CRP 4080: Introduction to Geographic Information Systems for planners

Lecture 8: Site Suitability analysis

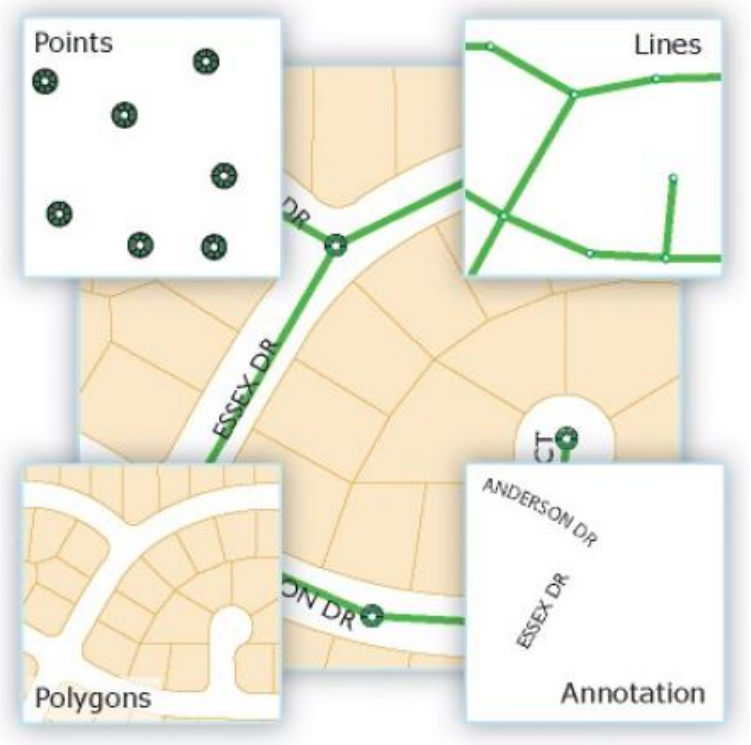
Wenzheng Li, Ph.D.
City and Regional Planning
Fall 2024

Overview

- Two data types
  - Vector data
  - Raster/imagery data
- Map Algebra
- Raster data function
  - local, neighborhood, and global levels
- Composite Suitability Analysis
- Model Builder

Types of Spatial Data in GIS – vector data

- **Vector/Feature Data**: Points, Lines, Polygons



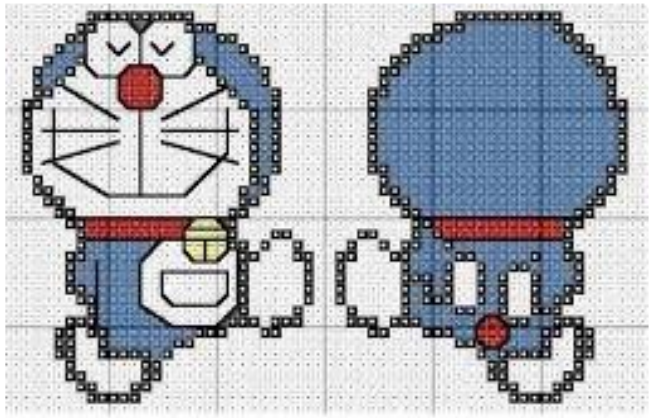
- Points are pairs of x,y coordinates.
- Lines are sets of coordinates that define a shape.
- Polygons are sets of coordinates defining boundaries that enclose areas.

Types of Spatial Data in GIS – vector data

Advantages of vector data:

- Accuracy & Aesthetically pleasing
- Ability to alter the scale of observation
- Topology (spatial relationship) is inherent in the vector model

Vector model

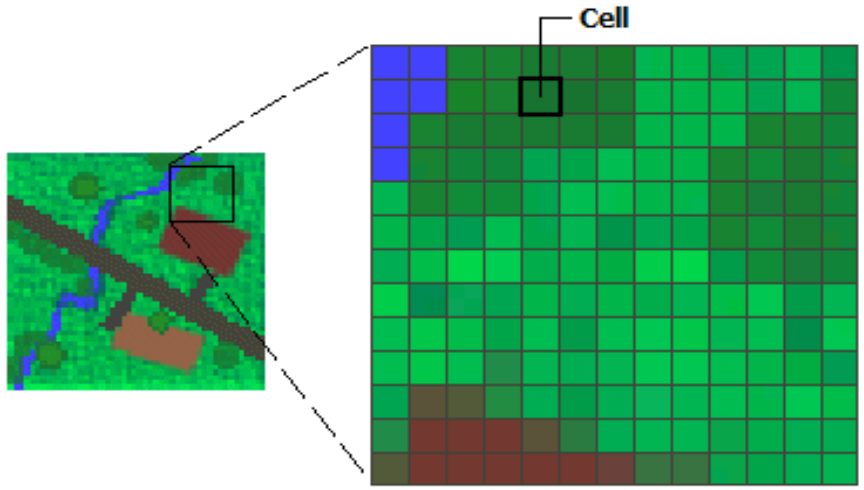


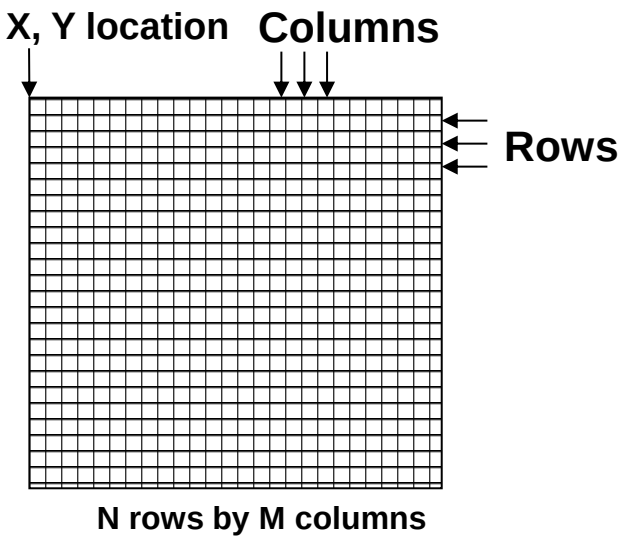
Raster model

Disadvantages of vector data:

- Storage and data structure much more complex: The location of each vertex needs to be stored explicitly
- Processing intensive: vector data must be converted into a topological structure
- Speed…
- Spatial limitations: Continuous data, such as elevation data, is not effectively represented in vector form

The raster data model

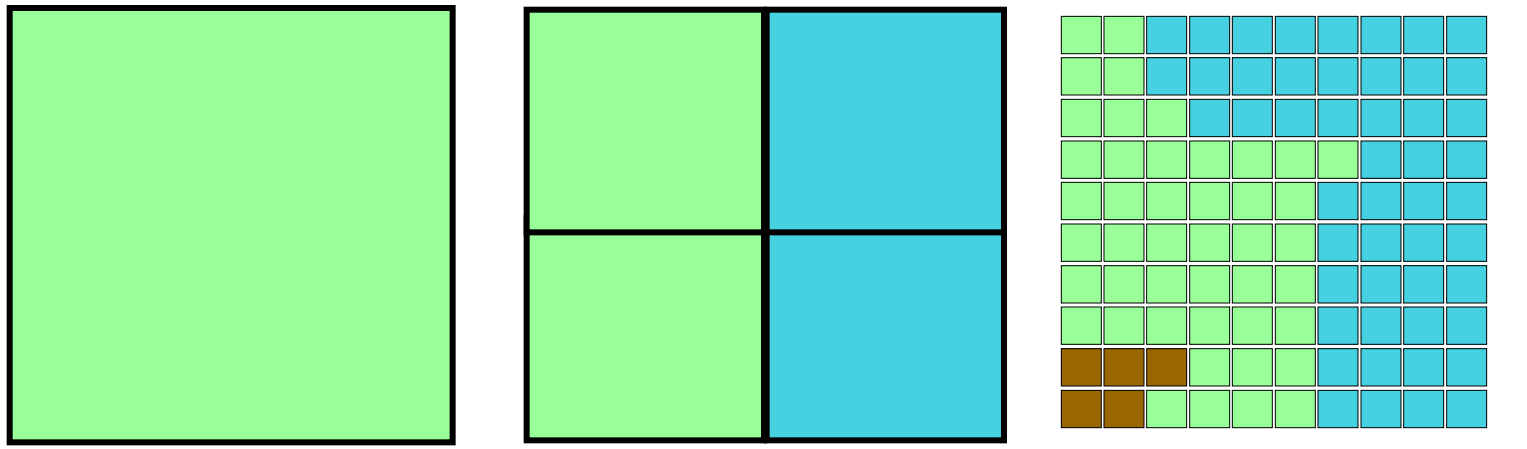




- A raster consists of a matrix of pixels (or cells) organized into rows and columns (or a grid) in which each **pixel** contains a **value** representing information.
- Raster data are collected by aircraft, drones, satellites, ground and water-based sensors, digital pictures, and scanned maps.
- Also includes aerial photographs and satellite imagery. File types: geoTIFF, TIFF, JPG, PNG, GIF, BMP, and other imagery file formats.

Cell Size & Resolution

- The **size** of the **cells** in the raster data model determines the **resolution** at which features can be represented
- Resolution - A measure of the accuracy or detail of a graphic display
- The resolution can have an effect on which features are represented in what locations:



