The Analysis of Discrete Entities in Space

Dr.Weerakaset Suanpaga
(D.Eng RS&GIS)

Department of Civil Engineering
Faculty of Engineering , Kasetsart University
Bangkok, Thailand

https://course.ku.ac.th/lms/files/resources_files/20939/84179/8discrete_6p.pdf

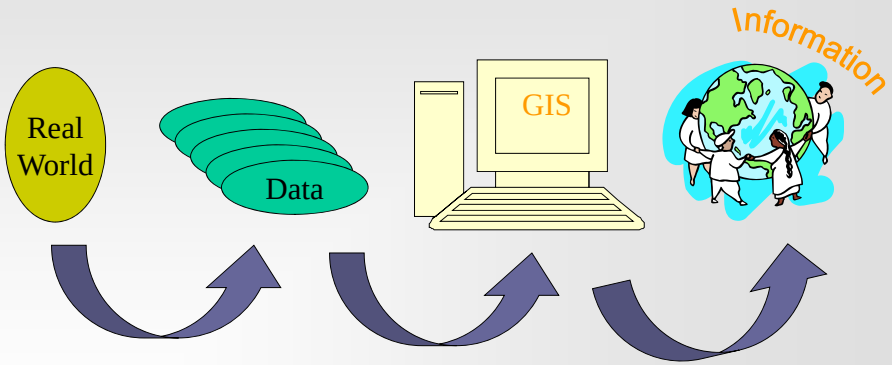
Aim of GIS ?

To create spatial and non-spatial database?

Not just this, but also

To facilitate query, retrieval and analysis of the data

What are discrete entities ?

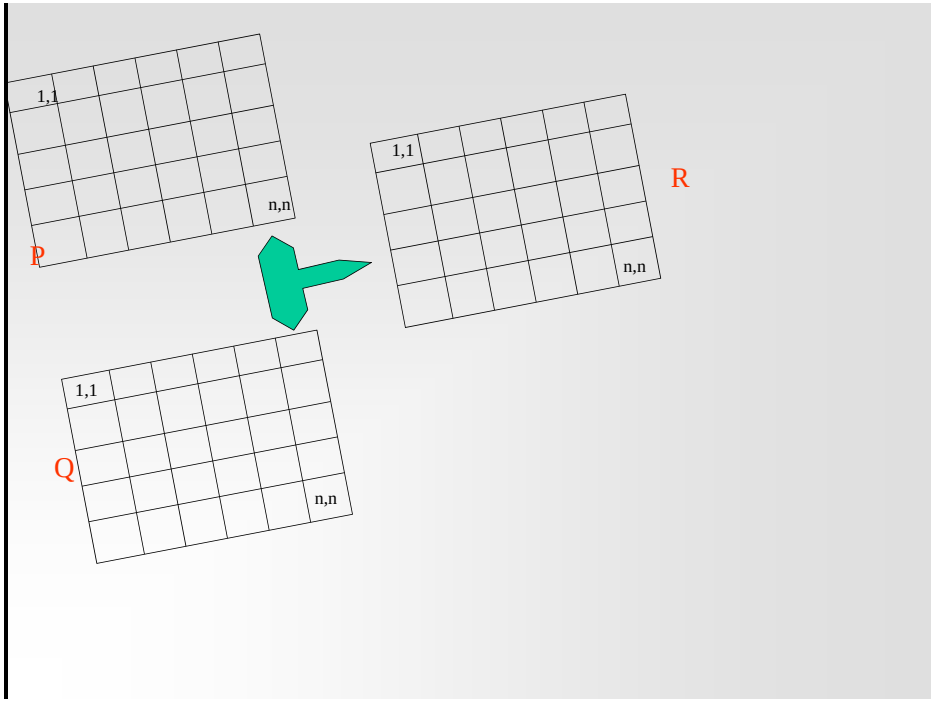


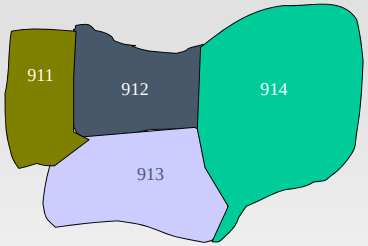
What are discrete entities ?

In case of entities – data retrieval and analysis concern the attributes, location, and connectivity of the entities and measures the way they are distributed in the space;

In the case of continuous fields, data analysis concerns the spatial properties of the fields.

The matter is many times complicated – when continuous fields are discretized and can be treated as individual entities.





