

## Introduction to GIS

Martin Dodge

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

[http://www.casa.ucl.ac.uk/martin/ss\\_methods/](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

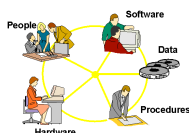
- why take a spatial perspective?
  1. space provides a convenient conceptual and practical frame for organising large amounts of data
  2. provides a means of linking together different data based on location
  3. provides a means of linking to secondary data source by location
  4. provides easy access 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

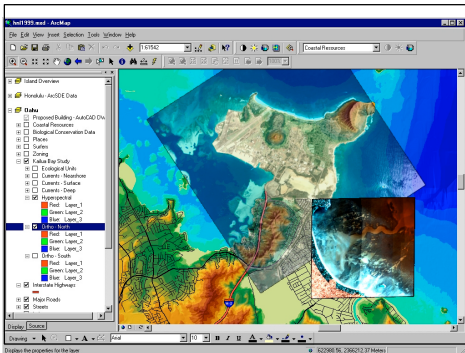
Definitions of GIS, and the groups who find them useful

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)

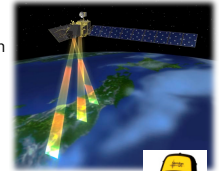
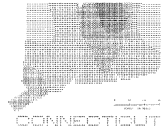
- GIS software is a major global business
- major software vendors include : Intergraph Corp., SmallWorld Systems Ltd., Environmental Systems Research Institute, Autodesk Inc., MapInfo Corp.
- also data, consultancy and training provided commercially. education - UCL's GIS courses for example

## 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

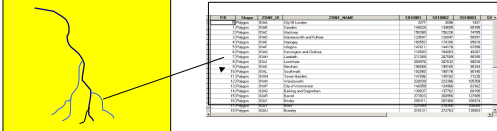
- 70-80% of the tasks undertaken by local government are geographically related
- still the biggest single group of GIS professionals
- drive to improve quality of products, processes and services
- inventory resources and infrastructure
- plan transportation routing
- improve service response time
- 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.'

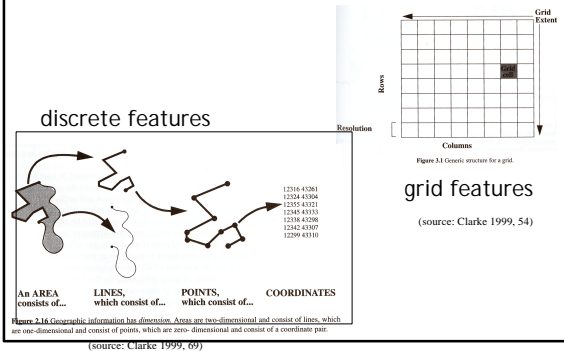
general purpose tool for particular kind of data  
just like word processor for text, paint program for photos

GIS -> data + map interface + tools = solution

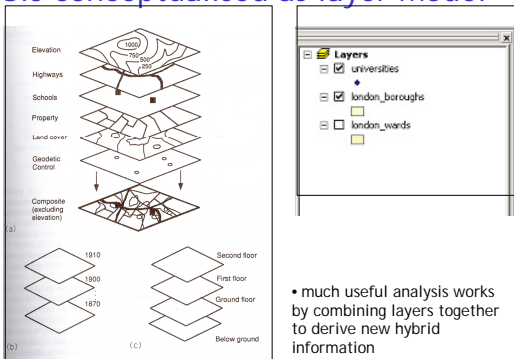


Crucially, the graphic elements of the map (e.g. line for a river) is linked to separate data. this is known attribute data and it is used to describe the nature of the map element (eg. flow rate, pollution level of river)

## Representing spatial features



## GIS conceptualised as layer model



## 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

	Random		Systematic	
	Points or Quadrats	Transects	Points or Quadrats	Transects
Homogeneous				
Stratified				

## Spatial analysis

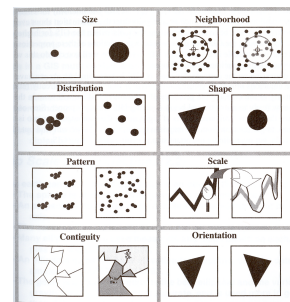
- spatial data analysis is the *quantitative* study of phenomena that manifest themselves in space, aiming to *describe* the nature of their distribution and understand the *processes* that give rise to it
- it is usually premised on scientific methods of enquiry, hypothetical-inductive methods
- exploring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



## Tobler's First Law of Geography

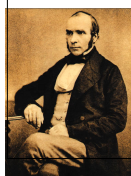
- to a large degree spatial analysis is premised on this
- Waldo Tobler (1979) noted,
- 'everything is related to everything else, but near things are more related than distant things'
- formalised as the concept of spatial autocorrelation
- thus where things are can be important aspect in any explanation
- 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)

