Intermediate Spatial Data Science Lab Report

Title: Lab 2 - Part 1: Exporting Files
Notice: Dr. Bryan Runck
Author: Andrew Arlt
Date: 10/29/2024

Project Repository: [andrewarlt/GIS5571/Lab2/]()
Google Drive Link: n/a
Time Spent: 10 hours

Abstract

Large datasets can be manually downloaded, sorted, and converted directly from database websites, like PRISM or Minnesota Geospatial Commons. This process can be automated and simplified by using ETL structures within Notebooks, which can download many files at once and convert each file through common methods. This laboratory activity aims to practice setting up data flows to convert data into usable data outputs: DEM/TIN layer and spacetime cube animation.

Problem Statement

The main problem in this activity is to establish a code structure that can collect, sort, and synthesize data into displayable formats. There are two main goals in Part 1:

1. Create a .LAS database from .LAZ files and then construct a DEM and TIN from the resulting point cloud, exported as PDFs.
2. Download the annual 30-Year Normals .bil files for precipitation from PRISM, convert the data into a spacetime cube, and export the spacetime cube as an animation file.

Data for each task were obtained from publicly available websites and downloaded using API interactions. Data was initially processed so that it was accessible, e.g. files unzipped, data fields created, etc. The general data processing is shown below in Figure 1, where blue is the initial dataset, yellow is the processed data, and green is the final data used in the products.

Figure 1. Table showing the transformation of data for each of the two tasks.

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	**MN Lidar API**	Download data via an API, then unzip the data, project coordinates	.LAZ Files (lidar data)	Point location data (x, y, z) [additional attribute data available, e.g. return time, frequency]	MN Lidar to “County Lidar”	API request for .laz file; unzip file; create .LAS database
2	**PRISM API**	Download data via an API, set .bil files into mosaic layer	.bil files (normal precipitation)	Location, Precipitation (mm)	PRISM to “Prism Mosaic”	API request for .bil files; add data for time (by title), convert .bil to mosaic
4	**County Lidar**	Create DEM from .LAS database, Create TIN from .LAS database	.ladb file	Point Cloud (x, y, z)	to “County DEM” and “County TIN”	Use las database to create DEM and TIN layer
5	**PRISM Mosaic**	Transform mosaic into ST cube	Mosaic Database	Location, Precipitation by Month	to “ST Cube”	Convert mosaic to ST cube, using precipitation and month data

6	**County DEM/TIN**	Project DEM and TIN layers in 2D and 3D	DEM, TIN	Location (x, y, z)	Layout PDFs (DEM and TIN)	Project in 2D and 3D maps, export as a PDF
7	**ST Cube**	Animate the layers and export as video	Spacetime Cube	Location, Precipitation, Date	ST Cube Animation	Export as animation video.

Input Data
The first goal of Part 1 was to create a set of map layouts for the DEM and the TIN lidar data, and to export an animation displaying the monthly normal precipitation values throughout the United States. Lidar data was obtained for Winona County, which consisted of 220 tiles. The lidar data was initially imported as zipped .laz files, which were then unzipped and merged to create an initial .las database (point cloud).

The second goal of Part 1 was to create an animation using monthly normal precipitation values from PRISM. PRISM data was downloaded as zipped .bil files, which were then unzipped. A new mosaic dataset was created to hold modified .bil files for all months.

Figure 2. Data layer(s) used to perform the described processes.

#	Title	Purpose in Analysis	Link to Source
1	**MN County Lidar Data**	Used to establish a .las database pointcloud for creating a DEM layer and a TIN layer.	LIDAR Data
2	**PRISM Climate Data**	Used to create a mosaic dataset and monthly spacetime cube that can be visualized as an animation.	PRISM

Methods
Figure 3 shows the ETL used to create the DEM and TIN PDF map layouts. The lidar database (point cloud) was used to create an initial DEM layer, which smooths the spaces between points using interpolation methods. This DEM layer can be used to create a TIN layer, which is a vector based elevation layer that is more suitable for irregularly spaced point data. Both the DEM and TIN layers were applied to a common map layout in order to compare the output formats.

