GIS DATA ANALYSIS & SPATIAL MODELLING

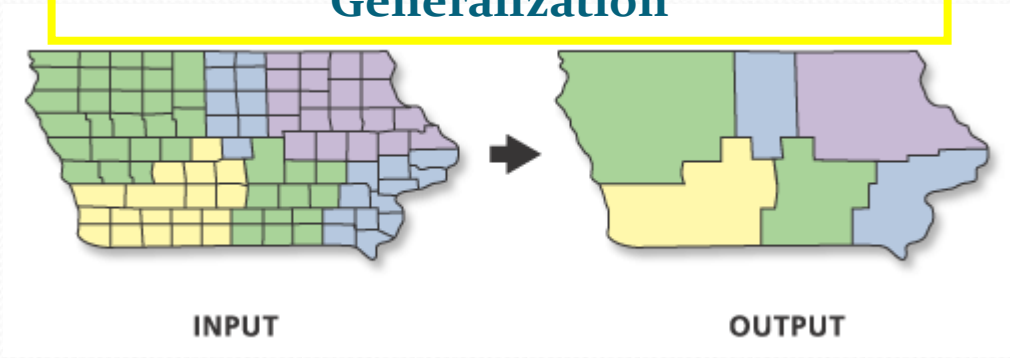
Dr. K.Palanivel
Assistant Professor
Centre for Remote Sensing
Bharathidasan University

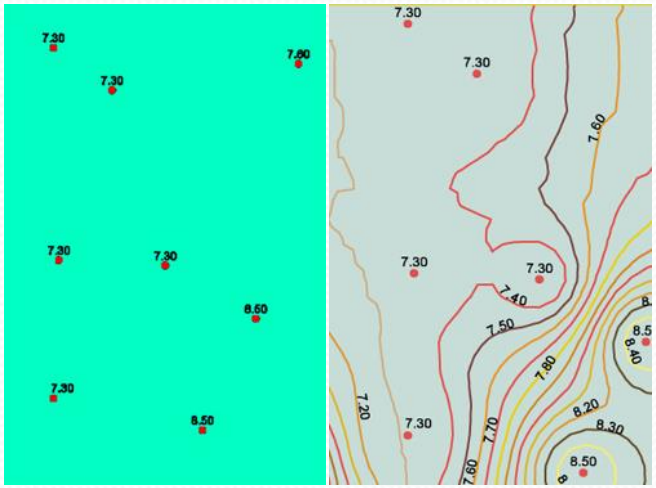
Unit-5

Data Analysis and Spatial Modeling: Simple data retrieval – Data retrieval through Boolean Logic – Map Overlaying and Cartographic Modeling (Two layers, Multiple layers, Binary, Index, Regression, and Process Models) – Overlay analysis, Capabilities (Point Operations, Regional Operations, Neighbourhood Operations) – Buffering – Cartographic Modeling using Natural Language Commands – Advantages and disadvantages of Carto modeling – Net work analysis. **12 Hrs.**

GIS DATA PREPARATION FOR ANALYSIS

Data Classification/Grouping , Regrouping,/Reclassification, Generalization





DISSOLVE

MERGE

UPDATE

CALCULATE RASTER ……

Classification Methods

Classifying the ranges of values of a single continuous variables

The different types of making class intervals are:

- **Exogenous class intervals** – according to threshold values that are relevant to, but not derived from the data set under study
  - Fail to resolve the variations in any specific area
  - If exogenous intervals are derived from qualitative class definitions, they may not be unique and may in fact overlap

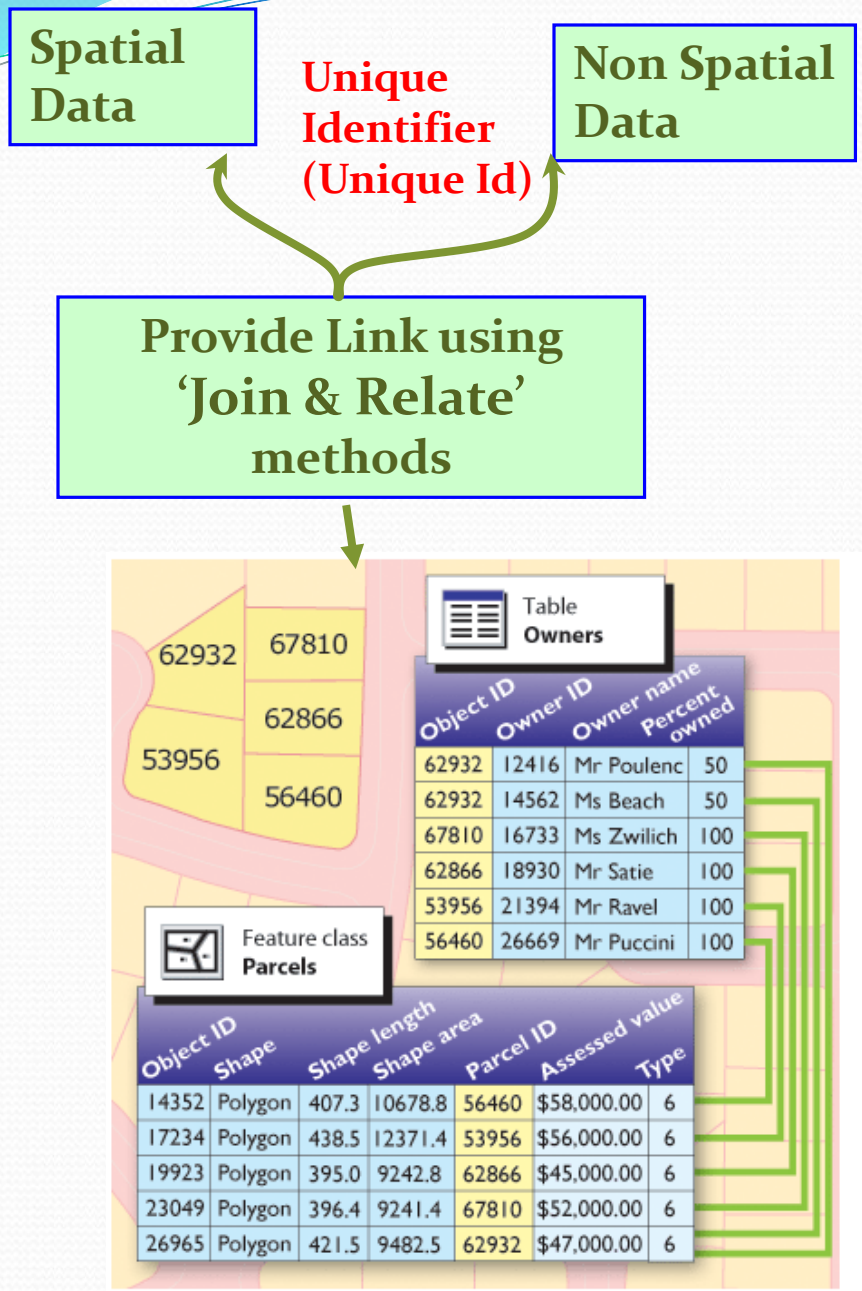
- **Arbitrary class intervals**
  - chosen without any clear aim in mind
  - very often the intervals are irregularly spaced and
  - Chosen without a proper examination of the data

- **Idiographic class intervals**
  - chosen with respect to specific aspects of the data set
  - They include methods that attempt to divide the data into multi-modal groupings
  - Using 'Natural Breaks', their location,
  - Objective comparison is difficult – not exactly followed from dataset to dataset

- **Serial class intervals**
  - have limits that are direct mathematical relation to one another.
  - intervals are chosen independent of the characteristics of the particular dataset,
  - then the results from different samples or datasets will be strictly comparable
    - **Normal percentiles** – class intervals subdivide a normal distribution with mean μ and standard deviation, with classes centred on the mean
    - **Equal arithmetic intervals** with no variation in class width, such as are normally used for elevation contours
    - **Equal intervals** on reciprocal, trignometric, geometric, or logarithmic scales – used when original data are not normally or rectangularly distributed

