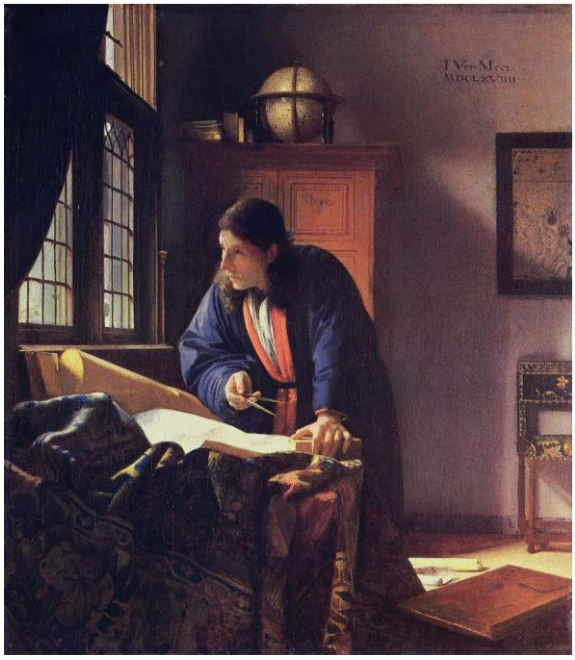
Considering Spatial Data

Learning objectives

- **Understand the basic components of spatial data**
- **Introduce common spatial data formats**
- **Describe the tabular attribute structure for spatial data**



Discuss components of spatial data:

Present common types of spatial data you may encounter and use

Discuss common table operations Perform table join and query operations

Spatial data?

Data with a location component:

- **Latitude/longitude**
- **Street address**
- **Zip code**
- **County**
- **State**

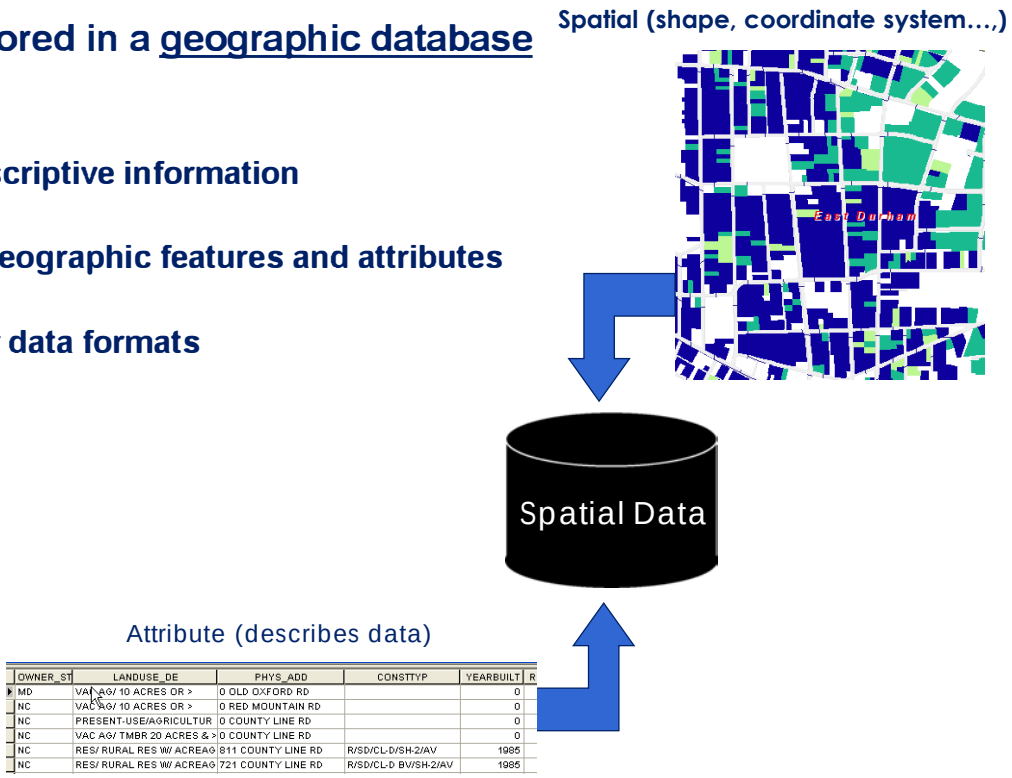
Spatial Data is data with any location component.
Including but not limited to: coordinates, street addresses, zip codes, county name or code, state name, abbreviation, or code.

If you really think about it all data are spatial.

