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

- Hans Koster → URBANECONOMICS.NL
  - Professor of Urban Economics and Real Estate
  - h.koster@vu.nl
  - Lectures
  - Programme director of STREEM (streem.sbe@vu.nl)

- Materials
  - All course materials, lecture slides, etc. can be accessed via www.urbaneconomics.nl/aese
  - If there is anything unclear, let me know!

1. Introduction

- This course
  - Learn about advanced tools and techniques important for spatial economics
    ➔ No theory – an applied course!

- Do not hesitate to ask questions during the class!

- Notation on slides
  - Most important concept are underlined
  - Questions (via Menti), exercises and applications
    → On red slides

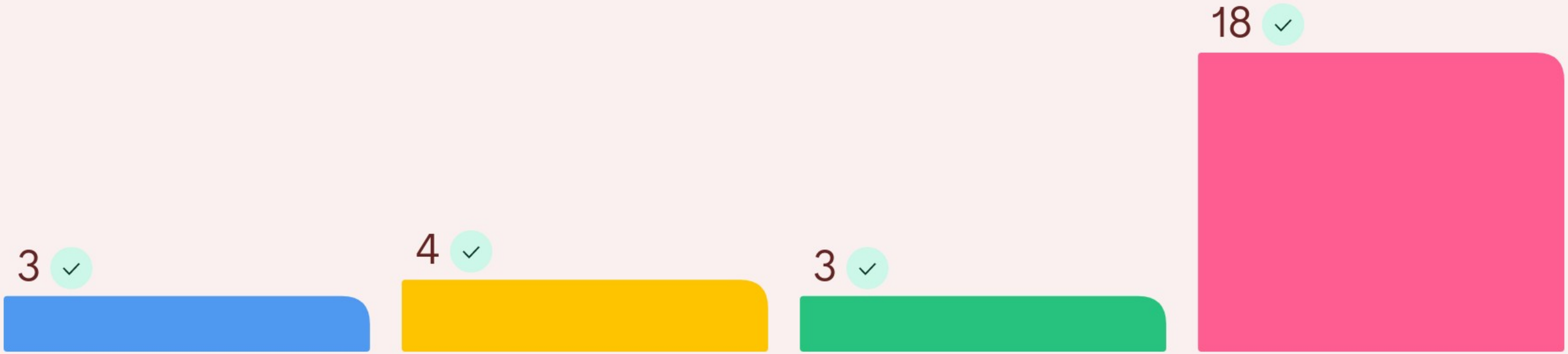
1. Introduction

- **Topics:**
  1. **Spatial econometrics**
     - **Spatial data, autocorrelation, spatial regressions**
  2. **Discrete choice**
     - **Random utility framework, estimating binary and multinomial regression models**
  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**

1. Introduction

- *Monday:*
  1. Spatial econometrics (1+2+3)
- *Tuesday:*
  2. Discrete choice (1+2+3)
- *Wednesday:*
  3. Identification (1+2+3)
- *Thursday:*
  4. Hedonic pricing (1+2)
  5. Quantitative spatial economics (1+2)

Test question: Who won the nobel prize in Economics in 2021? *(multiple answers may be correct)*



David Card

Guido Imbens

Joshua Angrist

Technically, there is no Nobel Prize in economics. Instead, there is the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel. Etc.

1

22

1. Introduction

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021

Ill. Niklas Elmehed © Nobel Prize Outreach.

David Card

"for his empirical contributions to labour economics"

Ill. Niklas Elmehed © Nobel Prize Outreach.

Joshua D. Angrist

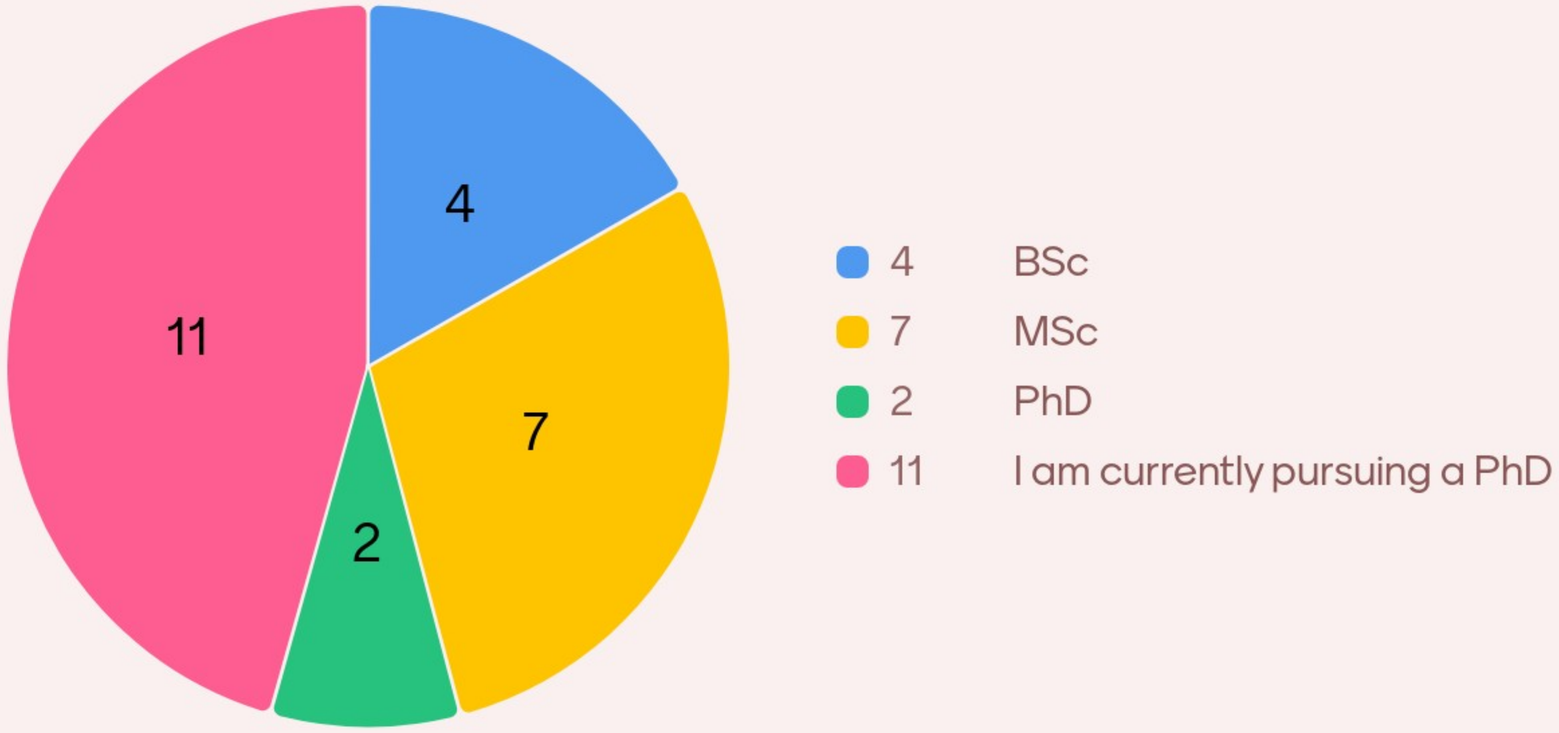
Ill. Niklas Elmehed © Nobel Prize Outreach.

Guido W. Imbens

"for their methodological contributions to the analysis of causal relationships."

