

GIS 310 Second Semester Test

6 May 2009

Question 1

- Discuss digitisers as a method of capturing data for the use in a GIS. Also explain how digitising is done, heads-up digitising as an alternative method, problem/errors that may arise and the advantages and disadvantages of digitising.(20)

Manual locating devices: Digitisers

Digitisers are the most common device for extracting spatial information from maps and photographs



- The map, photo, or other document is placed on the flat surface of the digitising tablet

The digitising operation

The map is affixed to a digitising table

Three or more control points ("reference points", "tics", etc.) are digitised for each map sheet

- these will be easily identified points (intersections of major streets, major peaks, points on coastline)
- the coordinates of these points will be known in the coordinate system to be used in the final database, e.g. lat/long, State Plane Coordinates, military grid
- the control points are used by the system to calculate the necessary mathematical transformations to convert all coordinates to the final system

Digitising Steps

- The trace features to be digitised with pointing device (cursor), using either:-
 - point mode: click at positions where direction changes
 - stream mode: digitiser automatically records position at regular intervals or when cursor moved a fixed distance

Heads-UP Digitising

- Raster-scanned image on computer screen
- Operator follows lines on-screen in vector mode



How Much to Input- Vector Data

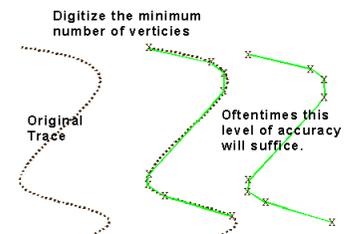
- Digitise vectors at a level needed for your objective or purpose.



Use the minimum number possible.

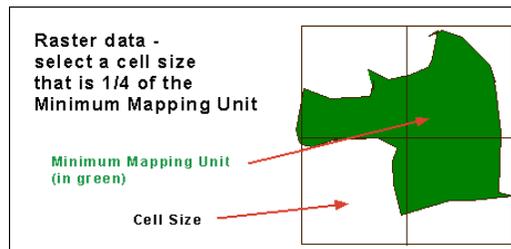
How Much to Input- Vector Data

- Here the X is a digitised point taken with "point mode" where instead one could have used "stream mode" and captured all the points represented by the dots, at a high cost in effort and storage without adding to the accuracy.



How Much to Input- Raster Data

- Raster - select a cell size 1/4 the size of the Minimum Mapping Unit



Problems with digitising maps

Arise since most maps were not drafted for the purpose of digitising

- paper maps are unstable: each time the map is removed from the digitising table, the reference points must be re-entered when the map is affixed to the table again
- if the map has stretched or shrunk in the interim, the newly digitised points will be slightly off in their location when compared to previously digitised points
- errors occur on these maps, and these errors are entered into the GIS database as well
- the level of error in the GIS database is directly related to the error level of the source maps

Problems with digitising maps

Maps are meant to display information, and do not always accurately record locational information

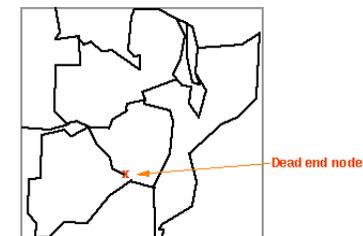
- For example, when a railroad, stream and road all go through a narrow mountain pass, the pass may actually be depicted wider than its actual size to allow for the three symbols to be drafted in the pass

Discrepancies across map sheet boundaries can cause discrepancies in the total GIS database

- e.g. roads or streams that do not meet exactly when two map sheets are placed next to each other

Problems with digitising maps

User error causes overshoots, undershoots (gaps) and spikes at intersection of lines

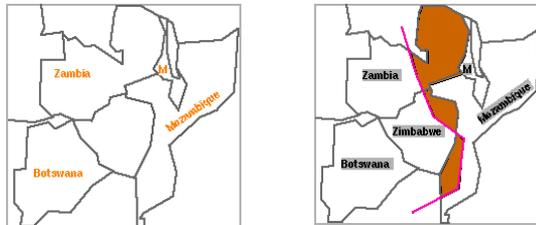


User fatigue and boredom

Editing errors from digitising

Some errors can be corrected automatically

Error rates depend on the complexity of the map, are high for small scale, complex maps



Advantages and disadvantages of manual digitising

- Advantages:
 1. can be performed on inexpensive equipment,
 2. requires little training,
 3. does not need particularly high map quality.