Radio call signs are spatial data: The letter K or W is the first letter of every radio station call sign. The K or W tell us which side of the Mississippi River the station is broadcasting from

Spatial data in GIS

- **Spatial data is stored in a geographic database in GIS software**
  - **Location and descriptive information**
  - **Unique IDs link geographic features and attributes**
  - **Vector and raster data formats**



OWNER_ST	LANDUSE_DE	PHYS_ADD	CONSTTYP	YEARBUILT
MD	VAC AG/ 10 ACRES OR >	0 OLD OXFORD RD		0
NC	VAC AG/ 10 ACRES OR >	0 RED MOUNTAIN RD		0
NC	PRESENT-USE/AGRICULTUR	0 COUNTY LINE RD		0
NC	VAC AG/ TMBR 20 ACRES & >	0 COUNTY LINE RD		0
NC	RES/ RURAL RES W/ ACREAG	811 COUNTY LINE RD	R/SD/CL-D/SH-2/AV	1985
NC	RES/ RURAL RES W/ ACREAG	721 COUNTY LINE RD	R/SD/CL-D BV/SH-2/AV	1985

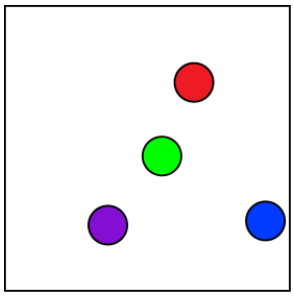
In GIS your data are stored in a **geographic database**

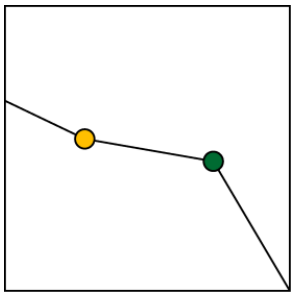
A geographic database contains spatial information (location) and descriptive information of real world entities.

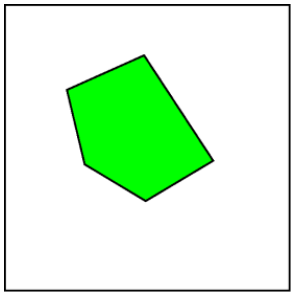
Data are stored in different models. We will focus on the vector model.

Shapes to represent the world

- **In vector-based spatial data, real world entities are abstracted into three basic shapes**
  - **Points:**
    - *facility locations*
    - *patient homes*
    - *fast food restaurants*
  - **Lines:**
    - *roads*
    - *rivers/streams*
    - *movements*
  - **Polygons:**
    - *census geographies*
    - *lakes/ponds*
    - *parcels/buildings*







To represent our world within a computer, GIS creates **geometric representations** of reality. There are three basic geometric shapes used: points, lines, and polygons.

These shapes are sometimes called geometric objects, geometric features, or feature types.

Coordinate systems

- A reference system used to represent the locations within a common geographic framework
- Allows use to use common locations for spatial datasets = integration
- With this information we can map and relate datasets!

Several hundred geographic coordinate systems and a few thousand projected coordinate systems are available for use. In addition, you can define a custom coordinate system.

Types of coordinate systems

The following are two common types of coordinate systems used in a geographic information system (GIS):

A global or spherical coordinate system such as latitude-longitude. These are often referred to as geographic coordinate systems.

A projected coordinate system such as universal transverse Mercator (UTM), Albers Equal Area, or Robinson, all of which (along with numerous other map projection models) provide various mechanisms to project maps of the earth's spherical surface onto a two-dimensional Cartesian coordinate plane. Projected coordinate systems are referred to as map projections.

Coordinate systems (both geographic and projected) provide a framework for defining real-world locations.

What is a spatial reference?

A spatial reference is a series of parameters that define the coordinate system and other spatial properties for each dataset in the geodatabase. It is typical that all datasets for the same area (and in the same geodatabase) use a common spatial reference definition.

A spatial reference includes the following:

The coordinate system

The coordinate precision with which coordinates are stored (often referred to as the coordinate resolution)

Processing tolerances (such as the cluster tolerance)

The spatial extent covered by the dataset (often referred to as the spatial domain)

Common vector data formats

- **Shapefile .shp**

http://www.digitalpreservation.gov/formats/fdd/fdd000280.shtml

- **File GeoDatabase .gdb**

http://www.digitalpreservation.gov/formats/fdd/fdd000294.shtml

ArcGIS PRO supports a number of vector data formats for storing point, line, and area features. We will focus on the most common: the **shapefile and the file** GeoDatabase

Although we will not cover them in detail, ArcGIS PRO supports several raster data formats for storing point line and area features.

There are a number of Geoprocessing tools associated with the raster data model within ArcGIS PRO as well.

