ITRC mistral

multi-scale infrastructure systems analytics

Spatial Training:

Introduction to spatial data

Newcastle University

Craig Robson

November 2016

EPSRC
Engineering and Physical Sciences Research Council

ITRC mistral
multi-scale infrastructure systems analytics

Outline

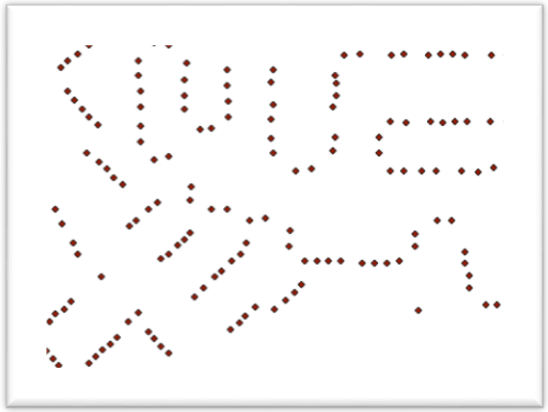
- Part 1:
  - Basics of spatial data
  - Coordinate systems
  - Data management
  - Common analysis methods
- Part 2:
  - MAUP
  - Working with data in different geographies
  - Networks
- ~10 minute discussion after each part
- Please ask questions as I go along!

Spatial data types

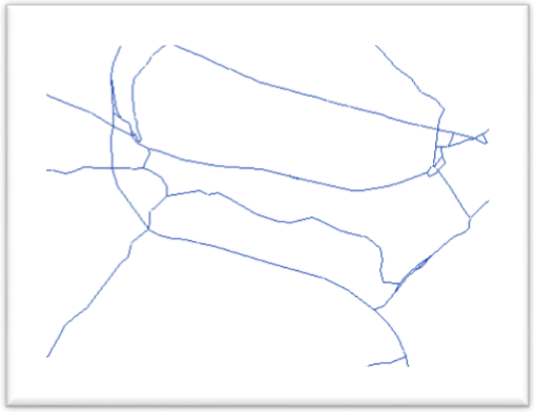
- 2 types of spatial data
  - Vector
    - Discrete data
    - Eg. Points, lines, polygons
  - Raster
    - Continuous data
    - Eg. Images, maps
- Both handled in GIS systems (Arc, QGIS)
- Libraries for different programming languages
- Types of analysis possible varies per data type

Spatial data types - vectors

Point



Line



Polygon

- Attributes
- Geometry

Table

Newcastle_Buildings

FID	Shape *	OID_	Toid	Featcode	Version	VerDate	Theme	CalcArea	Change	DescGroup
0	Polygon	0	1000030209612	10021	2	2001-11-05	Buildings	17.212736	1982-09-24 New	Building
1	Polygon	0	1000030209703	10021	2	2001-11-05	Buildings	70.0744	1982-09-24 New	Building
2	Polygon	0	1000030209696	10021	2	2001-11-05	Buildings	67.419344	1982-09-24 New	Building
3	Polygon	0	1000030209607	10021	2	2001-11-05	Buildings	77.077488	1982-09-24 New	Building
4	Polygon	0	1000030209584	10021	2	2001-11-05	Buildings	74.72728	1993-07-15 Modified	Building
5	Polygon	0	1000030209587	10021	2	2001-11-05	Buildings	11.4616	1982-09-24 New	Building
6	Polygon	0	1000030209572	10021	2	2001-11-05	Buildings	182.052464	1982-09-24 New	Building
7	Polygon	0	1000030209700	10021	2	2001-11-05	Buildings	68.233616	1982-09-24 New	Building
8	Polygon	0	1000030209614	10021	2	2001-11-05	Buildings	54.558992	1982-09-24 New	Building
9	Polygon	0	1000030209605	10021	2	2001-11-05	Buildings	9.432736	1982-09-24 New	Building
10	Polygon	0	1000030209695	10021	2	2001-11-05	Buildings	77.996256	1993-03-01 Modified	Building

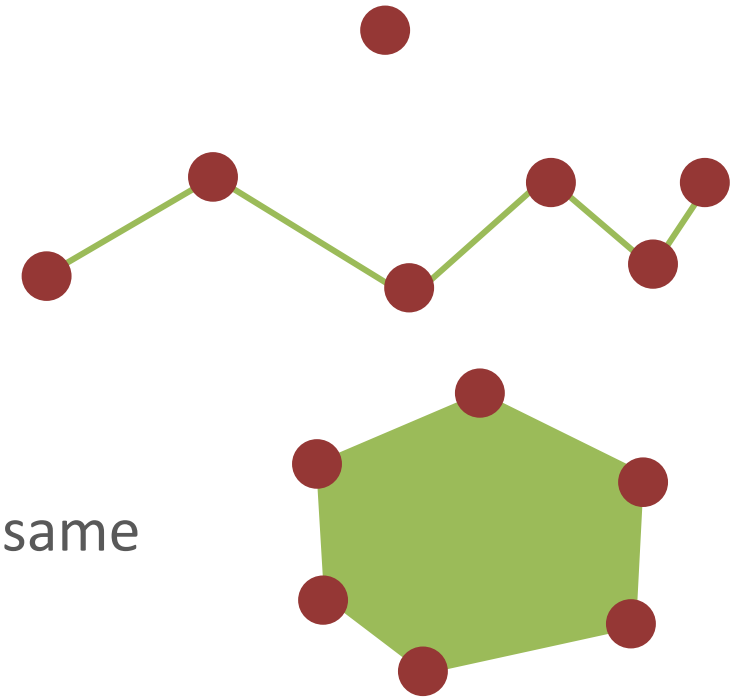
1 (0 out of 59127 Selected)

Newcastle_Buildings

EPSRC
Engineering and Physical Sciences Research Council

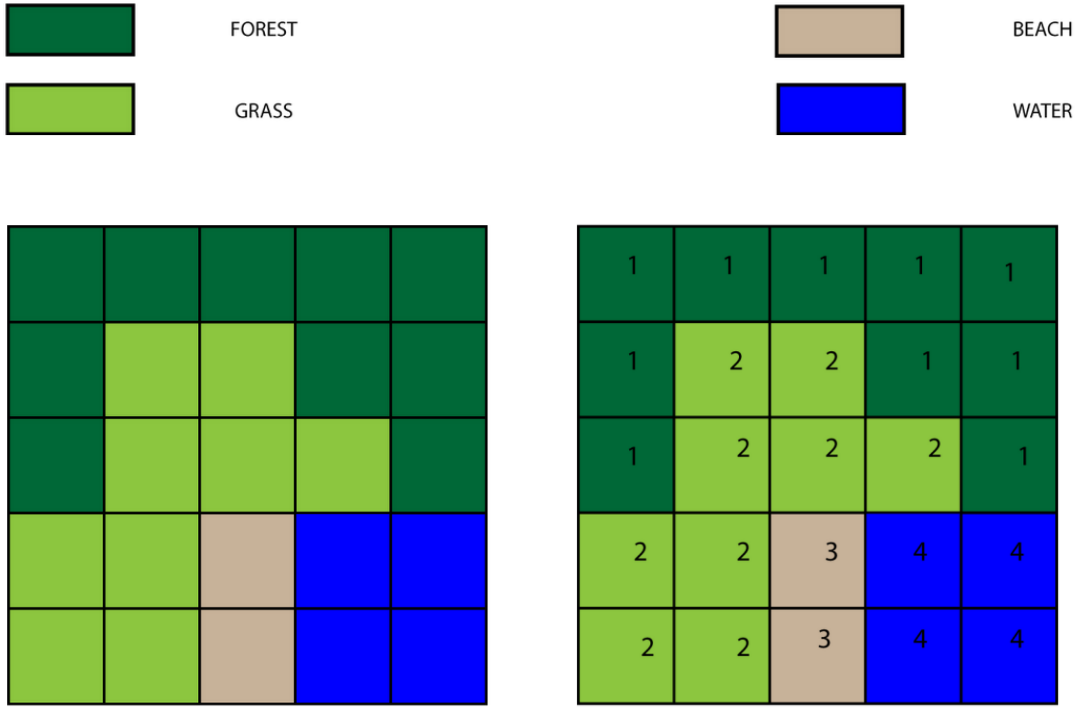
Spatial data types - vectors

- Points
  - Single x,y coordinate
- Lines (polylines)
  - A series points (x,y coordinates)
- Polygons
  - A single line
  - Start and end coordinates are the same



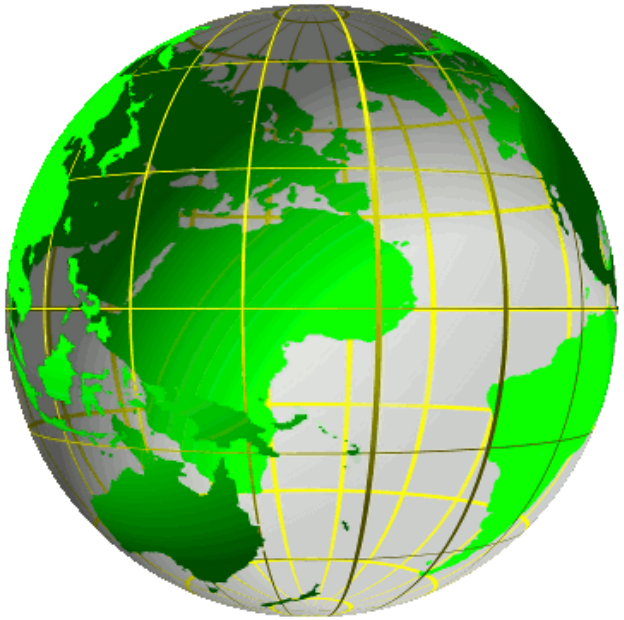
Spatial data types - raster's

- Continuous data
- Raster resolution: compromise between detail and storage size



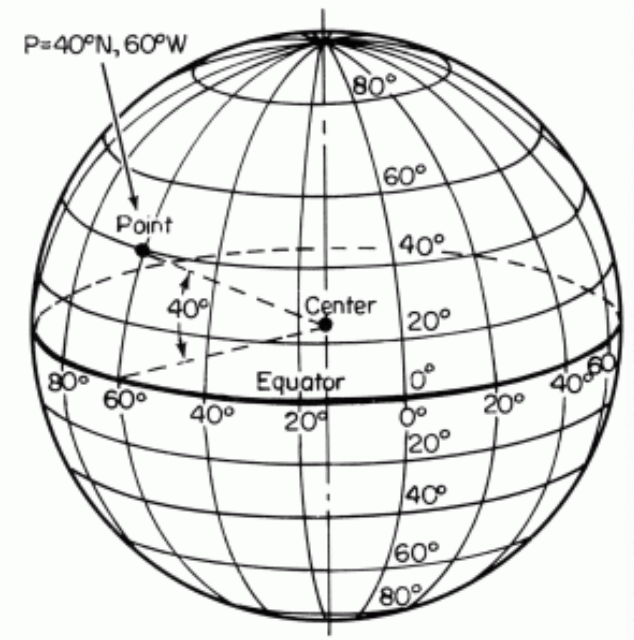
Coordinate systems

- A way of specifying a location on earth
  - E.g. X, Y, H
- Hundreds of coordinate systems
  - Each has a unique SRID
  - Spatial Reference ID
- Geographic and Projected systems
- Each coordinate system has its own datum for height

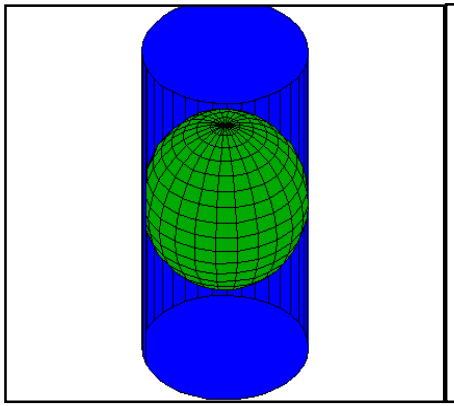


Coordinate Systems

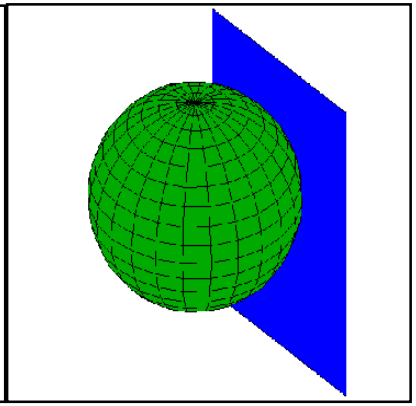
- Geographic
  - Based on a model of the surface
  - Latitude & longitude
  - Angular measurements
- Projected
  - 2 dimensional projection of the surface
  - Will always be distortions
  - Eastings & Northings



