Lab Report

Title: Examining Puerto Rico's Forest Transition with Spatial Statistics

Notice: Dr. Bryan Runck

Author: Mattie Gisselbeck

Date: December 21, 2022

Repository: https://github.com/mgisselbeck/GIS5571

Time Spent: 60 hours

Abstract

Between 1950 and 1990, Puerto Rico saw a dramatic shift in forest cover which reverted 9 to 37 percent of the island's land area back to forest cover (Rudel et al., 2004). This phenomena, Forest Transition Theory, is believed to have happened because of rural-to-urban migration, land abandonment, non-farm labor markets, and other environmental factors. In the early 1950s, Puerto Rico saw a decline in forest cover due large sums of previously forested land being transitioned to agriculture land for crop export. Overtime, natural disasters tied with over-farming led to decreases in soil viability; furthermore, changing markets caused a loss of income and livelihood for many families and individuals (Rudel et al., 2000). This agriculture shift Puerto Rico saw not only cause major rainforest loss, but also contributed to: (a) less fertile soils, (b) less product demand, (c) loss of export market, and (d) erosion and rain-washed communities. The objective of this research was to identify which of the factors thought to be contributing to forest transition are most significant and can predict forest cover given the explanatory variables. While the objective was achieved, there are some drawbacks to the results. Considering the time constraints of this project (i.e., switching topics 4 times), the results are a good starting point to further research on forest transition or reforestation/deforestation in general.

Problem Statement

The objective of this research was to identify which of the factors thought to be contributing to forest transition are most significant and can predict forest cover given the explanatory variables. Explanatory variables analyzed include slope, soil productivity, precipitation, poverty rate, population density, and land protection level. These variables will be visualized through a fishnet feature class with a 1 km and 5 km/per unit scale. The intent for the model's results were to cross-validate Forest Transition's hypotheses and to provide a basis for further spatial statistical research on factors contributing to deforestation on a global scale.

#	**Requirement**	**Defined As**	**(Spatial) Data**	**Attribute Data**	**Dataset**	**Preparation**
1	Puerto Rico Fishnet (1 km)	Fishnet, Puerto Rico (1 km)	.shp (Vector)	Slope, Precipitation, Population Density, Poverty, Tree Loss, Tree Gain	University of St. Thomas	ETL
2	Puerto Rico Fishnet (5 km)	Fishnet, Puerto Rico (5 km)	.shp (Vector)	Slope, Precipitation, Population Density, Poverty, Tree Loss, Tree Gain	University of St. Thomas	ETL
3	Tree Cover Canopy	NLCD 2016 USFS Tree Cover Canopy (CONUS), Puerto Rico	.shp (Vector)	Tree Cover	USFS	ETL

Table 1. Required Data

Input Data

The table below is a collection of data from Department of Geography at the University of St. Thomas. Data was scraped through an ETL in ArcGIS Pro via a Python notebook. All the data described below will be used to identify which of the factors most significant and can predict forest cover. The Digital Elevation Model (DEM) was included for visual reference (see Figure 1).

#	Title	Purpose in Analysis	Link to Source
1	Puerto Rico Fishnet (1 km)	This was used to run Hot Spot Analysis, Exploratory Regression, Generalized Linear Regression (GLR), and Geographically Weighted Regression (GWR).	ArcGIS Online – University of St. Thomas
2	Puerto Rico Fishnet (5 km)	This was used to run Hot Spot Analysis, Exploratory Regression, Generalized Linear Regression (GLR), and Geographically Weighted Regression (GWR).	ArcGIS Online – University of St. Thomas
3	NLCD 2016 USFS Tree Cover Canopy (CONUS)	This was used to complete change detection analysis on tree cover between 2000 and 2016.	USFS

Table 2. Input Data

Methods

Fishnet feature classes (1-km and 5-km scale) were used to assess the influence of explanatory variables on the Puerto Rico's Forest cover. Spatial Autocorrelation (Global Moran's), Hot Spot Analysis, Generalized Linear Regression, and Geographically Weighted Regression tools were used to determine an optimal model and resulting significance for forest cover (see Figures 1 through 3).

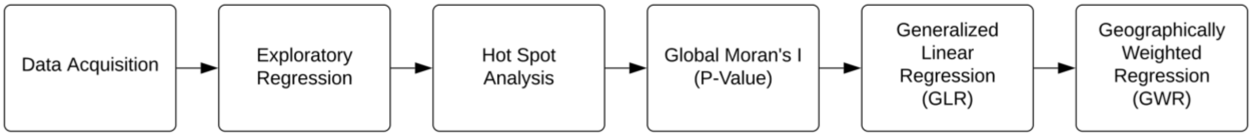

Figure 1. Forest Transition Model (Simplified).

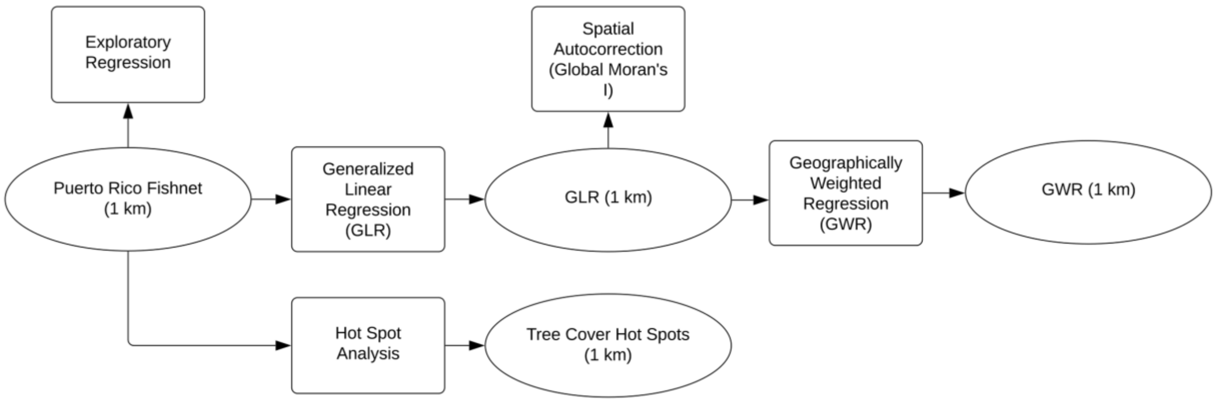

Figure 2. Forest Transition Model - 1 km.