Shape	ID	Population_Total	Num_School
Poly	911	23988	32
Poly	912	26980	42
Poly	913	32760	54
Poly	914	18700	8

Spatial analysis is more than JUST querying

It refers to deriving information based on attributes of discrete entities in the space

And depends on

attributes — location in space

distribution in the space — connectivity of entities

Real world

Decision criterion

Basic Classes of Operations for Spatial Analysis

Attribute operations

- Operations on one or more attributes of an entity
- Operations on one or more attributes of multiple entities that overlap in space
- Operations on attributes of entities that are contained by other entities (points in polygon)
- Operations on one or more attributes that are linked by directed pointers (object orientation)

Distance or location operations

Operations to locate entities with respect to simple Euclidean distance or location criteria

Operations to create buffer zones around an entity

Operations using in-built spatial topology

Operations to model spatial interactions over a connected net

These operations create new attributes

Operations on the attributes of geographic entities

What are attributes?

Location

latitude, longitude and altitude

Non-spatial property

qualitative or quantitative descriptors

Derived from spatial properties

based on spatial properties of the entity – perimeter, area, volume, nearness etc.

Process of selecting or creating a new attribute

For any given location *i*, *value of new attribute* U_i

$U_i = f(a, b, c, d)$

where a, b, c, d are the values of the attributes considered to evaluate new attribute U_i

The function *f (a, b, c, d) can be of several forms such as*

- Logical (Boolean) operations
- Simple or complex arithmetical operations
- Numerical models
- Univariate statistical analysis
- Multivariate statistical model
- Multicriteria methods, AI-based methods, Fuzzy logic, neural networks based

Logical (Boolean) Operations

Entities can be selectively retrieved or classified based on standard Boolean rules

Logical operators

AND, OR, XOR, NOT

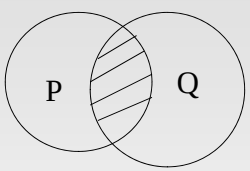
AND - intersection

XOR - exclusive OR

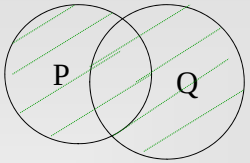
OR - union

NOT - difference operator

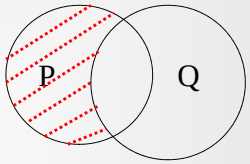
(Those entities that belong to one set but not to the other)



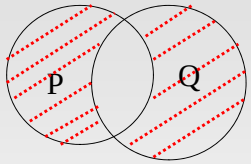
P AND Q



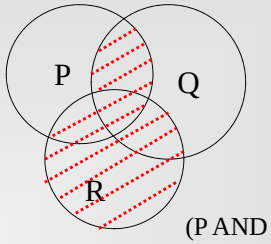
P OR Q



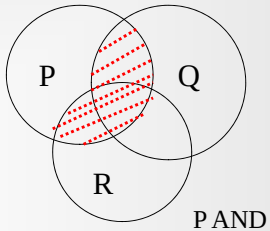
P NOT Q



P XOR Q



(P AND Q) OR R

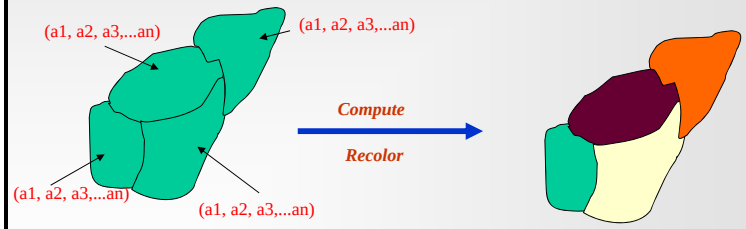


P AND (Q OR R)

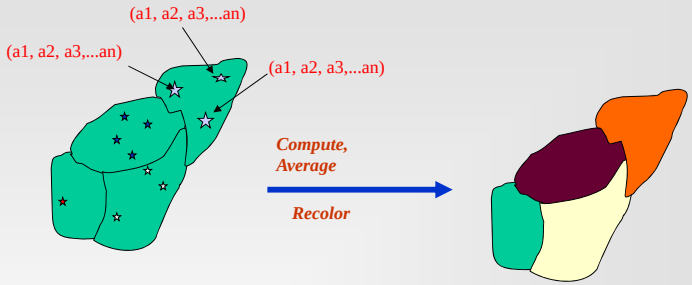
Spatial aspects of Boolean retrieval on multiple attributes of single entities

Logical operations and reclassification of non-spatial attributes of spatial entities has little effect on the map image, except in terms of symbolism and boundary removal

When selection leads to adjacent polygons receiving the same code then it may be desirable to dissolve the boundaries between them