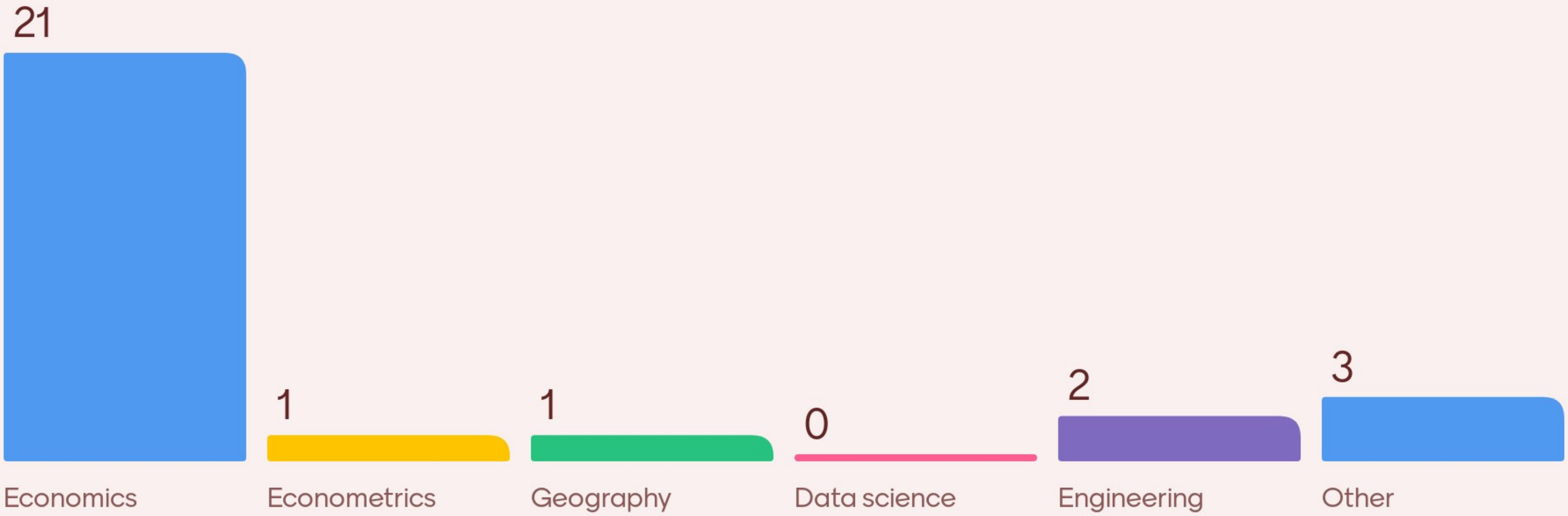
October 11, 2021

What is your highest achieved degree?

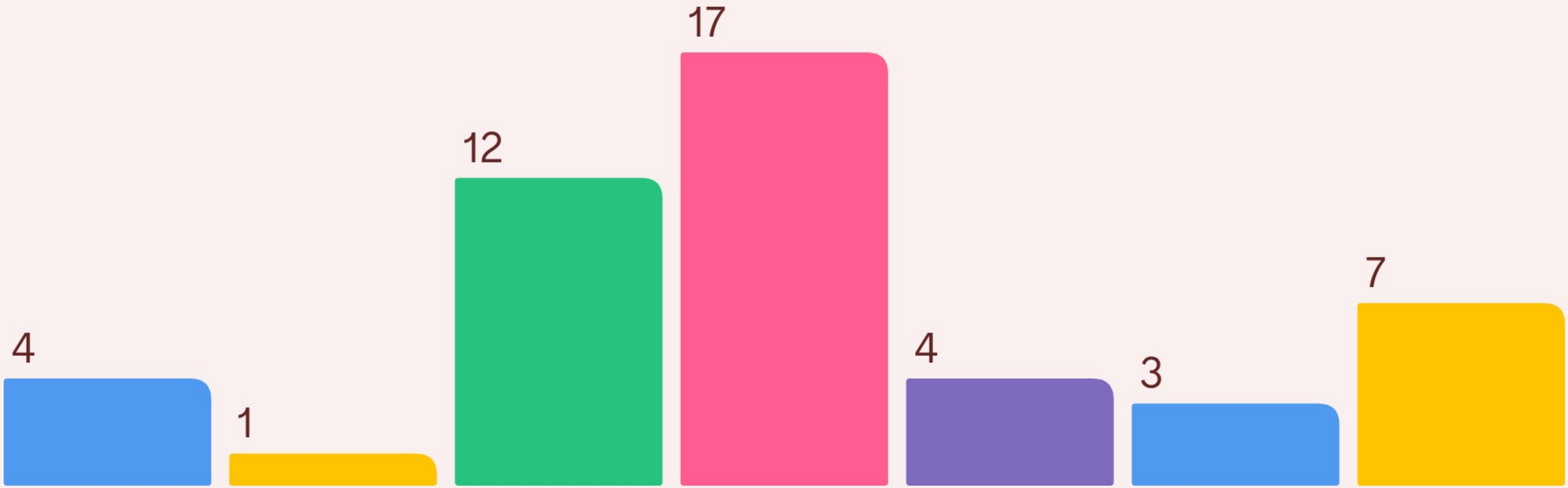


WHAT IS YOUR MAIN FIELD?

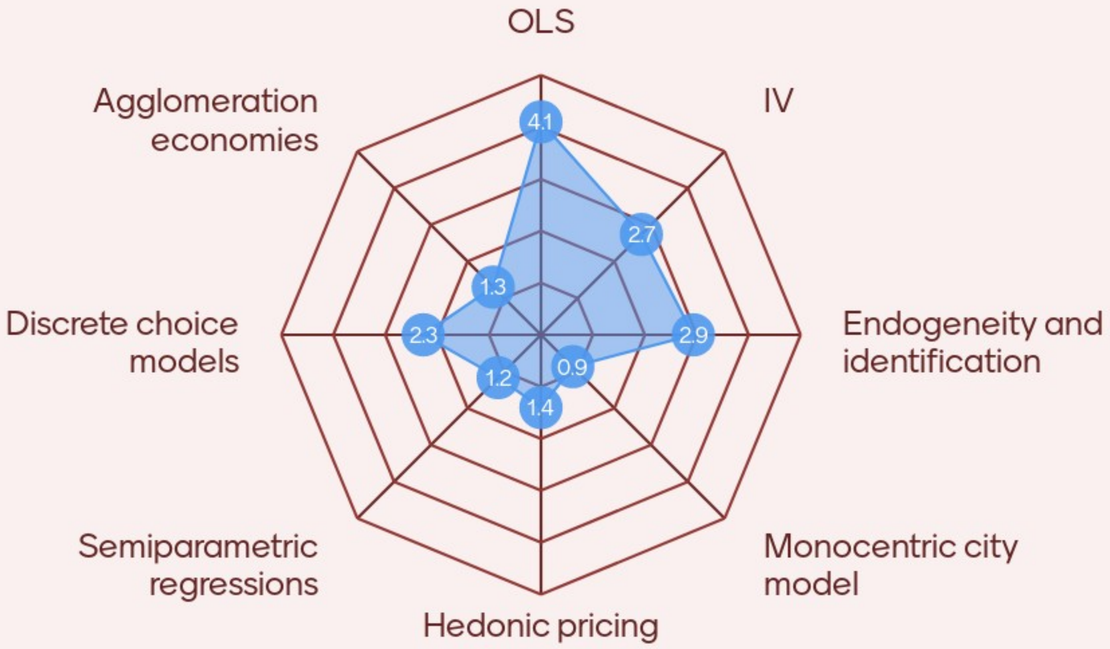
What is your main field?



