Lecture 9: Building a Geodatabase

Sergiy Kostrikov

Geographic Information Systems / Geodatabases for Human Geographers – PhD Courses

Spring 2018

Reminders

- You should look through the Basic Geodatabase textbook

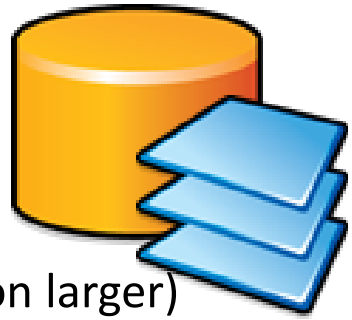
Objectives

- Become proficient in the use of GIS tools to conduct spatial analyses.
- Locate and obtain geospatial data resources from federal, state, and local data providers.
- Correctly convert data into a common coordinate system appropriate for a study site and objectives.
- Work with data in formats common to GIS, including shapefiles, rasters, and geodatabases.

Why are we covering geodatabases?

- It is the native data structure for ArcGIS
- Primary data format used for editing and data management
- Much of the powerful functionality of ArcGIS depends on geodatabase models

First, Some Design Considerations



- What kind of geodatabase do you need?

Single User

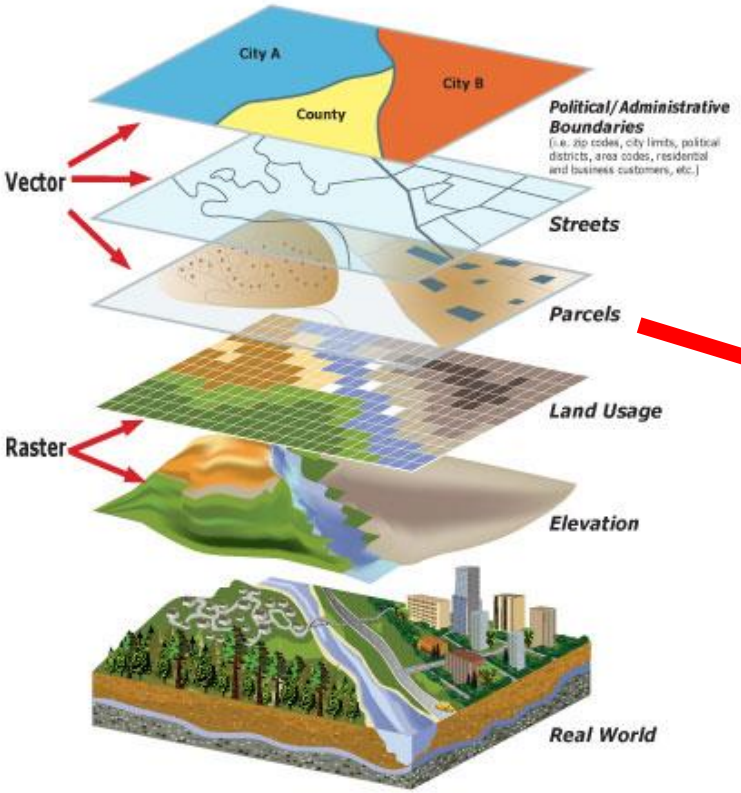
- File geodatabase
  - A folder of files that is cross platform
  - Datasets can be up to 1 TB in size (with configuration larger)
  - Best default choice
- Personal geodatabase
  - Stores data in Microsoft Access
  - Limited to ~250 – 500 MB for the entire geodatabase
  - Only supported on Windows

Multi User

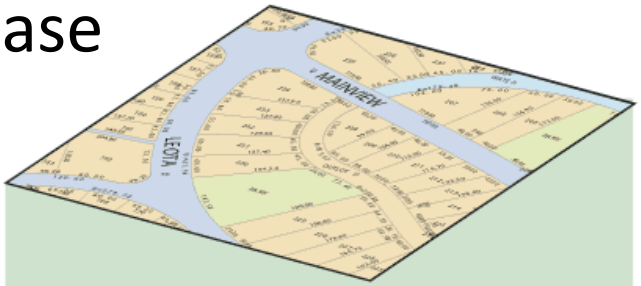
- Enterprise geodatabase
  - Required for multiple simultaneous users
  - Implemented within Relational Database Management System (RDBMS) software
  - Required for archive, replication, SQL querying, multi-user editing

What are the thematic layers I need to include?

- Roads, streams, soils, terrain?



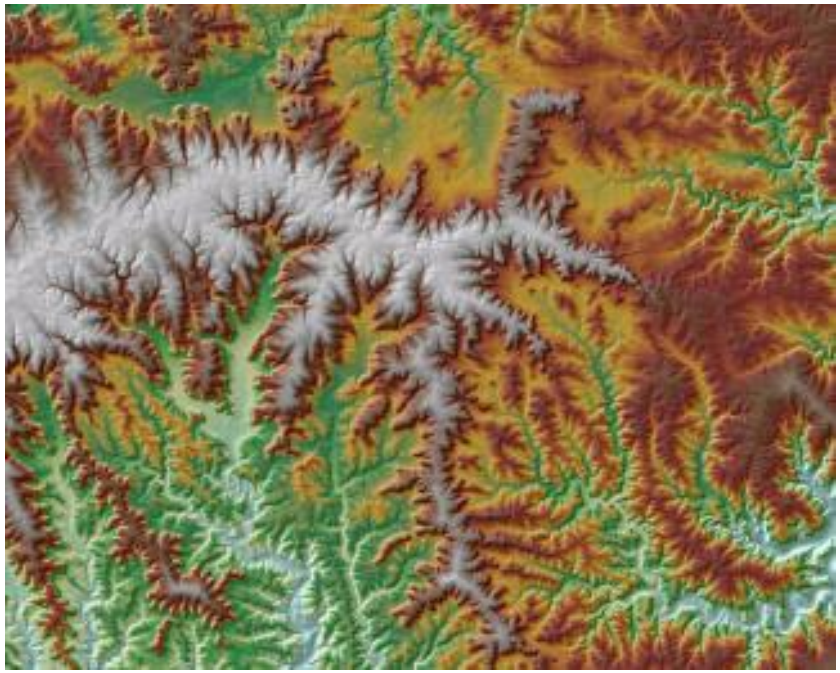
- Develop specification for representing the contents of each thematic layer in the database

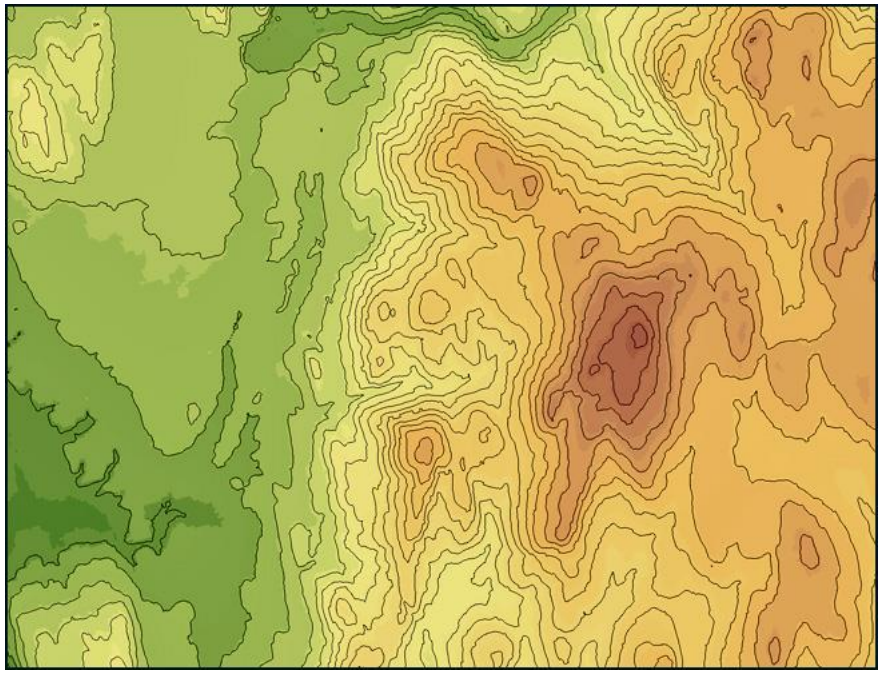


Layer	**Ownership parcels**
Map use	Parcels define land ownership and are used for taxation
Data source	Compiled from land ownership transactions at local government
Representation	Polygons in survey-aware feature classes and related annotation
Spatial relationships	Parcel Polygons cannot overlap and are covered by boundary lines
Map scale and accuracy	Typical map scales are 1:1200 and 1:2400
Symbology and annotation	Parcels often drawn using boundary features and related annotation

What is the correct geographic representation for each thematic layer?

