

An introduction to geographical referencing for social scientists

23 November 2006

*Southampton Statistical Sciences
Research Institute*



Welcome and introductions

- David Martin
- Samantha Cockings
- Samuel Leung
- Workshop participants...
- Workshop URL:



www.geog.soton.ac.uk/geo-refer/workshop1.html

Housekeeping

- Access cards
- Coffee and lunch upstairs
- Temporary login IDs
- Other facilities
- Travel assistance

Workshop structure

- A (very) little bit about the project
- Morning presentation of concepts and methods relevant to this group
- Online access to materials we have created so far
- Interactive afternoon workshop providing practice and advice on specific issues

Rules of engagement

- This is an informal workshop!
- Please interrupt...
- There is no such thing as a silly question

The Geo-Refer project

- ESRC Researcher Development Initiative
- Geographical referencing: a key data linkage tool
- Frequently needed by non-geographers but no obvious source of methodological guidance (“phone a friend?”)
- Recent experience in creation of online learning materials

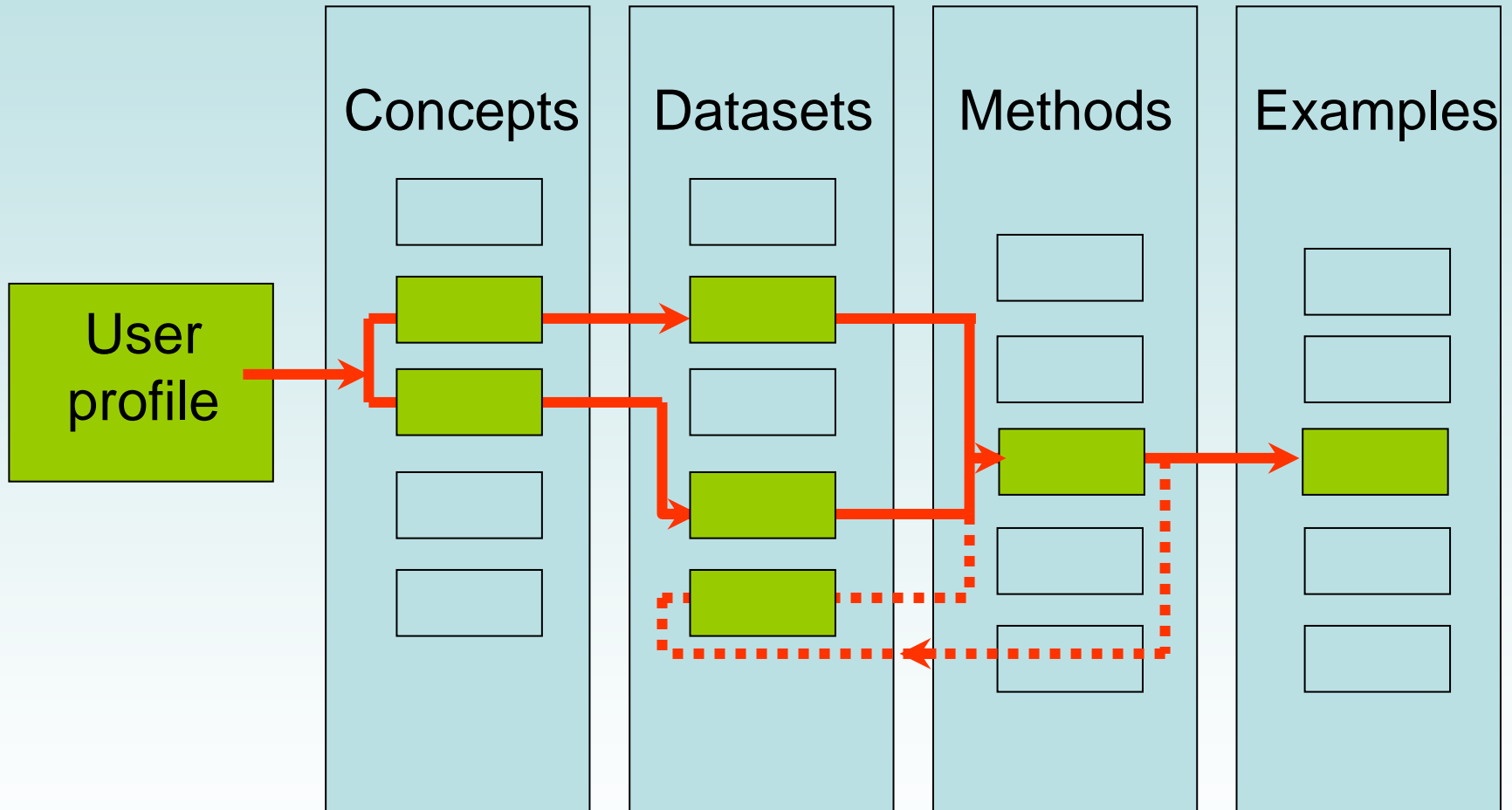
Georeferencing examples

- Link survey results to census data
- Associate postcoded patient records to indices of deprivation
- Allocate questionnaire responses to urban/rural neighbourhood types
- Identify which service delivery locations fall within different policy areas
- Use GPS to record survey locations

Researcher development...

- User needs
- Concepts (e.g. point and area objects)
- Datasets (e.g. postcode directory)
- Methods (e.g. tabular linkage)
- Examples (relevant to own needs!)

Online resource overview



User profiling

- Face-to-face experience, but...
- Importance of understanding how users are encountering georeferencing problems
- Desire to structure materials appropriately
- Need to harvest 'real world' user examples that will be of help to others
- Help!

Online user questionnaire


Geo-Refer User Profile Form - Microsoft Internet Explorer provided by the School of Geography, UoS

File Edit View Favorites Tools Help


Back Forward Stop Home Search Favorites Refresh Print Mail W Word Excel PowerPoint Outlook

Address <http://www.geog.soton.ac.uk/geo-refer/profile.php> Go Links Information Systems Services (ISS) ISS Subscribe

Google Search Web 102 blocked AutoFill Options

 **E·S·R·C**
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

GEO-REFER
User Profile form

 **RDI**
Researcher
Development
Initiative

Part 1 - Personal Information

Please answer all the following questions as fully as possible to help us identify the geographical referencing resources that are relevant to your project/study.

Discipline

Below is the list of subject areas used by ESRC. Please choose the nearest subject(s) to your own project/study. Your answer will help us to search for examples relevant to your interests.

- Area and Development Studies
- Demography
- Economic and social history
- Education
- Environmental Planning
- Human geography
- Linguistics
- Management and business studies
- Political science and international studies

Done Trusted sites

Workshop programme

- 11.15 Georeferencing concepts
- 11.45 Georeferencing methods
- 12.40 Really useful datasets
- 13.00 Lunch
- 14.00 Hands-on workshop
- 15.30 Presentation of results
- 16.00 Close

Georeferencing concepts

Geographical referencing of social science phenomena: 1

- Characteristics of people or events, recorded at mail addresses
- Geographical location of mail address



Photo: © David Martin

Geographical referencing of social science phenomena: 2

- Administration or policy related to defined areas, aggregations of individuals
- Geographical location of area boundaries



Geographical referencing of social science phenomena: 3

- Phenomena that are linear, describing routes or flows
- e.g. bus route, telephone conversation, social network link



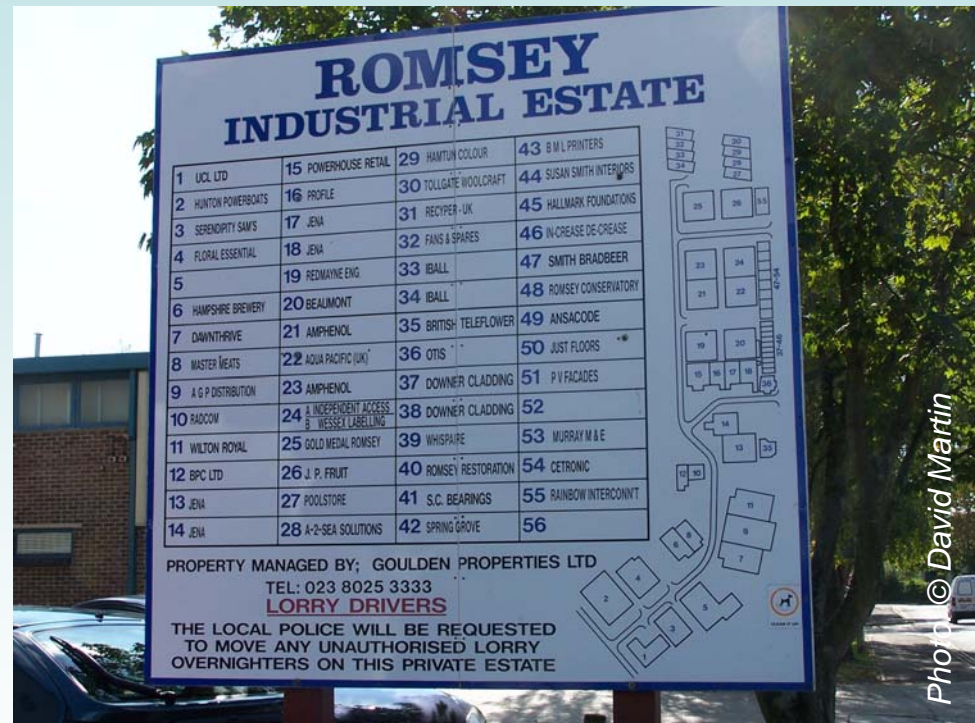
Geographical referencing of social science phenomena: 4

- Phenomena or events that have locations but do not fit standard descriptions
- e.g. road accident, environmental quality



Geographical referencing of social science phenomena: 5

- Complex phenomena not captured by a simple location
- e.g. business locations, catchment populations



Geographical object types

- Points: a single pair of coordinates
- Lines: an ordered sequence of coordinates
- Areas: closed, ordered sequences of coordinates
- Networks: complex line-based structures
- Surfaces: mathematically modelled representations of continuously varying phenomena

Focusing on points and areas...

