Final Report

Title: Integrated Assessment of Flooding Risk Factors and Optimal Rescue Routes
Notice: Dr. Bryan Runck
Author: Tzu Yu Ma
Date: December 19, 2023

Project Repository: *https://github.com/TzuYuMa/GIS5571/tree/main/Project*

Abstract

In the face of inevitable disasters, including wildfires, earthquakes, tornadoes, and others, there is an undeniable reality of tragedy. My primary objective is to establish a robust road network capable of planning optimal routes by avoiding high-risk areas, thereby empowering us to estimate suitable paths for effective disaster response. In my home, Taiwan, we frequently grapple with disasters, like typhoons, which can result in significant annual damage. Recognizing the severity of these events, it is crucial to acknowledge that, during rescue operations, certain roads or areas may be restricted, necessitating adaptive route planning by the rescue team to reach the incident site.

For instance, the flooding accompanying a typhoon or heavy rain might damage a bridge or road. While mainstream route platforms generally provide effective route estimates, they may not immediately account for these specific factors. In the current project, where predicting disasters is challenging and their impact is hard to foresee, my aim is to utilize historical precipitation records to analyze high-risk areas. By integrating the risk value into the road network, the goal is to test whether our process can identify optimal fire station locations and routes to the incident location when a disaster occurs. This approach aims to assist in preparing for the unpredictable nature of such events, providing a strategic means of response and mitigating potential damages.

Problem Statement

For this project, the study area will center on the southern part of Taiwan, known for its frequent incidents of flooding, and the process will be divided into three main parts: Analyzing the High-Risk Area, Building Network Dataset, and Estimating the Nearby Fire Stations and Routes.

Firstly, an analysis of the flooding high-risk area will be conducted, focusing on the numerous recorded instances of flooding, particularly during typhoon events in the study area. This analysis will encompass factors such as slope, Digital Elevation Model (DEM), distance to streams, Land Use and Land Cover (LULC), and precipitation.

Secondly, the project will leverage the ArcGIS REST API for network analysis, concurrent with the establishment of the network dataset. Three distinct modes, namely 'Distance,' 'Time,' and 'Risk,' will be configured for network analysis.

Thirdly, 100 random points will be generated within the study area, and the Google Places API will be utilized to identify nearby fire stations. Subsequently, drive-time buffer will be created for each point, determining the proximity of fire stations to potential incident locations. Routes from fire stations to incident locations will then be established, actively avoiding high-risk areas.

#	Requirement	Defined As	(Spatial) Data	Attribute Data (used in this project)	Dataset	Preparation
1	Slope	Slope elevation	Raster data		NASA SRTM	Download data using Google Earth Engine and then clip it to the study area.
2	Digital Elevation Model (DEM)	Digital Elevation 30m	Raster data		NASA SRTM	Download data using Google Earth Engine and then clip it to the study area.
3	Distance to streams	Distance to the streams, river	Raster data		Accumulated from DEM	
4	Land Use and Land Cover (LULC)	"a global land cover map for 2020 at 10 m resolution based on Sentinel-1 and Sentinel-2 data "(Google Earth Data Catalog).	Raster data		European Space Agency (ESA)	Download data using Google Earth Engine and then clip it to the study area.
5	Precipitation	Monthly Rainfall record from 1970-2000 (released in January 2020), resolutions 1km	Raster data		WorldClim	Use the July records, as July typically experiences the highest rainfall in the study area.

						Clip the data to the study area.
6	Road	Roads Dataset	Vector data	Route system	OpenStreet Map	Use SQL to categorize the road network
7	Nearby fire stations	Fire station point	Vector data	Fire station information	Google places API	Need Google places API
8	boundary	County boundaries	Vector data	County name	Department of household registration (Taiwan)	

Table 1. Requirement data

Input Data

The input data for this project comprises eight essential components: slope, DEM, distance to streams, LULC, and precipitation, roads, nearby fire stations, and boundaries. In the analysis of the high-risk area, LULC, precipitation, DEM, slope, and distance to streams will be utilized, requiring preliminary calculations and reclassification. Roads and nearby fire stations will play a role in setting up the road network dataset and conducting network analysis, while the boundary data will define the study area. Further details on these processes will be discussed in the method section.

#	Title	Purpose in Analysis	Link to Source
1	Slope	For analyzing the high-risk area	NASA SRTM
2	Digital Elevation Model (DEM)		NASA SRTM
3	Distance to streams		Accumulated from DEM
4	Land Use and Land Cover (LULC)		European Space Agency (ESA)
5	Precipitation		WorldClim
6	Road	Build the road network dataset	OpenStreetMap
7	Nearby fire stations	For network analysis	Google Places API
8	Boundary	Define the study area	National Land Surveying and Mapping Center (Taiwan)

Table 2. Input data used for this report.

Methods

The methodology can be divided into three parts (Figure 1)