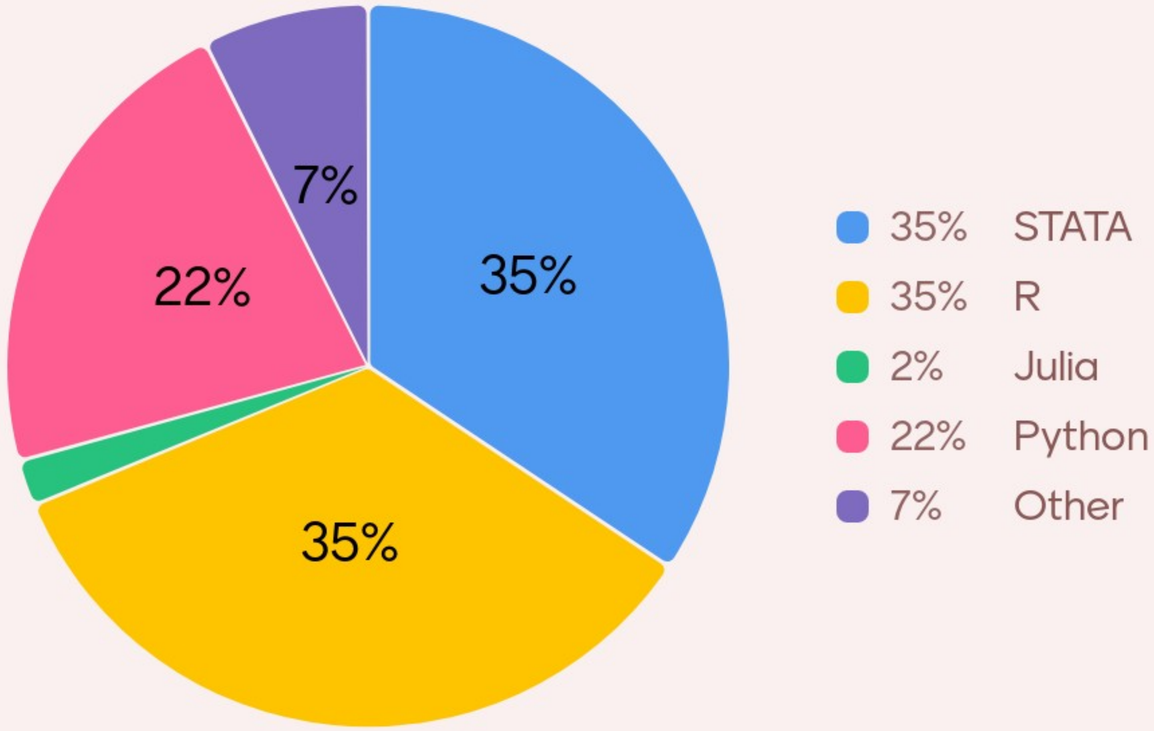
I have attended the following courses:



Please let me know how familiar you are with the following concepts:

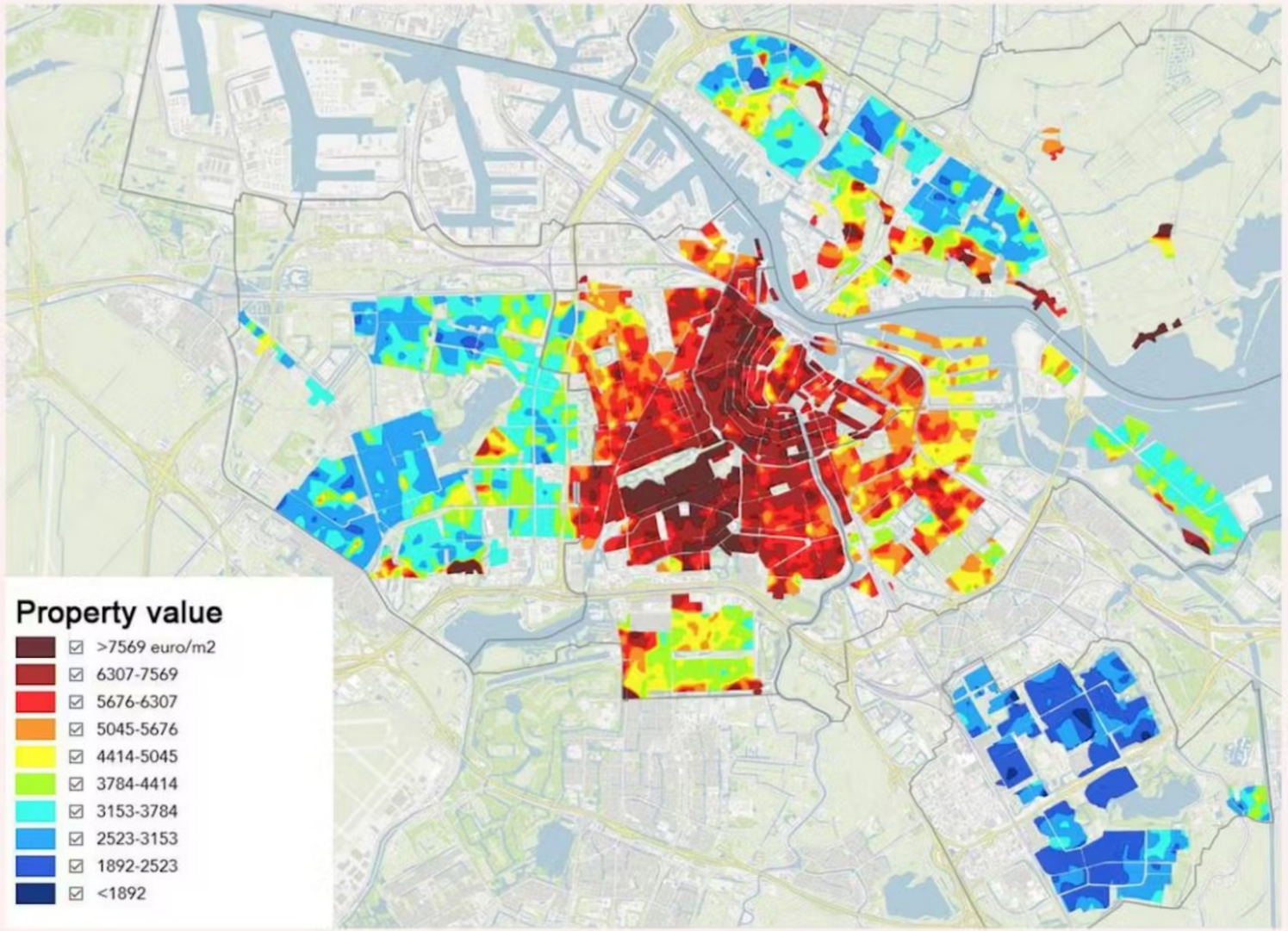


I have experience in:

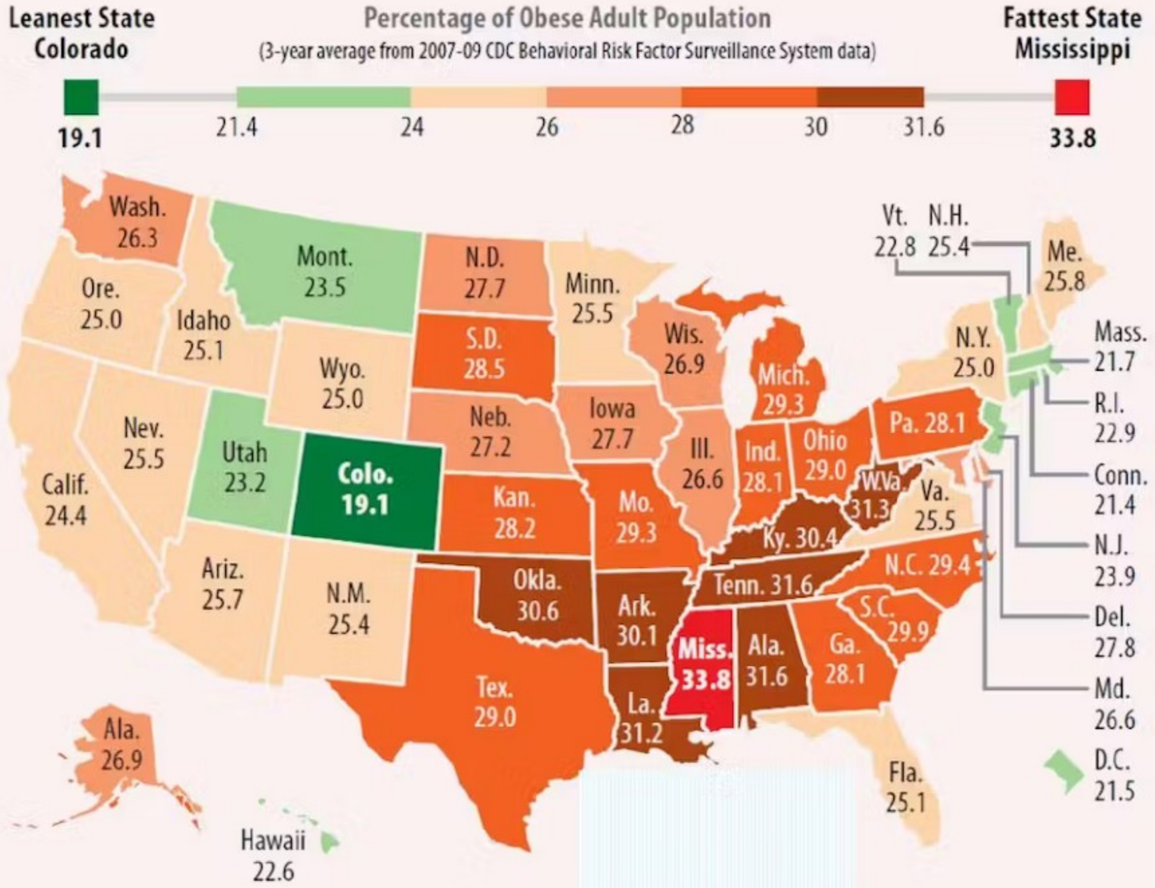


1. Introduction

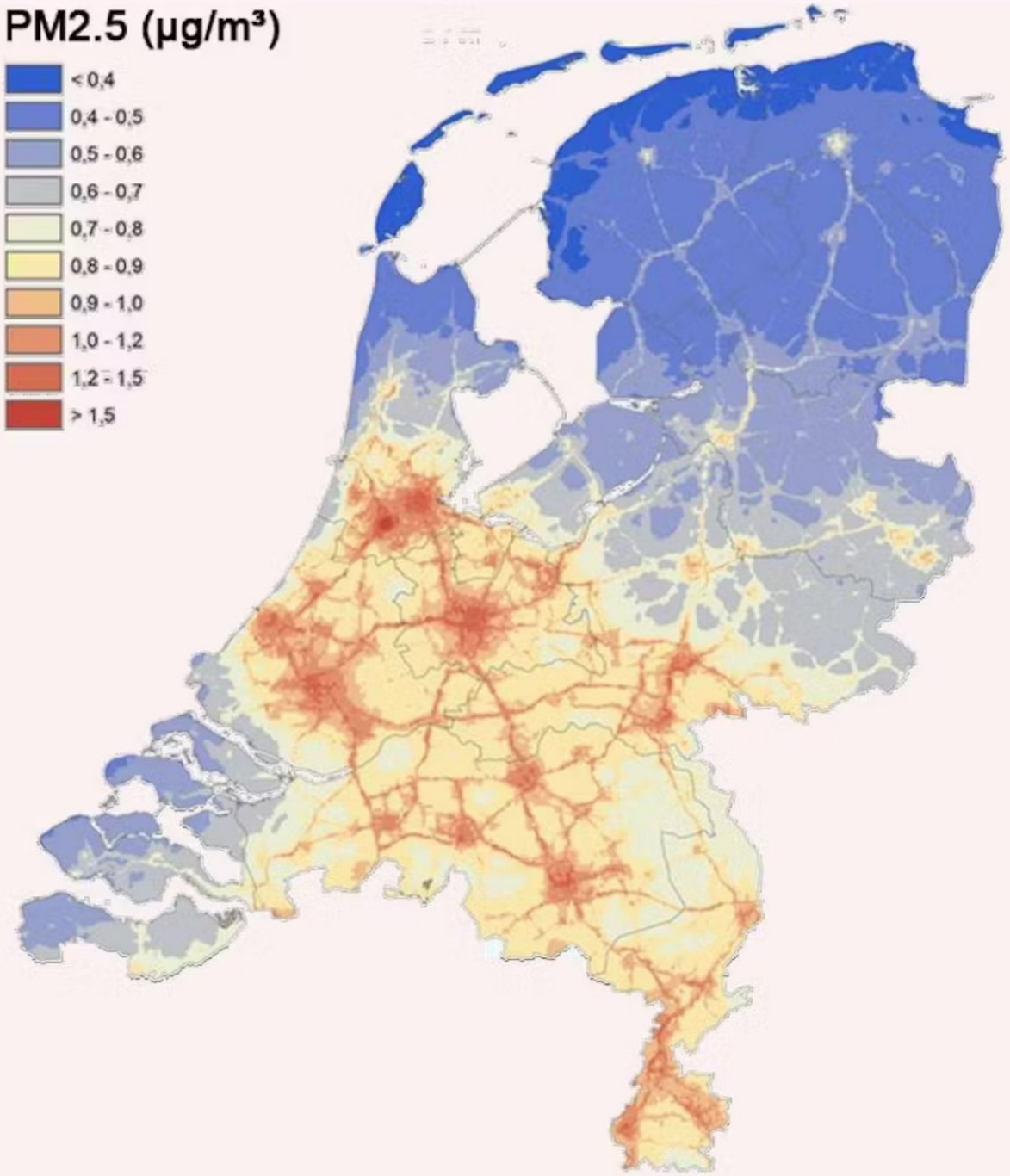
- What is the m^2 price in Amsterdam?



