ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MILESTONE 1

DATA VISUALIZATION

Alexandre Ben Ahmed Kontouli,
Aristotelis Dimitriou,
Konstantinos Spinakis

7th April 2023

CHAPTER 1

INTRODUCTION

1.1 PREAMBLE

In this project, we have chosen to use the data from the *Global Terrorism Database*™, an open-source database maintained by the *START* consortium (National Consortium for the Study of Terrorism And Responses to Terrorism). The definition of terrorism given by *START*, and the one this data is thus based upon, is as follows: "[Terrorism is] *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.*" START 2022a

The history of terrorism has not been a short one. Though this is a hotly debated topic by authorities on the matter, the roots of terrorism can be traced back to the 1st century and the *Sicarii*, a group of Jewish Zealots who strongly opposed the Roman's occupation of Judea, who would attack Romans and sympathizers alike during public gatherings using daggers. Since then, there has been no lack of groups attempting to promote their ideologies using such —or more— extreme means. In the modern era, the rampant technological progress has given terrorists more weapons, more ways to spread fear in the minds of the populace. Nowadays it seems like not one month can pass without hearing of some ideologically motivated attack in the news, by some terrorist group or other. It is with this in mind that the *START* consortium created the GTD™, "*...in an effort to increase understanding of terrorist violence so that it can be more readily studied and defeated.* " START 2022b

We wish to highlight that we understand that this constitutes an extremely sensitive topic. For this reason we wish to approach it with the utmost care, using nothing but the data found in the GTD™ and only for the purposes of visualizing it, without drawing any unnecessary conclusions, without any personal opinions, without any discussions on the ideologies of any of the groups mentioned in the data or the acts they committed.

A description of the data as well as further explanations on our motivation in choosing this specific dataset can be found in Sections 2 and 3.

CHAPTER 2

DATASET

In the context of the course of Data Visualization, we searched through a wide variety of datasets to find the most interesting one. The GTD™ is a well-documented open-source database, and on the topic of terrorism it is considered to be the most complete database in existence, with over 200'000 records of terrorist acts spanning from 1970 to 2020. It stores information about

1. Date
2. Region
3. Country
4. Perpetrator Group
5. Weapon Type
6. Attack Type
7. Target Type
8. Terrorism Criteria (successful and unsuccessful attacks[1], or ambiguous cases ...)
9. Casualties (Injuries, fatalities, numbers)

The dataset was originally created in 2005 by researchers at the university of Maryland. It was mostly based at the time on another dataset maintained by the PGIS (Pinkerton Global Intelligence Services) from 1970 to 1997. Due to an office move, all the data from 1993 was unfortunately lost and, according to the *START*, "Efforts thus far have been unsuccessful in fully recovering the 1993 data".

Since its foundation, a growing number of scholars have joined the GTD™ team to update its contents and maintain the project. It has since developped a much more rigorous methodology to gather data about each attack. Their website states that "More than 4,000,000 news articles and 25,000 news sources were reviewed to collect incident data from 1998 to 2017" START 2022b, with Machine Learning and modern data mining techniques being used to gather sources.

[1] "For an event to be included in the GTD, the attackers must be "out the door," en route to execute the attack. [...] However, in general if a bomb is planted but fails to detonate; if an arsonist is intercepted by authorities before igniting a fire; [...] the attack is considered for inclusion in the GTD, and marked success=0." START 2021

CHAPTER 3

PROBLEMATIC

In the context of the rising global instability accompanied by an increase in terrorist acts of all natures all over the world, we find it interesting to study the GTD™ and visualize it in ways that might reveal patterns such as regional trends, the frequency of different types of attacks, the relationship between terrorist groups and their targets. This is motivated by the need to understand the intricate and complex nature of the act itself and its motives. This study might additionally give us some insight on the factors that drive and nurture terrorist activities.

By presenting the data in a clear and easy way to understand, we allow anyone curious to learn more about the subject.

Given the attributes we listed earlier there are plenty of ways to represent the data. An advantage we have is that we could show a potential evolution in the character of the acts themselves as the records date from the 1970's to 2020. We could however also classify the data by the nature of the act itself which increases the dimensionality and modularity of our presentation.

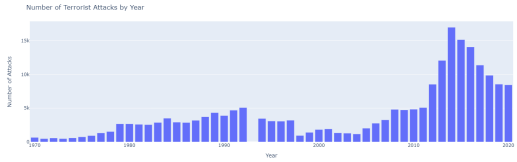
CHAPTER 4

EXPLANATORY DATA ANALYSIS

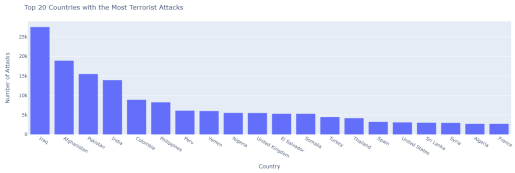
The primary goal of exploratory data analysis (EDA) is to uncover patterns, trends, and relationships in data. EDA offers insight into data and identifies areas requiring further investigation and visualization.