GIS ANALYSIS

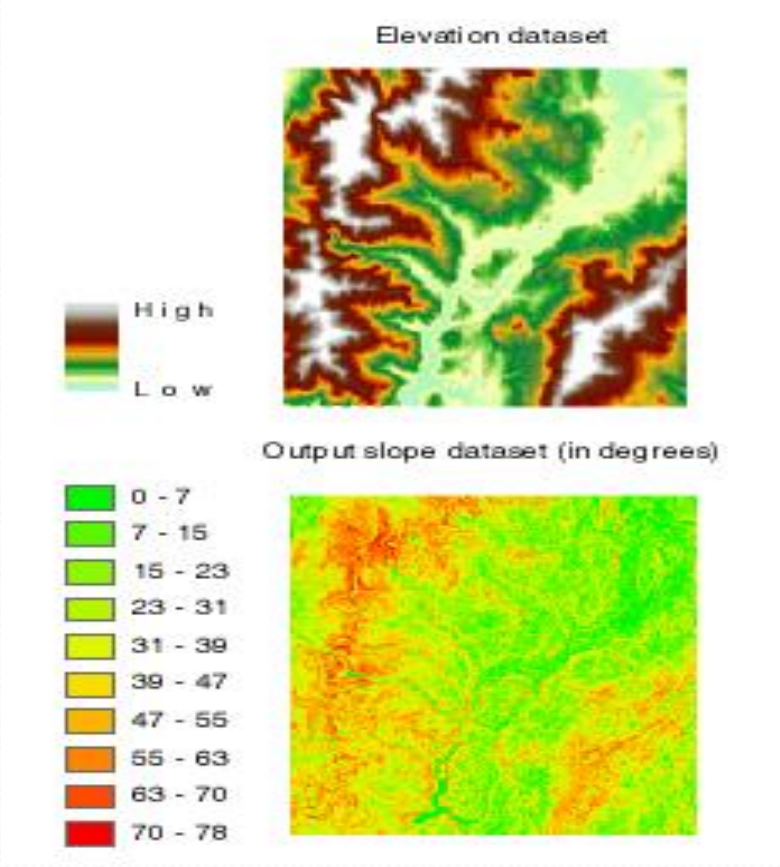
- GIS has the analysis capabilities ,that is not available with CAD
- Operates on Topology, spatial and/or non-spatial aspects combined
- The user can interactively work for getting the results
- Analysis capabilities range from
  - Retrieving subsets of information
  - up to univariate,
  - multivariate statistical and
  - spatial analysis using neighbourhood functions using interpolate methods
- Spatial modelling
- Simulation programs



- It is necessary to establish a link between database and output that provides answer in different forms – maps, tables, figures, etc. using "**Unique_Identifier**"
- That link is any function used to convert data from one or more input maps into an output / **Action Plan** map
- Define the database – nature of geodatabase – **vector** / **raster**, and whether it contains **points**, **lines**, **polygons** & **associated attributes**, having **same projection**

GIS Data Analysis ...contd...

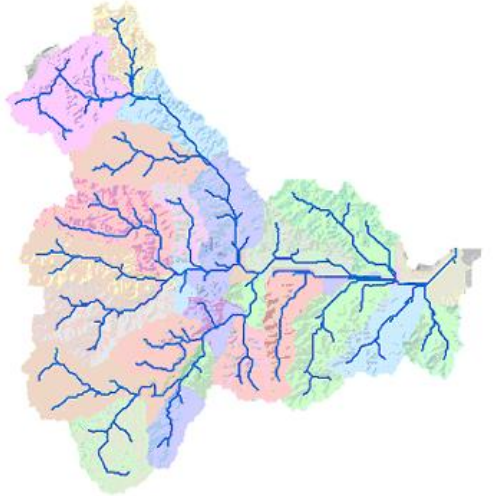
SLOPE ANALYSIS



Slope Categories – DEGREES , ASPECT, LENGTH

Intervisibility Analysis, etc.

DRAINAGE ANALYSIS



Delineation of Drainages,

Flow Direction, Flow Accumulation

Flow Length,

Demarcation of drainage basins,

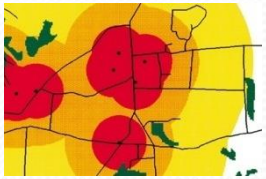
Runoff estimation,

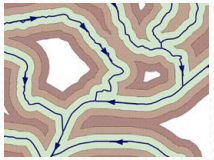
Quantum of soil erosion, etc.

Hydrological Modeling

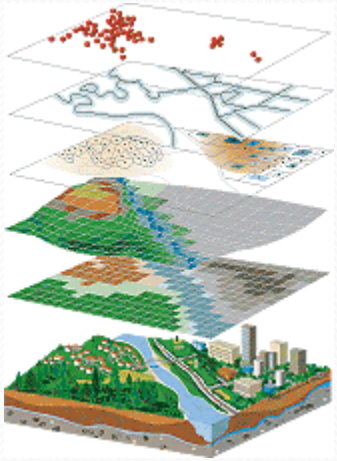
GIS Data Analysis & integration

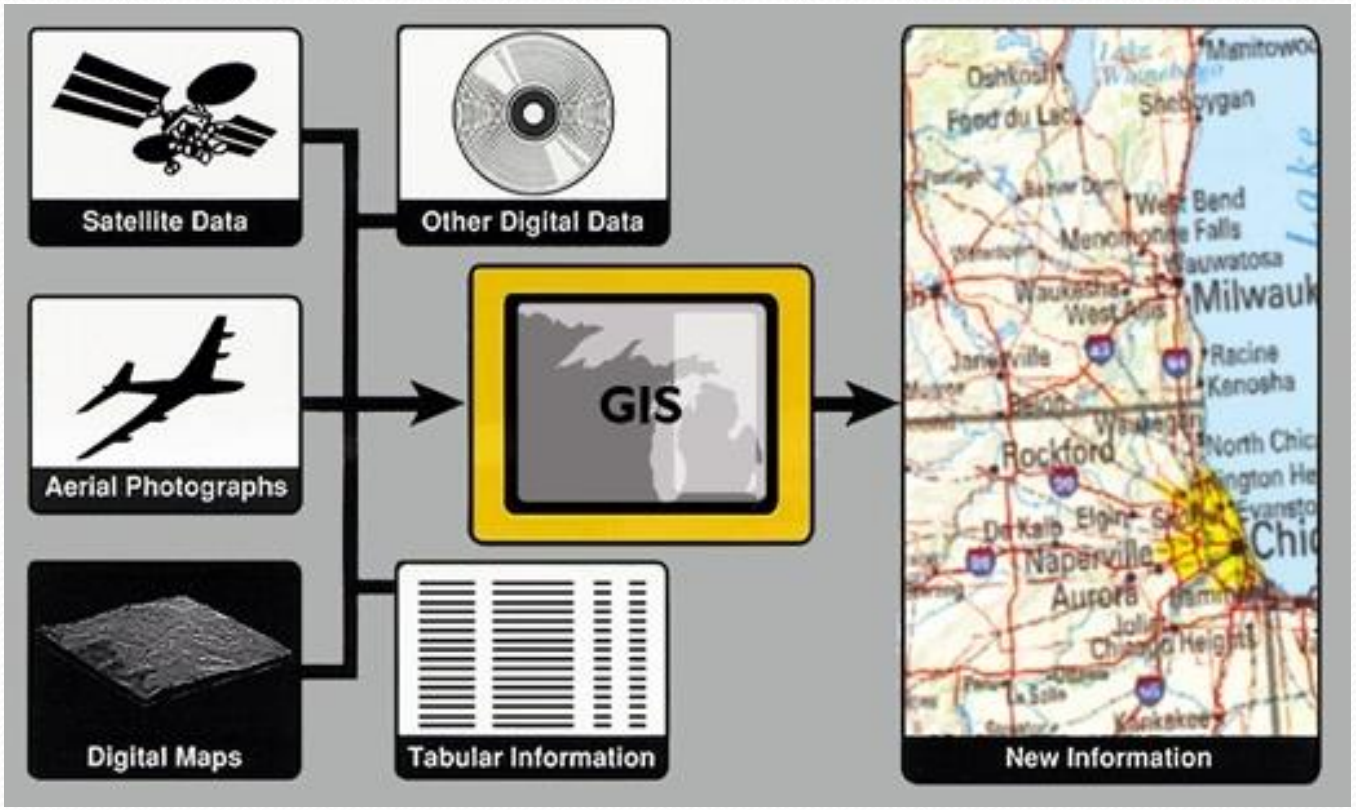
- Performing overlays
- Creating buffers
- Calculating statistics
- Merging datasets

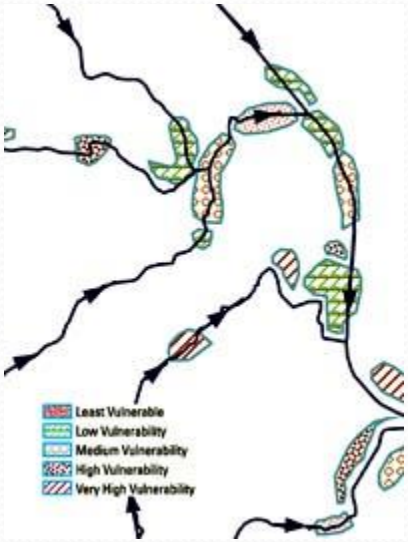




Data integration is the linking of information in different forms through a GIS







GIS Can do many operations / Analysis like add, subtract, multiply, divide, square, buffer, corridor, integrate multiple layers, etc.

