Center for Spatially Integrated Social Science

Spatial Tools for Econometric and Exploratory Analysis

Michael F. Goodchild
University of California, Santa Barbara
Luc Anselin
University of Illinois at Urbana-Champaign
http://csiss.org

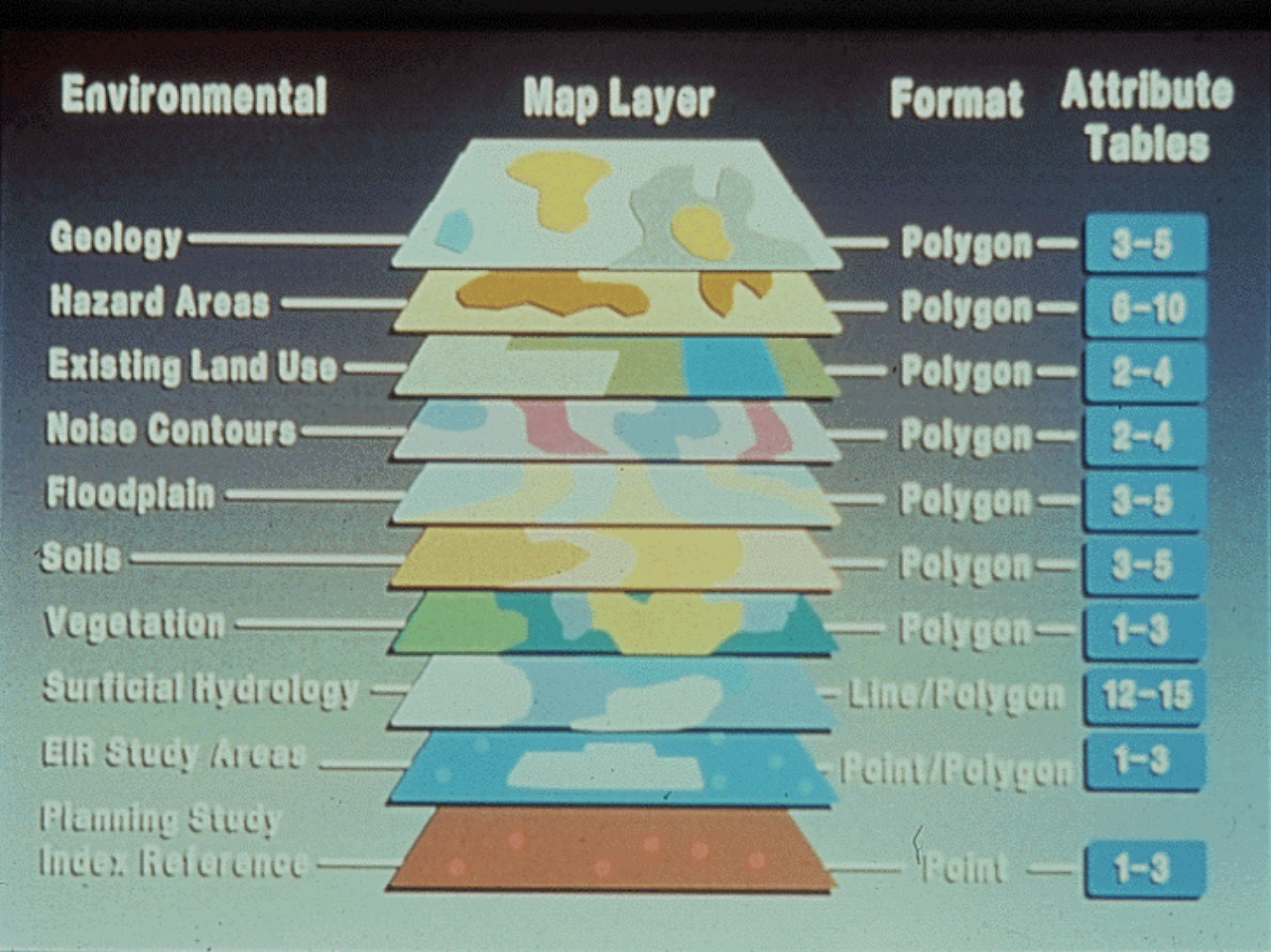
Outline

- A Quick Tour of a GIS
