# Lecture 14

Map Structures and Attributes: Vectors and Rasters

**GEOG 489** 

SPRING 2020

# Quiz 5

```
Write a function to remove "NA" from a string
For example:
Input: x <- "Programming NA GIS NA"
Output: y <- "Programming GIS "
RMNA <- function(x){
 y = gsub("NA", "", x)
 return(y)
RMNA <- function(x){
 y = paste(unlist(strsplit(x, "NA")), collapse = ")
 return(y)
x <- "Programming NA GIS NA"
RMNA(x)
```

# Assignment 3

Your goal is to write a function that calculates a set of descriptive statistics for a data frame, and return the output in a list format.

Input: x=data frame, probs=quantile probabilities, na.rm (logical)

**Output:** list, one list component per column, named after the data frame columns Sublist components (the components should be named "mean", "median", "sd", and "quantiles"):

```
# mean: the column's mean
```

- # median: the column's median
- # sd: the column's standard deviation
- # quantiles: a matrix of probs vs. the quantile of x for those probs.
- # The matrix should have one prob/quantile per row.
- # The first column should be named "prob" and the second "quantile".

Assignment 3 will be due by the end of today (March 10)

# **Basic Definitions: GIS**

### Geographic Information System (GIS):

- 1. A database.
- 2. Spatial information.
- 3. A way of linking the database and spatial information and using these links.

# **Data Models**

GIS uses two basic data models:

- Vector
- Raster

LEARN THE DIFFERENCE BETWEEN VECTOR AND RASTER.

# Features and Maps

- A GIS map is a scaled-down digital representation of point, line, area, and volume features.
- Most GIS can handle raster AND vector.
- However, only one at a time can be used for the internal organization of spatial data.
- Different display and analysis functions favor one or the other model; i.e. "vector functions" vs. "raster functions".

# **Basic Definitions: GIS**

### Geographic Information System (GIS):

- 1. A database.
- 2. Spatial information (vector or raster data model).
- 3. A way of linking the database and spatial information and using these links.

# Vector

Definition: "A map data structure using the point or node and the connecting segment as the basic building block for representing geographical features."

## **Vector Data Model**



• Entities: spatial information is described by coordinates.

• Attributes: descriptions of each of the features.

# **Vector Data Model**

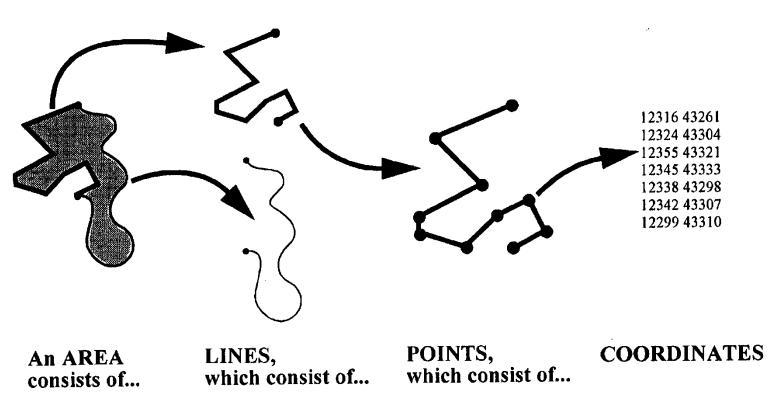
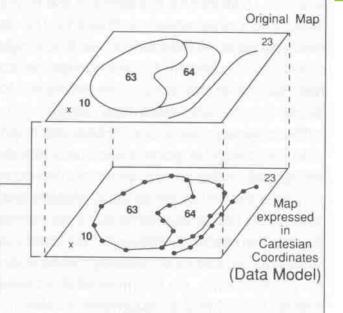


Figure 2.16 Geographic information has dimension. Areas are two-dimensional and consist of lines, which are one-dimensional and consist of points, which are zero-dimensional and consist of a coordinate pair.

