Lab 1. Getting started with geospatial data structures and attributes in ArcGIS Pro

ENV 978: GIS Applications for EOI Fall 2022

Instructor: Leif Brottem | brottem@wisc.edu

10 points: Due 11:59pm Thursday Sep 15

Objectives:

- *Become more familiar with the ArcGIS Pro interface and how to recognize certain fundamentals (coordinate systems) and adjust certain settings (home directory), and attribute table formats for vector layers*
- *Learn how to perform a straightforward Join that requires:*
    - *Knowing how unique identifiers work*
    - *Understanding how One-to-Many and Many-to-One joins function between relational database tables*
- *Become acquainted with raster data layers and how:*
    - *They have fundamentally different data structures compared to vector layers (including their attributes)*
    - *They are interoperable with vector data layers, which opens many spatial analysis possibilities*
- *Learn how to carry out basic data processes with raster and vector data layers.*

---

1. On the course Canvas site, locate the “Env978F22Lab1Data” directory and download it to your computer or workspace you will use for the lab.
2. Open ArcGIS Pro
3. Under “New”, select “Map”.
4. Save in your personal directory as “Lab1_*YourName*’.
5. On the map canvas, look to the bottom of the map viewer and take note of the coordinates that are in use: they should be Latitude/Longitude. Depending on the defaults of your software, it should have the “World Imagery” layer centered over the United States as the basemap.
6. On the Project tab go to Options Current Settings verify the Home directory is set to the project geodatabase (.GDB). You should do this each time you start a new project.
7. In the Catalog pane/Folders, expand the “Env978Lab1Data” directory and then click & draft the *Global Avg Temp.tif* to the Table of Contents pane.
8. Verify that the *Global Avg Temp.tif* raster layer fits over the basemap. Click on a few of the pixels and make sure they make sense as average temperatures somewhere on earth.
9. From the same directory, click & drag the states shapefile and then the counties shapefile *on top of the states* in the Contents Pane.
10. In the Contents pane again, right click to open the Attribute Table of each layer
    a. How many features in each?

   b. Sort by various attributes by right clicking on different column headers and choosing either "Sort Ascending" or "Sort Descending"

11. Sort by STATE_FIPS: this is the Federal Information Processing Standards code, which is a standard unique identifier for U.S. administrative units

One to Many and Many to One joins

Joining tabular data together based on a key field is a fundamental GIS operation because it allows you to link georeferenced data layers with non-spatial attributes (e.g. adding a population table to a map of the United States). The operation can get finicky based on data type (numerical versus strings) and how different software platforms function. Generally speaking, numerical codes are the best field for performing joins.

12. Notice that both the "states" and "counties" have their own FIPS codes *but they have one in common*. We will use this unique identifier to perform a *many to one* join. Remember which type of join this refers to.

13. Based on your understanding of relational table structures and *which kinds of joins* are possible, perform a join between the states and counties. NOTE: joining in one direction is logical and the other will work but result in a strange dataset. You need to figure out which one is the correct type of join to perform.

14. When you understand which kind of join is feasible, search for the Join Field tool in the Data Management directory in the toolbox ". Select "counties" and "states" in the correct order and use the correct fields as the unique identifier to serve as the *Input* & *Join* fields.

    a. In the Export Features dialogue, save the new layer in your exercise folder and call it "Counties_Joined". Leave everything else the same and, still in the Export dialogue pane, click on the "Environments" tab.
    b. NOTE: in ArcMap, you export your layer/shapefile in order to make the join *permanent*. This can lead to problems in ArcPro when joined attributes can get lost due to differences in ArcPro's geodatabase structure.

15. After you run the join, look at the Attribute Table and make sure that it was successful: you should see the fields from the other table.

16. **Question**: what is some kind of information that you now have that you didn't have before performing this Join?

17. Now, let's learn a new attribute table tool. Open the attribute table of the counties layer and right click on the SUB_REGION field header and select "Summarize". For the Statistics Field, choose "NAME" and for the Case field, select "SUB_REGION". For Statistics Type, choose "Count". Leave everything else and click "Run".

18. A table should automatically appear in your Contents pane. Open it up and take note of what the Summarize tool did for you and how it reflects vector attribute data structures.

19. **Question:** How many counties are in the Middle Atlantic region?

PRO TIP: You can use the Statistics tool in the table that was generated to verify that there are 3,141 records being used in your sub-region summary.

20. **Further**: use the Summarize tool to count up the number of counties by state and then *join* the output table back to the states layer. Explore the data and see if you can discover anything interesting; e.g. is there a relationship between state size and number of counties? What if you control for region?

21. **Extra challenge**: find an additional *point* layer on the internet that would make sense and fit geographically with the U.S. counties. Load it into your project and perform an additional GIS process with it (hint--one good one would be "Summarize Within").