- Spatial Data Analysis
- CSISS Tools

Spatial Data Analysis

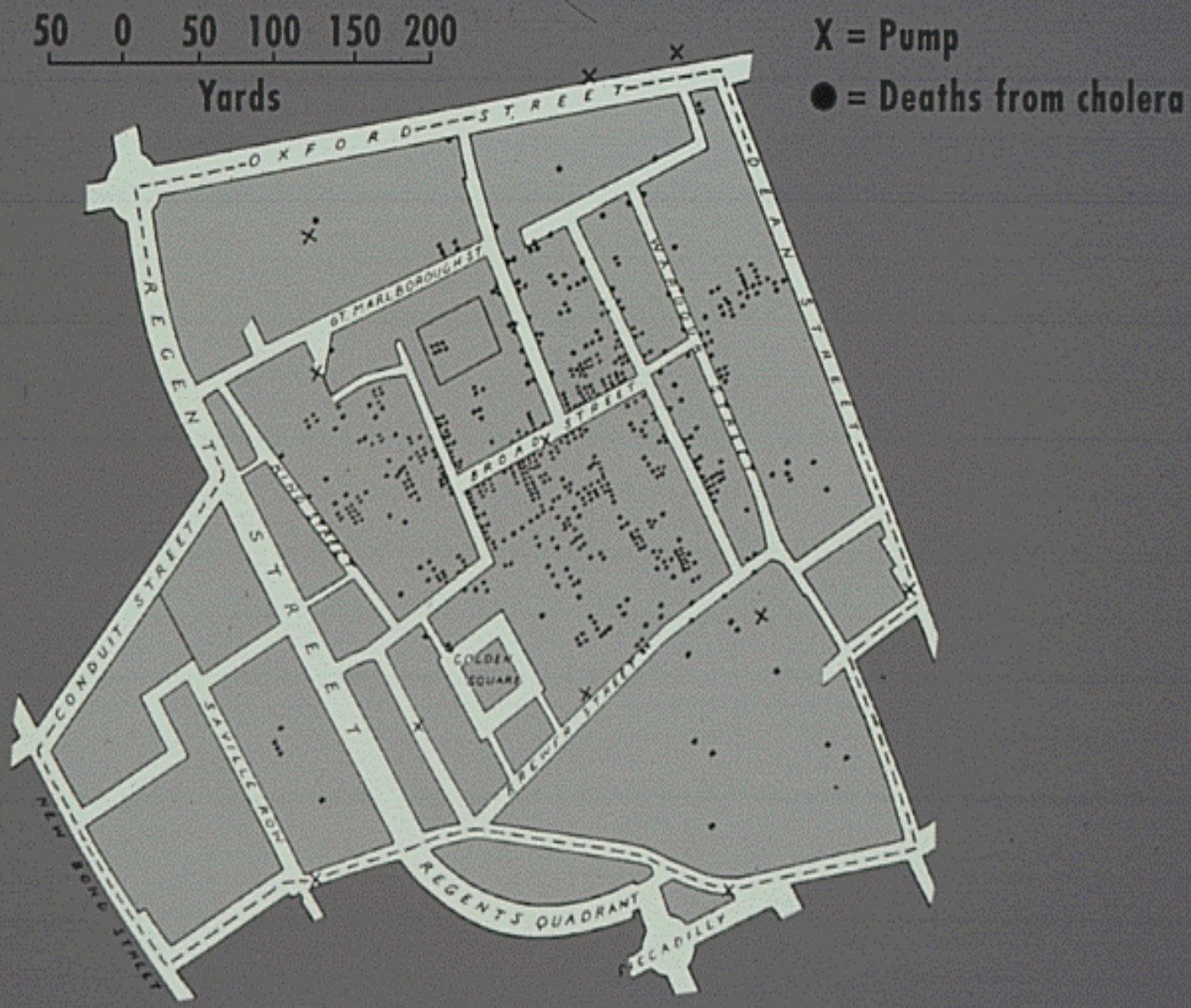
Principles: 1. Integration

- Linking data through common location
  - the layer cake
- Linking processes across disciplines
  - spatially explicit processes
  - e.g. economic and social processes interact at common locations