- Disadvantages
 1. tedious,
 2. time consuming.

Question 2

- Define and explain what the lineage of a data set is. Also refer to the benefits of lineage.(10)

Managing GIS error...

- Data quality information is frequently used to help construct a *data lineage*
- Lineage is a record of data history that presents essential information about the development of data from their source to their present format
 - Lineage information should provide the data user with details of their *source, method of data capture, data model, stages of transformation, editing and manipulation, known errors*

Lineage

- Lineage is a record of data history which is presented as a descent or ancestry

FAMILY



TREE

- Lineage lets the user know where problems with the data are likely to occur
- Lineage information can be recorded manually or automatically

Lineage

1. Manual Method

- User physically performs the actions that create the lineage
- Problems..???

 - A. Retrieval time
 - B. Complexity of linkages between different data files
 - C. Time-consuming process

- Requires dedicated user to maintain log book which would then need to be entered into a database

Lineage

2. Automatic Method

- Removes the need for user intervention to produce a lineage
- Often done automatically by a GIS software package e.g. IDRISI
- Lanter (1990), describes 3 computer methods of recording change:-
 1. **History files:** A history file maintains a record of commands
 2. **Audit files:** An audit file maintains a record of commands and input/output from the user
 3. **Log files:** Log files record commands, I/o and computer system details

Lineage

- Basic lineage requirements:-

1. Describe the source of data

- Information including the name, date, method of production, date of last modification, producer, reference, map scale, and projection

2. Transformation documentation

- Includes details of the data files used and the products generated

Lineage

3. Input/output specifications

- Includes descriptions of file formats, transfer formats, input/output procedures, media specifications etc

4. Application-dependent information

- Record information about the purpose for which a particular data set was generated
 - Helps the user determine for which applications the data are useful

Lineage

• Benefits of lineage:-

A. Error detection

- Lineage helps recreate analysis processes in which previous data sets containing errors have been used

B. Management accountability

- Lineage provides information from which accounting can be undertaken

Lineage

C. External accountability

- Lineage records the work of each GIS user
- Allows the assessment of the validity of the work undertaken

D. Quality reporting

- Lineage is required for GIS data quality reports
- Include information on data history, positional accuracy, attribute accuracy and completeness

Question 3

Explain what an 'error' in GIS data is. (2)

Explain the difference between accuracy and precision (2)

Name and discuss errors that may arise in GIS data. (16)

(20)

Definitions: Error

- Error is a general term that refers to both quantitative and qualitative information.
- Error can mean that the wrong value of a quantity is entered into a database
- For example longitude 35 E may be incorrectly inputted as -35 (this would be 35 west in most systems).
- Also the name or attribute of a specific location may be incorrectly entered into a database. Other errors result from the accuracy and precision of quantitative measurements.

What is an 'error' ??

- A difference between digital data and reality?
- A difference between digital data and reality at the time the data were collected?
- A difference between the data quality expectations as defined by the data collection method and the actual data quality?
- Failure of the user of the GIS data to understand the quality limitations of the data?
- ... flaws in the data sets

Definitions: Precision

- Precision is a measure of how exact a measurement is.
- Example:
 - Width of river = nearest metre
 - Width of stream = nearest centimetre
- True precision should also reflect how repeatable a given measurement is.

