Lab Report

Title: Lab 2 Part 1
Notice: Dr. Bryan Runck
Author: Alexander Edstrom
Date: 10/27/21

Project Repository: *AlexanderEdstrom/GIS5571 (github.com)*
Time Spent: 7 hours

Abstract

Part one of this lab covers the process of setting up an ETL pipeline to find, download, convert, and display a dataset. The dataset will be taken from a server in .LAS format, then be converted to a DEM and a TIN. This data will then be displayed in 2D and 3D for comparison. This section will also cover the production of a pipeline that downloads precipitation data, and creates a spacetime cube, which will then be displayed.

Problem Statement

The data for step one will be taken from the MN DNR as a .LAS, converted into both a DEM and a TIN. Each of these DEMs and TINs will then be exported as a PDF after being turned into a basic layout. After this, the original .LAS file will be opened in both 2D and 3D format to analyze the similarities and differences in the way that ArcGIS Pro lets you view and manipulate the data within each.

The data for the spacetime cube will be downloaded from PRISM and converted into a spacetime cube using ArcGIS's built in spacetime cube functionality.

The required materials are outlined in table 1.

Table 1. Requirements

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	.LAS file	Raw input dataset from MN DNR	Elevation Data		MN DNR	Pulled from the DNR server
2	Spacetime Cube Data	Raw input dataset from PRISM	Precipitation Data		PRISM	Pulled from the PRISM server

Input Data

The .LAS file being used in step one is a lidar point cloud covering an area on the bank of the Minnesota river below Blakeley, Minneota. This data is taken from the MN DNR and will be used in the subsequent steps in part one. The dat will be pulled directly from the MN DNR server, this particular file was found under the public, data, elevation, lidar, examples, samples section. This data is relatively small, only covering about 4.4 square kilometers of area, this small size will help speed up the visualization process by taking less computer resources to generate graphically.

The data for the spacetime cube in step two is from the PRISM Climate Group and contains the "average monthly and annual conditions over the most recent three full decades. This 30 year period is from 1981-2010. The climate variable used in this project is precipitation.

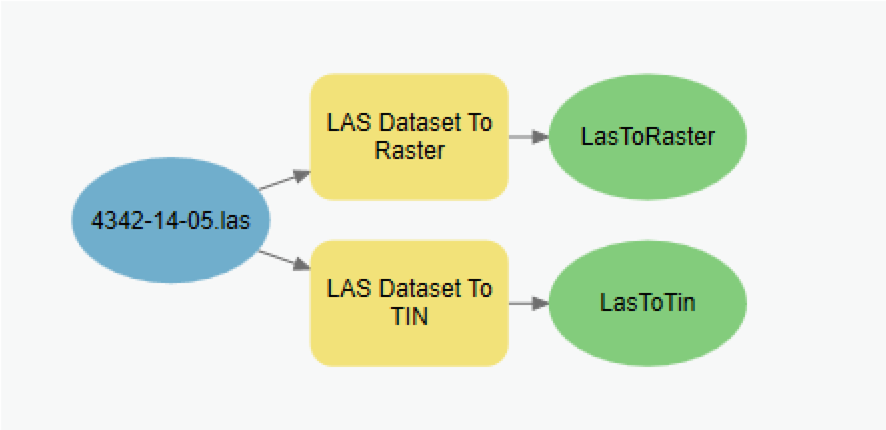
These datasets are outlined in table 2.

Table 2. Input Data

#	Title	Purpose in Analysis	Link to Source
1	Lidar Point Cloud	Raw input dataset for DEM and TIN creation from MN DNR	MN DNR
2	Normals Data	Raw input datasets for Spacetime Cube creation from PRISM Climate Group	PRISM

Methods
Section 1:
The process of getting the Lidar Point Cloud data downloaded and imported was very similar to the pipeline created in Lab 1. This consisted of a simple ETL pipeline that reached out to the MN DNR server with the URL that contained the data, pulling the data, and downloading it. Once downloaded the data was converted into both a TIN and a DEM with simple ArcPy functions (LAS Dataset To Raster and LAS Dataset To TIN). For the conversion to TIN I used random node count thinning and limited the nodes to 500,000, this was to speed up the process of working with the file. For these purposes, 500,000 nodes is plenty. This process is shown in figure 1. With the data downloaded and converted, it could be added to layouts for export. The creation of these layouts utilized the ArcPy Mapping library (arcpy.mp). After adding each layer to a separate layout, they could be exported as PDFs.

