Terrorism Research Project
An Investigation Into Global Terrorism Trends

Hrushik Reddy Tadvai

170007110

I. Problem Statement

The overview of this report is to find out how terrorism has developed in the modern world and illustrate the temporal patterns of terrorism. According to industrial statistics, terrorism costs the world around 30 billion dollars, and countless lives lost per year. The threat of terrorism is a worldwide burden, with several states experiencing continuous, frequent attacks.

Since the catastrophic events of 9/11, there has been a drive to collect and interpret terrorism data to identify characteristics and patterns of terrorist incidents. Many of the surveys on public opinion on terrorism indicate that a significant sector of the population is worried about such events. The ultimate goal is to identify which people based on geographical and other types of data are at most risk of acts of terrorism and identify key trends to understand how terrorism operates across the globe.

The following contexts are analyzed in this thesis:

- Terrorist attacks over time
- Apparent distinctive aspects of attacks
- Spread of attacks over the globe
- Activities by the deadliest groups

In this report, the analysis will enormously rely upon the Global Terrorism Database; This is the most comprehensive unclassified source of a database on these matters available to the public domain containing 137 features of the attack.The information is kept up to date and is maintained by respected professors of Maryland university increasing its credibility.

II. State of the art

Respected professors of Maryland University Gary LaFree, Laura Dugan, and Erin Miller combined to create a book to put terrorism into context and manipulate data from the same database. This book was inspired after 9/11, and the authors sought to further their understanding of global terrorism and explore how these impacts have shaped our behavior and attitude towards terrorism. These researchers demonstrate how even small-scale impacts significantly impact people's attitudes and policies in these matters. The study has also considered only the terrorism database available and then explored the evolution of these events. There is a clear American focus in some of these visualizations and reports, but the report delves into global affairs.

This publication explored the tactics and the casualties by the decade to illustrate the changes in terrorism over the past century. This inspired exploration of chronological elements further and visualized these groups' weaponry change in a time series. The report briefly examines the relationship between missing data and the time range of the incident and concludes that data in the past century is likelier to have inaccurate and missing values.

There was a further focus on regional data and groups where they compared regions' mortality count by bar charts. It was decided that this report will portray this information in a color-scaled world map as it is more visually compelling and makes it simpler to identify general patterns within the map.

Another reputable source for reference was the analysis of terrorism by Hannah Ritchie, Joe Hasell, Cameron Appel, and Max Roser. This study was more interested in laying out the relevant statistics and distributions to explore the data sets and primarily investigate the bias by Media to terrorism. The main questions asked were if there was a case of over-exaggeration by public media lobbies over terrorism incidents and public opinions using data collected by surveys.

This was a more global outlook and identified limitations in data collected by countries. A significant point raised was the different statistics of counts of killed and wounded and questioned this data's accuracy. The paper assumed the location of the incidents as the original risk point without taking into consideration the scale of the incident. Furthermore, attacks by the same group over multiple locations were considered as different attacks for simplicity. This report will group these attacks as one row.

Comparing to the aforementioned papers, this analysis will follow similar objectives to identify relevant trends in global terrorism. Clear limitations in these reports have been identified, and the analysis will explore possible methods to counteract these.

III. Properties of Data

A. Data structure

This study has been constructed by leveraging data from the global terrorism website containing relevant terror-related information. The information is the largest source of public data available on issues to terrorism, and it has been updated and maintained by the respectable professors of Maryland University.

The source included one enormous excel file containing 191465 terror attacks from 1970 to 2018 and 137 different information values about the attack, including kill count, weapons, target type. Twenty-five variables were selected to form a pandas data frame based on the relevance to the

objectives; This improves the linear time taken to upload and run algorithms, and the discounted columns provided no additional value to the analysis. Furthermore, many of these columns contained a large number of NaN values, which made it near impossible to identify accurate trends in this material

B. Variables

The following elements were considered:

- The Geo locations of each incident provided in city and country data and latitude and longitude points, making geographical analysis possible. Data without any positional or chronological information were deleted; this was the case for around 0.14 percent of data.
- Chronological data in the form of dates for periodical analysis. Missing rows were deleted.
- Categorical information including target type (Ambassador, government official, celebrities etc), Terrorist groups. Where information were missing, the cells was filled with Unknown string to compensate for the missing information.
- Numeric statistics including kill, wounded counts for categories of people. Null values were converted to the medians of the column based on corresponding categorical data where integer data was missing. This is more accurate than column means, and the median was chosen to diminish outliers' effect on data.
- Summary of attack in context.