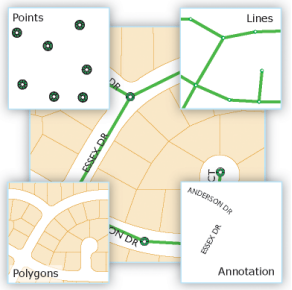
- Elevation as a raster of elevations or as a vector dataset of elevation contours? Both?
- Can be determined by how you intend to use the data



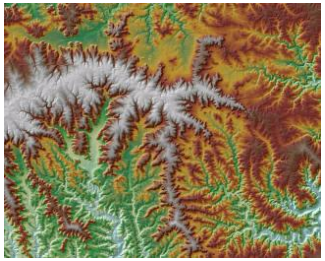


What object types do I need in my database?

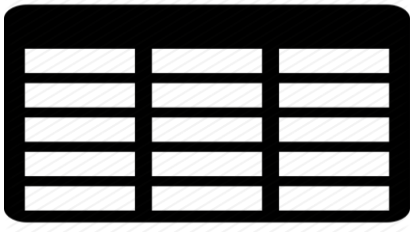
Vector Datasets



Rasters



Non-Spatial Tabular Data



Geodatabase

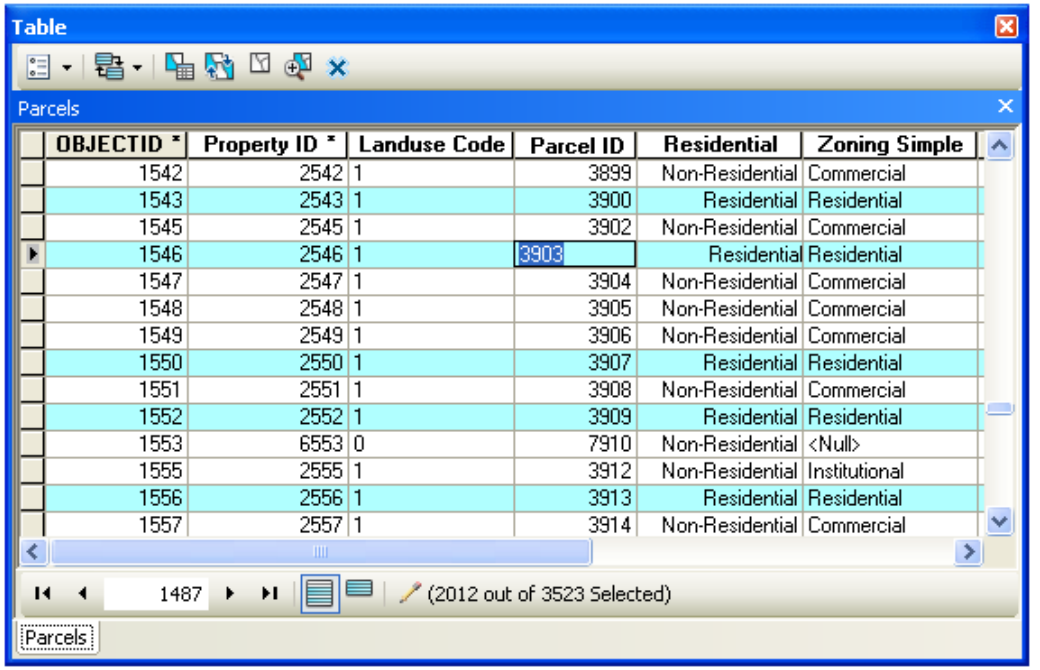
Feature Classes

Raster Datasets

Object Class (Table)

What are the attribute fields and their data types for each dataset?

- What columns do I need in the table for each dataset?
- What data type should each column in a table have?
  - Numbers (multiple)
  - Text
  - Dates
  - BLOBs
  - Object IDs (automatic)
  - Global IDs
  - Raster
  - Geometry



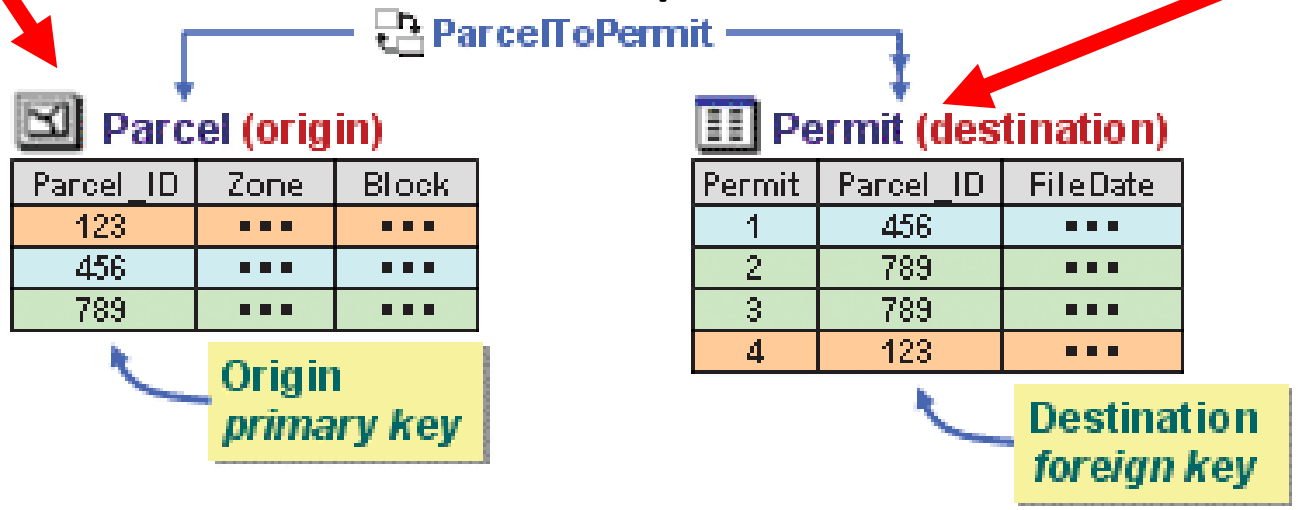
Are there important relationships I need to capture?

- Are there tables containing additional data related to the geospatial location data?

Parent Feature Class

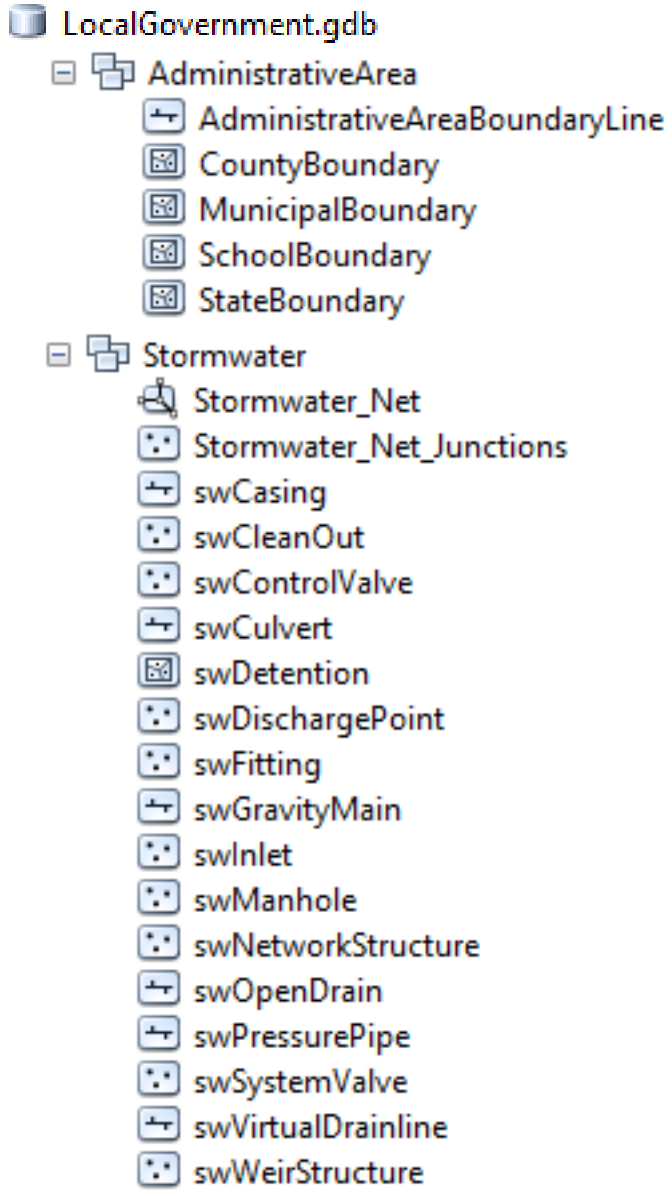
Table with related information

Relationship Class



How should my datasets be organized?

- Are there spatially related feature classes that should be grouped into a feature dataset?
- Advanced topics: are you using one of the following?
  - Topology
  - Network dataset
  - Terrain dataset (TIN)



Do you need geodatabase elements to facilitate data editing and manage data integrity?