Environmental
Map Layer
Format
Attribute Tables
Geology
Hazard Areas
Existing Land Use
Noise Contours
Floodplain
Soils
Vegetation
Surficial Hydrology
EIR Study Areas
Planning Study Index Reference
Polygon
Polygon
Polygon
Polygon
Polygon
Polygon
Polygon
Line/Polygon
Point/Polygon
Point
3-5
6-10
2-4
2-4
3-5
3-5
1-3
12-15
1-3
1-3

2. Spatial analysis

- Social data collected in cross-section
  - longitudinal data are difficult to construct
- Cross-sectional perspectives are rich in context
  - can never confirm process
  - though they can perhaps falsify
  - useful source of hypotheses, insights



The Snow Map of Cholera Incidence in the Area of Broad Street, London, in 1854.
The contaminated water pump is located at the center of the map, just to the right of the D in BROAD STREET.

3. Spatially explicit theory

- Theory that is not invariant under relocation
- Spatial concepts (location, distance, adjacency) appear explicitly
- Can spatial concepts ever *explain*, or are they always surrogates for something else?

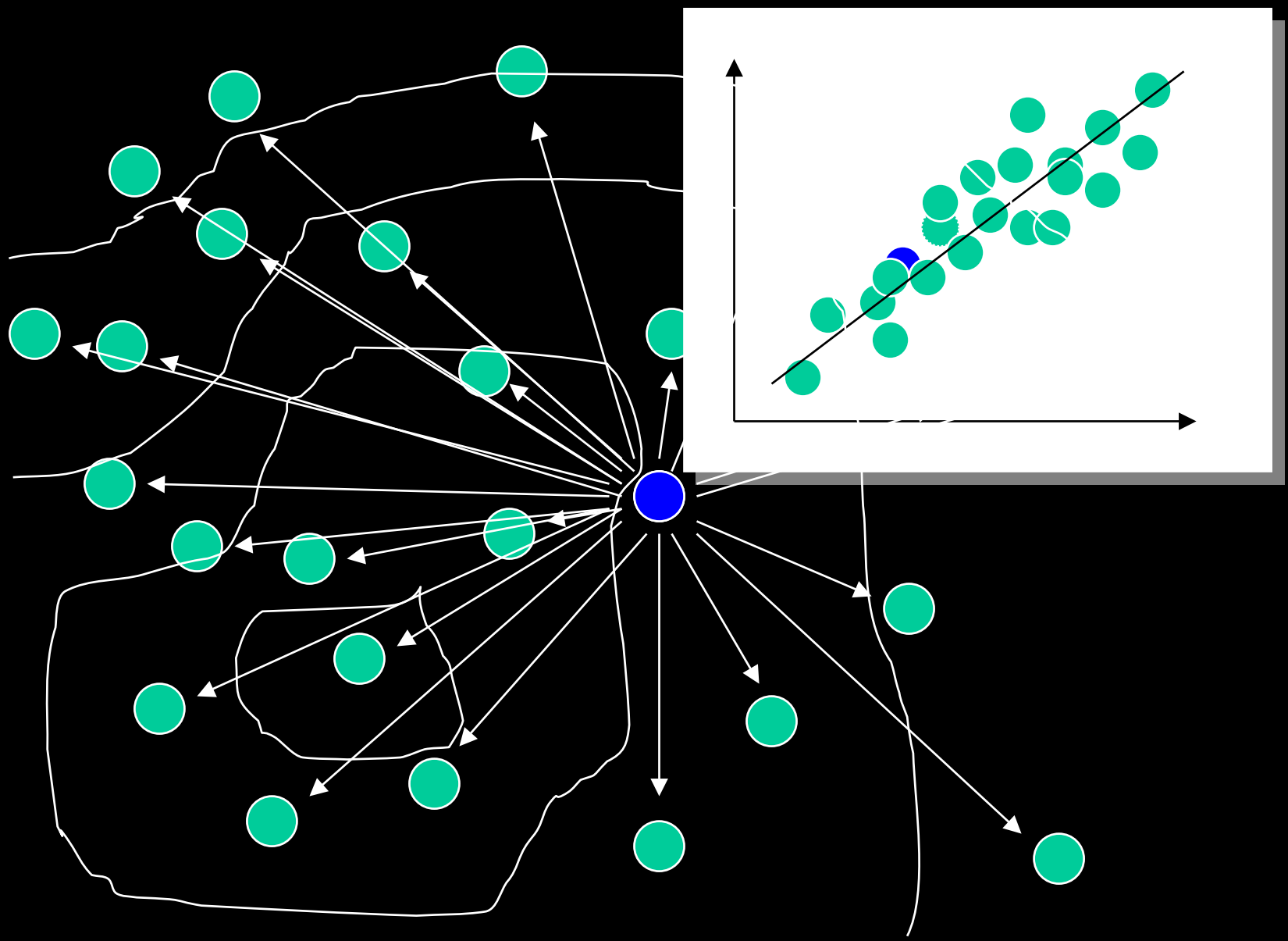
4. Place-based analysis

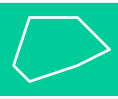
- Nomothetic - search for general principles
- Idiographic - description of unique properties of places
- An old debate in Geography

The Earth's surface

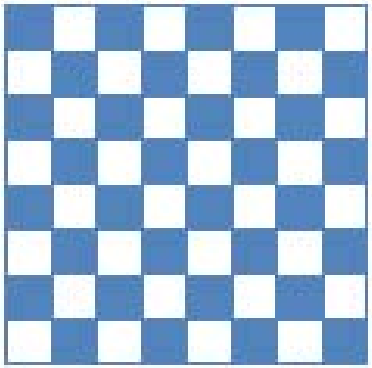
- Uncontrolled variance
- There is no average place
- Results depend explicitly on bounds
- Places as samples
- Consider the model:

  $y = a + bx$



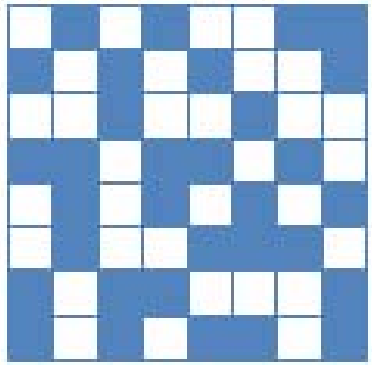
Tract	Pop	Location	Shape
1	3786	*x*,*y*	
2	2966	*x*,*y*	
3	5001	*x*,*y*	
4	4983	*x*,*y*	
5	4130	*x*,*y*	
6	3229	*x*,*y*	
7	4086	*x*,*y*	
8	3979	*x*,*y*	

(A)



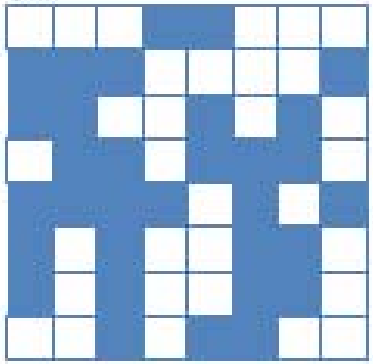
I = -1.000
n_{BW} = 112
n_{BB} = 0
n_{WW} = 0

(B)



