Social Science Research Methodologies

Introduction to GIS

Martin Dodge

Lecture 3, 9th March 2005, 2-3pm

http://www.casa.ucl.ac.uk/martin/ss_methods/

Course outline

- 23rd February
  - Visualisation for academic research
- 2nd march
  - Cartographic design and exploratory mapping
- 9th March
  - What is GIS and what is it good for?
- 16th March
  - No lecture. Time to work on the course assessment!

Today's session

- 2- 3.00: GIS 101
- Break (5 minutes)
- 3.05 - 4.00: mapping practical in lab next door

Spatial is special

- why take a spatial perspective?

1. space provides a convenient conceptual and practical frame for organising large amounts of data
2. provides a means of linking together different data based on location
3. provides a means to linking to secondary data source by location
4. provides easy access to useful spatial properties (relative location, distance) that can be explicitly analysed
5. provides a powerful means of visualisation

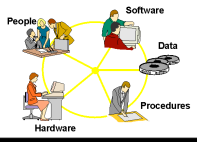
- most data will have some spatial elements and likely geographic locations. but these might not be obvious or easy to use
- if you are going to collect large volumes of data then think about GIS as the key tool for doing spatial analysis. the spatial properties of your data may not be 'accessible' (and thus not analysable) using other software
- although you can do spatial analysis using other tools, and maybe not using a computer at all. *GIS is not a panacea*

What is GIS?

- *Geographic Information Systems* (GIS) are not one thing, nor a single analytical technique; but a heterogeneous network of actors

Six Parts of a GIS

- Hardware
- Software
- Data
- People
- Procedures
- Network



GISs are simultaneously the telescope, the microscope, the computer, and the Xerox machine of regional analysis and synthesis of spatial data

The technology of spatial problem solving

GIS are:

- software products acquired to perform well-defined functions (*GIS software*)
- digital representations of aspects of world (*GIS data*)
- a community of people who use these tools for various purposes (the *GIS community*)
- the activity of using GIS to solve problems or advance science (*doing GIS*)

- everyone has their own favorite definition of GIS, and there are many to choose from

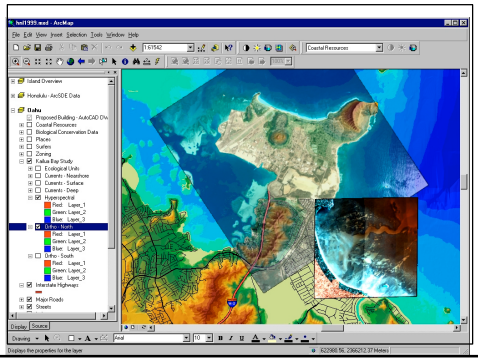
Definitions of GIS, and the groups who find them useful

a container of maps in digital form	the general public
a computerized tool for solving geographic problems	decision makers, community groups, planners
a spatial decision support system	management scientists, operations researchers
a mechanized inventory of geographically distributed features and facilities	utility managers, transportation officials, resource managers
a tool for revealing what is otherwise invisible in geographic information	scientists, investigators
a tool for performing operations on geographic data that are too tedious or expensive or inaccurate if performed by hand	resource managers, planners

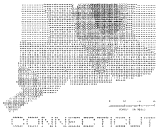
(source: Longley et al. 2001)

- GIS software is a major global business
- major software vendors include : Intergraph Corp., SmallWorld Systems Ltd., Environmental Systems Research Institute, Autodesk Inc., MapInfo Corp.
- also data, consultancy and training provided commercially. education - UCL's GIS courses for example

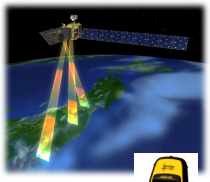
ArcGIS



History of GIS



- the first GIS was the Canada Geographic Information System
  - mid 1960s computerized mapping system
- late 1960s U.S. Bureau of the Census developed DIME (Dual Independent Map Encoding)
  - digital records of all U.S. streets, for automatic referencing and aggregation of census records
- separate needs of cartographers and mapping agencies
  - computers support for map editing
  - computerization by the late 1970s
- role of remote sensing
  - military satellites of the 1950s => early 1970s civilian systems
  - military needs also developed the GPS



- early 1980s take off (hardware prices could sustain software industry) $250,000 computers and $100,000 software (large resource managers)
- the modern history of GIS dates from the late 1980s, when the price of sufficiently powerful computers fell below a critical threshold

The utility of GIS in daily life

- the need to answer the fundamental question, *where*?
- not just a tool for military or researchers
- widely used in business and government
- GIS is used to improve many of our day-to-day working and living arrangements
- affects each of us, every day
- can be used to foster effective short- and long-term decision-making
- has great practical importance (e.g. efficient routing of the bin men)
- can be applied to many socio-economic and environmental problems (e.g. monitoring of GM crop trials)
- supports measurement, management, monitoring, and modeling operations
- rational, effective, and efficient allocation of resources
- understanding the difference that *place* makes
- generates measurable economic benefits

