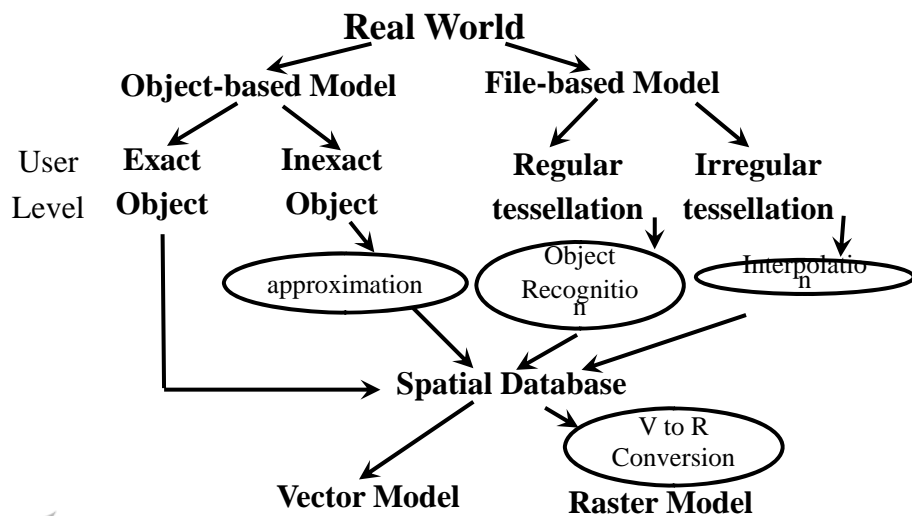


# GIS Data: Data Models

## Topics/Learning Objectives

- Methods used in digital representation of GIS data (GIS Data Models)
  - Raster
  - Vector
- The role of databases in GIS
- Data collection methods
- Storage, editing and management of GIS data

