Andrew Ferguson
SSCI 575 - Project 4
Dr. Qi
4 December 2022

Project 4

Spatial analysis of terror events in Myanmar

Introduction

Myanmar (Burma) is a southeastern Asian country that has been experiencing an ongoing humanitarian crisis, where Rohingya refugees are being systematically displaced and killed in an attempt at ethnically cleansing the nation. The systematic displacement often includes the burning of villages and the rape and murder of Rohingya people (Beyrer & Kamarulzaman, 2017).

This ongoing humanitarian issue lacks public awareness of other humanitarian crises. In an effort to learn more about the drivers of terrorism in the Myanmar, this study addresses the hypothesis that there is a significant spatial correlation between population density and the number of terror attacks in Myanmar. It also aims to understand if there are clusters or regions where the population-terror relationship are unique.

Understanding the spatial distribution of terror events in the country can help governmental (i.e. the US State Departement or the EU's EEAS board) and nongovernmental (i.e. Amnesty International) respond to the growing humanitarian crisis by effectively allocating resources where terrorist events are affecting the greatest number of people.

Study Area

This project focuses on the country of Myanmar, which is also known as Burma. It is in southeast Asia, and it is situated between India, Pakistan, China, Thailand, and Laos. Figure 1 shows the outline of Myanmar with each of its administrative districts colored by population density. The location of all terror events between 1 January 2015 and 26 November 2022 are plotted as well.

Data

This project utilizes data from 3 sources. First, the terror location data was queired from the Armed Conflict Location & Event Data (ACLED) website. ACLED is a non-profit NGO that analyzes political violence throughout the entire world. They host a free data export tool that allows SQL-Like queries from their database. Exporting data requires an account and access key, but the data itself is freely accessible to the public. The raw data from the website is provided in .csv format. The latitude and longitude columns of the csv file were used to project the location of terror events onto Figure 1 (Raleigh, et al.).

