Spatial econometrics (1)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute

1. Introduction

- **Topics:**
  1. Discrete choice
     - Random utility framework, estimating binary and multinomial regression models
  2. Spatial econometrics
     - Spatial data, autocorrelation, spatial regressions
  3. Identification
     - Research design, IV, OLS, RDD, quasi-experiments, standard errors
  4. Hedonic pricing
     - Theory and estimation
  5. Quantitative spatial economics
     - General equilibrium models in spatial economics

VU VRIJE UNIVERSITEIT AMSTERDAM

1. Introduction

Wednesday

09:30-10:30	Lecture 1	Discrete Choice I (The random utility framework)
10:45-11:45	Lecture 2	Discrete Choice II (Estimating discrete choice models)
12:00-13:00	Lecture 3	Spatial Econometrics I (Spatial data)
14:00-15:30	Tutorial 1	Assignment 1

Thursday

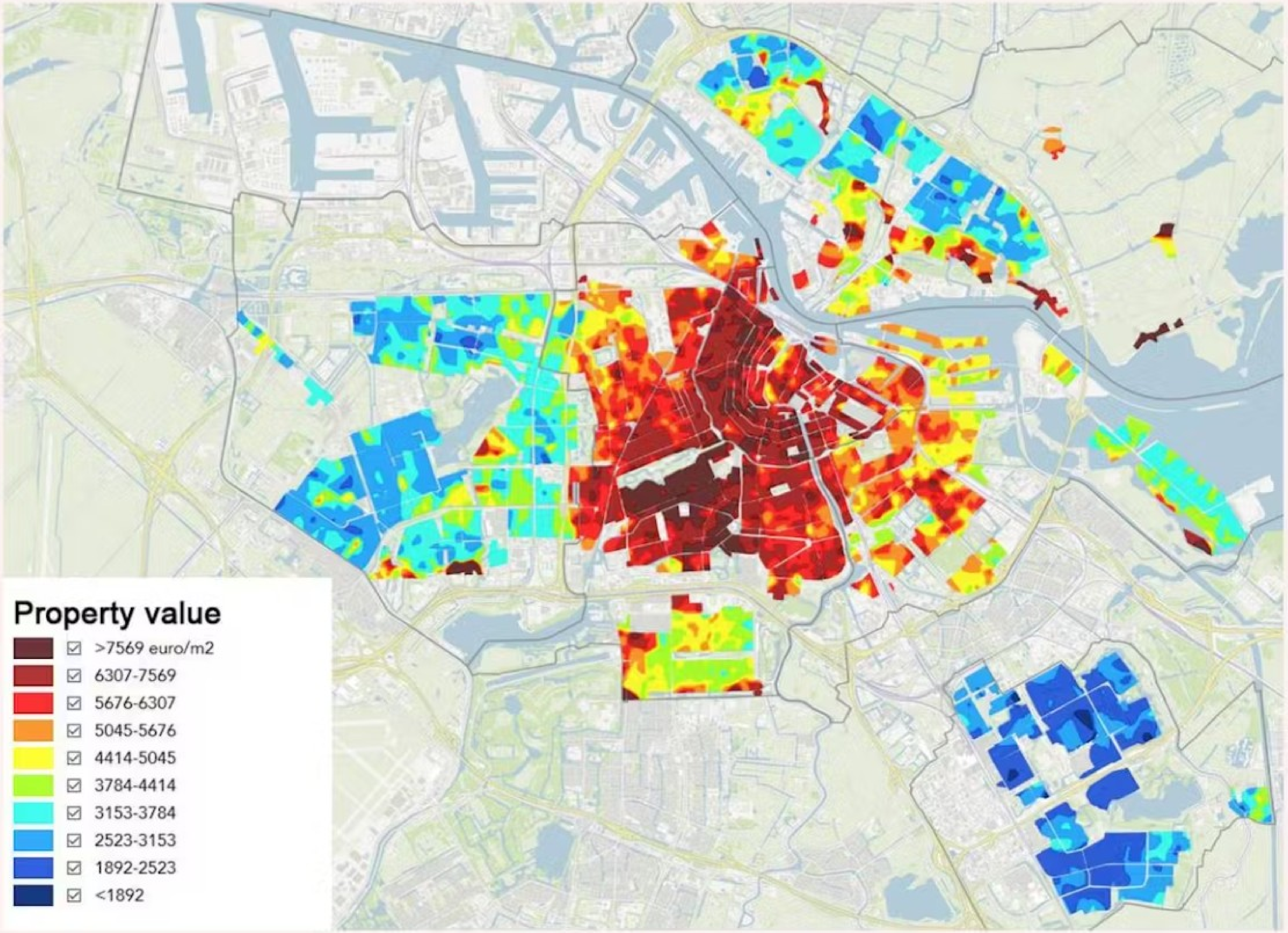
09:30-10:30	Lecture 4	Spatial Econometrics II (Spatial autocorrelation)
10:45-11:45	Lecture 5	Spatial Econometrics III (Spatial regressions)
12:00-12:30	Lecture 6	Identification I (Research design)
13:30-14:00	Tutorial 2	Discussion of Assignment 1
14:00-15:00	Tutorial 3	Assignment 2

Friday

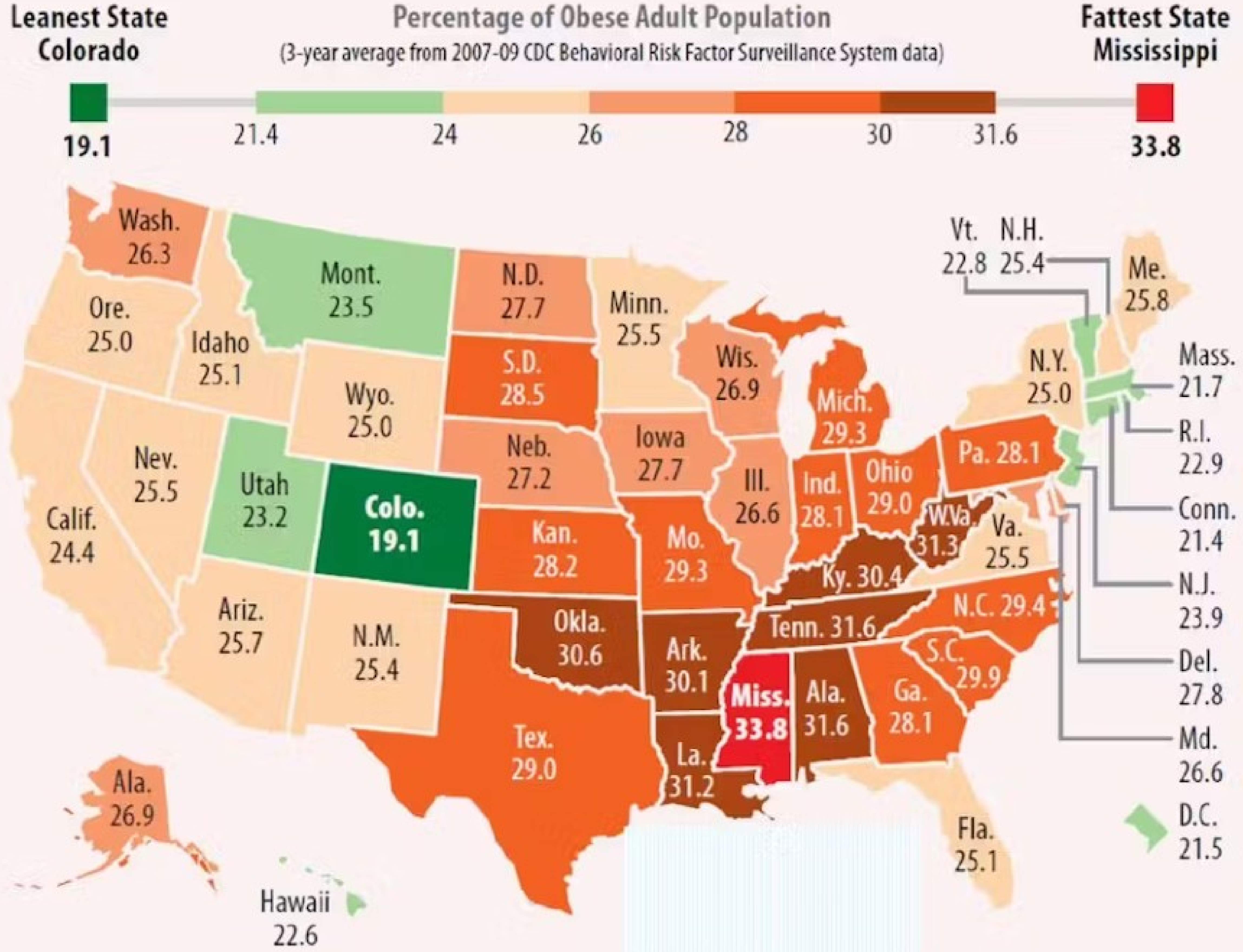
09:30-10:00	Lecture 7	Identification II (RCTs, OLS, IV, quasi-experiments)
10:00-10:30	Lecture 8	Hedonic pricing I (Theory)
10:45-11:45	Lecture 9	Hedonic pricing II (Estimation)
12:00-12:30	Tutorial 4	Discussion of Assignment 2