Figure 4 shows the ETLs used to create the PRISM animation of precipitation values. The PRISM animation was created by adding the extracted .bil files into the new PRISM dataset. These data files had two new fields added to extract the 30-year normal value by month and establish a datetime using the file name. These files were used to build a multidimensional raster layer, which could be transformed into a spacetime data cube. Spacetime data cubes can be visualized as an animation, and then exported in a video format.

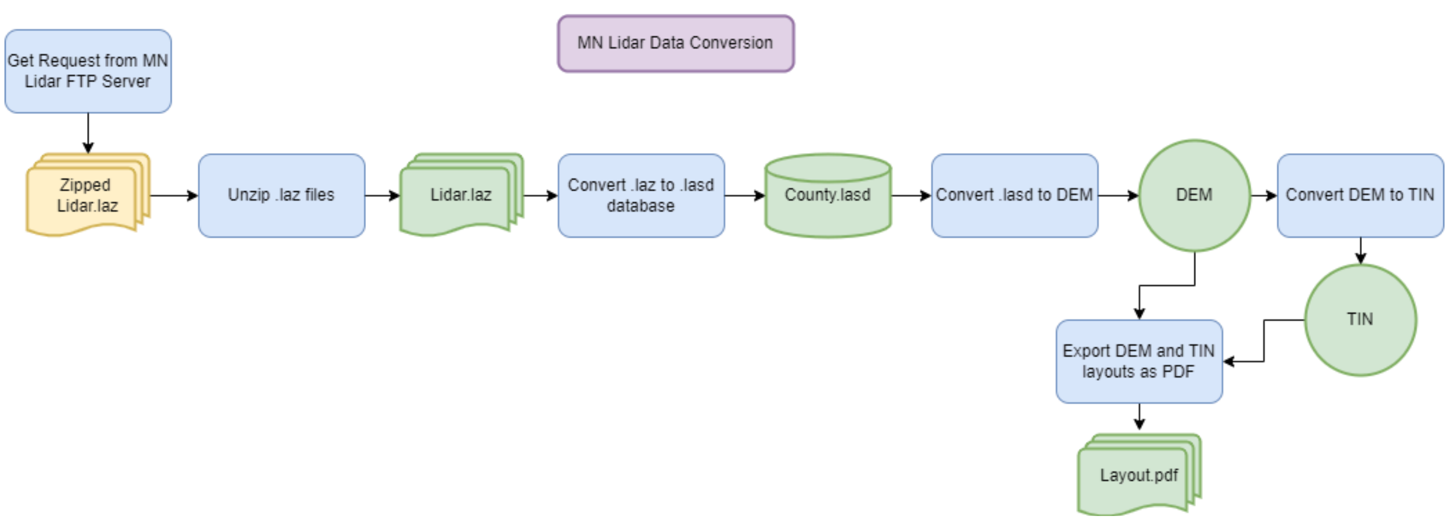

Figure 3. Data flow diagram showing the API interaction, data transformations, and export functions used in the ETL to create map layout PDFs for a DEM and TIN derived from a point cloud.

