DATA MODELS

Sandeep Talasila, GISP

COMPONENTS OF GEOGRAPHIC DATA

- Geometry
- Attributes
- Behavior

Table

Countries

ObjectID *	Shape *	ISO_NUM	CNTRY_NAME	LONG_NAME
1	Polygon	328	Guyana	Co-operative Republic of Guyana
2	Polygon	740	Suriname	Republic of Suriname
3	Polygon	780	Trinidad & Tobago	Republic of Trinidad and Tobago
4	Polygon	862	Venezuela	Bolivarian Republic of Venezuela
5	Polygon	32	Argentina	Argentine Republic
6	Polygon	68	Bolivia	Republic of Bolivia
7	Polygon	76	Brazil	Federative Republic of Brazil
8	Polygon	152	Chile	Republic of Chile
9	Polygon	218	Ecuador	Republic of Ecuador
10	Polygon	600	Paraguay	Republic of Paraguay
11	Polygon	604	Peru	Republic of Peru
12	Polygon	858	Uruguay	Oriental Republic of Uruguay
13	Polygon	124	Canada	Canada

1 (1 out of 249 Selected)

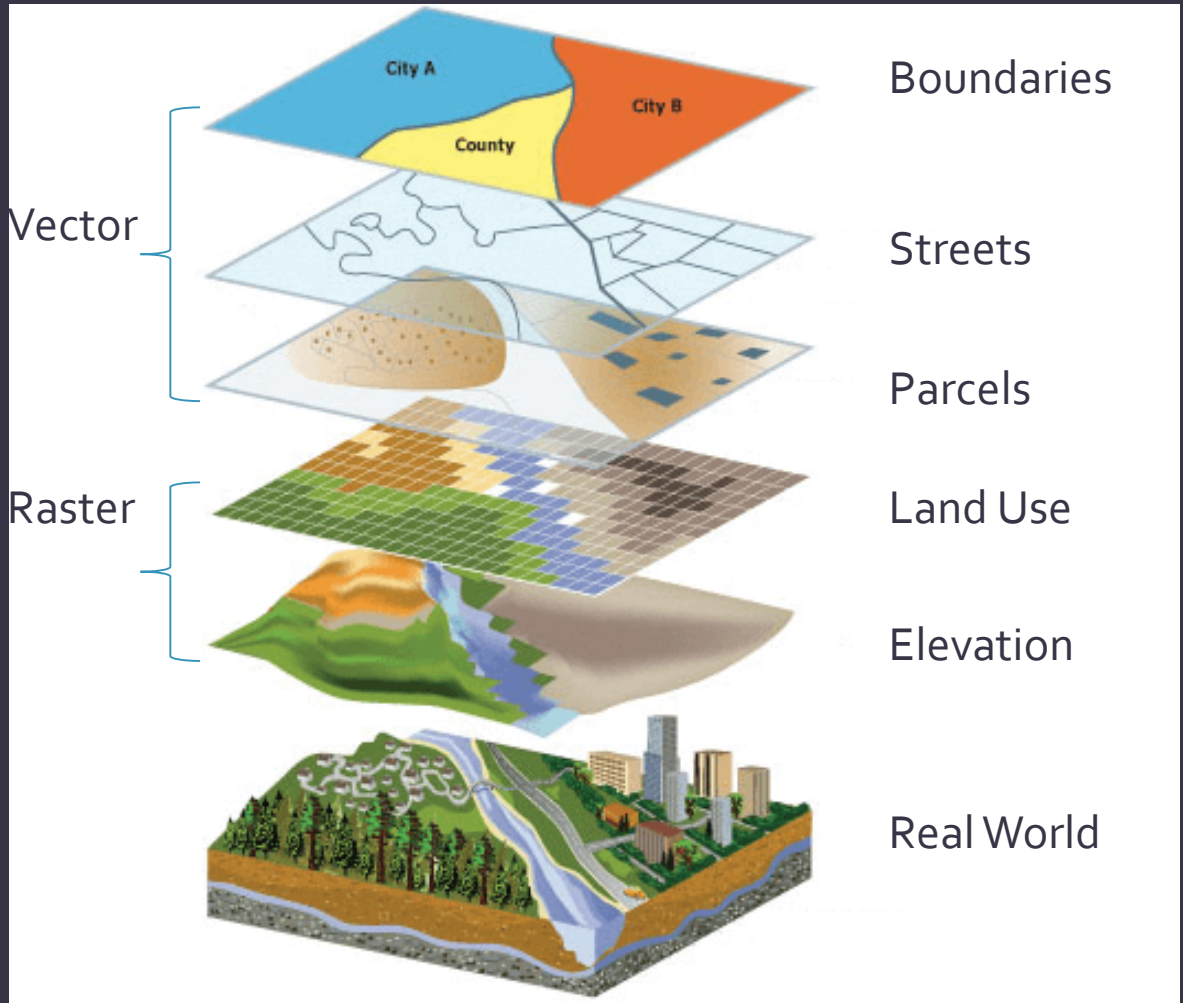
Countries

Rules

Country Boundary polygons cannot include gaps or form voids

SPATIAL DATA MODELS

- Representation of data
- Vector
  - Point, Line, Polygon
- Raster
  - Cells or pixels
- Triangulated Irregular Network (TIN)