## Hierarchy of Data Model

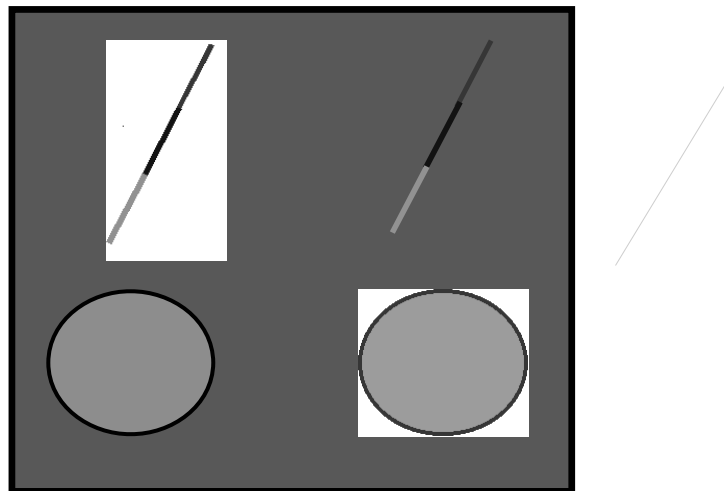


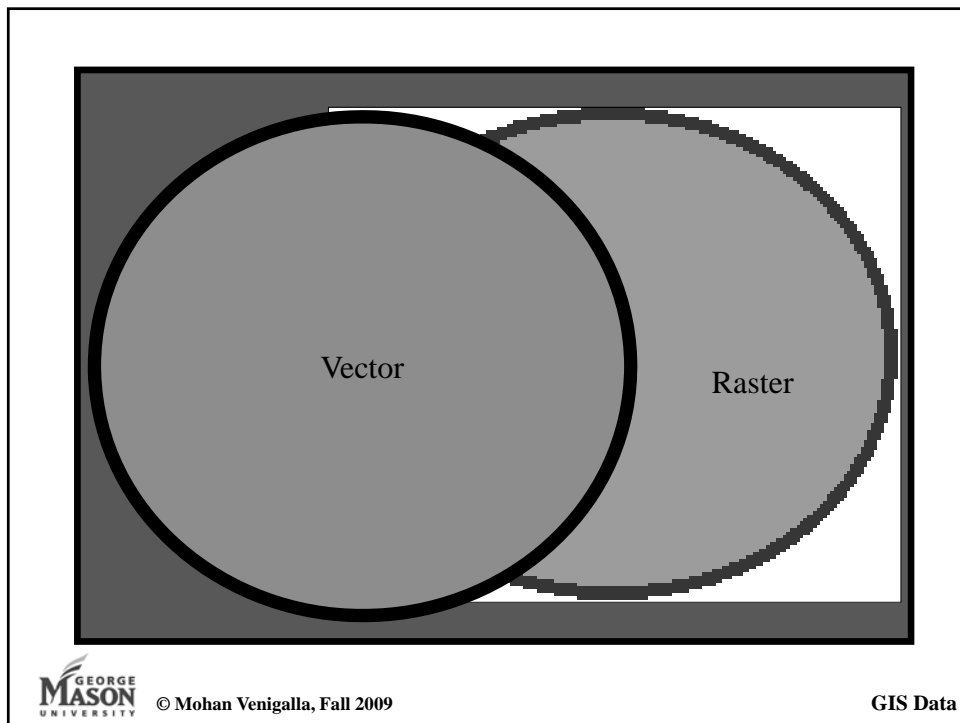
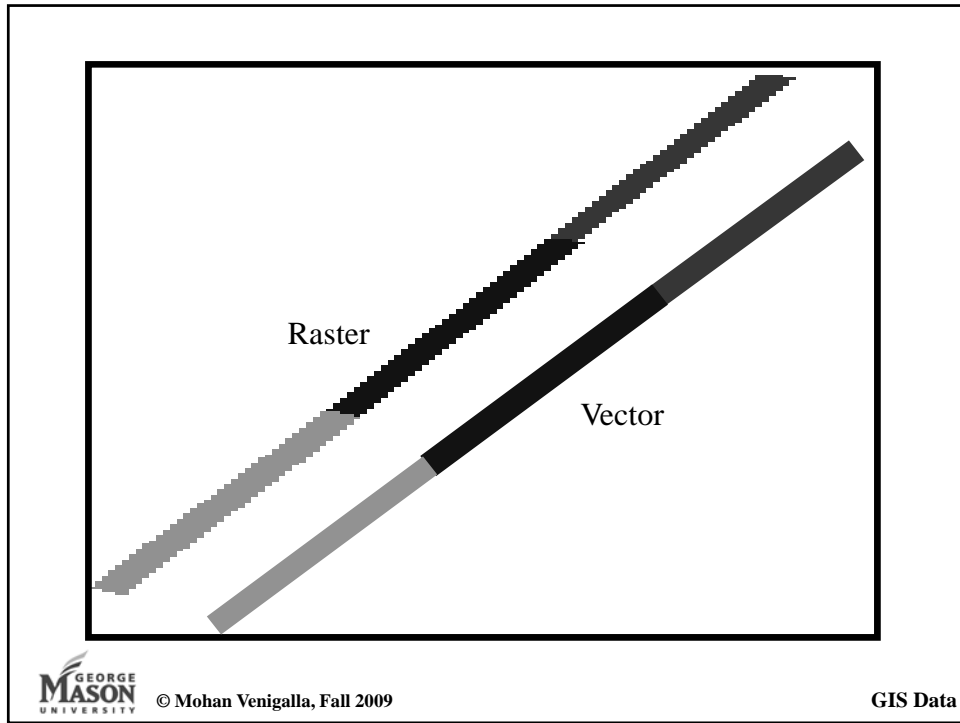
# GIS Data Models

## ➤ Two primary data models

- Vector Model
  - Edges and vertices are defined by a series of coordinate pairs:  $(x,y)$ ;
  - Intermittent lines are sprightliness or smoothed curves
- Raster Model
  - Map area is divided into grid cells
  - Each cell will contain a value

## Which is Raster & Which is Vector





## Digital Images



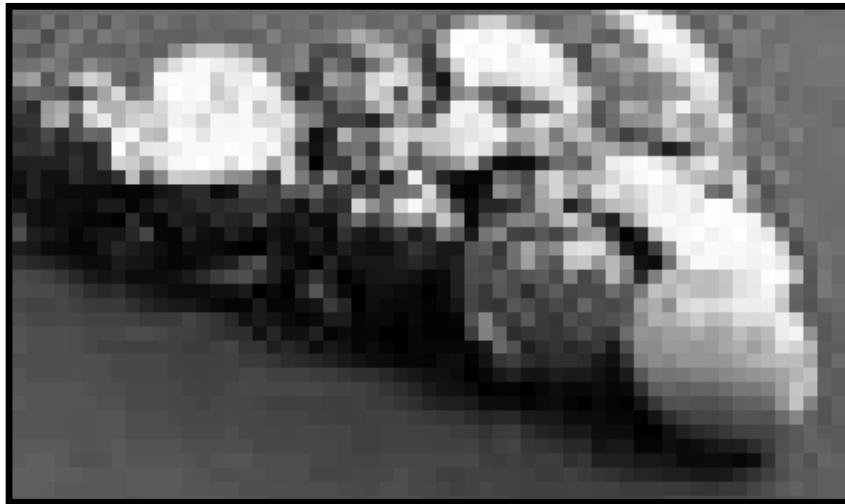
Note: not all digital images are useful as raster models



