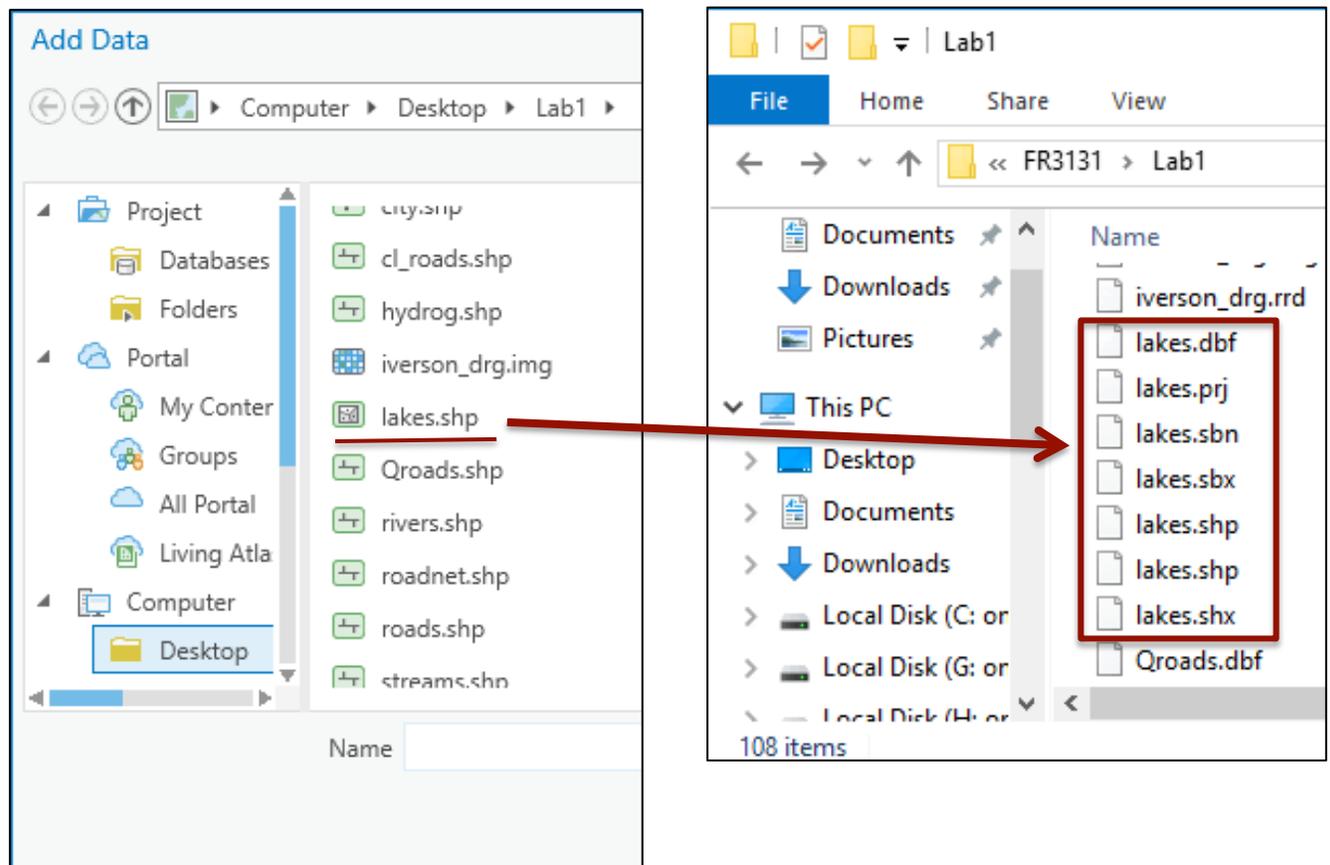


ArcGIS Pro Projects, Shapefiles, and GeoDatabases

A “shapefile” is actually a cluster of files. You can see this below. On the left is the window that displays when you add data in ArcGIS Pro. It shows just one entry for the lakes.shp layer. However, if we Windows File Explorer to display the same directory (below right), we can see several files with the name lakes, each with different extensions. In addition to the .shp, there is a .dbf, .prj, .shx, and others. A complete shapefile is made up of several files, at a minimum a .shp, .shx, and .dbf files. If you ever copy shapefiles using File Explorer or another generic file management tool, you need to copy ALL THE FILES, NOT JUST THE .SHP FILE. Otherwise, your copy will likely be incomplete, and useless.