Why resolution matters? Mixed-pixel issue

Water dominates

W	W	G
W	W	G
W	W	G

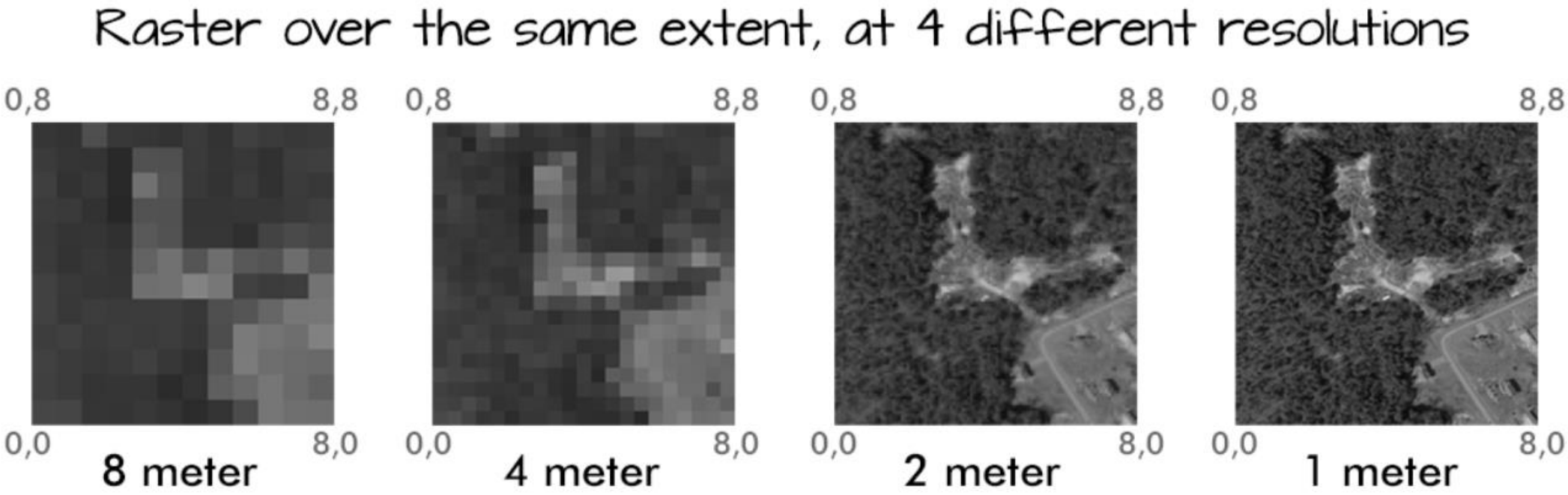
Winner takes all

W	G	G
W	W	G
W	G	G

Edges separate

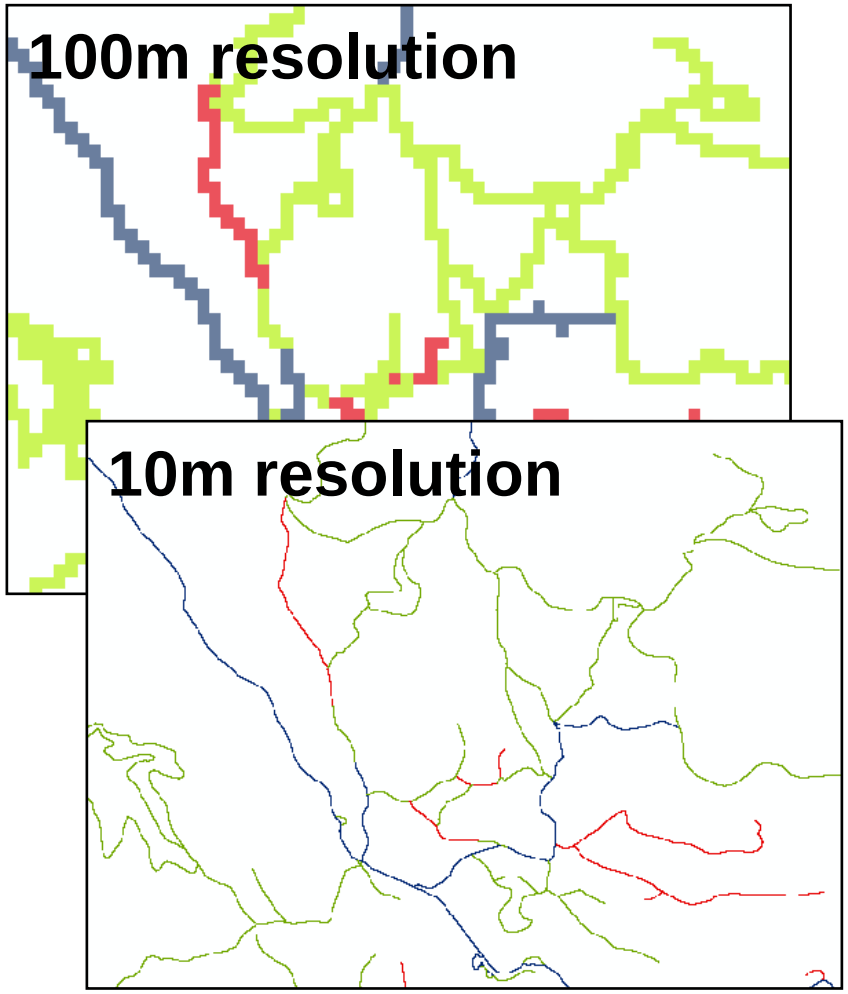
W	E	G
W	E	G
E	E	G

Solution? Increase spatial resolution



(Source: National Ecological Observatory Network (NEON)) {: .text-center}

Impact of resolution



- There is a **trade-off** between **spatial resolution** and **data storage** when we use the raster data model

- 60 km satellite image with 10m cell size
  - 6000 X 6000 = 36,000,000 cells
  - 1 byte of attribute value (i.e. values 0-255)
  - ~36 MB of disk storage!

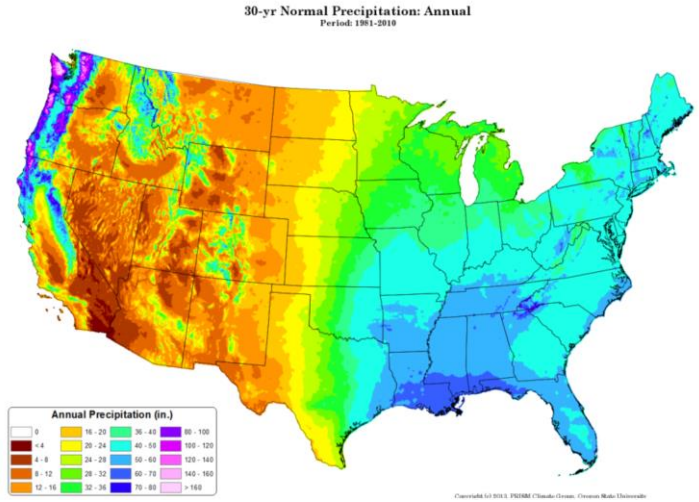
- 60 km satellite image with 100m cell size
  - 600 x 600 = 360,000 cells
  - 360 KB of data… **1%** the size of the other one

Discrete vs continuous: Utilizes numeric values

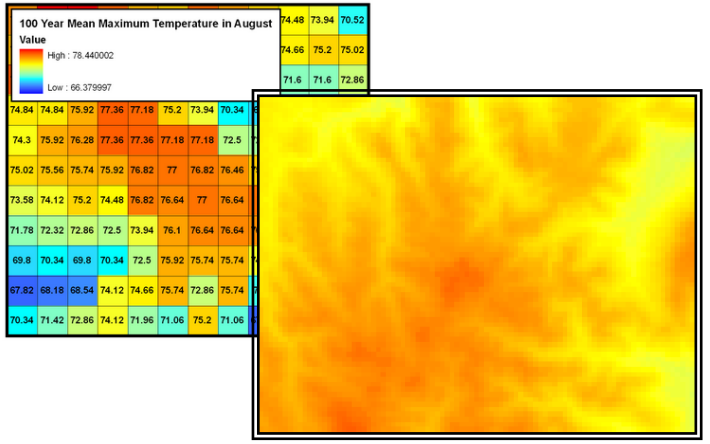
Continuous variation:

- **actual values** (Elevation, temperature, rainfall, noise levels)

Precipitation map



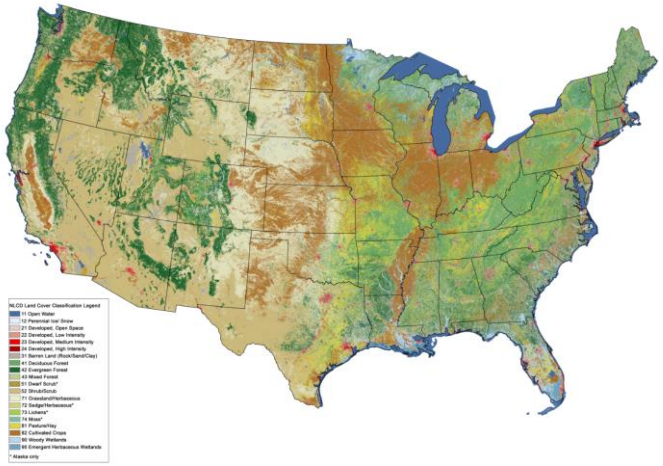
Temperature map



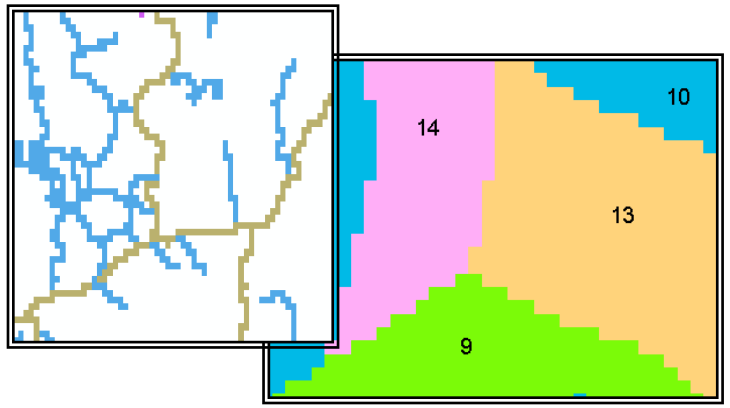
Discrete:

- **Codes** representing an **attribute** (land use, etc.)
- Elevation maps classified as low, medium, and high elevation.

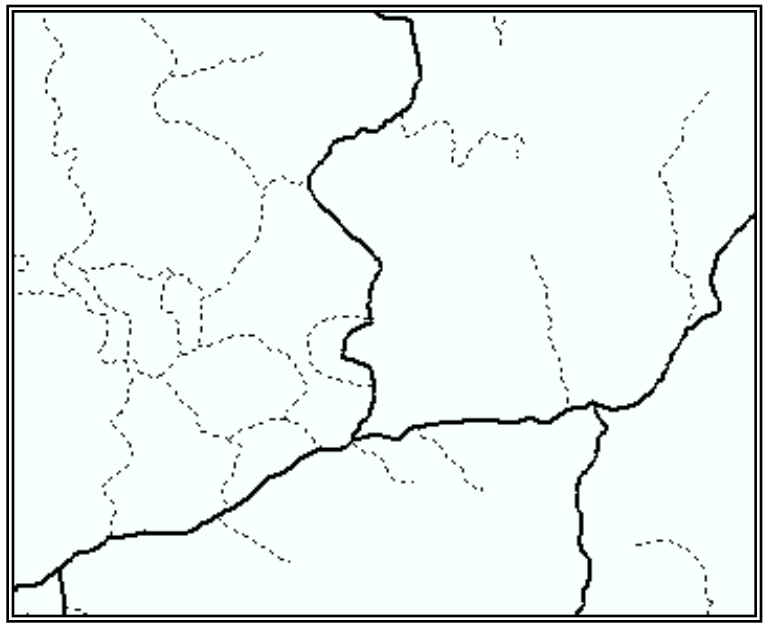
Landcover / land-use map (USA)