# THE "SPAGHETTI" DATA MODEL



#### **Data Structure**

FEATURE	NUMBER	LOCATION				
Point	10	X Y (Single Point)				
Line	23	$X_1 Y_1, X_2 Y_2, \dots, X_n Y_n$ (String)				
Polygon	63	$X_1 \ Y_1, X_2 \ Y_2 \ \dots \ X_1 \ Y_1$ (Closed Loop)				
rolygon	64	X <sub>1</sub> Y <sub>1</sub> , X <sub>2</sub> Y <sub>2</sub> , X <sub>1</sub> Y <sub>1</sub> (Closed Loop)				

# graphic Spaghetti

**Points** represented by an ID and a single coordinate

Lines represented by an ID and a set of ordered coordinates

**Polygons** represented by an ID and a set of ordered coordinates, and the last coordinate is the same as the first coordinate.

# **Vector Data Model**

Туре	Examples of Graphic Representation	Digital Representation				
Point	• * + <sup>Δ</sup> O	Coordinates: (x,y) in 2D; (x,y,z) in 3D				
Line	2005 ( )	(i) Ordered list of coordinates (chain) (ii) Mathematical function				
Area	090	(i) Line in which the first point equals the last (ii) Set of lines if an area has holes				
Surface		(i) Matrix of points (ii) Triangulated set of points (TIN) (iii) Mathematical functions (iv) Contour lines				
Volume		Set of surfaces				

# **Basic Definitions: GIS**

### Geographic Information System (GIS):

- 1. A database (attribute data).
- 2. Spatial information (vector or raster data model).
- 3. A way of linking the database and spatial information and using these links.

## **Attribute Data**

Definition: a characteristic of a feature that contains a measurement or value for the feature.

Can be labels, categories, numbers, dates, etc.

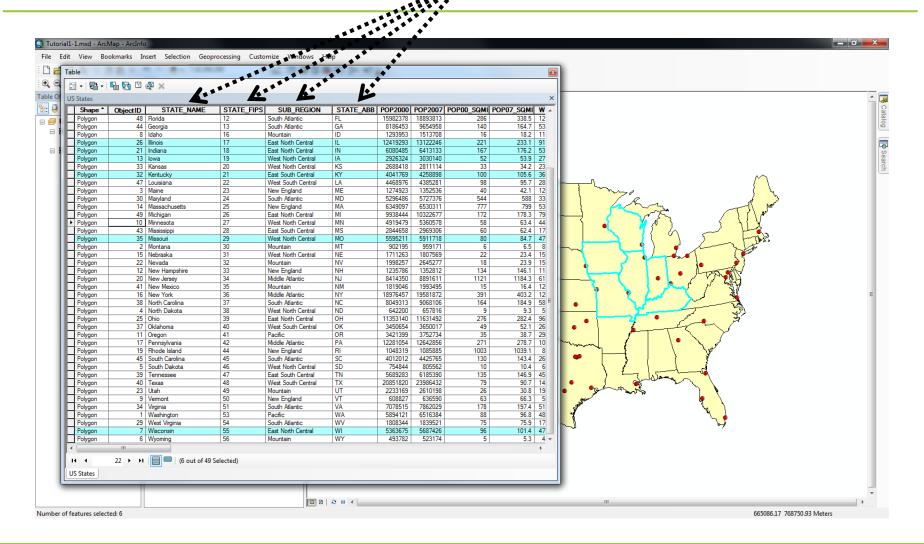
"Spreadsheet" definition: a column in a table.

The NON-SPATIAL part of a GIS feature.

Some (almost) synonyms you will hear:

- Spreadsheet
- Table
- Flat file
- Matrix

# Attribute Data



# Attribute Data Storage

Attribute data is often stored separately from the spatial data.

Simple attribute data is often stored as a table.

Complex attribute data is often stored in a database.

# **Basic Definitions: GIS**

### Geographic Information System (GIS):

- 1. A database (attribute data).
- 2. Spatial information (vector or raster data model).
- 3. A way of linking the database and spatial information and using these links.