Figure 1: Model Builder Diagram for conversion



Section 2:
This section of the lab compares and contrasts the process of visualizing and working with the 2D and 3D representations of a .las file. The 2D view of the area is limited as the only difference between the valley and non valley is the color of the points or the color of the surface. I think the surface view gives a much better representation when compared to the points but there is still very little that can be done to visualize the depth of the data in its true form. The 2D view with the data drawn as a surface is shown in figure 2. The best thing to represent depth in this situation is depth itself. This means that the 3D view is much better visually. The 3D view with the data drawn as a surface is shown in figure 3. The depth is immediately understood without having to refer to a legend to decode a color gradient. The 3D view is being rendered on the fly so it obviously takes more computing resources,

this may cut into performance elsewhere in the model on some machines. The 3D view also offers different options for analysis, the option of the exploratory 3D analysis tool opens up when looking at a the data in 3D. Tools under the visibility analysis tab can be used with 2D elevation data, but having the data presented in 3D can help better visualize things like line of sight and field of view.

Figure 2: 2D view with data drawn as a surface

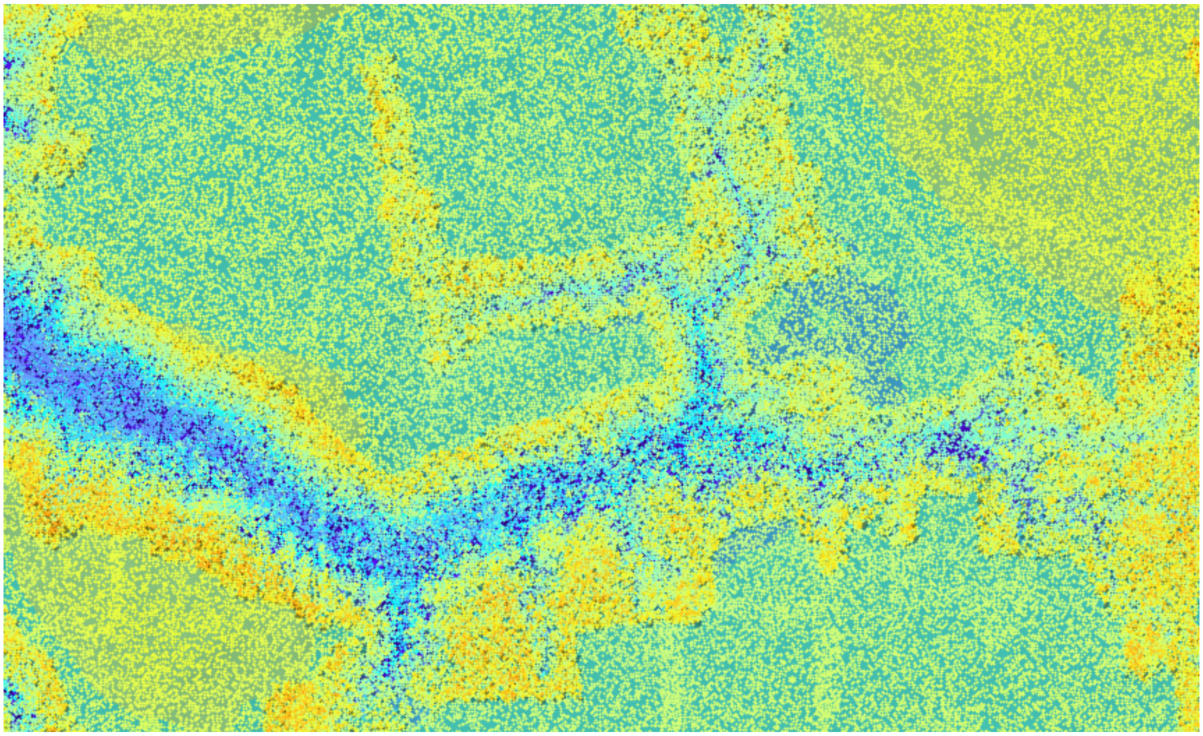
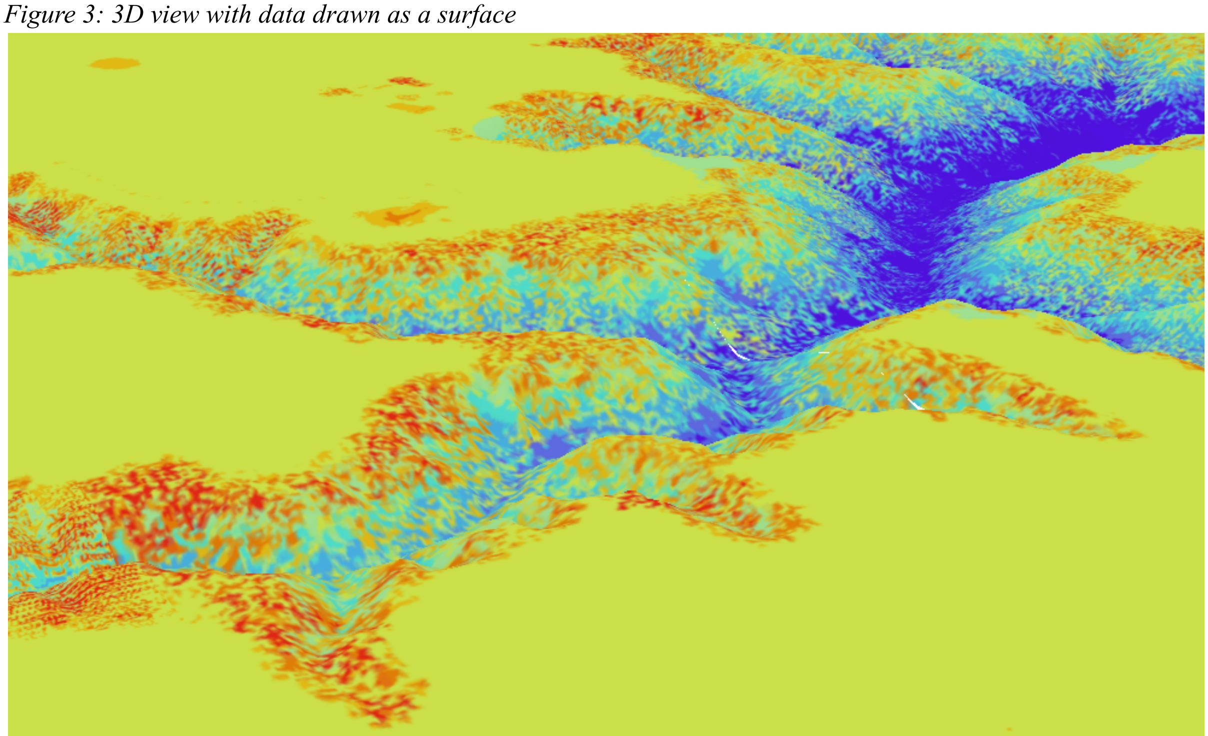
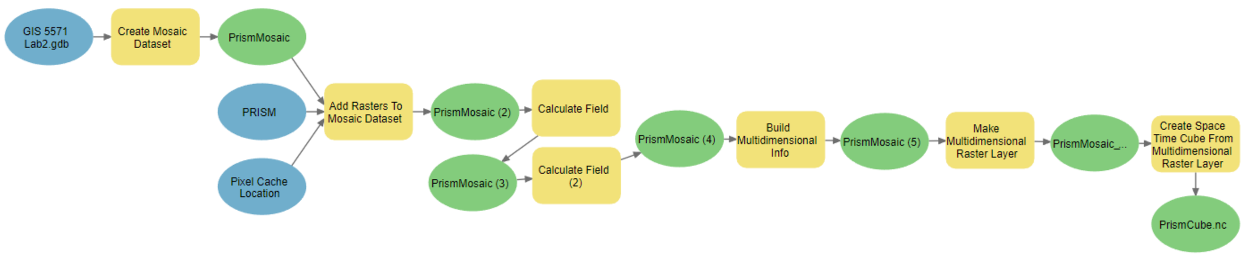