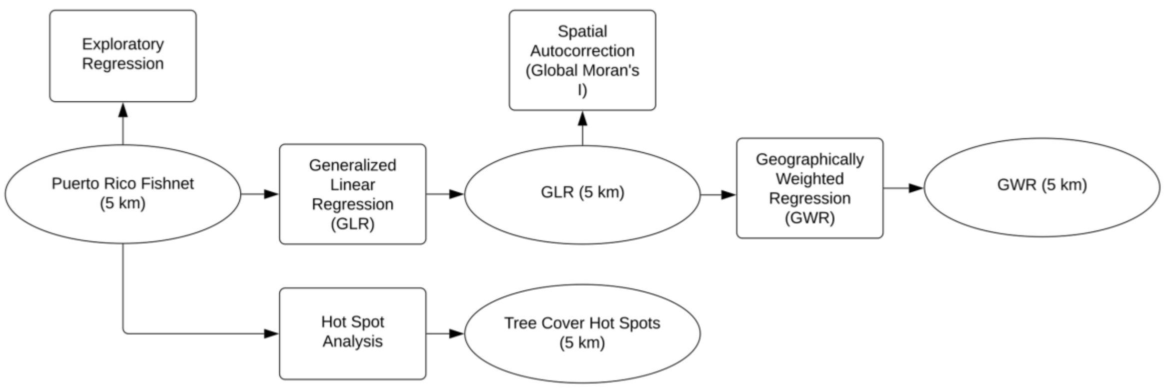

Figure 3. Forest Transition Model - 5 km.

Data Acquisition

To download the data from ArcGIS Online, I customized an Esri script to iteratively download shapefiles matching the item title. The module sequence is as follows: (1) signs into an ArcGIS, (2) download content if item title matches, (3) download file if item type is shapefile, and (4) define download format and specified output file path. For full script, see GitHub. Unless you want to download all of a user's content, this method of extracting data through a notebook is more time consuming and less efficient than directly downloading the content from your ArcGIS Online Portal. It's important to note that all identifiable text was removed for security purposes and was replaced with an asterisk (*).

```
# Download Feature Layers from ArcGIS Online via Search
# Sign into ArcGIS
import arcgis
from arcgis.gis import GIS
gis = GIS(None,'mgisselbeck', '*', verify_cert=False)

# Download File by Search
def downloadUserItems(item.title, downloadFormat):
    try:
              # Search Items by Username
             items = gis.content.search('*:{0}'.format(item.title))
              print(items)
              for item in items:
              if item.type == 'Feature Layer':
             result = item.export('sample {}'.format(item.title),
                 downloadFormat)
             result.download(r'C:\Users\gisse015\Documents\ArcGIS')
```

```
        result.delete()
    except Exception as e:
            print(e)

downloadUserItems(title='*', downloadFormat='Shapefile')
```

Figure 3. Script for Extracting Content from ArcGIS Online Portal.

Explanatory Variables

Explanatory variables analyzed include slope, soil productivity, precipitation, poverty rate, population density, and land protection level. Explanatory variables are shown in Figures 5 through 10.

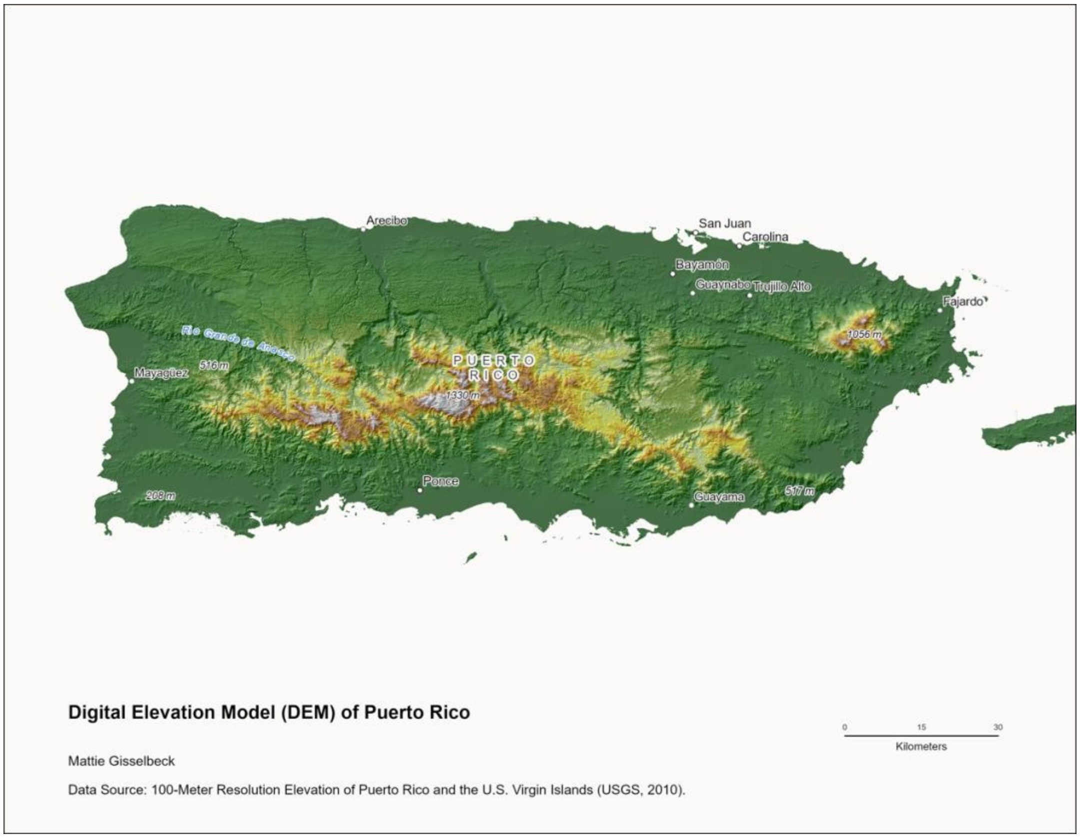

Figure 1. Digital Elevation Model of Puerto Rico (1:750,000).