Raster formats:

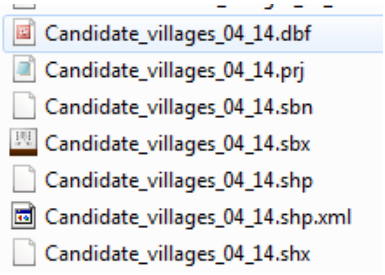
- Esri Grid
- ERDAS Imagine
- GeoTiff
- MrSid

Data format focus: shapefile

A Shapefile is a set of related files:

Mandatory:

shapefile.dbf-stores attribute data
shapfile.shx- index file
shapefile.shp- describes shape



Optional but common:

Shapefile.prj- stores coordinate system information
Shapefile.sbn- spatial index part 1
Shapefile.sbx- spatial index part 2
Shapefile.xml- stores metadata

What you see in the ArcGIS environment:

As stated previously, the shapefile is the most widespread. So, what is a shapefile? It's a collection of related files.

On the right we see what one shapefile: Candidate_villages looks like in the windows environment

On the bottom of the slide we see what the same shapefile looks like in the ArcGIS environment

Index files allows for accurate linking between attribute and shape files (data) (via a unique identifier)

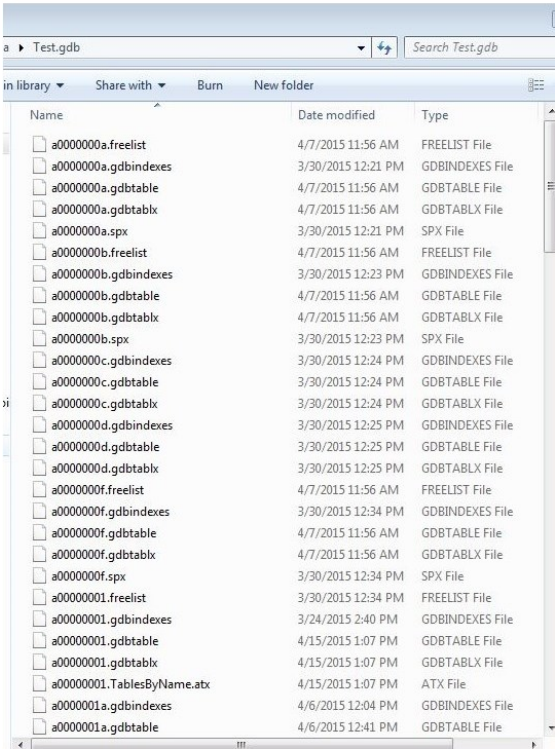
This common spatial data format is widely used and recognized making it useful for collaborative/cooperative projects.

Data format focus: file geodatabase

File Geodatabases allow users to thematically organize and store spatial databases: features, tables, and raster datasets

What you see in the ArcGIS environment:

What you see in the Windows Explorer:



1 TB of storage limit for each dataset (Geodatabases can contain many datasets)
Many users can view data inside the File Geodatabase while the geodatabase is being edited by another user.

The geodatabase can be compressed which helps reduce the geodatabases' size on the disk

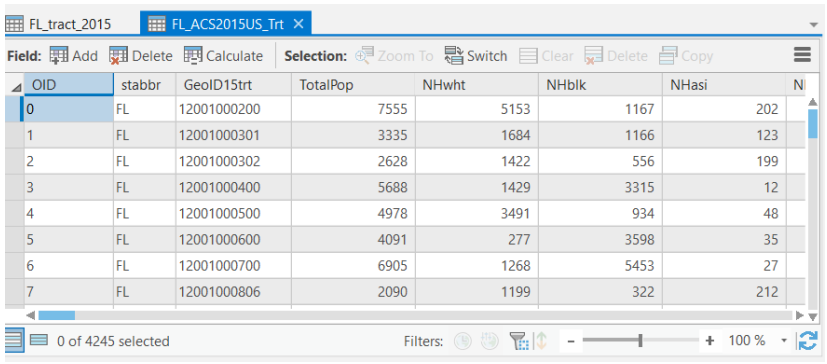
Feature datasets store Feature Classes (which are the equivalent to shapefiles) with the same coordinate system. Like shapefiles, users can create points, lines, and polygons with feature classes; feature classes also have the ability to create annotation, and dimension features.