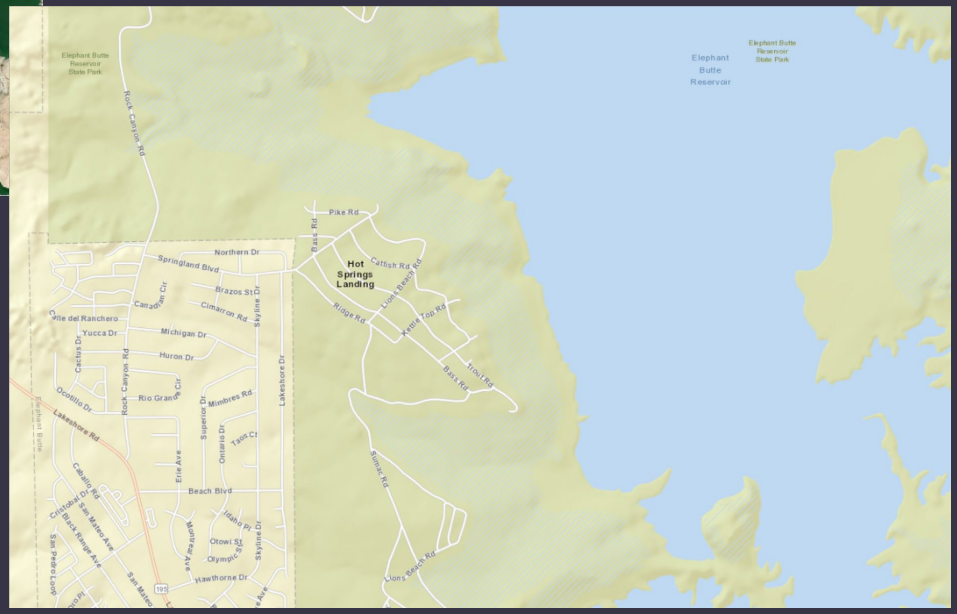


Source: ESRI

DATA ABSTRACTION

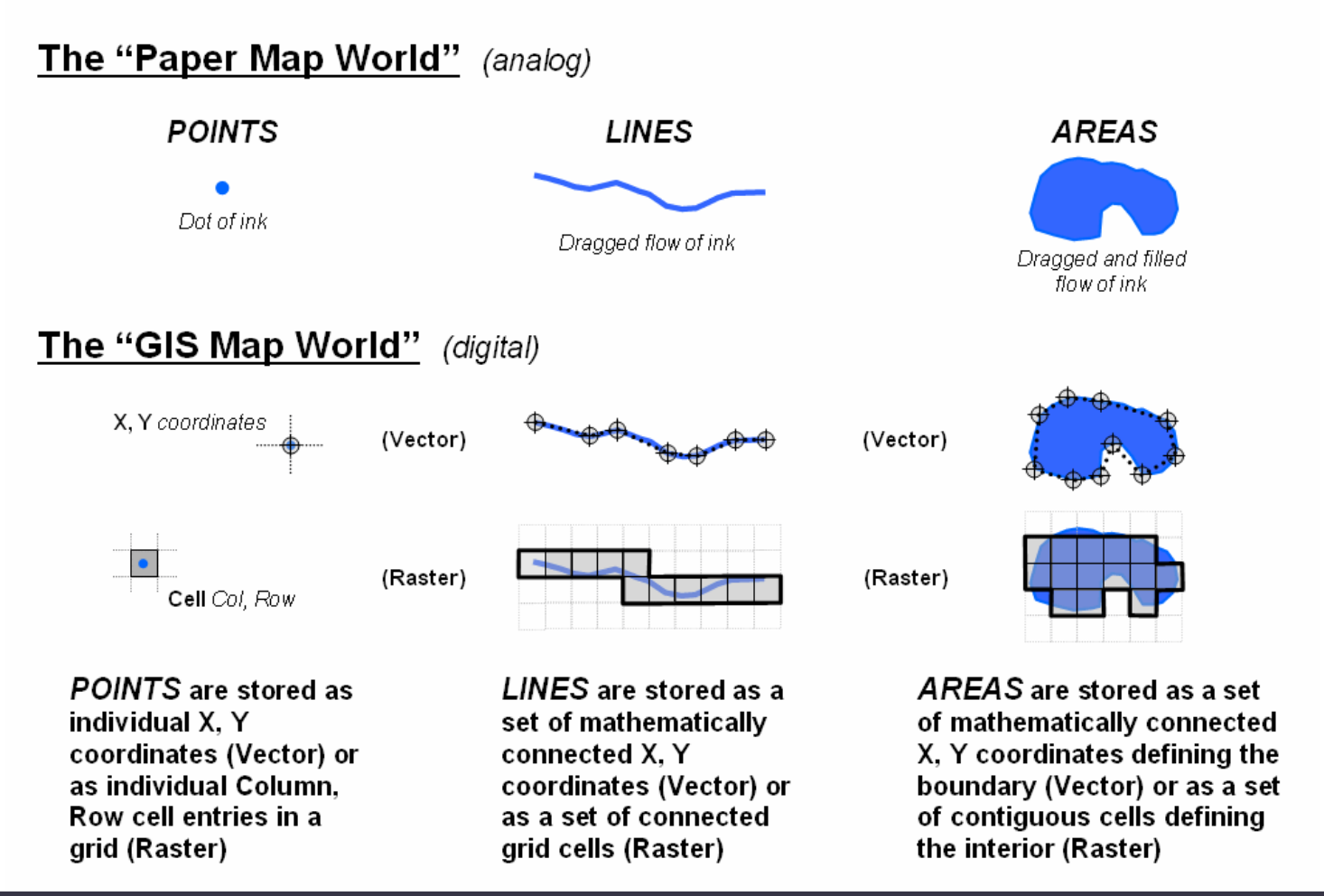
RASTER

City of Elephant Butte and Reservoir



VECTOR

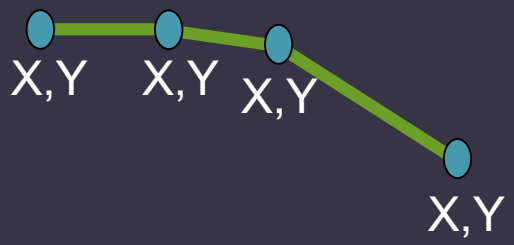
DATA REPRESENTATION



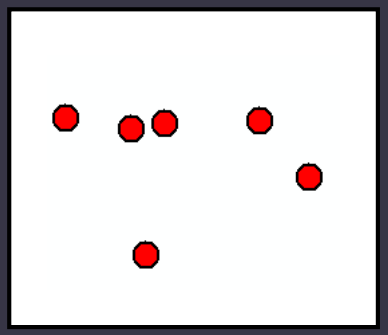
VECTOR DATA MODEL

- Discrete representations of reality
- Stores positional coordinates for each shape
- Real-world entities are abstracted into three basic shapes: Points, Lines, and Polygons

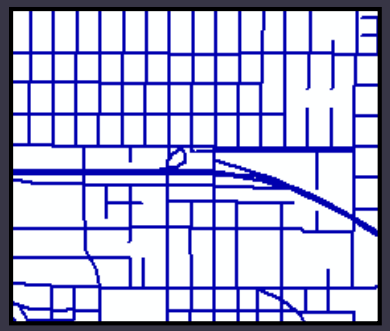
Reality
(A highway)



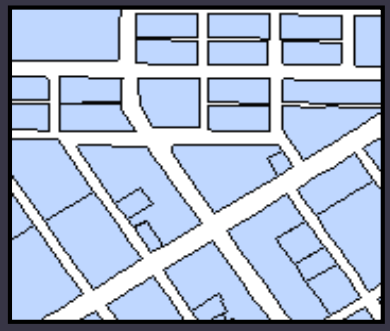
VECTOR DATA MODEL



Points
(Retail stores)

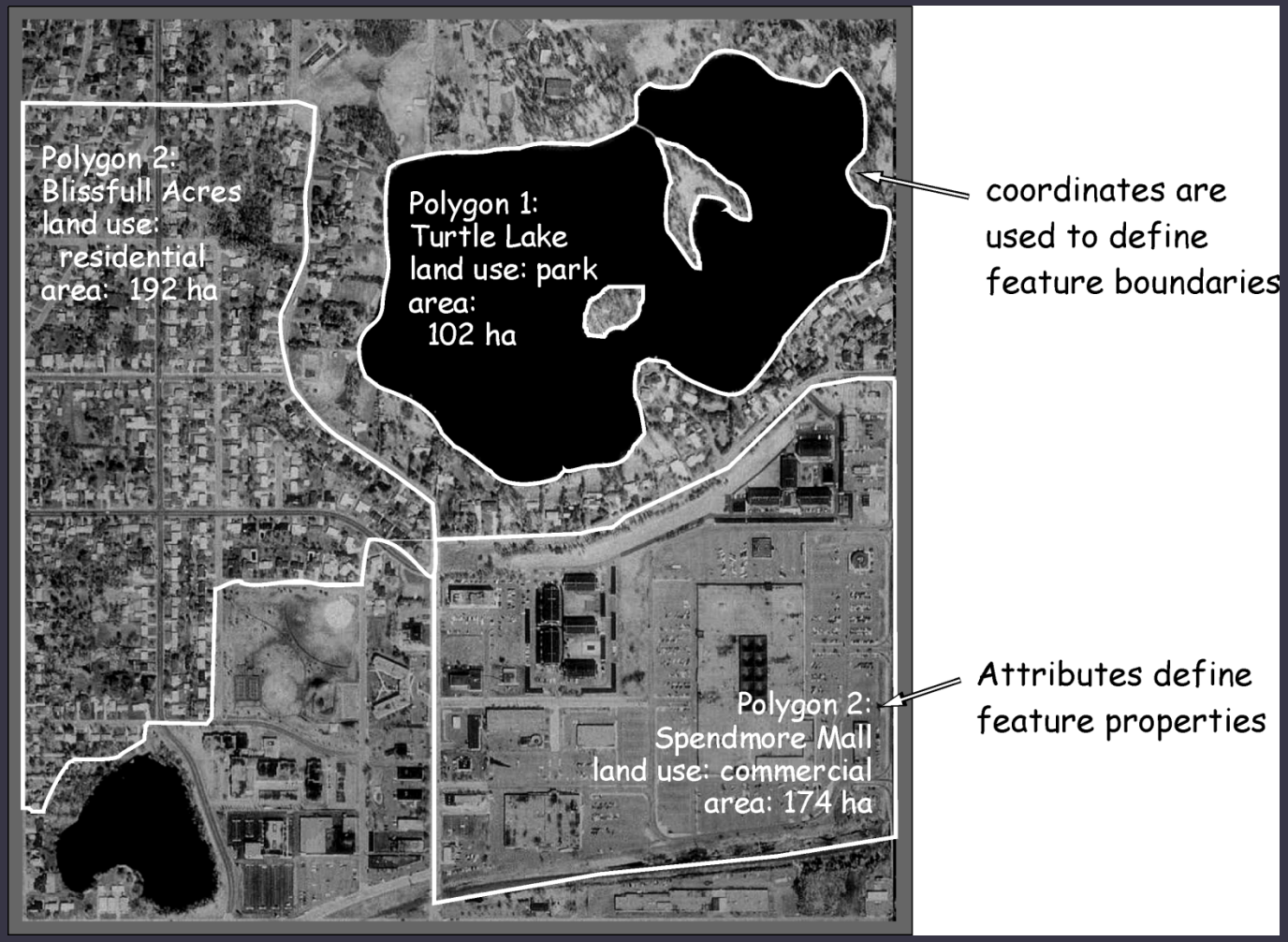


Lines/Arcs
(Streets)



