What is GIS?

Ian Gregory

Digital Humanities

Lancaster University

Structure of talk

- 1. Define GIS and related terms
  - Geographic Information
  - GIS as:
    - a type of software
    - a set of tools
    - an approach to analysis
- 2. Data in GIS
  - Spatial and attribute
  - Geo-referencing data
  - Raster and vector
  - Layers of data
- 3. Querying a GIS database
- 4. Integrating data with GIS

Geographical Information (GI)

- Information that refers to a location on the Earth's surface
  - Has both a spatial and a thematic component
    - Census data
    - Hospitals admissions data
    - Relief data (e.g. from contours)
    - Information on transport networks
    - A text about a specific place (egST Coleridge's tour of the Lake District)
    - A collection of photographs or paintings of buildings
  - Locational component can be an explicit (e.g. a co-ordinate or a precisely defined administrative unit) or vaguer (e.g.. "The area around London"or "In Gaelic speaking areas")

Geographical Information Systems (GIS)

- 1. GIS: A type of software
  - A computer system that allows us to handle information about the location of features or phenomena on the Earth's surface
  - Has the functionality of a conventional DBMS PLUS functionality to handle the spatial component of the data (manipulating, mapping analysing).
  - GIS as a DBMS that allows us to explicitly handle the spatial
  - Common examples:
    - ArcView
    - ArcGIS
    - MapInfo

Geographical Information Systems (2)

- 2. GIS: A tool-kit
  - Manipulate spatially:
    - Calculate distances and adjacencies
    - Change projections and scales
    - Integrate disparate sources
  - Analyse spatially:
    - Quantitative analysis
    - Exploratory spatial data analysis
    - Qualitative analysis
  - Visualise data:
    - Maps!
    - Tables, graphs, etc.
    - Animations
    - Virtual landscapes

Geographical Information Systems (3)

- 3. Approach:
  - Explore the database:
    - In conventional ways
    - AND geographically
  - Allows us to think about the implications of location
  - Allows us to think holistically
  - Should not be restricted by vendor-provided functionality
  - Should be used imaginatively taking into account :
    - the advantages and limitations of geographical information
    - the traditions of your discipline

Types of data

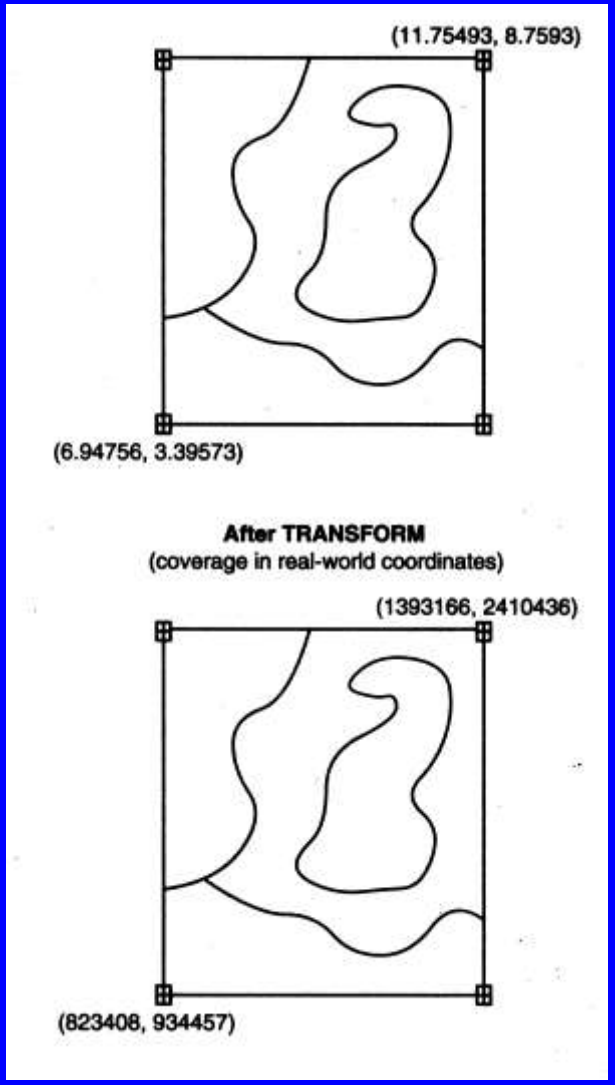
- Two types of data are stored for each item in the database

- 1. Attribute data:
  - Says ***what*** a feature is
    - Eg. statistics, text, images, sound, etc.
- 2. Spatial data:
  - Says ***where*** the feature is
  - Co-ordinate based
  - Vector data – discrete features:
    - Points
    - Lines
    - Polygons (zones or areas)
  - Raster data:
    - A continuous surface

Geo-referencing data

- Capturing data
  - Scanning: all of map converted into raster data
  - Digitising: individual features selected from map as points, lines or polygons
- Geo-referencing
  - Initial scanning digitising gives co-ordinates in inches from bottom left corner of digitiser/scanner
  - Real-world co-ordinates are found for four registration points on the captured data
  - These are used to convert the entire map onto a real-world co-ordinate system

Example of geo-referencing

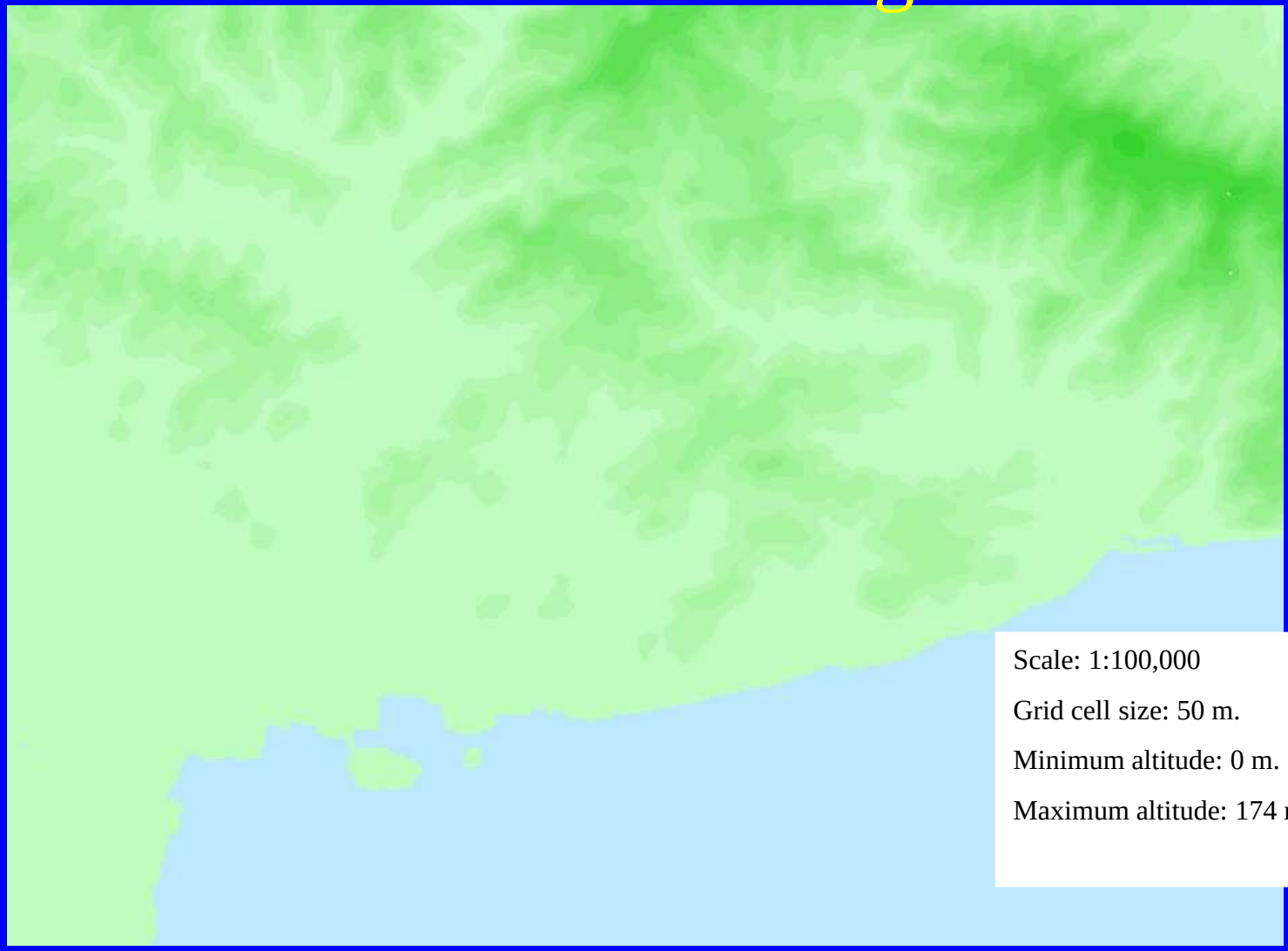


Source: ESRI (1997)

