Introduction to GIS: Previewing GIS Data

GIS Data
- Basic Data Models: Vector, Raster
- Data Types: Shapefiles, Feature Classes, Rasters, Tables

Exploring Data
- Previewing & Documenting Geographic Data
- Acquiring Geographic Data

ArcCatalog
- Using GIS to Model Real-world Data
- Browsing, Managing, and Documenting Geographic - based Data

Information systems help us to manage what we know by making it easier to organize, access, manipulate, and apply knowledge to the solution of problems. *Geographic* Information Systems (GIS) allow us to do all of these things and more. The following are a few useful definitions of GIS:

- A container of maps in digital form
- A computerized tool for solving geographic problems
- A tool for revealing what is otherwise invisible in geographic information

This tutorial will enable you to identify different forms of GIS data, preview the grographic extent of the data, examine the attribute table associated with a particular dataset, and view the metadata, or information about the GIS data. Through the activities, you will become familiar with ArcGIS software and to some of the basic concepts of GIS. ArcGIS is a software suite that is composed of three applications: **ArcCatalog**, for data management; **ArcMap**, for visualizing and analyzing data; and **ArcToolbox**, for advanced geoprocessing and data conversion.

This tutorial provides step-by-step exercises that will familiarize you with the core tools and functionality of ArcCatalog. More conceptual discussions – those in blue text – are interspersed throughout to provide context for the tools and to introduce you to GIS terminology.

Before you begin, make sure that you have a personal copy of all the tutorial data in a directory that you can modify and save to.

ArcCatalog

ArcCatalog is an application for **organizing** and **managing** the various datasets and documents used in a GIS. It provides an integrated and unified **view** of all the data files, databases, and GIS documents available to ArcGIS users. In this application you will find the necessary tools which allow you to:

- Browse and find geographic information
- Store, view, import, and manage metadata
- Define, export, and import data models and datasets
- Search for and discover GIS data on local networks and the Web

Geographic datasets, unlike other data (a photo or Word document), often consist of a set of files, rather than a single file. When listed in Windows Explorer, the datasets appear as a list of system folders and files, whereas ArcCatalog displays and manages the datasets as single entities. Whenever possible, use ArcCatalog to move and organize your geographic data.

1. Start ArcCatalog by double-clicking your ArcCatalog desktop icon or by clicking Start > All Programs > ArcGIS > ArcCatalog10.3.1.

The ArcCatalog window has a catalog display for looking at data, a catalog tree for browsing folders and data, and several toolbars (fig. 1). You will use all three to explore data in our project folder.