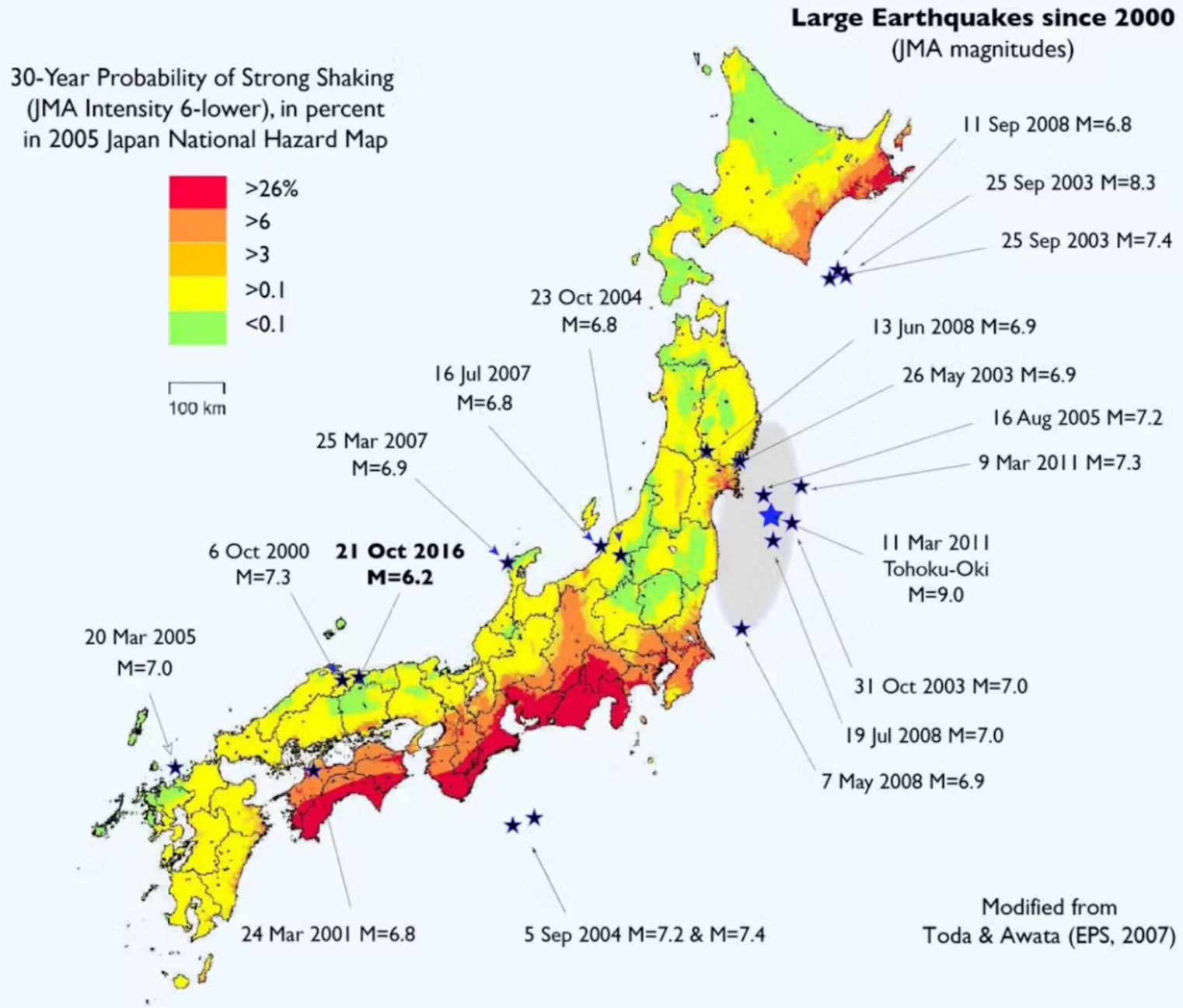
1. Introduction



1. Introduction



1. Introduction



1. Introduction

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

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

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

Please mention examples of spatial data

1. Introduction

- **Why not use OLS?**

- **Let's have the standard OLS equation**

$$y_i = \beta x_i + \epsilon_i \quad (1)$$

where i is the location

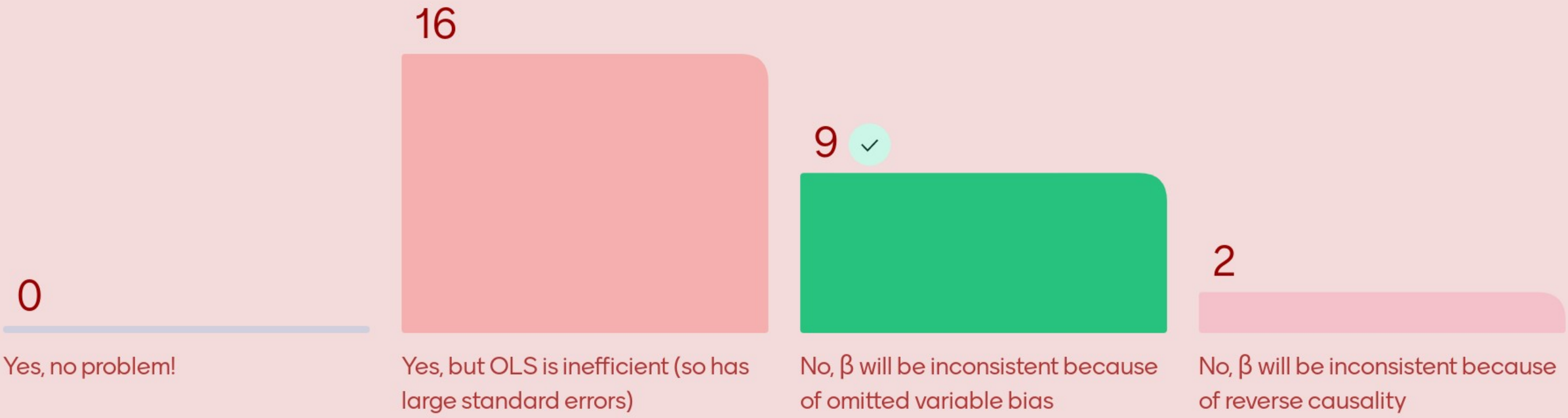
- **Let's consider the case where**

$\mathrm{cov}(x_i, x_j) \neq 0$ **and**

$$y_i = \beta x_i + f(x_j) + \epsilon_i \quad (2)$$

where j are all other locations

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



1. Introduction

- This week
  - Learn about how to deal with spatial data
  - … and spatial econometrics

- Plan:

Lecture #1:	Spatial data
Lecture #2:	Spatial autocorrelation and regressions
Lecture #3:	Spatial regressions (cont'd)
Assignment:	Open space and school quality

- **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**

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

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

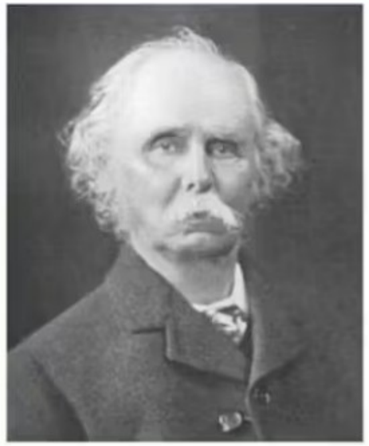
    $\rightarrow \boldsymbol{X^{-1}X} = \boldsymbol{XX^{-1}} = \boldsymbol{I}$
  - **More details in the appendix of the syllabus**

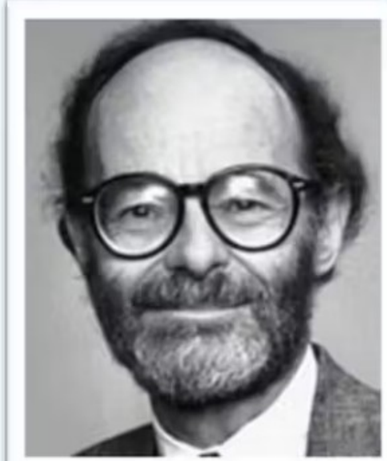
2. Space in economics

- 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

- **Economists and space**





YOU?