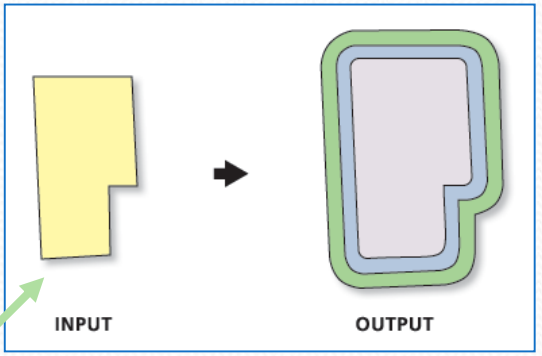
GO BACK

Raster Based Addition

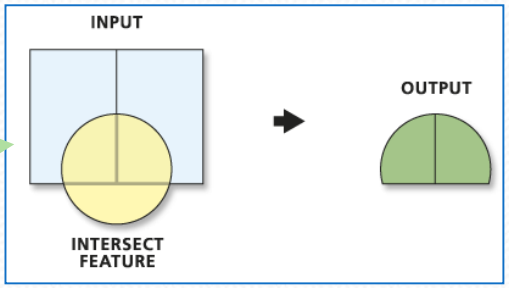
1	1	2	2	2	+	3	3	1	2	2	=	4	4	3	4	4
3	1	3	4	4		2	3	1	1	1		5	4	4	5	5
3	3	2	4	4		2	4	3	3	1		5	7	5	7	5
3	2	2	1	4		2	2	3	4	4		5	4	5	5	8
3	2	2	1	1		1	2	1	1	4		4	4	3	2	5

InRaster1 + InRaster2 = OutRaster

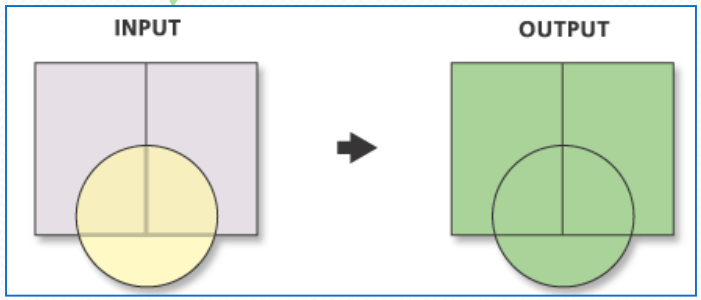
Vector Based Layer Integration



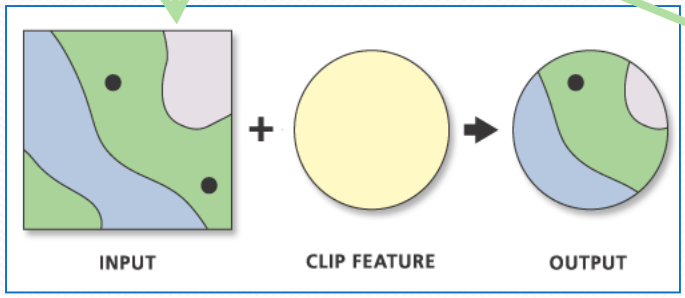
BUFFER



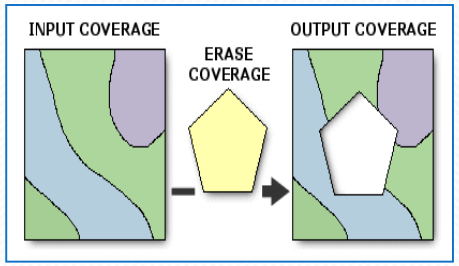
CLIP



UNION



INTERSECT



ERASE

Simple Data Retrieval

- **Boolean logic**
- **Reclassification and display**
- **Boolean operations on two or more maps**

Boolean logic

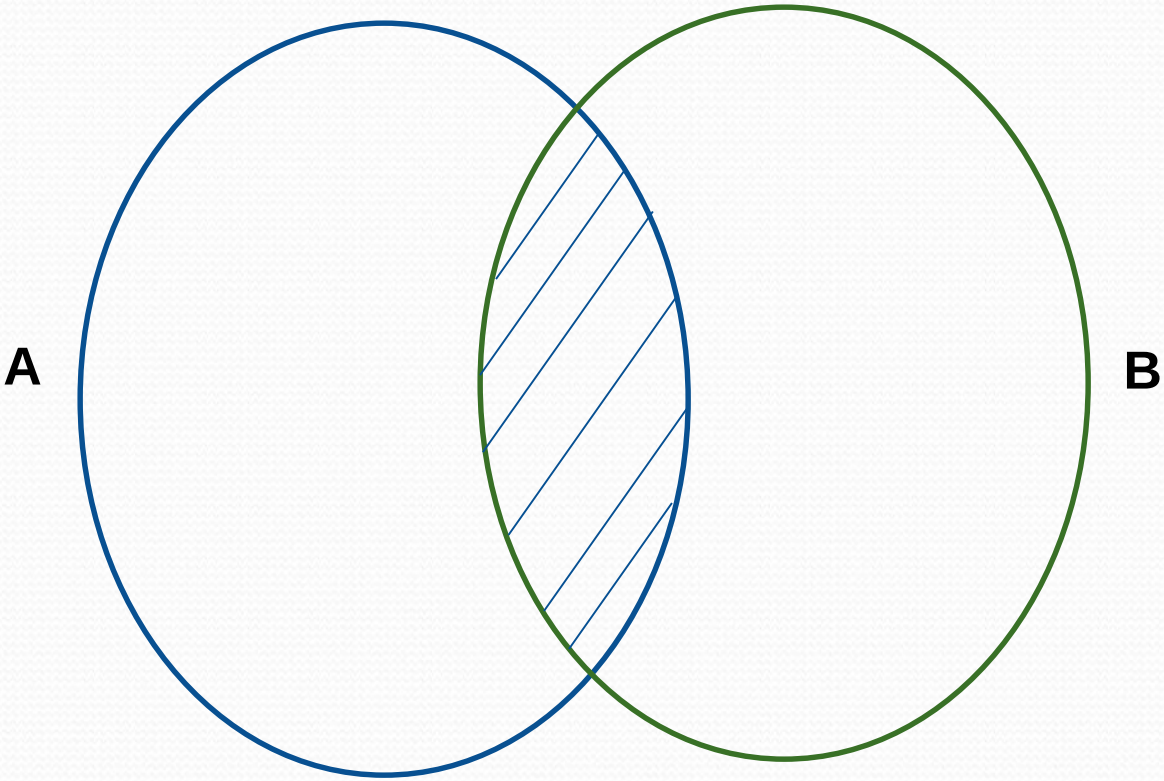
- Retrieve data by using the rules of Boolean logic to operate on the attributes and spatial properties
- Operators are : AND, OR, XOR, NOT
- The particular condition is 'true' or 'false'
- Visual portray of Boolean logic is in the form of 'Venn diagrams'
- A is the set of items having attribute 'A' and B is the set of items having attribute 'b', then the statements