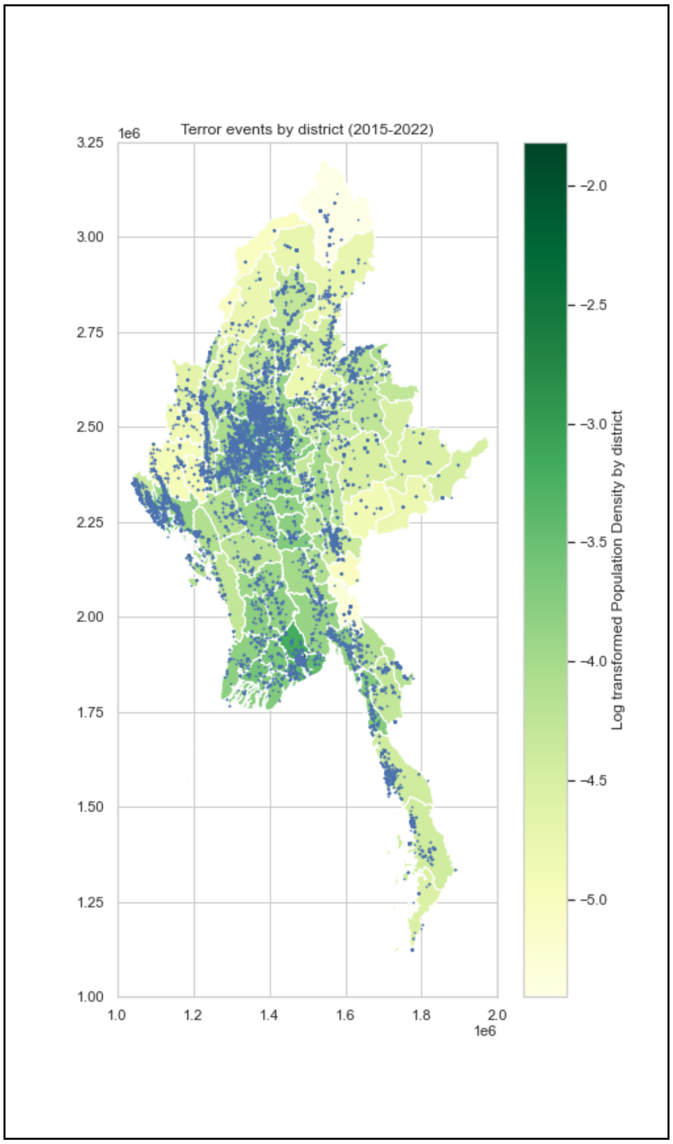

Figure 1: Study Area (Myanmar) with district population density and location of terror attacks between Jan 1, 2015 and Nov 26, 2022

The district boundary shapefile was downloaded directly from the Myanmar Information Management Unit Geonode website(Mimu-Gis). Population data was downloaded in .csv format from the Myanmar Data Grid on the Humanitarian Data Exchange (HDX) website (Myanmar). The HDX download UI provided a field mapping that could be used to join population data onto the shapefile (district boundary shapefile field “DT_PCODE” = population csv field “ADM2_PCODE”). These fields are aligned in a 1-to-1 manner making it simple to join the total population field from the csv into the shapefile/GeoDataFrame on the key field.

In order to include the number of terror events in each administrative district, first both the terror GeoDataFrame (point) and the district boundary GeoDataFrame (polygon) were projected into

EPSG 24305 (Klokan). Then, the GeoPandas Spatial Join tool was used to calculate the count of terror events in within each district. Finally, population density was calculated from total district population, and log-transformed to achieve a more normal distribution.

The population density of each administrative district and the count of terror events between 2015-present is satisfactory for addressing the hypothesis of whether population density is correlated with terror events in Myanmar.

The final step in preprocessing was to prepare the data to be input into a non-spatially constrained clustering algorithm. Unsupervised clustering requires data to be normalized. This project used min-max scaling to transform the terror count column and the log-transformed population density into normalized values between 0 and 1.

Methodology

The entire methodology in this project utilized open-source software. Development was written in Python using open-source libraries including os, pandas, geopandas, numpy, matplotlib, seaborn, scikit learn, shapely, esda, and libpysal. JupyterLab was used as the development platform. Attached to this document submission is the .ipynb that follows the entire methodology (from data cleaning through modeling and analysis), and the required data.

Figure 2 shows a workflow diagram of the spatial analysis.

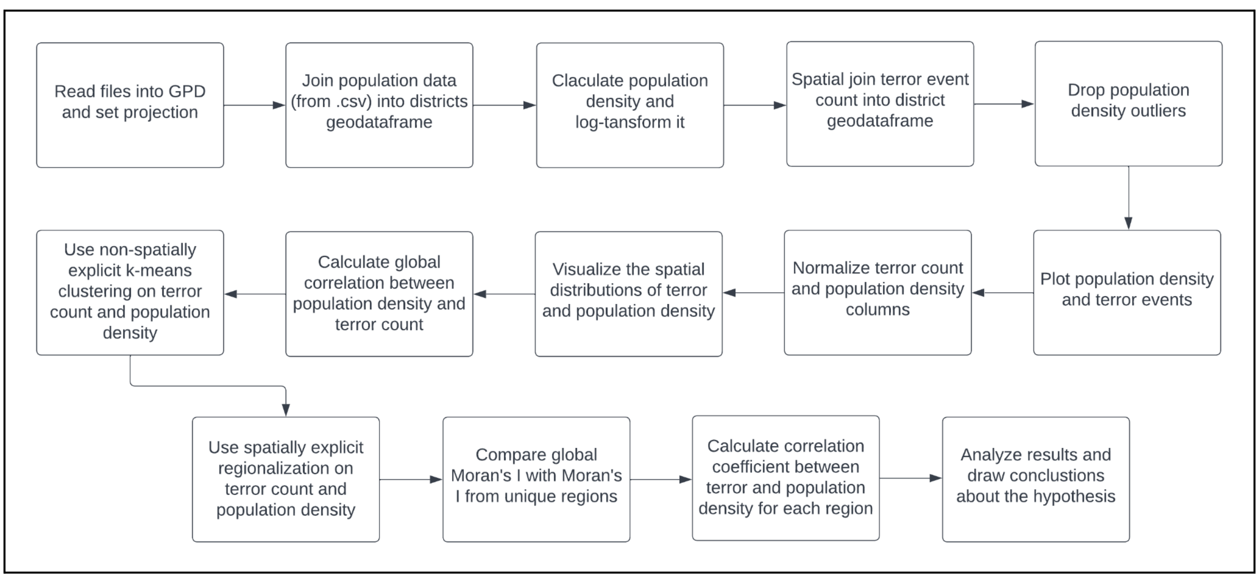

Figure 2: Workflow diagram for the project

Data Preparation and Exploration:

The first step in this project was to read all of the input data into a geopandas (gpd) GeoDataFrame (GDF) and set the projection to EPSG 24305. Figure 3 shows the distribution of population density by district (left), and the log-transformed population density by district (right). The data achieves a more normal distribution after the log transformation so it will be used in the upcoming analysis.

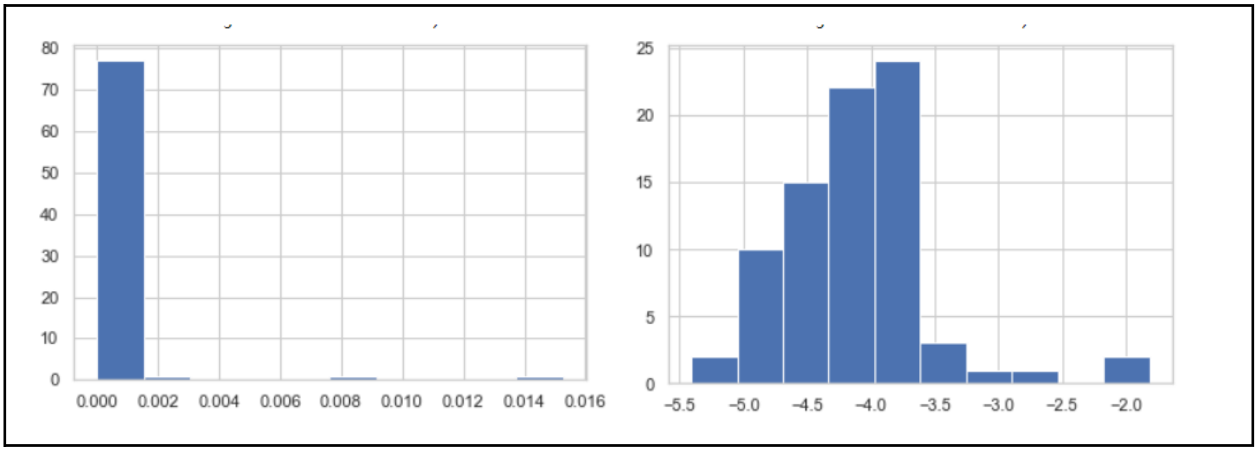

Figure 3: Left: Distribution of population density by district
Right: Log-transformed population density by district

Before transformation, population density has a few outliers. In this project, I chose to treat the outliers by dropping them from the dataset. The logic behind dropping the outliers is that I initially tried to keep them in the analysis, but the subsequent regionalization resulted in a few regions that only consisted of 1 district (the outliers). Dropping outliers resulted in a more logical and explainable result. Figure 4 shows the distribution of population density and log transformed population density after dropping outliers.

