Geospatial analysis

John Östh

This presentation

- Let's think about the *grammar* of geography
  - Presenting a few core features in spatial analysis that are fundamental for spatial analysis
  - And a few geographical levels
- The geographical scale, and scaling up computations
  - A few notes on opportunities and ethics
- We will look at a few geographical theories
  - Presentation of theories at different geographic scales
  - Questions that can be answered using Big Data

Let's think about the *grammar* of geography

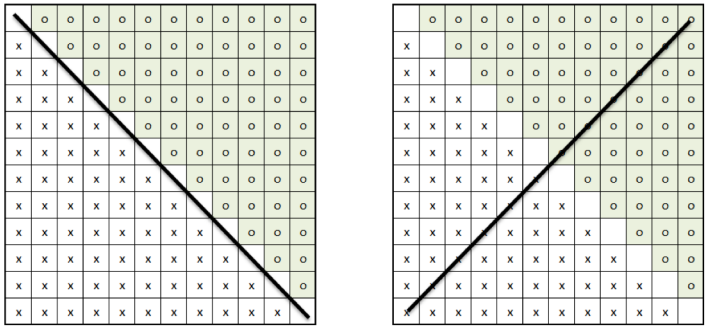
- SPACE
  - All the same – everywhere. Easy to measure, using cost, time, distance…
- PLACE
  - Any unique location. Place is difficult to model, since most of the qualities are culturally attributed (borders, feelings for, history, relation to other places, etc.)
  - Can be classified – but classifications will never be correct
    - Could be defined administratively – usually historically/culturally
    - Could be defined physically
    - …
- Geographical theory
  - ~ the intersection between place and space
  - Often where we start our work (spatial models intertwined with place)

Let's think about the *grammar* of geography

- Example of the intersection between space and place
  - Define an accurate measure of an individual's neighbourhood (why?: no one is an island, more later)
    - Three ways (crudely defined)
      - Area-based
        - » The usual method. All individuals in an area (object) are aggregated and used as a representative of the area. This model is very common since data is 'spatial' already. Municipalities, wards, blocks Counties,…
      - Radii-based
        - » A relatively common method. All individuals within a fixed distance from the reference point (line or polygon) are aggregated…
      - *K*-nearest-based
        - » A relatively uncommon method. The k-nearest individuals from any point (line or polygon) are aggregated…

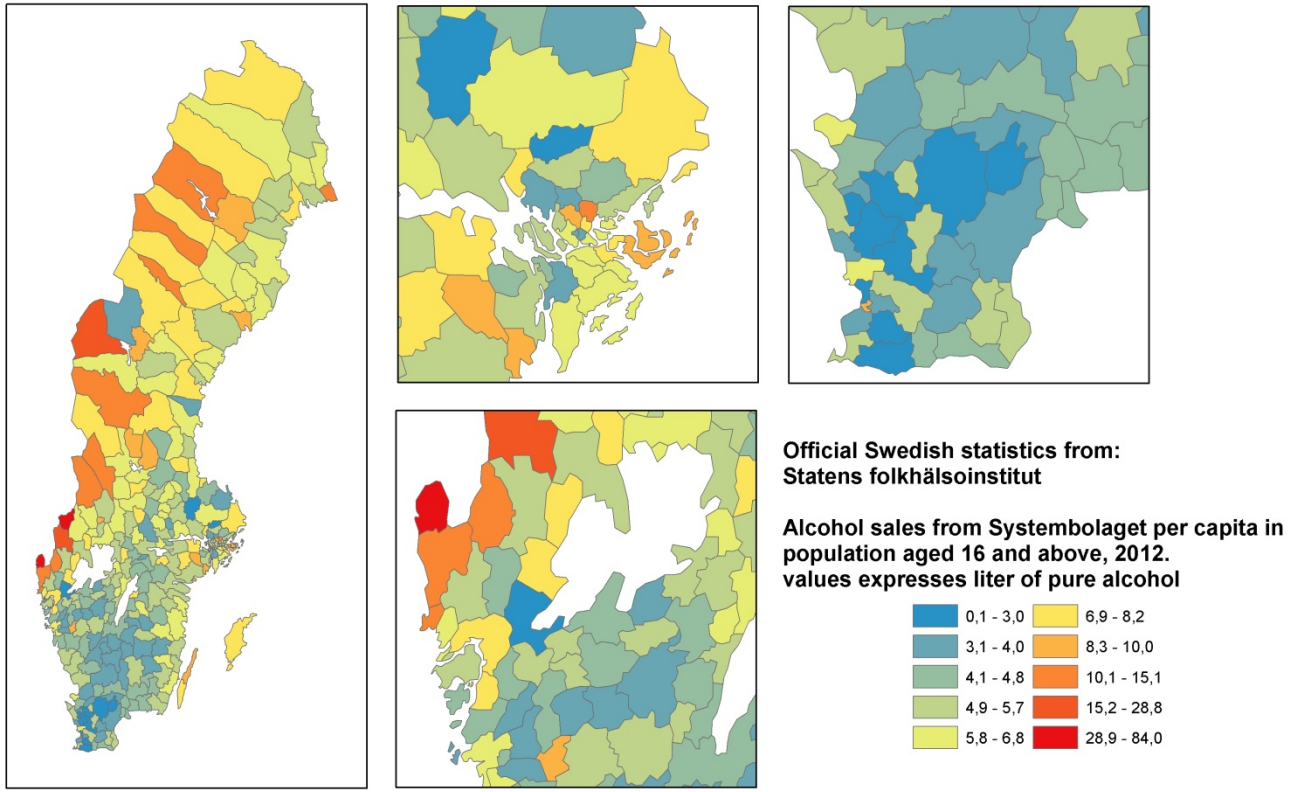
Let's think about the *grammar* of geography

- The use of **area** based aggregation techniques is associated with several potential biases:
  - Boolean border effects
    - Variables become over-simplified; within = 1, outside = 0.
  - MAUP (Modifiable Areal Unit Problem)