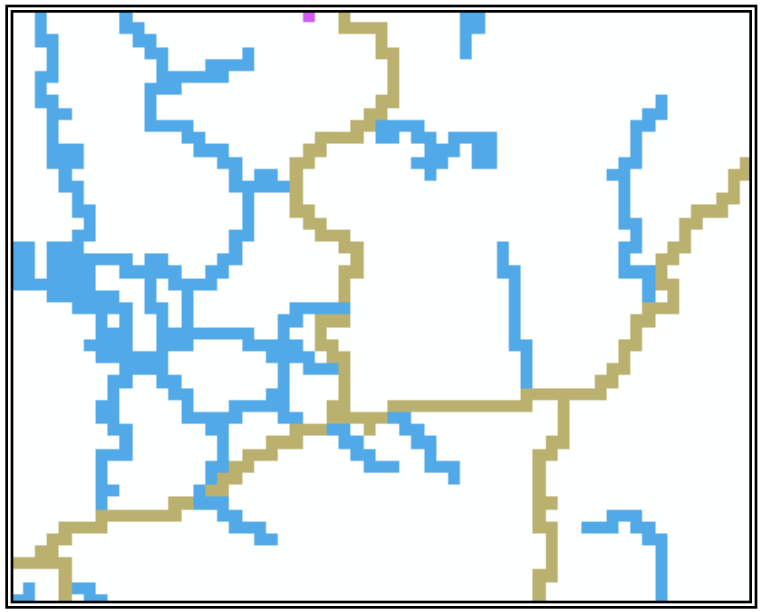
Landcover / land-use map (local)



Raster: Discrete



Vector model

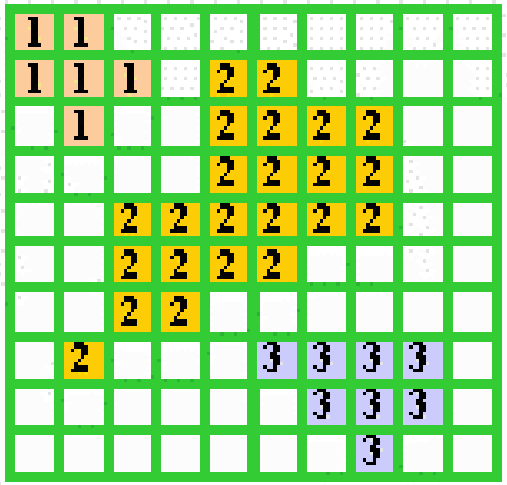


Raster model

Raster Data: Nominal vs. Interval/Ratio data

Nominal:

- Each cell is assigned a specific value representing the class (category)
  - Value Attribute Table (VAT): stores value assigned to each zone of a raster, a second field that shows the count - the number of cells that belong to that zone.



Treetype Grid

Value	Count	Type	Code
1	6	Maple	200
2	23	Oak	400
3	8	Pine	300

Treetype.vat

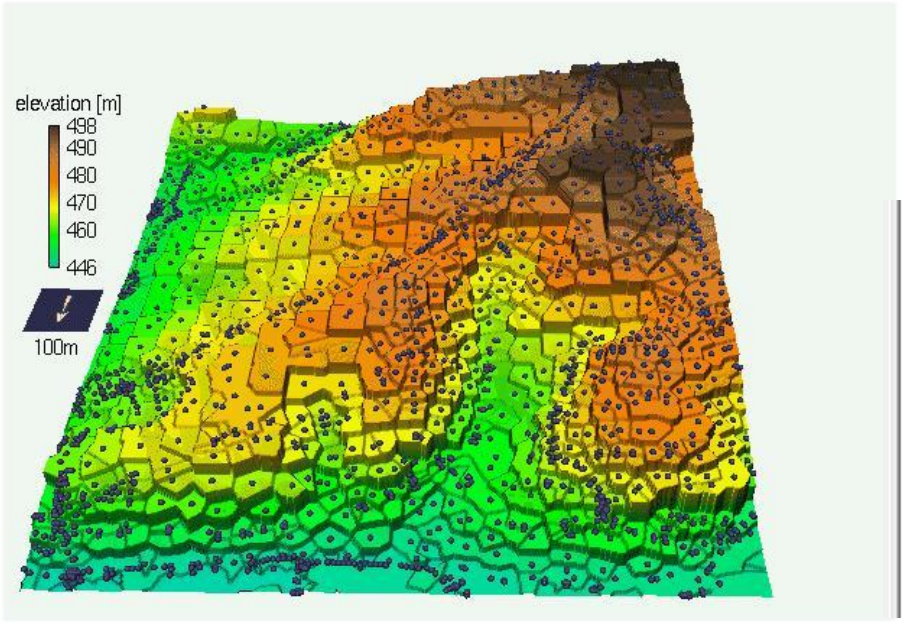
Attributes of reclass_slop1

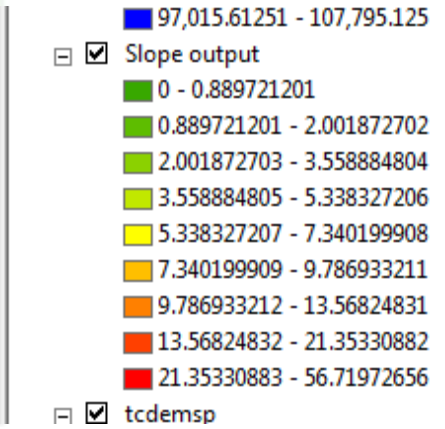
Rowid	VALUE *	COUNT
0	1	673
1	2	6074
2	3	24928
3	4	60442
4	5	155165
5	6	396790
6	7	882228
7	8	193837
8	9	363079
9	10	517860

Raster Data: Interval/Ratio:

represents a magnitude, distance or relationship of the cell on a continuous surface (e.g. elevation, slope).

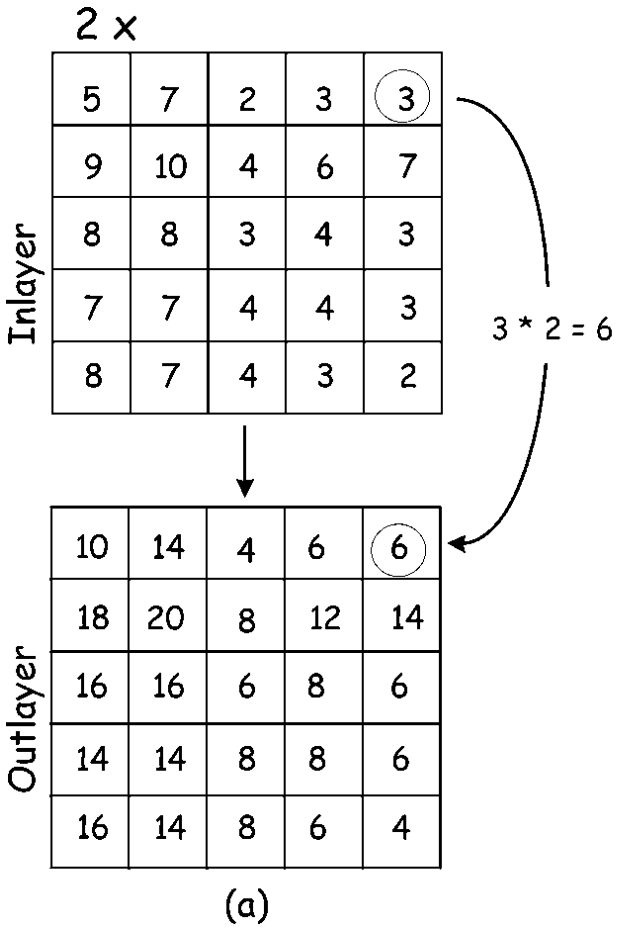
- unique data value for every cell in the grid.
- No VAT.

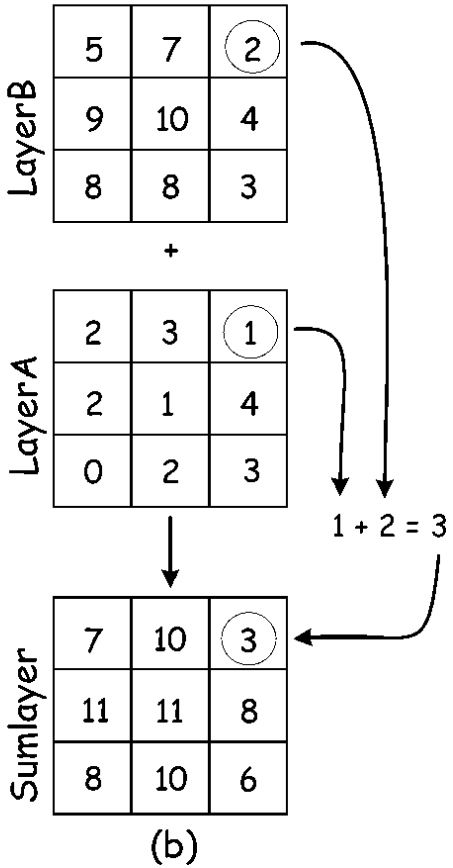




Map Algebra

- Map algebra is a cell-by-cell combination of raster layers using mathematical operations
  - Unary – one layer
  - Binary – two layers
- Basic Mathematical Operations
  - Addition, subtraction, division, max, min, virtually any mathematical operation you would find in an Excel spreadsheet
- But be careful of:
  - Layers that are not coincident
  - Different cell sizes

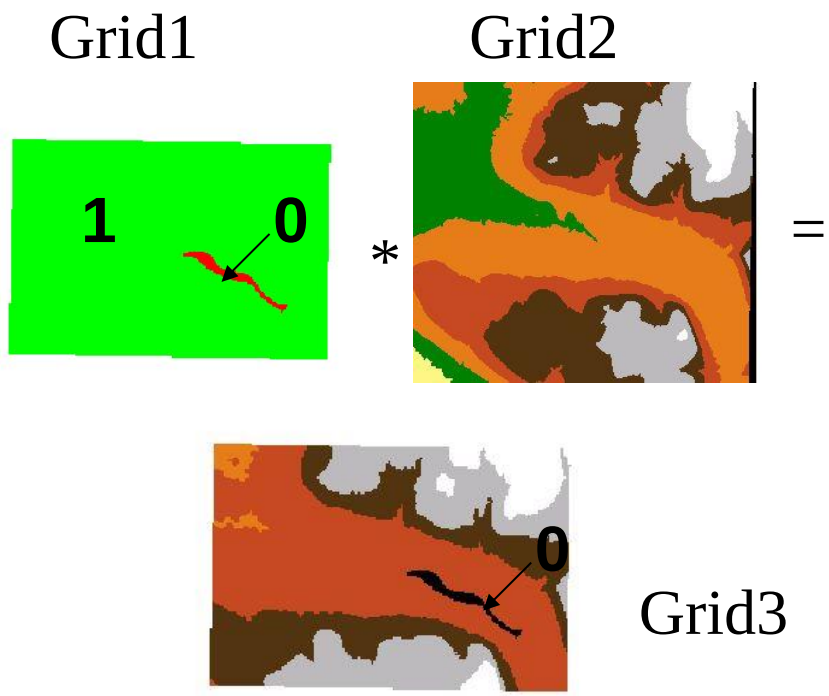




Map Algebra

- Map algebra can be extended to performing a number of mathematical operations.
- some will make sense, others won't.
  - Grid 1: water features = *0*, land values = *1*.
  - Grid 2: reclassified elevation map.
  - Grid 3: *0*'s where water existed (x * 0 = 0), and original elevation value where land existed (x * 1 = x)
  - Or, you can add Grid 1 and Grid 2, but it would be meaningless