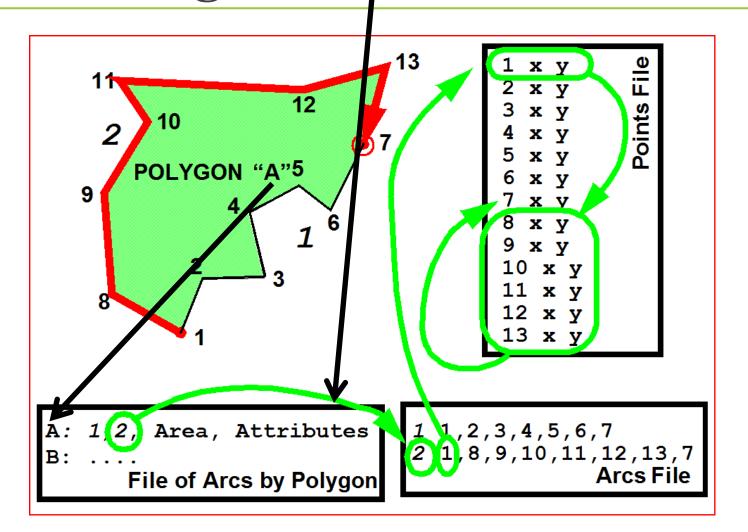
# Linking Attributes to Arc/Node

Recall that each vector type has a "feature ID" (aka "FID"):

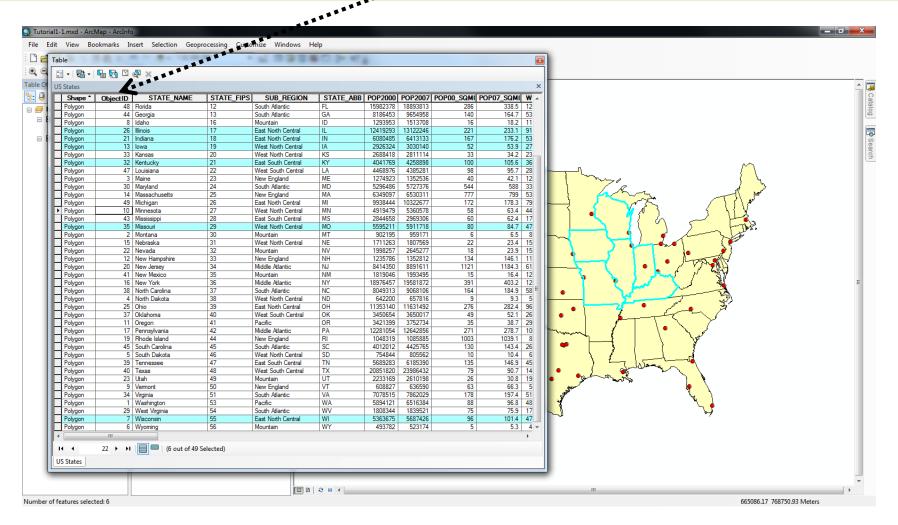
- Point ID
- Line ID
- Polygon ID
- Triangle ID (TINs)

These feature IDs can be used to both link points to arcs to polygons, but also to attribute data via the ObjectID field.

# Linking Attributes to Arc/Node



# Linking Attributes to Arc/Node



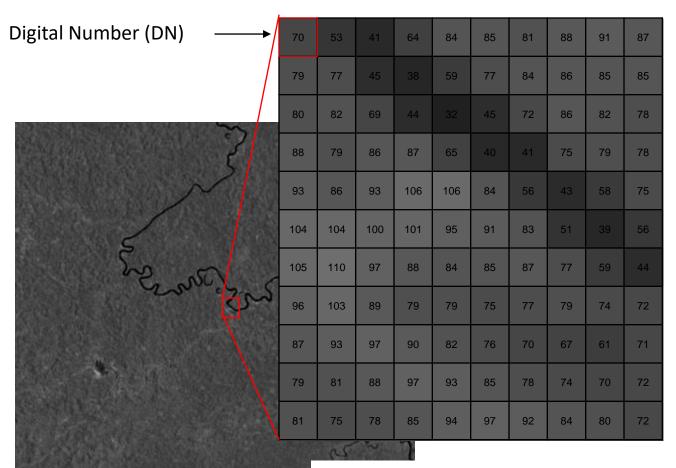
The ObjectID is a unique value for each feature in a layer.

### Rasters...

### First, some synonyms:

- Rasters
- Grids
- Images
- Matrices
- Arrays
- Digital images or digital photographs

