A Geospatial Analysis of Conflict in Afghanistan

Dalton Shaver

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

Faculty Advisor: Professor Steve Morris

ABSTRACT

The purpose of this analysis was to determine the spatial distribution of high concentrations of conflict in Afghanistan following the 2001 invasion by NATO coalition forces. The United States government and military, as well as NATO allies and international organizations, rely on accurate data analysis and multisource intelligence to make important decisions regarding policy-making and defense strategy. In my analysis, I use spatial autocorrelation and density clustering methods to find patterns of spatial distribution in the data. I also calculated the distance of each individual event and its proximity to the nearest provincial capital and the nearest national highway. Exploratory Spatial Data Analysis techniques show that there is a clear positive autocorrelation present within the spatial distribution of conflict events. A Density-Based Spatial Clustering of Applications with Noise (DBSCAN) model produces very similar results to the "hotspots" the Local Moran's *I* test produced. Furthermore, my analysis shows that conflict was more likely to occur closer to provincial capitals and major infrastructure rather in rural regions.

INTRODUCTION

Operation Enduring Freedom began on October 7, 2001 following the September 11, 2001 attacks. The United States and other NATO forces staged a swift invasion and takeover of the Taliban-ruled government in Afghanistan. The United States created, funded, and trained the Afghan National Army in an attempt to form a security force that would stabilize the country. Both the Taliban and al-Qaeda groups increasingly conducted guerrilla warfare up until NATO's formal withdrawal in 2014. Beginning in early 2015, the Taliban began staging large offensives against the Government of Afghanistan to seize major cities. The intensity of the war had significantly increased between all warring parties until the reinstatement of Taliban rule on August 15, 2021.

METHODS

The main dataset is Uppsala University's Georeferenced Event Dataset which contains information on each conflict event. The key variables used are Province, District, Year, Latitude, Longitude, and estimated Casualties. After cleaning, the dataset featured 31,658 observations consisting of events between 2001 and 2019. The two additional sources include a list of provincial capitals obtained from The World Bank Afghanistan Dashboard and the Afghanistan Road Network shapefile containing all marked roads, retrieved from World Food Program's GeoNode. The shapefiles for both regions and districts were retrieved from The University of Texas Online Library.

Histograms were used to examine the distribution of the calculated distance between each conflict event and the nearest provincial capital, as well as the distance between each event and the nearest highway, much of which forms Afghanistan's Ring Road.

Multiple Bar Chart was used to examine the relationship between the relative frequencies of events and casualties by province.

Single Bar Chart was used to assess the distribution of events for years 2001-2019.

Simple Linear Regression Model was used to determine the correlation between the number of casualties and conflict events in each province.

DBSCAN Model was used to detect clusters of conflict events based on the density of latitude/longitude points.

Moran Plot displays the correlation between the number of events per district and the number of spatially lagged events per district.

Moran's *I* Tests were used to determine both the global and local spatial autocorrelation of conflict events in Afghanistan.

Choropleth Maps were used to visualize quantiles, related quadrants, and statistical significance derived from Moran's *I* Tests.