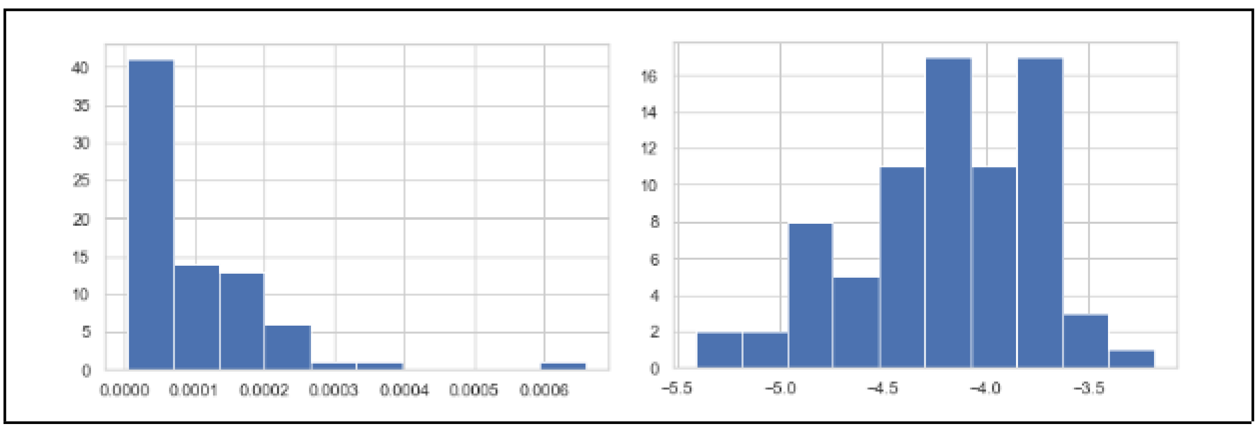

Figure 4: Left: Distribution of population density by district after outlier treatment
Right: Log-transformed population density by district after outlier treatment

The next step in the analysis was to use a spatial join to calculate the number of terror events occurred in each district. An inner join was used with predicate = 'intersects' in order to calculate the count of terror events in each district. One district had no terror events; the spatial join tool output a null value in that district. The pandas fillna() tool was used to impute the null with a value of 0.

In order to prepare the data for use in a non-spatially explicit clustering model, the next step was to normalize the data using Min-Max Scaling.

Before running the data through the non-spatially explicit clustering model was to visualize the spatial distribution of the number of terror events and the log-transformed population density next to each other. Visual inspection of Figure 5 suggests that areas with higher population density are not more prone to terror attacks. Clustering and Regionalization will be used to dig deeper and learn more about the relationship between the 2 variables.

The final data exploration step was to calculate the global correlation between number of terror events and population density. The global correlation coefficient between the 2 datasets is 0.1688.