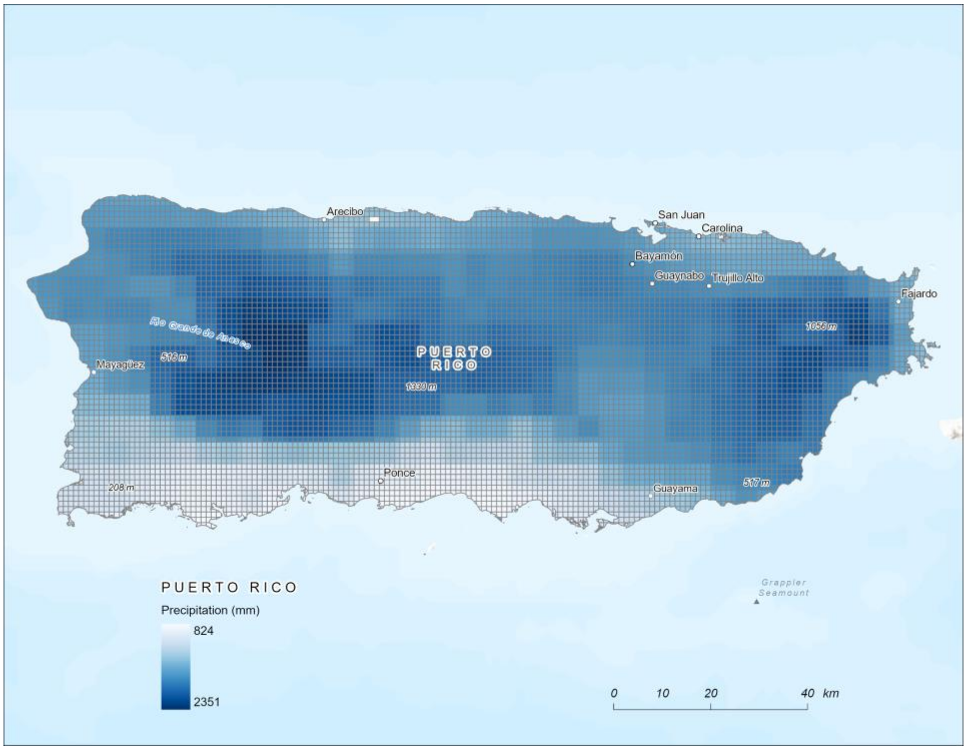

Figure 5. Precipitation of Puerto Rico – 1 km.

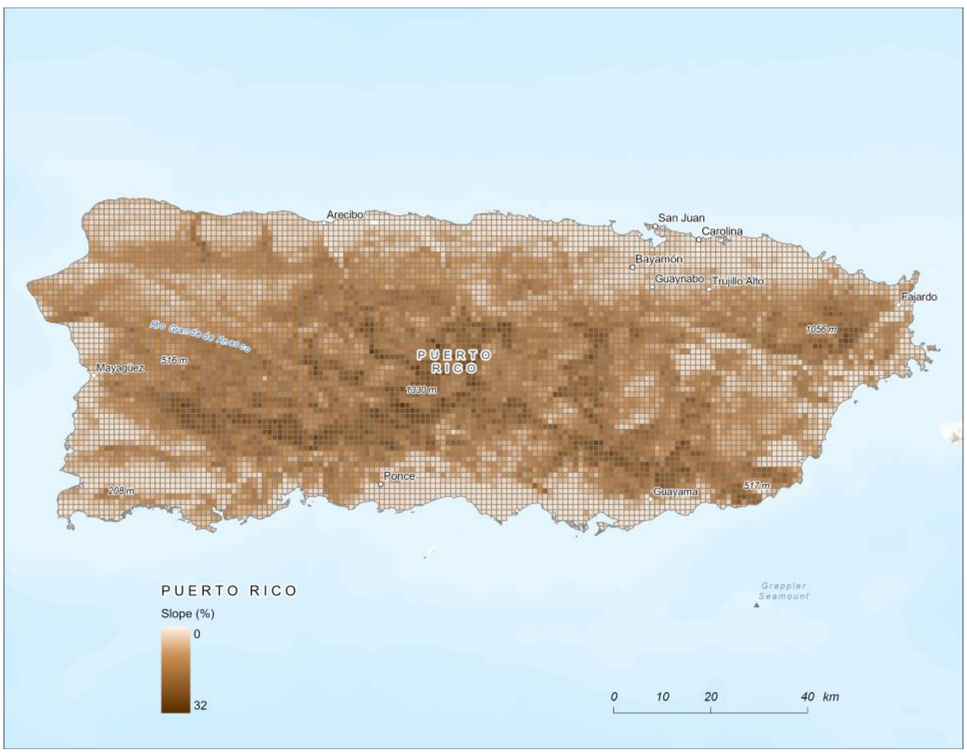

Figure 6. Slope of Puerto Rico – 1 km.

Figure 7. Soil Productivity of Puerto Rico – 1 km.

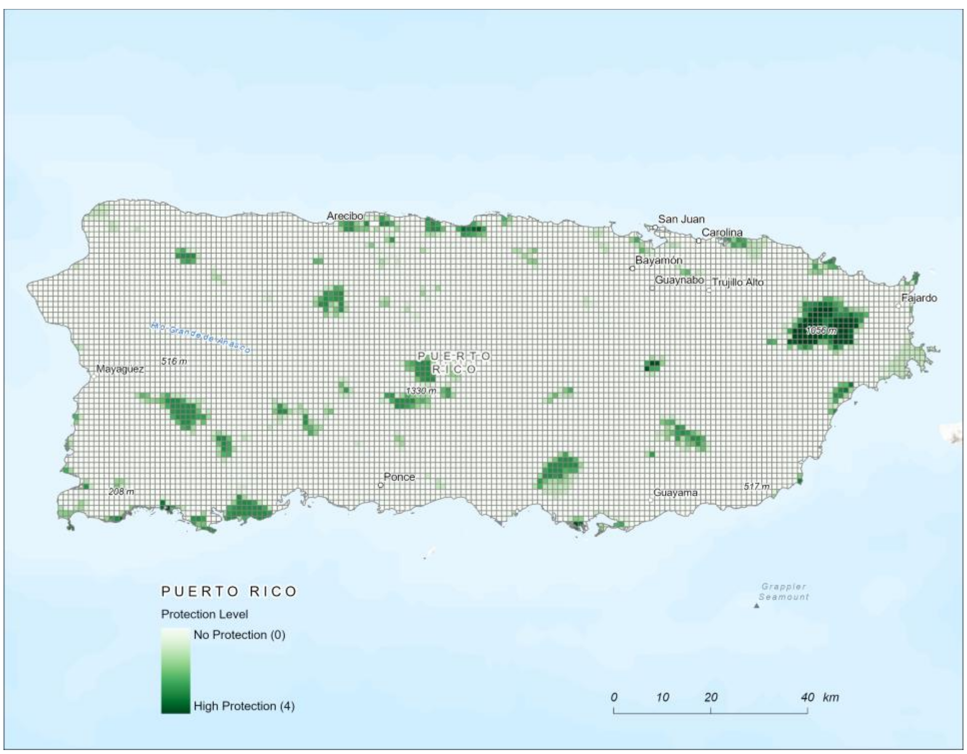

Figure 8. Forest Protection Level in Puerto Rico – 1 km.