1. Introduction

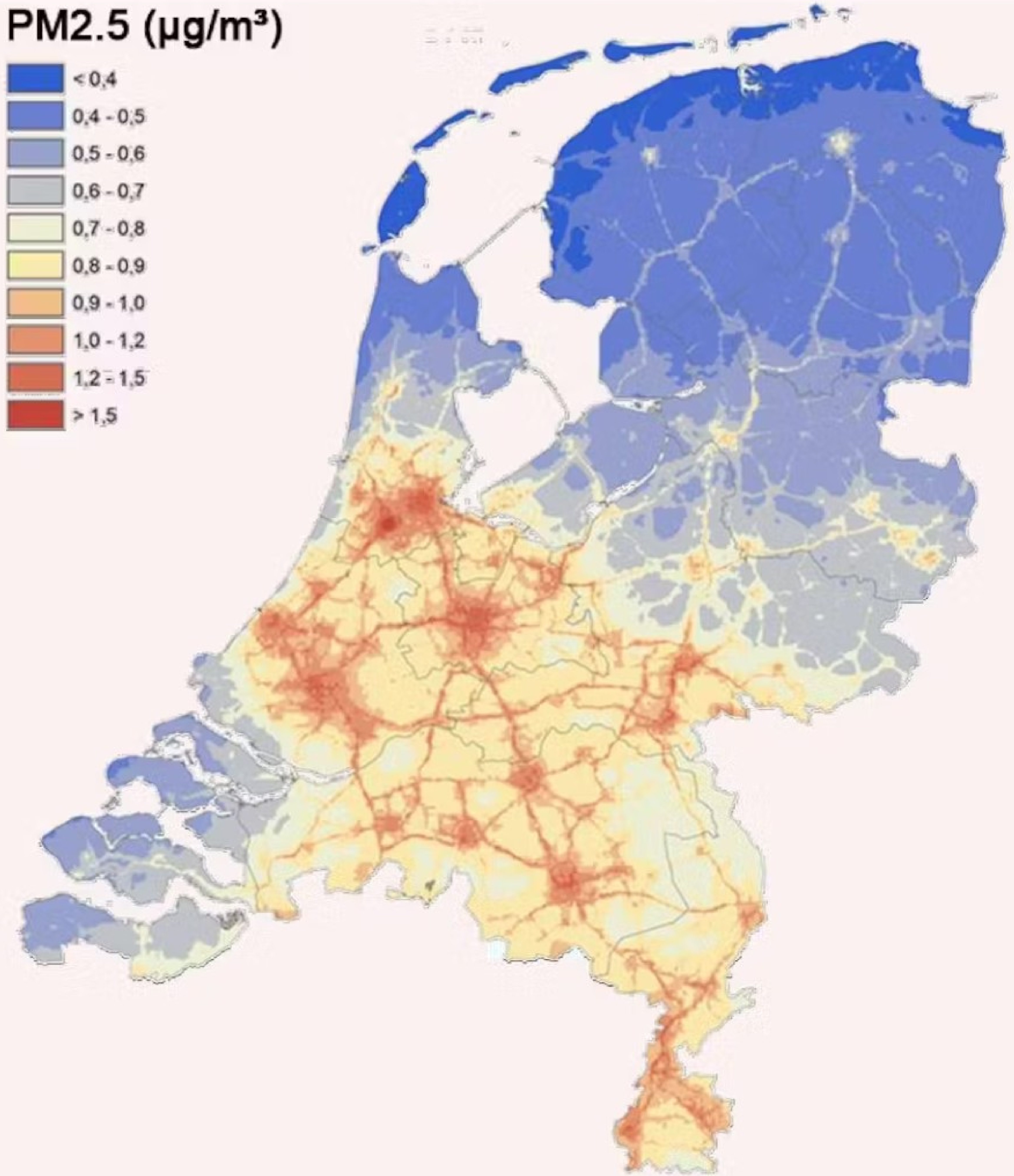
- **What is the m² price in Amsterdam?**



1. Introduction



1. Introduction



VU VRIJE UNIVERSITEIT AMSTERDAM

1. Introduction

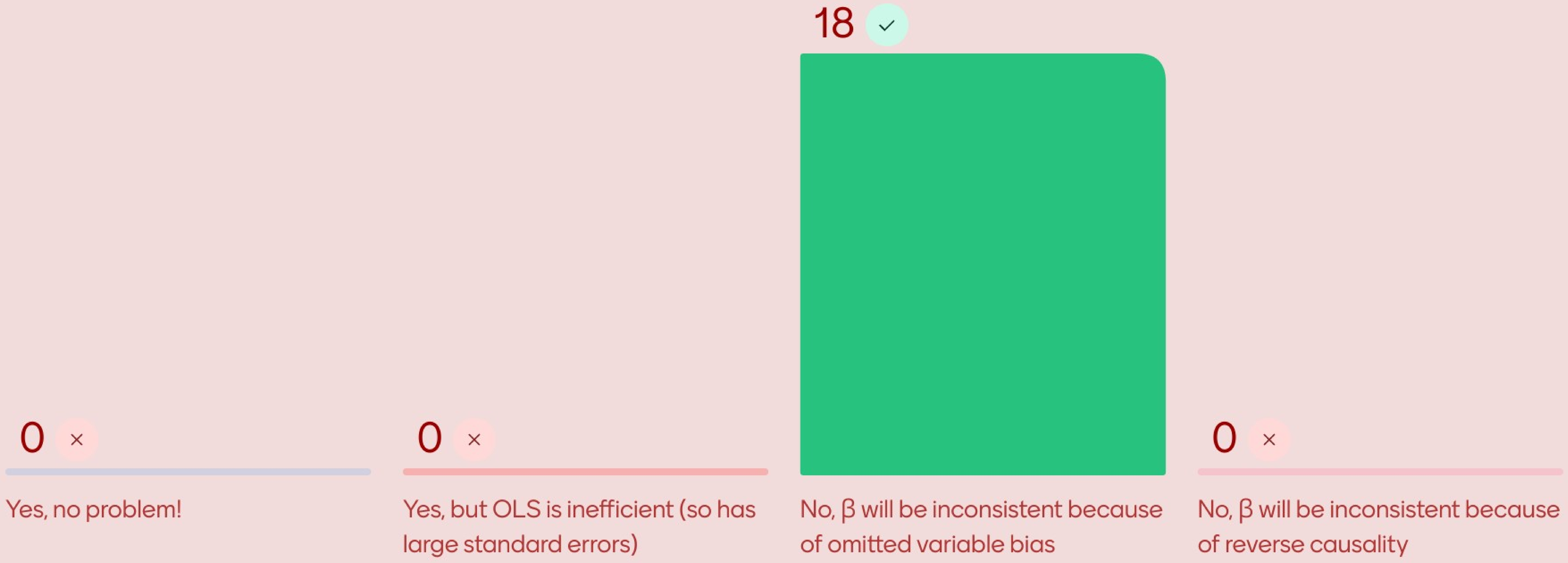
- **What is special about spatial data?**

- **Not only time component, but also spatial component:**

$$y_{t,i} = \beta x_{t,i} + \epsilon_{t,i} \qquad (1')$$

VU VRIJE UNIVERSITEIT AMSTERDAM

With $y_i = \beta x_i + f(x_j) + \epsilon_i$, will β be estimated consistently with OLS when estimating $y_i + \beta x_i + \epsilon_i$?