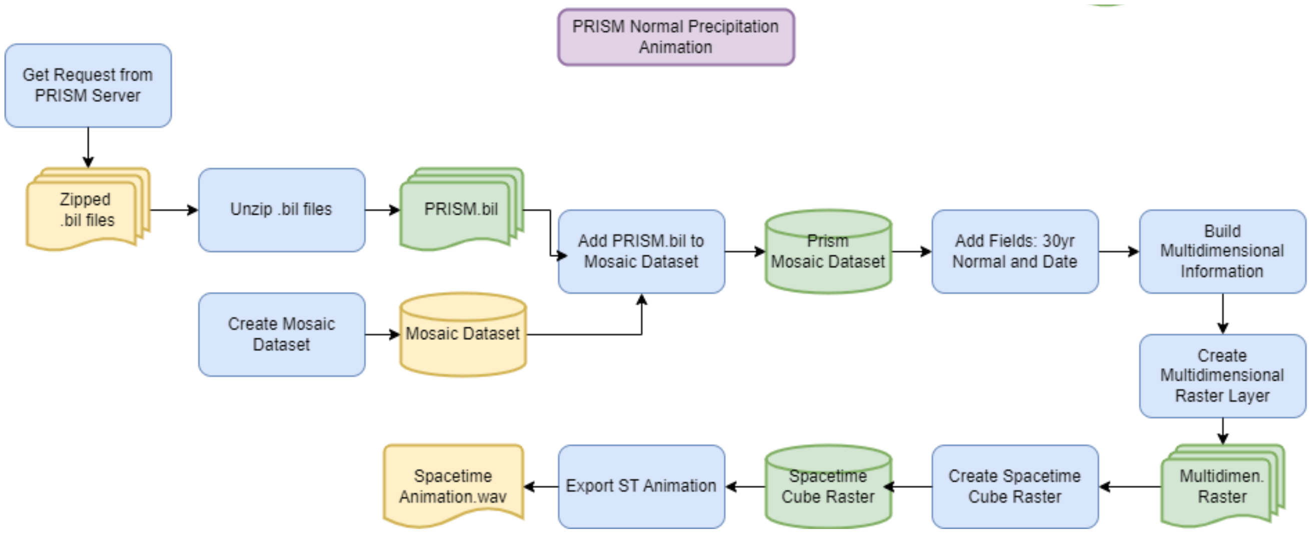

Figure 4. Data flow diagram showing the API interaction, data transformations, and export function used in the ETL to create and export a timelapse animation of normal precipitation values across the United States.

Results

The lidar imagery was successfully represented through a viable point cloud, raster DEM layer, and TIN vector layer. Each of the map layers were applied into separate layouts, but in the same orientation and scale. PDF files for these documents were exported and are attached separately to this report.

A spacetime animation for the precipitation was successfully produced, rendered, and exported from the ArcGIS Pro interface. The resulting animation video is attached separately to this report.

Results Verification

Figure 5 (below) shows the outputs of the DEM (left image) and point cloud layer (right image). From a nadir aerial view the images appear very similar, particularly with a similar color ramp applied to the elevation gradient. The main differences are apparent in the planar (3D) view since the point cloud image has actual relief beyond the color ramp variations. The 3D view is helpful for visualizing the landscape for EDA purposes and checking for rendering issues. The DEM (or TIN) images are helpful in creating a cost surface to be used in raster math or other cost surface analysis applications, such as buffers.

Figure 6 (below) shows a resulting map visualization of the spacetime cub produced from the multidimensional raster dataset for the normal monthly precipitation values in the PRISM .bil files.