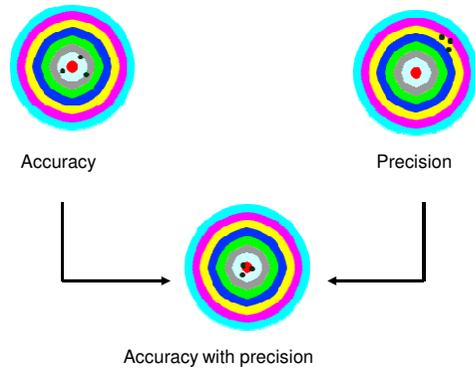
Definitions: Accuracy

- Accuracy is a measure of how closely the measurements approach the true value
- What is the true value???

Definitions: Error propagation

- In a GIS analysis such as map overlay errors in each data layer may propagate (be carried through) through the analysis
- Errors accumulate through multiplication

Definitions



Error in GIS data

1. **Age of data** - *The data may be old or out of date.*
2. **Partial or incomplete coverage**
 - *There may be incomplete coverage in a geological map, or very little known about a part of a mapped area.*
3. **Map scale**
 - *At what scale were the data originally mapped*
4. **Density of observations**
 - *The density of observations determine the level of detail in a map*

Error in GIS

5. **Relevance**
 - *The data that is really required isn't available so it is necessary to collect a proxy of the data*
6. **Format**
 - *physical storage media (disk, tape, ftp), data structure (raster or vector), data scale projection and classification*
7. **Accessibility**
 - *The best data may not be easily accessible, so a less detailed data set is used.*
8. **Cost**
 - *The cost of the most accurate data may be too high*

Question 4:

In both a raster and vector GIS there exist various methods that can be used for the analysis of data. Discuss the following analysis methods in terms of definitions, different types (if applicable) and usage. Choose your own application field and explain how the analysis methods can be applied to answer complex spatial questions

- Measurement of distances and areas (5)
- Attribute queries (15)
- Spatial Queries (10)
- Buffering and neighbourhood functions (10)
- Overlays (10)

(50)

Calculating distance

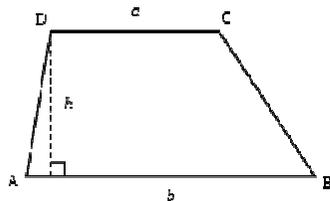
- Common application in GIS environments
- How does GIS go about calculating distance ??
- Vector vs. Raster ??
- Important to remember that all measurements in GIS will be approximations. Why ??

Vector GIS

- Uses co-ordinate values of start and end points to work out the length of a line
- Vector perimeter = line segment + line segment
- Fully topological vector GIS will calculate areas and perimeters of polygons at the time that topology is created
- Advantage = perimeter and area are stored in a database and can be used and queried
- Example = population density of electoral wards

Vector GIS - Area

- Most frequent method to find area is by constructing a set of trapezoids
- Works out area of each trapezoid – add together



A figure with two parallel sides is referred to as **trapezoid** in North America, and as a **trapezium** in Great Britain.

Raster GIS

- Berry (1993): Alternative to the Pythagorean approach – proximity
- Method: . Concentric equidistant zones are established around the location of interest

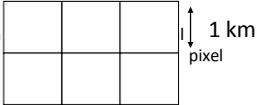
	1.00	2.00
1.00	1.414	2.414
2.00	2.414	2.828

 Origin cell

- The distance to a cell in the next 'ring' of cells is calculated based on information from previous ring
- Either orthogonal or diagonal distance

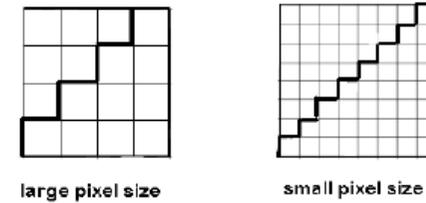
Raster GIS – perimeter and area

- Perimeter calculation in raster GIS ??
- Trace around polygon and count up the pixel sides
- Multiply: No. of pixel sides * Dimensions of pixel

