Global Terrorism: Terrorist attacks around the world

Team AbracaData

University of Maryland, College Park, MD

E-mail: [abiddata@umd.edu](mailto:abiddata@umd.edu), [dmir1@umd.edu](mailto:dmir1@umd.edu), [dnisha@umd.edu](mailto:dnisha@umd.edu), [kjain307@umd.edu](mailto:kjain307@umd.edu) [skotasth@umd.edu](mailto:skotasth@umd.edu)

## Motivation

Terrorism is a cause of global concern for civilians and government bodies alike. With a lot of terrorist attacks happening all over the world, it has become possible to collect the data about global terrorism and observe patterns from it. Global terrorism dataset contains over 181k incidents of terrorist attacks from all over the world. Apart from the city and the country, the dataset also specifies the latitude and longitude of the attacks thereby giving accurate locations that can help in data visualization. This dataset can help identify the solutions to different problems through data analysis and visualization techniques. With the help of the abundant features available in this dataset, we can find out the intensity of attacks by region for a particular year, whether there are any temporal or geographical trends in the attacks and the relationships between the characteristics of the attacks and their success and failure rates. This analysis will help the governmental organizations take appropriate decisions to increase the safety of citizens and take appropriate measures in preparation for any possible attacks.

Problems

A common problem found in most publicly available datasets is that it has incomplete, inconsistent or redundant data which could be attributed to reasons like unavailability of data at the time of collection, error in entering data or misinformed references. The Global Terrorism dataset that we collected from kaggle is no stranger to the above-mentioned issues. We have identified certain irregularities that can be cleansed and prepped for further analysis.

1. Month and Day columns have 0 as values which is an unacceptable representation of the exact date of the event.
2. Multiple Target and Weapon type/Subtype columns which are redundant.
3. “Unknown” values in each row will have to be dealt with.
4. The record of incidents for the year 1993 is missing in the dataset.
5. Group/Individual responsible and Motive for the attack is not known for many incidents.
6. This is a large dataset with 1,80,000+ records.

## Approach

The dataset is available on both Kaggle and Global Terrorism Database, University of Maryland.

<https://www.kaggle.com/START-UMD/gtd#globalterrorismdb_0718dist.csv>

[https://www.start.umd.edu/pubs/START\_GTD\_Overview2017\_July2018.pdf](https://www.kaggle.com/START-UMD/gtd#globalterrorismdb_0718dist.csv)

The dataset has information about terrorist attacks that happened around the world between 1970 to 2017, except in 1993. The first step in our study is to understand the meaning of each variable using the GTD codebook and start with the data cleaning process.

The column ranges from the year, month, incident summary, longitude, latitude, attack type, target type. The month and year column has missing values so we will fix the missing values with the data available in the summary column. For incidents before 1998, the date column has no data. There are a lot of redundant columns that add no value to data, we will take care of such columns. The columns target type and weapon type has subtypes, for our analysis, we would consider only different types, not on subtypes. Once the data cleaning is completed, we start the exploratory data analysis and find answers to the problems addressed in the problem section. We will also create a dashboard to better understand our analysis.

## Milestones

We will first explore the data, understand the discrepancies in the data set and correspondingly clean and prepare the data for analysis. Further, we will use the tools learned in class to draw inferences from the data set. Finally, we will develop a comprehensive and visual storyline for the project presentation.

## Extensions

This is an exhaustive dataset which has information of over 180,000 terrorist events from 1970 through 2017. The best part about the dataset is its reliability. People who manage the dataset do not add any information unless the [credibility of the sources is determined](https://www.start.umd.edu/gtd/faq/). So, government officials can actually use this information to draw some insights. Analysts can use it to know how terrorism impacts social, economic and cultural aspects (growth) of a region. International crime agencies can use the dataset to find patterns of terrorist attacks.

## Tools

Python, R, Open refine, Excel - Data cleaning, data analysis, data modeling

Tableau, ggplot - Data visualization

## References

<https://www.start.umd.edu/gtd/downloads/Codebook.pdf>

<http://visionofhumanity.org/app/uploads/2017/02/Global-Terrorism-Index-2016.pdf>