Layers

- Data on different themes are stored in separate “layers”
- As each layer is geo-referenced layers from different sources can easily be integrated using location
- This can be used to build up complex models of the real world from widely disparate sources

Raster data: Hastings



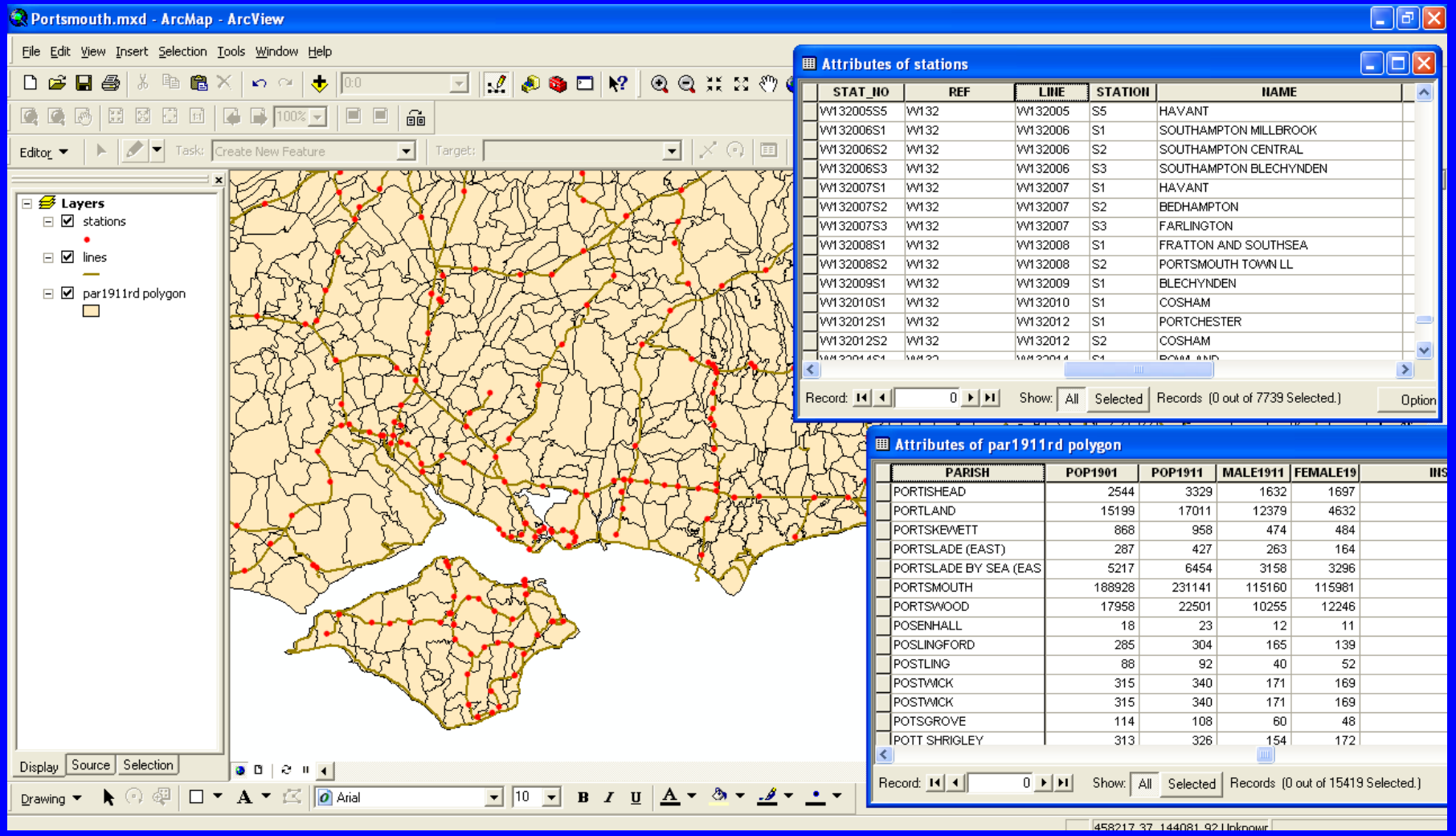
Scale: 1:100,000

Grid cell size: 50 m.

Minimum altitude: 0 m.

Maximum altitude: 174 m.

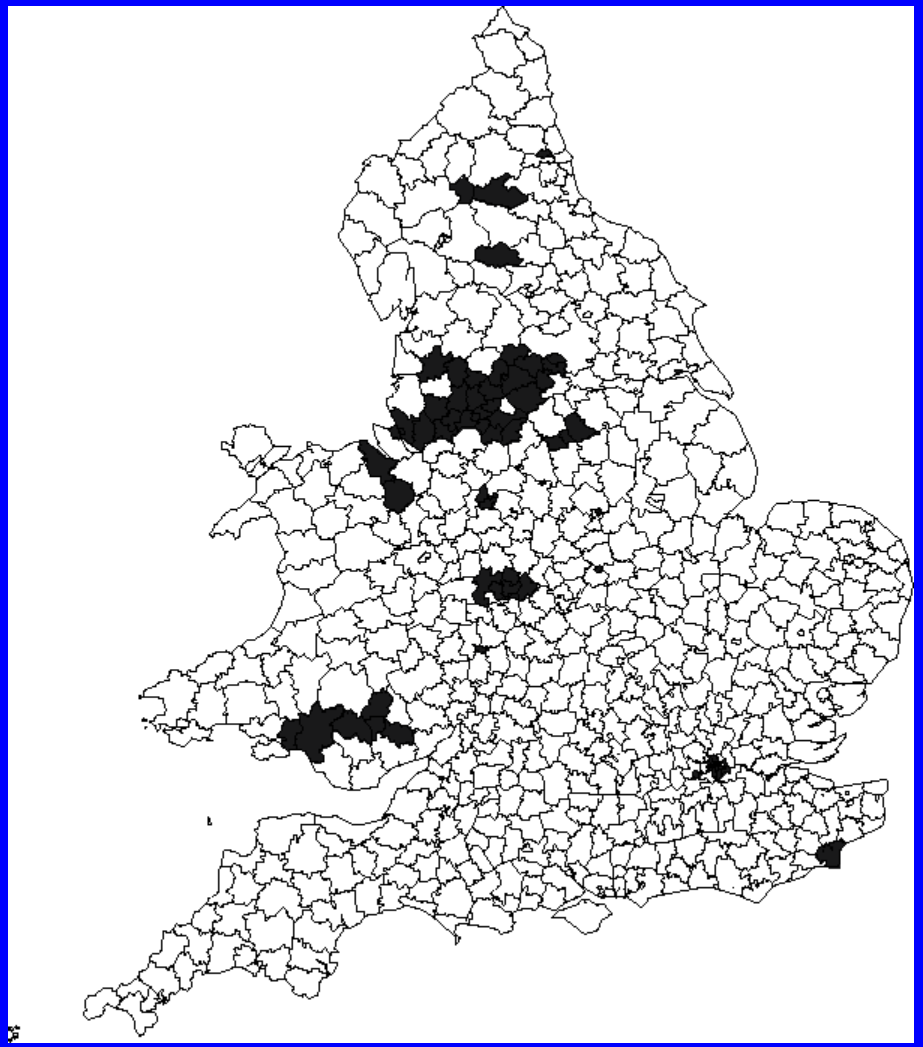
Example: Vector data



Querying GIS data

- Attribute query
  - Select features using attribute data (e.g. using SQL)
  - Results can be mapped or presented in conventional database form
  - Can be used to produce maps of subsets of the data or choropleth maps
- Spatial query
  - Clicking on features on the map to find out their attribute values
  - Which features on one layer intersect with features from another?
- Used in combination these are a powerful way of exploring spatial patterns in your data

