PENNSTATE POPULATION RESEARCH INSTITUTE
SOCIAL SCIENCE RESEARCH INSTITUTE

Contexts as Grids:

A discussion of some conceptual and methodological challenges and the opportunities presented by using gridded data in demographic and health research

Stephen A. Matthews
Liberal Arts Professor of Sociology, Anthropology and Demography
(courtesy Geography and School of International Affairs)
Director, Graduate Program in Demography
Interim Director, Online Programs in Applied Demography
Penn State, USA
E-mail: sxm27@psu.edu

May 22, 2018

Outline

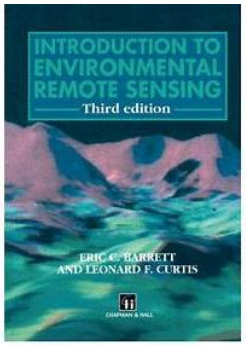
1: Grids in a Research Context (Personal overview)

2: Grids in My Research

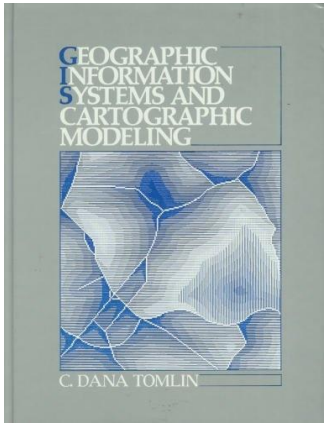
Across qualitative and mostly quantitative studies

3: Questions & Comments

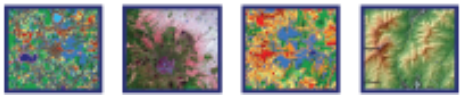
Grids in a Research Context (Personal overview)



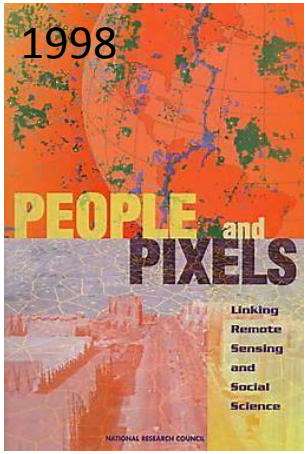
The pre-GIS 'spatial' data world was mostly raster – Remote Sensing - gridded data very popular in Environmental Sciences. **Dana Tomlin** (1983) **"map algebra"** the process of map combination for co-registered layers with rasters of identical size and resolution.



CLARK LABS



Raster software – IDRISI – Clark Labs (Ron Eastman) https://clarklabs.org



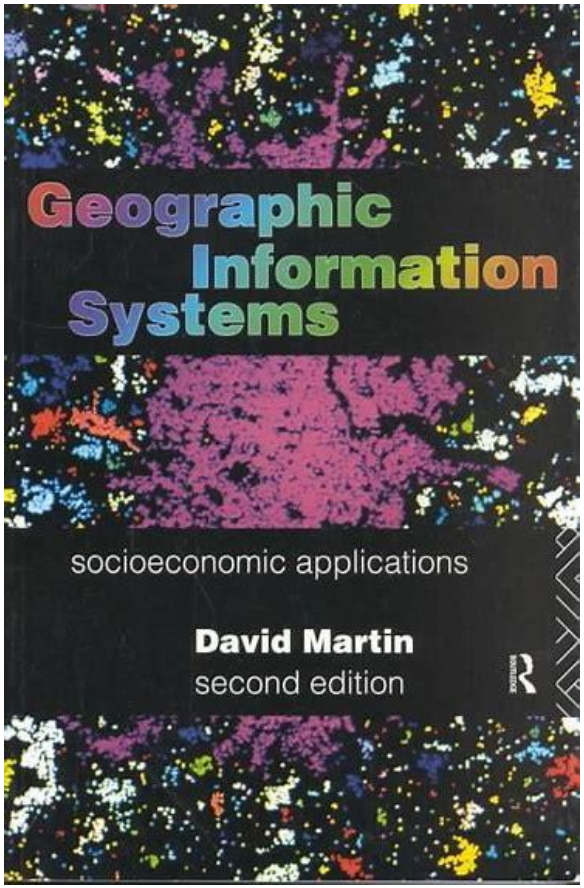
People & Pixels: Linking Remote Sensing & Social Science

("socializing the pixel") Linking environmental data with gridded human population data across all geographic scales.

EPERN

https://www.populationenvironmentresearch.org/cyberseminars

Grids in a Research Context (Personal overview)



1989, 1996

In socioeconomic applications of GIS the world was mostly **vector**.

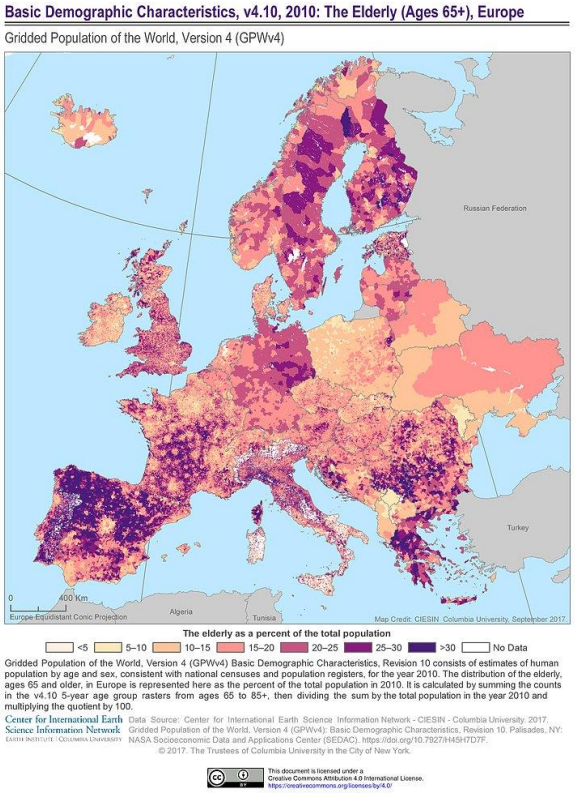
An early proponent of **raster data** structures in human geography and population studies was and is David Martin (University of Southampton)

David's core research interests concern the modelling of population in geographical information systems, particularly with regard to census and health applications. He has long standing interests in gridded population mapping, automated zone design, health care accessibility and, more recently, time-space population modelling.

https://www.southampton.ac.uk/geography/about/staff/djm1.page

Grids in a Research Context (Personal overview)

Gridded World Population

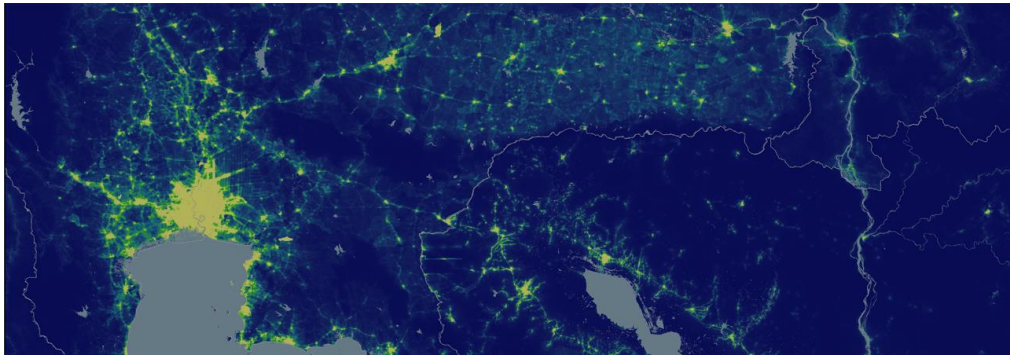


http://sedac.ciesin.columbia.edu/data/collection/gpw-v4

WorldPop

http://www.worldpop.org.uk

afripop asiapop ameripop

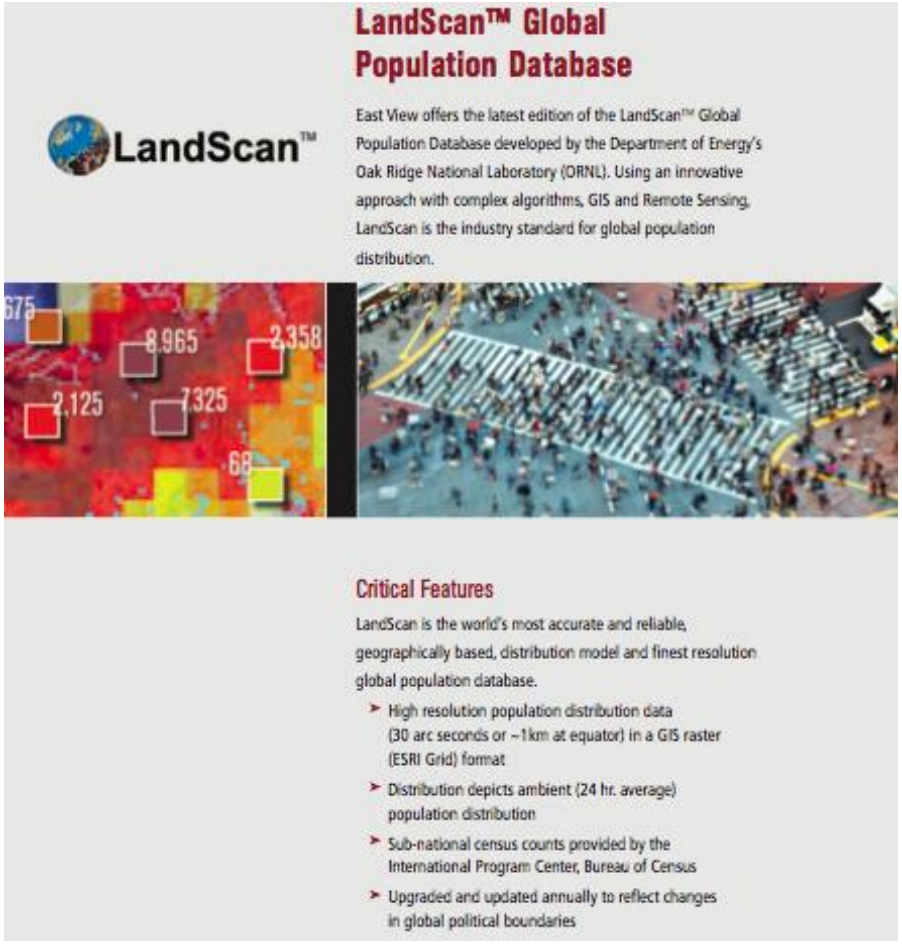


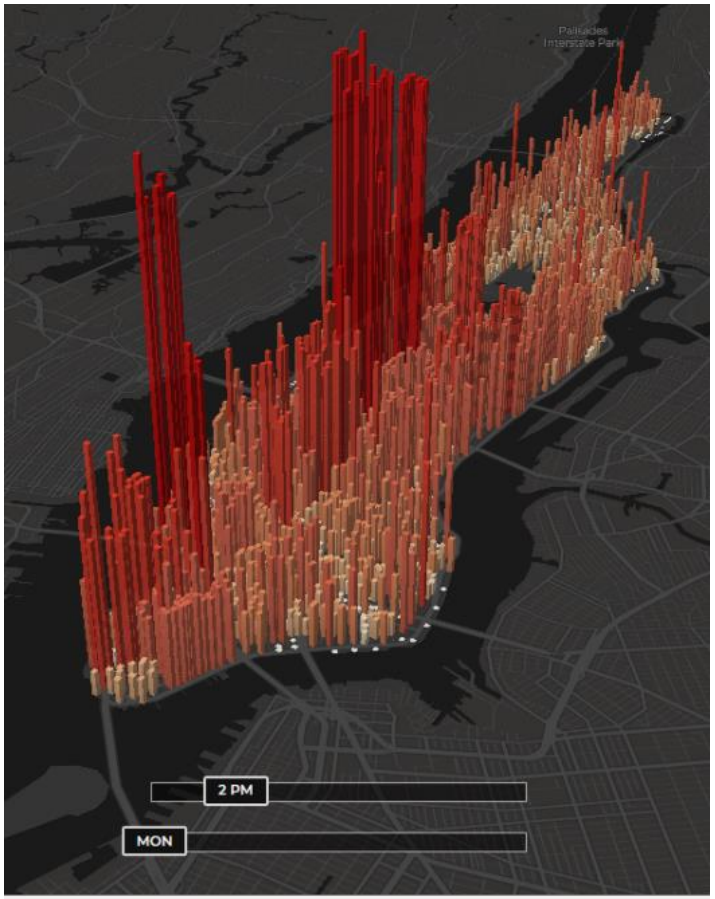
High spatial resolution, contemporary data on human population distributions are a prerequisite for the accurate measurement of the impacts of population growth, for monitoring changes and for planning interventions.

Grids in a Research Context (Personal overview)

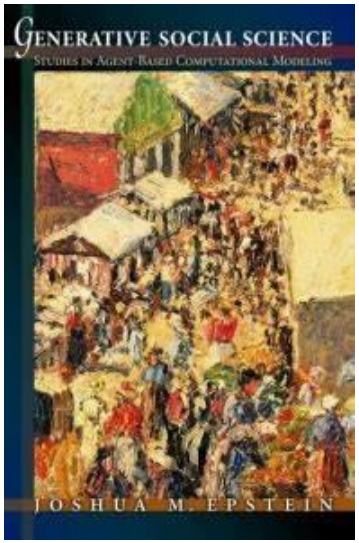
Landscan
https://landscan.ornl.gov

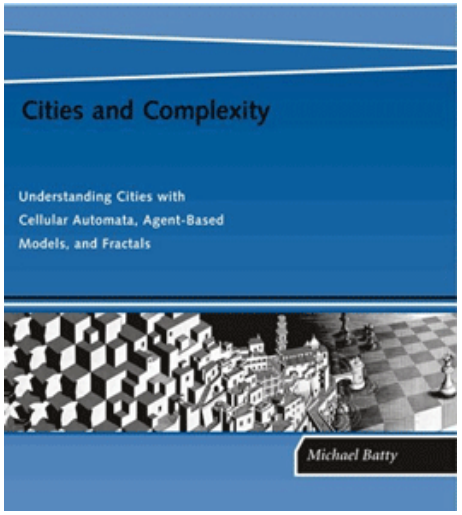
https://manpopex.us/

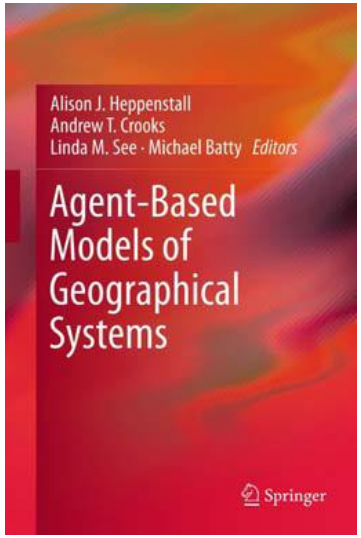


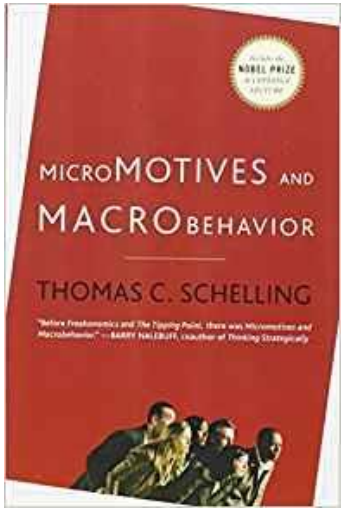


Grids in a Research Context (Personal overview)









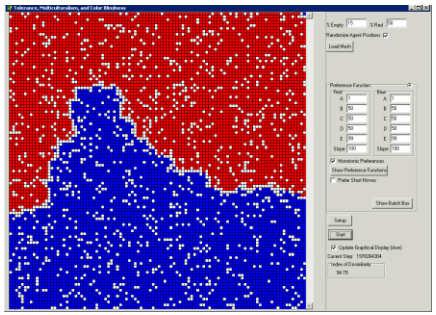
Demographers have begun to explore the potential for a simulation method called **agent-based modeling** (explicitly model the macro-level consequences of individual-level attributes and behaviors based on a set of assumptions about intervening mechanisms, and then observe the fit to empirically observed patterns).

https://www.ucl.ac.uk/bartlett/casa/
Their focus is multidisciplinary, building simulation models, visualising 'big data', developing bespoke applications and creating new methods for the collection, analysis and communication of data.

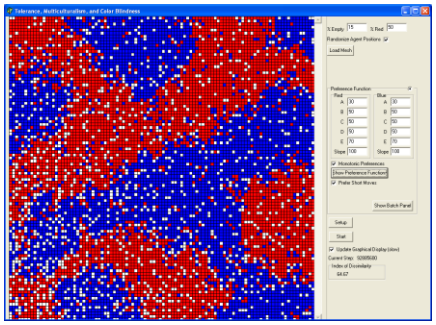
Grids in a Research Context (Personal overview)

Group preference

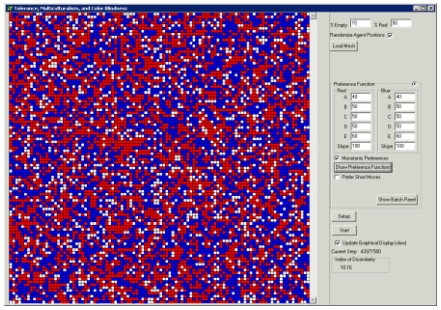
Strong



Moderate



Weak



Thomas Schelling

Micro Motives & Macro Behavior (1978)

The whole is more than the sum of its parts. Individuals influence one another in response to the influence they receive.

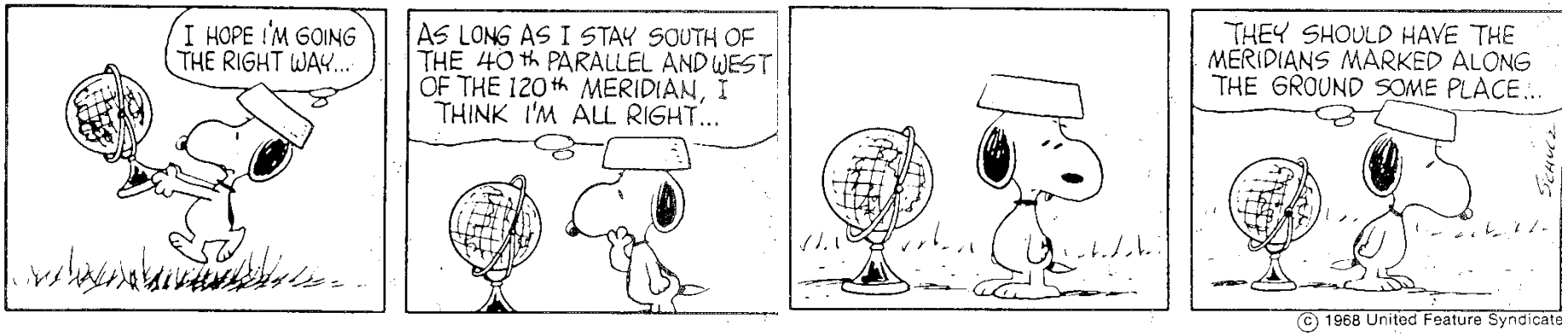
Agent-based modeling is a tool for exploring how individual decisions can aggregate with surprising consequences for population behavior.

Macro-level segregation need not reflect micro-level ethnic preferences.

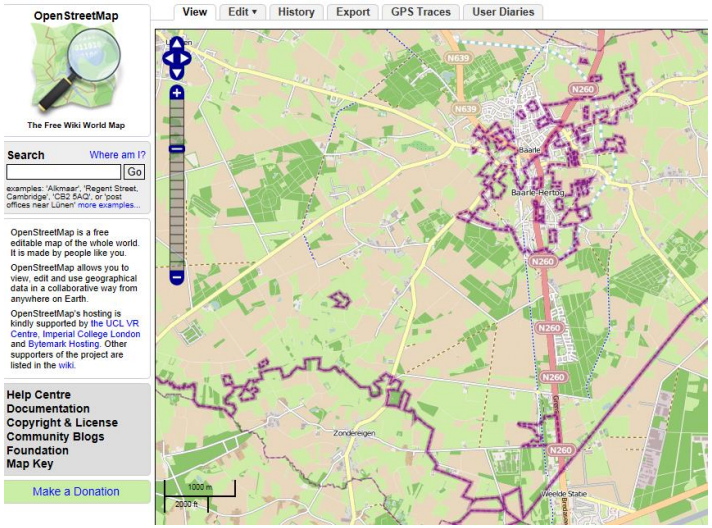
http://ccl.northwestern.edu/netlogo/models/Segregation

What do we know?

Objectively defined neighborhood boundaries are largely invisible (to everyone)
Subjective neighborhood boundaries are largely invisible (to the researcher).



One of the few visible boundaries



Grids in My Research

Spatial Scales

Global (macro)

National

Regional

Urban

Local (micro)

Accessibility
Supermarkets (Food Outlets)

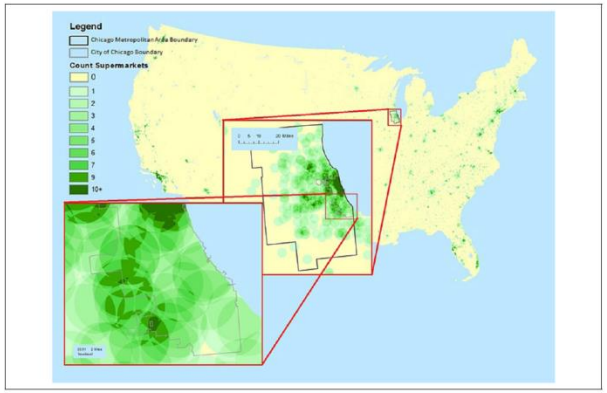


Figure 2. Accessibility of supermarkets within 3 miles for the continental United States.

Accessibility
Services (Tehran)

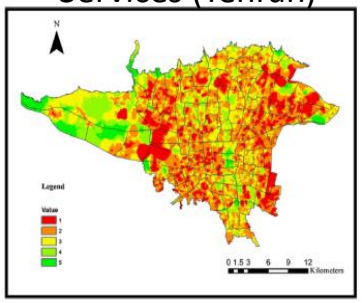
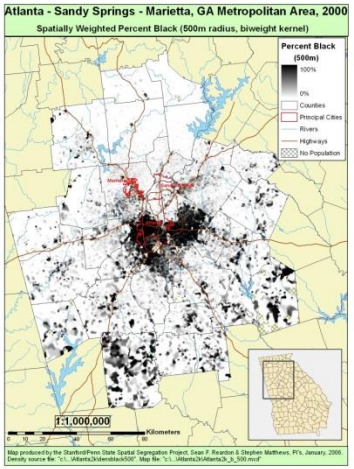
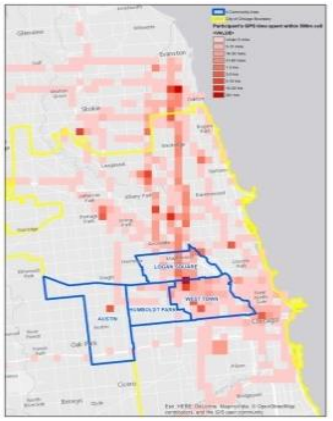


Fig. 9 Reclassification of accessibility

Measuring
Spatial Segregation



Measuring
Activity Space

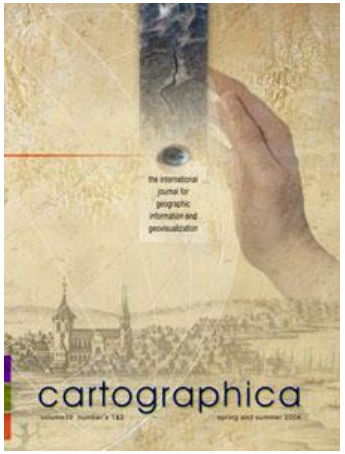


Past (other temporal epochs; short-long or 1 second-lifetime) Future

Temporal Scales (Intervals)

Qualitative Work – Geoethnography

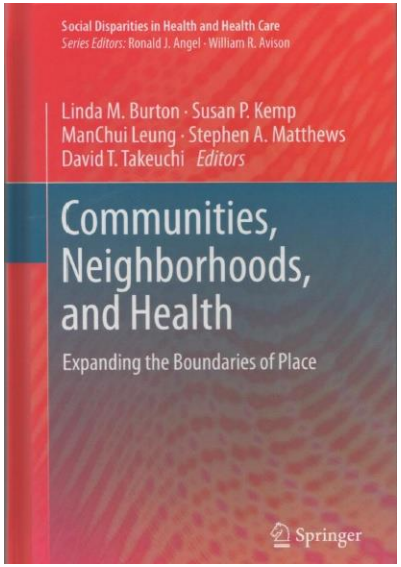
Research Team:
Linda Burton (Duke) is PI on all ethnographic projects referred to here, key investigators include Debra Skinner (UNC), and a **massive team of ethnographers** and coders.



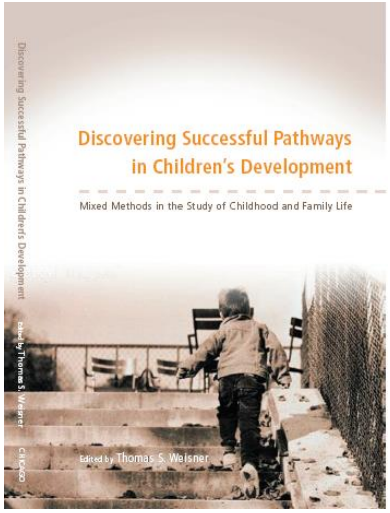
Stephen Matthews, James Detwiler, and Linda Burton (2005) Geoethnography: Coupling Geiographic Information Analysis Techniques with Ethnographic Methods in Urban Research

Cartographica

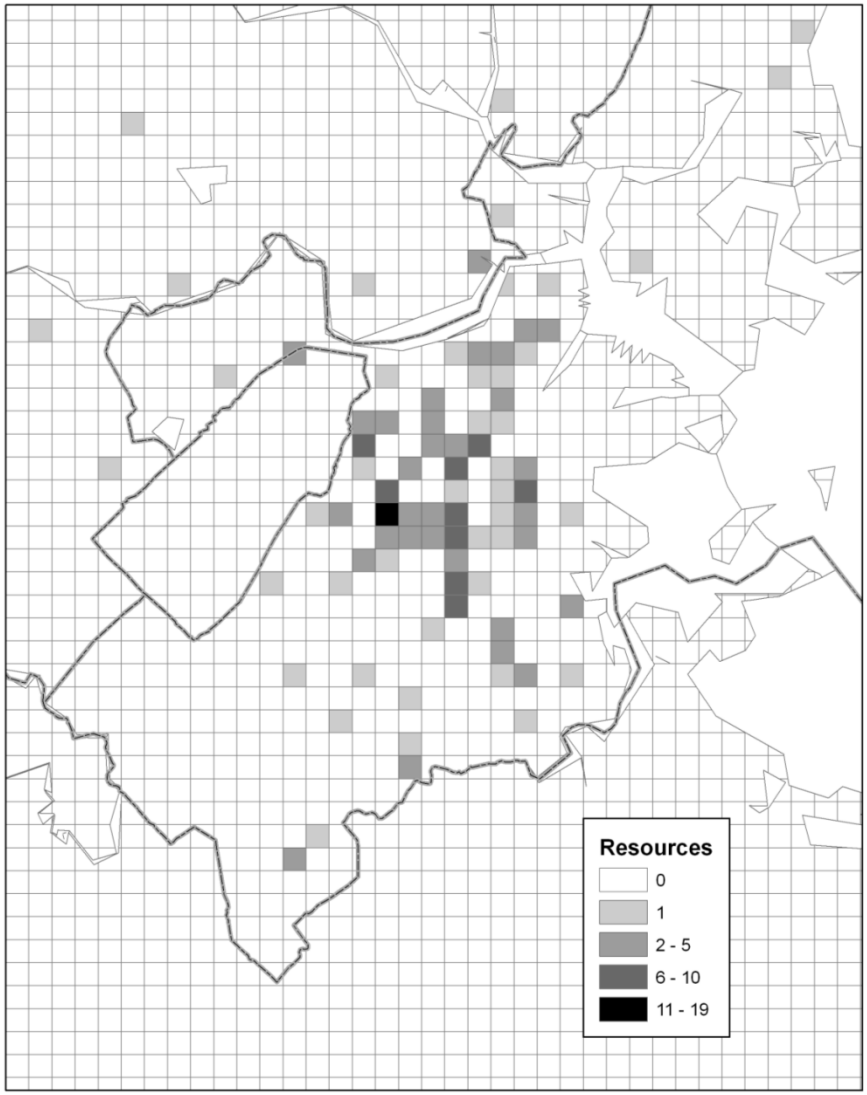
Stephen Matthews (2011) **Spatial Polygamy and the Heterogeneity of Place: Studying People and Places via Egocentric Methods** (Chapter 3)



Debra Skinner, Stephen Matthews and Linda Burton (2005) Using ethnography and GIS to examine constructions of developmental opportunities in contexts of poverty and disability. Pp. 223-239 (Chapter 7) in Thomas S. Weisner [Editor]



Basic Grids – Hiding Individuals



Families and shared definitions of neighborhoods. Where inconsistencies within and between residents of the same neighborhood undermine the 'salience' of neighborhood and the identification of mechanisms on how place gets in to people.

A gridded surface of total family activities (n = 222) based on ten families residing in one neighborhood. Each grid cell is 500 x 500 meters

Source:
Stephen A. Matthews (2011)
"Spatial polygamy and the heterogeneity of place: studying people and place via egocentric methods"

GPS study: Activity Space and Contextual Measures of Environmental Exposures (ASCMEE) Study

PIs: Shannon Zenk (UIC) & Stephen Matthews (Penn State)

National Cancer Institute (R21CA195543)

Purpose:

Activity Space (AS) research offers a promising approach to understanding environmental influences on health behaviors and outcomes.

An objective of the ASCMEE measurement grant is to develop tools to examine the **joint** spatial and temporal stability of AS.

(THEME: ***HETEROGENEITY*** IN USE OF SPACE ACROSS TIME)

Assumptions about Environmental Exposures

The role of **Time** in "Spatial" thinking (Exposure = S x T)

Static conceptualizations and mis-estimation

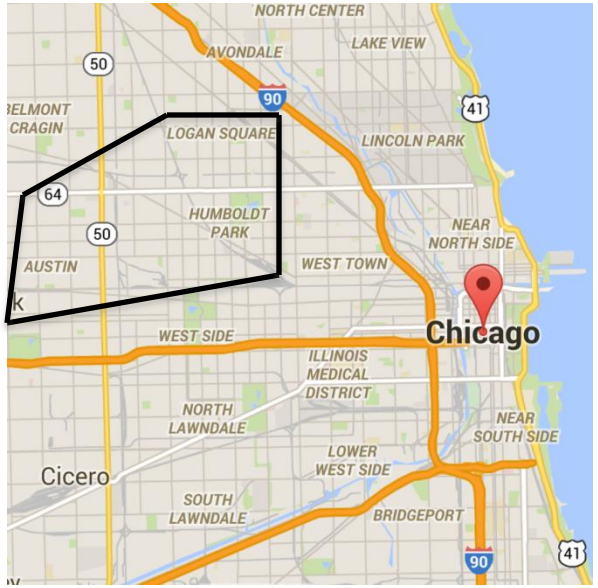
- residential neighborhood
- residential buffers
- LAFANS AS

GPS Studies and Time

- technology driving measurement window

But … how long do we need to collect data to have a reasonable understanding of human spatial behavior?

ASCMEE Study



The design facilitates a comparison of activity spaces and exposure to environmental attributes by race/ethnicity and SES.

Recruit a non-probability quota sample of adults ages 20-64 at T1 in contiguous Chicago Community Areas (CA): **Austin, Humboldt Park, Logan Square**. Overall, residents of the CAs are 39% AA, 33% H (50%+ MA), and 28% W. These community areas were selected to facilitate recruitment of a socioeconomically diverse sample of AA, MA, and W (CA poverty rates in 2010 vary between 17-32 percent) living in the same general area, roughly equidistant from the central business district.

Original Design

Table 3. Data collection elements and phases

Element	Time 1, n=144				Time 2 (6 months later),[a] n=36				
	W1	W2	W3	W4-5	W1	W2	W3	W4	W5
In-person visit	0[b]		15[c]	31	0		15[c]		36
GPS & accelerometer									
Activity questionnaire									
Dietary recalls									

[a] T2 data collection will be adjusted based on optimal number and combination of days found in Aim 1 [b] Numbers refer to day; [c] Visit will occur at approximately Day 13-17

GPS: Points … to Grids

- GPS data quality and handling GPS data

Some issues

slicing data by time (day, hour)

slicing data by space/grids (grid size)

python script to measuring both new and accumulated places (grids).

GPS: Points … to Grids via python

Simple python script (written by Jessie Li, IST at PSU)

Reads in GPS data (coordinates, date, time), user specifies *time unit* (hour or day) and *grid size* (e.g., 240m, 500m).

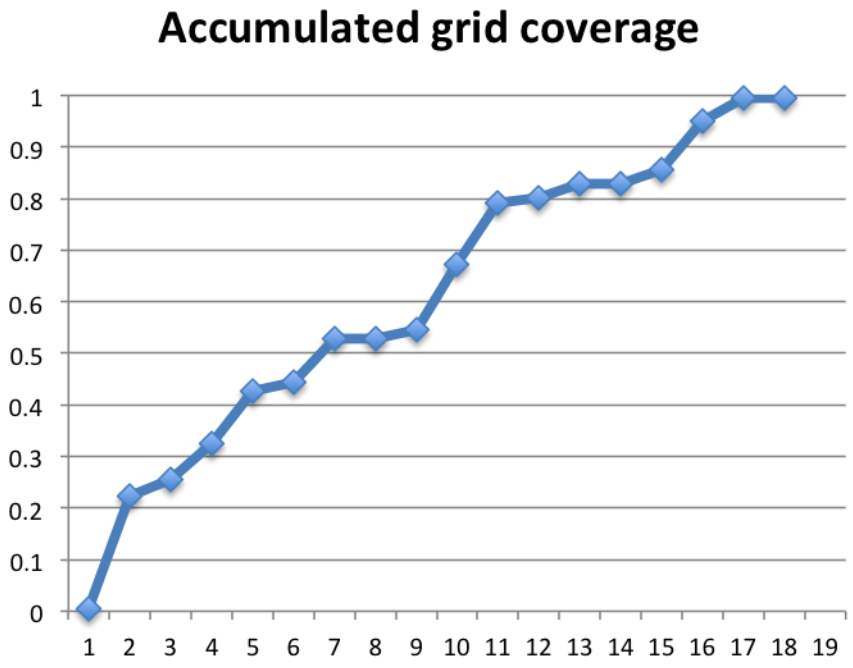
Calculates the number of output rows (time periods- e.g., days) and generates output that includes

- counts of points
- grids, new girds by time period
- number of new points within new grids
- proportion of points in new grids
- accumulation of grids/points based on the N time units

Accumulated Places (Grids)

An Sample Output

weekday	# of points Today	# new grids	# of points in new grids	Portion of points in new grids	Accumulated grid coverage
	1	1	1	1.000	0.00
Fri	754	41	671	0.889	0.22
Sat	1321	6	40	0.030	0.25
Sun	2787	13	72	0.025	0.32
Mon	1944	19	53	0.027	0.42
Tues	1912	3	3	0.001	0.44
Wed	2310	16	86	0.037	0.52
Thurs	1908	0	0	0.000	0.52
Fri	1850	3	19	0.010	0.54
Sat	2680	24	277	0.103	0.67
Sun	2708	22	57	0.021	0.79
Mon	1608	2	2	0.001	0.80
Tues	2789	5	29	0.010	0.82
Wed	1868	0	0	0.000	0.82
Thurs	1832	5	23	0.012	0.85
Fri	1656	18	46	0.027	0.95
Sat	2843	8	572	0.201	0.99
Sun	2256	0	0	0.000	0.99

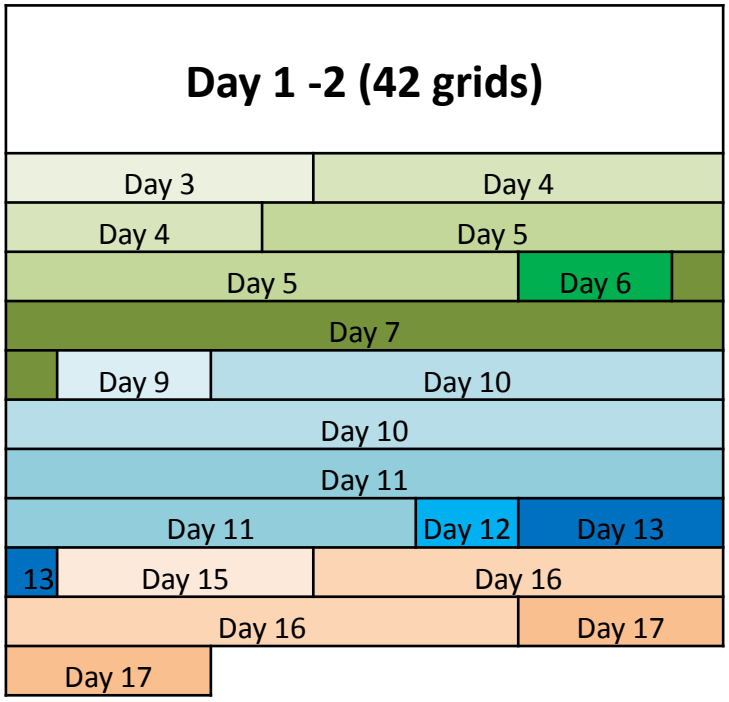


30 second epoch

A Conceptual Map of New and Accumulated Places (Grids)

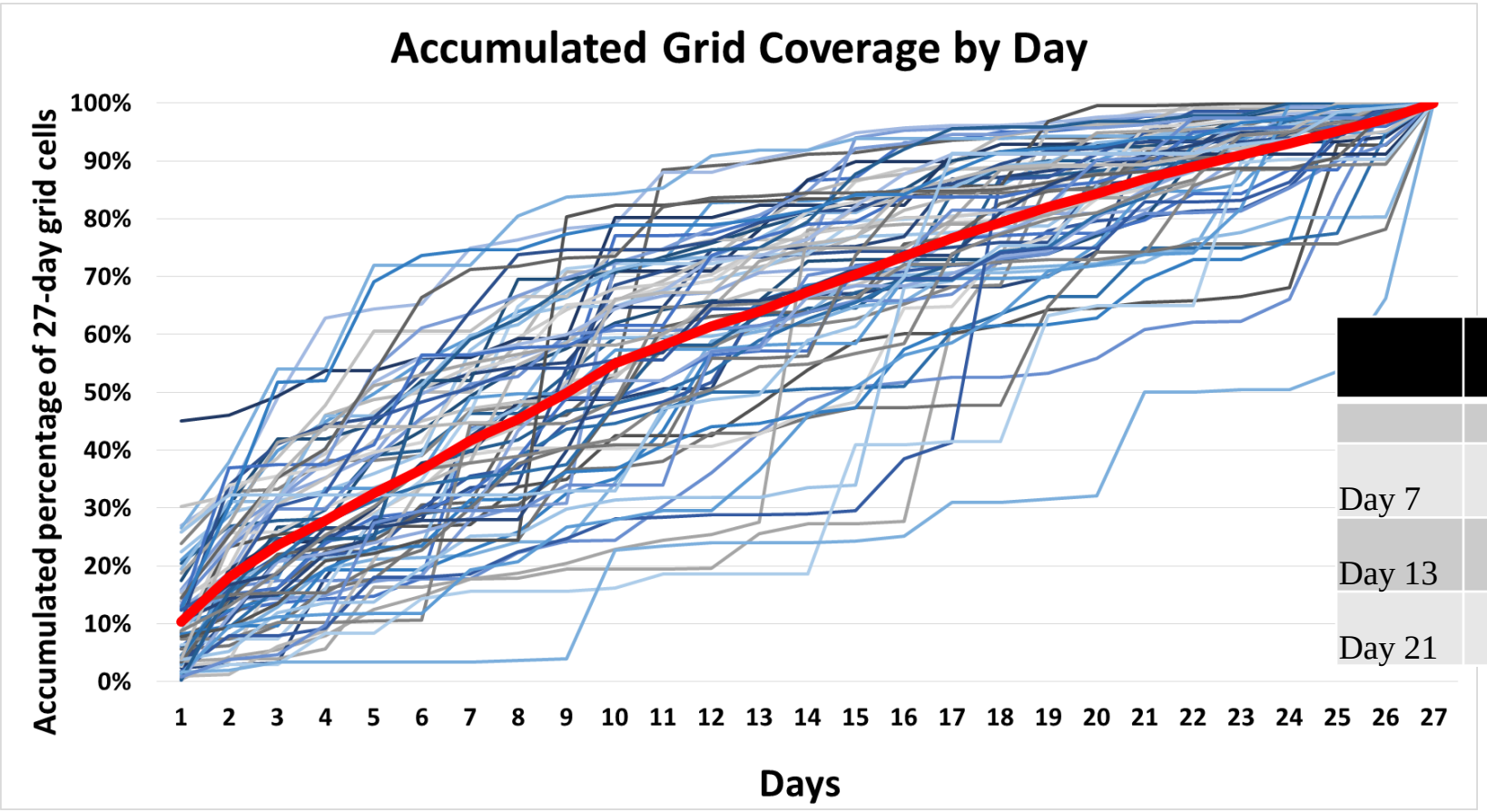
An Sample Output

weekday	# of points Today	# new grids	# of points in new grids	Portion of points in new grids	Accumulated grid coverage
	1	1	1	1.000	0.00
Fri	754	41	671	0.889	0.22
Sat	1321	6	40	0.030	0.25
Sun	2787	13	72	0.025	0.32
Mon	1944	19	53	0.027	0.42
Tues	1912	3	3	0.001	0.44
Wed	2310	16	86	0.037	0.52
Thurs	1908	0	0	0.000	0.52
Fri	1850	3	19	0.010	0.54
Sat	2680	24	277	0.103	0.67
Sun	2708	22	57	0.021	0.79
Mon	1608	2	2	0.001	0.80
Tues	2789	5	29	0.010	0.82
Wed	1868	0	0	0.000	0.82
Thurs	1832	5	23	0.012	0.85
Fri	1656	18	46	0.027	0.95
Sat	2843	8	572	0.201	0.99
Sun	2256	0	0	0.000	0.99



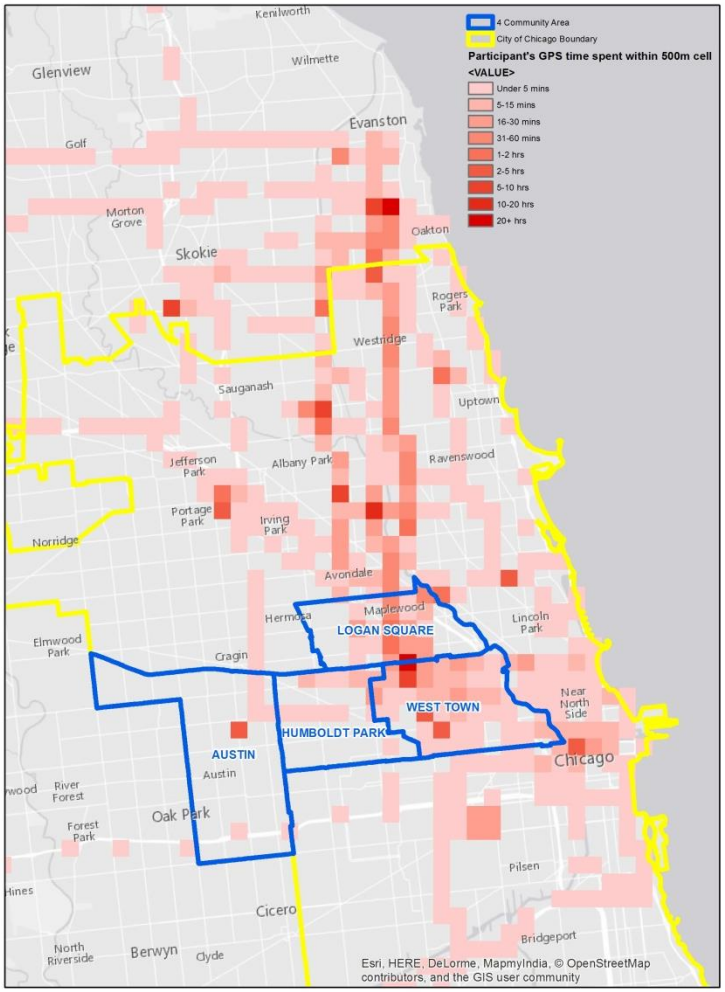
186 grid visited by this respondent

Percent of the total number of grid cells visited (across 27 days) by day and by participant, 240m grid



	Accumulated space	
	Mean	SD
Day 7	41.7%	16.0
Day 13	64.0%	17.1
Day 21	86.8%	9.9

Total Places (Grids)



67 participants
27 days

Based on 240m x 240m

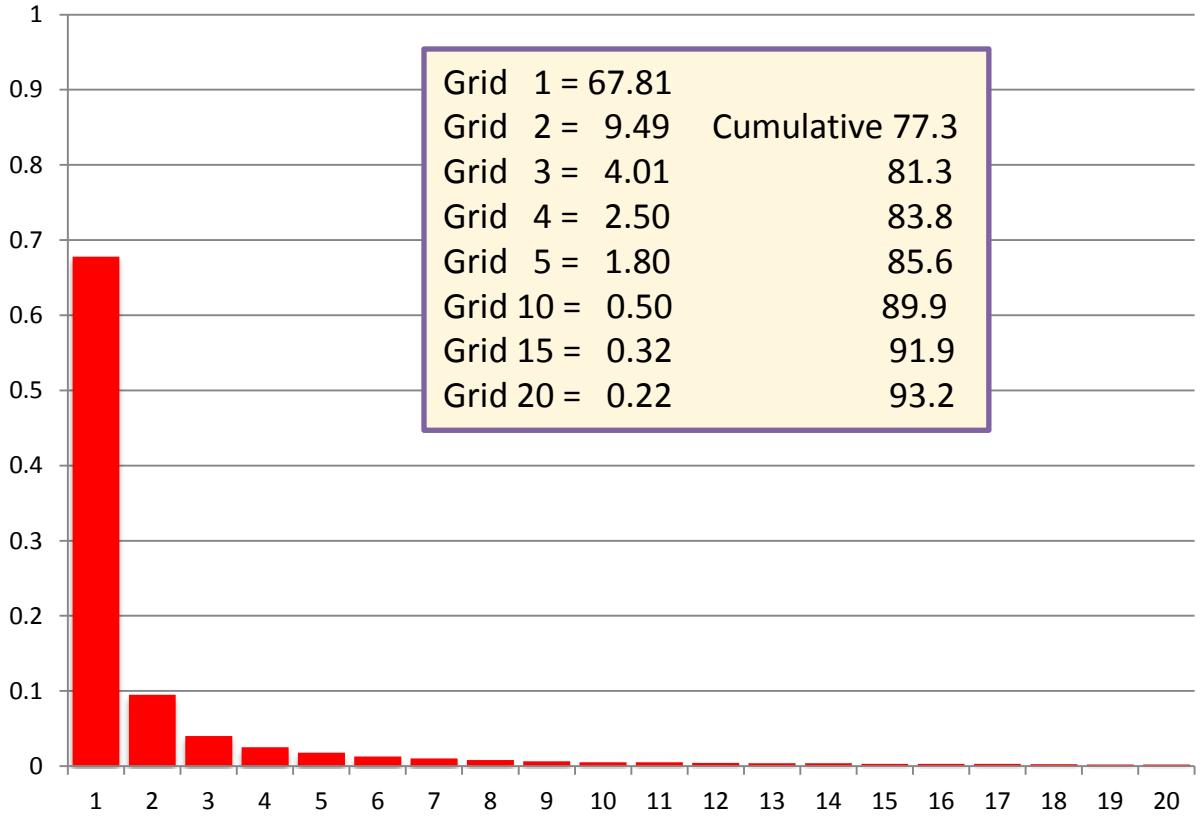
New Grids after Day 1
Range 91-1,873, mean = 571, med = 468,
LQ = 327; UQ = 701

New Points in New Grids after Day 1
Range 943-8,570, mean = 3,301, med = 3,105,
LQ = 2,313; UQ = 4,106

60 points = 1 hour
600 points = 10 hours

Frequently Visited Places (Grids)

Example: Percent of TIME in each of top 20 grids (500m x 500m)

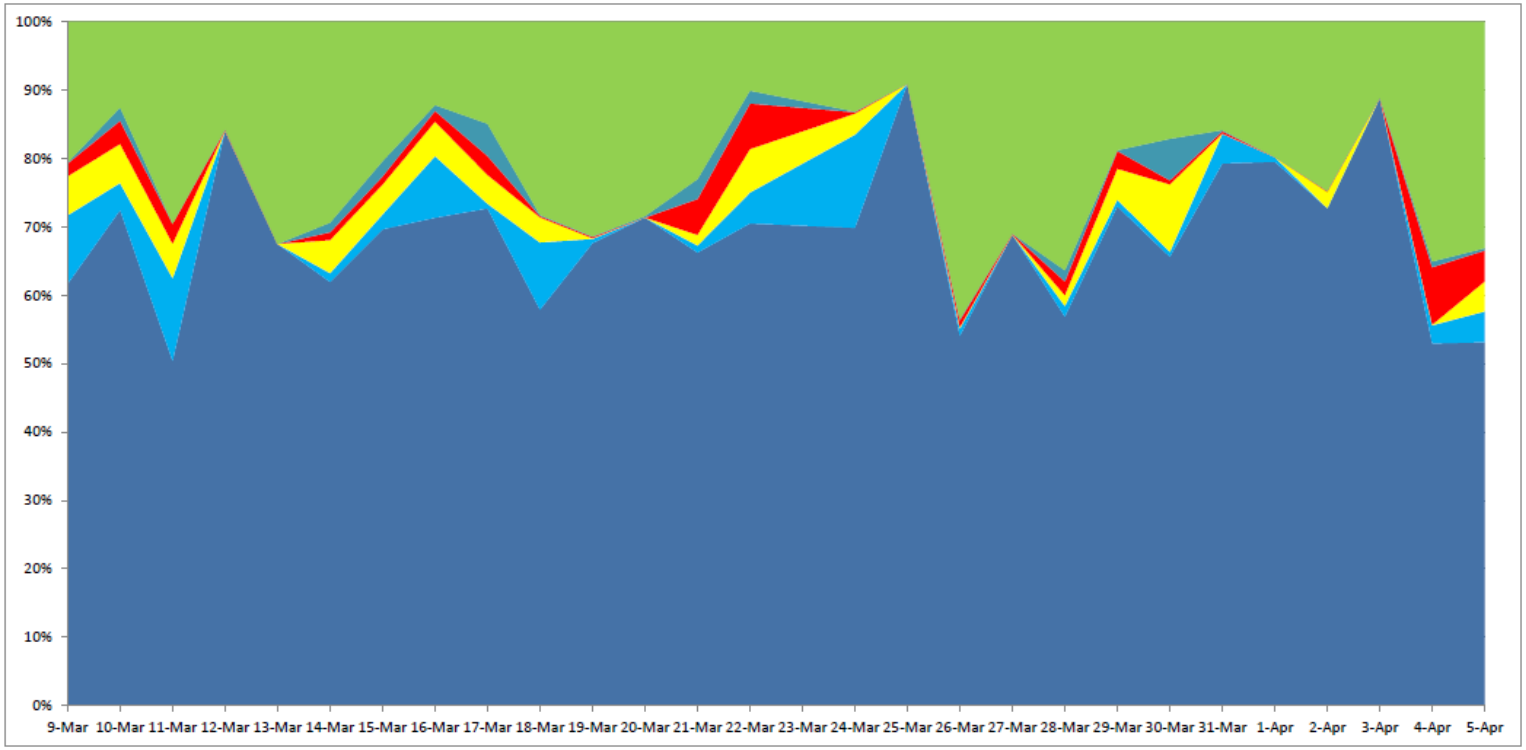


NB: The range of grid cells visited over 27 days can be very large

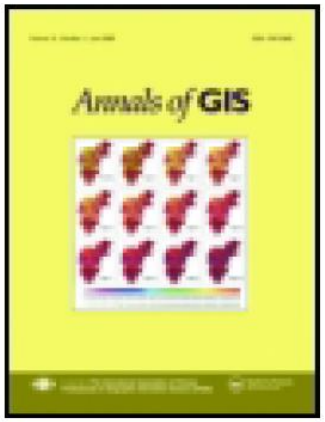
A 500m x 500m grid could be an area where multiple 'activity locations' may be found (e.g., home and place or worship; work and eating locations) and 'journeys.'

Top 5 + remainder
Frequently Visited Places (Grids)

Examples: Percent of TIME in each of top 5 grids **by day** (500m x 500m)



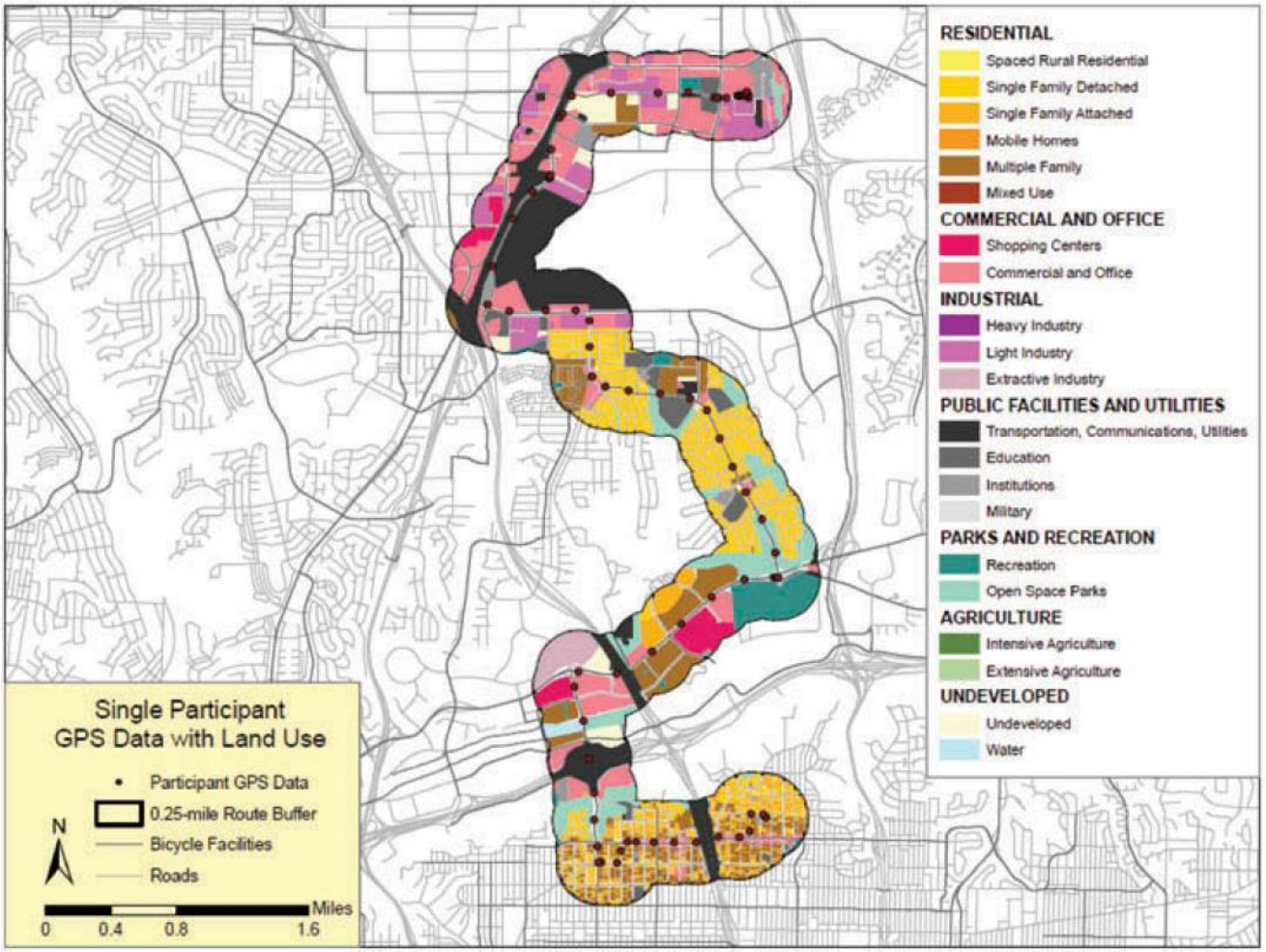
Series 1 = Grid cell with most points; Series 2 = Grid cell with second most points … to series 5; then Series 6 = remainder



Portion of the activity space of a person in San Diego associated with land use mix

Geospatial and Contextual Approaches to Energy Balance and Health

David Berrigan[a]*, J. Aaron Hipp[b], Philip M. Hurvitz[c], Peter James[d], Marta M. Jankowska[e], Jacqueline Kerr[e], Francine Laden[d], Tammy Leonard[f], Robin A. McKinnon[a], Tiffany M. Powell-Wiley[g], Elizabeth Tarlov[h,i], and Shannon N. Zenk[i] and the TREC Spatial and Contextual Measures and Modeling Work Group



ASCMEE Study first GPS publication

Importance of Time (Exposure)

Health & Place 51 (2018) xxx–xxx

Contents lists available at ScienceDirect

Health & Place

journal homepage: www.elsevier.com/locate/healthplace

How many days of global positioning system (GPS) monitoring do you need to measure activity space environments in health research?

Shannon N. Zenk[a,*], Stephen A. Matthews[b], Amber N. Kraft[c], Kelly K. Jones[a]

This study examined the number of days of global positioning system (GPS) monitoring needed to measure attributes of an individual's routine activity space. Multiple alternative activity space representations (cumulative, mean daily), measures (kernel density, route buffer, convex hull), and attributes (area size, supermarkets, fast food restaurants, parks) were examined. Results suggested wide variability in required GPS days to obtain valid estimates of activity space attributes (1–23 days). In general, fewer days were needed for mean daily activity space representations, kernel density measures, and densities of environmental exposures (vs. counts). While kernel density measures reliably estimated between-person differences in attributes after just a few days, most variability in environmental attributes for convex hull and route buffer measures was within-person. Based on these results, a minimum of 14 days of valid GPS data is recommended to measure activity spaces.

Measuring Spatial Segregation

Overview

PIs

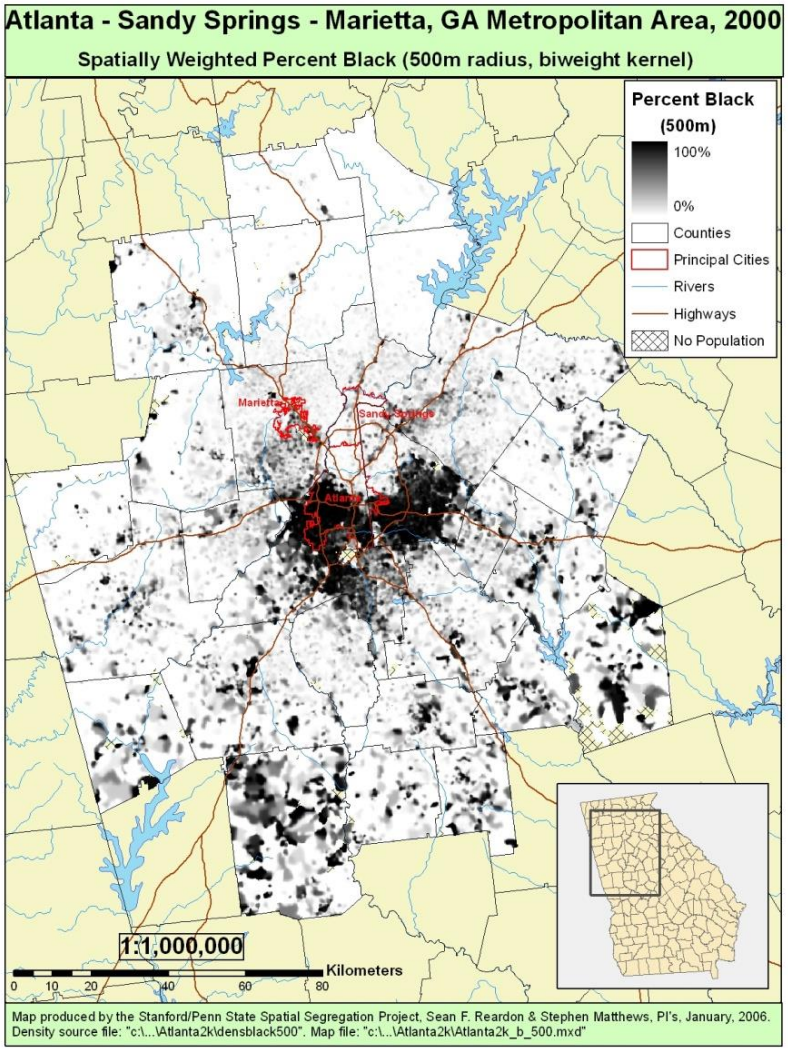
Sean Reardon (Stanford)
Stephen Matthews (Penn State)

Investigators

Barrett Lee (Penn State)
Glenn Firebaugh (Penn State)

Agency

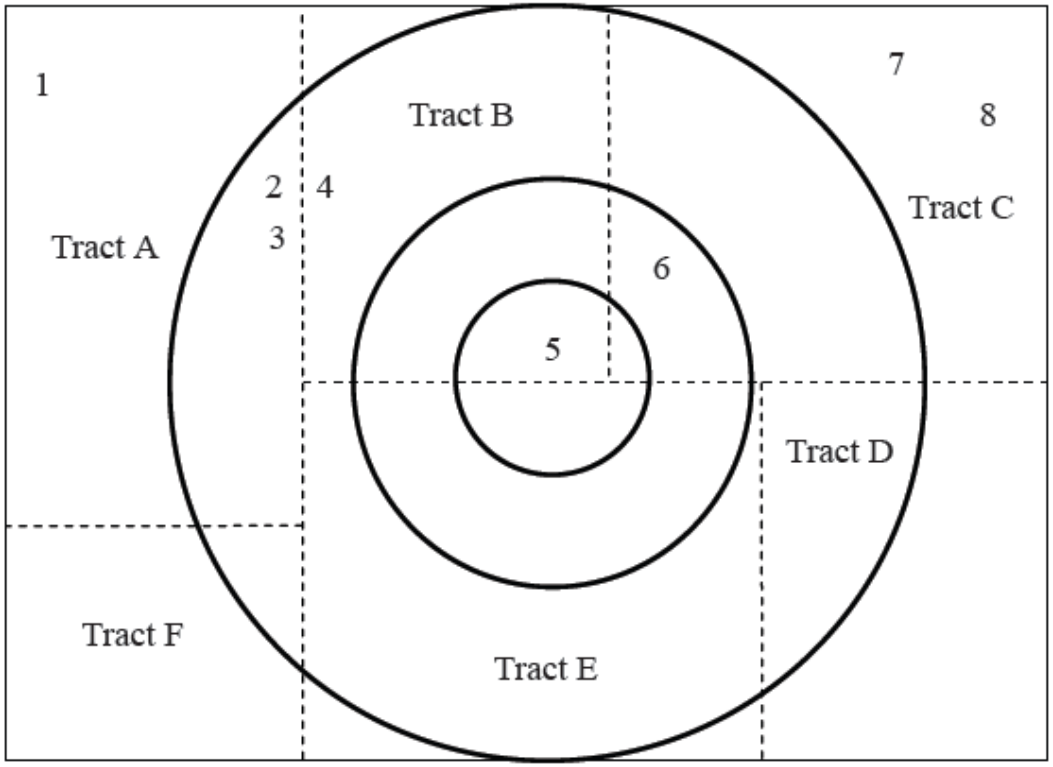
National Science Foundation (co-funded by the Sociology Division and the Methodology, Measurement and Statistics Division) and a consortium of federal statistical agencies



Publications in
Demography (2008)
American Sociological Review (2008)
Social Science Research (2009)

Figure 1. Census Tracts and Local Environments

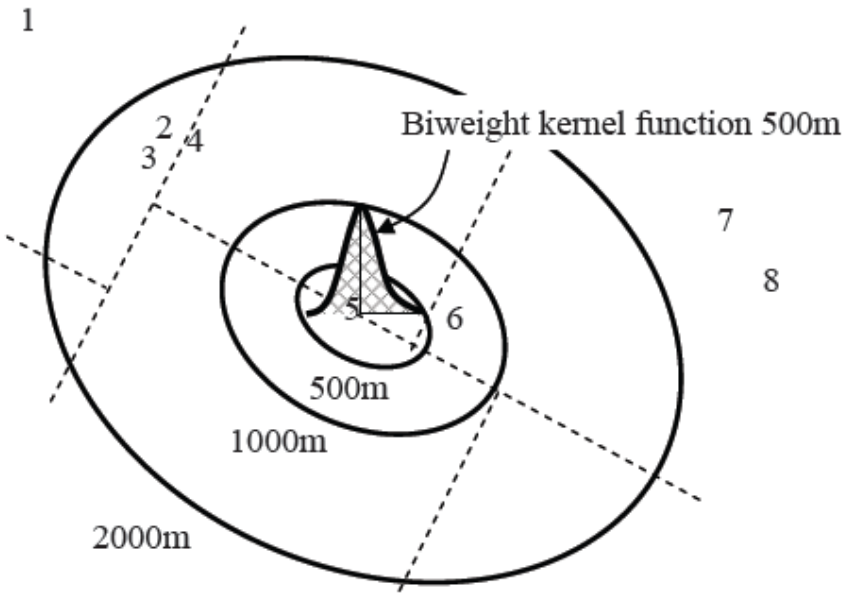
Map view



Segregation measures fail to distinguish among different locations within a tract. Person 1 is assumed to live as close to 2 and 3 as the latter do to each other.

The proximity assumption built in to segregation measures treats proximity is a binary condition. Persons living across the 'street' (e.g., 4 and 2) are judged to be more distant than individuals living in the same tract (e.g., 4 and 5).

Inset for person 5



We compute using a biweight kernel spatial proximity function (this weights nearer locations more than farther ones and has a finite maximum radius)

We use radii of 500m, 1000m, 2000m, & 4000m

500m radius ≈ pedestrian neighborhood
1000-2000m radius ≈ local institutional neighborhood
4000m radius ≈ maximum neighborhood radius (50 km^2 = 20 $miles^2$)
– local shopping, church, high school environment

This approach to measuring spatial segregation requires two types of information:

(1) an estimate of the population density of each group at each point in space;

(2) a measure of the spatial proximity between all pairs of points in a region.

In practice, we base our calculations here on estimates of race-specific population densities in a grid of 50-by-50 meter cells.

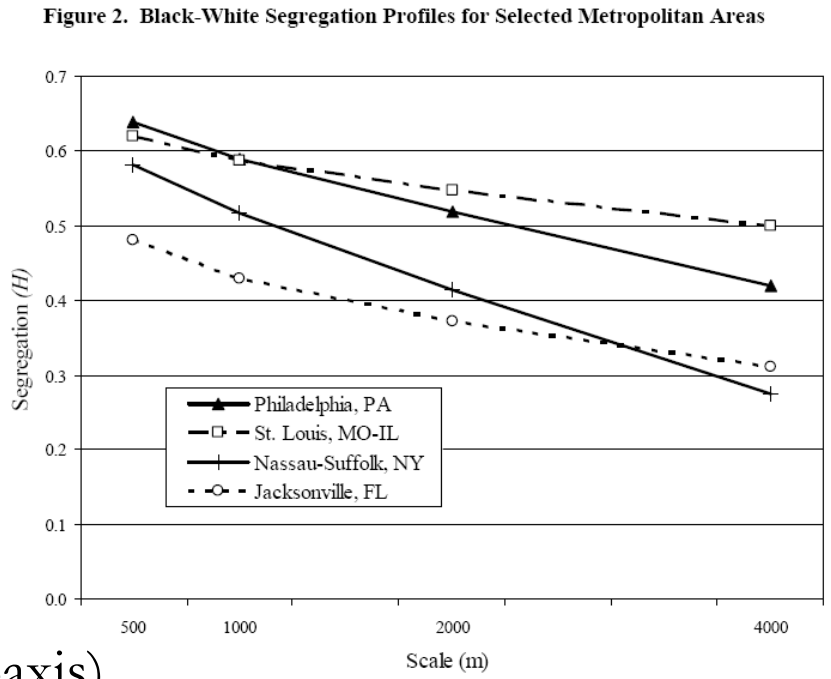
Data and procedures

- US Census 2000 data for 100 largest metro areas
- Block counts of non-HWs, Bs, As, all Hs
- Reconfigure/clean block data shapefiles with GIS software
- Divide block map into 50x50m cells
- Estimate racial group counts for each cell
- Smooth grid, preserving block counts
  (Tobler's ***pycnophylactic smoothing***)
- Apply proximity function to compute racial composition of LEs with nearby populations weighted more heavily

There are now faster routines available in R (seg) developed by David O'Sullivan and colleagues *PLOSOne* (2014)
https://doi.org/10.1371/journal.pone.0113767

The Segregation Profile

This is a scale-sensitive segregation measure.



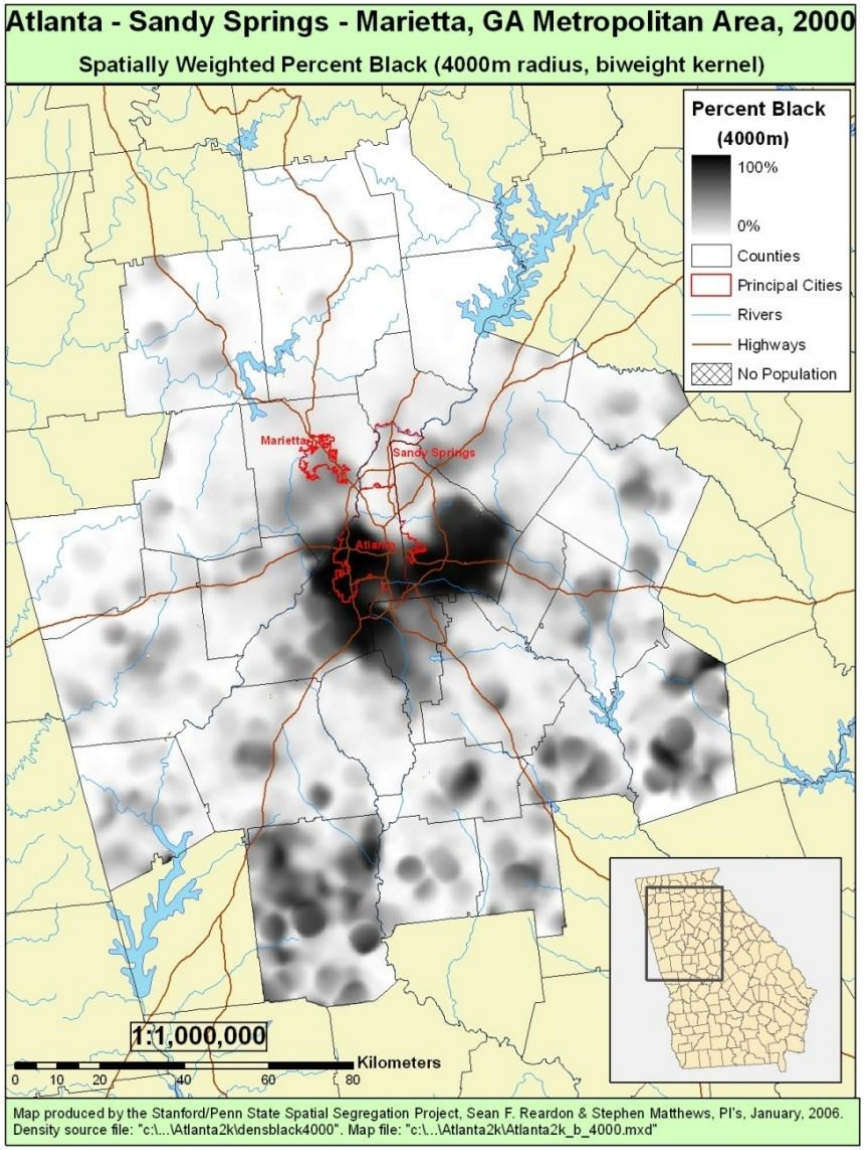
- We compute segregation at multiple scales (distance radii)
- We plot segregation (y-axis) v. scale (x-axis)
- Height of segregation profile at a given 'scale' indicates level of segregation **at that scale**
- The profile is monotonically decreasing
- **The slope of profile (the Macro/Micro Segregation Ratio) is <u>informative about scale of residential patterns</u>**

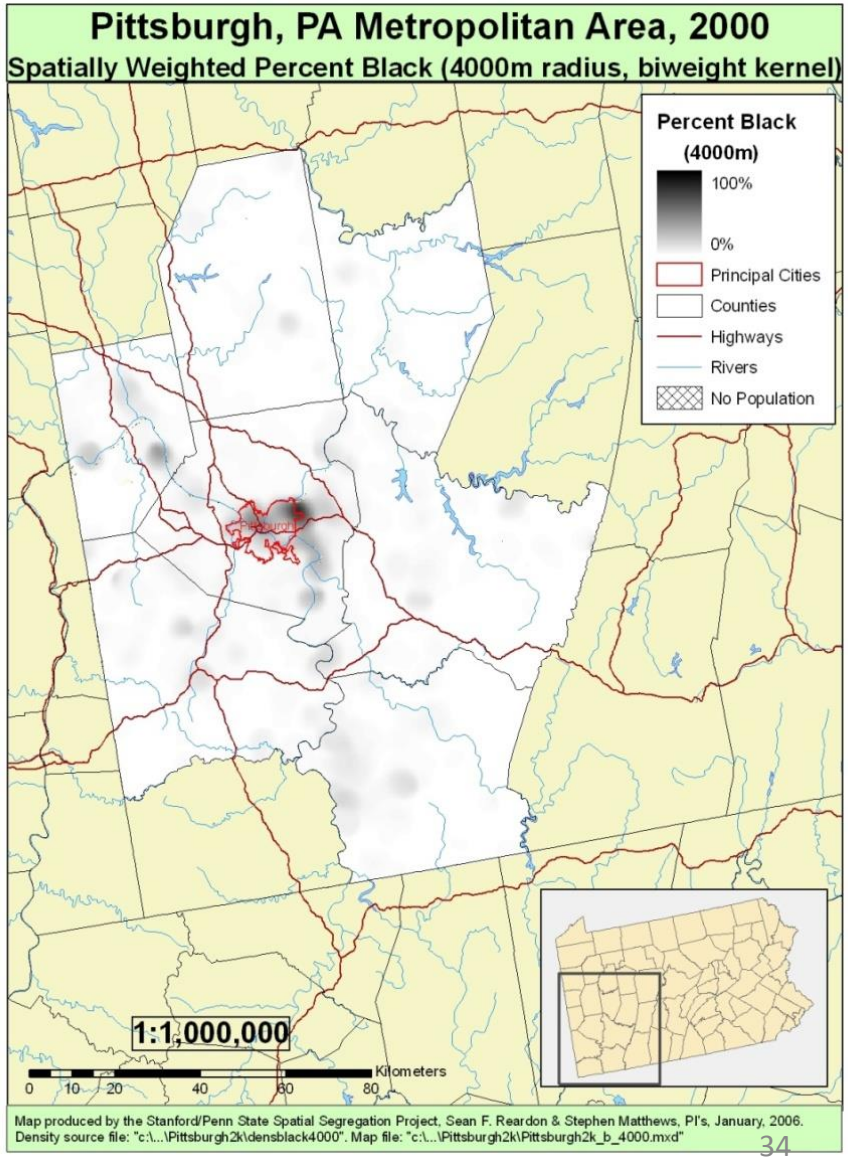
Illustrative Example: Black/White Segregation

	H_{500m}	H_{4000m}	H_{4000m}/H_{500m}
Atlanta	.554	.408	.75
Pittsburgh	.512	.260	.51

Atlanta macro-scale dominated: flat profile
(on average an individual's 4000m environment is only modestly more diverse than their 500m environment)

Pittsburgh micro-scale patterning: steep profile.


Atlanta - Sandy Springs - Marietta, GA Metropolitan Area, 2000
Spatially Weighted Percent Black (4000m radius, biweight kernel)
Percent Black
(4000m)
100%
0%
Counties
Principal Cities
Rivers
Highways
No Population
Marietta
Sandy Springs
Atlanta
1:1,000,000
0 10 20 40 60 80
Kilometers
Map produced by the Stanford/Penn State Spatial Segregation Project, Sean F. Reardon & Stephen Matthews, PI's, January, 2006.
Density source file: "c:\...\Atlanta2k\densblack4000". Map file: "c:\...\Atlanta2k\Atlanta2k_b_4000.mxd"


Pittsburgh, PA Metropolitan Area, 2000
Spatially Weighted Percent Black (4000m radius, biweight kernel)
Percent Black
(4000m)
100%
0%
Principal Cities
Counties
Highways
Rivers
No Population
1:1,000,000
0 10 20 40 60 80
Kilometers
Map produced by the Stanford/Penn State Spatial Segregation Project, Sean F. Reardon & Stephen Matthews, PI's, January, 2006.
Density source file: "c:\...\Pittsburgh2k\densblack4000". Map file: "c:\...\Pittsburgh2k\Pittsburgh2k_b_4000.mxd"

Opportunities

"Determining what "nearby" means in local environment terms for each type of resource (or risk) and social group strikes us as a crucial issue across a range of substantive topics.

Fortunately, the prospects for resolving it should continue to improve as spatially flexible concepts and methods of the sort introduced here, coupled with a growing supply of geocoded data, **permit sociologists to move beyond the census tract**."

Beyond the Census Tract: Patterns and Determinants of Racial Segregation at Multiple Geographic Scales

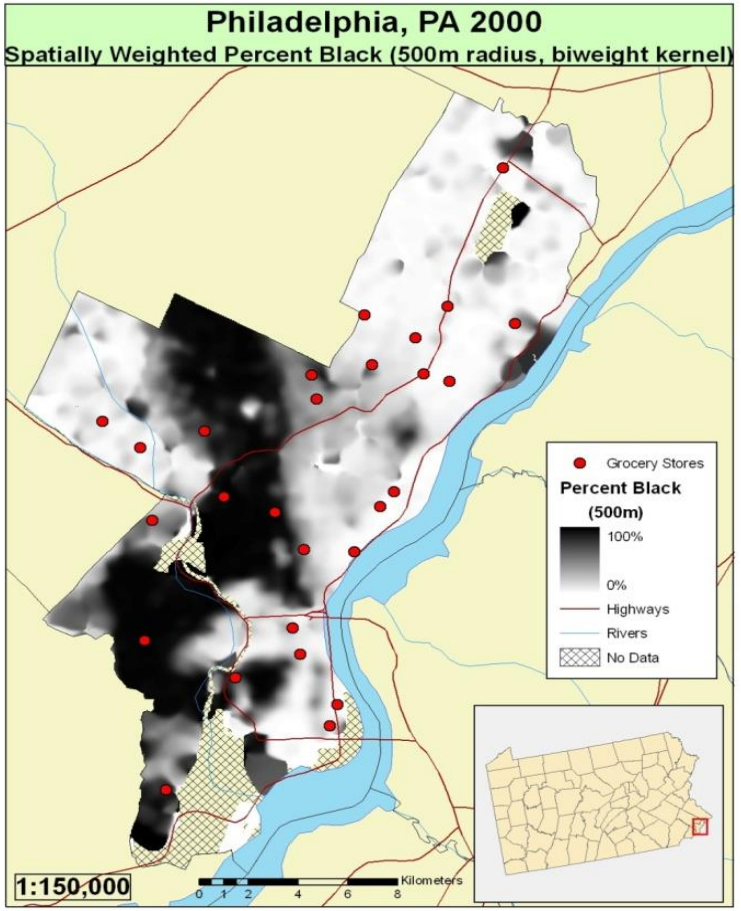
Barrett A. Lee, *The Pennsylvania State University*
Sean F. Reardon, *Stanford University*
Glenn Firebaugh, *The Pennsylvania State University*
Chad R. Farrell, *University of Alaska-Anchorage*
Stephen A. Matthews, *The Pennsylvania State University*
David O'Sullivan, *University of Auckland*

AMERICAN SOCIOLOGICAL REVIEW, 2008, VOL. 73 (October:766–791)

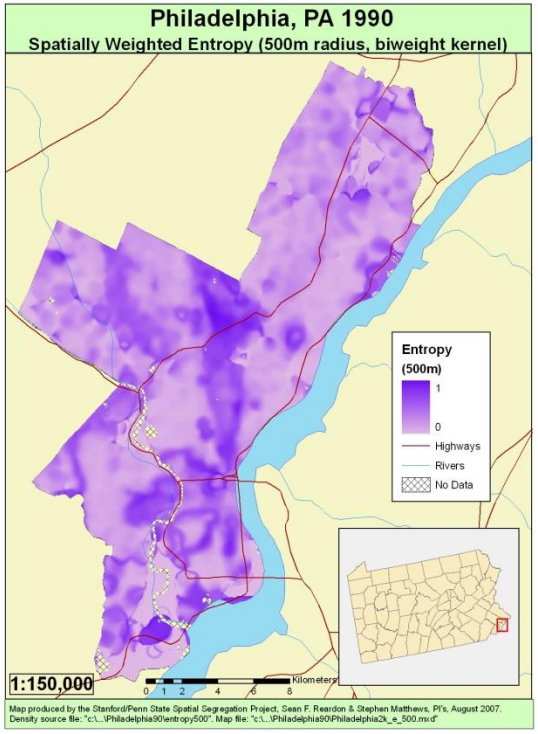
Segregation and Health Disparities

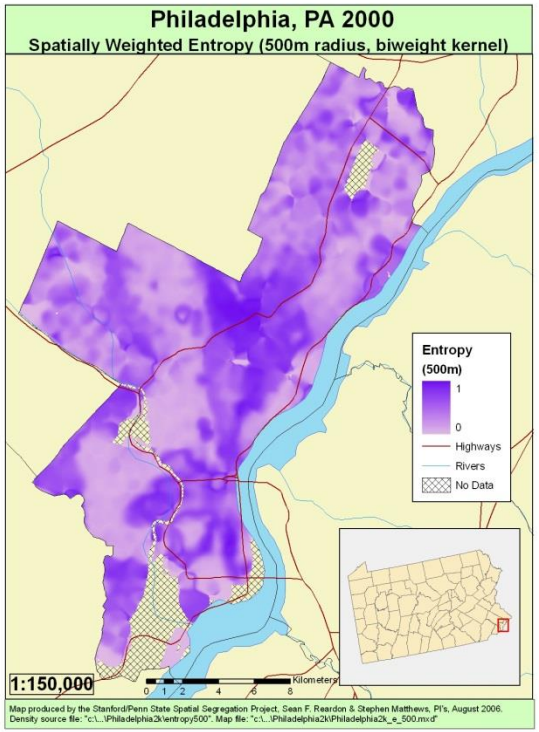
The replacement of tracts with egocentric local environments will allow investigators to think in a spatially sophisticated manner about the linkages between inequality and neighborhood circumstances: namely, how the settings that surround one's home can shape both **exposure to risks and access to resources and opportunities**.

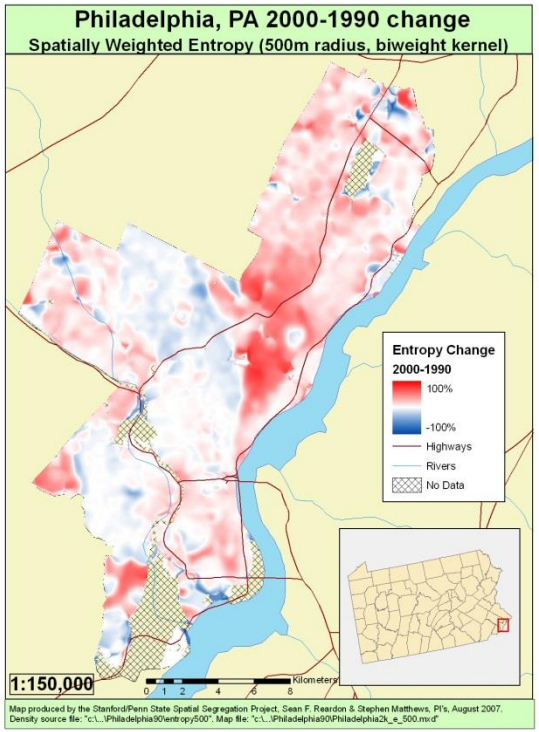
For example, Segregation and Grocery Store locations in Philadelphia



Segregation Change over Time (Map Algebra)







Accessibility: Creating a Grid Based Composite Index

Spat Demogr (2018) 6:17–34
https://doi.org/10.1007/s40980-016-0028-2

CrossMark

Measuring Spatial Accessibility to Urban Facilities and Services in Tehran

Hamidreza Rabiei-Dastjerdi[1] · Stephen A. Matthews[2,3] · Ali Ardalan[4,5,6]

The composite index of spatial accessibility was produced using the raster calculator and weighted overlay in ArcGIS.

Table 2 Urban facility and services maps

List	Facility and service	List	Facility and service	List	Facility and service
1	Ambulance	11	Local healthcare center	21	Red crescent
2	Disaster Management	12	Local hospital	22	Regional bus
3	Electricity	13	Metro line	23	Roads
4	Firefighting station	14	Metro station	24	Sewage
5	Food, vegetation, and fruit wholesale	15	Mosque	25	Sport field
6	Garbage collection	16	National Center for Natural Disaster	26	Storage
7	GAS	17	Park	27	Telecom center
8	Gas capsule	18	Petrol station	28	Traffic police
9	High school	19	Police	29	Vacant
10	Local government	20	Primary school	30	Water services

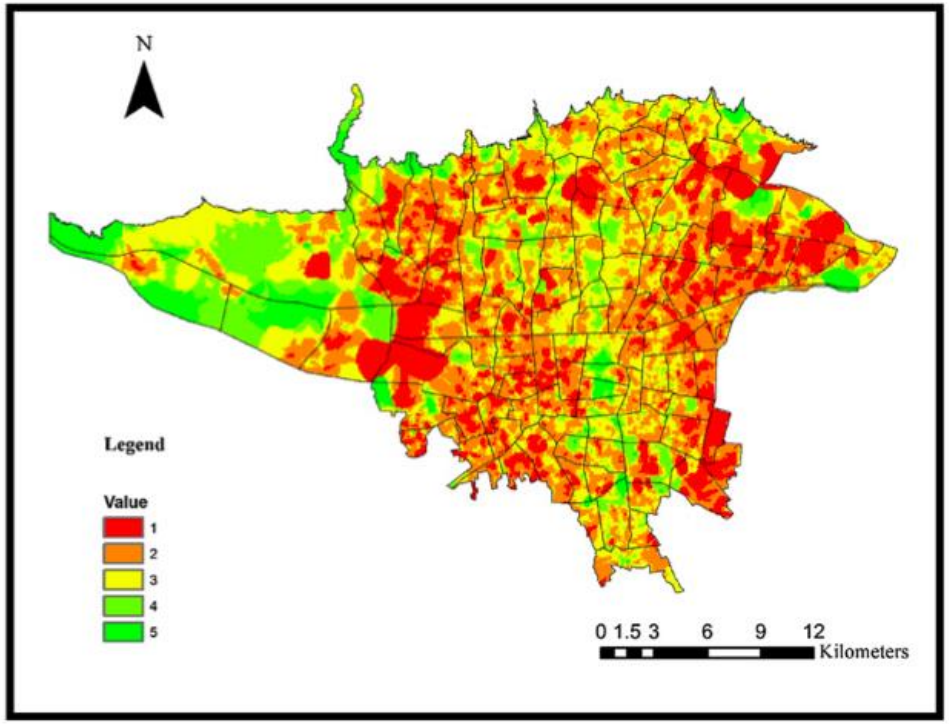


Fig. 9 Reclassification of accessibility

Measuring Spatial Accessibility to Urban Facilities and Services in Tehran

Hamidreza Rabiei-Dastjerdi[1] · Stephen A. Matthews[2,3] · Ali Ardalan[4,5,6]

Class	Area (%)	Population (%)
1	17	38.50
2	35	40.20
3	29	17.46
4	14	3.54
5	5	0.3
Total	100	100

Tehran's residents have low level of accessibility to urban amenities (i.e., class 1 and class 2), while less than four percent of the residents (i.e., those in class 4 and class 5) have the highest levels of accessibility. High levels of uneven accessibility thus exist within the city.

Map Algebra: National Gridded Database

Research Methods: Quantitative Research

American Journal of Health Promotion
2018, Vol. 32(3) 779-794
© The Author(s) 2017
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0890117117694448
journals.sagepub.com/home/ahp
SAGE

Weight and Veterans' Environments Study (WAVES) I and II: Rationale, Methods, and Cohort Characteristics

Shannon N. Zenk, PhD[1], Elizabeth Tarlov, PhD[1,2], Lisa M. Powell, PhD[3,4], Coady Wing, PhD[5], Stephen A. Matthews, PhD[6], Sandy Slater, PhD[3,4], Howard S. Gordon, MD[7,8], Michael Berbaum, PhD[4], and Marian L. Fitzgibbon, PhD[3,4]

Funding for WAVES

NIH/National Cancer Institute R01CA172726

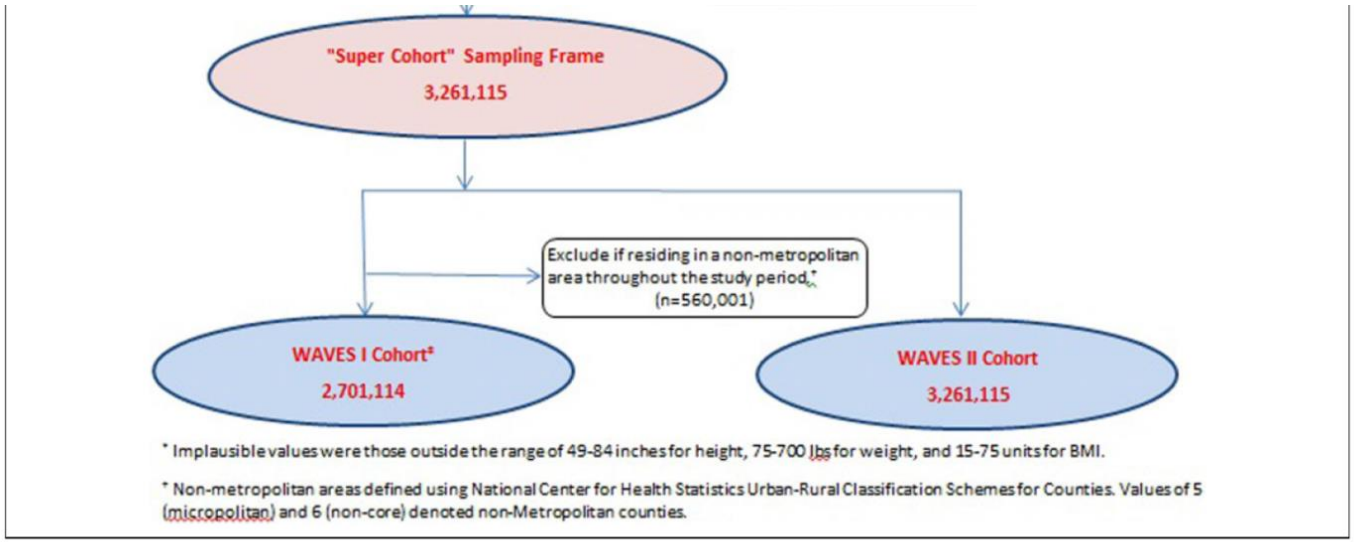
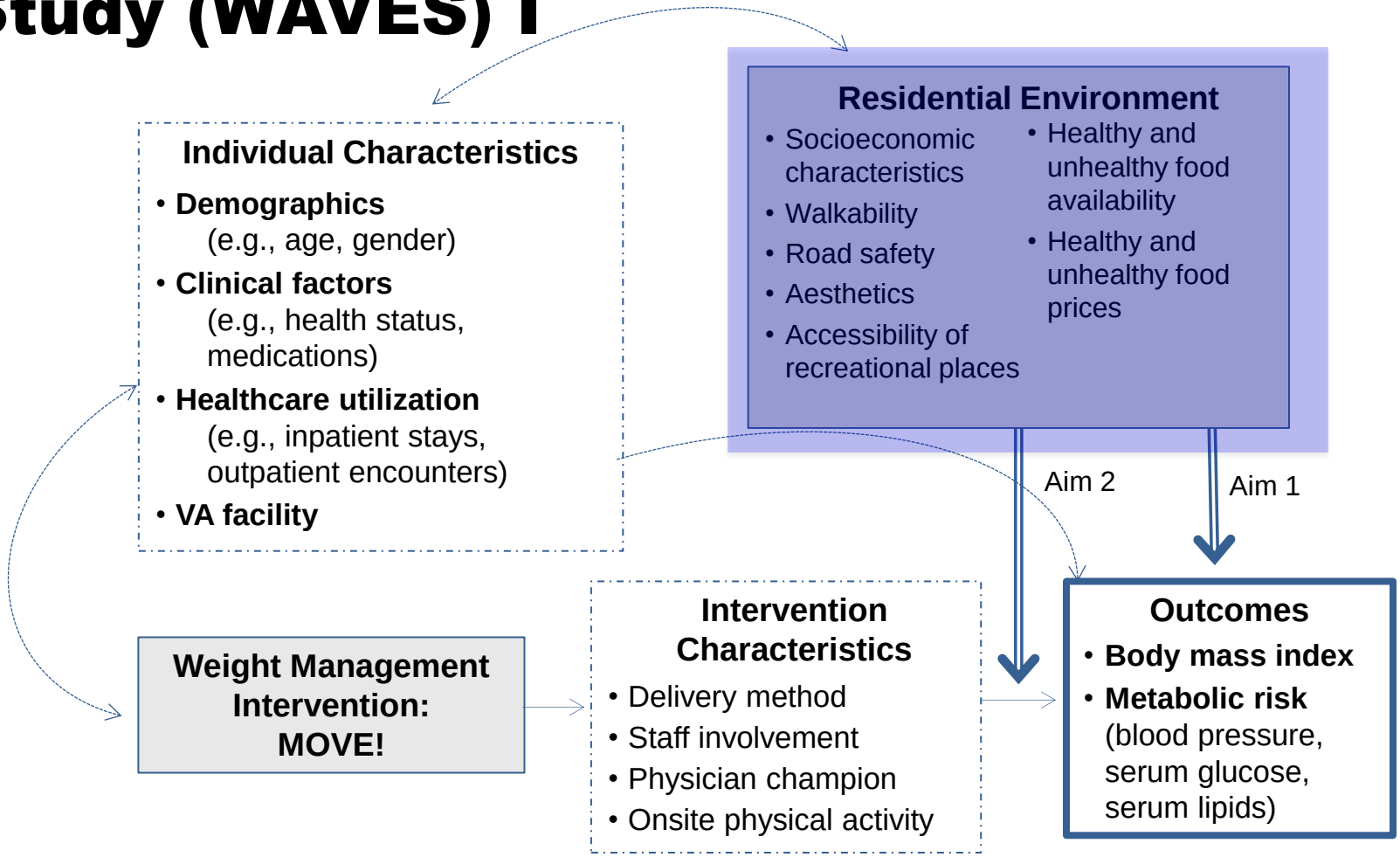


Figure 1. Sample derivation and sizes for weight and veterans' environments studies (WAVES) I and II.

Weight and Veterans Environments Study (WAVES) I



Conceptual model for WAVES I

WAVES I Strengths

- *Longitudinal* design (7 years)
- *National* coverage providing diversity in environments
- Health information, including clinically measured height, weight, and BP, medical diagnoses, medications, and lab results (lipids, glucose), from *electronic health records*
- *Precise* home address information
- Repeated environmental measures over time, both for "movers" and "stayers" (non-movers)
- Large sample (>200K) for weight management program participants

PENNSTATE POPULATION RESEARCH INSTITUTE
SOCIAL SCIENCE RESEARCH INSTITUTE

Jones et al. BMC Res Notes (2017) 10:35
DOI 10.1186/s13104-016-2355-1

BMC Research Notes

PROJECT NOTE Open Access

A step-by-step approach to improve data quality when using commercial business lists to characterize retail food environments

Kelly K. Jones[1*], Shannon N. Zenk[1], Elizabeth Tarlov[1,2], Lisa M. Powell[3], Stephen A. Matthews[4,5] and Irina Horoi[6]

Methods: We developed data improvement methods based on existing validation studies. These methods included purchasing records from commercial business lists (InfoUSA and Dun and Bradstreet) based on store/restaurant names as well as standard industrial classification (SIC) codes, reclassifying records by store type, improving geographic accuracy of records, and deduplicating records. We examined the impact of these procedures on food outlet counts in US census tracts.

Results: After cleaning and deduplicating, our strategy resulted in a 17.5% reduction in the count of food stores that were valid from those purchased from InfoUSA and 5.6% reduction in valid counts of restaurants purchased from Dun and Bradstreet. Locational accuracy was improved for 7.5% of records by applying street addresses of subsequent years to records with post-office (PO) box addresses. In total, up to 83% of US census tracts annually experienced a change (either positive or negative) in the count of retail food outlets between the initial purchase and the final dataset.

Table 6 Purchased and final business counts for InfoUSA (food stores) and Dun and Bradstreet (restaurants), 2007–2014

	Purchased data			Final data		
	1	2	3	4	5	6
	Count N	By SIC code N (%)	By name N (%)	Count	Unused[a]	Percent unused
InfoUSA	2,847,339	2,690,245 (94.5)	157,094 (5.5)	2,341,030	506,309	17.5
Dun and Bradstreet	2,143,147	2,104,369 (98.2)	38,778 (1.8)	2,023,032	120,115	5.6

[a] Records were unused in the final dataset if they had insufficiently accurate geocoding, were duplicates, or were purchased in error

Table 7 Store and restaurant counts before and after processing, overall and by store type, 2007–2014

	Supermarkets and grocery stores	Convenience stores	Pharmacies	Liquor stores	General merchandise stores	Limited service restaurants
Provisionally classified by SIC code	712,033	1,152,453	461,555	306,881	57,323	2,079,985
Reclassified by name	707,724	1,154,429	448,388	306,856	73,566	2,064,742
After cleaning for locational accuracy	643,306	1,038,993	424,842	286,261	69,592	2,023,032
After deduplication by name	624,845	977,178	408,979	280,473	N/A	N/A
After deduplication by address	624,700	977,165	408,935	280,473	N/A	N/A
Final count after cleaning and deduplication (excluding AK, HI)	621,343	972,735	407,270	278,895	60,787	2,023,032

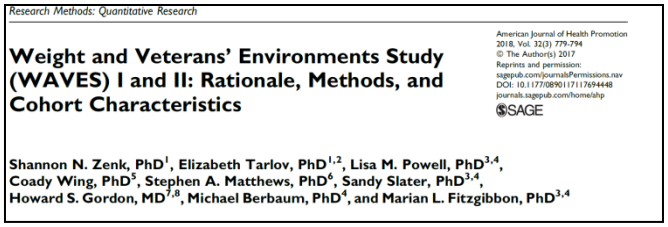
General merchandise stores were deduplicated using geographic deduplication. Count changes in the final step are due to exclusion of records in Alaska and Hawaii

Table 2. Residential Environmental and MOVE! Weight Management Program Exposure Measures.

Construct	Conceptual Definition	Operational Definition	Source
Residential environmental attributes			
Healthier food outlet accessibility[a]	Availability[b] of food outlets offering a large selection of healthier food products	Chain and nonchain supermarkets, grocery stores, supercenters and other mass merchandisers[c]	InfoUSA
Less healthy food outlet accessibility[a]	Availability[b] of food outlets primarily selling energy-dense, nutrient-poor food products	Convenience stores, chain and nonchain fast-food restaurants	InfoUSA, Dun & Bradstreet
Walkability[e]	Availability of nonresidential destinations and directness of walking routes	Street connectivity (eg, density of street intersections, ratio of 4-way to all street); population and housing unit densities; density of	NAVTEQ, ACS, InfoUSA and Dun & Bradstreet
Accessibility of recreational settings[a]	Availability[b] of indoor and outdoor settings to be physically active	Parks—count and land area, fitness facilities	NAVTEQ and TeleAtlas, InfoUSA
Aesthetics[e]	Attractiveness or appeal	Vacant housing unit density	ACS
Traffic safety[e]	Road features that slow or calm traffic	Presence of (a) median/dividers, (b) roundabouts, and (c) "low mobility" features (eg, low speeds and traffic volumes)	NAVTEQ
Socioeconomic and demographic characteristics	Deprivation and affluence	Percentage below poverty, median household income, percentage college educated	ACS
VA health-care accessibility	Proximity of VA health-care facility	Distance from patient's home to (a) nearest VA inpatient facility, (b) nearest VA outpatient facility, and (c) patient's most frequently used VA facility	
Urbanicity[f]	Degree of urbanization	Urban–rural Classification Scheme: metropolitan area (large central metro, large fringe metro, medium metro, and small metro) and nonmetropolitan area (micropolitan and noncore)	NCHS

Methods

- Map Algebra – or *Smartmaps* (Hurvitz & Moudon, 2012 *AJPM*)
- Divided country into 30m x 30m cells (1.68 Billion cells)
- Derived measures based on each grid cell's centroid for multiple spatial scales (0.25 mi to 5 mi) for all years separately
- Individuals veterans then assigned value for measure based on grid cell where home located



Research Methods: Quantitative Research

Weight and Veterans' Environments Study (WAVES) I and II: Rationale, Methods, and Cohort Characteristics

American Journal of Health Promotion
2018, Vol. 32(3) 779-794
© The Author(s) 2017
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0890117117694448
journals.sagepub.com/home/ahp
SAGE

Shannon N. Zenk, PhD[1], Elizabeth Tarlov, PhD[1,2], Lisa M. Powell, PhD[3,4], Coady Wing, PhD[5], Stephen A. Matthews, PhD[6], Sandy Slater, PhD[3,4], Howard S. Gordon, MD[7,8], Michael Berbaum, PhD[4], and Marian L. Fitzgibbon, PhD[3,4]

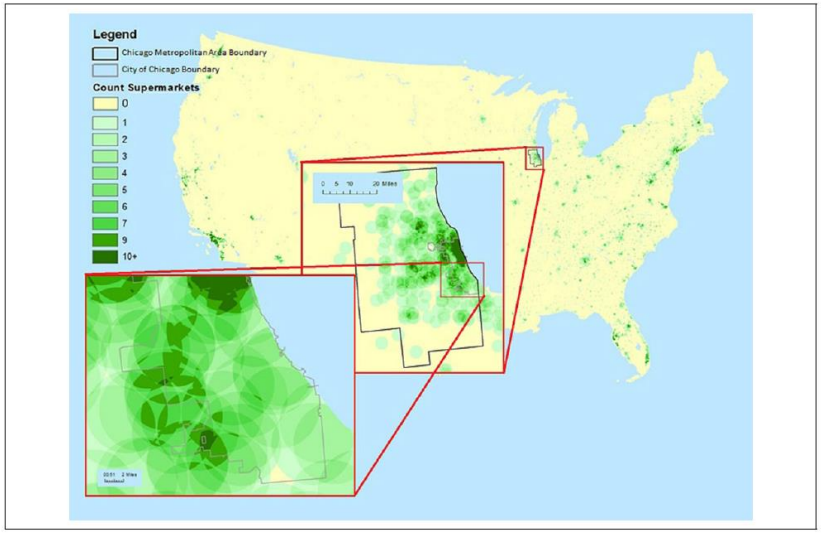


Figure 2. Accessibility of supermarkets within 3 miles for the continental United States.

POPULATION HEALTH

By Shannon N. Zenk, Elizabeth Tarlov, Coady Wing, Stephen A. Matthews, Kelly Jones, Hao Tong, and Lisa M. Powell

Geographic Accessibility Of Food Outlets Not Associated With Body Mass Index Change Among Veterans, 2009–14

DOI: 10.1377/hlthaff.2017.0122
HEALTH AFFAIRS 36,
NO. 8 (2017): 1433–1442
©2017 Project HOPE—
The People-to-People Health
Foundation, Inc.

ABSTRACT In recent years, various levels of government in the United States have adopted or discussed subsidies, tax breaks, zoning laws, and other public policies that promote geographic access to healthy food. However, there is little evidence from large-scale longitudinal or quasi-experimental research to suggest that the local mix of food outlets actually affects body mass index (BMI). We used a longitudinal design to examine whether the proximity of food outlets, by type, was associated with BMI changes between 2009 and 2014 among 1.7 million veterans in 382 metropolitan areas. We found no evidence that either absolute or relative geographic accessibility of supermarkets, fast-food restaurants, or mass merchandisers was associated with changes in an individual's BMI over time. While policies that alter only geographic access to food outlets may promote equitable access to healthy food and improve nutrition, our findings suggest they will do little to combat obesity in adults.

EXHIBIT 1

Descriptive statistics for the total sample of veterans who received health care services from the VA and those who were nonmigrants at baseline, by sex

	Men		Women	
	Total sample ($n = 1,522,803$)	Nonmigrants[a] ($n = 1,034,375$)	Total sample ($n = 183,618$)	Nonmigrants[a] ($n = 112,670$)
Mean age, years (SD)	51.8 (11.5)	52.5 (11.3)	43.4 (11.5)	44.5 (11.3)
BMI				
Mean (SD)	30.2 (6.0)	30.3 (6.0)	29.5 (6.4)	29.6 (6.4)
Underweight or normal weight (less than 25.0)	18.2%	17.6%	26.5%	25.9%
Overweight (25.0–29.9)	35.9	35.9	31.1	31.2
Obese (30.0 or more)	45.9	46.5	42.4	42.9
MARITAL STATUS				
Married	48.8	53.0	33.3	36.6
Separated or divorced	26.2	23.8	31.6	30.1
Widowed	1.8	1.7	2.2	2.2
Single	21.8	19.9	30.8	28.7
Unknown	1.4%	1.6%	2.1%	2.4%
RACE/ETHNICITY				
Non-Hispanic white	60.5%	60.9%	50.1%	49.8%
Non-Hispanic black	22.5	20.8	32.1	31.1
Hispanic	6.0	6.0	6.0	5.8
Other	2.5	2.6	3.3	3.3
Unknown	8.5	9.7	8.5	10.1

POPULATION HEALTH

By Shannon N. Zenk, Elizabeth Tarlov, Coady Wing, Stephen A. Matthews, Kelly Jones, Hao Tong, and Lisa M. Powell

Geographic Accessibility Of Food Outlets Not Associated With Body Mass Index Change Among Veterans, 2009–14

DOI: 10.1377/hlthaff.2017.0122
HEALTH AFFAIRS 36,
NO. 8 (2017): 1433–1442
©2017 Project HOPE—
The People-to-People Health
Foundation, Inc.

Other topics

Grids in Demography

Grids and data integration (across social, built and physical environment data – see also DHS)

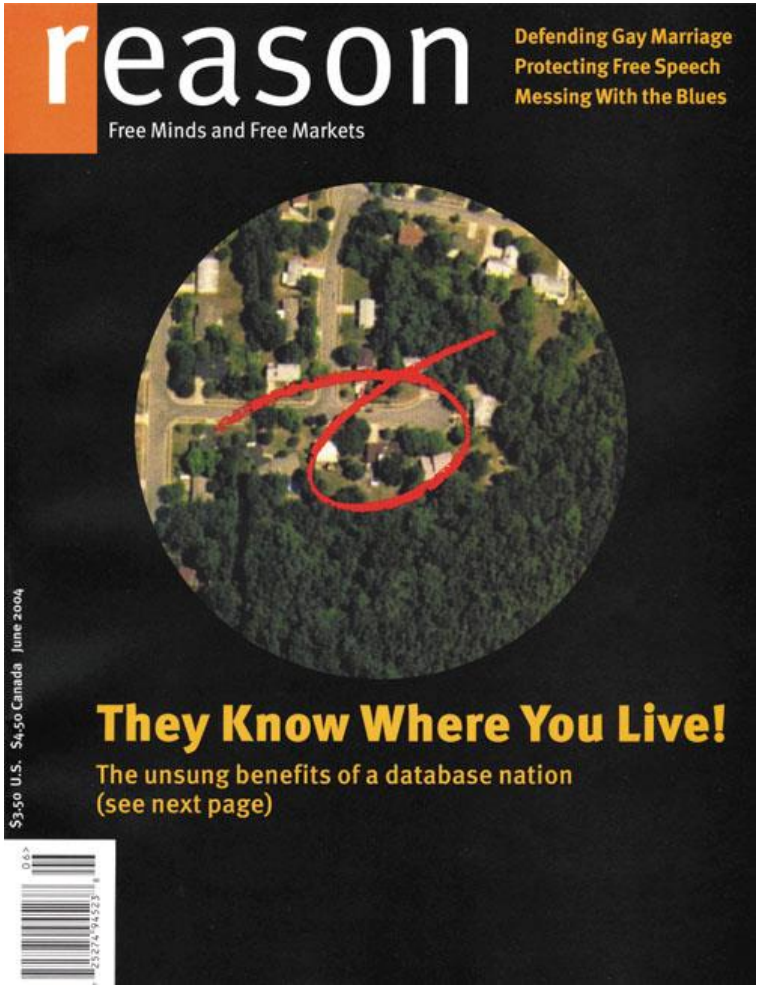
Grids and new (time-constant) geographies (moving beyond polygons and flexible data aggregation)

Grids, scale and privacy

DHS randomizes cluster points to 'preserve' privacy (but point must fall within correct district polygon)

Privacy and sexual health review paper

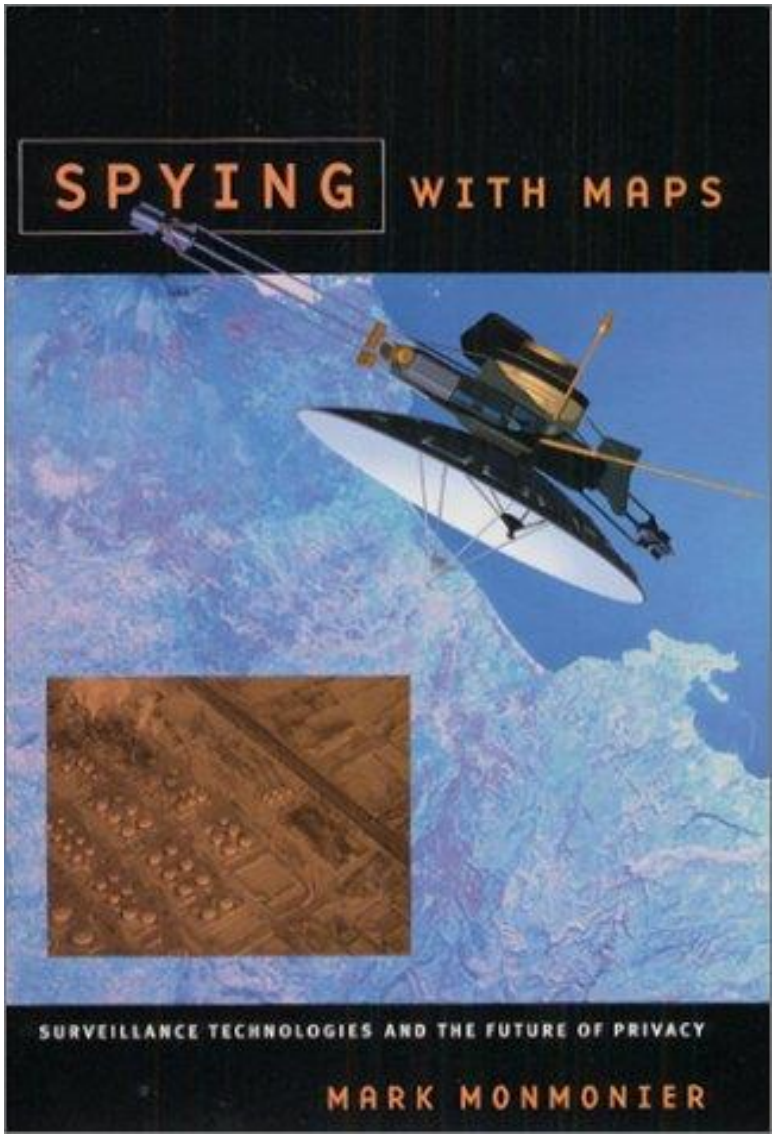
Collection vs. Analysis and Display



In June 2004 ***reason*** (a monthly magazine) delivered a unique copy to its 40,000 subscribers.

Each copy had a different cover. On each cover was an aerial photo of the subscriber's neighborhood, with a circle around the subscriber's house.

NPR covered the story
http://www.npr.org/templates/story/story.php?storyId=1870509



Mark Monmonier. 2002. Spying with Maps: Surveillance Technologies and the Future of Privacy. Chicago, IL: University of Chicago Press.

Following Monmonier (p. 1) note …

"the maps one looks at are less important than the spatial data systems that store and integrate facts about where we live and work (p. 1)"

In terms of the "maps we look at" there are mapping strategies to avoid the most obvious privacy violations (removing map layers, use of small-scale maps, aggregating data, masking techniques).

Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data

Panel on Confidentiality Issues Arising from the Integration of Remotely Sensed and Self-Identifying Data, National Research Council

ISBN: 978-0-309-10414-2, 176 pages, 6 x 9, paperback (2007)

Precise, accurate spatial information linked to social and behavioral data is revolutionizing social science by opening new questions for investigation and improving understanding of human behavior in its environmental context. At the same time, precise spatial data make it more likely that individuals can be identified, breaching the promise of confidentiality made when the data were collected. Because norms of science and government agencies favor open access to all scientific data, the tension between the benefits of open access and the risks associated with potential breach of confidentiality pose significant challenges to researchers, research sponsors, scientific institutions, and data archivists. Putting People on the Map finds that several technical approaches for making data available while limiting risk have potential, but none is adequate on its own or in combination. This book offers recommendations for education, training, research, and practice to researchers, professional societies, federal agencies, institutional review boards, and data stewards.

Thank you

E-mail: matthews@psu.edu or sxm27@psu.edu

Funding Agencies and Institutions

Past: NSF SES-0520400[405] Spatial Seg); NSF SES 041473;
Current: NCI R21 CA1695353 (UCSD: GIS/GPS); NCI R01 CA172726 (UIC/VA Project)
Institution-level: NICHD R24 HD41025 (PRI Center Grant) and Social Science Research Institute, Penn State; University of Illinois at Chicago.
And Valarie, Kayleigh and Celine

Finding the "Holy Grail" of Contextual Environments

School Districts

ZIP Codes

? Activity Spaces

Tracts

"fine" surfaces

AREAS

Egocentric Spaces

Scale Issues

POINTS

GRIDS & SURFACES

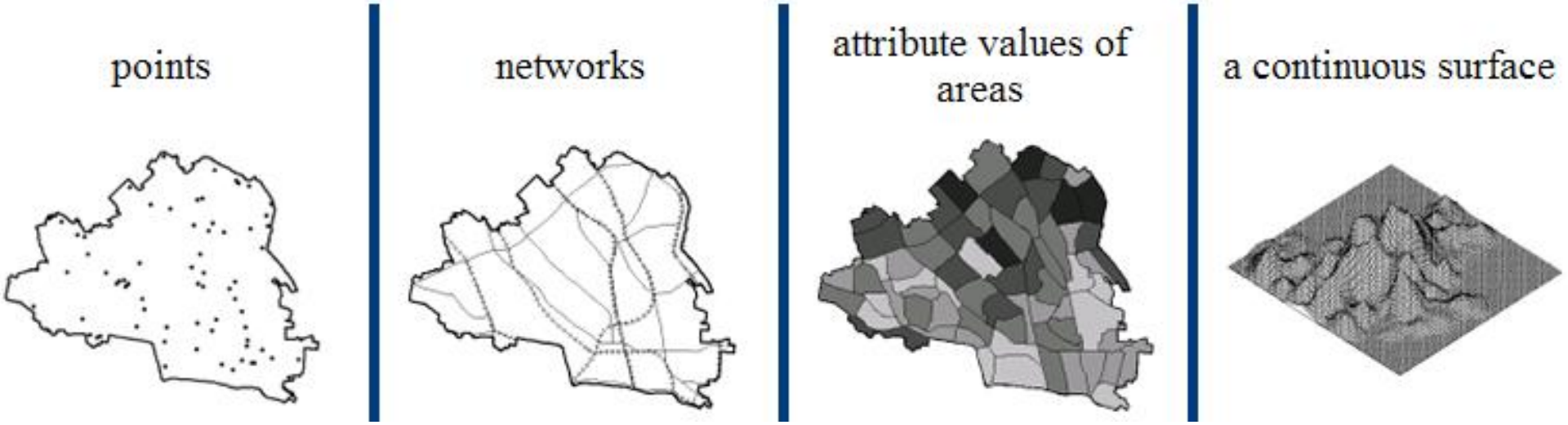
Geometric Buffers

"crude" surfaces

More than any other time in history, mankind faces a crossroads. One path leads to despair and utter hopelessness. The other, to total extinction. Let us pray we have the wisdom to choose correctly.
Woody Allen *My Speech to the Graduates*

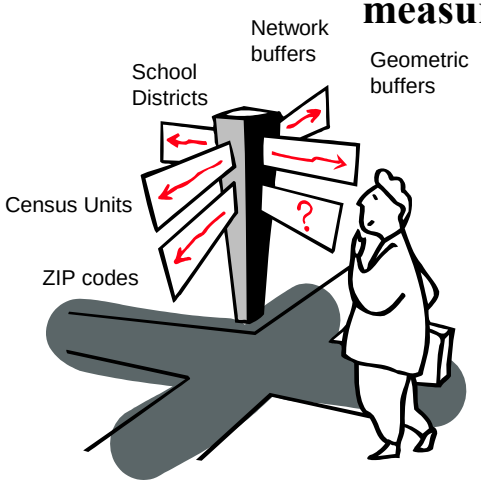
Implications for Spatial Analysis

The way that spatial data are stored – or how geographical phenomena are represented in a GIS – is increasingly important for subsequent analysis.



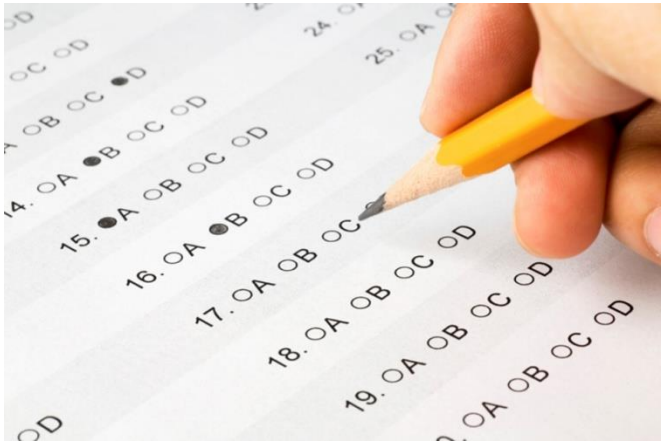
Thinking about place/context) … Some key points

1: There are **many blinkers on how we theorize**, and therefore measure, place. Mostly **binary** thinking.
2: **Different disciplines measure place differently**
3: **The measurement of place is not benign**



Q17: What is the best spatial unit to use to measure place/contexts/environments?

A: Census Blockgroup/Tract
B: School District
C: Grids
D: Geometric Buffers
E: Street Network Buffers



"Our results depend on our units" Yule & Kendal, 1950, p.313.