Raster attributes

A fundamental aspect of GIS is that raster (pixels) and vector (points/lines/polygons) are fundamentally different data structures that are nonetheless interoperable in a GIS environment. Each bring its own potentials and limitations that you will become familiar with.

22. Add the *Global Avg Temp.tif* to your project and use your cursor to look at the values. This is an average temperature in Celsius over some unit of time (I can't remember whether its monthly or yearly, which would be a *problem* if I wanted to use these data for anything other than an exercise! I'm pretty sure it's a winter month)

23. You might notice that these are *only* values for terrestrial areas, not the oceans. So, go ahead and switch on the World Imagery base layer and drag the US States over the temperature grid.

24. So, this is a *raster* grid: pixels with individual values! Each pixel represents a specific attribute over the geographic area represented by the pixel: Right click on *Global Avg Temp* in the Contents pane, select Properties and go to Raster Information. You will see the "Cell Size" dimensions for this grid (Note—you also need to check the Linear Units under Spatial Reference to know what these represent)

25. We use these for *specific kinds* of geospatial stuff, especially environmental features like temperature, water, and land cover.

26. Try to open the *Global Avg Temp.tif* attribute table. It doesn't have one because this layer's values are not integers. The *single* attribute that is assigned to each pixel is encoded within the raster data structure itself.

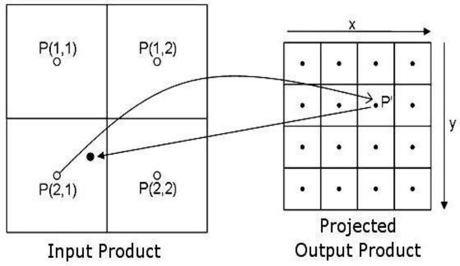
27. In terms of raster/vector interoperability, the most important thing is that they are in the same <u>coordinate</u> space. In this case, ArcGIS has *projected our temperature data on the fly* which "fits" it into the coordinate system of the data frame, which, in our case, was set by the "states" layer because that was the <u>first</u> one we added to the project.

28. First, we do not need temperature values for the entire globe, so let's "clip" this layer to the boundaries of the states layer. To do this, search for "Clip Raster" in the toolbox. Global Temp is the input, States is the Output Extent. Check BOTH boxes for "Use Input Features for Clipping Geometry" and "Maintain Clipping Extent". Be sure to save your output to your project geodatabase that was created when you started the project.

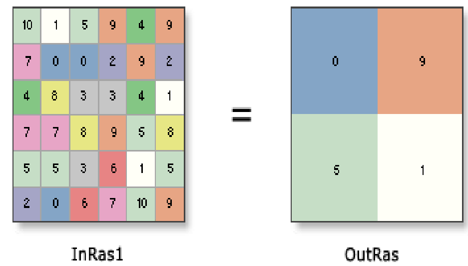
Resampling and neighborhood analysis

29. On your new raster layer, zoom in so you can see the individual pixels. Use the Measure tool on the Map tab/Inquiry group to measure how large the pixels are.
30. One important raster processing task is *resampling*: modifying the size and position of raster grid cells. As a rule, different raster grid layers that you work with should line up in the same coordinate space and have the *same resolution*. Another rule is that *output* raster grids will always have the cell size of the *lowest* resolution input and you cannot resample grid cells to higher resolutions (because where would the higher-resolution pixel values come from?)
31. There are four principal resampling methods: Nearest Neighbor and Majority for *categorical* data; Bilinear and Cubic
    a. Nearest neighbor: Performs a nearest neighbor assignment, is the fastest of the interpolation methods. used for categorical data
    b. Majority Performs a majority algorithm, determines the new value of the cell based on the most popular values. It is *not* appropriate for continuous data (e.g. elevation)