Cylindrical

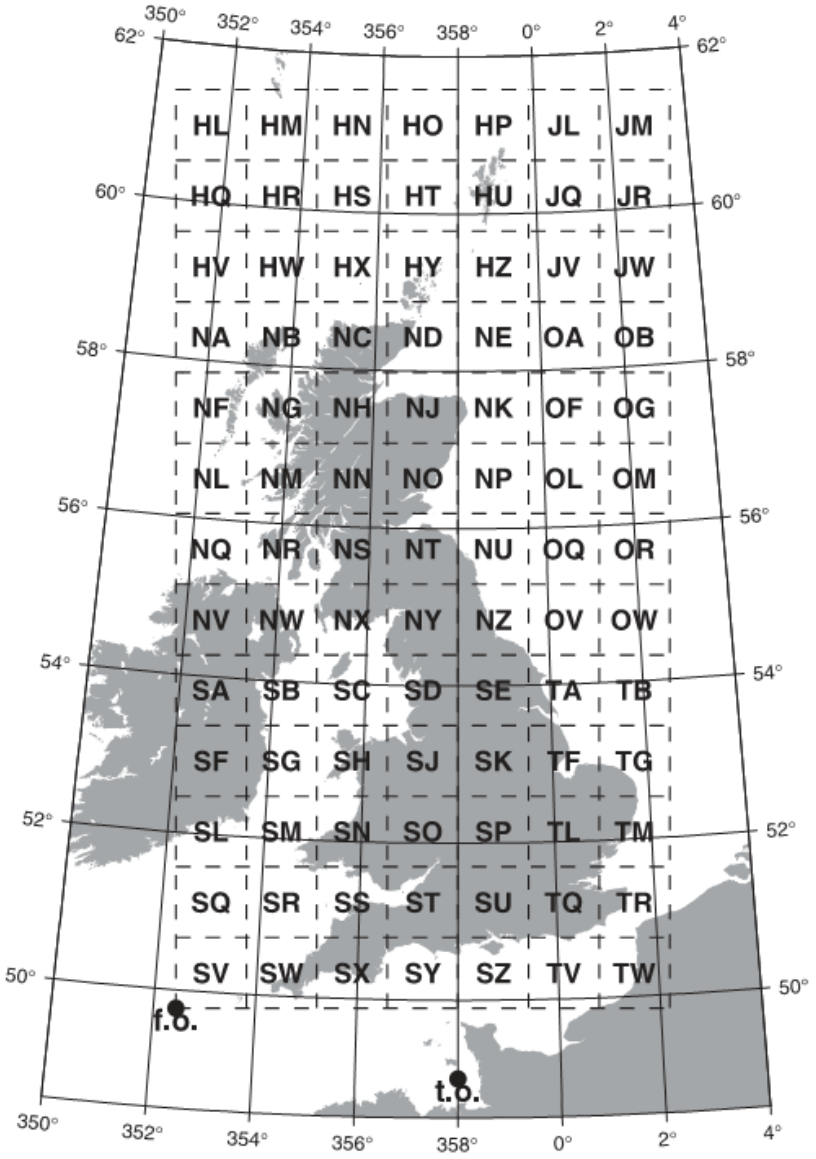


Azimuthal



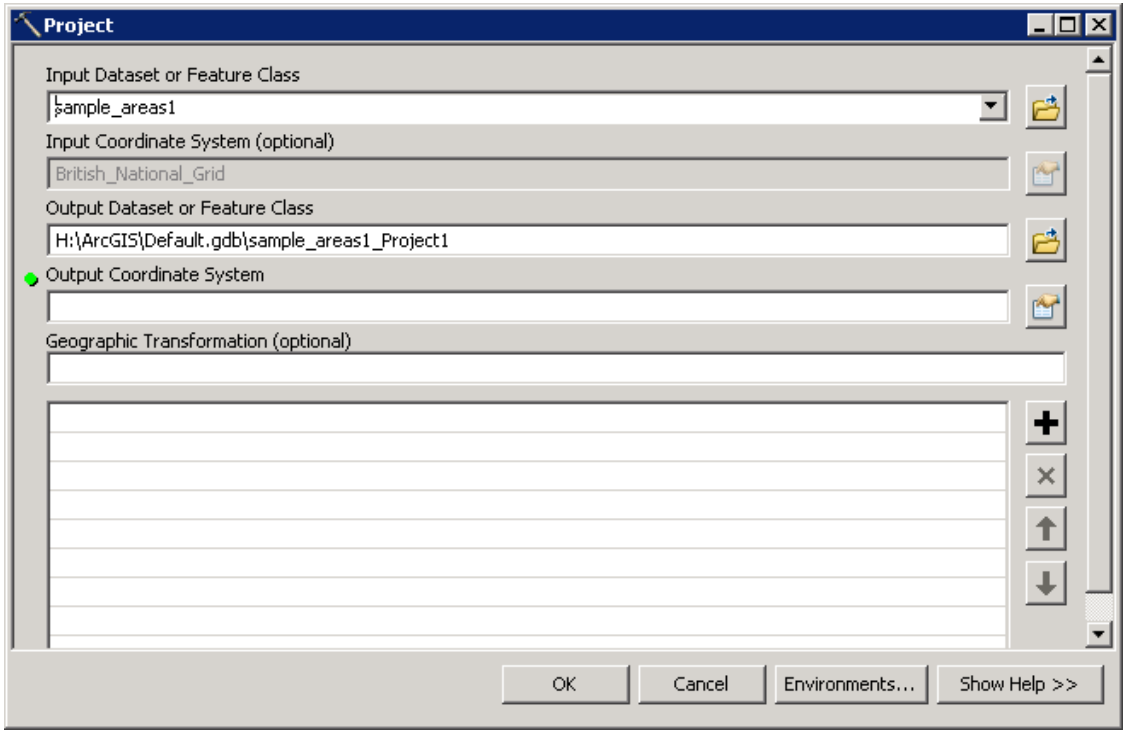
Coordinate Sys. - GB

- WGS84
  - Geographic system
  - SRID: 4326
  - GPS data
  - 51.758786, -1.2537852
- OSGB36
  - Projected system
    - Easting & Northing
  - SRID: 27700
  - Datum: Newlyn
  - 451601, 206941 (SP)



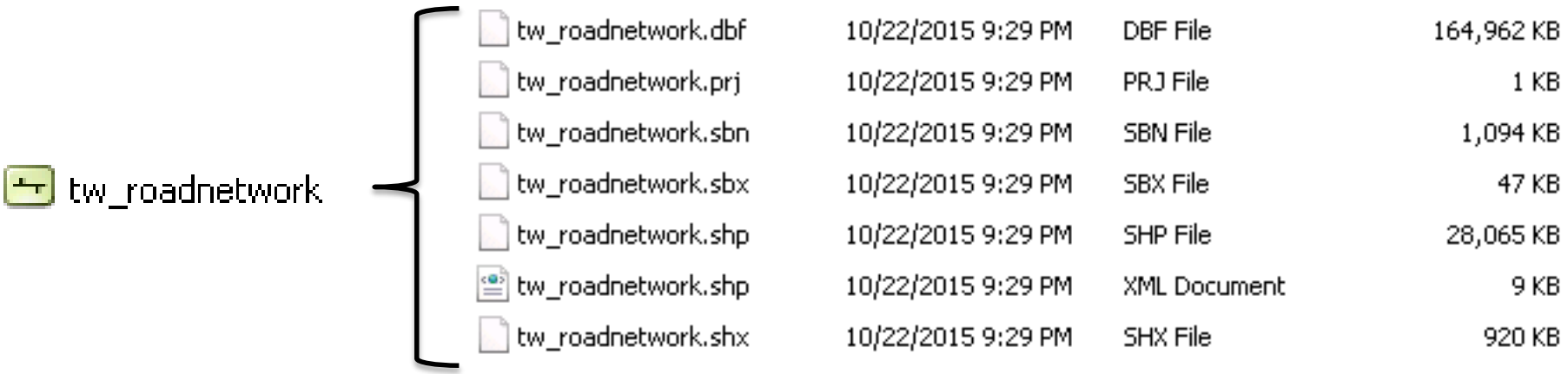
Coordinate Sys. - conversions

- Data can be converted between coordinate systems
  - Can introduce errors though
- Most GIS systems/tools allow conversions
  - Arc: Project tool
  - QGIS: Define projection when saving as new layer
  - GDAL…



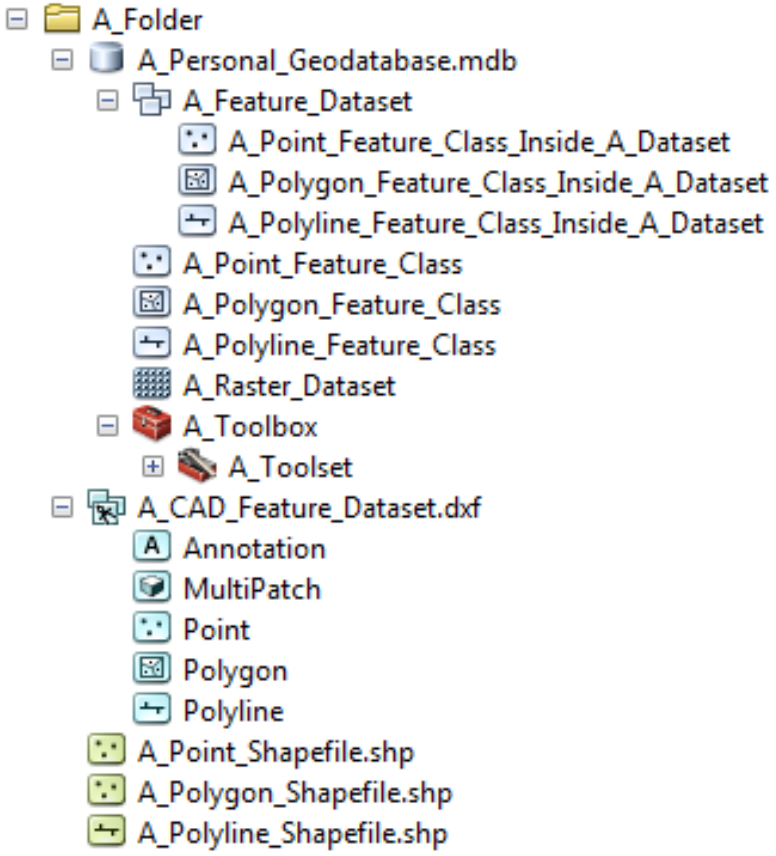
Data management

- Shapefiles
  - Store vector data
  - Points, lines and polygons
  - 4 core files: .dbf, .shp, .shx, .sbn
  - May also include others: e.g. .prj



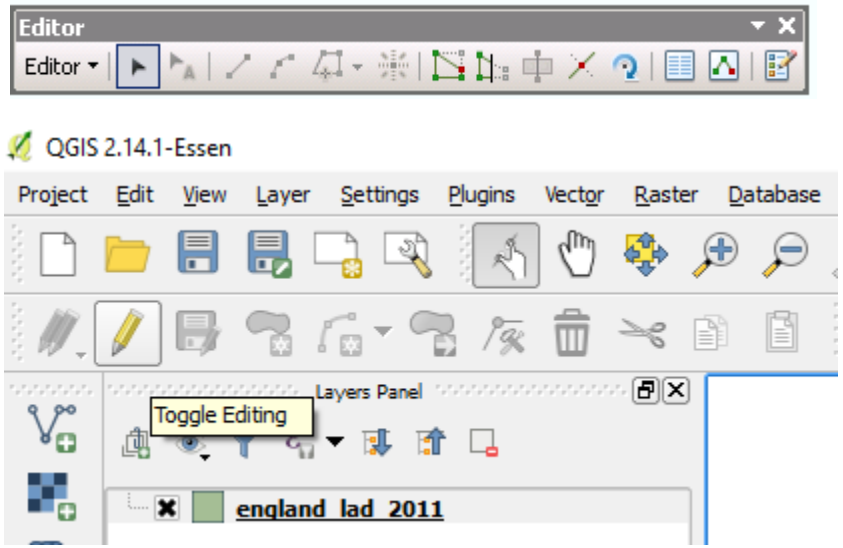
Data management

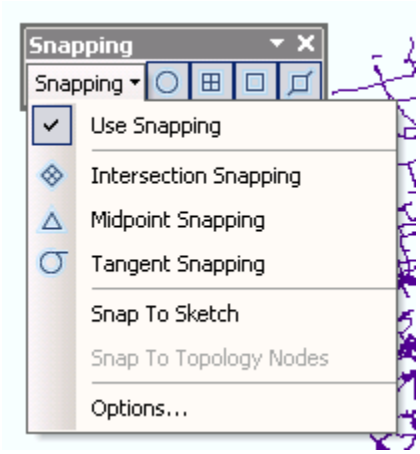
- Geodatabase's – Arc only
  - A folder for shapefiles
  - Feature class = shapefile
  - Feature dataset = sub-folder
    - Contains feature class's
    - All with the same coordinate system



Editing data (digitizing)

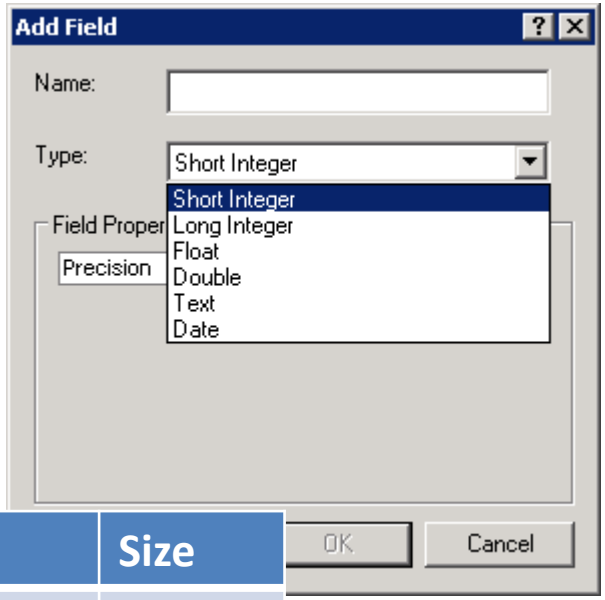
- Editing/Adding points/lines/polygons
  - Best done in a GIS package, but can be programmed
  - Editor toolbar in ArcMap
  - Edit button in QGIS
  - Move and modify existing features
  - Create new features
  - Snapping





Editing data

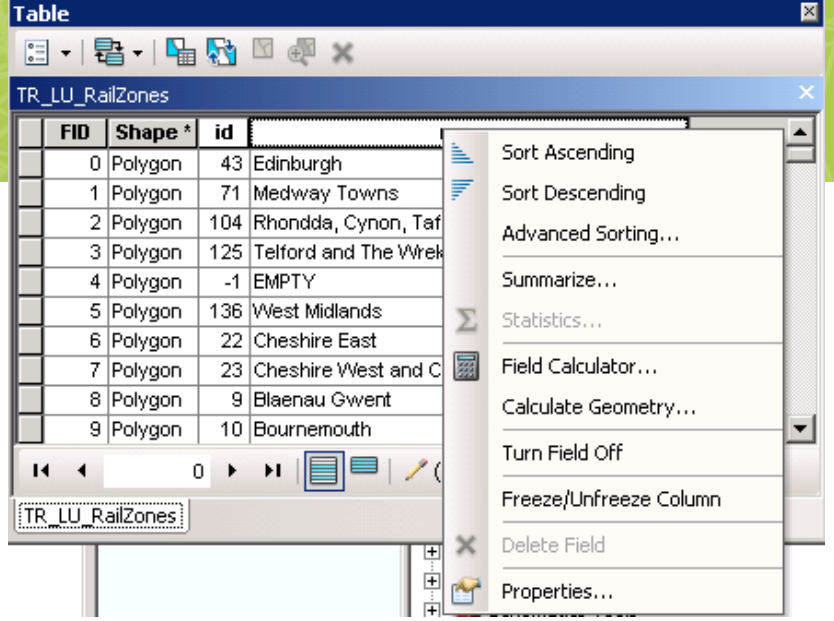
- Adding attributes
  - Can be done from attribute tables
  - Column data types restrict data stored e.g. ArcGIS:

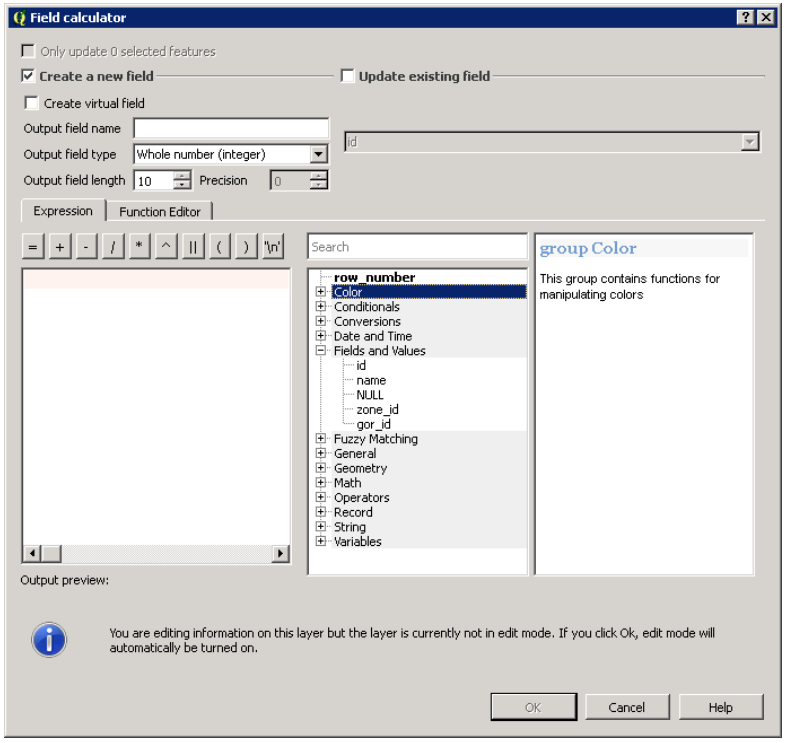


Data type	Data ranges	Size
Short integer	-32,768 to 32,767	2
Long integer	-2,147,483,648 to 2,147,483,647	4
Float	Approx. -3.4E38 to 1.2E38	4
Double	Approx. -2.2E308 to 1.8E308	8
text	Text	
date	Dates	

Editing data

- Editing attributes
  - Manually
    - Need to be in ‘editing mode’
  - Field calculator
    - Create more complex updates
- Calculating geometry
  - Area, length etc.
  - QGIS: Field calculator
  - Arc: Calculate geometry





ITRC mistral
multi-scale infrastructure systems analytics

Common spatial processes

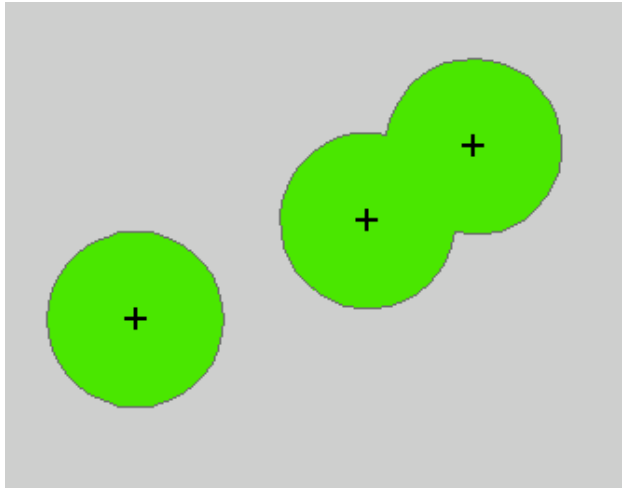
- Buffers
- Clip
- Intersect
- Tabular data
- Selections
- Joins

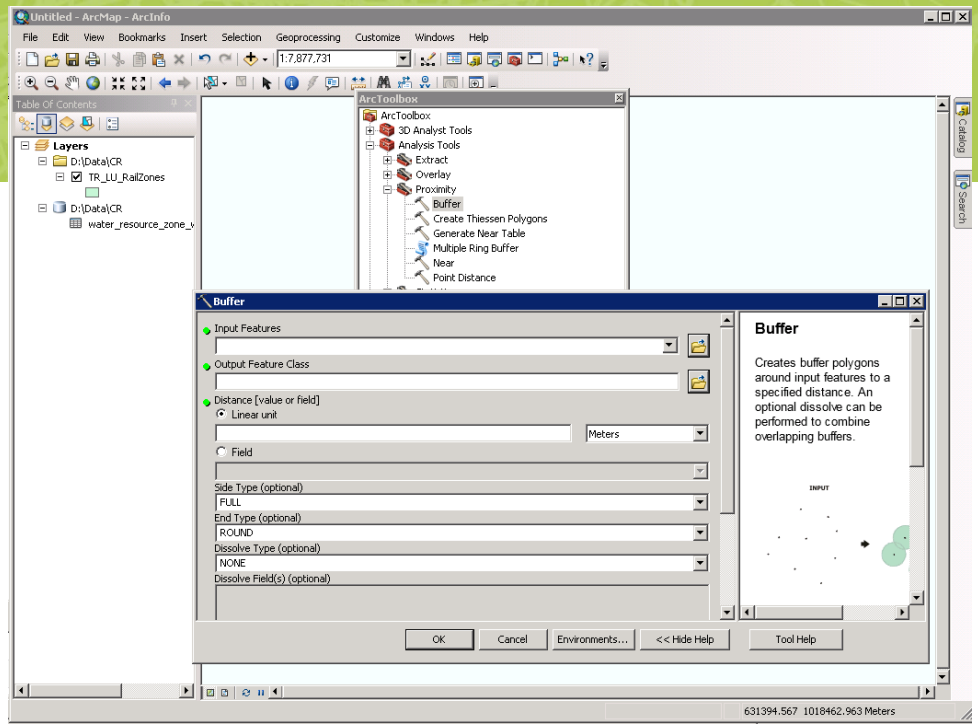
EPSRC
Engineering and Physical Sciences Research Council

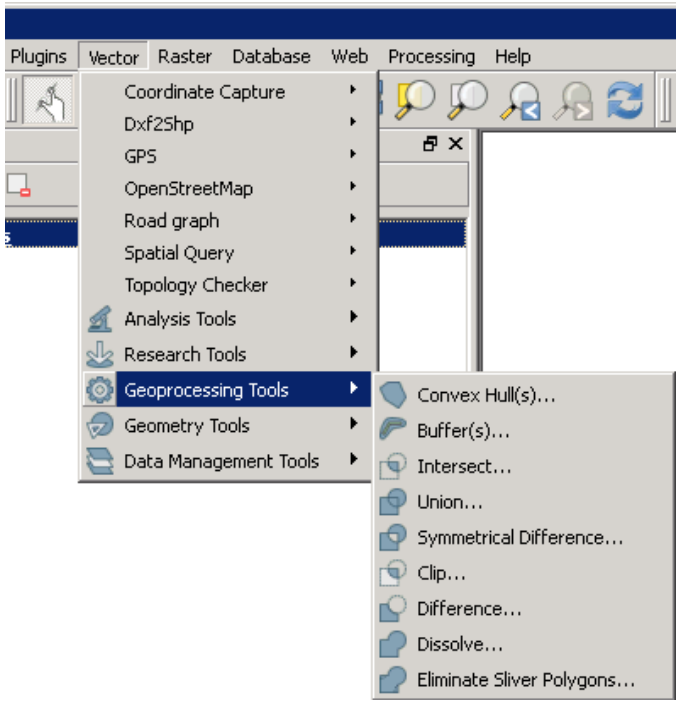
ITRC mistral
multi-scale infrastructure systems analytics

Buffers

- How to do a buffer
  - Create a polygon around existing features with a set distance
- Dissolving buffers
- Multiple (ring) buffers







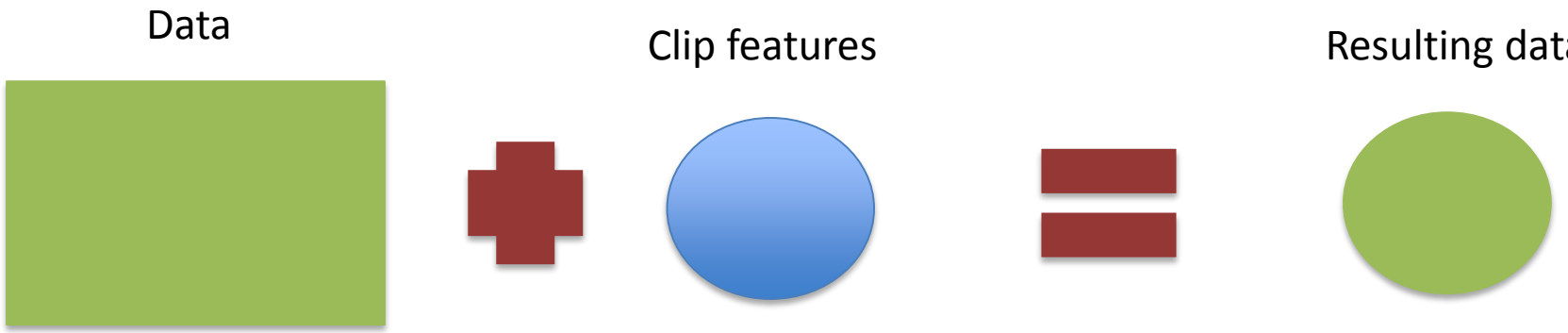
EPSRC
Engineering and Physical Sciences Research Council

Clip

- What does clip do?

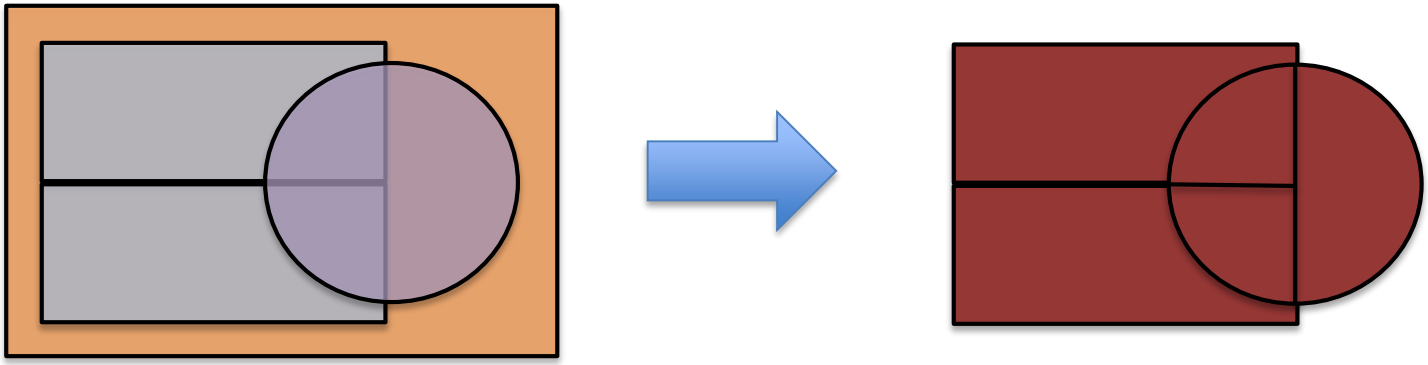
"Extracts input features that overlay the clip features. Use this tool to cut out a piece of one feature class using one or more of the features in another feature class as a 'cookie cutter'".