C. Limitations of Data

A quick examination of the data demonstrated that recent and western countries' data were more likely to be sufficiently complete with inferring the distribution of incidents for later analysis. A grouping by countries records' mean count of missing variables indicated by the figure below that a third world country is more likely to have missing data.

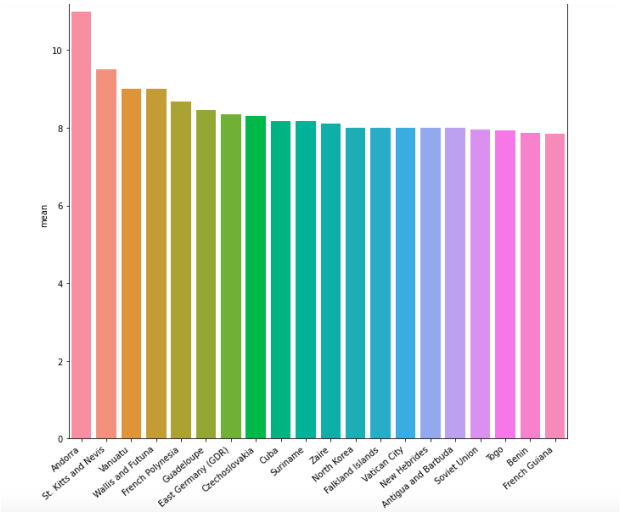

Figure 1.

There are limitations in information available when looking to identify trends in the past century. There is a greater count of missing data for the 1980s and the 1990s, as illustrated by the graph below; this is due to more inadequate data collection practices and the limited availability of technology required to store databases.

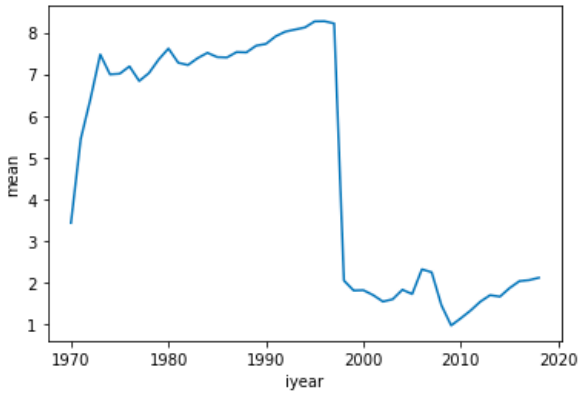

Figure 2.

IV. Analysis Plan

The primary tool used for the analysis and visualization is Python to synchronize parameters better and assess their effects on data analysis with greater control and account for human reasoning with data. Text in this section follows the chronological approach of addressing the objectives.

A. Data Preparation

The data was imported as a pandas Dataframe, which is the most popular method of handling data. Many of the planned use of algorithms do not consider blank values; there is a matter that needs to be addressed urgently. The dataset is extensive; therefore, it was acceptable to consider deleting rows with missing time and location data. Numeric and categorical data were handled, as mentioned in the data section of this report. The data frame had an explicit limitation in grouping attacks carried out in multiple locations; this was separated into different rows. This matter was addressed by grouping by time and the group responsible.

B. Initial Exploratory Data Analysis

Initial Exploratory analysis was carried out using histograms, density maps, scatter plots, and a correlation matrix. Clear trends and patterns were identified, and this helped decisions on what factors to visualize. Human interference was required to remove irrelevant data and those that showed no signs of patterns and identify those that required feature engineering.

C. Data Transformations

Some numeric values were normalized to eliminate anomalies that can make the analysis more complicated and minimize outliers' effect.

D. Visualizations

Next, visualizations were contrived to explore arguments in the objectives. There was a particular focus on the temporal and spatial distributions where clear patterns were found and visualized using an appropriate distribution decided by human reasoning. Chlorographs represented the geographical data of spreads of attacks, and time-series line graphs illustrate trends over time.

E. Discussion

The plots were interpreted to question the objectives, and clear trends were explained. Where possible, a background exploration was investigated to offer suggestions to these interpretations.

V. Analysis and Discussion (952 words)

Surprisingly, the data illustrates that mass fatality attacks are increasingly rare and that attacks that claim no casualties are more common than attacks that do. This brings the average mean for attack mortality to 2.4. A common misinterpretation is that the horrific rampage of 9/11 or the Paris attacks encourages us to think of terrorism as producing mass casualties on average.

The figure below describes the correlations for the whole dataset to identify any potential correlation trends between variables. The dataset was normalized to account for any outlier effects. There were no clear indications for any potential relationships discounting the relationship between the number killed, and if the perpetrators were also executed. This tells us that for larger terror attacks, the assailant is more likely to be killed.

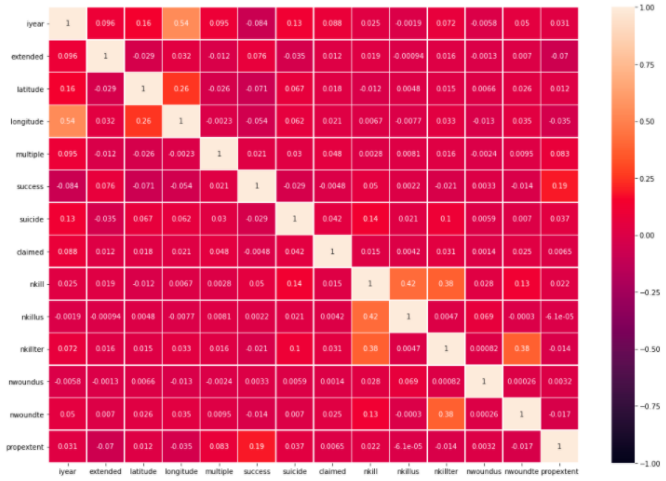

Figure 3. Correlation matrix of entire data set (*corr.png*)

A. Chronological Data

The figure below illustrates the number of assaults has been steadily increasing till 2010, where there was a sudden increase in the number killed; figures almost quadrupled in 5 years. Variables for this graph were calculated by taking the sum of deaths for each year to identify peak terrorism points. Over this decade, the death toll has ranged from 7,731 to 44,286. 2015 was the year where the attacks reached their peak. Most of these deaths were linked to the Middle East and the African region where ISIL and other groups were at its peak based on the latest figures. For later figures this graph tells us that a large proportion of the deaths in this analysis will be linked to the past decade. It is possible that data from earlier decades is missing or inaccurate therefore the report may be biased towards more recent events

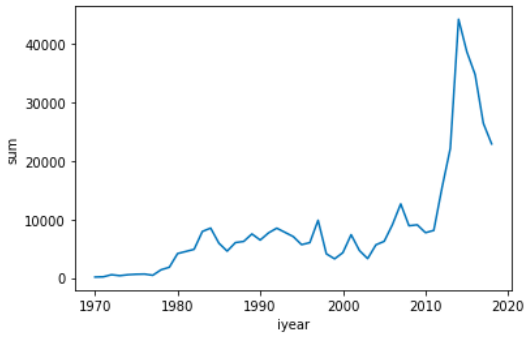

Figure 4.

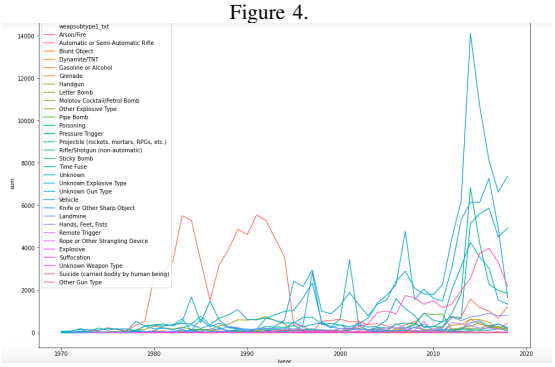

Figure 5.

With time, the deadliest types of weaponry used have evolved; in the 1980s, non-automatic guns or pipe bombs were the most common attacks. In the mid-2000s, there was a severe increase in the number died by explosives and automatic weapon attacks. We can observe that Bombing/Explosion was always the most frequent type of attack, followed by Armed Assault in modern times. The diagram was calculated by grouping by year and weaponry and visualizing each type of methods' sum kill count to identify the deadliest types. Sum was chosen over mean to discount bias of outlier solutions, this value would be heavily skewed by the count of people present at the incident.

B. Attack types

Following from the previous section, further analysis was conducted on weaponry types. For each weapon type the count and mean were calculated; summation was already explored in the data above and it was critical to identify which weaponry was deadlier on average to better plan against. There appears

to be a clear correlation where particular arsenal was used to a deadlier extent. Guns proved to, unfortunately, be the most common and one of the deadliest methods. The most common is an unknown explosive type, while the deadliest type of attack on average is surprisingly suicide. This plot was contrived by calculating the count and mean deaths from each weaponry type. Mean was chosen over the sum to explore which weaponry was the deadliest on average. In Europe, the most common were blunt objects like knives or vehicles, while gun types were more common in many other regions. This indicates a change in tactics based on geographical locations.

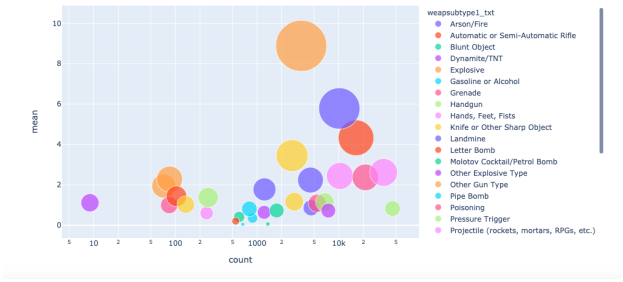

Figure 6.

C. Terrorist activity locations

There was a further focus on regional data. Information was portrayed in a color-scaled choropleth world map as it is more visually compelling and makes it simpler to identify general patterns within the map. Figure 7 shows the distributions of average terror attack death count against Geo locations. As expected, the deadliest attacks appear to be in the middle eastern and, surprisingly, America. This graph is misleading because it suggests America suffers from most terror impacts, but it highlights that an average terror attack in America, however low the number may be, will be expected to be deadlier on average compared to other countries. To construct these latitude and longitude points were used as they were the most accurate types of data possible to pinpoint locations of these incidents. The rainbow colour scheme selected makes it easier to distinguish between different points.

To fully understand the spread of terrorism, please refer to the graphs in the Analysis results section where the 2015 spread of attacks are represented—only the severe cases where a death count of greater than seven was recorded. This was to cluster the severe cases around the globe, and we immediately identify most of the attacks congregate into a few regions near Syria, central Africa, and near Afganistan and Pakistan. Clustering by KMeans was used to identify the epicenters of many of the attacks and the regions the clusters covered. There is evidence to support that a large number of these can be linked to specific groups.

D. deadliest groups

While the media often associates most terrorist attacks attributing to specific groups, nearly half of the incidents are

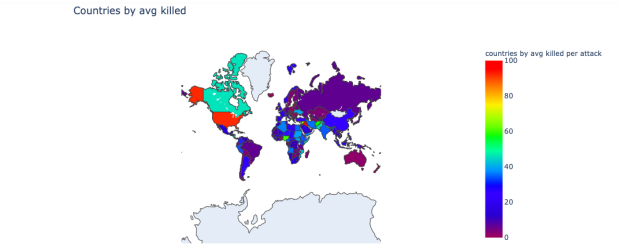

Figure 7.

never claimed, and the perpetrators remain anonymous. There is also a common misconception that the terrorist groups have existed for an extensive lifespan, which is greatly exaggerated; the average is around four years. This was found to be longer for religious groups and nationalists groups. These statistics were calculated by considering the differences between the first and last assaults.

Below are the 10 of the most notorious groups in the world. These afflictions account for 37 percent of the total number of deaths related to terrorism. To illustrate the number of causalities these groups account for, a bar chart was used with sum as the y aggregate. Some of the groups that have started lately in 2000's like the ISIL and Taliban, have shown a great increase in the number of attacks in the past years. The total sum for casualties afflicted to these groups is shockingly high in the tens of thousands, clearly illustrating these groups' shocking deadliness. This data again can be questionable due to the methods of linking these attacks to the said groups. The authors admitted that these numbers are likely to be inflated for earlier decades, and groups may claim ownership of events they have not contributed.

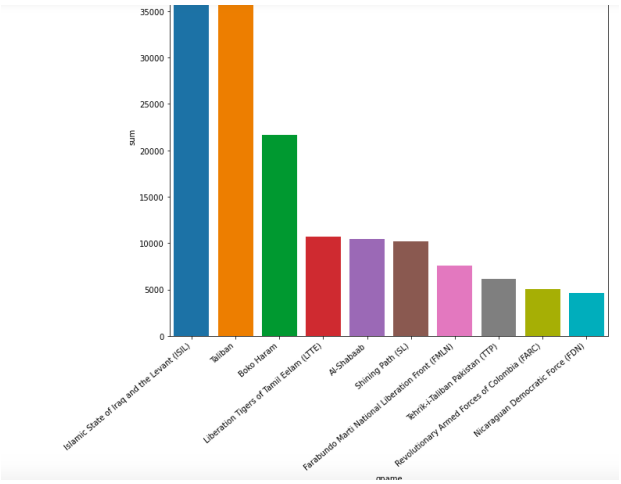

Figure 8.

E. Target types

The piechart has been constructed to identify the most common target types. It is important to note that much of this data for the earlier decades is missing and is only relevant for

more recent times. The data used to construct this pie chart is each targeted type, and each group's count to identify those most vulnerable to terrorism. The most significant portions of the targets were the government officials, army, and police workers. Terrorism is often designed to disrupt and discredit the government and military services by weakening them with targeted attacks. Attacks on private citizens are likely to be more random in densely populated areas. Grouping by terrorist groups has shown that these groups are more likely to attack citizens over others to create panic and fear amongst the general population.

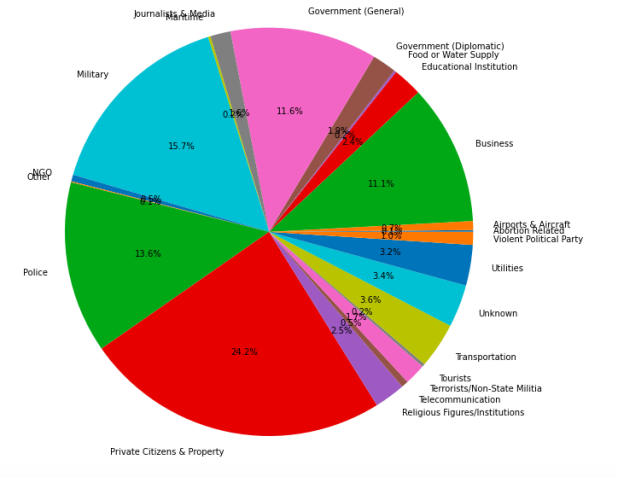

Figure 9.

F. *Wordcloud*

The Wordcloud generated highlights the most common keywords used to describe a terror attack. Worlds like damaged, property, and vehicles were frequently used to highlight these events' extensive damage. It indicates that the purpose is to create as much chaos, panic, and destruction as possible. Positions and titles such as the army, politicians, and celebrities were also specified, foreshadowing these incidents' targets.

Figure 10.

VI. Analysis Results

This publication explores the tactics and the casualties by terrorism acts. The chronological visualizations indicate that terrorism has been on a constant rise since the 1970s, and the most significant peak was reached in 2015 where most attacks were clustered in the African and middle eastern regions. There is also a clear link between significant terrorist groups and the number of terror-related deaths they attribute to their respective countries. We have learned that terror attacks have fallen sharply in western countries since the 1970s and have a lower expected death count except for America, possibly due to relaxed gun legislation.