Figure 1: Conflict Events Quantiles

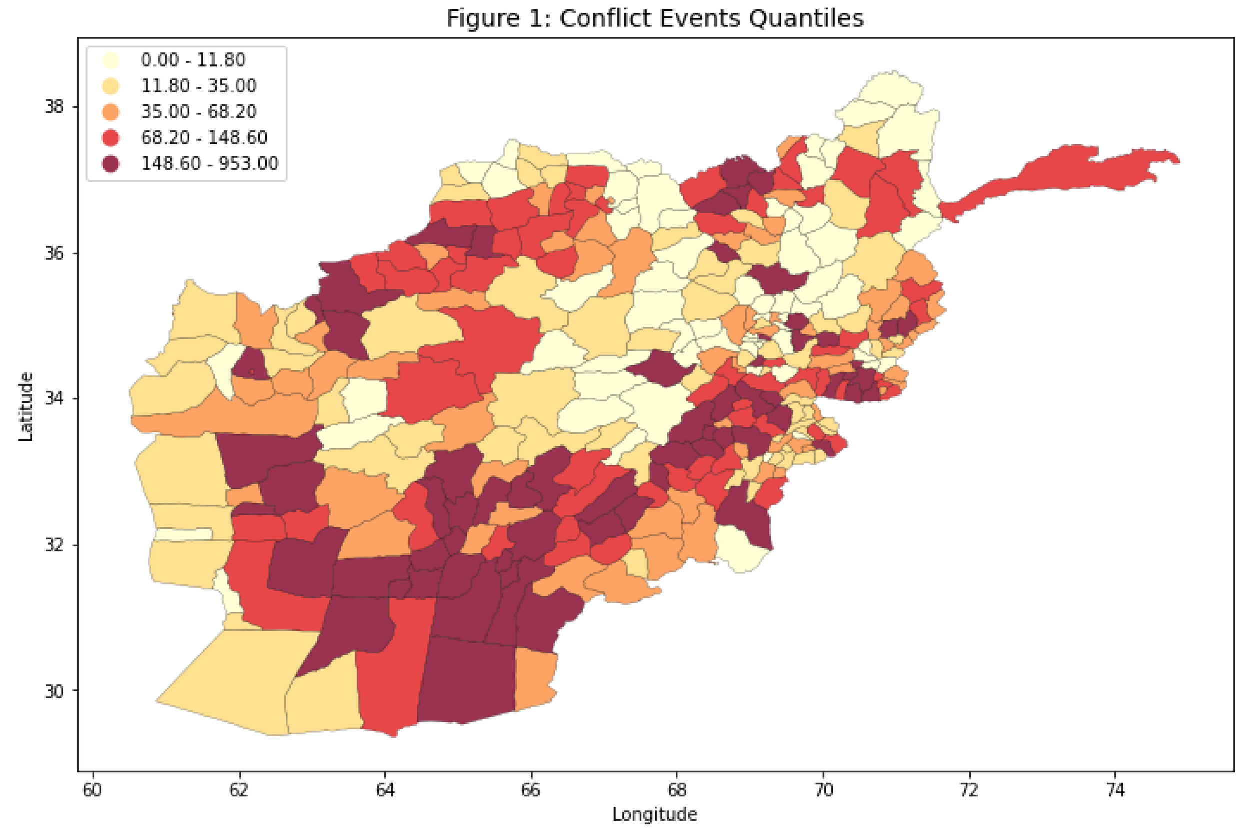


Figure 2: Spatially Lagged Conflict Events