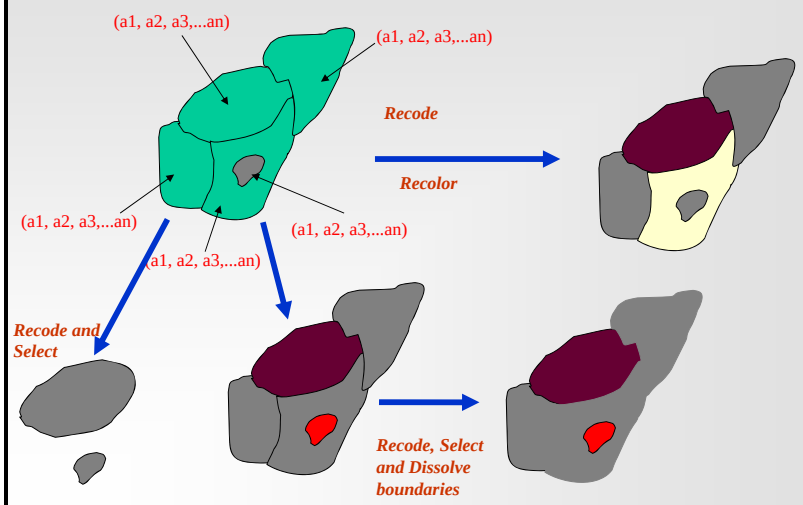


Points in polygon operation



$U_i = f(a, b, c, d)$

Boundary dissolve and map generalization



Statistical analysis of attributes

GIS Packages do not provide more than few basic statistical tools

Most statistical packages provide at least few procedures for statistical analysis

- Basic statistics- means, std. deviations, variance, skewness, kurtosis, maxima and minima
- Non-parametric statistics – median and mode
- Histogram, 2D and 3D scatter plot
- Linear regression and correlation
- Discriminant analysis
- Principal components and factor analysis

Deriving new attributes for spatial entities

Empirical regression model

A simple statistical model is the linear regression of the temperature in Swiss Alps as a function of elevation

$T = 5.697 - 0.00443 * E$

where T is in degrees Celsius and

E is elevation in metres

This relation is applied to a grid matrix of elevation to get the output image as temperature