# What Is A Digital Image?



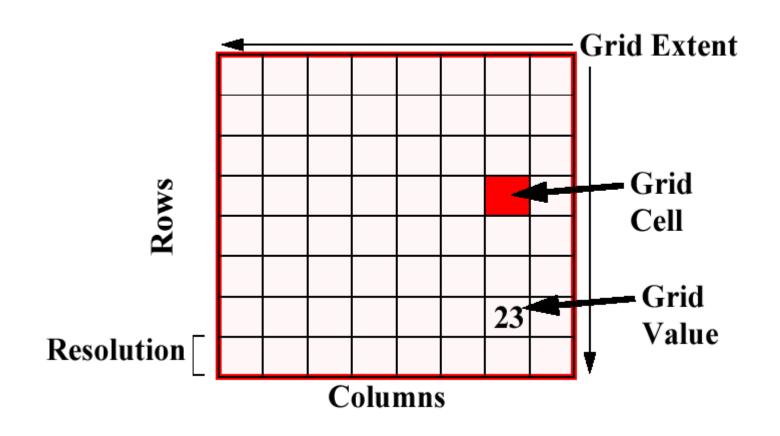
Digital numbers (DNs) typically range from 0 to 255; 0 to 511; 0 to 1023, etc. These ranges are binary scales: 28=256; 29=512; 210=1024.

What your computer sees...

# Rasters (Grids)

- Rasters are rectangular (fixed number of columns and rows).
- Rasters can have several "bands" (layers) to add a third dimension.
- The "building block" of a raster is the **cell** (aka "pixel").
- Each cell (pixel) represents an area.
- Each cell holds one numerical attribute.
- Every cell in the raster has a value, even if the value is "missing".

# Structure of a 2-D Raster (Grid)



# Some Parameters of a Raster

Number of columns (ncol) and number of rows (nrow)

**Number of cells**= ncol x nrow

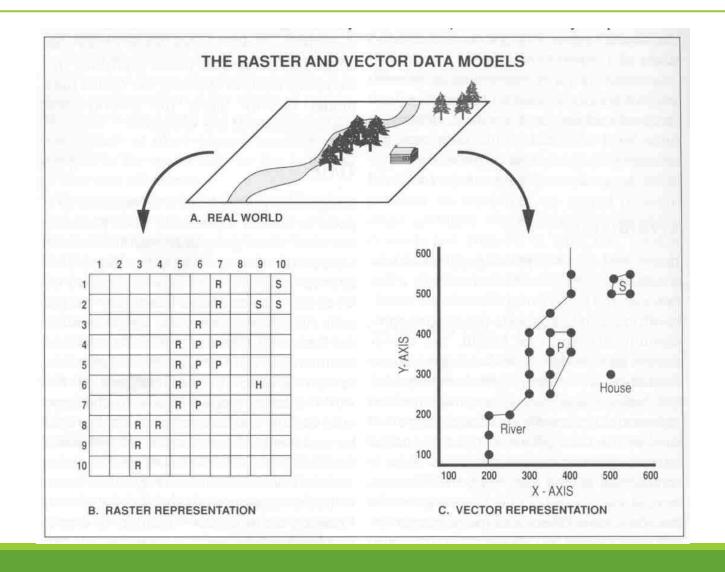
**Cell size** = the **resolution** of the cell.

- Either described in terms of area of the cell (e.g. m<sup>2</sup>)...
- ...or in terms of a length if the cell is square (e.g. m)

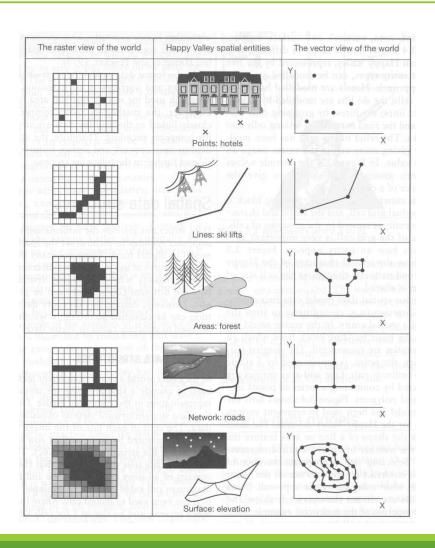
#### **Extent:**

- Either the total area of the raster (number of cells \* cell area)...
- ...or the corner coordinates of the raster (the **bounding box**).

# Raster vs. Vector



### Raster vs. Vector



Information must be coerced to a raster grid cell.

Points, lines and areas are not explicitly stored; they are implicit.

# Rasters: the Good and the Bad

#### The Good:

Represent surfaces well.

Localized topology.

Scanned or remotely sensed data is "natively" raster.

Easy to read/write.

Easy to display.

Maps directly onto programming structures (array); many raster operations are easily parallellized.