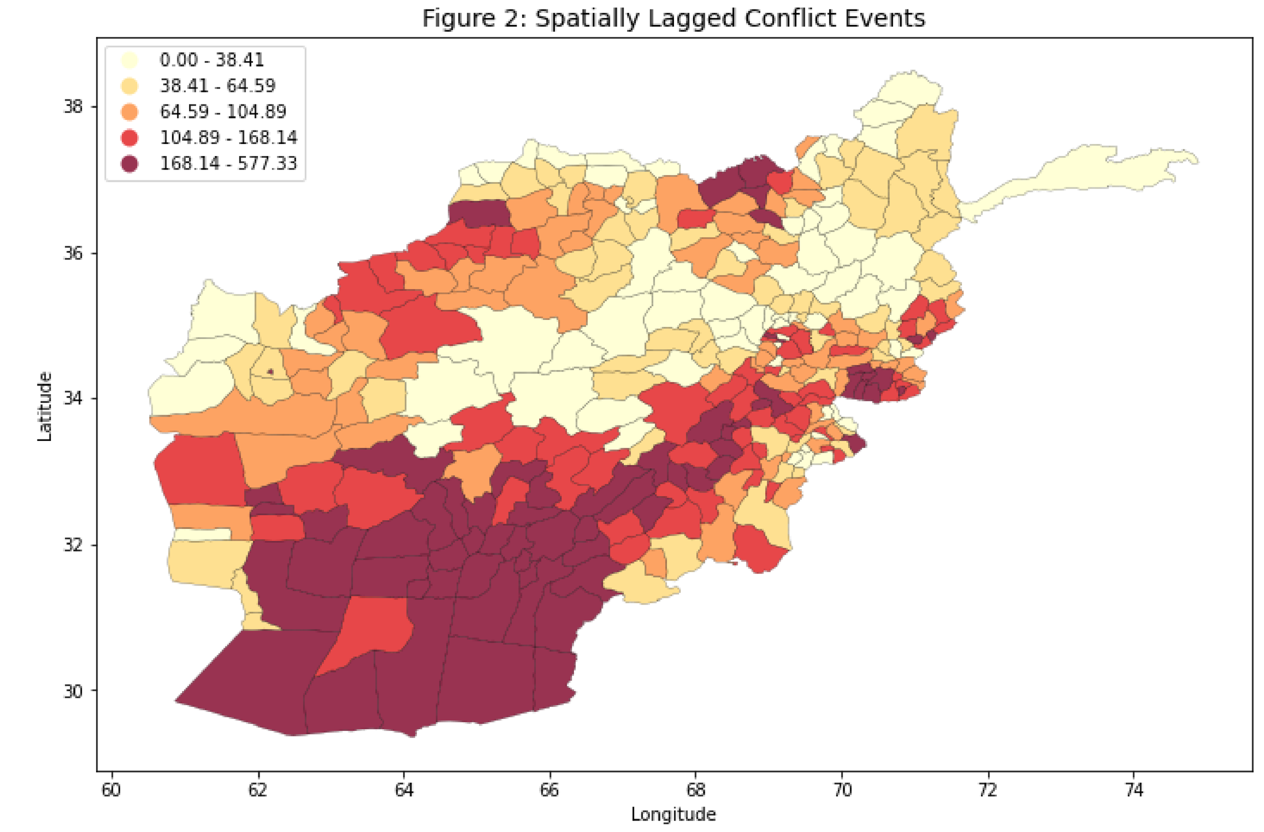


Figure 3: Local Moran's I Coefficient Map

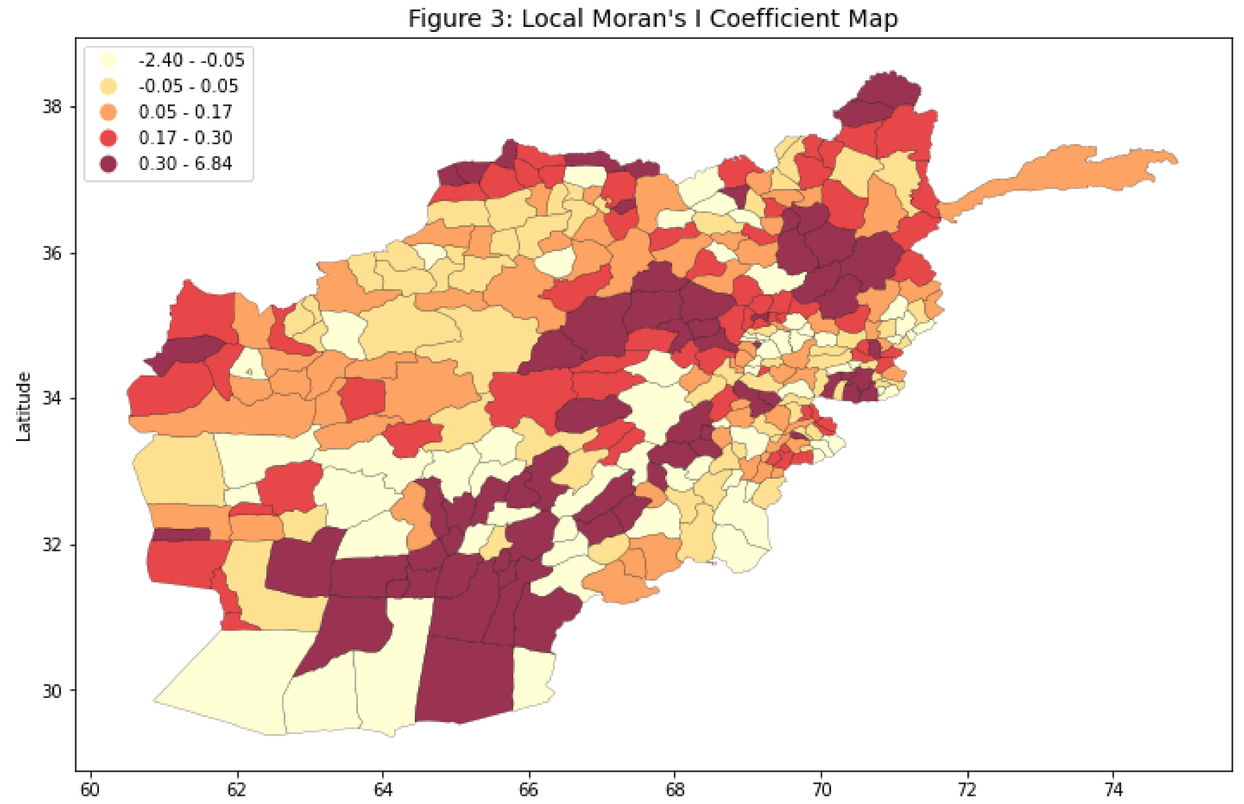


Figure 4: DBSCAN Model of Conflict Events

