Comprehensive Analysis of Global Terrorism Database

CSE 5520 Fall 2021, Data Visualization of Communication

Shubhangi Shubhangi

12/09/2021

Contents

1 Introduction

Terrorist attacks pose a great threat to global security, and their analysis and prediction are imperative.Considering the high frequency of terrorist attacks and the inherent difficulty in finding related terrorist organizations, this project is to visualize the analysis of different aspects of terrorism around the world.

The analysis and prediction of terrorist attacks support targeted attacks on terrorist groups and provide valuable information for anti terrorism and terrorism prevention. The occurrence of terrorism activities involves various social, political, and economic factors, which differ from country to country and vary over time. The characteristics of such terror acts, including the specific tactics, weapons, target types, or fatalities inflicted, also change across time and space, highly used words which depicts the motive behind the terrorism.

In this project the visualization present various types of patterns and thus to facilitate explorations of the incident data from different perspectives The scope of this project is to drill down the terrorist events around the world from 1970 through 2017. The primary objectives are to identify and highlight the geographical and temporal patterns of the terrorism, to analyze the causalities with respect to region, countries and year to analyze the most active gangs killed lots of people with respect to region to discover the main parameters of a attack type, weapons used and targets to allow the user to customize the analysis and to explore the data in the most interactive way. Tried multiple models to predict the number of kills due to terrorism in upcoming years around the globe. Also, analyzed that the happiness of the people definitely gets impacted with terrorism in the region.

The idea behind the project is to find out how the terrorism has developed in the Western world and whether we need to build tall walls to protect ourself against future threats. The topic has been chosen to be more global oriented, because It enables aggregation on many geographical levels including the globe, regions, countries, states, and cities and It is very diversified and encapsulates many interesting attributes Analysis and interpreting the patterns will be beneficial for anti-terrorism group to increase the security and make some rules/ policies to control the occurrence of events.

2 Dataset Description

The dataset is very comprehensive and contains a lot of terrorism-related information. I have downloaded the entire dataset Global Terrorism Database, available from Gtd Homepage. It contains 156,772 terrorist attacks x 137 features, and takes 142.3 MB of disk space. It's worth to mention that it is almost completely encoded (strings/long numbers to short numbers). To decode the dataset i have looked at the codebook available Code Book. After exploring the codebook i have

discovered some columns having lots of null values, some columns data was skewed or not relevant, which i have removed. See the corresponding notebook Cleaning Data for further details on how we approached.

I have used multiple combinations of features to extract the best information, patterns, which contain the quantitative as well as the qualitative information of the main interest. After decoding, cleaning, filtering, and encoding steps, we've got 156,772 rows x 15 columns like region, country, weapontype, attacktype, targettype, gname, iyear, latitude, longitude, events, nkills, nwound, motive To analyse the happiness is related the terrorism i have picked the happiness data from kaggle World Happiness Report.

3 System Functionality

3.1 Figure 1.

0pt12pt plus 4pt minus 2pt0pt plus 2pt minus 2pt The geospatial map is interactive and change according to the selection of region, country and time. The location in the map also get changed with the selection of region and country. The number of incidents or the markers gets changed according to the number of event in selected years with respect to region and country. The size and color of the markers also varies, markers gets prominent if the number of attacks are more. If we put the cursor on one of the marker it will give more information like Death, State, Injured, Attacks. Geospatial map is one of the best way to visualize the number of incidents around the world

3.2 Figure 2.

0pt12pt plus 4pt minus 2pt0pt plus 2pt minus 2pt The Scatterplot and barchart is interactive and change according to the selection of region, country and year. The x-axis is year and y-axis is count and the bar plot with yellow color is depicting the number of attacks and red bar is depicting number of the injured . The markers are scatter and those are showing the variation in the total number of death in the chosen region and country in the selected time frame.

3.3 Figure 3.

Pie Chart is an interactive graph which showing the total count of death, attack and injured with respect to region, country and year selected

3.4 Figure 4.

When we click on second link which is "Causality Analysis" then first figure is Sunburst Chart, which is an interactive chart. This chart change with the selection of region. Till now we saw the overall picture of death, attacks and injured. Now, we are majorly focusing on number of kills with respect to the attack types and target type according to the selection of region. The area coverage

in the chart shows the frequency of the attack type and target type and when we keep or cursor on any of the section it will show related causalities.