Can be highly compressed.

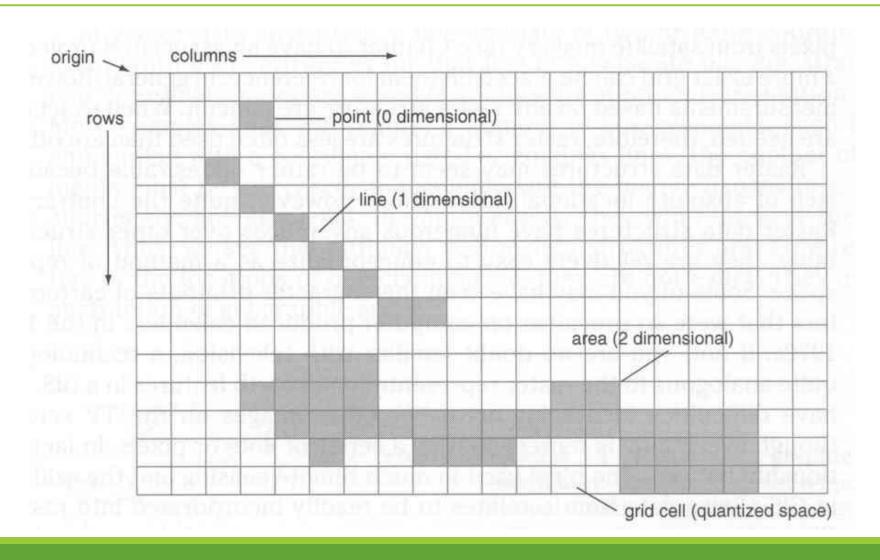
#### The Bad:

Represent points, lines and areas poorly.

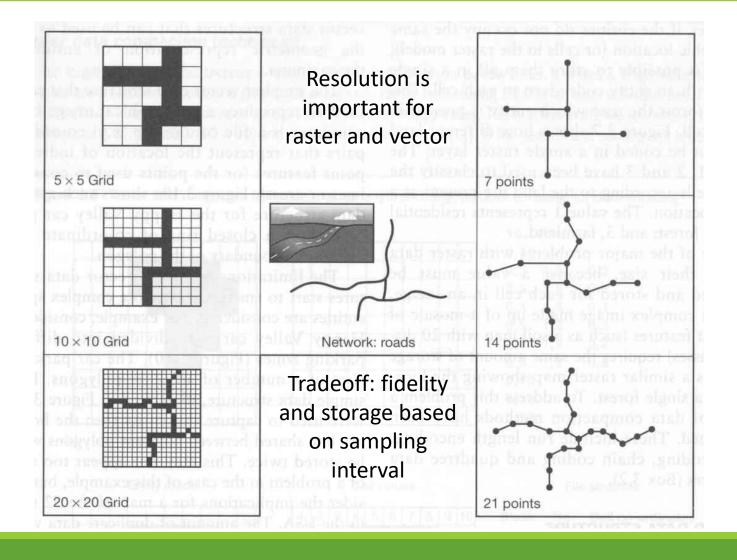
Mixed pixel problem.

Lots of redundant or missing data that must be stored.

# Rasterizing Points, Lines and Areas



# **Spatial Resolution**



# 3-d Rasters

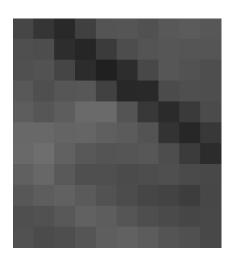
- Rasters can store multiple "bands".
- Each band has the same number of rows and columns (and data type, pixel size, etc.)
- The header will add a new parameter: **number of bands**
- A digital camera stores three bands, representing "red", "green" and "blue".
- Some remote sensing devices can collect over 200 (!) different colors and store them in a single raster image.

