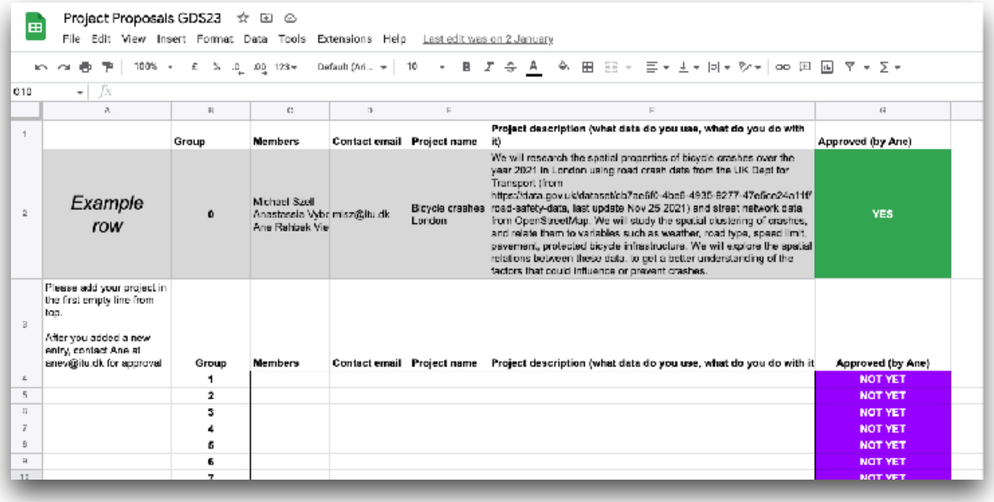
https://geodacenter.github.io/data-and-lab/

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

Åpne data - Oslo Bysykkel (open data on city bikes in Oslo)

Ask questions if you’re stuck or in doubt!

# Project Exam



Project Proposals GDS23

File Edit View Insert Format Data Tools Extensions Help Last edit was on 2 January

| | Group | Members | Contact email | Project name | Project description (what data do you use, what do you do with it) | Approved (by Ane) |
|---|---|---|---|---|---|---|
| *Example row* | 0 | Michael Szell<br>Anastasios Vyb<br>Ane Rahbek Vie | m.sz@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/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/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.<br><br>After you added a new entry, contact Ane at anev@itu.dk for approval | Group | Members | Contact email | Project name | Project description (what data do you use, what do you do with it | Approved (by Ane) |
| | 1 | | | | | NOT YET |
| | 2 | | | | | NOT YET |
| | 3 | | | | | NOT YET |
| | 4 | | | | | NOT YET |
| | 5 | | | | | NOT YET |
| | 6 | | | | | NOT YET |
| | 7 | | | | | NOT YET |

Deadline for project proposals: March 31

Deadline for project approvals: April 11

Exam hand-in: May 26

# Intended Learning Outcomes (ILOs)

- Demonstrate GIS/GDS concepts and be able to use relevant Python libraries programmatically to import, manipulate and analyze spatial data in different formats. Apply a number of spatial analysis techniques and explain how to interpret the results, in a process of turning data into insights.
- Reflect on the motivation and inner workings of the main methodological approaches of GDS, both analytical and visual.
- Critically evaluate the suitability of a specific GDS technique, what it can offer and how it can help answer questions of interest.
- Apply a number of spatial analysis techniques and explain how to interpret the results, in a process of turning data into insights.
- When faced with a new data-set, work independently using GIS/GDS tools programmatically to extract valuable insight.

https://learnit.itu.dk/local/coursebase/view.php?ciid=1170

# Intended Learning Outcomes (ILOs)

- **Demonstrate GIS/GDS concepts and be able to use relevant Python libraries** programmatically to import, manipulate and analyze spatial data in different formats. Apply a number of spatial analysis techniques and explain how to interpret the results, in a process of turning data into insights.

- **Reflect on the motivation and inner workings** of the main methodological approaches of GDS, both analytical and visual.