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)**

2. Space in economics

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

3. Spatial data structure

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

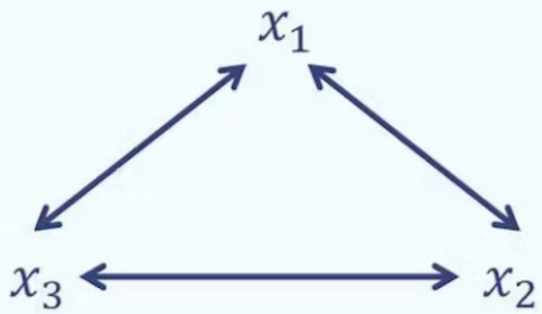
$$x_{t-3} \longrightarrow x_{t-2} \longrightarrow x_{t-1} \longrightarrow x_t$$

- **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

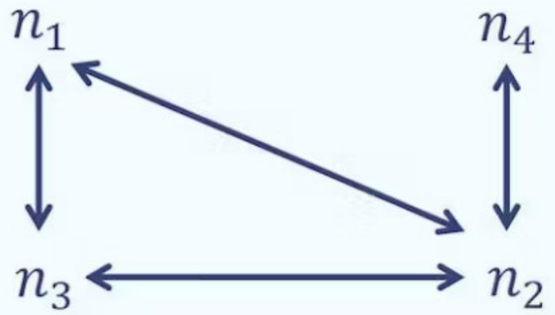
- How to define spatial matrices?

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

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

3. Spatial data structure

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



to

from 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

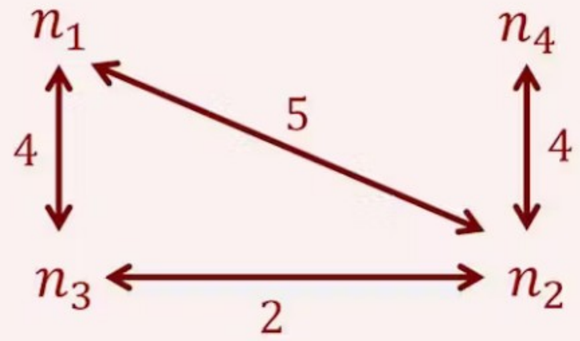
- **Matrices can be standardised**
  - **Different principles can be used**
  - **Most common: *row-standardisation*:**

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

    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**

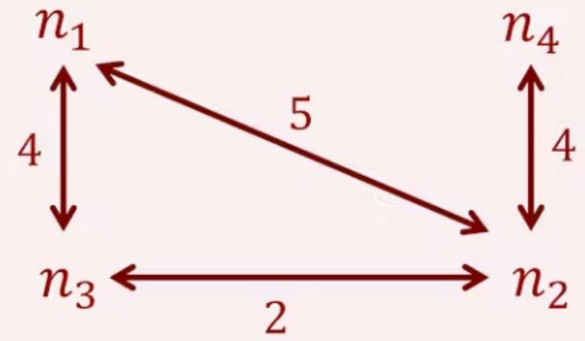


to

from / W	n_1	n_2	n_3	n_4
n_1				
n_2				
n_3				
n_4				

Create an inverse distance weight matrix with row-standardised weights

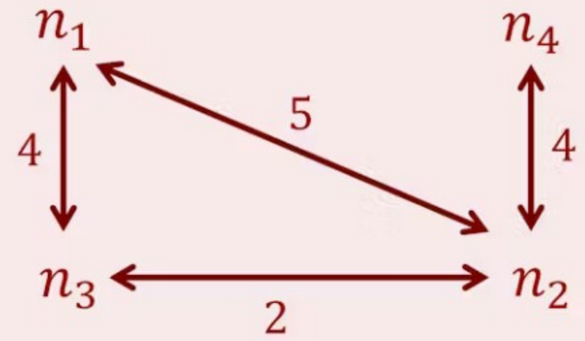
- **Create an *inverse* distance weight matrix with row-standardised weights**



from \ to	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

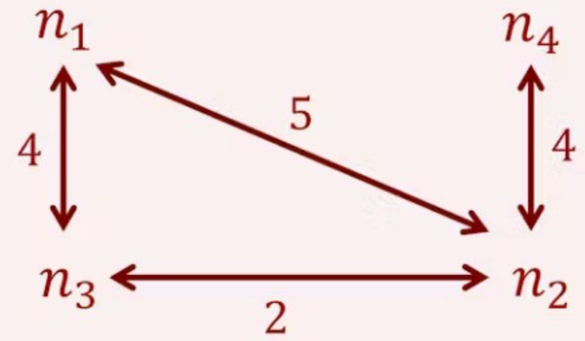
- **Create an *inverse* distance weight matrix with row-standardised weights**



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**



from \ to	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

3. Spatial data structure

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

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

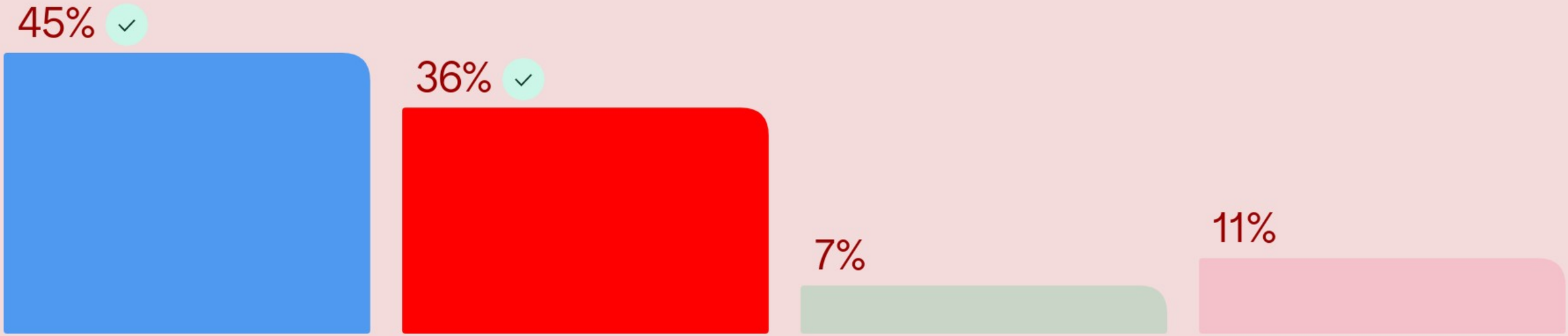
$$\boldsymbol{y} = \beta \boldsymbol{e} + \boldsymbol{W}\boldsymbol{e}'\gamma + \boldsymbol{\epsilon} \qquad (3)$$

$\boldsymbol{y} =$ income; $\boldsymbol{e} =$ education

Say that $\boldsymbol{W}$ depends on *the number of friends you have*

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?



