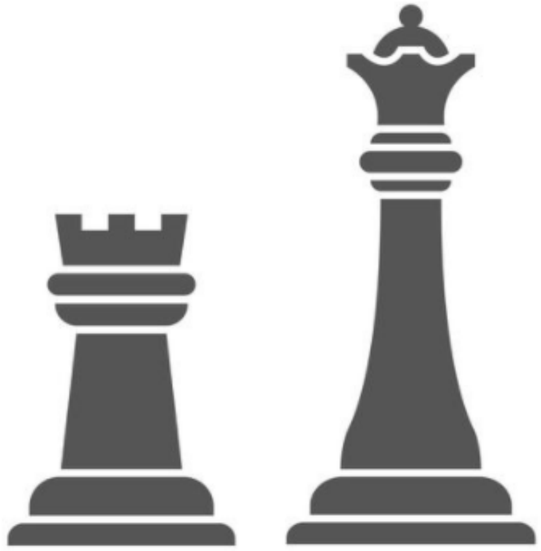
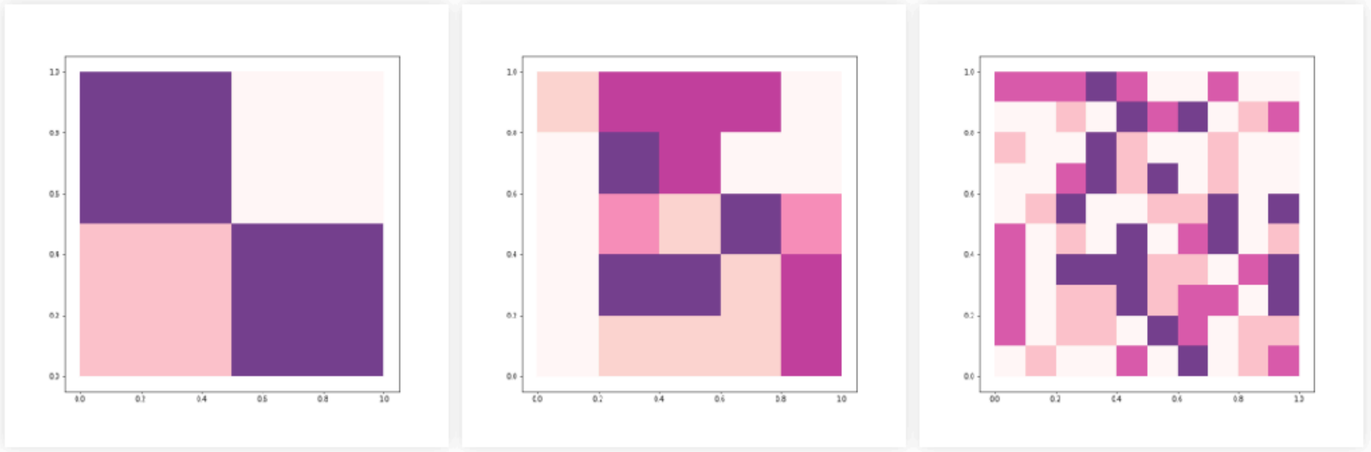
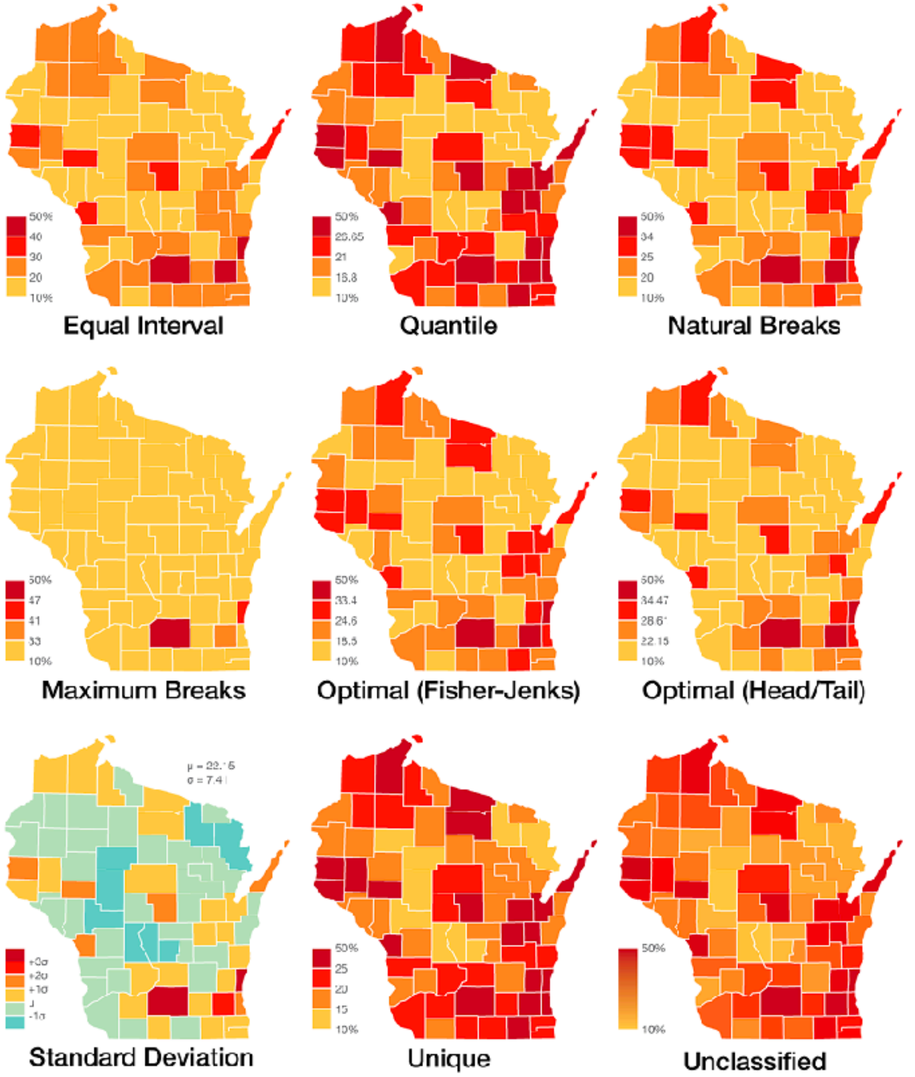
Lecture 4: Spatial weights

Instructor: Ane Rahbek Vierø
Feb 20, 2022



Questions from last week?



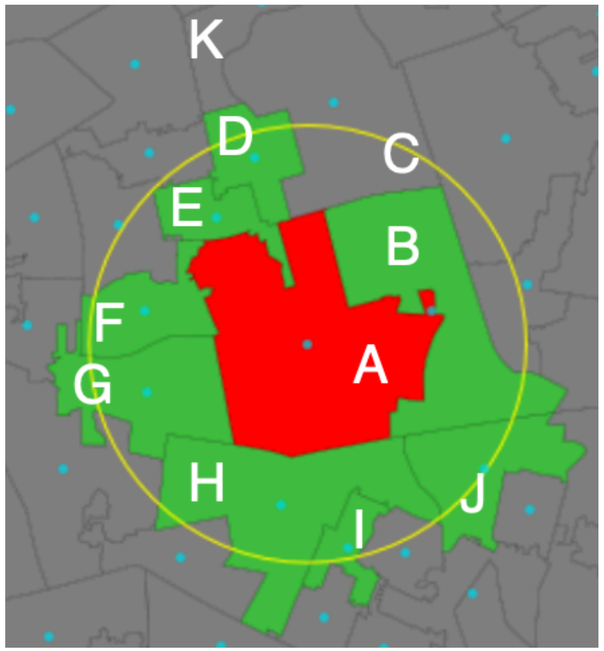


GDS asks: How do things relate in space?

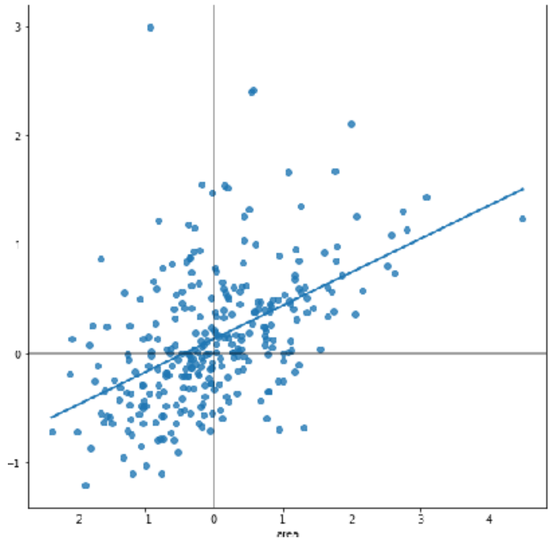
> Everything is related to everything else, but near things are more related than distant things.
>
> Tobler's 1st law of geography

Today you will learn about spatial weights

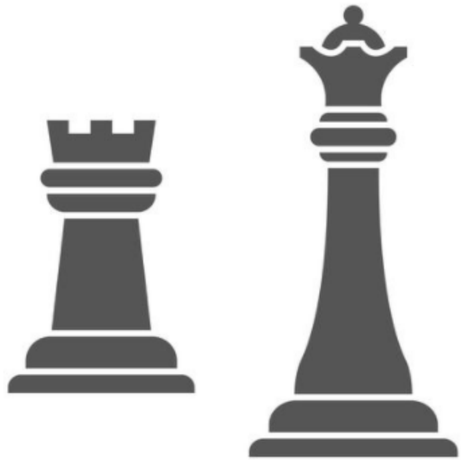
What they are,
why they are important



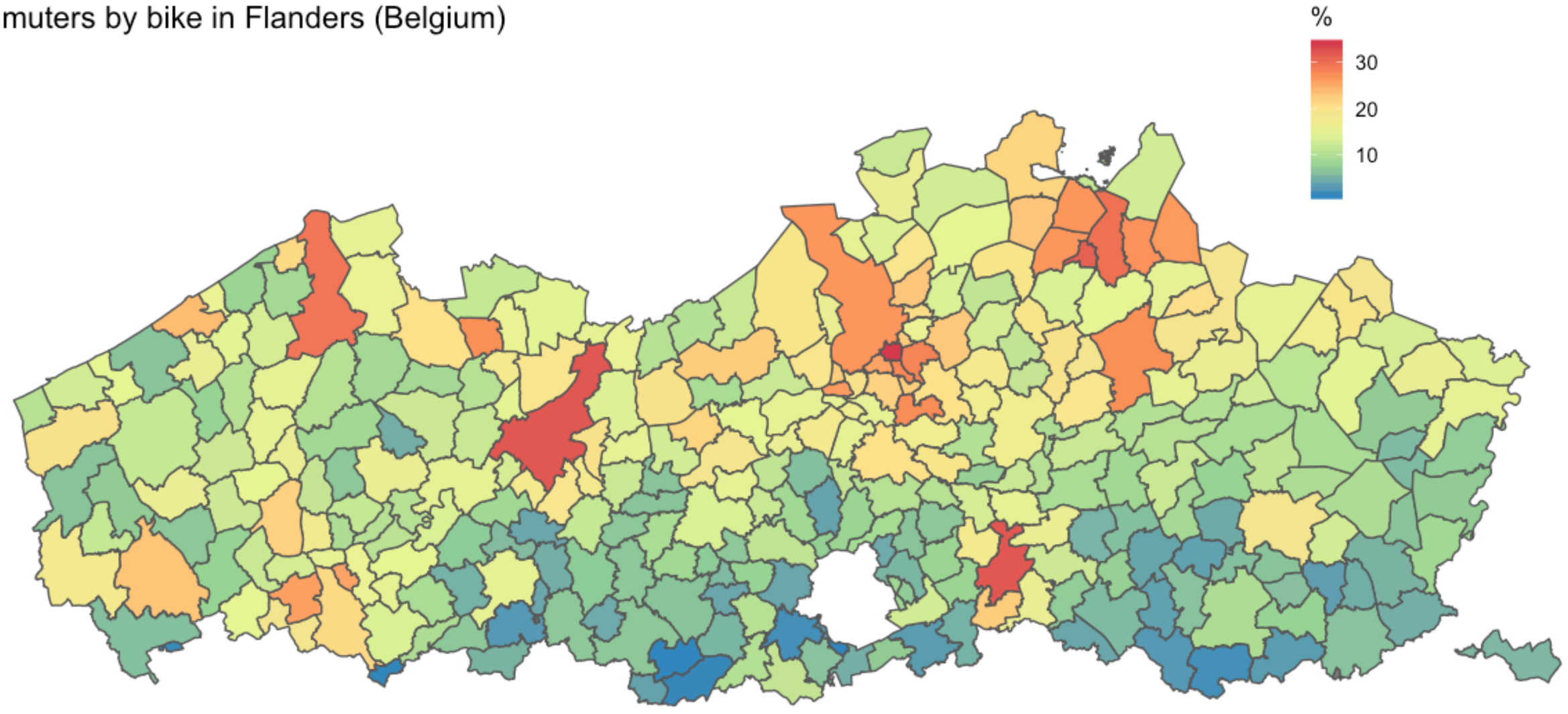
How to use them
in Python



Ways of defining them



Commuters by bike in Flanders (Belgium)



Data: Gemeentemonitor 2017

“Spatial weights … are widely used constructs that represent geographic relationships between the observational units in a spatially referenced dataset.”

Rey, Arribas-Bel, Wolf

How are objects related to each other in space?

The spatial weight matrix W encodes the spatial relation between N objects

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

N times N, positive

$$w_{ii} = 0$$

Can be symmetrical $w_{ij} \neq w_{ji}$

How are objects related to each other in space?

The spatial weight matrix W encodes the spatial relation between N objects

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

N times N, positive

$$w_{ii} = 0$$

Discrete values are used for a binary definition of neighbors (0/1)
Continuous values indicate the strength of the neighbor relationship

How are objects related to each other in space?

The spatial weight matrix W encodes the spatial relation between N objects

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

N times N, positive

$$w_{ii} = 0$$

Generally, all non-zero elements in a row *i* are called the neighbors of object *i*. How to define ‘neighbor’?

How are objects related to each other in space?

The spatial weight matrix W encodes the spatial relation between N objects

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

N times N, positive

$$w_{ii} = 0$$

Contiguity Is object 2 "next to" object 1?