3.5 Figure 5.

The Volin and the box plot shows the top most gang according to the number of kills in the selected region. The box and volin plot shows the min, max and median of gang names and number of people killed by those gang.

3.6 Figure 6.

The Bar plot depicts the frequency of weapons used in the selected region

3.7 Figure 7.

In the 2-D density plot the x-axis is showing the years and the y-axis is showing the count of attacks. The plot is providing the information about the years with maximum number of attacks and count of attacks.

3.8 Figure 8,9.

To predict the number of kills the prediction models like Regression, Decision, K Nearest Neighbours has been implemented with respect to the year. Decision tree showed outstanding results.

3.9 Figure 10.

In the drop down there are two options, first is default in which all the incidents has been scattered on geospatial map and the second, is KMEAN clustering in which the cluster size has been decided by silhouette score according to the incident happened and displayed the cluster with the mean

3.10 Figure 11.

The most frequent words has been grabbed according the frequency from the motive feature and then displayed in word cloud

3.11 Figure 12.

The data has been taken from happiness report in the world and joined on the feature type region where the hypothesis testing was carried out that the people in the region with less terrorism are happier than those who lives in terrorism prone region. The hypothesis was accepted for similar terrorism regions and rejected for opposite.

4 Dashboard Figures

4.1 Figure 1

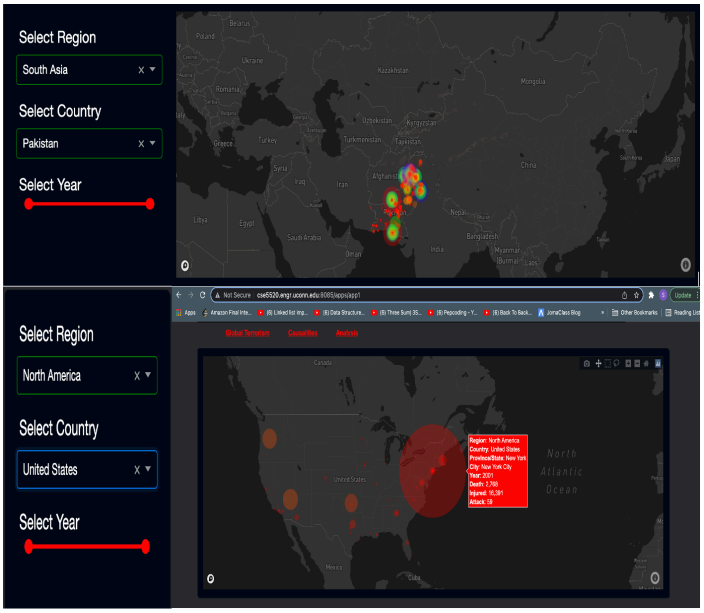

Figure 1: Geospatial Map: Terrorist Incidents

4.2 Figure 2

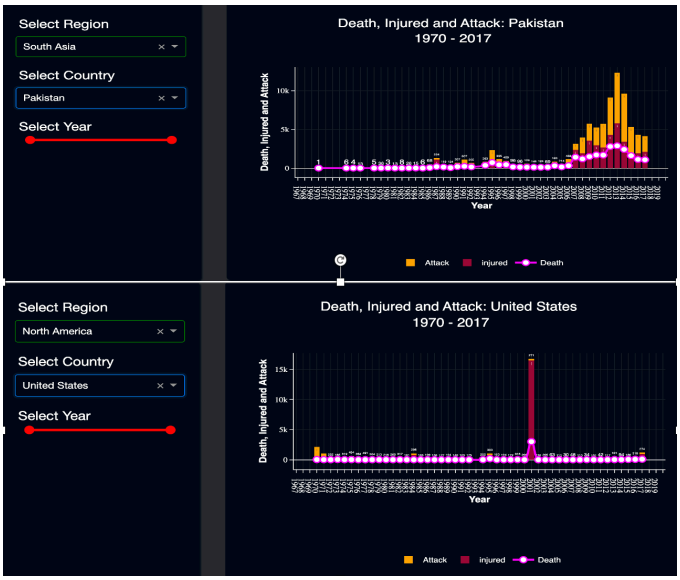

Figure 2: Line and Barchart: Death,Injury and Attack

4.3 Figure 3

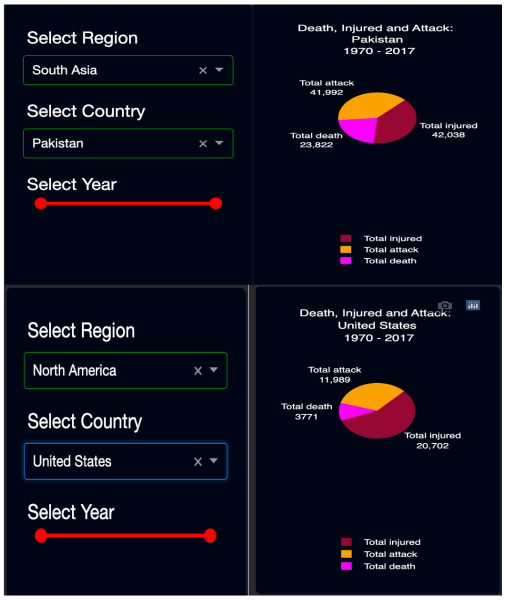

Figure 3: Piechart: Total Count Of Death,Injury and Attack

4.4 Figure 4