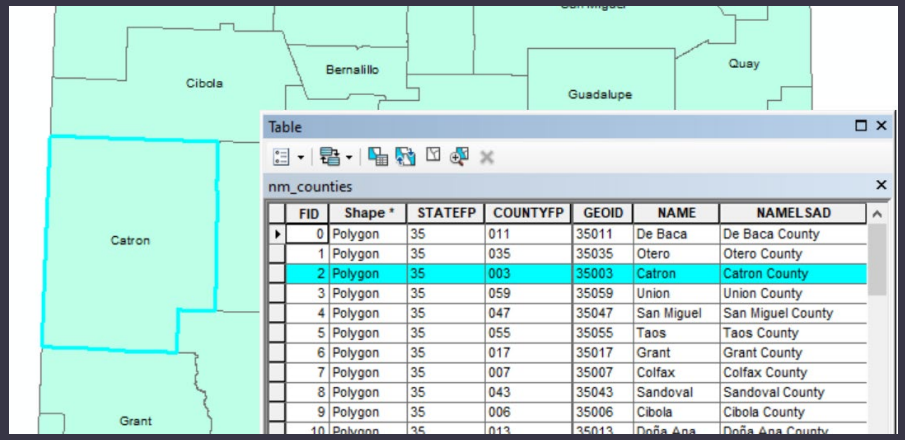
Polygons/Areas
(Land uses)

VECTOR POLYGONS



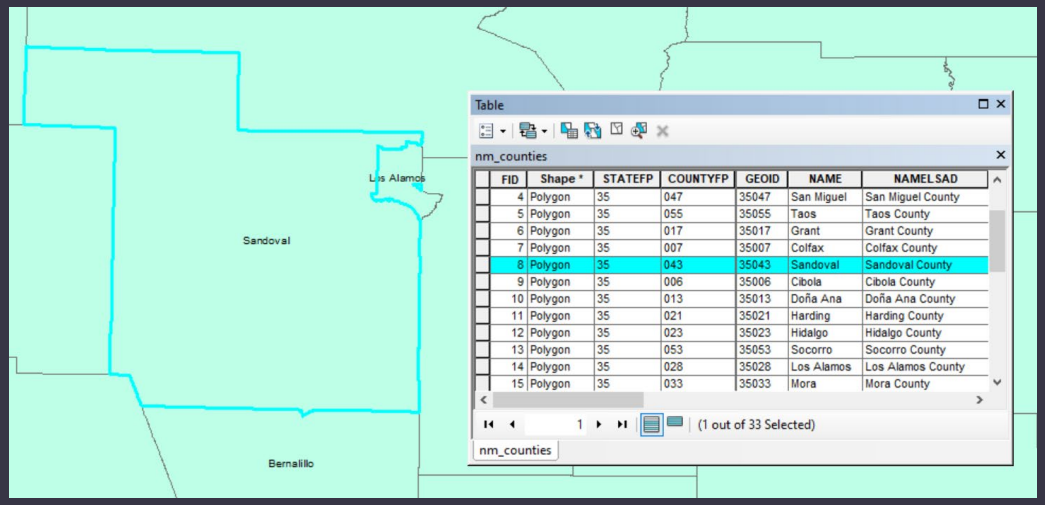
VECTOR ATTRIBUTES

- One-To-One



FID	Shape *	STATEFP	COUNTYFP	GEOID	NAME	NAMELSAD
0	Polygon	35	011	35011	De Baca	De Baca County
1	Polygon	35	035	35035	Otero	Otero County
2	Polygon	35	003	35003	Catron	Catron County
3	Polygon	35	059	35059	Union	Union County
4	Polygon	35	047	35047	San Miguel	San Miguel County
5	Polygon	35	055	35055	Taos	Taos County
6	Polygon	35	017	35017	Grant	Grant County
7	Polygon	35	007	35007	Colfax	Colfax County
8	Polygon	35	043	35043	Sandoval	Sandoval County
9	Polygon	35	006	35006	Cibola	Cibola County

- Many-To-One (multipart features only)

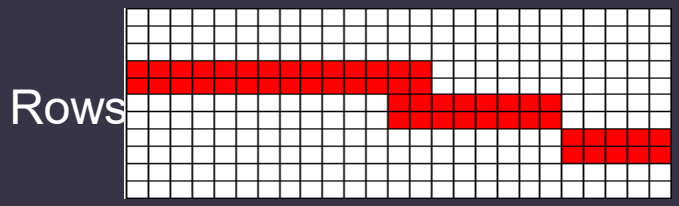


FID	Shape *	STATEFP	COUNTYFP	GEOID	NAME	NAMELSAD
4	Polygon	35	047	35047	San Miguel	San Miguel County
5	Polygon	35	055	35055	Taos	Taos County
6	Polygon	35	017	35017	Grant	Grant County
7	Polygon	35	007	35007	Colfax	Colfax County
8	Polygon	35	043	35043	Sandoval	Sandoval County
9	Polygon	35	006	35006	Cibola	Cibola County
10	Polygon	35	013	35013	Doña Ana	Doña Ana County
11	Polygon	35	021	35021	Harding	Harding County
12	Polygon	35	023	35023	Hidalgo	Hidalgo County
13	Polygon	35	053	35053	Socorro	Socorro County
14	Polygon	35	028	35028	Los Alamos	Los Alamos County
15	Polygon	35	033	35033	Mora	Mora County

RASTER DATA MODEL

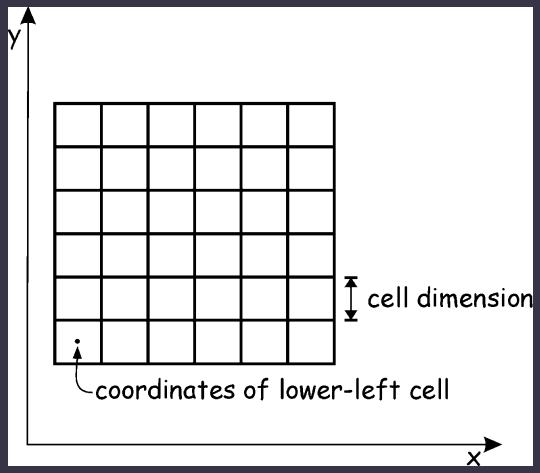
- Continuous and discrete representations of reality
- Use square cells to represent real world objects

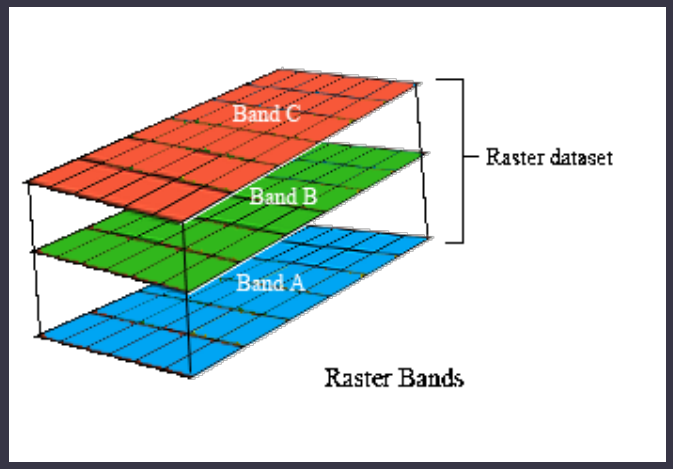
Reality
(A highway)