18

- **Some remarks on matrix notation**
  - **Use bold symbols for vectors**

$$\boldsymbol{x} = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix}$$

  - **Use bold symbols and capitals for matrices**

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

  - **Identity matrix**

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\rightarrow \boldsymbol{I}\boldsymbol{X} = \boldsymbol{X}$

  - **Inverse $\boldsymbol{X}^{-1}$ is matrix equivalent of $1/x$**

$\rightarrow \boldsymbol{X}^{-1}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{X}^{-1} = \boldsymbol{I}$

  - **More details in the appendix of the syllabus**

VU VRIJE UNIVERSITEIT AMSTERDAM

2. Space in economics

- Many economic processes are spatially correlated
  - Tobler's first law of geography
- Most economics models are "topologically invariant"
- New economic fields have emerged
  - Urban economics
  - New economic geography (NEG)
- Synergy with other fields
  - Economic geography
  - Regional science
  - GIS

2. Space in economics

- Spatial econometrics
- 40-50s – mainly domain of statisticians
- Cliff and Ord (1973): "Spatial autocorrelation"
- Paelinck and Klaassen (1979): "Spatial Econometrics"
- Rapid growth since Anselin (1988)
- New estimators, tests and interpretation
  - *e.g.* Kelejian and Prucha (1998, 1999, 2004, 2007, 2010)

VU VRIJE UNIVERSITEIT AMSTERDAM

2. Space in economics

- Spatial modelling is becoming increasingly important
  - New and geo-referenced data
  - Advanced software
  - *New methods and regression techniques!*

VU VRIJE UNIVERSITEIT AMSTERDAM

3. Spatial data structure

- Time is simple
  - Natural origin
  - No reciprocity
  - Unidirectional

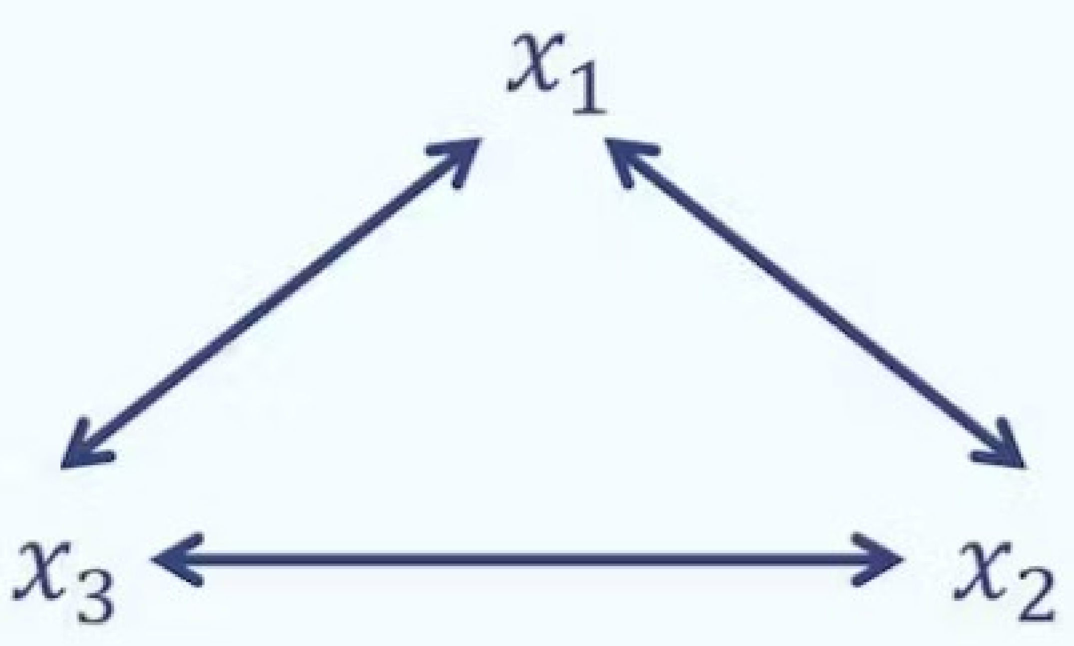
$$x_{t-3} \longrightarrow x_{t-2} \longrightarrow x_{t-1} \longrightarrow x_t$$

- Linear space (*e.g. beach*) is different
  - No natural origin
  - Reciprocity
  - Unidirectional

$$x_1 \longleftrightarrow x_2 \longleftrightarrow x_3 \longleftrightarrow x_4$$

3. Spatial data structure

- **Two-dimensional space becomes even more complex**
  - **No natural origin**
  - **Reciprocity**
  - **Multidirectional**



- $i = 1,2,3$ **can refer to point data, areas, grids**

3. Spatial data structure

- First, we have to define the spatial structure of the data

- Specified through a spatial weights matrix

- Spatial weights matrix W:
  - Consists of $n \times n$ elements
  - Discrete or continuous elements

- How to define weights?
  - Euclidian distance
  - Network distance
  - Spatial interactions
  - Social networks

3. Spatial data structure

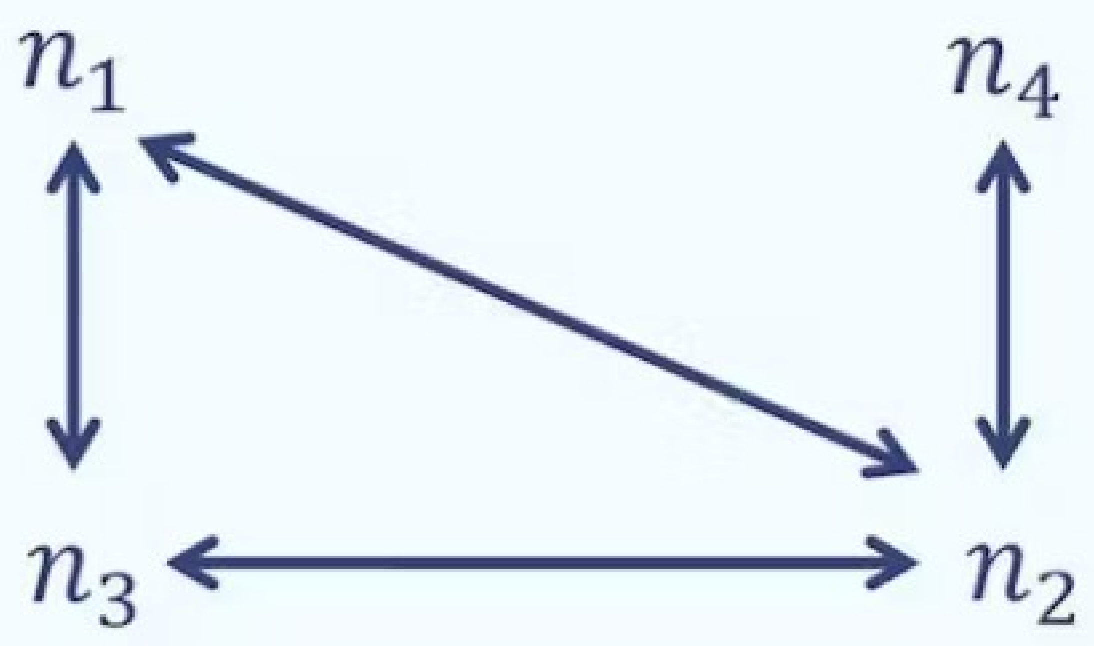
- How to define spatial matrices?