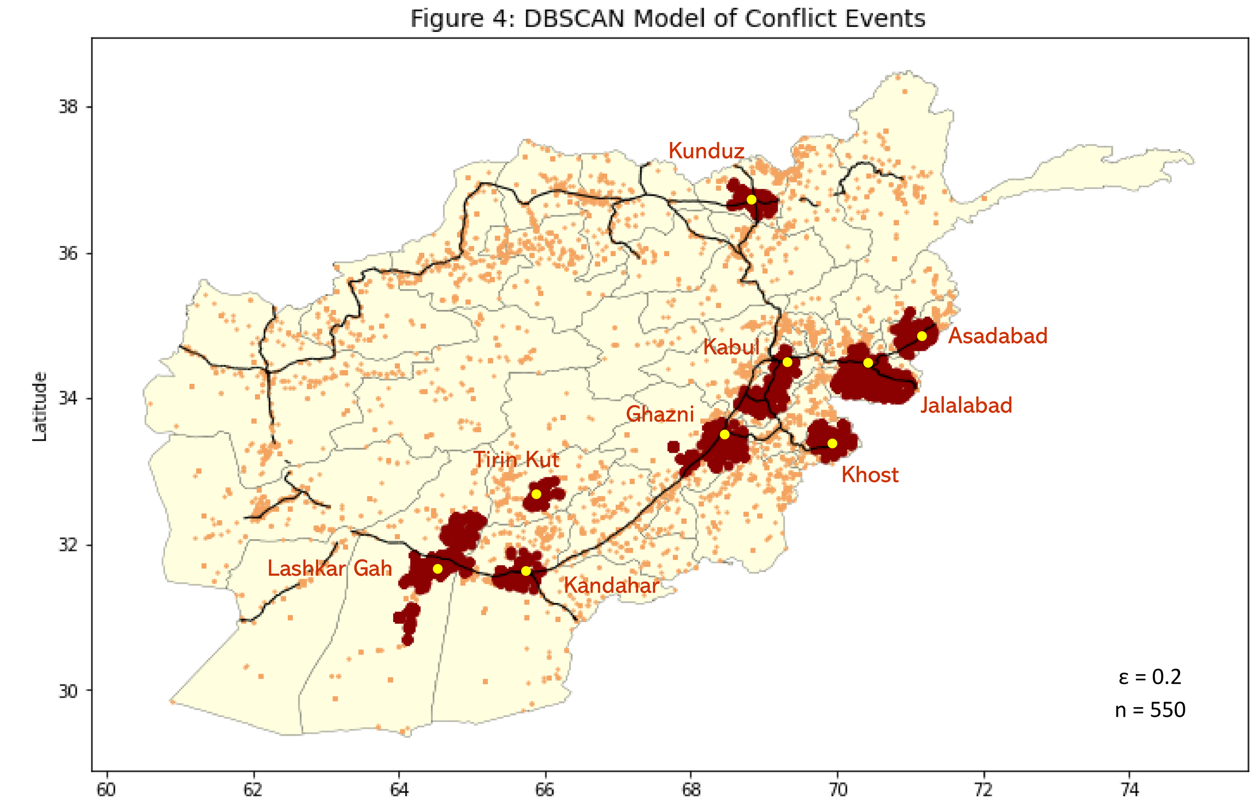


Figure 5: Local Moran Quandrant Map

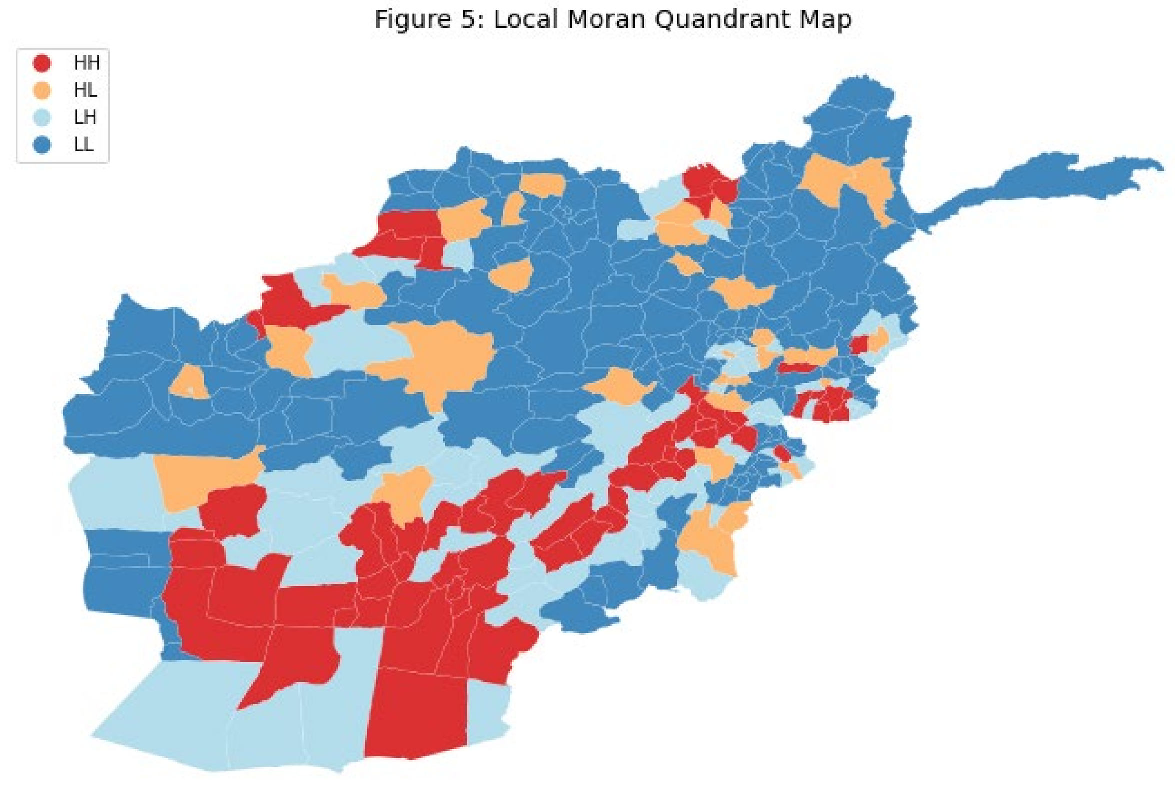


Figure 6: Local Moran Statistically Signficant Quandrant Map

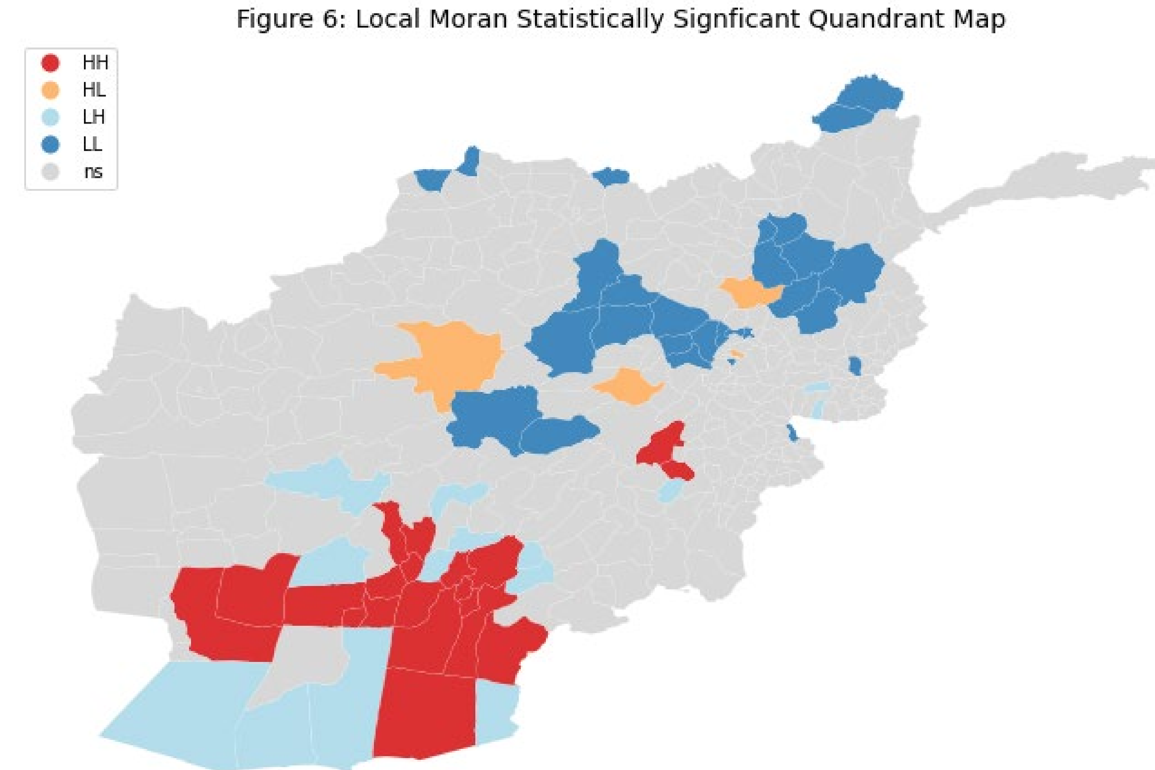


Figure 7: Moran Plot of Conflict Events