A **AND** B, A **OR** B, A **XOR** B, A **NOT** B,

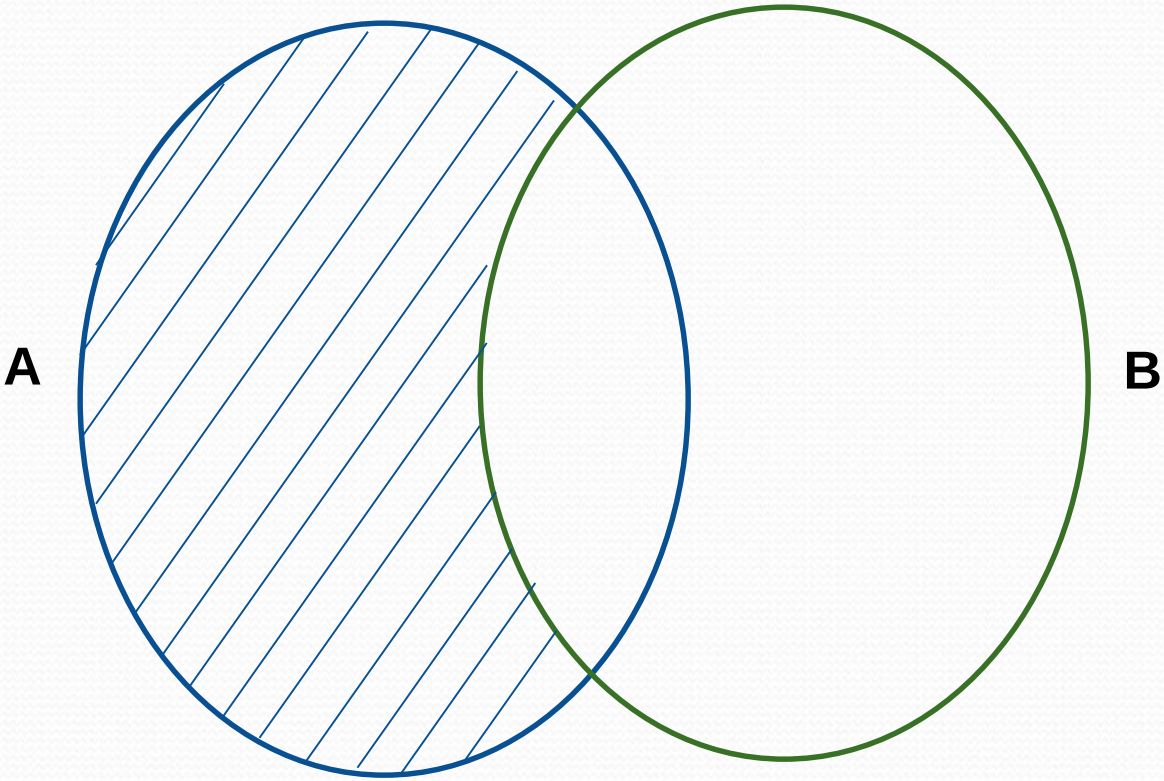
(A **AND** B) **OR** C, A **AND** (B **OR** C)

Will return items that have attribute combination covered by the **shaded portions** respectively

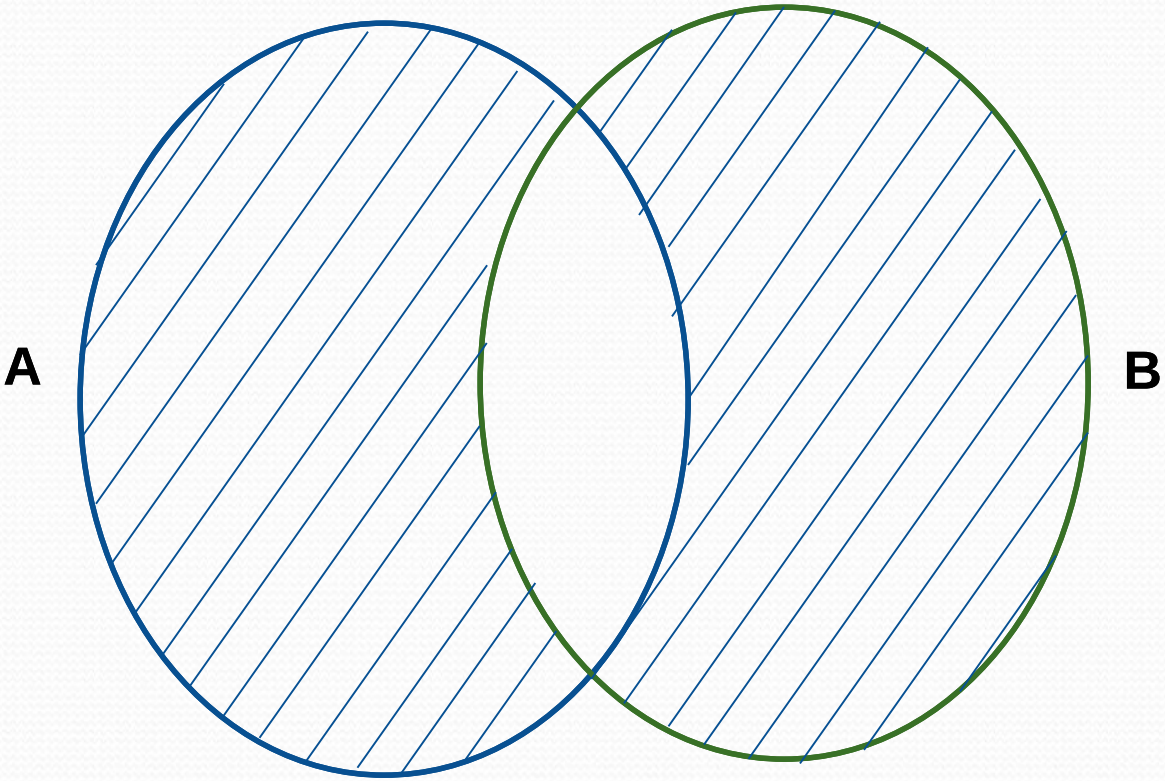
- On the basis of priority of AND with respect to OR
- Parentheses are usually used to represent the order of evaluation
- More complicated searches involve the shapes of areas, the properties of the boundaries of areas or the neighbouring areas.
  - All areas of woodland bordering urban areas
  - All open areas that have a particular shape and area bounded by hedgerow (a row of bushes)