Attribute query: Lung disease in the 1860s



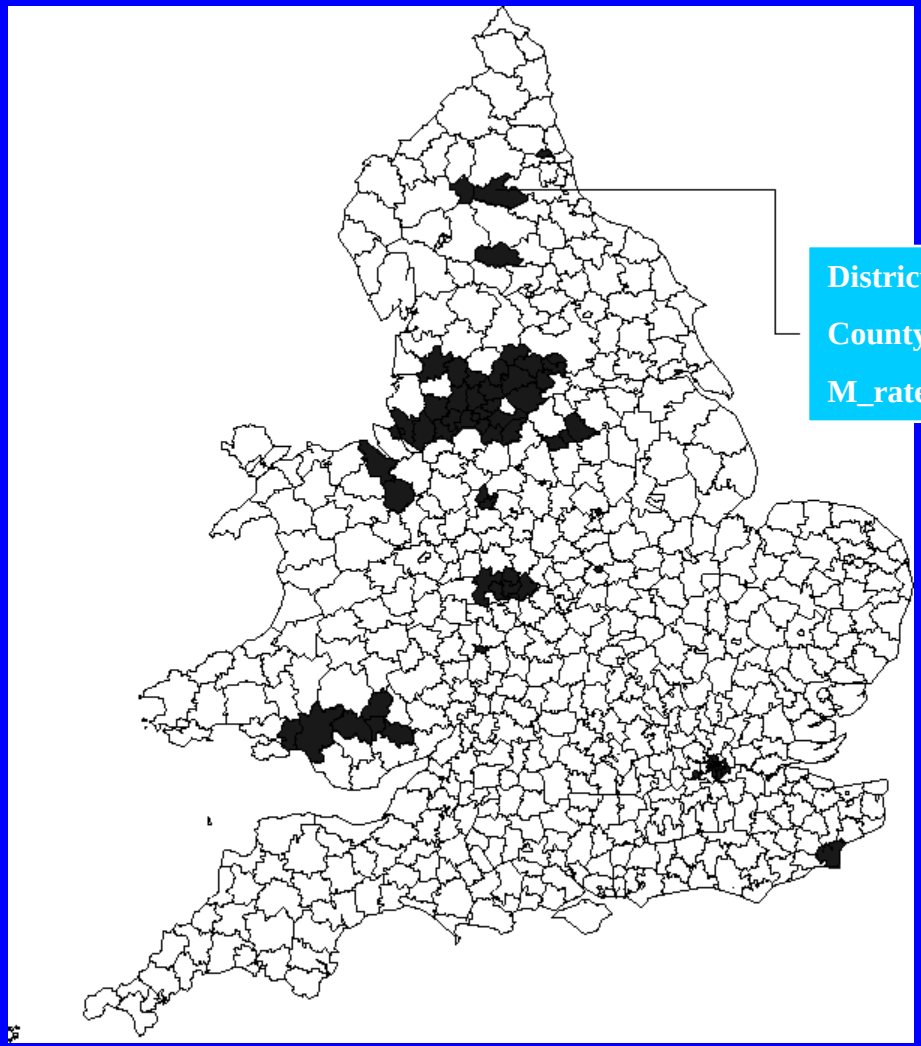
Spatial data: Registration Districts, 1/1/1870