Let's think about the *grammar* of geography

More errors #1

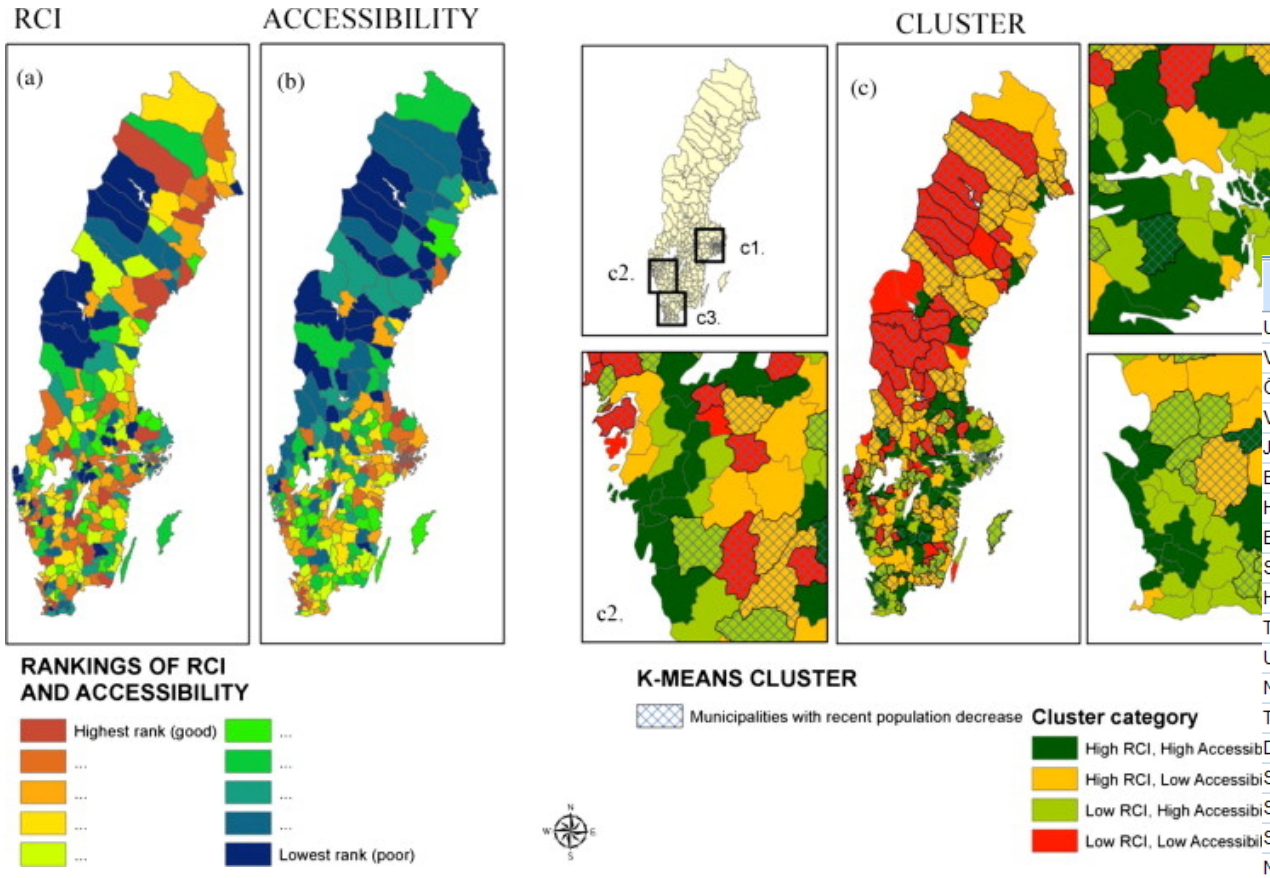


Let's think about the *grammar* of geography

More errors #2

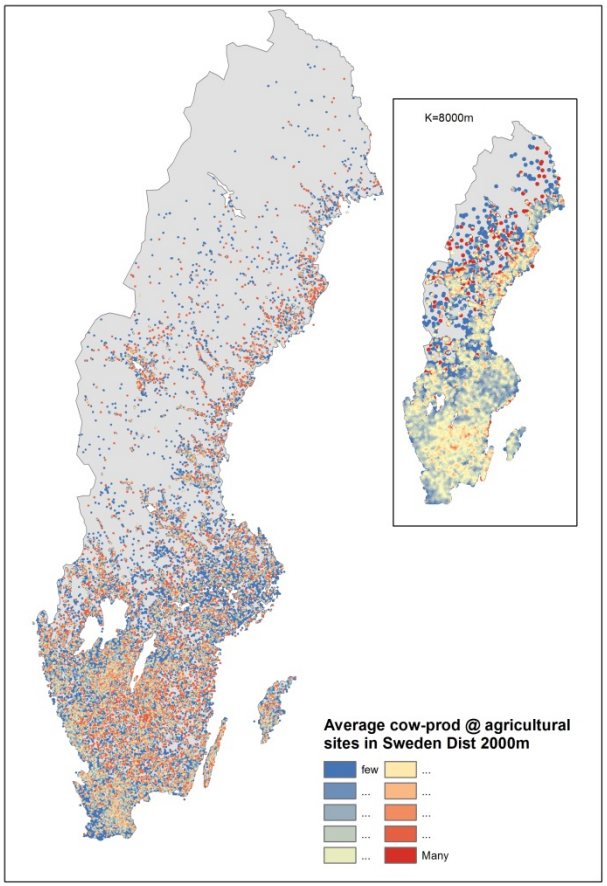
	County 1	County 2	County 3	County 4	County 5	
REGIONAL CAPITAL	5%	5,5%	4,8%	5%	5,5%	BETWEEN-MUNICIPALITY TYPE VARIATION = SIGNIFICANT
SUBURB MUNICIPALITY	21%	20%	18%	22%	19%	
COMMUTING MUNICIPALITY	12%	11%	13%	10%	10%	
INDUSTRY MUNICIPALITY	7%	6%	7%	8,5%	9%	
RURAL MUNICIPALITY	19%	22%	18,5%	13%	13,5%	
	BETWEEN-COUNTY VARIATION = INSIGNIFICANT					

Let's think about the *grammar* of geography



NameKommun	Kkod	RCIZ	RCIrang	AccRang
Upplands Väsby	114	-,131443	194	21
Vallentuna	115	-,028042	146	28
Österåker	117	,051237	113	32
Värmdö	120	-,220132	228	37
Järfälla	123	,108494	89	13
Ekerö	125	,301732	31	23
Huddinge	126	,074851	104	6
Botkyrka	127	-,369099	270	14
Salem	128	-,054889	156	22
Haninge	136	-,308826	252	24
Tyresö	138	,230996	47	16
Upplands-Bro	139	-,221791	230	38
Nykvarn	140	-,425088	277	46
Täby	160	,926556	3	12
Danderyd	162	1,946322	1	4
Sollentuna	163	,563241	12	10
Stockholm	180	,818352	5	3
Södertälje	181	,027508	128	18
Nacka	182	,617642	10	7
Sundbyberg	183	-,117950	191	2
Solna	184	,340726	25	1
Lidingö	186	1,059402	2	5
Vaxholm	187	,286325	35	27

Let's think about the *grammar* of geography



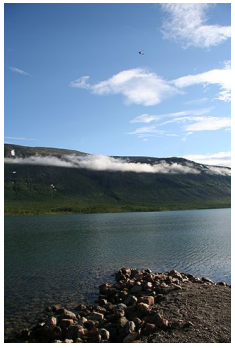
Example of radii-based statistics

Distance	Meaning
100m radii	Farm
200m radii	Still farm
800m radii	Small village
2000m	Roads, water, epidemics…
8000m	Shared service market (agglomeration services)

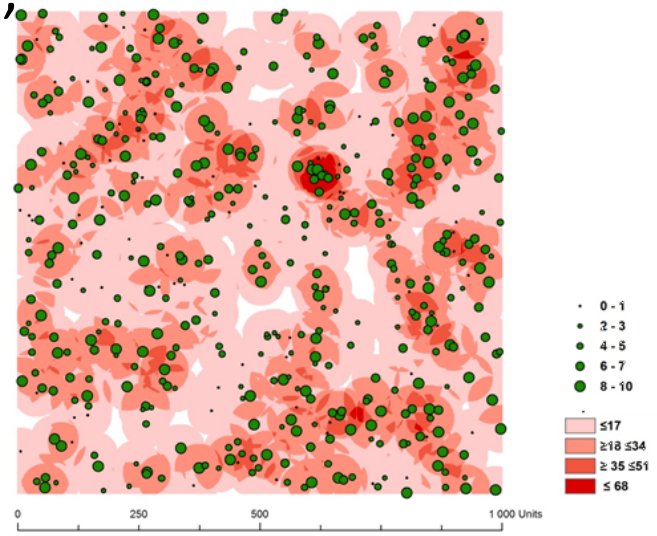
Let's think about the *grammar* of geography

- The use of radii based aggregation for the creation of neighbourhoods is associated with:
  - Boolean border effects
    - Variables become over-simplified; within = 1, outside = 0.
    - Risk for *"high-variation"* datasets

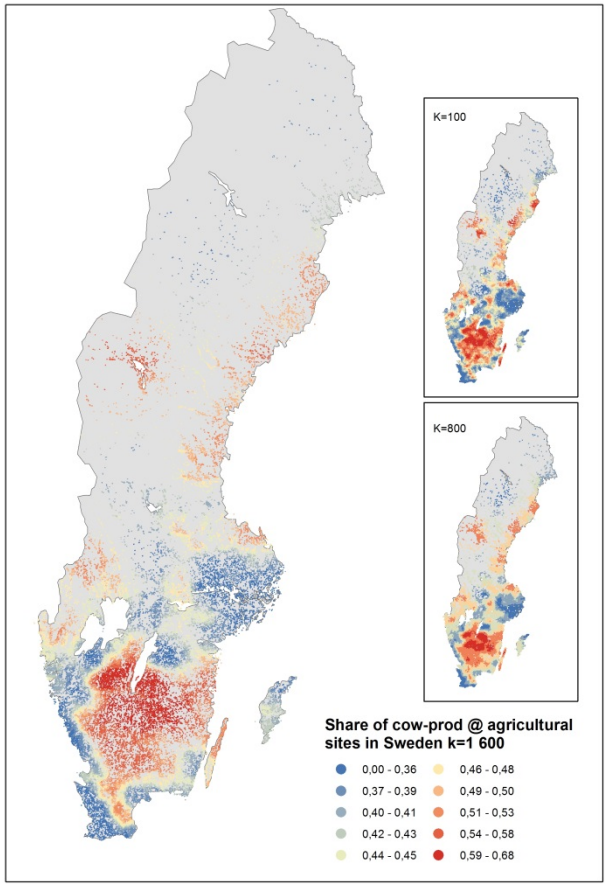
https://upload.wikimedia.org/wikipedia/commons/e/e9/Stockholms-stadshus-fran-sodermalm.jpg



https://upload.wikimedia.org/wikipedia/commons/thumb/4/40/Svenska_fj%C3%A4llen_-_panoramio.jpg/320px-Svenska_fj%C3%A4llen_-_panoramio.jpg



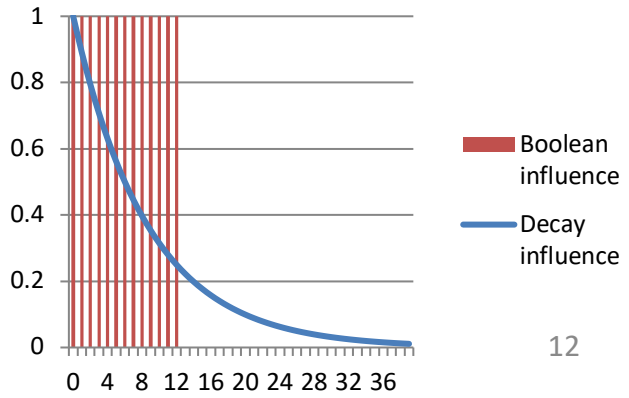
Let’s think about the *grammar* of geography



Example of *k*nn based statistics

Let's think about the *grammar* of geography

- The use of *k*-nearest-based aggregation is associated with:
  - Computationally demanding techniques
  - Distant object may be less influential than near objects – potential need for decay profiles
    - How to think when it comes to distance decay?



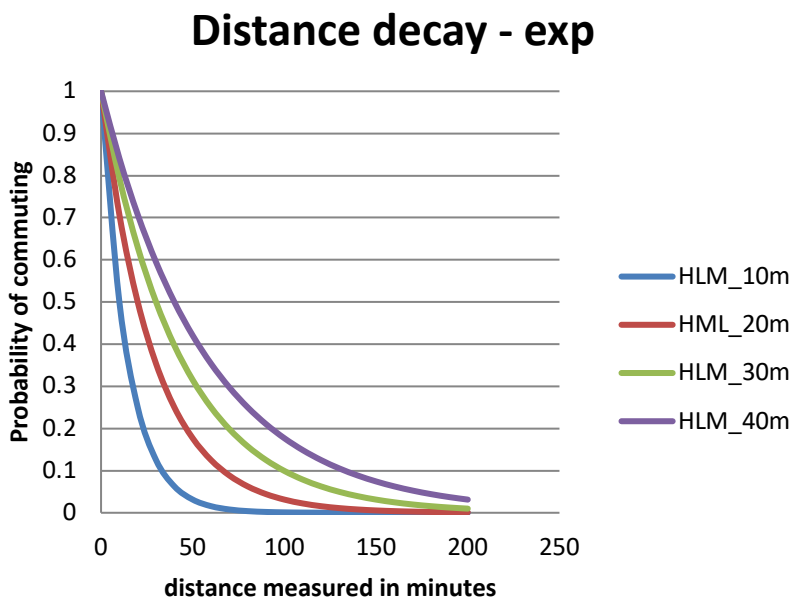
Let's think about the *grammar* of geography

Distance/cost/time-decay

- Think of it as economists think of reservation wage
  - Where do your expectation meet offers?
    - Now add space.
      - Commuting
        - » Would you commute 1h for a well paid job?
        - » 2h?
        - » 2h – less well paid?
        - » 2h & well paid but you have kids at day-care…
      - Would you move?
      - Would you migrate?

Observations suggest that the form of decay is dependent on activity:

Decay parameter can be found statistically and mathematically
Kids movement = -exponential normal
Commuting = -exponential (Carbon 14) = ln(0,5)/median
Migration = -power



A new way of determining distance decay parameters in spatial interaction models with application to job accessibility analysis in Sweden, Östh John, J Lyhagen, A Reggiani European Journal of Transport and Infrastructure Research 16 (2), 344-362

Let's think about the *grammar* of geography

- Geographical scale (in the next few slides)
  - Contexts and neighbourhoods
    - Individual in the centre
  - Urban/rural
    - The company, service provider in the centre
  - The region/nation
    - The state, firm in the centre

individual

Let's think about the *grammar* of geography

- Contexts and neighbourhoods

  Main message – there is not ONE geography

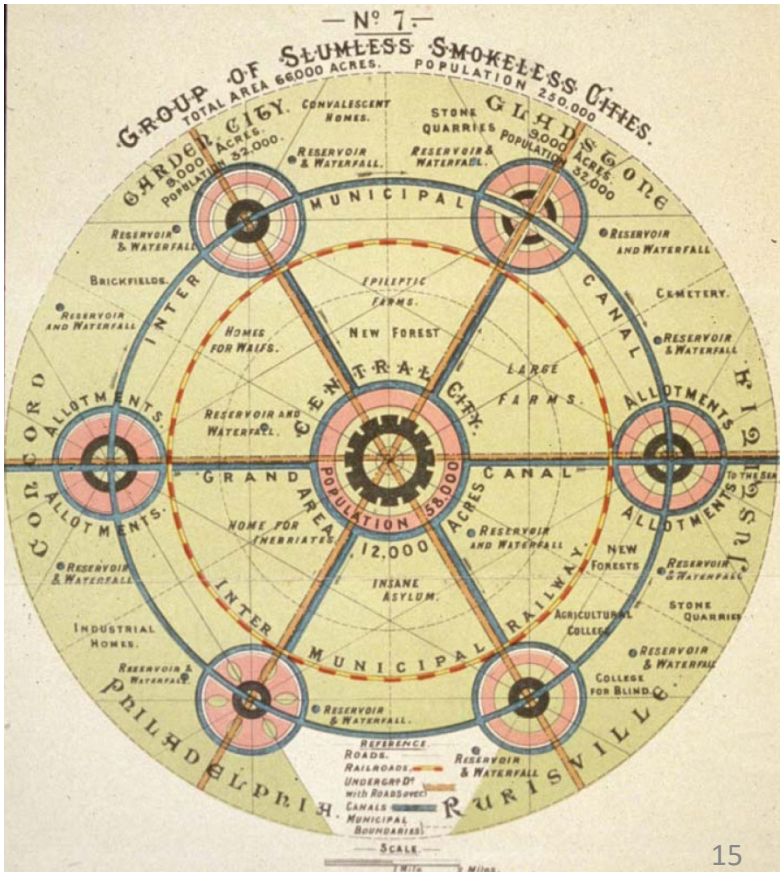
  – Individual in the centre

  Read for instance (Galster, 2001, On the nature of Neighbourhood, DOI: 10.1080%2F00420980120087072)

  – Or Garden city

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

k-neighbours	Meaning
12	Stairs in building
25	Building
50	Sharing laundry room, bicycle basement, garbage recycling bins, etc.
100	Block
200	Buss stop
400	Kiosk, familiar with topology, recognize all neighbours
800	Football field
1 600	Small shop
3 200	Day care, school
6 400	Local square, different retail stores, dentist…
12 800	Upper secondary school, Big stores, communities (sports, religion)
25 600	Hospital, place-belonging, municipality

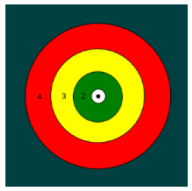


Urban/rural

Let's think about the *grammar* of geography

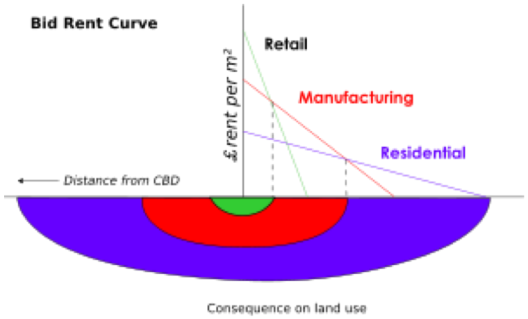
Economic sorting of landscapes

- Urban /rural
  - Von Thünen (OLD)



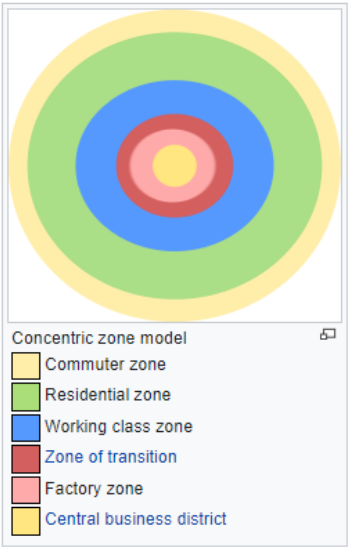
1 (white) dairy and market gardening; 2 (light green) forest for fuel; 3 (yellow) grains and field crops; 4 (red) ranching; the outer, dark green area represents wilderness where agriculture is not profitable

  - Alonso, Bid rent (NEWER)



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

  - Concentric model
  - Sector model
  - Multi-nuclei model



- Affects house pricing, migration patterns, demographics, etc.

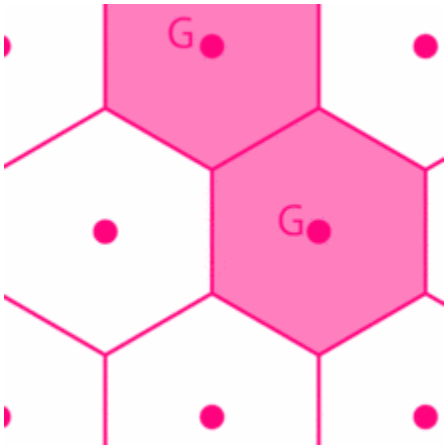
https://en.wikipedia.org/wiki/Concentric_zone_model

Region/nation

Let's think about the *grammar* of geography

- Central Place Theory

  (Christaller, Walter (1933): Die zentralen Orte in Süddeutschland. Gustav Fischer, Jena.)
  - Self-organisation of activities
- The theory relies on two concepts: *threshold* and *range (has since been developed)*
  - Threshold is the minimum market (population or income) needed to bring about the selling of a particular good or service.
  - Range is the maximum distance consumers are prepared to travel to acquire goods - at some point the cost or inconvenience will outweigh the need for the good.
    - Thought experiment, how far are you willing to go for an ice-cream?, and for a medical treatment?
- Theory has been used to design distribution of Swedish municipalities



Why should we scale up the geographical studies?
And what do we mean by scale?

- Geographical scale & computational scale
  - statistical significance
    - Place (not space) makes it difficult example:
      - Transport study – (usually solved through survey)
        - » Age, SEI, gender, availability to transport = a few categories
        - » If you include OD (origin-Destination) for Swedish municipalities you have 290^2 unique links
        - » Loss of DFs
  - If we use vast amounts of observations from longer time periods and if we enrich the observations with contextual statistics – then:
    - Place becomes less of a problem and we can start testing geographical hypotheses "for real"
- Ethics – a warning
  - Again place (not space) makes it all difficult:
    - Discrete statistics (census data or similar – annual registers in Sweden) makes it easy to identify individuals
    - Spatial distribution makes it even easier
    - Observation of trajectories makes it even more easy.
- And an opportunity
  - By designing models that observes trajectories for shorter time periods, or recode contents of keys – the data may be used
  - Universities have greater freedom than do companies (GDPR)
  - i.e. be smart in selecting and coding data, or be sorry…

Geographical theories in which scalable spatial models can be employed – a few examples

- Departing from the same spatial scales as listed in the grammar of geography section
  - Time geography
    - Individual level -> neighbourhood level.
  - Spatial mismatch
    - neighbourhood level -> urban level
  - Functional regions
    - Urban level -> regional/national levels

Geographical theory and Big Data Time-Geography

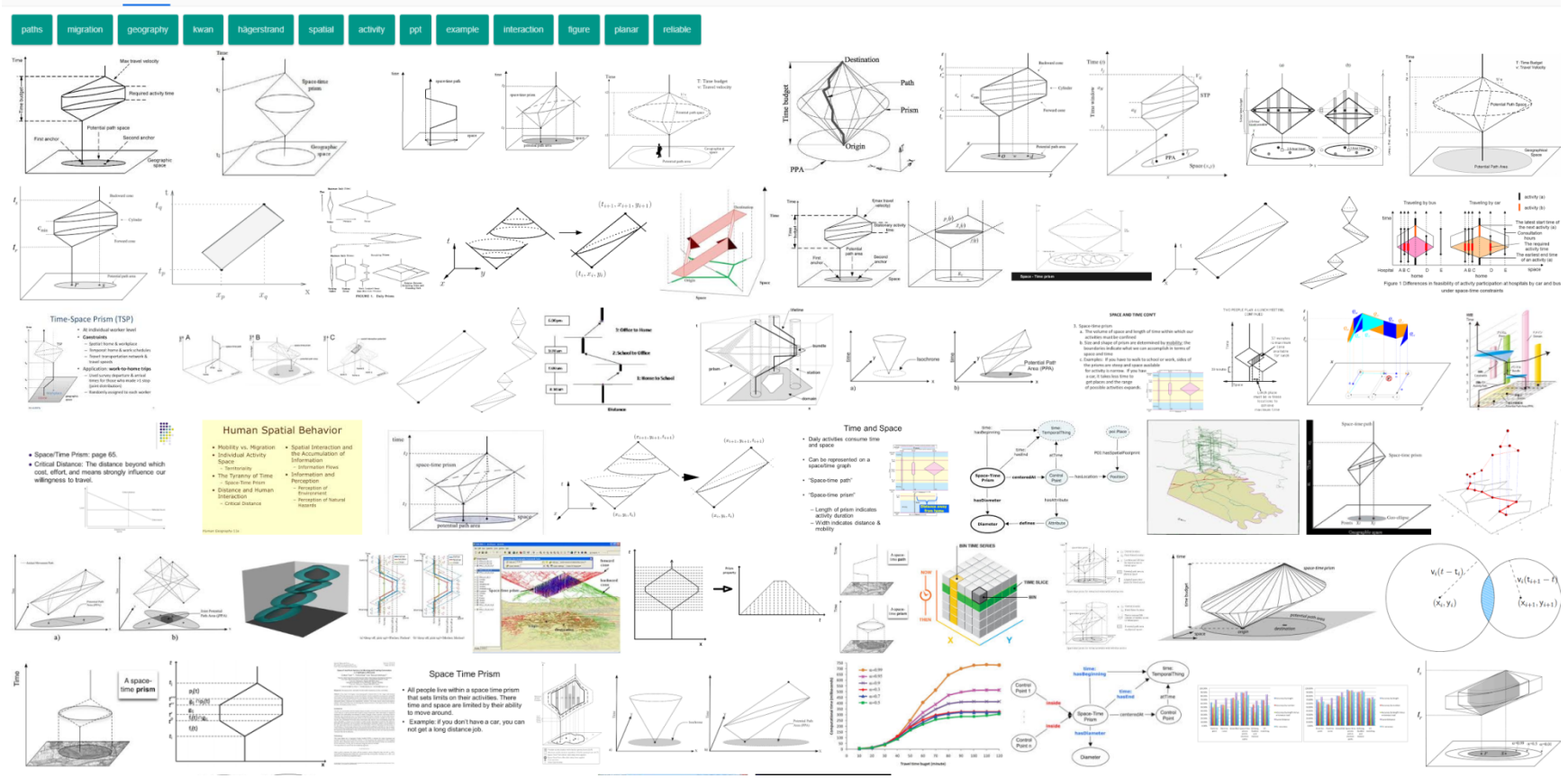
With inspiration from Wikipedia (https://en.wikipedia.org/wiki/Time_geography):

- Hägerstrand's earliest formulation of time geography informally described its key ontological features:
  - "life paths become captured within a net of constraints, some of which are imposed by physiological and physical necessities and some imposed by private and common decisions“
- He described three important classes of constraints:
  - *capability constraints* — limitations on the activity of individuals because of their biological structure and/or the tools they can command,
  - *coupling constraints* — limitations that "define where, when, and for how long, the individual has to join other individuals, tools, and materials in order to produce, consume, and transact“
  - *authority constraints* — limitations on the domain or "time-space entity within which things and events are under the control of a given individual or a given group“
- Hägerstrand illustrated these concepts with novel forms of graphical notation such as the space-time prism
- While this innovative visual language is an essential feature of time geography, Hägerstrand's colleague Bo Lenntorp emphasized that it is the product of an underlying ontology, and "not the other way around.

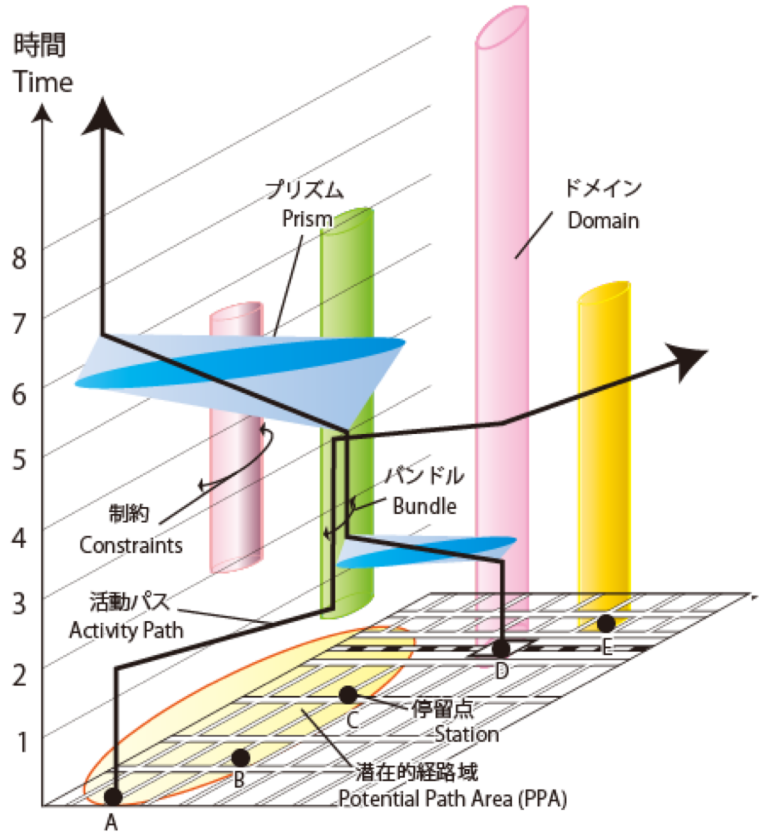
REFERENCES

- Hägerstrand, T. (1970, December). What about people in regional science?. In *Papers of the Regional Science Association* (Vol. 24, No. 1, pp. 6-21)
- *Kwan, Mei-Po (2013). "Beyond space (as we knew it): toward temporally integrated geographies of segregation, health, and accessibility“. Annals of the Association of American Geographers. **103** (5): 1078–1086. doi:10.1080/00045608.2013.792177.*
- Boman, Magnus; Holm, Einar (2004). "Multi-agent systems, time geography, and microsimulations". In Olsson, Mats-Olov; Sjöstedt, Gunnar. *Systems approaches and their application: examples from Sweden*. Dordrecht; Boston: Kluwer Academic Publishers. pp. 95–118. doi:10.1007/1-4020-2370-7_4

Space-time prism



Space-time prism



- Prism illustrates the potential area of interaction with others.
- Result of constraints
  - Social, biological, etc.
- Affects our possibilities for integration, interaction and accessibility to opportunities

https://upload.wikimedia.org/wikipedia/commons/e/e4/Sample_of_time_geographical_description.png

Space-time prism vs. activity spaces

SPACE-TIME

- Geographical grammar
- Depicts potential for interaction

ACTIVITY SPACE

- Observed geography
- Depicts the observed trajectory of individuals as functions
  - Standard deviation ellipse
  - Minimum convex polygon
  - Path-buffer
  - Etc.

Space-time offers a model for understanding variation and activity space depicts activities

Questions

Data

- Time-coded GPS trajectories or equivalent (GSM, Public Transport data, etc.) from a vast number of users (millions on a daily basis)
- GIS-layers depicting land-use, networks, objects (schools, supermarkets, restaurants, etc.)
- Contextual discrete data describing populations

Questions

- Do people from different walks of life meet?
  - When, where,?
  - How do we express this statistically?
- What kinds of systematic variation in activity spaces can be observed?
  can these be connected to time-geography constraints?
- How do different categories of individuals use space?
  - Can we talk about segregation?

Spatial mismatch hypothesis

the neighbourhood level -> urban level

- John F. Kain (1968), the concept:
  - conditions
    - Poor residential areas in down-town Chicago experience job-loss.
    - Job growth is observed in remote suburbs
    - To meet the demand, long commutes are necessary.
  - Development
    - The negative spiral begins with a reduction in accessibility to jobs -> reduces commuting -> observed commuting decreases -> service is decreased since fewer travel-> reduce commuting…
    - The development leads to increasing poverty, and the closing of local services and jobs -> leads to isolation -> lower competitive force – los of social capital

Spatial mismatch hypothesis

the neighbourhood level -> urban level

Symptoms

- Reduction in out-commuting
  - From few to fewer
- Marginal in-commuting
- Little variation in destinations as places
- Commuters may have several lags of travels
  - -> trip to station -> trip to destination
- Over-time reduction in local jobs and service
- Limited in-migration and limited out-migration (levels are subject to variations in immigration levels, etc.)

Questions

- Can we detect neighbourhood mobility patterns that are pointing towards spatial mismatch?
- Can we detect an order of events?
  - For instance: do local job-loss start a negative spiral?

Functional regions

urban level -> regional/national level

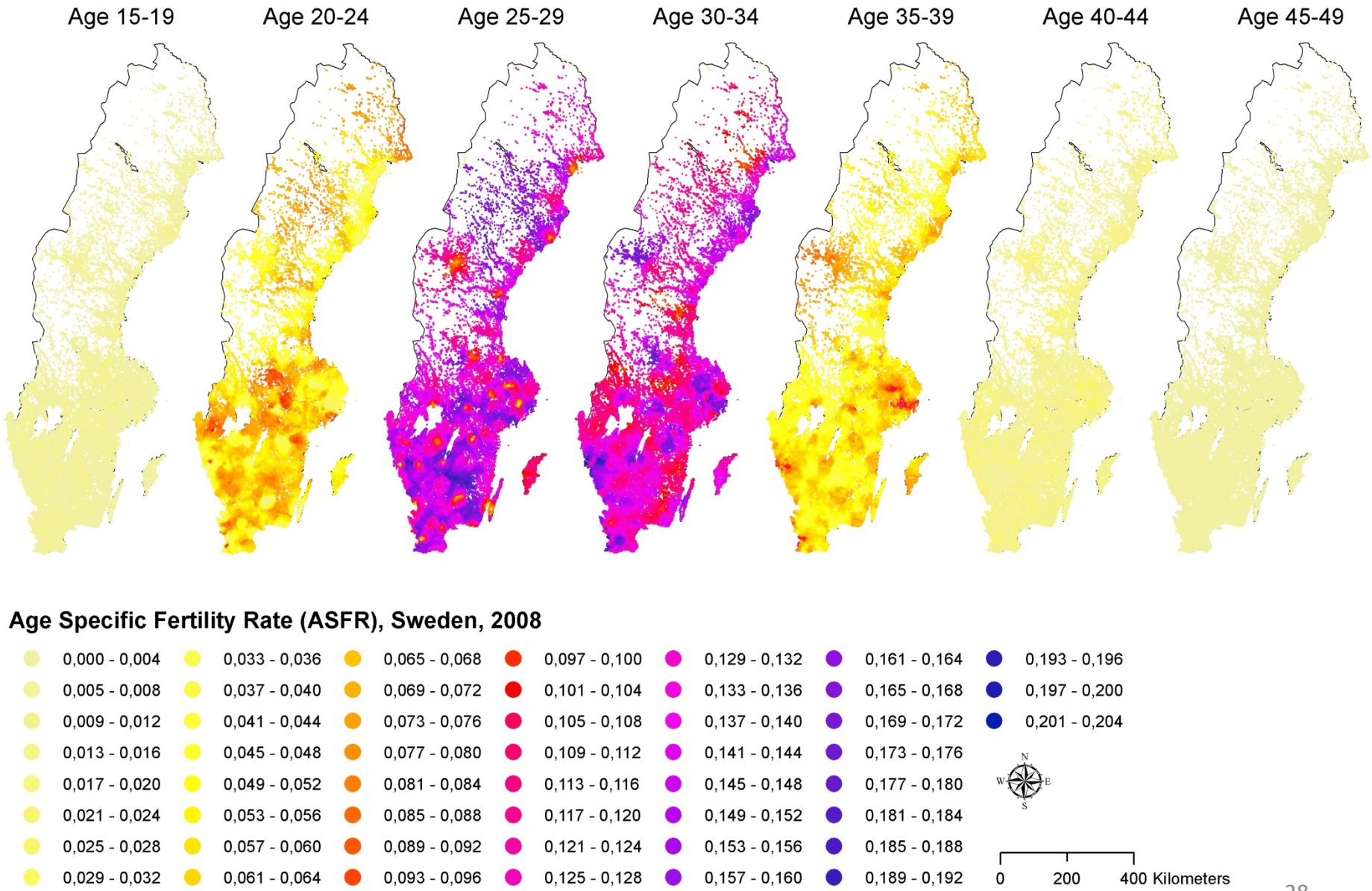
Functional regions

- Functional regions should be constructed from observed interaction
  - In SE, usually based on data from OD-matrixes derived from taxation registers (Reside @ O and work @ D)
  - Municipalities are clustered to LA or FA
    - https://www.scb.se/hitta-statistik/statistik-efter-amne/arbetsmarknad/sysselsattning-forvarvsarbete-och-arbetstider/registerbaserad-arbetsmarknadsstatistik-rams/produktrelaterat/Fordjupad-information/lokala-arbetsmarknader-la/

Questions

- Can we depict regions using the mass-movement of things during the day?
  - That are more appropriate in spatial-unit composition than administrative borders
- Are there time, and seasonal differences?
- Can we associate destinations with activities, and consequently refine our knowledge about how CPT (Christaller) works?

I stop there – with maps that can be used to exemplify our fascination for our field

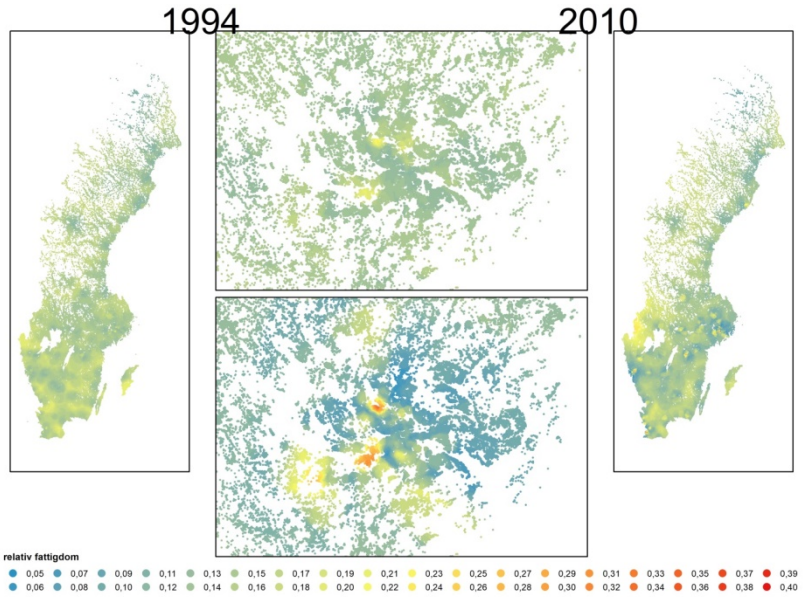


http://equipop.kultgeog.uu.se/

Östh J., (2015), Superlocal spatial variations in fertility, Environment and planning A 47 (5)

