

**11.520: A Workshop on Geographic Information
Systems**

11.188: Urban Planning and Social Science Laboratory

Lecture 9: Geocoding & Network Analysis

November 2, 2005, Joseph Ferreira, Jr.
(based, in part, on Fall 2003 notes by Visiting Prof. Zhong-Ren
Peng)

Administrative

- Homework #2: Question #1 due today (filling in the table)
- Homework #2: Question #2 due at start of Monday's lab
 - Downhill from here - lots of parts to HW#2 that draw on various exercises to date. Once you are comfortable with the HW#2 elements, you are past the steep part of the semester's learning curve
- Last Homework set (#3) posted tomorrow - adding one criterion to HW#2 based on a raster layer
- Project Proposal due next Wednesday - see notes on 'project' page
- Current lab exercise #7 (raster analysis) is due Friday, Nov. 11
- Two more lab exercises: #8 (geocoding) and #9 (web mapping)
- One in-lab test

Raster Analysis revisited

- Review key raster analysis steps in Lab #7
- Develop housing value 'surface' for Cambridge based on:
 - Using the Census block group data:
 - Using the sales89 data:
 - Combine the two estimates

Geocoding

- Converting textual addresses and names to X,Y locations
 - Address matching - develop point map from mailing list
 - Lookup place names in a 'gazeteer' to find lat/lon, zip, place boundary, voting district, etc.
- General 'service' to translate among geographic identifiers
- Examine ArcGIS geocoding services
 - Focus on address matching using TIGER-style street centerline data (with address ranges)
 - In class - examine the exercises in Chapter 17 of "Getting to Know ArcGIS Desktop" by Tim Ormsby, et al. (one of the class books)
 - In lab (next Monday) - address match some of the Cambridge sites in the Mass Dept. of Environmental Protection database of oil spills

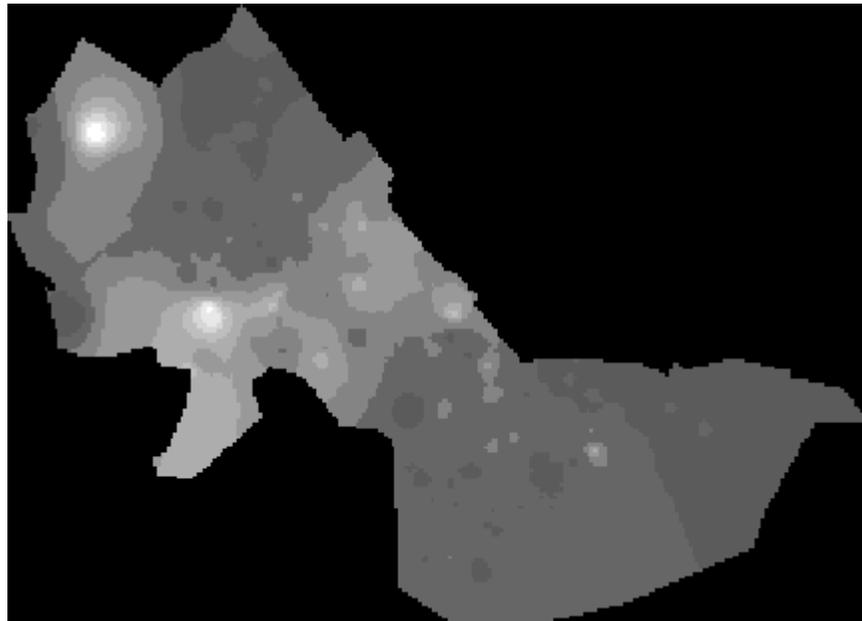
Network Analysis

- Encoding *proximity* using a *network* (or graph) model, facilitates certain types of *connectivity* analyses
 - Find shortest path along streets from Point A to Point B
 - Find shortest path through N cities (Traveling Salesman problem)
 - How far can you get in 30 minutes
 - Many transportation analyses use network data models
 - Many hydrological analyses use network data models (runoff, flow, ...)
-

Raster Analysis Review key raster analysis steps in Lab #7

- Computing a housing value 'surface' for Cambridge
 - Using the Census block group data:
 - Rasterize Cambridge into 100 meter grid cells
 - Vector-to-raster conversion of (median) housing value for Cambridge block groups
 - Smooth the 'surface' using neighborhood averages
 - Using the sales89 data:
 - Compute cell values from sales within the cell

- Adjust the 'surface' based on neighboring sales
- Combine the two estimates:
 - Average the two cell estimates (does this make sense? use weights?)
 - Understand use and limitations of map algebra models and the reasoning behind an interpolation method.



Interpolated Grid of Inflation-Adjusted 1989 Cambridge, MA Housing Sales Prices
(note 'hot spots' resulting from individual sales with unusually high price).