When the data is drilled down, we find that multiple terrorist attacks are highly concentrated in a small number of cities or states; The clustering below in the global map indicates the regions of high attacks during the 2015 period where the attacks were an all-time high. The figure indicates that most of the attacks were in areas mentioned above and India; There is evidence suggesting that groups like ISIL, Lashkar-e-Tayyiba, and Boko Haram account for most of these attacks.

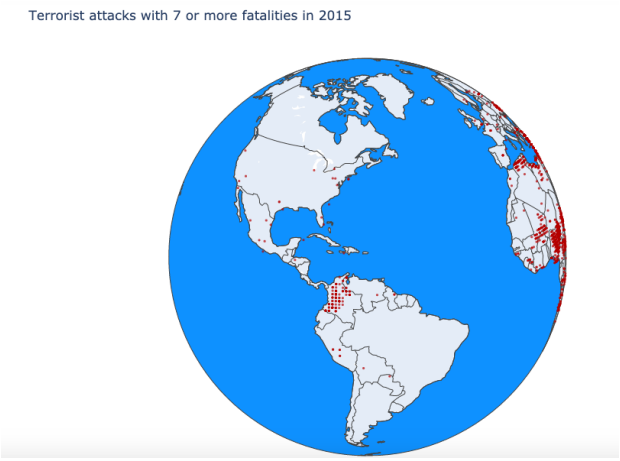

Figure 11.

Figure 12.

VII. Critical Reflection

Understanding the sources and frequency of terrorism can undoubtedly benefit society and policies, including immigra-

tion, global relations, and counter methods. This research heavily relies on the Global Terrorism Database as a critical data source on incidents and fatalities from terrorism. It is the most comprehensive database of incidents to date available to the public domain. The methodology in this paper was sufficient to produce viable steps into the creation of descriptive visualizations. Further steps were taken in normalizing data after an initial exploratory analysis to counteract any effects of outliers, and this analysis has dealt with missing values judiciously. Points regarding changes over time and the spread over geography were considered in an illustrating way. The chlorograph and the clustered map both successfully indicated the spread of terrorism in the modern world and which regions are the most affected by terrorism. Tableau would have been a better software for producing these graphs but it does not offer python's ability to transform and synronise data. Time-series graphs produced effectively visualize the changes in terrorism over time and methods used by said perpetrators. For each pattern/trend, background research has been conducted to offer possible explanations.

However, it does have certain limitations as identified in earlier sections with missing data from the earlier decades and some countries. There are reservations about the completeness of earlier data and for data in certain regions; therefore, caution is thrown against inferring trends over this entire period. The data set also has limitations regarding multiple locational attacks, as discussed, and this has been corrected by using grouping appropriately. However, the database has made assumptions on what terrorism is, which can be subjective to different researchers. This raises queries in the relevance of some of the data used in this analysis.

Factors that could have been potentially further explored were target types and the political stances on some of these groups; this data was either unavailable or a case that much of the data in the columns were missing. These elements could not be considered in these columns because there is a large likelihood that there would be a bias in data collected regarding certain information; some countries may decide not to store these statistics. A possible method to counteract this could have been to look for alternative sources to fill in the missing blanks. It would have also been beneficial to get exact attacks' locations to identify if an attack is more likely in a city's hotspots. Terrorist groups was a factor which could have been investigated further; What is the size of their group, their aims?

There was no data about these incidents' public opinions, which would have been a fascinating factor to consider alongside media coverage to investigate possible media bigotry. There was no data available online about these matters that could have effectively merged with this own dataset and that were reliable.

A. References

- Putting Terrorism in Context: Lessons from the Global Terrorism Database by LaFree, Gary, and Laura Dugan, Erin Miller. 2015
- https://ourworldindata.org/terrorism
- https://plotly.com/python/choropleth-maps/
- https://github.com/hrushikreddy/visuals-project

B. Word Count

problem statement	223
State of the art	424
Properties of the data	381
Analysis Approach	341
Analysis Process	1212
Analysis Results	180
Critical reflection	478