- Coded value domains – require attribute values to be selected from a list

- Subtypes – require features to be one of a set of available subtypes

- Relationships – allow records in one table only when related records exist in another

Coded value domain
TSDataType

Description
Field type *Long integer*
Split policy *Default value*
Merge policy *Default value*

Code	Description
1	Instantaneous
2	Cumulative
3	Incremental
4	Average
5	Maximum
6	Minimum

Line feature class
Boundary

Subtypes are Water, Right of Way, Subdivision Boundary, Parcel Split, Lot Line, Parcel, Private Road

Create a New Geodatabase

- Three methods:
  1. Design and create a new empty geodatabase
  2. Copy and modify an existing geodatabase schema (empty database)
  3. Create a copy of the schema and contents from an existing geodatabase

ArcGIS Data Models

- ESRI has developed a series of geodatabase model templates
- These can be used as-is or as the basis for a more advanced implementation
- To use an existing ESRI data model:
  1. Download the appropriate data model http://support.esri.com/datamodels
  2. Create an empty file geodatabase
  3. Import the schema and set up the appropriate spatial reference for its contents. See http://desktop.arcgis.com/en/desktop/latest/manage-data/geodatabases/copying-the-schema-of-a-geodatabase-about-copying-.htm
  4. Load some data
  5. Test and refine

Existing ESRI Data Models

- Address
- Agriculture
- Atmospheric
- Basemap
- Biodiversity
- BroadbandStat
- Building Interior Space
- Carbon Footprint
- Census – Admin Boundaries
- Defense – Intel
- Energy Utilities
- Environmental Regulated Facilities
- Fire Service
- Forest Service
- Forestry
- Geology
- Groundwater
- Health
- . . .

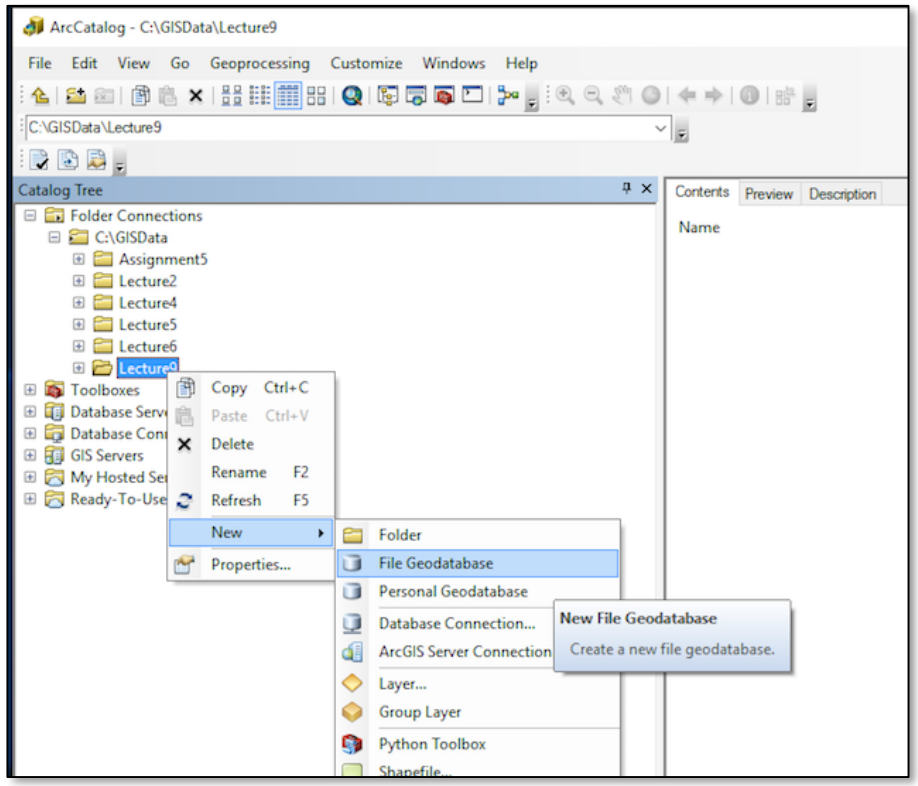
Learn More: http://support.esri.com/en/knowledgebase/techarticles/detail/40585

Creating a Geodatabase

- A useful and more in depth tutorial: http://desktop.arcgis.com/en/desktop/latest/manage-data/geodatabases/a-quick-tour-of-the-building-geodatabases-tutorial.htm

Create a New File Geodatabase

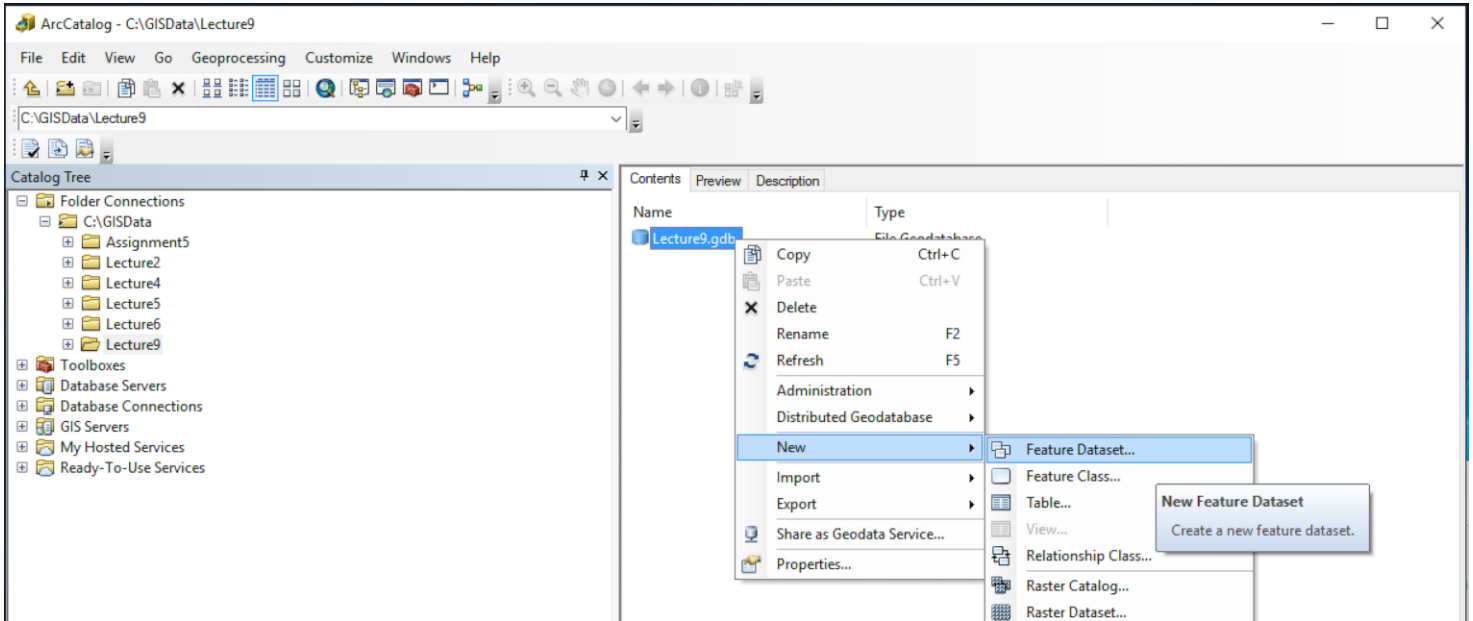
1. Open ArcCatalog (or the Catalog Window in ArcMap)
2. Navigate to folder where you want to create the geodatabase in the Catalog tree
3. Right click on the folder where you want to create the geodatabase, point to “New”, and click “File Geodatabase”



NOTE: You can rename a geodatabase by right clicking on the geodatabase and selecting “Rename” from the context menu

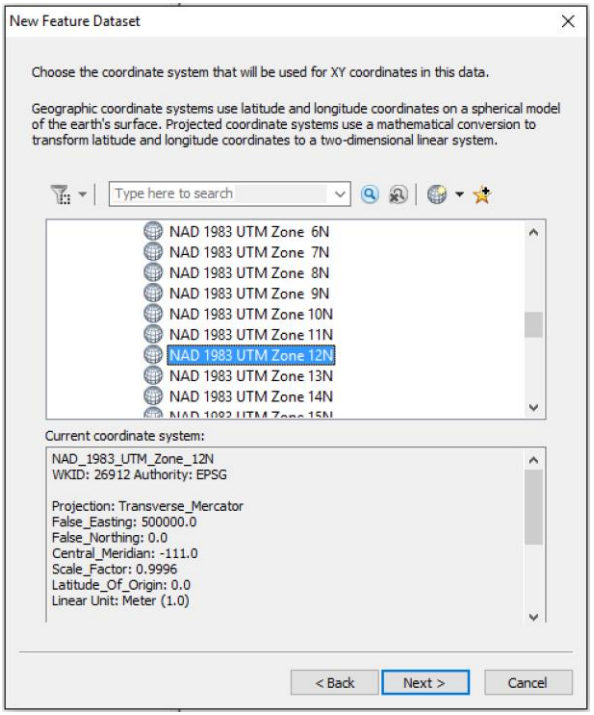
Create a Feature Dataset

1. Right click on the geodatabase
2. Click New --> Feature Dataset
3. Give your Feature Dataset a name
4. Choose a coordinate system (all of the feature classes in your feature dataset must use the same coordinate system)