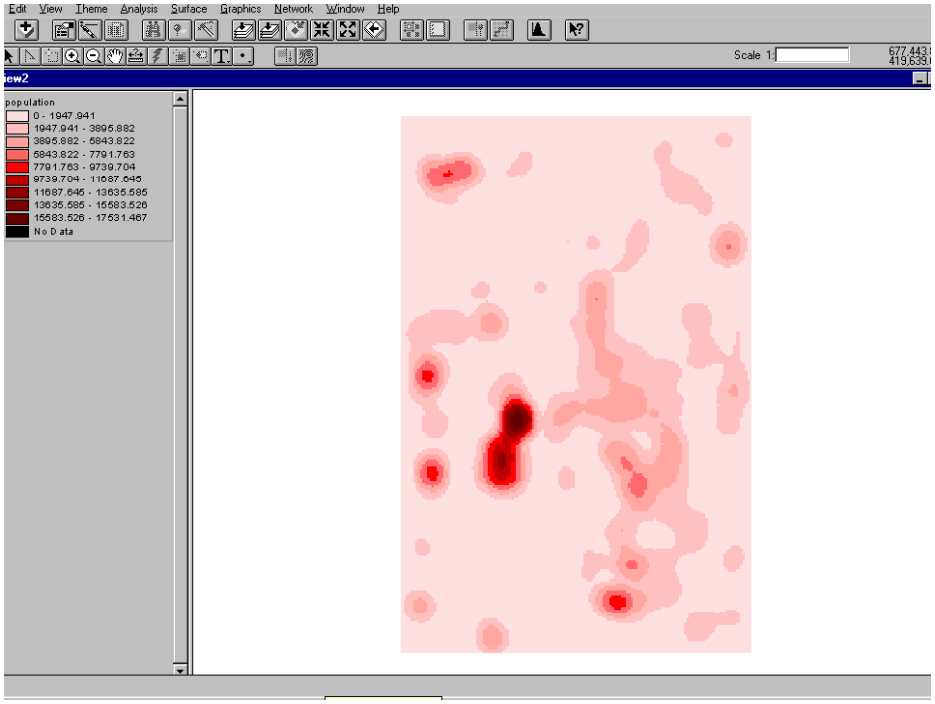
Mapping Classes

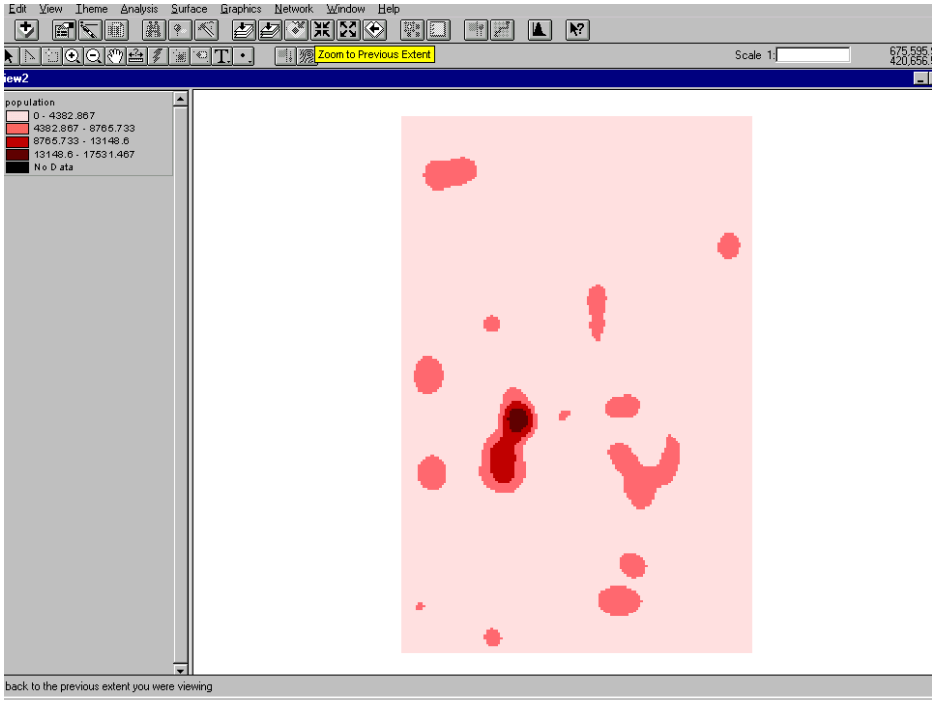
Classes group features with similar values, by assigning them same symbol

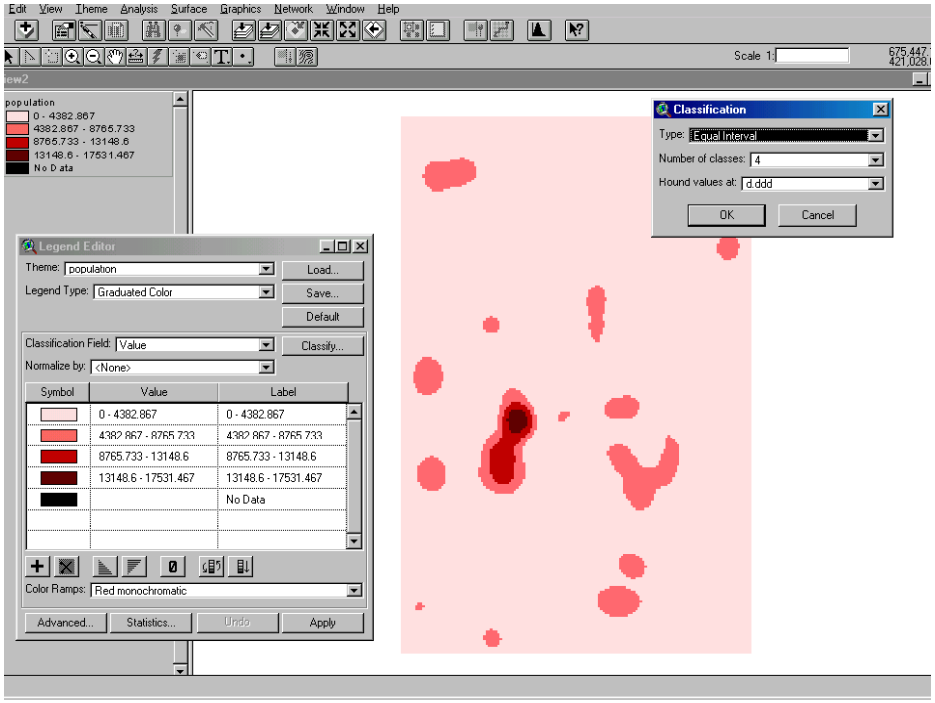
From same data – different class assignments possible – different map appearance possible

Why classes ?

– to identify similar group in large dataset







Features of similar values are in one class

Other class values should be distinctly different

You can classify manually

or

Use a standard classification scheme

Reclassify

Allows you to set standard classification methods and the parameters for that classification.