Attribute data: Mortality rate per 1,000 from lung disease among men aged 45-64

Source: Registrar General's Decennial Supplement, 1871

Query: Select areas where mortality rate > 58.0

Spatial query: Lung disease in the 1860s

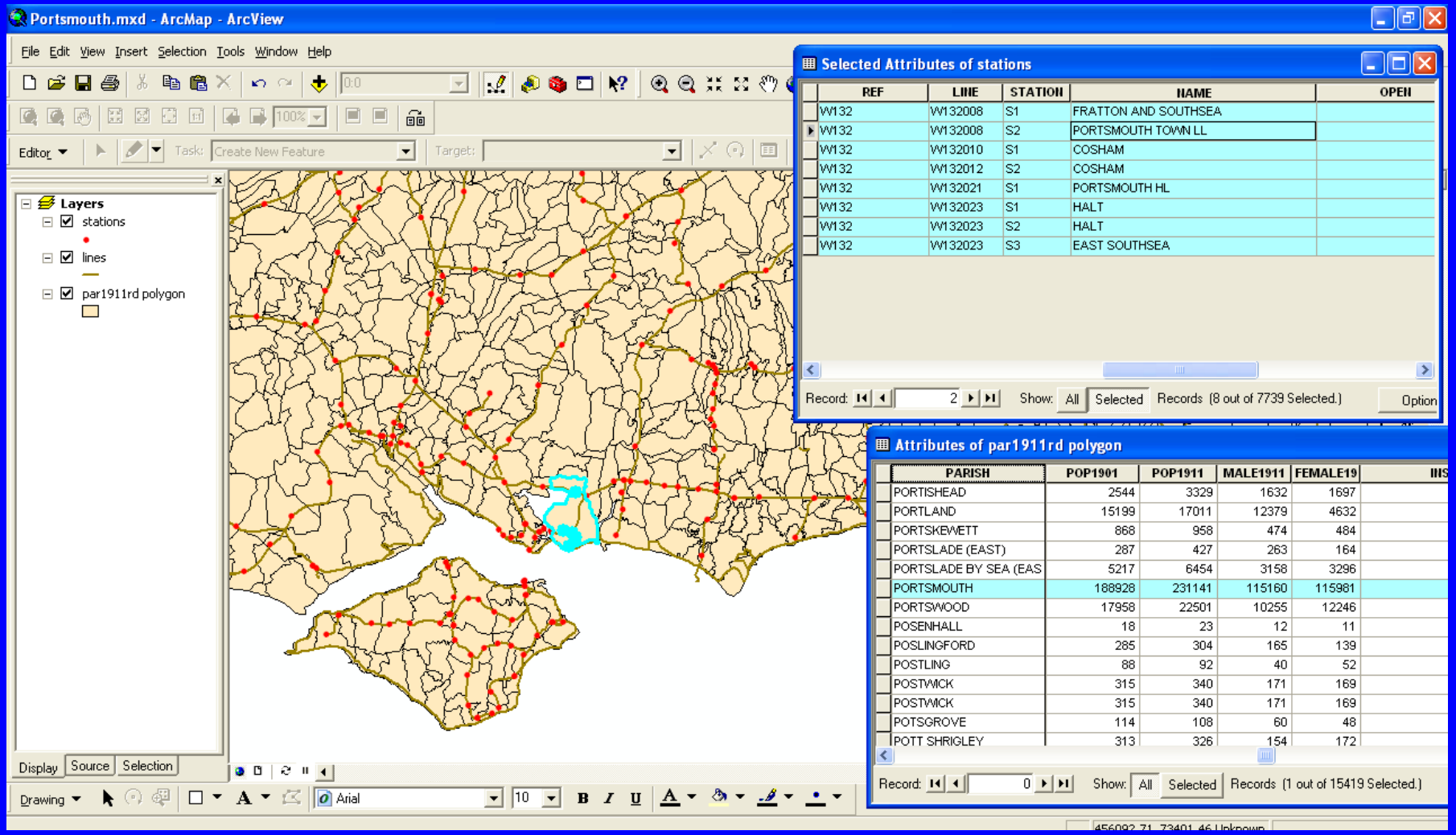


District: Alston with Garrigill

County: Cumberland

M_rate: 68.4

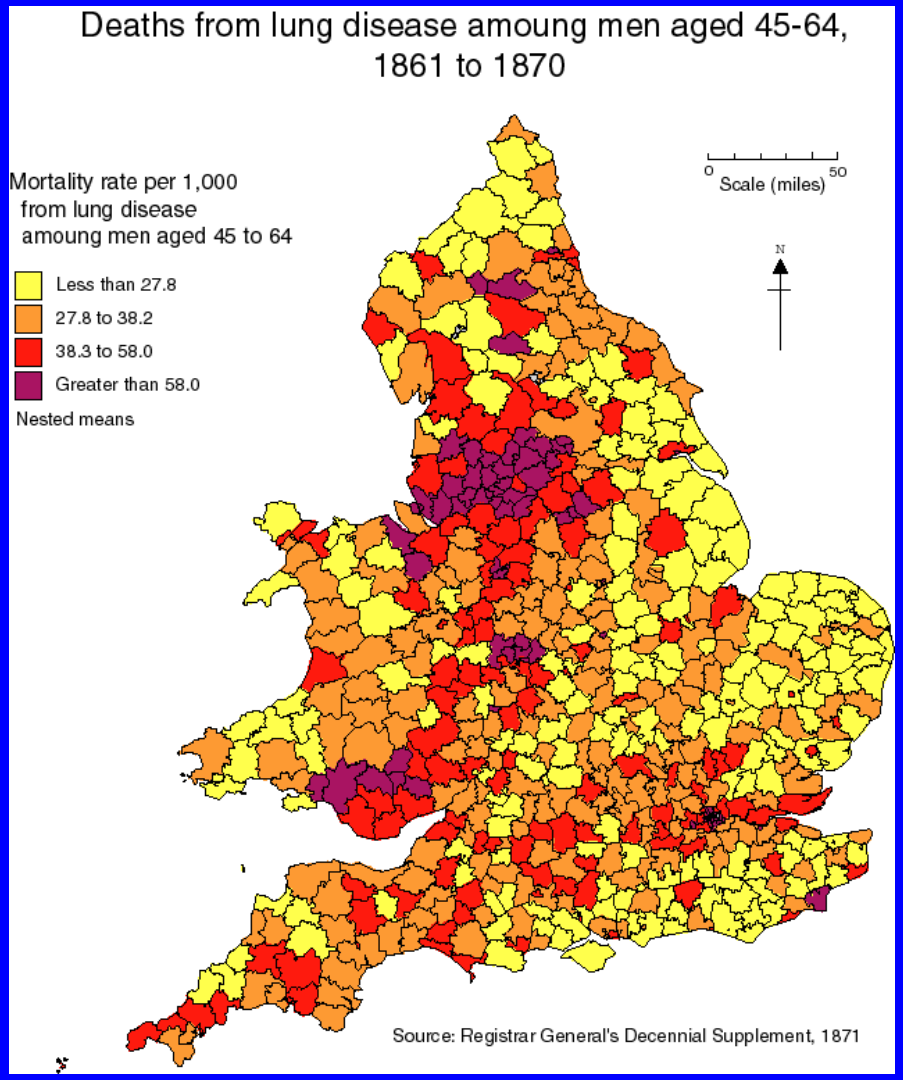
Combined spatial and attribute querying



Par1911: Select parish = ‘PORTSMOUTH’ (Attribute)

Stations: Select stations that intersect with Par1911 (Spatial)

Mapping through attribute query



Conclusions

- Advantages of GIS
  - Exploring both geographical and thematic components of data in a holistic way
  - Stresses geographical aspects of a research question
  - Allows handling and exploration of large volumes of data
  - Allows integration of data from widely disparate sources
  - Allows analysis of data to explicitly incorporate location
  - Allows a wide variety of forms of visualisation
- Limitations of GIS
  - Data are expensive
  - Learning curve on GIS software can be long
  - Shows spatial relationships but does not provide absolute solutions
  - Origins in the Earth sciences and computer science. Solutions may not be appropriate for humanities research

From GIS to Historical GIS

Components of data

- 1. Attribute (theme)
  - Says what the data is
    - Statistical, textual, image, etc.
- 2. Temporal
  - Says when the data existed/are relevant for
- 3. Spatial
  - Say where the data refer to
    - Can be precisely defined “at grid reference (x,y)”
    - Can be vague “in the west of Ireland”
  - Difficult to handle
    - On paper
    - In a database
  - Neglected

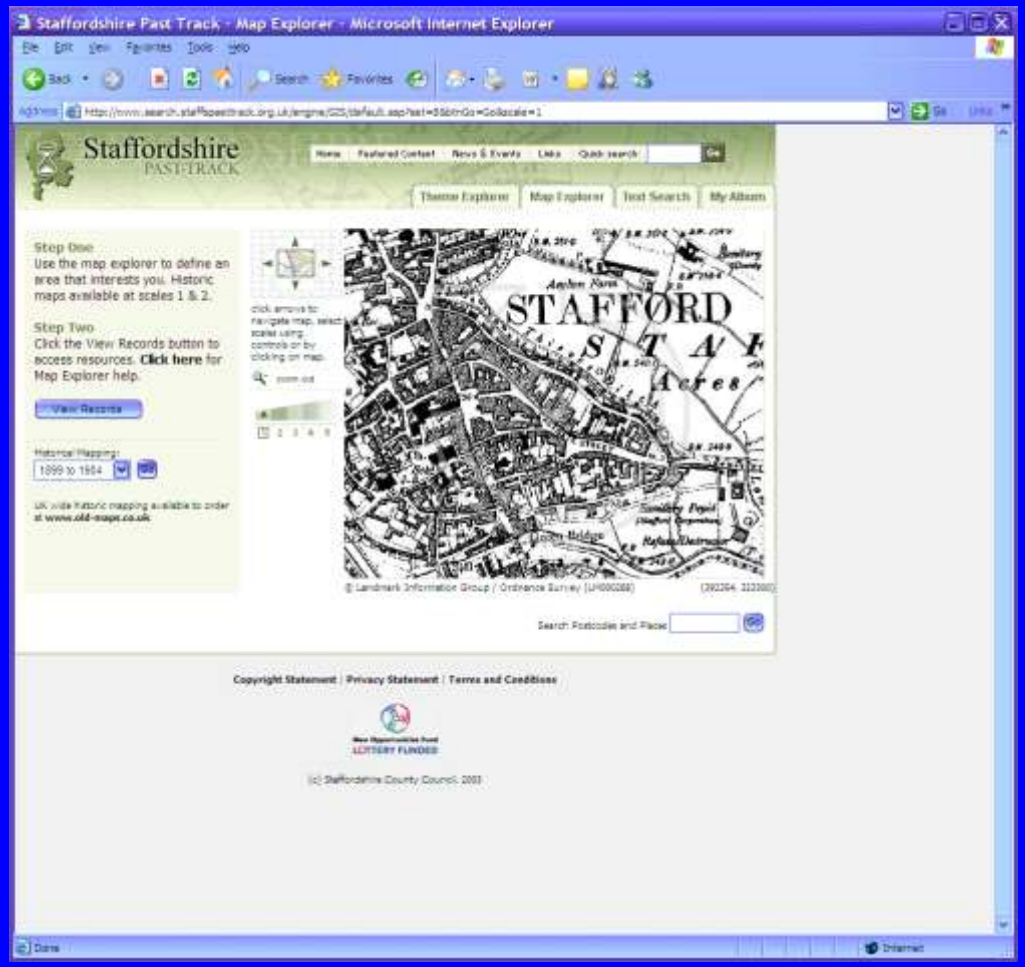
Advantages of GIS

- 1. Structures a database
- 2. Data integration
- 3. Data visualisation
- 4. Spatial analysis