- Contiguity matrix
  - Adjacent → 1^{st} order contiguous
  - Neighbours of neighbours → 2^{nd} order contiguous

- Distance matrix
  - k-nearest neighbours
  - Inverse distance weights ($1/distance$)
  - Cut-off distance

VU VRIJE UNIVERSITEIT AMSTERDAM

- **Let's provide an example of a contiguity matrix**



to (columns), from (rows)

W	n_1	n_2	n_3	n_4
n_1	0	1	1	0
n_2	1	0	1	1
n_3	1	1	0	0
n_4	0	1	0	0

3. Spatial data structure

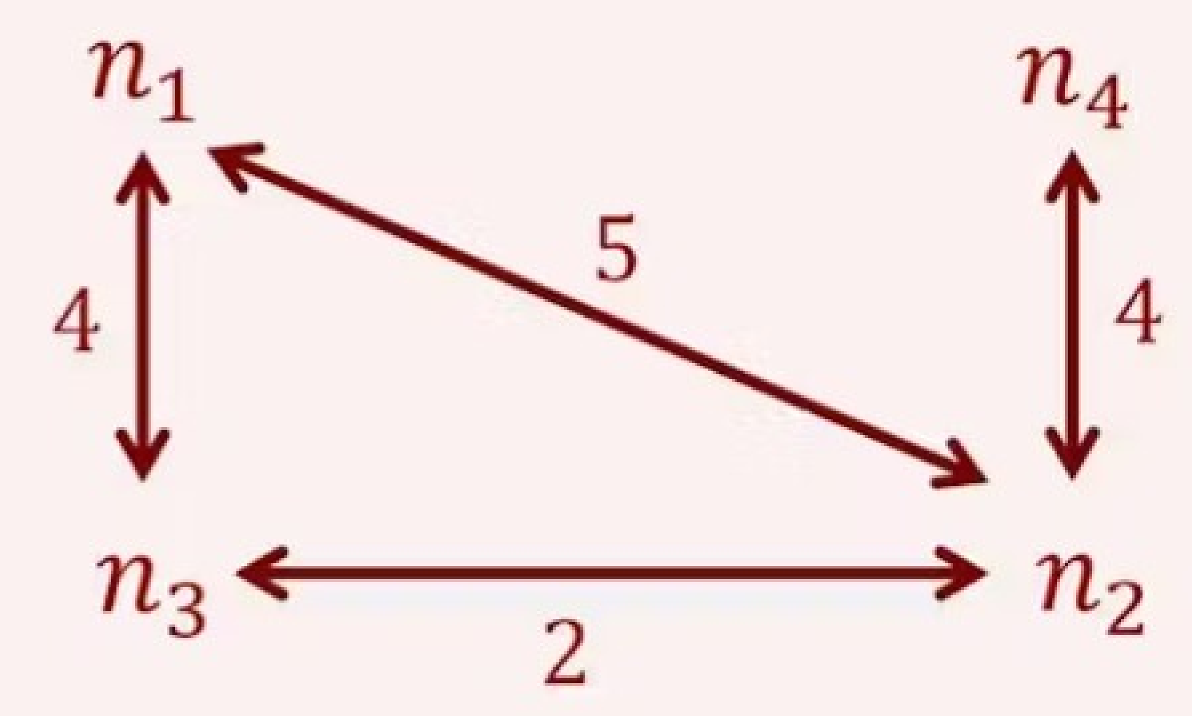
- **Matrices can be standardised**
  - **Different principles can be used**
  - **Most common: *row-standardisation*:**

    $$w_{ij}^* = \frac{w_{ij}}{\sum_{k=1}^{n} w_{ik}}$$

    where k are other locations

- **Interpretation of**
  - $\sum_{j=1}^{n} w_{ij}$**: sum of connections to neighbours**
  - w_{ij}^* **denotes the share of connections to neighbours**

- **Create an *inverse* distance weight matrix with row-standardised weights**



to

from

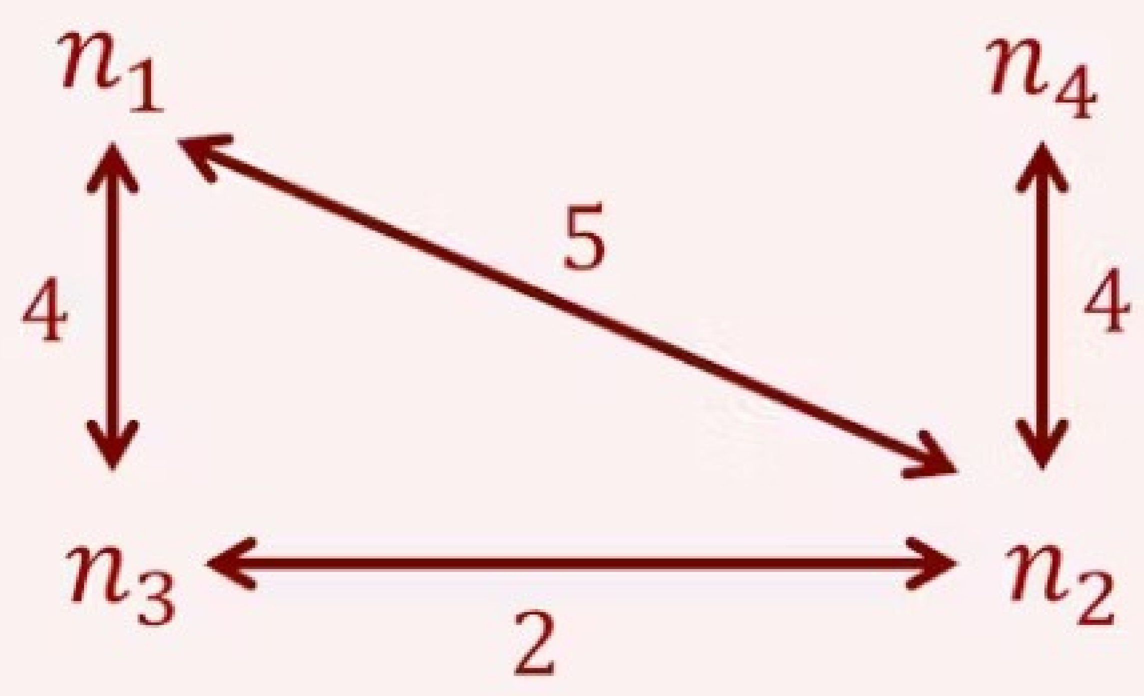
W	n_1	n_2	n_3	n_4
n_1				
n_2				
n_3				
n_4				

VRIJE UNIVERSITEIT AMSTERDAM

Create an inverse distance weight matrix with row-standardised weights

I am ready!

- **Create an *inverse* distance weight matrix with row-standardised weights**



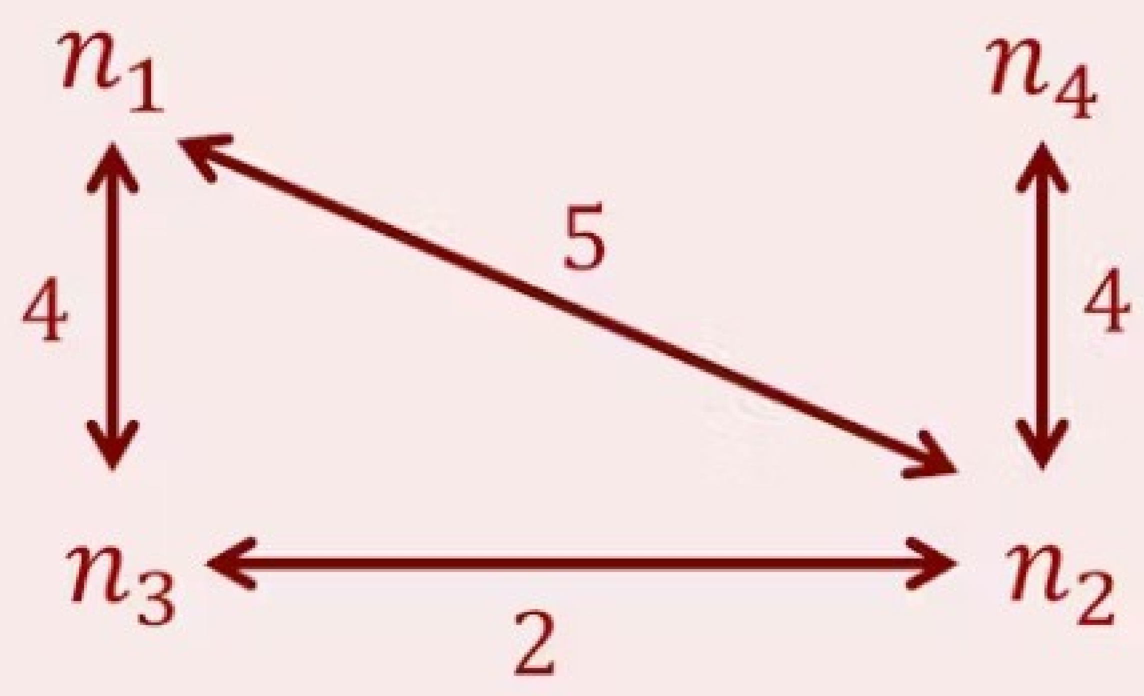
to

from

W	n_1	n_2	n_3	n_4
n_1	0	1/5	1/4	1/9
n_2	1/5	0	1/2	1/4
n_3	1/4	1/2	0	1/6
n_4	1/9	1/4	1/6	0