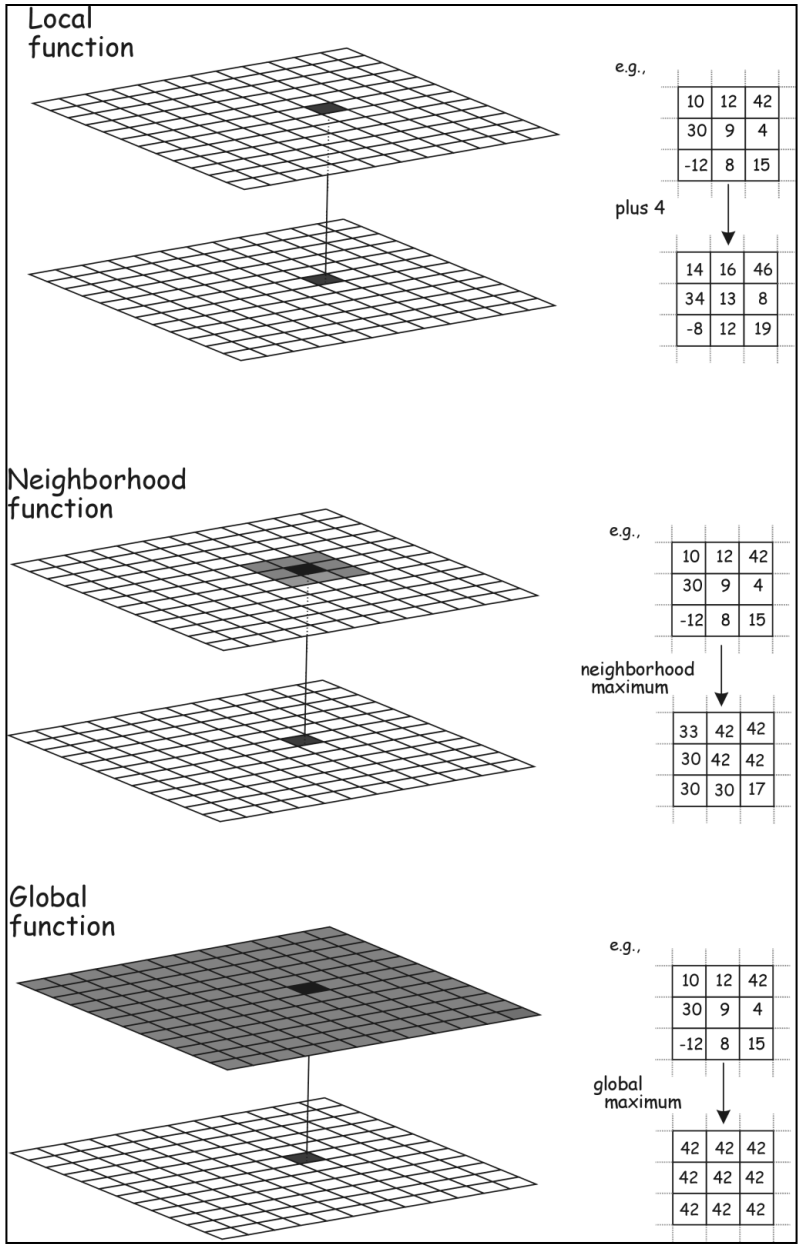
Grid1 * Grid2 = Grid3



Can't build in the 0 area since there is water

Raster Functions

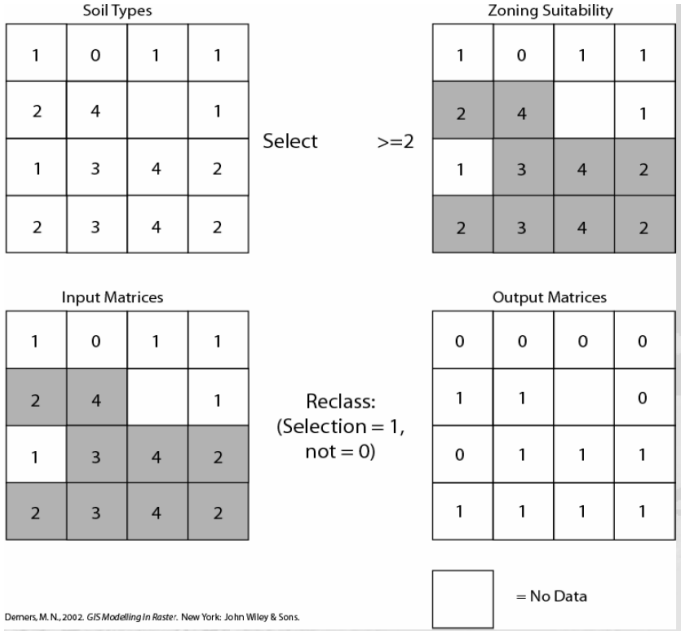
- **Local**: only use data in a single cell to calculate an output value (map algebra)
- **Neighborhood** (Focal): use data from a set of cells, most often a “kernel”
- **Global**: use all data from a raster layer



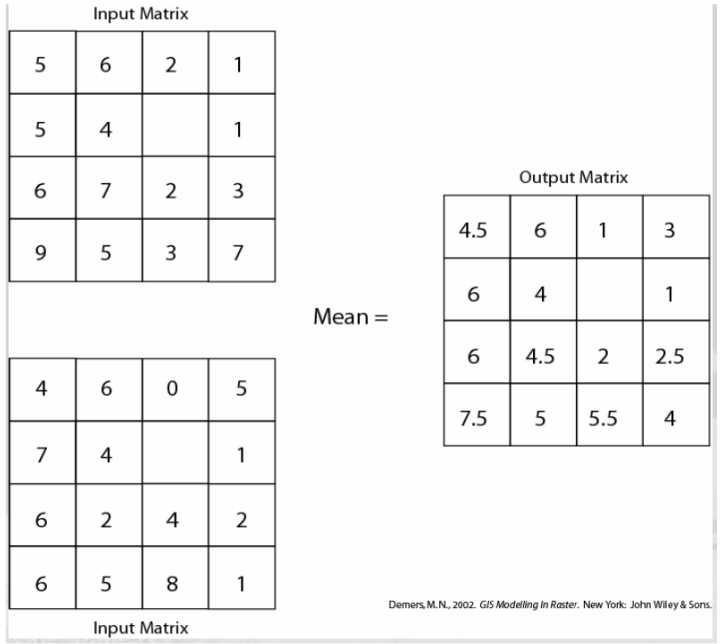
Local Operations

- Based on cell-by-cell analysis: map algebra
- **Reclassification**: Subset of cells chosen for reclassification (i.e. all cells that represent a high level of urbanization, agricultural land use, etc.)
- **Overlays**: comparing 2 or more input grids (minimum, maximum, median, mean, etc values for each grid cell)

Reclassify

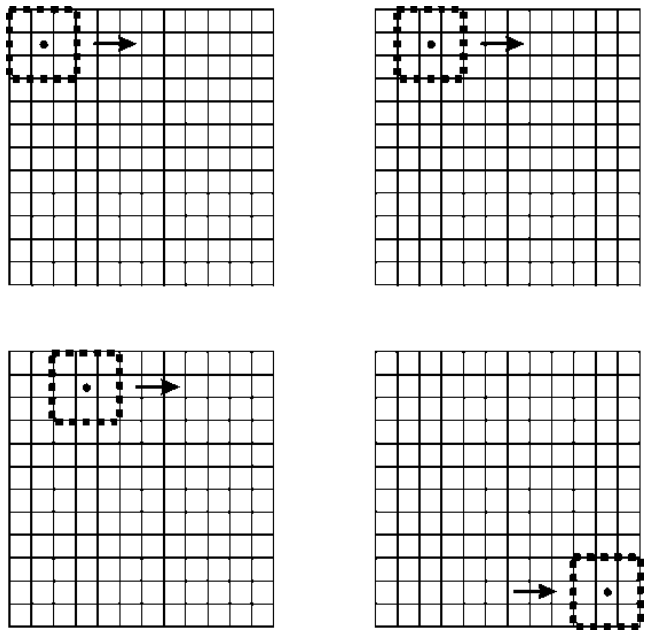


Overlay



Neighborhood Functions

- Also called ‘focal’ functions
- Examine target cell and immediate neighbors
- Neighborhood operators (min, mean, avg, max, etc.) use surrounding cells to assign values



Kernal (Moving window)

Input Matrix

4	7	2	1	9
7	2	3	2	7
3	2	5	3	5
4	1	2	2	4
9	5	4	6	2

FOCALMAJORITY (Grid, Neighbourhood, Rectangle, 3, 3)

Output Matrix

4	7	2	1	9
7	2	3	2	7
3	2	2	3	5
4	1	2	2	4
9	5	4	6	2

Input Matrix

4	7	2	1	9
7	2	3	2	7
3	2	5	3	5
4	1	2	2	4
9	5	4	6	2

FOCALMIN (Grid, Neighbourhood, Rectangle, 3, 3)

Output Matrix

4	7	2	1	9
7	2	3	2	7
3	2	1	3	5
4	1	2	2	4
9	5	4	6	2

Input Matrix

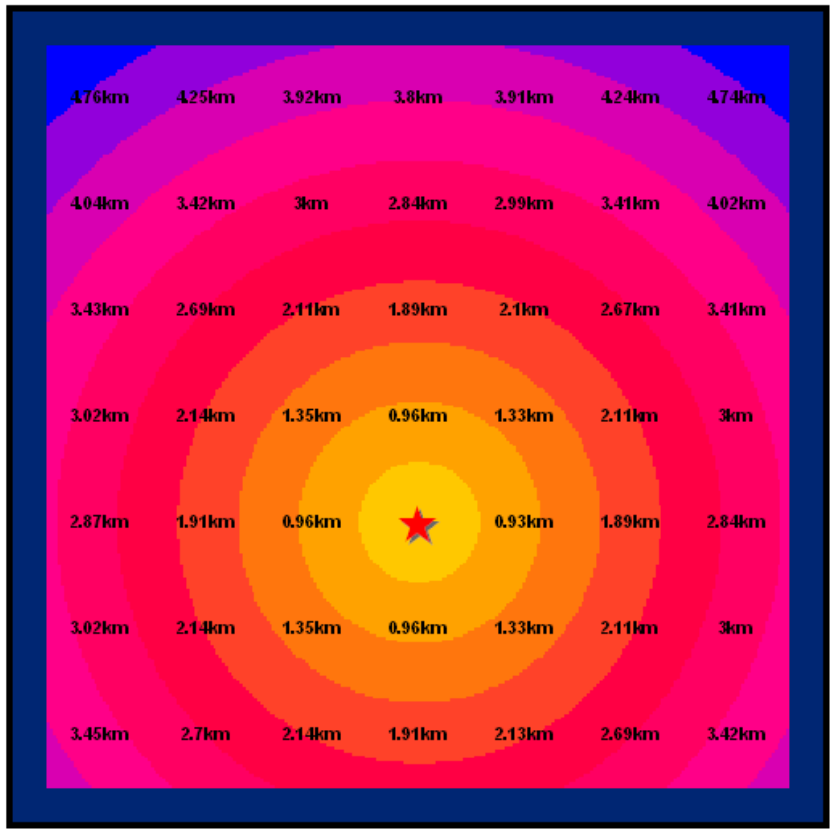
4	7	2	1	9
7	2	3	2	7
3	2	5	3	5
4	1	2	2	4

FOCALMEAN (Grid, Neighbourhood, Rectangle, 3, 3)

Output Matrix

4	7	2	1	9
7	2	3	2	7
3	2	2.4	3	5
4	1	2	2	4

Global functions: Creating Raster Data



- For example, you can create a Raster Dataset that contains the Distance from some feature to every other point in the area of interest.