Dialog box options

Type Classification method to reclassify the data with. Type can be

Equal Area - Reclassifies so there is an equal number of cells in each of the output classes.

Equal Interval - Divides the range of cell values into equal sized sub-ranges.

Natural Breaks - Identifies breakpoints between classes using a statistical formula (Jenk's optimization). The Jenk's method minimizes the sum of the variance within each of the classes. Natural Breaks finds groupings and patterns inherent in your data.

Quantile - Reclassifies so there is an equal number of features in each of the output classes. For a grid theme this is exactly the same as Equal Area.

Standard Deviation - Finds the mean cell value and then places class breaks above and below the mean at intervals of either 1/4, 1/2, or 1 standard deviations until all the cell values are contained within the classes. Any cell values that are beyond three standard deviations from the mean are put into two classes, greater than three standard deviations above the mean ("> 3 Std Dev.") and less than three standard deviations below the mean ("< -3 Std. Dev.").

Only the Equal Interval and Standard Deviation methods are available for floating point grid themes.

Number of classes Shows the number of classifications that will be generated when you classify the values in the Classification Field. This is not used for the Standard Deviation classification.

Round values at Shows the rounding scheme that will be used to determine where class breaks are placed.

Break Classes At This option is shown for Standard Deviation classification instead of Round values at. Sets class breaks at 1, ½, or ¼ standard deviations.

Comparing classification schemes

Natural Breaks

Advantages

Mapping data values that are not evenly distributed

Disadvantages

1. Since class ranges are specific to the individual dataset, it is difficult to compare the map with other maps

2. Choosing the optimum number of classes is difficult

Quantile

Advantages

1. Comparing areas that are approximately same size
2. Mapping data in which values are evenly distributed

Disadvantages

1. Features with close values may end up in different classes

2. If areas vary greatly in size, a quantile classification can skew the patterns on the map

Standard Deviation

Advantages

1. Seeing which features are above or below an average value
2. Displaying data which has many values around the mean and only few away from the mean

Disadvantages

1. The map does not show actual values of the features, only how far they exist from the mean

2. Very high and low values can skew the mean so that most features will fall in the same class

Equal Interval

Advantages

1. Information to a non-technical audience
2. Mapping continuous data, rainfall and temperature

Disadvantages

If the data values are clustered rather than evenly distributed, there may be many features in one or two classes and some classes with no feature

Deciding how many classes

After deciding classification scheme ⟶ how many classes

Based on classification criteria and no. of classes – a map of class ranges with break is created

Changing the number of classes should not change the basic appearance of the map – only make pattern more/ less distinct

Most map readers can distinguish upto seven colors on the maps

Usually 4 to 5 classes reveal the pattern in the data

Make the classes easier to read

After finalizing classification scheme and number of classes -> make the map easier to read

If showing exact data values is optional then round off the minimum and maximum of class values for easy interpretability