3. Spatial data structure

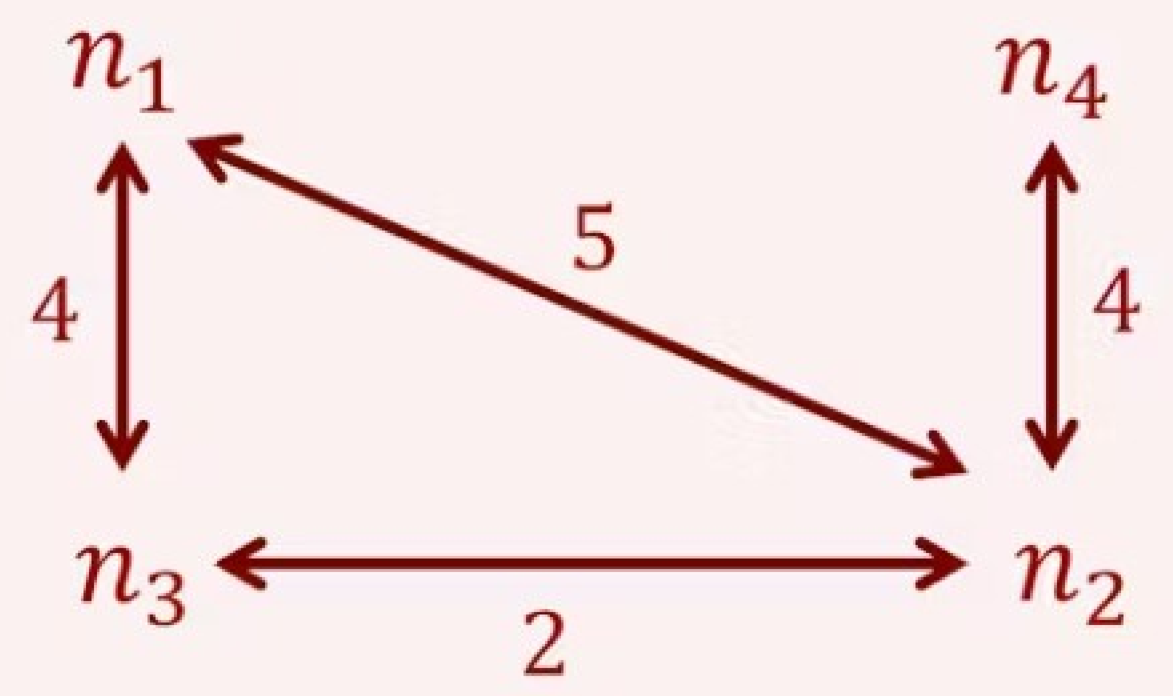
- **Create an *inverse* distance weight matrix with row-standardised weights**



to / from

W	n_1	n_2	n_3	n_4
n_1	0	$\frac{1/5}{1/5+1/4+1/9}$	$\frac{1/4}{1/5+1/4+1/9}$	$\frac{1/9}{1/5+1/4+1/9}$
n_2	$\frac{1/5}{1/5+1/2+1/4}$	0	$\frac{1/2}{1/5+1/2+1/4}$	$\frac{1/4}{1/5+1/2+1/4}$
n_3	$\frac{1/4}{1/4+1/2+1/6}$	$\frac{1/2}{1/4+1/2+1/6}$	0	$\frac{1/6}{1/4+1/2+1/6}$
n_4	$\frac{1/9}{1/9+1/4+1/6}$	$\frac{1/4}{1/9+1/4+1/6}$	$\frac{1/6}{1/9+1/4+1/6}$	0

- **Create an *inverse* distance weight matrix with row-standardised weights**



to (columns), from (rows)

W	n_1	n_2	n_3	n_4
n_1	0	0.36	0.45	0.20
n_2	0.21	0	0.53	0.26
n_3	0.27	0.55	0	0.18
n_4	0.21	0.47	0.32	0

VU VRIJE UNIVERSITEIT AMSTERDAM

- **Let's say you aim to create a spatial weight matrix**

→ **What could be a problem with the following weight matrix?**

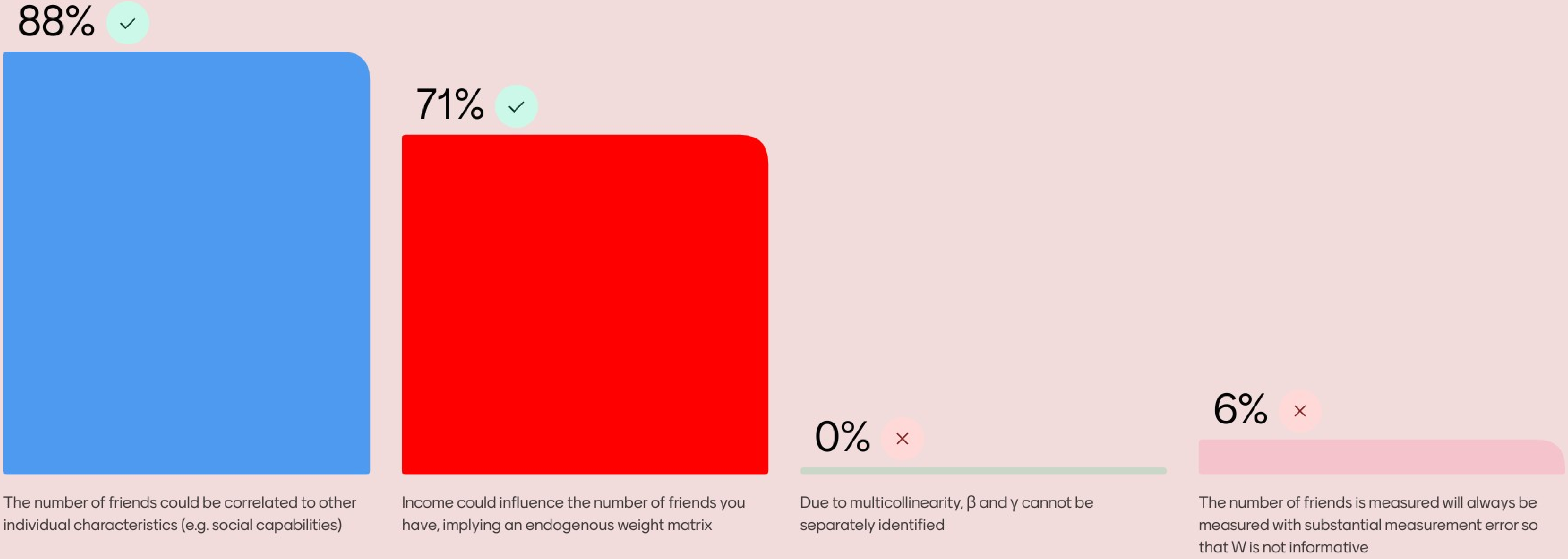
$$y = \beta e + W e' \gamma + \epsilon \quad (3)$$

y = income; e = education

Say that W depends on *the number of friends you have*

VRIJE UNIVERSITEIT AMSTERDAM

What could be a problem with: $\mathbf{y} = \beta\mathbf{e} + \mathbf{W}\mathbf{e}'\gamma + \varepsilon$, where $\mathbf{W}$ depends on the number of friends?