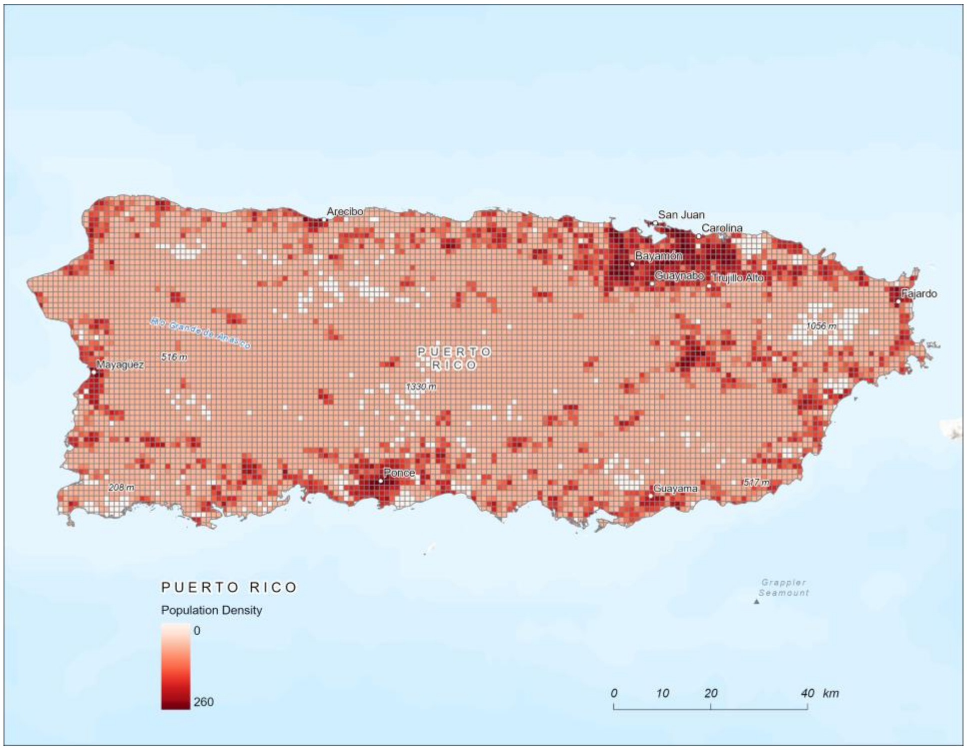

Figure 9. Population Density of Puerto Rico – 1 km.

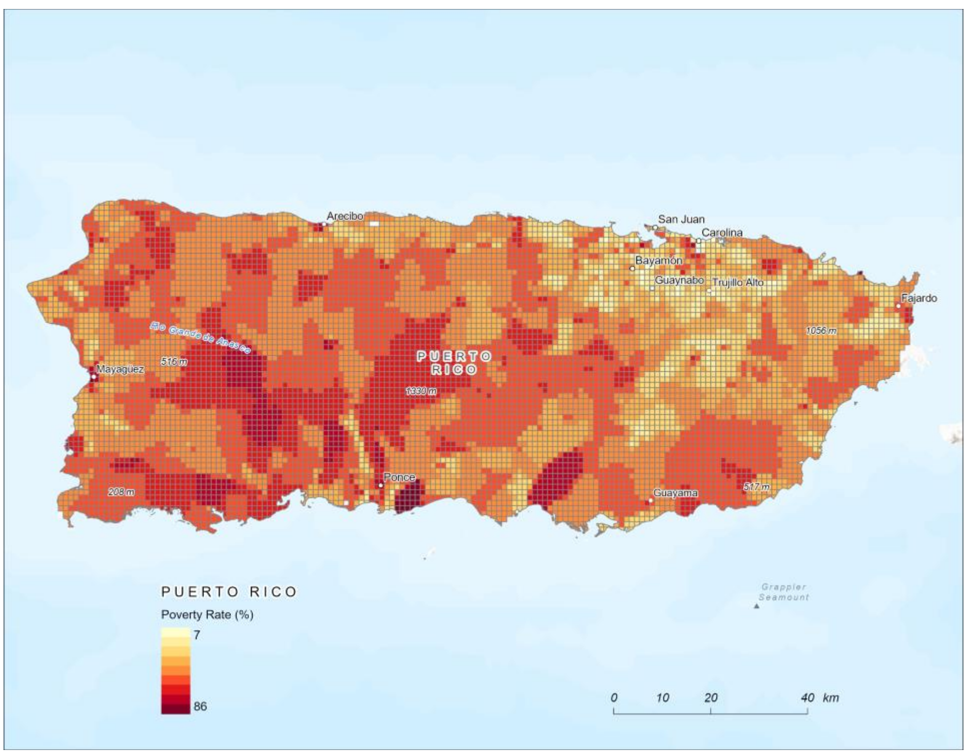

Figure 10. Poverty Rate in Puerto Rico – 1 km.

Results

Results have found high Adjusted R-squared values at both the 1-km and 5-km scales implying up to 90% of the local variance can be explained by the model. The results should be interpreted with caution due to judgements made by the author when determining the explanatory variables of the model. Otherwise, the results suggest a strong influence and predictive potential of the slope, population, soil productivity, precipitation, poverty, and tree loss on the forest cover of Puerto Rico. The results are shown in the figures below (see Figure 2 through Figure 7).

Spatial Autocorrelation (Global Moran's)

Spatial Autocorrelation (Global Moran's) was used to measure the spatial autocorrelation of forest cover by accounting for the nearest neighbor interactions (see Figure 11). In this situation, getting a "random" result is desires because it suggests that the residuals are independent of each other. This "random" result suggests that forest cover at one location in Puerto Rico is dependent on its surrounding forest cover.