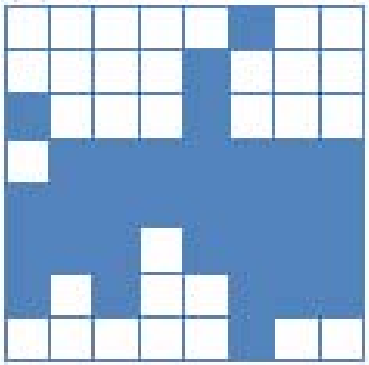
I = -0.393
n_{BW} = 78
n_{BB} = 16
n_{WW} = 18

(C)



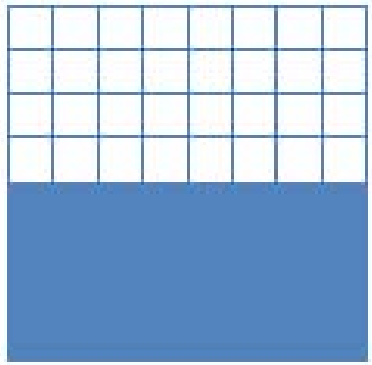
I = 0.000
n_{BW} = 56
n_{BB} = 30
n_{WW} = 26

(D)

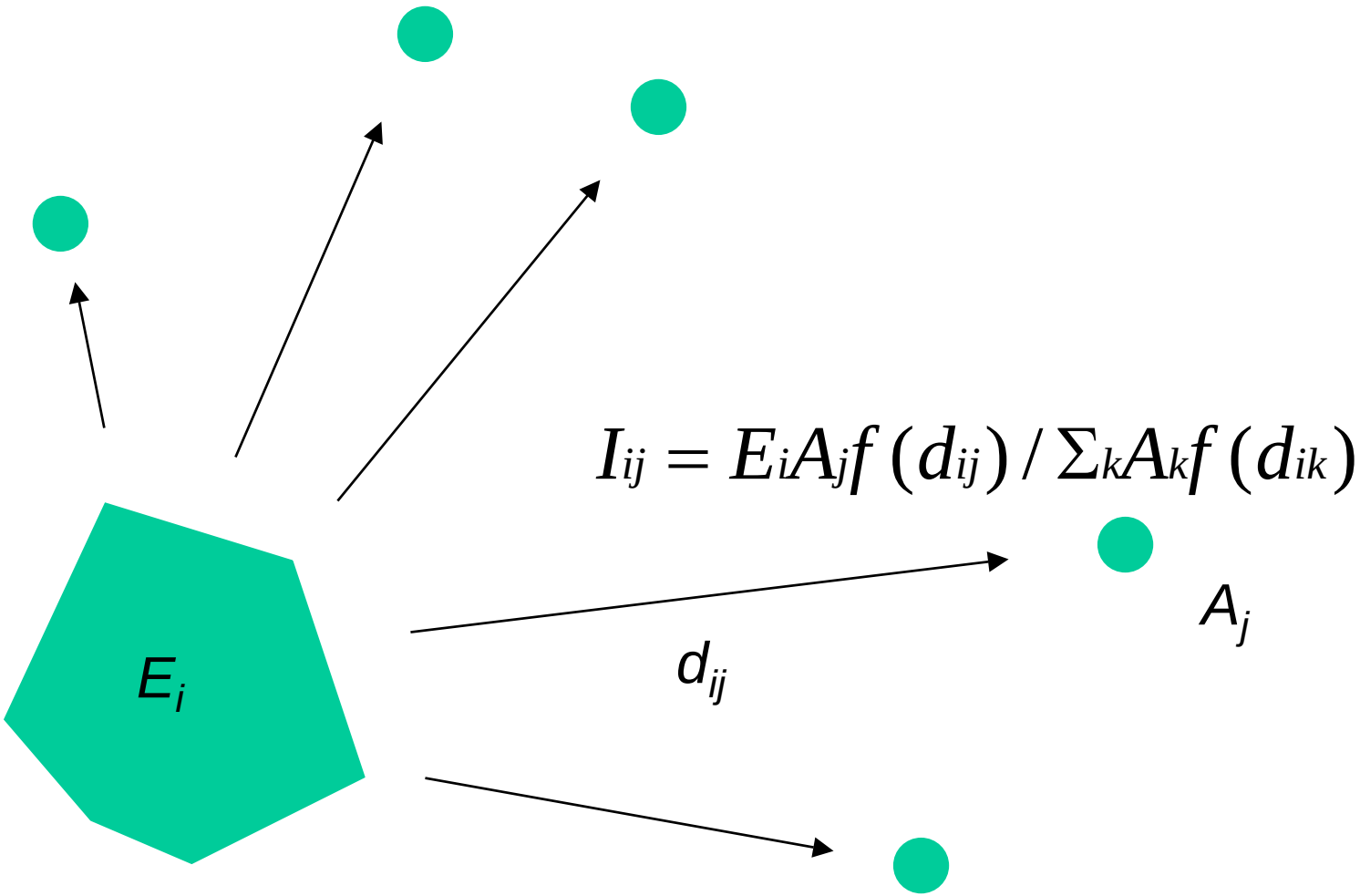


I = +0.393
n_{BW} = 34
n_{BB} = 42
n_{WW} = 36