Major GIS users

- central and local government
- utilities (e.g. where are the pipes)
- developing (mid 1990s) in business (e.g. retail for store location planning), transportation logistics, real estate, market analysis
- new, e.g. online services, mobile location-based services, consumer applications

Local government

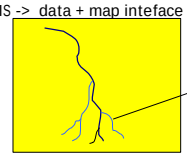
- 70-80% of the tasks undertaken by local government are geographically related
- still the biggest single group of GIS professionals
- drive to improve quality of products, processes and services
- inventory resources and infrastructure
- plan transportation routing
- improve service response time
- manage land development
- generate revenue through increased economic activity

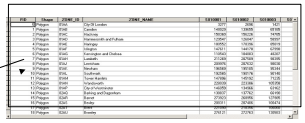
GIS as spatial processing toolbox

• GIS as a *'computerized data management system designed to capture, store, retrieve, analyze and report geographic and demographic information.'*

general purpose tool for particular kind of data
just like word processor for text, paint program for photos

GIS -> data + map inteface + tools = solution





Crucially, the graphic elements of the map (e.g. line for a river) is linked to separate data. this is known attribute data and it is used to describe the nature of the map element (eg. flow rate, pollution level of river)

Representing spatial features

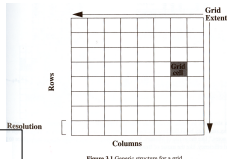


Figure 3.1 Generic structure for a grid.

grid features

(source: Clarke 1999, 54)

discrete features

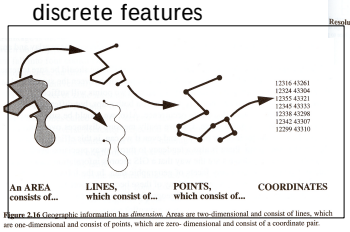


Figure 2.16 Geographic information has *dimension*. Areas are two-dimensional and consist of lines, which are one-dimensional and consist of points, which are zero- dimensional and consist of a coordinate pair.

(source: Clarke 1999, 69)

GIS conceptualised as layer model

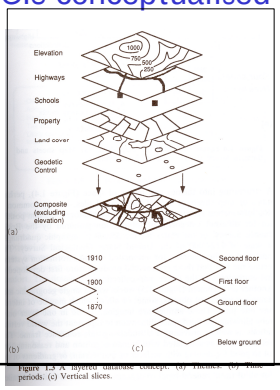
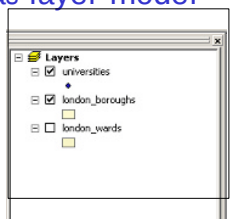


Figure 1.3 A layered umbrella concept. (a) Themes. (b) Time periods. (c) Vertical slices.



• much useful analysis works by combining layers together to derive new hybrid information

Spatial data handling

- dimensions of geographic data quality

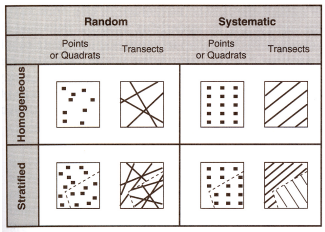
[e.g. think about in terms of universities data we will use next]

  - lineage (authenticity, metadata)
  - positional accuracy
  - attribute accuracy
  - logical consistency
  - completeness (geographic extent; thematically)
  - currency

- error and uncertainty. relative accuracy.
- random versus systematic errors

Spatial sampling methods

- to insure accurate (i.e. fit for purpose and scientifically defensible) collection of spatial data
- most cases you are using a sample from a whole population



Spatial analysis

- spatial data analysis is the *quantitative* study of phenomena that manifest themselves in space, aiming to *describe* the nature of their distribution and understand the *processes* that give rise to it
- it is usually premised on scientific methods of enquiry, hypothetical-inductive methods

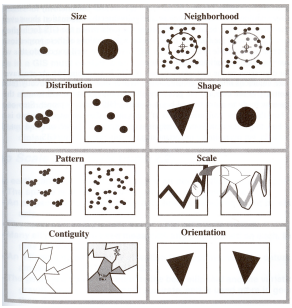
- exploring data
- looking for patterns
- hypothesising underlying spatial processes
- testing the validity of the theory

Tobler's First Law of Geography

- to a large degree spatial analysis is premised on this
- Waldo Tobler (1979) noted,
- 'everything is related to everything else, but near things are more related than distant things'