- Used to cut datasets down e.g. to your area of interest



Intersect

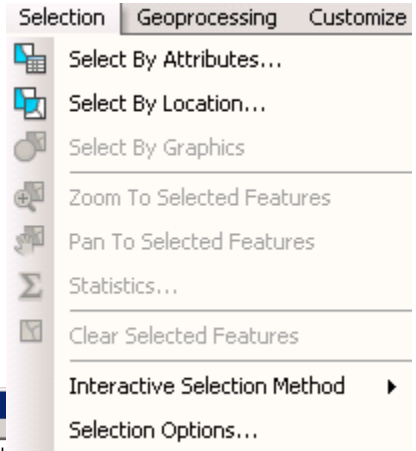
- What does intersect do?
  - Returns the features which intersect, with overlaps forming new features

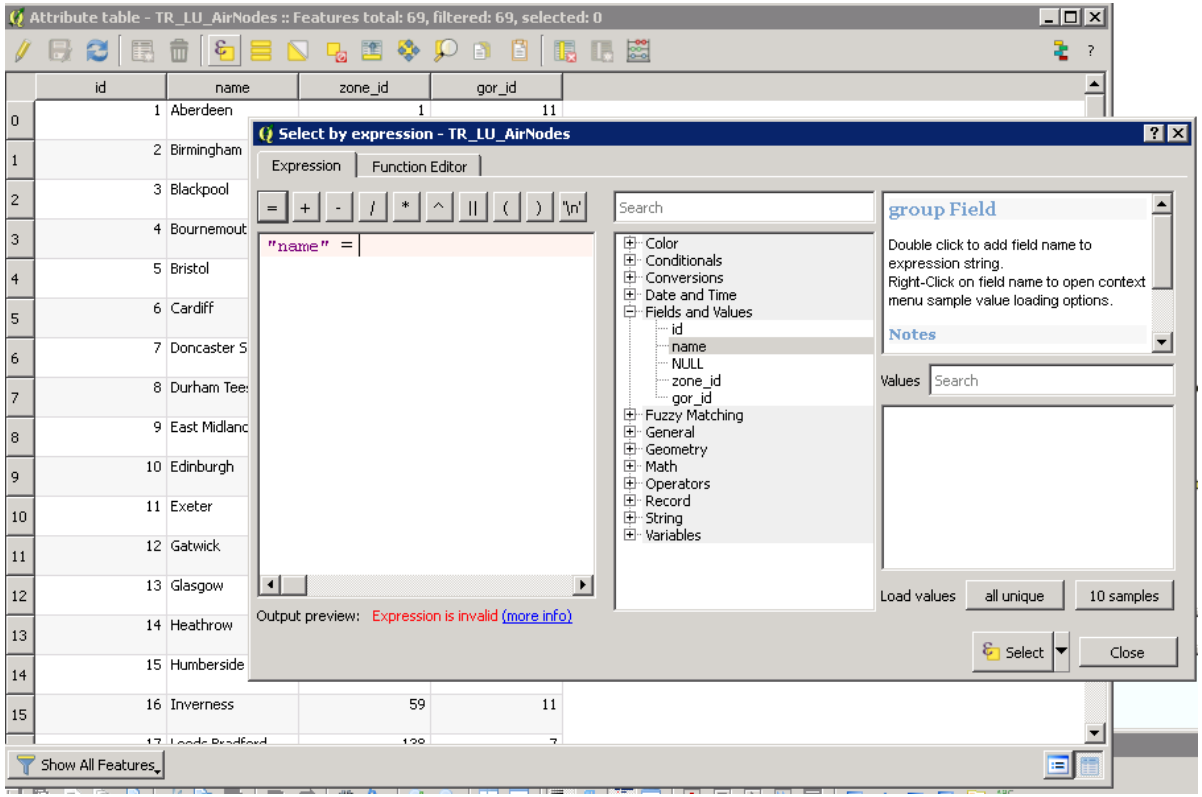


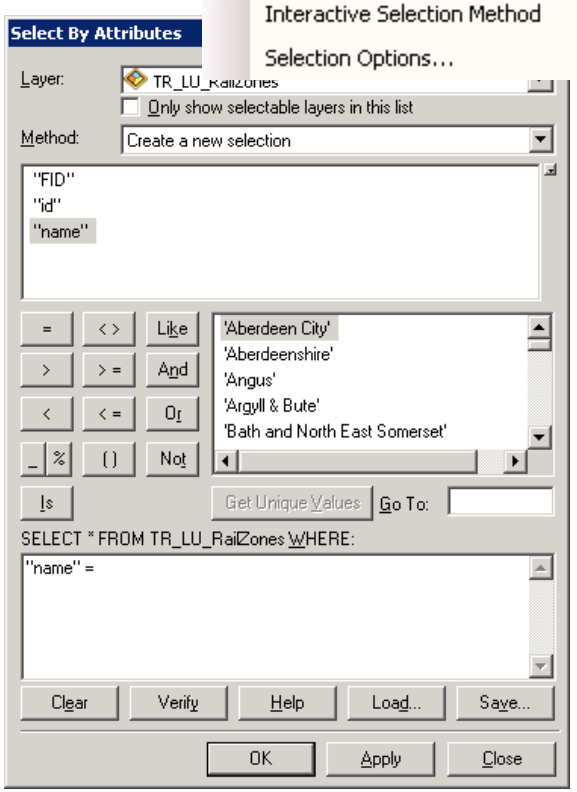
- Useful for finding areas which fall within multiple areas

Selections

- Selecting a subset of a dataset
- Select By Attribute
  - Select features on a set of rules based on attribute values

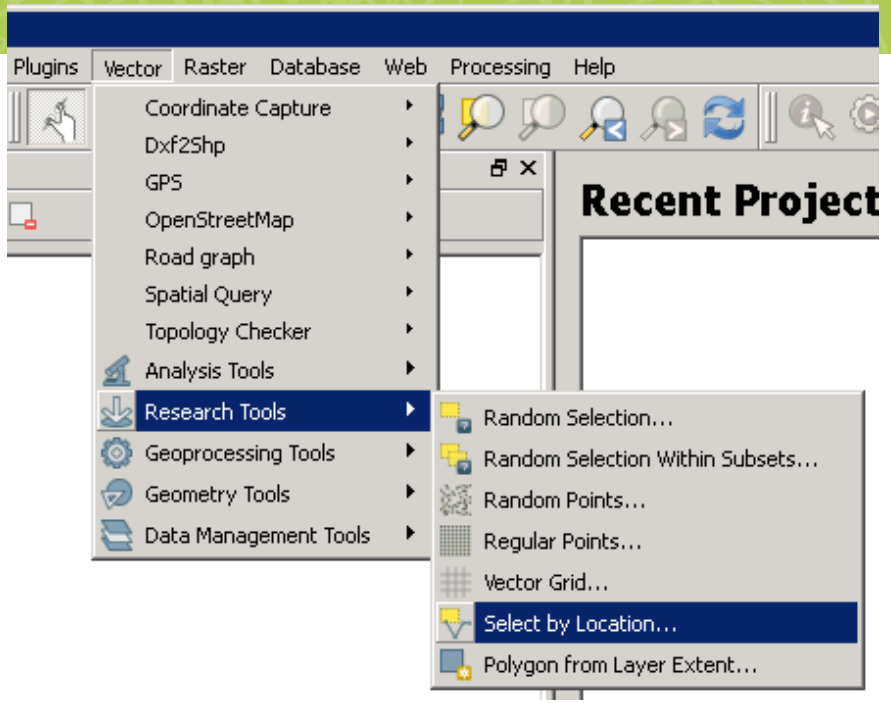


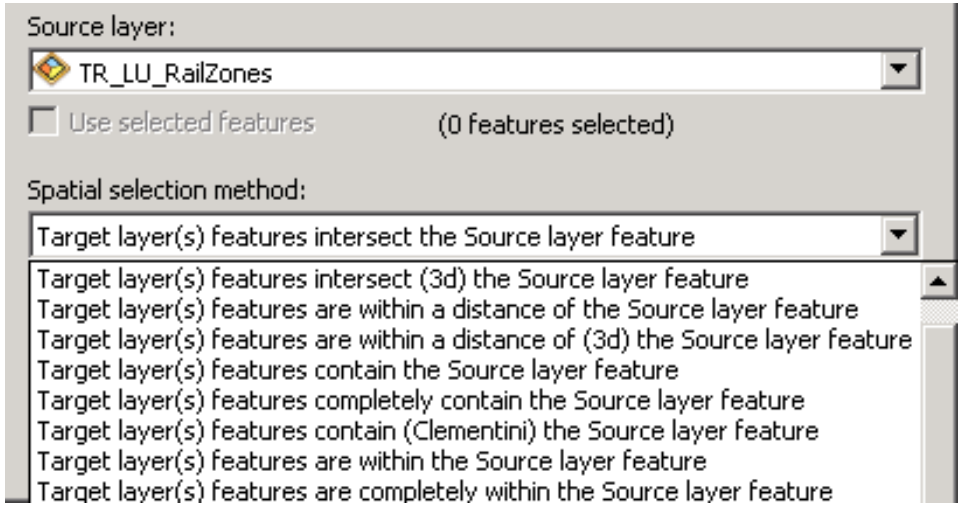


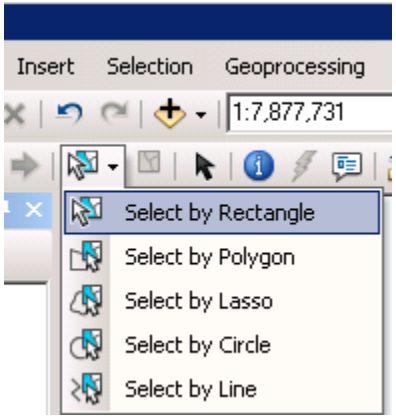


Selections

- Select By Location
  - Select features based on their spatial location with regard to another layer
- Manual selection

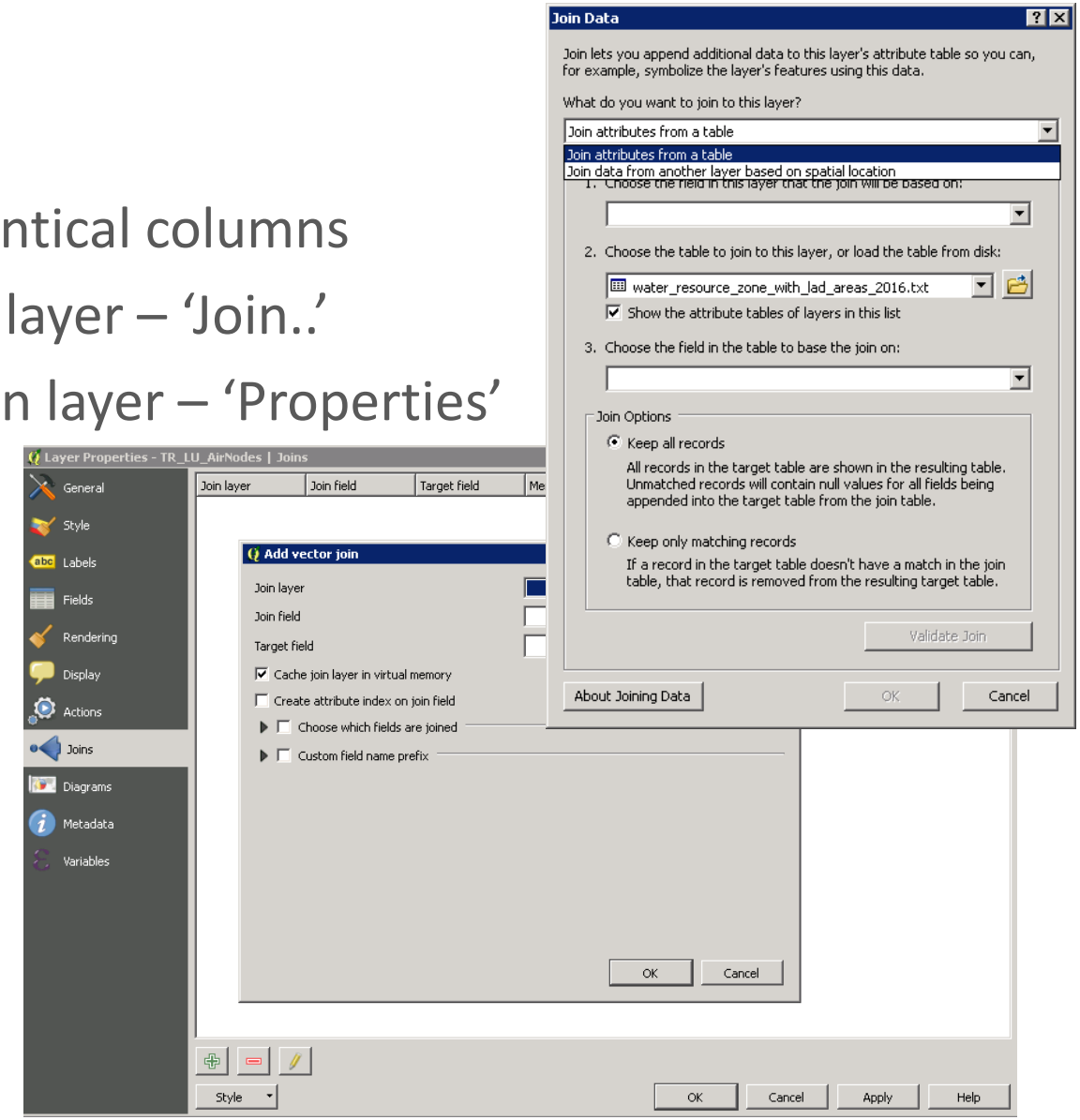






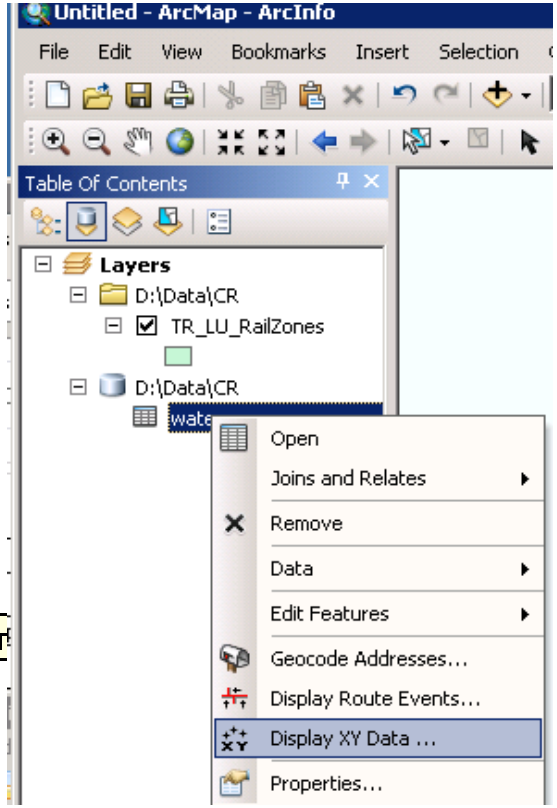
Joins

- Join by data
  - Two files have identical columns
  - Arc: right-click on layer – ‘Join..’
  - QGIS: right-click on layer – ‘Properties’
- Spatial Join
  - Based on a rule



Tabular data

- Use a join to add to another dataset with geography
- Tell the GIS what the spatial columns are
  - Arc: right-click – 'Add XY'
  - QGIS: add a csv layer
  - Select columns with X and Y data in




```
1  FID, point_att_1, point_att_2, easting, northing
2  1, 240, 'Operational', 54123.34, 12348.78
3  ...
4
```

ITRC mistral
multi-scale infrastructure systems analytics

Data management

- Databases
  - External to GIS systems
  - Spatially enabled databases allow spatial and non-spatial data to be stored in a generic format
  - GIS systems can connect directly e.g. QGIS
  - Database stores all spatial information as well as attributes

e.g. Water Pumping stations

gid integer	objectid numeric(10,0)	unique_ref numeric(10,0)	name character varying(254)	county character varying(254)	postcode character varying(254)	geom geometry
1	286354	18229077	Pump (Disused)	Dumfries and Galloway	DG12	0101000020346C0000000000002C6413410000000070
2	327397	18334772	Pump (Disused)	Cornwall	TR19	0101000020346C00000000000088C80041000000008(
3	327402	18334782	Wind Pump	Cornwall	TR19	0101000020346C00000000000C0D900410000000080
4	328508	18341519	Pump	Devon	EX7	0101000020346C00000000000E0F81141000000B(
5	328510	18341527	Pump	Devon	EX7	0101000020346C000000000008CF8114100000003(
6	328515	18341539	Pump	Devon	EX7	0101000020346C00000000000C8EF1141000000050
7	328533	18341569	Pumping House	Devon	EX7	0101000020346C00000000000014DC1141000000D(
8	583751	18334771	Wind Pump (Disused)	Cornwall	TR19	0101000020346C00000000000060BC0041000000040
9	669761	18334773	Wind Pump (Disused)	Cornwall	TR19	0101000020346C00000000000088C8004100000000(
10	741245	18334759	Hydraulic Ram	Cornwall	TR19	0101000020346C0000000000003865004100000008(
11	754037	18341524	Wind Pump	Devon	EX6	0101000020346C000000000088F51141000000D(
12	327364	18334703	Pumping House	Isles of Scilly	TR24	0101000020346C00000000000030C6F540000000008(

EPSRC
Engineering and Physical Sciences Research Council

ITRC mistral
multi-scale infrastructure systems analytics

Data sources

- Main spatial data sources
  - Ordnance survey
  - Open street map (volunteer generated)
  - Government departments (data.gov.uk etc.)
- Open source data v known data
  - Limitations of open source data
    - …..
  - Advantages of open source data
    - …..