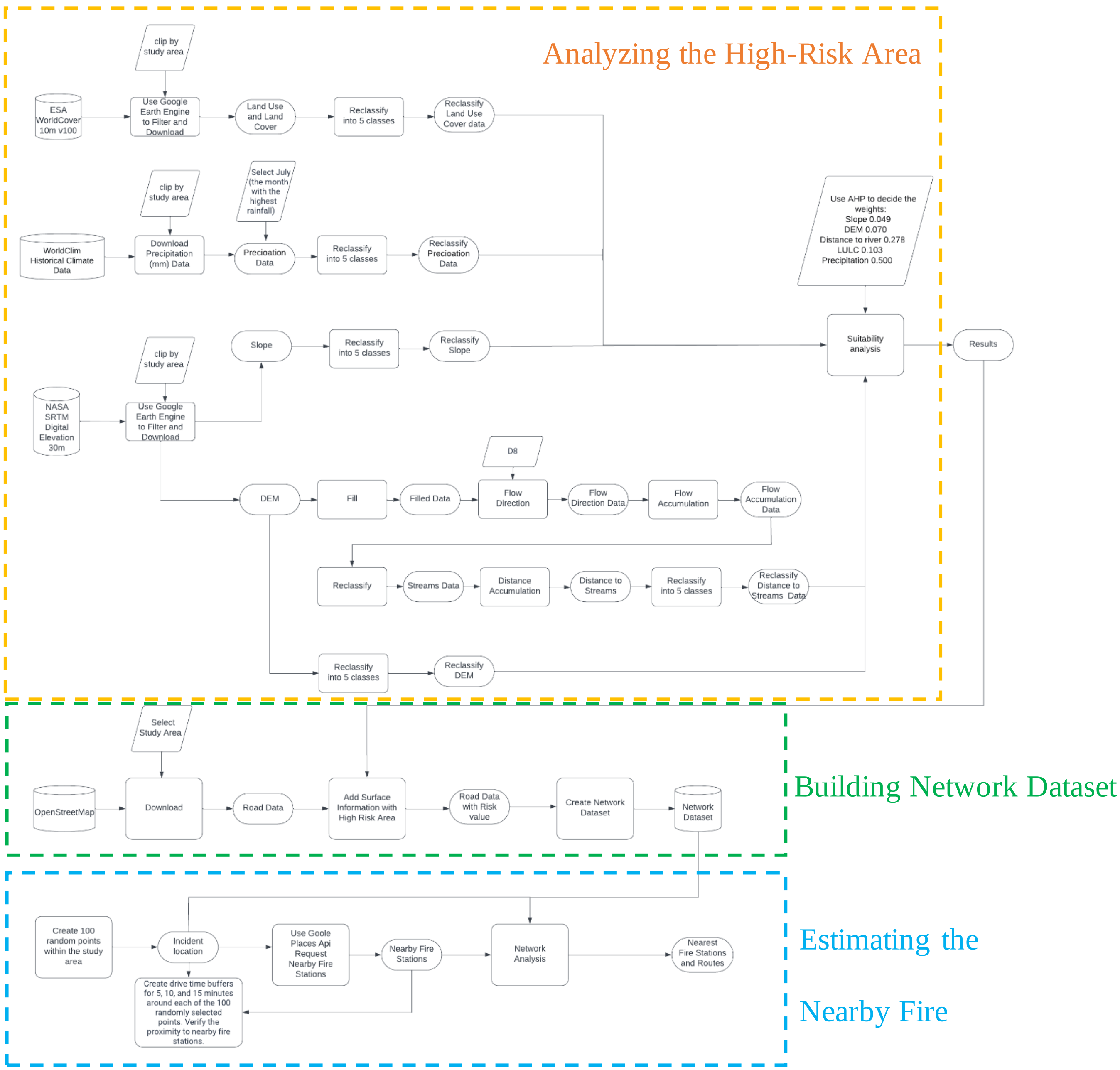


Figure 1. Data-flow diagram

1. Analyzing the High-Risk Area

Download the monthly precipitation data from WorldClim, selecting July as it represents the month with the highest rainfall in the study area. This data will be utilized for analyzing flooding during the period of highest risk. Utilize Google Earth Engine to retrieve LULC information from the European Space Agency (ESA), as well as DEM and slope data from NASA. Subsequently, employ the DEM data to calculate the cumulative distance to streams. Finally, reclassify all pertinent factors into five distinct classes (Table 3, the results are in Figure 2).

Factors	**Reclassified Category**					
	Unit	**2**	**4**	**6**	**8**	**10**
Slope	degrees	20 to 78.986473	10 to 20	5 to 10	2 to 5	0 to 2
DEM	meters (m)	285.588235 to 3917	1524.262745 to 2285.588235	825.086285 to 1524.262745	234.670588 to 825.086275	-45 to 234.670588
Distance to streams	meters (m)	0.072795 to 0.152153	0.049524 to 0.072795	0.031027 to 0.049524	0.014320 to 0.031027	0 to 0.014320
LULC	categories	10 (Tree cover)	20 (Shrubland) 30 (Grassland)	40 (Cropland)	50 (Built-up) 60 (Bare / sparse vegetation) 90 (Herbaceous wetland) 95 (Mangrove)	80 (Permanent water bodies)
Precipitation	millimeters (mm)	88 to 266	266 to 370	370 to 497	497 to 655	655 to 1093