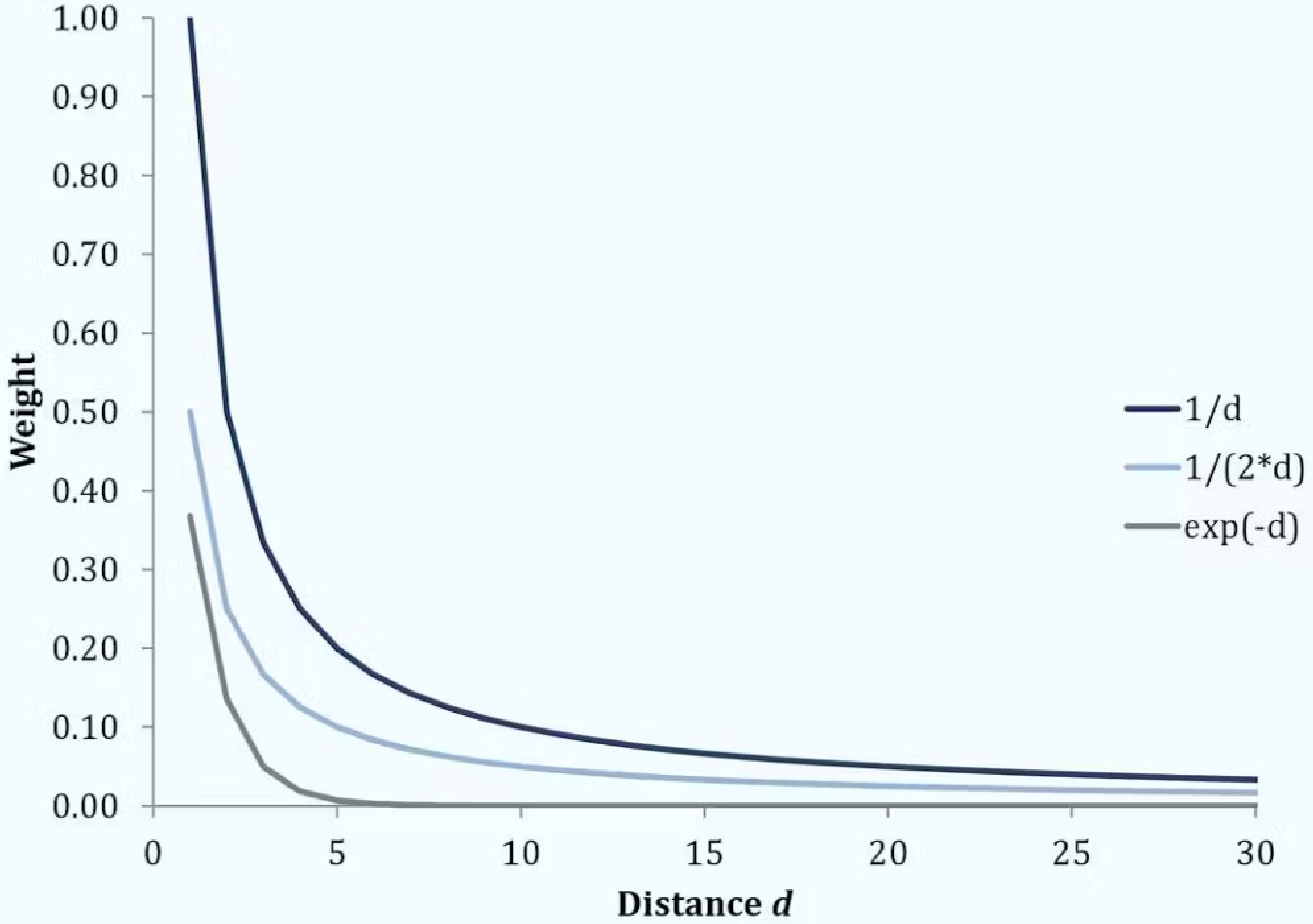
3. Spatial data structure

- Remarks regarding distance weight matrices
  - Check for exogeneity of matrix
  - Connectivity
  - Symmetry
  - Standardisation
  - Distance decay

VU VRIJE UNIVERSITEIT AMSTERDAM

3. Spatial data structure

- **Choice of distance decay is arbitrary**
  - Sometimes theory may help
  - May also try to find the optimal decay parameter empirically

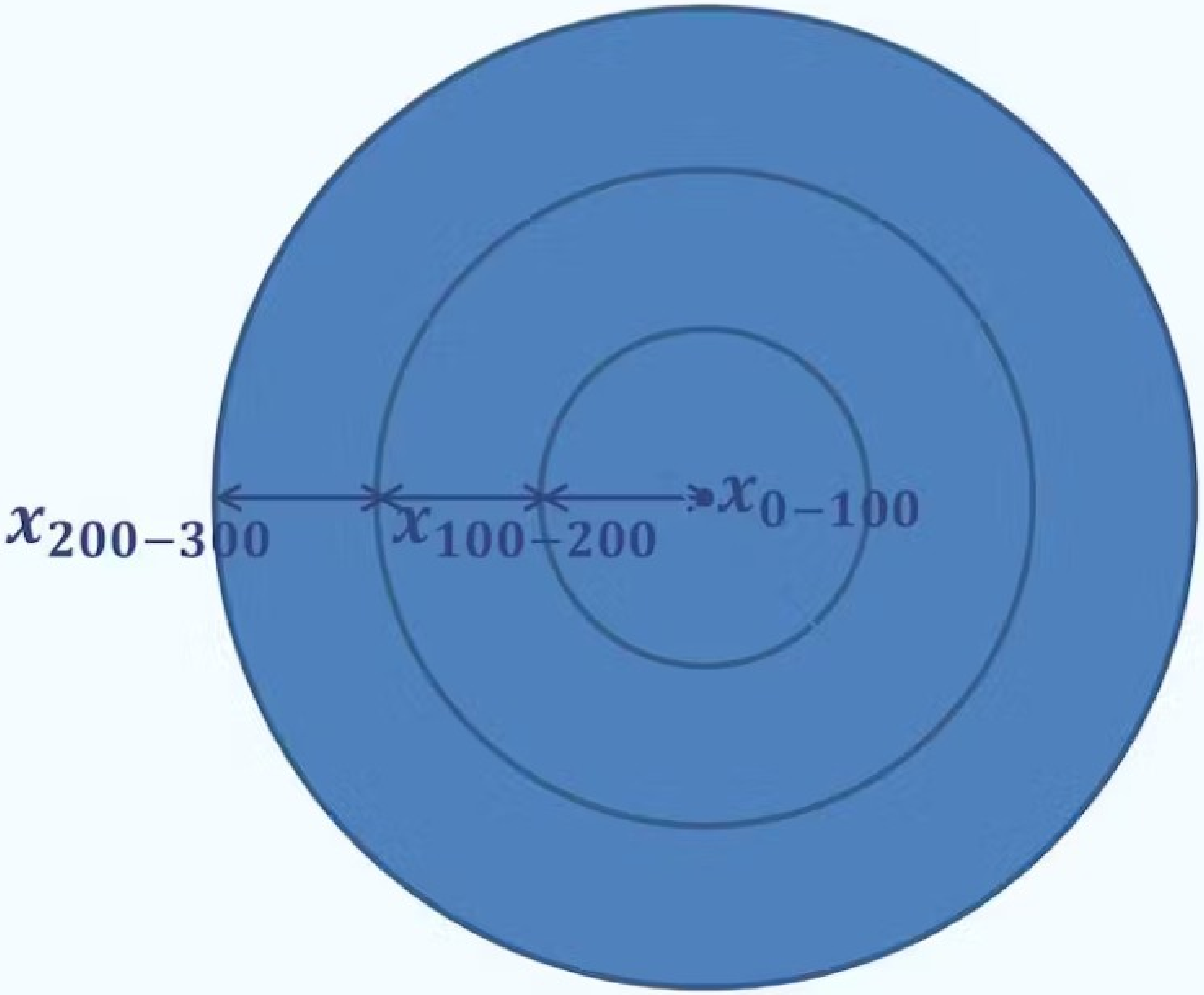


3. Spatial data structure

- Choice of distance decay is arbitrary
  - An alternative is to forget about specifying W
  - Alternatively, use different x-variables capturing concentric rings
  - Average of x-variable for different distance bands

3. Spatial data structure

- **Choice of distance decay is arbitrary**
  - e.g. $y = \alpha x_{0-100} + \beta x_{100-200} + \gamma x_{200-300} + \epsilon$



3. Spatial data structure

- **How to define spatial weight matrix using software**
  - SPATWMAT **in STATA, based on geographic coordinates**
  - SPWEIGHT **in STATA**
  - **Geoda**
  - SPATIAL STATISTICS TOOLBOX **in ArcGIS**
  - SPDEP **in *R***

- **Concentric rings should be calculated manually**

3. Spatial data structure

- **Koster, Van Ommeren, Rietveld (2014, *Economica*)**
  - **The effect of employment density on rents of commercial properties**

- **We have a dataset with**
  - **127,439 locations/firms**

- **Calculate spatial weight matrix and use that to calculate the *weighted employment (density) for each location***
  - **Hence, Wx**