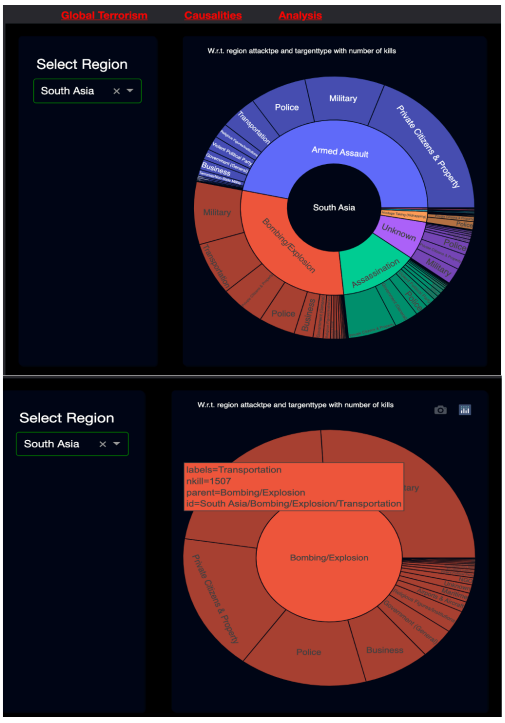

Figure 4: Sunburst Chart: Kills, Targets and Weapons w.r.t Region

4.5 Figure 5

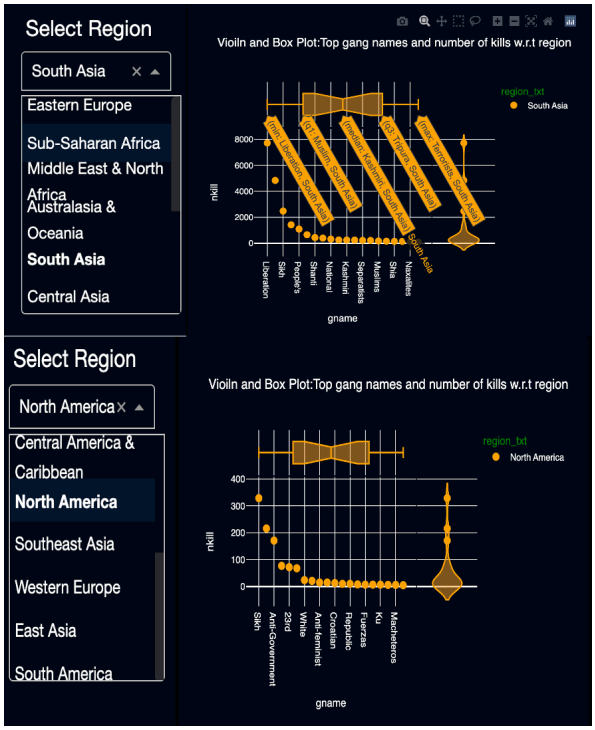

Figure 5: Volin and Box Plot:Most Active Terrorist Groups w.r.t Region

4.6 Figure 6

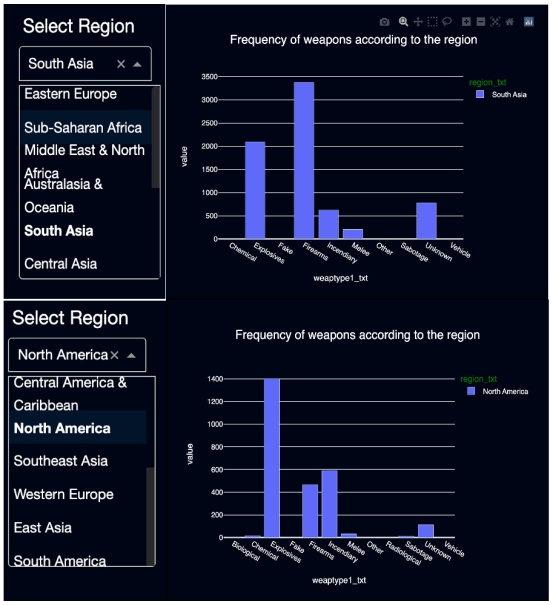

Figure 6: Barchat: Frequency of Weapons Used w.r.t region

4.7 Figure 7

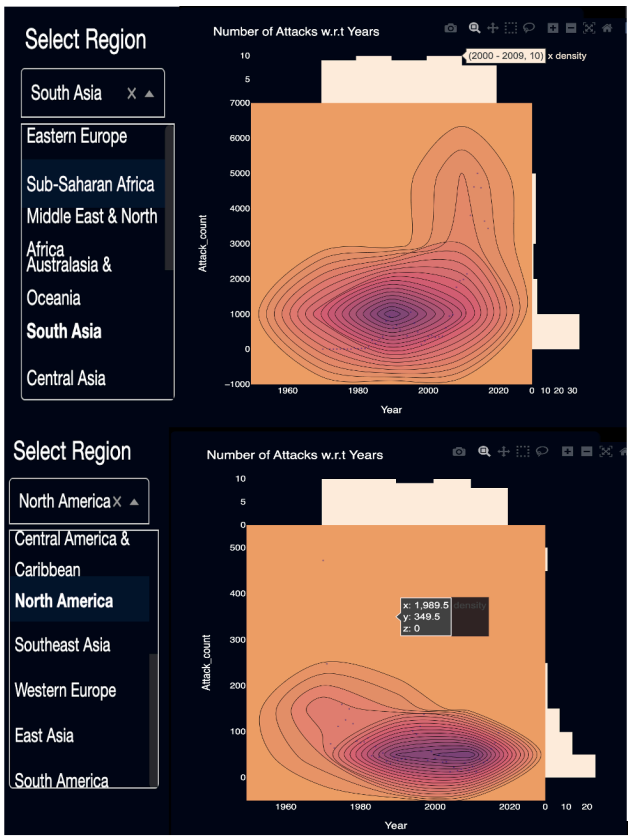

Figure 7: 2-D density plot: Density of attacks from 1970-2017 w.r.t region

4.8 Figure 8

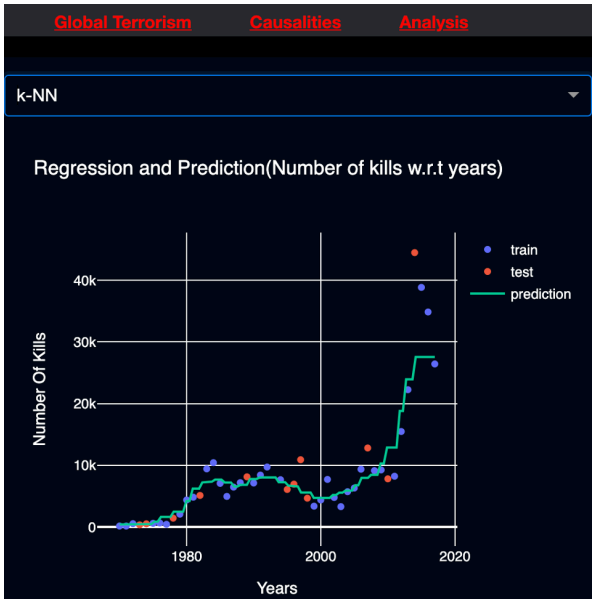

Figure 8: KNN: Predicting the number of kills w.r.t time

4.9 Figure 9

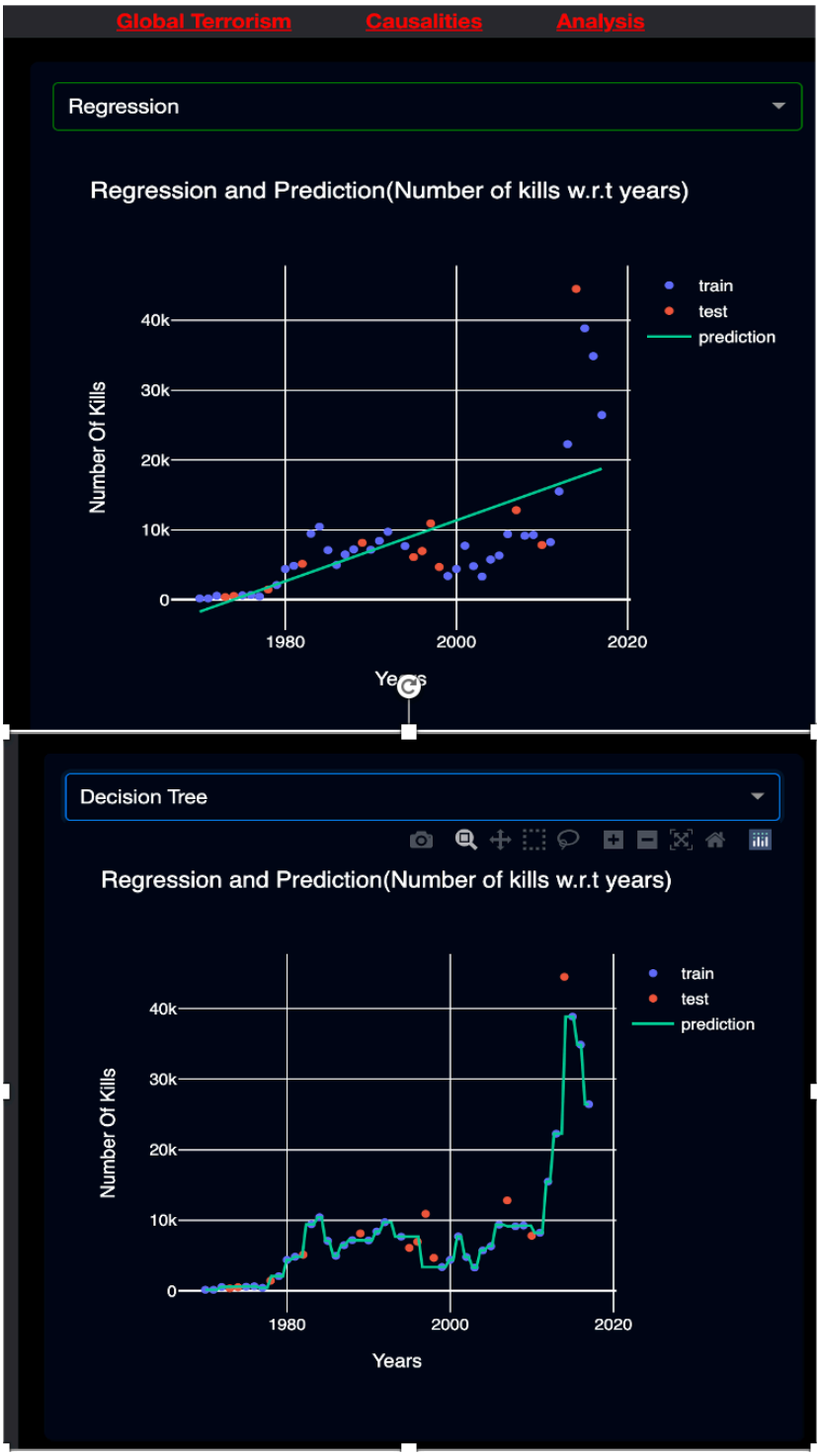