ESRI also has another, newer file storage system, called a Geodatabase. This usually stores a cluster of files that are related by geography, or type, or some other criteria. It is a bit complicated, and takes some time to set up and interpret, and has both advantages and disadvantages relative to shapefiles. Since much data are stored and distributed as geodatabases, you should be familiar with them if you are going to use ArcGIS.

Geodatabases are great for organizing data you will often use together, for data sets stored as tiles, or for complex mixes of data. You can store rasters, vectors, points, polygons, and topology all together in geodatabases, and you can specify a common coordinate system, extent, and other characteristics for sub-components of a geodatabase.

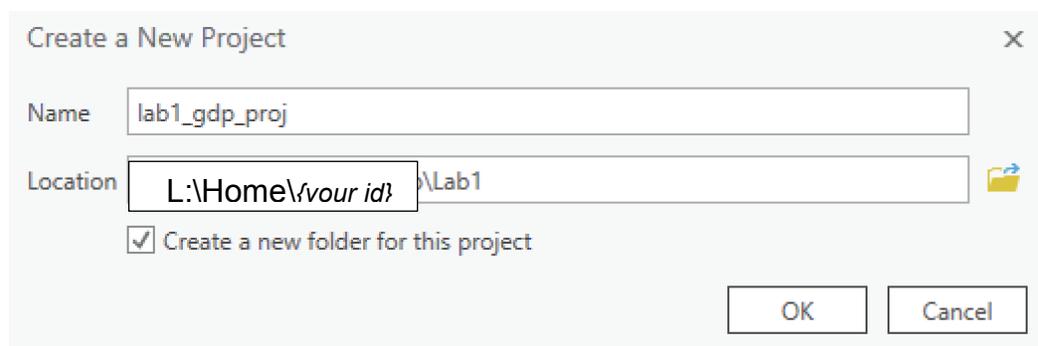
A geodatabase appears as a directory or folder to file explorer, and has a .gdb extension. If you look inside the directory, it has a number of component files. You should NOT copy, delete, or edit these component files with tools other than ArcGIS Pro, as it will likely corrupt your geodatabase and render it useless.

A geodatabase usually contains something called a **Feature Dataset**. In ESRI's parlance, a Feature Dataset is a collection of related spatial data, usually data layers and other geometric constructs. You can have several feature datasets in your geodatabase. You may want to hold clusters of data together, for example, data layers on river locations, lakes along those rivers, and dams associated with each lake. These layers are called **Feature Classes**. It is a bit convoluted, and takes some getting used to, but since all ESRI's documentation uses these conventions, you should be familiar with them. Remember, the geodatabase is the overall container, a feature dataset is a sub-container of data, and a feature class is a layer of data.

You don't have to store a feature class in a feature dataset, but it is still rather cumbersome if you have a single layer you wish to treat and store separately, because you have to create a geodatabase, then a feature class, so some folks use shapefiles for single layers.

Exit from ArcGIS Pro, and restart it.

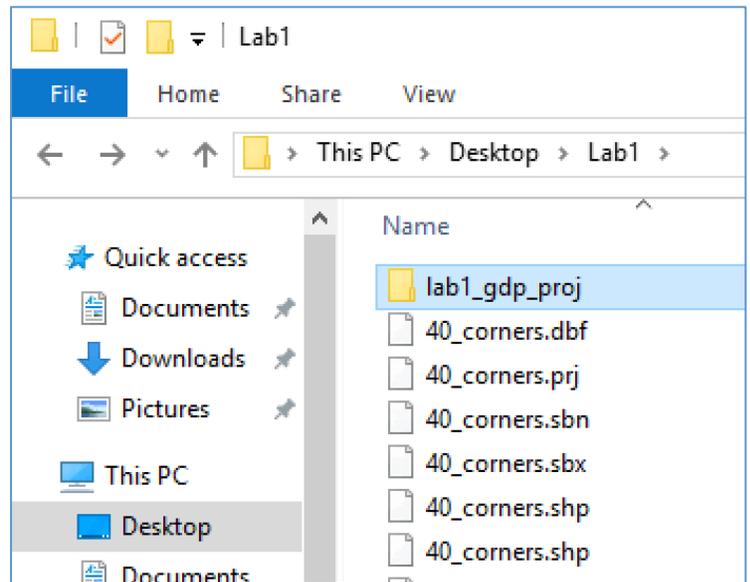
I can create a new project, here stored in a Lab1 directory, and name it *lab1_gdp_proj* (where gdp is my shorthand for geodatabase practice):



I include proj in the name just to remind myself that it is a project.

I also specify I want to create a new directory (check box, image above).