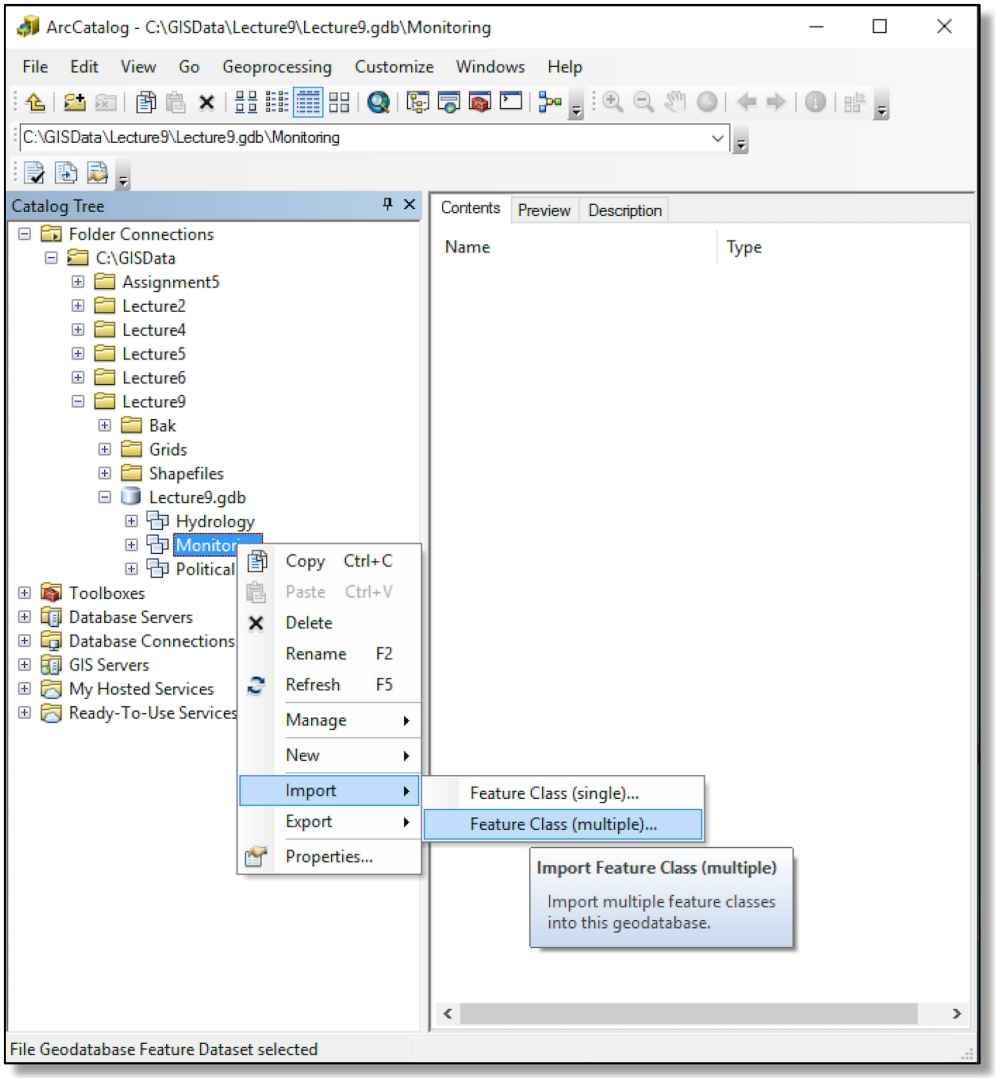
Choose Coordinate Systems

- XY Coordinate System: NAD 1983 UTM Zone 12N
- Z Coordinate System: NAVD 1988



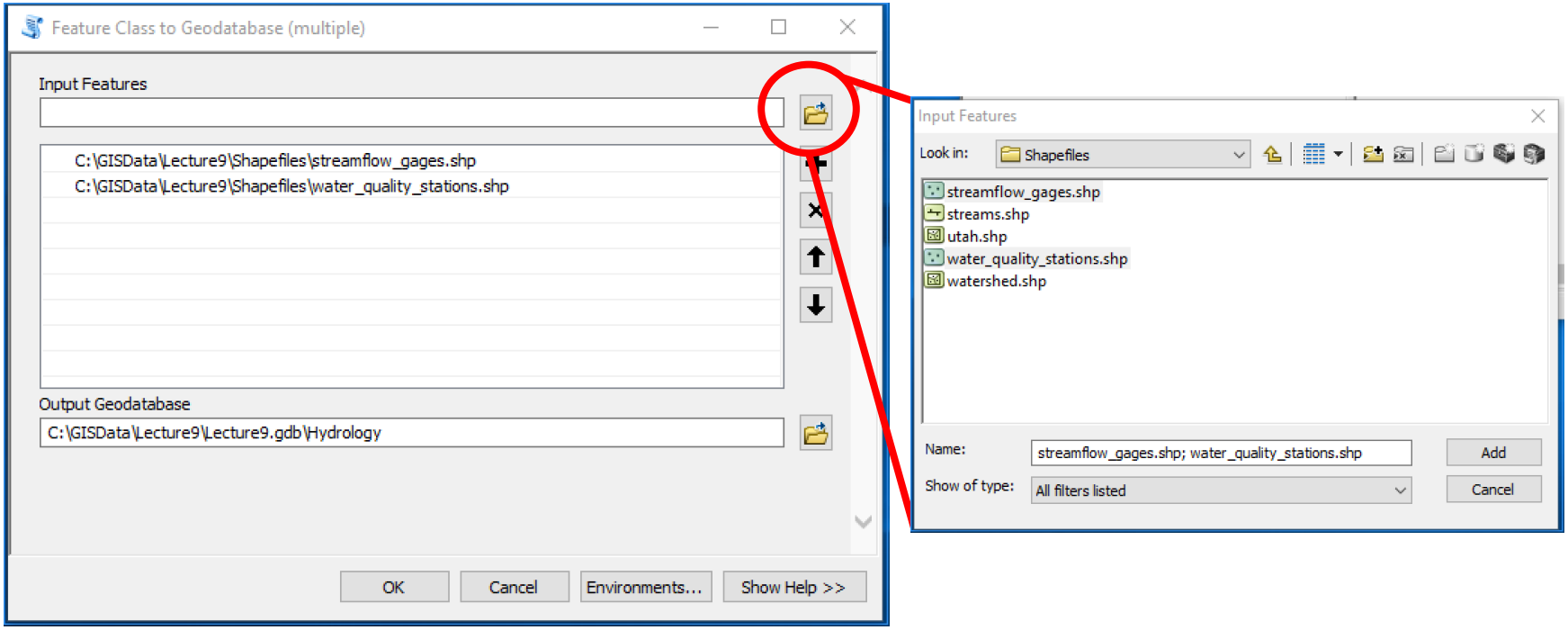
Create Feature Classes

- Two methods
  - Create new, empty feature class
  - Import feature class from shapefiles or layer files
- Right click on Feature Dataset
  - New
  - Import