• AREA:
Multiply: No. of pixels in 

Raster GIS – problems

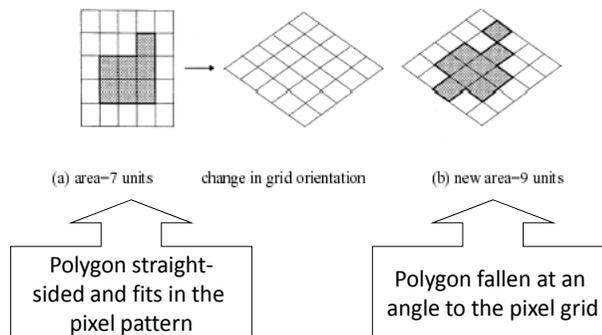
1. One factor which influences distance, perimeter and area in raster GIS is the **resolution** of the data



- Although the length is the same under both pixel sizes the precision of the line in the second case is better

Raster GIS – problems

2. Another problem with length, perimeter and area calculations result from the orientation of the raster grid with respect to the polygon of interest



Overview

- Selection (according to attribute criteria)
 - Set Algebra
 - Boolean Algebra
- Classification
- Dissolve

1. SELECTIONS

- Identifies features that meets one to several conditions
- Attributes are checked against criteria
- Answer usually shown in a selection
- Can be written to a new layer or used for further manipulation

On Screen queries

- Simplest form of selections is **on-screen query**
- Layer is displayed and feature is selected by the operator
- Operator uses a pointing device to identify the feature.
- Used for once of queries or update/editing of attribute data.

Set Algebra

- Selection conditions often formalised using **set algebra**:

- =
- <>
- < } Cannot be applied on nominal data
- > }

- Can be applied alone or in combination to select features from a set.

Boolean Algebra

- **Boolean Algebra** uses the conditions:
 - OR
 - AND
 - NOT
- Used to combine set algebra conditions to create compound spatial selections
- Are evaluated by assigning an outcome (true or false) to each condition.
- Order of application of boolean operators are important
- Parenthesis () are used to specify order of application

2. CLASSIFICATION

- Is very common and often generated in GIS analysis
- Often used in conjunction with selection
- Also known as reclassification or recoding
- Will categorise features based on a set of conditions
- Classification may add to or modify attribute data which may be used for further analysis
- Can also be used to group objects for display or map production – similar objects are identified as a group
- Display colours are assigned base on the attributes
- Classes may be assigned automatically or manually

EXAMPLES:

Classification

- **Binary Classification** – simplest form of classification
Place objects into two classes – 0 and 1, true or false etc.
Often used to store the results of complex selection operation

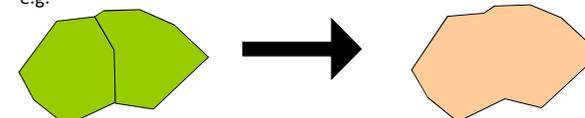
Manual definition may be tedious – rather do it automatically
Automatic classification uses some rule to assign classes
Assigning of boundaries to classes may be a problem.
Mathematical calculation or algorithm assign classes.
Groups may be split
Time can be saved but we may have to adjust the boundaries manually
The way classes are assigned will change the appearance of the map

Choosing a Classification Method

- Very often you will define your own classes rather than making use of the existing classification methods. The better you know your data set the better you can define the classes. If you are not sure which classes to use you must consult a specialist and find out which classes to use in a project. E.g. high traffic volumes in Johannesburg will differ from high traffic volumes in Delmas, high population density in KwaZulu Natal will differ from high population density in the Kalahari.
- To choose the correct classification method you need to know how the data values are distributed in a data set:
- If the data are evenly distributed or there are gaps between groups of values you will make use of natural breaks
- If your data is evenly distributed and you want to emphasise the differences between the features you will make use of equal interval or standard deviation.
- If your data is evenly distributed but you want to emphasise the relative differences you will make use of quantile

Dissolve

- **Dissolve** – combining features based on attribute data
- Adjacent polygons may have the same values for an attribute
- Are useful to remove unneeded information
- Very often follows on classification (depending on your application)
- Dissolve reduce volume size – increased processing speeds
- May lose important attribute data when doing dissolve function
- **ALWAYS KEEP A COPY OF THE ORIGINAL**
- e.g.