# 3-d Rasters

70	53	41	64	84	85	81	88	91	87		bang	٧,
79	77	45	38	59	77	84	86	85	85	87		3
80	82	69	44	32	45	72	86	82	78	85	87	
88	79	86	87	65	40	41	75	79	78	78		
93	86	93	106	106	84	56	43	58	75	78		
104	104	100	101	95	91	83	51	39	56	75		
105	110	97	88	84	85	87	77	59	44	56		
96	103	89	79	79	75	77	79	74	72	44		
87	93	97	90	82	76	70	67	61	71	72		
79	81	88	97	93	85	78	74	70	72	71		
81	75	78	85	94	97	92	84	80	72	72		
	81	75	78	85	94	97	92	84	80	72		
		81	75	78	85	94	97	92	84	80	72	

rows (y)

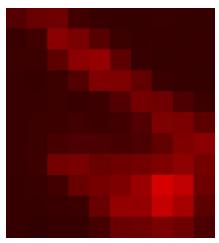
# Greyscale vs. RGB

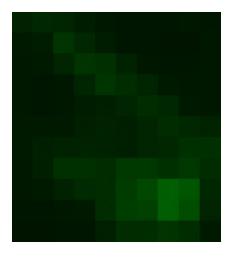


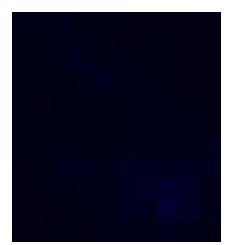
**Greyscale** is typically used to display a *single band*...

...while **RGB** ("Red", "Green", "Blue") images can display *3* bands, corresponding to the red, green and blue phosphors on a monitor. Computer monitor colors are additive, meaning "true" red + green + blue = white.









### Raster Attributes

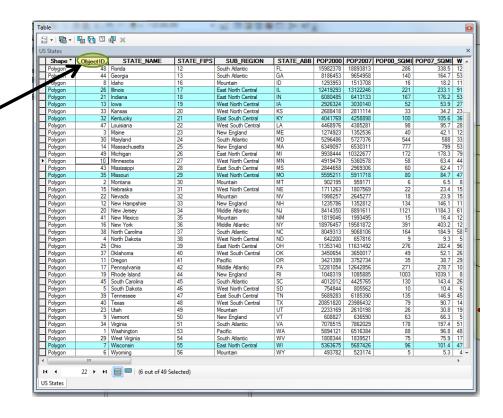
Rasters can be linked to attribute tables just like vectors.

The raster cells must represent "IDs".

Often, however, rasters encode a single attribute without the need for a separate attribute table.

## Raster Attributes

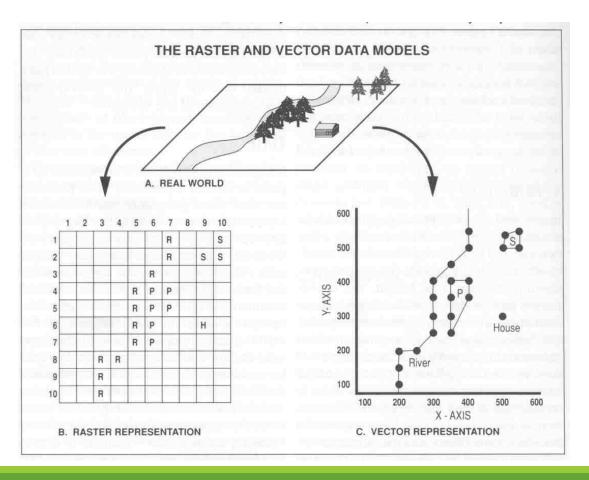
70	53	41	64	84	85	81	88	91	87
10	55		04	04	00	0.1	00	91	01
79	77			59	77	84	86	85	85
80	82	69	44			72	86	82	78
88	79	86	87	65			75	79	78
93	86	93	106	106	84	56	43	58	75
104	104	100	101	95	91	83	51		56
105	110	97	88	84	85	87	77	59	44
96	103	89	79	79	75	77	79	74	72
87	93	97	90	82	76	70	67	61	71
79	81	88	97	93	85	78	74	70	72
81	75	78	85	94	97	92	84	80	72



Rasters can be linked by their cell values

# Review of Rasters vs. Vectors

Both formats are abstractions of the real world.



Both formats store information about their components' positions in two-dimensional space (coordinates).