How are objects related to each other in space?

The spatial weight matrix W encodes the spatial relation between N objects

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

N times N, positive

$$w_{ii} = 0$$

Contiguity — Is object 2 "next to" object 1?

Distance — Is object 2 "close" to object 1?

How are objects related to each other in space?

The spatial weight matrix W encodes the spatial relation between N objects

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

N times N, positive

$$w_{ii} = 0$$

Contiguity	Is object 2 "next to" object 1?
Distance	Is object 2 "close" to object 1?
Block	Is object 2 in the same "place" as object 1?

How are objects related to each other in space?

The spatial weight matrix W encodes the spatial relation between N objects

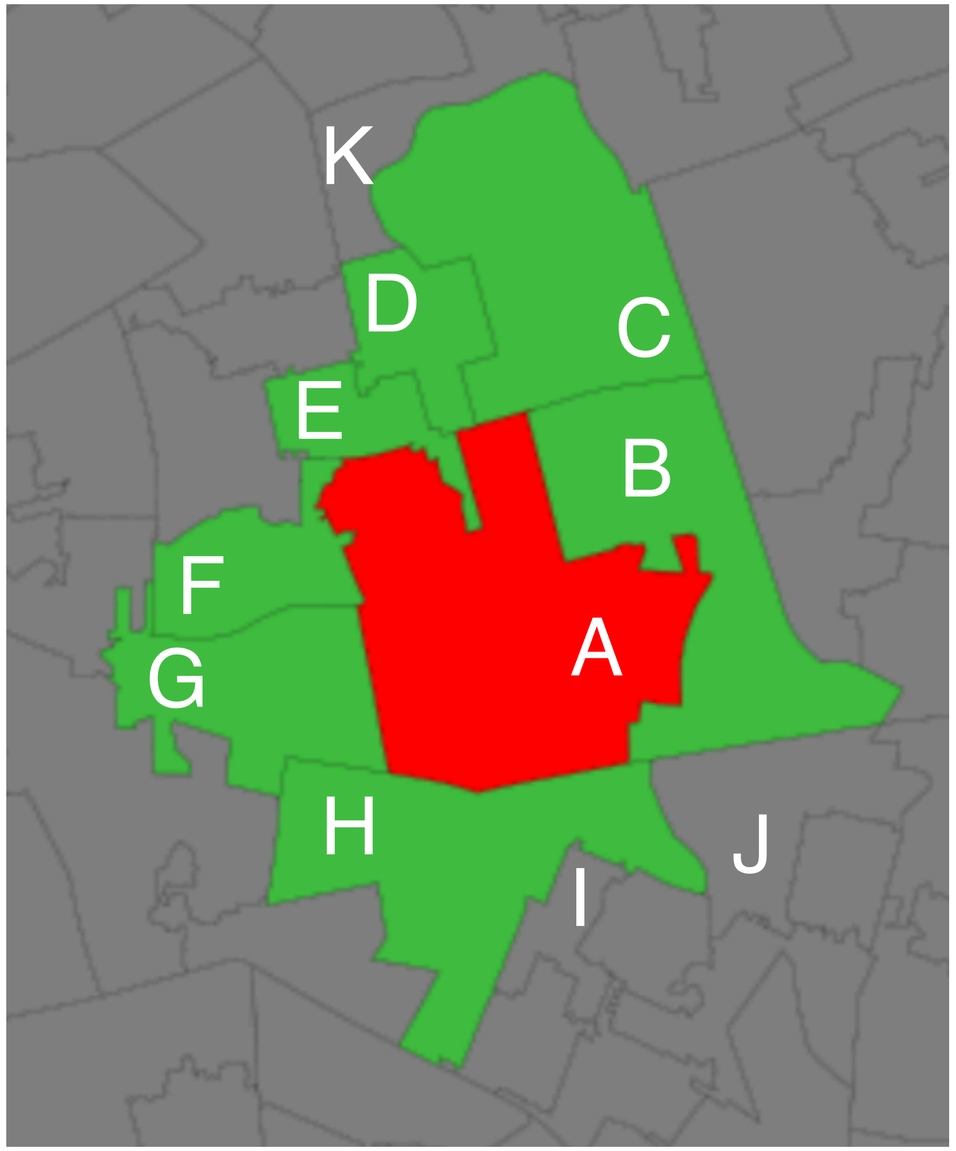
$$W = \begin{pmatrix} 0 & w_{12} & \cdots & w_{1N} \\ w_{21} & \ddots & w_{ij} & \vdots \\ \vdots & w_{ji} & 0 & \vdots \\ w_{N1} & \cdots & \cdots & 0 \end{pmatrix}$$

N times N, positive

$$w_{ii} = 0$$

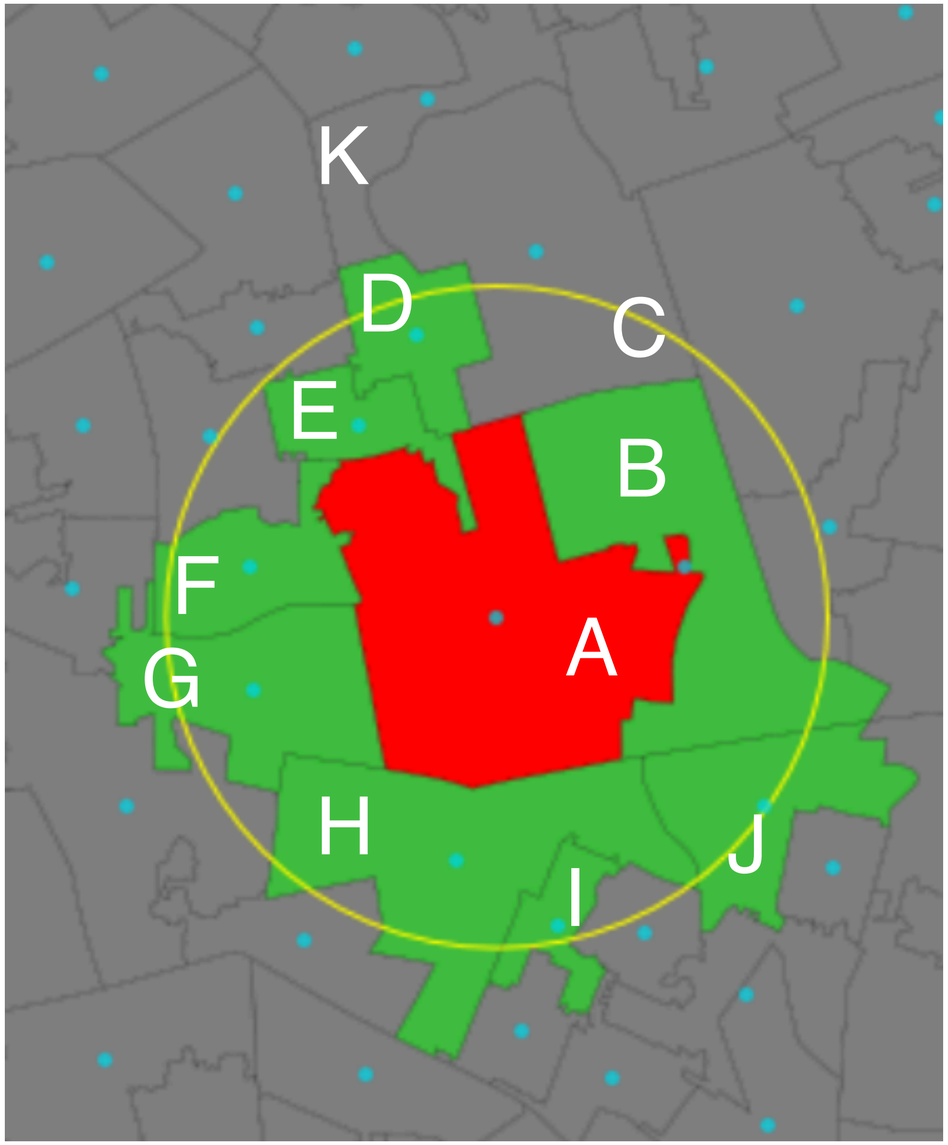
Similar to a graph adjacency matrix

Contiguity-based W: sharing a boundary



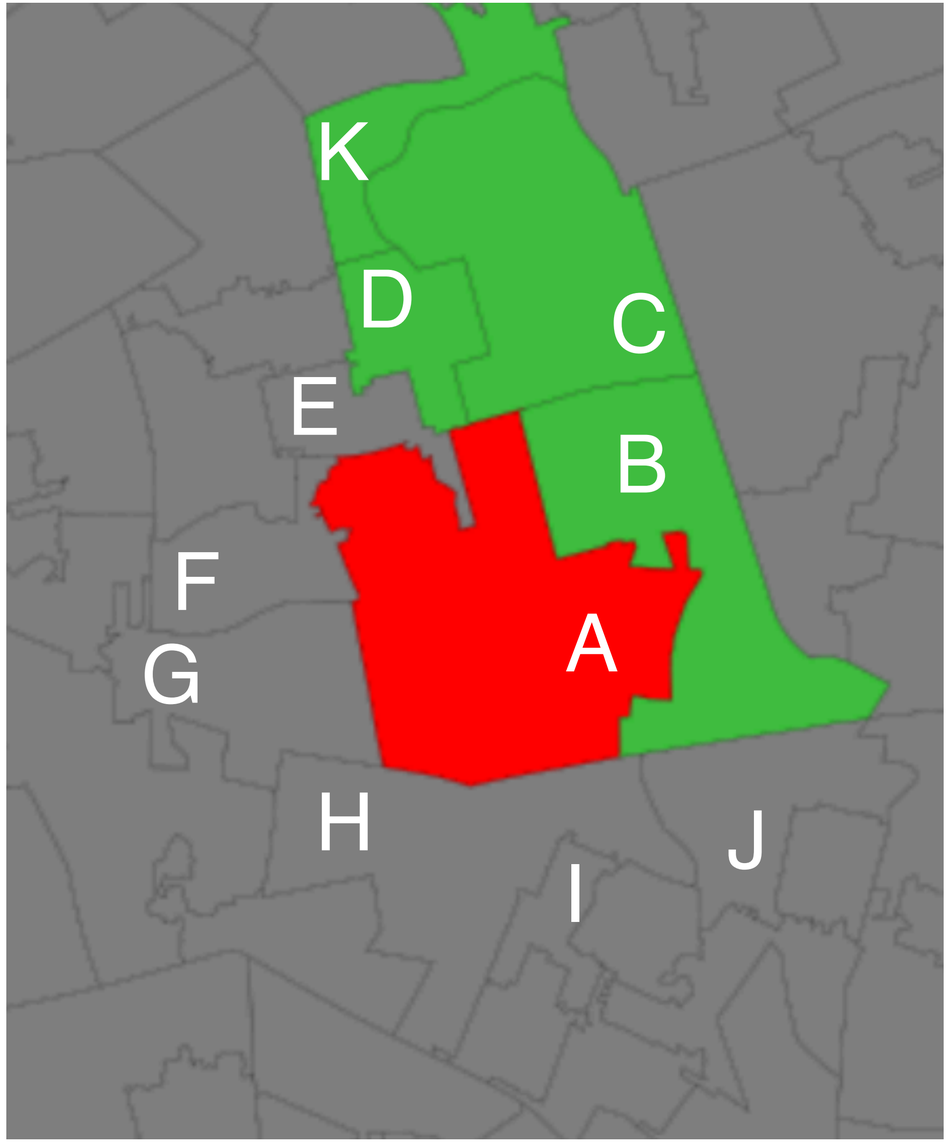
	A	B	C	D	E	F	G	H	I	J	K
A	0	1	1	1	1	1	1	1	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...

Distance-based W: being closer than a threshold



	A	B	C	D	E	F	G	H	I	J	K
...	...	...	...	...	...	...	...	...	...	...	...
A	0	1	0	1	1	1	1	1	1	1	0
...	...	...	...	...	...	...	...	...	...	...	...

Block-based W: being in the same administrative unit



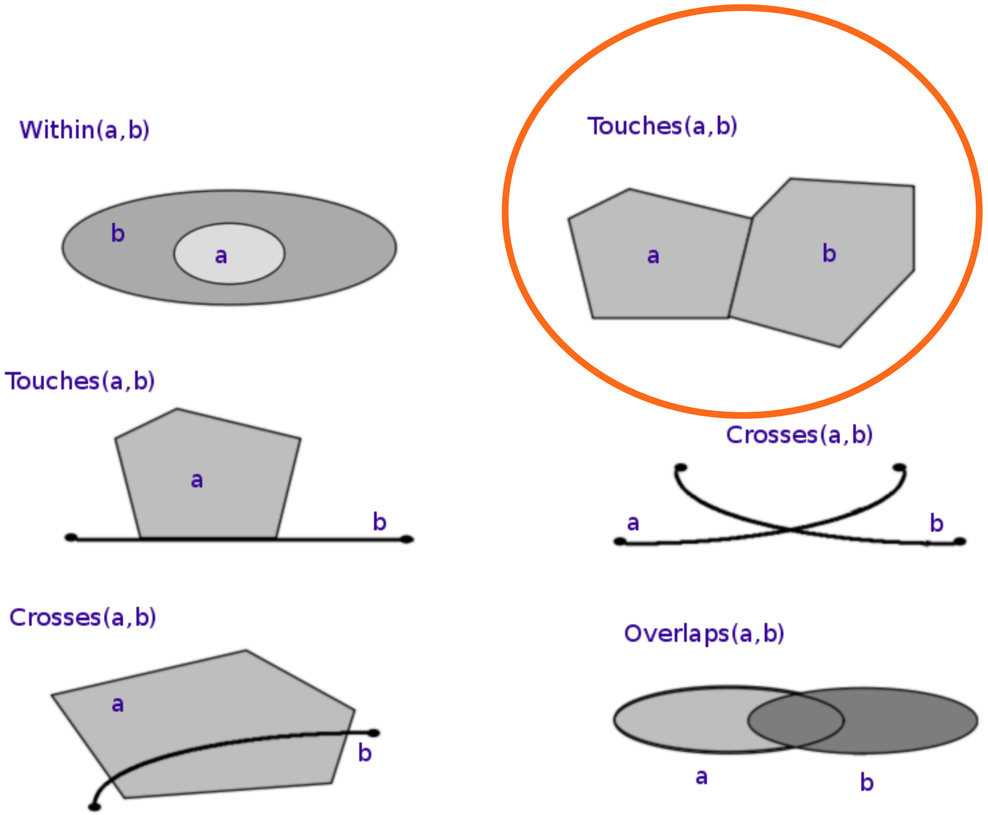
	A	B	C	D	E	F	G	H	I	J	K
...	...	...	...	...	...	...	...	...	...	...	...
A	0	1	1	1	0	0	0	0	0	0	1
...	...	...	...	...	...	...	...	...	...	...	...