Spatial Autocorrelation Report

Global Moran's I Summary

Moran's Index	0.417802
Expected Index	-0.00015
Variance	0.000001
Z-Score	401.585602
P-Value	0.000000

Given the z-score of 401.585602, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Spatial Autocorrelation Report

Global Moran's I Summary

Moran's Index	0.762143
Expected Index	-0.00015
Variance	0.000029
Z-Score	421.752091
P-Value	0.000000

Given the z-score of 421.752091, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Figure 11. Spatial Autocorrelation Report – 1 km (Left); Spatial Autocorrelation Report – 5 km (Right).

Hot Spot Analysis (Getis-Ord Gi)

Hot Spot Analysis (Getis-Ord Gi) was used to identify if local areas of forest cover are significantly above or below the island's mean. The results of the Hot Spot Analysis are shown below in Figure 12. The areas marked as hot spots have a greater forest cover than the island's average whereas the cold spots have less forest cover than average.

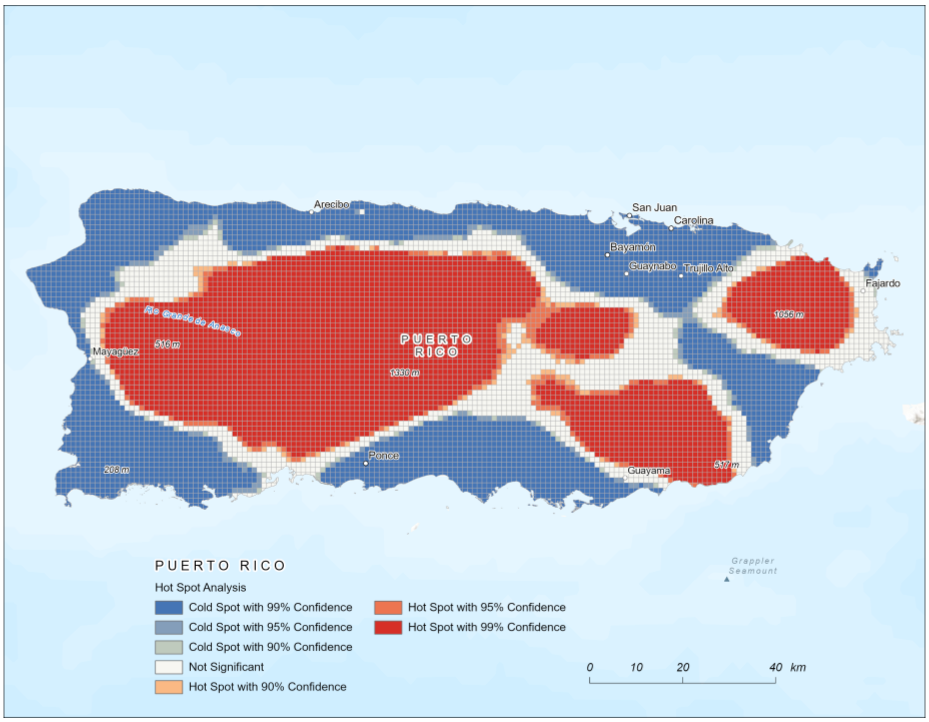

Figure 12. Hot Spot Analysis – 1 km.

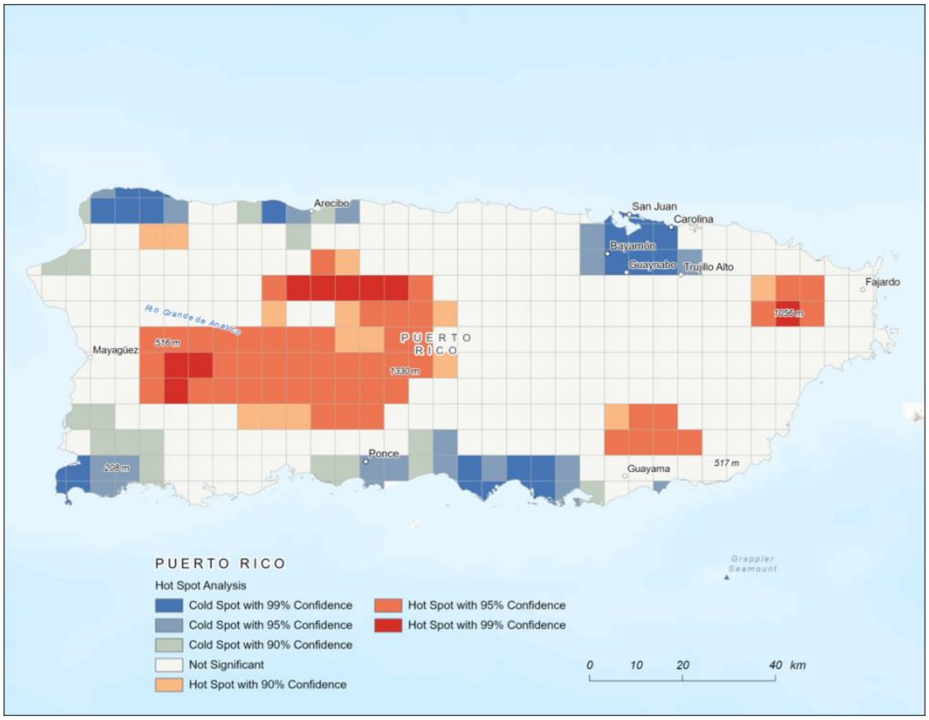

Figure 13. Hot Spot Analysis – 5 km.

Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) was used to determine the local relationships on a 1 km and 5 km scale. The high values outputted for both scales (see Figure 14 and Figure 15) suggest that forest cover in Puerto Rico is strongly spatially patterned by the model factors. It's important to note that the Forest Transition Model (1 km) includes poverty rate whereas the Forest Transition Model (5km) includes protection level instead (see Table 3).

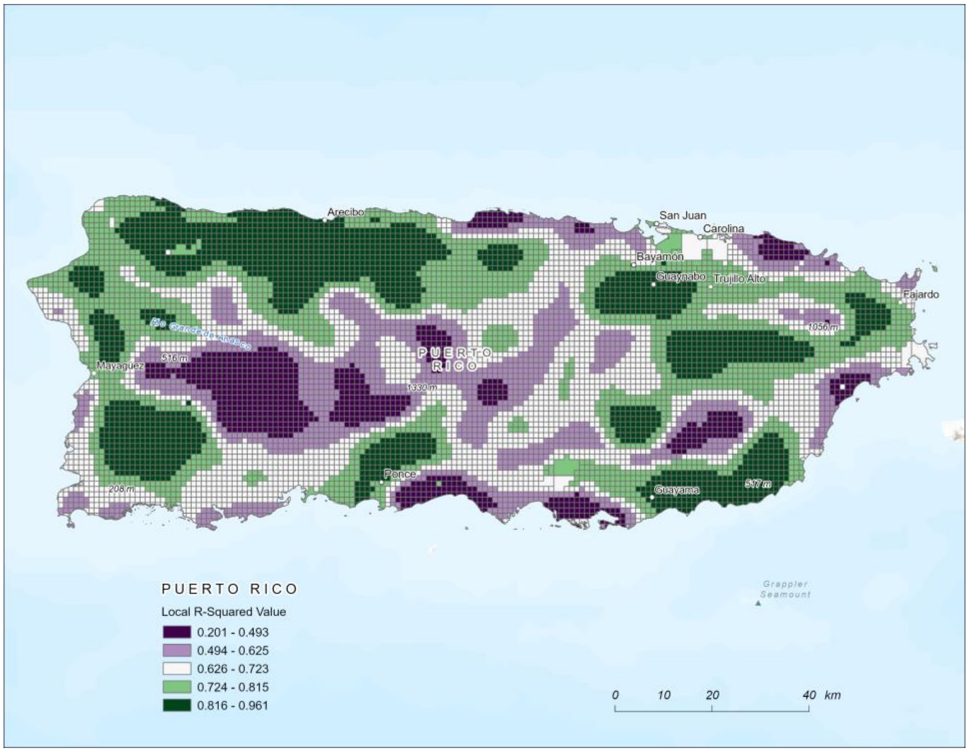

Figure 14. Local R-Squared Values – 1 km.

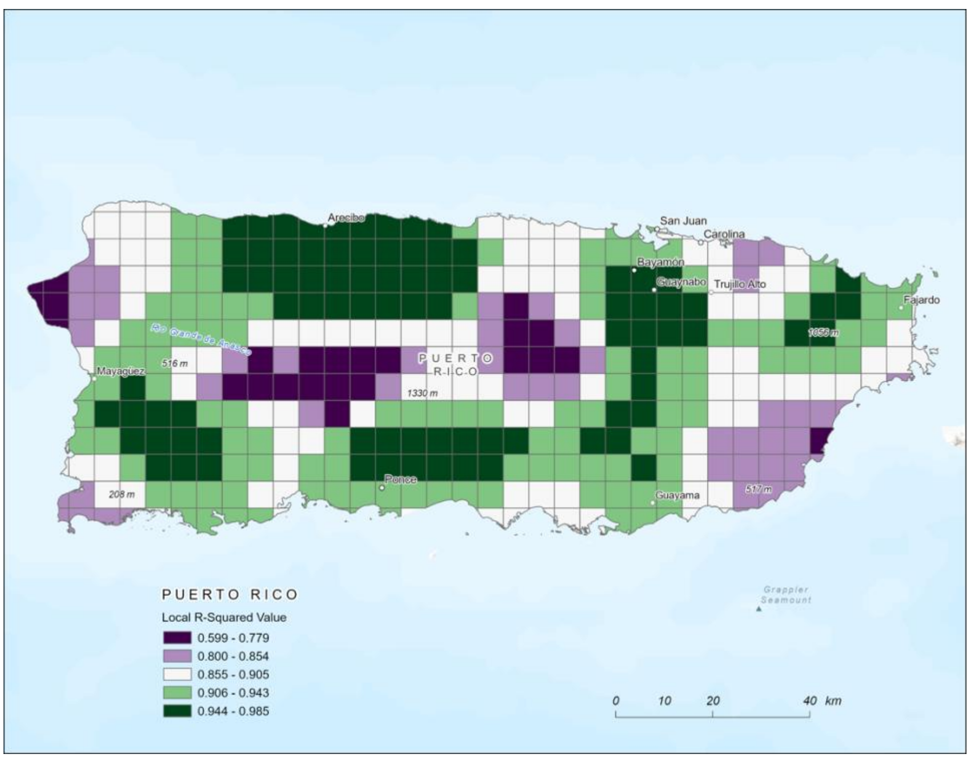

Figure 15. Local R-Squared Values – 5 km.

Scale	Equation
1 km	*Forest Cover = β_0 + β_1 Precipitation+β_2 SoilProductivity+ β_3 PopulationDensity+ β_4 PovertyRate+ β_5 Slope*
5 km	*Forest Cover = β_0 + β_1 Precipitation+β_2 SoilProductivity+ β_3 PopulationDensity+ β_4 ProtectionLevel+ β_5 Slope*

Table 3. Geographically Weighted Regression (GWR) Model Equations.

Predicted Forest Cover

Generalized Linear Regression (GLR) generated a predicted forest cover value for 2000 based on the explanatory variables chosen. The visualizations of the predicted forest cover values (shown in Figure 15) in comparison to the actual forest cover in 2000 (shown in Figure 17), show the efficacy of the model.

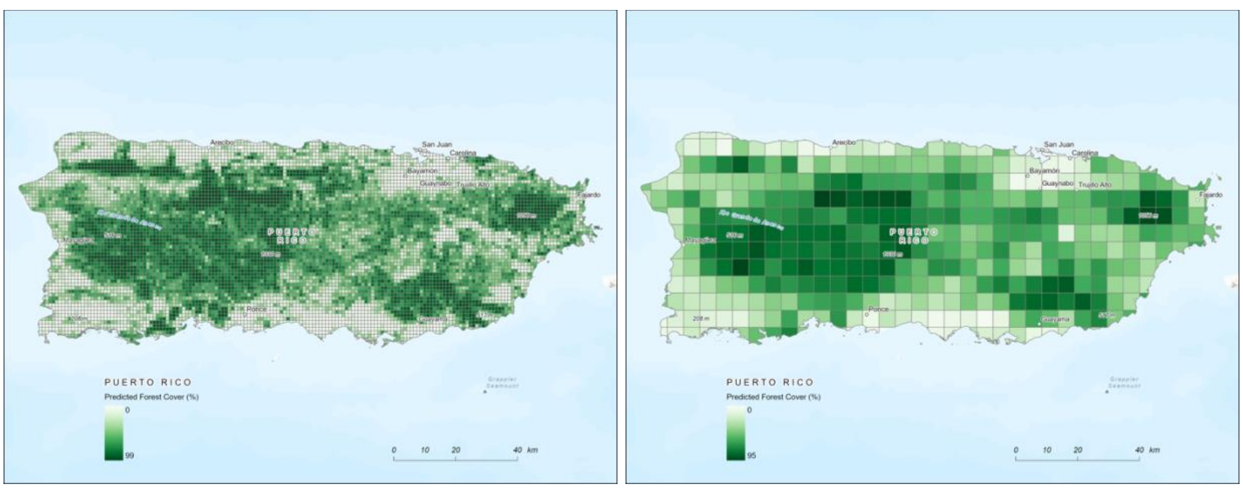

Figure 16. Predicted Forest Cover - 1 km (Left); Predicted Forest Cover - 5 km (Right).