2 forest areas
Attribute data:
Soil types the same
Dissolve boundaries

Overview

- Spatial Data Queries:
 - Spatial vs. non-spatial queries
 - by Cursor
 - by Graphic
 - by Spatial Relationship
- Methods of spatial data enquiry
 - Containment
 - Intersect
 - Proximity
- Querying a vector GIS
- Combining queries - example
- Querying a raster GIS

Spatial Data

- When we talk about spatial data, we mean data that has a spatial component
- Both spatial and non-spatial (i.e., descriptive, or also alphanumeric, attribute) aspects are associated with this kind of data. For example, a river will have attribute information associated with it (such as its name and length) together with its geometric description (the spatial component)
- Often we talk about spatial data to mean its spatial component

Spatial Data Queries

- Spatial data query is the process of retrieving data from a map theme (or themes) by working with map features.
- Spatial data query is the geographic interface to the attribute database and is, therefore, useful for tasks that cannot be easily accomplished through attribute data query.

1. Selection by Cursor

- * Simplest method of spatial data query.
- * Select feature(s) using pointing device
 - individual selection using cursor
 - group selection using area

2. Feature Selection by Graphic

- Select features from the active theme that are contained within or are intersected by a graphic object.
 - line
 - circle
 - box
 - Polygon

E.g.

- ❖ selecting restaurants within a one-km radius of a hotel,
- ❖ selecting land parcels that intersect a proposed highway

3. Feature Selection by Spatial Relationship

- Selects map features based on their spatial relationships to other features.
 - a) Select features in same theme
 - Find the roadside rest areas within a radius of 50 km of a selected rest area.
 - b) Select features within another theme
 - Find rest areas within a certain county.

Methods of Spatial Data Query

– by Spatial Relationship

• Containment

- selects features that fall completely within features used for selection
- E.g. finding schools within a selected county

• Intersect

- selects features that intersect features used for selection
- E.g. selecting land parcels that intersect a proposed road

Methods of Spatial Data Query

• Proximity

- selects features that are within a specified distance of features used for selection
- E.g. finding national parks within 10 km of an highway, and finding pet shops within one km of selected streets
- If features to be selected and features used for selection share common boundaries in a theme and if the specified distance is 0, then proximity becomes *adjacency*

Querying a vector GIS

- Types of vector GIS queries according to Haining (1994)

1. Univariate

a. Aspatial

Describe the pH of the soil type

- numerically (e.g. mean, median, etc)
- graphically (normal distribution, skewed etc.)

b. Spatial

Map cases of soil where the pH is higher than the average

Querying a vector GIS

2. Multivariate

a. Aspatial

Describe the properties of the soil and rainfall

(e.g. what relationship exists between soil pH and rainfall)

b. Spatial

Where are the cases with greater than average pH values and high rainfall?

Querying a vector GIS

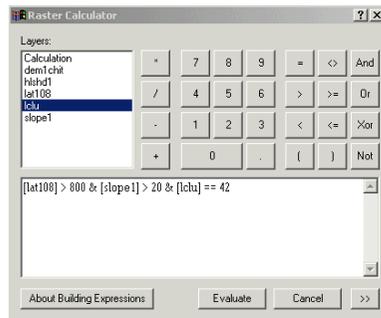
- The ease with which these queries can be implemented is a function of the relationship between the graphic data and the attribute data
- With most vector GIS a link between these two data types is set up as topology is created
- I.e. each graphical entity has unique identifier = references a database table containing the related attributes

Querying a raster GIS

- Queries are generally one of two options:-
 1. What is at ...?
 2. Where is ...?
- For raster the equivalent is:-
 1. What is the content of a given pixel?
 2. Which pixels have this value?