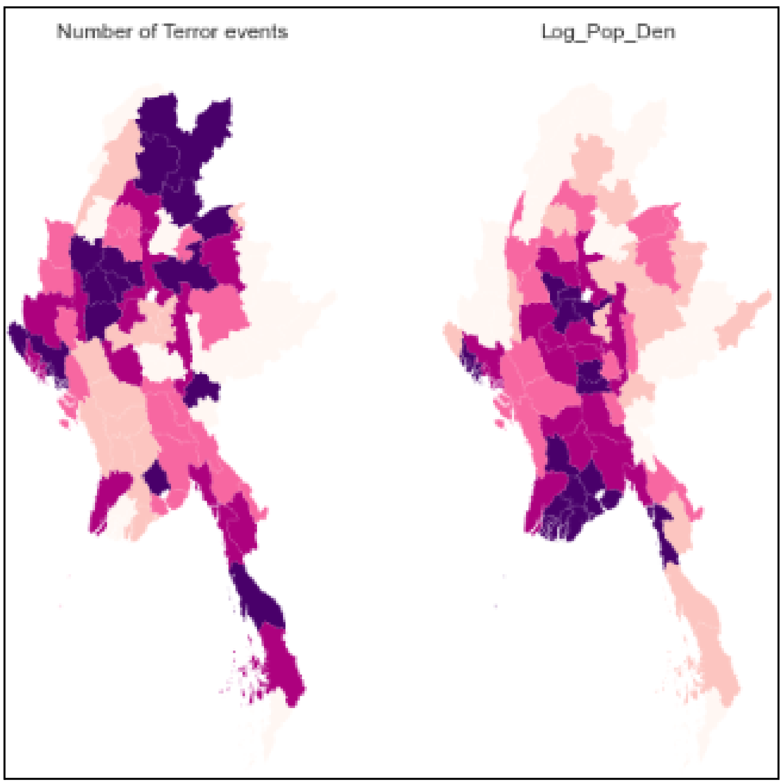

Figure 5: Spatial distribution of terror events and population density in Myanmar

Modeling

In order to determine the optimal number of clusters for the k-means non-spatially explicit clustering, the within-cluster sum of squares (WCSS) was calculated for all k in the range 1-30. Figure 6 shows the WCSS plot and a vertical line at k=6, which is where the plot began to level off.

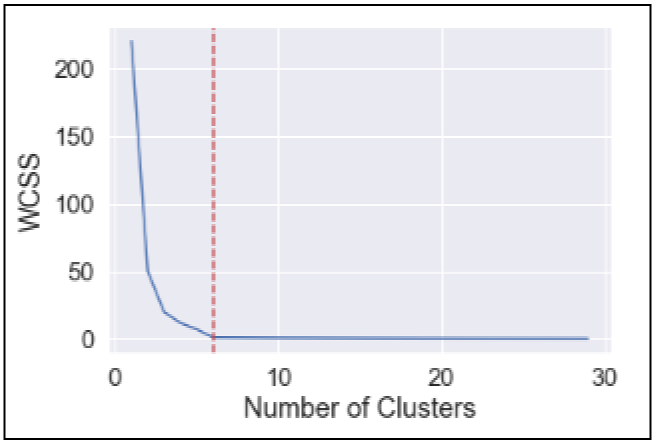

Figure 6: WCSS vs Number of Clusters

Figure 7 shows the results of the non-spatially explicit k-means clustering model fit to the data with k=6. The spatial distribution of clusters appears to be rather random while lacking the contiguity constraint.

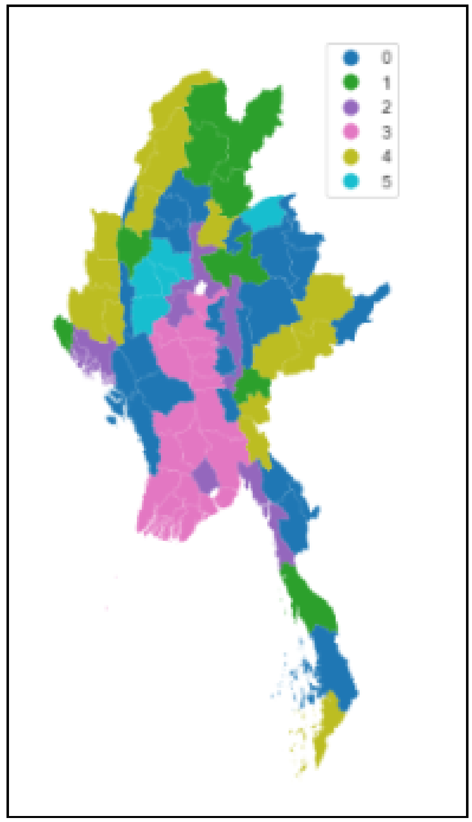

Figure 7: Results of k-means clustering algorithm with k=6

Next, a seaborn facet plot was developed to visualize the distribution that each cluster had along the input variables. Figure 8 shows the facet plot. The number of terror events in most clusters appears to reach its maximum at unique values. The population density is more similar across all clusters than the number of terror events. There doesn't seem to be a particularly close alignment between the distribution of population density and the number of terror events in each k-means cluster.