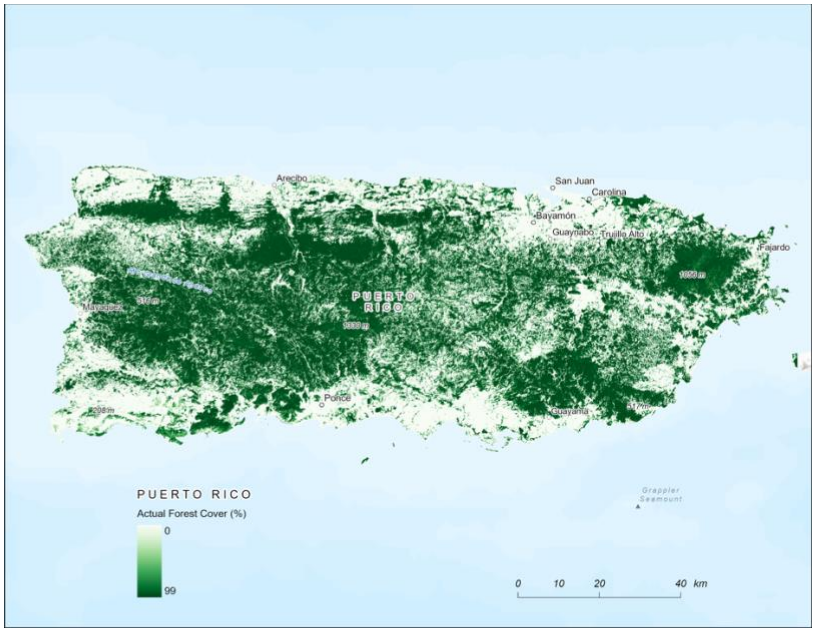

Figure 17. Actual Forest Cover - 1 km.

Predictors for Forest Cover

The strongest predictors for forest cover at a 1 km and 5 km scale were slope and soil productivity. Figures 18 and 19 depict the strong relationship slope and soil productivity have with forest cover at 1 km and 5 km.

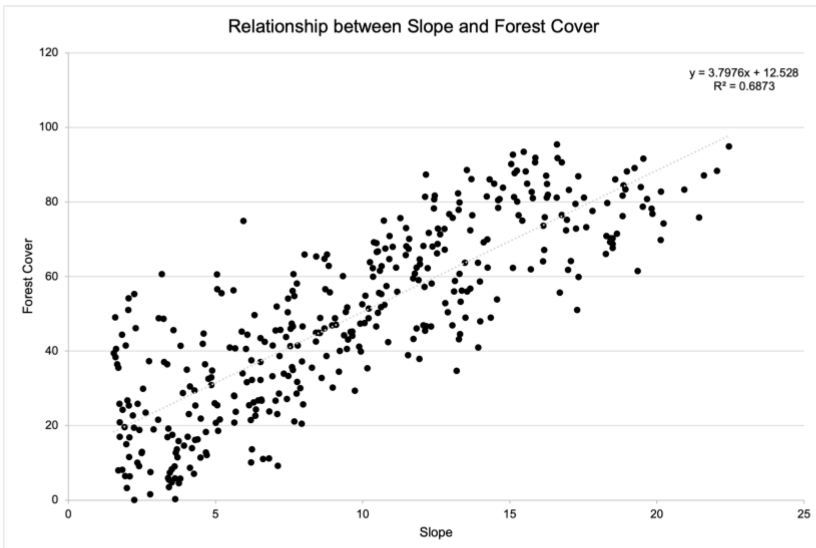

Figure 18. Relationship between Slope and Forest – 5 km.

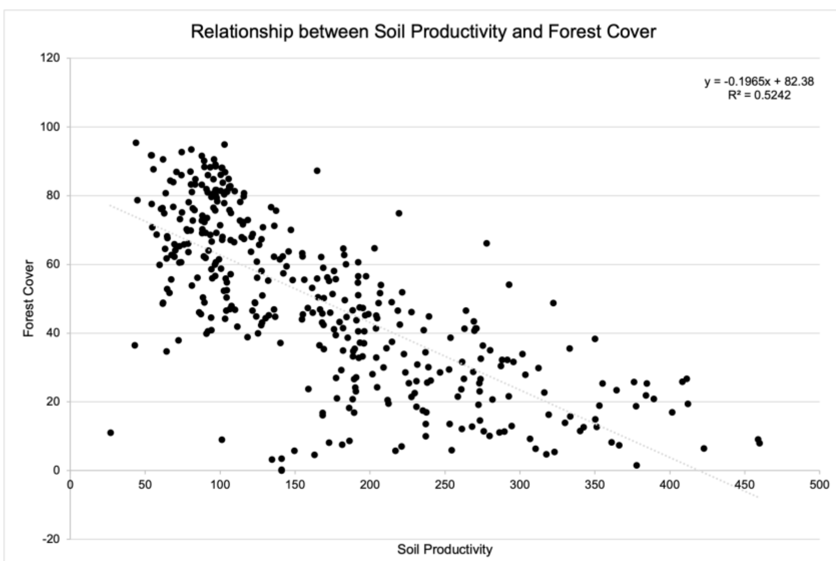

Figure 19. Relationship between Soil Productivity and Forest Cover – 5 km.