- Here focusing on point and area geographical references
- Centroid points frequently used to locate areas where boundaries are unavailable or uncertain
- Maybe geometric, population-weighted, other...

Direct and indirect georeferencing

- Direct georeferencing: explicit coordinate system
 - Lat/Long, Ordnance Survey grid reference
- Indirect georeferencing: any type of area code or name relating to a known location
 - Postcode, Zip code, County, Census output area/enumeration district/tract; local government district; health authority, etc...

Why georeference?

- Massive growth in geographical data
- Adds analytical value
- Fundamentally, geographical referencing leads to either:
 - Data linkage, potentially for aspatial analysis
 - Mapping, and other forms of spatial analysis

Where do coordinates come from?

- Surveys for mapping
- Remote sensing
- Digitising paper source documents
- Direct capture using Global Positioning System (GPS) receiver
- Generally contributing to reference datasets created for re-use

Accuracy and precision

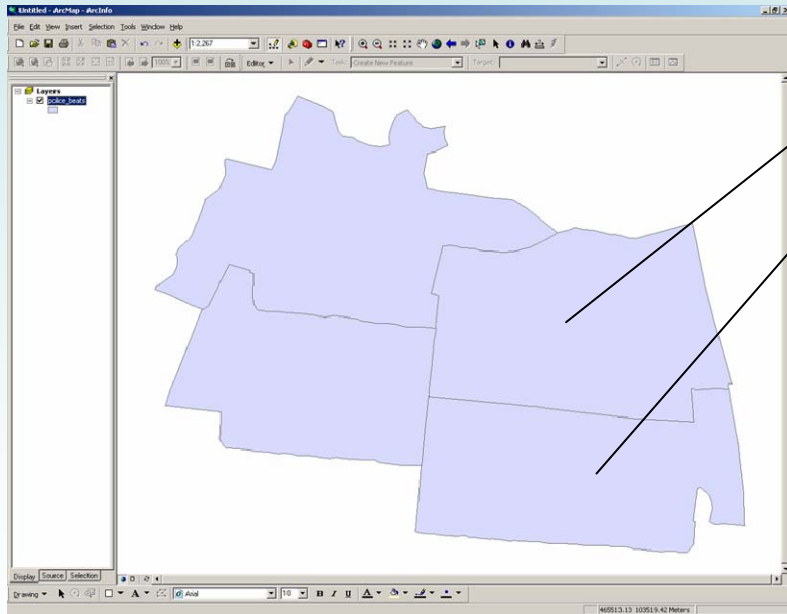
- Accurate location – free from locational bias (respondent lives in Hampshire)
- Precise location – provides detail (respondent lives at 55a Highfield Lane) but not necessarily accurate
- Need to assess and adopt different strategies according to purpose

Scale and projection

- Map scale
 - Explorer map: 1:25,000
 - Road atlas: typically 1:400,000
- Leads to inclusion/generalization of different features, boundary details, etc.
- Projection systems:
 - Very important when using lat/long datasets,
 - GB National Grid sidesteps the issues...

Locations and attributes

- Locations: points, area boundaries



OID	BEATID	NAME	HOFFS	AUTH
0	1	Westend	30	Anywhere
1	2	Southend	20	Anywhere
2	3	Eastend	5	Nowhere
3	4	Northend	10	Nowhere

Record: 1 Show: All Selected Records (0 of 4)

- Attributes: values, characteristics

Geoprivacy issues

- Detailed location is potentially disclosive
- Postcode is generally considered to be sensitive information
- Explicit subject consent in data collection?
- Subject identification may be possible through linkage and mapping
- All usual rules apply!

What about geographical information systems (GIS)?

- Massive growth in use of GIS since early 1980s: complex information systems
- GIS growth promoting data standards, growth of geoinformation industry etc.
- GIS provide useful georeferencing tools
- Lots of other functions not needed here...
- This is not a GIS course!

Georeferencing methods

Resources

- This session:
General overview
- Workshop website:
Sample resources
- This afternoon:
Try them out!

The screenshot shows a web browser window titled "GEO-REFER Workshop 1: An Introduction to Geographical Referencing for Social Scientists - Microsoft Internet Expl...". The address bar shows the URL "http://www.geog.soton.ac.uk/geo%2Drefer/sampleresources.html". The page features the GEO-REFER logo (a stylized 'GR' with a globe) and the title "GEO-REFER Workshop 1". Below the title, there are "Useful links" for ESRC, RDI, NCRM, and NCeSS. The main heading is "An Introduction to Geographical Referencing for Social Scientists". A search box is present on the left. A navigation menu includes links for Project homepage, Programme, Profiling form, Sample resources, Useful datasets, Hands-on session, and Evaluation form. A sidebar provides event details: Date (Thu 23, November 2006), Time (10:30 - 16:30), and Venue (Southampton Statistical Sciences Research Institute). The main content area lists "Sample Resources for Geo-Refer Workshop 1" with categories and links to concept overviews and methods examples:

- Matching lists together**
 - Concept overview
 - Methods example: Matching two lists using Microsoft Access
- Mapping point information**
 - Concept overview
 - Methods example: Mapping points using ArcGIS
- Joining attributes and boundaries**
 - Concept overview
 - Methods example: Joining attribute and boundary data for mapping purposes in ArcGIS
- Creating a choropleth map**
 - Concept overview
 - Methods example: Creating a choropleth map in ArcGIS
- Spatial data linkage: Joining point data to polygons**
 - Concept overview
 - Methods example: Joining point data to a set of polygons in ArcGIS
- Spatial data linkage: Joining polygons to points**
 - Concept overview
 - Methods example: Joining polygon data to a set of points in ArcGIS

Georeferencing methods

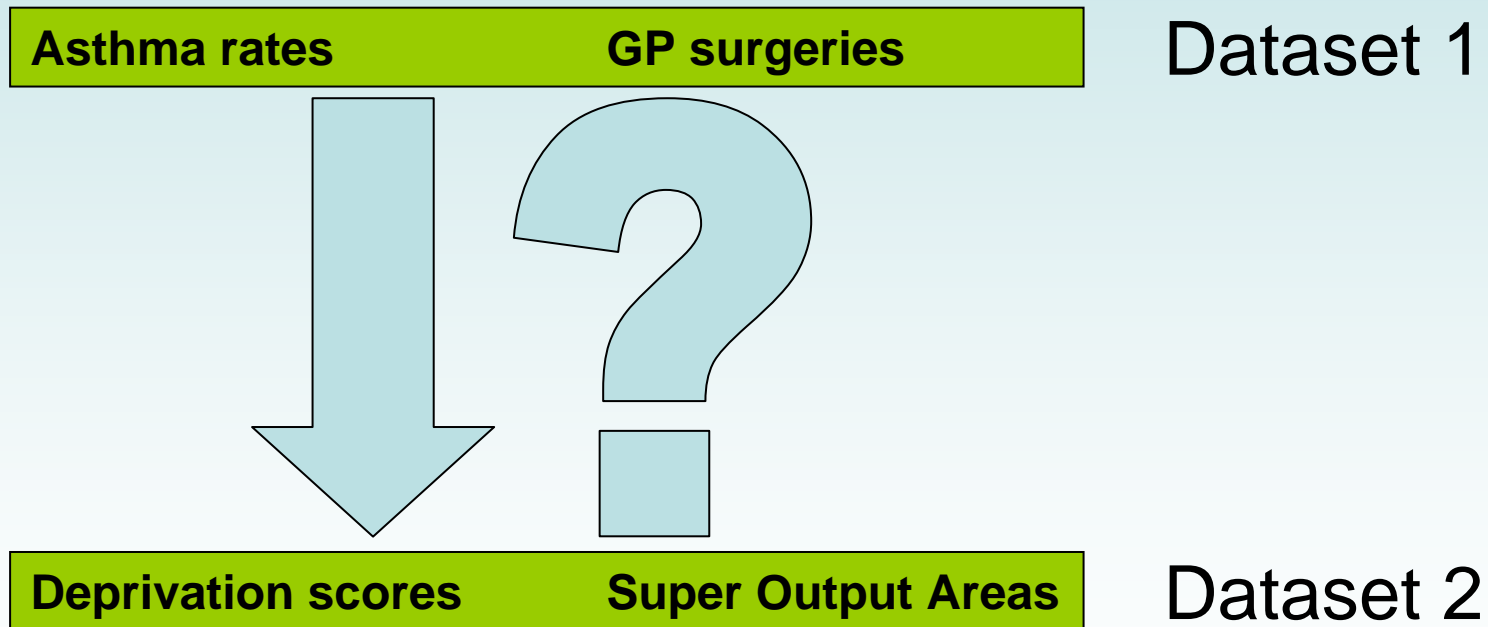
- Data linkage
 - Tabular
 - Spatial
 - Allocation and aggregation
- Mapping
 - Points
 - Areas
 - Choropleth/thematic maps