Table 3. Reclassified Category

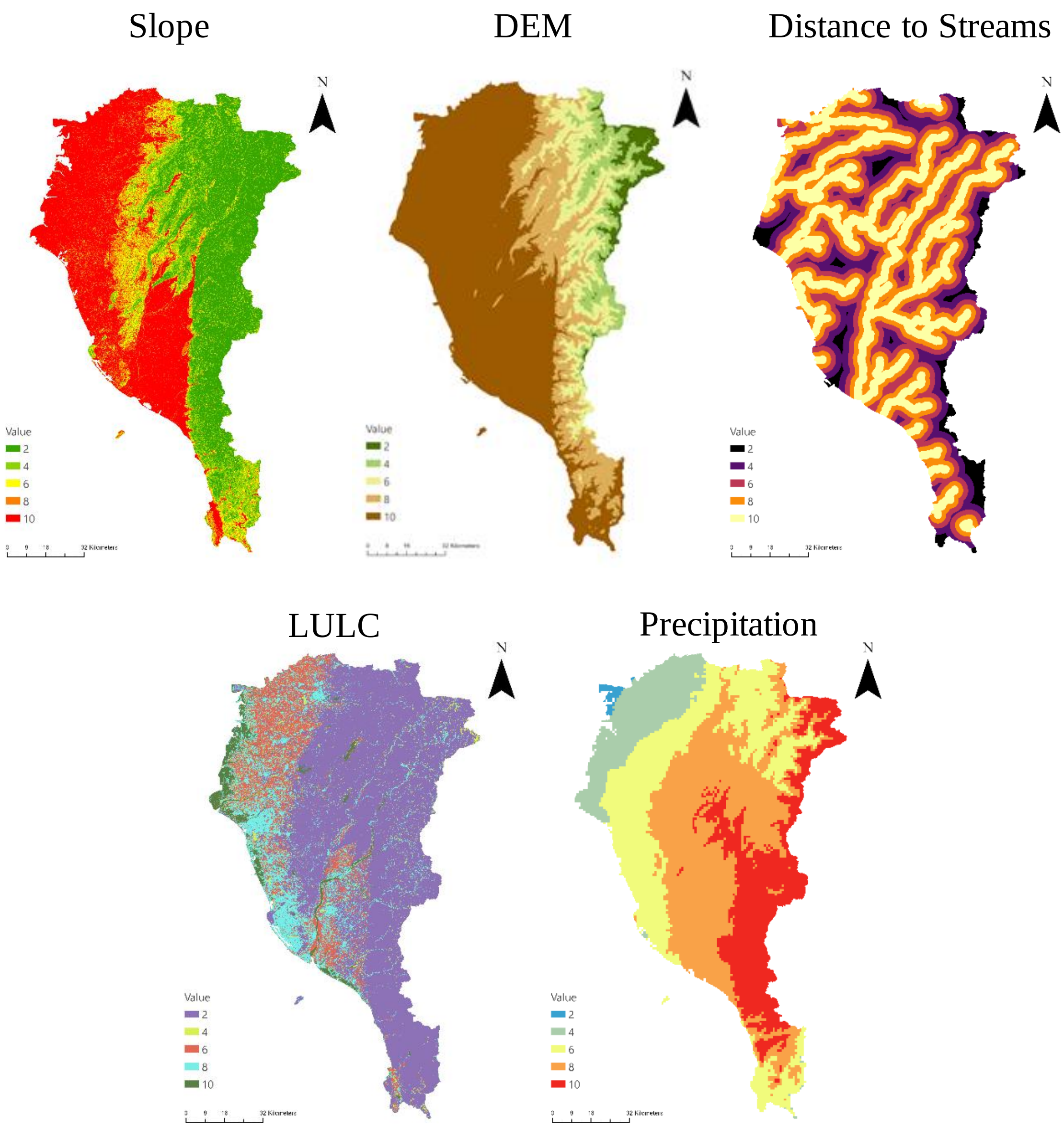


Figure 2. Reclassified Category results.

After reclassifying all the factors, employ the Analytic Hierarchy Process (AHP) to determine the probability weights for analyzing high-risk areas. Developed by Thomas L. Saaty in the 1980s, AHP is a multi-criteria decision analysis method. The AHP process consists of the following steps:

(1) Hierarchy Construction

The primary goal is to analyze the high-risk area for flooding, which can be broken down into five essential factors: LULC, Precipitation, DEM, Slope, and Distance to Streams.

(2) Pairwise Comparisons

Utilize a numerical scale based on Saaty's ratio scale and definitions (Table 4) to compare each factor in pairs, considering their relative importance to the criterion or sub-criterion to which they belong (Table 5).

Numeric Scale	Definition
1	Equal Importance
3	Moderate Importance
5	Strong Importance
7	Very Strong Importance
9	Extreme Importance
2,4,6,8	Intermediate Values
Reciprocal of 1~9	Inverse Comparison

Table 4. The fundamental scale for pairwise comparisons.

	Slope	DEM	Distance to streams	LULC	Precipitation
Slope	1	1/2	1/4	1/3	1/7
DEM	2	1	1/5	1/2	1/6
Distance to streams	4	5	1	5	1/3
LULC	3	2	1/5	1	1/5
Precipitation	7	6	3	5	1

Table 5. The pairwise comparisons used in this report.

(3) Consistency Check

Conduct a consistency check for the pairwise comparisons. According to Saaty's research, if the ratio is less than or equal to 0.1, the comparisons are considered consistent. If not, the pairwise comparison matrix needs to be revisited and adjusted. After calculation, the consistency ratio for the pairwise comparisons in this report is 0.060201261, falling within the acceptable range.

Consequently, the obtained weights (Slope 0.049, DEM 0.070, Distance to streams 0.278, LULC 0.103, Precipitation 0.500) are used to perform the high-risk analysis (Figure 3).

2. Building Network Dataset

Download road data from OpenStreetMap, or any other open-source road data, and construct the network dataset using ArcGIS Pro. Configure three modes: Distance, Time, and Risk. For Distance and Time, employ road length and speed attributes for calculations. For Risk, perform a Spatial Join with high-risk area results onto the road network to assign a high-risk score to each road segment. Set the cost for high-risk, then compute, prioritizing roads with lower high-risk scores.

3. Estimating the Nearby Fire Stations and Routes

   Generate 100 random points as potential incident locations and establish drive-time buffers of 5, 10, and 15 minutes for each random point, corresponding to distances of 4.17, 8.33, and 12.5 km, based on a speed of 50 km/hr (Figure 4). Leverage the Google Places API to gather information on nearby fire stations (Figure 5). Identify 3 to 4 fire stations within each buffer area. The selection of 3 to 4 fire stations allows for strategic placement, ensuring a swift response to incidents. This optimization takes into account factors such as proximity, road network conditions, and response times. Apply network analysis to determine the route depending on the chosen mode (Figure 6 to 8).

Results

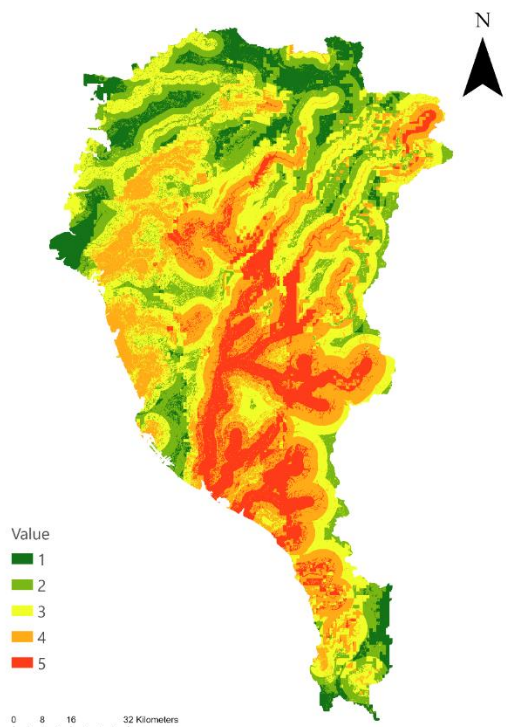


Figure 3. High-risk area result. The value 1 to 5 means lower risk to higher risk.