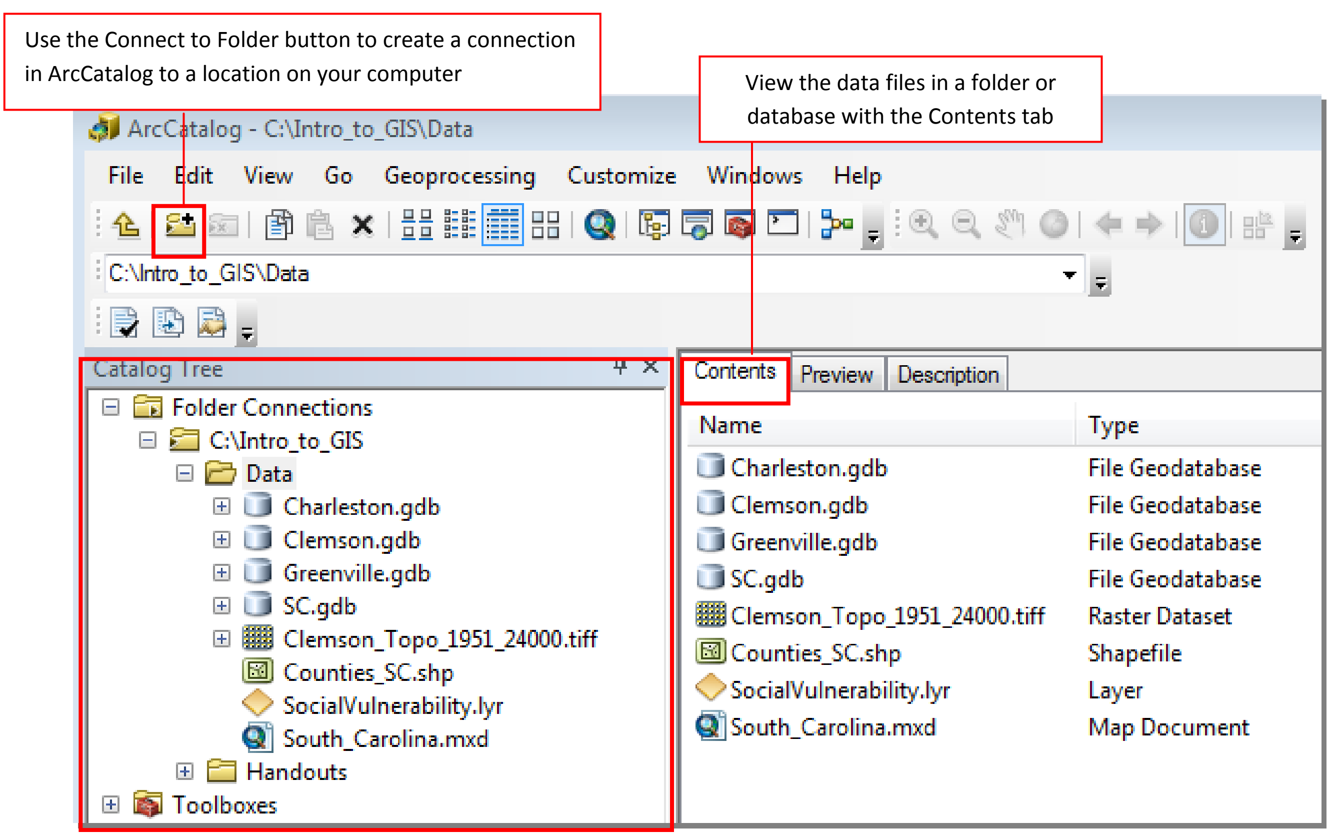


Figure 1: The ArcCatalog interface

1. The first step in using ArcCatalog is to establish connections to the folders and datasets you will be working with. Hit the **Connect to Folder** button .

2. In the Connect to Folder window that opens (fig. 2), **navigate to the location of the tutorial data**.

3. Select the folder and hit OK to close the Connect to Folder window.

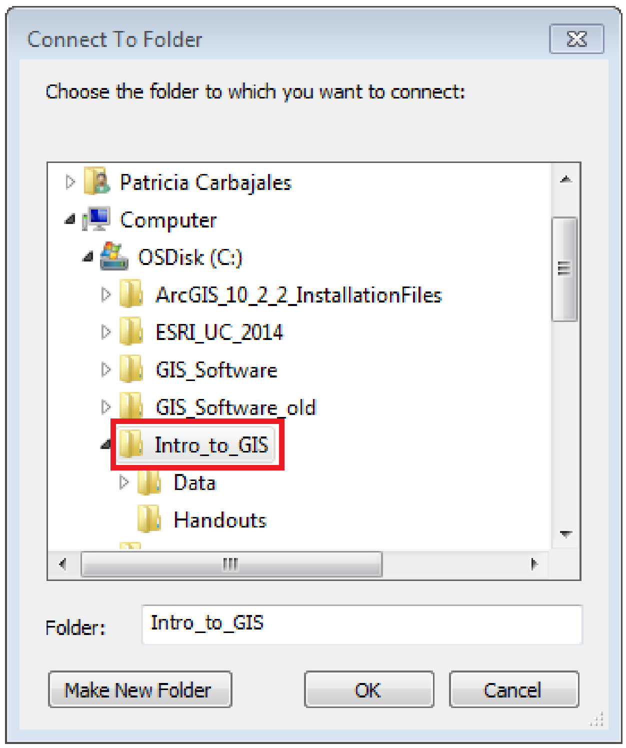
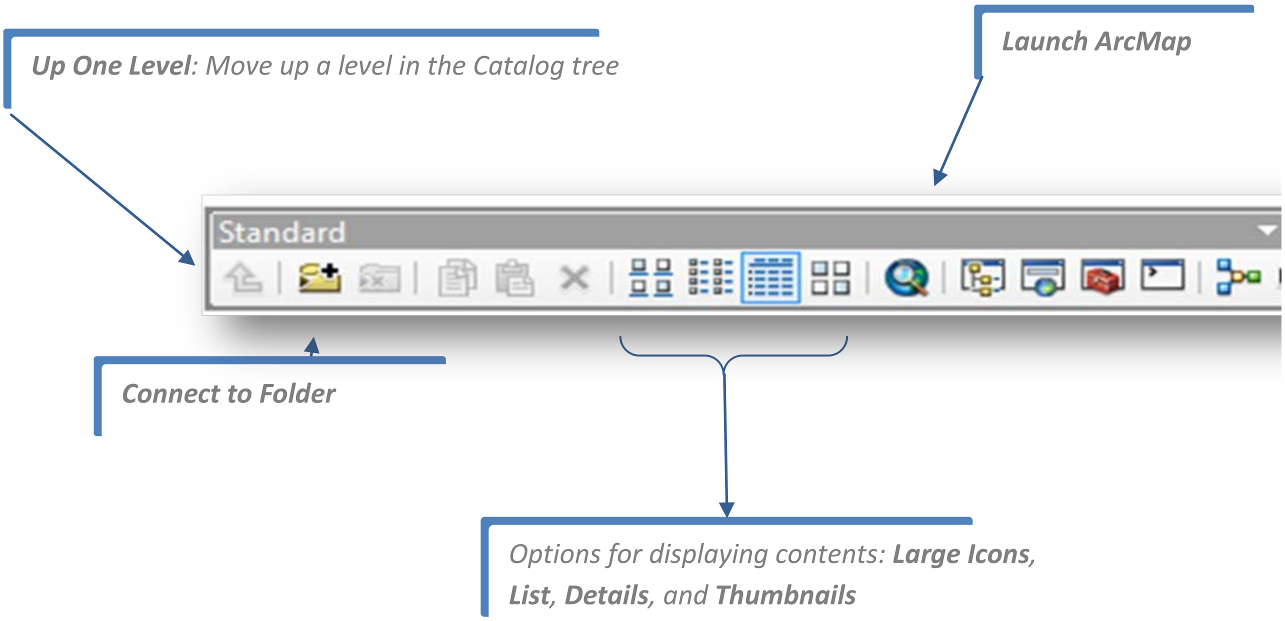