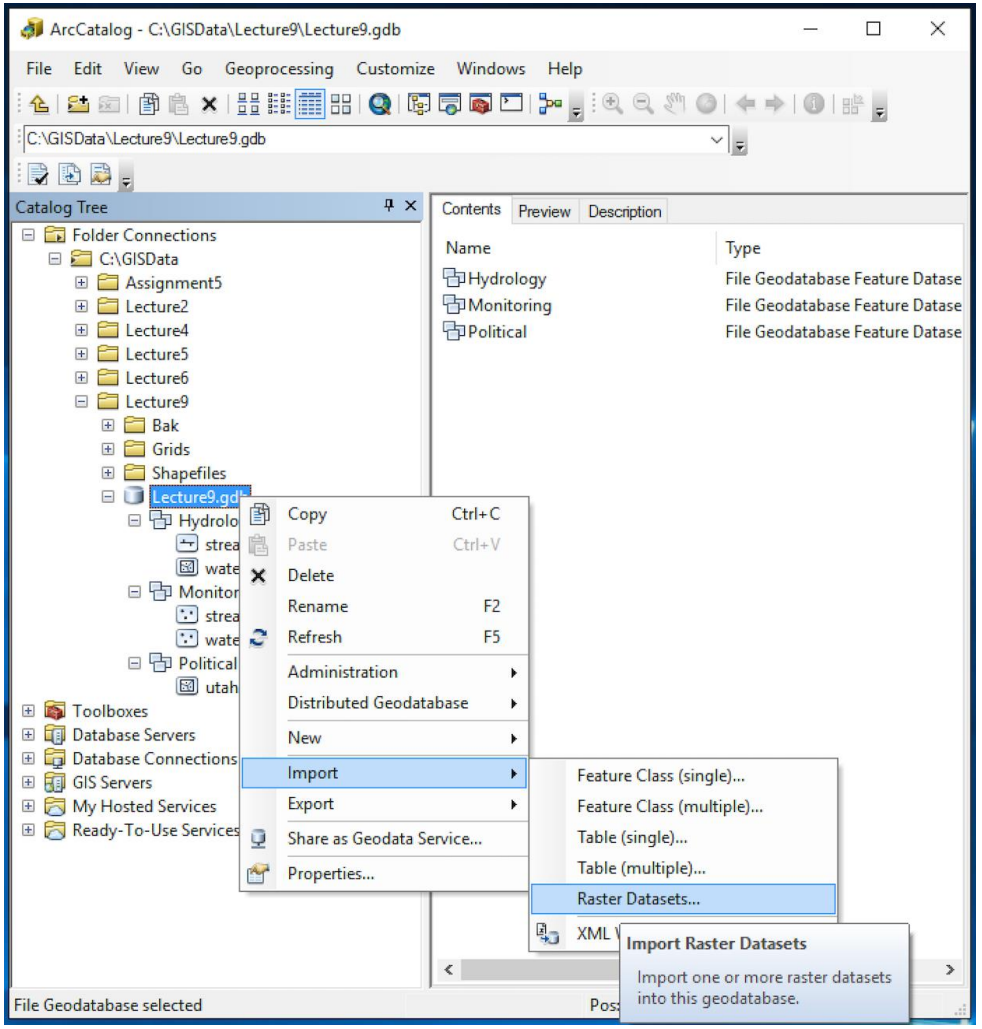
Import Feature Classes

- Import Feature Class (multiple)
- Choose feature classes to import



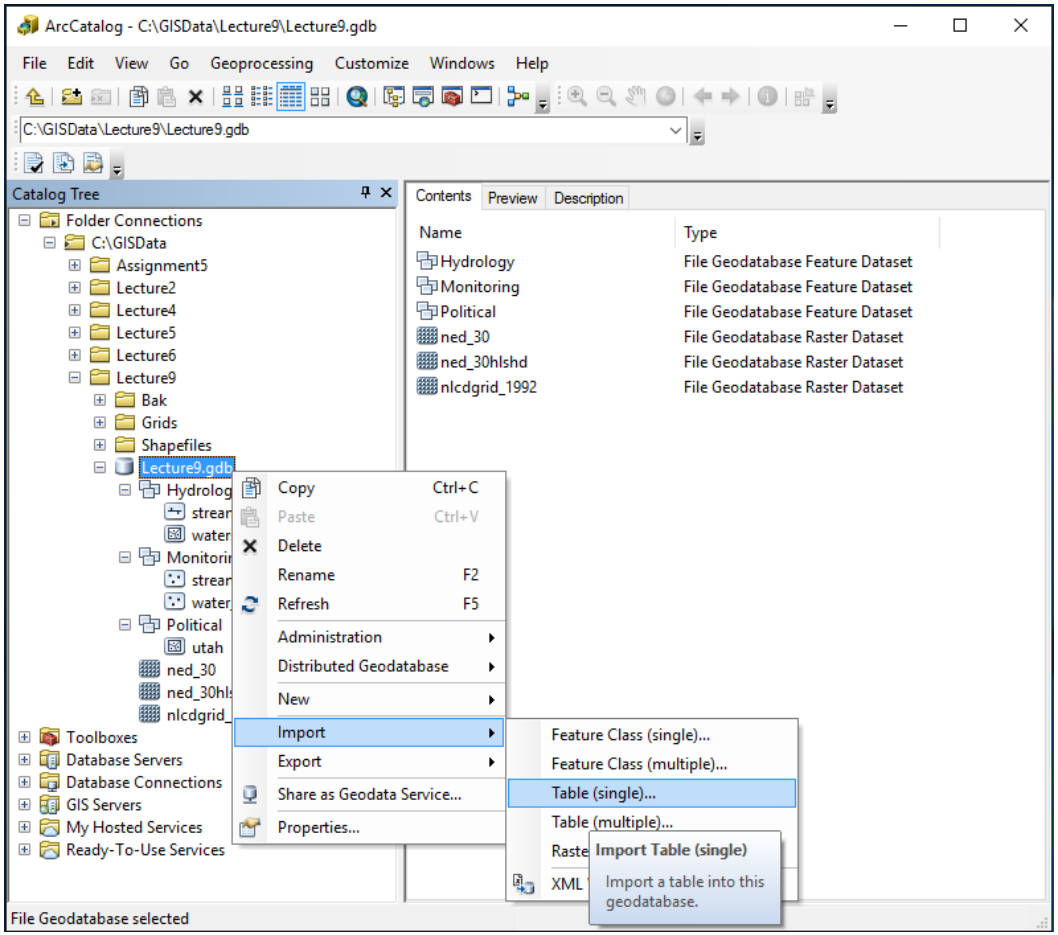
Create a Raster Dataset

- Right click on geodatabase
- Import --> Raster Datasets…



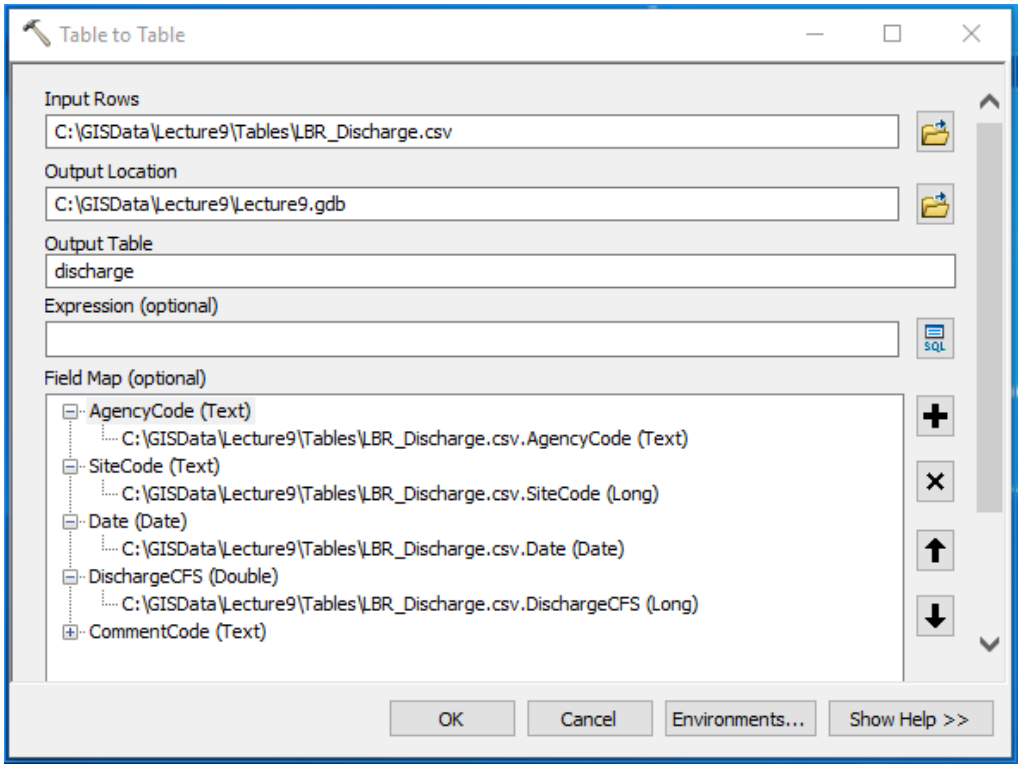
Create a Table

- Right click on geodatabase
  - New --> Table: create a new empty table
  - Import --> Table: import a table from an existing file



Import Table from File

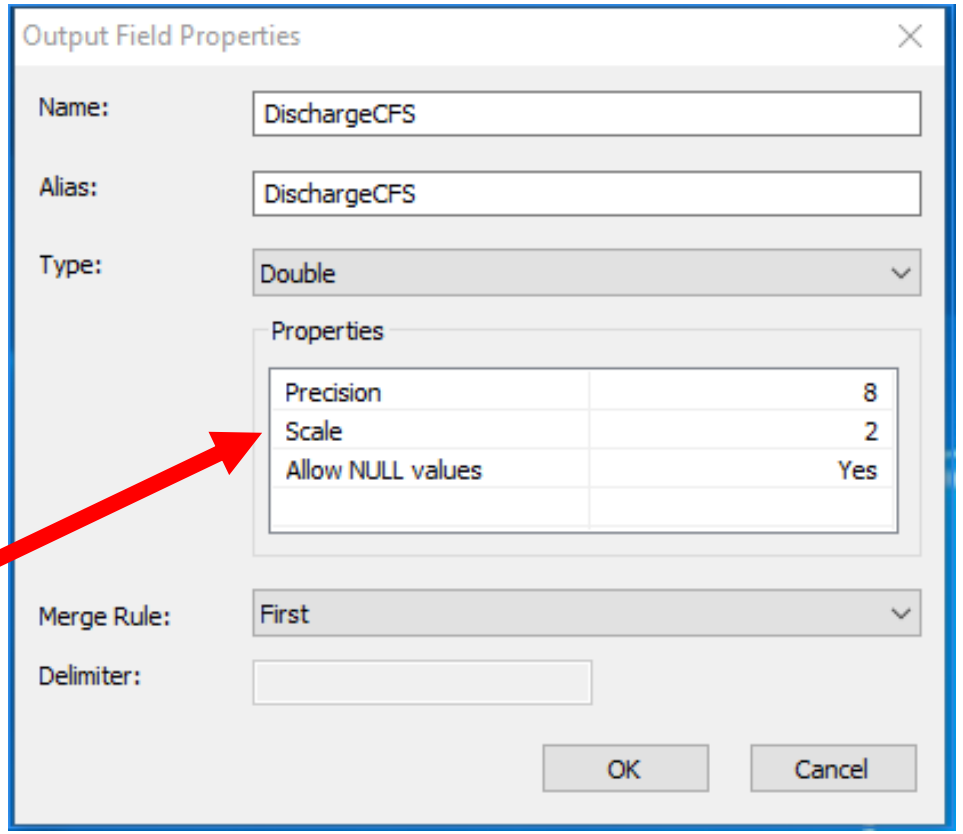
- Supported formats:
  - Excel (if you have it installed) (.xlsx)
  - Comma separated values (.csv)
  - dBASE (.dbf)
  - A couple of other less common formats
- Must specify:
  - Input rows (the input file)
  - Output location (the geodatabase)
  - Output table (name)
  - An expression (optional)
  - Field data type mapping (optional)