The number of friends could be correlated to other individual characteristics (e.g. social capabilities)

Income could influence the number of friends you have, implying an endogenous weight matrix

Due to multicollinearity, β and γ cannot be separately identified

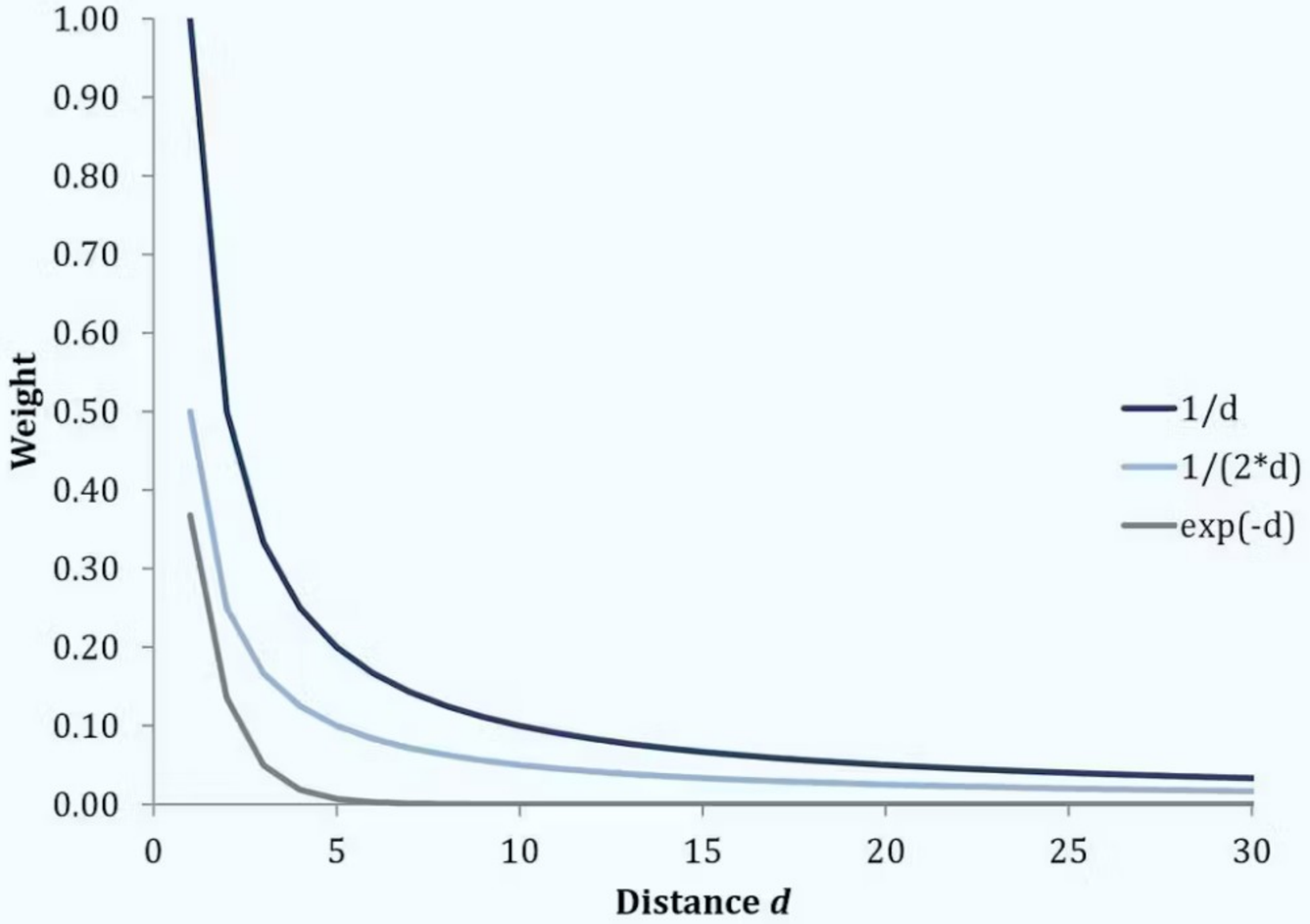
The number of friends is measured will always be measured with substantial measurement error so that W is not informative