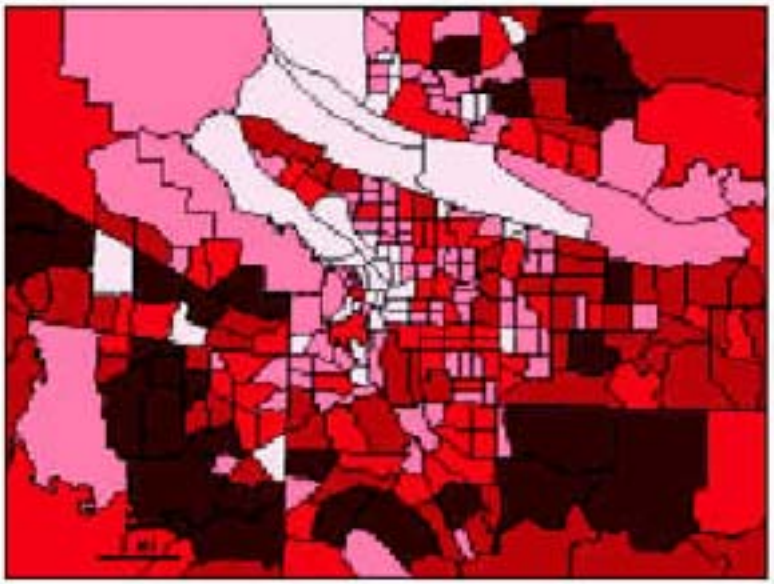
(E)



I = +0.857
n_{BW} = 8
n_{BB} = 52
n_{WW} = 52


$I_{ij} = E_i A_j f(d_{ij}) / \Sigma_k A_k f(d_{ik})$
A_j
d_{ij}
E_i

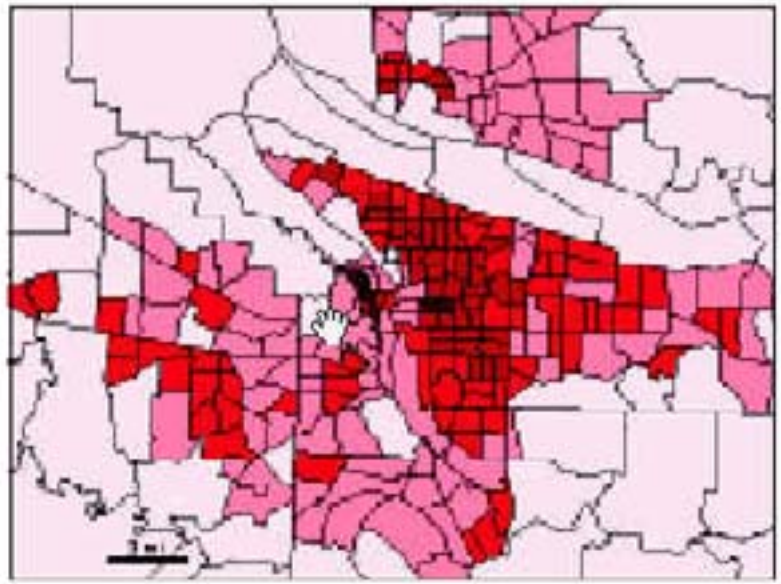
If you want to know approximately how many people each census tract has, map total population.