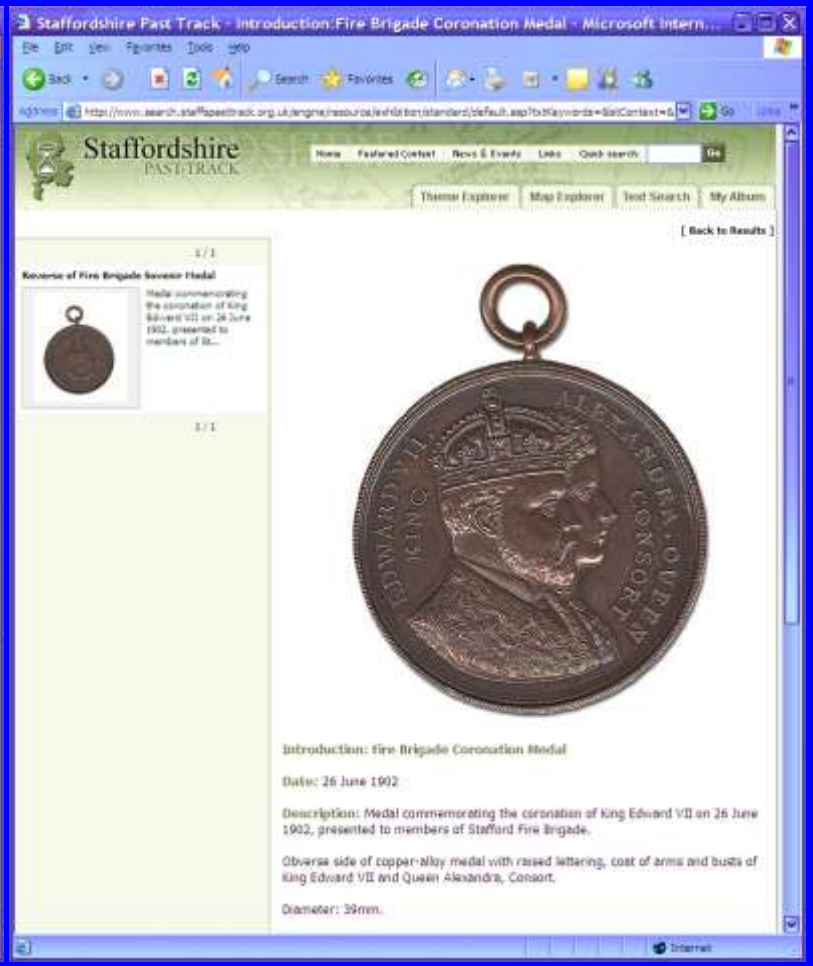
1. Structuring a database

- The Staffordshire Past-Track
  - http://www.staffspasttrack.org.uk/
  - Database of images of Staffordshire's past
    - 7,000 photographs
    - 3,000 images of artwork
    - 75 videos
    - 3 audio files
  - Can be searched by:
    - Theme: eg. Agriculture, Health & Welfare, Parks & Gardens
    - Text: Key words, dates, media types, etc.
    - Maps

Staffordshire Past-Track

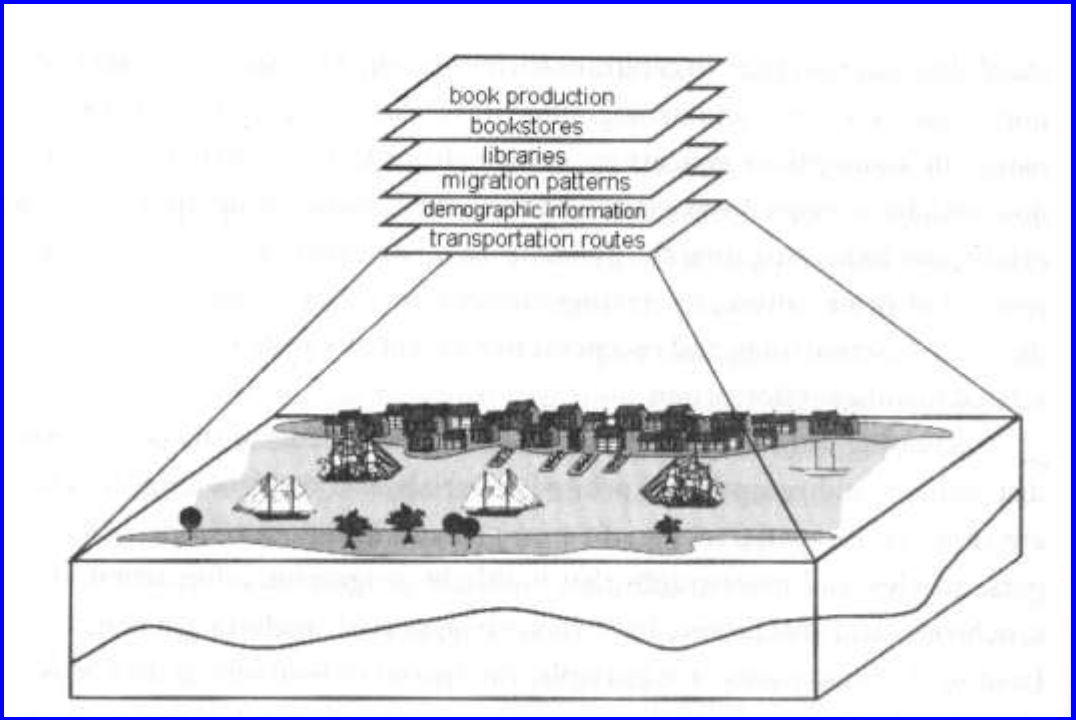


Staffordshire Past-Track



2. Data Integration

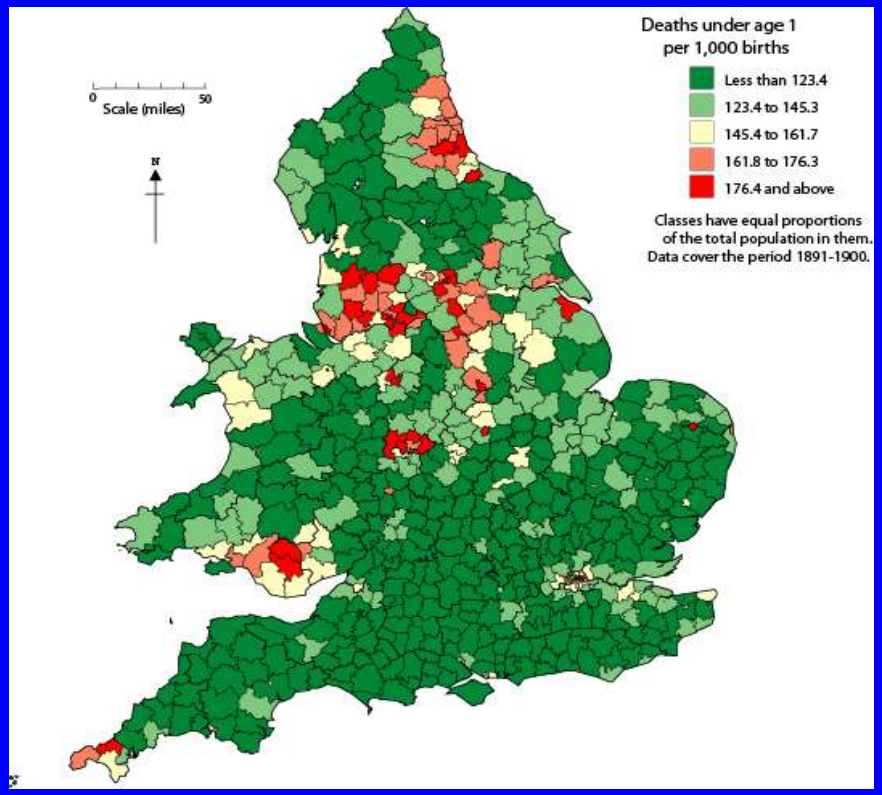
History of the book



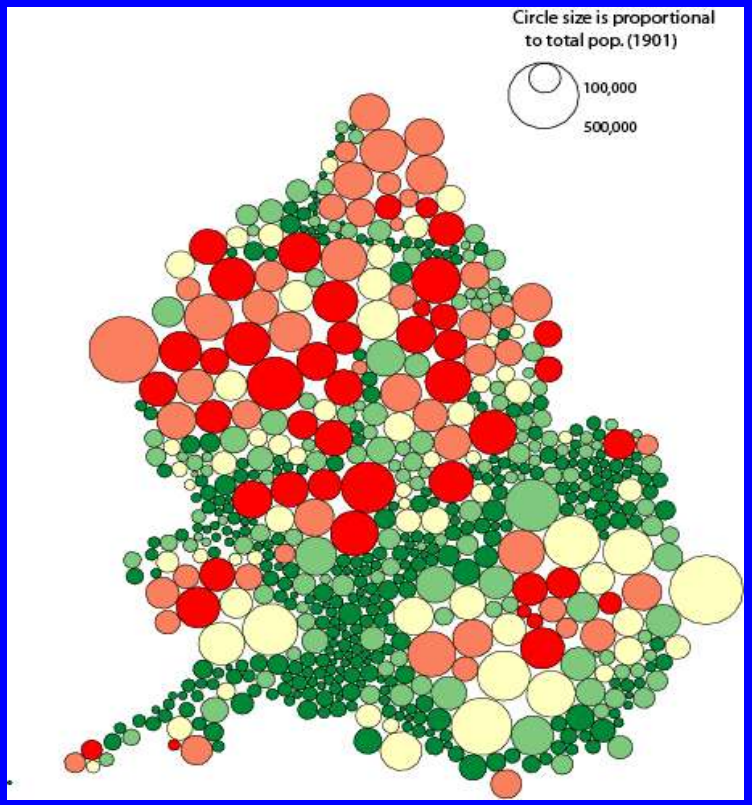
Source: MacDonald B and Black F (2000) "Using GIS for spatial and temporal analyses in print culture studies" Social Science History, 24, pp. 505-536