I stop there – with maps that can be used to exemplify our fascination for our field

Relative poverty – multiscales EquiPop



Spatial Isolation – multiscales EquiPop

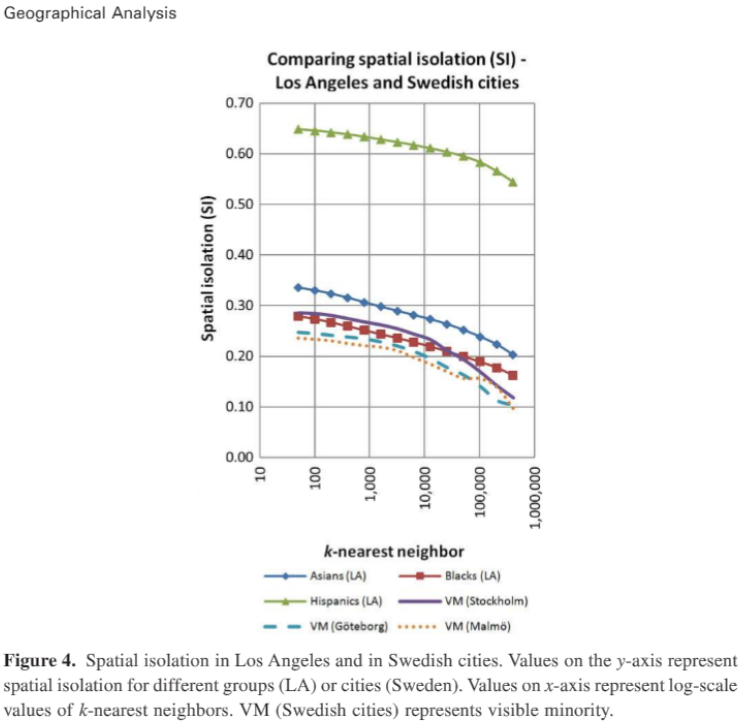
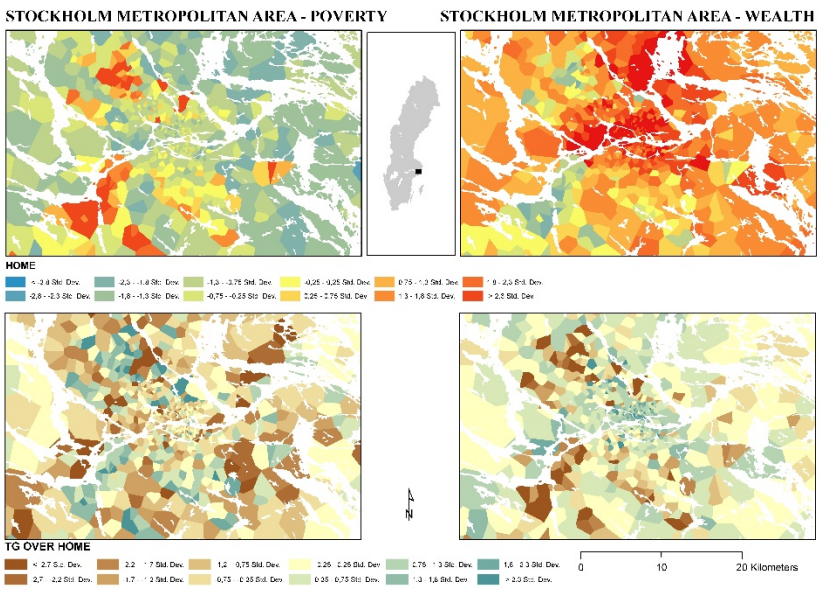


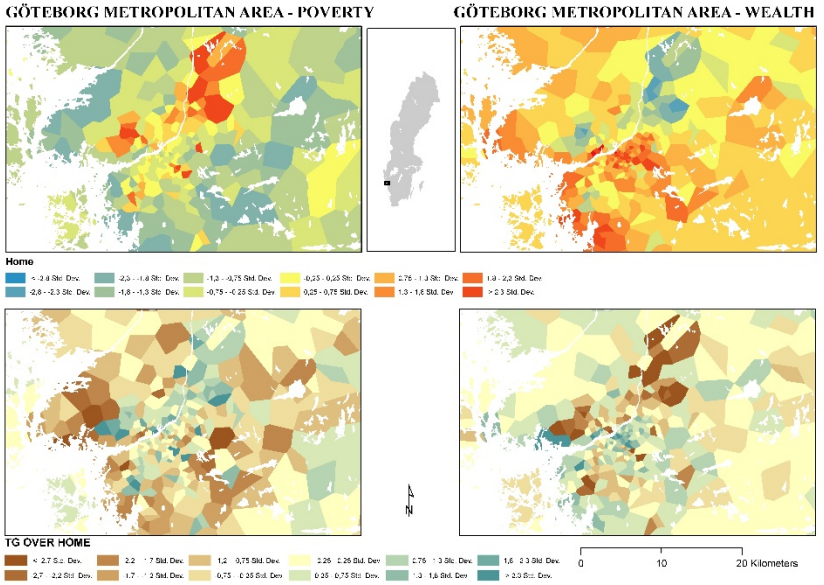
Figure 4. Spatial isolation in Los Angeles and in Swedish cities. Values on the *y*-axis represent spatial isolation for different groups (LA) or cities (Sweden). Values on *x*-axis represent log-scale values of *k*-nearest neighbors. VM (Swedish cities) represents visible minority.

http://equipop.kultgeog.uu.se/

Östh John, Clark A. W. & Malmberg Bo, (2015), Measuring the scale of segregation using k-nearest neighbor aggregates, Geographical Analysis, DOI: 10.1111/gean.12053

I stop there – with maps that can be used to exemplify our fascination for our field





Top row show maps over the distribution of residential (Home) poverty (left map) and wealth (right map) where strong concentrations are depicted red and the opposite blue. Lower row shows TG over Home calculations for poverty (left) and wealth (right) where blue values indicates concentration effects due to mobility and brown indicates a reduction due to mobility.

Östh, J., Shuttleworth, I., & Niedomysl, T., (2018),Spatial and temporal patterns of economic segregation in Sweden's metropolitan areas: a mobility approach, E&PA: Economy and Space, DOI: https://doi.org/10.1177/0308518X18763167