Some Notes about Data Type Mapping

- AgencyCode = Text
- SiteCode = Text (Length =10)
- Date = Date
- DischargeCFS = Double (Precision = 8, Scale = 2)
- CommentCode = Text

Right click on the Field name and select “Properties” to modify the data type mapping

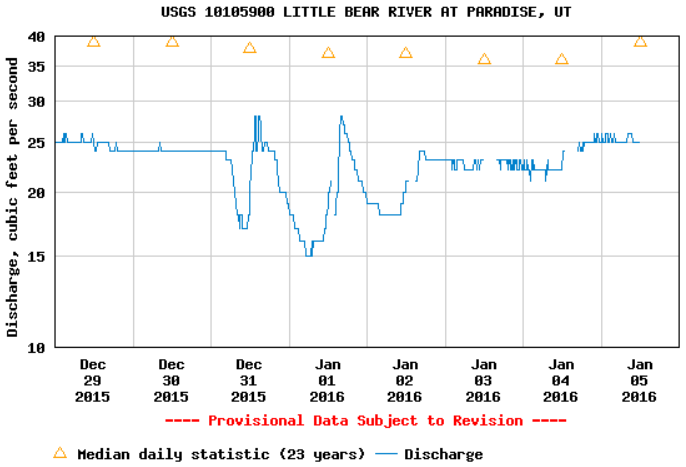


Length: Total number of characters
Precision: The total number of digits stored (total field length)
Scale: The total number of decimal places
Allow Nulls: Can the field be empty?

Create a Relationship Class

- Right click on geodatabase
- New --> Relationship Class

Connect streamflow gage locations to discharge values measured at those sites



SitesToDischarge

Table

streamflow_gages

OBJECTID *	Shape *	STATION *	STATION_NA
1	Point	10104600	SOUTH FORK LITTLE BEAR RIVER NEAR AVON, UT
2	Point	10104700	LITTLE BEAR RIVER BELOW DAVENPORT CREEK NEAR AVON, UT
3	Point	10104900	EAST FORK LITTLE BEAR RIVER ABOVE RESERVOIR NEAR AVON, UT
4	Point	10105000	EAST FORK LITTLE BEAR RIVER NEAR AVON, UT
5	Point	10105900	LITTLE BEAR RIVER AT PARADISE, UT
6	Point	10106000	LITTLE BEAR RIVER NEAR PARADISE, UT
7	Point	10107500	LITTLE BEAR RIVER NEAR HYRUM, UT

0 (0 out of 9 Selected)

streamflow_gages | discharge

Table

discharge

OBJECTID *	AgencyCode	SiteCode *	Date	DischargeCFS	CommentCode
1	USGS	10104600	7/1/1966	13	A
2	USGS	10104600	7/2/1966	13	A
3	USGS	10104600	7/3/1966	13	A
4	USGS	10104600	7/4/1966	13	A
5	USGS	10104600	7/5/1966	13	A
6	USGS	10104600	7/6/1966	13	A
7	USGS	10104600	7/7/1966	12	A
8	USGS	10104600	7/8/1966	12	A

0 (0 out of 66305 Selected)

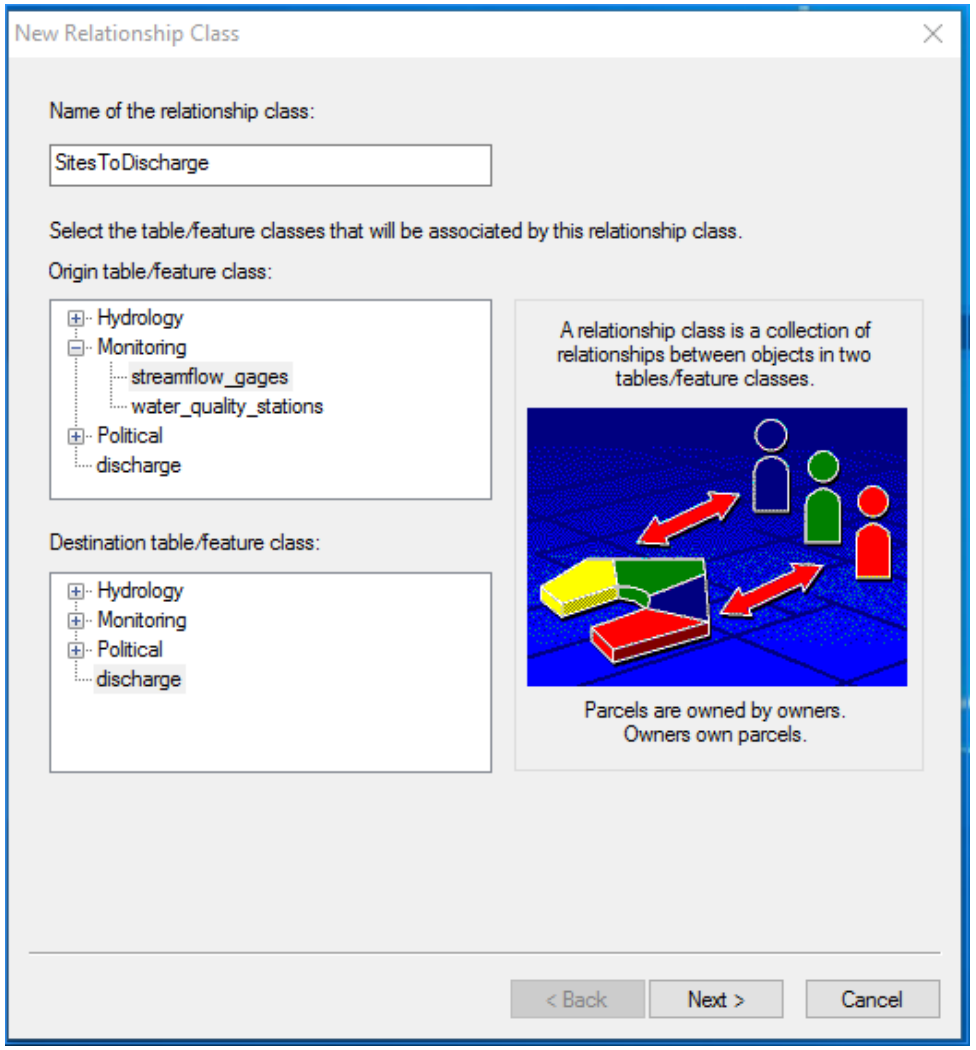
streamflow_gages | discharge

streamflow_gages (origin)

discharge (destination)

Origin and Destination

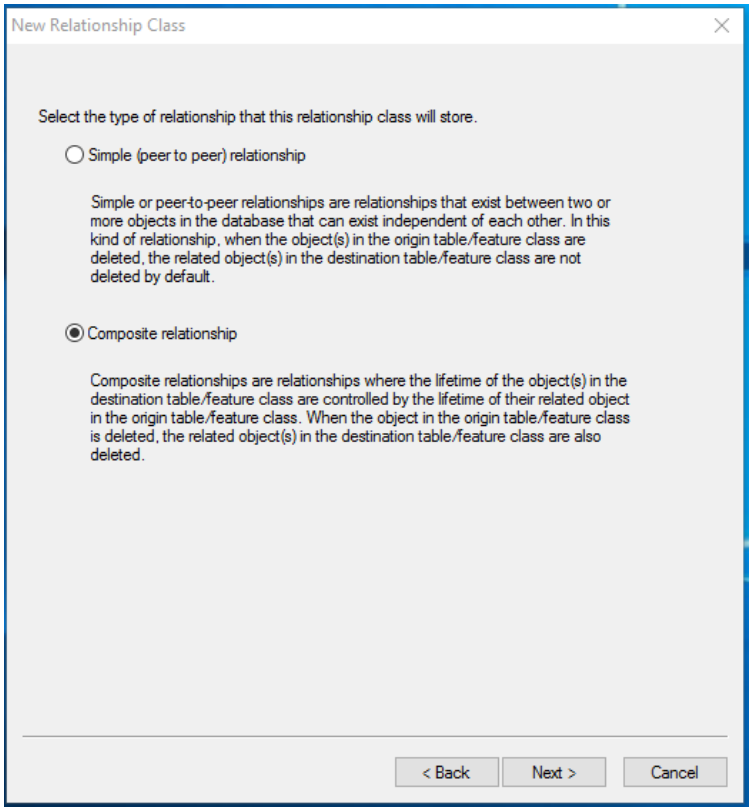
- Origin table/Feature class: “streamflow_gages”
- Destination table/Feature class: “discharge”