Follow the link for more information about geodatabases: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/administer-file-gdbs/file-geodatabases.htm#:~:text=A%20file%20geodatabase%20is%20a,Feature%20class

Table components

Fields (columns) ↓

Records (rows) →



OID	stabbr	GeoID15trt	TotalPop	NHwht	NHblk	NHasi	N
0	FL	12001000200	7555	5153	1167	202	
1	FL	12001000301	3335	1684	1166	123	
2	FL	12001000302	2628	1422	556	199	
3	FL	12001000400	5688	1429	3315	12	
4	FL	12001000500	4978	3491	934	48	
5	FL	12001000600	4091	277	3598	35	
6	FL	12001000700	6905	1268	5453	27	
7	FL	12001000806	2090	1199	322	212	

The one-to-one relationship between geometry and attributes is based on record number (OID)

Both formats include tabular storage of tabular and attribute data..

dbf columns

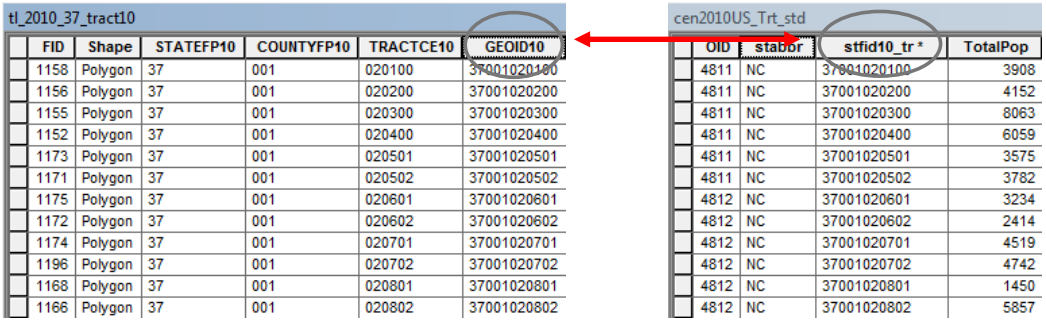
- Must have unique names
- Can contain different data type (text, numeric, date)
- Column names limited to 10 characters
- Allowable column characters: all lower and uppercase alpha, all numeric digits, and the underscore '_' charact

Geodatabase columns

- Must have unique names- supports use of aliases
- Can contain different data type (text, numeric, date)
- Column names limited to 64 characters
- Allowable column characters: all lower and uppercase alpha, all numeric digits, and the underscore '_' charac'

Table joins

- Dynamic connection between tables in your project
- Based on a common ID in each table
  - Must be the same data type
  - Do not need to have the same name



tl_2010_37_tract10

FID	Shape	STATEFP10	COUNTYFP10	TRACTCE10	GEOID10
1158	Polygon	37	001	020100	37001020100
1156	Polygon	37	001	020200	37001020200
1155	Polygon	37	001	020300	37001020300
1152	Polygon	37	001	020400	37001020400
1173	Polygon	37	001	020501	37001020501
1171	Polygon	37	001	020502	37001020502
1175	Polygon	37	001	020601	37001020601
1172	Polygon	37	001	020602	37001020602
1174	Polygon	37	001	020701	37001020701
1196	Polygon	37	001	020702	37001020702
1168	Polygon	37	001	020801	37001020801
1166	Polygon	37	001	020802	37001020802

cen2010US_Trt_std

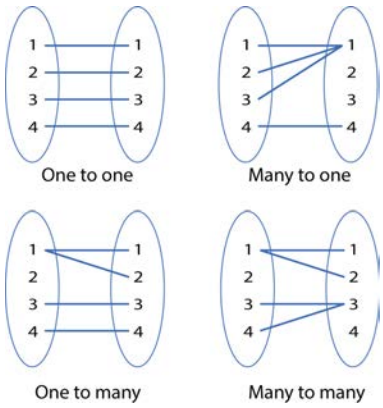
OID	stabbr	stfid10_tr *	TotalPop
4811	NC	37001020100	3908
4811	NC	37001020200	4152
4811	NC	37001020300	8063
4811	NC	37001020400	6059
4811	NC	37001020501	3575
4811	NC	37001020502	3782
4812	NC	37001020601	3234
4812	NC	37001020602	2414
4812	NC	37001020701	4519
4812	NC	37001020702	4742
4812	NC	37001020801	1450
4812	NC	37001020802	5857

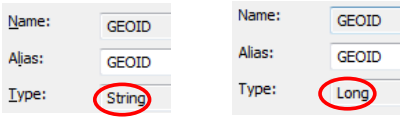
<Target Table>

<Join Table>

Key table join considerations

- **Does your table have a primary key or a unique identifier?**
- **What type of join are you doing?**
- **Are the data types the same?**





Follow the link for more information on table joins: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/tables/essentials-of-joining-tables.htmhtm