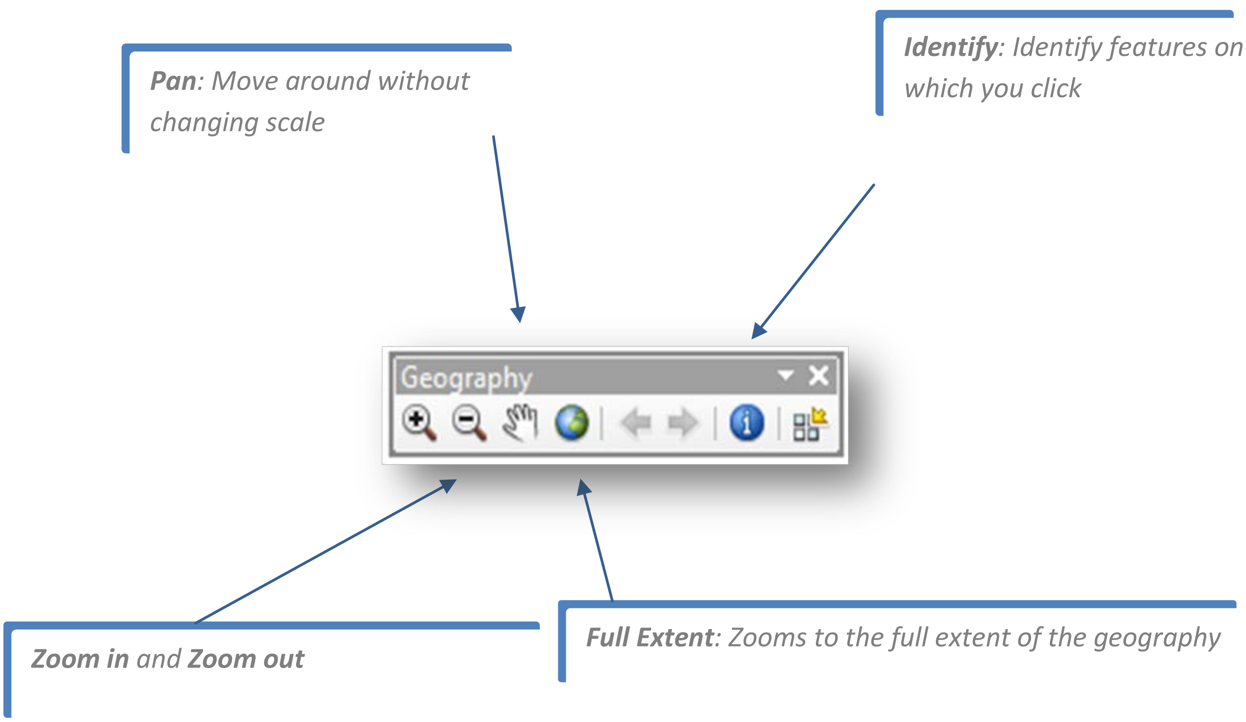
Figure 2: Connect to Folder window

4. Find the new connection listed in the catalog tree.

5. Select your connection and view the files in the display under the **Contents tab**. You should see files similar to those listed in Figure 1.

There are two main toolbars that you will utilize frequently in ArcCatalog: the *Standard* and the *Geography* toolbars. The **Standard toolbar** provides tools for connecting to data (you used Connect to Folder from this toolbar), for changing the display of data, and for opening other GIS applications. The **Geography toolbar** provides tools for zooming in and out, for panning, and for returning to full extent view of the data.





When ArcCatalog is first opened, the default view under the Contents tab is a list of small icons. You can change this view using the Standard toolbar, much like you can change the view of folders in Windows Explorer.

6. Change your view of the data in your project folder by toggling through the display options in the **Standard toolbar**. These display options are available only when you are viewing multiple data files under the **Contents tab**. Your four options include the default list of small icons, large icons, a detail view showing file type, and a thumbnail view. Choose the view you prefer to work with.

7. When trying to find GIS data, it is much more efficient to work on the Catalog Tree side (left side) than on the contents side. Any ⊞ symbol on the Catalog Tree next to a folder means that there are data or folders inside that folder. When you click on that ⊞ sign, you expand its contents. Click on the on the ⊞ sign next to **Intro to GIS > Data** (indicates it is a folder), then expand **Charleston** (indicates it is a Geodatabase) and then expand **Administrative** (indicates it is a Feature Dataset). Select **CityCouncil** in the catalog tree (represents a polygon Feature Class). Under the contents tab you will see a thumbnail of this file (fig.3).

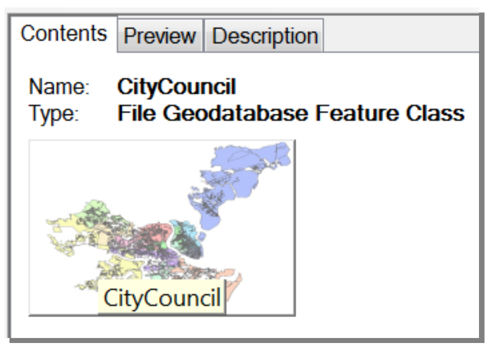


Figure 3: Single file viewed in Contents tab

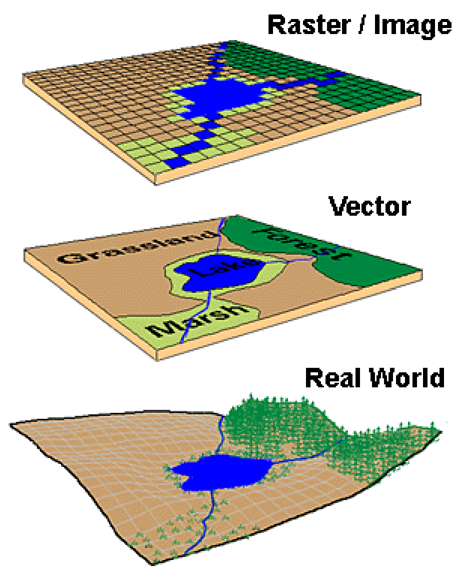
8. Now move to the **Preview tab**. Here you see that CityCouncil is a file storing information about the Council District boundaries within the City of Charleston. The display under the Preview tab is where you can preview any geographic data you obtain.