Relationship Types

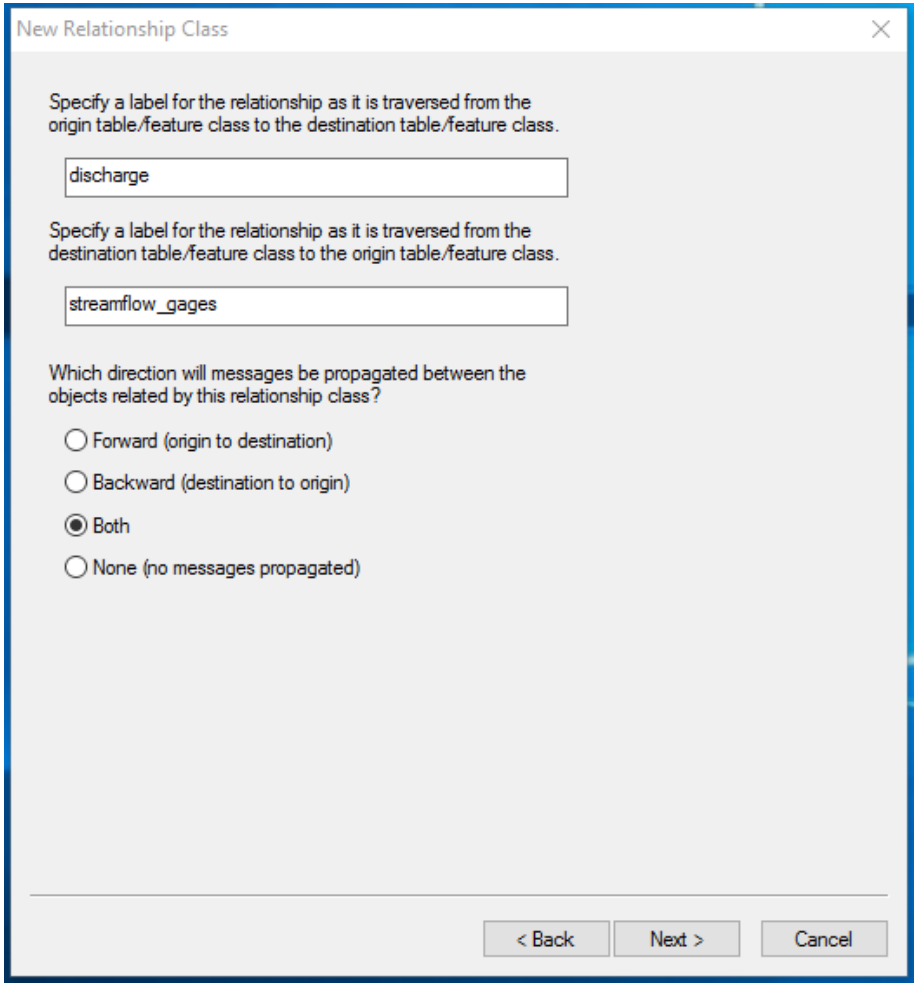
- Simple
  - No referential integrity
  - Both tables can exist independent of the other
- Composite
  - Existence of records in the destination table depend on a related record in the source table

Example: Can discharge values exist in the database if there is no related gage site where they were measured?



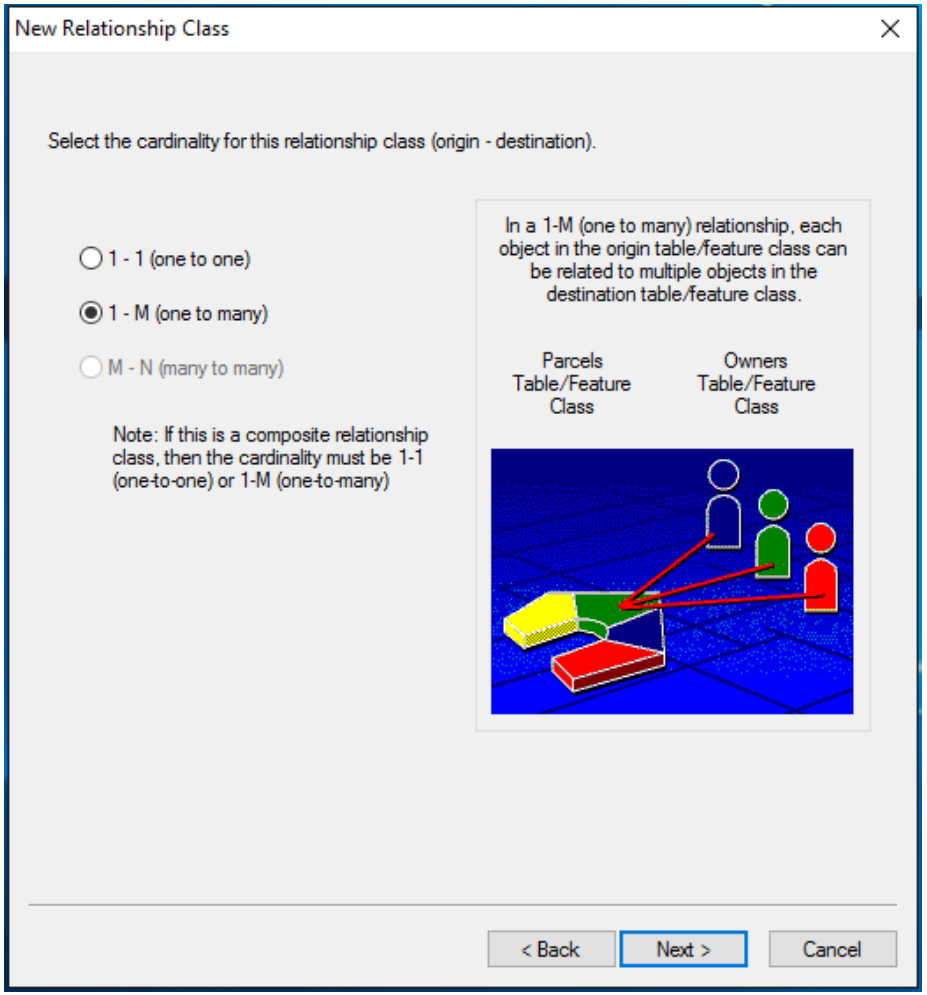
Labels and Message Propagation

- When you select a gage site, do you want related discharge values to be automatically selected?
- When you select discharge values, do you want the related site(s) to be automatically selected?



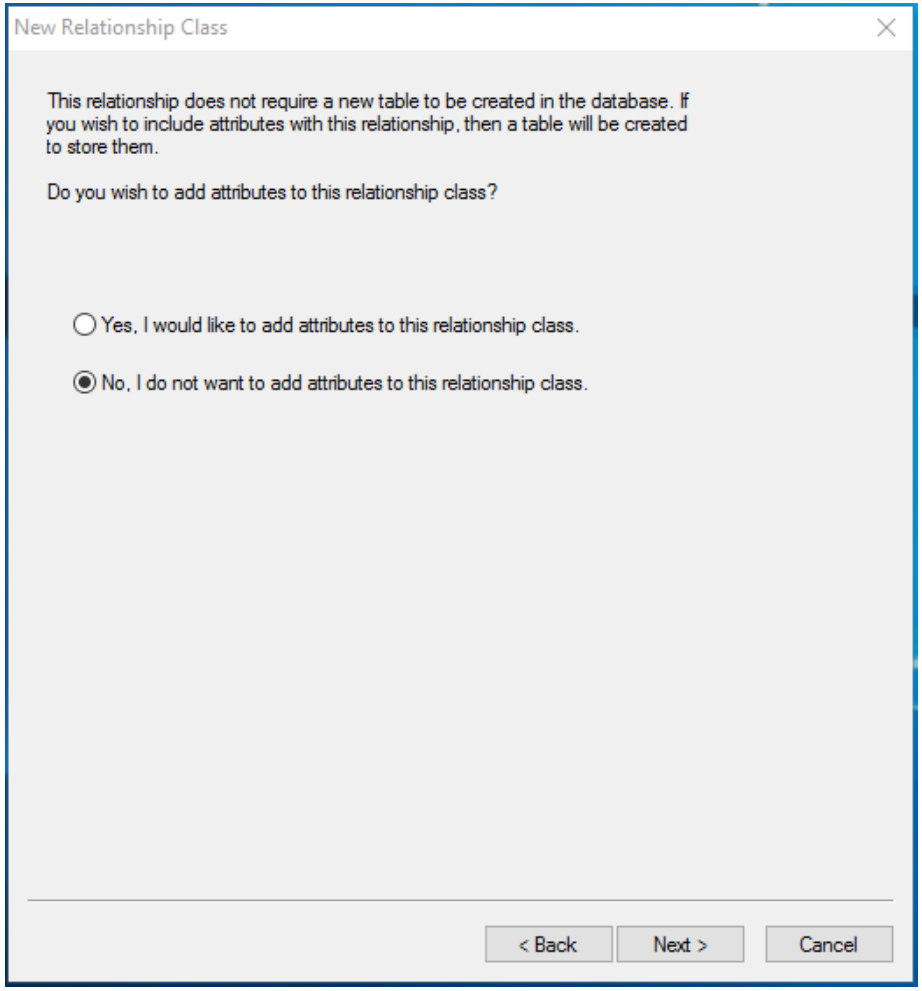
Cardinality

- **1 – 1 (one to one)**: There is one related record in each table on both sides of the relationship
- **1 – M (one to many)**: One record in the table on one side of the relationship is related to many records in the table on the other side
- **M – N (many to many)**: A complex relationship that requires a bridge table