I can use Windows Explorer or the Mac Finder to verify the new folder/directory is created, within my Lab1 folder/directory:

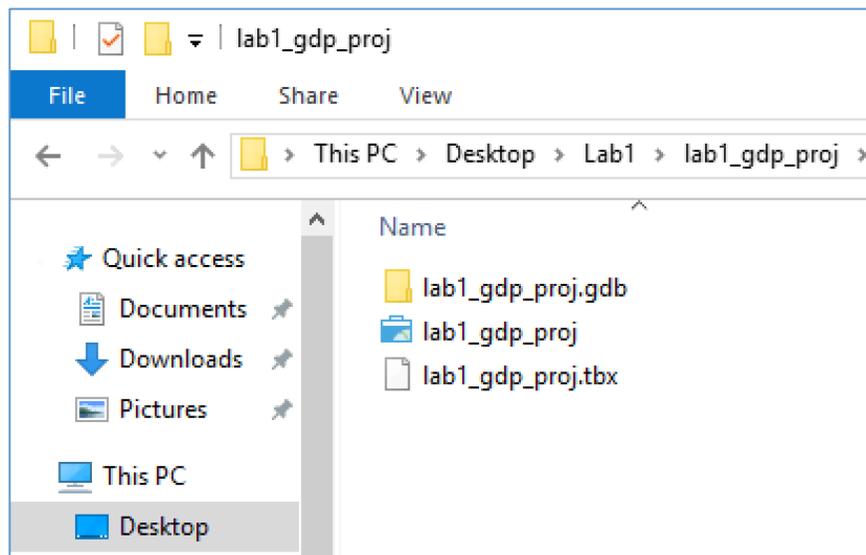


I can then look inside the lab1_gdp_proj, again using Windows File Explorer, and see there is a project with the .aprx extension, and a geodatabase folder/directory with the .gdb extension, and a .tbx file, all with the same first part name, lab1_gdp_proj.

The geodatabase (top in the list) is the default location for new data layers and temporary data layers.

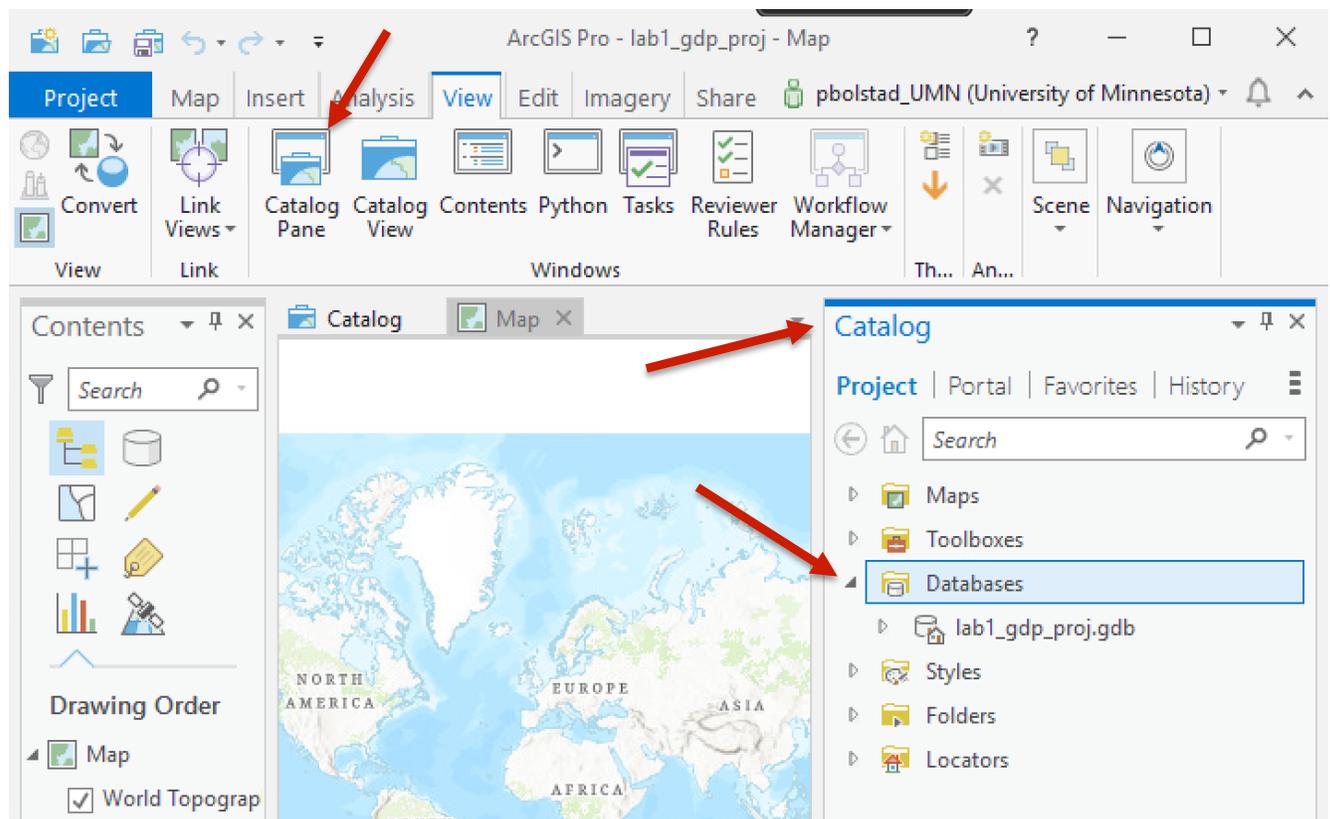
The .aprx project file (middle in the list, with a map/colored folder) is an organizing file with instructions on how you display your data, and where to find the data. It doesn't hold any data, just the instructions on what to display, and how.