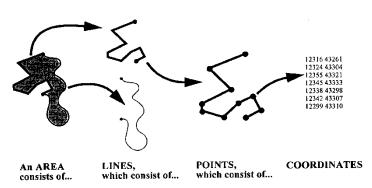
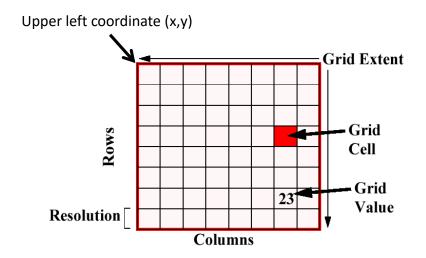


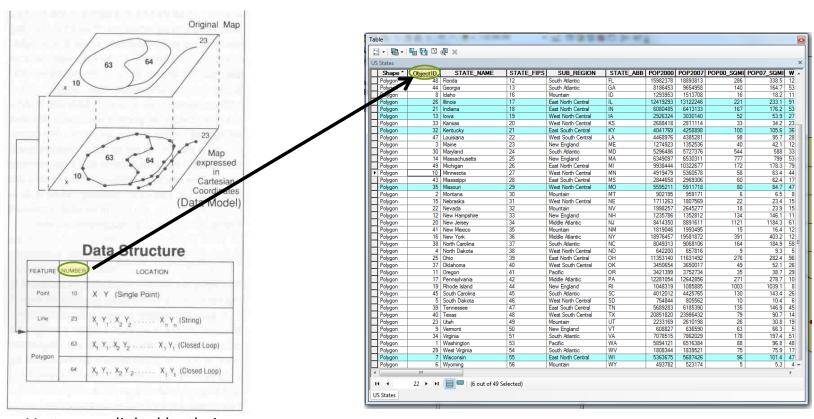
Figure 2.16 Geographic information has dimension. Areas are two-dimensional and consist of lines, which are one-dimensional and consist of points, which are zero-dimensional and consist of a coordinate pair.

Vector coordinates are determined by knowing the coordinates of the nodes.



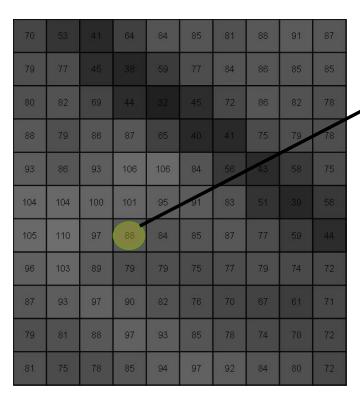
Grid cell coordinates are determined by knowing the coordinates of the upper left corner and the cell resolution.

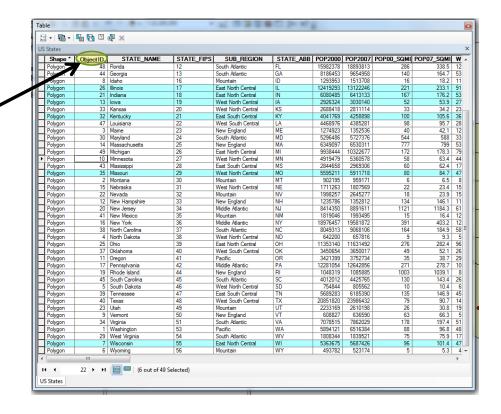
Both formats can be linked to an attribute table.



Vectors are linked by their feature IDs (FIDs)

Both formats can be linked to an attribute table.

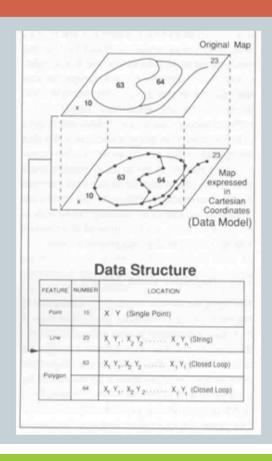


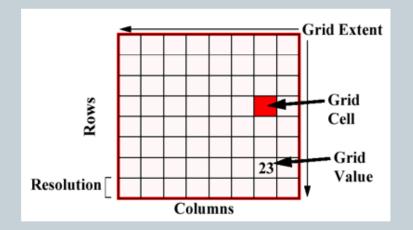


Rasters are linked by their cell values

The building block of a vector is the node.

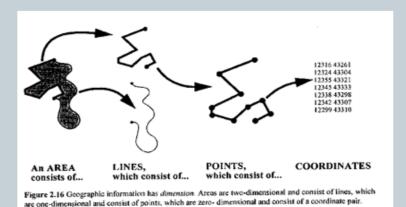
The building block of the raster is the cell.