You may even label the classes ->

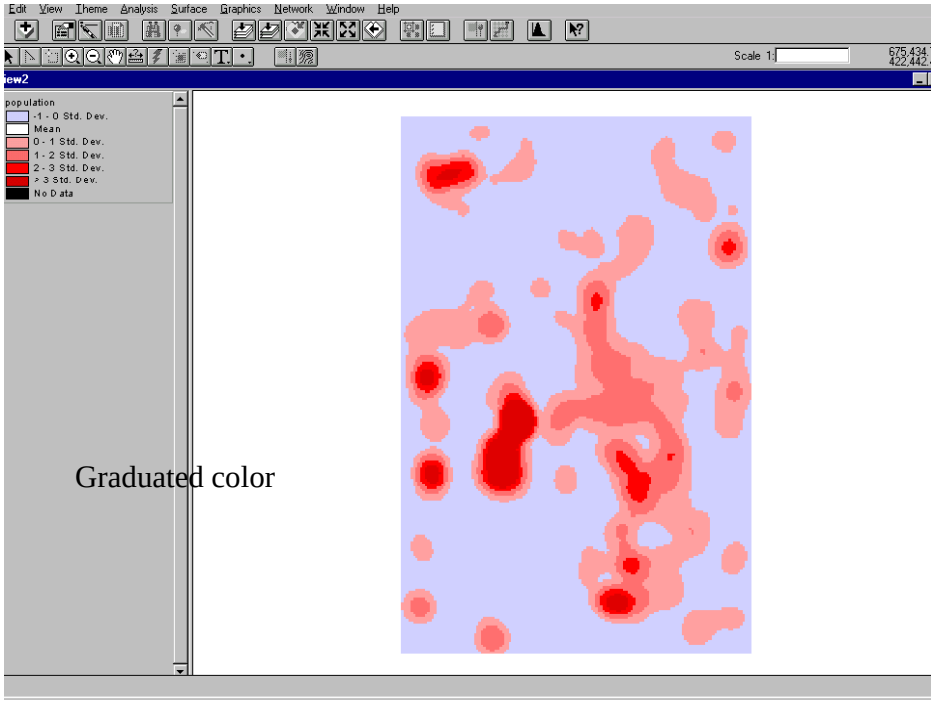
low, medium, high and very high

Choose an appropriate color scheme – ramp

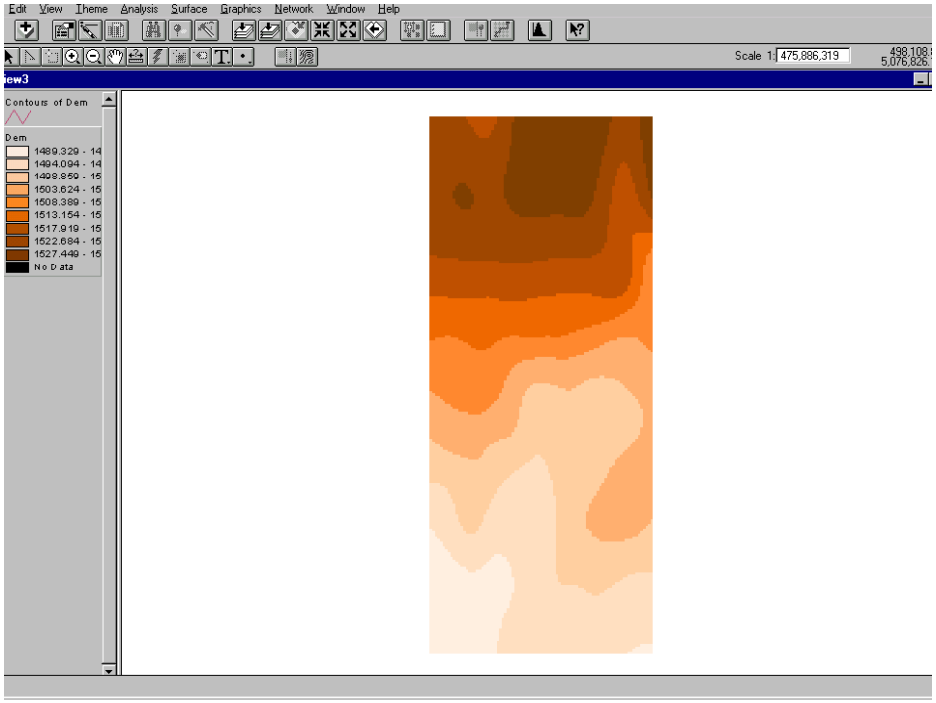
Low values are brighter and high values darker