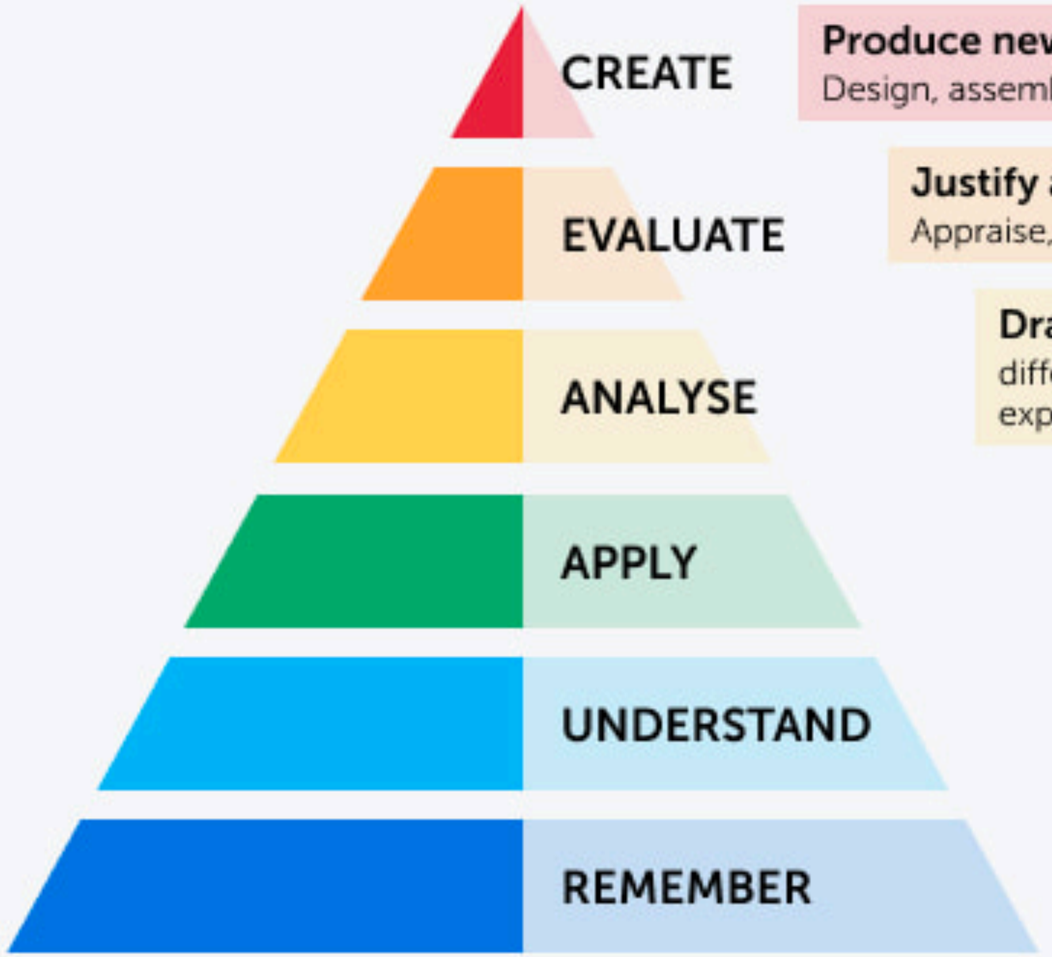
- **Critically evaluate the suitability** of a specific GDS technique, what it can offer and how it can help answer questions of interest.

- **Apply a number of spatial analysis techniques and explain** how to interpret the results, in a process of turning data into insights.

- When faced with a new data-set, work independently using GIS/GDS tools programmatically to **extract valuable insight.**

https://learnit.itu.dk/local/coursebase/view.php?ciid=1170

# Bloom’s Taxonomy




**Produce new or original work**
Design, assemble, construct, conjecture, develop, formulate, author, investigate

**Justify a stand or decision**
Appraise, argue, defend, judge, select, support, value, critique, weigh

**Draw connections among ideas**
differentiate, organise, relate, compare, contrast, distinguish, examine, expertiment, question, test

**Use information in new situation**
Execute, implement, solve, use, demonstrate, interpret, operate, schedule, sketch

**Explain ideas or concepts**
Classify, discribe, discuss, explain, identify, locate, recognize, report, select, translate

**Recall facts and basic concepts**
define duplicate, list, memorise, repeat, state

# 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

https://mgimond.github.io/Spatial/spatial-autocorrelation.html

# Next week: Spatial Clustering