EPSRC
Engineering and Physical Sciences Research Council

Discussion (15mins)

- ~10mins in groups, 5mins open
- Data sources/reliability:
  - Volunteered data e.g. Open street map v open v commercial etc.
  - Data verification
- Data management:
  - Folders/databases?
  - Version control?
  - Is everyone using the same version?
  - How often should be data be updated?

Lunch!

- 2pm restart

ITRC mistral

multi-scale infrastructure systems analytics

Spatial Training:

Introduction to spatial data Part 2

Newcastle University

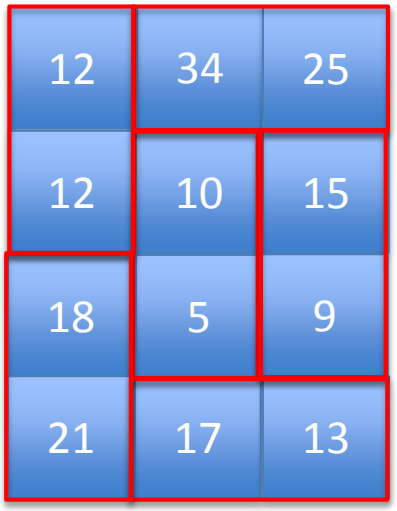
Craig Robson

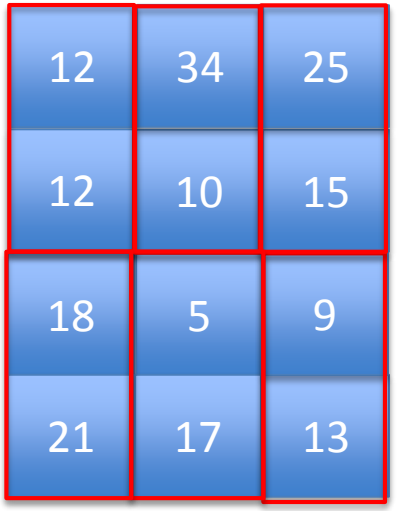
November 2016

EPSRC
Engineering and Physical Sciences Research Council

Modifiable areal unit problem

- Problem caused by using spatial areas
  - E.g. population density
  - The density of a city changes on how you draw the boundary of the city
  - How do you split a geographical space into areas where data is continuous
- Census example
  - If a deprived area of a city is a zone itself, it will be seen
  - If the same area is split amongst other zones, it might not bee seen
  - Neither zoning pattern is wrong





ITRC mistral
multi-scale infrastructure systems analytics

Modifiable areal unit problem

- For us the problem is exacerbated by using different geographies
- Given the same data, you can get different results depending on how you aggregate it
- MISTRAL
  - buildings, super output areas, postcode areas, telephone exchange areas, local authority district areas, council areas, government office region areas, water resource zone areas, substations.....

EPSRC
Engineering and Physical Sciences Research Council

ITRC mistral
multi-scale infrastructure systems analytics

Modifiable areal unit problem

- 2 aspects
  - Zone
    - The shape of the zone's being used change
    - E.g. from 2001 census boundaries to 2011 census boundaries
    - E.g. electoral boundaries
  - Scale
    - Different levels of scale are used for different results (or inputs in our case)
    - E.g. local authority district (380+) areas and government office regions (11)

EPSRC
Engineering and Physical Sciences Research Council

ITRC mistral
multi-scale infrastructure systems analytics

Modifiable areal unit problem

- There is no ‘right’ solution
- Each solution will give a different answer
- Need to think carefully
- Case by case basis
- Further reading
  - S. Openshaw (1984)
  - Fotheringham and Wong (1991)

Converting between geographies

- Area density values
  - Estimate values based on density and zone sizes
- Spatial interpolation
  - ‘the procedure of predicting the value of attributes at unsampled sites from measurements made at locations within the same area (Burrough & McDonnel, 1998)

ITRC mistral
multi-scale infrastructure systems analytics

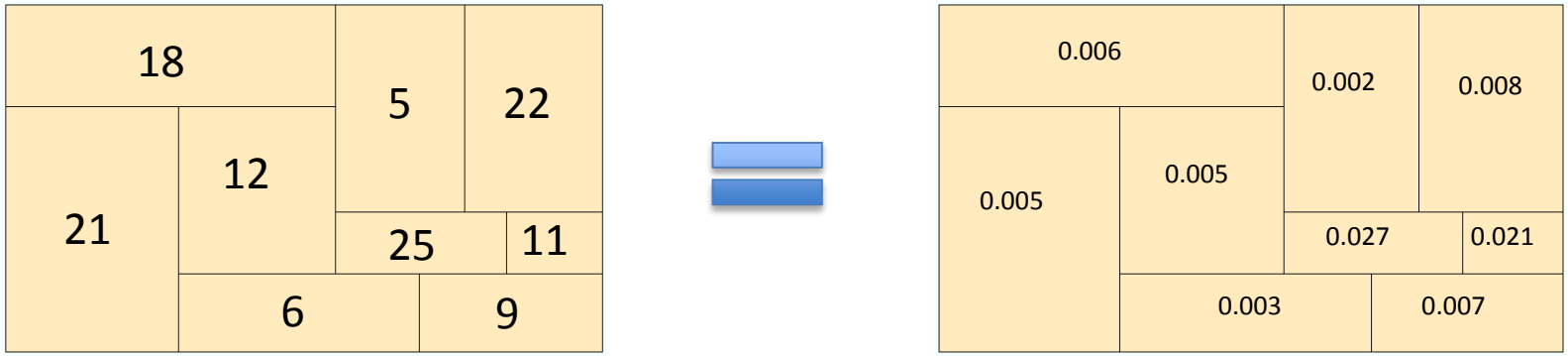
Density approach

- Calculate the density of a variable for each zone in the data set
- Intersect the current data zones with the target data zones
- Calculate the areas of the resulting zones
- Calculate the values in each of the zones using the densities and areas
- Sum for the values for each target zone to get a total value

EPSRC
Engineering and Physical Sciences Research Council

Density approach

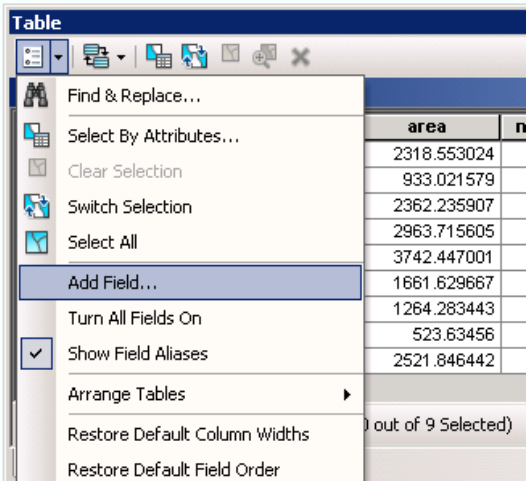
- Calculate the density of a variable for each zone in the data set



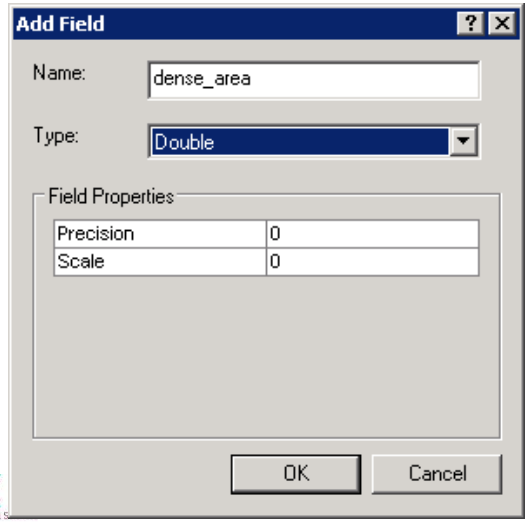
- Add Field > Field calculator

Density approach

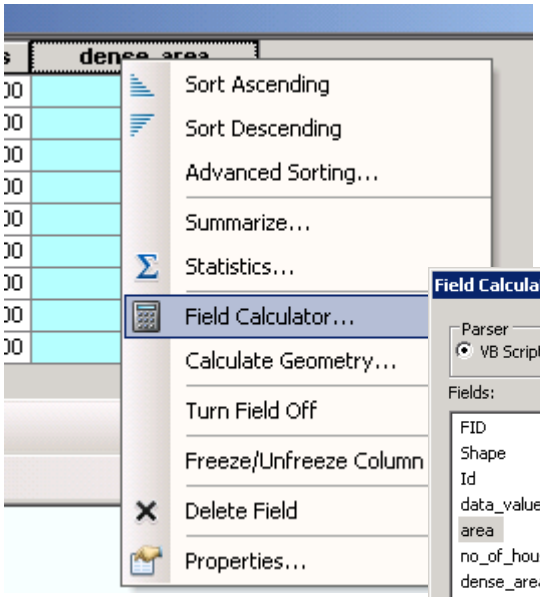
1) Add Field



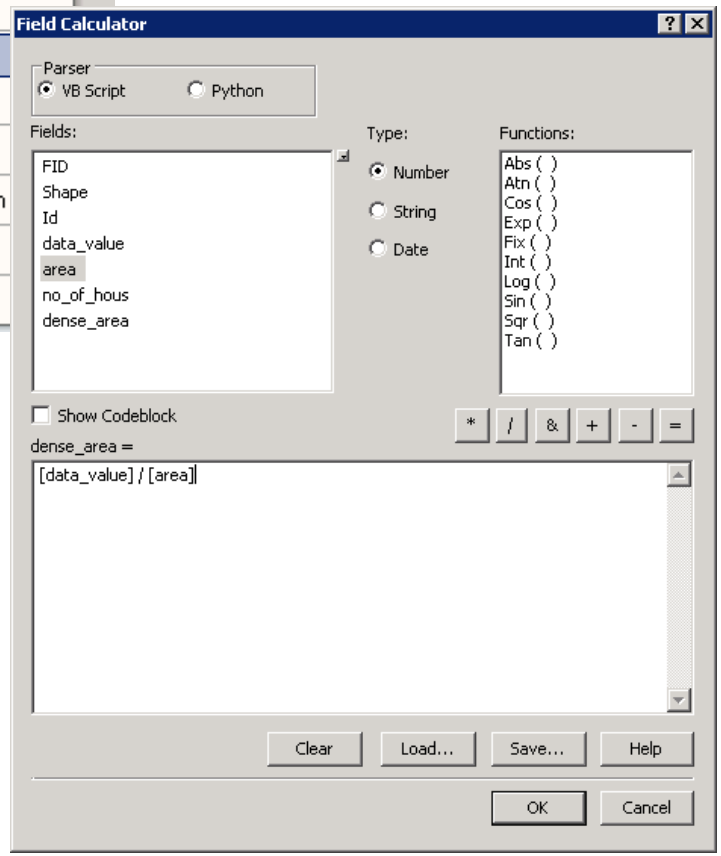
2) Define field



3) Calculate value – Field Calculator

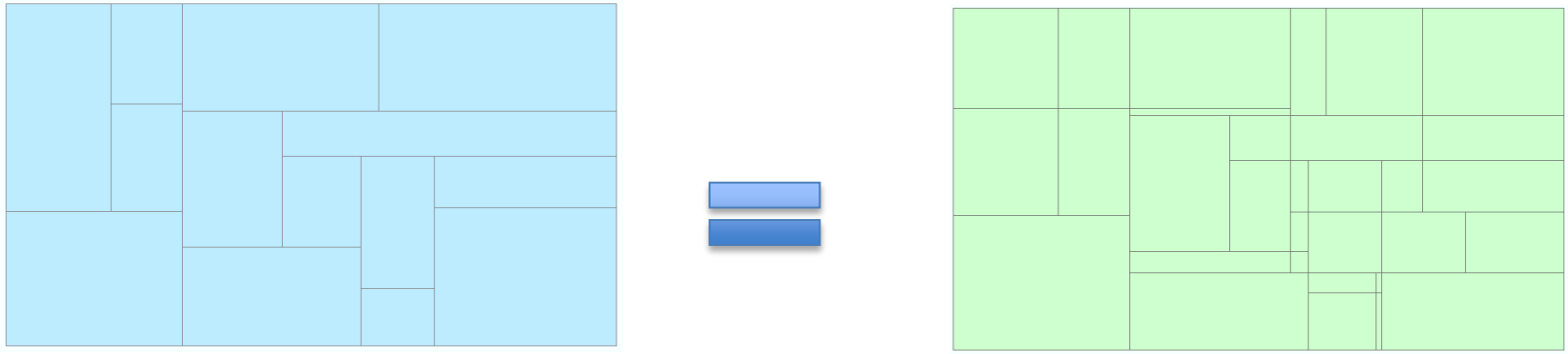


4) Field Calculator – enter formulae



Density approach

- Intersect the current data zones with the target data zones – attributes are copied as well