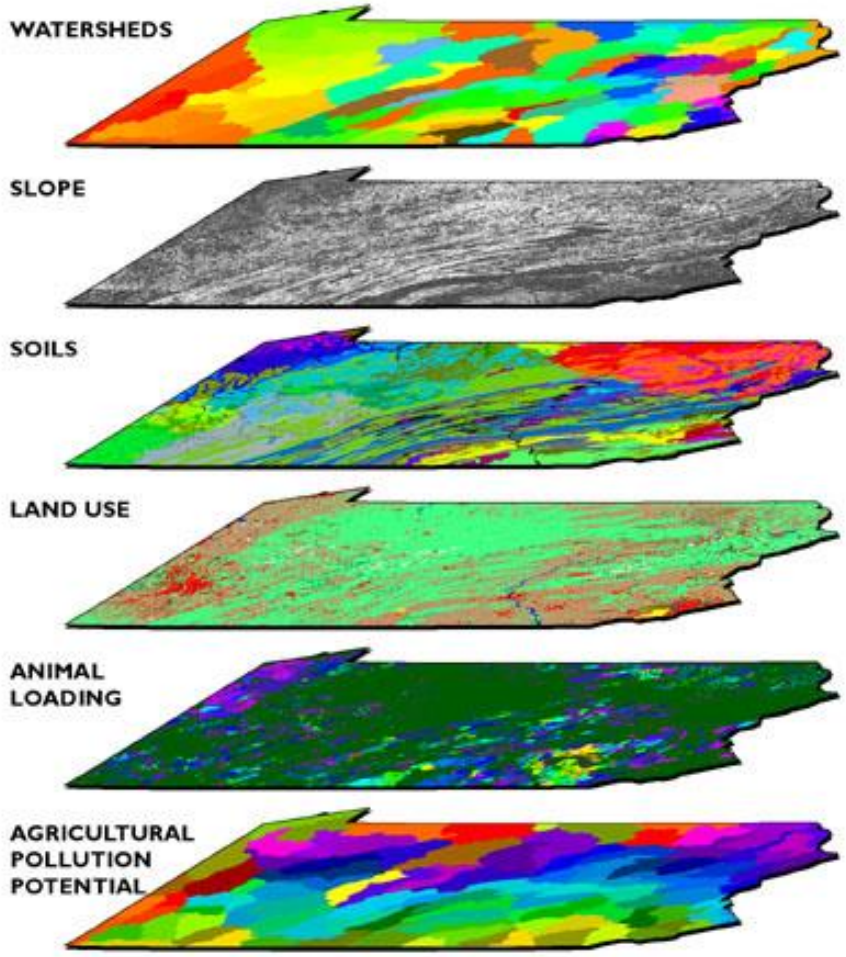
- The distance values can then be extracted to other datasets for use as an attribute for analysis.

Composite Suitability Analysis

- To rank potential sites according to suitability for a proposed activity

Requirements

- A set of “factor” or criteria maps, organized to rate sites relative to one or more characteristics
- A technique for appropriately combining factors



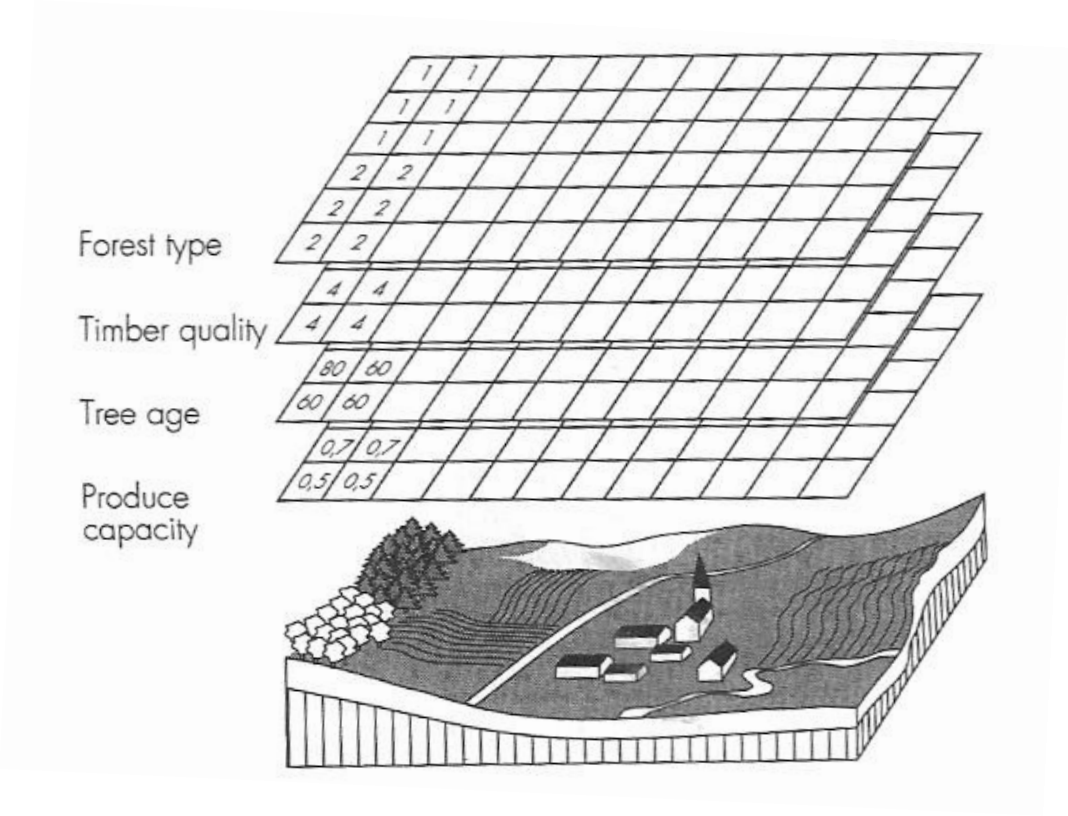
Suitable For Whom?

- Suitability models have a “point of view”
  - Audience can be human
    - “Affordable housing”
    - Best sites for High-end commercial
  - Audience can be environmental
    - Best habitat for black bear
    - Most suitable multispecies conservation areas

Common Units??

Variables have “apples and oranges” characteristics. How do you “combine”

- a map representing “distance to nearest road”
  - Units = meters
- with another representing “land cost”?
  - Units = dollars



Short Answer: find or create common units

- Easiest: Likert scale “preference” units
  - A range of values: 1 to 5, 1 to 9, etc., etc.
  - Polar opposites on both sides of range
    - i.e. “Best”/”Worst”, “Most Suitable”/”Completely Unsuitable”

  - With multiple factors, must make sure that scale consistently applied

    example: We want to be near streams and far from roads, using 1-9 scale with 9 = best
    - Calculate distance to streams, distance to roads
    - Reclassify stream distance to preference units
      - Closest = 9 distance = 0
    - Reclassify road distance to preference units
      - Closest = 0 distance = 9
    - In other words, may need to “flip” values when reclassifying

Weighted rating:

- assigns a weight to each feature under consideration (weight x rating) (Wa x Ra) + (Wb x Rb) + (Wc x Rc)…
- Weights allow us to reflect the relative importance of a combination of features

Direct assignment rating:

- assigns a suitability rating of a land area based on combined data
  - Areas with slope<5% in close proximity to a road: high suitability
  - Slopes 5 – 15% in close proximity to a road: moderate suitability
  - Slopes>15% not suitable, but are considered suitable if they lie within close proximity to a road

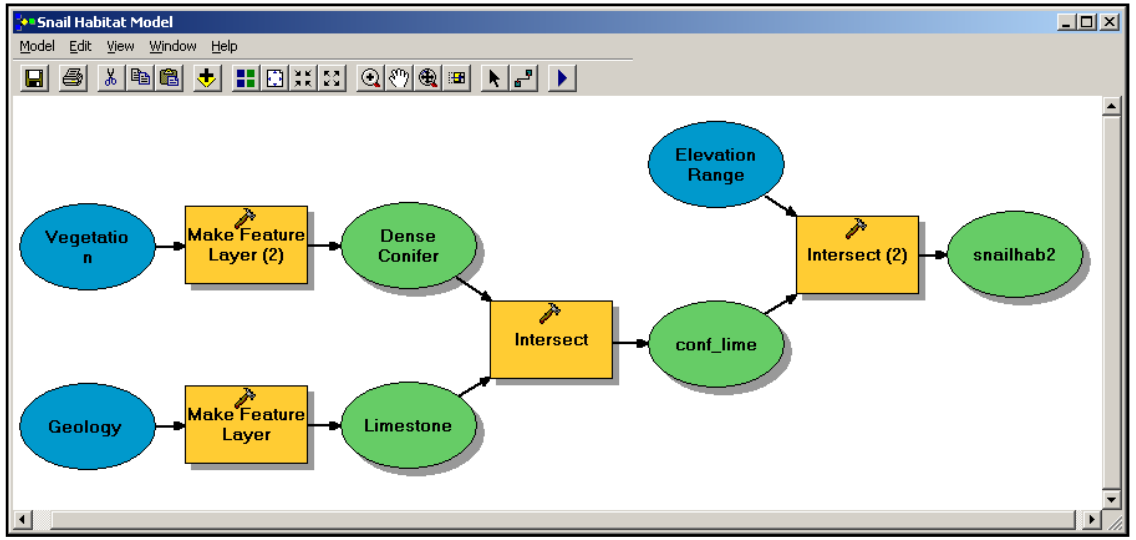
Suitability analysis Criteria and layer creation