- **Use spatial weights approach**

3. Spatial data structure

- **How do we determine $w_{ij}, \forall i, j$?**
  - $w_{ij} = I(d_{ij} < d_T)$
    where $d_T = 2.5$
- **Multiply w_{ij} by the employment x_j at $j, \forall i, j$ to get $\boldsymbol{Wx}$**

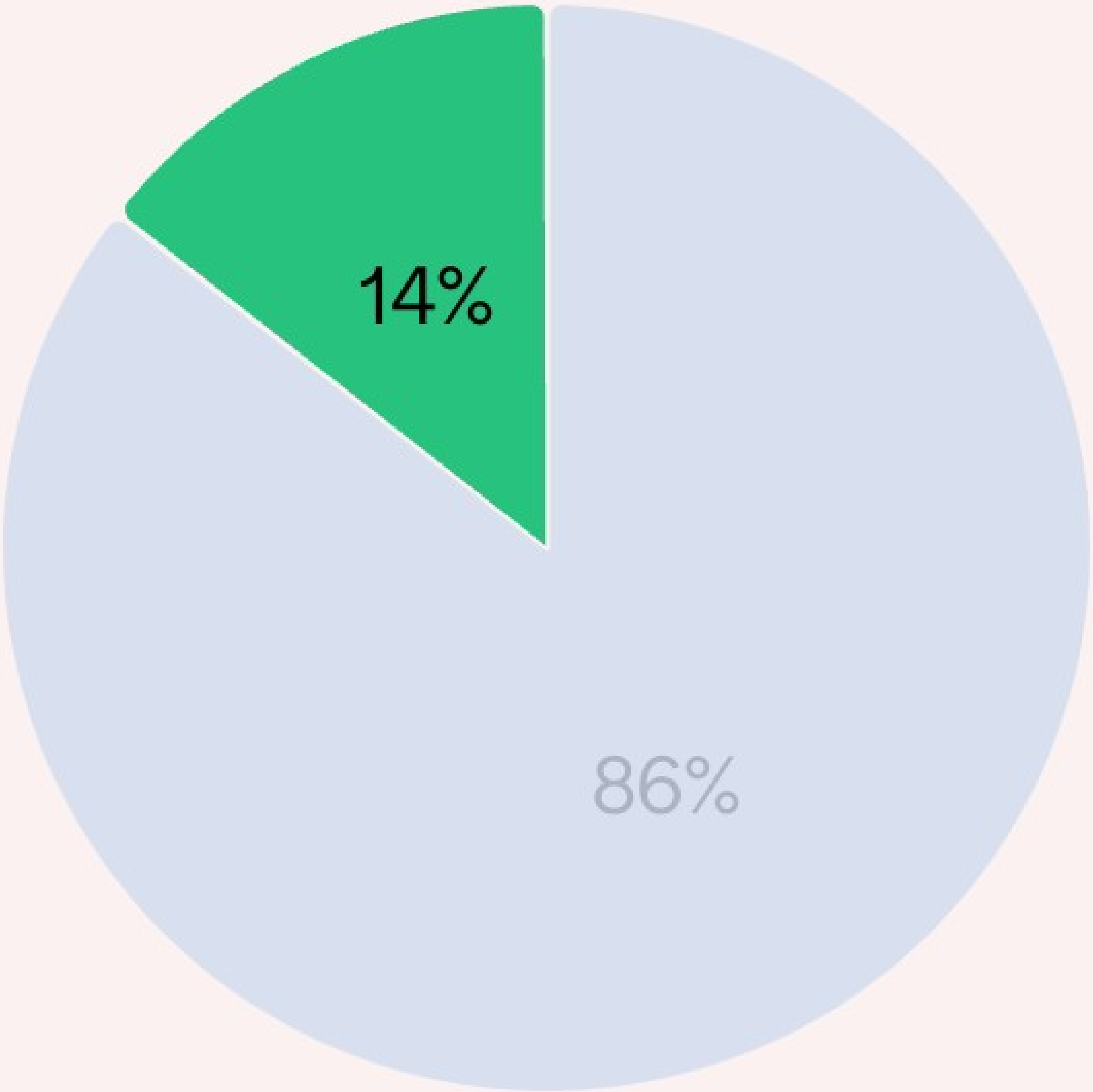
Distance matrix

from / to	1	2	3	4
1	0	2	3	5
2	2	0	1	2
3	3	1	0	5
4	5	2	5	0

Weight matrix

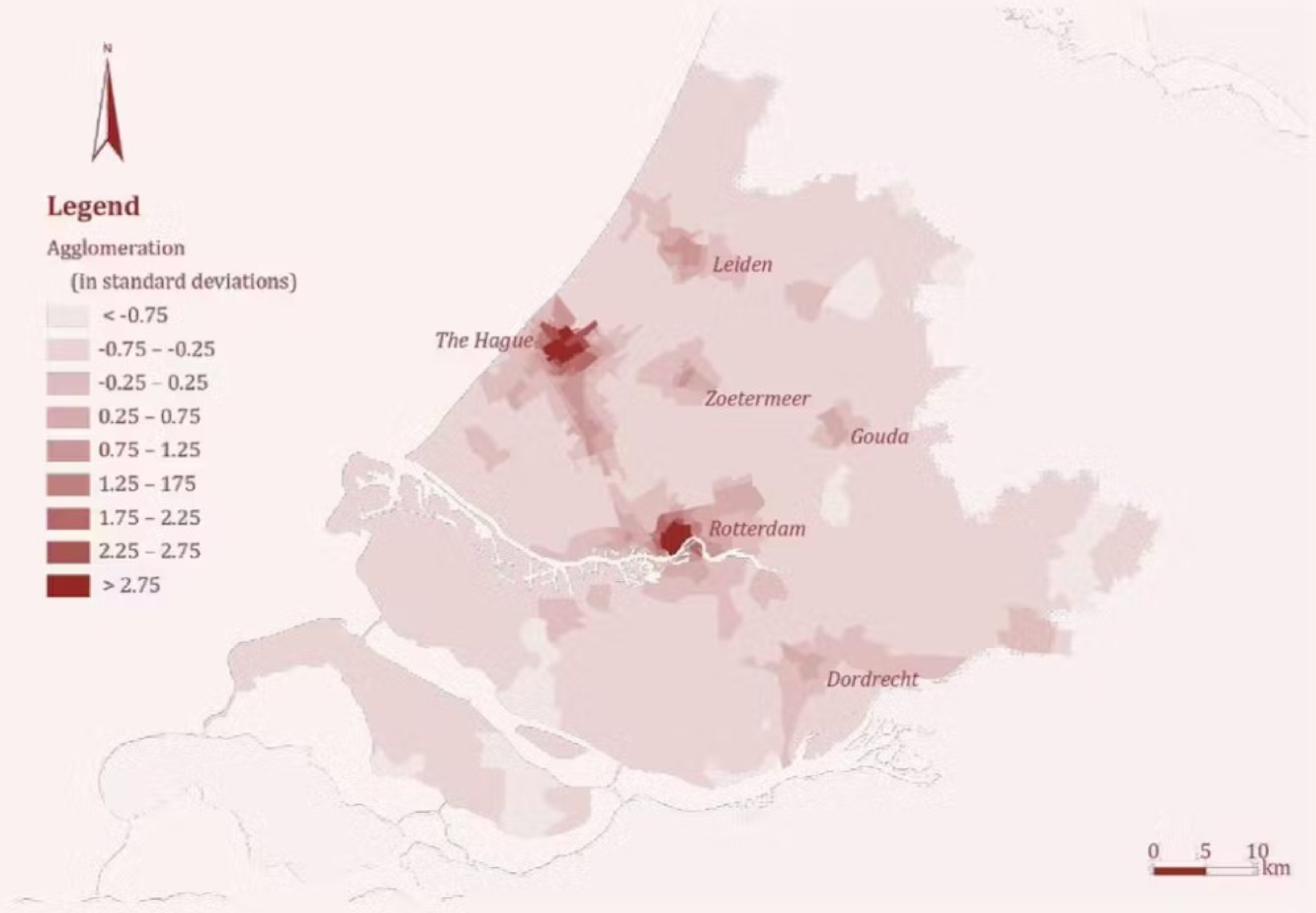
from / to	1	2	3	4
1	0	1	0	0
2	1	0	1	1
3	0	1	0	0
4	0	1	0	0

We use *unstandardised* weights in the weight matrix. Is this an issue in this application?