Special cases and variations

Geospatial topology is the study and application of qualitative spatial relationships between geographic features

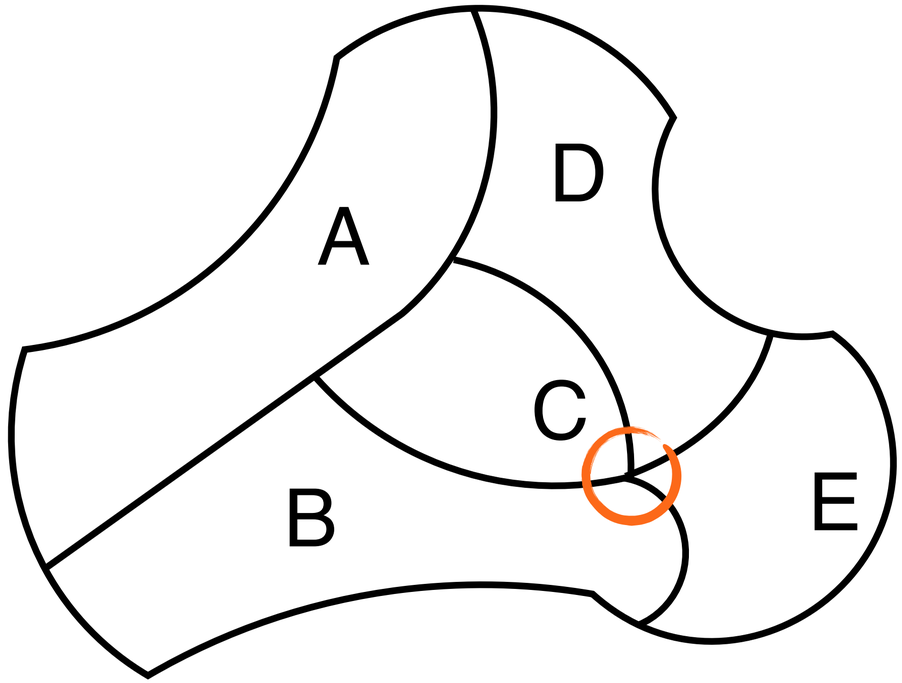
https://en.wikipedia.org/wiki/Geospatial_topology

Geospatial topology is the study and application of qualitative spatial relationships between geographic features



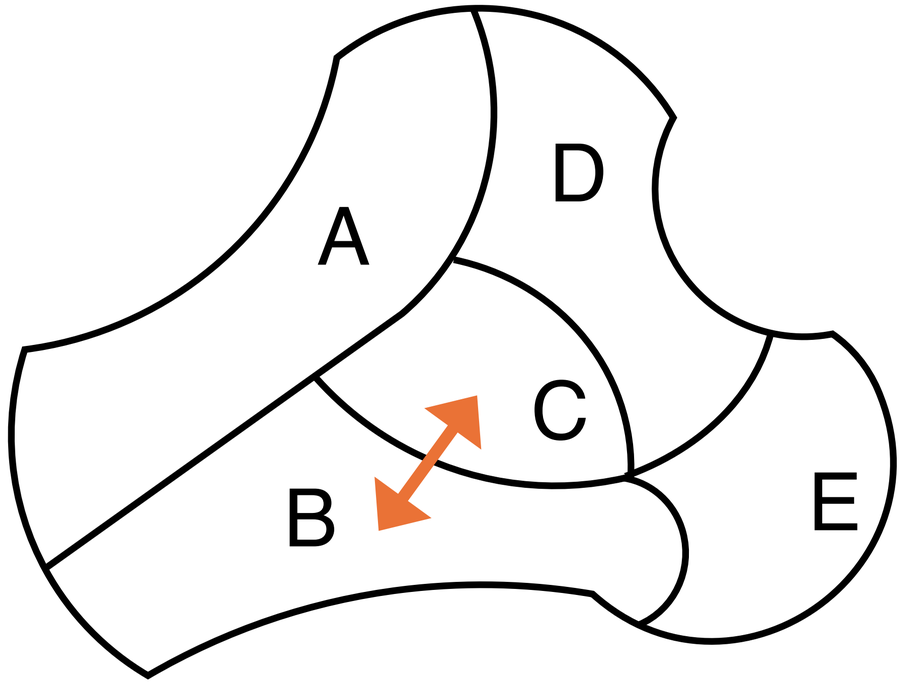
https://en.wikipedia.org/wiki/Geospatial_topology

Contiguity-based W: What is a boundary?



	A	B	C	D	E
A	0	1	1	1	0
B	1	0	1	?	1
C	1	1	0	1	?
D	1	?	1	0	1
E	0	1	?	1	0

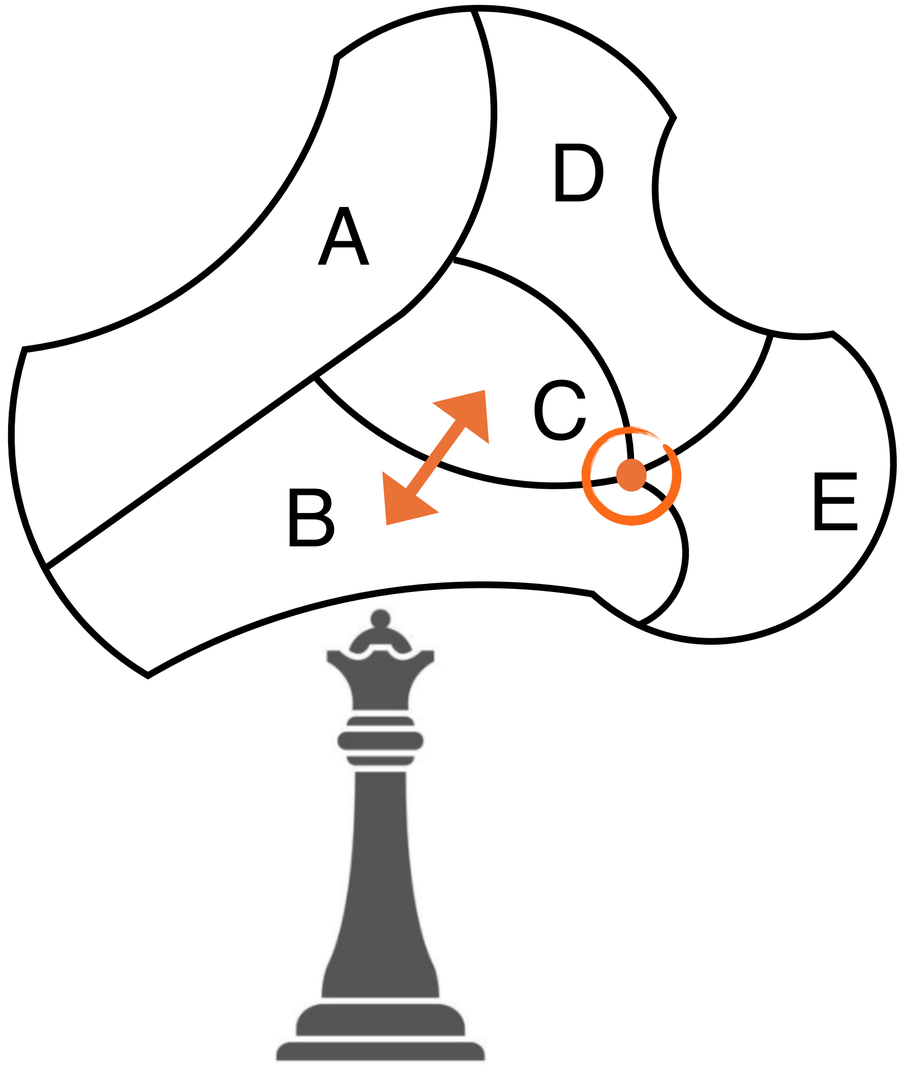
Contiguity-based W: What is a boundary?



ROOK

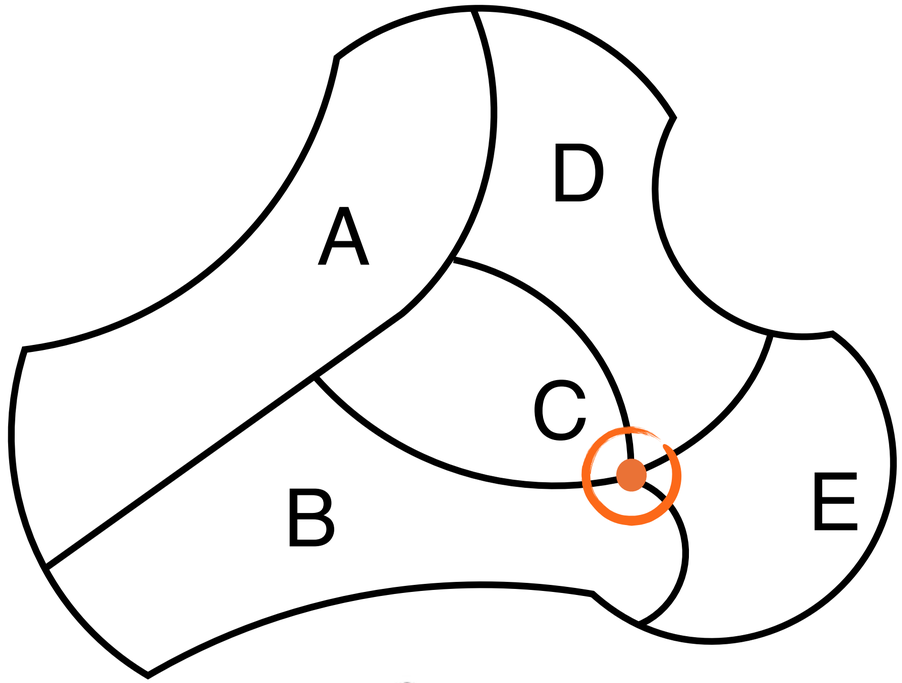
	A	B	C	D	E
A	0	1	1	1	0
B	1	0	1	0	1
C	1	1	0	1	0
D	1	0	1	0	1
E	0	1	0	1	0

Contiguity-based W: What is a boundary?



	A	B	C	D	E
A	0	1	1	1	0
B	1	0	1	1	1
C	1	1	0	1	1
D	1	1	1	0	1
E	0	1	1	1	0

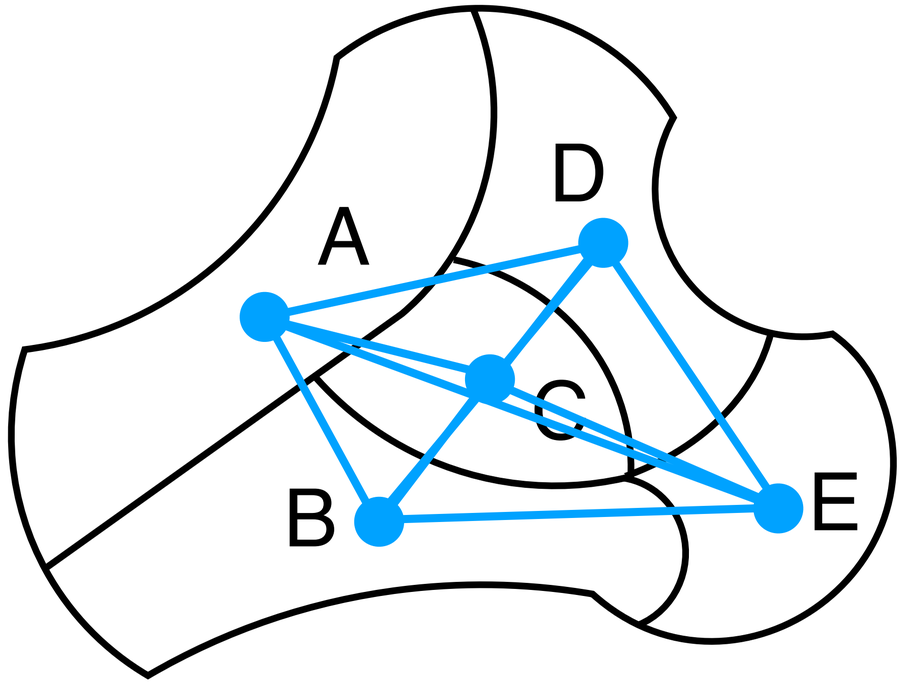
Contiguity-based W: What is a boundary?