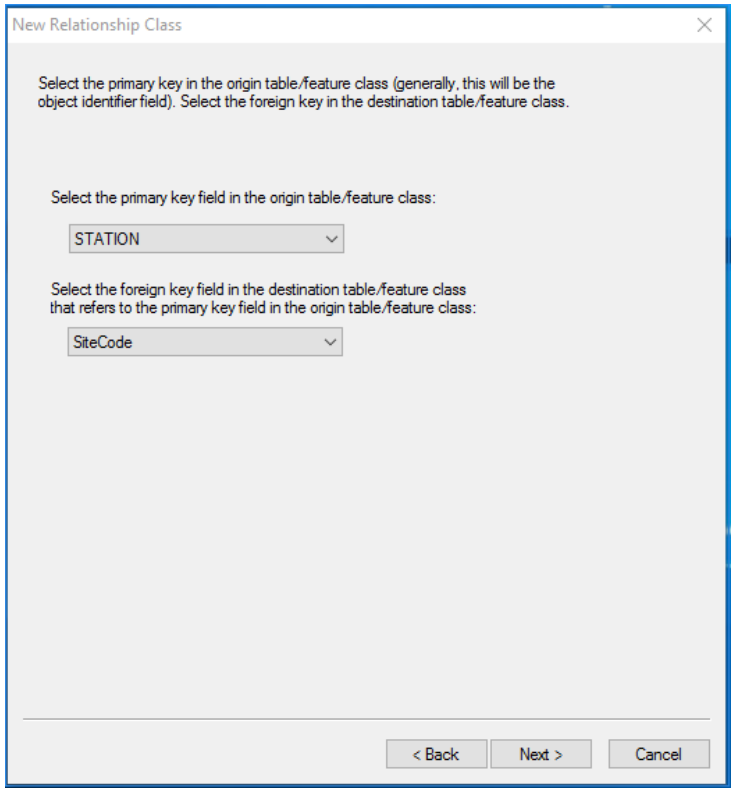
Attributes for the Relationship

- Can be stored in an additional table
- Not required for most relationships
- A new “bridge” table is required for “many-to-many” relationships



Primary and Foreign Keys

- The attribute columns on which the tables are linked
- Do not have to have the same name
- **Must** have the same data type



Origin Primary Key

SitesToDischarge

Destination Foreign Key

Table

streamflow_gages

OBJECTID *	Shape *	STATION *	STATION_NA
1	Point	10104600	SOUTH FORK LITTLE BEAR RIVER NEAR AVON, UT
2	Point	10104700	LITTLE BEAR RIVER BELOW DAVENPORT CREEK NEAR AVON, UT
3	Point	10104900	EAST FORK LITTLE BEAR RIVER ABOVE RESERVOIR NEAR AVON, UT
4	Point	10105000	EAST FORK LITTLE BEAR RIVER NEAR AVON, UT
5	Point	10105900	LITTLE BEAR RIVER AT PARADISE, UT
6	Point	10106000	LITTLE BEAR RIVER NEAR PARADISE, UT
7	Point	10107500	LITTLE BEAR RIVER NEAR HYRUM, UT

0 (0 out of 9 Selected)

streamflow_gages | discharge

Table

discharge

OBJECTID *	AgencyCode	SiteCode *	Date	DischargeCFS	CommentCode
1	USGS	10104600	7/1/1966	13	A
2	USGS	10104600	7/2/1966	13	A
3	USGS	10104600	7/3/1966	13	A
4	USGS	10104600	7/4/1966	13	A
5	USGS	10104600	7/5/1966	13	A
6	USGS	10104600	7/6/1966	13	A
7	USGS	10104600	7/7/1966	12	A
8	USGS	10104600	7/8/1966	12	A

0 (0 out of 66305 Selected)

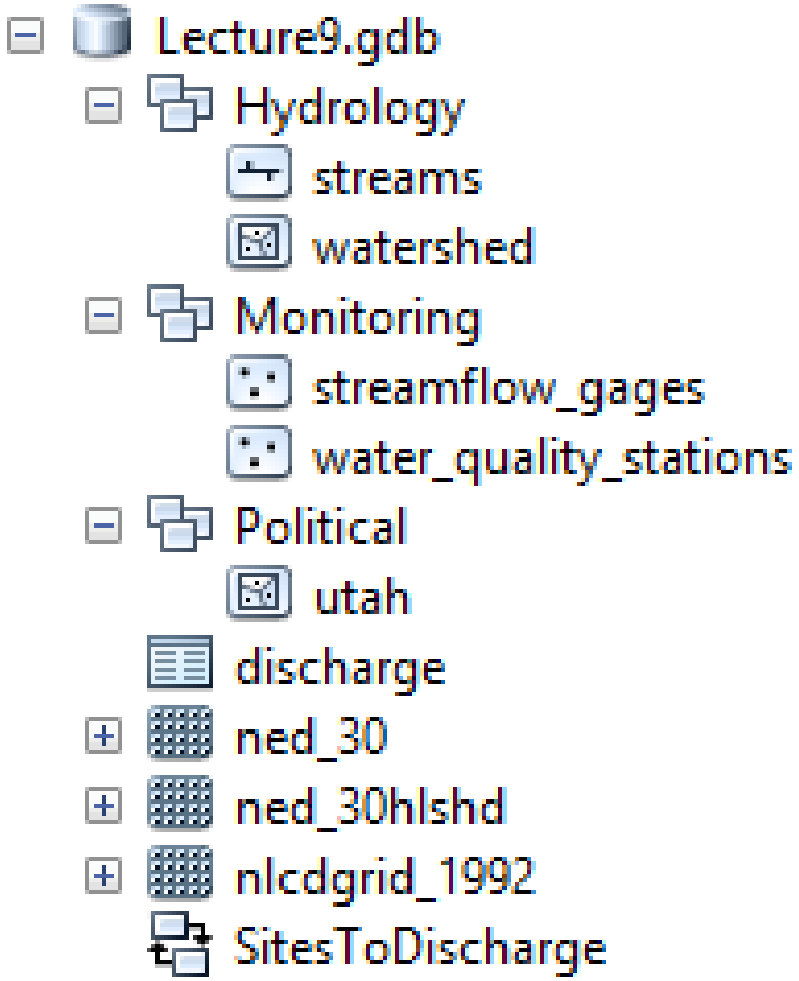
streamflow_gages | discharge

streamflow_gages (origin)

discharge (destination)

Result

- Three Feature Datasets
- Five Feature Classes
- Three Raster Datasets
- One Table
- One Relationship Class



Advantages of File Geodatabases vs. Personal Geodatabase and Shapefiles

- Compatible with Windows and Linux platforms
- Optimized performance (outperform in both size limitations and speed)
- Storing both vector and raster datasets together
- Less restrictive editing locks (per table rather than entire database)
- Can use data compression to reduce storage requirements
- Data integrity tools (e.g., coded domains, relationships)
- Simpler to create relationships

Summary

- Design considerations for geodatabases include
  - Geodatabase type
  - Thematic layers to include
  - Geographic representation of thematic layers
  - Selection of object types, attributes and relationships
- Geodatabases enable organization of thematic layers (e.g., Feature Datasets)
- ArcGIS provides tools for creating geodatabases and importing data
- A number of geodatabase data models are already available for use
- Geodatabase design and implementation is an iterative process that may ned to be refined over time

Additional Resources

- Geodatabase design steps: http://desktop.arcgis.com/en/desktop/latest/manage-data/geodatabases/geodatabase-design-steps.htm

Assignment 3 – Building a Geodatabase

- Finish building the geodatabase we began today in class
  - Feature datasets and feature classes
  - Raster datasets
  - Object class (table) and relationship
- Choose a study area of some sort
- Create a new file geodatabase for your study area
- Create 1 or more Feature Datasets
- Create or load at least 2 feature classes
- Download the NED DEM for your study area and load it into your geodatabase