Part 2: Using GIS to Model Real-World Data

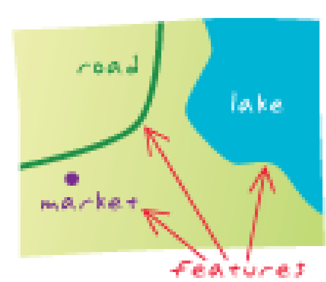
Geographic data link place, time, and attributes. Geographic Information Systems represent these three properties into two basic data types: spatial and attribute data. **Spatial data** describe the relative and absolute locations of geographic features. **Attribute data** provide additional context and meaning to spatial data. Attribute data can include any piece of information relevant to a location, including a date or range of dates. In GIS, attribute data are typically appended to spatial data.

Spatial Data: the Vector and Raster Data Models

Spatial data are displayed in GIS using both vector and raster data models. You may be familiar with the terms vector and raster because they refer to computer graphics formats. In GIS, these two data models allow us to view and analyze spatial data very differently.



The **vector data model** is used for spatial data that can be represented as one of three geometry types – points, lines, or polygons – and each has its own unique set of characteristics. These are well suited for representing features with discrete boundaries, such as wells, streets, rivers, states, and parcels. In GIS, geometric objects with vector geometry are referred to as features. A feature[1] is simply an object that stores its geographic representation, which is typically a point, line, or polygon.



[1] ESRI defines feature as a representation of a real-world object on a map. In ArcGIS, a collection of geographic features with the same geometry type (such as point, line, or polygon), the same attributes, and the same spatial reference are called a feature class.

9. Return to ArcCatalog, and go back up one level to see the contents of the Charleston geodatabase in the catalog tree.

10. Select the **Addresses** feature class. Under the **Preview tab** in the catalog display, you will see the points that compose this feature class. Points are used for objects too small to be polygons, such as trees, city locations, and schools. The Addresses feature class represents all of the street addresses in the City of Charleston.

11. Now expand the **Transportation** feature dataset, and select the **Streets** feature class. This feature class comprises lines that represent the road centerlines in Charleston.

12. Select **Buildings**. In this feature class, the boundary area for each building in Charleston is represented by a polygon.

The **raster data model** stores data in a rectangular grid of identically-sized cells. Surfaces and information about surfaces – like elevation and slope – are modeled using rasters. These spatial data represent geographic objects with numeric values, as opposed to a distinct shape. What they contain are measurable values for any particular location on the earth's surface. Therefore rasters are useful way of representing quantifiable values at a location. Phenomena such as rainfall, temperature and wind speed are also well-suited to the raster format. Aerial photography, satellite imagery and scanned maps are examples of common raster data in GIS.

Print maps are the foundation of historical GIS[2]. They are scanned and stored as digital images, and eventually entered into a GIS as raster data. There are a few steps that must be taken before a scanned map can be utilized in GIS; we will explore those steps in a later tutorial. For now, you will view a few rasters in the Data folder.

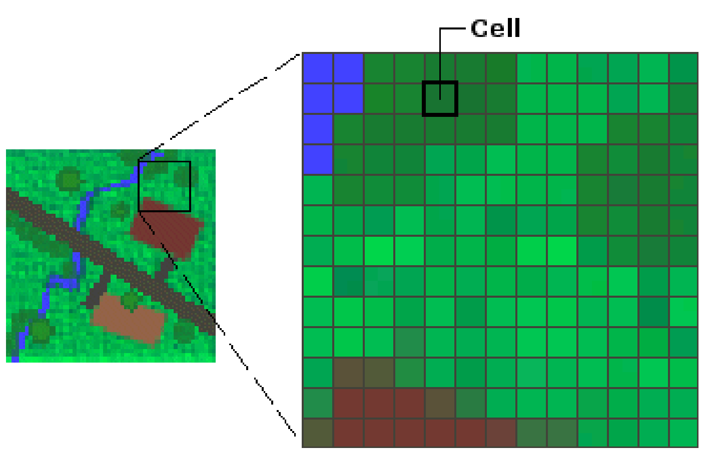


Figure 4: Raster data model

[2] Robert Churchill and Amy Hillier, among others, confirm this in *Placing History*. See their chapter, "Teaching with GIS," for more information about finding historical spatial data.

13. Return to ArcCatalog and select **Clemson_Topo_1951_24000.tiff** in the Catalog Tree. This is a raster dataset in a TIFF format, which is a common format for raster datasets. The original file was a pdf that was exported as an image file (TIFF). This format stores raster data with no reduction in image quality, but the file size is quite large.

14. View and explore this raster in the preview tab. It is very large so give the image a few moments to load.

As the image loads, take a minute to find the Geography toolbar in ArcCatalog. You will use the tools in this toolbar to examine the image.

15. Use the **Zoom In**  tool to view the map in greater detail. You use this tool by drawing a rectangular box around the area that you want to zoom to. Try using the **Zoom Out**  tool to change your view. Instead of clicking, get used to drag windows with these tools. You will end up saving a lot of time.

16. Zoom in to the bottom right corner of the image. Here you will find the title of the map, the year the map was published, and legend information. Also notice at the right corner of the map the two sets of map coordinates to help locate these geographic data on the globe (fig. 5).

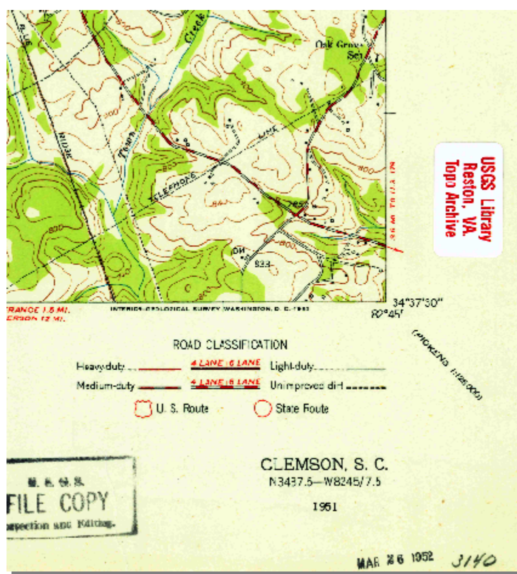


Figure 5: Map data in detail

17. Use the **Pan** tool  to scroll over to the left corner of the image. Along the way, you should see other information about the map: the map scale, information about the contour lines and about the standards used in the making of this map. The far left corner of the image (fig. 6) has information about the creation and symbology of the map. Notice that the left corner of the map also has geographic coordinates.

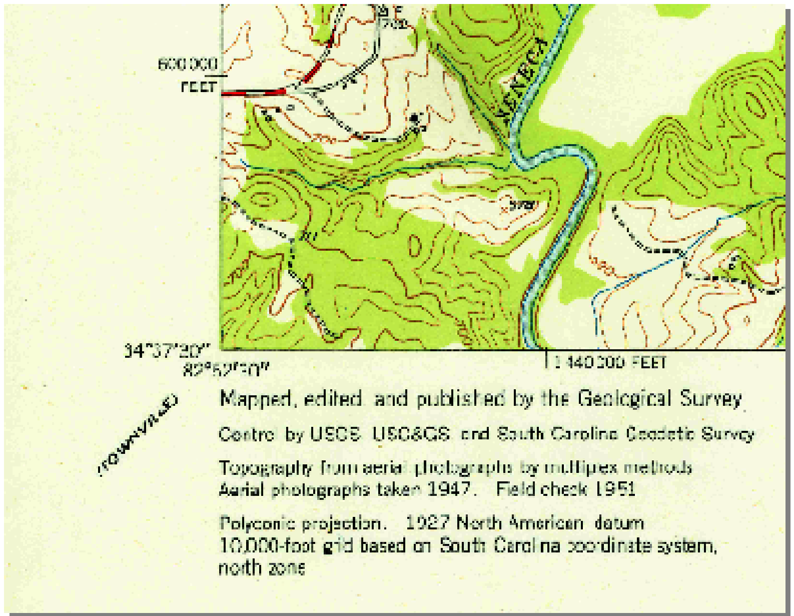
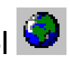


Figure 6

18. Now return to the original view of the whole map. Use the **Full Extent** tool  to do this.

Part 3: Data Formats

Vector representations of geographic objects are referred to as features. Vector data are referred to in a GIS most frequently as feature classes, but the term shapefile is also used. Both terms refer to vector data formats even though there are technical differences between the two formats.

In your Data folder, the vector data are stored in several geodatabases called **Charleston, Clemson, Greenville and SC**. A **geodatabase** is a type of database that allows for the storage, query, and manipulation of spatial data. With a geodatabase, you can provide a single spatial reference for multiple datasets, which makes it a useful tool for managing spatial data.

27. Return to ArcCatalog and select the **Clemson_University_Buildings_2011** feature class in the **Clemson** geodatabase.
28. View the data in the Preview tab. Select the **Identify** tool . Use this to click on any building feature in the display area to learn more about that feature (fig. 7).