Methods 1: Data linkage

- Datasets with common items
- Generally, multiple research datasets and one or more lookup tables
- Example:
 - (i) Survey responses with postcodes
 - (ii) Super Output Areas with deprivation scores
 - Requires postcode to Super OA linkage

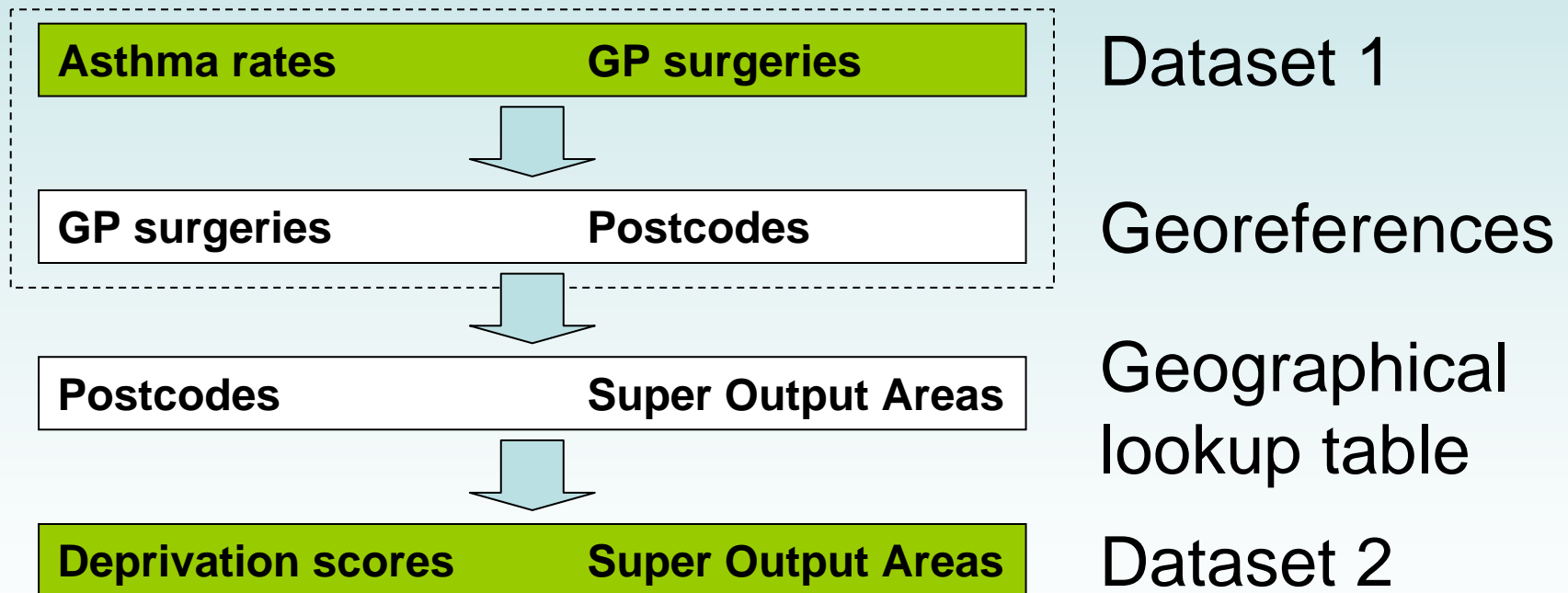
Tabular data linkage example

Research question: What is the relationship between asthma and deprivation?



Tabular data linkage example

Tabular linkage requires lookup between georeferences on two or more datasets?



Spatial data linkage example

Research question: What is the relationship between car crime and policing policies?




Police beat areas: not defined according to any recognised administrative geography



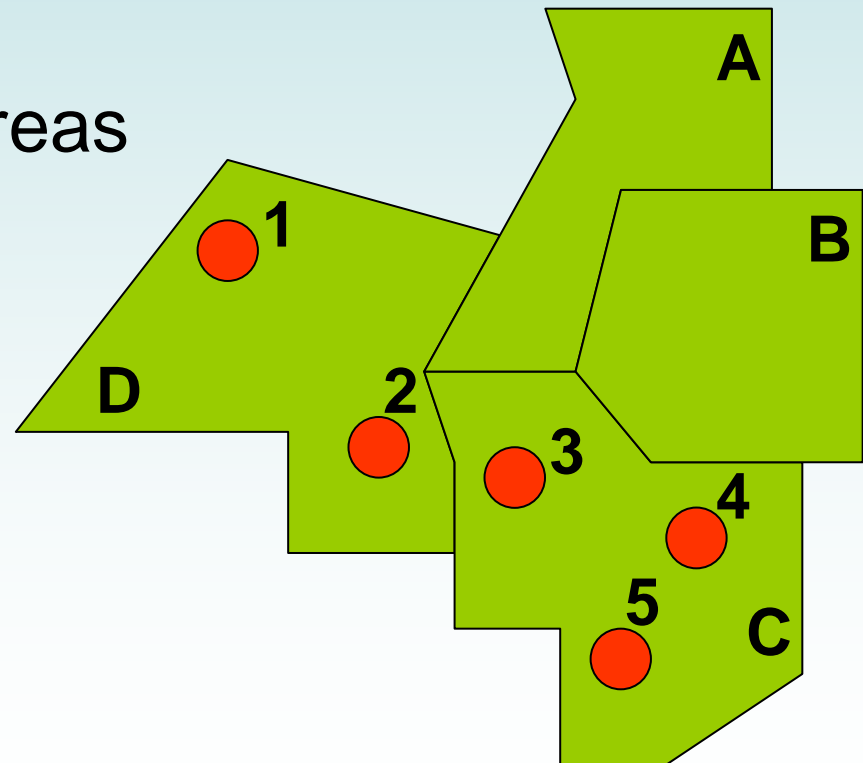
Car crimes: occur at many locations which cannot be addressed or postcoded

Spatial data linkage example

Spatial linkage requires intersection of area and point coordinates to associate crimes with beat areas

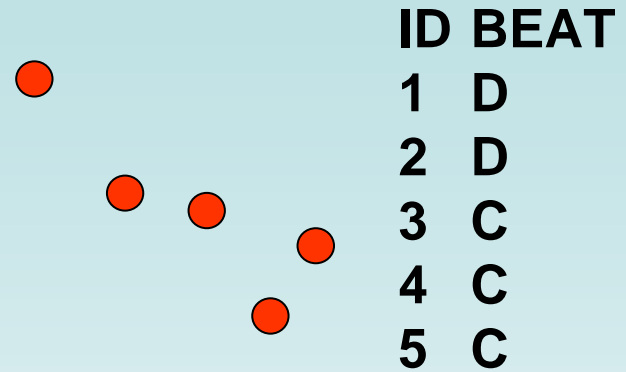
 Police beat areas

 Car crimes

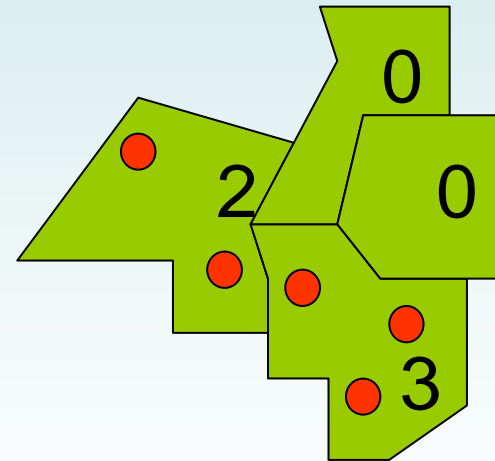


Allocation and aggregation

- Allocate
 - E.g. allocate attributes of police beats to car crime locations



- Aggregate
 - E.g. aggregate number of car crimes in each police beat



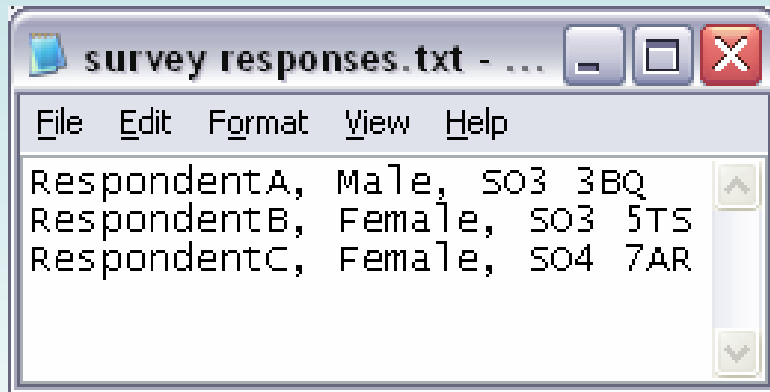
- Same spatial linkage operation, but different outputs

Tabular linkage tools

- Access, SPSS etc. – general purpose database or statistical packages for matching lists with common data items
- GeoConvert – online tool specific to recent UK postcode-based lookup