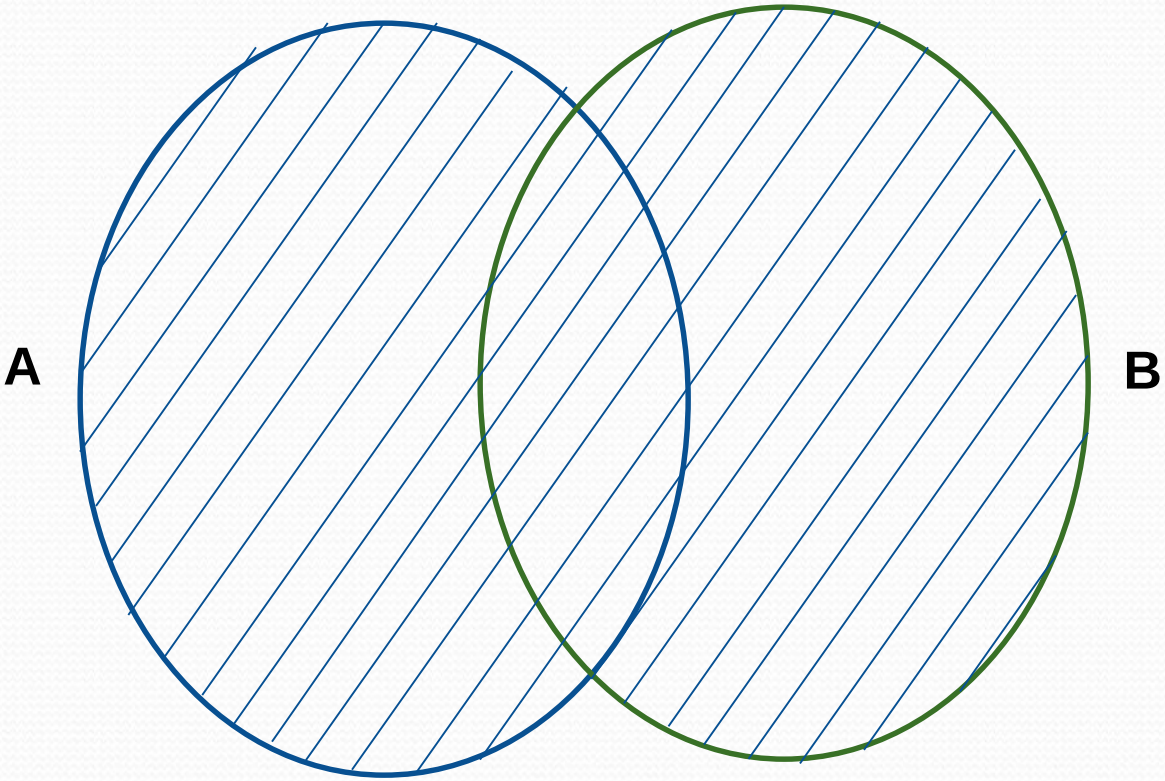
A AND B



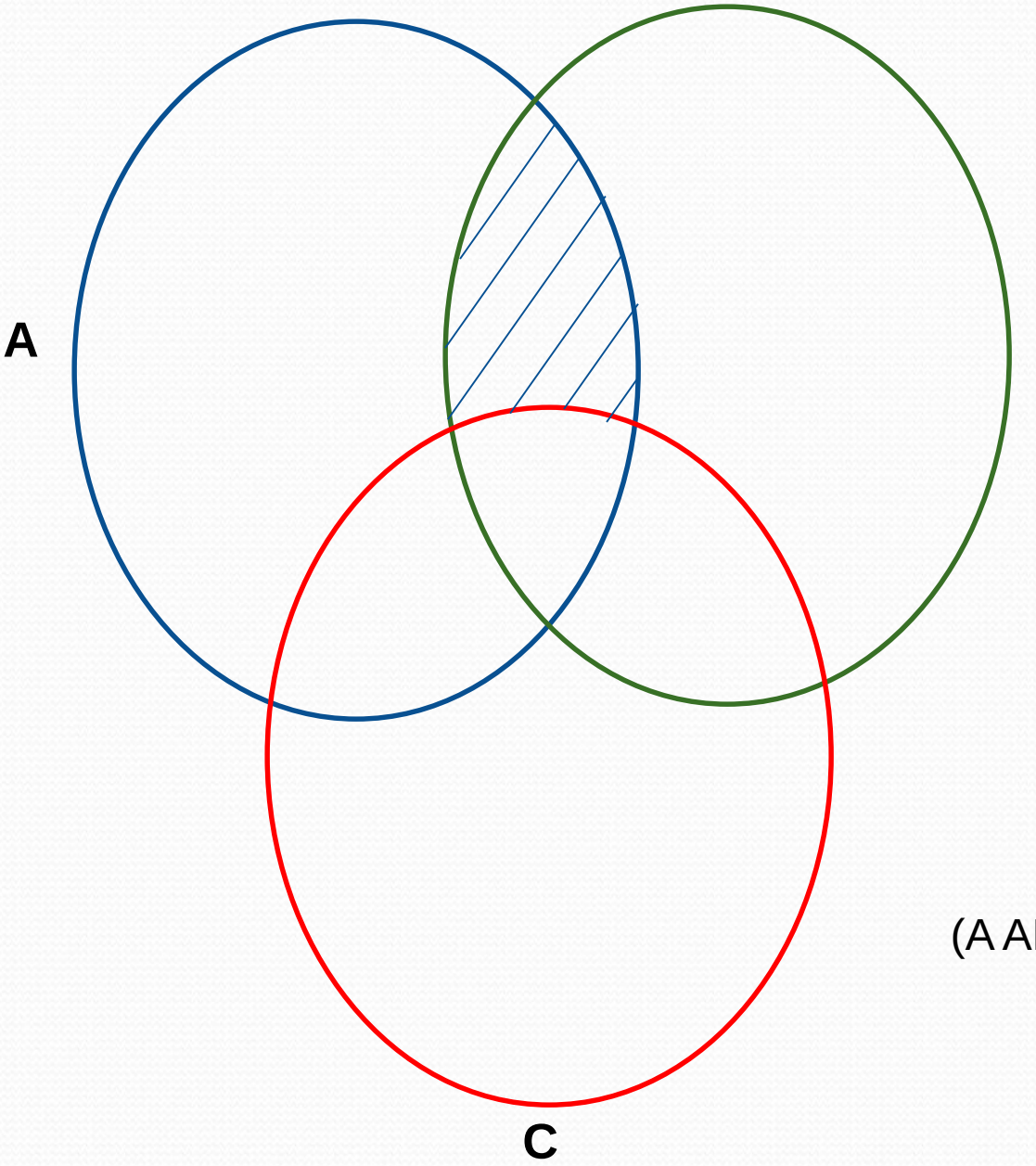
A NOT B



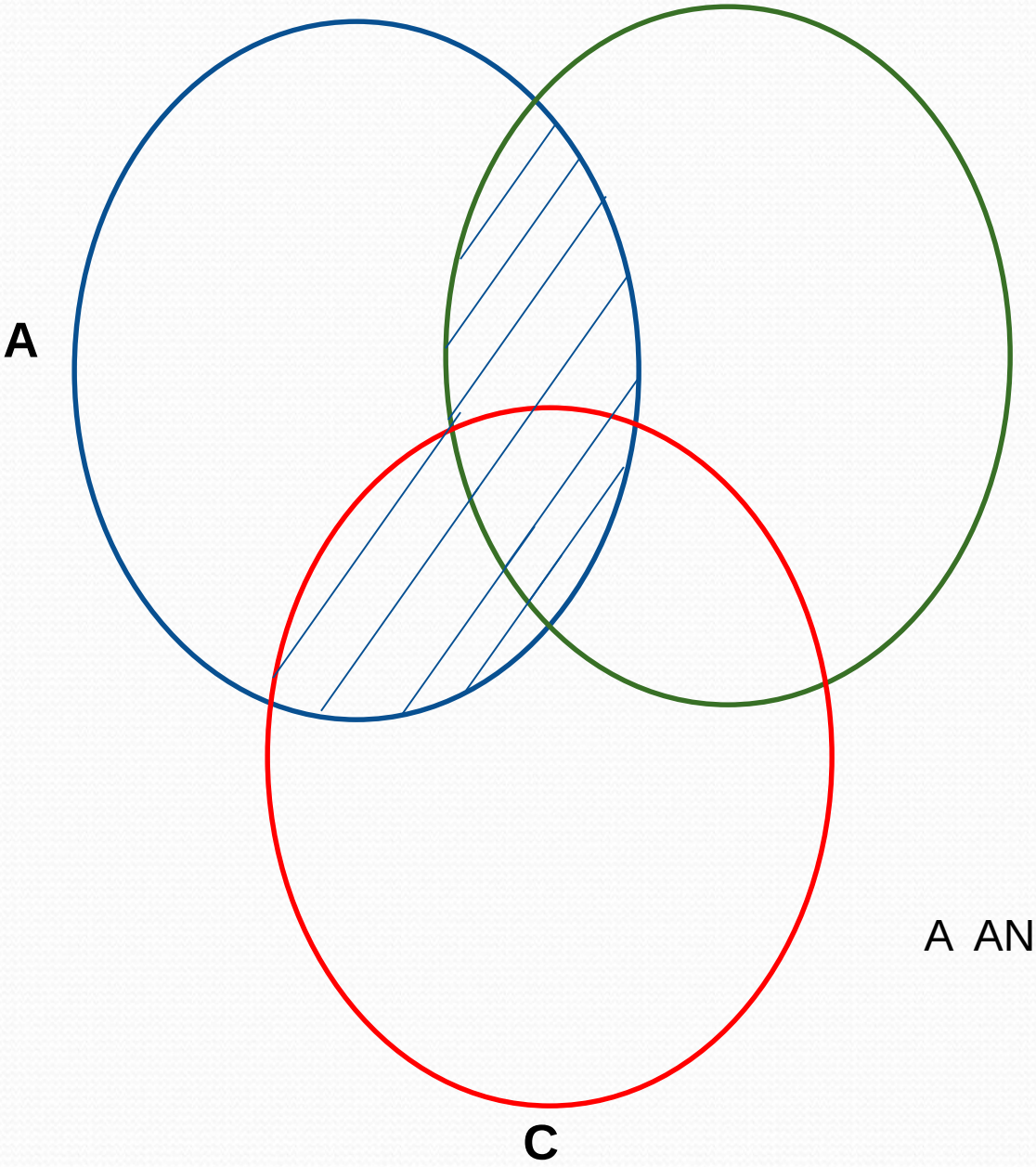
A XOR B



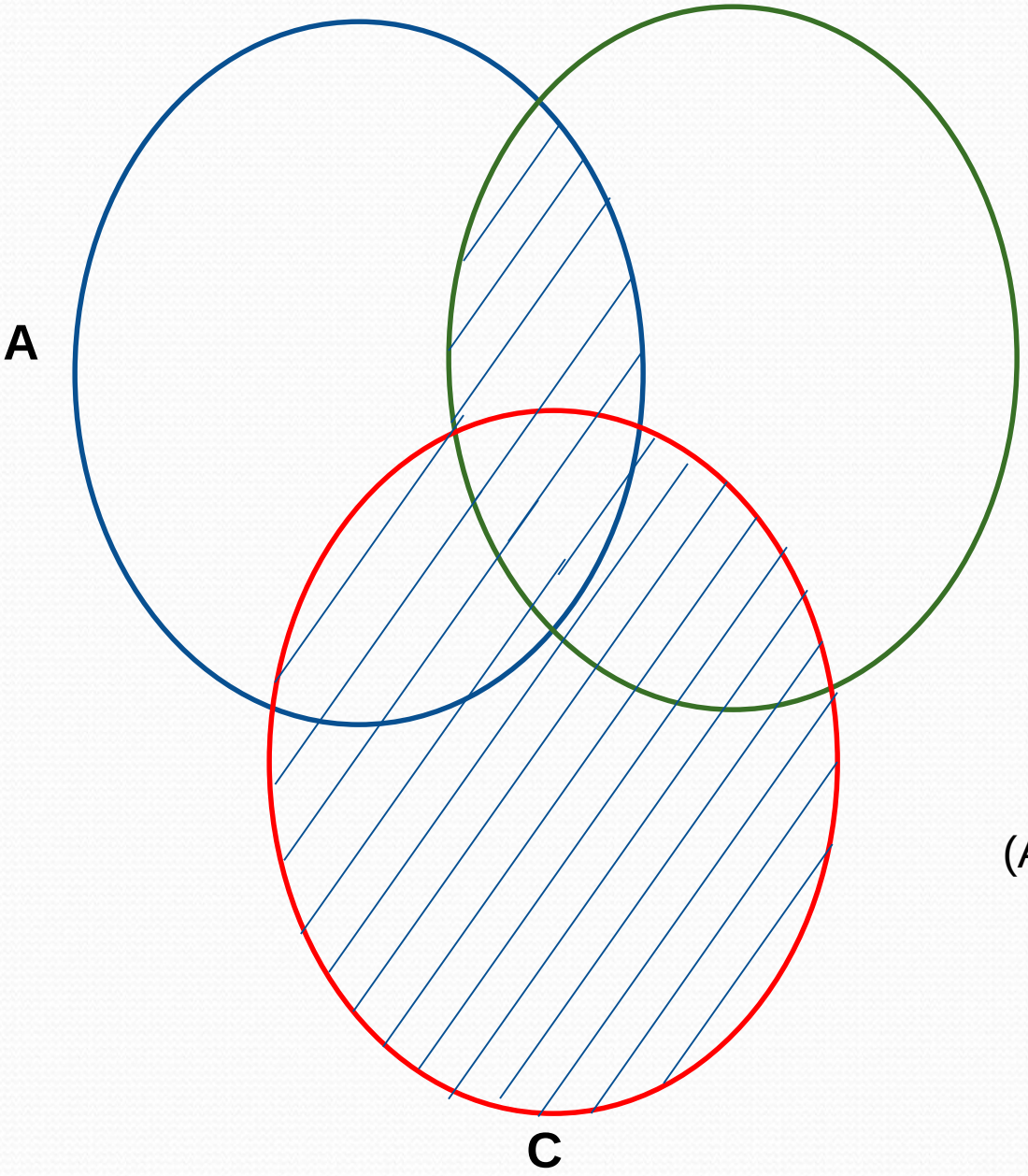
A OR B



(A AND B) NOT C



A AND (B OR C)



(A AND B) OR C

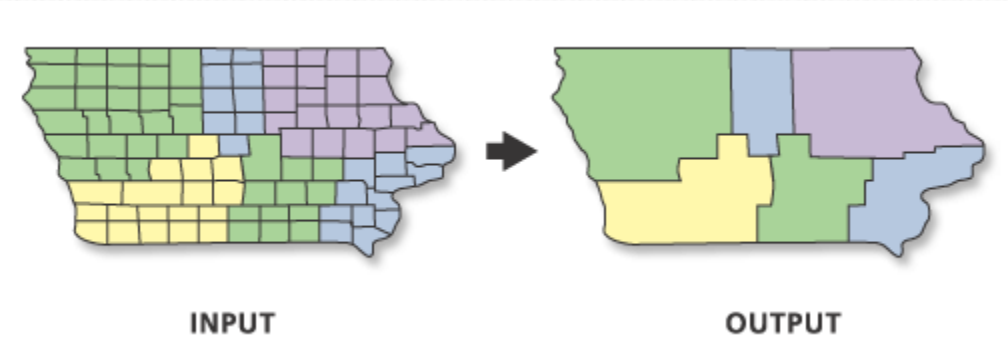
Reclassification and display

- After the retrieval of required data, reclassification is necessary
- So that redisplay of the retrieved data can be done which will exactly reflect the user's perception
- The similar polygons in adjacent places / boundary shared polygons having same codes / values / attributes,
- Then Dissolving has to be done
- Then, recompute the polygon structure again

Due to the multiple fractionated classes of polygon groups with 1000's of features, it is important to provide clarity for representing the output.