3. Spatial data structure

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

3. Spatial data structure

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

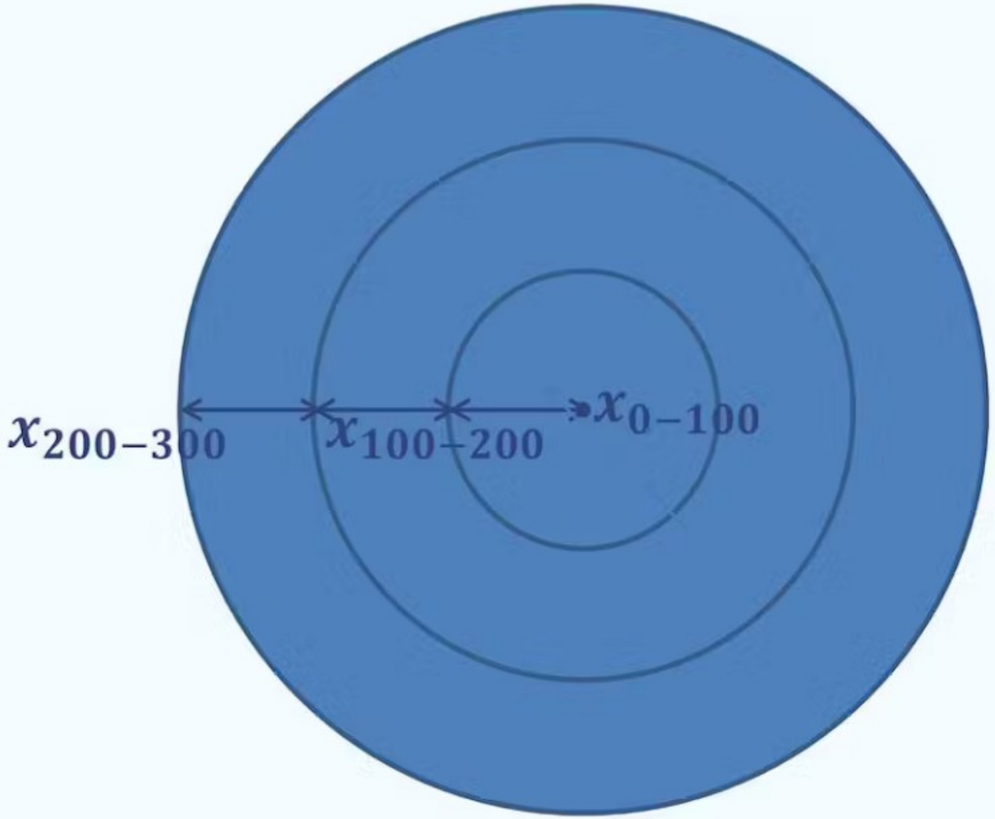


3. Spatial data structure

- **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 $\boldsymbol{Wx}$**

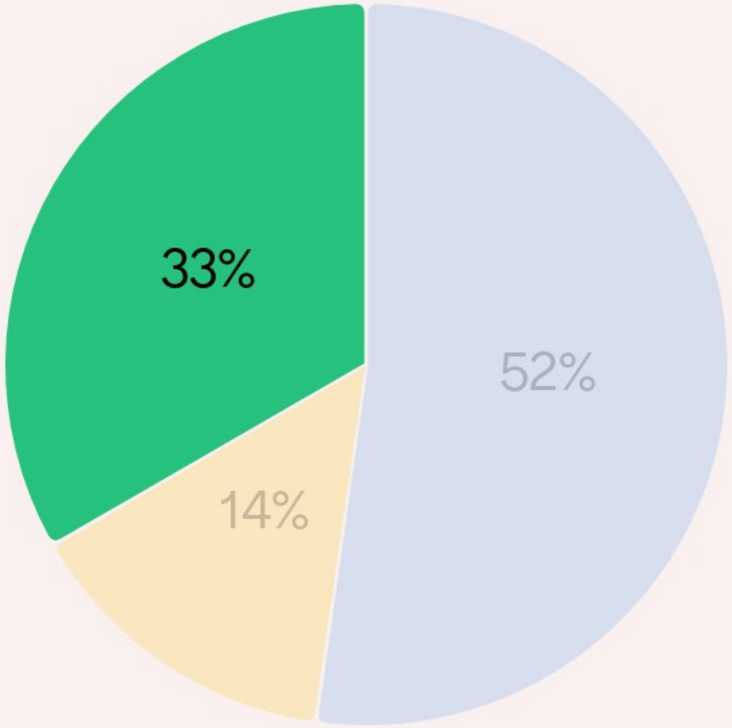
Distance matrix

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

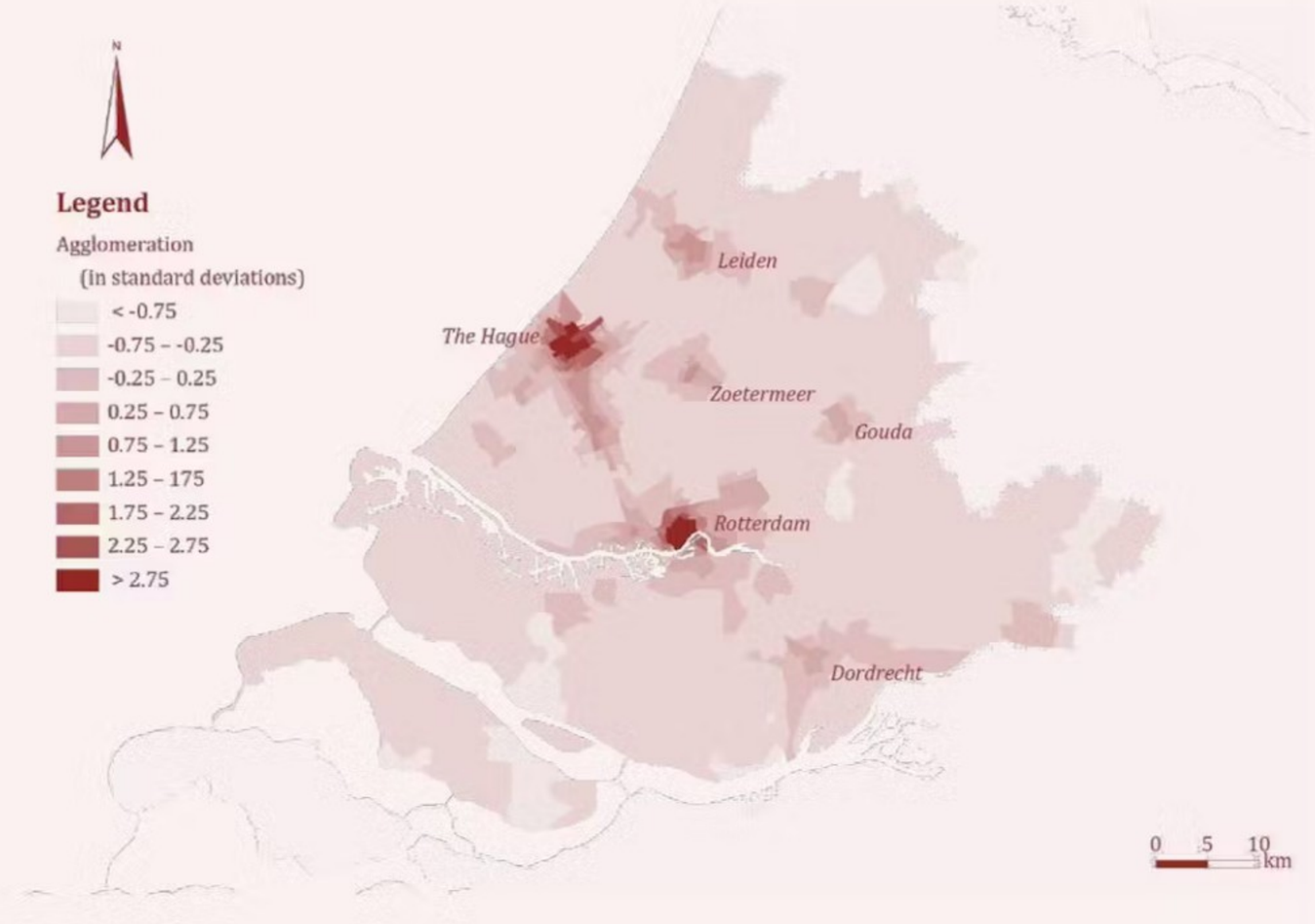
from / to	1	2	3	4
1	0	1	0	0
2	1	0	1	1
3	0	1	0	0
4	0	1	0	0

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



- 52% Yes, because for some areas employment will be weighted much more heavily
- 14% Yes, because we are interested in average employment, rather than total employment at a location
- 33% No, we are interested in total employment, rather than average employment at a location

Weighted employment in Zuid-Holland

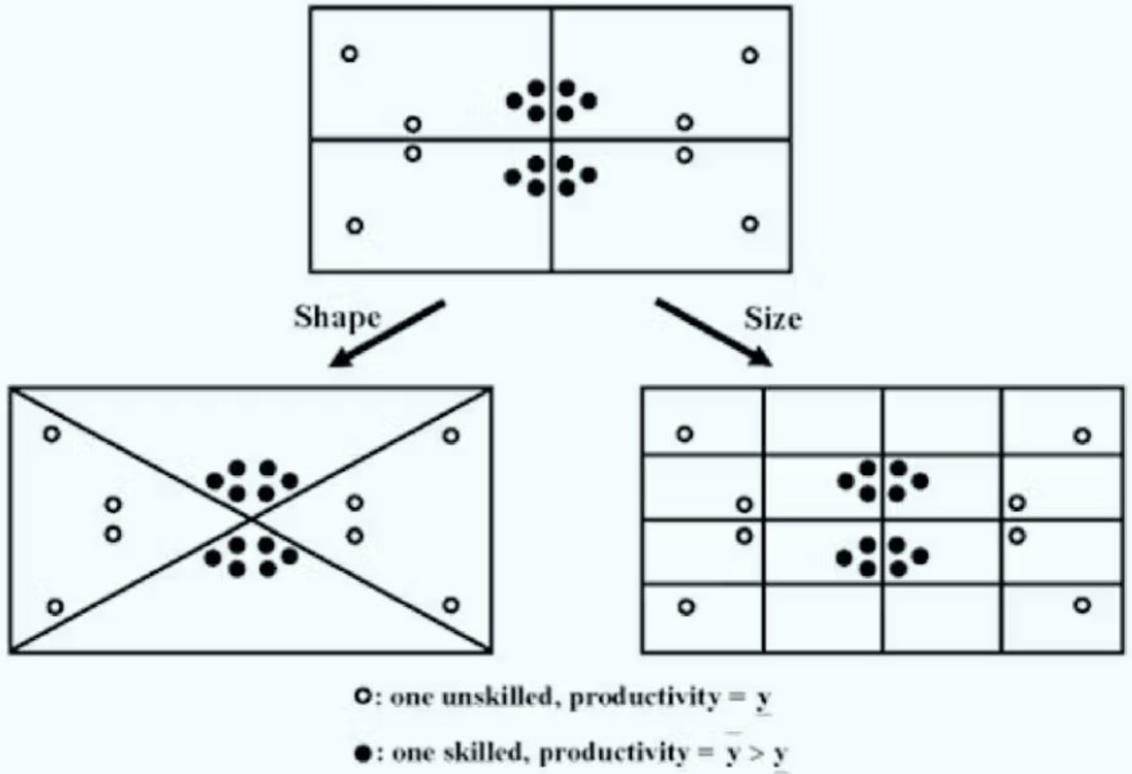


4. Modifiable areal unit problem

- 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)

4. Modifiable areal unit problem

- **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**

341 Employment Areas (EA)

341 Small squares (SS)

21 Régions (RE)

22 Large squares (LS)

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)*

0	0	0	0 ✓
Block-groups	Traffic-analysis zones	Census tracts	Zipcodes

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

Spatial econometrics (1)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute