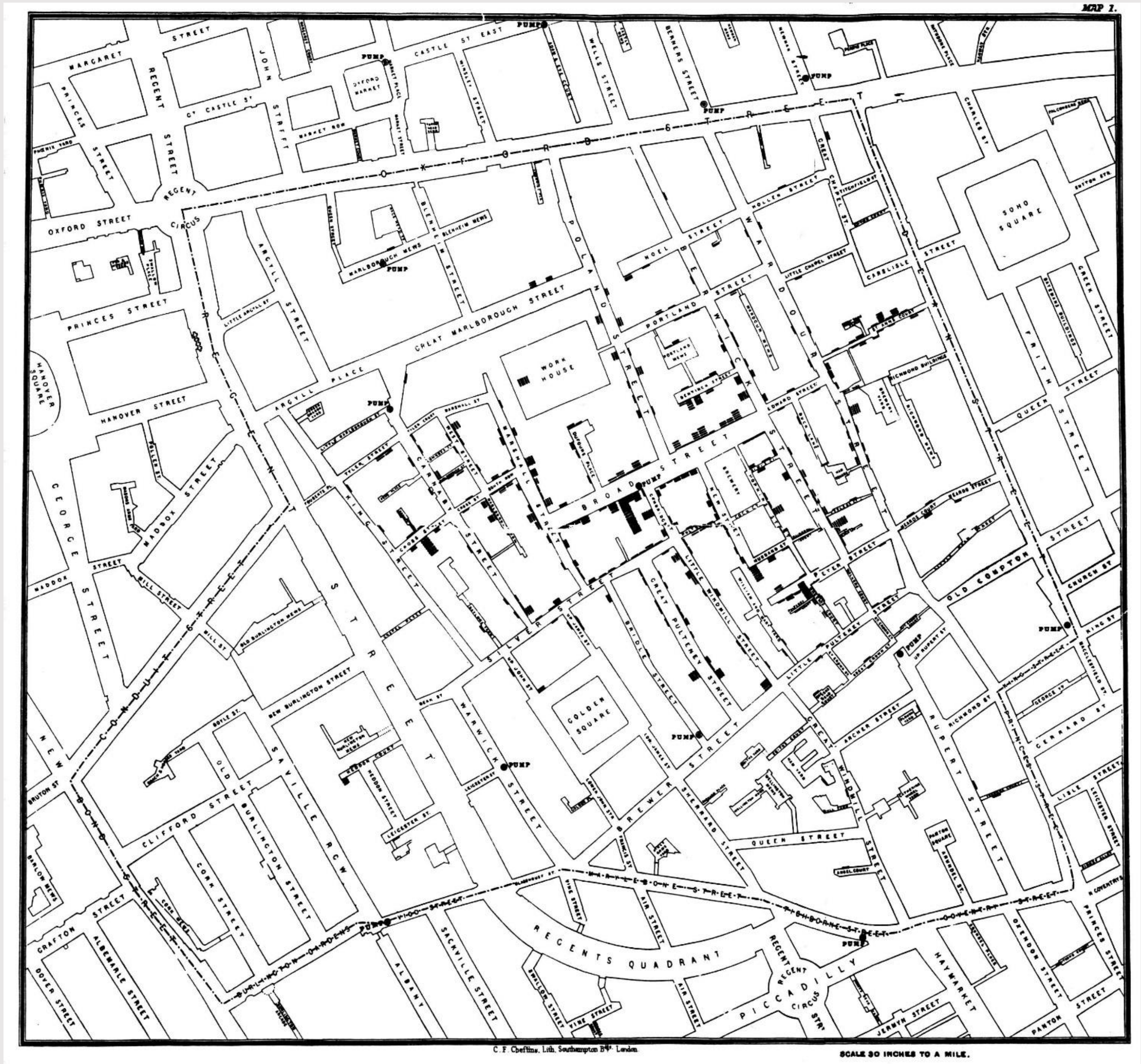
Applied spatial analysis in perinatal and pediatric epidemiology

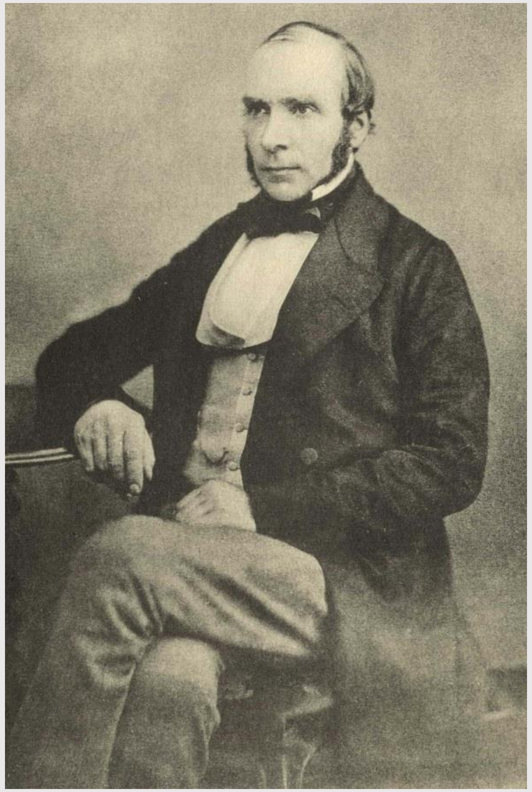
Michael Kramer
Emory University
mkram02@emory.edu
@mkramer_atl

SPER 2015
Methods Workshop
Denver, CO

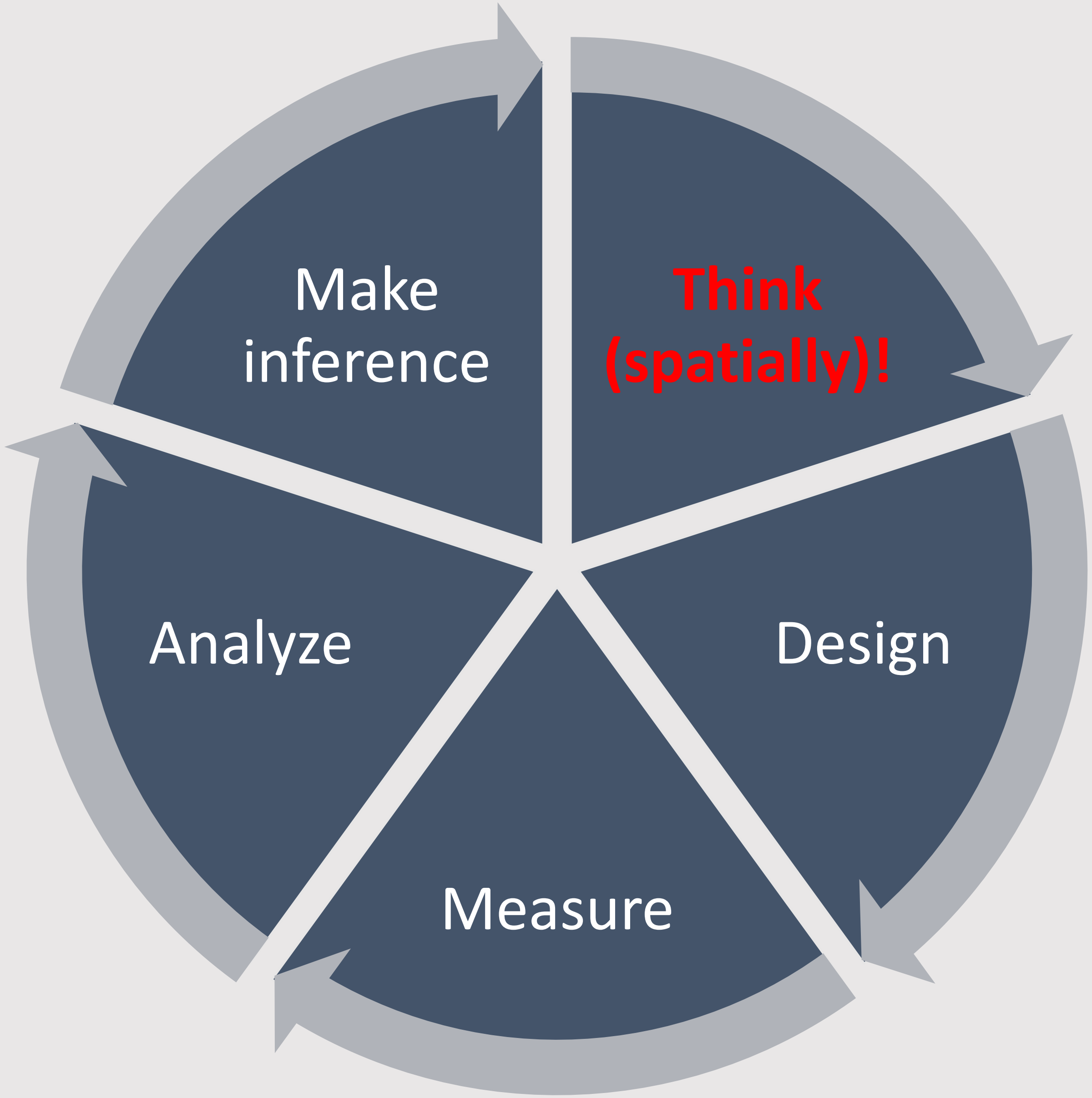
Primary objectives

1. **Spatial thinking!** When is spatial analysis useful for pediatric and perinatal epidemiologists? What does it add to our knowledge?
2. **Spatial structure**: do like events/exposures cluster?
3. **Discrete or continuous space**: within which boundaries does health happen?
4. **Spatial scale**: how big IS a neighborhood?
5. **Next steps**: further learning and software tools


MAP 1.
OXFORD STREET
REGENT CIRCUS
PRINCES STREET
HANOVER SQUARE
HANOVER STREET
GEORGE STREET
ARGYLL PLACE
GREAT MARLBOROUGH STREET
WORK HOUSE
BROAD STREET
PUMP
POLAND STREET
BERWICK STREET
WARDOUR STREET
SOHO SQUARE
FRITH STREET
GREEK STREET
OLD COMPTON STREET
DEAN STREET
GOLDEN SQUARE
WARWICK STREET
NEW BURLINGTON STREET
SAVILLE ROW
CLIFFORD STREET
BURLINGTON STREET
CORK STREET
CONDUIT STREET
NEW BOND STREET
QUEEN STREET
BREWER STREET
SHERRARD STREET
GREAT PULTENEY STREET
RUPERT STREET
REGENTS QUADRANT
PICCADILLY
REGENT CIRCUS
HAYMARKET
JERMYN STREET
SACKVILLE STREET
ALBEMARLE STREET
DOVER STREET
GRAFTON STREET
C. F. Cheffins, Lith, Southampton Bldgs London
SCALE 30 INCHES TO A MILE.

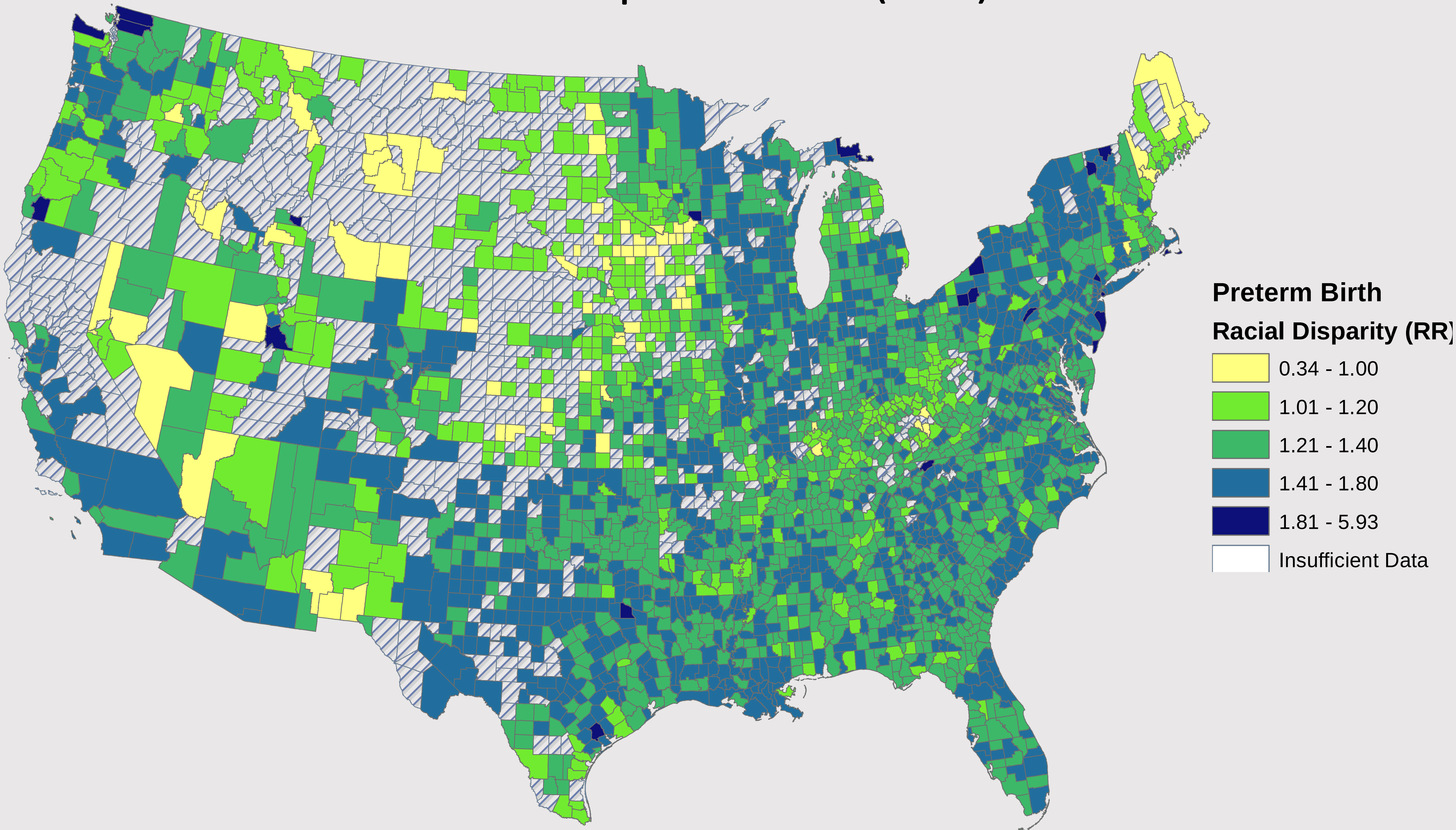


Epidemiology Work Flow



1. Spatial thinking

Black-White Racial Disparities (RR) in PTB



CASE #1 – Preterm Birth in U.S. Counties

1. Describe the spatial patterns…
2. What are pros and cons of the spatial scale (counties)?
3. What causes spatial variation in risk?
4. Does this map communicate something about risk that would be missed in a table?

Space as a storyteller

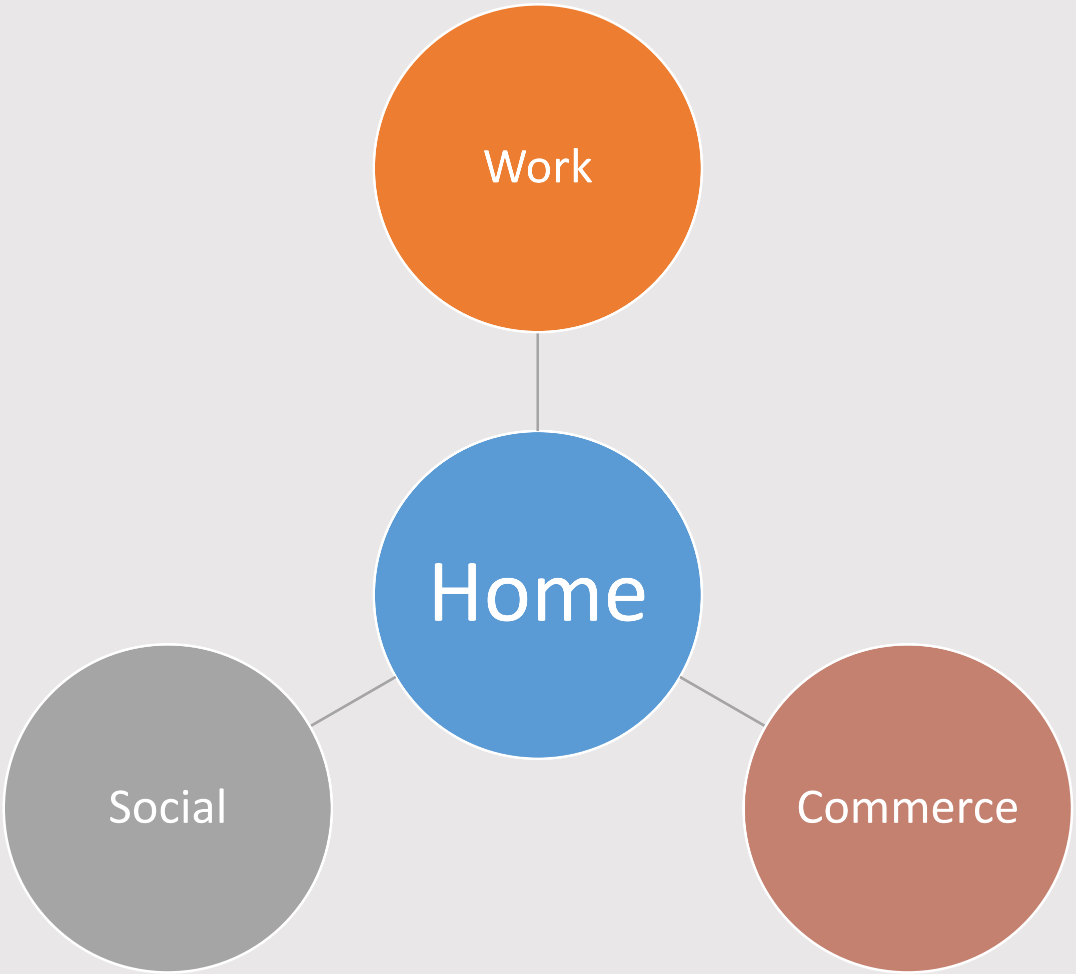
Uses	Examples
• Surveillance	• Map health outcomes
• Descriptive analysis	• Overlay social and health resources
• Hypothesis generating	• Test whether exposure proximity affects health
• Resource allocation	• Describe service catchment area

Space as a conduit

Spatial ≈ Social

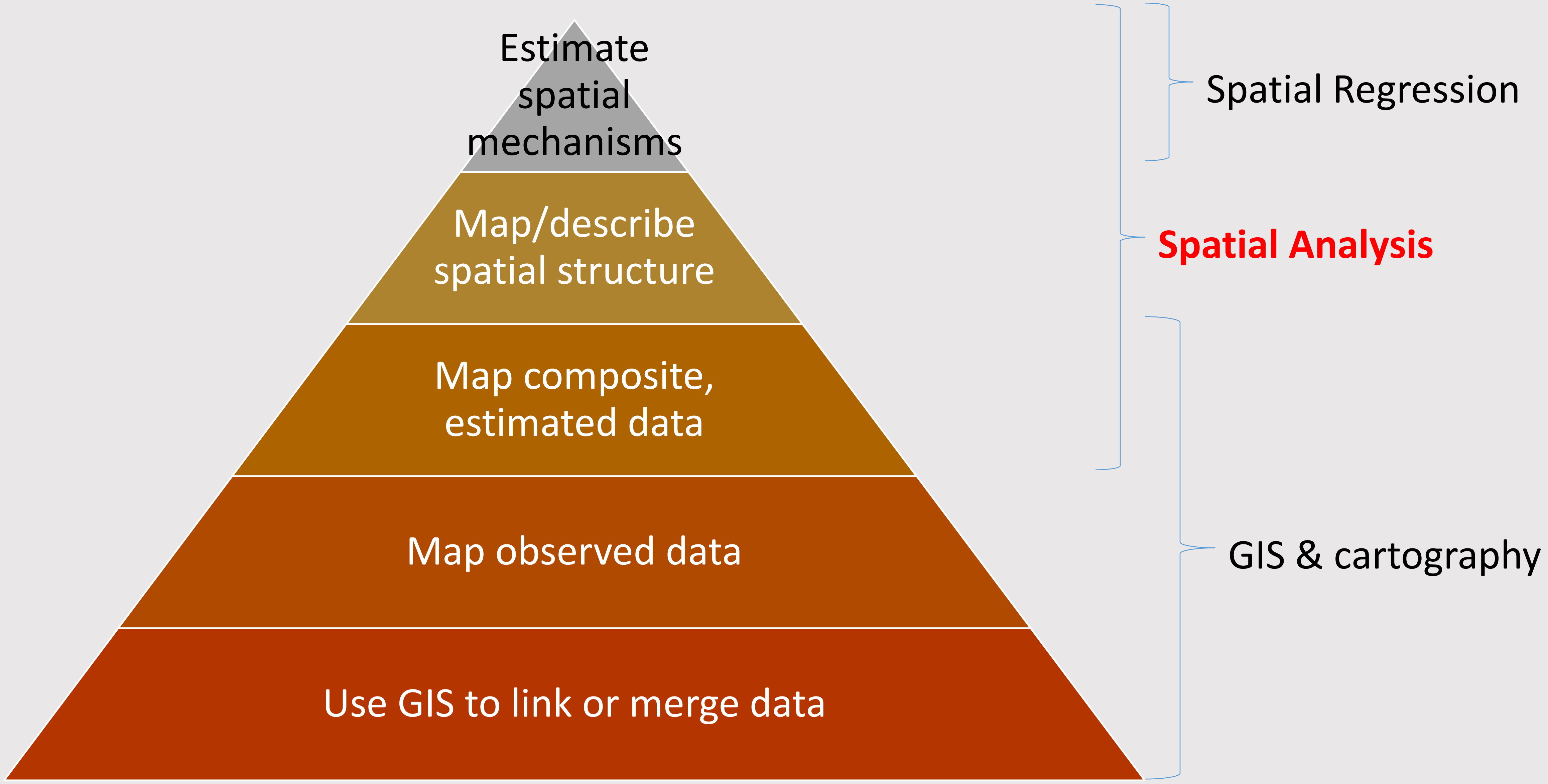
- Spatial patterns exist because social processes (choice, stratification, economics, culture) sort us into locations and networks
  - Residential segregation
  - Neighborhood deprivation
  - School quality
  - Social support

What places matter?

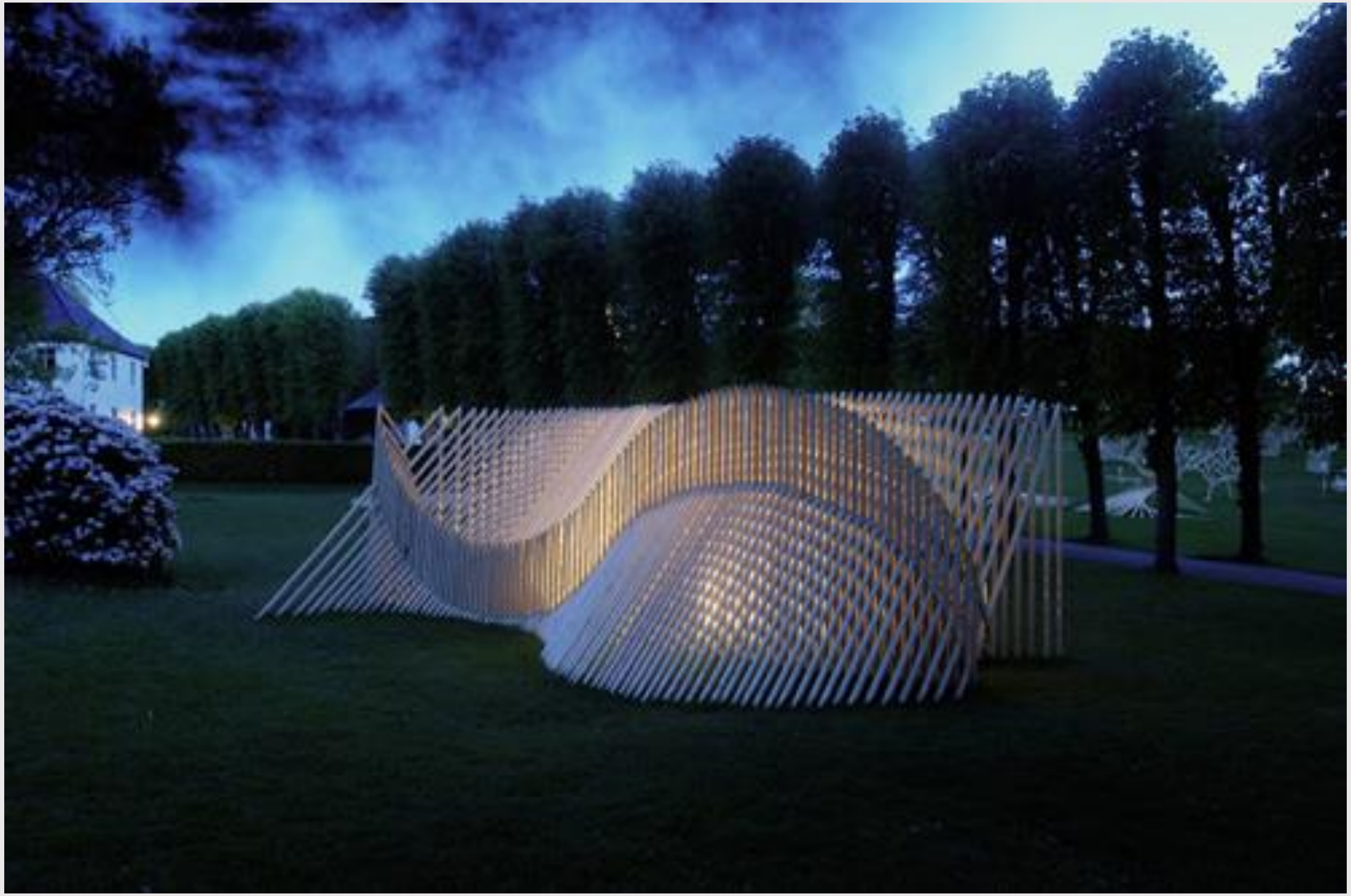


Spatial Polygamy*

*Matthews SA, Yang TC. *American Behavioral Scientist*. 2013.


Estimate spatial mechanisms
Map/describe spatial structure
Map composite, estimated data
Map observed data
Use GIS to link or merge data
Spatial Regression
Spatial Analysis
GIS & cartography

2. Spatial Structure

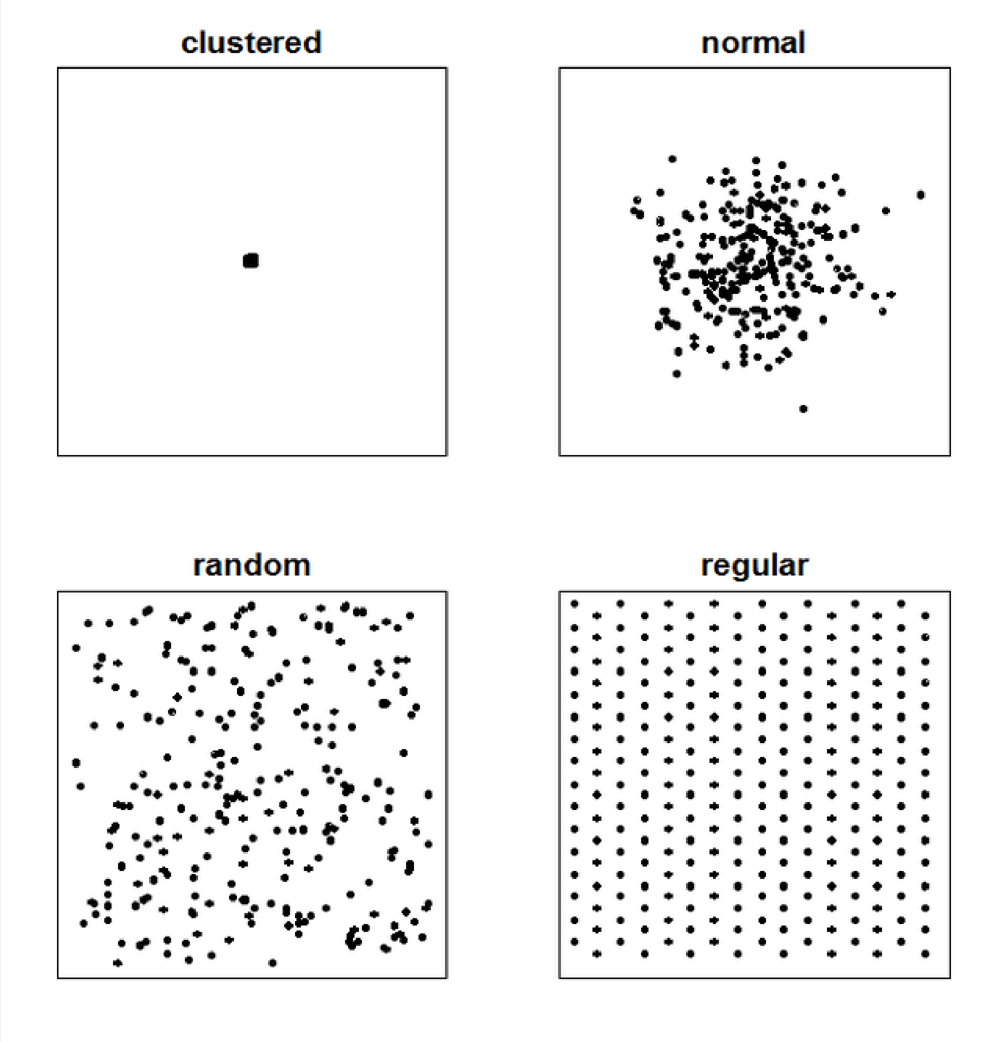


http://www.bergenwood.no/wp-content/media/images/snas.jpg

Is there spatial structure?

- What is 'structure'?

Departure from 'complete spatial randomness'



Is there spatial structure?

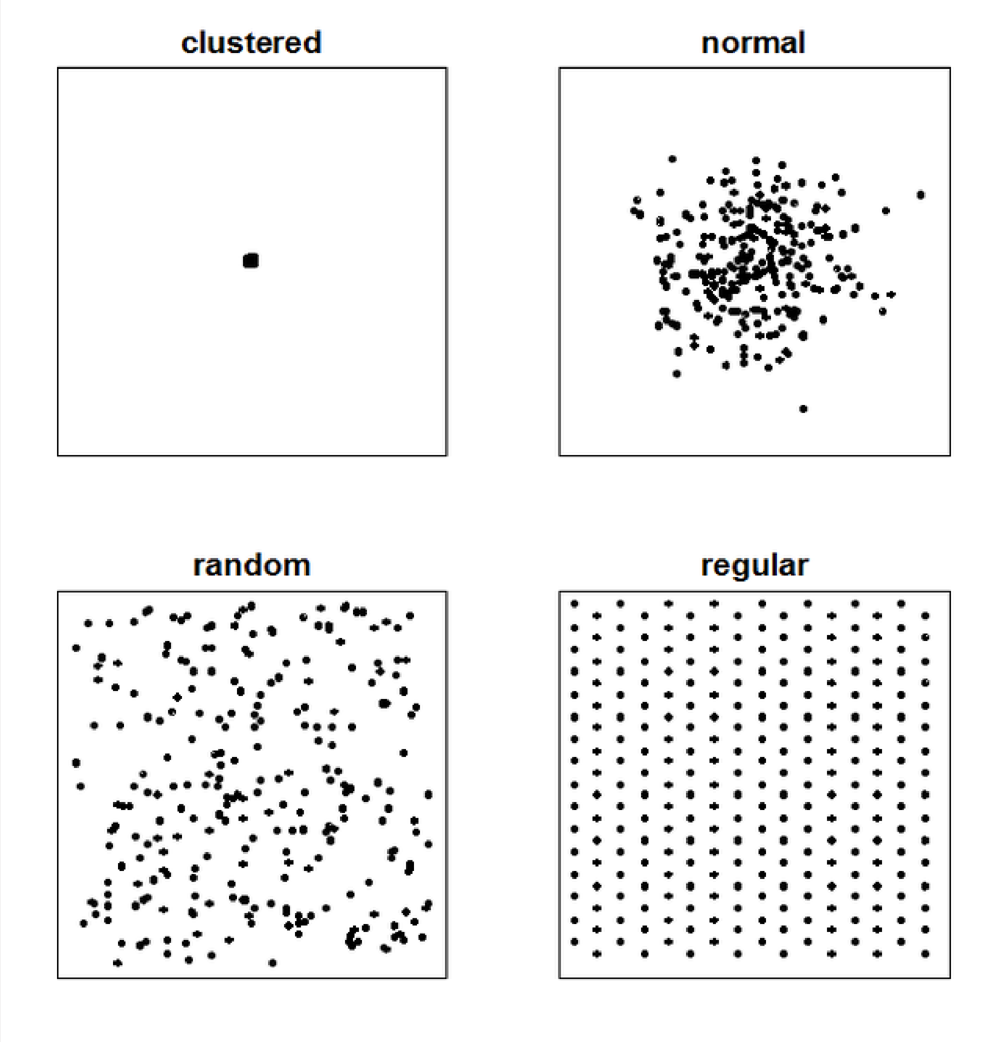
- What is 'structure'?

  Departure from 'complete spatial randomness'

- How can we measure it?

  Spatial autocorrelation

  "Are values for near objects more correlated than values of far objects?"



Tobler’s 1st Law of Geography

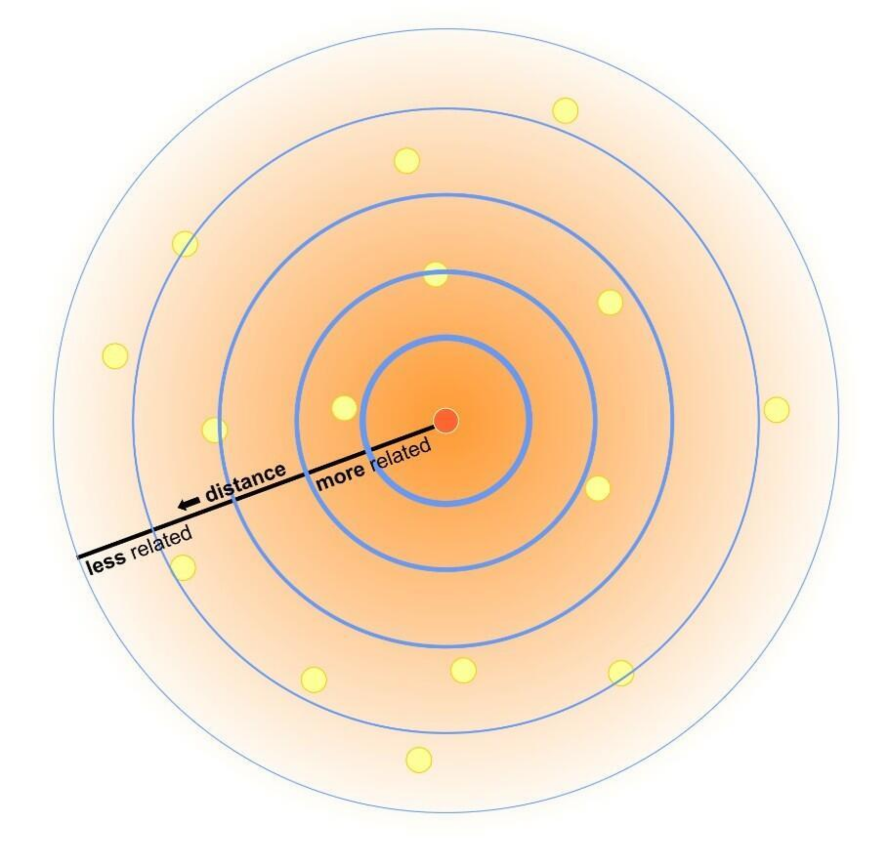


Photo Credit: Anthony C. Robinson

“Everything is related to everything else, but near things are more related than distant things.”*

* Tobler W., (1970) "A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46(2): 234-240.

Spatial Structure

Nuisance?

- Violation of assumptions of statistical independence
- Failure to account for correlation could produce biased estimates

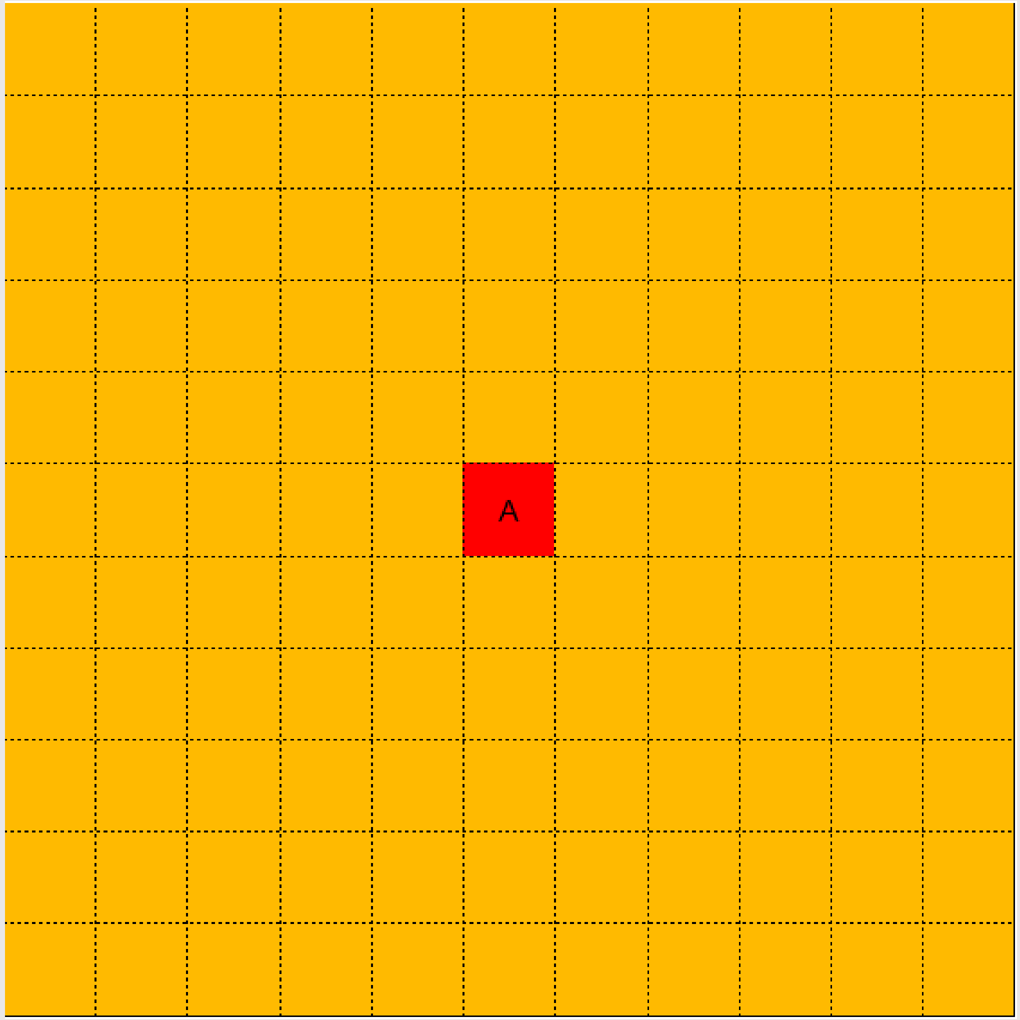
Feature?

- What is producing clustering?
  - Missing variables
  - Spatial interaction
  - Common source of exposure

How can you measure spatial structure?

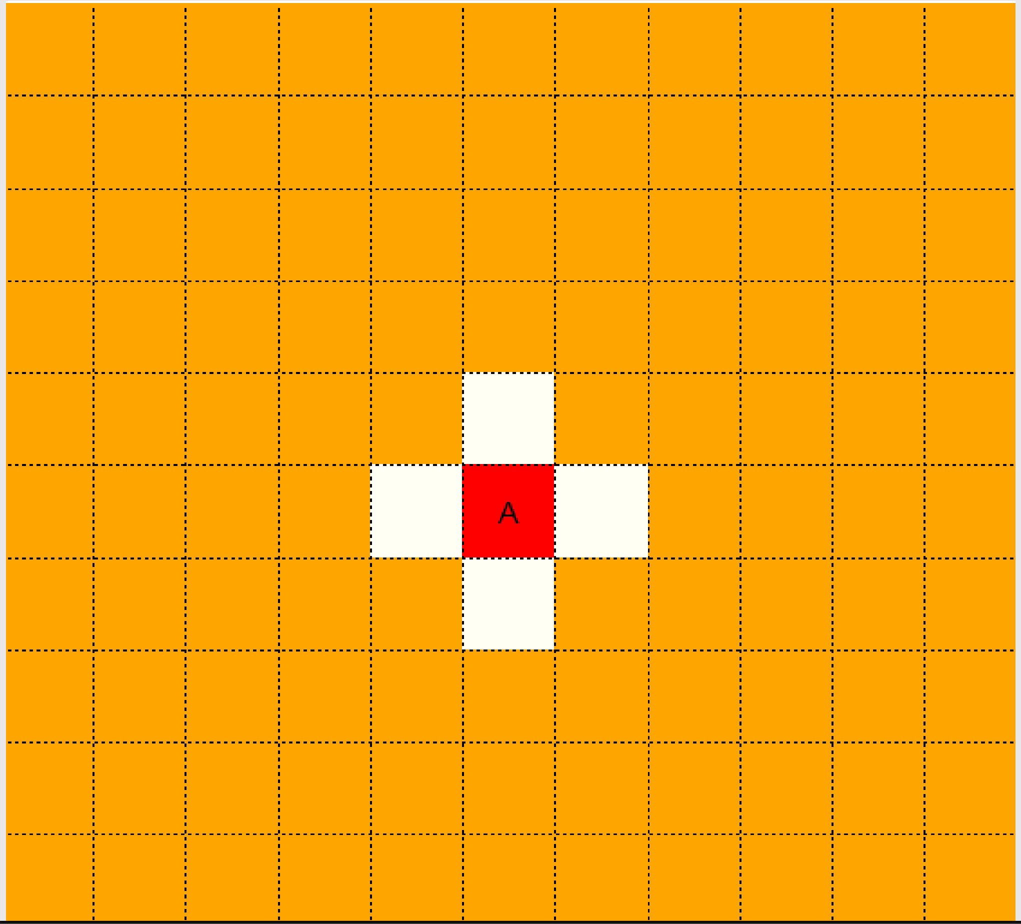
1. Define NEIGHBORS (who is close, who is far?)
2. Make spatial relationships mathematical with WEIGHTS

What is ‘close’ and who is a ‘neighbor’?

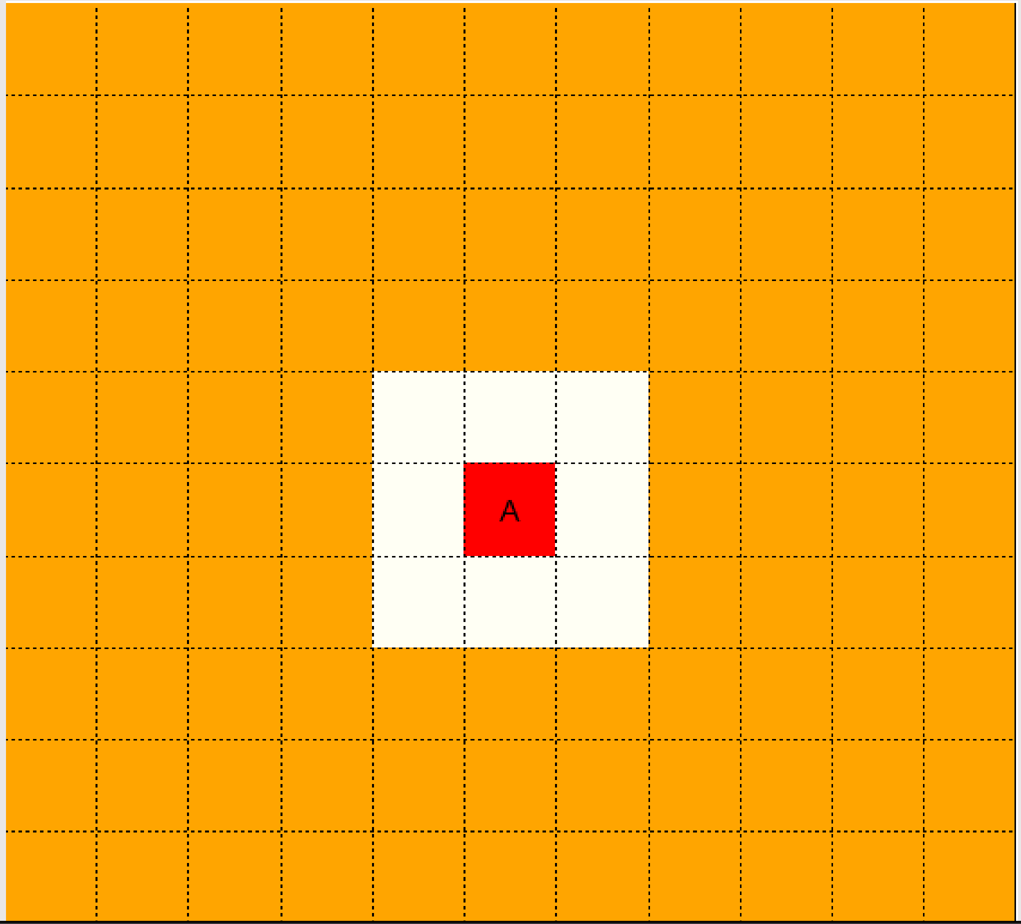


Who are the neighbors of A?

Edge (Rook) contiguity

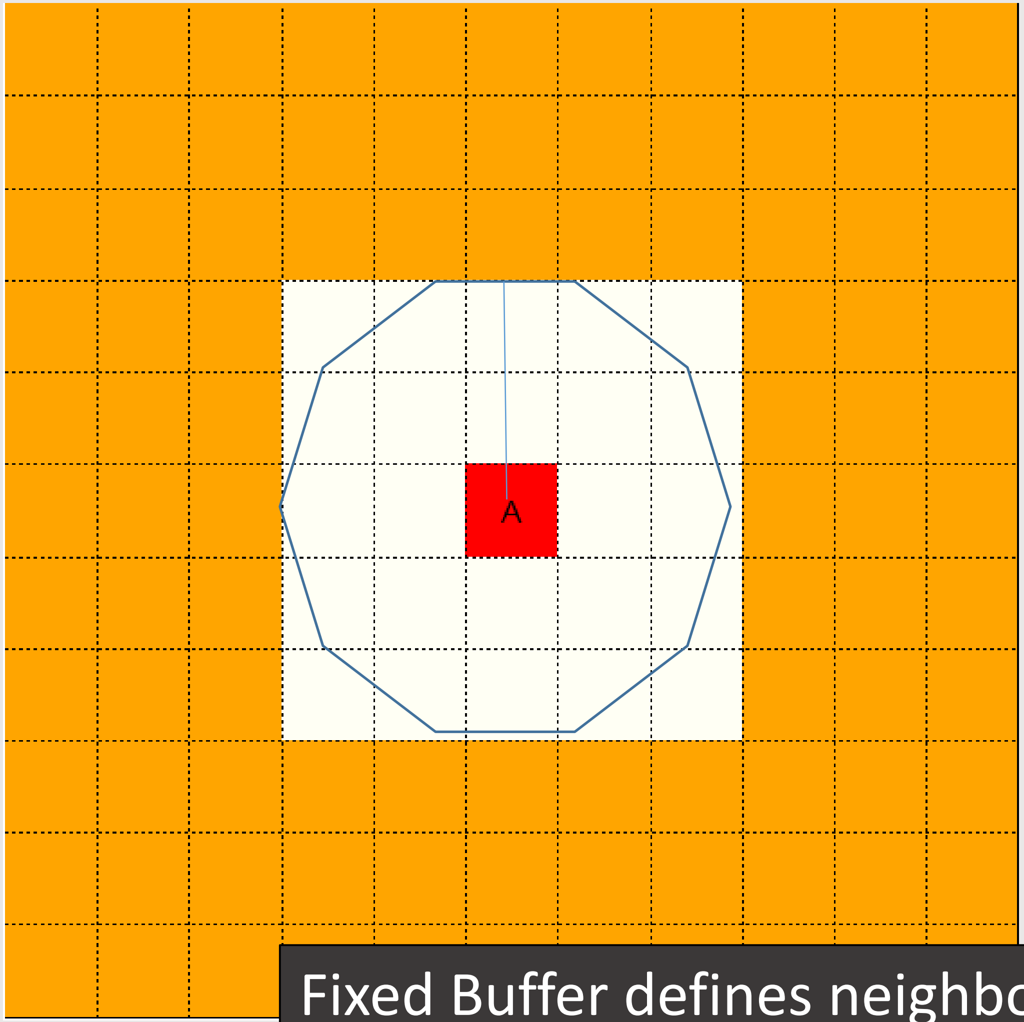


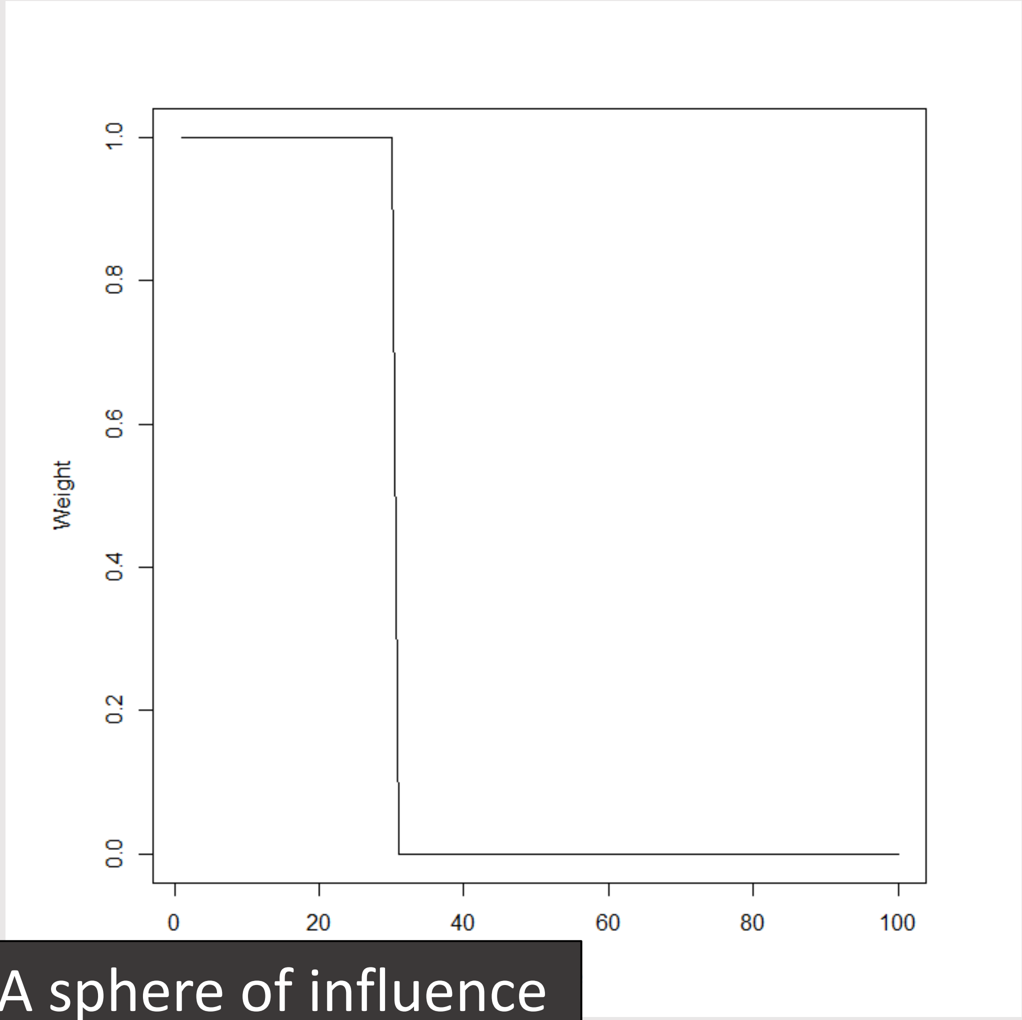
Edge + Corner (Queen) contiguity



Sharing boundaries implies interaction with immediate neighbors, diffusion, spillover

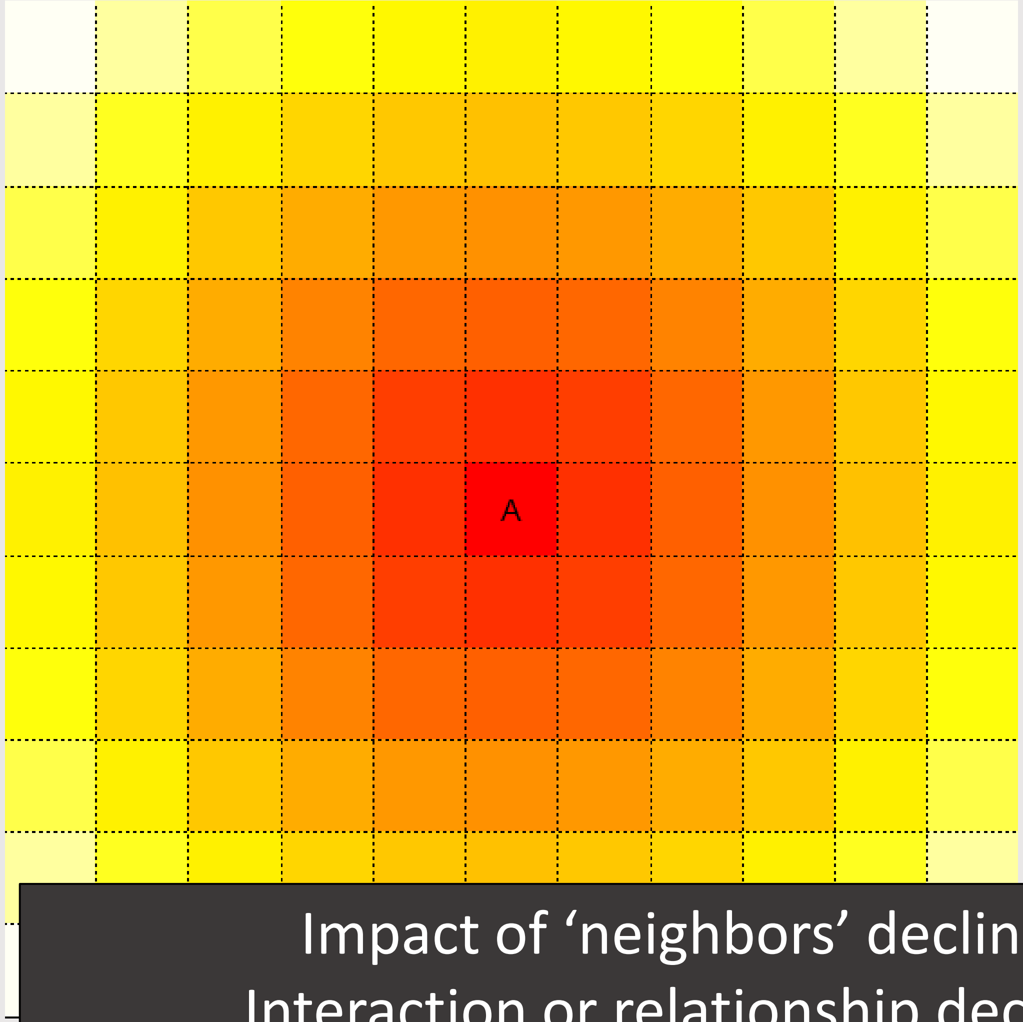
Fixed Distance (Buffer)

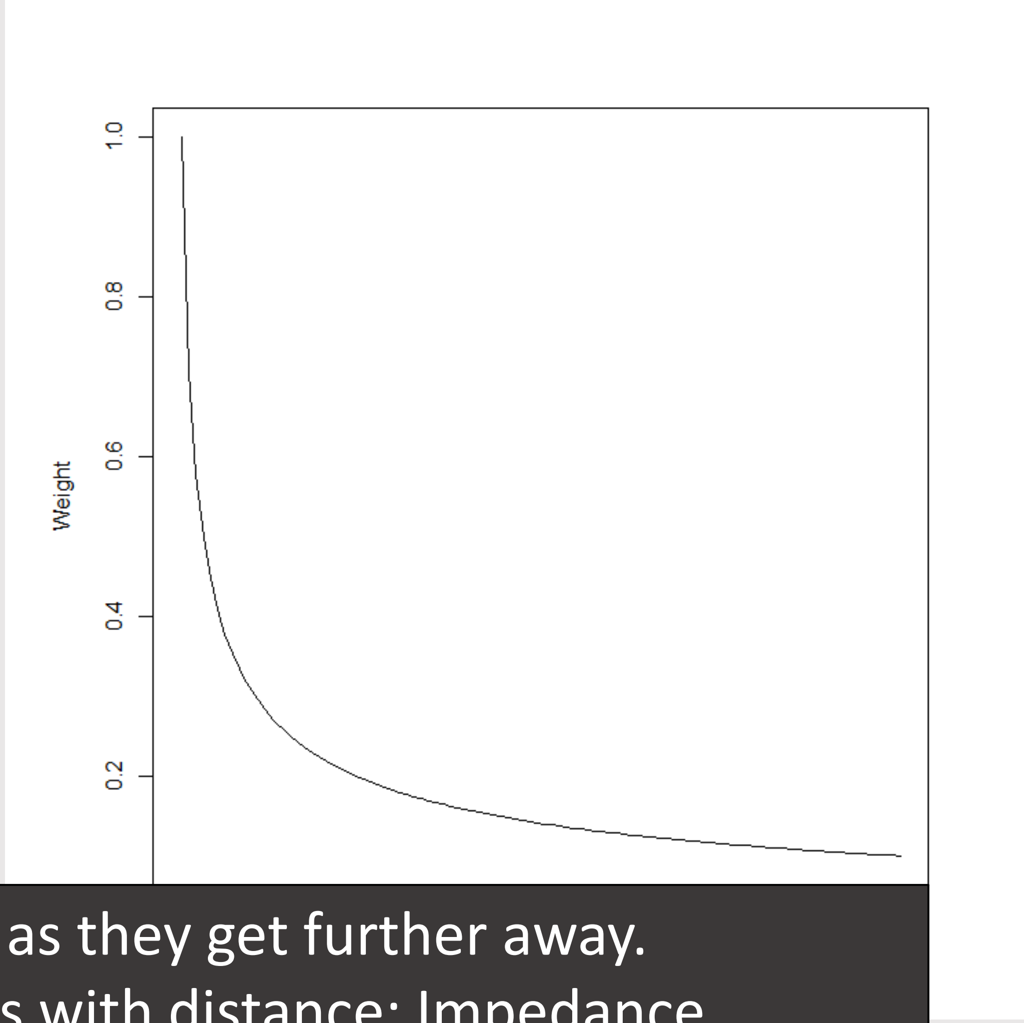




Fixed Buffer defines neighbors; A sphere of influence

Inverse Distance Weight





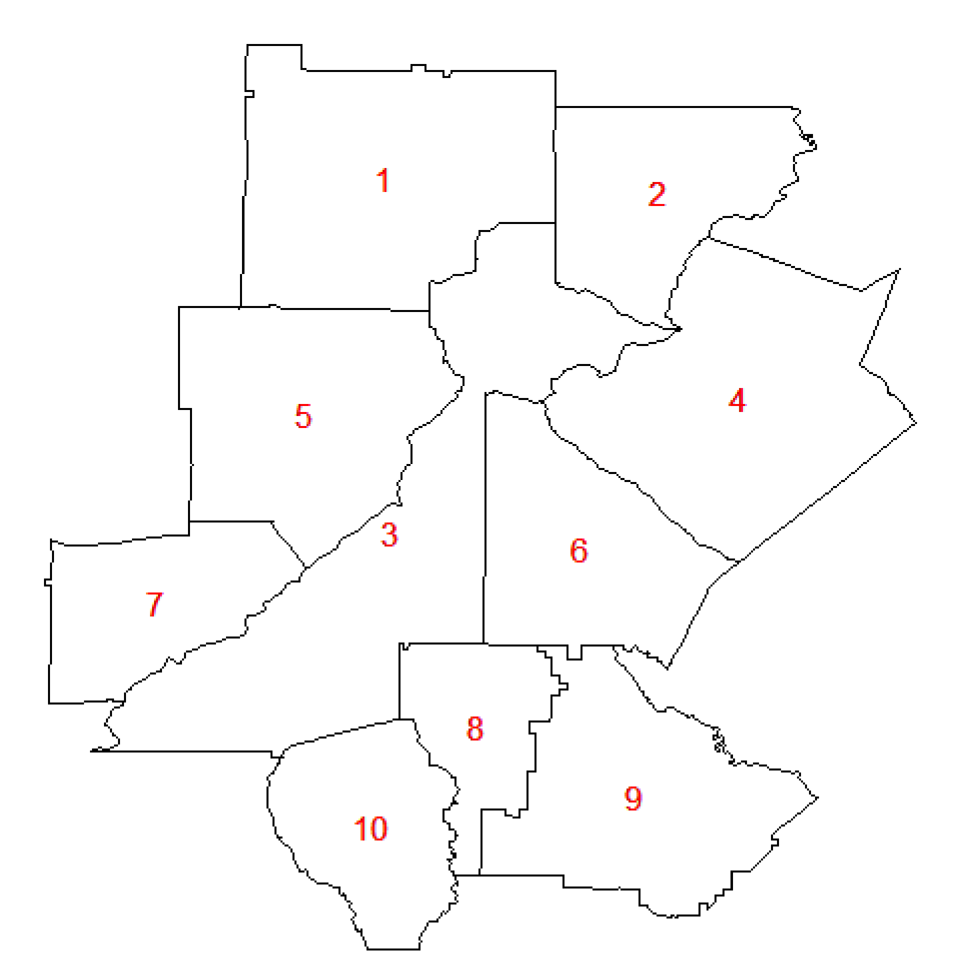
Impact of ‘neighbors’ declines as they get further away.
Interaction or relationship decays with distance; Impedance

Defining Spatial Weights

- Binary weights (contiguity and buffer)
  - Neighbors = 1
  - Non-neighbors = 0
  - Row-standardized: weights of all neighbors sum to 1

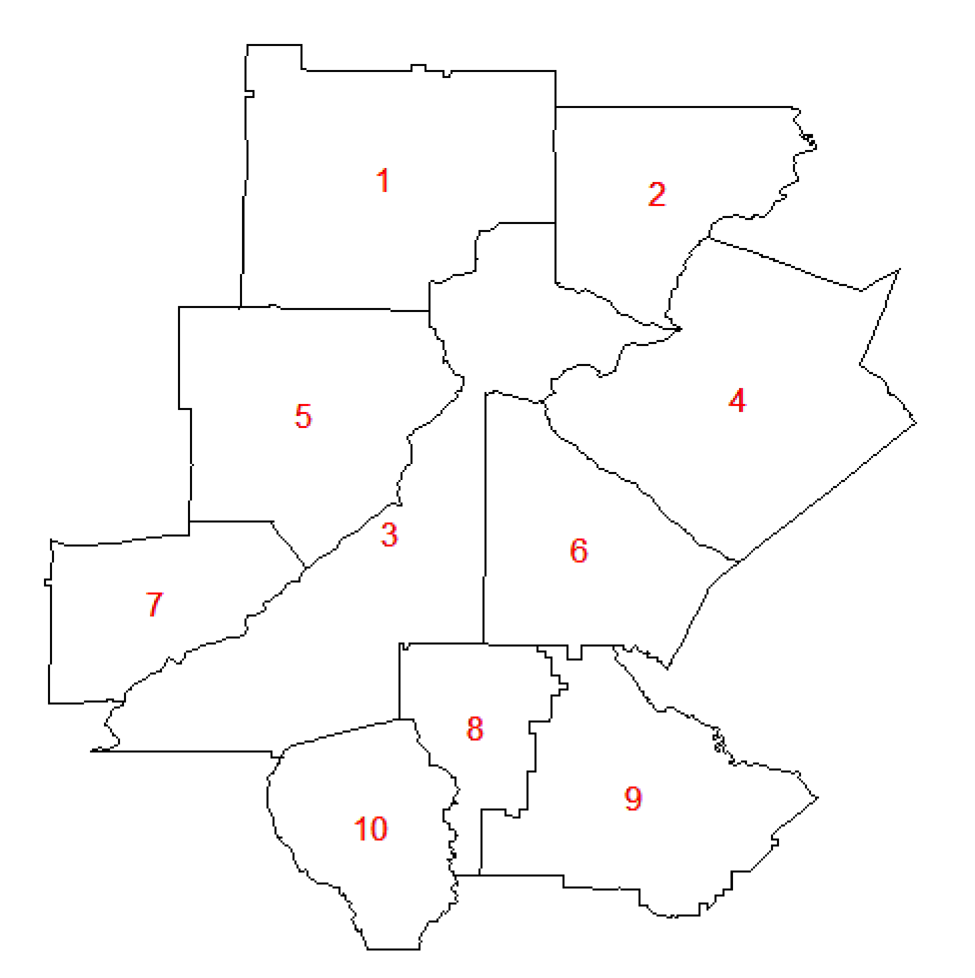
- Continuous weights (Inverse Distance)
  - Weight is a function of distance from index location
  - Weight (neighborliness) declines with distance

Queen contiguity matrix (binary)



	1	2	3	4	5	6	7	8	9	10
1	**0**	1	1	0	1	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	1	1	0	1	1	1	1	1	0	1
4	0	1	1	0	0	1	0	0	0	0
5	1	0	1	0	0	0	1	0	0	0
6	0	0	1	1	0	0	0	1	1	0
7	0	0	1	0	1	0	0	0	0	0
8	0	0	1	0	0	1	0	0	1	1
9	0	0	0	0	0	1	0	1	0	0
10	0	0	1	0	0	0	0	1	0	0

Queen contiguity matrix (row-standardized)



	1	2	3	4	5	6	7	8	9	10
1	0	0.33	0.33	0	0.33	0	0	0	0	0
2	0.33	0	0.33	0.33	0	0	0	0	0	0
3	0.13	0.13	0	0.13	0.13	0.13	0.13	0.13	0	0.13
4	0	0.33	0.33	0	0	0.33	0	0	0	0
5	0.33	0	0.33	0	0	0	0.33	0	0	0
6	0	0	0.25	0.25	0	0	0	0.25	0.25	0
7	0	0	0.5	0	0.5	0	0	0	0	0
8	0	0	0.25	0	0	0.25	0	0	0.25	0.25
9	0	0	0	0	0	0.5	0	0.5	0	0
10	0	0	0.5	0	0	0	0	0.5	0	0

From NEIGHBORS…

To WEIGHTS…

To TESTS of spatial autocorrelation

What is this?

Product of difference (of X and Y) from mean
This is the SIMILARITY between points

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}\,(X_i - \bar{X})^2}\,\sqrt{\sum_{i=1}^{n}\,(Y_i - \bar{Y})^2}}$$

Product of the SD

Pearson's Correlation Coefficient

What is this?

Similarity between two measures of Y

$$I = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\,(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}}\left(\frac{1}{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2}\right)$$

Weights

SD

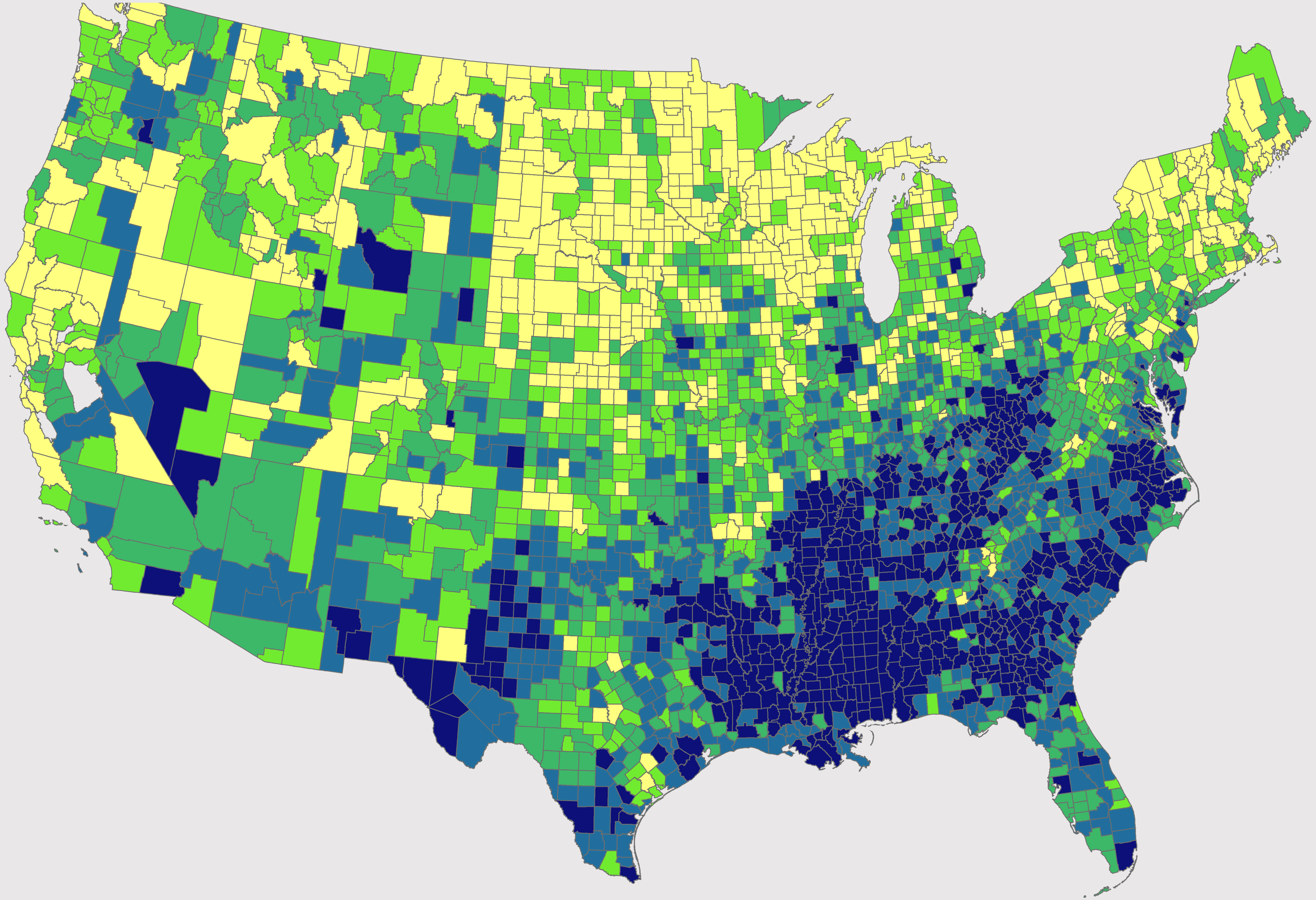
Moran’s I

Moran’s I

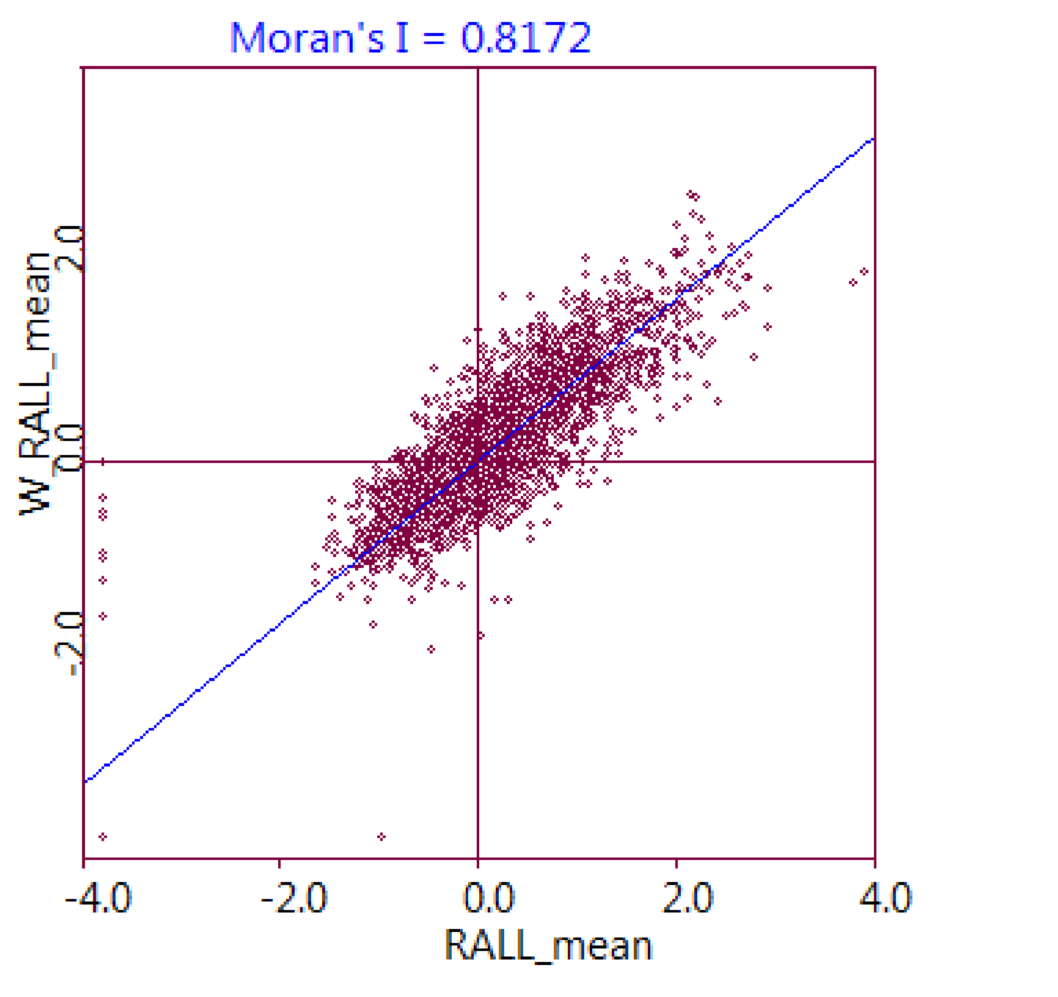
- Similar to Pearson: compares the similarity of values
- Different from Pearson: includes a spatial weight

- Range from -1 to 1 (mostly)
  - Positive values: neighboring observations are more similar (clustered)
  - Negative values: neighbors are different (dispersed or regularly patterned)
  - Zero: no relationship between self and neighbors

Preterm birth risk, total population, 2004-2006



Moran's I scatter plot



- X-axis: index place value (PTB %)
- Y-axis: the weighted average of the 'neighbors' PTB % (e.g. lagged)

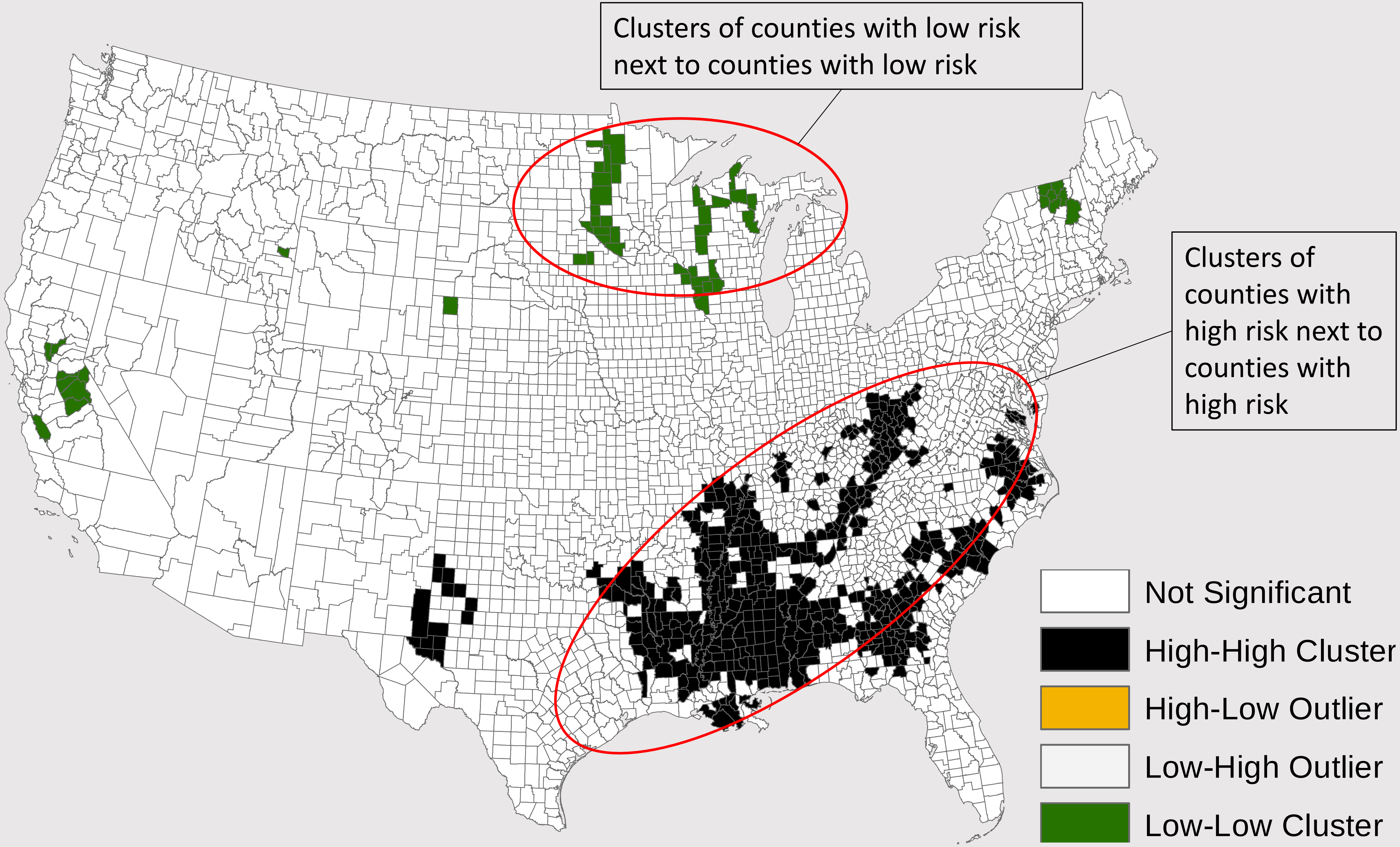
Global versus Local Spatial Autocorrelation

Global structure

- Tests null hypothesis that the entire dataset is *completely spatially random*
- Requires specification of spatial weights
- Single test statistic (e.g. global Moran's I)

Local structure

- Tests null that local values are spatially randomly distributed
- Requires specification of spatial weights
- Separate test statistic for every unit (e.g. local Moran's I or LISA)


Clusters of counties with low risk next to counties with low risk
Clusters of counties with high risk next to counties with high risk
Not Significant
High-High Cluster
High-Low Outlier
Low-High Outlier
Low-Low Cluster

Other measures of spatial clustering

- Moran’s I
  - Global: do the data depart from complete spatial randomness
  - Local Indicators of Spatial Association (LISA): where are high/low clusters?
- Getis-Ord G statistic
  - Global measure of clusters or either high or low values
  - G-i* statistic – statistic for mapping local clusters
- Ripley’s K-function
  - Cluster analysis at multiple distances

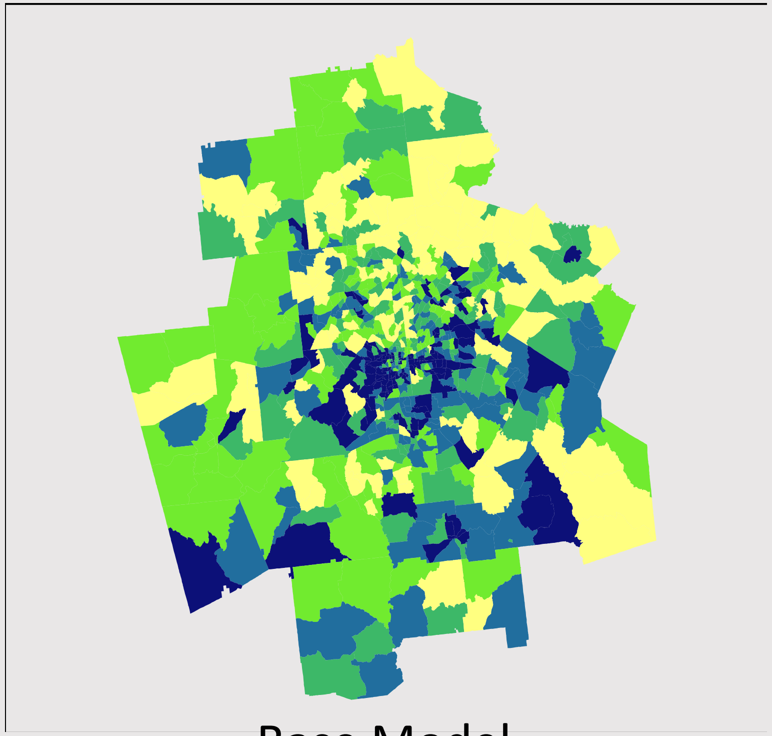
What does clustering mean?

- **Selection**
  - People with high (low) risk select into or are constrained within some areas
- **Causal exposure varies spatially**
  - Causal characteristics (social, environmental) are unevenly distributed in space
- **Spurious**
  - Type I error from multiple comparisons (can use false discovery rate or Monte Carlo simulation of null distribution)
  - Modifiable Areal Unit Problem (MAUP): Ecologic fallacy from choice of scale

Extension of autocorrelation: clustered regression residuals?

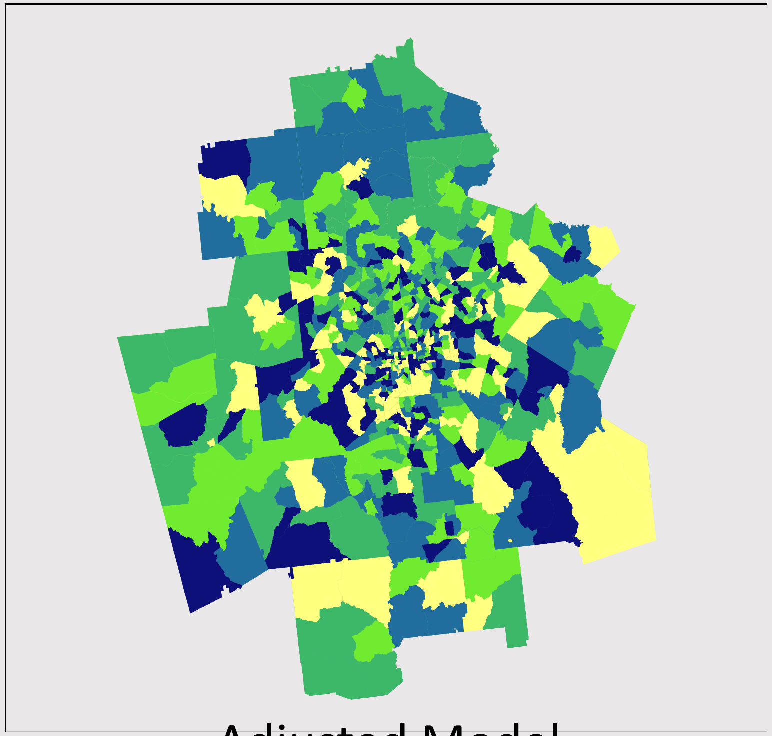
- Map model output: residuals, predicted risks

Moran’s I: 0.29



Base Model

Moran’s I: -0.01



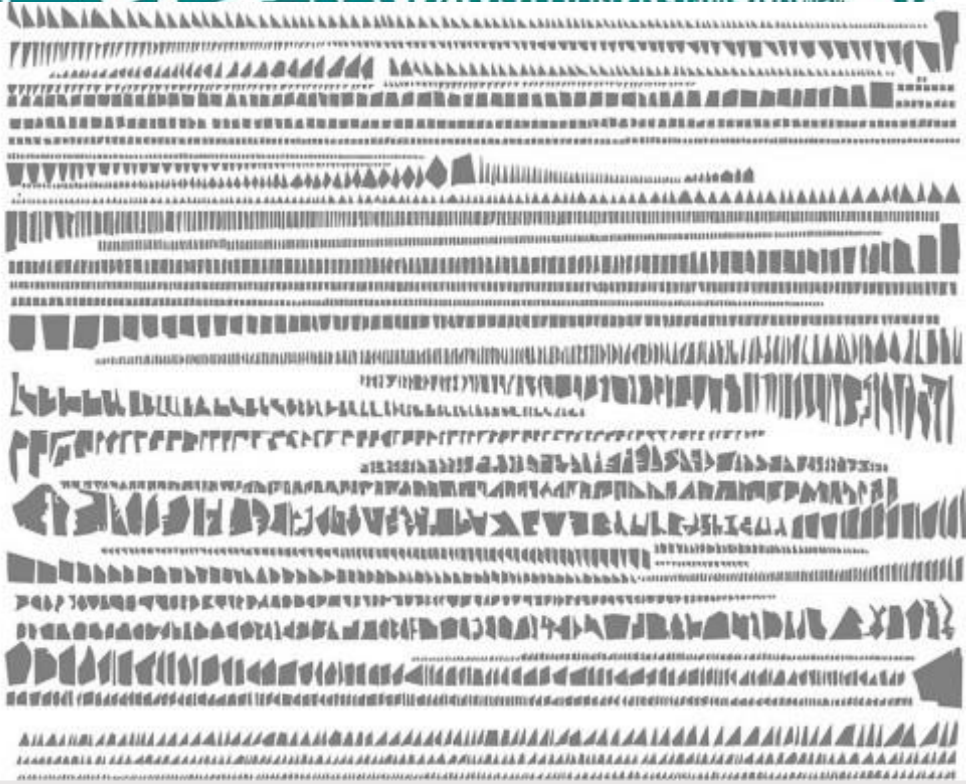
Adjusted Model

3. Discrete or continuous space?

Scale, zoning, and surfaces

New York City

Paris



http://www.armellecaron.fr/works/les-villes-rangees/

CASE #2 – Space or Place?

1. When is a ‘place’ a container and when is the spatial orientation of multiple places important?

2. When do boundaries matter for health?

Modifiable Areal Unit Problem: MAUP

- Ecologic fallacy
  - Inference from aggregate may not apply to individual
- Problem when…
  - Boundaries are not naturally tied to process of interest
- Bias from arbitrary…
  - Zoning (location, shape of boundaries)
  - Scale (level of aggregation)

Total Births

100	100	400
200	300	300
500	400	100

Preterm Births

10	20	10
50	10	5
60	10	20

Preterm Birth %

10%	20%	3%
25%	3%	2%
12%	3%	20%

Neighborhoods within a town

Adapted from: http://www.openmedicine.ca/article/view/286/359

Total Births

100	100	400
200	300	300
500	400	100

Preterm Births

10	20	10
50	10	5
60	10	20

Preterm Birth %

10%	20%	3%
25%	3%	2%
12%	3%	20%

MAUP:
Scale Effect

7%
8%
9%

Total Births

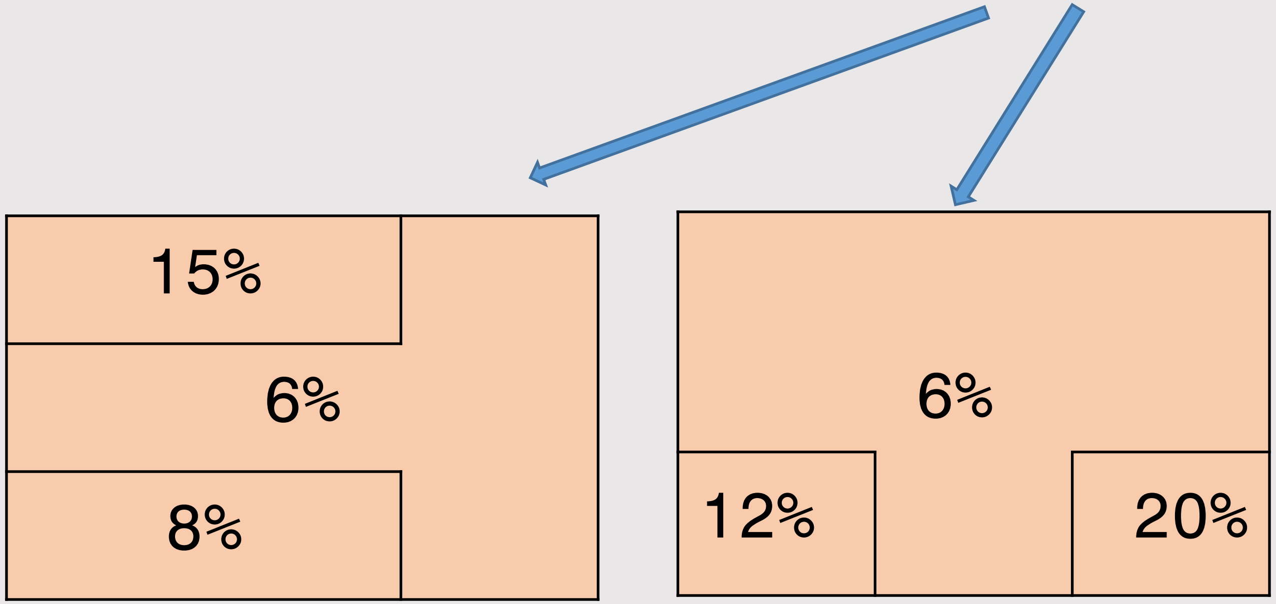
100	100	400
200	300	300
500	400	100

Preterm Births

10	20	10
50	10	5
60	10	20

Preterm Birth %

10%	20%	3%
25%	3%	2%
12%	3%	20%



MAUP: Zoning Effect

MAUP: Solutions

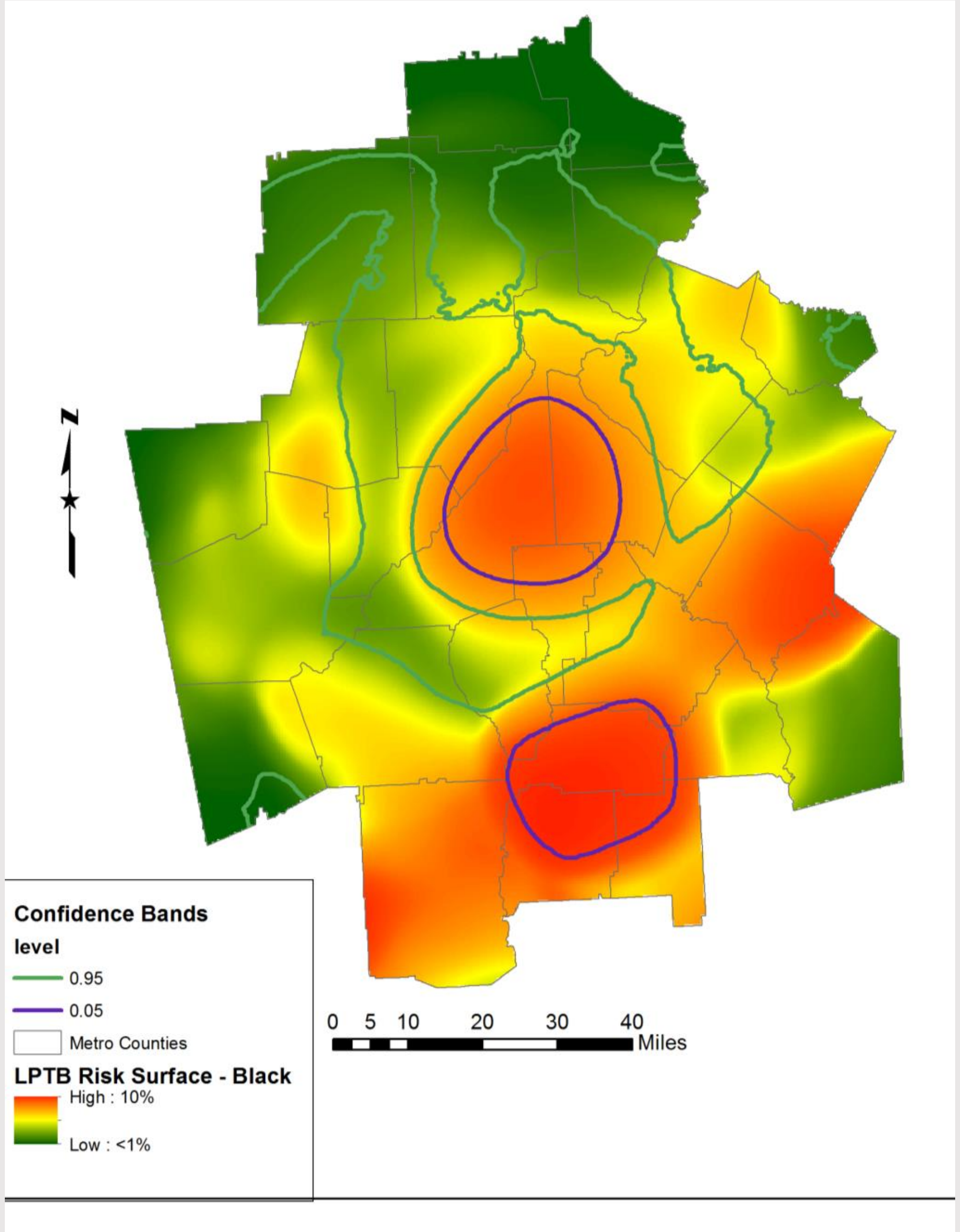
1. Choose meaningful boundaries
2. Multi-scale analysis
   - Repeated for blocks, block groups, tracts, counties
3. Autozoning algorithms
   - AZTool (http://www.geodata.soton.ac.uk/software/AZTool/)
   - Aggregate sub-areas into new meta-zones as a function of composition (income, race/ethnicity)
4. Convert to surfaces and use egocentric 'neighborhoods'

Spatial risk surfaces

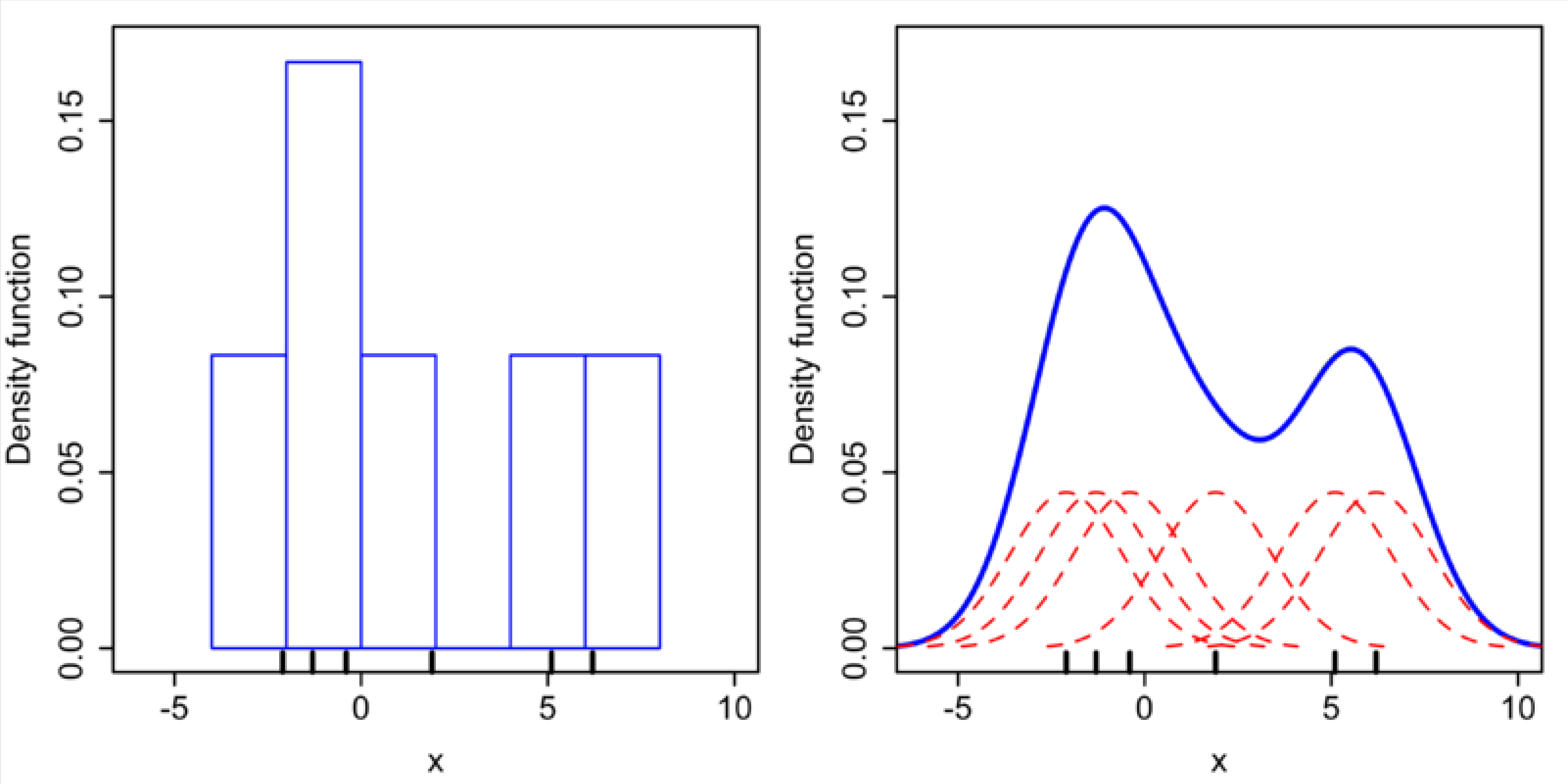
- Spatial intensity/density
  - "1st order process"
  - Local mean or average per unit area

$$\lambda(x,y) = \frac{n}{|A|}$$

  - λ(x,y): intensity at location x,y
  - n: # events
  - |A|: area
- Start with POINTS (or convert polygons to points)

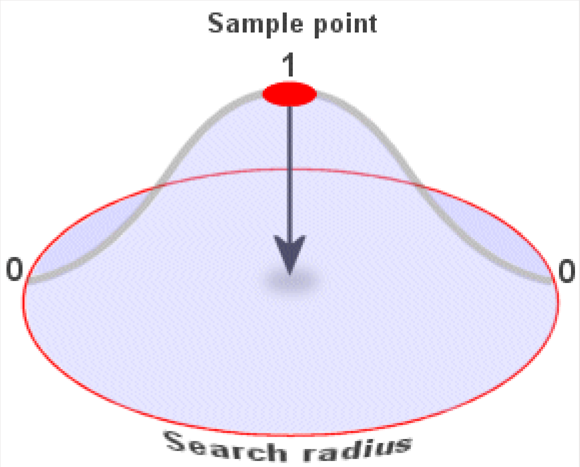


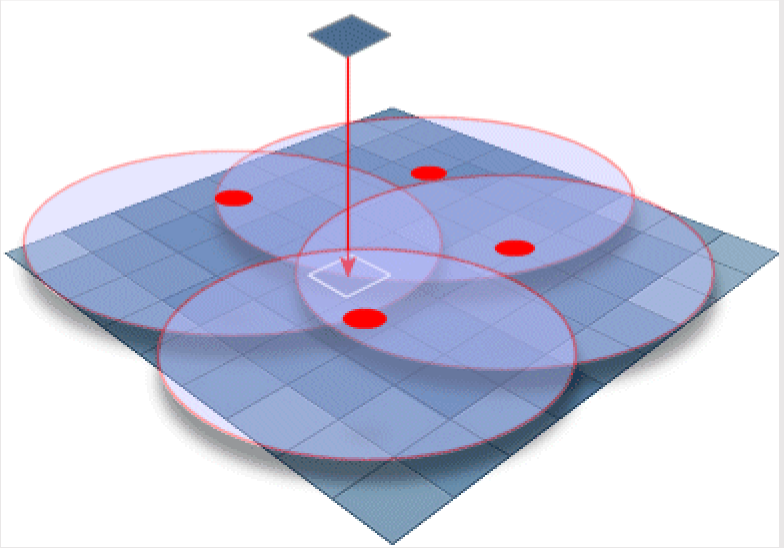
Kernel Density Estimators



http://en.wikipedia.org/wiki/File:Comparison_of_1D_histogram_and_KDE.png

Spatial KDE





http://www.geography.hunter.cuny.edu/~jochen/GTECH361/lectures/lecture11/concepts/Kernel%20density%20calculations.htm

4. Spatial scale:

How big IS a neighborhood?

Scale problem

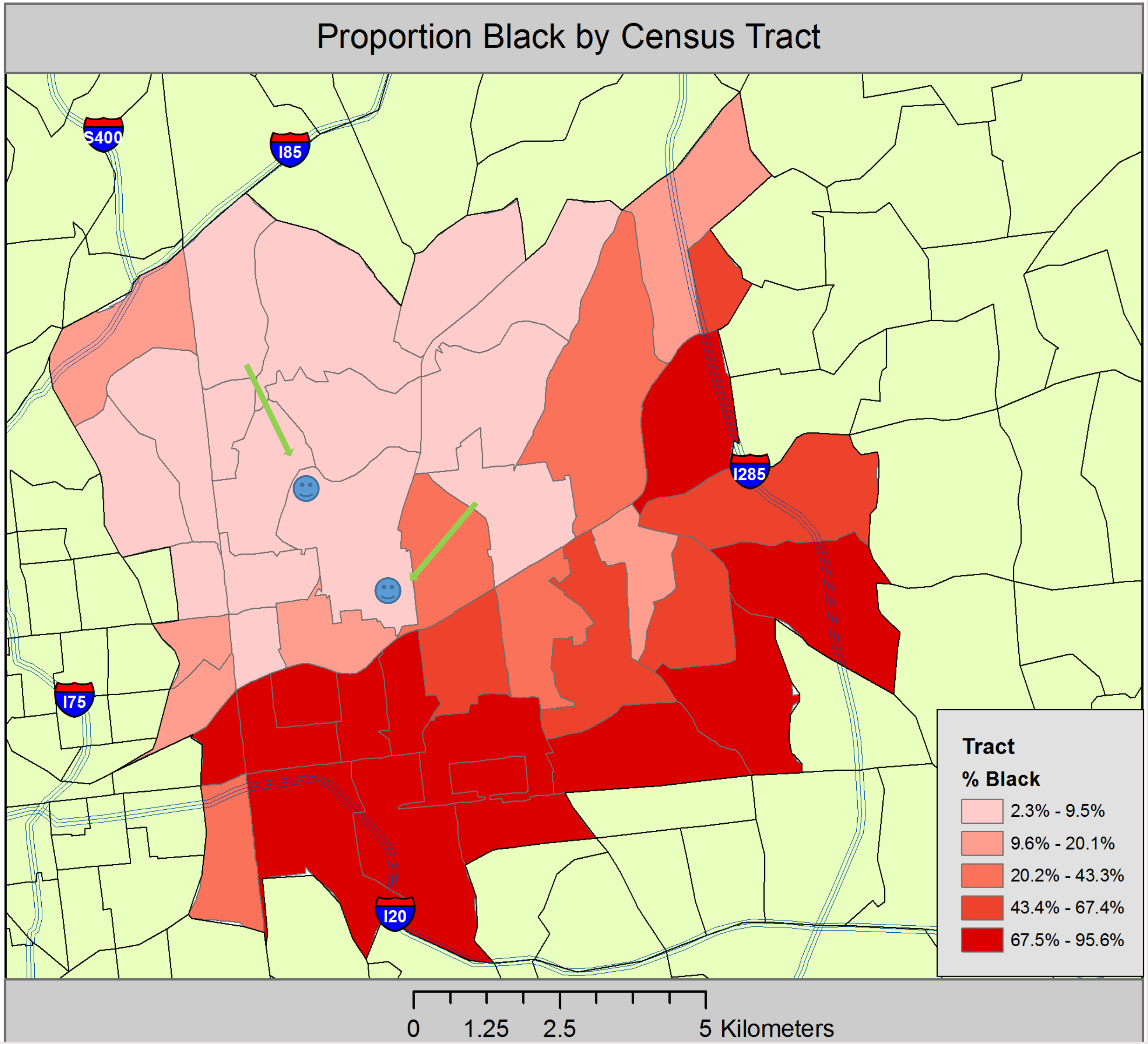
- What is the RIGHT scale to measure
  - Neighborhood health?
  - Neighborhood opportunity?
  - Humans are spatially polygamous

Small	Medium	Large
Violent crime, safety Housing quality Exposure to toxic release	Green space/park access Food environment Primary care health access	Employment opportunity Secondary education Specialty care health access

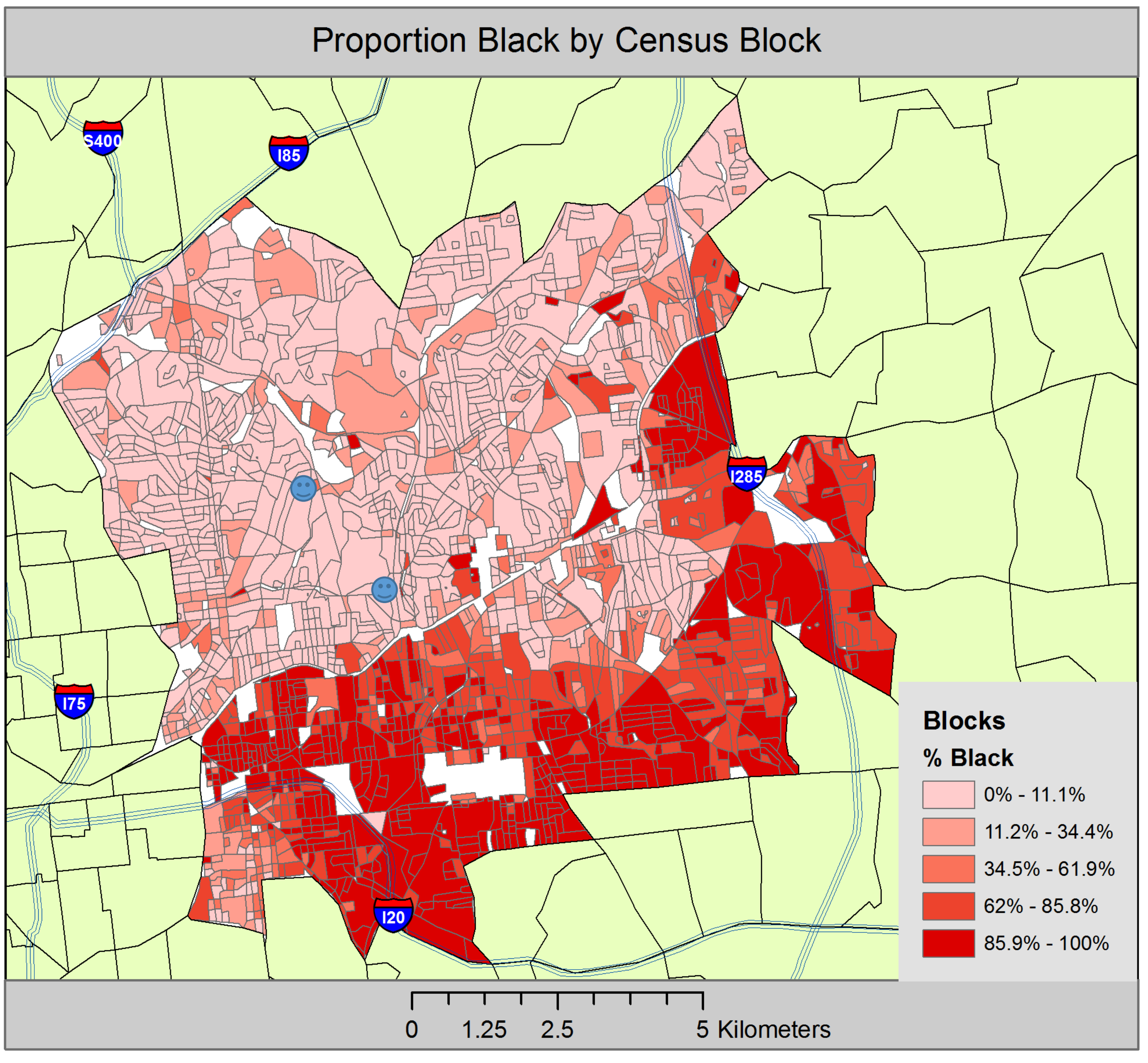
Matthews, 2013

Egocentric exposure measures

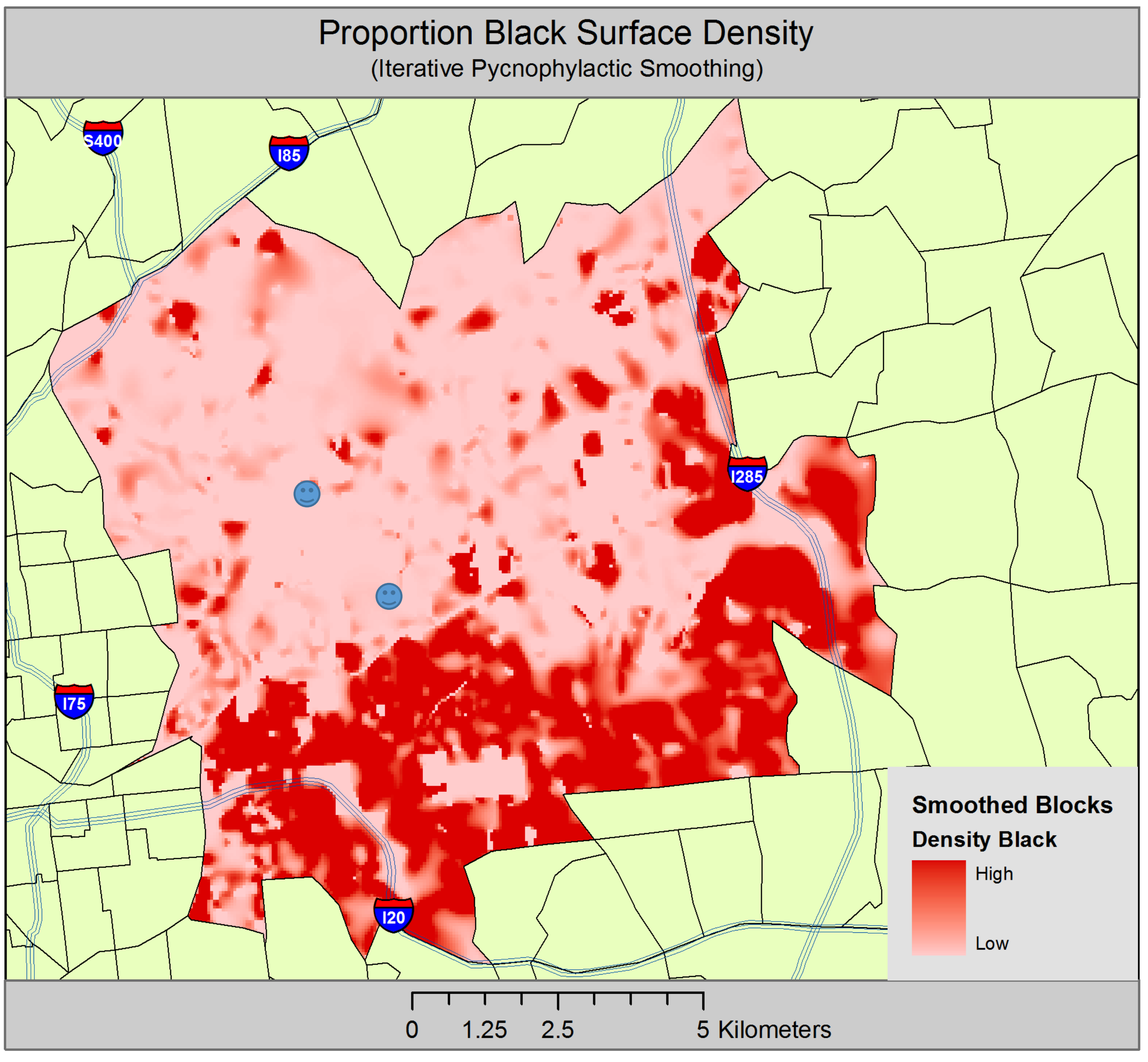
1. **Input**: highest resolution data available
   - Points, census block, block groups
2. **Convert**: polygon to regularly gridded points
   - Each point represents density of event per grid point
3. **Calculate**: egocentric neighborhoods
   - Kernel density estimators with varying bandwidth
   - Minimum scale depends on resolution of base data
4. **Contrast multiple scales**: How does pattern change with scale?



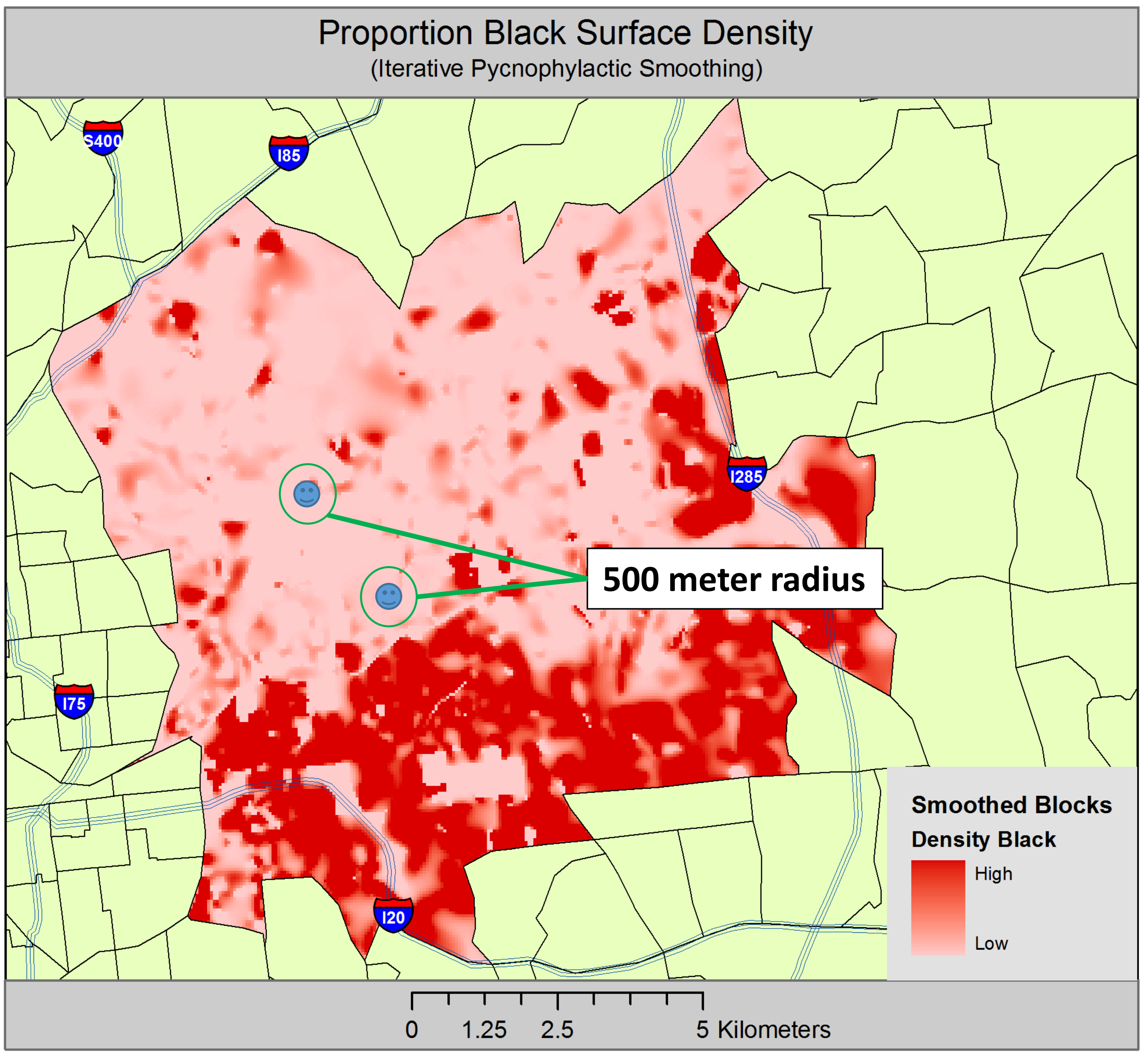
Reardon, O'Sullivan, et al, 2004
Kramer, et al, 2010



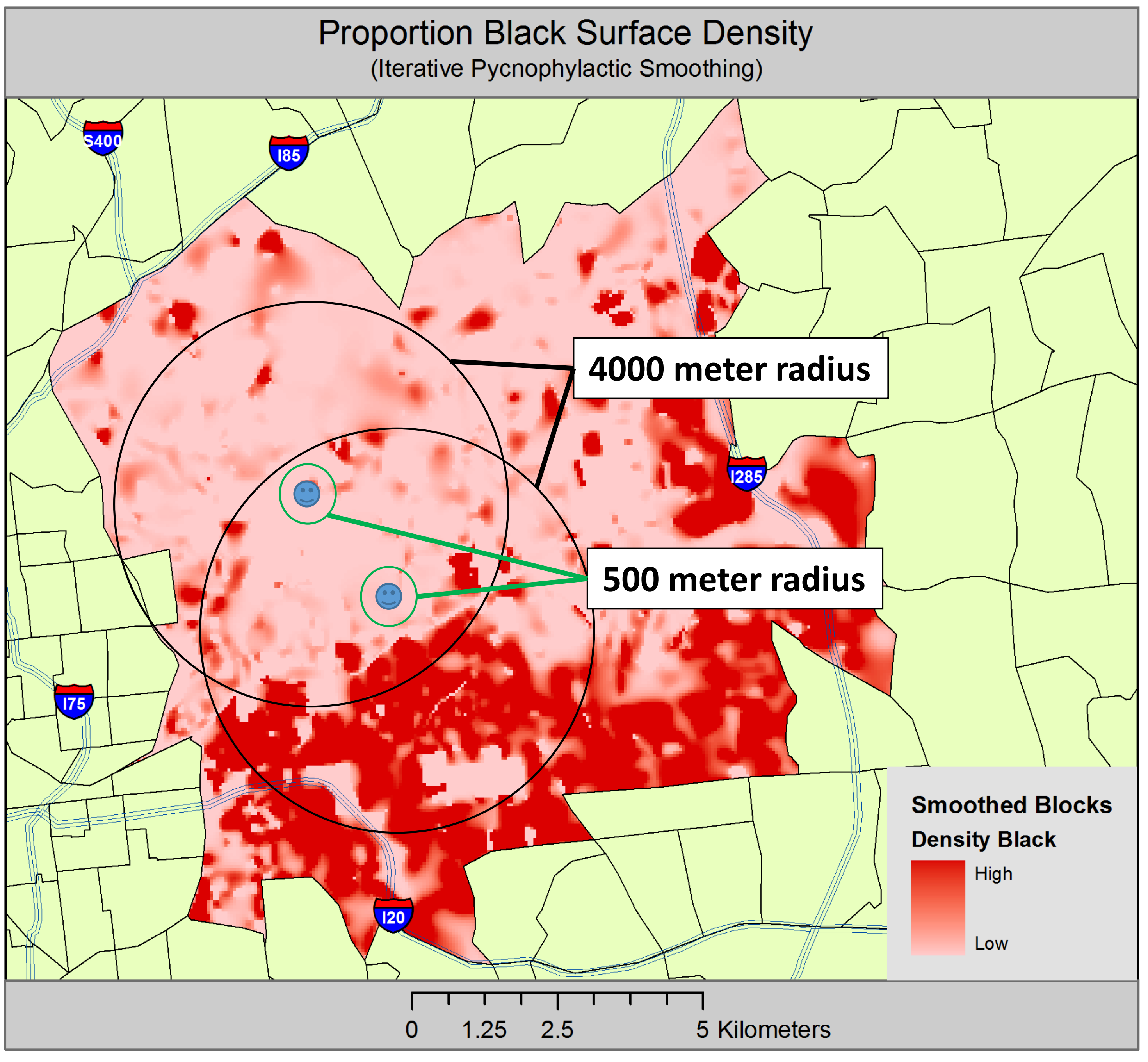
Reardon, O'Sullivan, et al, 2004
Kramer, et al, 2010



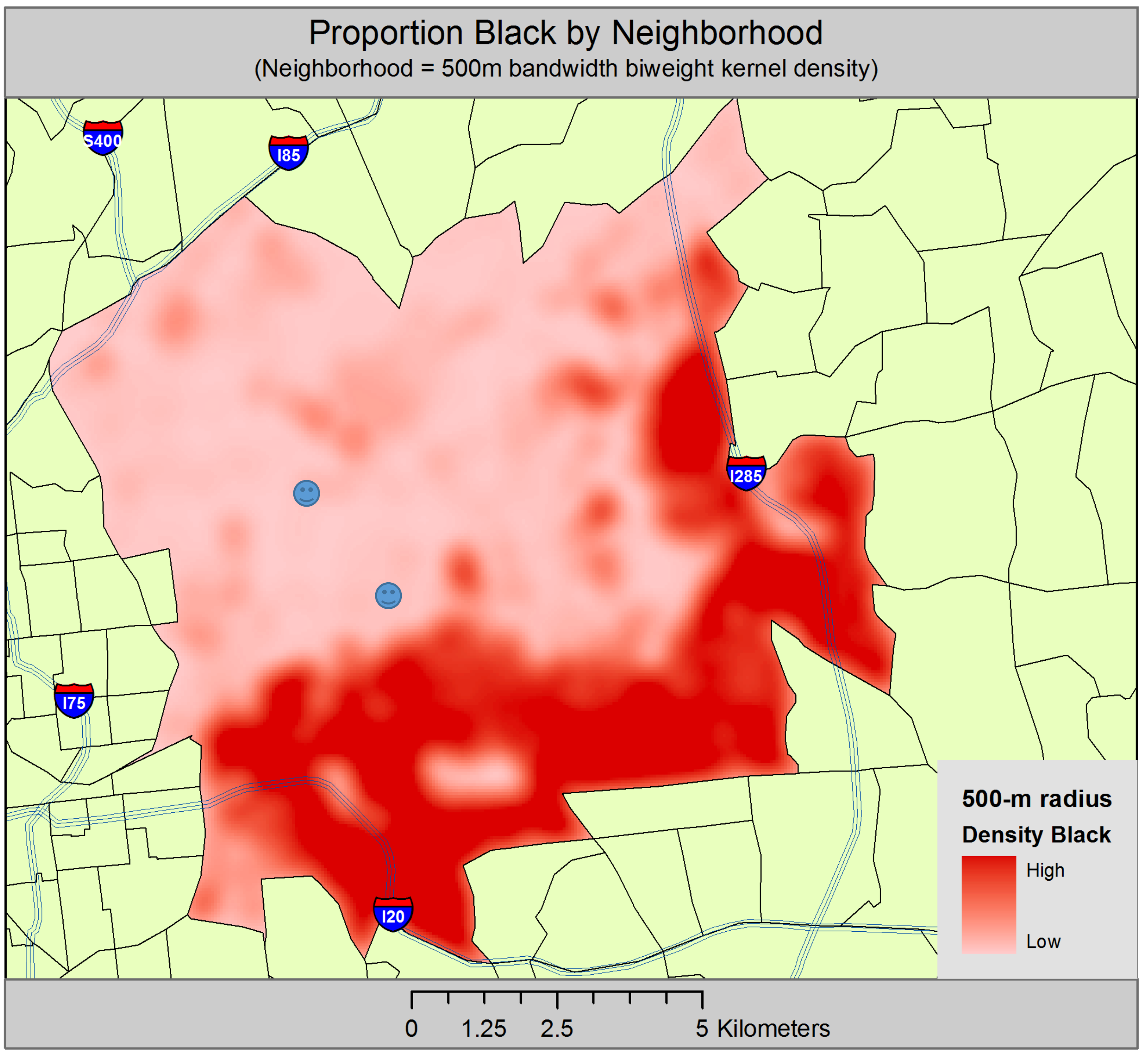
Reardon, O'Sullivan, et al, 2004
Kramer, et al, 2010