The next step in this analysis was to compare the results of the k-means clustering model to the results of a spatially-explicit regionalization model. This project chose to use Queens case spatial contiguity. Given the non-uniform district shape in Myanmar, Queen's case and Rook's case contiguity will result in very similar cases. Queen's case was ultimately chosen so that, in the event that 2 administrative districts only share a vertex and not an edge, they will still be considered neighbors.

Using Queen's case spatial contiguity, spatial weights were developed and global Moran's I values were calculated for each of the variables (number of terror events and population density). Then, scikit learn's Agglomerative Clustering tool was used to calculate 6 spatially contiguous regions in the dataset. Figure 9 shows the results from the regionalization (Agglomerative Clustering).

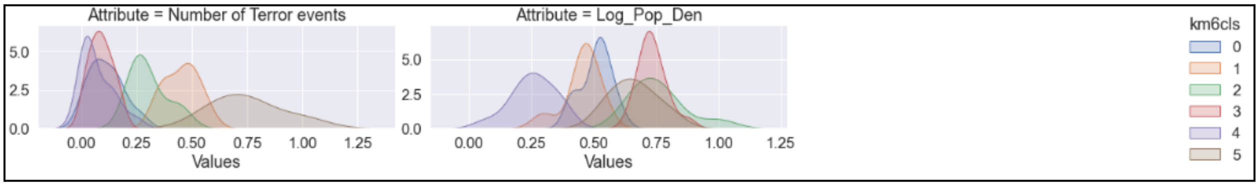

Figure 8: Facet plot

To understand whether there is stronger spatial autocorrelation within individual regions than the global results, each region was extracted into its own geodataframe. Then, the correlation coefficient between number of terror events and the population was calculated for the 6 regions.

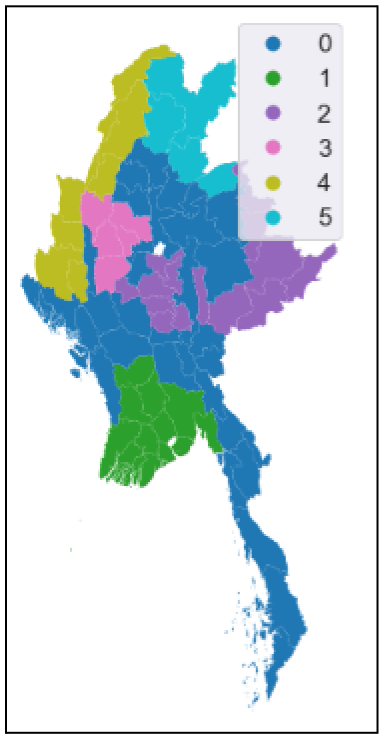

Figure 9: Map of spatially contiguous regions

Results

Figure 10 shows the value of Moran's I and the P-value (statistical significance) for the global dataset, and separated by region. Relative to global values, Regions 2 and 4 show stronger spatial autocorrelation on number of terror events. These are regions where terrorism is clustered (with p-values < 0.1, showing significance). Region 2 also shows a high Moran's I value for population density. This is the only region where spatial autocorrelation of terror and population are more clustered than global results. Regions 3 and 5 show have negative values for spatial autocorrelation among number of terror events suggesting dispersion of terror events. The p-value in region 3 is 0.053, which is significant at a p-value of 0.1. The p-value for region

5, however, is too high to determine significance. The Moran’s I values for population density and number of terror events in regions 0 and 1 are closer to 0 than global results. The spatial distribution of these phenomena are more random in regions 0 and 1 than the global trends.