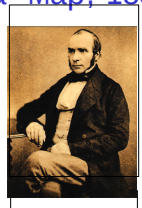
- formalised as the concept of spatial autocorrelation
- thus where things are can be important aspect in any explanation
- do you think there is a 'where' process involved in the phenomena you are studying?
- but be wary of simplistic geographic determinism

Descriptive analysis of geographic features



(source: Clarke 1999, 55)

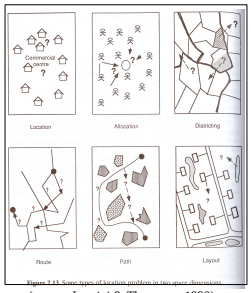
Dr John Snow's 'Cholera' Map, 1854



- combined layers of water pump locations and cholera deaths
- geographic space as backdrop. space is part of the explanation
- testing for *spatial patterns* in cases and *proximity* to likely sources. Spatial properties of data reveal the underlying processes of disease transmission
- analysis is done 'manually' using the visualisation power of eye-brain
- could calculate a measure of clustering in GIS to formerly test the hypothesis (cholera caused by polluted water)
- medical geography is key area for the application of spatial analysis. epidemiologists are heavy GIS users

Spatial relationships

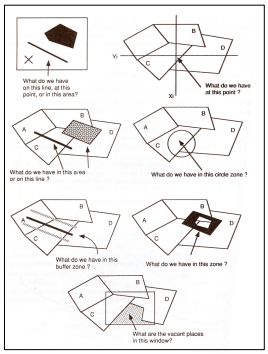
• look at the data in terms of spatial relationships (what is where and how are they related across space)
• go beyond looking to measuring and description to analysing nature and strength of spatial relationships, proscribing changes



(source: Laurini & Thompson 1992)

Queries & reasoning

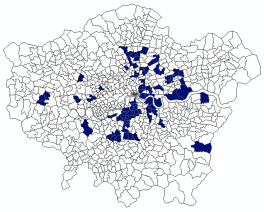
- asking questions of the data
- We used the 'i' information tool to look at attributes of specific objects on the map
- how many universities are in Camden?
- no changes are made to the data. no new data layers are produced



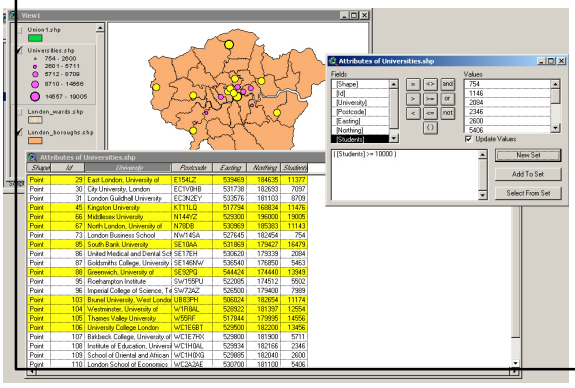
Using a subset of features

- often want to work with a specific selection of theses object, e.g. map only those universities with more than 10,000 students
- this is done by Selections on the attribute database
- 3 main types of selections
  - attribute query,
  - geographic query,
  - interactively

Where are the 'worst' places?



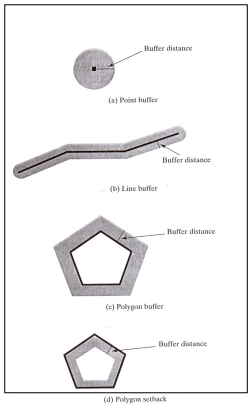
Attribute query, ([Students] >= 10000)



Transformations

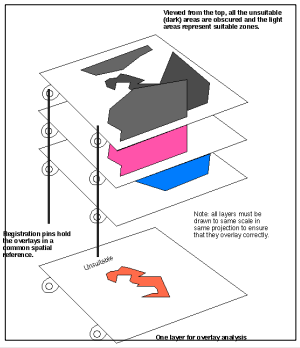
- analytical operations that change the data and usually result in new data layers
- buffering is a common and useful example
- create a user specified zone around a given object
- buffer zone is a notional 'area of influence' and allows you to do selections
- e.g. draw a 5 km buffer around universities and calculate how many people life within the zone
- buffer is usually a uniform linear distance and takes shape of the original object

Buffering types



Overlay analysis

- manual example from school geography of tracing paper layered onto a map



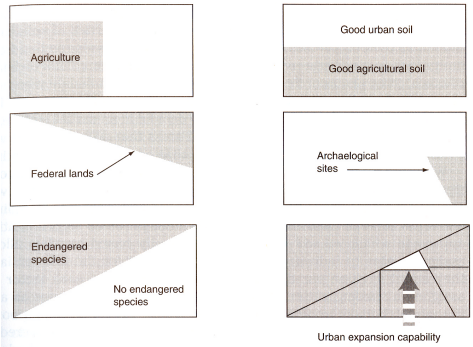


Figure 12.3 Logical "or" overlay. Shaded polygons showing relationship among polygons that are to be zoned against urban expansion. This illustrates the use of the logical "or" search in raster overlay operations.