- 86% Yes, because for some areas employment will be weighted much more heavily ✗
- 0% Yes, because we are interested in average employment, rather than total employment at a location ✗
- 14% No, we are interested in total employment, rather than average employment at a location ✓

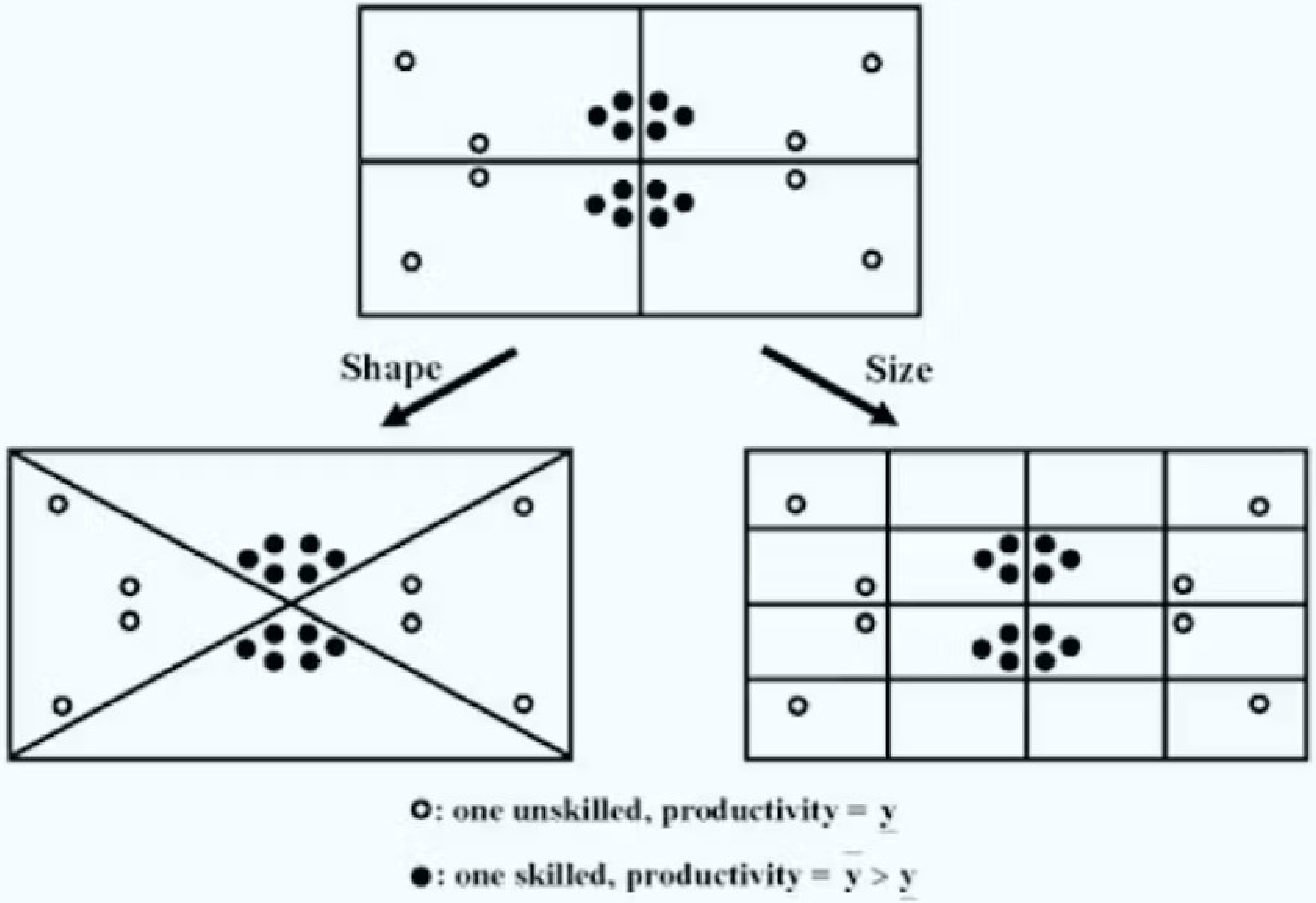
Weighted employment in Zuid-Holland



4. Modifiable areal unit problem

- Usually we do not have space-continuous data
  - ‘Dots’ to ‘boxes’
- Data is aggregated at
  - Postcode areas
  - Municipalities
  - Regions
  - Countries
- Problems:
  - Aggregation is often arbitrary
  - Areas are not of the same size
- This may lead to distortions
  - Modifiable areal unit problem (MAUP)

- **An illustration:**



Briant, Combes and Lafourcade (2010, JUE)

- **Aggregation seems to be important!**

4. Modifiable areal unit problem

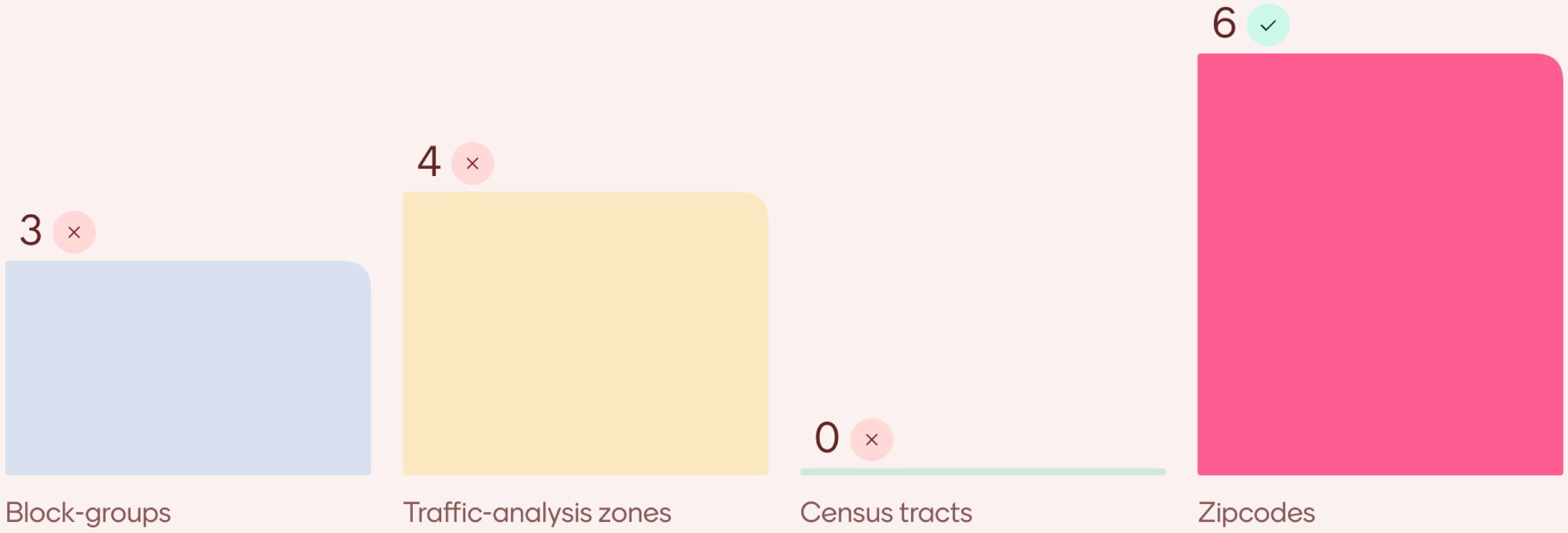
- **Briant et al. (2010) investigate whether choice matters for regression results**

4. Modifiable areal unit problem

- MAUP is of secondary importance
  - If y and x are aggregated in the same way
  - Matters more for larger areas (*e.g.* regions)
  - Use meaningful areas if possible

- Specification issues are much more important

In what of the below maps on accidents in Tampa FL, you think the MAUP could be the most pronounced? *(from Xu et al., 2018)*



5. Summary

Today:

- 'Space' in economics is becoming more and more important
- Incorporating space in econometric applications is not straightforward
- Important to define the spatial structure of the data
  - Spatial weight matrices
  - Modifiable areal unit problem

VU VRIJE UNIVERSITEIT AMSTERDAM

Spatial econometrics (1)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate