Lab 2 Report: Cost Surface

Title: Cost Surface Model
Notice: Dr. Bryan Runck
Author: Diego Osorio
Date: 11/2/2022

Project Repository: https://github.com/osori050/GIS5571/tree/main/Lab2
Google Drive Link:
Time Spent: 17

Abstract

This lab's purpose consists of creating an ETL pipeline to retrieve data from Google Places and the Minnesota Geospatial Commons (MGC) to develop cost surface models highlighting the cheapest routes for Dory to get from her house to the North Picnic area. The cost of the surface increases due to steep slopes, muddy landcover types, crossing water bodies without bridges or waders, and deviation from the shortest distance between the house and the park. Map algebra was performed by rasterizing all the variable datasets and reclassifying them on a scale from 1 to 4 where 4 is the most expensive value. Additionally, the pipeline was developed so the user could create as many models as necessary by inputting different weighting factors to each variable. The models generated are highly accurate when compared with the expected results.

Problem Statement

This lab aims to find an optimal route to go from Dory's house (outside of Whitewater State Park) to the North Picnic area. As several variables are added, such as streams, bridges, slopes, etc., the cost surface model will show the places where Dory would prefer to walk to get to the park based on different weights assigned by the user to each variable. Figure 1 illustrates the conceptual cost surface model.

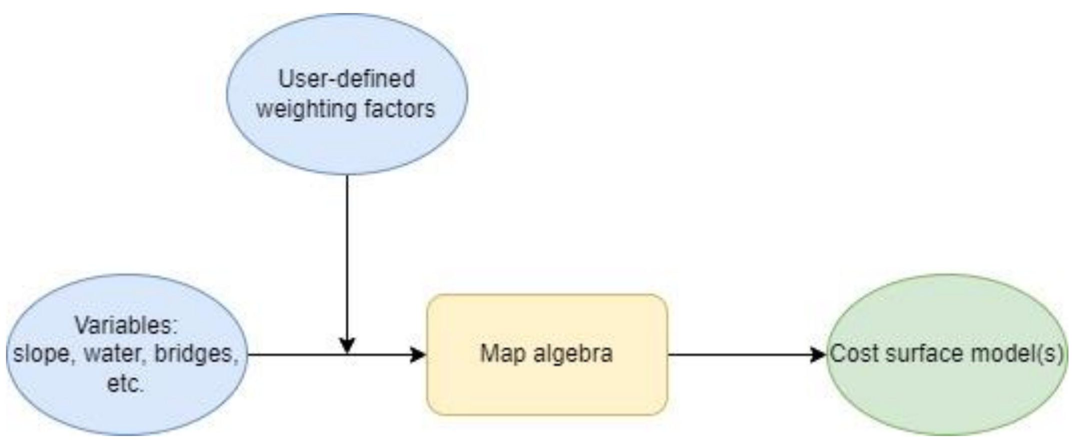

Figure 1. Conceptual cost surface model

Table 1. Data required for cost surface analysis

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Start and end points	Pair of coordinates	Point geometry		Google Places	Convert coordinates from strings to points, project to NAD 1983 UTM Zone 15N, create a bounding box, and generate a buffer around the bounding box (AOI).
2	Elevation dataset	Minnesota DEM	Raster	Elevation	Mn GeoSpatial Commons	Clip to AOI, and generate slope dataset
3	Farm fields	Cropland data	Polygon	Land type	Mn GeoSpatial Commons	Clip to AOI
4	Water bodies	Streams or watercourses	Line		Mn GeoSpatial Commons	Clip to AOI
5	Bridges	Bridges	Point		Mn GeoSpatial Commons	Clip to AOI, and snap points to streams (lines)

Input Data

First, the start point coordinates were given in the lab instructions, while the end point was obtained from Google Places through a personal API key. From the Minnesota Geospatial Commons portal, 4 datasets were downloaded: i) a state 30-m DEM, ii) a geodatabase containing the cropland layer 2021 raster with information about the different landcover types (including farm fields) in the state, iii) a line shapefile encompassing the state streams, and iv) a point shapefile representing the state bridges.

Table 2. Input data

#	Title	Purpose in Analysis	Link to Source
1	End Point	Raw input strings containing the coordinates of the North Picnic area	https://maps.googleapis.com/maps/api/place/findplacefromtext/json?input=North%20Picnic%20area%20St%20Charles%20Minnesota&inputtype=textquery&fields=formatted_address%2Cname%2Crating%2Copening_hours%2Cgeometry&key=YOUR_API_KEY

2	Minnesota Digital Elevation Model - 30 Meter Resolution	Raw input dataset for slope generation	https://gisdata.mn.gov/dataset/elev-30m-digital-elevation-model
3	Cropland Data Layer 2021, Minnesota	Extract landcover information, especially farm fields	https://gisdata.mn.gov/dataset/agri-cropland-data-layer-2021
4	Public Waters (PW) and Watercourse Delineations	Dataset with streams	https://gisdata.mn.gov/dataset/water-mn-public-waters
5	Bridge Locations in Minnesota	Locate bridges over water bodies	https://gisdata.mn.gov/dataset/trans-bridges

Methods

To begin with, the North Picnic area (end point) location was obtained through a personal API key from Google Places and stored in a JSON dictionary. These together with those of the start point (coordinates given in the lab instructions) were then stored in lists from which a point shapefile was created. The shapefile was projected to the coordinate system NAD1983 UTM Zone 15 N and an envelope bounding box was used to generate a polygon shapefile. An 8-km buffer was applied to this polygon to create the area of interest (AOI) since the optimal route may not necessarily be within the bounding area.

Additionally, a line between the two points was created, which represents the shortest distance. Then, the Euclidean Distance tool was used to generate a raster with cells showing the shortest distance values from the line to anywhere within the AOI. This raster was reclassified from 1 to 4 by using the geometric interval. This was done to penalize more severely the changes in the distance when going further from the line to keep the optimal route as close as possible to the line. That is, for example, going from 1200 to 1300 m away from the main line is more expensive than going from 200 to 300 m away even when the difference is the same (100 m). Figure 2 shows the workflow diagram of this process.

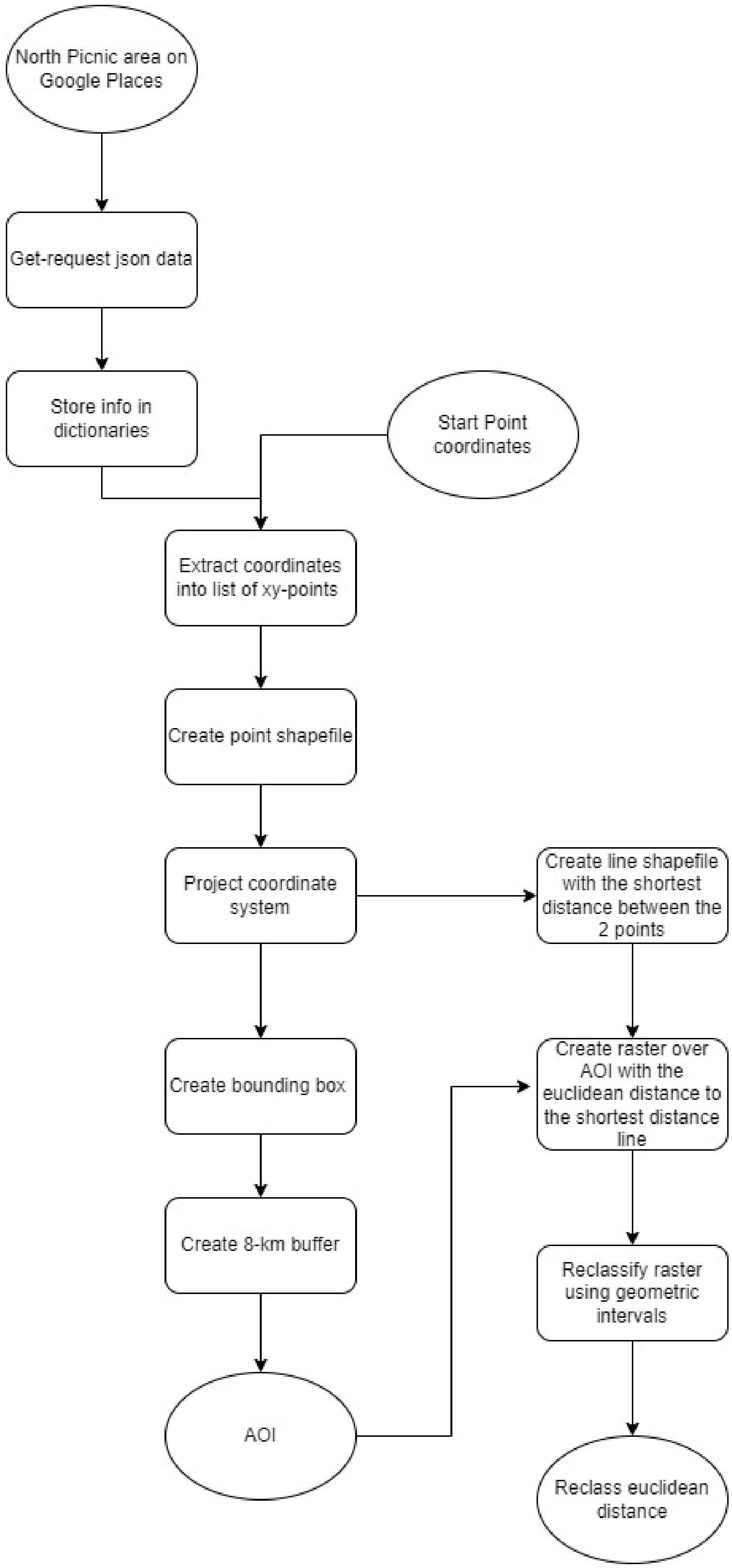

Figure 2. Workflow to create AOI and Euclidean distance variable

Regarding the datasets from Minnesota Geospatial Commons (MGC), first, the DEM was clipped to the AOI. Later, it was used to generate the slope dataset which was eventually reclassified from 1 to 4 by utilizing the geometric interval as well to keep the optimal path the most gradual in terms of the slope. The cropland layer was also clipped to the AOI and reclassified from 1 to 4 as follows: i) 4 for all types of crops (farm fields), ii) 3 for grasslands as they may be muddy but not as much as the farm fields, iii) 2 for forests and shrublands since the leaves intercept rainwater acting as an umbrella which reduces the amount of mud, and iv) 1 for barren and developed areas which are the least muddy land types.

Likewise, the streams were clipped to the AOI, converted to a raster, and reclassified by assigning 4 to all of them (the least wanted condition). Then, as the bridges were represented with a point geometry and were not exactly over the stream-line layer, the stream vector layer clip was used to create a 30-m buffer to include the nearby bridges. Later, the bridges were snapped to the closest stream edges and converted to raster. Said raster was reclassified by assigning a value of 4, the same as the streams, so they can be multiplied, if desired, by the same weighting factor as the streams but negative to counterbalance the cost of crossing water bodies. Figure 3 illustrates the above workflow. It is worth mentioning that all the raster datasets had the same cell size: 30 m.

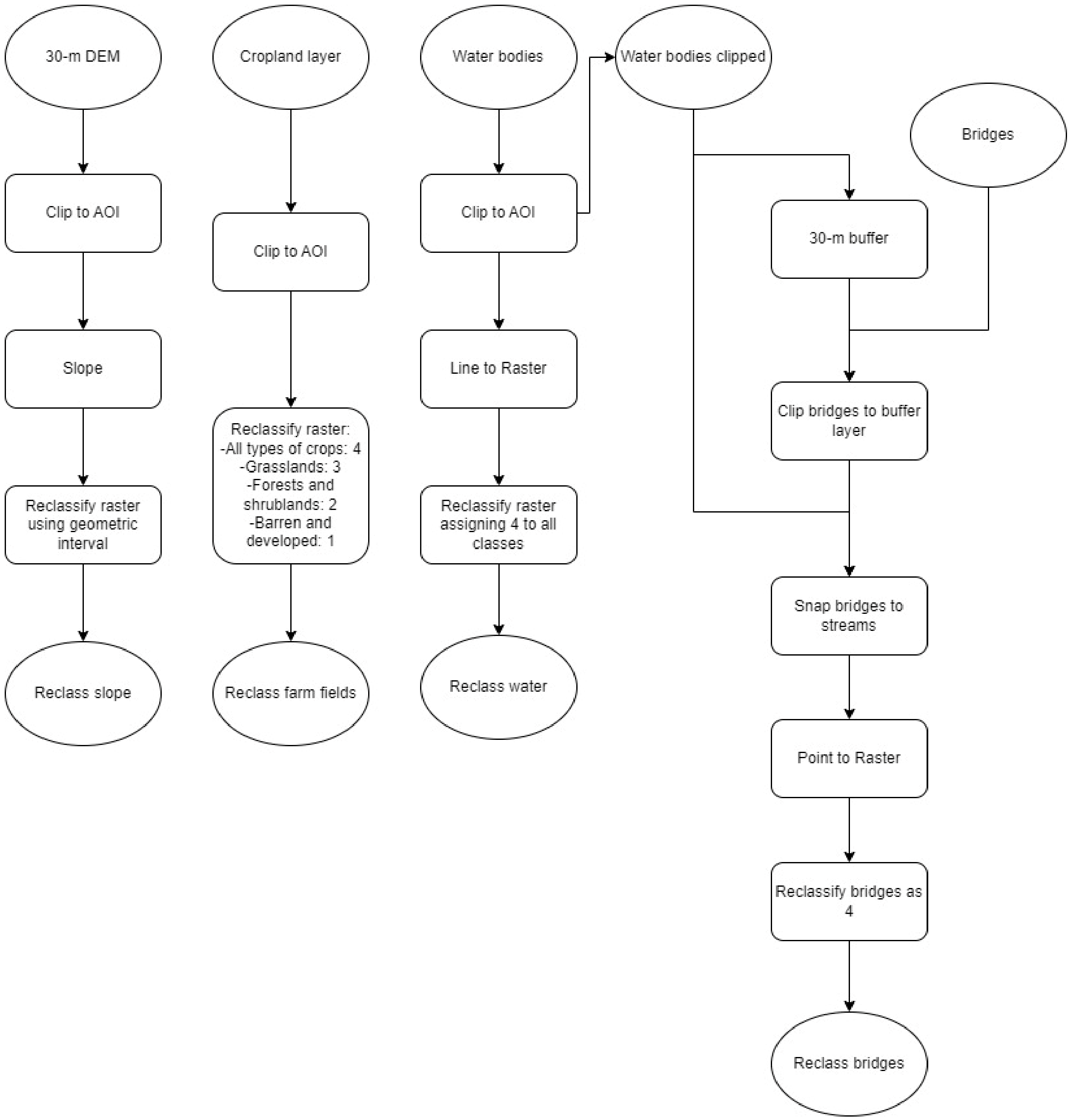

Figure 3. Workflow for MGC datasets

Finally, a loop was created to generate cost surface models based on different weighting factors assigned to each variable. Here, the user is asked whether they want to create a new cost surface and then they input each weighting factor which is then multiplied by each raster. As mentioned above, the user is recommended to assign a negative weighting factor to the bridges to counterbalance the cost of crossing the streams. All the reclassified rasters are then added up by using map algebra and the output is saved to disk. The user is asked again whether a new cost surface is needed: if they reply "yes" the process starts over; if they reply "no", the infinite loop is broken. The cost surface created in each loop is saved with the sequence number of the said loop so as not to overwrite the outputs. This process is summarized in Figure 4.

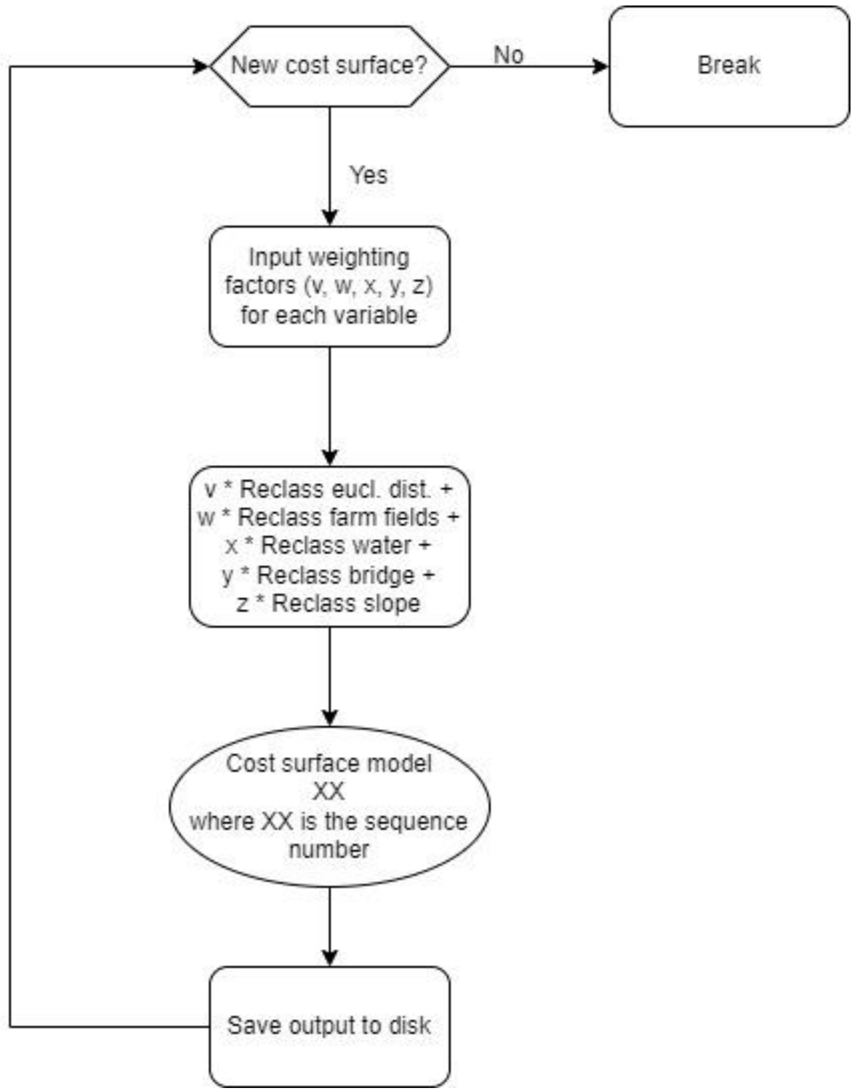

Figure 4. Workflow diagram for cost surface model generation.

Results

The script creates successfully cost surface models based on different weighting factors. For instance, Figure 5 was generated by inputting v=0.5, w=2, x=3, y=-2, and z=1. The scale shows negative values as some bridges, after being rasterized, did not align perfectly over the rasterized streams.

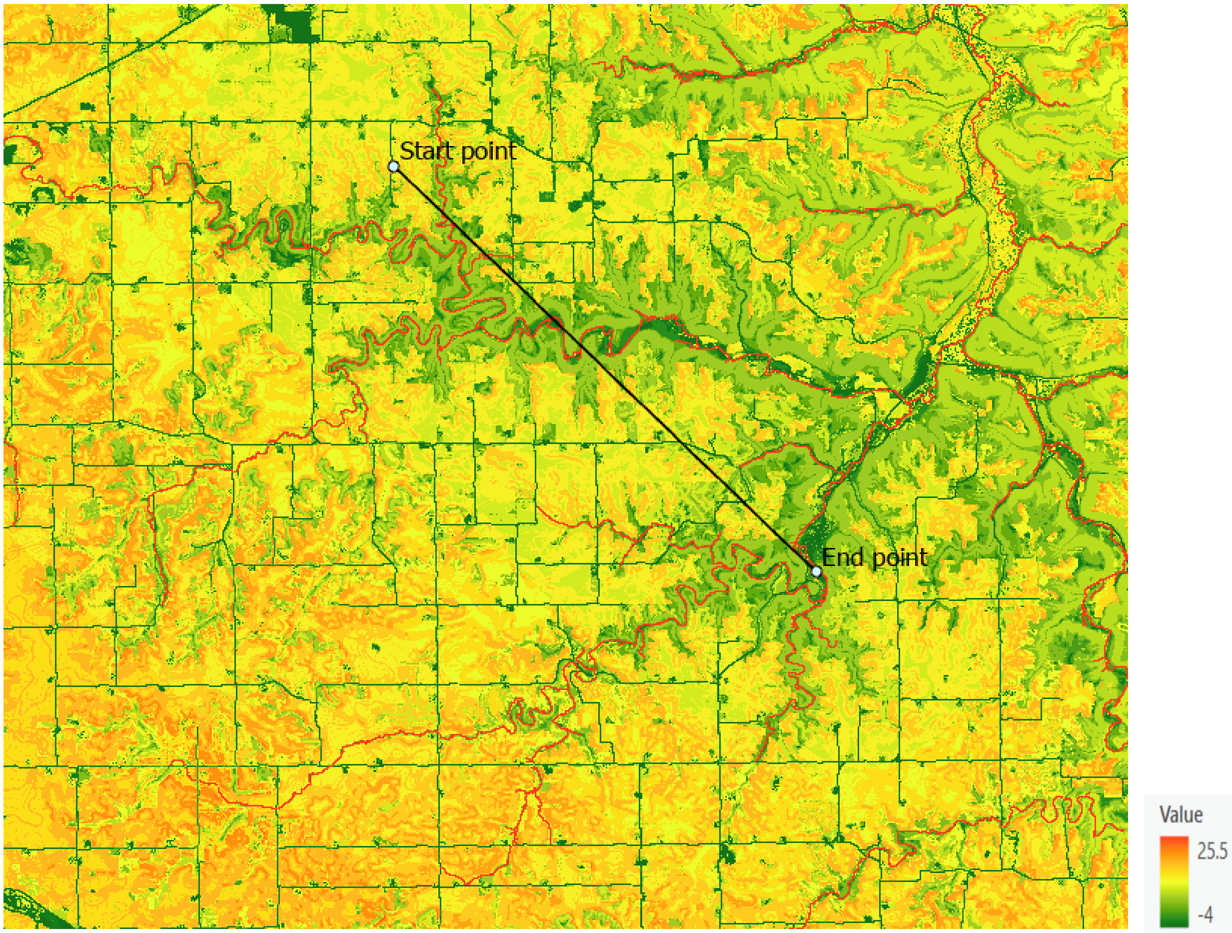

Figure 5. Cost surface model example

Results Verification

The results obtained here are highly accurate as the features in the cost surface models are displayed as expected. For example, in Figure 5 the streams have the highest cost since Dory is not willing to cross any water body as she is not wearing her waders. The lowest-cost routes are the roads since they are not muddy and are generally built on very gradual slopes. Similarly, the farm fields have a medium cost, and the furthest areas from the line connecting the two points are more expensive than the closest ones.

Discussion and Conclusion

Part 2 of Lab2 was a fun exercise. As it was built on Lab1, the data retrieval from the APIs was straightforward to carry out. However, it was very time-consuming due to the several datasets needed to perform the models. Particularly, developing the conceptual model to transform the data was a back-and-forth process due to several bugs. For instance, using the raster calculator tool in Python got me stuck for hours until I found a different approach to receive the user's inputs and perform the map algebra. Likewise, I was having issues when using fiona to create the point shapefile as the Jupyter Notebook did not find the Projections database. That is why I had to use os.environ['PROJ_LIB'] to debug it.

All in all, the result was a very good script that uses an ETL pipeline to create high-accurate cost surface models. Indeed, in further research, the “infinite” loop could be incorporated into a GUI.

References

Google. (2022). *North Picnic area*. Retrieved from Google Places: https://maps.googleapis.com/maps/api/place/findplacefromtext/json?input=North%20Picnic%20area%20St%20Charles%20Minnesota&inputtype=textquery&fields=formatted_address%2Cname%2Crating%2Copening_hours%2Cgeometry&key=YOUR_API_KEY

MNDNR. (2008). *Public Waters (PW) Watercourse Delineations*. Retrieved from Minnesota Geospatial Commons: https://gisdata.mn.gov/dataset/water-mn-public-waters

MNDOT. (n.d.). *Bridge Locations in Minnesota*. Retrieved from Minnesota Geospatial Commons: https://gisdata.mn.gov/dataset/trans-bridges

USDA. (2022). *Cropland Data Layer 2021, Minnesota*. Retrieved from Minnesota Geospatial Commons: https://gisdata.mn.gov/dataset/agri-cropland-data-layer-2021

USGS. (2004). *Minnesota Digital Elevation Model - 30 Meter Resolution*. Retrieved from Minnesota Geospatial Commons: https://gisdata.mn.gov/dataset/elev-30m-digital-elevation-model

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