What is Geocoding

Geocoding is a process of creating map features from addresses, place names, or similar textual information based on attributes associated with a referenced geographic database, typically a street network that has address ranges associated with each street segment or 'link' running from one intersection to the next. (Derived from Longley, Goodchild,

Maguire and Rhind, *Geographic Information Systems and Science*, 2001.)

- 1.
1. Geocoding typically uses *Interpolation* as a method to find the location information about an address. (If the address along one side of a block range from 1 to 199, then Street Number = 66 is about one-third of the way along that side of the block.)
1. Data needed for geocoding
 1. A list of addresses stored as a database table or a text file;
 2. Georeferenced features linked to the address database (such as a street centerline shapefile with street names and address ranges stored as attributes of the shapefile)
 3. A geocoding service, which is a configuration file that specifies the georeferenced feature layer and its relevant attributes, and various rules and tolerance for use in the matching.
1. The output of the geocoding is a point file stored as either a shapefile or a geodatabase in ArcGIS.
- 2.
- 3.

Example: using US Census Bureau, TIGER Line Files

- Geocoding Strategy using TIGER
 - Encode road network as street centerlines
 - Attach address information to each street segment
 - Use 'in reverse' to match street address to street segment to get approximate X,Y location
- TIGER: **T**opologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing system
 - <http://www.census.gov/geo/www/tiger/>
 - US Census Bureau TIGER line file 2000, technical documentation

- at Census:
http://www.census.gov/geo/www/tiger/rd_2ktiger/tgrrd2k.pdf

Geocoding Example: Chapter 17, exercises a, b, and c in Ormsby

- Convert mailing list of neighborhood address to a point shapefile
 - Exercise 17a: create a 'geocoding service' using
 - Street centerline shapefile with address ranges (TIGER-format)
 - Attribute information consistent with address information in mailing list data table
 - Exercise 17b: use the geocoding service to 'address match' all the easy cases
 - Exercise 17c: use the interactive tools of the geocoding service to handle the tough cases
- 'Crib Sheet' of steps for Exercises 17a: create geocoding service
 - Add street shapefile and mailing list table to your ArcMap session
 - Examine layers to see what they have
 - Don't need to open this shapefile and table in order to create geocoding service
 - Open ArcCatalog and create a new geocoding service
 - Select US Streets with Zone style
 - Select street shapefile that contains address ranges
 - Match attribute columns of street file and mailing list in order to cross-reference
 - Note that the name of the geocoding service includes your user name
- 'Crib Sheet' of steps for Exercises 17b: use the automatic geocoding service
 - Add street shapefile and mailing list table to your ArcMap session
 - Tools/Geocoding/Geocoding-Services-Manager/Add-geocoding-service
 - Add the service you created in 17a
 - Use **Find** button on toolbar to locate individual addresses
 - Enter address by hand
 - Create graphic annotation showing location (but don't save as shapefile)
 - Use **Select Elements** button on toolbar to

- Automate address matching of easy-to-match addresses (Note use of quality-of-match index)
 - Save results as a shapefile of point features
- 'Crib Sheet' of steps for Exercises 17c: tweak the automatic results
 - Add street shapefile and 17b output to your ArcMap session
 - Tools/Geocoding/Review&Rematch-Addresses
 - Use Geocoding-Options and Interactive-Review to
 - tweak settings/rules for automated matching
 - edit mailing list addresses and match tough cases as best you can

What is a Network?

A network is a system of linear features connected at nodes, e.g, nodes could be where three or more street segments intersect.

The linear feature connecting any given pair of nodes is called an arc, or network link.

Each arc on a network is represented as an ordered pair of nodes, in the form from node i to node j , denoted by (i, j) , and thus has direction.

A network representation that is good for transportation modeling may differ from a geographically accurate representation of the physical road (e.g., street centerline, handling exit ramps, 3D overpasses, etc.)

- Other basic elements of a network:
 4. A shortest path is the shortest (or least 'cost' path) from a source node (origin) to a destination node. In practice, pathfinding seeks the shortest or most efficient way to visit a sequence of locations.
 5. A tour is an enclosed path, that is, the first node and the final node on the path are the same node on the network.

6. A stop is a location visited in a path or a tour.
7. Events or locations may be viewed as collection points (e.g., 'origins' or 'destinations') where certain resources are supplied or consumed.
8. A turn on a network is the transition from one arc to another arc at a node (there are 16 ways in which two intersecting roads can allow vehicle flow among the 4 links that 'connect' to the one node).
9. 'Location-allocation' models often use network representation of connected places in order to determine the optimal locations for a given number of facilities (e.g., stores, restaurants, banks, factories, warehouses, libraries, hospitals, post offices, and schools) based on some criteria, assign people to the the 'nearest' facility.

- Reference for further information about GIS and networks:

Source: <http://www.ncgia.ucsb.edu/giscc/units/u064/>

Last modified 2 November 2005.