Table

sample_areas_intersect

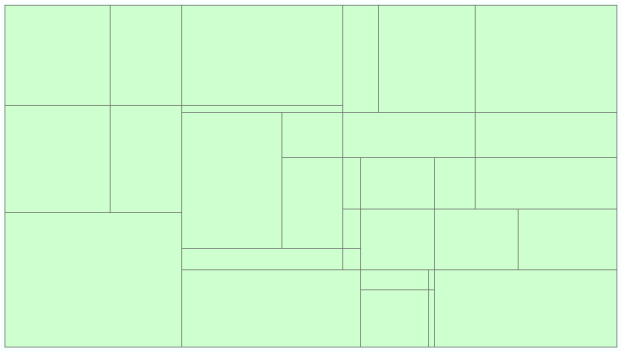
FID	Shape *	FID_sample	Id	data_value	area	no_of_hous	dens_area	data	FID_samp_1	Id_1	data_val_1	area_1	no_of_ho_1
0	Polygon	0	0	12	2318.553024	50000	0.005176	4	1	0	0	1540.101647	20000
1	Polygon	0	0	12	2318.553024	50000	0.005176	4	2	0	0	1191.758656	29000
2	Polygon	0	0	12	2318.553024	50000	0.005176	4	6	0	0	1848.121976	5000
3	Polygon	0	0	12	2318.553024	50000	0.005176	4	7	0	0	628.361472	8000
4	Polygon	0	0	12	2318.553024	50000	0.005176	4	8	0	0	1309.0864	24000
5	Polygon	1	0	25	933.021579	36000	0.026795	4	1	0	0	1540.101647	20000
6	Polygon	1	0	25	933.021579	36000	0.026795	4	7	0	0	628.361472	8000
7	Polygon	1	0	25	933.021579	36000	0.026795	4	10	0	0	849.576072	19000
8	Polygon	1	0	25	933.021579	36000	0.026795	4	13	0	0	2199.265151	21000

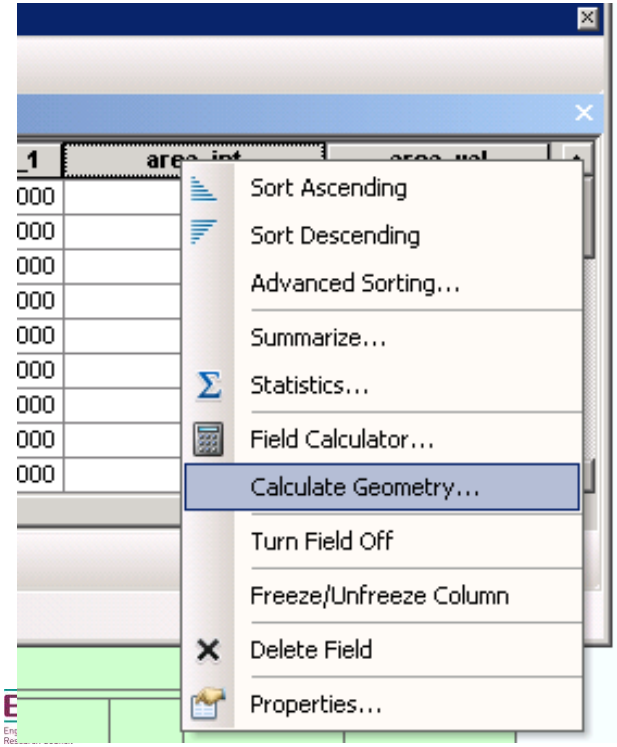
1 (0 out of 31 Selected)

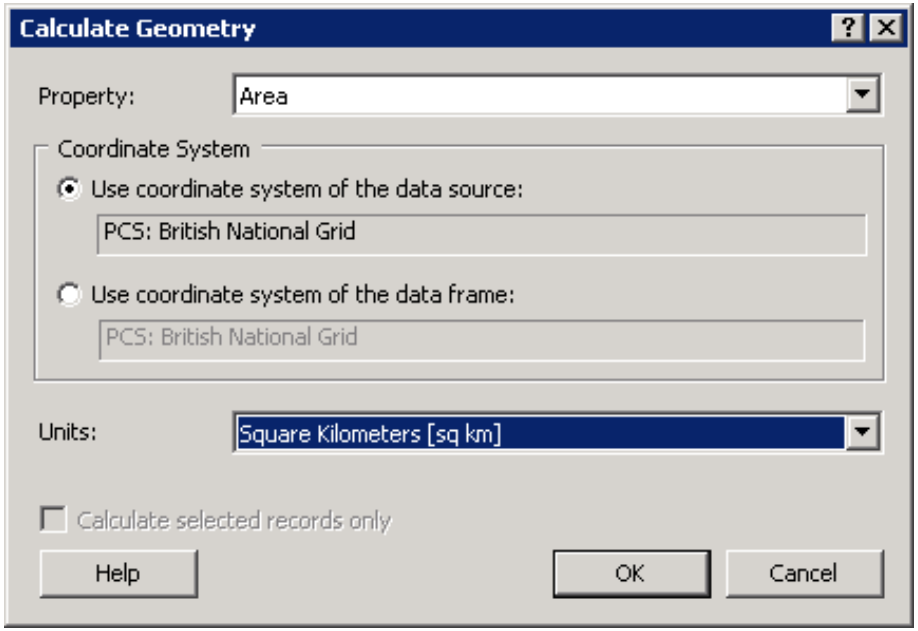
sample_areas_intersect

Density approach

- Calculate the areas of the resulting zones
- Add field > Calculate geometry

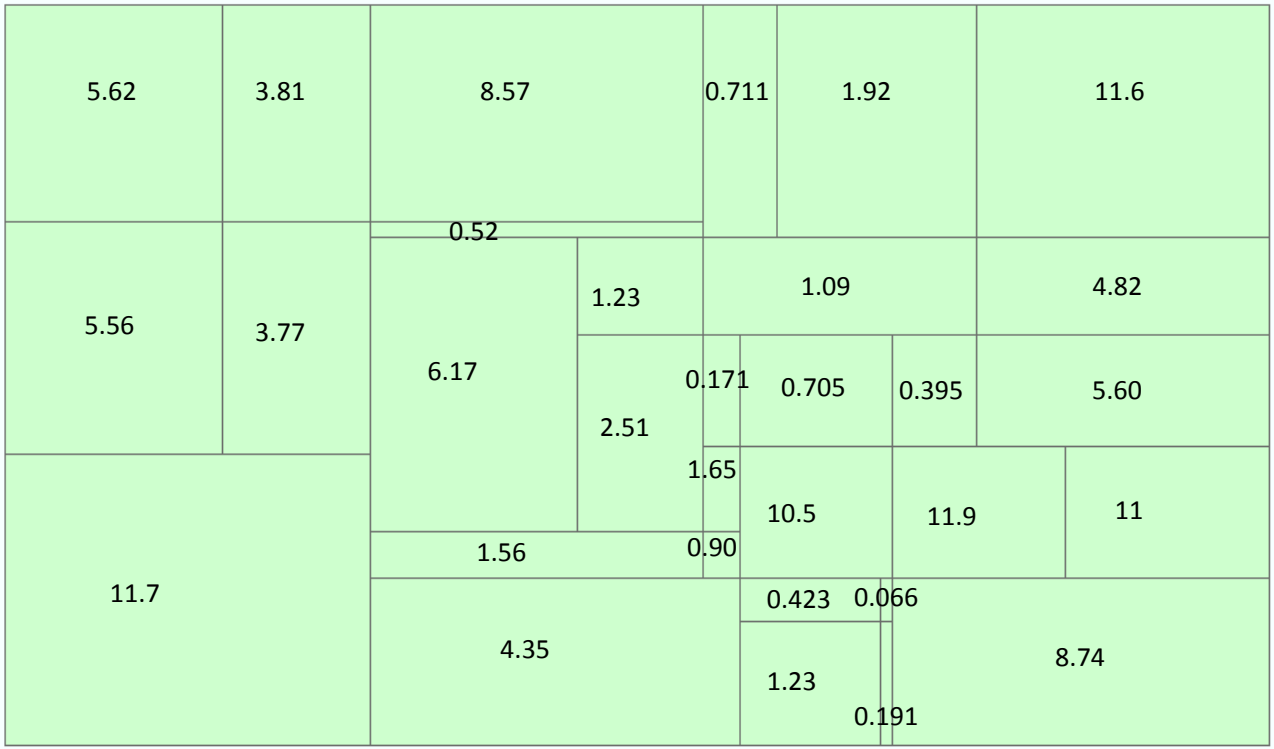






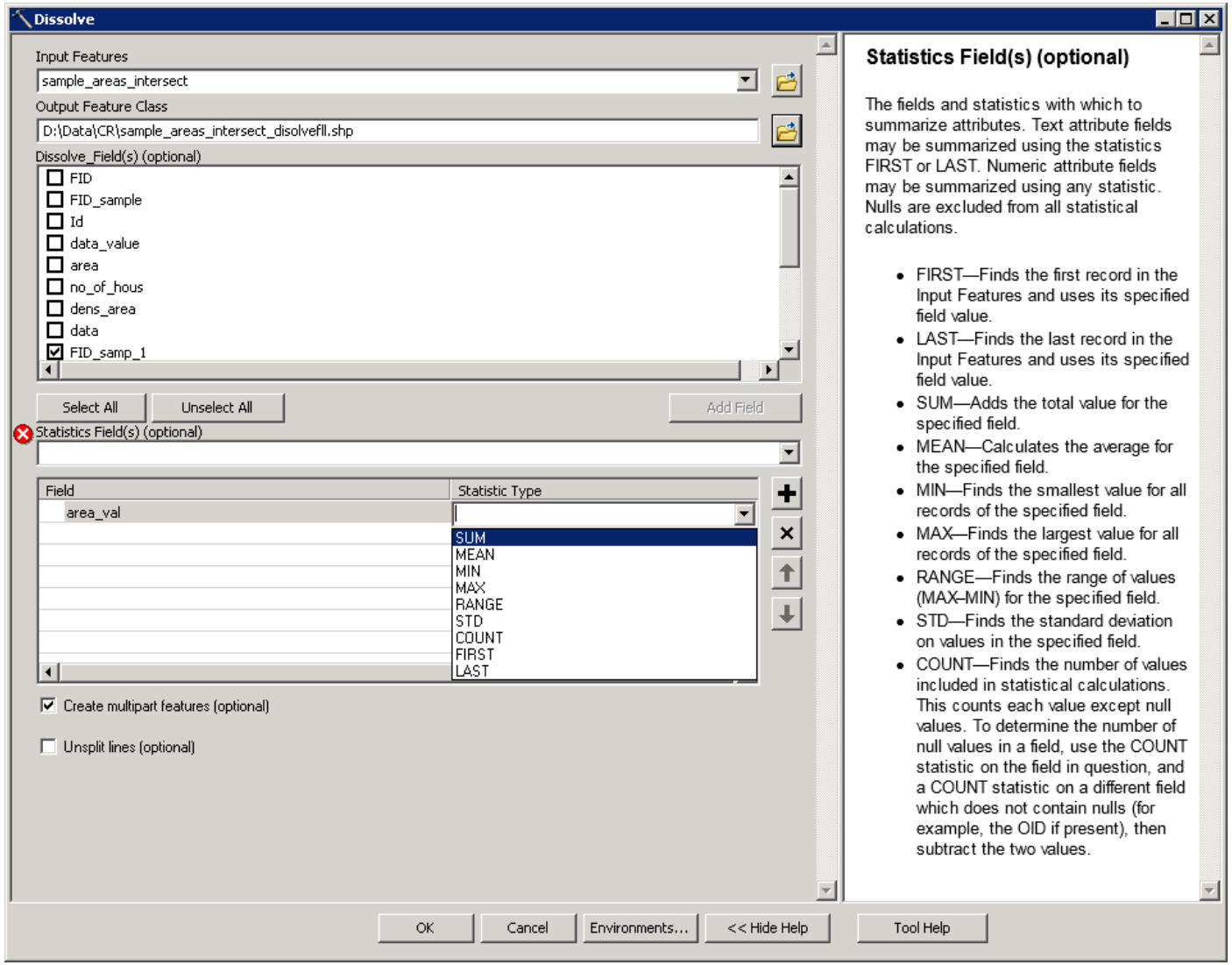
Density approach

- Calculate the values in each of the zones using the densities and areas
- Add field > field calculator



Density approach

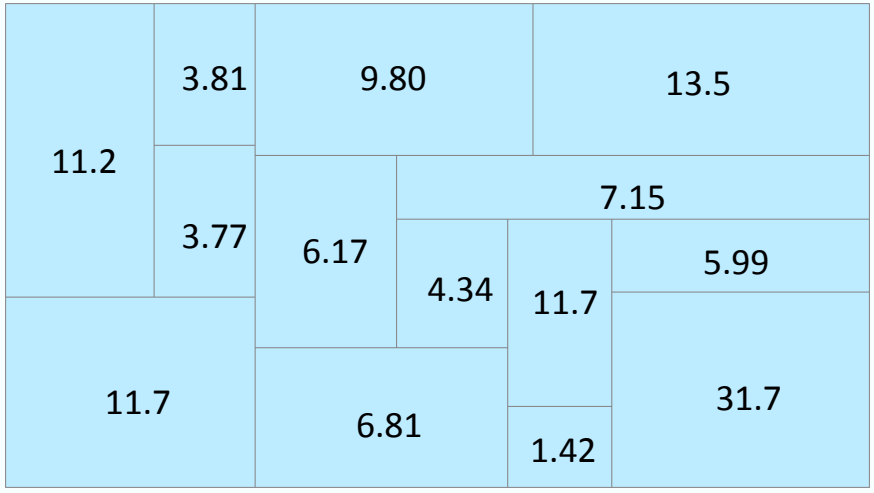
- Sum for the values for each target zone to get a total value (dissolve function)