BISHOP

	A	B	C	D	E
A	0	0	0	0	0
B	0	0	0	1	0
C	0	0	0	0	1
D	0	1	0	0	0
E	0	0	1	0	0

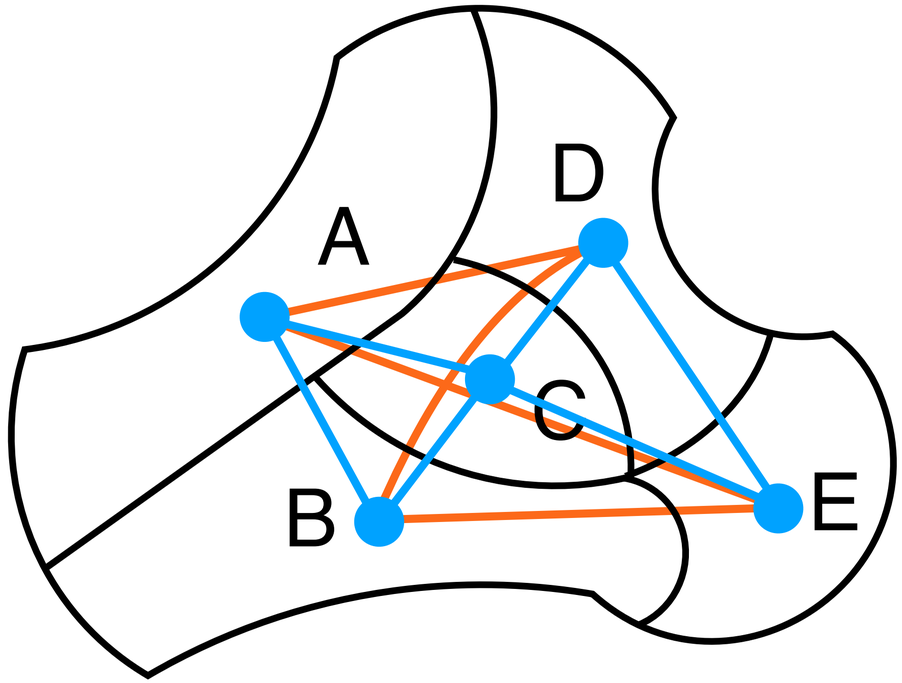
Distance-based W: Values can be continuous



$$w_{ij} = \frac{1}{d_{ij}} \quad \text{or} \quad w_{ij} = \frac{1}{d_{ij}^{P}}$$

	A	B	C	D	E
A	0	0.58	0.57	0.39	0.25
B	0.58	0	0.76	0.38	0.33
C	0.57	0.76	0	0.76	0.43
D	0.39	0.38	0.76	0	0.42
E	0.25	0.33	0.43	0.42	0

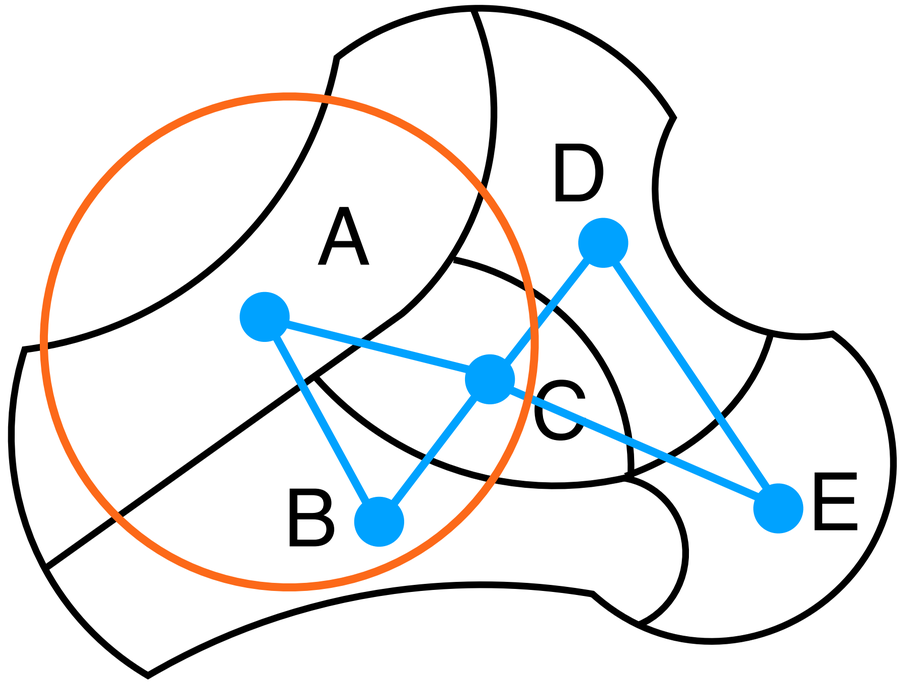
Distance-based W: Values can be continuous



We want a threshold
to keep W sparse!

t=0.4	A	B	C	D	E
A	0	0.58	0.57	0	0
B	0.58	0	0.76	0	0
C	0.57	0.76	0	0.76	0.43
D	0	0	0.76	0	0.42
E	0	0	0.43	0.42	0

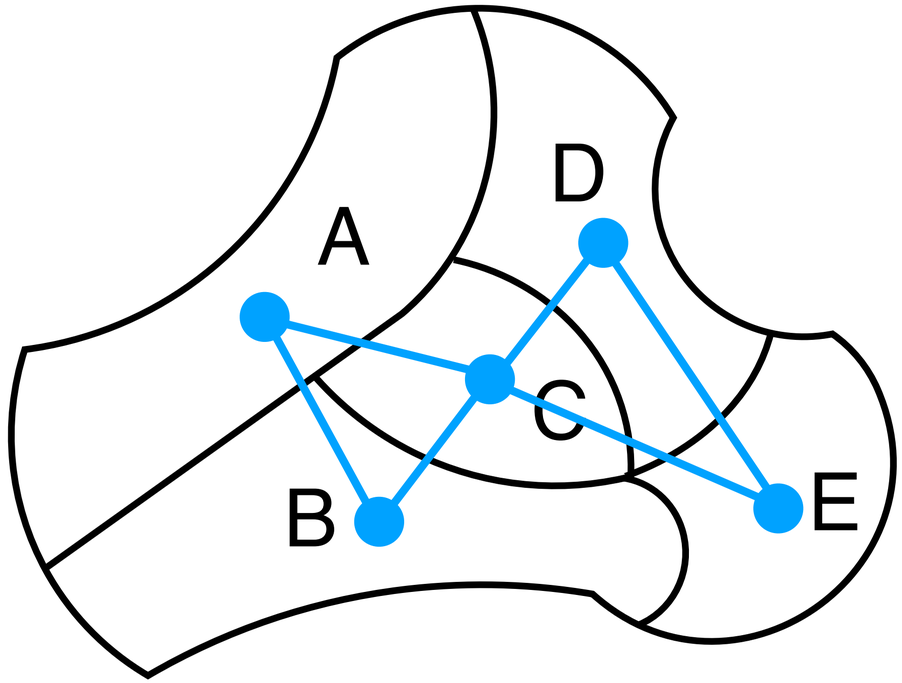
Distance-based W: Values can be discrete



Distance bands are used for binary distance classifications

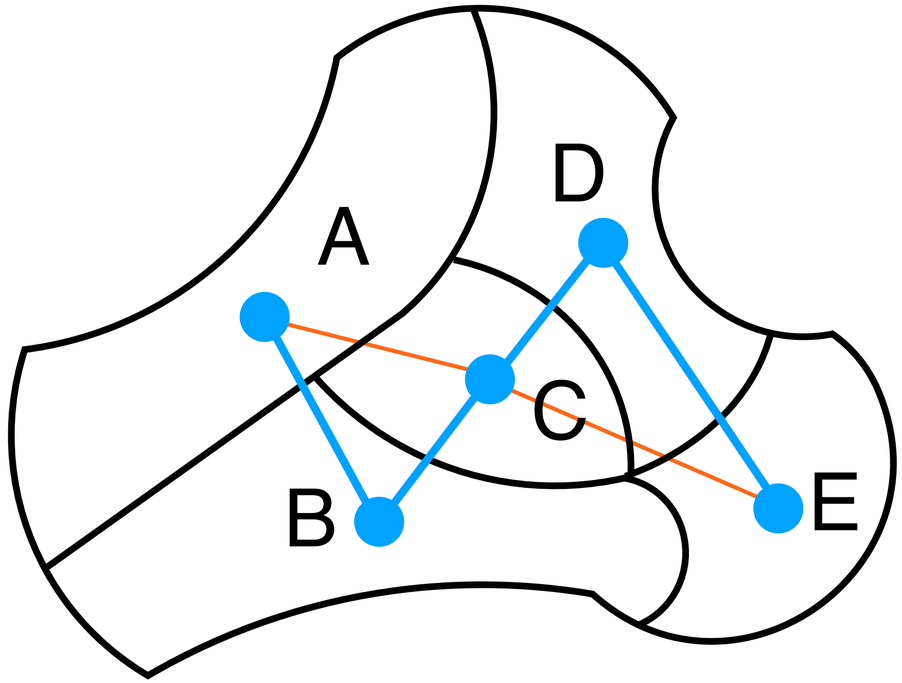
d=3	A	B	C	D	E
A	0	1	1	0	0
B	1	0	1	0	0
C	1	1	0	1	1
D	0	0	1	0	1
E	0	0	1	1	0

Distance-based W: KNN (k nearest neighbors)



k=2	A	B	C	D	E
A	0	1	1	0	0
B	1	0	1	0	0
C	0	1	0	1	0
D	0	0	1	0	1
E	0	0	1	1	0

Distance-based W: KNN (k closest neighbors)

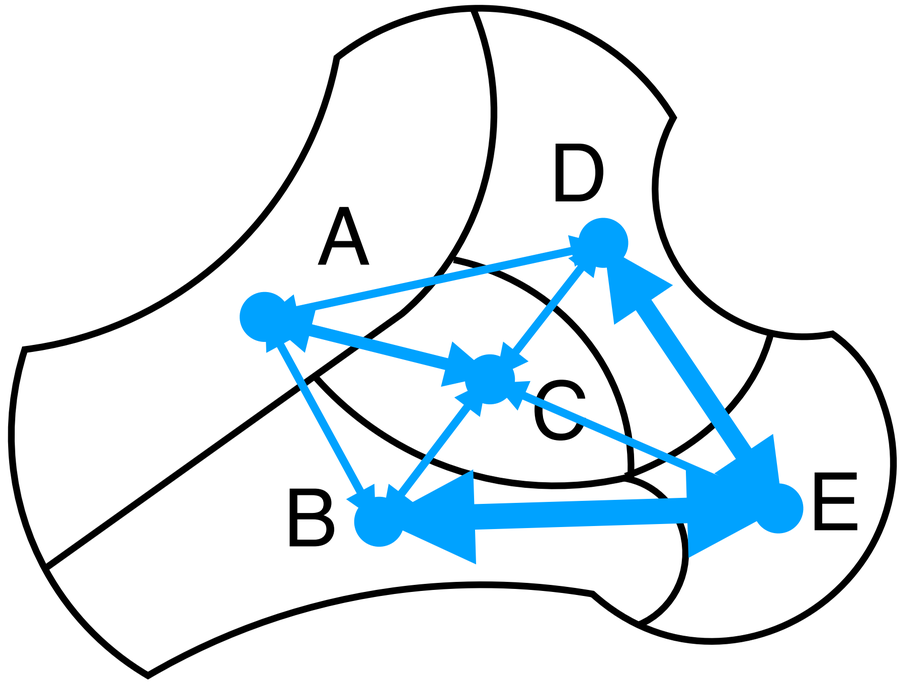


k=2	A	B	C	D	E
A	0	1	0.5	0	0
B	1	0	1	0	0
C	0.5	1	0	1	0.5
D	0	0	1	0	1
E	0	0	0.5	1	0

To make it symmetric, you could use:

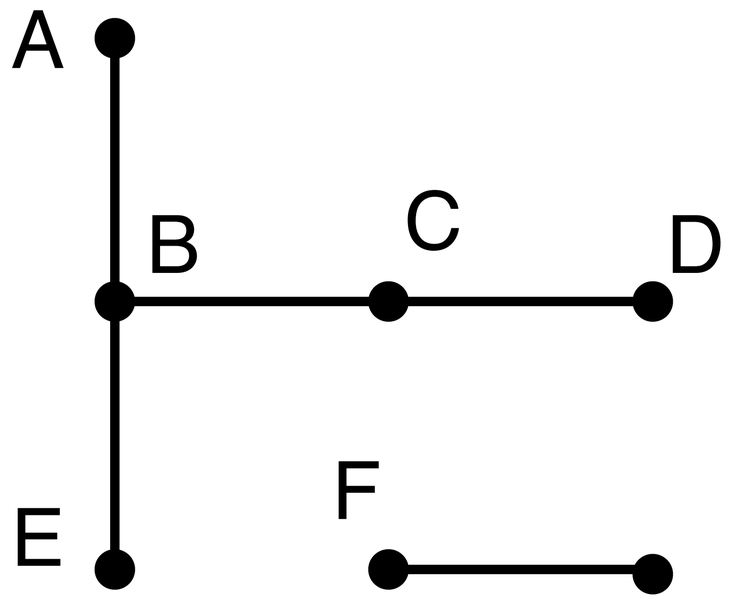
$$(W + W^T)/2$$

Interaction-based W: Flows



	A	B	C	D	E
A	0	1	2	1	0
B	1	0	1	0	5
C	2	1	0	1	1
D	1	0	1	0	2
E	0	4	1	3	0

The structure can be a network

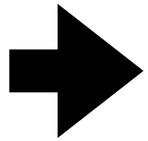


$$w_{ij} = \frac{1}{l_{ij}}$$

	A	B	C	D	E	F
A	0	1	0.5	0.3	0.5	0
B	1	0	1	0.5	1	0
C	0.5	1	0	1	0.5	0
D	0.3	0.5	1	0	0.3	0
E	0.5	1	0.5	0.3	0	0
F	0	0	0	0	0	0

It is common to standardize W: divide all by sum of row

	A	B	C	D	E
A	0	1	2	1	0
B	1	0	1	0	5
C	2	1	0	1	1
D	1	0	1	0	2
E	0	4	1	3	0

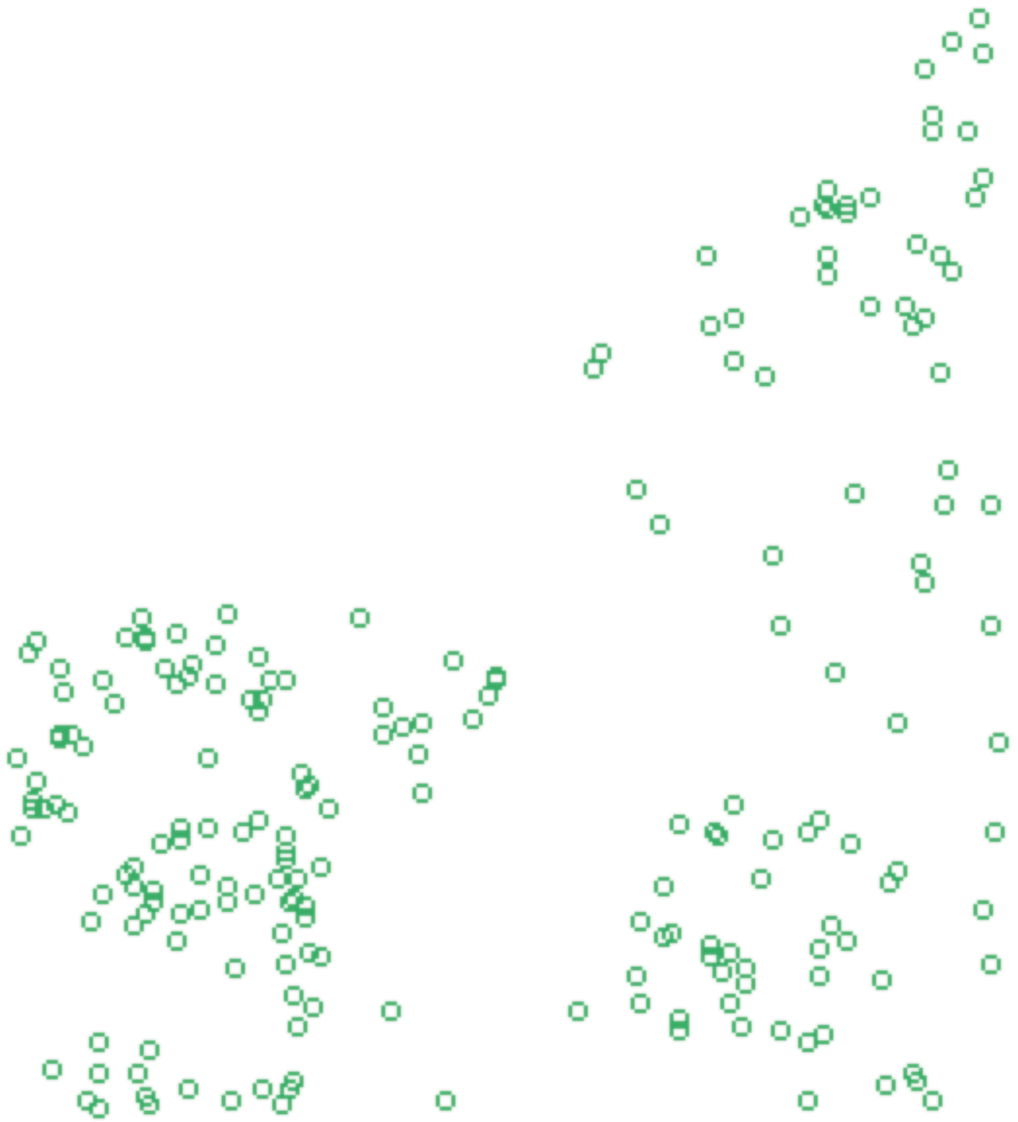


	A	B	C	D	E	Σ
A	0	1/4	1/2	1/4	0	1
B	1/7	0	1/7	0	5/7	1
C	2/5	1/5	0	1/5	1/5	1
D	1/4	0	1/4	0	1/2	1
E	0	1/2	1/8	3/8	0	1

The choice of W should reflect the studied interactions

Spreading processes like a virus	→	Contiguity, Flow
Accessibility	→	Distance
Effects of laws on counties	→	Block

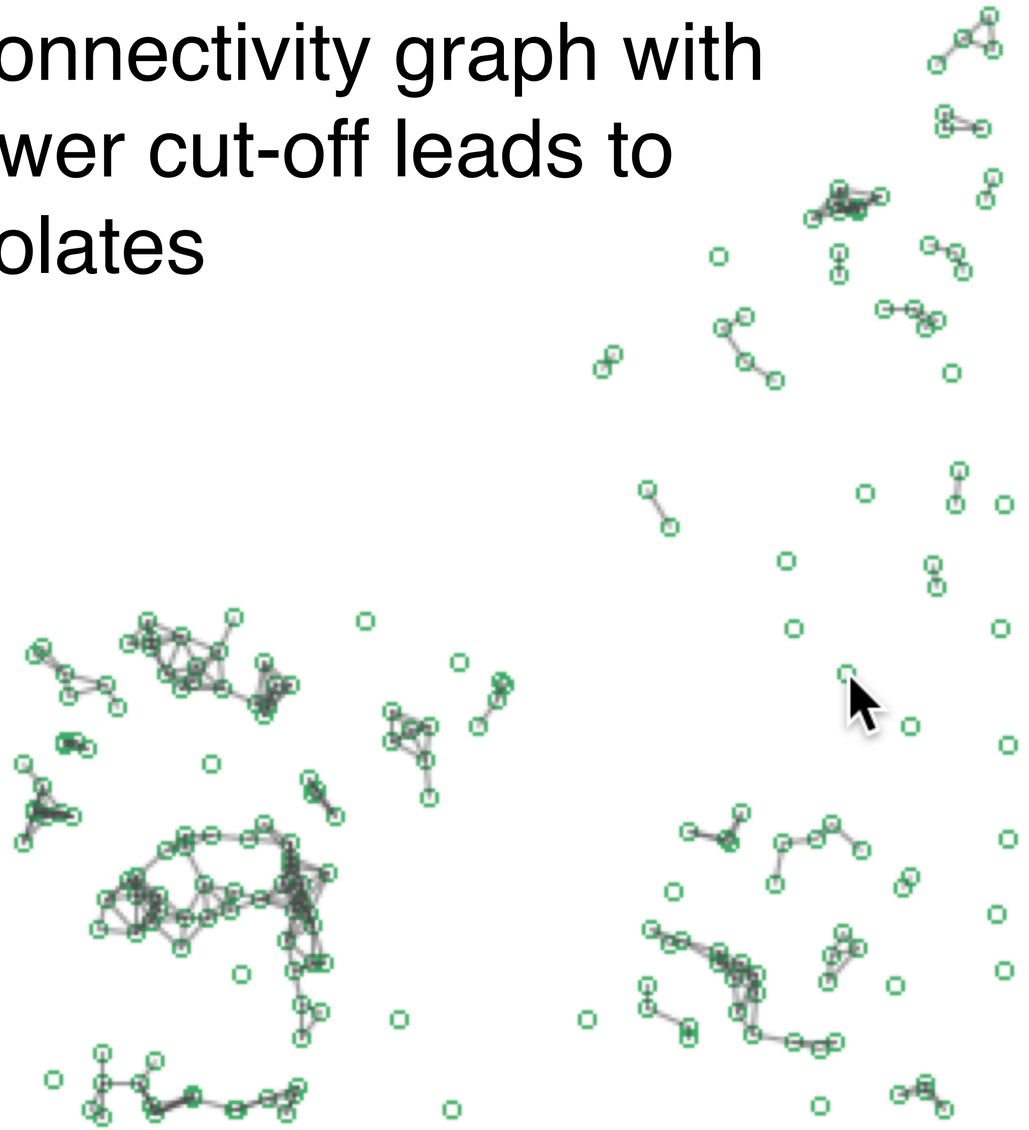
The problem with distance-bands



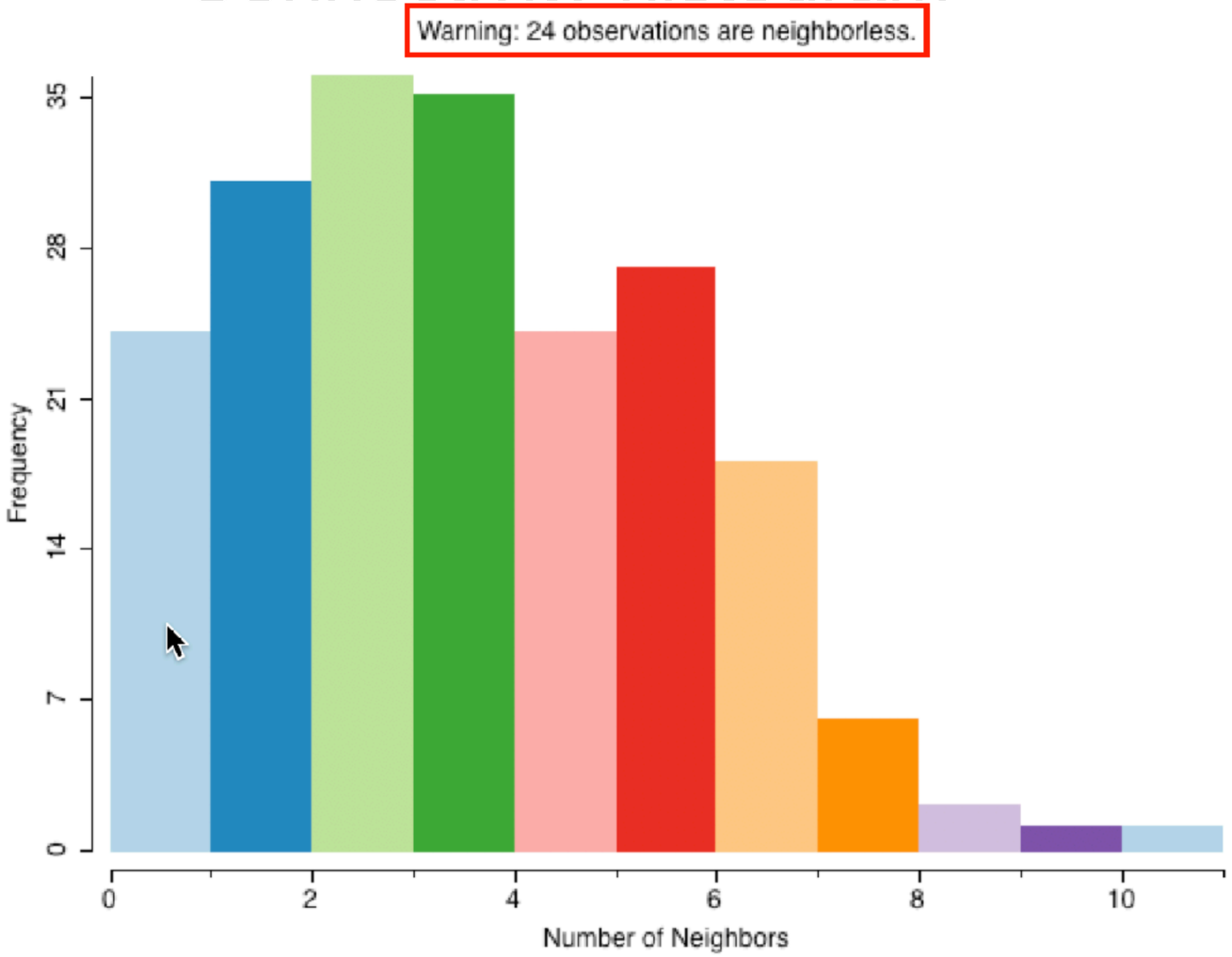
https://geoda.gitee.io/workbook/4b_dist_weights/lab4b.html