The .tbx file (last in the list, bent-eared paper icon) is a toolbox, a container for any tools you might create associated with the project.



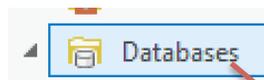
Creating a New Geodatabase

I can look at the current working directory and geodatabases through the Catalog Pane, found via the View tab:



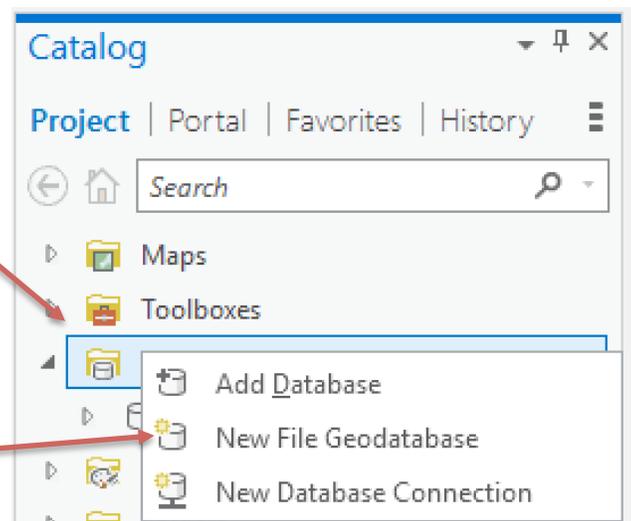
The catalog pane is usually shown by default on the right-hand side of the main window, but may be hidden below another tab, so you may use View-Catalog Pane, or the tabs on the bottom left of the main window to display.

The Catalog lets you inspect the Databases in your current project (lower arrow, figure above).



You can create and add a new geodatabase by right clicking on the databases icon,

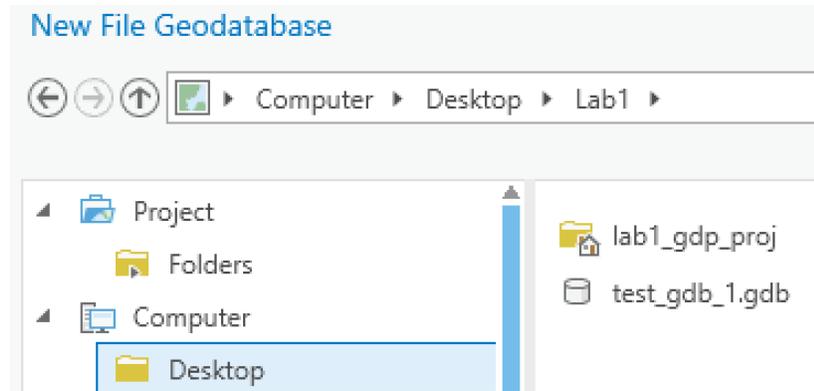
and then left clicking on New File Geodatabase



from the dropdown menu:

This should display a menu to create a new database. Set the directory to where you have stored your current project (in our example, Lab1), and create a new geodatabase, named something like test_gdb_1:

This may take a bit of time, be patient.



This will open a Geoprocessing tab:

We need to specify the input features. Here I clicked on the folder next to the Input Features slot and navigated to a data directory (here, Lab1), and specified the lakes.shp layer as the Input Features.

I then set test_gdb_1.gdb as your output location, and the name and other parameters.

On clicking Run at the lower right, a processing progress bar appears, and eventually, a message that the process completed successfully.

