Lab 1: GIS Fundamentals + Data Organization

1. Introduction

Geographic Information Systems (GIS) are tools and methods for working with spatial data—data tied to locations on Earth. GIS lets you ask questions like: *Where are things? What is near what? How does something vary across space? How do patterns change through time?* GIS is used in environmental science, public health, geology, planning, emergency management, business logistics, and many other fields because "location" is often a key piece of the problem. In this mini-course you will use **ArcGIS Pro**, a GIS application made by the company Esri. Esri products are the most popular GIS applications, but there are many others, including free and open-source options like QGIS and GRASS GIS. While these applications look different and have slightly different toolset, the fundamentals are the same, so if you are proficient in ArcGIS you can translate those skills to a different application (similar in concept to going from Microsoft Word to Google Docs).

A GIS program like ArcGIS Pro is not just a "map maker." It is a full workflow environment where you:

- bring in data from different sources,
- check and fix data problems (especially coordinate systems),
- visualize patterns with maps,
- measure and compare spatial relationships (buffers, overlays, distances, densities),
- and export results (maps, tables, and analysis layers) that can be reproduced and shared.

A major difference between GIS and many other kinds of software is that GIS projects depend on **many files working together**. A map usually references multiple datasets stored on disk. ArcGIS Pro does *not* always "copy everything into the project." Note that this is very different from, say, a PowerPoint or Google Slides presentation, where the images, text, etc. are part of the file. Instead, an ArcGIS file points to where the files live. That means **organization matters**:

- If you move or rename files, ArcGIS can lose the link ("broken data source").
- If you scatter data across Downloads/Desktop/email attachments, your project becomes fragile and hard to reproduce.
- If you don't keep track of what you changed, you won't know which layer is the "final" one.

Good organization solves these problems. A clean folder structure and consistent naming make your work faster, reduce errors, and make it possible for someone else **(or future you**) to open the project and understand exactly what you did. **<u>Good organization is a core professional skill in GIS.</u>**

Today’s Goals
By the end of this lab, you will be able to:

- Define a Geographic Information System and its analytical functions.
- Recognize vector, raster, and tabular data structures.
- Set up a reproducible project folder structure.
- Create an ArcGIS Pro project, load data, and confirm file paths are stable.
- Document workflows in a README file.

Key Concepts

- **GIS:** A system for storing, managing, analyzing, and visualizing spatial data linked to real-world coordinates.
- **ArcGIS Pro:** Desktop GIS software used to view, edit, analyze, and map spatial data.
- **Vector data:** Data represented as discrete geometric features defined by vertices.
- **Point:** Zero-dimensional features (e.g., stream gages, structural failure locations).
- **Line:** One-dimensional features (e.g., river channels, road networks).
- **Polygon:** Two-dimensional enclosed features (e.g., flood zones, building footprints).
- **Feature class:** A vector dataset stored inside a geodatabase containing the geometry and an attribute table.
- **Raster data:** Data represented as a continuous grid of cells (pixels), used for elevation models or imagery.
- **Attribute table:** A database table linked to a vector layer where each row represents a feature and each column represents a variable.
- **Geodatabase (.gdb):** The primary data container in ArcGIS. It stores feature classes and tables efficiently.
- **Coordinate Reference System (CRS):** The mathematical framework that defines how coordinates project onto the Earth's surface. Mismatched CRS definitions invalidate spatial analysis.
- **Metadata:** Documentation regarding dataset provenance, CRS, and limitations.

Project Organization
Create a main folder for this lab. Inside it, establish the following subdirectories:

- raw: Original source data. Do not edit these files.
- working: Intermediate outputs generated during processing.
- outputs: Final deliverables (exported layers, figures, tables).
- docs: Notes and documentation, including the README.

Why you need multiple folders:

- Prevents accidental edits to source data (raw stays raw).
- Keeps intermediate clutter out of your final deliverables.
- Makes it obvious what is “final” versus “in progress.”
- Makes your project portable (easy to zip/share and reopen).

General rules to make life easier:

- Create (or use the default) **project geodatabase (.gdb)** inside your main folder.

- Save GIS layers you create (feature classes, cleaned layers) in the geodatabase rather than scattering shapefiles everywhere.
- Keep *everything* for this lab inside this folder. For example, if you download a file, instead of dragging it from your Downloads folder into ArcGIS, first move it into your working folder

Step-by-Step ArcGIS Pro workflow
1. Initialize the Project Structure
- Create a root directory named `GIS_Lab1`.
- Inside GIS_Lab1, create the subdirectories: `raw`, `working`, `outputs`, and `docs`.
- Extract the provided Ellicott City data package into the raw folder.

2. Start a New ArcGIS Pro Project
- Open ArcGIS Pro.
- Under "New Project", select **Map**.
- Name the project: `Lab1_Organization`.
- Set the location to your GIS_Lab1 folder. Check the box for "Create a folder for this project".