Data linkage: input tables

Coded survey responses



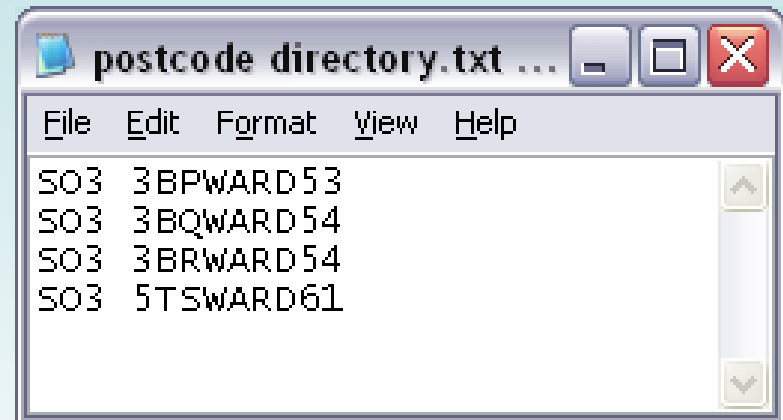
```
File Edit Format View Help
RespondentA, Male, SO3 3BQ
RespondentB, Female, SO3 5TS
RespondentC, Female, SO4 7AR
```

Respondent ID

Gender

Postcode

Postcode directory

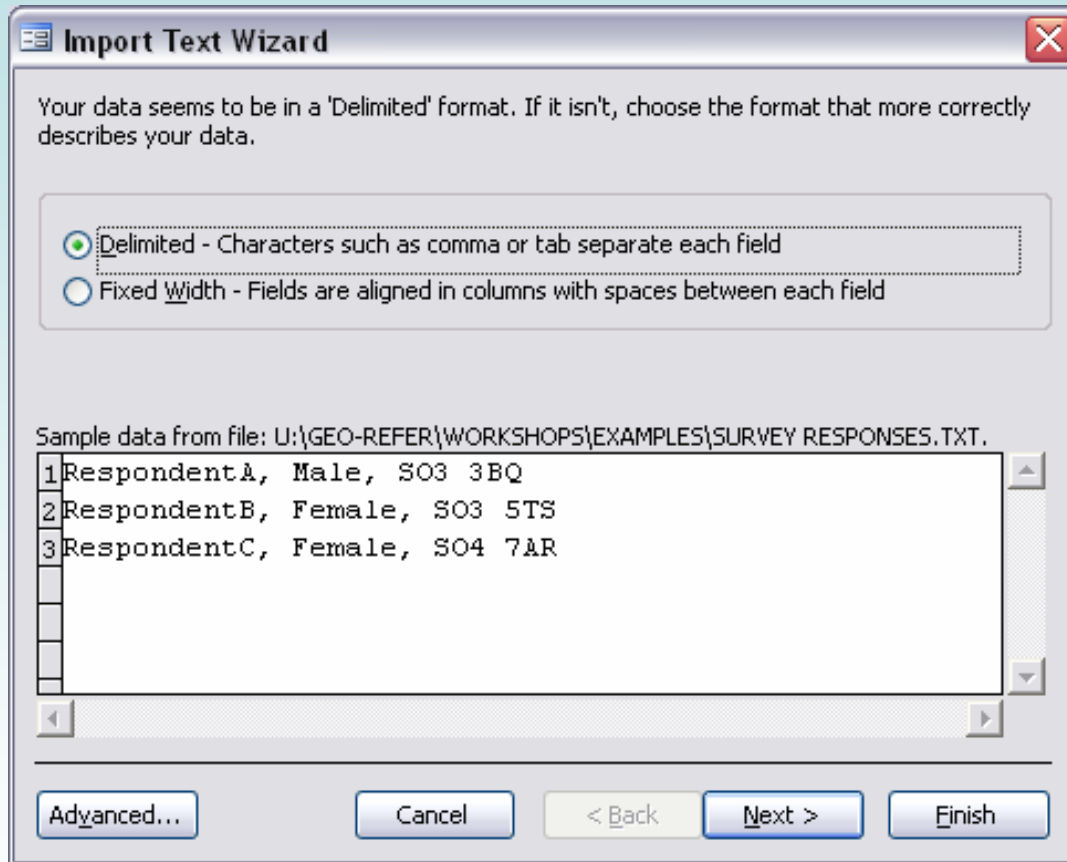


```
File Edit Format View Help
SO3 3BPWARD53
SO3 3BQWARD54
SO3 3BRWARD54
SO3 5TSWARD61
```

Postcode

Ward code

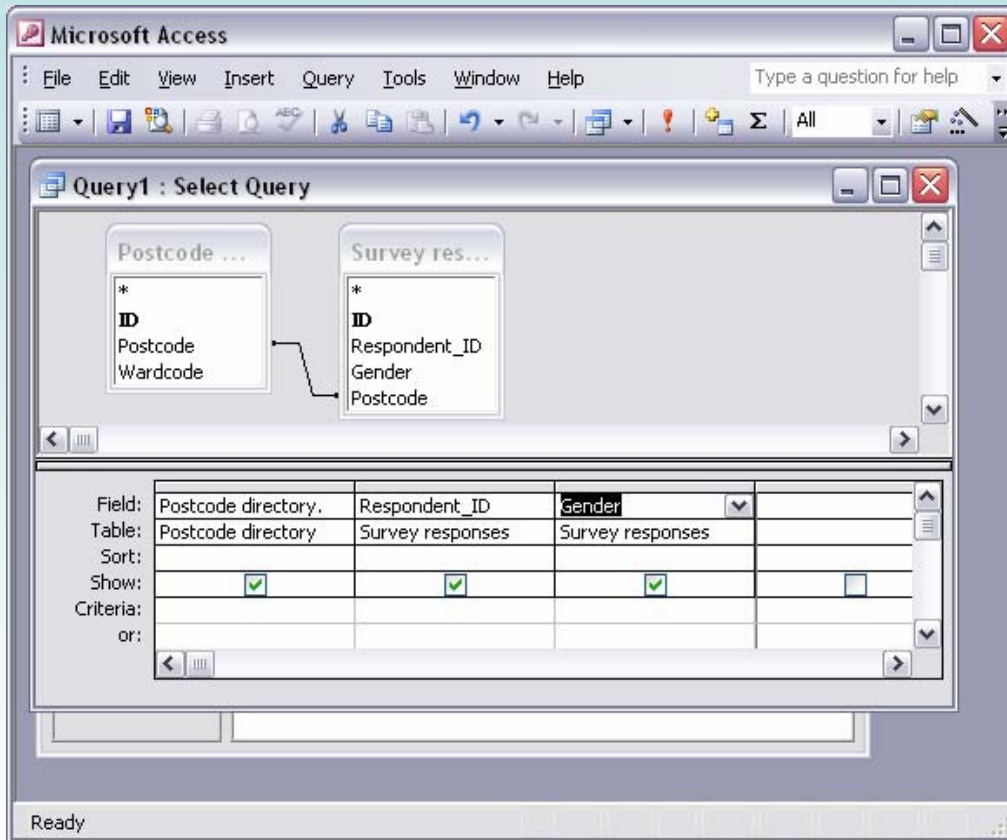
Import external data files



Get external data > import

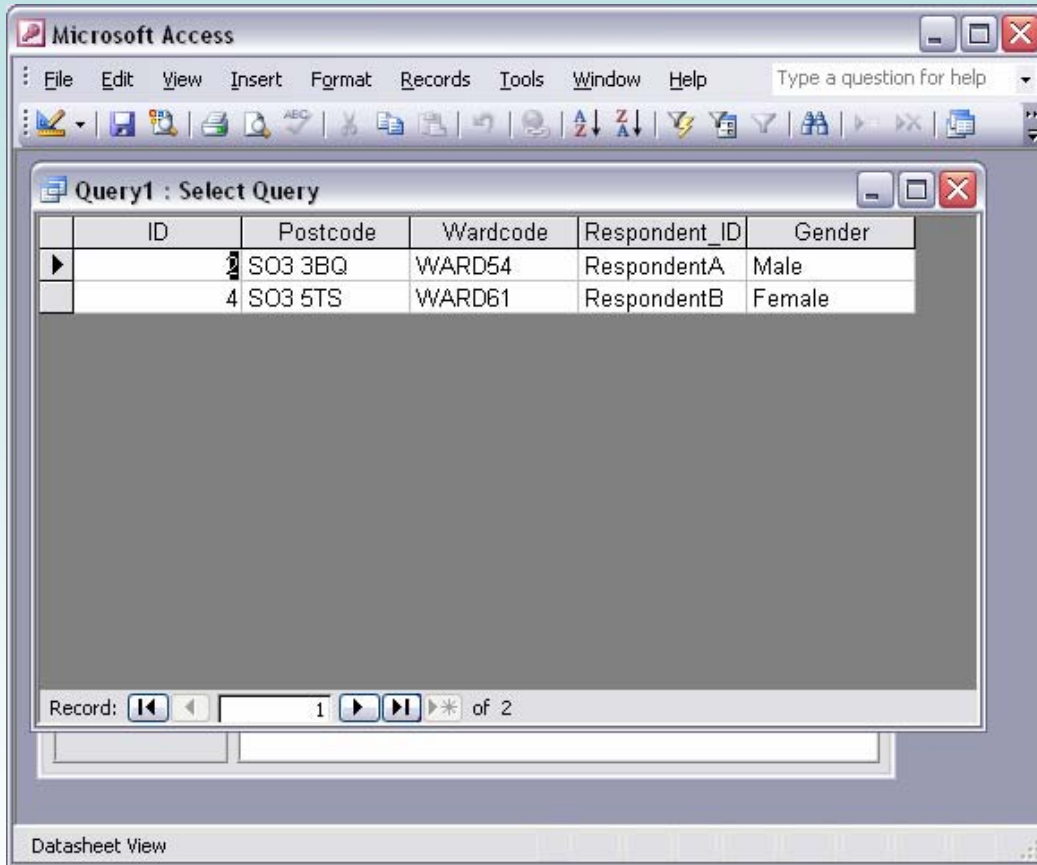
We need to describe the data structure (here comma separated)