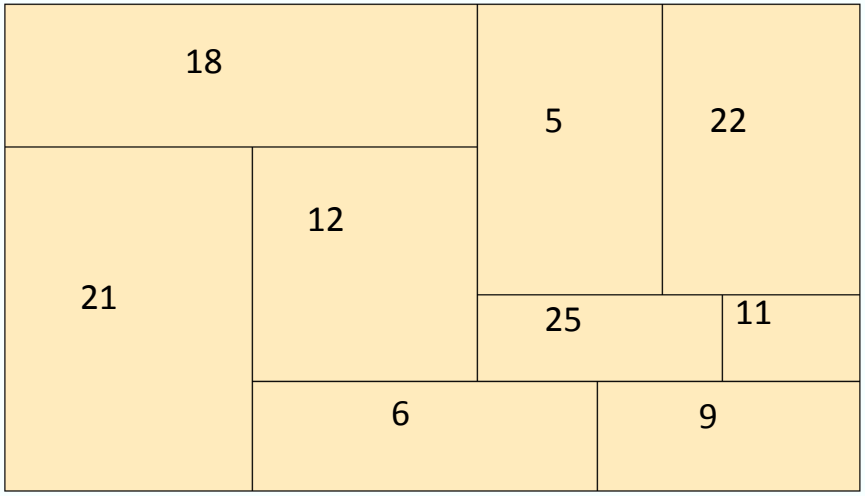
Density approach

- Sum for the values for each target zone to get a total value (dissolve function)

Values in new areas



Original areas and values



ITRC mistral
multi-scale infrastructure systems analytics

Density approach

- Limitations
  - Assumes uniform distribution across the zone
  - Assumes variable distribution is a function of the chosen parameter e.g. area or number of houses

- Advantages
  - Quick method of switching between geographies
  - Computationally simple
  - Can be automated

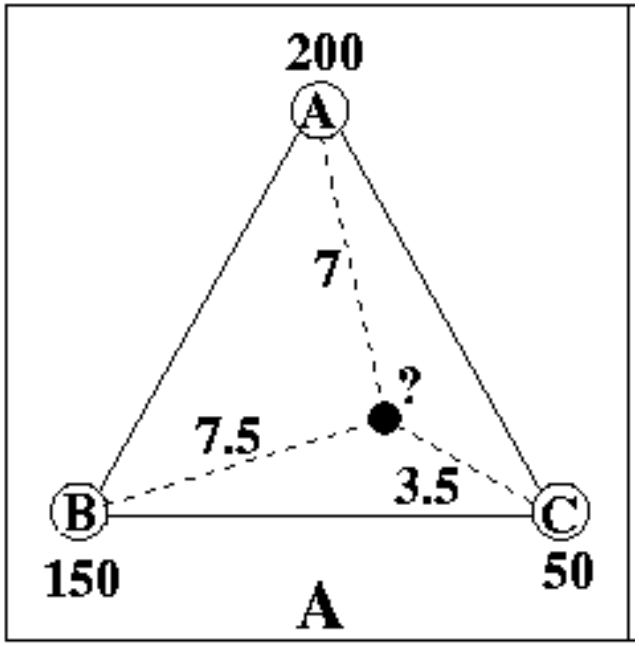
ITRC mistral
multi-scale infrastructure systems analytics

Interpolation

- Filling in gaps in data and generating a surface of values
- 3 common methods
  - IDW (Inverse distance weighted)*
  - TIN (Triangular Irregular Networks)
  - Global trend surfaces

Inverse Distance Weighted

- Based on distance from the unknown to the know
- Distance used to weight each know value's relationship for the unknown
- Weights used to estimate the unknown



Inverse Distance Weighted

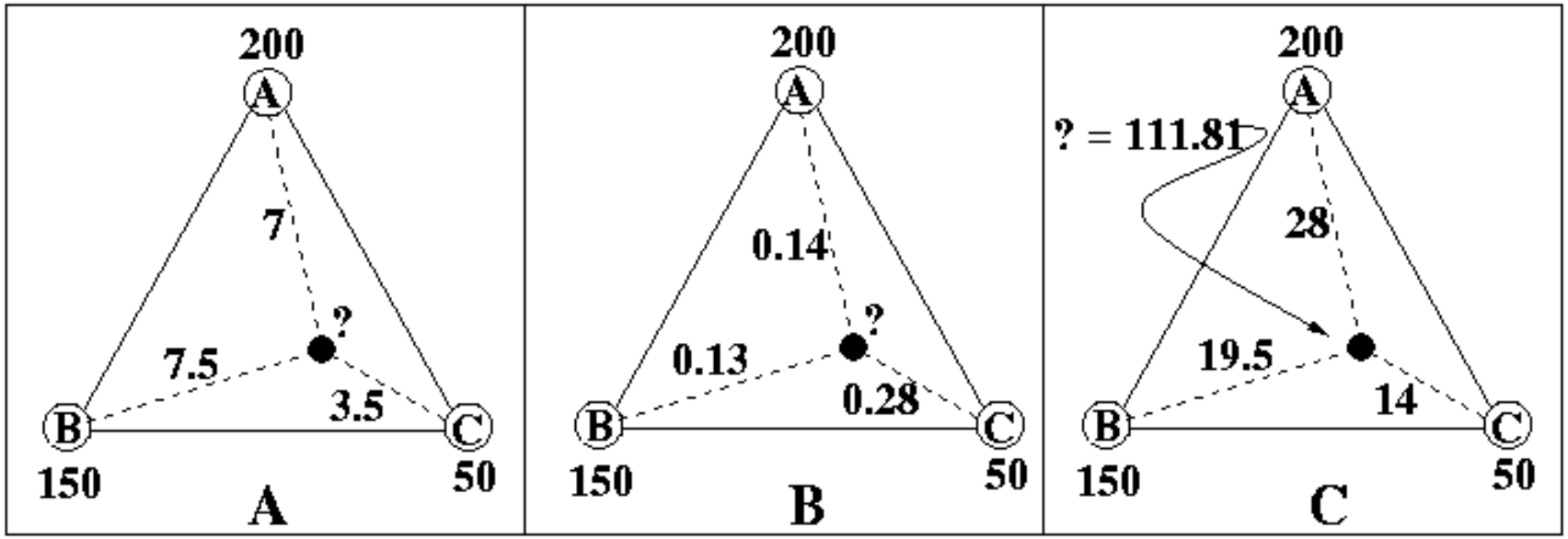
$$Z(x_j) = \sum_{i=1}^{n} z(x_i) \bullet d_{ij}^{-r} \Big/ \sum_{i=1}^{n} d_{ij}^{-r}$$

Where:

$Z(x_j) = Z(x, y)$ = the unknown point to be interpolated

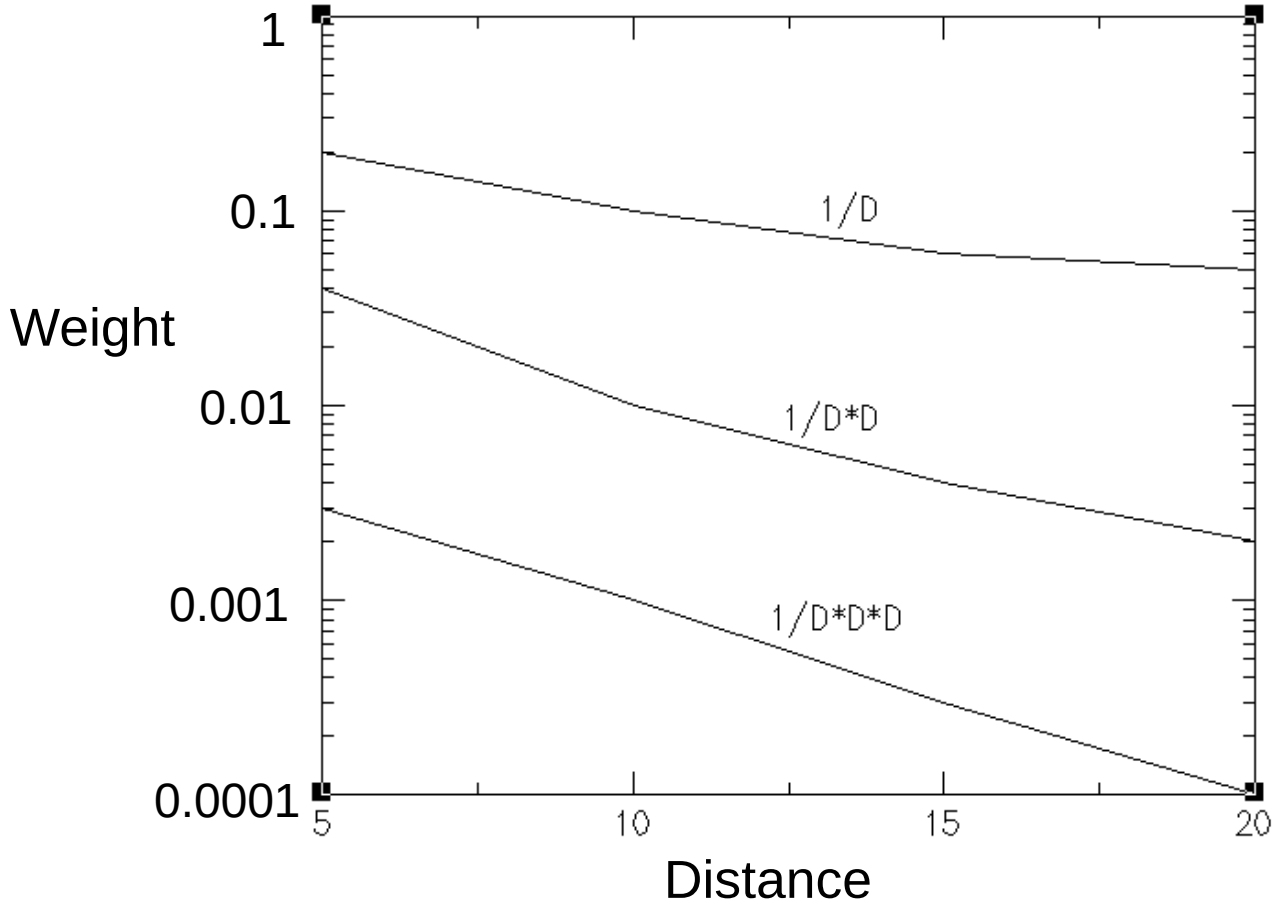
$z(x_i) = z(x, y)$ = the known points used to derive the interpolated point

d_{ij}^{-r} = the distance between a known point and the unknown weighted by a reciprocal



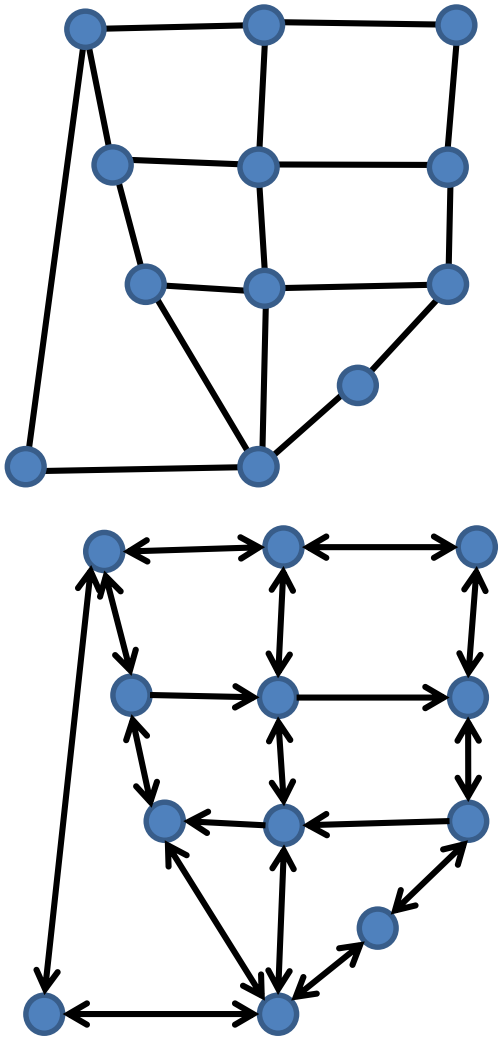
Inverse Distance Weighted

- The reciprocal for the weight calculation can vary



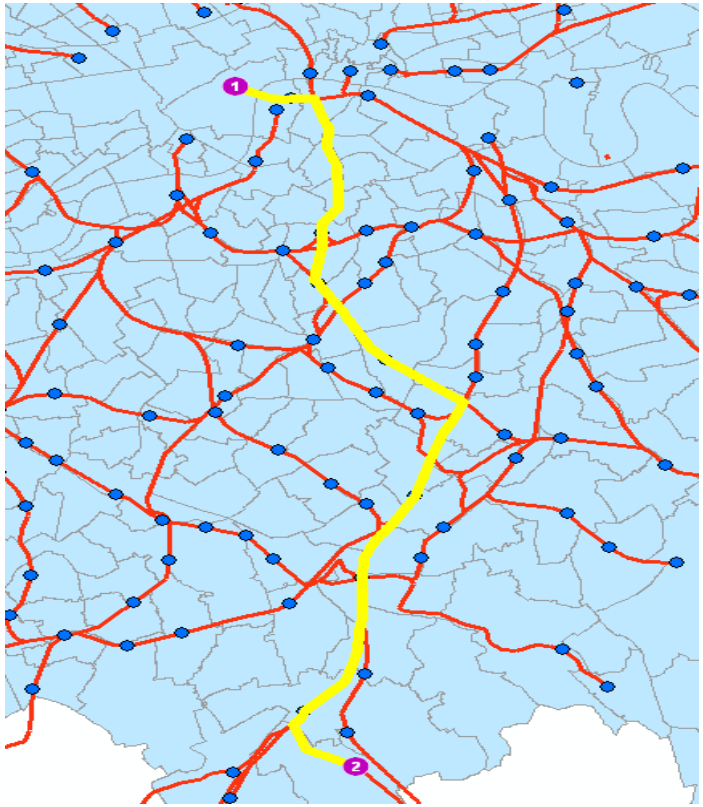
Networks

- What is a network (or graph)
  - A series of nodes and edges
  - Standard graph (undirected)
  - Directed graph:
    - Each edge has a direction set
  - Multigraph:
    - Multiple edges between the same node pair
    - E.g. one for each lane on a motorway