The problem with distance-bands

Connectivity graph with lower cut-off leads to isolates



Connectivity histogram

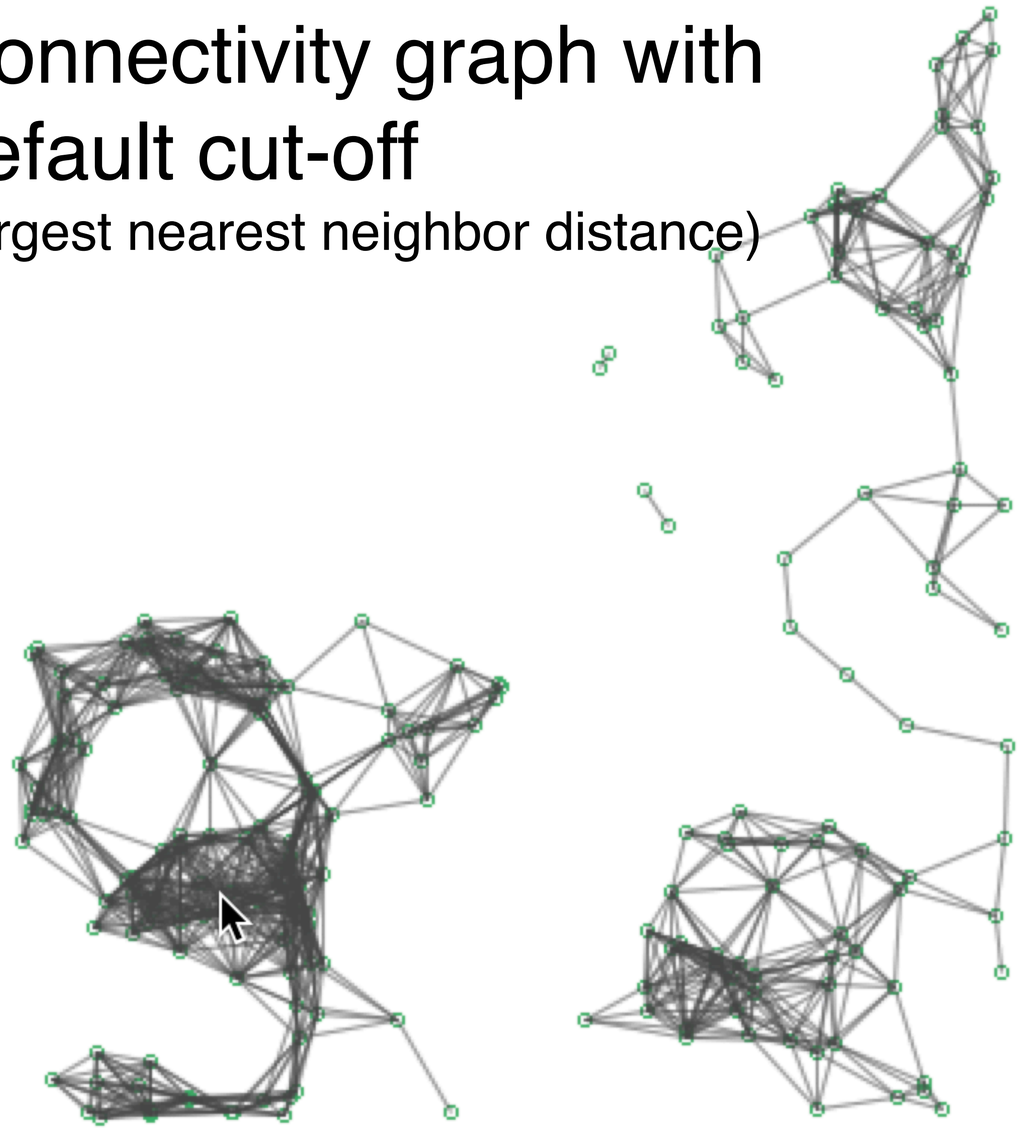


https://geoda.gitee.io/workbook/4b_dist_weights/lab4b.html

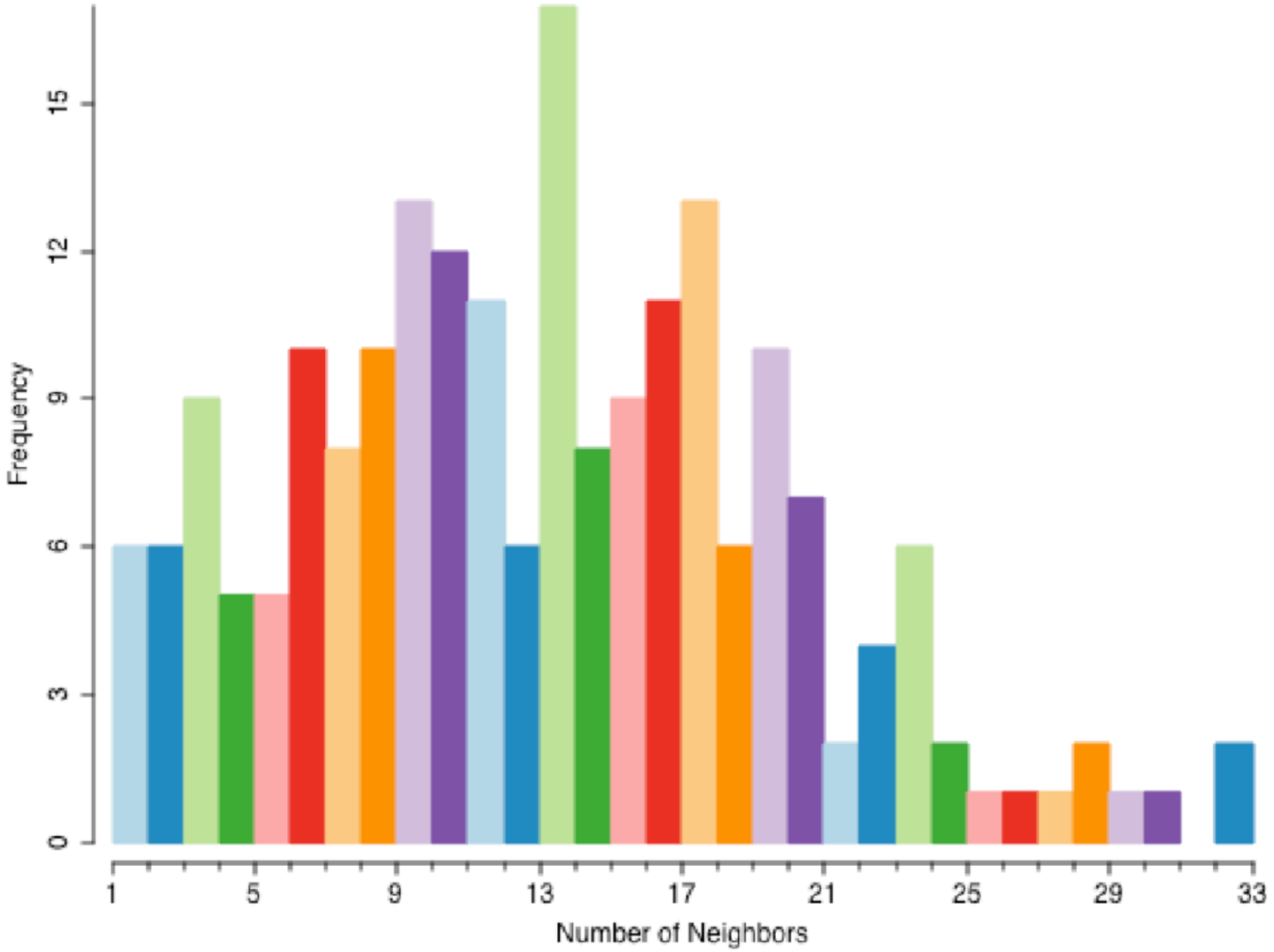
The problem with distance-bands

Connectivity graph with default cut-off

(largest nearest neighbor distance)

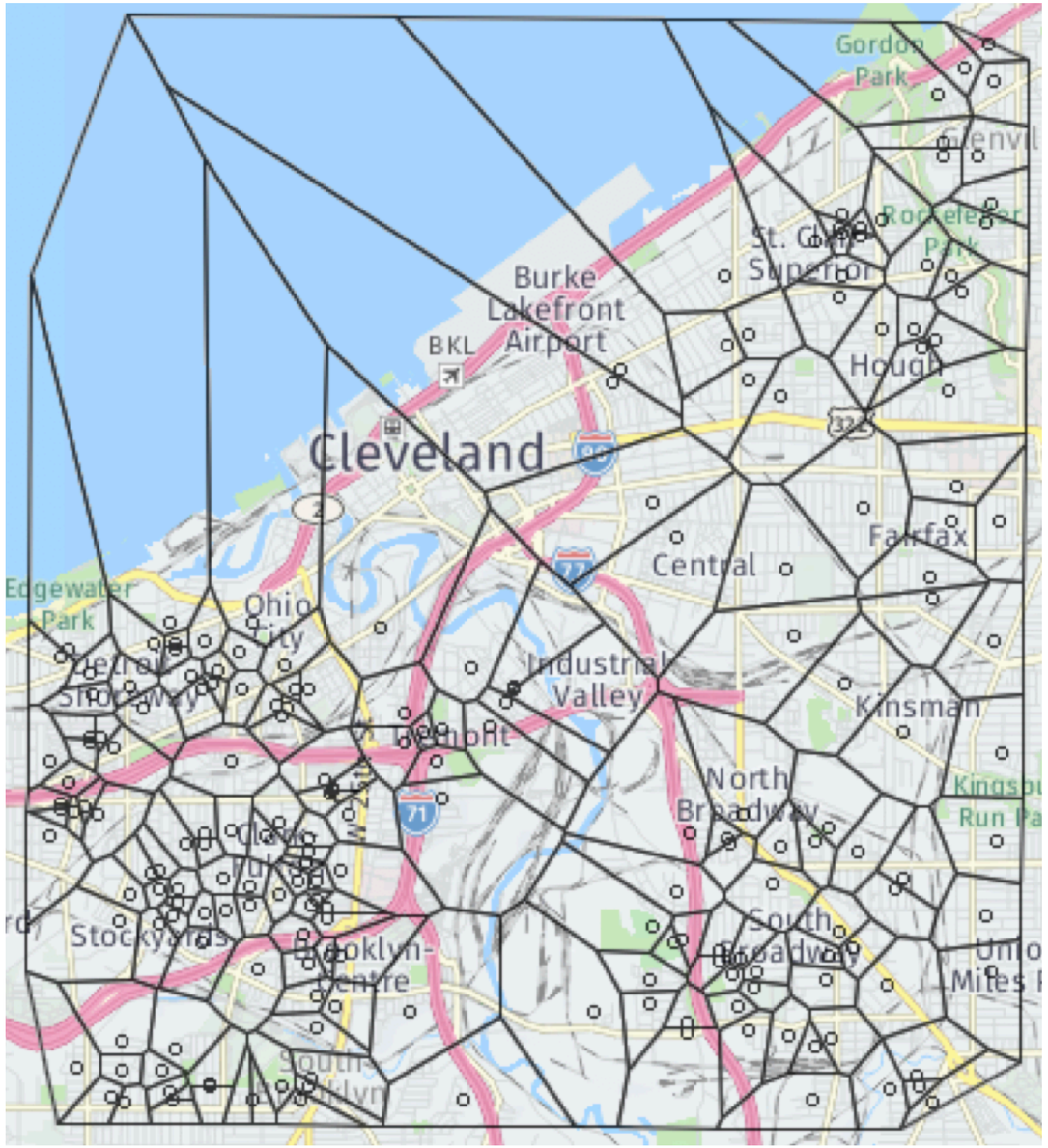


Connectivity histogram

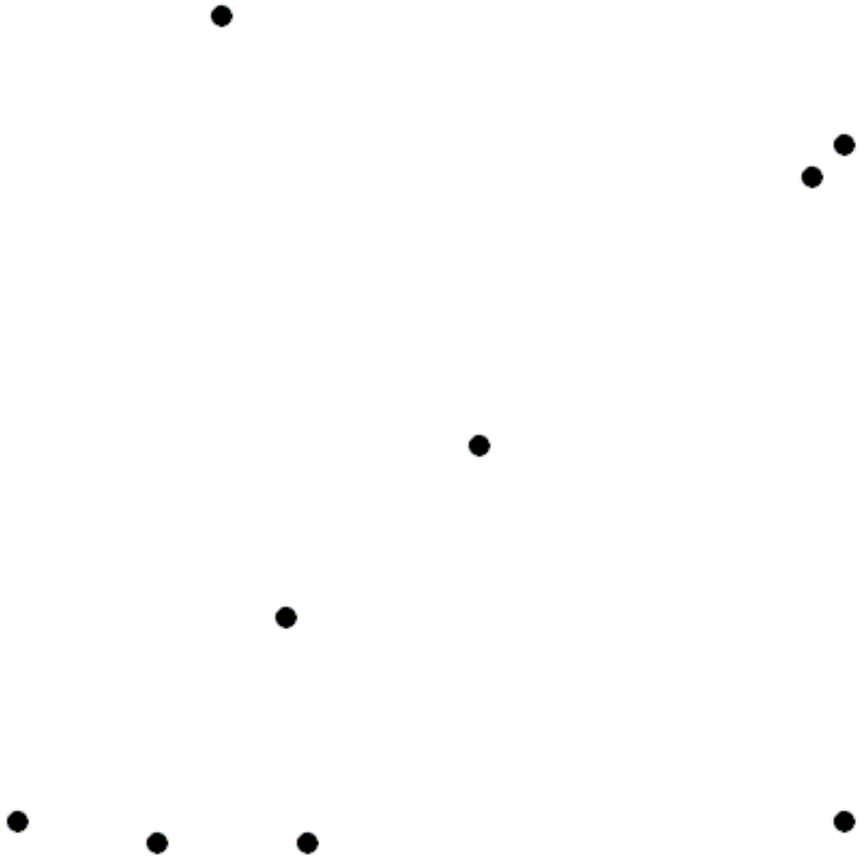


https://geoda.gitee.io/workbook/4b_dist_weights/lab4b.html

The problem with distance-bands



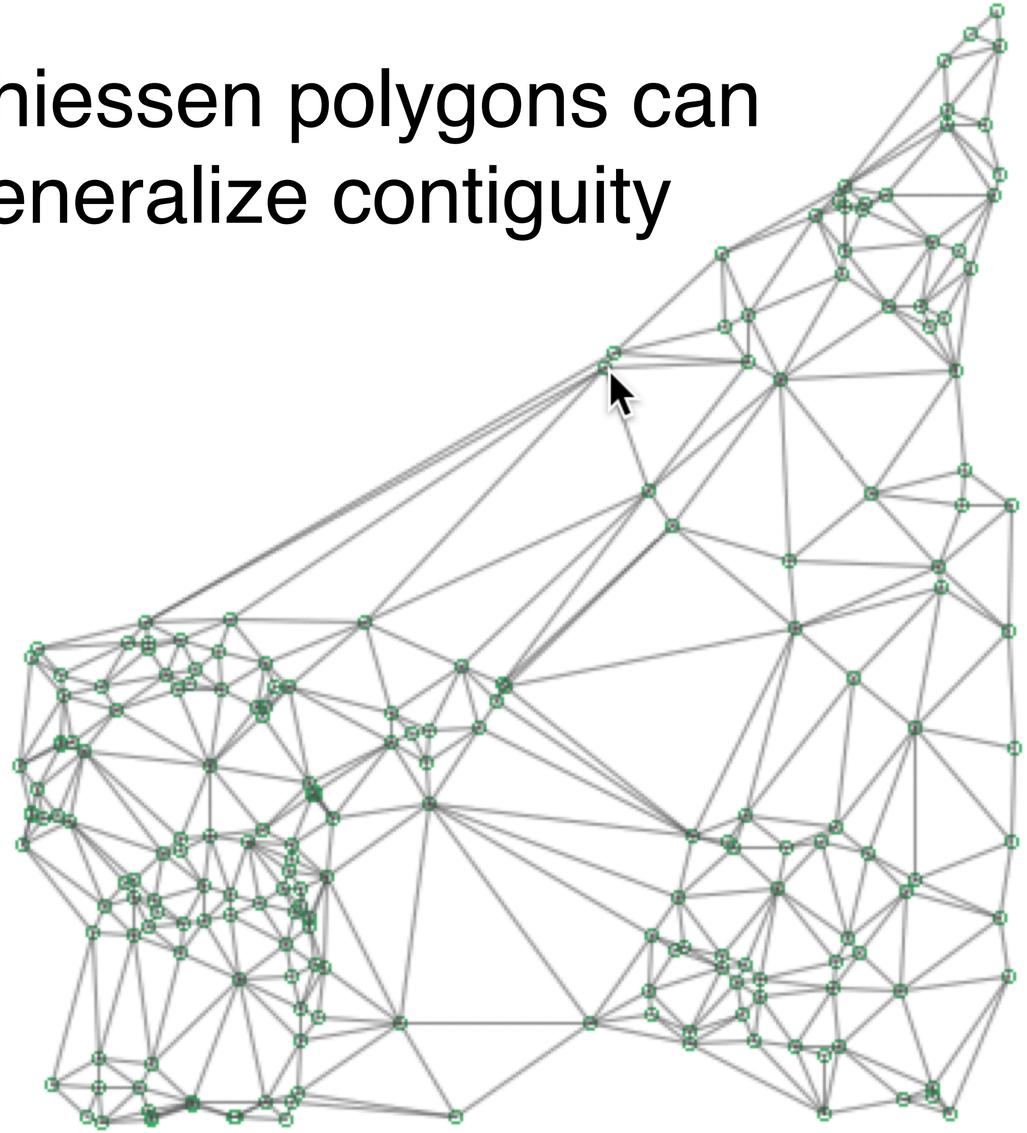
Thiessen polygons can generalize contiguity