132 - 2224
2225 - 3900
3901 - 5507
5508 - 7732
7733 - 12611

Census tracts by total population.

If you want to know where most of the people are concentrated. map population density.



8 - 1882
1883 - 4409
4410 - 6906
6907 - 11111
11112 - 21334

Census tracts by people per square mile.

Types of Spatial Data Analysis

- Exploratory Spatial Data Analysis
  - exploring the structure of spatial data
  - determining the nature of spatial dependencies
  - spatial data as context
- Confirmatory Spatial Data Analysis
  - testing data against spatial models
  - spatial econometrics
  - econometric models that incorporate spatial effects

Four Elements of Spatial Econometrics

- Specifying the Structure of Spatial Dependence
  - which locations/actors interact
- Testing for the Presence of Spatial Dependence
  - what type of dependence
  - what alternative
- Estimating Models with Spatial Dependence
  - spatial lag, spatial error, higher order
- Spatial Prediction
  - interpolation, missing values

Abstraction of geographic space

- Cartograms

- Invariance under rotation, displacement, reflection

Space as a matrix

- $\mathbf{W}$ where w_{ij} is some measure of interaction
  - adjacency
  - decreasing function of distance
  - invariant under rotation, displacement, reflection
  - readily obtained from GIS
  - a measure of spatial vs social interaction

Specification = Spatial Weights

- Spatially Lagged Variables
  - Wy, Wε, WX
- Data Structure
  - W not as N by N matrix
  - sparse structure
    - 3000+ counties 5/6 neighbors
- Computation
  - distance based: easy
    - uses point coordinates
    - generalized distance
  - contiguity: requires GIS interface
    - uses boundary file or tessellation

Software Classification

- Mainstream Statistics Packages
  - SAS, SPSS, Systat
- Mainstream Econometric Packages
  - Stata, TSP, Rats, Limdep, E-Views, Shazam
- Toolboxes
  - Matlab, S-Plus, Bugs/WinBugs
- Open Source Toolboxes
  - R, XLispstat
- Self-Contained Specialized
  - SpaceStat

Spatial Regression Functionality

- Rare in Commercial Software
  - spatial functionality tends to be geostatistics, point pattern analysis
    - variogram, kriging
    - e.g., SAS, Systat
- Specialized Scripts/Macros
  - routines with specific functionality
  - constrained by data format, size, speed
- “Comprehensive”
  - SpaceStat, S+Spatialstats, Spdep (R), LeSage-Pace (Matlab), Pisati (Stata Ado)

Review (not comprehensive)

- SpaceStat
  - linear spatial regression
    - weights construction, diagnostics, ML, IV/GM
    - outdated architecture and interface
- S+Spatialstats (Splus)
  - linear spatial regression
    - weights construction (ArcView bridge), ML
    - no diagnostics
- Spdep (Bivand)
  - R package, open source
  - linear spatial regression
    - weights construction (from boundary file), diagnostics
    - ML estimation only
  - other R packages: Venables-Ripley, GWR, etc.