RASTER DATA MODEL

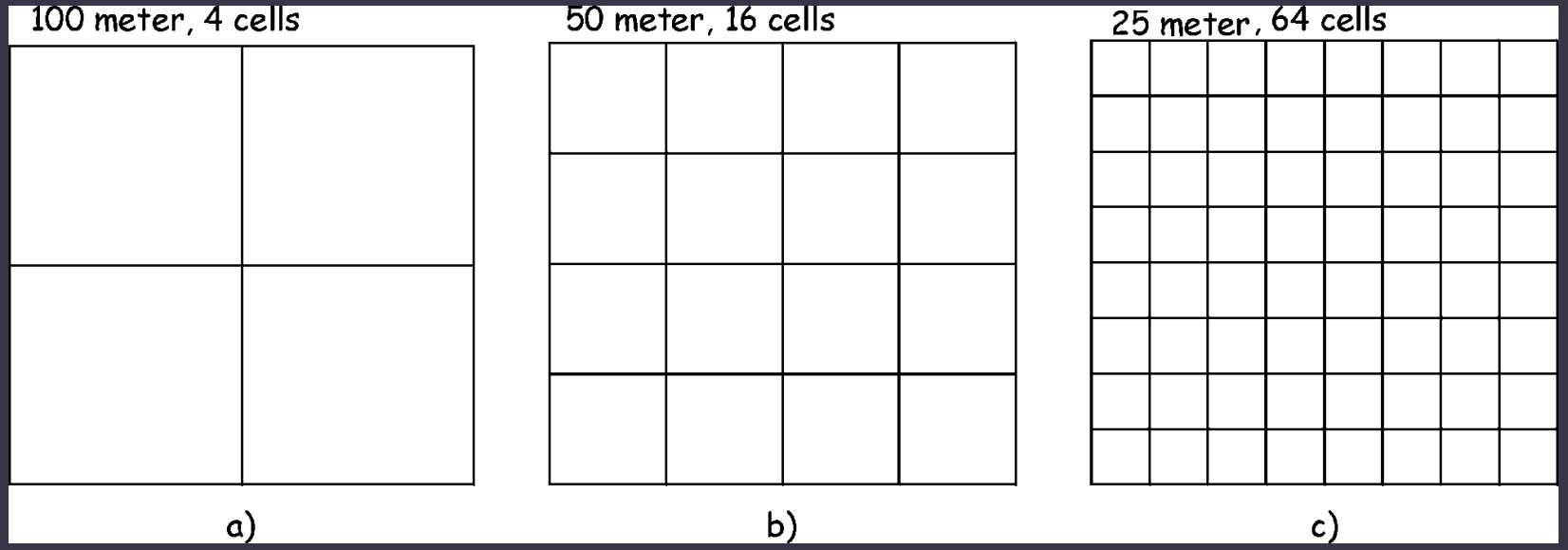
- Cell dimension defines the size of the cell
- A raster contains one or more layers – bands
- Each band contains the cell values and other properties





RASTER CELL DIMENSION

- Smaller cells provide greater detail and increase file size.
- Cell coordinate is defined as the center point of the cell and the coordinate applies to the entire area of cell, which affects positional accuracy.



RASTER ATTRIBUTES

- One-To-One

Raster, one-to-one

A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	B	B	B	B	B	B	B
A	A	A	C	C	B	B	B	B	B
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	E	E
C	C	C	C	C	D	D	E	E	E

attribute table
(cell 1 is upper-left corner)

cell-ID	IDorg	class	area
1	A	10	0.8
2	A	10	0.8
3	A	10	0.8
4	A	10	0.8
5	B	11	0.8
6	B	11	0.8
7	B	11	0.8
⋮	⋮	⋮	⋮
100	E	10	0.8

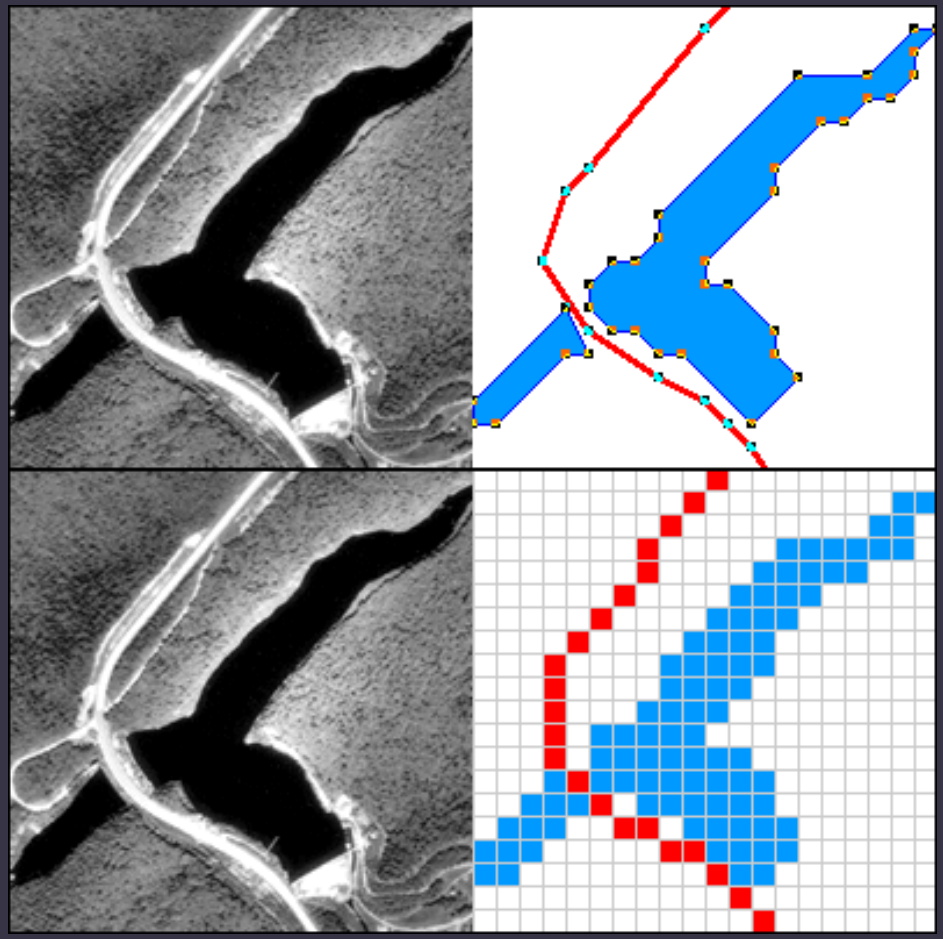
- Many-To-One

	4	4	3	3	3
4	4	4	4	1	1
4	3	3	1	1	1
4	3	3	1	1	1
4	3	3	2	2	1
4	4	2	2	2	

OID	VALUE	COUNT	TYPE	AREA	CODE
0	1	9	Forest land	8100	FL010
1	2	5	Wetland	4500	WL001
2	3	9	Crop land	8100	CL301
3	4	11	Urban	9900	UL040