3. Data Visualisation

Infant mortality, 1900s



Choropleths

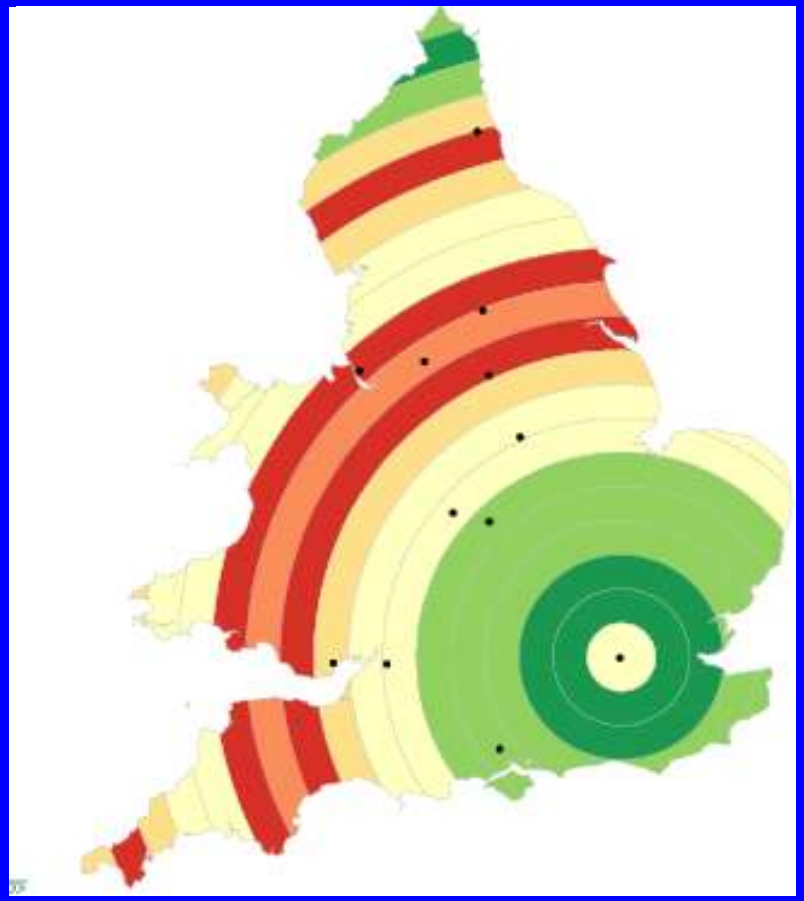


Cartograms

4. Spatial analysis

Infant mortality and the Core-Periphery Divide

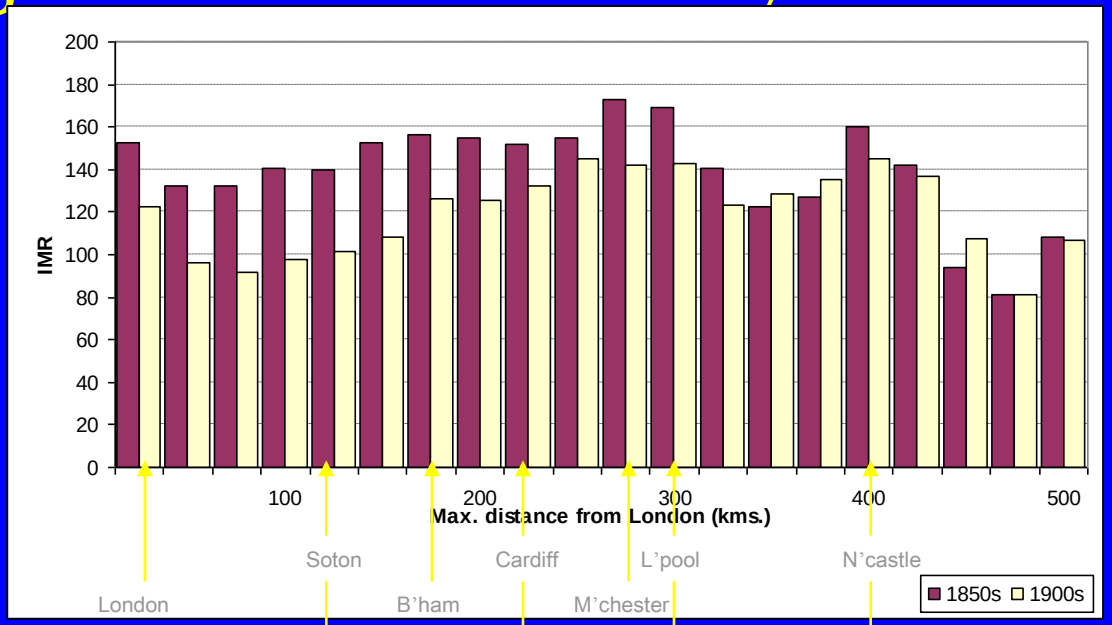
1850s



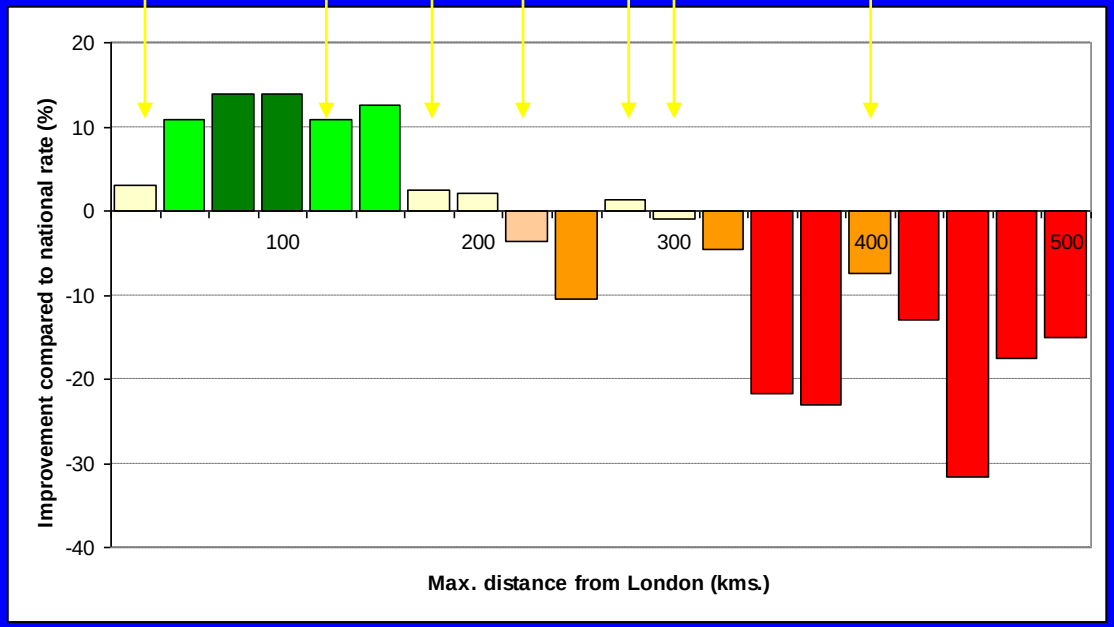
1900s

Change in Infant Mortality 1850s-1900s

Overall



Proportional to national rate



National rate dropped from 153.25 to 127.59 – an improvement of 16.74%

Shading indicates population deciles

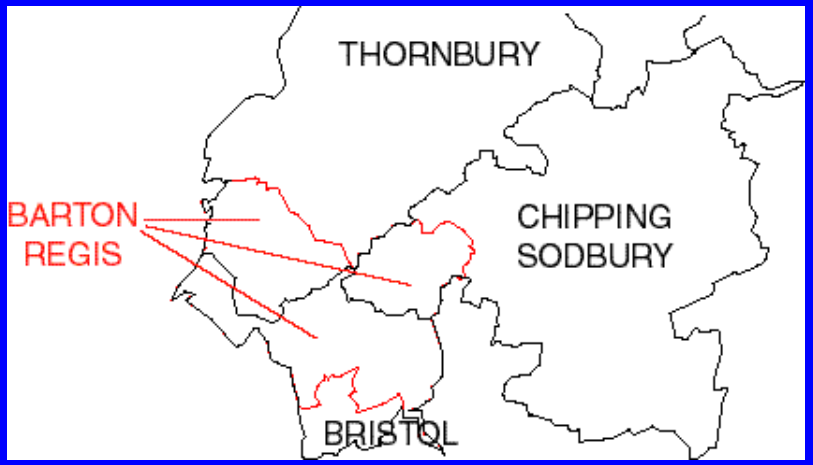
Areal interpolation

- Standardising administrative geographies over time
- Areal interpolation: "the transfer of data from one set (source units) to a second set (target units) of overlapping, non-hierarchical areal units" (Langford *et al*, 1991: p. 56)

