Global Terrorism - Final Project

MATH 4685/6685 (Fall 2019)

**Group Members:**

James Willson (*Data Scientist*)

Sean Kugele (*Data Scientist/Team Lead*)

# Introduction

Our objective for this project was to create statistical models for the purposes of prediction and inference using the Global Terrorism Database[[1]](#footnote-1) as our data set. Each data scientist independently selected several questions of interest. The questions selected included:

(1) predicting if an attack will be successful based on a variety of different factors (**Section 3.1**)

(2) estimating the number of casualties resulting from a successful terrorist attack (**Section 3.2**)

(3) predicting the terrorist group responsible for perpetrating a terrorist attack (**Section 3.3**)

(4) estimating the risk of attack based on temporal and geospatial variables (**Section 3.4**)

The remainder of the paper is organized into the following sections: Section 2 provides additional detail on the Global Terrorism Database (our dataset). Section 3 describes the modeling efforts for each research question, including question-specific data cleansing, generated statistical models, model comparisons, and summaries of findings. Section 4 concludes the paper with a discussion of our findings, general observations about our data modeling efforts, and possible directions for future work.

# Background

The Global Terrorism Database (GTD) contains over 180,000 observations and 135 variables. Each observation corresponds to a terrorist event that occurred between 1970 and 2017 (excluding 1993). Terrorism is broadly defined as “the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation." Data in the GTD is sourced from unclassified media articles (e.g., electronic news archives), and it is currently maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) group at the University of Maryland.

The type of information available for each event includes, but is not limited to, *temporal variables* (e.g., year, month, day, etc.), *geospatial variables* (e.g., latitude and longitude, region, country, city, province, etc.), *event descriptive variables* (e.g., attack type, duration of incident, weapons used, targets, etc.), *perpetrator descriptive variables* (e.g., terrorist group name, number of perpetrators, etc.), and *event outcome variables* (e.g., total number of fatalities, total number injured, extent of property damage, success or failure indicator, etc.).

## Geospatial and Temporal Distribution of Data

As expected, the spatial and temporal distribution of the terrorist events in the GTD are highly non-uniform.

Figure 1 shows the temporal distribution of the events from 1970 to 2017. The number of events appears to increase dramatically from 1970 to the 2010s; however, this is likely driven by differences in media coverage and data collection rather than a reflection of actual underlying trends.

A screenshot of a video game

Description automatically generated

Figure – Histogram showing frequency of terrorist events between 1970 and 2017.

Figure 2 shows the geospatial distribution of terrorist incidents. The highest density of events occurred in the Middle East & North Africa () followed by South Asia (). The lowest density of events occurred in North America () followed by Eastern Europe (). The countries with the greatest number of terrorist events were Iraq (), Pakistan (), Afghanistan (), India (), and Colombia ().

|  |  |
| --- | --- |
| Terrorist Events (1970 – 2017) | |
| A close up of a map  Description automatically generated | A close up of a map  Description automatically generated |
| A close up of a map  Description automatically generated | |
|  | |

Figure – Geospatial distribution of terrorist events.

Figure 3 shows changes in the geospatial distribution of terrorist events over time. Note that the areas of highest concentration have shifted considerably. Between 1970 and 1989, the greatest number of attacks occurred in South America () and Western Europe (). Between 1990 and 2009, the Middle East & North Africa emerged as the region with the highest density of attacks () and that trend continues into the 2010s where the Middle East & North Africa accounts for terrorist events. By comparison, South America accounted for only terrorist events in the same time period.

|  |  |  |
| --- | --- | --- |
| Changes in Terrorist Event Density Over Time | | |
| A close up of a map  Description automatically generated | | |
|  |  |  |
| *Time* | | |

# Modeling

## Q1—Predicting Successful Attacks

## Q2—Estimating Number of Casualties

## Q3—Predicting Responsible Terrorist Groups

|  |  |
| --- | --- |
| Terrorist Group | # of Attacks (percentage) |
| Al-Qaida in Iraq | 606 (3.35%) |
| Al-Qaida in the Arabian Peninsula (AQAP) | 877 (4.84%) |
| Al-Qaida in the Islamic Maghreb (AQIM) | 225 (1.24%) |
| Hamas (Islamic Resistance Movement) | 297 (1.64%) |
| Hezbollah | 107 (0.59%) |
| Islamic State of Iraq and the Levant (ISIL) | 4,274 (23.61%) |
| Kurdistan Workers' Party (PKK) | 1,074 (5.93%) |
| Liberation Tigers of Tamil Eelam (LTTE) | 614 (3.39%) |
| New People's Army (NPA) | 1,573 (8.69%) |
| Revolutionary Armed Forces of Colombia (FARC) | 1,121 (6.19%) |
| Salafist Group for Preaching and Fighting (GSPC) | 182 (1.01%) |
| Taliban | 5,912 (32.66%) |
| Tehrik-i-Taliban Pakistan (TTP) | 1,240 (6.85%) |

A close up of a map

Description automatically generated

|  |  |
| --- | --- |
| *With* Geospatial and Temporal Variables | *Without* Geospatial and Temporal Variables |
| A picture containing screenshot  Description automatically generated | A screenshot of a social media post  Description automatically generated |

|  |  |
| --- | --- |
| *With* Geospatial and Temporal Variables | *Without* Geospatial and Temporal Variables |
| A screenshot of a social media post  Description automatically generated | A screenshot of a cell phone  Description automatically generated |

## Q4—Estimating Attack Risk Using Temporal and Geospatial Features

# Discussion

1. https://www.kaggle.com/START-UMD/gtd [↑](#footnote-ref-1)