Identify a suitable location for an educational facility

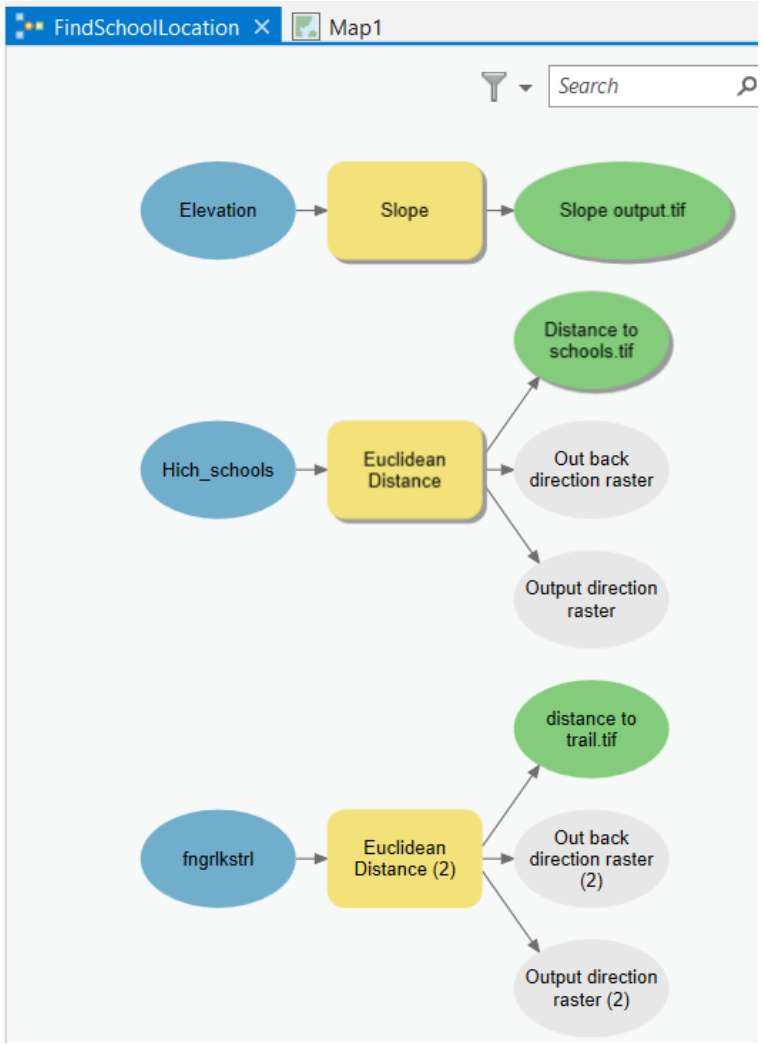
- Distribute high schools evenly (should be far from other schools)
- Provide access to the Fingerlakes trail (closer is better, based on distance to Fingerlakes trail)
- Relatively flat (create slope layer from DEM)
- Appropriate parcel type (vacant, public owned lands) - create raster layer of parcel types

ModelBuilder

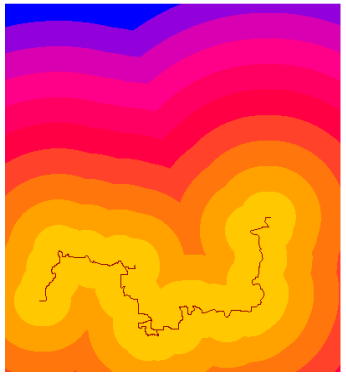
- Create models built from sequences of tools
- Store processing steps for later reference
- Execute models repeatedly with different inputs
- Share models with others



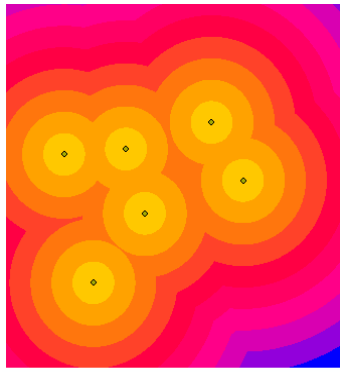
Step 1: Prepping Data…



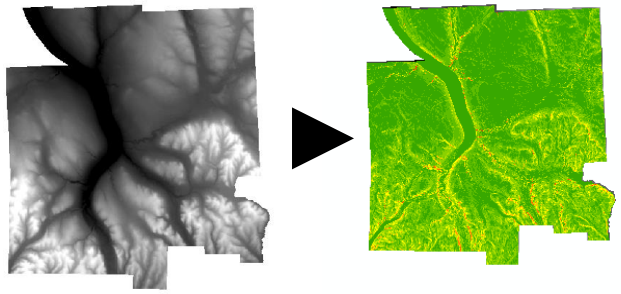
1. Create Schools layer based on Landmarks
2. Distance-based measures:



Create raster depicting Distance from Finger Lakes Trail



Create raster depicting Distance from schools



3. Elevation Model: Create slope based on DEM

Step 1: Prepping Data…

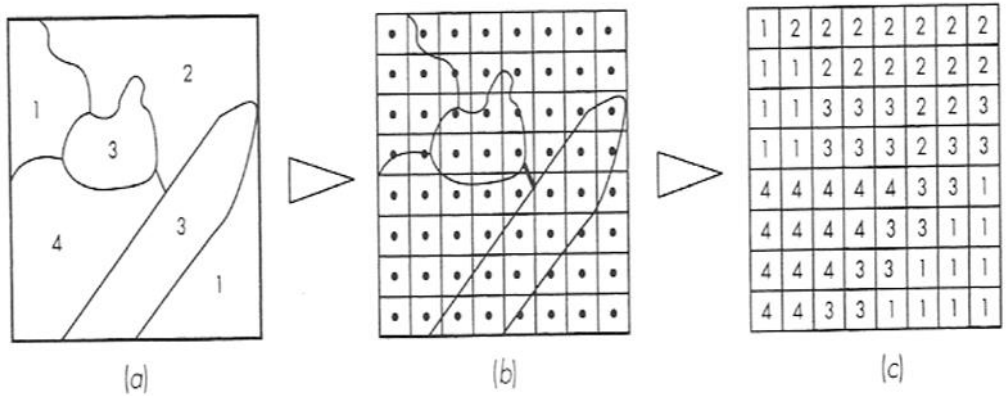
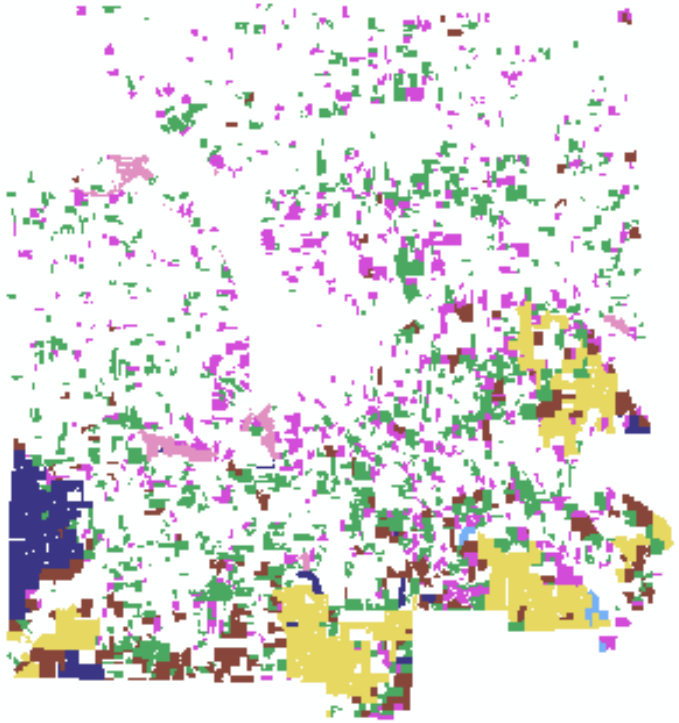
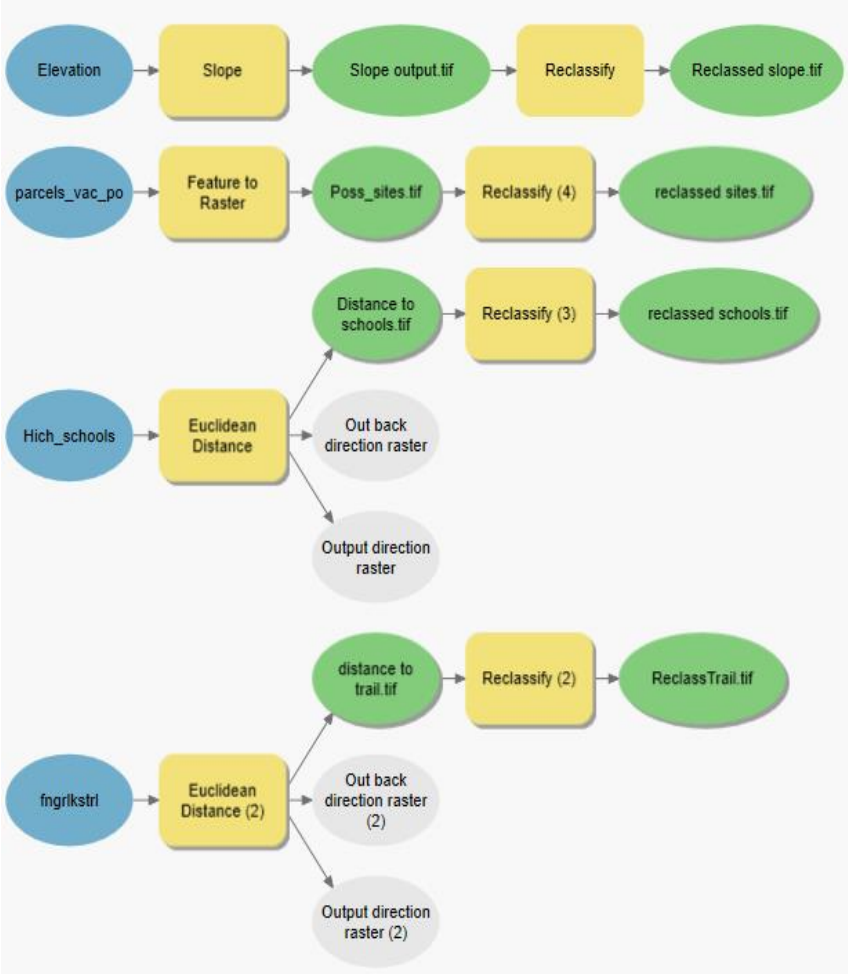


Figure 4.31
Conversion of vector data to raster data: (*a*) coded polygons; (*b*) a grid with the right cell size overlays the polygons (the polygons that contain the center of the individual cells are identified); (*c*) each cell is assigned the attribute code of the polygon to which it belongs.