Map Query Examples

- Let's say our criteria are elevation >800, slope >20% and land use category= coniferous forest (42)



The role of error

- Map and attribute data errors are the data producer's responsibility, but the GIS user must understand error.
- Accuracy and precision of map and attribute data in a GIS affect all other operations, especially when maps are compared across scales.
- Error can propagate as queries and overlays continue through a project

Buffers

- A prime single map layer operation
- Involves the creation of new information
- Has a strong spatial element
- Is most common operation

What is a buffer?

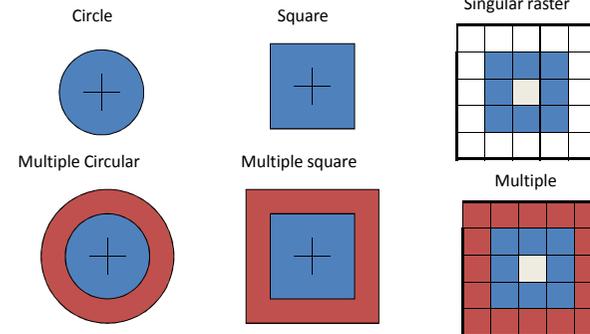
- Definition:** A zone of a specified distance around features in a coverage. Buffers can be set at constant or variable distance based on feature attributes. The resulting buffer zones form polygonal coverages. (GIS Lounge – glossary)

Buffering

- Possible in both raster & vector formats.
- Distance calculations are just performed differently.
- Other terms :
 - Spread
 - Search
 - Corridor
 - Zones

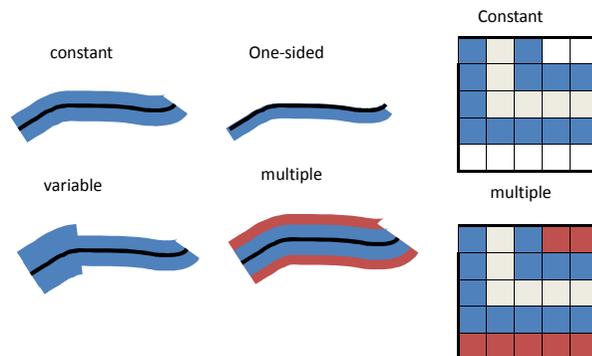
What types of buffers are there?

Buffers around points



What types of buffers are there?

Buffers around Lines



Examples

- Use buffers as part of analysis for siting a radio-active waste repository
- Repository has to be away from transport routes
 - All roads = (3km)
 - Highways = (1,5km)
 - Railway = (0,5km)

Rook's case vs. Queen's case (Raster GIS)

- Buffers allow a pixel or group of pixels to "grow" out in all directions.
- Buffers can be accomplished two ways:
 - the 'rook's case' and
 - 'queen's case'.

Buffering distances:

- Instigated on guess work
 - Should the boundary around a construction site be 1 m, 5 m or 20m ?
- Based on prior knowledge about the area within which the buffer is produced.
 - Rate at which water permeates through different soils
- Selected on the basis of a definable, measurable value.
 - Rate at which noise dissipates
 - Average distance people are willing to walk to a bus stop.
- Legal or otherwise mandated measures exist and must be adhered to.
 - Flood lines
 - Not building within certain distance of a road.

What should the buffer distance be?

- Are there any petrol stations **near** my site?
- The site **close to** the freeway.
- Is the site **far from** the river?

The Use of everyday language to describe distance.

Overlays

Overlay operations – powerful spatial analysis tools. Important driving force behind development in GIS

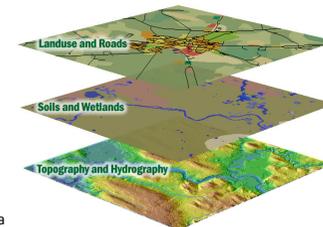
Combining **spatial and attribute** data from two or more spatial data layers

E.g.