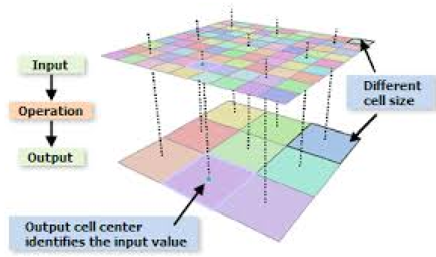
Nearest Neighbor



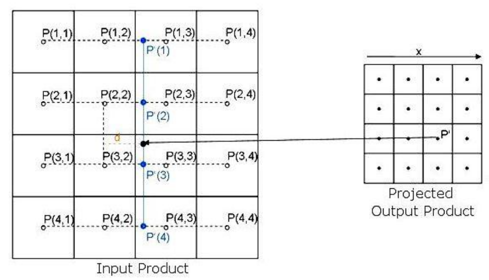
Majority



Bilinear

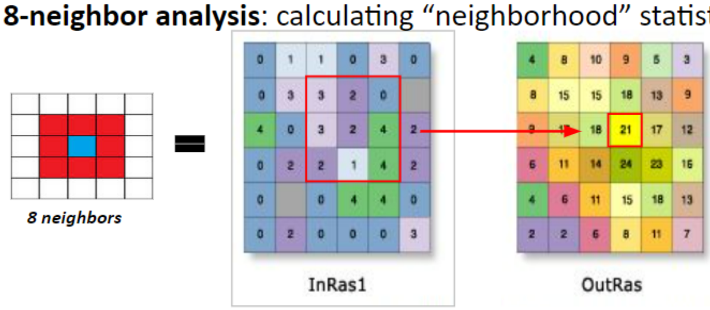


Cubic Convolution



32. Resample your temperature raster grid to 200,000 meters for the X and Y dimensions using the Bilinear technique. It calculates the value of each pixel by averaging (weighted for distance) the values of the surrounding four pixels. It is suitable for continuous data. Take note about how the pixel values have changed based on in the neighborhood of pixels and their resolution. Remove the resampled grid
33. Using the clipped (un)resampled temperature grid, we will now look at *neighborhood* analysis, which can only be done with raster data.
34. Resampling is a form of "neighborhood" analysis, which involves calculating and measuring values for "neighborhoods" of pixels. In this way, one can also perform various kinds of "moving window" analysis.

8-neighbor analysis: calculating "neighborhood" statistics

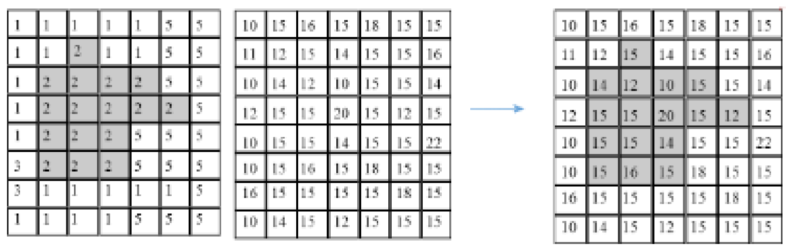


35. Above, the output is sum of values of <u>8 neighbors</u> plus the target grid cell for every cell in the input grid
36. Other descriptive statistics (mean, min, max, etc) possible
37. "Neighborhoods" can be defined with other shapes as well (circle, wedge, user-defined shape..)

38. Use the “Focal Statistics” tool to find the average temperature using a neighborhood of eight pixels. You’ll notice that this is also a *smoothing* filter; akin to what engineers use in various fields, including sound. Before you close the tool, look at the different neighborhood parameters you can set.

Zonal analysis that combines raster and vector data

39. In most cases you will have both raster and vector data that you want to use. *Zonal Statistics* are a handy and fundamental GIS tool that uses an overlay zone layer with a value layer to calculate statistics for the values within each zone



40. On the Analysis tab, select Tools and search for “Zonal Statistics” in the Spatial Analyst toolbox. This is a tool for performing calculations on raster data values within *areas defined by the boundaries of another layer*. In this case, we will calculate the average temperature in each state by using the Zonal Statistics tool to calculate the average of *all the pixel values* within the boundaries of each state. Get it?

41. In the Zonal Statistics tool, choose “states” for your Input feature zone data and the STATE_NAME as the Zone field. The Input value raster is the “Global Avg Temp.tif”. Name the output “AvgTempByState”. For Statistics type, choose “Mean” and click on “Run”.

42. *Turn off* the original temperature grid to see the zonal statistics output. Click on the output raster in one of the states.

43. **Question**: What is the average temperature in your home state or another random one?

44. **Question**: Briefly explain what you did and how it demonstrates differences in the vector and raster data structures?

45. Note that the Zonal Statistics raster grid is not necessarily the most useful data format. Find “Zonal Statistics as Table”, run the same analysis using STATE_FIPS as the Zone field, and save the output table as “StateFIPSAvgTemp”. This shows you how different data formats can be amenable to different types of analysis. Any surprises regarding the coldest states?

46. But wait! Note the number of records in the Zonal States temp is a little off. Join it to the original states shapefile to figure out why.

47. Question: Which states are missing data and what explains the discrepancy in the number of states and the zonal statistics for temperature?

48. **Question:** Find one more public data layer, add it to your project, and perform some kind of data description or analysis with.

49. **Further**: Find and add a U.S. Cities layer or shapefile from Arc online or the internet. Figure out which *tool* to use in order to extract the *temperature value* from the raster grid to the corresponding U.S. point features. In your report give an example city, briefly describe how you did the operation, and explain one limitation of this approach given the data you used.

Deliverable

A 1-2 page report that includes the answers to the questions, a brief description of the public data layer you found and added (plus any issues or concerns), and a screenshot of your project that includes all the layers.