© Mohan Venigalla, Fall 2009

GIS Data

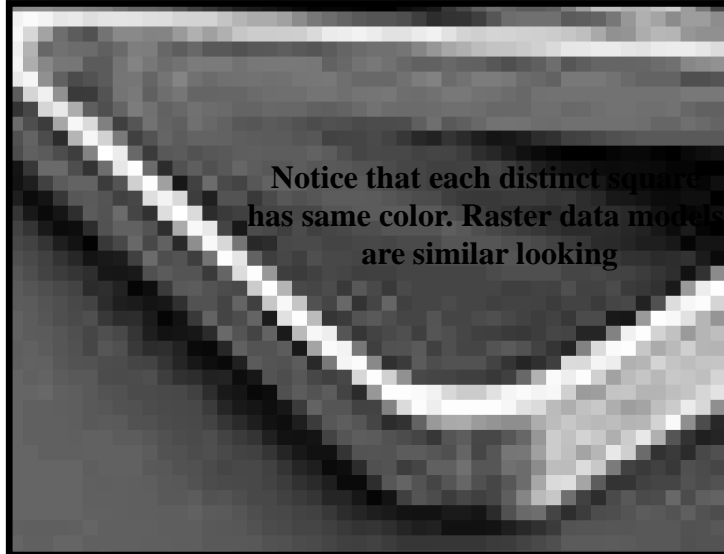
## Digital Images in Close-up (sphere look no longer spherical)



© Mohan Venigalla, Fall 2009

GIS Data

## Digital Images in Close-up (lines no longer look straight)

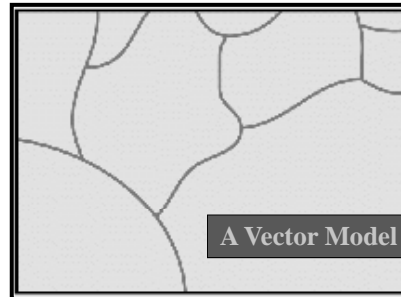


© Mohan Venigalla, Fall 2009

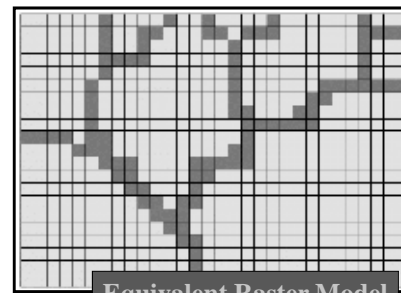
GIS Data



© Mohan Venigalla, Fall 2009



A Vector Model



Equivalent Raster Model

GIS Data

## Vector Models

- Stick model
- Spaghetti vector model
- Topological vector model



© Mohan Venigalla, Fall 2009

GIS Data

## Common Vector Models

- GBF/DIME – developed by US Census Bureau
- TIGER – improvement over GBF/DIME. First used in 1990 census.
- DLH – digital line graphs (used for USGS topographic maps)
- AutoCAD DXF – de facto standard for export format in many GIS
- ArcInfo E00 – Arc Info Export file
- Shape files (fn.dbf, fn.shx, fn.shp)
- CGM – Computer graphics metafile (ISO standard)



© Mohan Venigalla, Fall 2009

GIS Data

## Raster Data

- Raster data are applied in at least four ways
  - Models describing the real world
  - Digital image scans of existing maps (e.g. DRG)
  - Compiling digital satellite image data
  - Automatic drawing driven by raster output units
- Nature and characteristics of raster data
  - Spatial resolution
  - Column nos. and Row nos.
  - Cell values
  - Coordinate reference of beginning and end points



© Mohan Venigalla, Fall 2009

GIS Data

## Raster Models

- Purpose:
  - Storing and referencing individual grid cell values, their attributes, coverage names and legends
- Three popular models
  - GRID/LUNR/MAGI model
  - IMGRID GIS data model
  - The map analysis package (MAP) GIS data model