Where are properties for sale within walking distance from schools?

Where are sinkholes on dolomite?

Which properties will be affected by the expansion of the airport?



Data layers must use same **coordinate system**

May be seen as the vertical stacking and merger of spatial data and attribute data

1. Raster Overlay

- Cell by cell combination of two or more data layers (can be any number of layers)
- Cell values are combined in some way - output always in a new layer.
- Typically applied to nominal or ordinal data
- Each value corresponds to specific category
- Combined using a mathematical calculation and map algebra
 - Map Calculator

Map Algebra operators

- Arithmetic operators:
 - +
 - -
 - *
 - /
- Logical operators
 - =
 - >
 - <
 - <> etc.
- Boolean operators
 - AND
 - OR
 - XOR
 - NOT

Overlay using weights

- One theme more significant than another
- Degree of importance expressed as weight
- Either recode accordingly
- Or use weighting factor in overlay process

Difficulties in raster overlay

- Resolution & scales of measurement
 - Data problems
 - Not technical

2. Vector Overlay

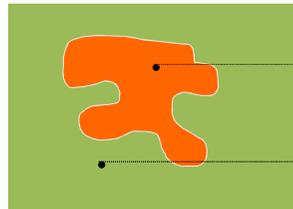
- Involves combining point, line and polygon geometry.
- Overlay can be any combination (Point-on-point and Point-on-Line intersections have rarely any results)
- Most common overlays are Polygon-on-polygon
- Time consuming and computational intensive process

Point in Polygon

- In simplest form:
 - Determines whether a given point lies inside or outside a polygon
- Can be extended:
 - Many points
 - Many polygons
 - Overlapping polygons

How point in polygon analysis works

- Many methods
- Simplest
 - ‘Plumbline’ or Jordan method
 - Extend horizontal (or vertical line)
 - Count crossings
 - ODD → inside
 - EVEN → outside



Line in polygon

- Used to determine if a line feature is in a polygon feature.
 - What administrative area a road is in?
 - What rock type a river flows over?
- Which polygon is this line *“contained in”*

Line in polygon

- Split line into segments where intersect with polygons.



Create new layer and attribute tables

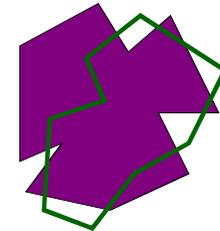
ID	Name
1	N1



ID	Name
11	N1
12	N1
13	N1

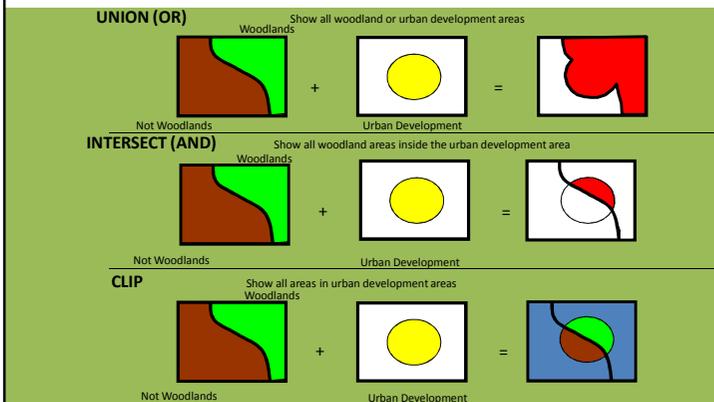
Polygon Overlay

- Complexity of polygon overlay was one of the greatest barriers to the development of vector GIS.
- Examples:
 - How much of a proposed clear-cut (felled plantation) lies in a riparian zone?
 - What proportion of SA Land is owned by the Department of Land Affairs?



Clip, Intersect and Union

- 3 common ways overlays are applied
- Answers and way in which attribute data are combined differs



A Problem in Vector Overlay

- Common features are represented in both output layers. Does not always fit exactly
- Is know as slivers
- Longer calculation time and bigger attribute tables.