Similar to processing of data prior to analysis, post processing of out put is also necessary.

Data Classification/Grouping , Regrouping,/Reclassification



Finally using option like,

DISSOLVE

MERGE

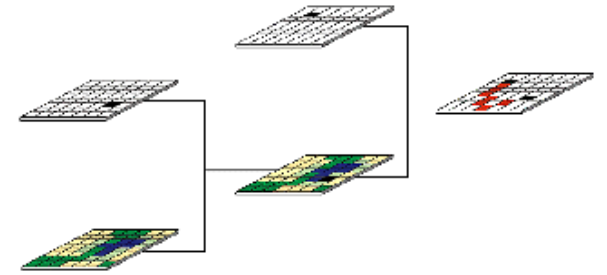
UPDATE

CALCULATE RASTER ……

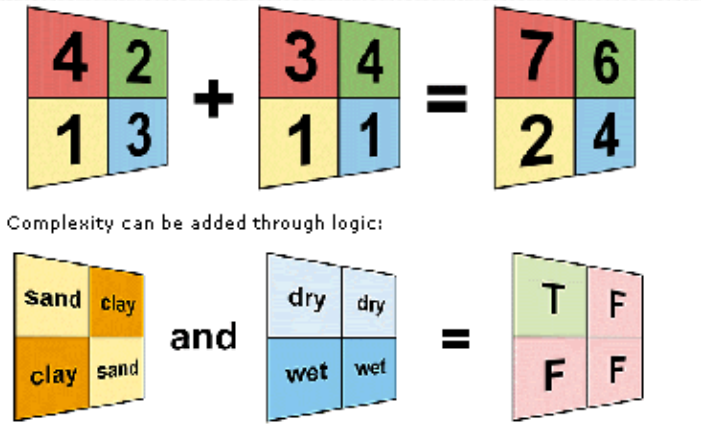
Simplified user friendly outputs can be presented.

GIS Modeling

- A model is a representation of reality.
- Due to the inherent complexity of the world and the interactions in it, models are created as a simplified, manageable view of reality.
- Models help to understand, describe, or predict how things work in the real world.

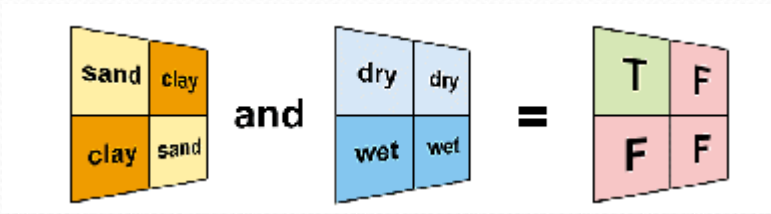


- **Binary model**
- **Process model**
- **Index model**
- **Regression model**
- **…**
- **…**



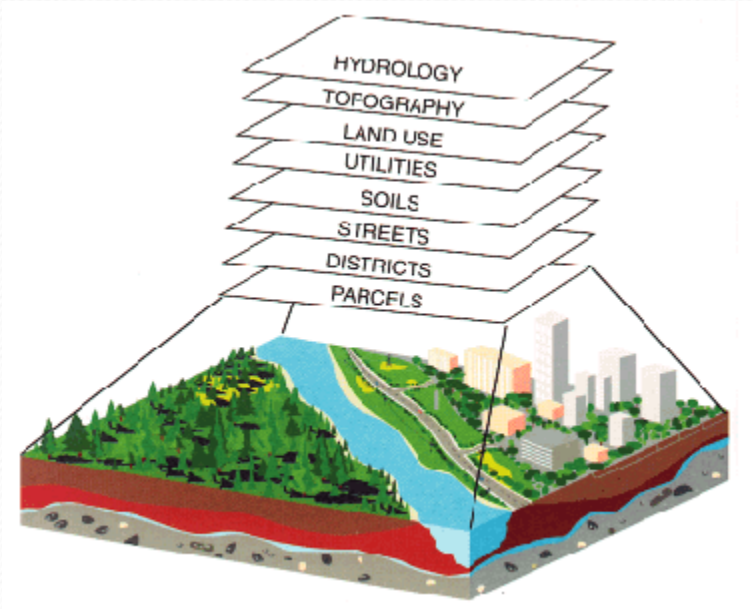
GIS MODELLING ...CONTD...