© Mohan Venigalla, Fall 2009

GIS Data

## Raster Data Compression

- Run-length encoding
  - Reduce a row of cell with the same value to a single unit having specific value and quantity
- Chain codes
  - Store only the cells that lie on the outer border of homogeneous areas/polygons
- Block codes
  - Homogeneous quadratic areas are stored as one unit
- Quad-tree model
  - Kind of like PLSS



© Mohan Venigalla, Fall 2009

GIS Data

## Raster File Formats

- BMP – usually larger file size
- PCX – proprietary (Zsoft)
- TIFF – Tagged Image File Format
- GeoTIFF – an extension of TIFF with georeferencing information at file header
- GIF – graphic interchange Format (cross-platform)
- JPEG – joint photographic experts group
- MrSID – Multi-resolution Seamless Image Database
  - A raster format using compressing technology



© Mohan Venigalla, Fall 2009

GIS Data

## Technical Issues

- Data classification and classification scheme
  - Example: Land use/Land cover
    - Main class: urban, agricultural, etc
    - Subclass: (residential, commercial etc) and (cropland, orchards etc)
- Feature codes and feature coding
- Geographic data precision
- Data organization in computer

## Preparing Data

- Digital Classification and Classification Scheme
- Feature Codes and Feature Coding
- Data Precision
- Data Organization
  - Database management systems
    - ER diagrams

## Data Organization

- Flat files
- Databases
  - Hierarchical
  - Network
  - Relational



© Mohan Venigalla, Fall 2009

GIS Data

## Data Collection

- Learning Objectives
  - List different methods of producing digital map data
  - Outline principles and issues related to these methods



© Mohan Venigalla, Fall 2009

GIS Data

# Producing Data

- Four methods
  - Digitizing existing maps
  - Scanning existing maps
  - Digital photogrammetric map production
  - Entry of computed coordinates from field measurements
- In general, done in three phases
  - Preparation
  - Digitizing/scanning
  - Editing and improvement of data quality



© Mohan Venigalla, Fall 2009

GIS Data

# Digitizing



© Mohan Venigalla, Fall 2009

GIS Data

## Scanning

- Maps are scanned to:
  - Use digital image data as bases for other map information
  - Convert scanned data to vector data for use in vector GIS
- Two operations
  - Scanning
  - Binary encoding



© Mohan Venigalla, Fall 2009

GIS Data

## Storage and Editing

- Learning Objectives
  - List the type of data stored for vector and raster models
  - Outline the purpose of data editing
  - List and describe the type of errors that are:
    - Encountered with vector and raster models
    - Associated with points, lines and polygons
  - Most importantly, avoid these errors in your project and other GIS work



© Mohan Venigalla, Fall 2009

GIS Data

## Data Storage

### ➤ In what format the GIS data are stored?

- Various data structures
  - Dependent on GIS data model
    - e.g.

### ➤ How the data are stored?

- All data as one big chunk?
- Bits and pieces?
  - Tiling
    - Edited and cleaned data before archiving as 'tiles'
    - Editing done on individual tiles – requires edge matching



© Mohan Venigalla, Fall 2009

GIS Data

## Data Storage - 2

### ➤ What is stored in a GIS database?

- Vector – entities and attributes as:
  - Individual tables, or
  - Separate databases that can be linked to entities in GIS
- Raster – attribute values of raster grid cells
  - Grid cell locations are catalogued relative the location of the other grids
  - An associated cell value (numeric)



© Mohan Venigalla, Fall 2009

GIS Data

# Editing

## ➤ Why edit?

- To fix error and to make changes for
  - Entities themselves (the GIS objects)
  - Attributes, and
  - Databases

## ➤ In general – errors can be

- Internal to data layers, or
- Mismatch with other data sources



© Mohan Venigalla, Fall 2009

GIS Data

# Editing - 2

## ➤ Vector

- Concerns with separate editing for
  - Entities
  - Attributes, and
  - Databases

## ➤ Raster

- Primarily concerned with correct relative positions of each grid cell
- If raster allows linkages to RDBMS, editing can be complicated



© Mohan Venigalla, Fall 2009

GIS Data

## Errors

- Three types of errors
  - Entity error
  - Attribute error
  - Entity-attribute error
- Entity (or positional) error
  - Most common in vector models
  - Relatively easy to locate and fix
  - 3 types
    - Missing entities
    - Improper placement
    - disordered



© Mohan Venigalla, Fall 2009

GIS Data

## Errors - 2

- Attribute errors
  - Occurs in both raster and vector models
  - Most common reason:
    - Bad typing/misspelling
  - Several types of errors occur – examples:
    - Raster
      - Use of wrong code for grid, or
      - Wrong grid location
    - Vector
      - Use of wrong code for *attributes*, misspelling, oversight etc.