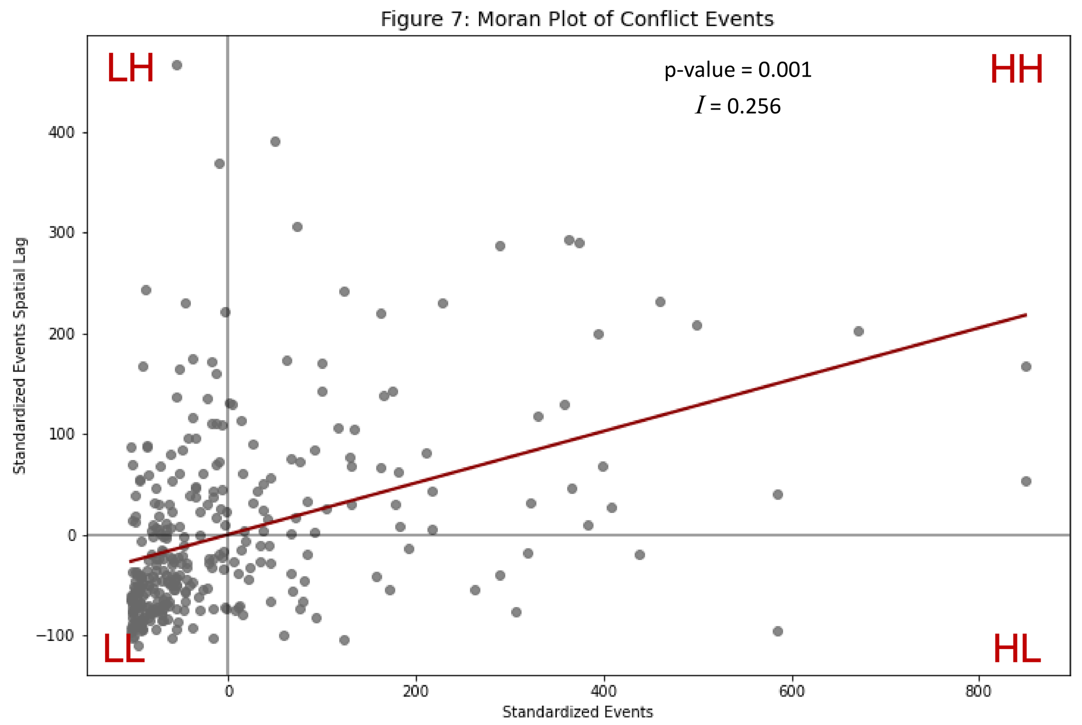


Figure 8: Distance From Nearest Provincial Capital

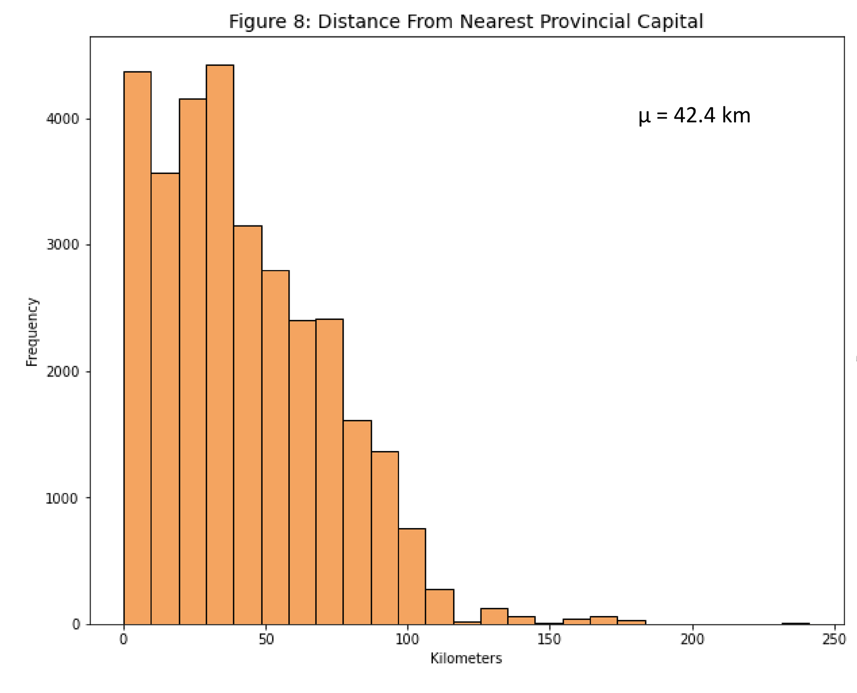