4.1 DESCRIPTIVE STATISTICS

- Total incidents: **209,706**
- Incidents per year, country/region: **Figures 4.1a, 4.1b**



(A) Histogram of incidents per year



(B) Histogram of incidents per country/region

FIGURE 4.1
Histograms of incidents per year and incidents per country/region

4.2 ATTACK, TARGET, AND WEAPON TYPES

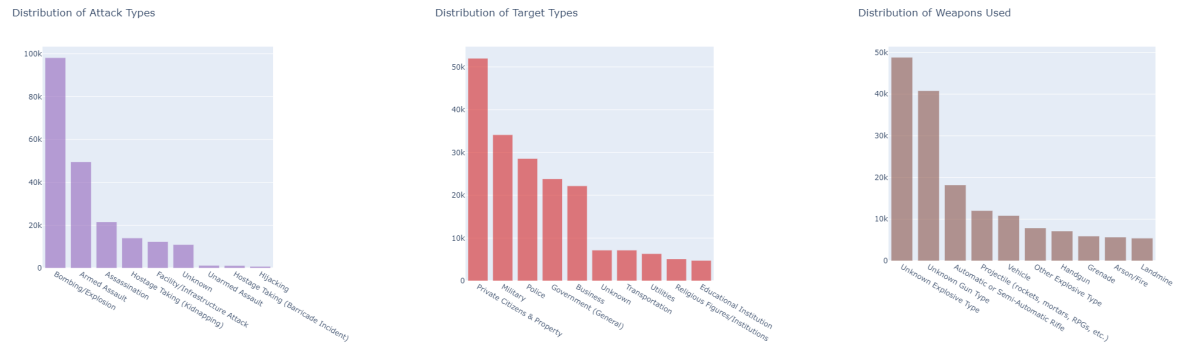
Frequency and percentage visualized in Figures 4.2a, 4.2b, and 4.2c.

4.3 PERPETRATORS

Top 10 most active terrorist groups in Table 4.1.

4.4 CASUALTIES AND FATALITIES

- Total: **479,348**
- Average per incident: **2.43**
- Distribution: **Figures 4.3a, 4.3b, 4.3c**

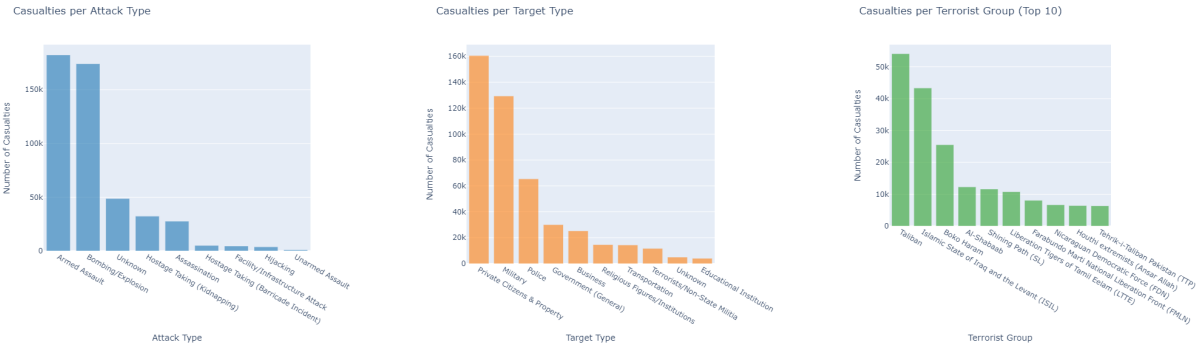


(A) Attack types (B) Target types (C) Weapon types

FIGURE 4.2
Frequency and percentage of attack types, target types, and weapon types

Terrorist Group	Number of Incidents
Taliban	11982
Islamic State of Iraq and the Levant (ISIL)	7254
Shining Path (SL)	4564
Al-Shabaab	4419
New People's Army (NPA)	3395
Farabundo Marti National Liberation Front (FMLN)	3351
Boko Haram	3320
Houthi extremists (Ansar Allah)	3196
Irish Republican Army (IRA)	2670