Figure 3: 3D view with data drawn as a surface



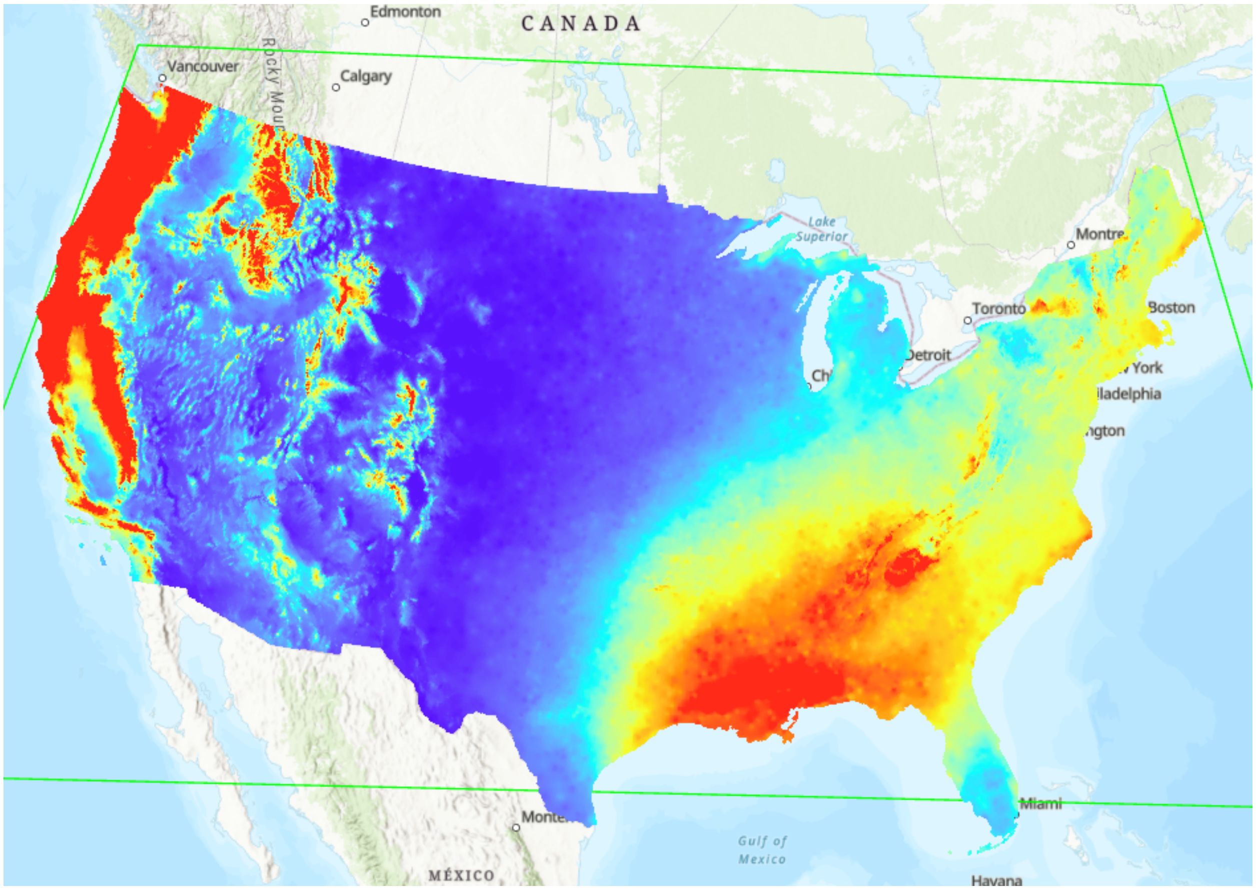
Section 3:
This section of the lab goes over the process of creating an ETL that downloads the 30 year normals data from PRISM and creates a space time cube with the dataset. The general workflow is shown in figure 4. The data cell labeled as “PRISM” is the raw output data from the pipeline created to pull the precipitation data from its server. The only thing done to the data after it was retrieved with a simple requests.get() function was unzipping the file. The PRISM input into the Add Rasters to Mosaic Dataset is the entire folder of .bil files. The first calculate field adds the precipitation (ppm) field to the mosaic’s Footprint table. The second calculate field adds the timestamp date field to the Footprint table. This timestamp is just a field that specifies what month each of the rasters being used is from. These both enable the mosaic to be represented in a multidimensional raster layer. Finally with the multi dimensional raster layer created, the Space Time Cube can be made.

Figure 4: ModelBuilder model of workflow



Results
The creation of TIN or DEM from lidar point cloud data was an important process to learn. The creation of a pipeline is very important in forestry lidar data in particular, as it gives you the ability to automate the process of making surfaces rather than having to go in and make a new one by hand each time.
Section 2 and section 3 are closely related, each going into detail showing the process of displaying data in different dimensions. The analysis shows that certain data favors a particular display method over another. While 2D vs. 3D view is largely the same, space time cubes enhance analysis and visualisation.Figure 5 shows one of the layers of the space time cube, the first raster in the series. I found that percent clip stretch type is a better fit for this dataset than the default setting of Min-Max. Having a space time cube dataset allows for all sorts of new methods of analysis and tools that can not be used on a traditional raster. Options like emerging hotspot analysis now become available.

Figure 5: Single layer of the space time cube



Results Verification

I'm confident that the data pipelines for each section 1 and 3 are correct as I was able to simply verify that the process gave identical results as when I did it by hand and downloaded the data from the source GUI and imported the data via the ArcPro GUI. For section 3, the data shown in figure 5 matches the preview that the PRISM website shows for the corresponding month. I'm able to verify that the space time cube worked as I was able to use it as an input for the emerging hotspot analysis tool as a test to check validity.

Discussion and Conclusion

As stated in the results, the creation of a pipeline for lidar is a great tool to learn as it will allow for the automation of the repetitive task. Simply retrieving the lidar data and creating a visualization of it is often very useful in many forestry applications. That is often the first step in a more comprehensive analysis. As for the space time cube section, looking at the space time cube alone does not make much sense as a visualization technique, as you can't see anything more than the outside layers of data. Time is not traditionally shown as the 3rd dimension so it may not be as easy to interpret as 2D raster data. This is when an animation comes into play, the change of time can be abstracted out of the 3rd dimension and insead shown as an animation that flips through the rasters each cycle of the clip. In many cases, animating a space time cube in 2D makes for a better visualization than simply displaying the space time cube.

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	**27**
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	**22**
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	**19**
		100	**96**