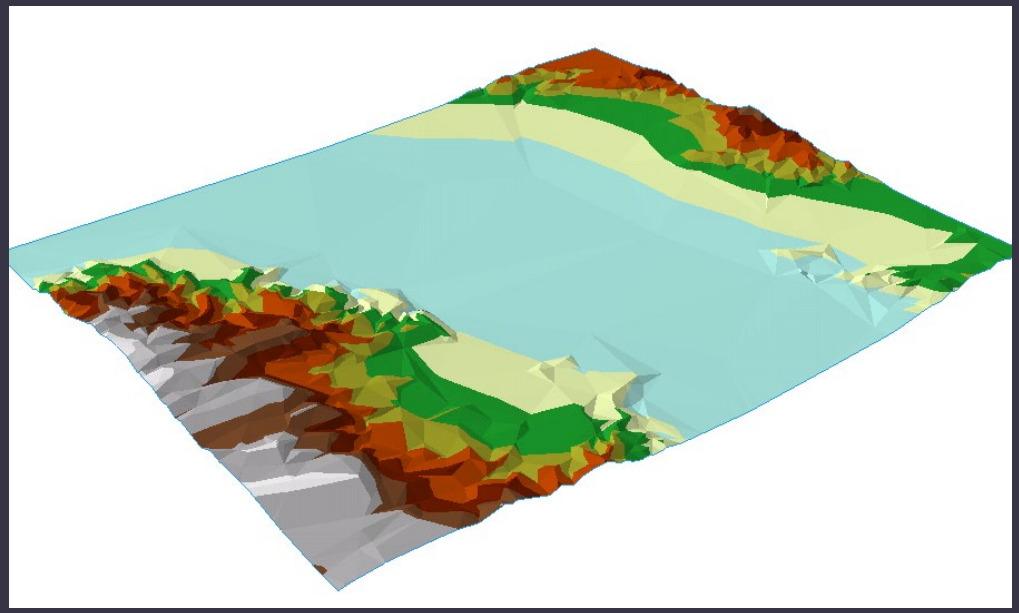
NoData

VECTOR VS. RASTER



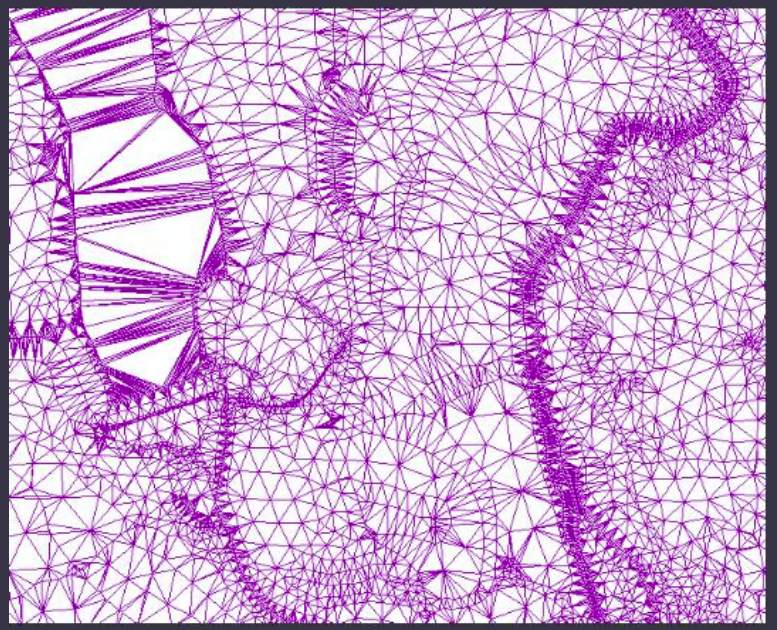
TIN – TRIANGULATED IRREGULAR NETWORK

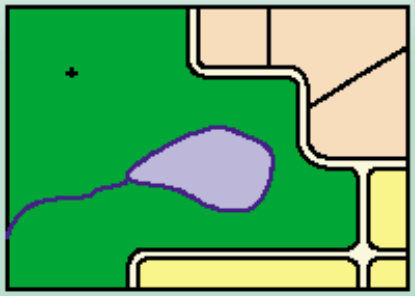
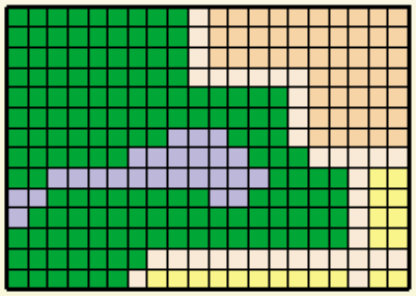
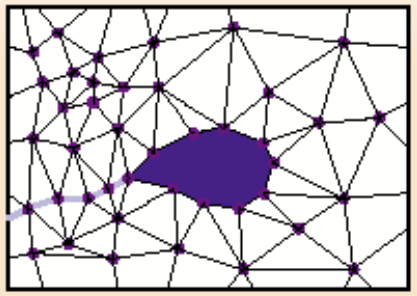
- Represents terrain heights
- Connects 3 closest points into a series or mesh of triangles using linear interpolation
- Triangles cannot overlap
- Used in 3D analysis

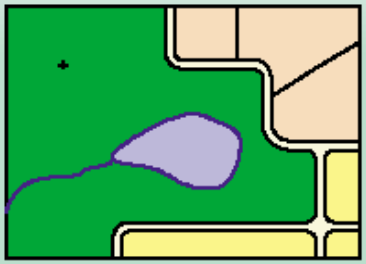
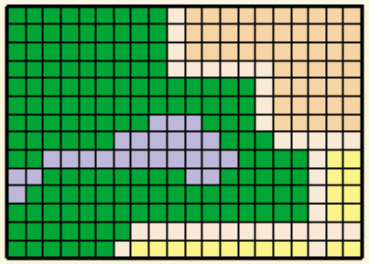
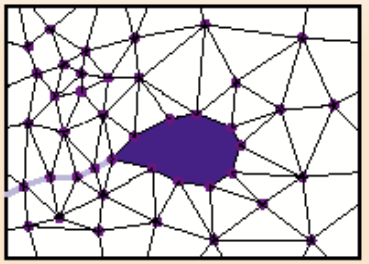


TIN DATA MODEL

- Composed on nodes, edges, and faces
- Represents continuous data
- Complicated data model
- Accommodate different sampling densities
- Preserves each input measurement point



	Vector data representation	Raster data representation	Triangulated data representation
			
Focus of model	Vector data is focused on modeling discrete features with precise shapes and boundaries.	Raster data is focused on modeling continuous phenomena and images of the earth.	Triangulated data is focused on an efficient representation of a surface that can represent elevation or other quality, such as concentration.
Sources of data	Compiled from aerial photography Collected from GPS receivers Digitized from map manuscripts Sketched on top of raster display Vectorized from raster data Contours from triangulation Reduced from survey field data Imported from CAD drawings	Photographed from an airplane Imaged from a satellite Converted from a triangulation Rasterized from vector data Scanned blueprints, photographs	Compiled from aerial photography Collected from GPS receivers Imported points with elevations Converted from vector contours
Spatial storage	Points stored as x,y coordinates. Lines stored as paths of connected x,y coordinates. Polygons stored as closed paths.	From a coordinate in the lower-left corner of the raster and cell height and width, each cell is located by its row and column position.	Each node in a triangle face has an x,y coordinate value.

	Vector data representation	Raster data representation	Triangulated data representation
			
Feature representation	Points represent small features. Lines represent features with a length but small width. Polygons represent features that span an area.	Point features are represented by a single cell. Line features are represented by a series of adjacent cells with common value. Polygon features are represented by a region of cells with common value.	Point z values determine the shape of a surface. Breaklines define changes in the surface such as ridges or streams. Areas of exclusion define polygons with the same elevation.
Topological associations	Line topology keeps track of which lines are connected to a node. Polygon topology keeps track of which polygons are to the right and left sides of a line.	Neighboring cells can be quickly located by incrementing and decrementing row and column values.	Each triangle is associated with its neighboring triangles.
Geographic analysis	Topological map overlay Buffer generation and proximity Polygon dissolve and overlay Spatial and logical query Address geocoding Network analysis	Spatial coincidence Proximity Surface analysis Dispersion Least-cost path	Elevation, slope, aspect calculations Contour derivation from surface Volume calculations Vertical profiles on alignments Viewshed analysis
Cartographic output	Vector data is best for drawing the precise shape and position of features. It is not well suited for continuous phenomena or features with indistinct boundaries.	Raster data is best for presenting images and continuous features with gradually varying attributes. It is not generally well suited for drawing point and line features.	Triangulated data is best for rich presentation of surfaces. This data can be viewed by using color to show elevation, slope, or aspect or in a three-dimensional perspective.

DATA FORMATS

What is that file?