Figure 9: Regression,Decision: Predicting the number of kills w.r.t time

4.10 Figure 10

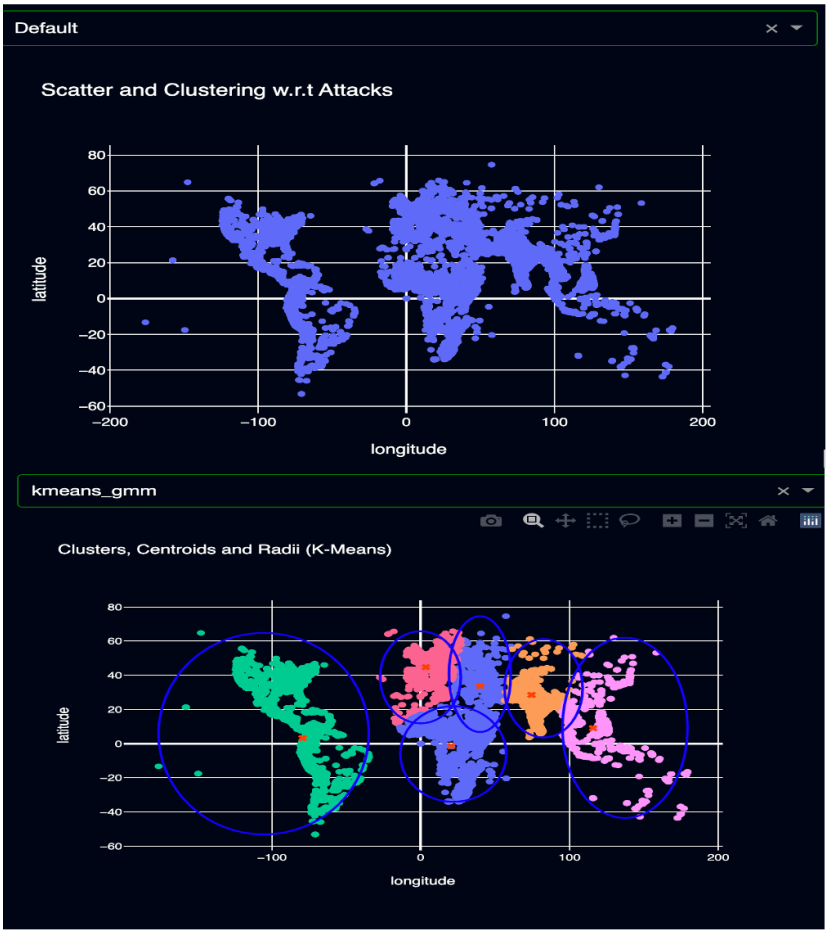

Figure 10: Scatter plot and KMEAN: Terrorist incidents around the world

4.11 Figure 11

Figure 11: Word cloud: Most frequent motivating words behind the incidents

4.12 Figure 12

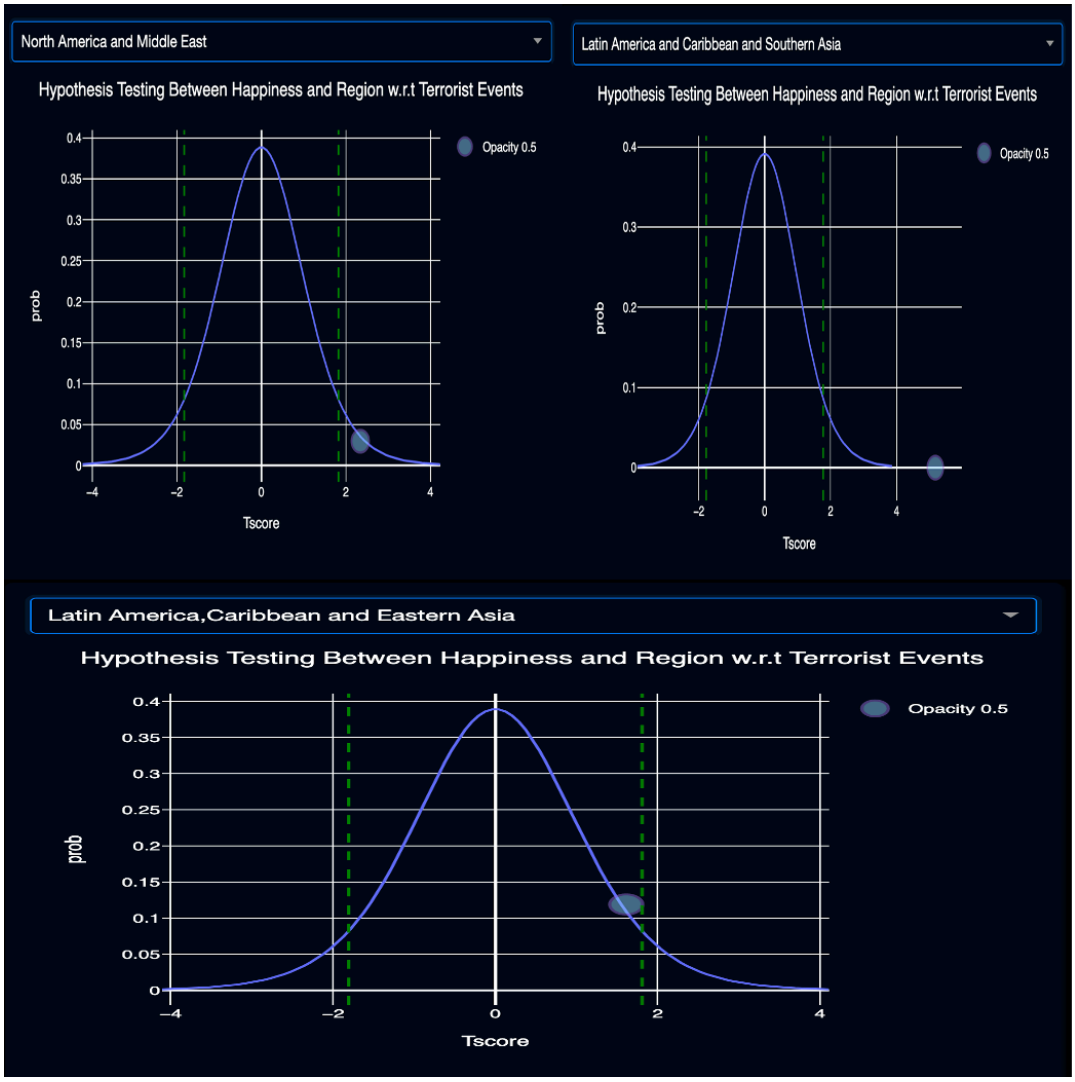

Figure 12: Hypothesis Testing: People in less terrorism regions are happier comparative to higher terrorism regions

5 Accomplishment Summary

	Your Name	Shubhangi		
(Note: You can add additional Categories and Visualization Methods if necessary)				
		Level of sophistication		
Categories	Visualization Methods	Basic	Medium	Intensive
Clustering:	Hierarchical			
	K-means			yes
Classification:	Gaussian mixed model			
Network Analysis:	Network Visualization			
Correlation Analysis:	Linear regression			yes
	Pearson correlation			
	Kernel Density Estimate			
Hypothesis Testing:	t-test			yes
	p-value			yes
Statistics:	Boxplots		yes	
	Violin plots		yes	
	Histogram			yes
Geospatial Analysis:	Cartogram map			yes
NLP/Text Mining:	WordCloud		yes	
	Barplot			yes
	N grams			
Basics:	Line graphs			yes
	Pie Chart			yes
Survival Analysis:	KM-Plot			
Other:	Sunburst			yes
	2-D density plot			yes
	Scatter Plot			yes

Figure 13