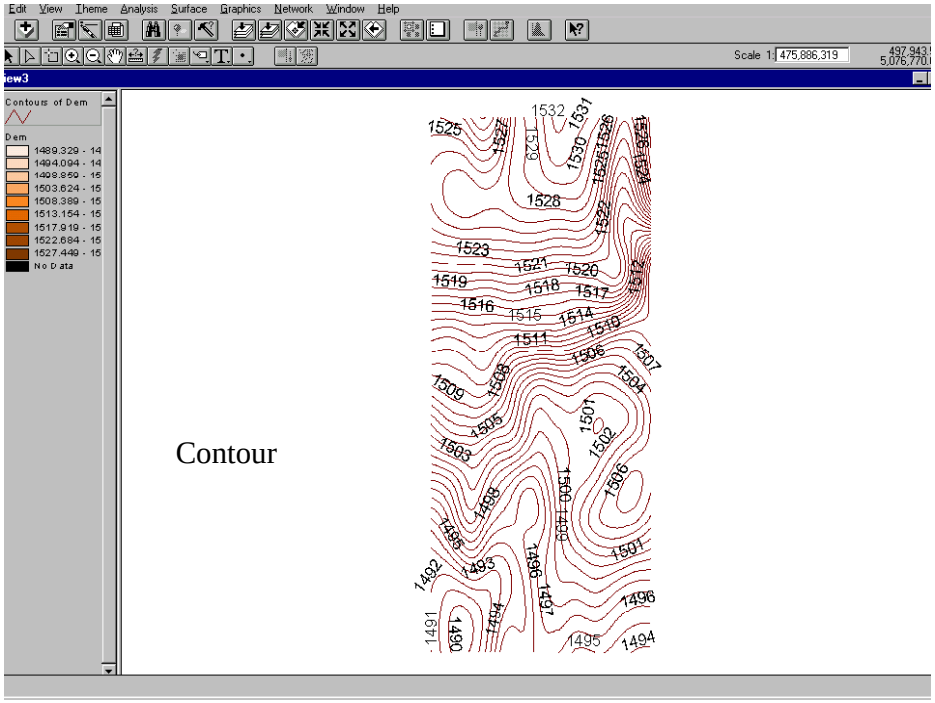
Suitable Map

1. Graduated symbols
2. Graduated colors
3. Charts
4. Contours
5. 3-D perspective view

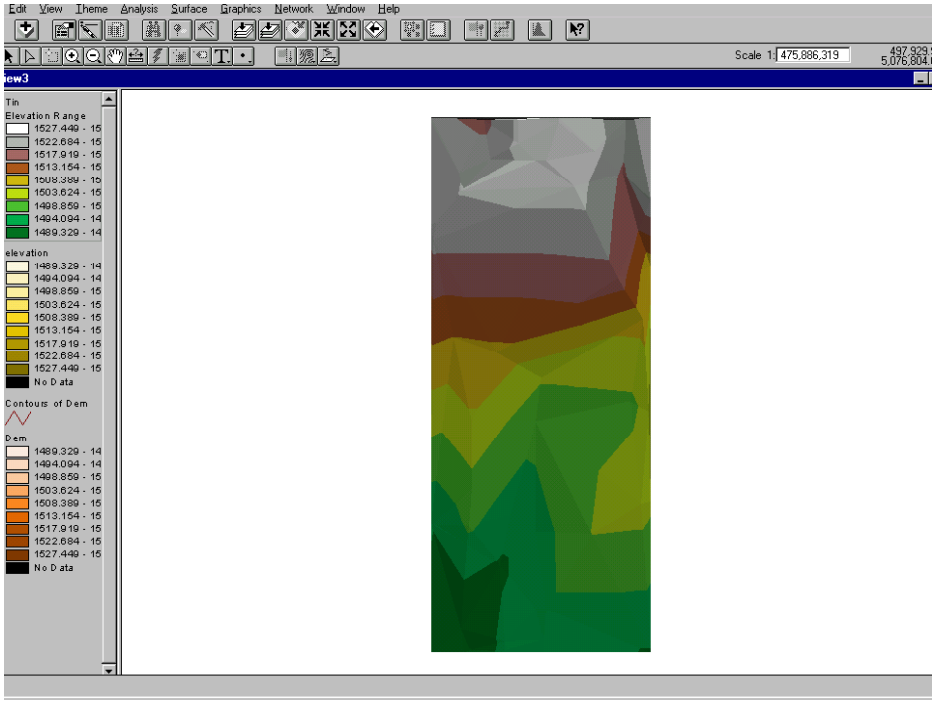


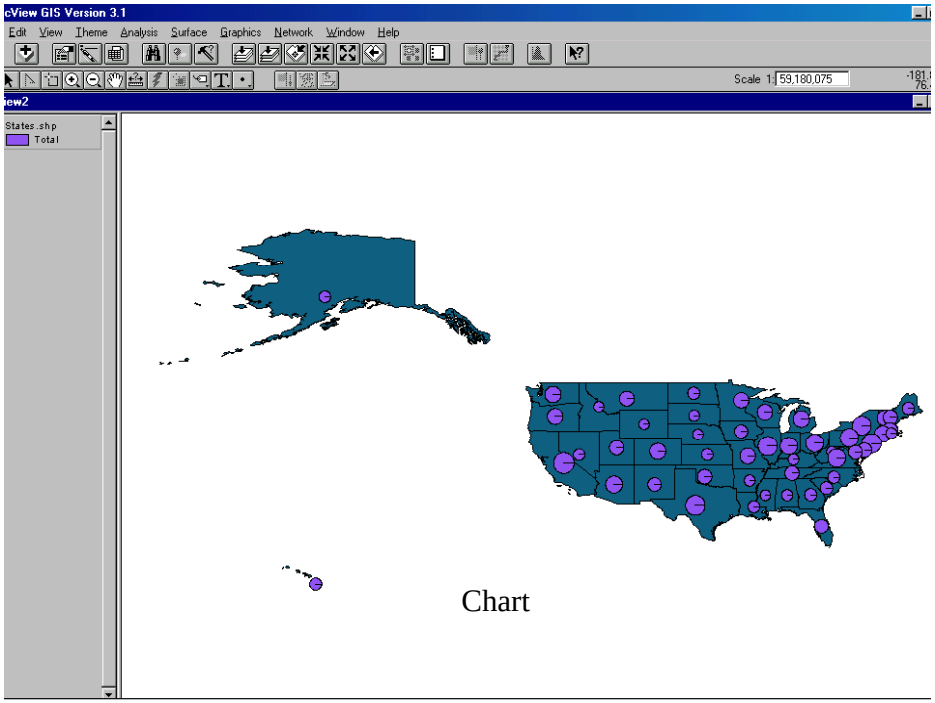
Graduated color





Contour





Chart

Choosing a map type

If you have discrete locations or lines, use

Graduated symbols or value ranges

Charts to show both categories and quantities

A 3-D view shows relative magnitude in 3rd Dimension

If you have discrete areas, or data summarized by area, use

Graduated symbols or value ranges

Charts to show both categories and quantities

A 3-D view to show relative magnitude

If you have spatially continuous data, use

Graduated symbols or value ranges

Contours to show rate of changes

A3-D view to show relative magnitude

The ESRI Guide to

GIS Analysis

- Andy Mitchell

Creating 3D View

Viewer Location

Z-Factor

Light Source

Displaying a perspective view

Models using discrete entities in space

Model 1: Real estate agent

- Spatial database used by a real estate agent
  - Property map (cadastre map)
  - Housing map
  - Cost map

A typical query

- Please show me the locations of all houses costing between Baht 2 Million and Baht 3 Million with 3 bedrooms, 1 Kitchen, and plot measuring at least 200 sq. m.

Please show me the locations of all houses costing between Baht 2 Million and Baht 3 Million with 3 bedrooms, 1 Kitchen, and plot measuring at least 200 sq. m.

- Model

```
IF COST GE 2Million AND COST LT 3Million
   AND NBEDROOM = 3 AND NKITCHEN = 1
   AND PLOTAREA GE 200
THEN
   ITEM = 1
ELSE
   ITEM = 0
```

Please show me the locations of all houses costing less than Baht 2 Million and with 2 bedrooms, No Kitchen, and plot measuring at least 100 sq. m.

- Model

?

Land Suitability Classification

- Database
- Soil mapping units
  - Each Mapping Units
    - Texture
    - pH

- Set A is set of mapping units called Oregon Loam
- Set B is set of mapping units for which soil pH equals or exceeds 7.0
- How the data retrieval statement will be?

Query

- X=A AND B