Vectors store positions precisely

Raster positions are "snapped" to the cell



Grid Extent
Grid Cell

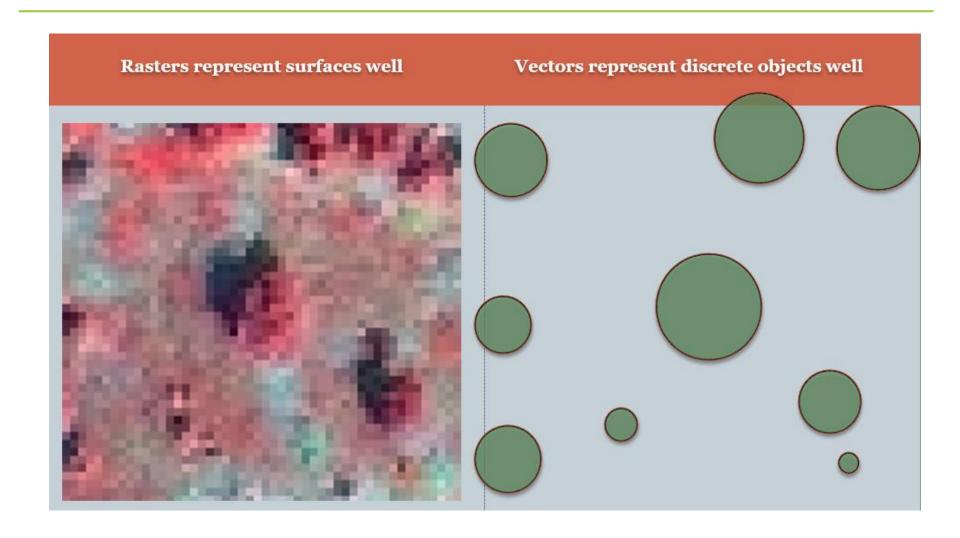
Columns

Grid

Value

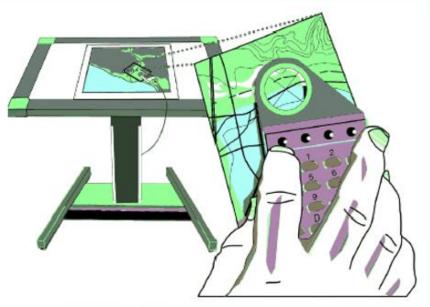
Upper left coordinate (x,y)

Resolution



Vectors often come from digitizing

Rasters often come from remote sensing

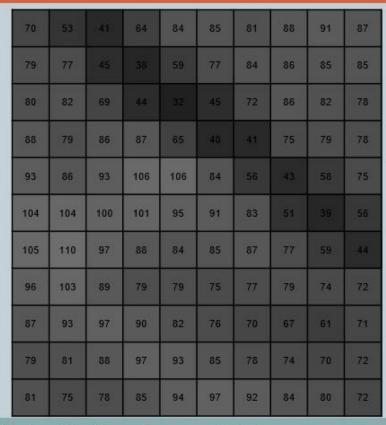


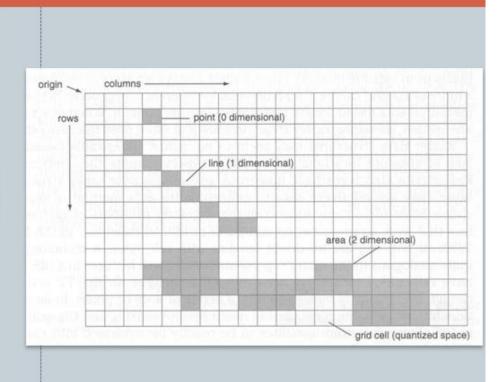
- Digitizer cursor transmits a pulse from an electromagnetic coil under the view lens.
- 2. Pulse is picked up by nearest grid wires under tablet surface.
- 3. Result is sent to computer after conversion to x and y units.



#### Vectors can precisely represent rasters

#### Rasters cannot precisely represent vectors





Each grid cell can be converted to a square polygon.

All vectors must be coerced to grid cells.

Vector formats are relatively complex

Raster formats are relatively simple

Multiple files
 representing arcs,
 nodes, polygons,
 topology, projection,
 and attribute
 information.

- Header file
- Raster data

#### Vector - HPC more difficult

Raster - HPC simple

 High performance computing (HPC) with vectors requires advanced topological and database algorithms (complex). • High performance computing (HPC) with rasters usually just requires sending "chunks" of the raster to different processors (simple).