What’s Special about Spatial?

Michael F. Goodchild
University of California
Santa Barbara
Outline

- GIS
- The Center for Spatially Integrated Social Science (CSISS)
- The $W$ matrix
Definitions

- **Spatial data**
  - information about phenomena organized in a spatial frame
  - the geographic frame

- **Spatial analysis**
  - methods applied to spatial data that
    - add value
    - reveal patterns and anomalies
    - support decisions
The role of the GIS

- The infrastructure for handling data types
  - to spatial data as Excel is to tables, as S-Plus is to statistical data, as Word is to text
  - spatial data or geographic data?
  - the housekeeper
  - the editor

- The visualization tool
The GIS data types

- Discrete geographic features
  - points, lines, areas
  - the contents of maps
  - with associated attributes
  - countable
  - conceived as tables with associated feature geometry
Scottish Munros

1. Ben Hope
2. Ben Klibreck
3. Ben More Assynt
4. An Teallach
5. Seana Bhragh
6. Ben Wyvis
7. Slioch
8. Sgurr Ruadh
9. Moruisg
10. Sgurr na Ruaidhe
11. Bia Bheinn
12. Sgurr na Lapalch
13. Ben Atow
14. The Saddle
15. Creag a’ Mhaim
16. Ladhar Bheinn
17. Coirechan
18. Ben Nevis
20. Ben Starav
21. Braeriach
22. Ben Avon
23. Meall Chualach
24. Mt Keen
25. Deinn Dearg
26. Glas Maol
27. Driesh
28. Schiehallion
29. Ben Chonzie
30. Ben Lawers
31. Ben Challum
32. Ben Lomond
Fields

- Geography as a collection of continuous variables
  - measured on nominal, ordinal, interval, ratio scales
  - vector fields of direction and magnitude
  - exactly one value per point
  - $z=f(x)$
  - population density, land ownership, zoning
If you want to know approximately how many people each census tract has, map total population.

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

Census tracts by total population.

Census tracts by people per square mile.
Taxonomies of spatial analysis

- Thousands of methods
  - every one a command, menu item, icon, ...
- Based on data type
  - point pattern analysis
  - area (polygon) analysis
  - analysis of interactions
  - Bailey and Gatrell, Haining, Unwin
A six-way conceptual classification

- Query and reasoning
- Measurement
- Transformation
- Descriptive summary
- Optimization
- Hypothesis testing
Queries and reasoning

- Real-time answers to geographic questions
  - Where is...?
  - What is this?
  - How do I get from here to here?
- Based on alternative views of a database
<table>
<thead>
<tr>
<th>Directions</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Start out going East on HENLEY ST towards WARREN ST</td>
<td>0.1 miles</td>
</tr>
<tr>
<td>2: Turn RIGHT onto WARREN ST.</td>
<td>0.0 miles</td>
</tr>
<tr>
<td>3: Turn RIGHT onto CHELSEA ST.</td>
<td>0.0 miles</td>
</tr>
<tr>
<td>4: CHELSEA ST becomes CHELSEA ST/CITY SQ.</td>
<td>0.1 miles</td>
</tr>
<tr>
<td>5: Turn RIGHT onto CITY SQ/NEW RUTHERFORD AVE/SR-99 N.</td>
<td>0.0 miles</td>
</tr>
<tr>
<td>6: Stay straight to go onto NEW RUTHERFORD AVE/SR-99 N.</td>
<td>0.2 miles</td>
</tr>
<tr>
<td>7: Turn SLIGHT LEFT onto SR-99 N.</td>
<td>0.4 miles</td>
</tr>
<tr>
<td>8: Turn SLIGHT LEFT onto SR-99 N/RUTHERFORD AVE.</td>
<td>0.1 miles</td>
</tr>
<tr>
<td>9: Turn SLIGHT LEFT onto SR-99 N.</td>
<td>0.3 miles</td>
</tr>
<tr>
<td>10: Turn SLIGHT LEFT onto SULLIVAN SQUARE OPAS.</td>
<td>0.4 miles</td>
</tr>
<tr>
<td>11: Turn SLIGHT LEFT onto MYSTIC AVE.</td>
<td>0.7 miles</td>
</tr>
<tr>
<td>12: MYSTIC AVE becomes MYSTIC AVE/SR-38 N.</td>
<td>1.2 miles</td>
</tr>
<tr>
<td>13: Turn LEFT onto HARVARD ST.</td>
<td>0.6 miles</td>
</tr>
<tr>
<td>14: HARVARD ST becomes WARNER ST.</td>
<td>0.2 miles</td>
</tr>
<tr>
<td>15: Turn RIGHT onto POWDER HOUSE SQ.</td>
<td>0.1 miles</td>
</tr>
<tr>
<td>16: Turn RIGHT onto BROADWAY.</td>
<td>1.0 miles</td>
</tr>
<tr>
<td>17: Turn LEFT onto ALEWIFE BROOK PKWY/SR-16.</td>
<td>0.4 miles</td>
</tr>
<tr>
<td>18: ALEWIFE BROOK PKWY/SR-16 becomes ALEWIFE BROOK PKWY/SR-16/US 3.</td>
<td>0.4 miles</td>
</tr>
<tr>
<td>19: Take CONCORD TURNPIKE/SR-2 W.</td>
<td>4.7 miles</td>
</tr>
<tr>
<td>20: Take the WALTHAM ST. exit, exit number 54B, towards LEXINGTON.</td>
<td>0.2 miles</td>
</tr>
<tr>
<td>21: Merge onto WALTHAM ST.</td>
<td>1.9 miles</td>
</tr>
<tr>
<td>22: Turn RIGHT onto MASSACHUSETTS AVE/MASS AVE/SR-225</td>
<td>0.0 miles</td>
</tr>
</tbody>
</table>