1. Select out those tax parcels with the appropriate property code (i.e., available for development)
2. Convert vector to raster

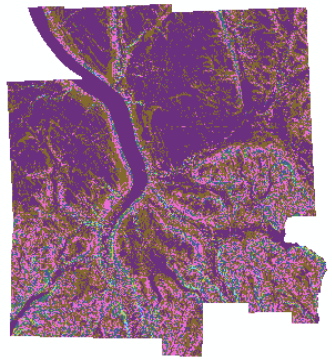
Step 2: Reclassifying Datasets:1 (worse) – 10 (best) ranking



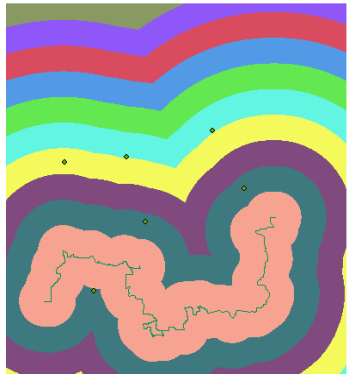
- Reclassify slope → Flatter is better!
- Reclassify distance to trail → Closer is better!
- Reclassify distance to schools → Further is better!
- Reclassify parcel types → Some types are better!

School: 1 - 10

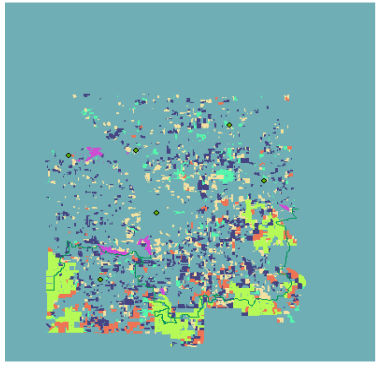
Slope: 1 - 10

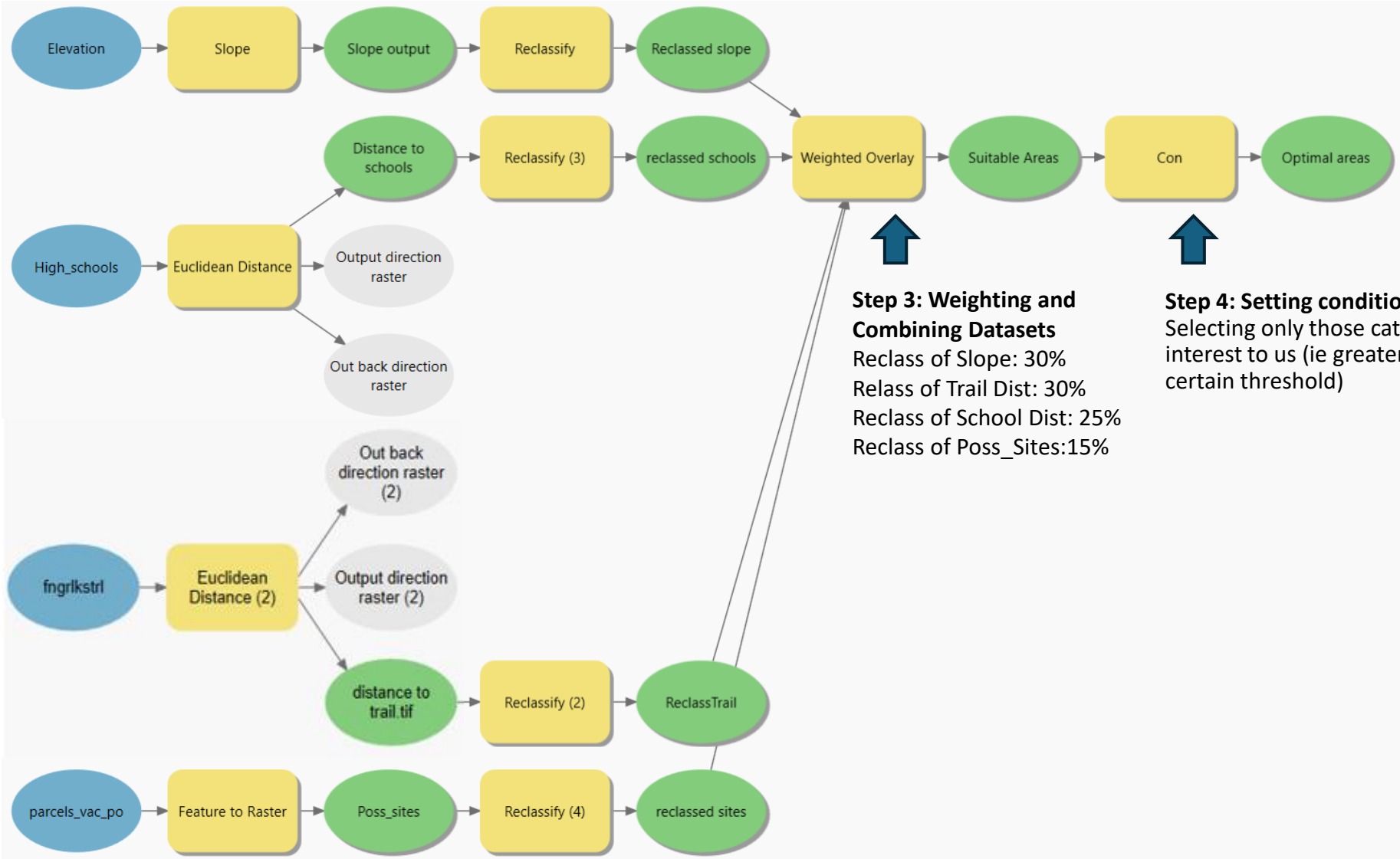


FingerLake Trail: 1 - 10



Parcel types: 1 - 10



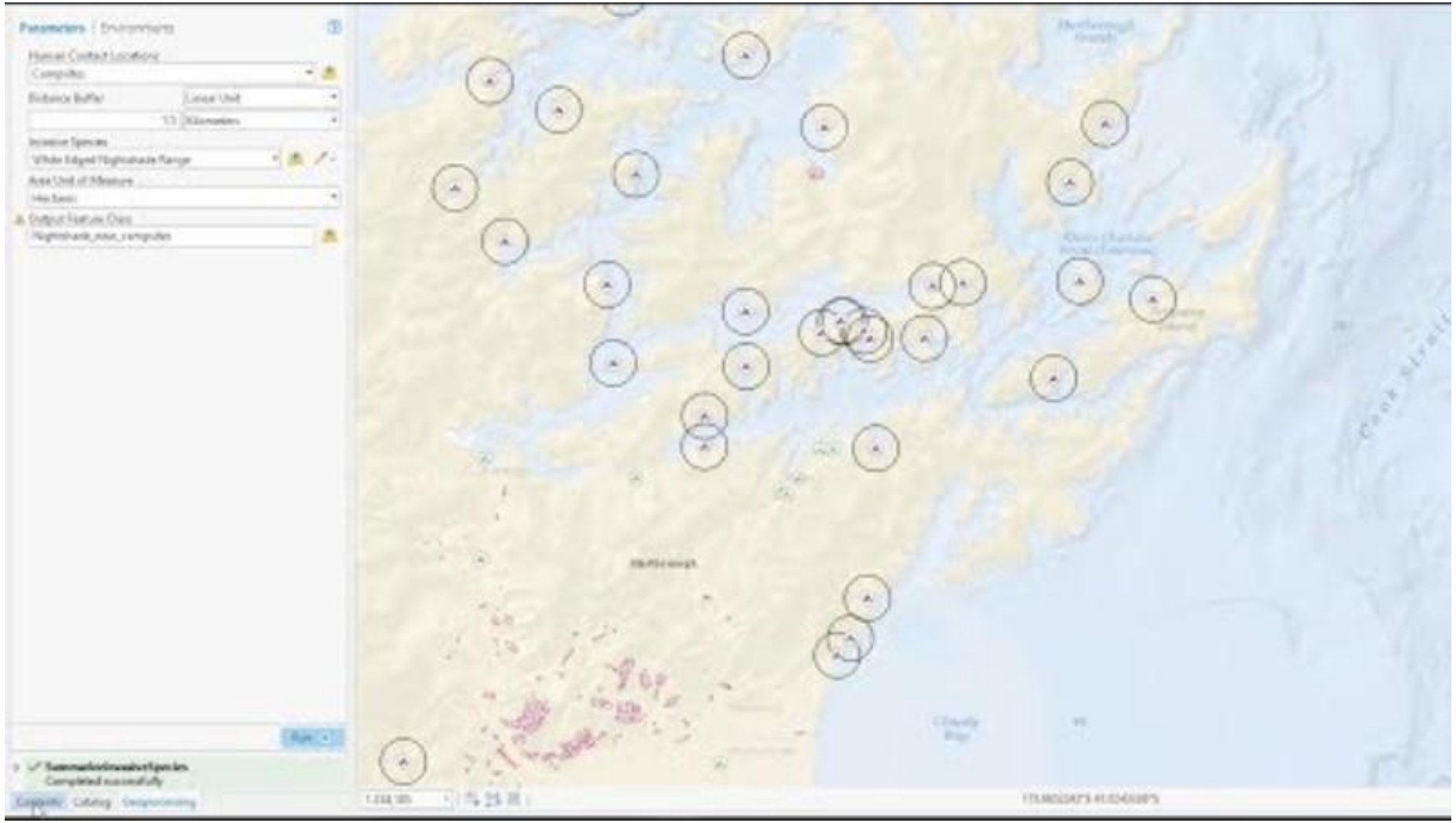


Step 3: Weighting and Combining Datasets
Reclass of Slope: 30%
Relass of Trail Dist: 30%
Reclass of School Dist: 25%
Reclass of Poss_Sites:15%