© Mohan Venigalla, Fall 2009

GIS Data

## Errors – 3

- Entity-attribute mismatch
  - Mostly in vector
  - Attributes, though correct, are attached to wrong entities
- Level of difficulty
  - Entity errors – easy to spot and easy to correct
  - Attribute errors - difficult to spot but doable
  - Entity-attribute errors – nightmares!



© Mohan Venigalla, Fall 2009

GIS Data

## Spotting Entity Errors

ESRI's six-rules for spotting entity errors – based on negative cases for:

1. All entities that should have been entered are present
2. No extra entities have been digitized
3. The entities are in the right place and are of the correct shape and size
4. All entities that are supposed to be connected to each other are connected as they should
5. All polygons have only single label point to identify them
6. All entities are within the outside boundary identified with registration marks.



© Mohan Venigalla, Fall 2009

GIS Data

## Entity Errors - Vector

- Node errors
  - Pseudo nodes
  - Dangling nodes
    - Undershoot
    - Overshoot
  - Acceptable dangling nodes



© Mohan Venigalla, Fall 2009

GIS Data

## Entity Errors – Vector (2)

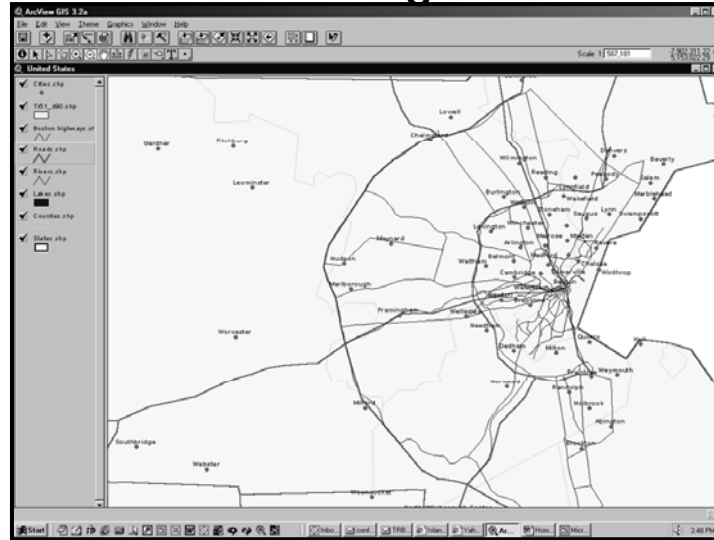
- Line data errors
  - Mismatch with nodes (or endpoints or *a-node* and *b-node*)
  - Missing and crooked shape points (or intermediate points)
  - Missing connections (nodes) with other lines



© Mohan Venigalla, Fall 2009

GIS Data

## Mismatching *Lines*



© Mohan Venigalla, Fall 2009

GIS Data

## Entity Errors – Vector (3)

- Polygon errors
  - Missing labels or too many labels
    - Location of the label-point
  - Edge mismatch
  - Sliver polygons
    - Problems with common boundaries
    - Important for your project
- Other editing topics



© Mohan Venigalla, Fall 2009

GIS Data

## Attribute Errors

- Most difficult to find
- Attribute errors with discernible patterns
  - Relatively easy to fix
- If there is no pattern
  - Tough luck
  - Do it the hard way
    - Manual adjustment is almost mandatory



© Mohan Venigalla, Fall 2009

GIS Data

## Attribute Errors (2)

- Missing attributes
  - RDBMS + GIS can help you spot the missing values
    - You still have to fix them
  - Misspelling
    - This is the MOST COMMON problem for vector models
    - Some are easy and some are difficult to fix
    - E.g. how many different ways can Sioux City can be misspelled in a survey sample of 1000?
      - 12? 20? 100? 200?
      - One possible way to fix these errors
- Incorrect attribute values
  - Difficult to spot and fix



© Mohan Venigalla, Fall 2009

GIS Data