Build new Access query



We now add both tables to a new query, link them by the postcode field and select the items required in output

Matched lists



The screenshot shows a Microsoft Access window titled "Microsoft Access" with a menu bar (File, Edit, View, Insert, Format, Records, Tools, Window, Help) and a search bar. Below the menu is a toolbar with various icons. The main window displays a query named "Query1 : Select Query" in Datasheet View. The data is presented in a table with the following columns and rows:

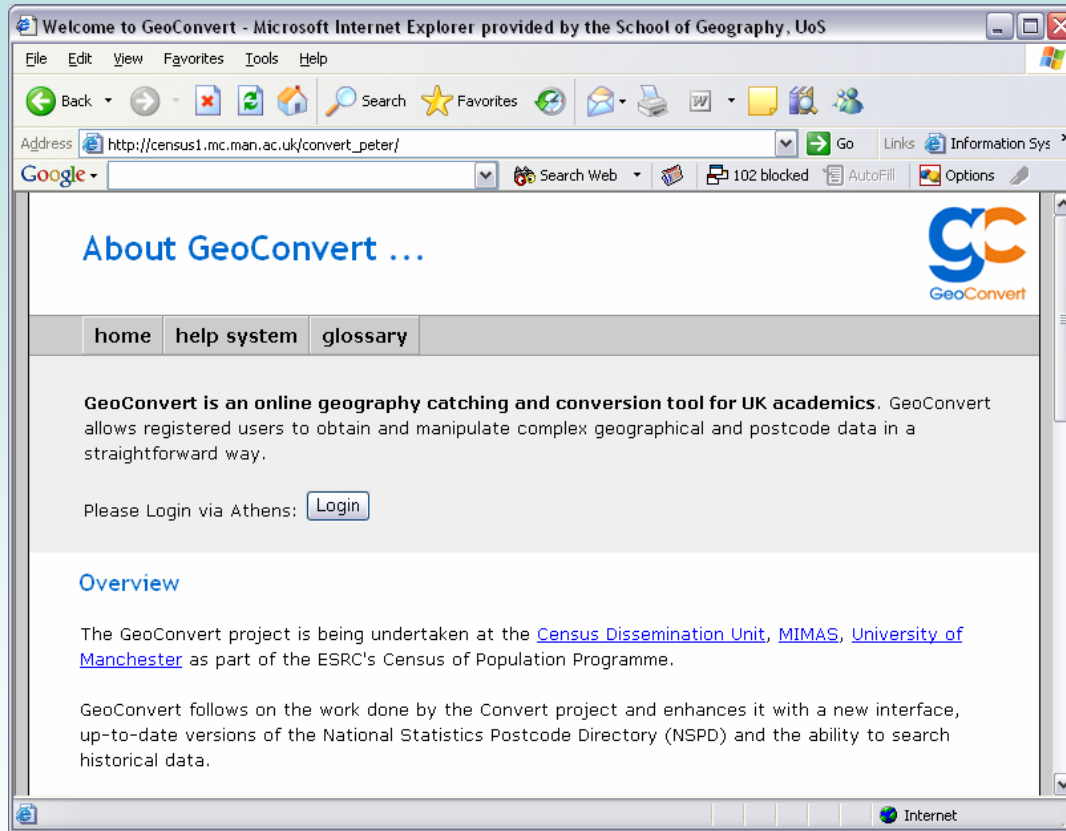
ID	Postcode	Wardcode	Respondent_ID	Gender
2	S03 3BQ	WARD54	RespondentA	Male
4	S03 5TS	WARD61	RespondentB	Female

At the bottom of the table, there is a record navigation bar showing "Record: 1 of 2".

The resulting query contains requested fields for all matching rows and can be exported

NB not all are matched!

Data linkage with GeoConvert



Developed by
MIMAS

Online tools for
NSPD
extraction and
matching

GeoConvert functions

- Obtain postcode-related information
- Identify spatial relationships between geographies
- Convert data between geographies
- Obtain a subset of rows from a postcode directory, together with supplementary information

Using GeoConvert: general structure

- Specify version of directory to use
- Specify to and from geographies
- Specify other fields of interest
- Enter code or upload file
- Download matched file

Downloaded file

	A	B	C	D	E	F	G	H
1	Line numb	Output Area	proportion	Postcode	OSEAST1	OSNRTH1	ADDRCT01	var1
2	1	24UNGF0006		SO5 0BH	430800	118900		1
3	1	24UNGF0006		SO5 0BJ	430600	118800		1
4	1	24UNGF0006		SO5 0BL	430700	118900		1
5	1	24UNGF0006		SO5 0BQ	430600	118900		1
6	1	24UNGF0006		SO5 0EU	430600	118700		1
7	1	24UNGF0006		SO5 0EW	430600	118700		1
8	1	24UNGF0006		SO5 0EX	430600	118700		1
9	1	24UNGF0006		SO5 0EY	430600	118700		1
10	1	24UNGF0006	0.0278	SO516AZ	430532	118670	3	1
11	1	24UNGF0006	0.0741	SO516EW	430675	118818	8	1
12	1	24UNGF0006	0.0741	SO516EX	430630	118814	8	1
13	1	24UNGF0006	0.0741	SO516BL	430728	118936	8	1
14	1	24UNGF0006	0.0926	SO516BQ	430590	118909	10	1
15	1	24UNGF0006	0.1019	SO516EU	430654	118723	11	1

Sample output:
specified fields
for specified
postcodes

Watch the time!

- Extreme care needs to be taken when matching between datasets relating to different dates
- Postcodes, boundaries, area names all subject to change
- All changes on different timescales
- Another time = another geography!

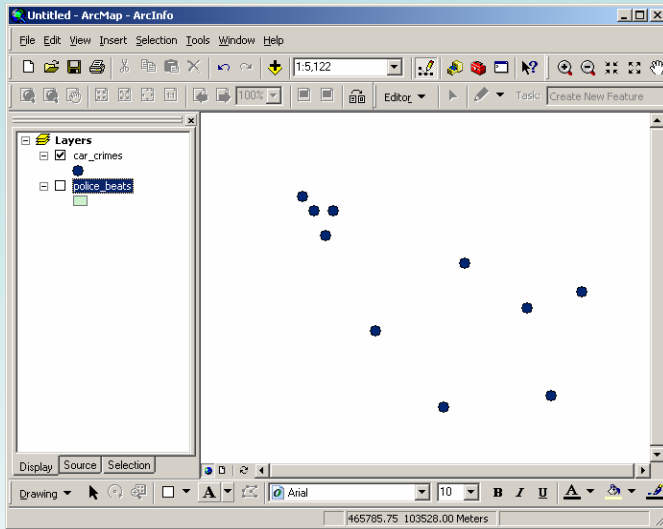
Spatial linkage tools

- Spatial linkage usually done using a GIS
 - Various GIS software packages
 - ESRI (ArcGIS, ArcInfo, ArcView)
 - MapInfo
 - IDRISI
- Remember: This is not a GIS course!
- Current worked examples use ArcGIS

Allocation: joining poly data to points data

Car crimes: points

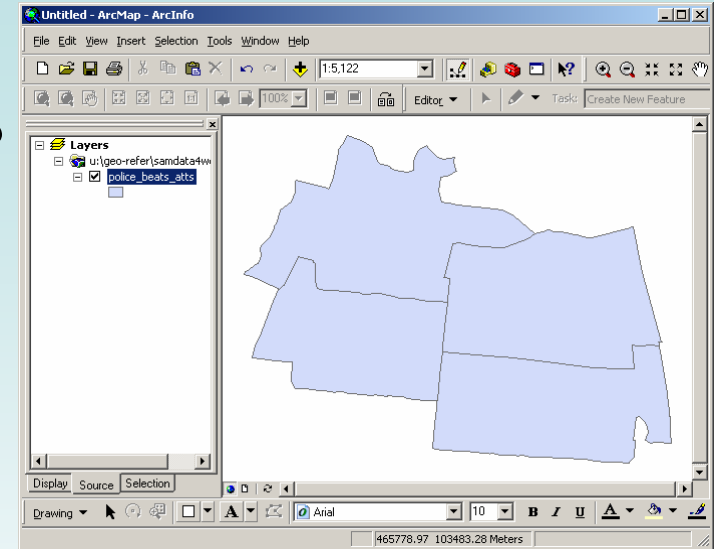
Police beats: areas



Spatial link?



Point-in-polygon



OID	CRIMEID	X	Y
0	1	465474	103429
1	2	465490	103409
2	3	465517	103409
3	4	465507	103374
4	5	465703	103334
5	6	465791	103271
6	7	465869	103294
7	8	465577	103238
8	9	465825	103146
9	10	465673	103129

FID	Shape*	BEATID	NAME	HOFFS	AUTH
0	Polygon	1	Westend	30	Anywhere
1	Polygon	2	Southend	20	Anywhere
2	Polygon	3	Eastend	5	Nowhere
3	Polygon	4	Northend	10	Nowhere



Allocate polygon attributes to points

Allocation: joining poly data to points data

Spatial join: polygons to points

Sample output

Join Data

Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.

What do you want to join to this layer?

Join data from another layer based on spatial location

1. Choose the layer to join to this layer, or load spatial data from disk:

police_beats_atts

2. You are joining: Polygons to Points

Select a join feature class above. You will be given different options based on geometry types of the source feature class and the join feature class.