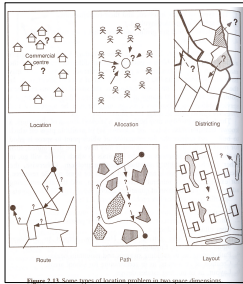
## Dr John Snow's 'Cholera' Map, 1854



- 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

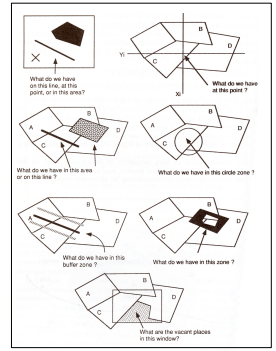
- look at the data in terms of spatial relationships (what is where and how are they related across space)
- go beyond looking to measuring and description to analysing nature and strength of spatial relationships, proscribing changes



(source: Laurini & Thompson 1992)

## Queries & reasoning

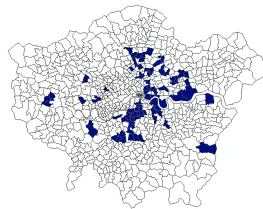
- asking questions of the data
- We used the 'I' information tool to look at attributes of specific objects on the map
- how many universities are in Camden?
- no changes are made to the data, no new data layers are produced



## Using a subset of features

Where are the 'worst' places?

- often want to work with a specific selection of these 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



## Attribute query, ([Students] >= 10000)

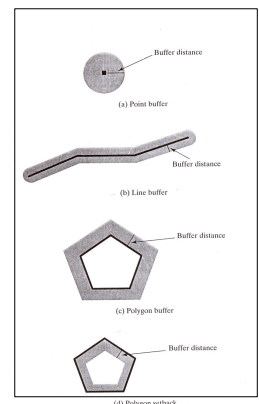
Screenshot of a GIS software interface showing an attribute query for universities with 10,000 or more students. The query is: `[Students] >= 10000`. The results table is as follows:

Point	City	University	Postcode	Enroll	Month	Student
Point 25	East London	University of E1541Z	E1541Z	53980	18453	11377
Point 30	City	University, London	EC1W9B	53179	18203	7097
Point 31	London	Goldsmith University	E32QJY	53276	18103	8929
Point 45	Kingston	University	K113LD	51759	18884	11478
Point 52	Medicine	University	N144JZ	52000	18020	18000
Point 67	North London	University of N708B	53960	18533	11143	
Point 73	London Business School	NW41E3	52145	18245	754	
Point 85	South Bank	University	SE18AA	51183	17927	16479
Point 86	United Medical and Dental Sci	SE17SN	53020	17939	2084	
Point 97	Goldsmith College	University	SE149BA	53540	17890	5463
Point 98	Imperial College of Science	SW72BZ	54444	17448	1948	
Point 99	Imperial College of Science	T4 SW72AZ	52000	17940	789	
Point 103	Imperial University	West London	UB8PH	53024	18204	11174
Point 104	University	University of V419AA	52925	18157	12564	
Point 105	Thomas Telford	University	W99FF	51744	17996	4356
Point 106	University College	University of V4218BT	50963	18020	13964	
Point 107	Birkbeck College	University of V421E79C	52900	18190	5711	
Point 108	Institute of Education	University of V421H4AL	52913	18216	2441	
Point 109	School of Oriental and African	University of V421H45D	52905	18240	300	
Point 110	London School of Economics	V421D34E	53700	18100	546	

## Transformations

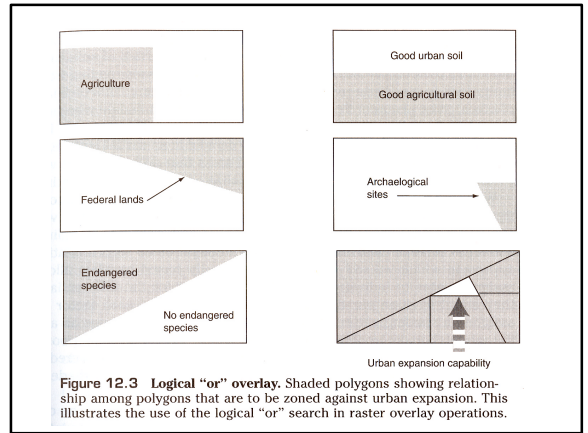
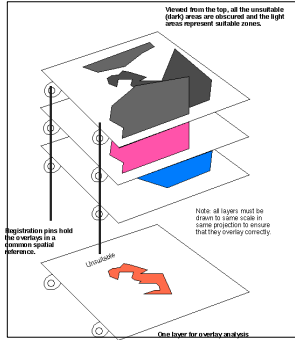
- analytical operations that change the data and usually result in new data layers
- buffering is a common and useful example
- create a user specified zone around a given object
- buffer zone is a notional 'area of influence' and allows you to do selections
- e.g. draw a 5 km buffer around universities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