**Total Distance:** 12.9 miles (20.8 km)

**Estimated Time:** 24 minutes
Measurements

- Area
- Distance
- Length
- Perimeter
- Slope, aspect
- Shape
Transformations

- Buffering
- Points in polygons
- Polygon overlay
- Spatial interpolation
- Density estimation
City limits

Areas reachable in 5 minutes

Areas reachable in 10 minutes

Other areas
Search radius = 200 feet

Search radius = 1000 feet
Descriptive summary

- Centers
- Measures of spatial dispersion
- Spatial dependence
- Fragmentation
- Fractional dimension
MASS MOVEMENTS OF LAND USE SURFACES
LONDON, ONTARIO
1850 - 1960

1850 - 1875

1875 - 1950

1950 - 1960

Land Use Surfaces

Residential (*)
Public - Institutional (*)
Commercial (*)
Industrial (*)

Mean locations of land use surfaces
Direction of mass movements
Peak value intersection: 1960

Industrial land includes commercial and commercial properties.
Public and institutional land included with residential land for 1850 - 1875 and 1875 - 1950.
Optimization

- Design to achieve specific objectives
- Location of central point-like facilities to serve dispersed demand
- Location of linear facilities
- Design of boundaries for elections
Hypothesis testing

- Geographic objects as a sample from a population
  - what is the population?

- The independence assumption
  - the First Law of Geography
  - failure to find spatial dependence is always a Type II error
  - hell is a place with no spatial dependence
Information lost to the representation

- All sub-polygon spatial variation
- All within-decade temporal variation
- All identities
  - instead of <xy, person> we have
    <R, number>
  <xy, xy, xy, xy, ..., R>
Challenges of GIS

■ How to characterize what is missing?
  – error, accuracy, uncertainty

■ How to choose the best representation?
  – confounding influences

■ How to support many data models in a single software package
Weaknesses of GIS

- There are too many possible data models
  - special-purpose GIS
  - lack of interoperability
- Difficult to add data models retroactively
Objectives, Structure, and History of CSI-SS
Research infrastructure

- Facilities that serve generic needs
  - economies of scale
- The Hubble Telescope
  - high fixed costs distributed over many users
- Infrastructure funding at NSF
  - South Pole
  - advanced computation
Generalizing the concept

- Shared computational facilities
- Shared data archiving and access
- Software tools
  - shared licenses
  - shared development
- Education and training
  - investments in skills
  - high leverage
General 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
<table>
<thead>
<tr>
<th>Environmental</th>
<th>Map Layer</th>
<th>Format</th>
<th>Attribute Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geology</td>
<td>Polygon</td>
<td>3–5</td>
<td></td>
</tr>
<tr>
<td>Hazard Areas</td>
<td>Polygon</td>
<td>6–10</td>
<td></td>
</tr>
<tr>
<td>Existing Land Use</td>
<td>Polygon</td>
<td>2–4</td>
<td></td>
</tr>
<tr>
<td>Noise Contours</td>
<td>Polygon</td>
<td>2–4</td>
<td></td>
</tr>
<tr>
<td>Floodplain</td>
<td>Polygon</td>
<td>3–5</td>
<td></td>
</tr>
<tr>
<td>Soils</td>
<td>Polygon</td>
<td>3–5</td>
<td></td>
</tr>
<tr>
<td>Vegetation</td>
<td>Polygon</td>
<td>1–3</td>
<td></td>
</tr>
<tr>
<td>Surficial Hydrology</td>
<td>Line/Polygon</td>
<td>12–15</td>
<td></td>
</tr>
<tr>
<td>EIR Study Areas</td>
<td>Point/Polygon</td>
<td>1–3</td>
<td></td>
</tr>
<tr>
<td>Planning Study Index Reference</td>
<td>Point</td>
<td>1–3</td>
<td></td>
</tr>
</tbody>
</table>
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?
\[ I_{ij} = \frac{E_i A_{ij} f(d_{ij})}{\sum_k A_{ik} f(d_{ik})} \]
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:

  \[ y = a + bx \]
5. Knowledge and policy

- Policy requires the projection of general knowledge in spatial context
  - the implications of this process in this location
  - alternative futures visualized under local circumstances

- GIS combines the general (processes, models, algorithms) with the specific (database of local details)
6. Place-based search

- Location as an organizing dimension to information
  - much information can be georeferenced
  - much more than maps and images

- The Geolibrary
  - what have you got about there?
  - impossible physically, feasible digitally
Prototype geolibraries

- National Geospatial Data Clearinghouse
  - www.fgdc.gov
- Microsoft's Terraserver
  - terraserver.microsoft.com
- Alexandria Digital Library
  - alexandria.ucsb.edu
Map Browser

ADL Search Results

The query that produced these results can be found at the bottom of this page.

1. DRG o33117g6, Digital Raster Graphic of BLACK STAR CANYON, CA.
   [HIGHLIGHT IN MAP] [COMPLETE DESCRIPTION] [BROWSE GRAPHIC] [ACCESS/DOWNLOAD]

2. DRG o33117g7, Digital Raster Graphic of ORANGE, CA.
   [HIGHLIGHT IN MAP] [COMPLETE DESCRIPTION] [BROWSE GRAPHIC] [ACCESS/DOWNLOAD]

3. DRG o33117g8, Digital Raster Graphic of ANAHEIM, CA.
   [HIGHLIGHT IN MAP] [COMPLETE DESCRIPTION] [BROWSE GRAPHIC] [ACCESS/DOWNLOAD]
The CSISS mission recognizes the growing significance of space, spatiality, location, and place in social science research. It seeks to develop unrestricted access to tools and perspectives that will advance the spatial analytic capabilities of researchers throughout the social sciences.
Seven CSI S programs

- National Workshops
- Software Tools
- Virtual Community
- Best Practice Examples
- Place-Based Search
- Learning Resources
- Specialist Meetings
The W Matrix
Abstraction of geographic space

- Cartograms
- Invariance under rotation, displacement, reflection
- Reconstruction from a distance matrix
- Reconstruction from ranked distances
  - ordered metric data (Coombs)
Space as a matrix

- $W$ where $w_{ij}$ is some measure of interaction
  - adjacency
  - decreasing function of distance
  - invariant under rotation, displacement, reflection
  - readily obtained from GIS
  - reflects a discrete object conceptualization
The Modifiable Areal Unit Problem

- Openshaw and Taylor
  - 99 counties of Iowa
  - % Republican voters, % over 65
- 48 regions: -.548 to +.886
- 12 regions: -.936 to +.996

Solutions:
- manipulate to determine range
- strengthen theoretical framework
Applications of the $W$ matrix

- Spatial regression
  - add spatially lagged terms weighted by $W$
  - Anselin’s SPACESTAT

- Moran and Geary indices of spatial dependence

$$
C = \frac{(n-1) \sum_i \sum_j w_{ij} (x_i - x_j)^2}{2 \sum_i \sum_j \sum_i (x_i - a)^2}
$$
Social versus spatial

- $W$ estimated
- Well-defined discrete object conceptualization
- Direct measures of interaction

- $W$ calculated
- Discrete objects as arbitrary regions, MAUP
- Surrogate measures of interaction
Cross-Product Statistics

- Let $C$ be a matrix of similarities between objects.
- Consider the cross-product

\[
\Gamma = \sum_i \sum_j C_{ij} W_{ij}
\]
Properties of $\Gamma$

- Generalizes Moran and Geary
- Measures the correlation between social and spatial
- Simple randomization tests based on permutation
Social versus Spatial

- The ability to explain
  - empirical estimates versus measurements
- Additional arguments supporting spatial
  - context
  - integration
  - implementation
- A fruitful basis for collaboration