- Areal Weighting:

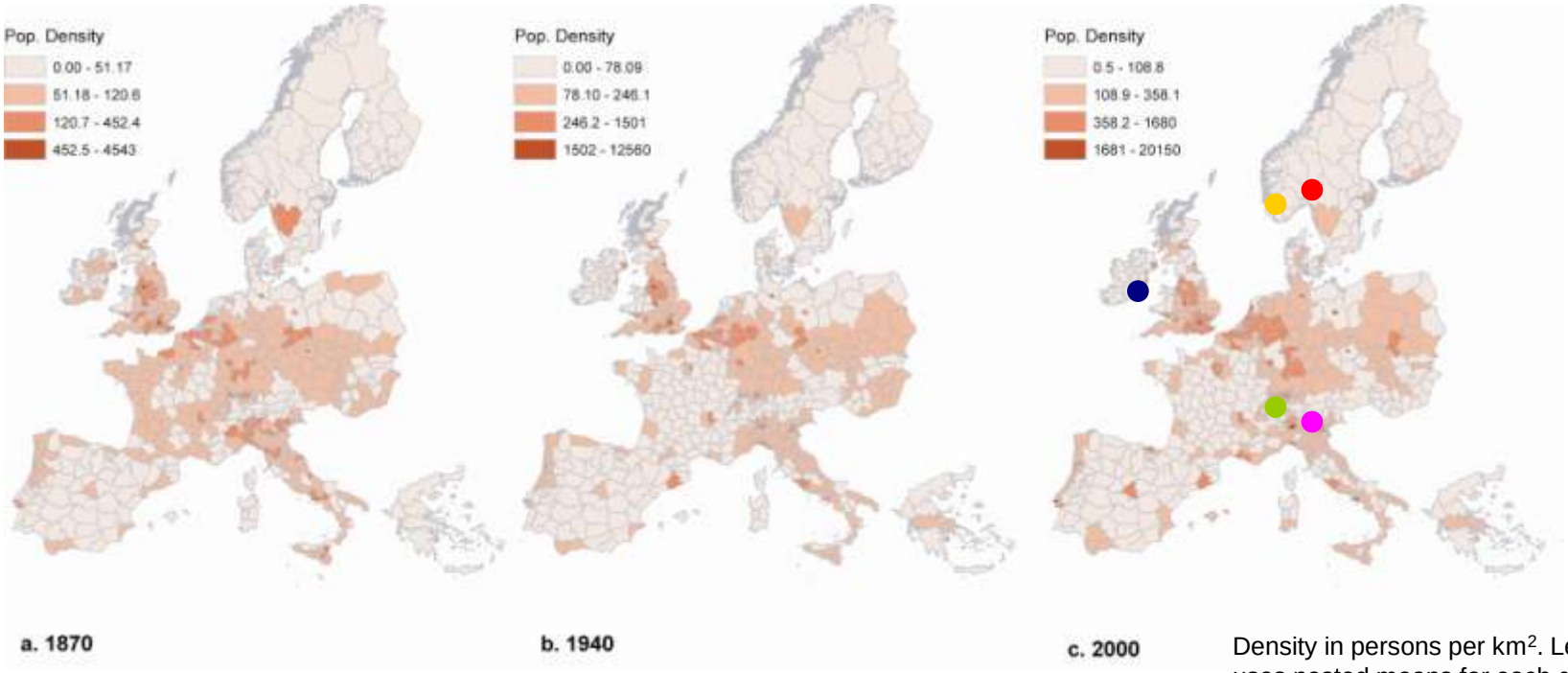
$$\hat{y}_t = \sum_s \left(\frac{A_{st}}{A_s} \times y_s \right)$$

  - Assumption – Variable *y* is homogeneously distributed across the source zones
  - Using this:

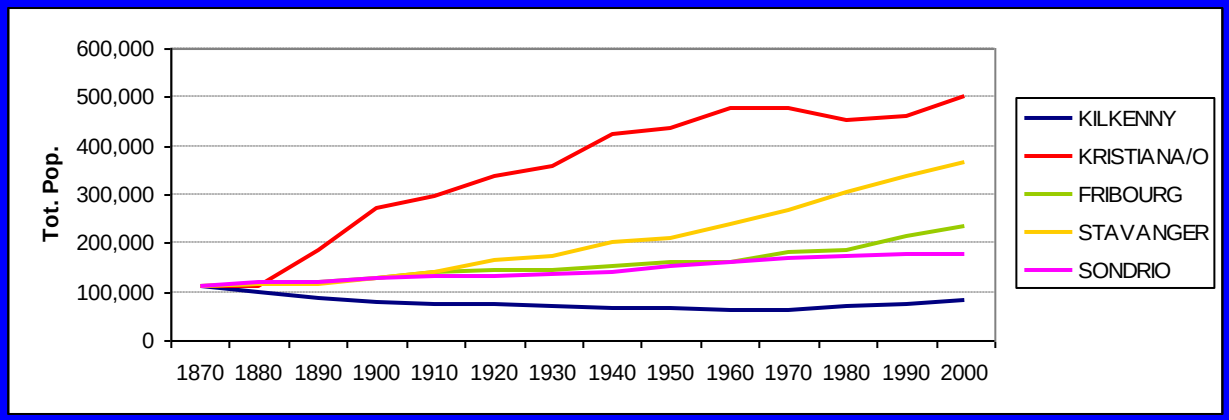


  - BUT: Very unrealistic assumption.