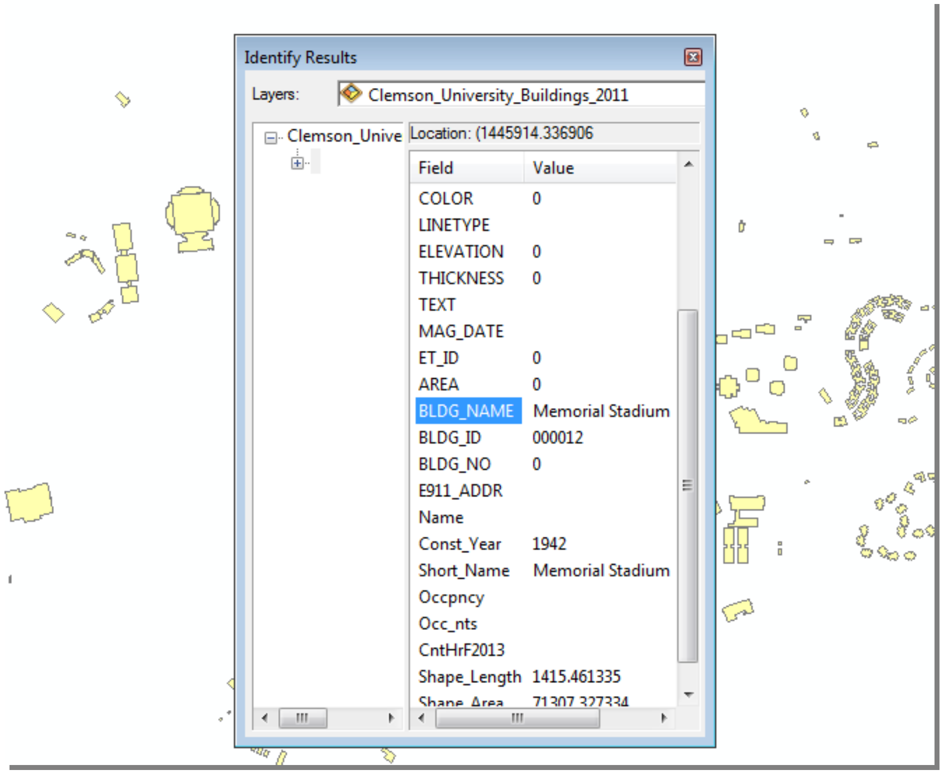


Figure 7: Using the Identify Tool to identify features

Attribute Data: Linking Features to Information

As discussed above, there's often more to a geographic feature than its location and shape—there is also all of the information associated with that feature, known as attribute data. For a road, this might include its name, speed limit, and whether it is one-way or two-way. For a city, this might include its population, demographic characteristics, and average monthly temperatures.

These additional data exist in a GIS as **tabular data**. If a table is directly associated with a feature class, the attribute information is stored in an attribute table.

29. Select the **Counties_SC** shapefile in the Preview tab. Make sure to view the whole extent of this layer (hint: use the Full Extent button)

30. Find the **Preview drop-down menu** at the bottom of the catalog display. Select **Table** from this menu (fig: 8).

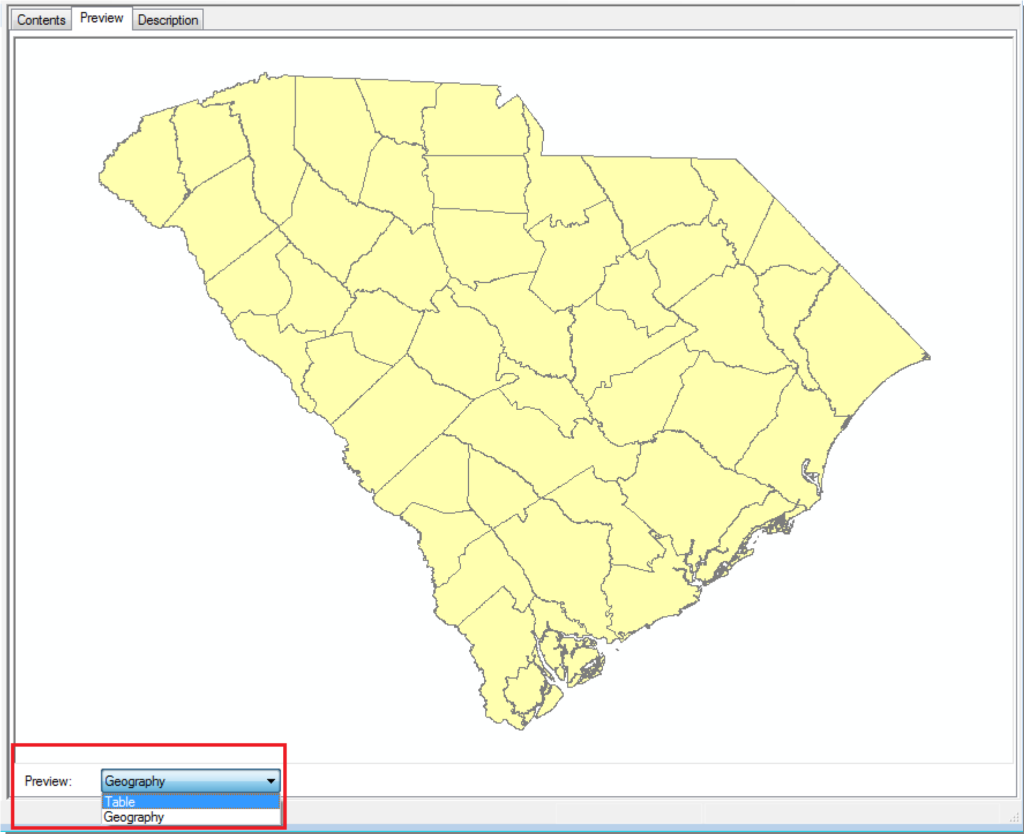
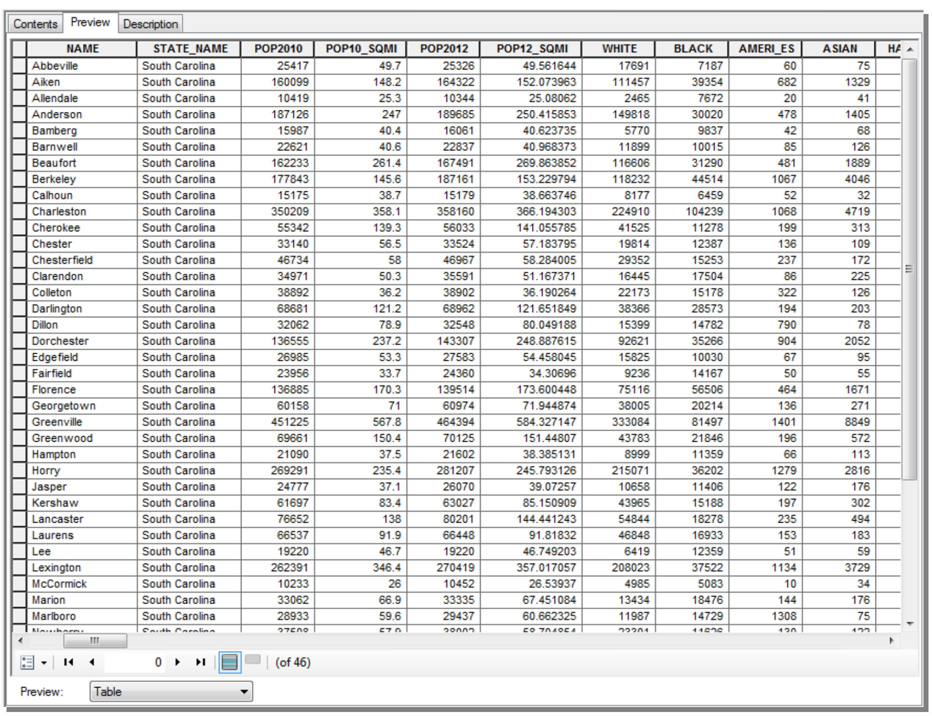