# GIS software availability for university research



- ESRI products, [www.esri.com](http://www.esri.com), [www.esriuk.com](http://www.esriuk.com)
  - ArcGIS (arcview or arcinfo)



- MapInfo, [www.mapinfo.com](http://www.mapinfo.com)
  - Mapinfo Professional



- Microsoft's MapPoint, [www.microsoft.com/mapoint](http://www.microsoft.com/mapoint)



# 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](http://www.edina.ac.uk)

**EDINA**  
enabling better online services for education and research

**Map and Data Place**  
Resources at EDINA | [Related resources in the UK electronic library](#) | [Further useful links](#)

**Resources at EDINA**  
EDINA provides key geographic data, especially for the UK, such as digital maps and digital boundaries. These services can be used in a wide variety of disciplines.

- **Digimap** - Delivers Ordnance Survey online maps and data for mapping - customisable, printable, downloadable.
- **UKBOUNDRIES** - Provides boundary data for the UK, both current and historical, for thematic and statistical mapping.
- **Historical Accounts of Scotland** - Digital version of the first two Historical Accounts of Scotland, collected in the 1790s and the 1830s.
- **Geo-Que** - A geo-spatial resource discovery tool available as a trial service
- **UK Agricultural Census** - Grid-square data in resolutions down to 2 kilometres from 1969 - present.
- **Scottish Register of Employment, 1950-1993** - A set of industrial time-series data for Scotland.

**EDINA Digimap**  
enabling better online services for education and research

**Products Available**

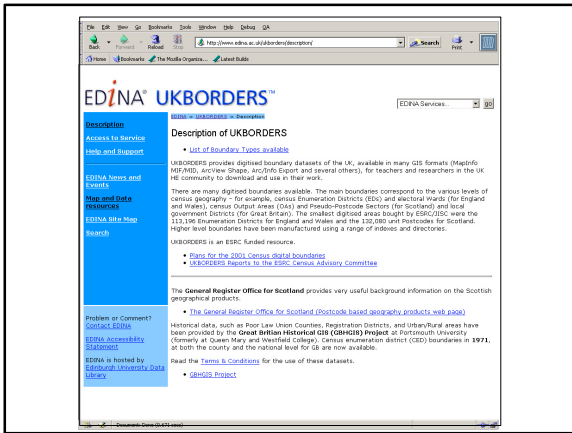
**Background**  
The following Ordnance Survey products are available through Digimap. This page provides access to further information about each of the products. More extensive information about the data and its use is available to registered users within the Digimap service.

**Ordnance Survey Products Available through Digimap**

Product Name	Description
<a href="#">Land Line Plans</a>	Large scale (1:1,250, 1:1,500, 1:10,000 depending on area), comprehensive data depicting an extensive range of both man-made and natural features.
<a href="#">Mediterranean</a>	Comprehensive road network, railway lines, urban areas, boundaries, water features, woodland and place names, with a nominal scale of 1:10,000.
<a href="#">1:50,000 Colour Master</a>	The definitive raster dataset providing national coverage for Great Britain. This dataset mirrors the popular paper OS Landranger® series.
<a href="#">Land Form</a>	Contours and digital terrain model (DTM) data at 1:10,000 scale.
<a href="#">Mediterranean</a>	"Road start" scale mapping, showing major settlements, roads, railways, water features and land use.
<a href="#">1:50,000 Single Channel</a>	Containing over 260,000 placenames, derived from 1:50,000 Landranger® mapping.
<a href="#">Code Point with Polygons</a>	National Grid coordinates for a point within each postcode unit in Great Britain, and the digital postcode unit boundaries for use in a GIS.

**Update on Products available**

- updates since 5.0.0.1
- the Ordnance Survey policy on updating the data and distributing updates to digimap



## MIMAS (Manchester Information & Associated Services)

www.mimas.ac.uk

### MIMAS Services Overview

Bibliographic Reference	Electronic Journals	Scientific Data	Learning & Teaching	Research & Development
Archives Hub	JSTOR	CrossFire	NLM Learning Materials Delivery Project	Digital Preservation
COPAC				e-Science
ISI Web of Knowledge	Services for Librarians	Socio-economic Data	Spatial Data	Information Environment
LILink	Arnel	Census	Satellite & Digital Map Data	Learning & Teaching
zotoc	NESL2	ESDS International		

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

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

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

## 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*