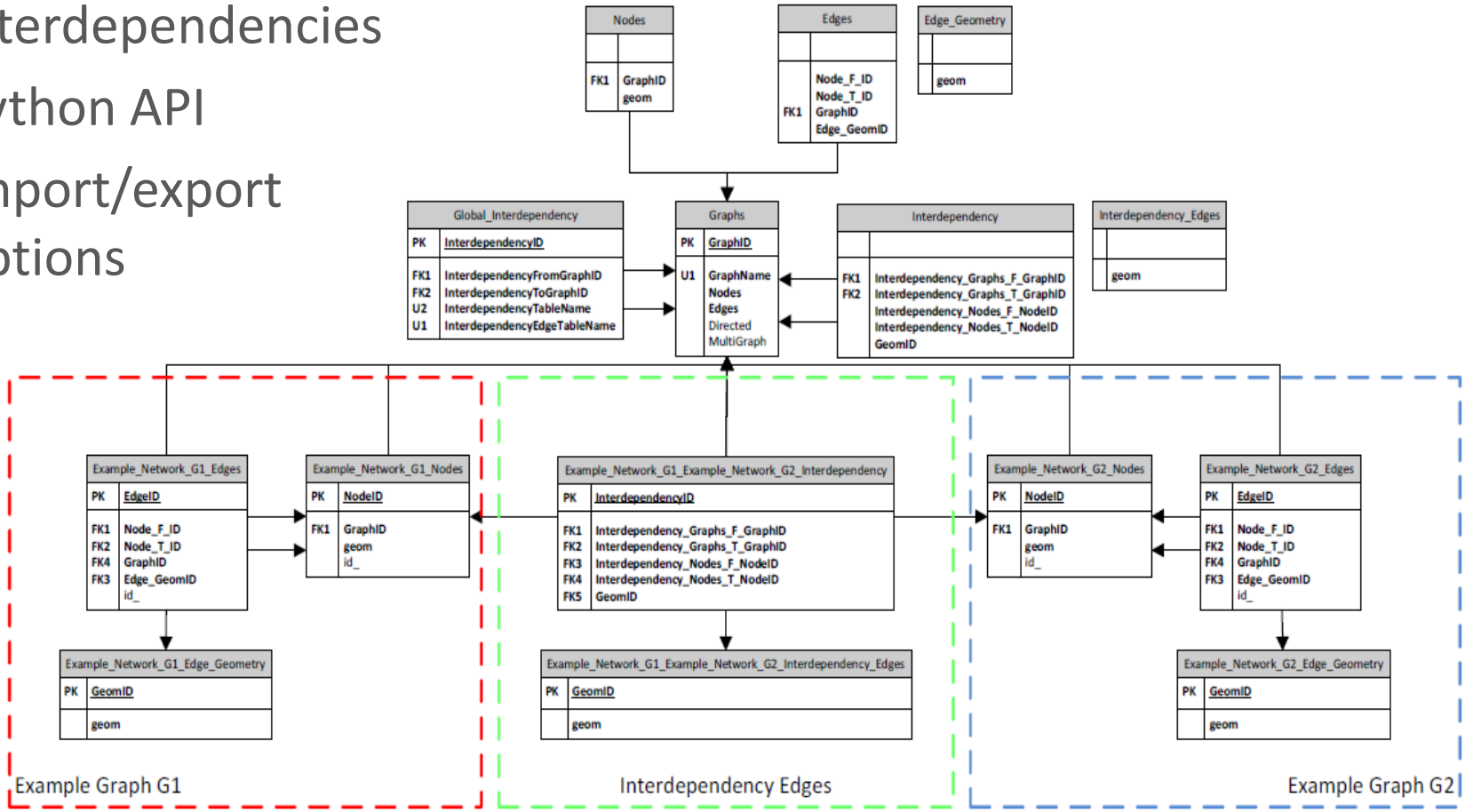
Networks

- Can run routing in ArcGIS and QGIS
- But…
- More options using programming solutions
- E.g. python – NetworkX, i-graph…
- E.g. postgreSQL – pgrouting



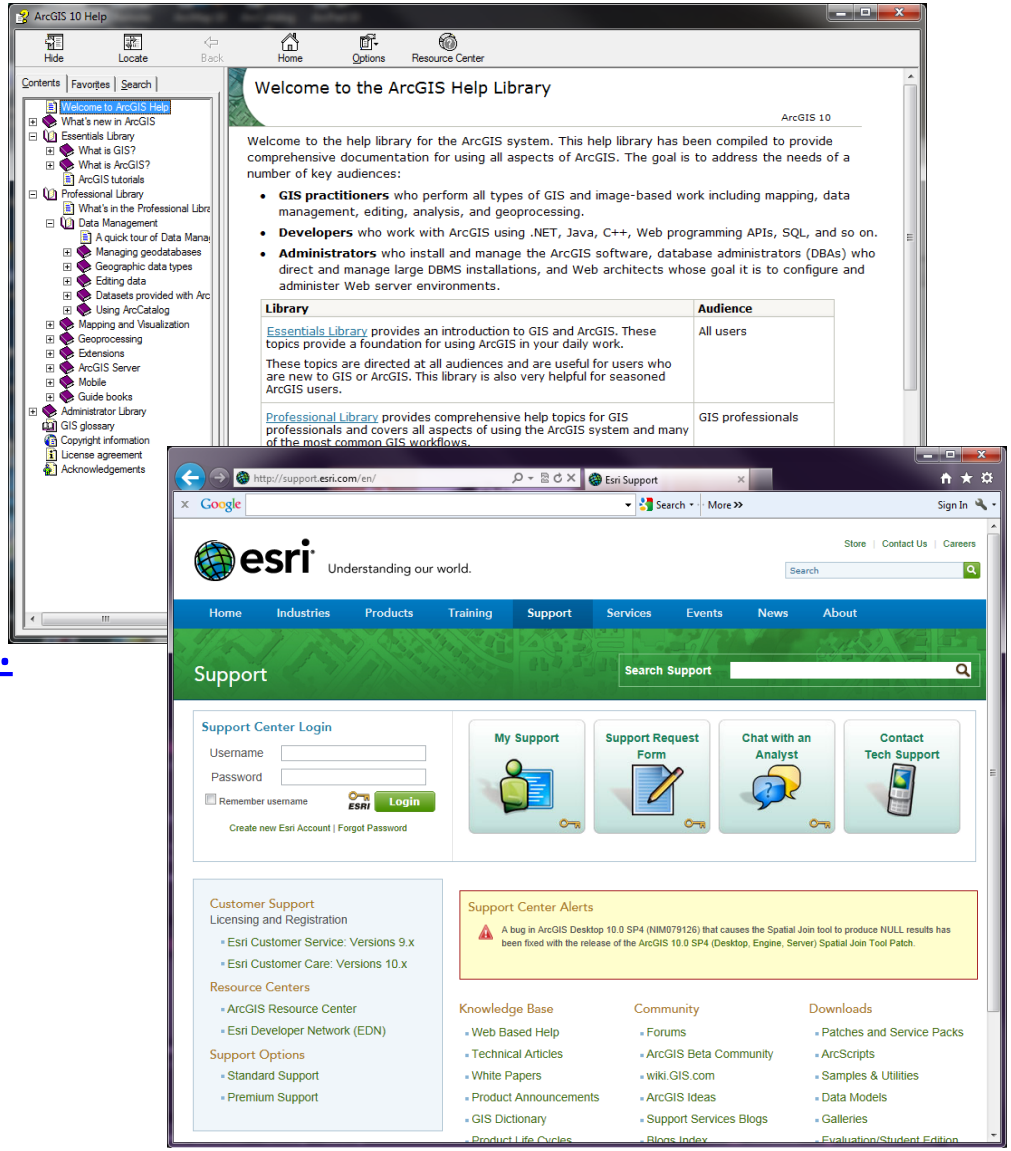
Networks - handling

- Database schema
  - Stores networks, dependencies and interdependencies
  - Python API
  - Import/export options



Help

- Plenty of help on the desktop software
- ArcMap help guides
  - www.
- QGIS help online
  - Eg:http://docs.qgis.org/2.2/en/docs/user_manual/
  - Forums etc.



ITRC mistral
multi-scale infrastructure systems analytics

Discussion (15mins)

- Converting between geographies
  - How to approach this
  - Can all cases be identified
- Network generation
  - Standardized methods
  - Standardized storage methods

EPSRC
Engineering and Physical Sciences Research Council

Extra slides

Extra slides on a few topics follow

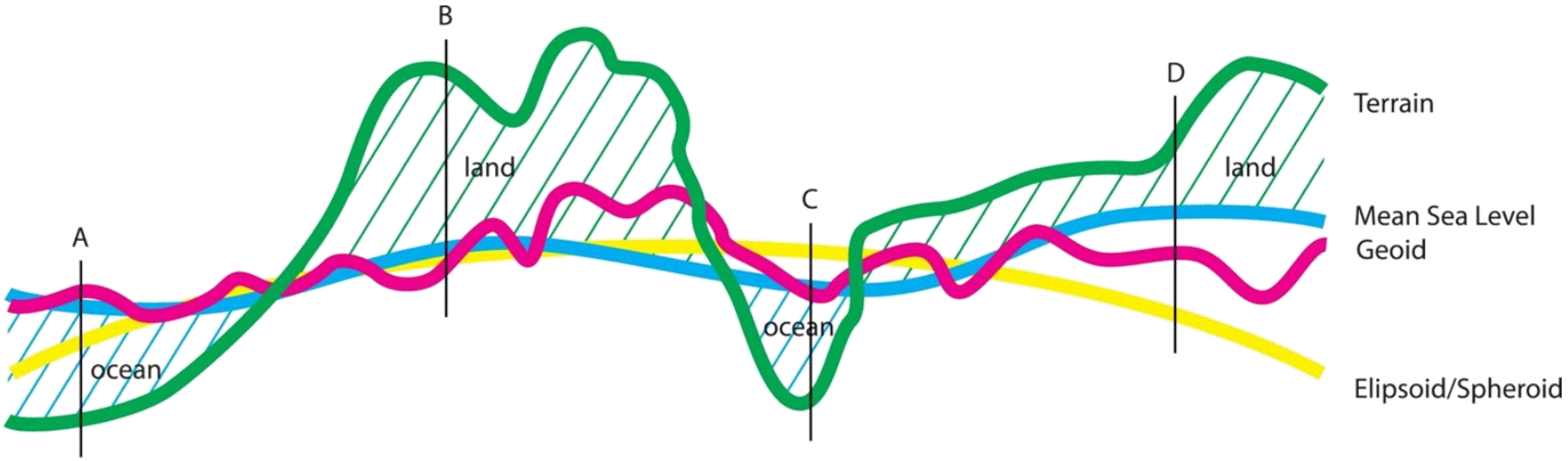
ITRC mistral
multi-scale infrastructure systems analytics

Questions I haven't answered

- Version control for spatial data
  - Discussed in discussion one
- Mapping objects to a point on a network
- Compute multi-modal commuting times between two points
- Division of areas in areas of influence
- Raster processes

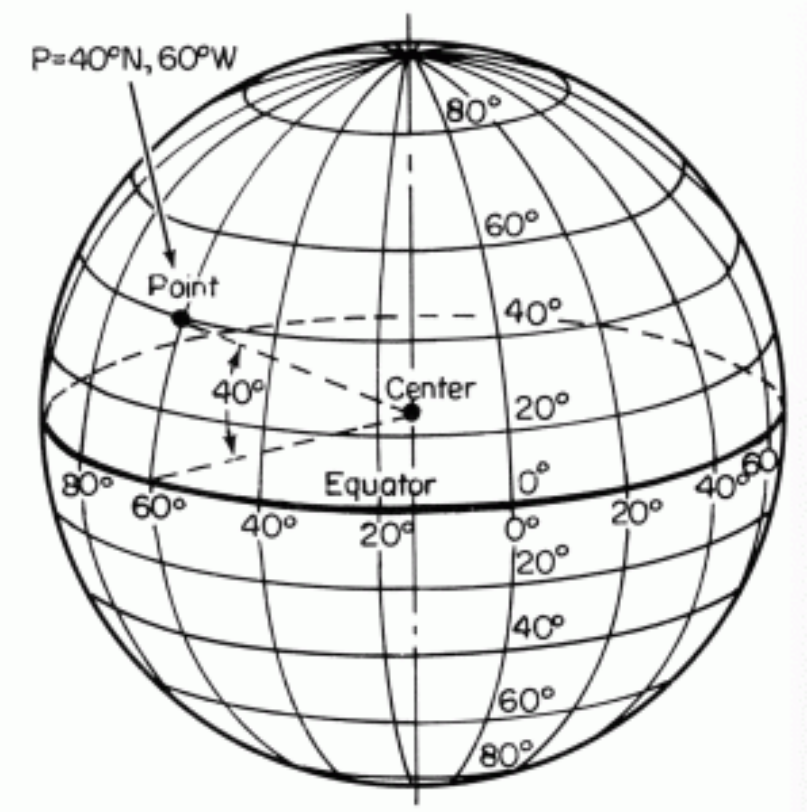
Datum's

- How do you model the Earth's surface?



Coordinate Sys. - Geographic

- Based on a model of the surface
  - the ellipsoid
- Latitude & longitude
  - Angular measurements
- Global
- Can include height



Coordinate Sys. - Projected

- Based on 2 dimensional projection of the surface
  - Will always be distortions
  - Global or local
- Eastings & Northings
- E.g. Mercator and Transverse Mercator (UTM)