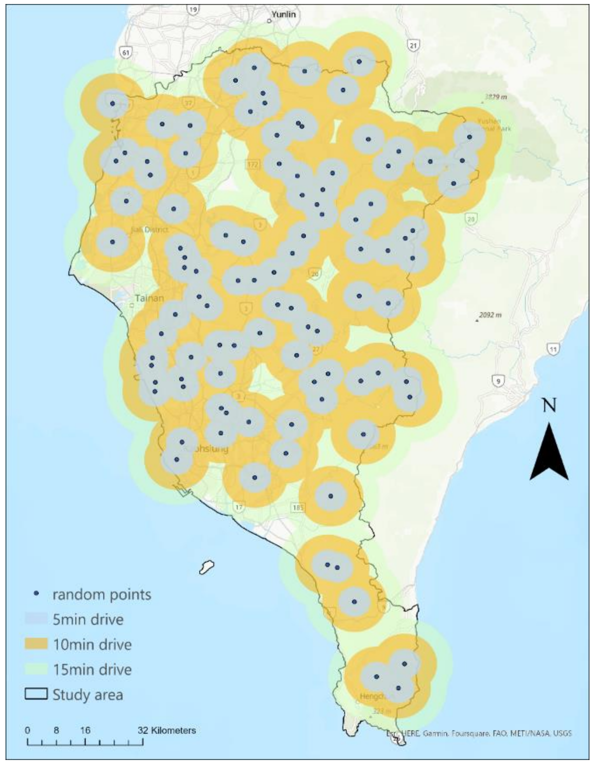


Figure 4. Drive-time buffers of 5, 10, and 15 minutes for each random point.

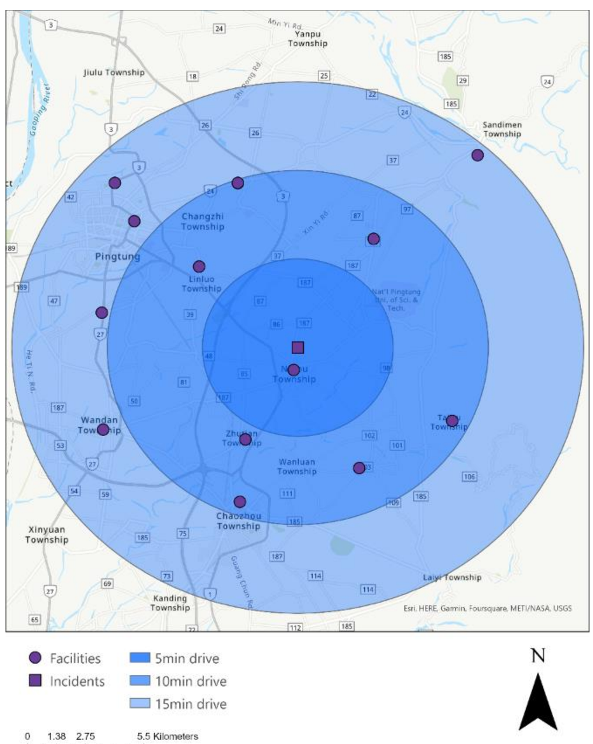


Figure 5. A single random point and create a drive-time buffer, encompassing nearby fire stations (facilities) within the buffer zone.

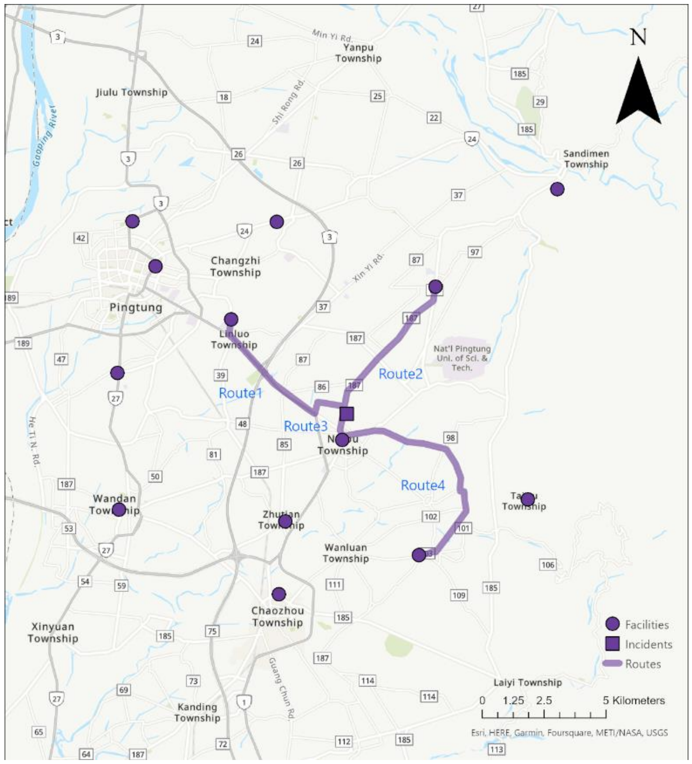


Figure 6. The results for the 'Distance' mode.

Mode	Distance (km)	Time (min)	Risk
Route1	7.009144	7.828099	29.321737
Route2	6.420666	8.833328	35.99979
Route3	1.094717	2.82175	11.024621
Route4	10.522353	14.134888	49.307456

Table 6. Comparison of routes in 'Distance' mode

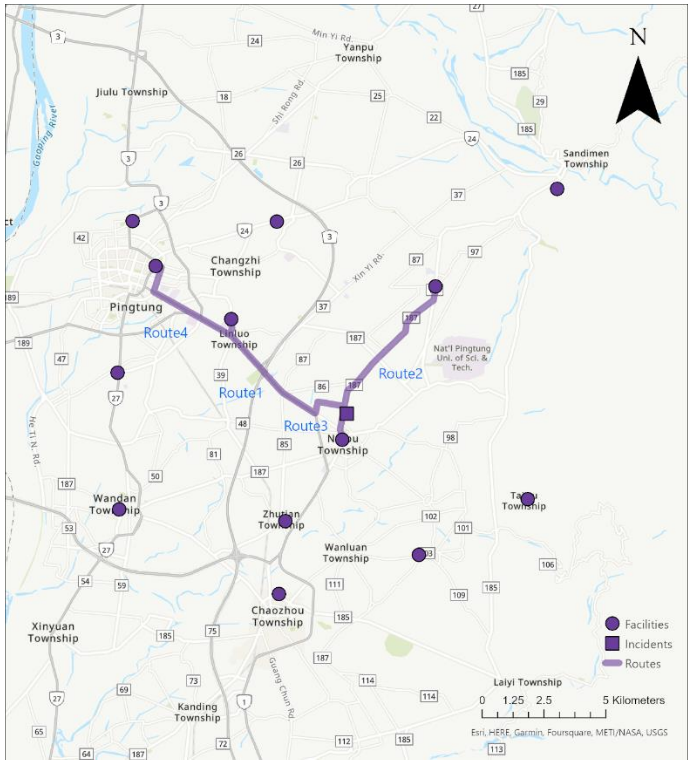


Figure 7. The results for the 'Time' mode.

Mode	Distance (km)	Time (min)	Risk
Route1	7.009144	7.828099	29.321737
Route2	6.420666	8.833328	35.99979
Route3	1.094717	2.82175	11.024621
Route4	10.902212	12.49978	44.194017

Table 7. Comparison of routes in ' Time' mode

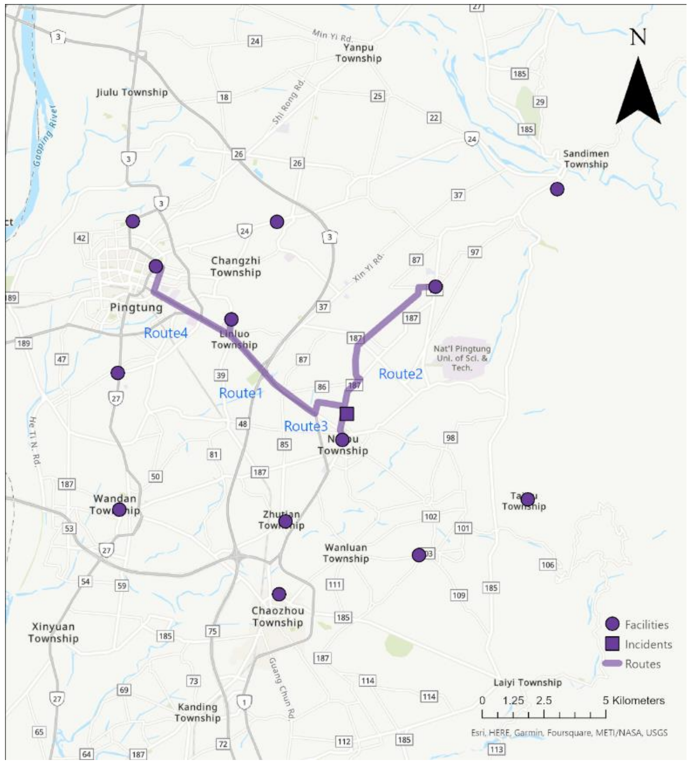


Figure 8. The results for the 'Risk' mode.

Mode	Distance (km)	Time (min)	Risk
Route1	7.009144	7.828099	29.321737
Route2	7.043735	9.58101	30.849531
Route3	1.094717	2.82175	11.024621
Route4	10.902212	12.49978	44.194017

Table 8. Comparison of routes in ' Risk' mode

Results Verification

Utilize the chart to compare three modes based on distance cost, time cost, and risk cost, and verify if the results meet the requirements outlined in the results section. Each mode includes four nearby fire stations with corresponding routes to the incident location. Total the route costs and check if the sum aligns with the chosen mode's criteria. For example, in the 'Distance' mode, the anticipated result should be the shortest distance; in the 'Time' mode, the expected result should be the fastest time; and in the 'Risk' mode, the desired outcome should be the lowest risk value. Some routes may overlap because they fulfill criteria such as the shortest distance, fastest time, or lowest risk value for the four nearby fire stations to the incident location. Therefore, I will assess the sum of all four routes in each mode.

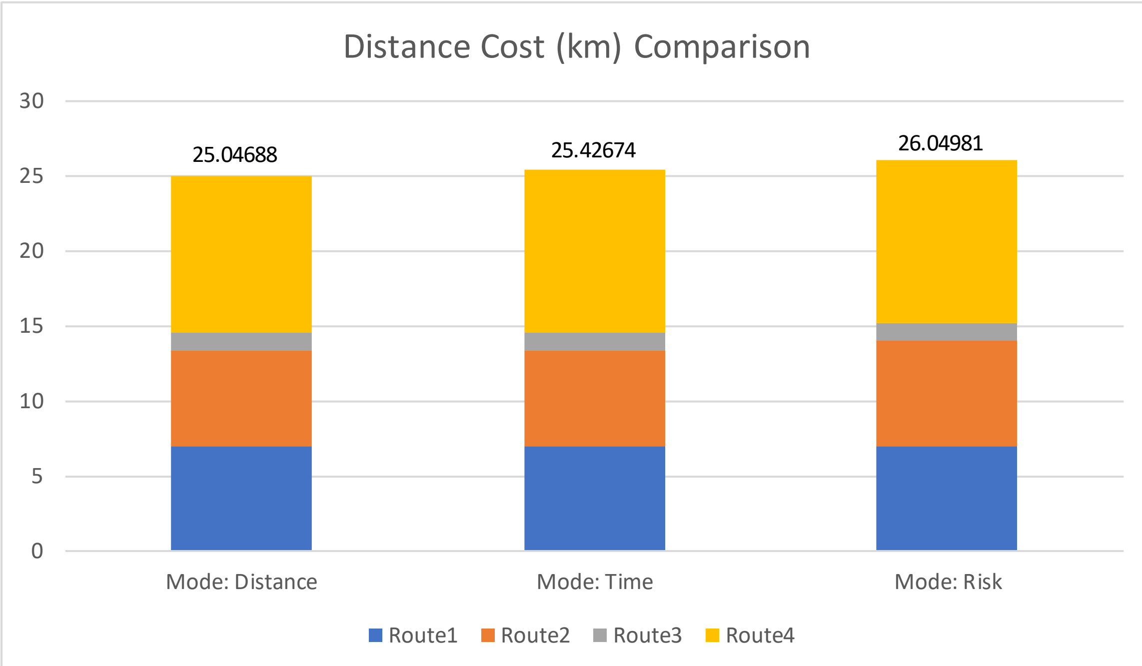


Figure 8. Chart of Distance Cost Comparison, Unit: Kilometers.