Figure 8: Preview drop-down menu with Geography and Table options

31. Now the display changes to show the **attribute table** associated with the Counties shapefile. As you can see, there are several fields in this attribute table. Each field contains different types of information (fig. 9).



NAME	STATE_NAME	POP2010	POP10_SQMI	POP2012	POP12_SQMI	WHITE	BLACK	AMERI_ES	ASIAN
Abbeville	South Carolina	25417	49.7	25326	49.561644	17691	7187	60	75
Aiken	South Carolina	160099	148.2	164322	152.073963	111457	39354	682	1329
Allendale	South Carolina	10419	25.3	10344	25.08062	2465	7672	20	41
Anderson	South Carolina	187126	247	189685	250.415853	149818	30020	478	1405
Bamberg	South Carolina	15987	40.4	16061	40.623735	5770	9837	42	68
Barnwell	South Carolina	22621	40.6	22837	40.968373	11899	10015	85	126
Beaufort	South Carolina	162233	261.4	167491	269.863852	116606	31290	481	1889
Berkeley	South Carolina	177843	145.6	187161	153.229794	118232	44514	1067	4046
Calhoun	South Carolina	15175	38.7	15179	38.663746	8177	6459	52	32
Charleston	South Carolina	350209	358.1	358160	366.194303	224910	104239	1068	4719
Cherokee	South Carolina	55342	139.3	56033	141.055785	41525	11278	199	313
Chester	South Carolina	33140	56.5	33524	57.183795	19814	12387	136	109
Chesterfield	South Carolina	46734	58	46967	58.284005	29352	15253	237	172
Clarendon	South Carolina	34971	50.3	35591	51.167371	16445	17504	86	225
Colleton	South Carolina	38892	36.2	38902	36.190264	22173	15178	322	126
Darlington	South Carolina	68681	121.2	68962	121.651849	38366	28573	194	203
Dillon	South Carolina	32062	78.9	32548	80.049188	15399	14782	790	78
Dorchester	South Carolina	136555	237.2	143307	248.887615	92621	35266	904	2052
Edgefield	South Carolina	26985	53.3	27583	54.458045	15825	10030	67	95
Fairfield	South Carolina	23956	33.7	24360	34.30696	9236	14167	50	55
Florence	South Carolina	136885	170.3	139514	173.600448	75116	56506	464	1671
Georgetown	South Carolina	60158	71	60974	71.944874	38005	20214	136	271
Greenville	South Carolina	451225	567.8	464394	584.327147	333084	81497	1401	8849
Greenwood	South Carolina	69661	150.4	70125	151.44807	43783	21846	196	572
Hampton	South Carolina	21090	37.5	21602	38.385131	8999	11359	66	113
Horry	South Carolina	269291	235.4	281207	245.793126	215071	36202	1279	2816
Jasper	South Carolina	24777	37.1	26070	39.07257	10658	11406	122	176
Kershaw	South Carolina	61697	83.4	63027	85.150909	43965	15188	197	302
Lancaster	South Carolina	76652	138	80201	144.441243	54844	18278	235	494
Laurens	South Carolina	66537	91.9	66448	91.81832	46848	16933	153	183
Lee	South Carolina	19220	46.7	19220	46.749203	6419	12359	51	59
Lexington	South Carolina	262391	346.4	270419	357.017057	208023	37522	1134	3729
McCormick	South Carolina	10233	26	10452	26.53937	4985	5083	10	34
Marion	South Carolina	33062	66.9	33335	67.451084	13434	18476	144	176
Marlboro	South Carolina	28933	59.6	29437	60.662325	11987	14729	1308	75

Figure 9: Previewing the attribute table behind the Counties dataset

32. Take a look at the **FID** field. This field is automatically generated by ArcGIS when a new shapefile is created. Each item in this shapefile – in this case, each county – is assigned a unique identifier.