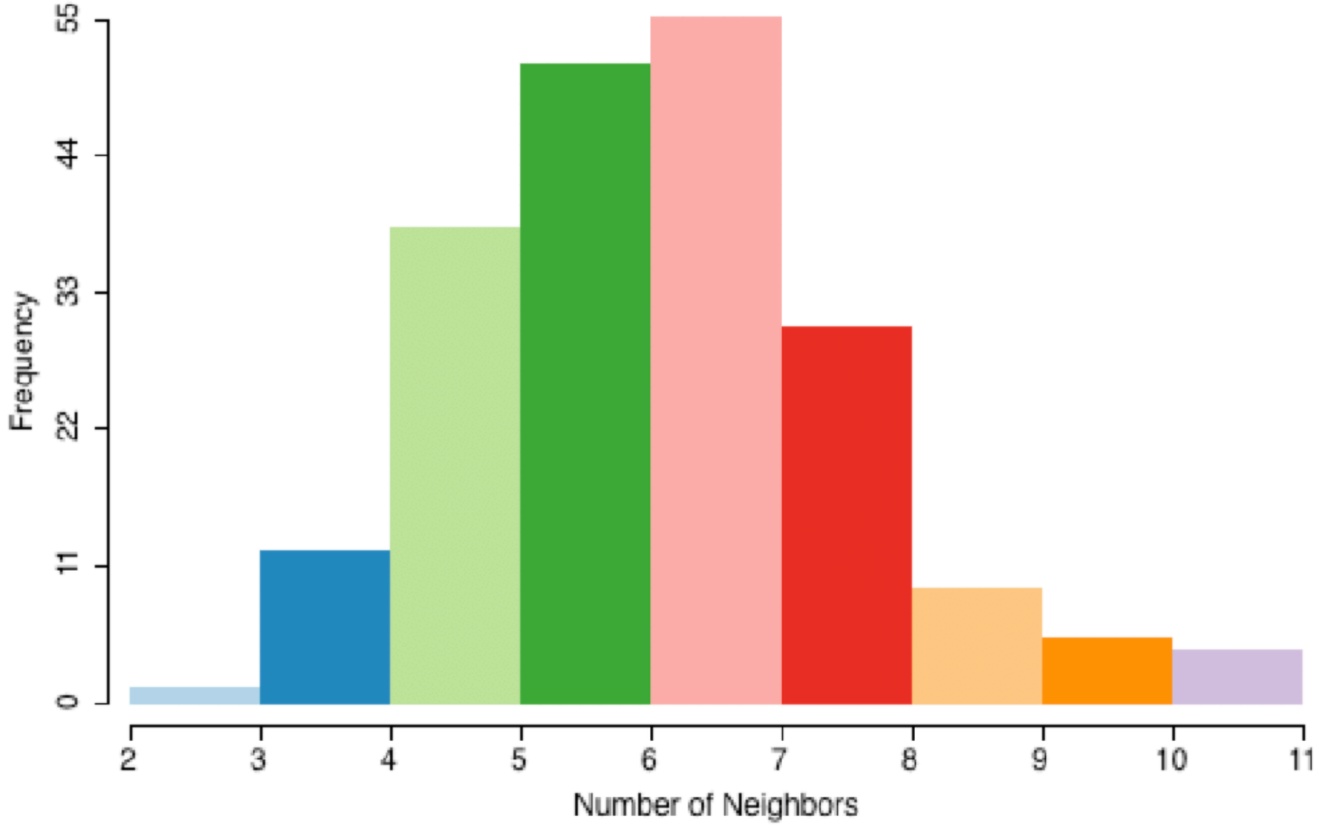
https://geoda.gitee.io/workbook/4b_dist_weights/lab4b.html

The problem with distance-bands

Thiessen polygons can generalize contiguity

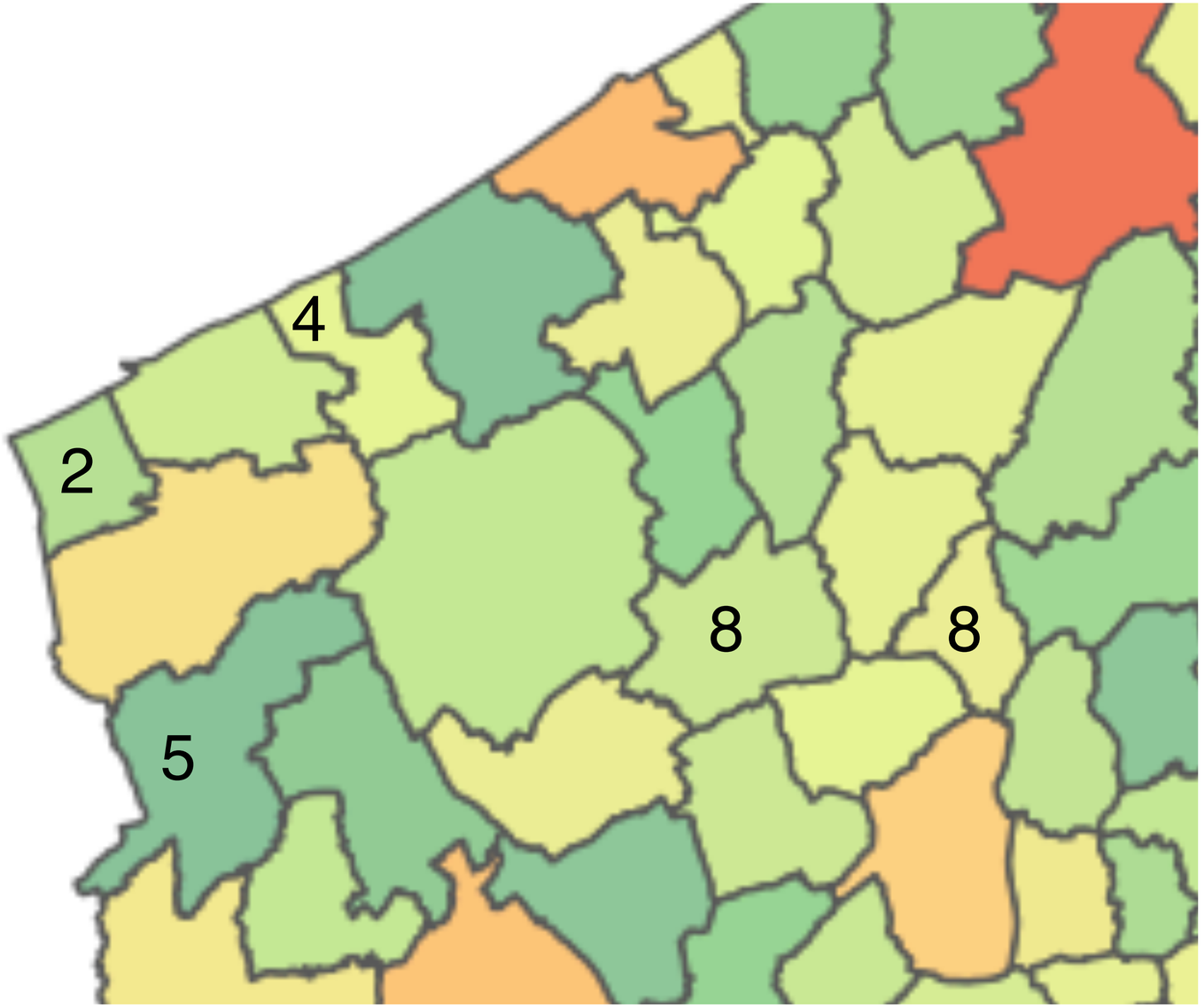


Connectivity histogram



https://geoda.gitee.io/workbook/4b_dist_weights/lab4b.html

It is important to consider edge effects!