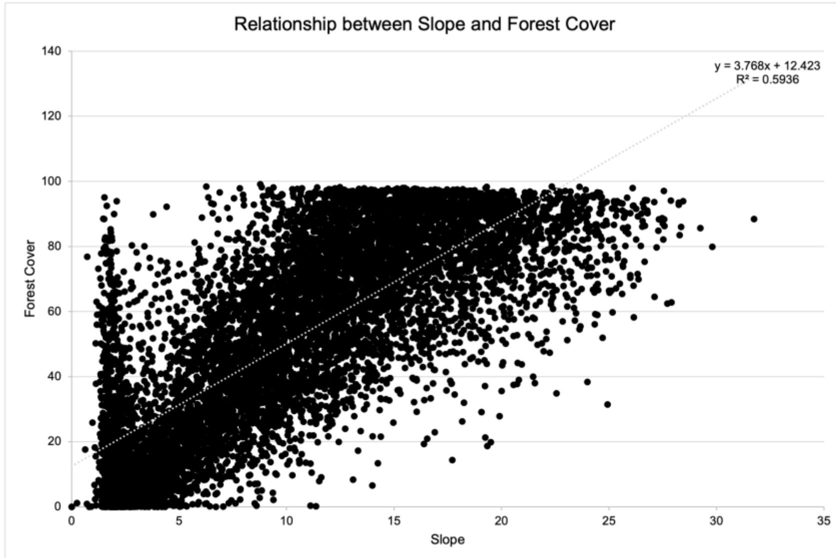

Figure 20. Relationship between Slope and Forest Cover – 1 km.

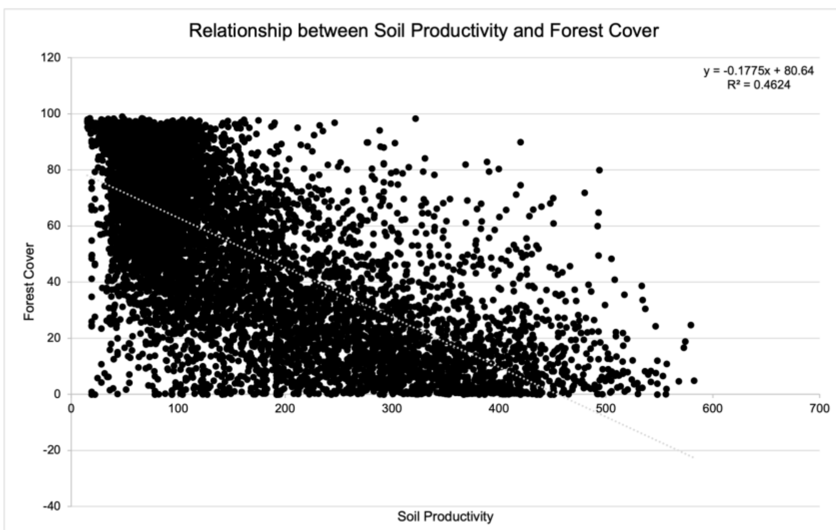

Figure 21. Relationship between Soil Productivity and Forest Cover – 1 km.

Results Verification

The results were verified qualitatively with a linear regression model (see Figure 22). The results were verified by plotting the model's predicted values for forest cover in 2000 against the actual values of forest cover. The linear regressions below show high r-squared values for both models (*1 km scale = 0.8548, 5 km = 0.9465*).

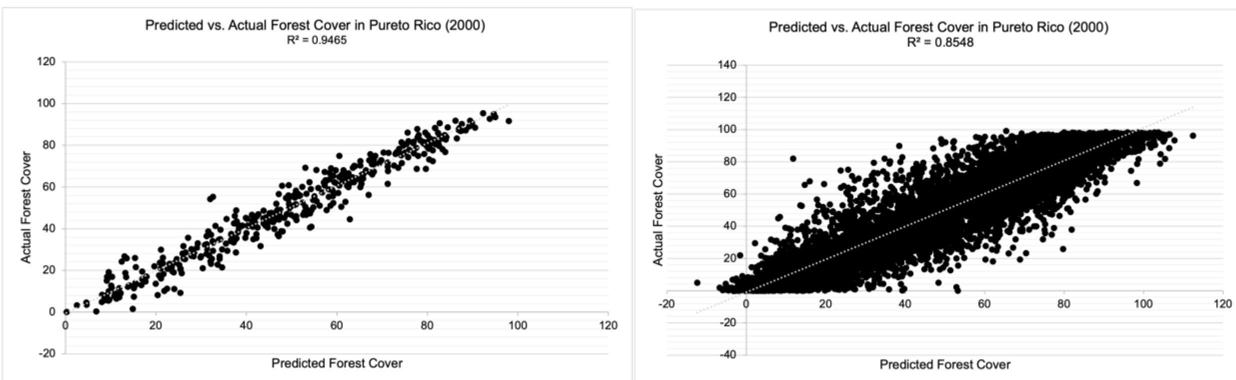

Figure 22. Predicted vs. Actual Forest Cover – 5 km (Left); Predicted vs. Actual Forest Cover – 1 km (Right).

.

Discussion and Conclusion

While the objective was achieved, there are some drawbacks to the results. First, the available data was not comprehensive enough to run statistical model validation. To run the model, some of the data had to be removed due to its invalid values, which led to gaps in the output map. Second, the model includes human dominated variables such as poverty and population. This shows no certainty in their significance across different study areas.

Considering the time constraints of this project (i.e., switching topics 4 times), the results are a good starting point to further research on forest transition or reforestation/deforestation in general. For future research, I plan to use a machine learning method like RandomForest to compare results and find an optimal model.

Throughout this project, I was able to gain a wealth of knowledge on spatial statistic application with ArcPy and open-source packages.

References

Rudel, T. K., Perez-Lugo, M., & Zichal, H. (2000). When Fields Revert to Forest: Development and Spontaneous Reforestation in Post-War Puerto Rico. *The Professional Geographer*, *52*(3), 386–397. https://doi.org/10.1111/0033-0124.00233

Helmer, E., Ruzycki, T., Wilson, B., Sherrill, K., Lefsky, M., Marcano-Vega, H., Brandeis, T., Erickson, H., & Ruefenacht, B. (2018). Tropical Deforestation and Recolonization by Exotic and Native Trees: Spatial Patterns of Tropical Forest Biomass, Functional Groups, and Species Counts and Links to Stand Age, Geoclimate, and Sustainability Goals. *Remote Sensing*, *10*(11), 1724. https://doi.org/10.3390/rs10111724

Esri (2022). How To: Download feature service items from ArcGIS Online using ArcGIS API for Python. https://support.esri.com/en/technical-article/000018909

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	**23**
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	**95**