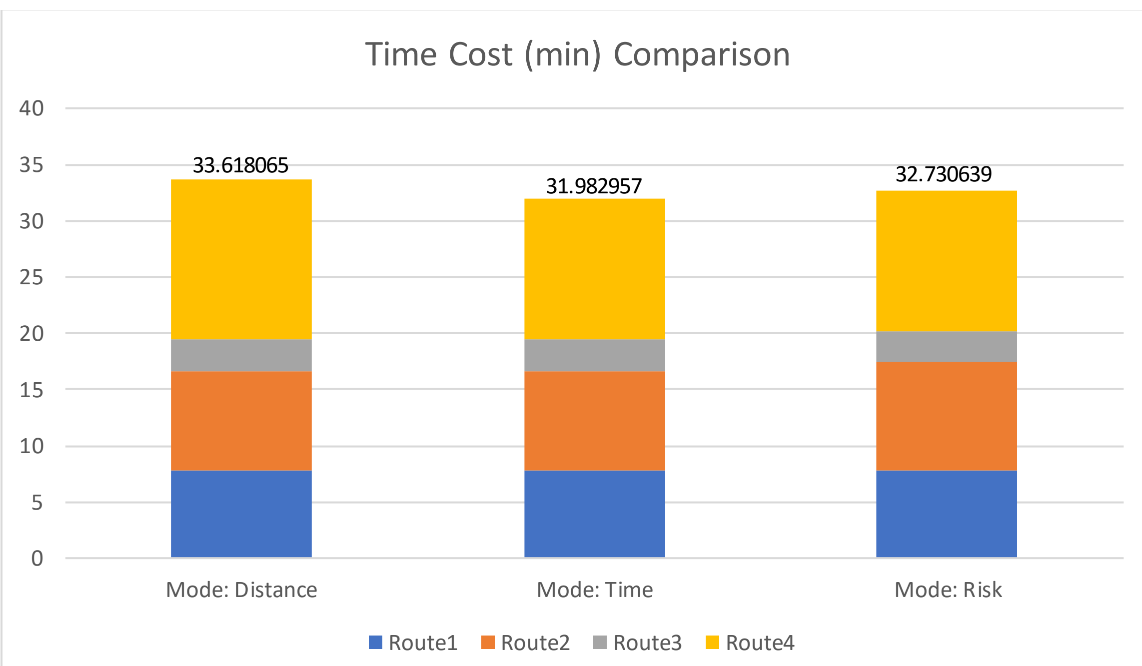


Figure 9. Chart of Time Cost Comparison, Unit: minute.

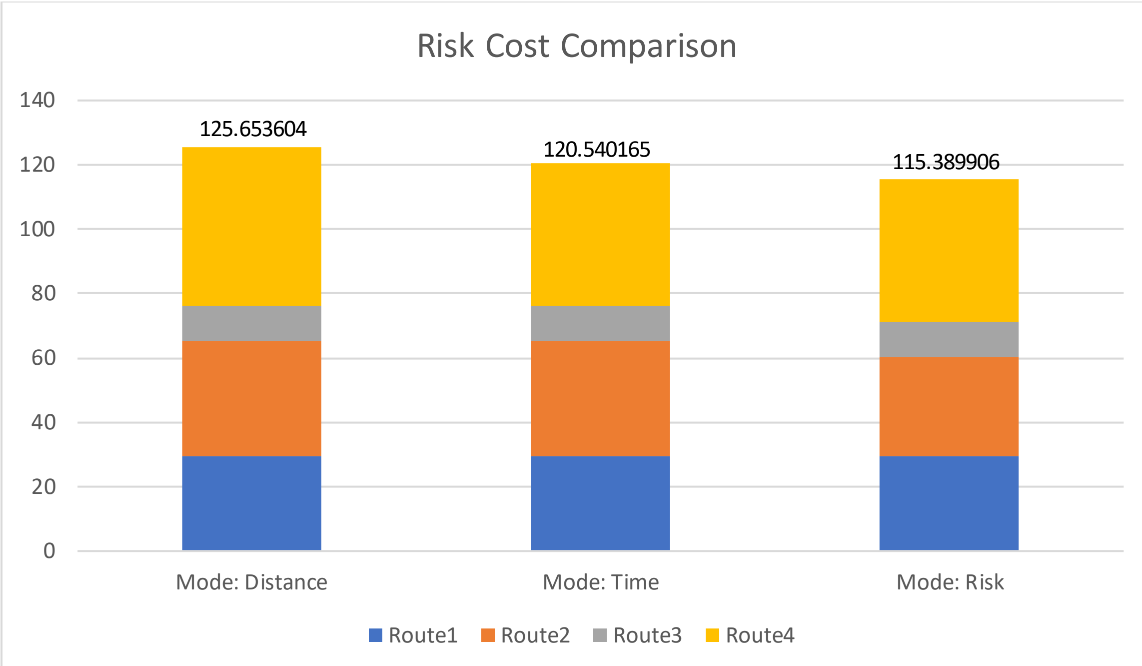
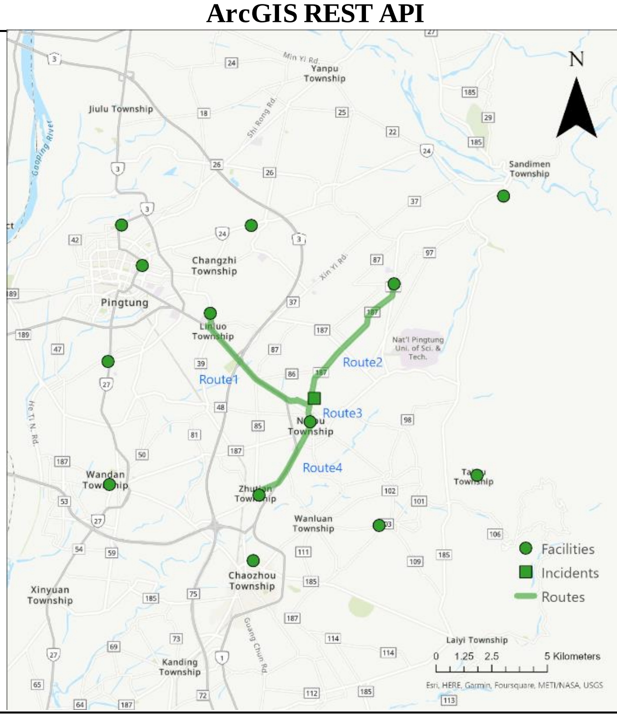
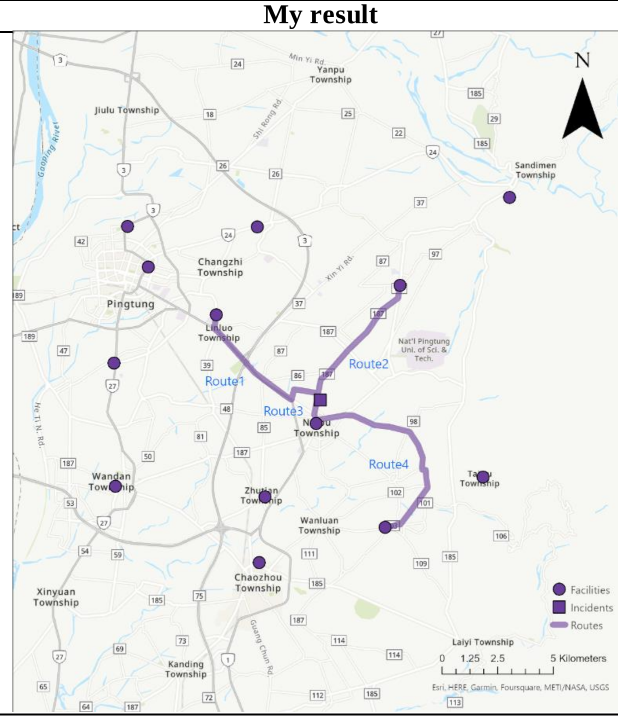