33. The **Shape** field indicates the geometry type of the feature class. As you can see, Counties_SC is a polygon shapefile. This field is also automatically generated by ArcGIS when the shapefile is created.

34. The remaining fields in this attribute table were created by the author. The **Name** field is clear enough – it stores the name of each county. The **POP2010** field stores population information, but it is difficult to intuit anything more about its significance. You will take a look at the metadata in the next section to find more information about this field.

35. Scroll to the right and right-click on the field name **POP2012**. A context menu will pop up with a few tools and options (fig. 10).

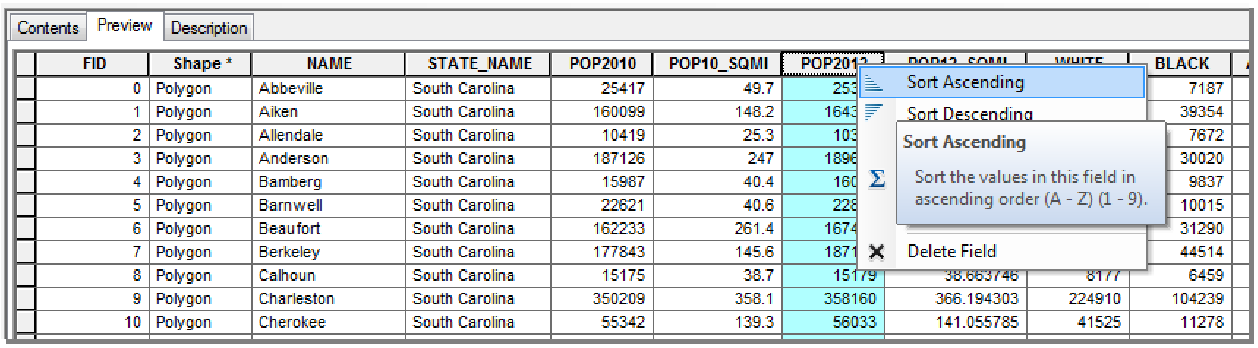


Figure 10: Sorting tabular data

36. Choose **Sort Ascending** from this menu. This field is now sorted numerically in ascending order.

Metadata

The word 'metadata' is broadly defined as data about data. One can also think of it as a historical record of the dataset itself. A metadata record for a GIS dataset can be very detailed, but it typically includes information about

- why the dataset was created or collected (its purpose)
- what geographic area the dataset covers (its geographic extent)
- who created/collected the dataset and when
- what processes were performed on the dataset

Metadata is a critical component of sharing data—people searching for a dataset first view its metadata in order to determine whether it is appropriate for the project. If a dataset lacks metadata, it may be used inappropriately—and any analysis or resulting measurements based on the data may be inaccurate. It is very important to know the projection and/or coordinate system of your data, which can be found in the metadata.

37. Let's return to the CityCouncil dataset in the Charleston geodatabase to investigate its metadata. Select the **Description tab** with the CityCouncil feature class selected in the catalog tree.

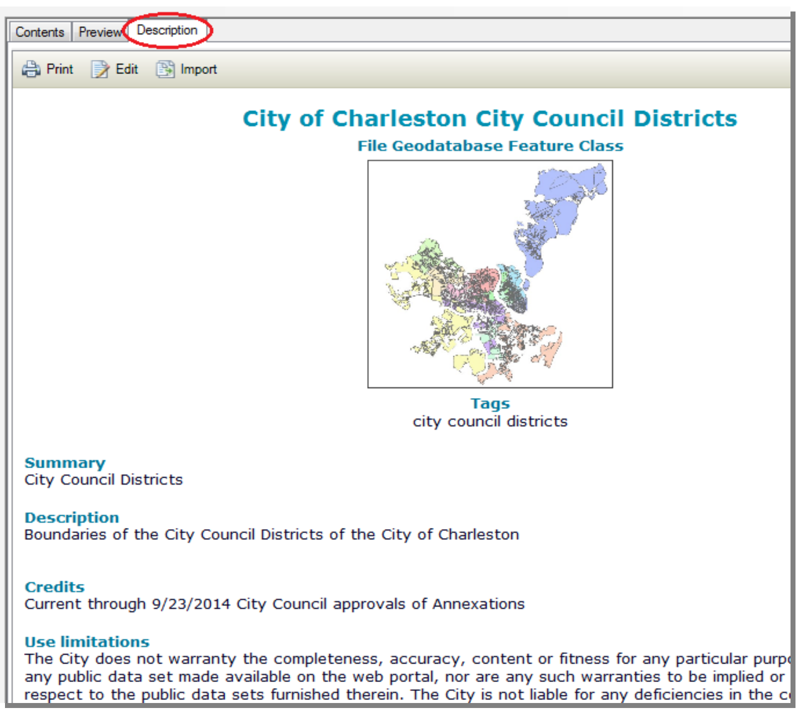


Figure 11: Options under Metadata Tab

38. A great deal of information about this dataset is displayed under the Metadata tab. There are three main sections under the Metadata tab: Summary, Description, and Access and use limitations (fig. 11). **Tip**: go to Customize – ArcCatalog Options – Metadata and change the Item to FGDC (you will see all the metadata fields for any GIS dataset).

39. It is the trademark of a good GIS professional to always dedicate time at creating and updating metadata for any GIS dataset created.