SPATIAL DATA FORMATS

Raster

- Digital Raster Graphic (DRG)
- ECW
- ESRI Grid
- GeoTIFF
- IMG
- JPEG2000
- MrSID
- …

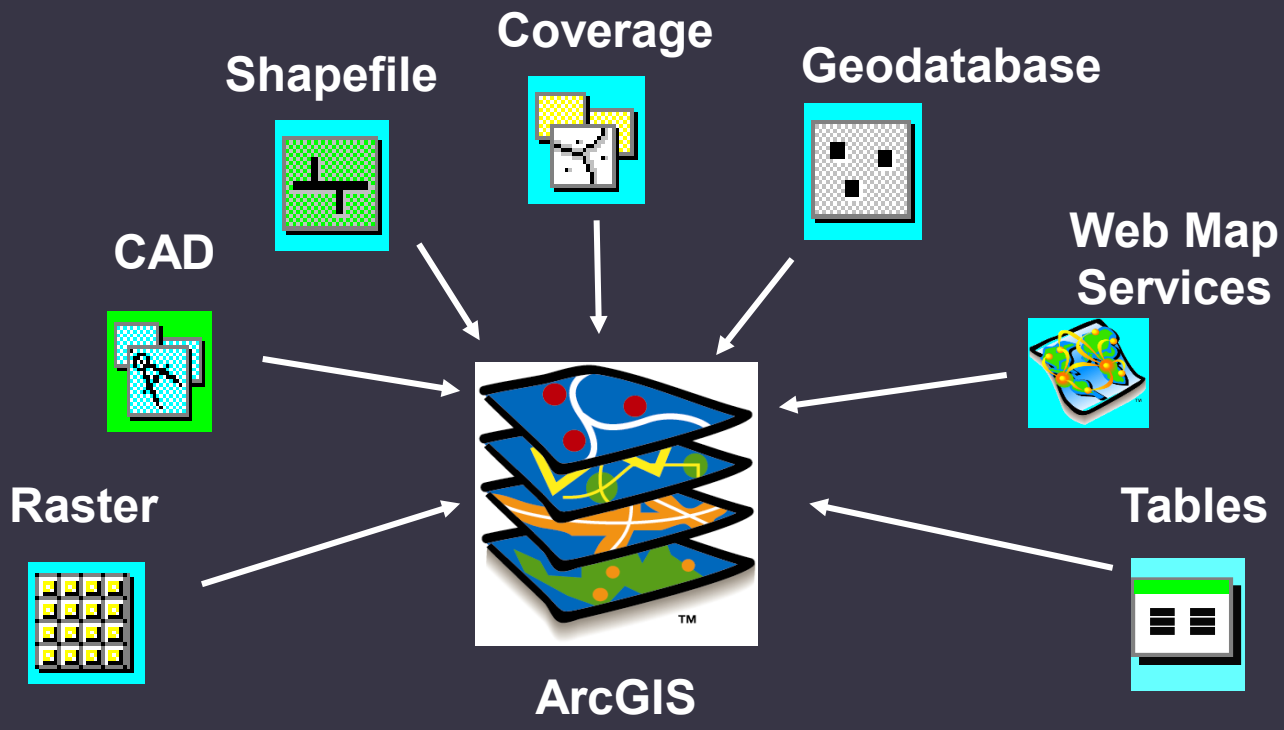
https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/supported-raster-dataset-file-formats.htm

Vector

- Shapefile
- Auto CAD DXF
- Digital Line Graph
- KML
- GeoJSON
- GeoPackage
- TIGER
- MapInfo TAB
- …

DATA FORMATS IN ARCGIS

- ArcGIS can work with spatial data in several formats
- Other data formats may require a conversion before they can be used.



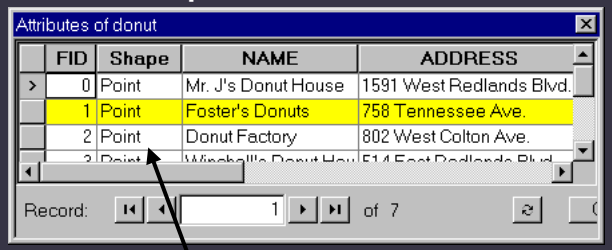
SHAPEFILE

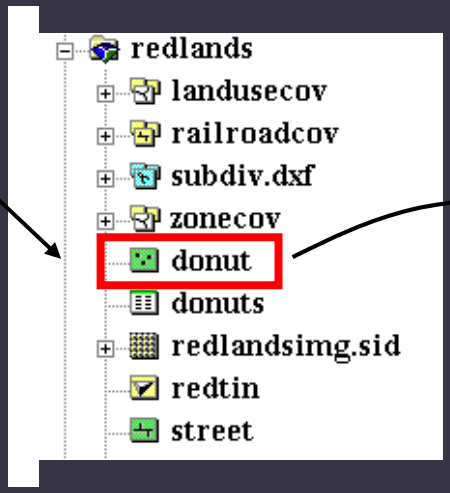
- Native format of ArcView GIS 3.x
- A simple, non-topological format for storing the geometric location and attribute information of geographic features.
- Actually made up of several files:
  - Always SHP, SHX, DBF
  - Sometimes PRJ, SBN, SBX, FBN, FBX, AIN, AIH, AVL, XML

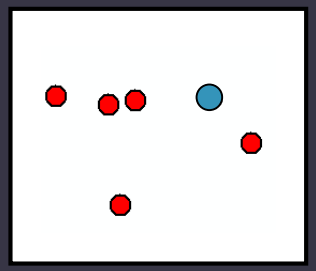
SHAPEFILE

- Single feature class
- Attributes stored in dBASE table

Shapefile .dbf table







Donut shapefile

"Shape" field accesses separate coordinate files

SHAPEFILE EXTENSIONS

- **.shp—The main file that stores the feature geometry; required.**
- **.shx—The index file that stores the index of the feature geometry; required.**
- **.dbf—The dBASE table that stores the attribute information of features; required.**
- .sbn and .sbx—The files that store the spatial index of the features.
- .fbn and .fbx—The files that store the spatial index of the features for shapefiles that are read-only.
- .ain and .aih—The files that store the attribute index of the active fields in a table or a theme's attribute table.
- .atx—An .atx file is created for each shapefile or dBASE attribute index created in ArcCatalog. ArcView GIS 3.x attribute indexes for shapefiles and dBASE files are not used by ArcGIS. A new attribute indexing model has been developed for shapefiles and dBASE files.
- .ixs—Geocoding index for read/write shapefiles.
- .mxs—Geocoding index for read/write shapefiles (ODB format).
- **.prj—The file that stores the coordinate system information; used by ArcGIS.**
- **.xml—Metadata for ArcGIS—stores information about the shapefile.**
- .cpg—An optional file that can be used to specify the codepage for identifying the character set to be used.

- **Each file must have the same prefix, for example, roads.shp, roads.shx, and roads.dbf**.

When viewing shapefiles in ArcCatalog (or any ArcGIS program), you will only see one file representing the shapefile; however, you can use Windows Explorer to view all the files associated with a shapefile. When copying shapefiles, it is recommended that you do so in ArcCatalog or by using a geoprocessing tool. However, if you do copy a shapefile outside ArcGIS, be sure to copy all the files that make up the shapefile.