Binary Model



Yes / No, True / False, 1 / 0, etc.

Process Model



Back

Lookup Table (LUT)

Combined Attribute table containing the values / classes available from the both thematic layers involved in GIS integration.

1st Thematic Map / Layer having 3 classes – e.g. Thickness of Top soil (TTS) High, Moderate & Low

+

2nd Thematic Map / Layer having 3 classes – e.g. Water level (WL) - Deep, Moderate & Shallow

	TTS-H	TTS-M	TTS-L
WL-D	DH	DM	DL
WL-M	MH	MM	ML
WL-S	SH	SM	SL

In Binary model, Thematic layers can be classified with only two classes, Yes - No. In this case, once the GIS integration is performed then, the number of combined classes in the resultant layers can be easily determined by a simple formula,

$$N_{rc} = 2^n - 1 \ ,$$

Where,

n = no. of layers having Y/N classes involved in GIS integration analysis.

$\mathbf{N_{rc}}$ = resultant number of classes

GIS Integration of **Two layers** in binary model

	TTS-1	TTS-0
WL-1	11	10
WL-0	01	00

	SLP-1	SLP-0
11	111	110
10	101	100
01	011	010
00	001	000

	GEM-1	GEM-0
111	1111	1110
101	1011	1010
011	0111	0110
001	0011	0010
110	1101	1100
100	1001	1000
010	0101	0100
000	0001	0000

GIS Integration of **Three layers** in binary model

First step of GIS integration in Binary model will result 3 + 1 (1-unwanted) classes.

In the **Second step** of GIS integration, 7 +1(unwanted) classes are resulted.

$$N_{rc} = 2^3 - 1$$

Similarly, for the GIS Integration involving **Four layers** in binary model will result into ($2^4 - 1$) = **15** + 1 (1-unwanted) classes.

Network Analysis / Model

- **Network** – Consists of connected linear features
  - E.g. Road, Railway, Bicycle paths, Streams & Shorelines.
- **Dynamic Segmentation**
  - A data model that is built upon lines of a network and allows the use of real-world co-ordinates with linear measures such as mileposts
  - To link linearly referenced data such as accidents & pavement conditions to geographically referenced road network
- For this, the networks need to be segmented and labelled by inserting nodes in proper locations
- For both **NW/DS** - Appropriate attributes are to be collected as per the real-world applications and entered in computer for 'Linking'.
  - For e.g., to use road network for Path Finding / Resource Distribution, etc., - we need the following attributes:

- Travel time
- Impedance at turn
- One-way streets
- Linear measures – to be added with arcs and nodes of the network – can be directly linked to road segments.

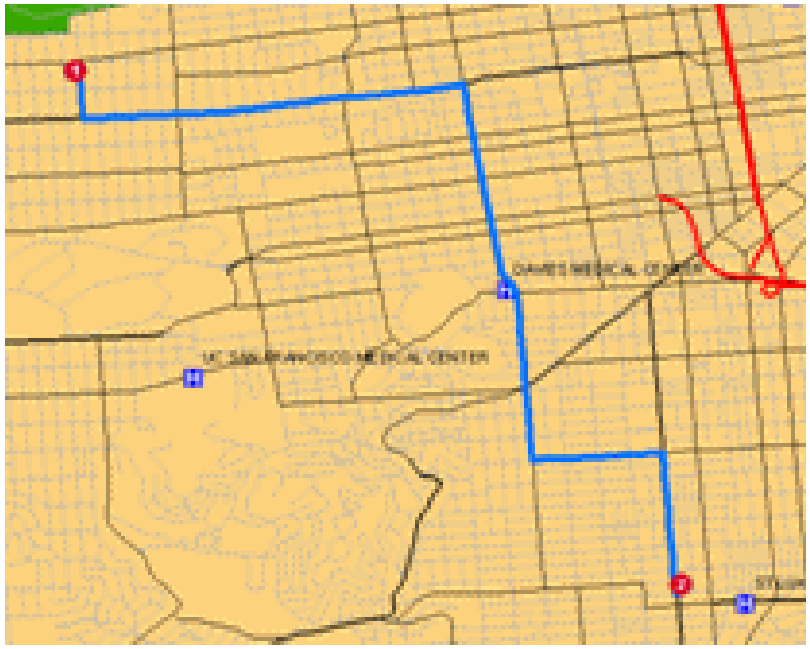
- 1. Network – is a line coverage, which is topology based and has the appropriate attributes for the flow of objects such as traffic.
  - Linear Feature + Attribute data, such as
    - Impedance values assigned to
      - network links – link impedance
      - Turns – turn impedance
    - One-way streets and
    - Overpasses and under passes.

Link impedance

- A link refers to a segment separated by two nodes in a road network

NETWORK ANALYSIS

Route Identification



DATA SORTING And PERFORMING STATISTICAL OPERATIONS

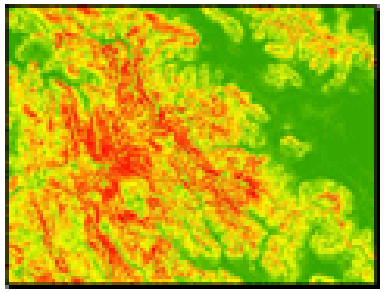
- Data Sorting
- Mean, Mode, Std. Deviation, Correlation, Factor analysis, Multi variate analysis, Multiple Regression analysis, etc.

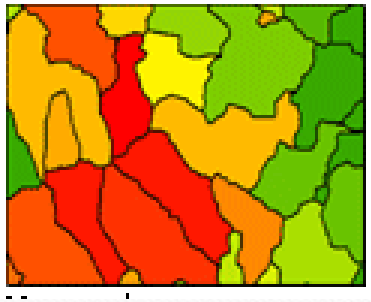
In order to perform statistical analysis in GIS, it is necessary to have samples of uniform length (equal number of samples) for all the thematic maps.

Hence, it would be better to convert the vector data to Raster format for statistical analysis.

DATA CONVERSION

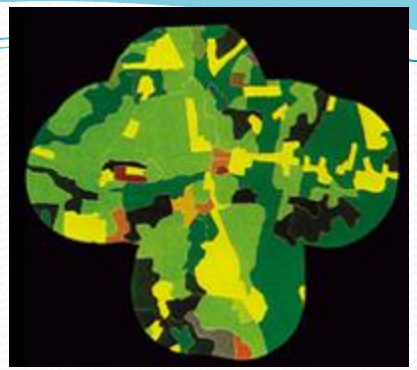
Raster ⇄ Vector



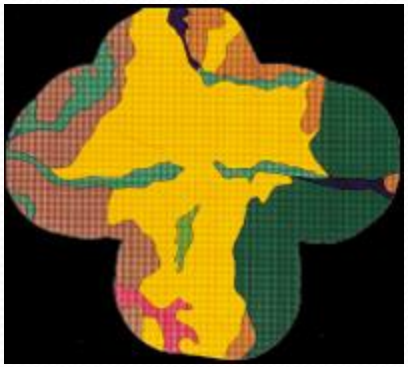


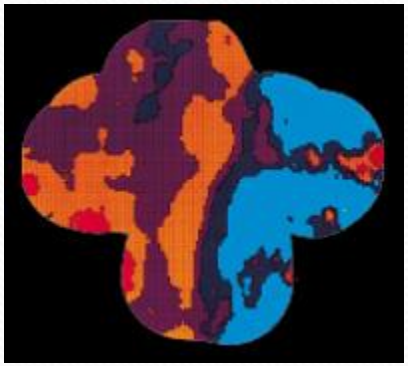
Identification of Potential groundwater well sites

Land use and land cover data for the area bounded by a half-mile buffer zone around the water company service area.