TABLE 4.1
Top 10 most active terrorist groups

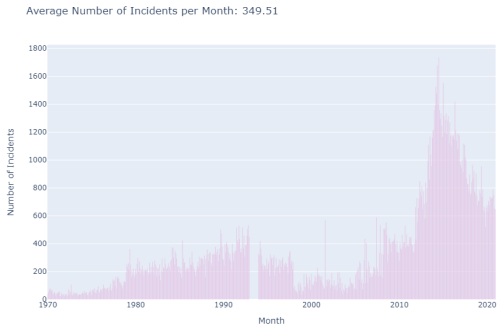


(A) Casualties per attack types (B) Casualties per target types (C) Casualties per group

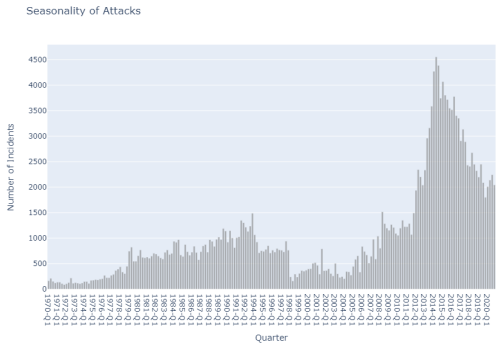
FIGURE 4.3
Distribution of casualties and fatalities by attack type, target type, and terrorist group

4.5 TEMPORAL PATTERNS

- Average incidents per month: **349.51** (Figure 4.4a)
- Seasonality: Analyze quarterly incidents data (average: **1048.53**) to discuss patterns (Figure 4.4b).



(A) Number of incidents per month



(B) Number of incidents per quarter

FIGURE 4.4
Temporal patterns of terrorist incidents

4.6 ADDITIONAL VISUALIZATIONS

To enhance EDA, we can consider additional visualizations:

- **Geographic Distribution**: Map incident locations to identify hotspots or regions with high terrorist activity.
- **Success Rates**: Visualize effectiveness of terrorist groups or vulnerability of target types to identify areas requiring increased security.
- **Variable Relationships**: Investigate correlations between attack types, target types, and terrorist groups to reveal hidden patterns or connections.

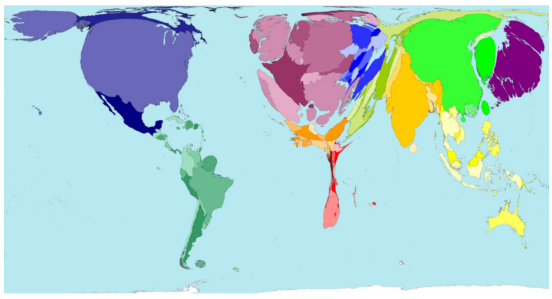
CHAPTER 5

RELATED WORK

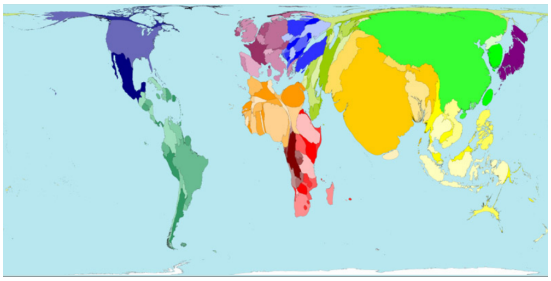
The GTD™ is a reference in the study of terrorism. It provides valuable data crucial for conducting research in the field. As such it has been used by many scholars to study terrorism in a extensive and unbiased manner. The target demographic of this database, however, is not limited to scholars as it has also been used by students in the past for various projects.

The originality of our approach lies in the simplicity of the interface we strive to provide. We aim at displaying information that does not require advanced statistical or prior geopolitical knowledge. Vast amounts of data, like the ones in the GTD™ , can quickly become complicated to comprehend due to the sheer amount of information that they contain.

One of our main sources of inspiration is a website [1] we found while browsing data visualization forums. The website presented different maps (seen below) where countries are sized proportionally to either the population's wealth, or to the number of citizens. These visualizations inspired us to look for more (and other) ways to represent big datasets on the world map, while keeping in mind the form of the GTD™ data.



(A) Map of the world with countries' sizes proportional to their population's wealth



(B) Map of the world with countries' sizes proportional to their number of citizens

FIGURE 5.1
Sources of inspiration

[1] http://mappemonde-archive.mgm.fr/num17/articles/art08105.html

BIBLIOGRAPHY

START (2022a). *Using GTD - Data Collection Methodology*. University of Maryland. URL: https://www.start.umd.edu/gtd/using-gtd/ (visited on 4th Apr. 2023).
– (2022b). *About GTD - Overview*. University of Maryland. URL: https://www.start.umd.edu/gtd/about/ (visited on 4th Apr. 2023).
– (2021). *GTD Codebook*. University of Maryland. URL: https://www.start.umd.edu/gtd/downloads/Codebook.pdf (visited on 7th Apr. 2023).