Lab 2 Report

Title: ETL in ArcPro Jupyter Notebooks and arcpy for downloading files and determining an optimal path between two points using Cost Raster.
Notice: Dr. Bryan Runck
Author: Zain Ul Abdin Siyal
Date: 11/02/2022

Project Repository: https://github.com/Zain1443/GIS5571.git
Google Drive Link:
Time Spent: 50 hours

Part 01

Abstract

With hundreds of libraries available for the ETL and other requirements, notebooks can provide a consistent environment for team members to execute the codes. This lab focuses on developing ETLs in ArcPro Jupyter Notebooks and arcpy to streamline the process of downloading different file formats, converting them using python modules and packages, producing pdfs, and creating a timeseries animation. Several python modules and packages are imported to ensure the reusability of code. The two datasets (.las and .bil) used in this lab are from MN DNR and PRISM Climate Group.

Problem Statement

With access to hundreds of libraries for the ETL and other needs, notebooks can offer a uniform environment for team members to run the same code. Downloading various geodata from different sources manually is relatively straightforward; however, creating an ETL will streamline this task. This lab is about creating ETLs in ArcPro Jupyter Notebooks and arcpy that can download LAS and BIL files, converts them into different file formats, export pdfs with correct visualization, and export an animation of the timeseries.

Table 1. Required Data

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Elevation data	LAS dataset from Minnesota Department of Natural Resources	Raster	Pixels (LIDAR)	MNDNR	No
2	Precipitation data	Annual 30-Year Normals .bil files for precipitation from PRISM Climate Group	Raster	Pixels (Precipitation value)	Prism Climate Group	No

Input Data

The input datasets required to create ETLs in ArcPro Jupyter Notebooks and arcpy are tabularized below. The first dataset is from the Minnesota Department of Natural Resources (MN DNR) and it is in LAS format, an industry-standard binary format for storing airborne LIDAR data.

The second dataset used for this part of the lab is the monthly precipitation over a specified area for a 30-year period from PRISM Climate Group. They generate geographic climate datasets to reveal short- and long-term climate patterns by assembling climate observations from a variety of monitoring networks, employing advanced quality control procedures. The PRISM data set covers the whole continental United States and has a cell raster size of 4km.

Table 2. Input Data for this lab.

#	Title	Purpose in Analysis	Link to Source
1	LAS dataset containing lidar point cloud data from MN DNR	To convert .LAS files into both TIN and DEM files and export pdf of DEM and Tin with correct visualization.	MNDNR
2	Annual 30-Year Precipitation data in .bil files from PRISM Climate Group	To convert data into a spacetime cube, export to disk, and export animation to timeseries.	Prism Climate Group

Methods

In order to build an ETL in ArcPro Jupyter Notebooks and arcpy that downloads .LAS files from the Minnesota Department of Natural Resources (MN DNR); first, the working directory is defined. Several python modules and packages such as arcpy, requests, os, io, and sys are imported to ensure the workability and reusability of code. Subsequently, the framework import requests (a versatile HTTP library) are used to download the .LAS file from the MN DNR webpage using the file URL. The methodology to perform said task is summarized in the data flow diagram, shown in Figure 1.

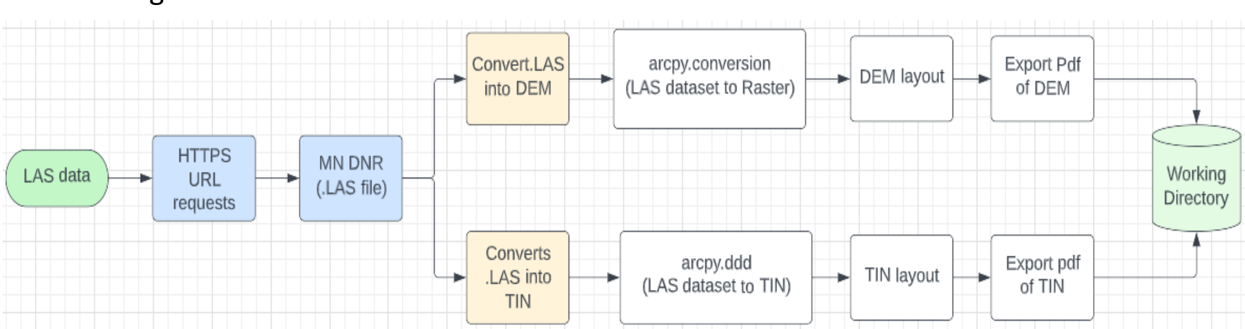


Figure 1: Flow Diagram of the Methodology for Lab 02 (part 1.1).

Once the .LAS file is downloaded, arcpy.conversion.LasDatasetToRaster and arcpy.ddd.LasDatasetToTin tool is used to convert the .LAS file into DEM and TIN, respectively. Next, the created DEM and TIN files are saved to the disk and then pdf of both page layouts or data frames of a map document is downloaded with correct visualization using lyt.exportToPDF. The python code of all the series of tasks is submitted with this lab report.

The methodology to download the annual 30-Year Normals .bil files for precipitation from PRISM Climate Group using an ETL in ArcPro Jupyter Notebooks and arcpy is shown in figure 2. Similar to .LAS part, the framework import requests library is used to download the .BIL file from the PRISM webpage using the file URL. The BIL data is then converted to TIFF using the arcpy convert file to tiff tool. A blank Mosaic dataset is created via arcpy.management.CreateMosaicDataset and raster data is added to the mosaic dataset, a pointer to the location of the data is stored within the mosaic dataset. Lastly, space time cube is created and animation.gif of the timeseries is exported.

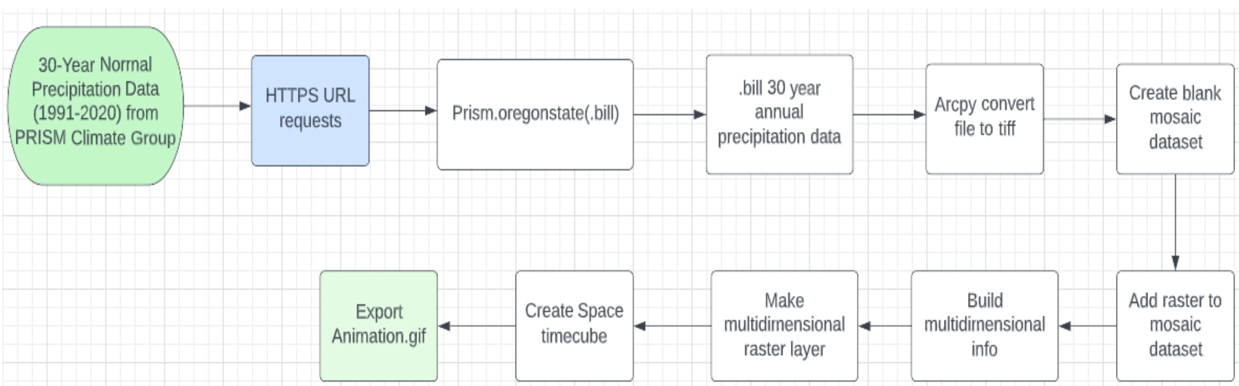


Figure 2: Flow Diagram of the Methodology for Lab 02 (part 1.3).

Results

The methodology described in figure 1 is sternly followed to get the outputs/results, as shown in figure 3. The coding done in ArcPro Jupyter Notebook to get these results is submitted with this lab report on GitHub and Canvas.

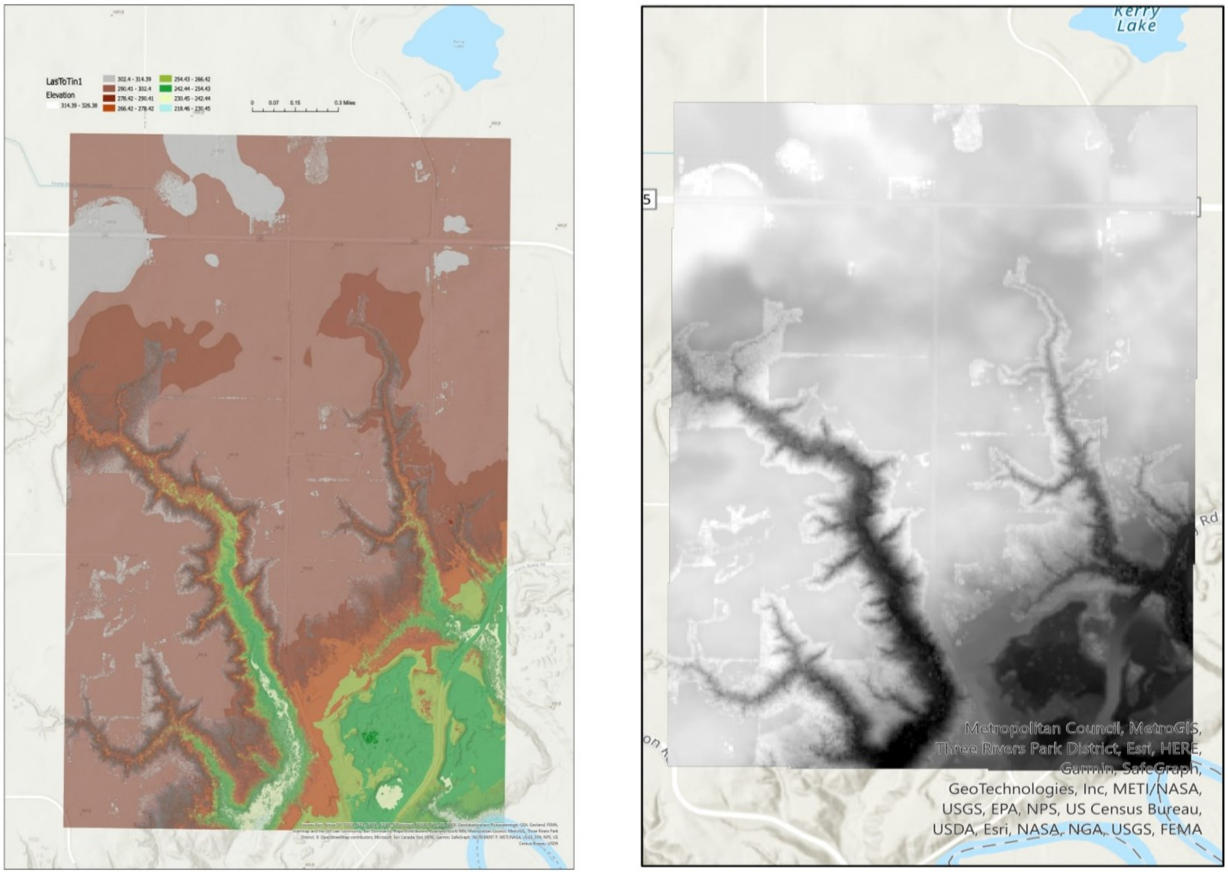
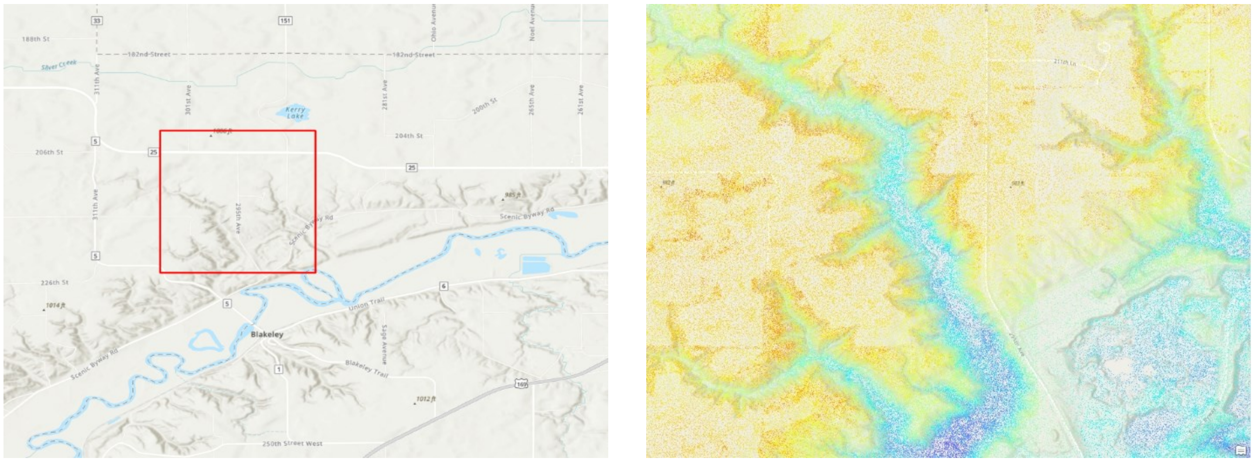


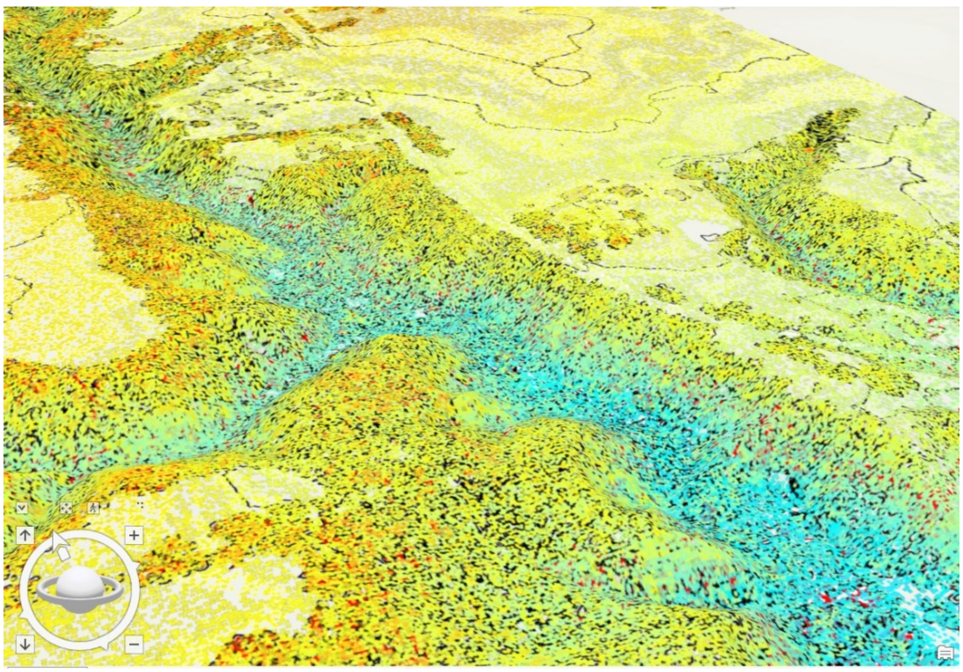
Figure 3: The downloaded TIN and DEM files.

A LAS dataset is visualized in ArcGIS in both 2D and 3D using ArcMap and ArcScene in order to do the side-by-side exploratory data analysis. In both cases, the LAS data is shown as a gradient of height or properties that are connected to elevation. It has been observed that the initial display of a LAS dataset is different between ArcMap and ArcScene.

The point count of LAS datasets is frequently too high to display at first when they are added to ArcMap. As a result, the smallest bounding boxes for each LAS file in the LAS collection are displayed. When a LAS dataset is added to ArcScene, the scene will always display the minimal bounding boxes that correspond to the extent of each LAS file in the LAS dataset. Figure 4 shows the 2D map and 3D Scene of the .las file.



(a.)



(b.)

Figure 4: (a.) 2D map and (b.) 3D Scene of the .las file.

Space-time cubes illustrate the temporal evolution of phenomena within geographic space. By using time-series analysis, integrated spatial and temporal pattern analysis, and 2D and 3D visualization techniques, spatiotemporal data can be viewed and analyzed by building a space-time cube. The results obtained by following the methodology shown in figure 2 are illustrated below. Figure 5 shows the mosaic dataset created using data management tools via an ETL, while figure 6 depicts the Space Time Cube of 30-year precipitation normals.

Figure 5: A multidimensional Mosaic dataset.

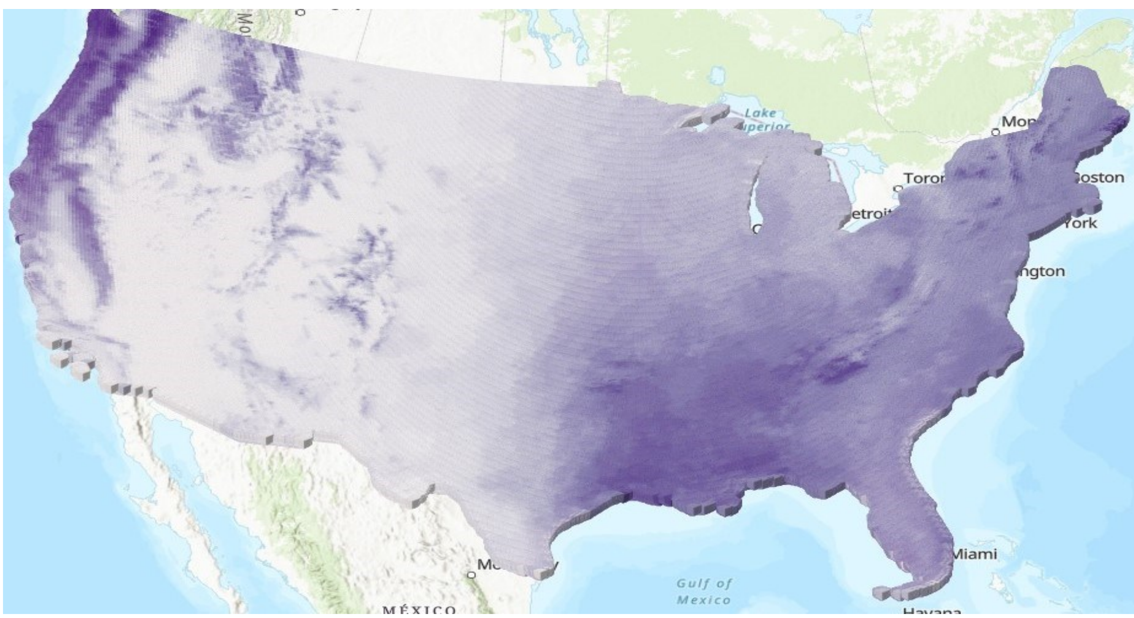
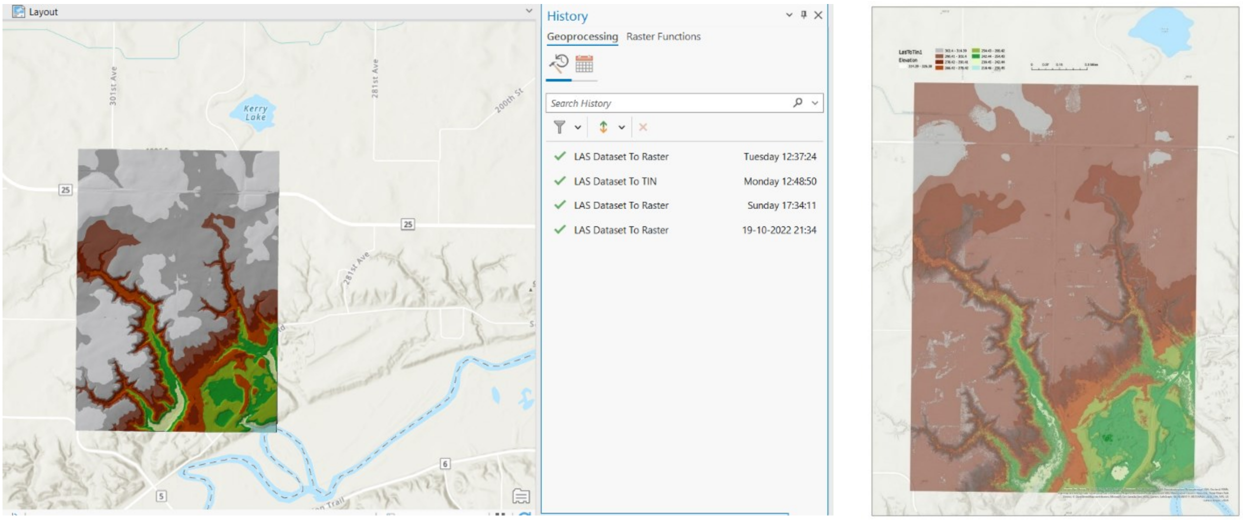


Figure 6: Space Time Cube of 30-year precipitation normals.

Results Verification

For the verification of results, I manually downloaded the subject DEM and TIN data and compared them with the ones I got through coding. I did a side-by-side visual comparison and thought that both documents looked identical.



(a) *(b)*

Figure 3: Side-by-side comparison of TIN data through (a) ETL (b) Manual Download.

For 1.2, I made the same visual comparison as I compared the 2D and 3D output with the TIN version from LAS, obtained in 1.1. I struggled slightly to validate my results for the last part of this lab (1.3).

Discussion and Conclusion

In this lab, I learned about different data formats and how to build an ETL in ArcPro Jupyter Notebooks and arcpy that can download LAS dataset, converts it into Digital Elevation Model (DEM) and Triangulated Irregular Network (TIN), and exports their pdf with correct visualization. Moreover, I also created an ETL in ArcPro Jupyter Notebooks that downloads Band interleaved by line (BIL) files for precipitation from PRISM. Additionally, I come to know about how LAS dataset is visualized in ArcGIS Pro in both 2D and 3D using ArcMap and ArcScene in order to do the side-by-side exploratory data analysis. All in all, I found 1st part of this lab slightly challenging, but I learned several new things. Lastly, Michael was kind enough to help several times.

References

1. Lidar Elevation Data for Minnesota: 2008-2012. MN Geospatial Commons. Website. October 19, 2022. https://www.mngeo.state.mn.us/chouse/elevation/lidar.html#data
2. Explore your raster data with Space Time Pattern Mining. Environmental Systems Research Institute. Website. October 18, 2022. https://www.esri.com/arcgis-blog/products/arcgis-pro/analytics/explore-your-raster-data-with-space-time-pattern-mining/

3. Introduction to arcpy.mp. Environmental Systems Research Institute. Website. October 18, 2022. https://pro.arcgis.com/en/pro-app/latest/arcpy/mapping/introduction-to-arcpy-mp.htm
4. Create Space Time Cube From Multidimensional Raster Layer (Space Time Pattern Mining). Environmental Systems Research Institute. Website. October 18, 2022. https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/createcubefrommdrasterlayer.htm
5. 30 years Normal. Northwest Alliance for Computational Science & Engineering. Website. October 19, 2022. https://prism.oregonstate.edu/normals/
6. An overview of displaying LAS datasets in ArcGIS. Environmental Systems Research Institute. Website. October 21, 2022. https://desktop.arcgis.com/en/arcmap/10.7/manage-data/las-dataset/an-overview-of-displaying-las-datasets.htm
7. ExportToPDF. Environmental Systems Research Institute. Website. October 21, 2022. https://desktop.arcgis.com/en/arcmap/10.3/analyze/arcpy-mapping/exporttopdf.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	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	26
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	18
		100	94