Figure 10. Chart of Risk Cost Comparison, Unit: Not Specified.

In Figure 8 to 10, we can observe that the chosen mode successfully meets the specified requirements. To ensure the accuracy of the route, I also compared the results with ArcGIS REST API below. The outcomes obtained using ArcGIS REST API depict the 'Driving Distance' mode, contrasting with the 'Distance' mode, and the 'Driving Time' mode compared with the 'Time' mode.

	ArcGIS REST API	**My result**
Mode: Distance		
	Distance (km)	Distance (km)

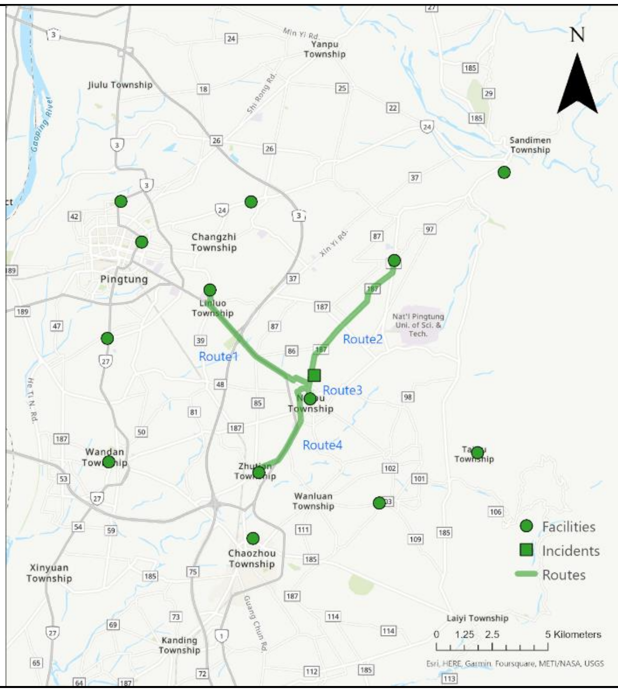
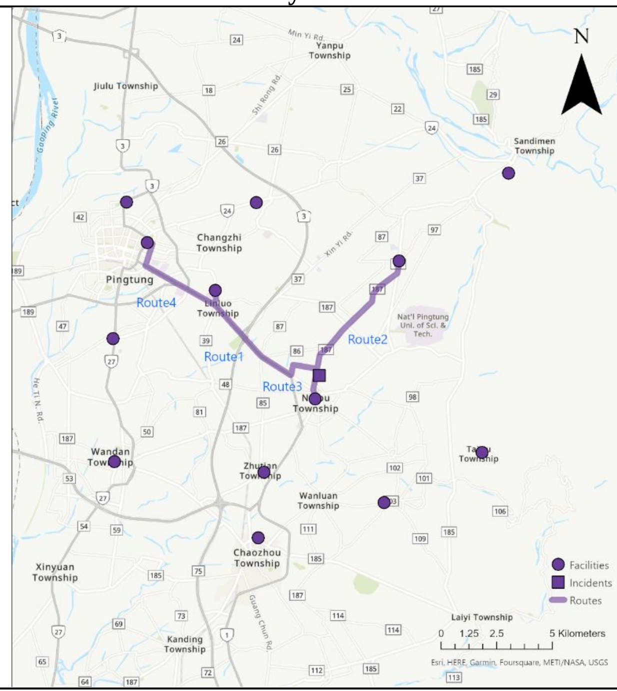
Route1	6.675919	7.009144
Route2	6.410181	6.420666
Route3	1.095845	1.094717
Route4	8.616311	10.522353
	ArcGIS REST API	**My result**
Mode: Time		
	Time (min)	Time (min)
Route1	8.945903	7.828099
Route2	8.641496	8.833328
Route3	2.12001	2.82175
Route4	8.352395	12.49978

Table 9. Compare with ArcGIS REST API in both 'Distance' and 'Time' modes.

In Table 9, it's evident that the 'Distance' mode and 'Time' mode, in comparison to the ArcGIS REST API result, exhibit a notable difference in route 4. However, it's important to note that the time and distance costs are generally very close. While the results may not be identical, they can still serve as valuable reference points.

As ArcGIS REST API lacks a specific mode related to the 'Risk' mode I configured in my network, I verified the results by overlaying the high-risk area outcomes. This comparison was conducted to assess whether the route planning effectively avoids areas identified as high risk.