3. Verify Project Architecture
You now have an ArcGIS Pro file (.aprx). In the Catalog pane (press Alt-Q and search for "Catalog" if it is hidden), check that there is also a geodatabase (it will be named either `Lab1_Organization.gdb` or `default.gdb`

Save the project (do this often!)

4. Load Data
Add the vector files from your raw folder to the Map frame:
- `EllicottCity_Floodplains.shp`
  - polygon file showing the FEMA-designated 100 and 500 year flood zones in Ellicott City
- `EllicottCity_HistoricDistrict.shp`
  - polygon file showing the Ellicott City historic district (downtown)
- `Patapsco_River_Streams.shp`
  - polyline file showing rivers in Maryland
- `EllicottCity_USGS_gages.csv`
  - locations of several USGS gages near downtown Ellicott City

These are all local files – they live on your computer. If you are connected to the internet, you can also use cloud-hosted files. These are great (particularly for large rasters) because you don't need to download huge volumes of data onto your computer. One example of a cloud-hosted file is the basemap – try switching between the different options.

The Imagery layer is good, but it uses data from a range of sources and dates which isn't always ideal (for example, it may mix images from summer and winter). You can search for other data sources in ArcGIS Online. In the Catalog, click ArcGIS Online (the cloud) and search for "md six inch imagery". The cloud-hosted data is high resolution aerial imagery data collected by the state of Maryland.

The first file loaded into an ArcGIS Map frame establishes the coordinate system (CRS) for that specific Map view. Verify that the Map is using a projected coordinate system appropriate for Maryland (e.g., NAD 1983 UTM Zone 18N or Maryland State Plane). Geographic coordinate systems define locations by their angular coordinates (e.g., latitude and longitude), while projected coordinate systems define locations using linear units (e.g., meters). You'll learn more about coordinate systems in the next lab.

Observe the data structures in the Table of Contents. The .shp files are vectors, the image service are rasters, the and the .csv is a standalone non-spatial table. Right a click a vector in the Table of Contents and click "Attribute Table". This shows the various features within that file. Note the different number of features in Patapsco_River_Streams.shp, EllicottCity_HistoricDistrict.shp, and EllicottCity_Floodplains.shp.

5. Generate Point Vectors from Tabular Data

Often you will need to work with data that does not have a spatial projection. For example, you might collect samples in the field and note the latitude and longitude, and now you want to see where those points are on your map.

The CSV file contains USGS stream gage locations in decimal degrees (latitude and longitude). You must convert this tabular data into spatial points.

- Click **Add Data → XY Point Data**.
- Input Table: EllicottCity_USGS_gages.csv.
- Select the appropriate columns for the Longitude and Latitude fields.
- For **Coordinate System**, select **Current Map**.
- Execute the tool.
- Right-click the resulting point layer and select **Zoom to Layer**.

Diagnostic Check: The points will fail to render near Ellicott City! Because the Map uses a projected CRS measured in meters, the tool interpreted the decimal degree values (e.g., -76.8, 39.2) as Cartesian coordinates. The points were plotted 76.8 meters west and 39.2 meters north of the projection's origin.

To correct this geometric failure:

- Rerun the XY Table to Point tool.
- Change the **Coordinate System** parameter to **WGS 1984** (a geographic coordinate system).

Run the tool. The points will now render correctly in Ellicott City.

6. Symbology

- Select a layer in the Table of Contents, and click Symbology under the Feature Layer or Raster Layer
- You can modify how vectors or rasters are shown on the map. More on this later.

7. Verify Data Sources

For each layer in the Table of Contents:

- Right-click the layer and select **Properties**.
- Navigate to the **Source** tab.
- Confirm the file path on your disk.
- Identify the spatial reference (CRS) for each dataset. Right click the layer in the Table of Contents and click Properties. Look at the spatial reference and the Cell Size for the two raster layers (NAIP and MD Six Inch). What differences do you notice between these?

8. Data Duplication Protocol

Do not edit raw data! If a layer requires modification, export it to your project geodatabase.

- Right-click the dataset -> **Data** -> **Export Features**.
- Set the Output Location to your default .gdb.
- Use a clear naming scheme (e.g., floodplain_working).

9. System Reboot Test

- Save the ArcGIS Pro project.
- Close the application completely.
- Reopen the .aprx file.
- Confirm all layers render without broken source links (indicated by red exclamation points in the Table of Contents).

Questions

Explain the functional difference between storing raw data in an organized project directory versus linking to files located in a temporary Downloads folder. What specific error occurs in the GIS software if a source file is moved?

Identify the exact geometric reason the USGS stream gages plotted near the equator during the initial execution of the XY Table to Point tool. Think about how the software interpreted the numeric values in the latitude and longitude columns based on the map's coordinate reference system.

When you duplicated the data in 4.8, you exported the feature class to the default geodatabase rather than creating a new shapefile. What are some technical advantages of using a geodatabase over a directory of independent shapefiles?