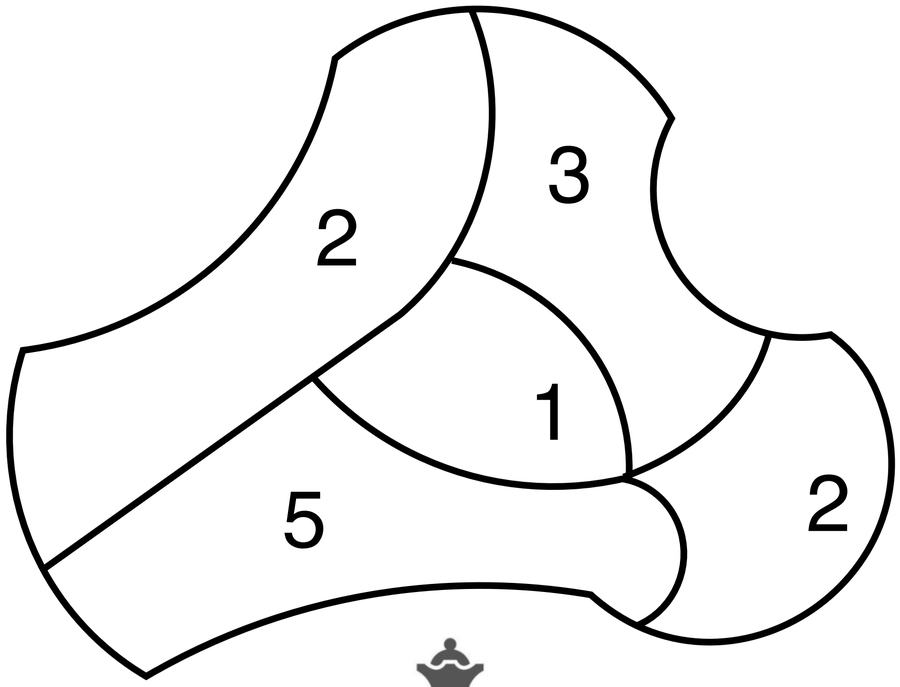


Spatial lag

The spatial lag is the weighted average value of neighbors

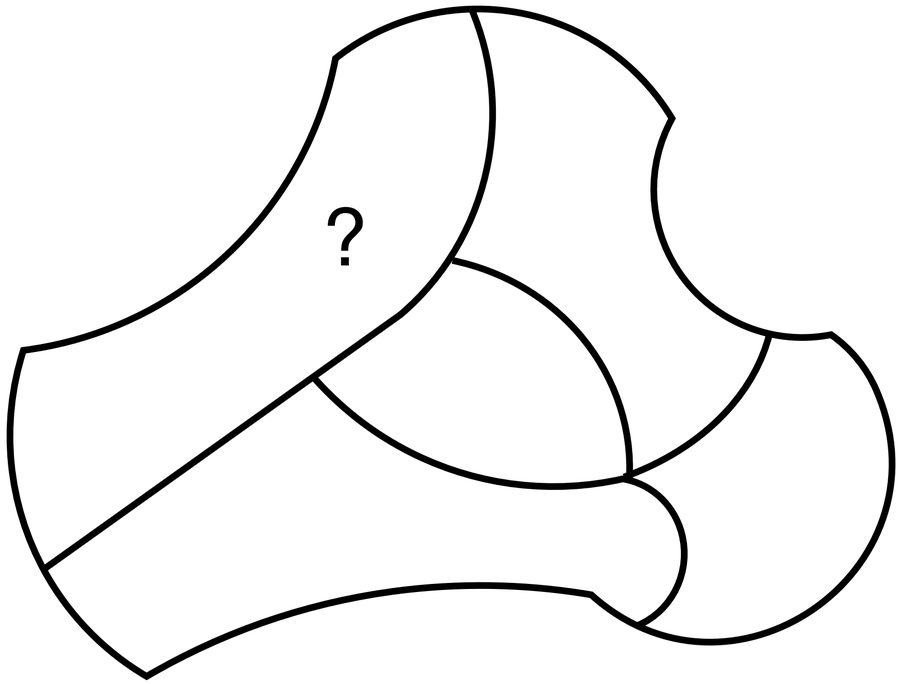
if W is standardized

Values $\mathbf{y}$



QUEEN

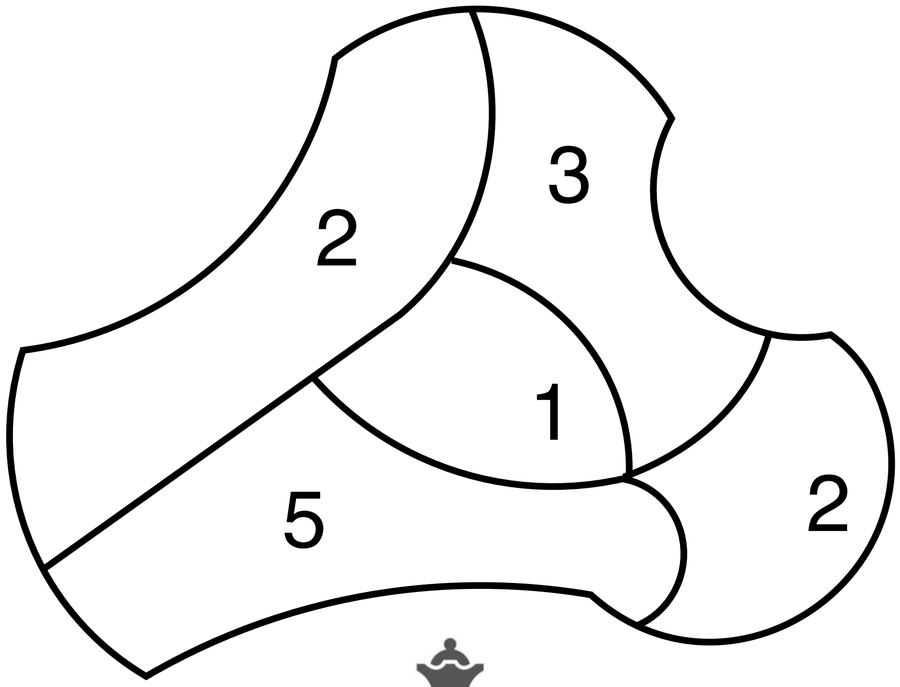
Spatial lag $\mathbf{y}_{\text{lag}}$



The spatial lag is the weighted average value of neighbors

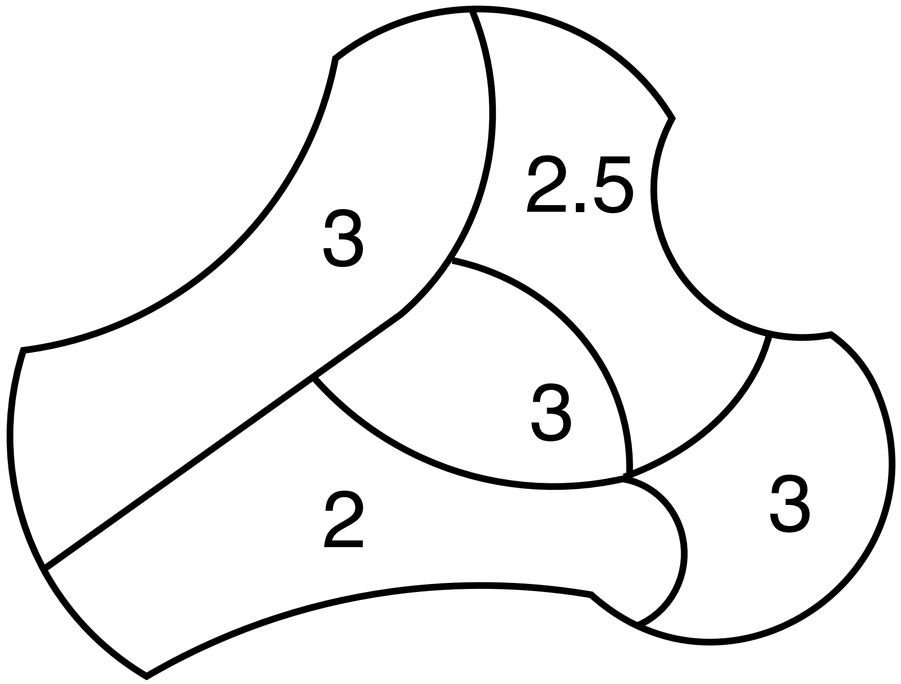
if W is standardized

Values $\mathbf{y}$



QUEEN

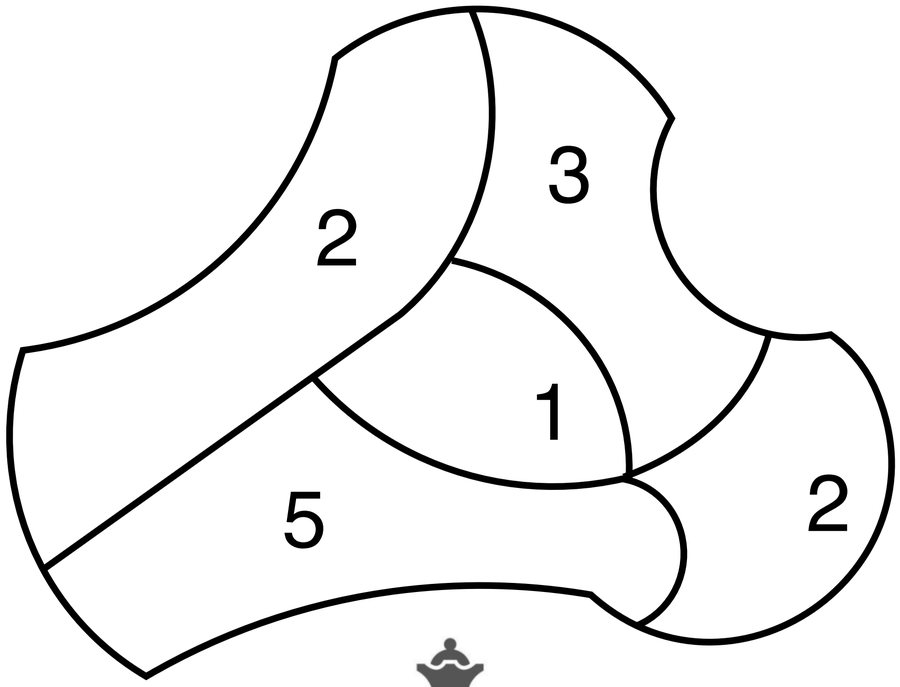
Spatial lag $\mathbf{y}_{\text{lag}}$



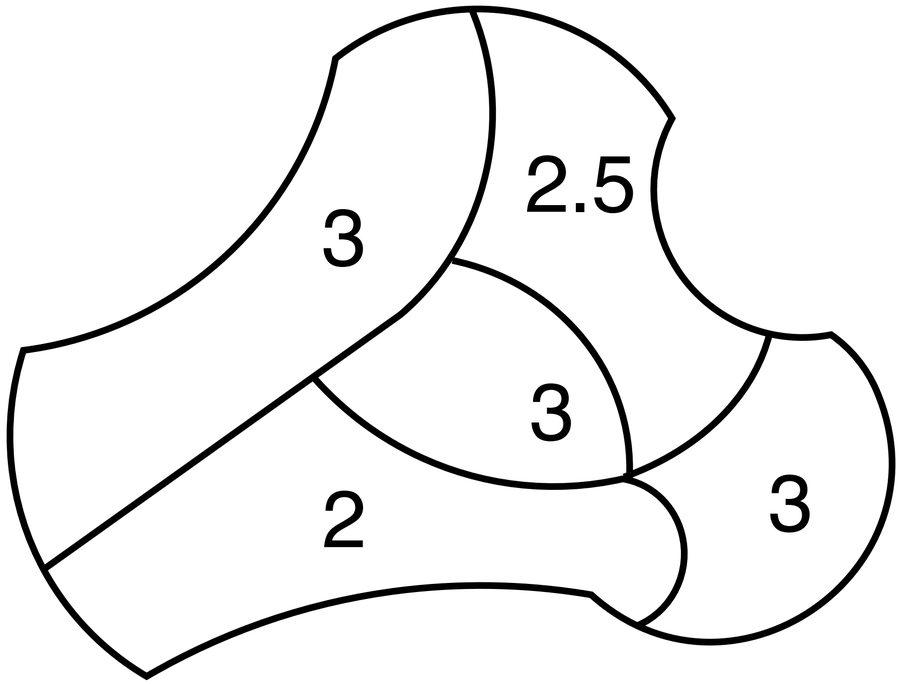
The spatial lag is the weighted average value of neighbors

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

The spatial lag is the sum of products of weights and values

$$y_{\text{lag},i} = w_{i1}y_1 + w_{i2}y_2 + \cdots + w_{in}y_n = \sum_{j=1}^{n} w_{ij}y_j$$

$$y_{\text{lag},1} = 0.25 * 3 + 0.25 * 2 + 0.25 * 2 + 0.25 * 5 = 3$$

The spatial lag is the sum of products of weights and values

$$y_{\text{lag},i} = w_{i1}y_i + w_{i2}y_2 + \cdots + w_{in}y_n = \sum_{j=1}^{n} w_{ij}y_j$$

$$\mathbf{y}_{\text{lag}} = \left(\sum_{j=1}^{n} w_{ij}y_j \right)_i = W\mathbf{y}$$

The spatial lag appears in many tools and models

Early evaluation

(Mutual alignment of expectation)

https://www.menti.com/ale3ebe9wzx6

We use PySAL to define spatial weights

```
from pysal.lib import weights

w_queen = weights.Queen.from_dataframe(db, idVariable="area_id")
```

spatial weight criteria

your GeoDataframe

column to index weights on

We use PySAL to define spatial weights

```
w_queen['708518']
```

```
{'706412': 1.0,
 '700870': 1.0,
 '709938': 1.0,
 '707896': 1.0,
 '718916': 1.0,
 '703634': 1.0,
 '711110': 1.0,
 '701970': 1.0,
 '703720': 1.0}
```

New library: splot

```
from splot.libpysal import plot_spatial_weights

plot_spatial_weights(w_queen, db, indexed_on="area_id");
```

Spatial weights created with PySal

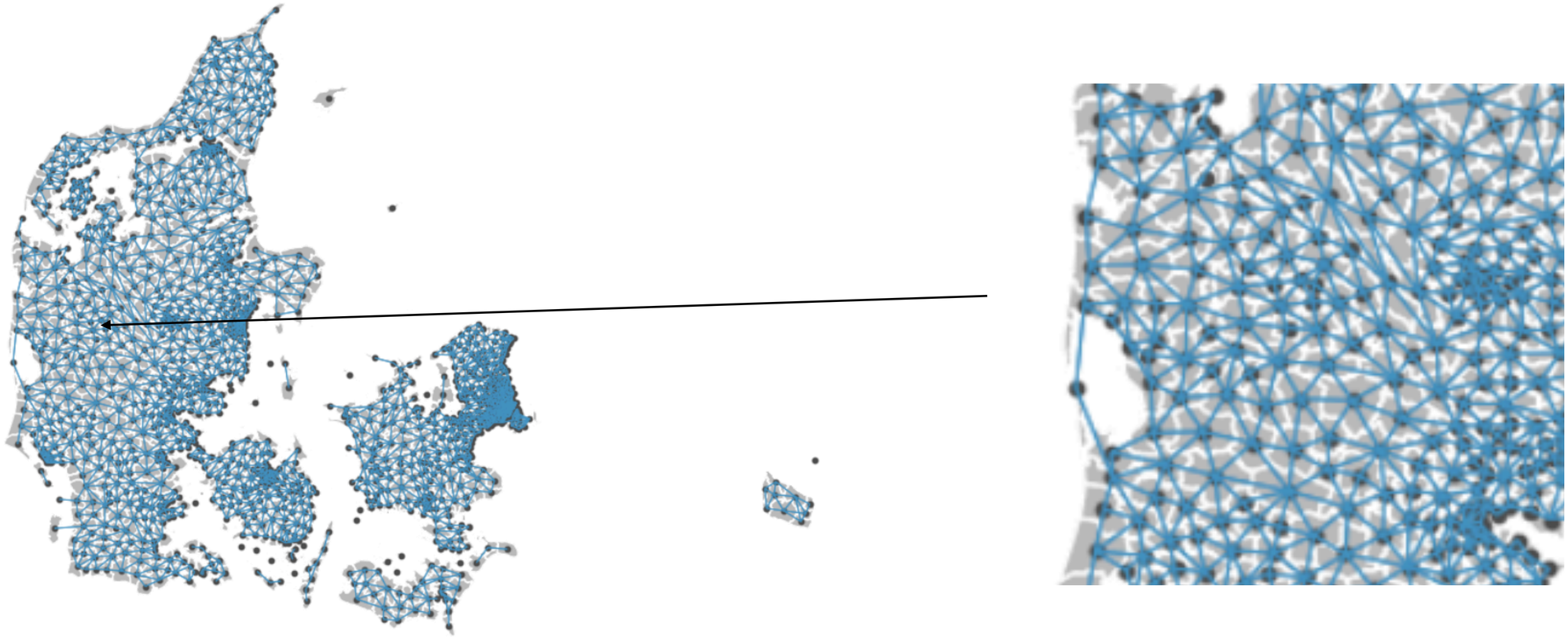
Your GeoDataframe

Column used to index spatial weights

https://splot.readthedocs.io/en/latest/

New library: splot

```
plot_spatial_weights(w_queen, db, indexed_on="area_id");
```



Jupyter

Sources and further materials for today's class

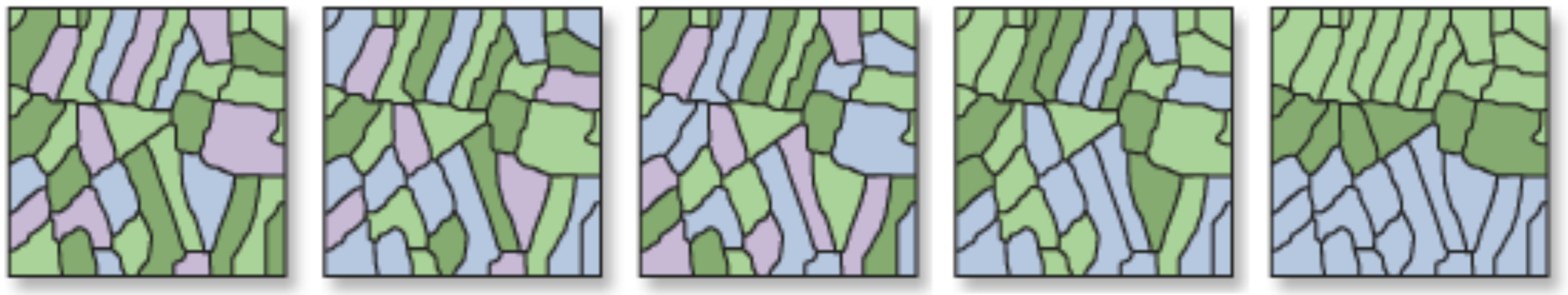
Geographic Data Science with Python

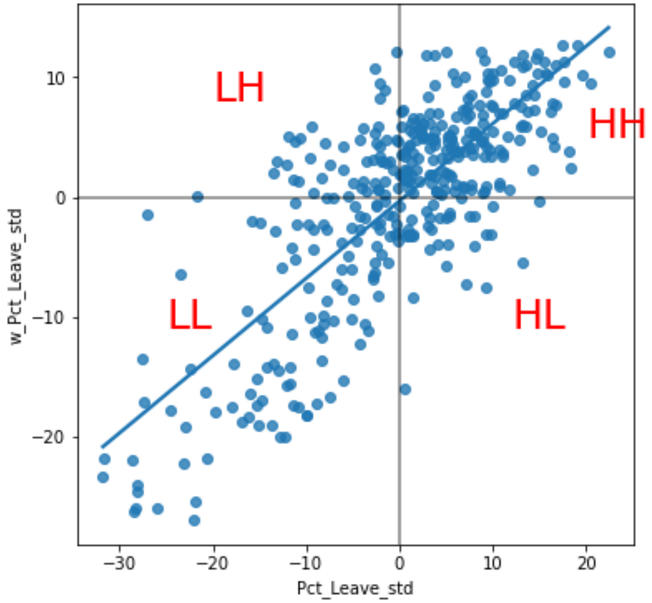
https://geographicdata.science/book/notebooks/04_spatial_weights.html

https://darribas.org/gds_course/content/bE/concepts_E.html

https://geoda.gitee.io/workbook/4b_dist_weights/lab4b.html

Next week: Spatial autocorrelation





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