Global

Variable	Moran's I	P-value
Number of Terror events	0.347220	0.002
Log_Pop_Den	0.526152	0.001

Region 0

Variable	Moran's I	P-value
Number of Terror events	0.126557	0.159
Log_Pop_Den	0.293257	0.011

Region 1

Variable	Moran's I	P-value
Number of Terror events	0.030366	0.207
Log_Pop_Den	0.066947	0.176

Region 2

Variable	Moran's I	P-value
Number of Terror events	0.475496	0.015
Log_Pop_Den	0.650342	0.001

Region 3

Variable	Moran's I	P-value
Number of Terror events	-0.491602	0.053
Log_Pop_Den	-0.004950	0.152

Region 4

Variable	Moran's I	P-value
Number of Terror events	0.639405	0.003
Log_Pop_Den	0.293474	0.046

Region 5

Variable	Moran's I	P-value
Number of Terror events	-0.423941	0.407
Log_Pop_Den	-0.240163	0.398

Figure 10: Moran’s I for Number of Terror Events and Population Density globally and by region

The global correlation coefficient between number of terror events and population density by district is 0.1688. Figure 11 shows the regional correlation coefficient. After regionalization, the average correlation coefficient is 0.2472.

	region	correlation coefficient
0	0	0.270824
1	1	0.584739
2	2	0.557586
3	3	-0.200672
4	4	-0.242876
5	5	0.513417

Figure 11: Regional correlation coefficient between population density and number of terror events

Regions 0, 1, 2, and 5 have a stronger correlation coefficient between number of terror events and population density than the global value. In these regions, there is a positive relationship between number of terror events and population density.

Regions 3 & 4 have a negative relationship correlation coefficient between number of terror events and population density. In these regions, increasing population density decreases the prevalence of terror events.

Conclusions

Globally, there is a weaker correlation between terror and population density than there is from a regional look. Given the results of this study, we fail to reject the null hypothesis that there population density has no statistically significant effect on the spatial distribution of terror events in Myanmar between 2015 and now. However, we the results do suggest that certain regions (0, 1, 2, and 5) maintain a positive relationship between the input variables.

One major shortcoming of this project was the lack of demographic data available in Myanmar. Specifically, I predict that including income and/or education data by administrative district would have improved the model performance. Unfortunately, in unstable/underdeveloped countries like Myanmar, complete census data is not always available.

The raw ACLED data contains date/time and event information for each terror event. One interesting follow up study would be to conduct a spatial time-series analysis to see if the terror events move in a certain definable direction over time. This could help organizations prepare vulnerable communities for attacks and potential violence in the ongoing ethnic cleansing event.

Ultimately, Spatially-Constrained Regionalization was an appropriate approach to addressing the hypothesis of whether there exists a spatial relationship between population density and number of terror events because it allowed me to compare the global relationship between the variables with the regional relationship.

On a personal note, this project gave me a great opportunity to expand out of my “comfort zone” of spatial analysis - ArcGIS Pro. This is the first time I’ve ever completed a spatial analysis project using completely open source software. I learned a lot about new libraries and modeling techniques using Python, rather than relying on proprietary software.

References

Beyrer, C., & Kamarulzaman, A. (2017). Ethnic cleansing in Myanmar: The rohingya crisis and human rights. *The Lancet*, 390(10102), 1570–1573. https://doi.org/10.1016/s0140-6736(17)32519-9

Klokan Technologies GmbH. (n.d.). *Projected coordinate systems for "Myanmar"*. EPSG.io: Coordinate Systems Worldwide. Retrieved November 20, 2022, from https://epsg.io/?q=myanmar+kind%3APROJCRS

Mimu-Gis. (2021, May 31). *Myanmar district boundaries mimu v9.3*. MIMU Geonode. Retrieved November 20, 2022, from https://geonode.themimu.info/layers/geonode%3Ammr_polbnda_adm2_250k_mimu

Myanmar - Subnational Population Statistics. Humanitarian Data Exchange. (2022). Retrieved November 20, 2022, from https://data.humdata.org/dataset/cod-ps-mmr

Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. (2010). “Introducing ACLED-Armed Conflict Location and Event Data.” Journal of Peace Research 47(5) 651660.