Each point will be given all the attributes of the polygon that:

it falls inside.

If a point falls inside more than one polygon (for example, because the layer being joined contains overlapping polygons) the attributes of the first polygon found will be joined.

is closest to it.

A distance field is added showing how close the polygon is (in the units of the target layer). A polygon that the point falls inside is treated as being closest to the point (i.e. a distance of 0).

3. The result of the join will be saved into a new layer.

Specify output shapefile or feature class for this new layer:

U:\Geo-Refer\SamData4\Workshop\crimes_beats.shp

About Joining Data OK Cancel

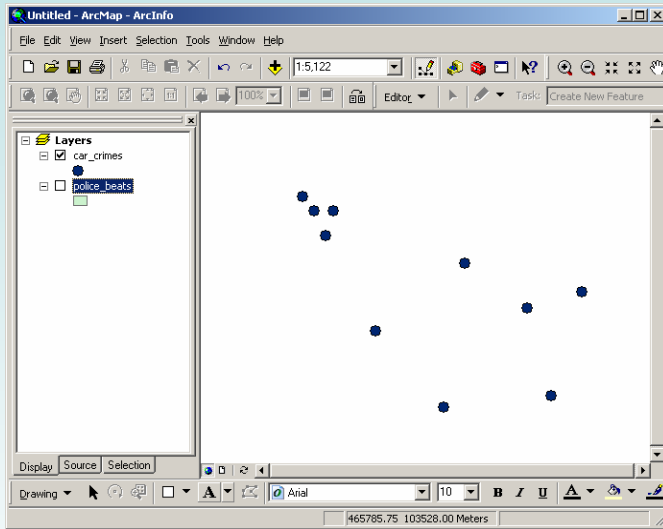
Attributes of crimes_beats

	FID	Shape	FI	CRIMEID	X	Y	FI	BEATID	NAME	NOFFS	AUTH
▶	0	Point	7	8	465577	103238	0	1	Westend	30	Anywhere
	1	Point	8	9	465825	103146	1	2	Southend	20	Anywhere
	2	Point	9	10	465673	103129	1	2	Southend	20	Anywhere
	3	Point	4	5	465703	103334	2	3	Eastend	5	Nowhere
	4	Point	5	6	465791	103271	2	3	Eastend	5	Nowhere
	5	Point	6	7	465869	103294	2	3	Eastend	5	Nowhere
	6	Point	0	1	465474	103429	3	4	Northend	10	Nowhere
	7	Point	1	2	465490	103409	3	4	Northend	10	Nowhere
	8	Point	2	3	465517	103409	3	4	Northend	10	Nowhere
	9	Point	3	4	465507	103374	3	4	Northend	10	Nowhere

Record: 1 Show: All Selected Records: (0 out of 10 Selected.) Option

Aggregation: joining points data to poly data

Car crimes: points

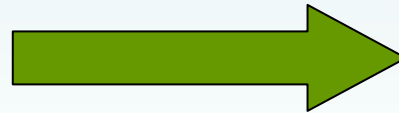
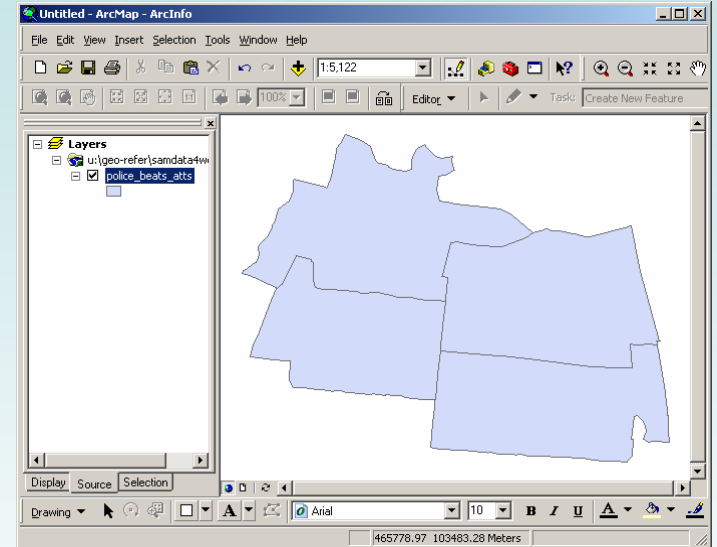


Spatial link?



Point-in-polygon

Police beats: areas



**Aggregate points
and/or their attribute
data to polygons**

Aggregation: joining points data to poly data

Spatial join: points to polygons

Sample output

Join Data

Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.

What do you want to join to this layer?

Join data from another layer based on spatial location

1. Choose the layer to join to this layer, or load spatial data from disk:

car_crimes

2. You are joining: Points to Polygons

Select a join feature class above. You will be given different options based on geometry types of the source feature class and the join feature class.

Each polygon will be given a summary of the numeric attributes of the points that fall inside it, and a count field showing how many points fall inside it.

How do you want the attributes to be summarized?

Average Minimum Standard Deviation
 Sum Maximum Variance

Each polygon will be given all the attributes of the point that is closest to its boundary, and a distance field showing how close the point is (in the units of the target layer).

Note: A point falling inside a polygon is treated as being closest to the polygon, (i.e. a distance of 0).

3. The result of the join will be saved into a new layer.

Specify output shapefile or feature class for this new layer:

U:\Geo-Refer\SamData4\Workshop\beats_crimes.shp

About Joining Data OK Cancel

Attributes of beats_crimes

FID	Shape	FID_1	BEATID	Count_
0	Polygon	0	1	1
1	Polygon	1	2	2
2	Polygon	2	3	3
3	Polygon	3	4	4

Record: 1 Show: All Selected Records (0)

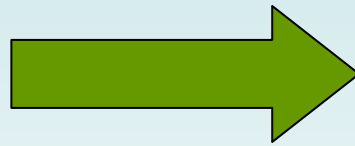
Methods 2: Mapping

- Map locations of points and/or areas
 - E.g. grid-refs of survey responses
 - E.g. boundaries of Local Authority Districts (LADs)
- Map attributes of areas
 - E.g. Link survey responses on attitudes to recycling to LADs then map response rates by LADs

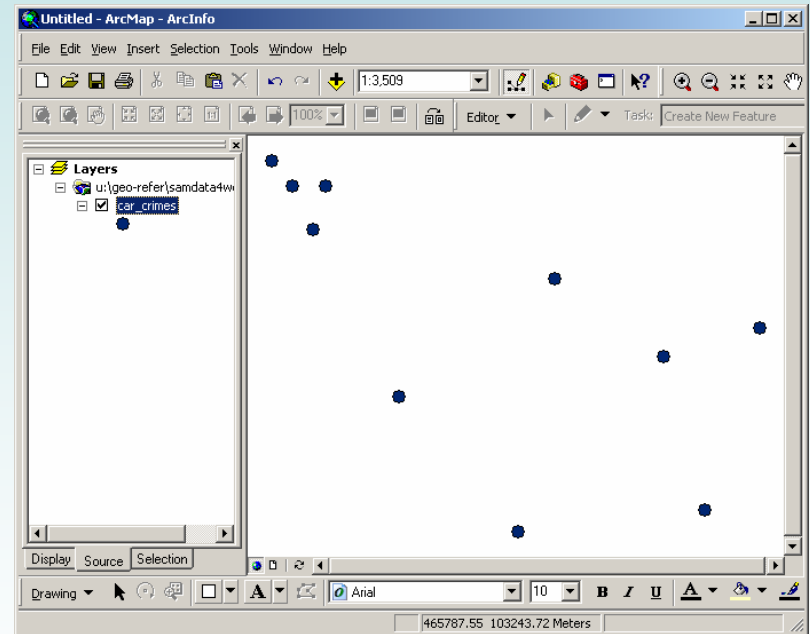
Mapping a set of grid-referenced points

Grid-refs of car crimes

465474,103429
465490,103409
465517,103409
465507,103374
465703,103334
465791,103271
465869,103294
465577,103238
465825,103146
465673,103129



How do I map these?



Mapping a set of grid-referenced points

Get data into format compatible with import to ArcGIS

Grid-refs (.csv, .tab, .xls)

Import into Access and define fields
e.g. ID, X, Y

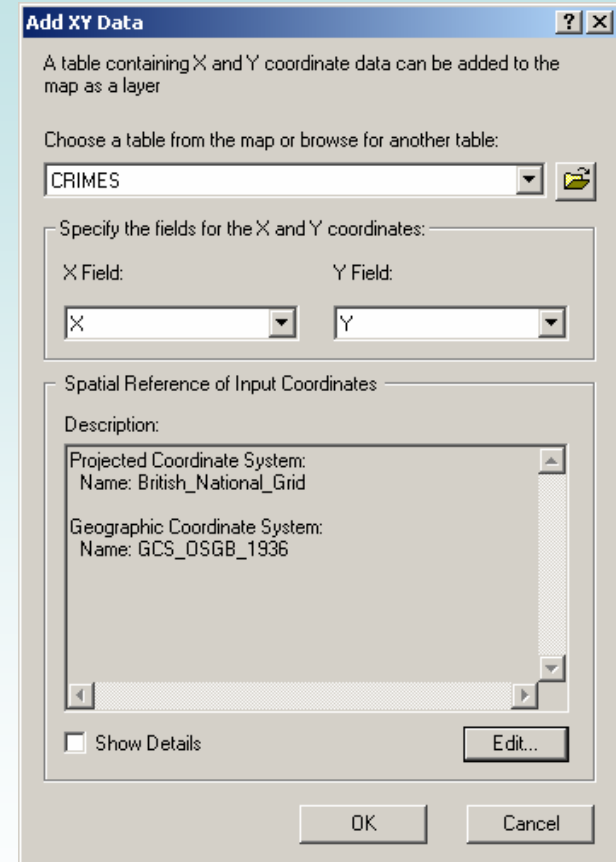
Save in dBASE (IV) format (.dbf)