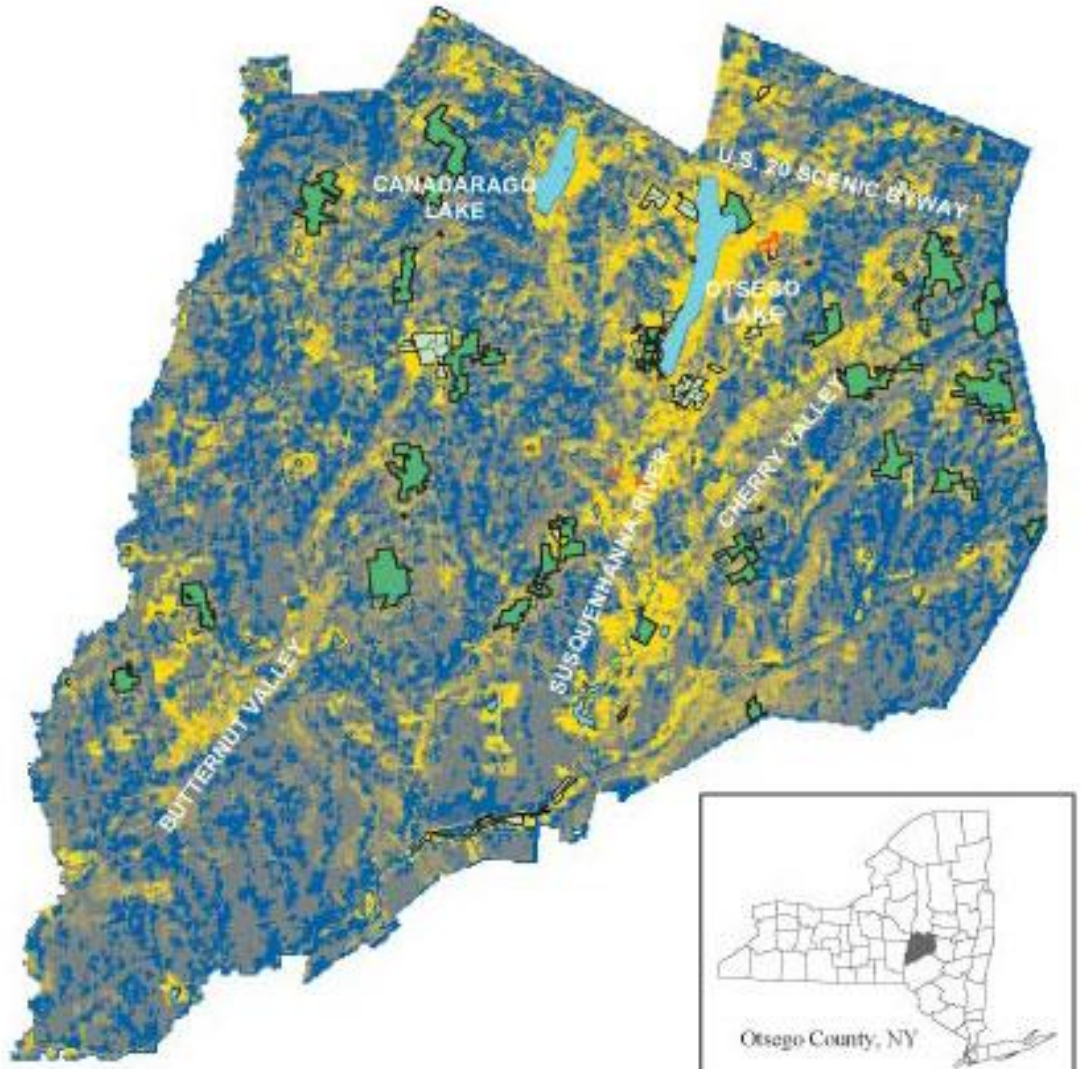
Step 4: Setting conditionals:
Selecting only those categories of interest to us (ie greater than a certain threshold)

Make a Geoprocessing Model in ArcGIS Pro

Make a Geoprocessing Model in ArcGIS Pro (youtube.com)



Map 21: Land Protection Suitability Analysis



Otsego County, New York

OTSEGO LAND TRUST, INC.

CLT Parcels

All Protected Parcels

Overlay Ranking

Value

High : 15

Low : 0

Test Parcels

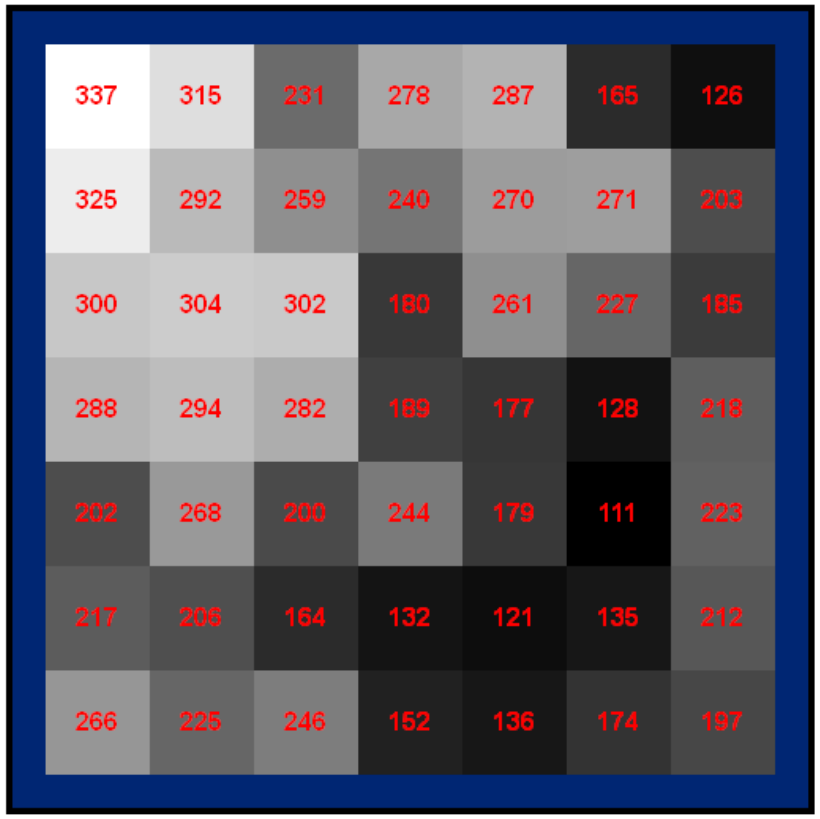
0 2.5 5 10

Miles

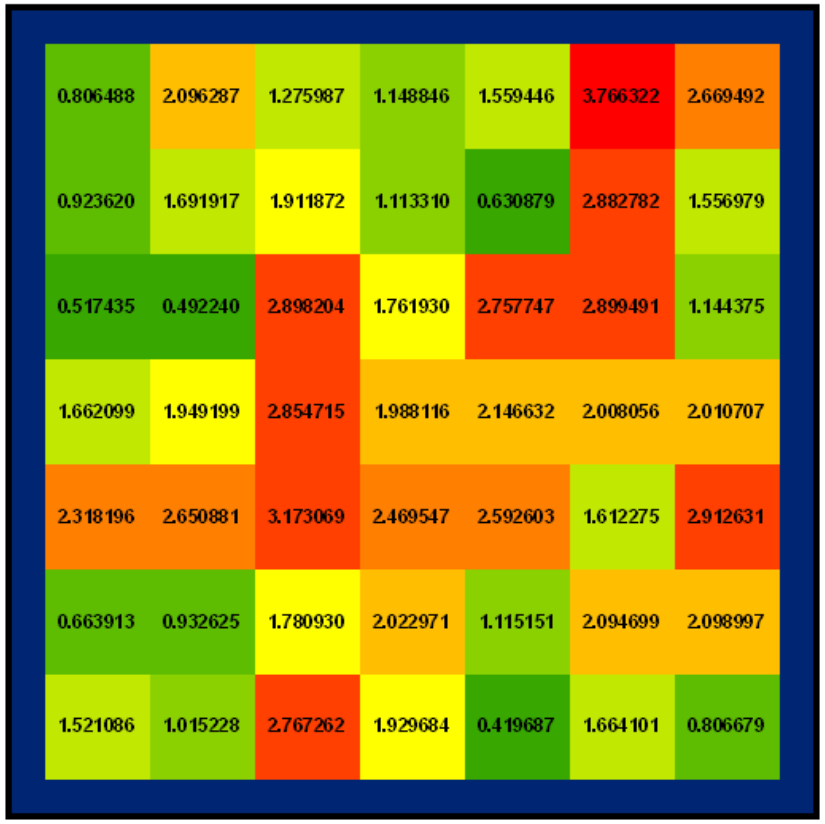
Sources: Test parcel boundaries, Protected parcels, Otsego Lakes provided by Otsego Land Trust, Inc. Suitability Analysis overlay created by CRP 558, Spatial Analysis Team, Cornell University, November 2004.

Cornell University

Focal function:



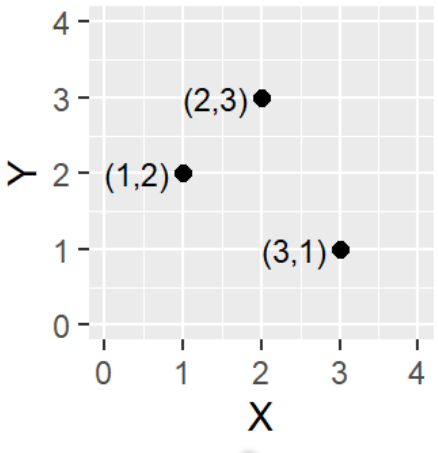
Digital Elevation Model (DEM)



Slope Raster contains the **Maximum change in elevation** for every elevation point, measured in degrees.

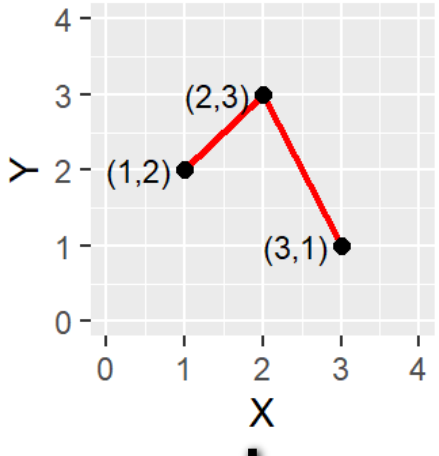
Types of Spatial Data – vector data

- Points



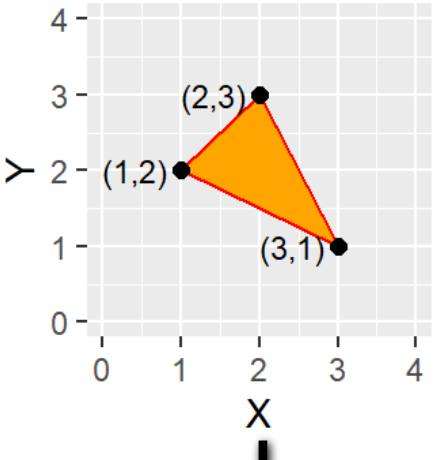
Points – tree planting

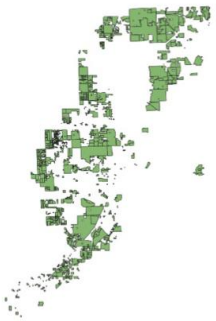
- Lines/polylines



Lines – Bike lanes

- Polygons





Polygon – Special tax district

Creating your model…

Selecting your inputs:

Layers – yellow diamond

Variables - blue arrows – shared between processes