Review (continued)

- Spatial Toolbox (LeSage, Pace)
  - Matlab and Fortran routines
  - linear spatial regression
    - some weights construction (Thiessen), ML estimation (sparse weights), Bayesian estimation (Gibbs sampler)
  - spatial probit/tobit: Gibbs sampler
- Stata (Pisati, Conley)
  - Spatreg
    - diagnostics, ML estimation, no weights from polygons
  - GMM Conley estimator
- GeoBugs
  - MCMC for conditional spatial model

Miscellaneous

- SAS
  - spatial regression estimation (Griffith)
- SPSS
  - spatial autocorrelation (Tiefelsdorf)
- Xlispstat
  - “Ord” ML estimation using eigenvalues
- TSP
  - ML estimation spatial error, spatial lag (based on Upton and Fingleton)
- RATS
  - Driscoll-Kraaij estimator

Issues

- Performance Problems
  - generic optimization routines
  - inefficient matrix data structure
  - slow loops
- Little/No Diagnostics
  - focus tends to be on estimation
- No Interoperability
- No Standards

CSISS Tools

CSISS Tools Project Mission

- Goals:
  - facilitate dissemination of spatial analysis software tools to social scientists
  - develop a library/libraries of spatial data analysis modules
  - develop prototypes implementing state of the art methods
  - initiate and nurture a community of open source developers

http://www.csiss.org/clearinghouse/index.php3

Center for Spatially Integrated Social Science

CSISS Tools Clearinghouse

The ***CSISS Tools Clearinghouse*** is intended to grow into a robust collection of spatial analysis software, software links, and links to information about tools for spatial analysis. The development of these tools is a lively research area and the goal of this clearinghouse is to provide up-to-date information on available tools. The clearinghouse is comprised of:

Search Engine

Search a continuously updated, comprehensive index of the CSISS Select Tools and Links to Portals.

Select Tools

Browse through tools particularly suited to the analysis of spatial phenomena.

Portal Links

A listing of useful collections of software tools for anyone interested in Spatial Analysis, or those looking for specific tools.

CSISS Tools (offsite)

The home of the software tools development efforts under CSISS, carried out in the Spatial Analysis Laboratory of the Department of Agricultural and Consumer Economics at the University of Illinois, Urbana-Champaign.

New - GeoDa 0.9, beta release software for ESDA with dynamically linked windows.

Your help is requested in suggesting tools, collections of tools, other portals, and methods that should be represented in this collection - please send these to the Tools Manager, Luc Anselin at anselin@uiuc.edu.

Address http://sal.agecon.uiuc.edu/csiss/index.html Go Links

CSISS Software Tools Project

Mission

The Center for Spatially Integrated Social Science (CSISS) is a five-year project funded by the U.S. National Science Foundation under its program of support for infrastructure in the social and behavioral sciences. CSISS promotes an integrated approach to social science research that recognizes the importance of location, space, spatiality and place.

One of the CSISS programs is devoted to "Spatial Analytic Tools" for the social sciences. It is directed by Luc Anselin and housed in the Spatial Analysis Laboratory of the Department of Agricultural and Consumer Economics at the University of Illinois, Urbana-Champaign. The Tools Project aims to develop and disseminate a powerful and easy to use suite of software for spatial data analysis, to advance methods of statistical analysis to account for spatial effects, and to integrate these developments with GIS capabilities.

Current Activities

- Openspace Mailing List
- Spatial Software Tools Clearinghouse
- GeoDa - Exploratory Spatial Data Analysis with Dynamically Linked Windows
- OpenSpace - Java Applets and Applications for Spatial Data Analysis
- PySpace - Spatial Statistical Analysis in Python
- Large Data Set SAR
- Supporting Materials

GeoDa

- ESDA with Dynamically Linked Windows
  - freestanding
  - reads ESRI shape files
    - points and polygons
  - MapObjects LT2 technology
  - free
- Replaces ArcView Extensions
  - SpaceStat Extension and DynESDA extension now obsolete
- Download
  - http://sal.agecon.uiuc.edu/csiss/geoda.html

Supporting Materials

Requirements and Challenges

- Interoperable
  - need for standards
    - data structures, model formulation
    - XML
  - adhere to/contribute to OGC
- Open
  - open source promotes quality control
  - standards allow a modular approach
- Fast
  - need for efficient (new) algorithms
  - large data set latent variable models
  - space-time models