GEODATABASE

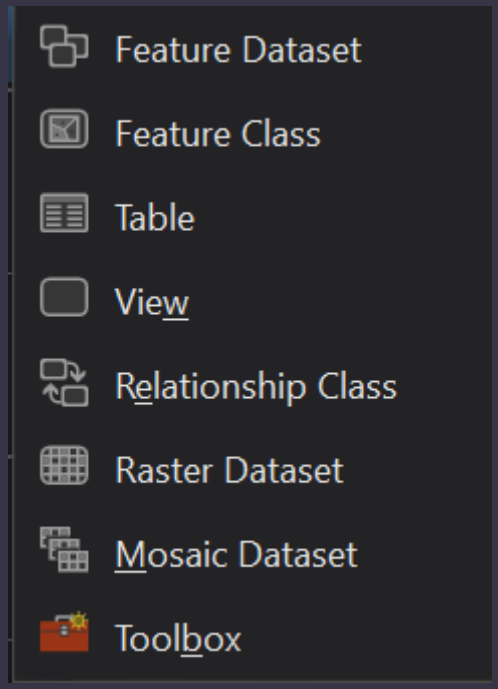
- A collection of geographic datasets of various types held in a common file system folder, or a multiuser relational database management system (DBMS).
- Native data storage and data management framework for ArcGIS.
- Object-Relational Database Management Systems (ORDBM) to store graphic and attribute data.
- Multiuser editing through versioning.
- Implement subtypes and domains.
- Build relationships

TYPES OF GEODATABASE

- File geodatabases – Stored as folders in a file system. File extension .gdb
- Personal geodatabases – All datasets are stored within a Microsoft Access data file. File extension .mdb
- Enterprise geodatabases – Also called ArcSDE. Stored in a relational database using Oracle, Microsoft SQL Server, IBM DB2, IBM Informix, or PostgreSQL.

INSIDE A GEODATABASE

- Feature Dataset – A collection of related feature classes.
- Feature Class – A collection of geographic features of same geometry type – point, line, polygon.
- Table – non spatial table
- View – database view on tables
- Relationship class – Create an association between two tables.
- Raster Dataset – A single raster or image data layer.
- Mosaic Dataset – A collection of rasters stored as a catalog and viewed as a mosaicked image.
- Toolbox – User created geoprocessing tools, scripts, and models.



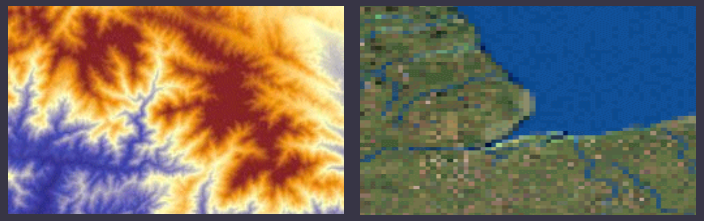
TABULAR FORMATS

- Geodatabase
- Database
- Feature layer attribute table
- dBASE (dbf)
  - One of the files which make up a shapefile is in DBF format
  - Commonly used format to get tabular data into ArcGIS from other application
- Microsoft Excel
- Text, ASCII, and csv

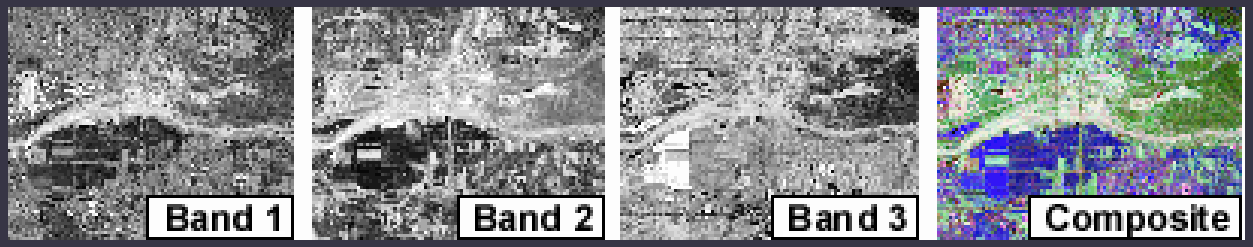
* ArcGIS does read Excel files directly but sometimes the dbf format is easier to use when joining. Excel does not export to dbf format from MS Office 2010. Open Office can be used to export to dbf.

RASTER FORMATS

- Grids (ESRI native raster format)



- Images (TIF, BMP, SID, JPG, ERDAS, ECW)



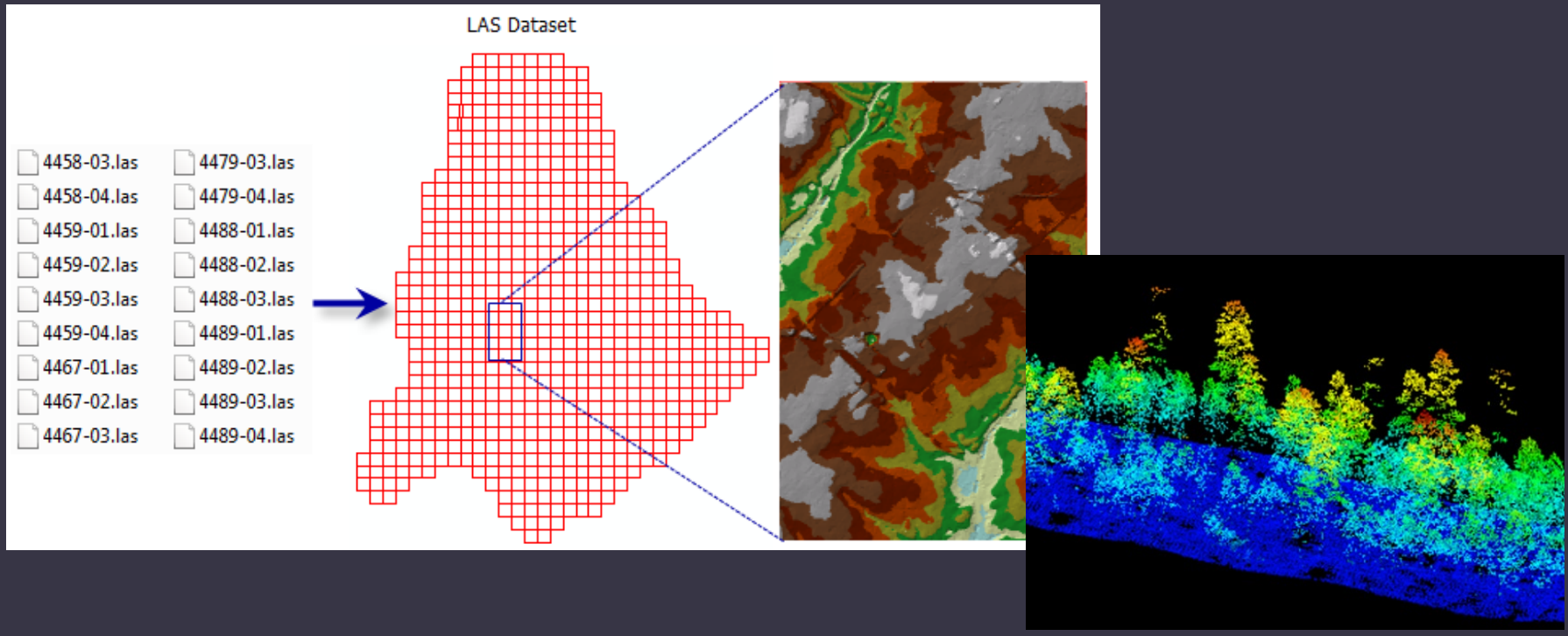
ESRI Supported Raster Formats: https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/supported-raster-dataset-file-formats.htm

GEOPACKAGE (GPKG)

- An open standard format designed to overcome limitations of shapefiles
- Uses SQLite database
- Store multiple types of geographic data
- No storage limitations and suitable for largeoscale projects
- Relatively new format (released in 2014) and adoption is not as widespread as shapefiles

LIDAR

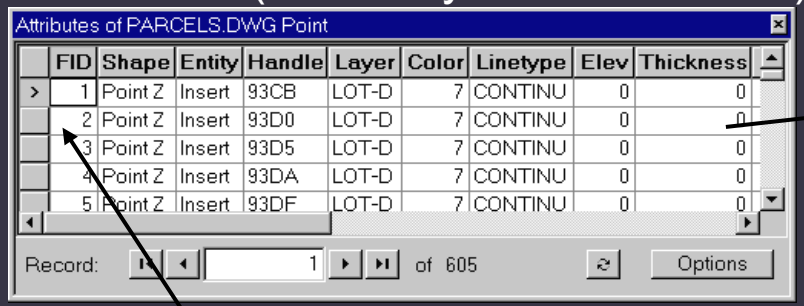
- LAS Format - an open/published standard file format for the interchange of lidar data. It is a binary file format that maintains specific information related to lidar data.