Import into ArcGIS

Add into ArcMap as Table

Create event theme (Add XY data)
using X and Y fields

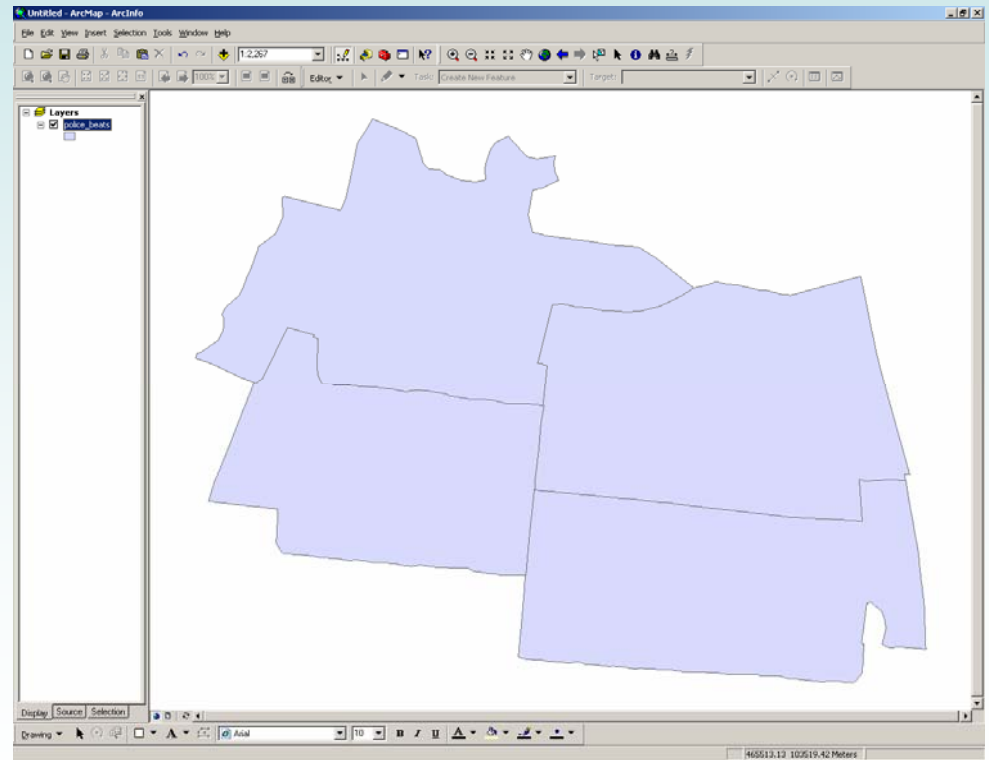
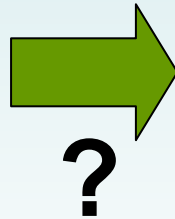
Export as shapefile



Linking attribute data to boundary data

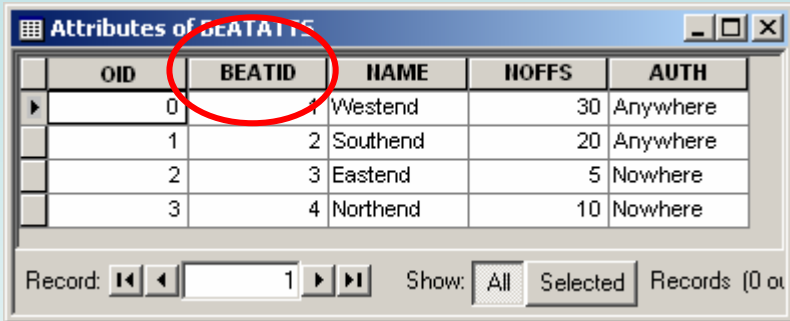
I've got boundary data and attribute data
but how do I link them together?

```
beatatts.txt - Notepad
File Edit Format View Help
BEATID, NAME, NOFFS, AUTH
1, westend, 30, Anywhere
2, southend, 20, Anywhere
3, Eastend, 5, Nowhere
4, Northend, 10, Nowhere
```

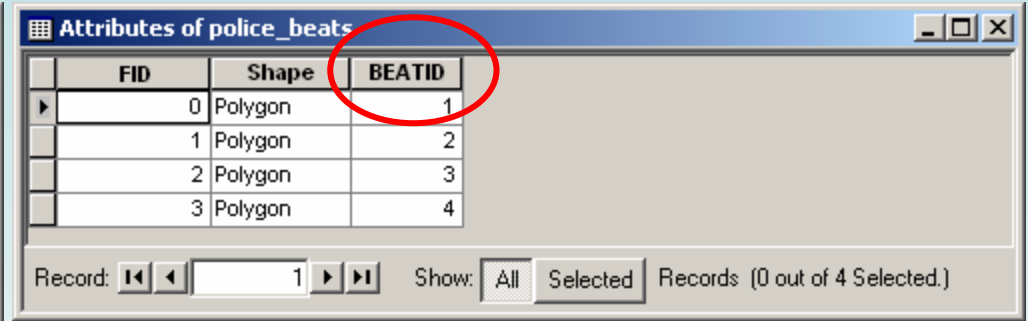


Linking attribute data to boundary data

Define common field in attribute table and in shapefile attribute table

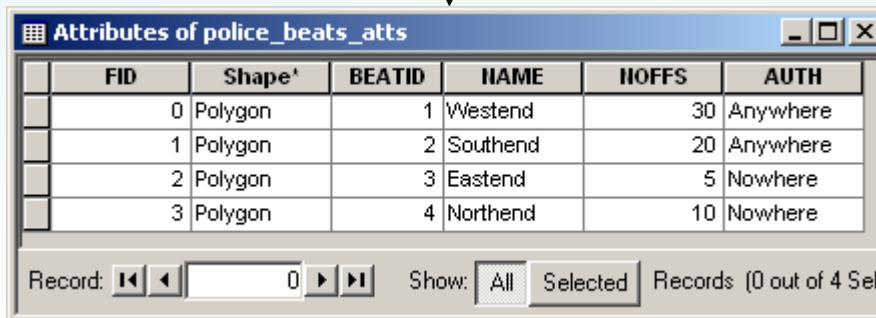


OID	BEATID	NAME	HOFFS	AUTH
0	1	Westend	30	Anywhere
1	2	Southend	20	Anywhere
2	3	Eastend	5	Nowhere
3	4	Northend	10	Nowhere



FID	Shape	BEATID
0	Polygon	1
1	Polygon	2
2	Polygon	3
3	Polygon	4

JOIN attributes from a table (based on BEATID)



FID	Shape*	BEATID	NAME	HOFFS	AUTH
0	Polygon	1	Westend	30	Anywhere
1	Polygon	2	Southend	20	Anywhere
2	Polygon	3	Eastend	5	Nowhere
3	Polygon	4	Northend	10	Nowhere

Attributes linked to boundaries

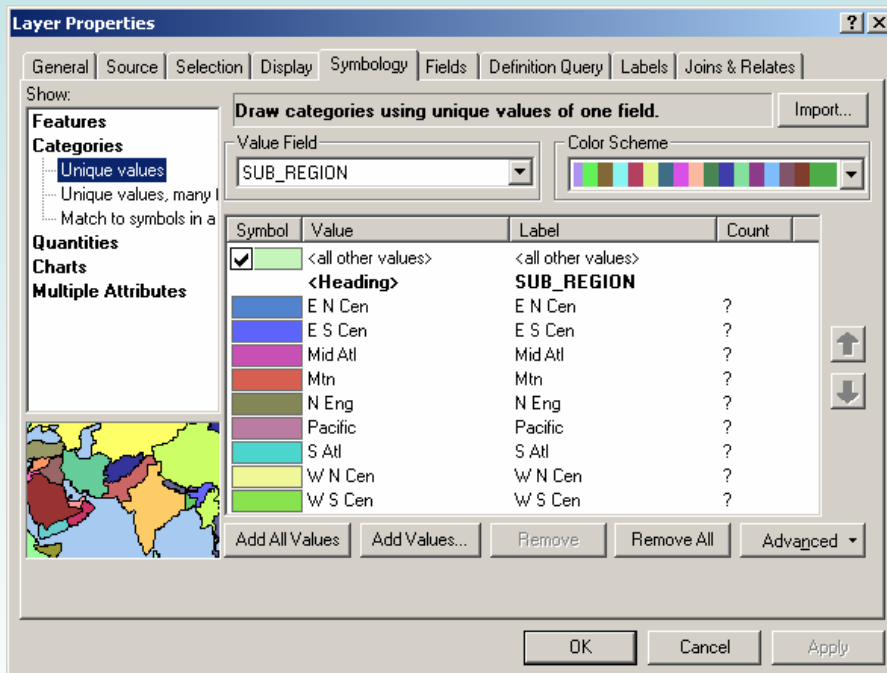
Creating a choropleth/thematic map

- I've got my attributes linked to boundaries, but how do I map them?
- Choropleth/thematic/area-shaded maps
- Categorical or continuous data?
- Be careful about choice of classification scheme, class breaks and colour schemes