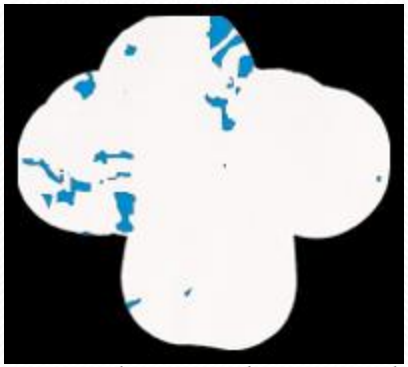
Map of surficial geology of the water service area





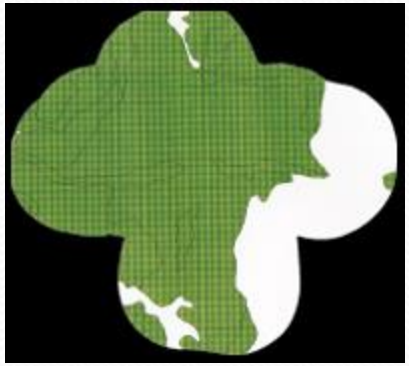
A bedrock elevation subtracted from water level elevation by a GIS to show the thickness of water-saturated sediment

A half-mile buffer zone drawn around the service area

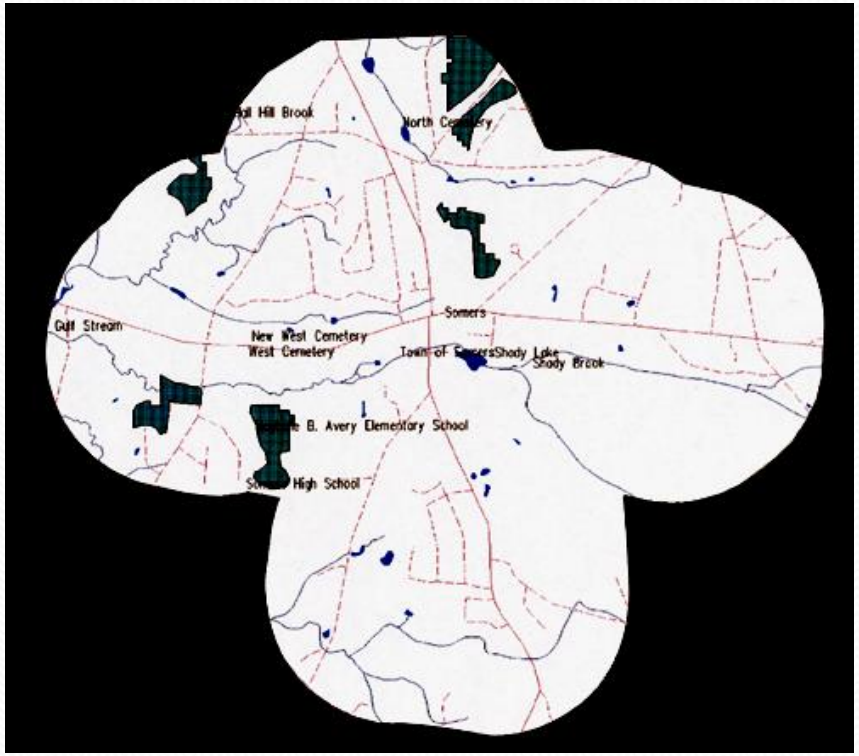


Potential sites with saturated thickness of sediments greater than 40 feet.

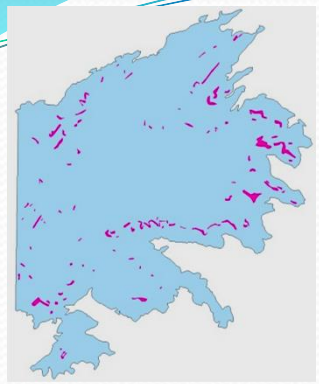
Selection areas of sand and gravel from the map of surficial geology



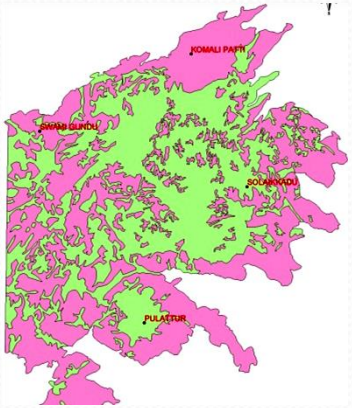
Buffer zones of 500 meters are drawn around the point sources of pollution.



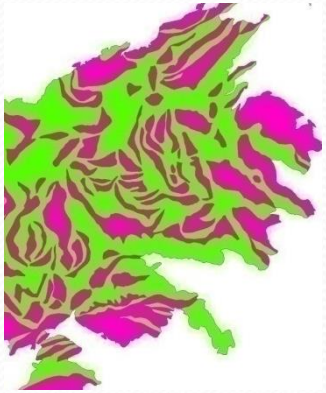
LANDSLIDE HAZARD ZONATION AND INDUCING PARAMETER IDENTIFICATION



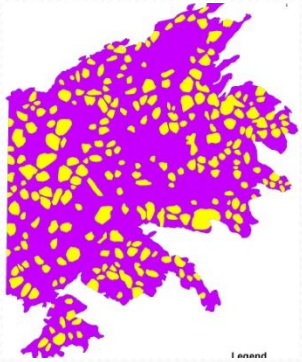
Mapping of Escarpments



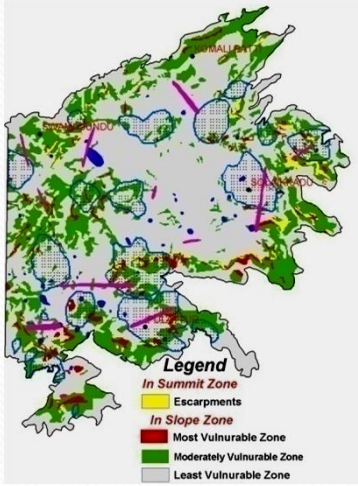
Identification of Active slope areas



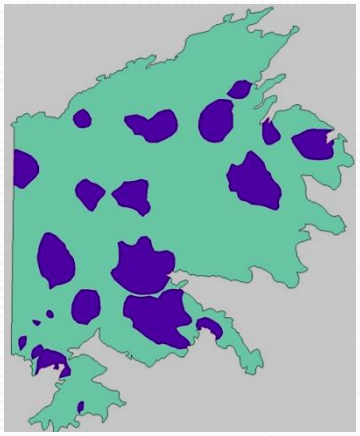
Interpretation of dip & obsequent slopes



Map out dissected slopes

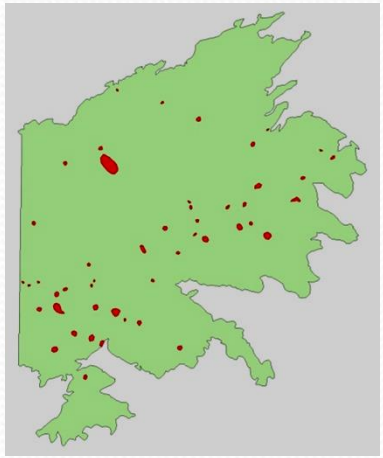


Landslide Hazard Zones

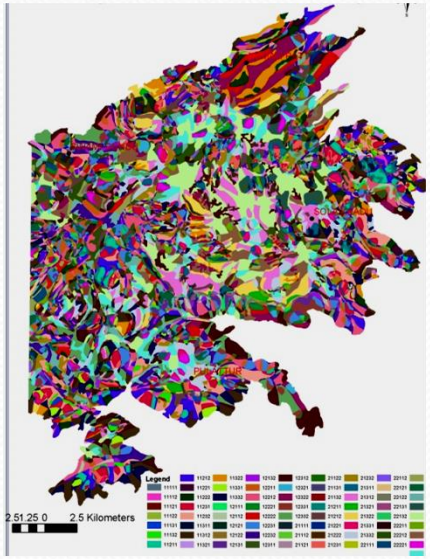


Zones of dendritic drainages

Filter out Convex & plain slopes



Buffer out sensitive toes



Landslide inducing parameters


Thank you
For your kind cooperation
Patient listening & Learning GIS
Let us enjoy using GIS more efficiently & effectively...