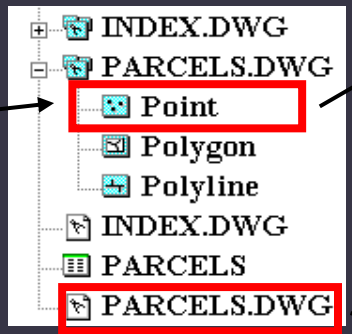
CAD FILES

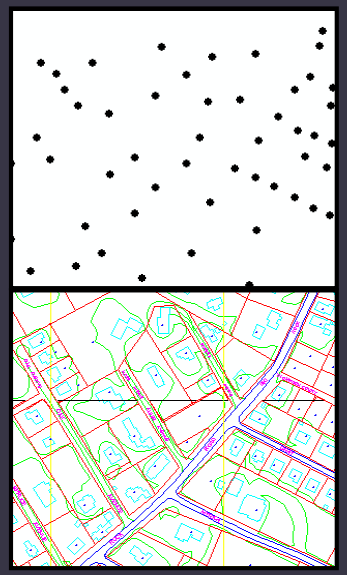
- Computer Aided Design files - DXF, DWG, DGN (AutoCAD/Microstation)
- Logical collection of features classes - access one or all feature class(es) at a time
- Data must be converted to coverage or geodatabase feature class to edit.

CAD file (read-only attribute tables)

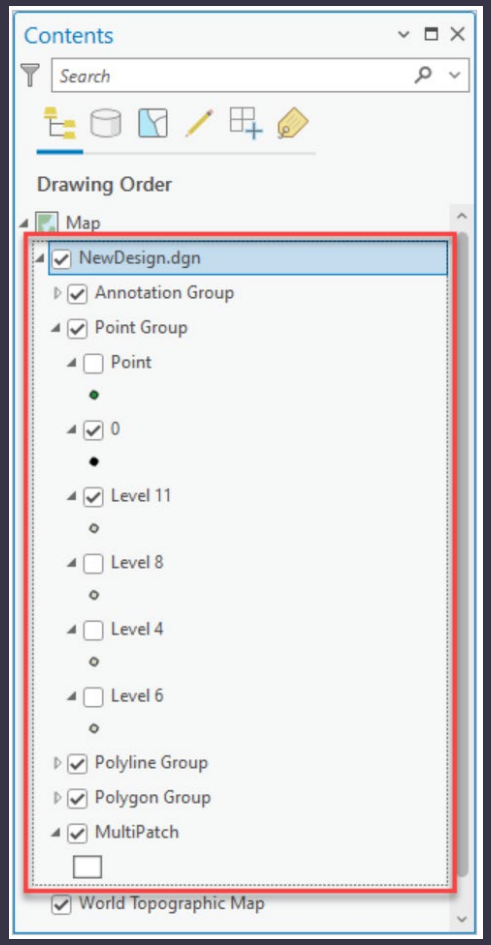


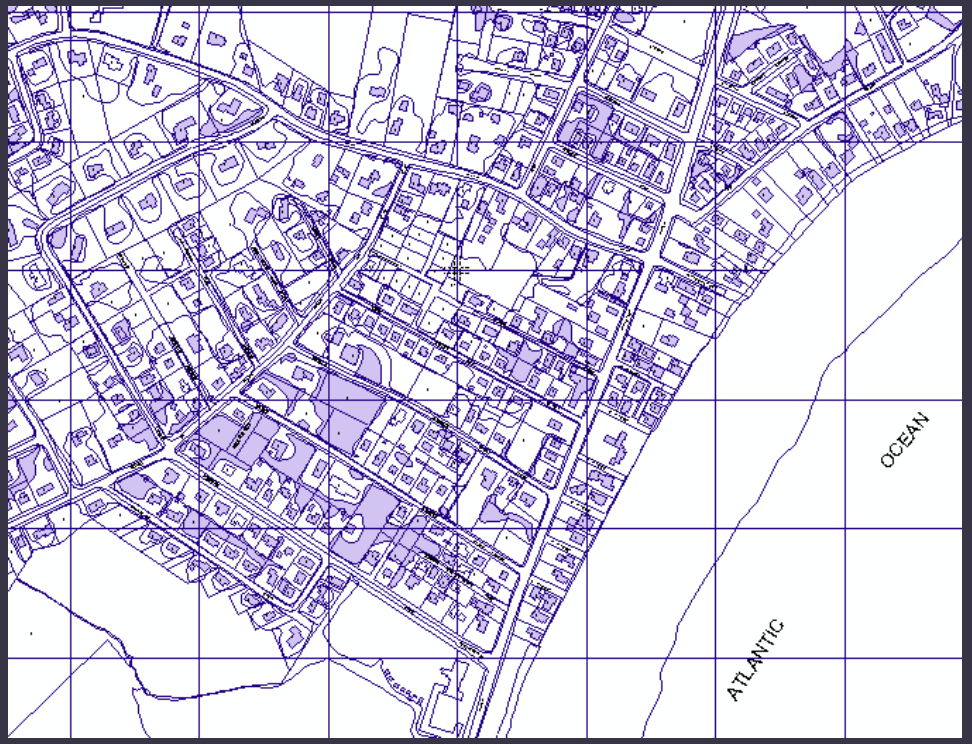
"Shape" field
accesses read-only coordinates





CAD FILES


Contents
Search
Drawing Order
Map
NewDesign.dgn
Annotation Group
Point Group
Point
0
Level 11
Level 8
Level 4
Level 6
Polyline Group
Polygon Group
MultiPatch
World Topographic Map


OCEAN
ATLANTIC

AGENCY FILE FORMATS

- DLG – Digital Line Graph
- DRG – Digital Raster Graphic
- DEM – Digital Elevation Model
- TIGER – Topologically Integrated Geographic Encoding and Referencing (developed for Census)
- GIRAS – Land Use Land Cover Data (outdated)
- SDTS – Spatial Data Transfer Standard (outdated)
- More…

COMPRESSED FORMATS

- Used to reduce file size as well as make a single file out of a large number of files
- Most common one:
  - ZIP
  - Can be created and extracted by WinZip or similar programs
- Others:
  - GZ – created by GZIP program
  - Z – created by Compress command
  - Sid -Mrsid
  - Ecw
  - Others: ARJ, LZH, ARC, TAR, TAZ, TGZ, EXE, UUencoded, XXencoded, BinHex, MIME, TXT, CSV, …

WEB FORMATS

- KML/KMZ
  - KML is a file format for rendering geographic features in certain web-browser two- and three-dimensional map viewers, such as Google Earth. Files are often delivered in zipped, or KMZ format.
- OGC Services
  - Web Map Service(WMS), Web Feature Service (WFS), and Web Coverage Service (WCS). These standards complement each other in terms of their capabilities, with WMS delivering images representing data (i.e. maps), and WFS and WCS delivering actual geospatial data.
- ESRI GIS Services (Web Layers)
  - REST API – RESTful web services
  - Map image layers, web tile layers, vector tile layers, web feature layers, web scene layers

ESRI DATA SHARING FORMATS

- Project packages
  - (.ppkx files) allow you to share complete projects. A project package is a file that contains all maps and the data referenced by its layers, as well as folder connections, toolboxes, geoprocessing history, and attachments.
- Map Package
  - (.mpkx files) allow you to share complete maps. A map package is a file that contains a map (.mapx) and the data referenced by its layers.
- Layer package
  - (.lpkx) includes both the layer properties and the dataset referenced by the layer. With a layer package, you can save and share everything about the layer—its symbology, labeling, table properties, and the data.

FILE NAMING CONVENTIONS

- Files should not be named with spaces or any other special characters in its name.
- ESRI grid multiband raster dataset cannot have more than 9 characters in its filename, and a single-band raster dataset cannot have more than 13 characters.
- File geodatabase – file system limit
  - Feature class or table name – 160 characters
  - Field name – 64 characters
  - File Name cannot start with a number
- Shapefile – file system limit
  - Field name – 10 characters