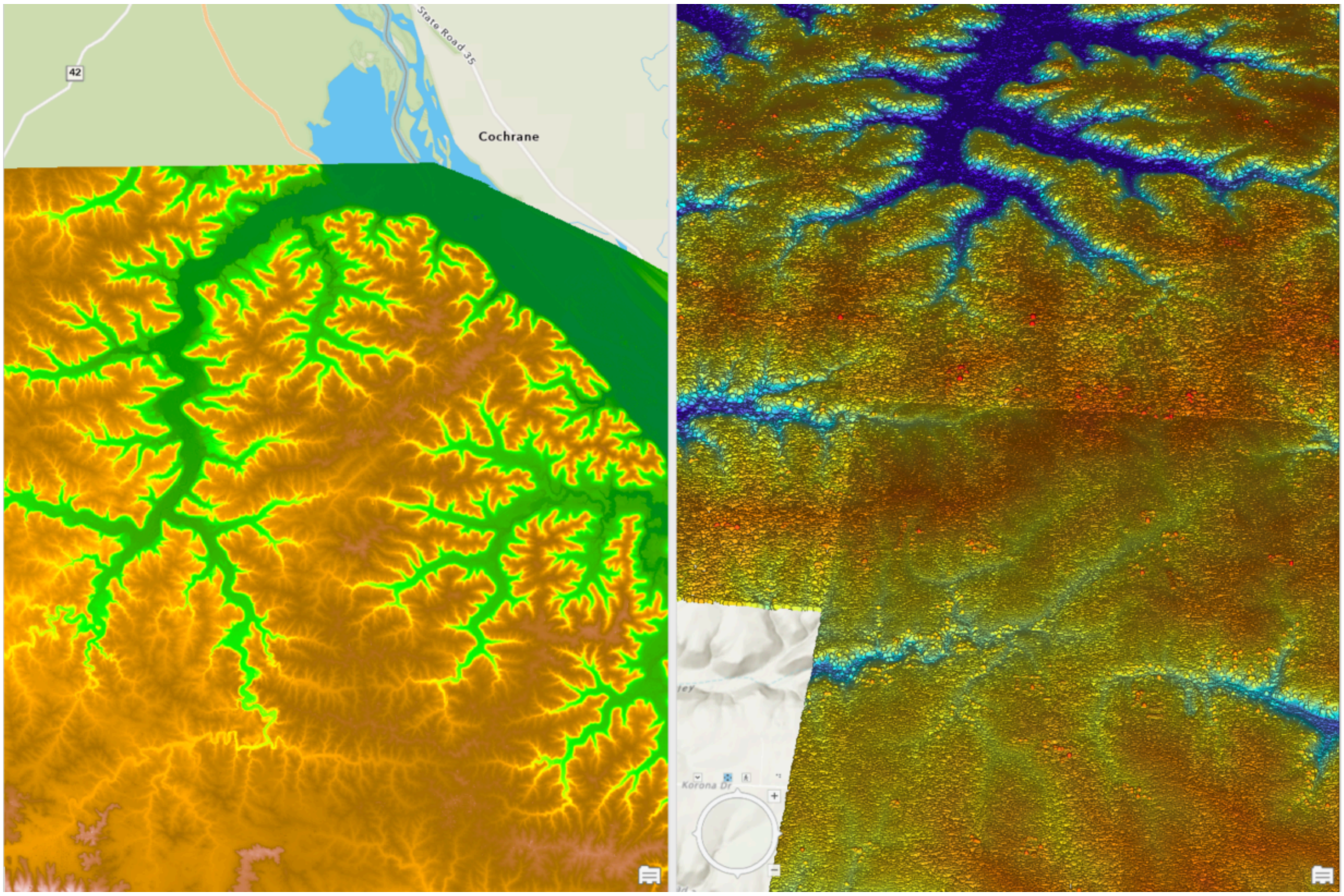

Figure 5. Comparison of the outputs of a 2D (left) and 3D (right) rendering of the lidar data, where the 2D image shows the DEM raster layer and the 3D image shows the rendered point cloud.