  - Finds all occurrence of Oregon loam with pH >= 7.0
- X= A OR B
  - Finds all occurrence of Oregon loam and all mapping units with pH >=7.0

Query

- X=A XOR B
  - Finds all mapping units that are either Oregon loam, or have a pH >= 7.0, but not in combination
- X= A NOT B
  - Finds all mapping units that are Oregon loam where pH <7.0

Query with more than 2 entities

- C is mapping units of poorly drained soils
- X=(A AND B)OR C
  - Finds all mapping units that are either a) Oregon loam with pH>= 7.0, or b) units of poorly drained soils
- X= A OR (B AND C)
  - Finds a) all Oregon loam mapping units and b) those mapping units with a combination of pH >=7.0 and poorly drained soils

Query with more than 2 entities

- C is mapping units of poorly drained soils
- X=A AND (B OR C)
  - Finds all mapping units that are either a) Oregon loam with pH>= 7.0, or b) units of poorly drained soils
- X= (A OR B) AND C
  - Finds a) all Oregon loam mapping units and b) those mapping units with a combination of pH >=7.0 and poorly drained soils

Query with more than 2 entities

- C is mapping units of poorly drained soils
- X=A AND (B OR C)
  - ?
- X= (A OR B) AND C
  - ?

Query with more than 2 entities

- C is mapping units of poorly drained soils
- X=A AND (B OR C)
  - Finds all mapping units that are either a) Oregon loam with pH>= 7.0, or b) units of poorly drained soils
- X= (A OR B) AND C
  - Finds a) all Oregon loam mapping units and b) those mapping units with a combination of pH >=7.0 and poorly drained soils?

Query with more than 2 entities

- C is mapping units of poorly drained soils
- X=A AND (B OR C)
  - Finds all mapping units that are a) Oregon loam and b) have pH>= 7.0, or units of poorly drained soils
- X= (A OR B) AND C
  - Finds a) poorly drained soils and b) Oregon loam mapping units or those mapping units having pH >=7.0

Text Reference

Chapter 7:	The analysis of discrete entities in space
Book :	Principles of Geographical Information Systems
Authors:	P. A. Burrough and R. A. McDonnell

Question?

Thank you for your attention