Figure 9: Distance From Nearest National Highway

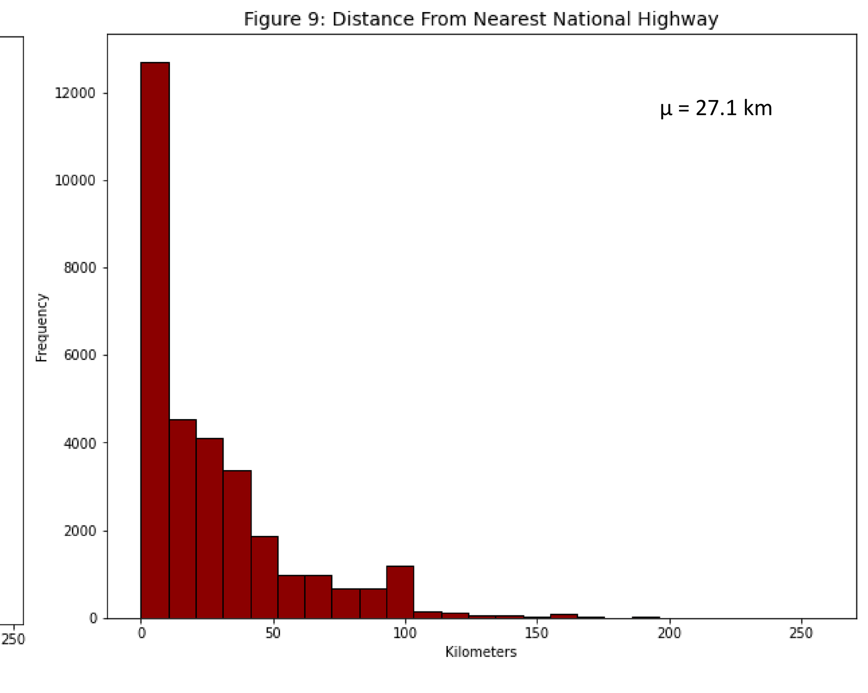
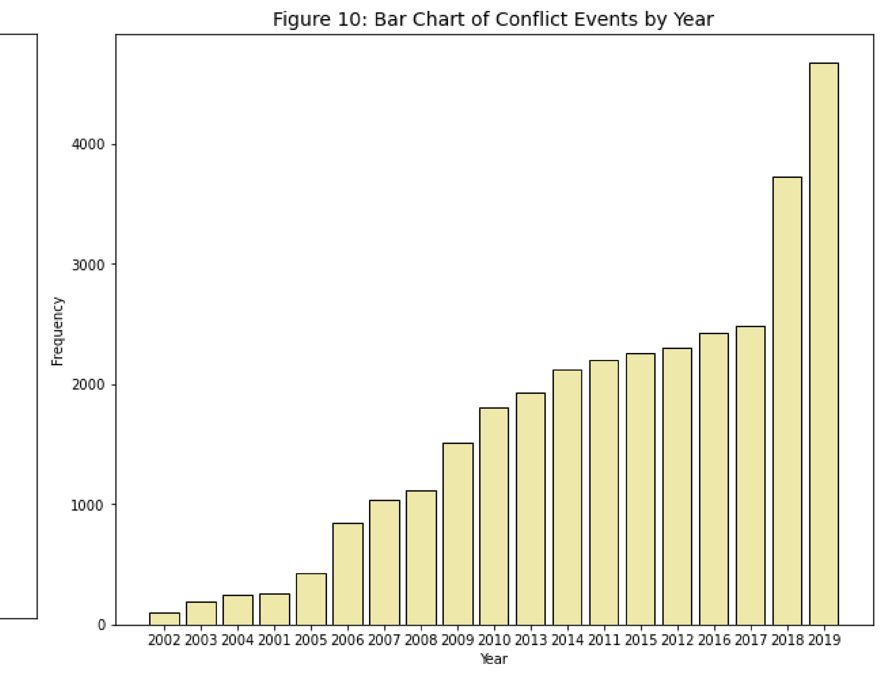