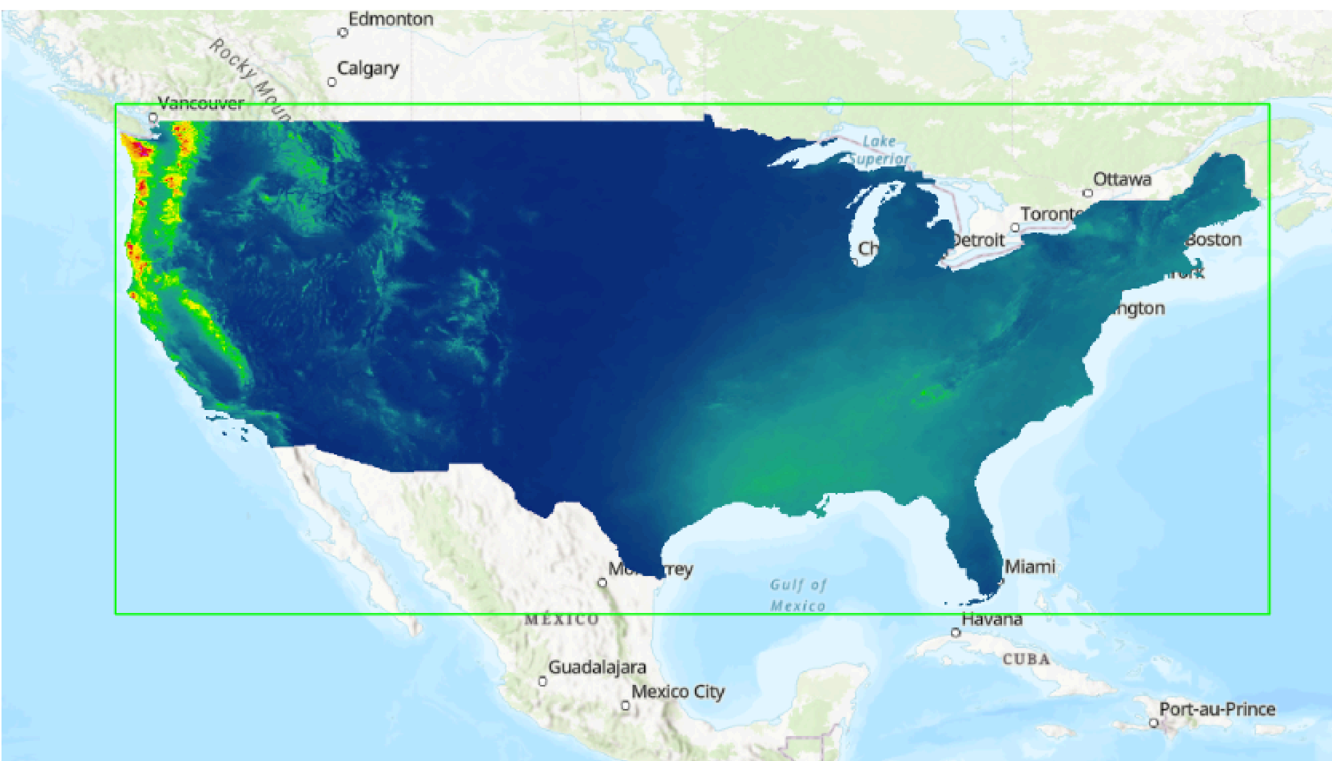

Figure 6. Image showing the output of the colorized normal precipitation time cube multidimensional layer, where red equals high precipitation and dark blue equals low precipitation (Note: January values are shown).

Discussion and Conclusion

The lab activity involved learning how to control a dataflow for many different files returned from an API get request. The files required some form of transformation before the final products could be created and then exported using Arcpy. This particular section required working with two different data formats, .las and .bil.

It was helpful having prior knowledge in working with .laz files and creating point clouds and DEMs. The challenge was in determining the range and extent of the lidar tiles to download, extract, and transform. I ended up acquiring all of the Winona County tiles, which consisted of 220 lidar tiles. While the process was slow, it ended up being a chance to practice extracting specific files from a folder on an FTP server like the Minnesota lidar site. I reduced the number of tiles when creating the TIN layer to save on processing power and run time.

The exporting PDF function was relatively simple to execute using the ArcGIS documentation, but did not work well when applying the maps to a premade layout template. This is something that I think would be really helpful to figure out, though, since this would simplify workflow and output speed, especially if a consistent style is used by an organization or report. I would like to find out if there is a way to access a tile map and extract the necessary lidar tiles using a bounding box, rather than manually needing to visually inspect the tile map, record the tile numbers, then call specific files.

Working with the PRISM data was also fairly straightforward in terms of the API request and unzipping requirements. The ArcGIS documentation on creating the time lapse animation (spacetime cube visualization) was very clear and easy to understand. The biggest challenge was in determining that the date field needed to be in a specific format in order to be understood by the multidimensional raster converter and spacetime cube function.

References

Explore your raster data with Space Time Pattern Mining. (n.d.). *ArcGIS Blog*. Retrieved October 29, 2024, from https://www.esri.com/arcgis-blog/products/arcgis-pro/analytics/explore-your-raster-data-with-space-time-pattern-mining/

PRISM Climate Group at Oregon State University. (n.d.). Retrieved October 29, 2024, from https://prism.oregonstate.edu/FAQ/

Raster To TIN (3D Analyst)—ArcGIS Pro | Documentation. (n.d.). Retrieved October 29, 2024, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/raster-to-tin.htm

Self-Score

Category	Description	Points Possible	Score
Structural Elements	All elements of a lab report are included **(2 points each)**: Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score	28	**28**
Clarity of Content	Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**.	24	**24**
Reproducibility	Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified.	28	**28**
Verification	Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**.	20	**20**
		100	**100**