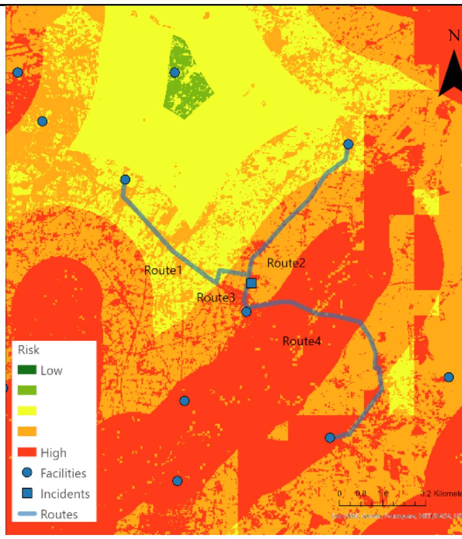
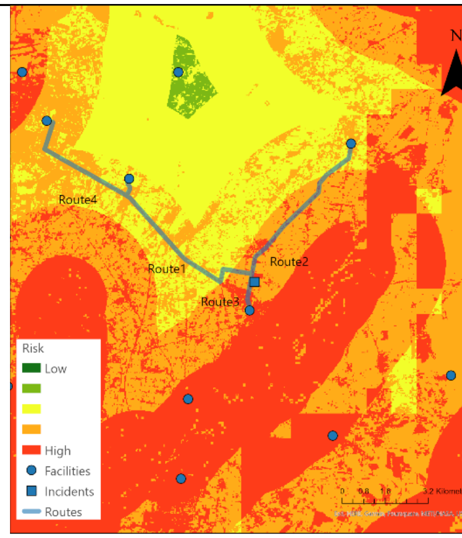
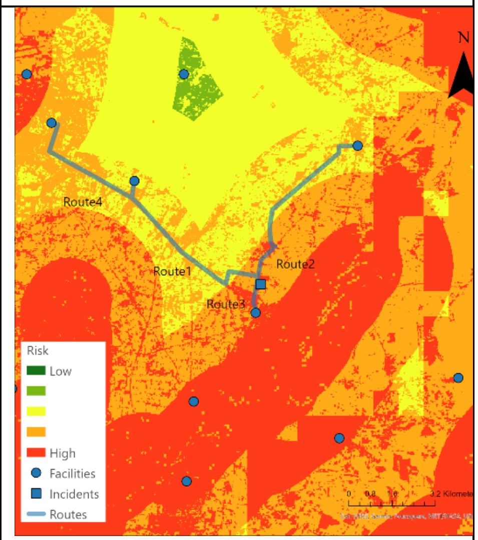
<table>
<tr><td>Mode</td><td>Distance</td><td>Time</td><td>Risk</td></tr>
<tr><td>Result overlay high risk area</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>Describe</td><td>As we can see, Route 4 was crossing more red areas, which identifies the highest risk.</td><td colspan="2">The 'Time' mode and 'Risk' mode results are similar, but we can observe a slight difference with Route 2. In 'Time' mode, Route 2 crosses more orange areas, contributing to an increased risk value.</td></tr>
</table>

Table 10. Overlay the high-risk area results and compare the three modes.

In Table 10, we observe that when the 'Risk' mode is chosen, the route avoids the higher-risk areas compared to the other two modes. This aligns with my expectation.

Discussion and Conclusion

In the beginning, I contemplated making ArcGIS REST API my primary method. Upon experimenting with it, the API performed well and offered a valuable feature—the ability to create barriers to avoid specific areas. This functionality is crucial for my project as it allows me to steer clear of compromised bridges or roads during a disaster. I also explored the Google Routes API, but it fell short in avoiding high-risk areas. Leaflet was considered as well, but its result interface lacked user-friendliness. Unfortunately, Open Route Service is not supported in my study area.

However, the reason for building my road network instead of using ArcGIS REST API is that using ArcGIS REST API involves credit consumption; I would need to purchase credits to access certain functions. Moreover, I am exploring the idea of incorporating high-risk considerations into my road network. This would enable me to establish a 'Risk' mode that anticipates optimal routes, proactively avoiding high-risk areas before any damage occurs. This proactive approach eliminates the need to wait for damage to happen before implementing barriers, a feature not provided in ArcGIS REST API. Additionally, setting up the road network allows me to predict optimal routes before incidents occur, and I can use the 'create barriers' function to avoid unpredictable areas, creating a win-win situation.

Although the results align with my expectations, I encountered challenges during the road network setup process, particularly in configuring road attributes. Some crucial information, such as the max speed column, is incomplete, and certain types of roads are inaccurately categorized. This makes it difficult to establish a hierarchy based on road type in the network. The incomplete road attributes and the lack of setup in the road network may be reasons why the results differ from ArcGIS REST API. When using Service Area in network analysis, some points didn't display the results. While the results satisfy my expectations, further improvement is needed to fully achieve the objectives of using my road network instead of ArcGIS REST API.

References

1. ESRI. (n.d.). Create a network dataset | Documentation. https://pro.arcgis.com/en/pro-app/latest/help/analysis/networks/how-to-create-a-usable-network-dataset.htm
2. ESRI. (n.d.). How Flow Accumulation works | Documentation. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-flow-accumulation-works.htm
3. López-Caudana, E., Ruiz, S., Calixto, A., Nájera, B., Castro, D., Romero, D., ... & Lara-Prieto, V. (2022). A Personalized Assistance System for the Location and Efficient Evacuation in Case of Emergency: TECuidamos, a Challenge-Based Learning Derived Project Designed to Save Lives. Sustainability, 14, 4931. https://doi.org/10.3390/su14094931
4. Qi, X., & Zhou, M. (2020). Integrated energy service demand evaluation based on AHP and entropy weight method. In E3S Web of Conferences (Vol. 185, p. 01046). EDP Sciences.
5. Xianjun Qi, Mucong Zhou, "Integrated energy service demand evaluation based on AHP and entropy weight method", ICEEB 2020
6. Yon Sugiarto, Perdinan, Tri Atmaja and Shalsa Nurhasanah, "Evaluation of the Use of Data Reanalysis for Climate Regionalization" IOP Conference Series: Earth and Environmental Science, October 2017

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**