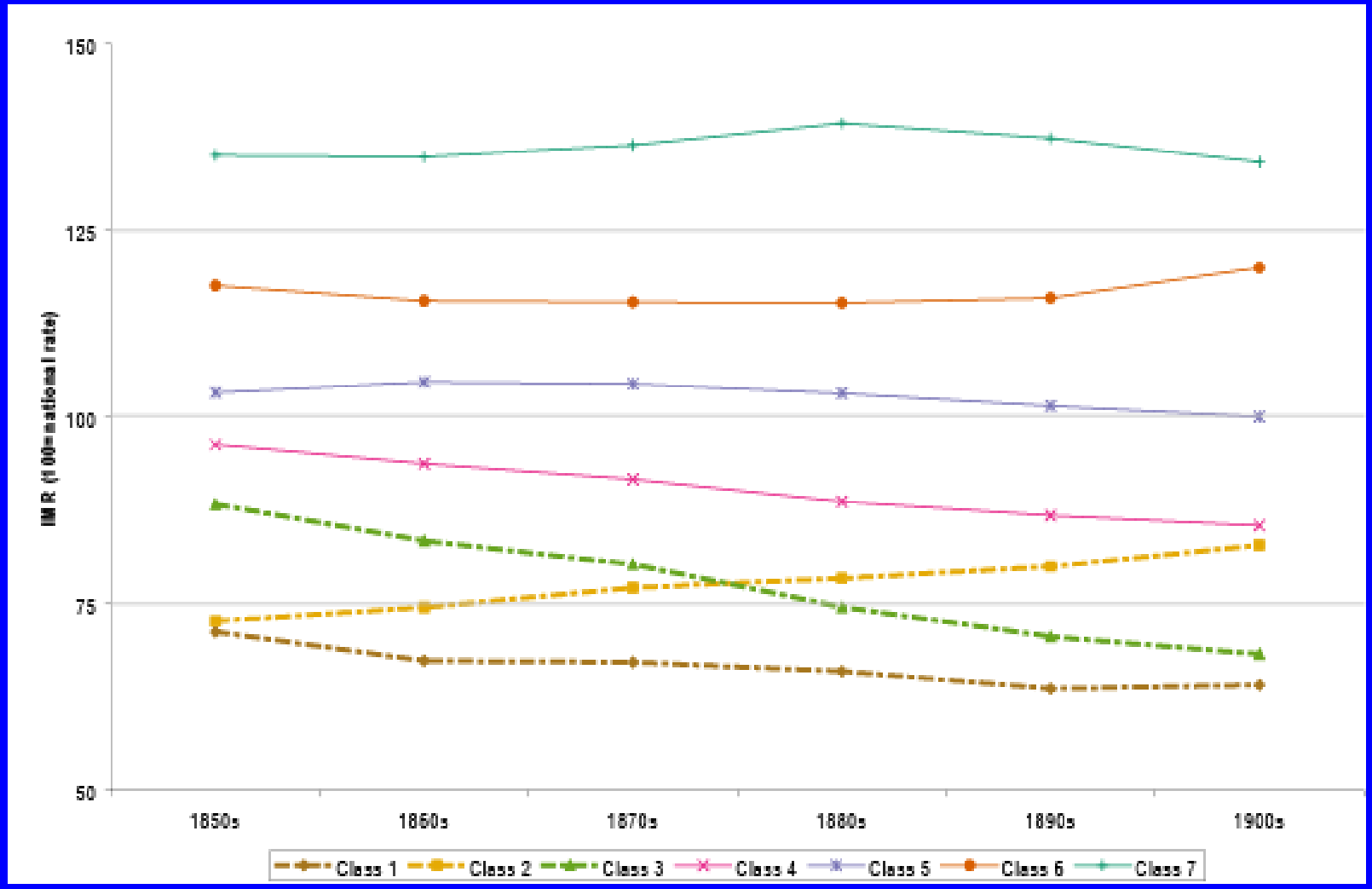
Long-term population change in Europe



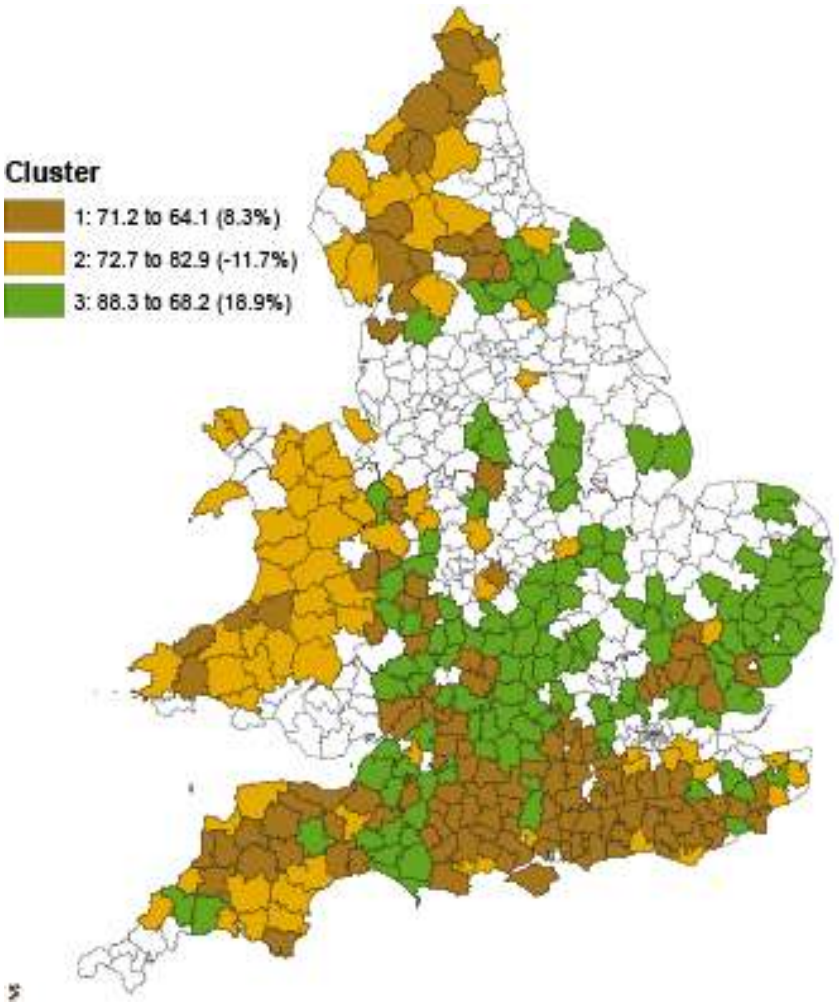
Density in persons per km^2. Legend uses nested means for each date.

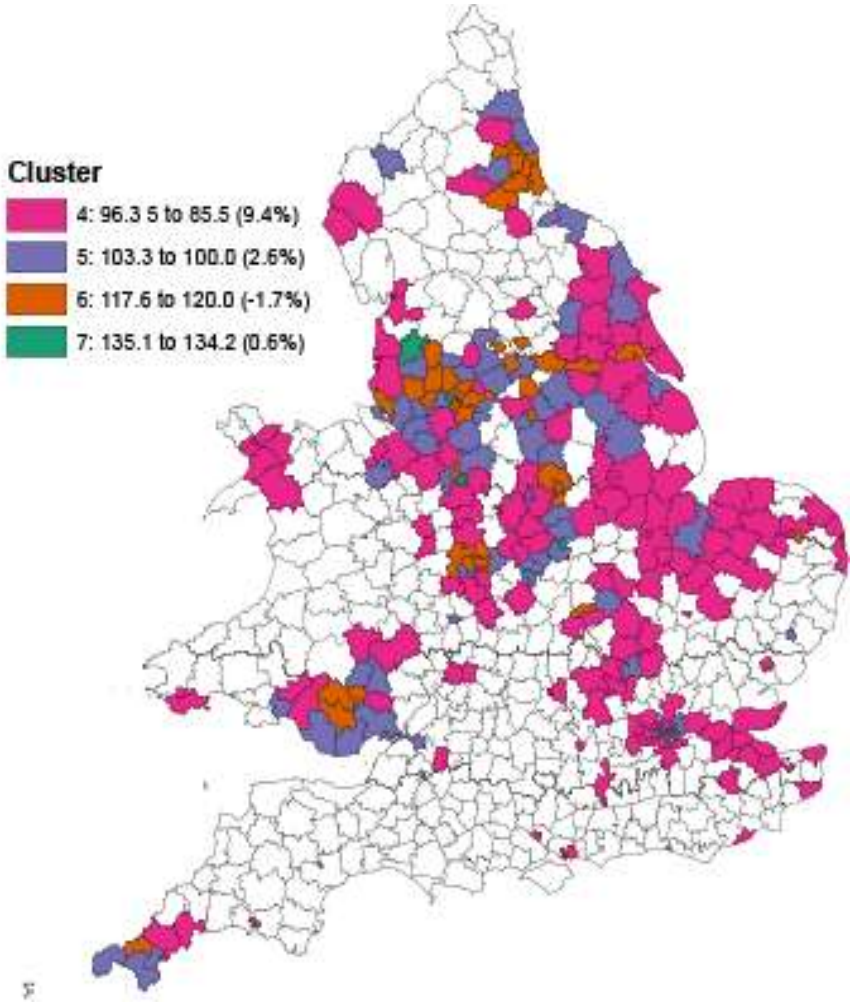


Cluster Analysis: Infant mortality in England & Wales



Maps of Clusters





1. Low – slight improvement
2. Low – gets worse
3. Below average – big improvement

4. Average – slight improvement
5. Average – stays the same
6. High – stays the same
7. Very high – stays the same

Conclusions

- GIS allows improved handling of spatially-referenced data
- Improved understanding of space allows:
  - Structuring
  - Integration
  - Visualisation
  - Analysis
- It does not solve all the problems – in fact it creates many new ones
- "With experience, GIS becomes simply an extension of one's analytical thinking… The system has no inherent answers, only those of the analyst. It is a tool for thought" Eastman J.R. (1992) IDRISI Users' Guide (p. 32)