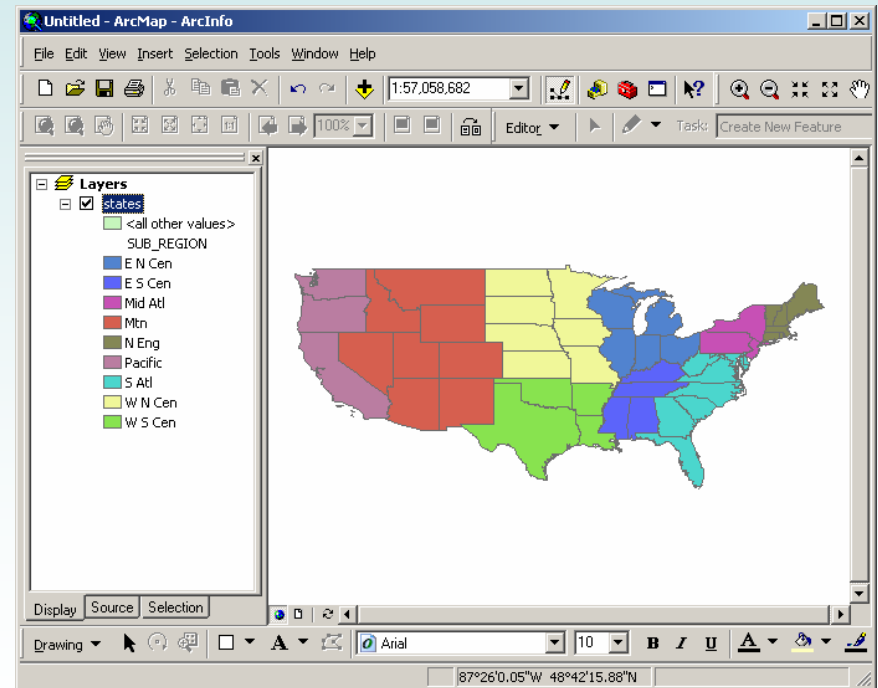
Creating a choropleth/thematic map

- Add layer (shapefile)
- Select variable to map
- Select data type
- (Select classification scheme/class breaks)
- Select colour scheme

Creating a choropleth/thematic map: Categorical data

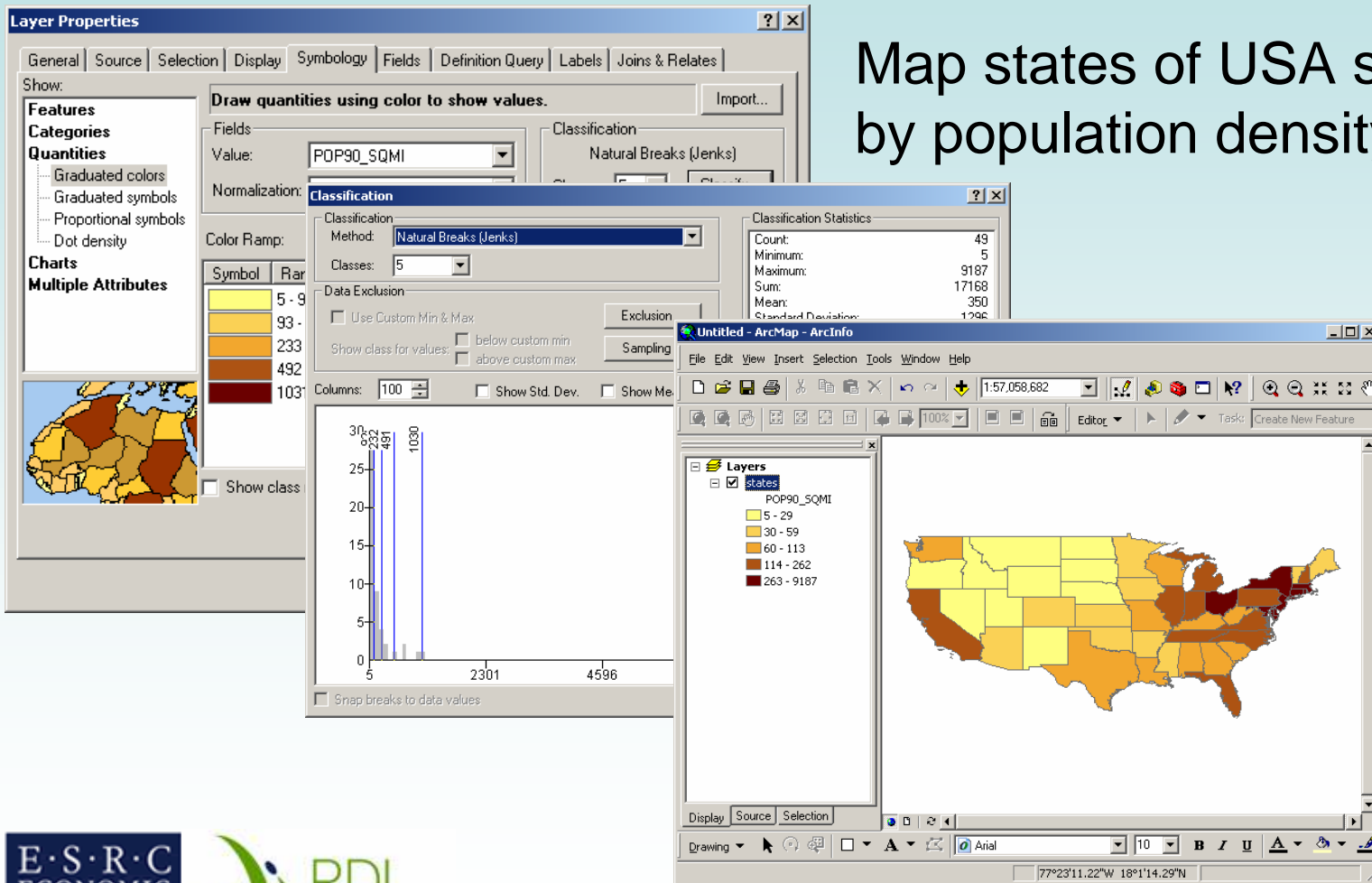


Map states of USA
shaded by sub-region



Creating a choropleth/thematic map: Continuous data

Map states of USA shaded
by population density



Really useful datasets



Some really useful datasets...

- National Statistics Postcode Directory
- UKBORDERS Digital Boundary Data
- Neighbourhood Statistics Services
- US Bureau of the Census
- ONS Beginners Guide to UK Geography

National Statistics Postcode Directory

Postcode

Ordnance Survey Grid
Reference

Dates of introduction and
termination (if applicable)

Delivery point count

2001 Census geography
codes

1991 Census geography
codes

Numerous other codes, esp.
changing health geographies

Numerous derived lookup
tables

AKA All Fields Postcode
Directory, Central Postcode
Directory, Postzon file...

Used in GeoConvert

Downloadable from
UKBORDERS; registration
required

UKBORDERS Digital Boundary Data

Coordinates of area boundaries downloadable in various GIS and mapping formats

Primarily census-derived, covering whole UK

Census, administrative, health and electoral geographies

Includes some historical (pre-1971 boundary sets e.g. 1951 local government

Need to understand which zones are needed

Need suitable software to load the boundaries

Large data volumes and further manipulation often required

No attribute data

Downloadable from UKBORDERS; registration required

Neighbourhood Statistics Services

Wide range of easily
downloadable social data

Includes basic 2001 census
datasets and administrative
data derived from government
departments

Searchable by many different
georeferences

Some online mapping and
analysis

Simple – general audience

Only covers most basic
datasets

No lookup tables or boundary
data downloadable, just
statistics for areas

Data formatting can be
frustrating for serious user

Freely accessible from ONS,
GROS, NISRA websites;
registration increases
functionality

US Bureau of the Census

Freely downloadable census and boundary data files

Large numbers of different administrative systems available as boundary files

Mostly covering 1990 and 2000 censuses

NB free online access to all US census results

American Factfinder allows retrieval of census tract from a street address

Landview product allows simple mapping and includes geocoding of addresses

Freely accessible from US Bureau of the Census website

ONS Beginners Guide to UK Geography

Online explanatory material about contemporary UK boundary systems

Includes census, postal, administrative, electoral, health and other geographies

Useful links to other resources

Very clear – written for general audience

Good reference for students!

Does not cover historical datasets: essentially describes the post-2001 census situation

No downloadable data, just explanatory guidance

Freely accessible from ONS website

www.statistics.gov.uk/geography/beginners_guide.asp

Other useful sites...

Royal Mail Postcode Finder
and Address Finder

Digimap – Ordnance Survey
mapping, inc. historical

Google Earth and Microsoft
Local Live

Registration required and
limited number of uses

JISC-funded subscription
service at Edina

Street mapping and aerial
photography (and many
others)

Lunchtime!





Linking International Macro and Micro Data Workshop

11 December; 25 April, University of Manchester, Free!

For further information:

peter.whitton@manchester.ac.uk



Practical workshop session



Datasets provided...

- Quality and outcomes framework 2004/5 data for general practitioners
 - Achievement data at general practice level
 - Practice code-name lookup table

www.ic.nhs.uk/services/qof/data/
- Link practices to postcodes, postcodes to locations, locations to map...

Presentation of examples

Workshop evaluation form

- Immediate feedback on workshop
- We would like to find out more about your projects so that we can use them as examples to help others...
- Thank you!