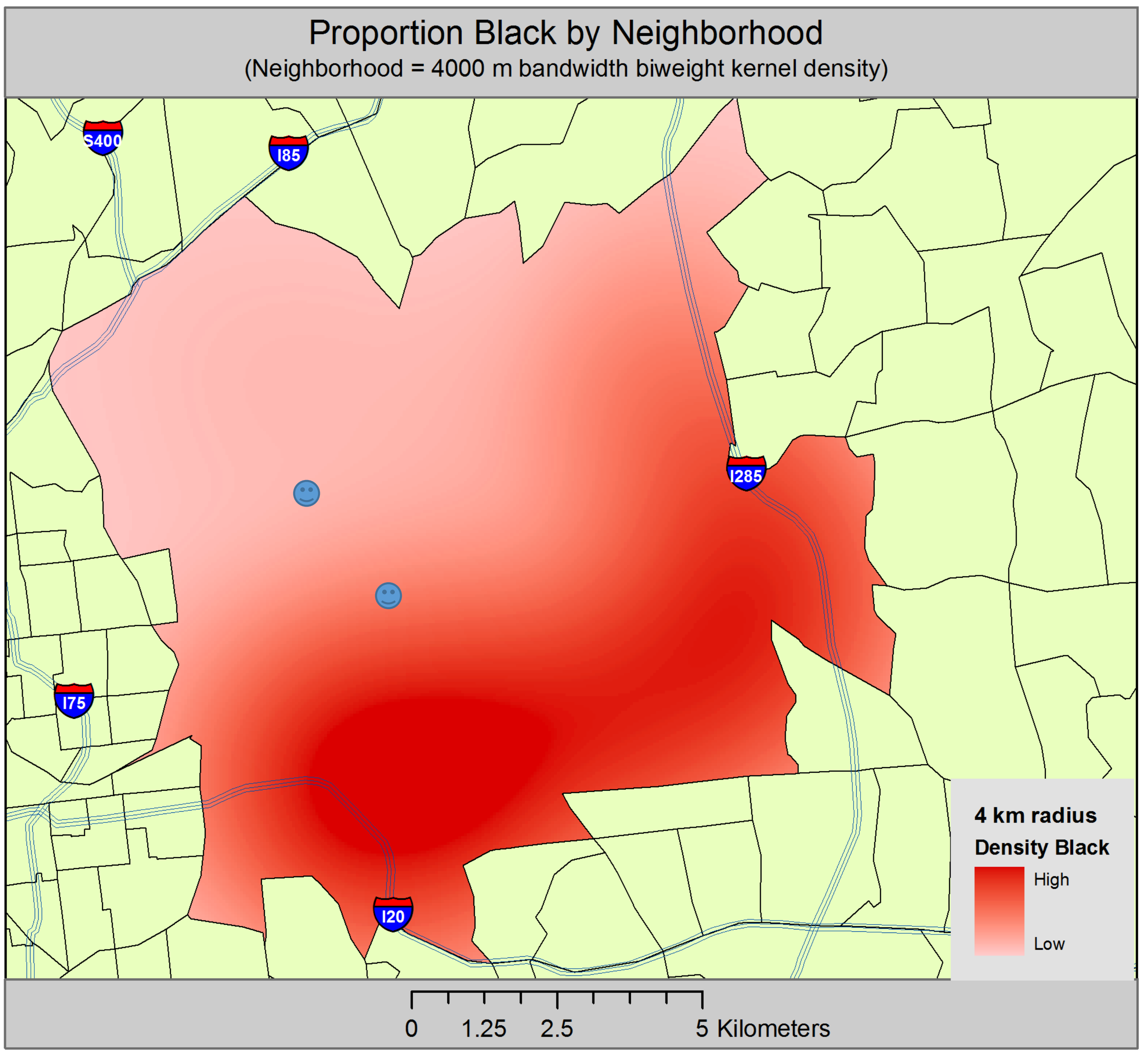
Reardon, O'Sullivan, et al, 2004
Kramer, et al, 2010



Reardon, O’Sullivan, et al, 2004
Kramer, et al, 2010

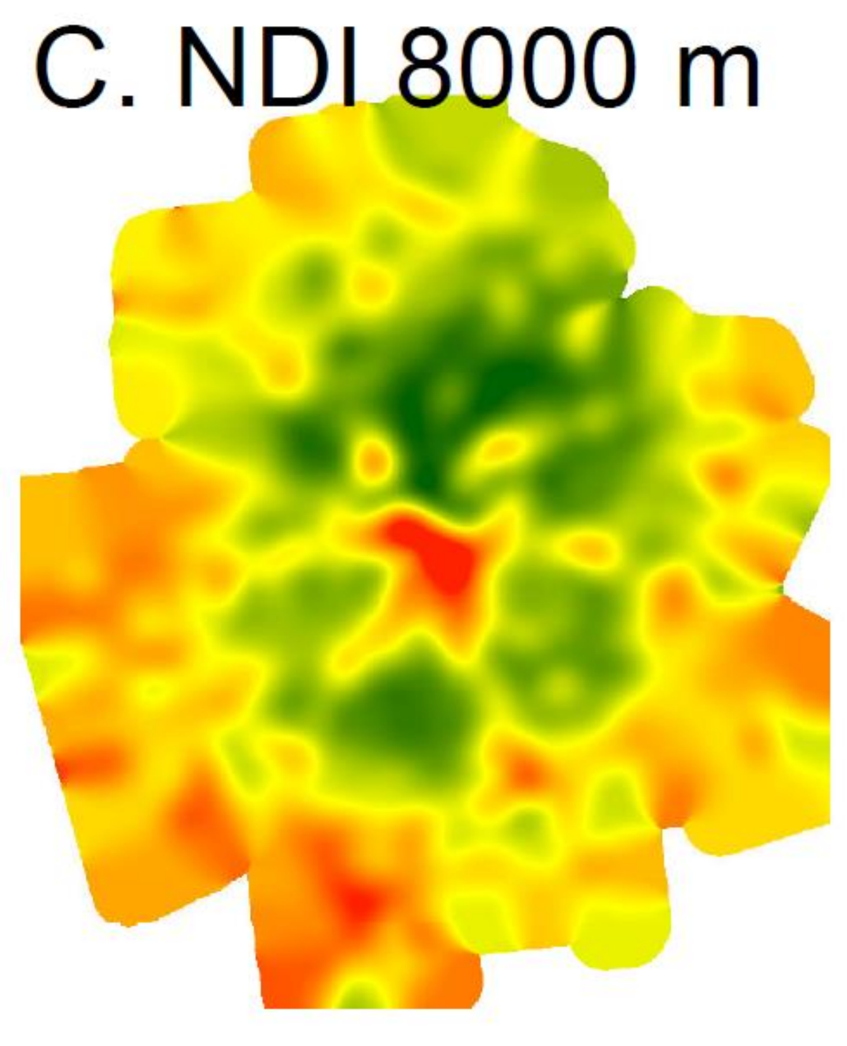


Reardon, O'Sullivan, et al, 2004
Kramer, et al, 2010

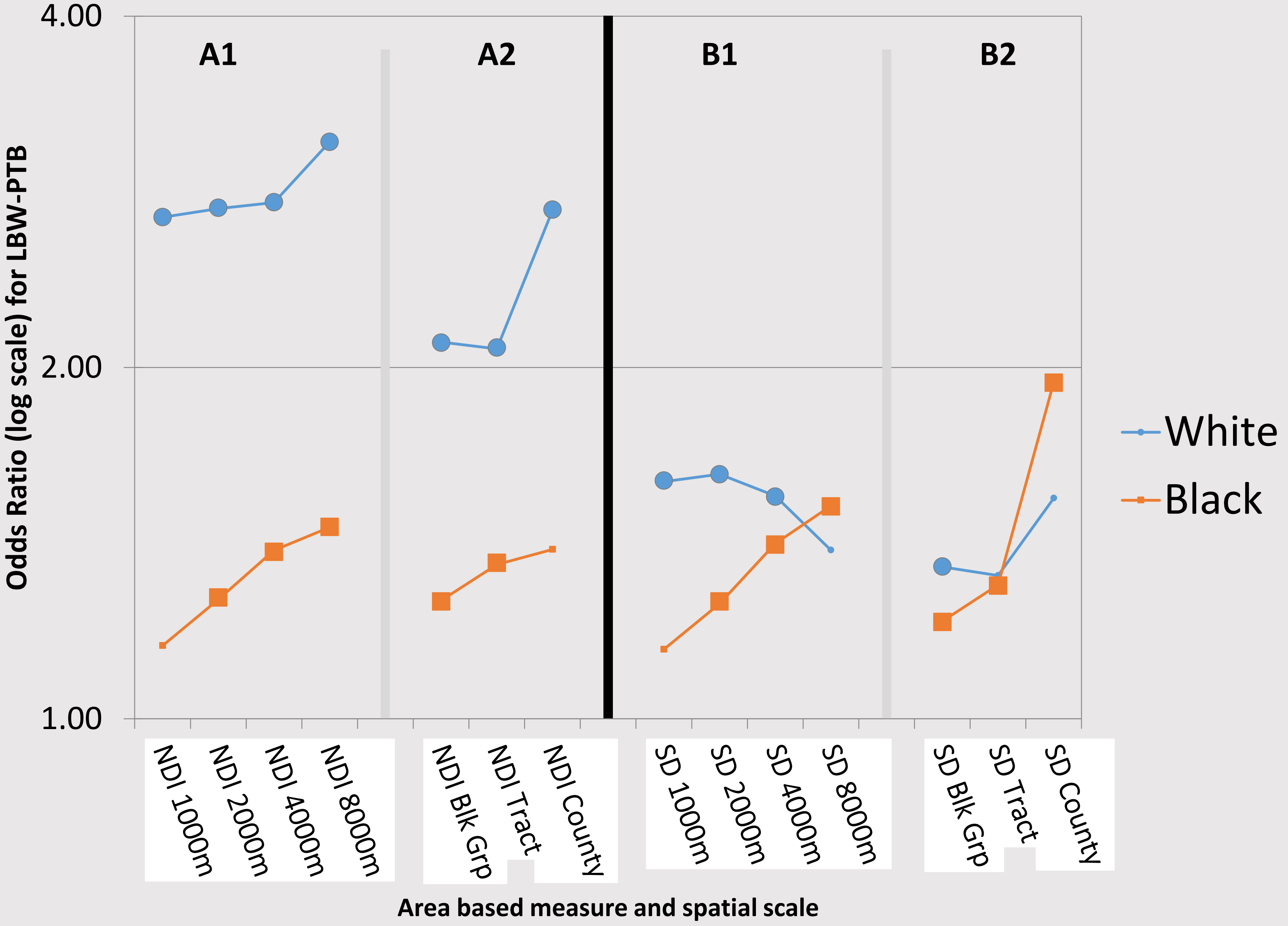


Reardon, O'Sullivan, et al, 2004
Kramer, et al, 2010

Neighborhood Deprivation Index, 2007



Kramer, in press, 2015
Messer, et al, 2006.



Kramer, in press, 2015

Recap

Think Spatially!

- What does space mean for your purpose?
- What tool answers your question?
- Avoid gratuitous use of spatial analysis

Spatial structure

- Who is your neighbor? Choose weights specification
- Use spatial autocorrelation (global and local) to describe structure

Place containers and spatial surfaces

- MAUP means think about choice of units!
- Spatial density surfaces can generalize some point and polygon data

Spatial scale: How big is your neighborhood?

- There is no single correct scale!
- Choice depends on data available and hypothesized process or mechanism
- Multi-scale analysis can be informative

5. Where from here?

Software

- ArcGIS
  - Data management and some analysis (Moran's I, Getis G*, Kernel densities)
- QGIS, GRASS
  - Open source GIS platforms
  - Learning curve, not all ESRI functionality, customizable
- GeoDa
  - Exploratory spatial data analysis: Moran's I, LISA, Getis-Ord, simple spatial regression
  - Empirical Bayes spatial smoothing for small area estimation
- R
  - Rapidly growing suite of spatial analysis packages including clustering, kernel density estimation, spatial regression, mapping

Thank you!

Michael Kramer
Emory University
mkram02@emory.edu
@mkramer_atl

Cited references

1. Kelsall, Julia E., and Peter J. Diggle. "Non-parametric estimation of spatial variation in relative risk." Statistics in medicine 14, no. 21-22 (1995): 2335-2342.
2. Reardon, Sean F., and David O’Sullivan. "Measures of spatial segregation." *Sociological methodology* 34, no. 1 (2004): 121-162.
3. Matthews, Stephen A., and Tse-Chuan Yang. "Spatial Polygamy and Contextual Exposures (SPACEs) Promoting Activity Space Approaches in Research on Place And Health." *American Behavioral Scientist* 57, no. 8 (2013): 1057-1081.
4. Kramer MR. Race, place, and space: Ecosocial theory and spatiotemporal patterns of pregnancy outcomes. In: Howell FM, Porter JR, Matthews SA, eds, Recapturing space: New middle-range theory in spatial demography. Springer. In press
5. Messer, Lynne C., Barbara A. Laraia, Jay S. Kaufman, Janet Eyster, Claudia Holzman, Jennifer Culhane, Irma Elo, Jessica G. Burke, and Patricia O’campo. "The development of a standardized neighborhood deprivation index." *Journal of Urban Health* 83, no. 6 (2006): 1041-1062.

Software Info

- QGIS
  - http://www.qgis.org/en/site/
- GRASS
  - http://grass.osgeo.org/
- GeoDa
  - http://geodacenter.asu.edu/
- R
  - General: http://cran.us.r-project.org/
  - Spatial Packages: http://cran.r-project.org/web/views/Spatial.html

Additional Resources

- Spatial Texts
  - Waller, Lance A., and Carol A. Gotway. *Applied spatial statistics for public health data*. Vol. 368. John Wiley & Sons, 2004.
  - Bivand, Roger S., Edzer J. Pebesma, and Virgilio Gómez-Rubio. *Applied spatial data analysis with R*. Vol. 747248717. New York: Springer, 2008.
  - Anselin, Luc. *Spatial econometrics: methods and models*. Vol. 4. Springer, 1988.
  - Elliot, P., Jon C. Wakefield, Nicola G. Best, and D. J. Briggs. *Spatial epidemiology: methods and applications*. Oxford University Press, 2000.
  - Lawson, Andrew B. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press, 2013.
  - Lai, Poh-Chin, Fun-Mun So, and Ka-Wing Chan. *Spatial epidemiological approaches in disease mapping and analysis*. CRC Press, 2010.

Additional Resources

- Articles -- *Methods*

  - Anselin, Luc. "Local indicators of spatial association—LISA." *Geographical analysis* 27, no. 2 (1995): 93-115.
  - Rushton, Gerard. "Public health, GIS, and spatial analytic tools." *Annual review of public health* 24, no. 1 (2003): 43-56.
  - Carlos, Heather A., Xun Shi, James Sargent, Susanne Tanski, and Ethan M. Berke. "Density estimation and adaptive bandwidths: a primer for public health practitioners." *International journal of health geographics* 9, no. 1 (2010): 39.

Additional Resources

- Articles – *MCH* (this is a non-representative sample of dozens or relevant papers)

  - Morello-Frosch, Rachel, and Edmond D. Shenassa. "The environmental “riskscape” and social inequality: implications for explaining maternal and child health disparities." *Environmental Health Perspectives* 114, no. 8 (2006): 1150.
  - Culhane, Jennifer F., and Irma T. Elo. "Neighborhood context and reproductive health." *American journal of obstetrics and gynecology* 192, no. 5 (2005): S22-S29.
  - Charreire, Hélène, and Evelyne Combier. "Poor prenatal care in an urban area: a geographic analysis." *Health & place* 15, no. 2 (2009): 412-419.
  - Huynh, M., and A. R. Maroko. "Gentrification and Preterm Birth in New York City, 2008–2010." *Journal of Urban Health* 91, no. 1 (2014): 211-220.
  - Kramer MR, Dunlop AL, Hogue CR. Measuring women's cumulative neighborhood deprivation exposure using longitudinally linked vital records: A method for life course MCH research. Maternal and Child Health Journal. 2014;18(2):478-487.