Lecture 5
Maps as Numbers
Clarke, Chapter 3
both data and maps must be numbers.

- computer memory in a physical data structure (i.e. files and directories).

Files in binary or ASCII

- Binary is faster to read and smaller, ASCII can be read by humans and edited but uses more space.
Data Model & Data Structure

- Data model - conceptual description (mental model) of how data are organized for use by the GIS
- Data structure - implementation in computer
  - “raster” or “vector” data structures for maps and the “flat-file” data structure for attributes.
Attribute data

- Attribute data are stored logically in flat files.
- i.e., matrix of numbers and values stored in rows and columns, like a spreadsheet.
- DBMSs use many different methods to store and manage flat files.
Run Length Encoding

- based on spatial autocorrelation
  - nearby things tend to be more similar than distant things
- data entered as pairs
  - run length & value
- 40 items instead of 70
Vector Data Structure
Featuring Arcs and Nodes
Arc-Node Combo Eliminates Redundancy of Storage
Vectors and Topology

- Vectors without topology are spaghetti structures.
- Points, lines, and areas
  - stored in their own files, with links between them.
  - stored w/ topology (i.e. the connecting arcs and left and right polygons).
- Relationships are computed and stored
Topology, GIS, and You

- Topological data structures dominate GIS software.
- Must BUILD topology from unconnected arcs — rarely are maps topologically clean when digitized or imported.
- “Tolerances” important -features can move or disappear — “snapping”, elimination, merging, etc.
Nodes that are close together are snapped.
Slivers due to double digitizing and overlay can be eliminated.
The bounding rectangle

(xmin, ymin)

(xmax, ymax)
Why Topology Matters

- allows automated error detection and elimination.
- makes spatial analysis possible.
- makes map overlay feasible.
- allows many GIS operations to be done without accessing the (x,y) files.
“Rasters are faster, but vectors are correcter”

- Vectors can represent point, line, and area features very accurately.
- far more efficient than grids.
- work well with pen and light-plotting devices and tablet digitizers.
- not as good at continuous coverages or plotters that fill areas.
  - TIN can be used to represent surfaces.
TIN: Triangulated Irregular Network

- Based on the Delaunay triangulation model of a set of irregularly distributed points.
- Way to handle raster data with the vector data structure.
- Common in most GIS.
- More efficient than a grid.
FORMATS

- Most GISs can import different data formats, or use utility programs to convert them.
- Data formats can be industry standard, commonly accepted or standard.
ESRI Spatial Data Access Model

Product

Arc/Info

ArcView

Map Objects

Data Model

Library

ArcStorm

Cover

SDE

Shape
Vector Data Formats

- Vector formats are either page definition languages or preserve ground coordinates.
- Page languages are HPGL, PostScript, and Autocad DXF.
- True vector GIS data formats are DLG and TIGER, which have topology.
The TIGER data structure
(Topologically Integrated Geographic Encoding & Referencing)
Most raster formats are digital image formats.

Most GISs accept
- TIF (tagged interchange format)
- GIF (graphics interchange format)
- JPEG (joint photographic experts group)
- encapsulated PostScript
- none are georeferenced

DEMs are true raster data formats.
Vector to raster exchange errors

vector to raster is easy
raster to vector is hard
Data Exchange

- Data also are often exchanged or transferred between different GIS packages and computer systems.
- The history of GIS data exchange is chaotic and has been wasteful!