Figure 10: Bar Chart of Conflict Events by Year



RESULTS

Figure 1 reveals that conflict events are concentrated in the southern and eastern regions of Afghanistan. Small clusters are also present within the northern and western parts of the country.

Figure 2 examines the spatial lag of all 328 districts. The value in each district represents the local average standardized count of conflict events in its locale. It is clear that regions with similarly high or low values are closer together, implying positive autocorrelation.

Figure 3 shows how similar each district is compared to its neighbors. Multiple clusters of similar values appear in the southern and eastern regions of the country.

Figure 4 shows nine of the total 13 clusters predicted intersect a national highway. Conflict is less likely to occur where there is a lack of major infrastructure, most notably in central Afghanistan.

Figure 5 captures multiple clusters not recognized in the DBSCAN model. Most hotspots (HH) of conflict are seen in southern and eastern Afghanistan, with three additional hotspots present in the north. The map shows 61 districts are designated as having high concentrations of conflict.

Figure 6 displays 64 districts that are statistically significant ($\alpha = 0.05$) based on Local Moran's *I* values. 19.5% of districts in Afghanistan present a higher spatial concentration of conflict than what could be randomly generated.

Figure 7 shows positive spatial autocorrelation throughout the country. Districts with similar values of conflict events tend to be closer to each other. Most districts lie in Quadrant 3, indicating there are significantly more coldspots than hotspots of conflict.

Figure 8 shows that conflict events tend to be in the surrounding outskirts of provincial capitals. This could potentially be the result of surrounding roads and villages holding more tactical significance over city centers.

Figure 9 shows that conflict events are more likely to occur close to National Highway infrastructure.

Figure 10 indicates a moderate increase in conflict each year from 2001 to 2017. Conflict significantly increased by 89% from 2017 to 2019.

Figure 11 shows that Hilmand is the most dangerous province of Afghanistan with the highest event-to-casualty ratio. Ghazni, Kandahar, and Nangarhar are also provinces featuring a high concentration of conflict.

Table 1 signifies there is an extremely high correlation between the number of events and the number of casualties in a province. We can assume that the results of a global spatial autocorrelation analysis of casualties would produce similar results as seen in this analysis.

CONCLUSIONS

This analysis aimed to provide insights into the spatial distribution of conflict events over 18 years of war in Afghanistan. The global and local spatial autocorrelation proved to be positive using the Moran's *I* statistic, showing districts with similar amounts of conflict are more likely to be clustered together. A greater number of hotspots were found in southern and eastern Afghanistan, particularly close to national highways. Further research in analyzing the spatial distribution of casualties and the specific involvement of belligerents would provide more effective insight into where and how war was conducted in Afghanistan since 2001.

Table 1: Simple Linear Regression Model of Events and Casualties

r	R^2	p-value
0.974	0.948	3.984e-22

Figure 11: Multiple Bar Chart of Casualties and Events by Province