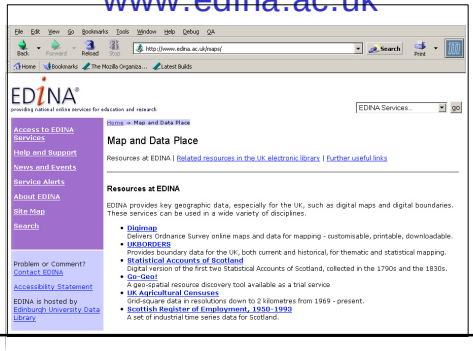
GIS software availability for university research

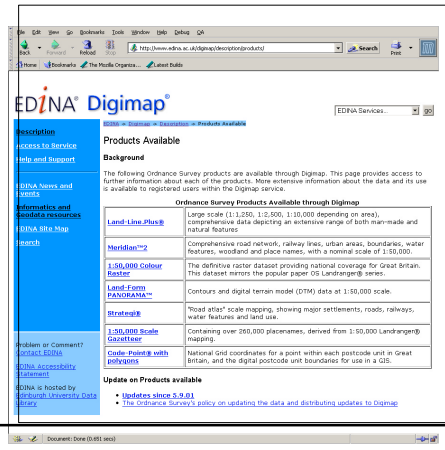
- ESRI products, www.esri.com, www.esriuk.com
  - ArcGIS (arcview or arcinfo)
- MapInfo, www.mapinfo.com
  - Mapinfo Professional
- Microsoft's MapPoint, www.microsoft.com/mappoint

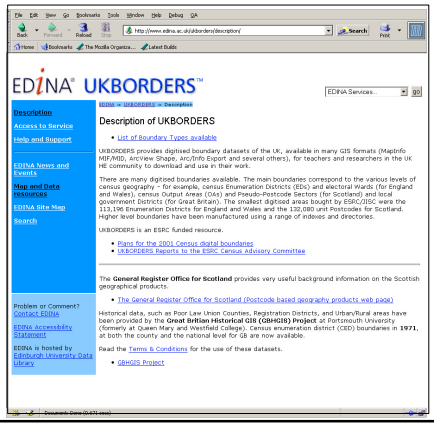
Useful sources of spatial data for social science research

- i.e. data that can be mapped easily in a GIS
- typically means, point location with co-ordinates (e.g. schools, towns), or areas with boundaries (e.g. wards)
- large amount of socio-economic data from government statistics and census
- easily obtain in digital format
- Ordnance Survey base map data
- a lot of environmental data as well
- free for academics (registration, Athens access)

Digimap service at EDINA www.edina.ac.uk

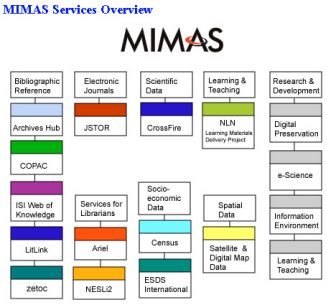




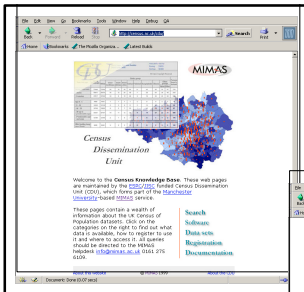


MIMAS (Manchester Information & Associated Services)

www.mimas.ac.uk

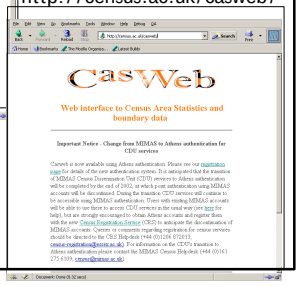


- 1981, 1991, 2001 Census data. Other larger govt. data sets.



http://census.ac.uk/cdu/

http://census.ac.uk/casweb/



Conclusions

- If you do decide to explore further, remember that GIS *analysis* requires more than just "pushing buttons"
- a critically thinking person must be operating the system. It is essential that you
- understand your data and know which questions to ask during the analysis process
  - how and when were the data collected? are they accurate? Is the scale, or precision, of the data appropriate for this type of analysis?
  - If you see a relationship between two mapped variables, which is the cause and which is the effect? Is there some other variable, or many variables, affecting a distribution that is not in the database?
- remember the dangers of GIGO

Further reading

- Paul Longley, et al. (2001) *Geographic Information: Systems and Science*
- Andy Mitchell (1999) *The ESRI Guide to GIS Analysis, Volume 1: Geographic Patterns & Relationships*
- David O'Sullivan and David Unwin (2002) *Geographic Information Analysis*