

# Proposal: Analysis of Global Terrorist Attacks

DATA 450 Capstone

Shreehar Joshi

2/5/23

## 1 Introduction

Global terrorism has been a major concern for countries across the world for many decades. Terrorist attacks are defined as acts of violence committed by individuals or groups with the intention of causing harm and destruction to a large number of people. These attacks can take many forms, including bombings, mass shootings, hijackings, and hostage situations, and can occur in any location, from urban centers to remote rural areas.

Terrorist attacks have a profound impact on both the victims and the larger society. They cause physical harm and loss of life, as well as emotional trauma and psychological distress. They can also have long-lasting socio-economic consequences, disrupting trade and commerce, causing job losses, and decreasing investor confidence.

In recent years, some of the most devastating terrorist attacks have occurred in cities across the globe, such as Paris, Madrid, London, and New York. These attacks have resulted in the loss of thousands of lives and have had a profound impact on the global community. In response to these attacks, many countries have increased their efforts to prevent and counteract terrorism, including by strengthening their security measures, enhancing their intelligence gathering and analysis capabilities, and working together with international partners to share information and coordinate efforts.

Despite these efforts, the threat of terrorism continues to persist, with new groups emerging and existing ones adapting their tactics. As the world becomes more interconnected, the risk of terrorism spreading across borders increases, making it all the more important for countries to work together to address this global challenge.

## 2 Dataset

In this project, three different datasets will be used to analyse the trends of terrorist activities and their impact on social and economic health of different countries that are plagued the most by such activities.

The first dataset - [Global Terrorism Database \(GTD\)](#) - is an open-source database including information on more than 180,000 terrorist attacks around the world from 1970 through 2017. The database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism ([START](#)), headquartered at the University of Maryland. Prepared by reviewing more than 4,000,000 news articles and 25,000 news sources, it is the most comprehensive unclassified database on terrorist attacks in the world. The variables that will be use in the analysis from this dataset are as follows:

- iyear: The year in which the incident occurred
- imonth: The month in which the incident occurred
- iday: The day in which the incident occurred
- country: The country code in which the incident occurred
- country\_txt: The country name in which the incident occurred
- region\_txt: The geographic region in which the incident occurred
- city: The city in which the incident occurred
- latitude: The latitude of the city in which the incident occurred
- longitude: The longitude of the city in which the incident occurred
- suicide: A flag indicating if the incident was a suicide attack
- attack\_type1\_txt: The general method of attack and broad class of tactic used
- targtype1\_txt: The general type of the target/victim
- gname: The name of the group that carried out the attack
- weaptype1\_txt: The general type of the weapon used in the incident
- nkill: The total number of confirmed fatalities for the incident

Similarly, the second dataset - [World Population](#) - is an open-source dataset in kaggle consisting of different information of the world population from 1955 to 2020. It was created by Nguen Thi Cam Lai using the [Countries in the world by population](#) dataset from the Worldometer. The variables that will be used from this dataset are as follows:

- Year: The year for which the population data is recorded
- Population: The total population of a country
- Migrants(net): The net number of migrants in a country
- FertlityRate: The fertility rate of a specific country

The third dataset - [World, Region, Country GDP/GDP per capita](#) - is an open-source dataset in kaggle consisting of the Gross Domestic Product of different different countries of the world from 1960 to 2021. It was created by Todor Mishinev using the [World Bank national accounts data](#). The variables that will be used from this dataset are as follows:

- Country Name: The name of the country
- Country Code: The three letter codes for the countries
- year: The year for which the GDP data is recorded
- GDP\_per\_capita: GDP per capita for a specific country in the given year

### 3 Data Acquisition and Processing

[In this section, if applicable, describe how you will obtain the data (if it's anything more complicated than a simple download). Discuss what data processing steps will be needed, such as recoding variables, data cleaning, data tidying, imputing missing values, etc. See sections 1c, 1d, 1e in the “Good Enough Practices” paper.]

### 4 Research Questions and Methodology

[In this section, list each of the questions you will explore. Following each question, provide a detailed and specific plan for how you plan to answer the question. Include the specific steps you will take, what form the answer will take (a number? table? visualization? model? Give all the specifics), and estimate how many hours each question will take to complete.]

1. Is smoking correlated with diabetes? To answer this, I will create a filled bar plot, with the left bar representing non-smokers, the middle bar representing people who smoke moderately, and the right bar representing heavy smokers. The bars will be the same height, and each bar will be colored two colors based on the proportion of patients in the group who do or do not have diabetes.
2. Question 2? Plan for question 2.
3. Question 3? Plan for question 3.
4. etc.

### 5 Work plan

[Fill in the list below with a plan for what you will do each week. You should have around 7 hours worth of work each week. Writing work counts. Several tasks have already been filled in for you.]

**Week 4 (2/6 - 2/12):** [Just an example:

- Data tidying and recoding (4 hours)
- Question 2 (4 hours).]

**Week 5 (2/13 - 2/19):**

**Week 6 (2/20 - 2/26):**

**Week 7 (2/27 - 3/5):**

- Presentation prep and practice (4 hours)

**Week 8 (3/6 - 3/12):** *Presentations in class on Thurs 3/9.*

- Presentation peer review (1.5 hours)

**Week 9 (3/20 - 3/26):**

- Poster prep (4 hours)

**Week 10 (3/27 - 4/2):** *Poster Draft 1 due Monday 3/27. Peer feedback due Thursday 3/30.*

- Peer feedback (2.5 hours)
- Poster revisions (2 hours)

**Week 11 (4/3 - 4/9):** *Poster Draft 2 due Monday 4/3. Final Poster due Saturday 4/8.*

- Poster revisions (1 hour).

**Week 12 (4/10 - 4/16):**

**Week 13 (4/17 - 4/23):** [All project work should be done by the end of this week. The remaining time will be used for writing up and presenting your results.]

- Draft blog post (5 hours).

**Week 14 (4/24 - 4/30):** *Blog post draft 1 due Monday 4/24. Peer feedback due Thursday 4/27. Blog post draft 2 due Sunday 4/30.*

- Peer feedback (2.5 hours)
- Blog post revisions (2 hours)
- [Do not schedule any other tasks for this week.]

**Week 15 (5/1 - 5/7):** *Final blog post due Tuesday 5/2.*

- Final presentation prep and practice.
- [Do not schedule any other tasks for this week.]

**Final Exam Week (5/8):** *Final Presentations during final exam slot, Monday May 9th 3:20-6:40pm.* [Do not schedule any other tasks for this week.]

## 5.1 Some cool Quarto stuff

[You can delete this section from your proposal.]

For your reference, here's an example of a Python code cell in Quarto, along with a figure that gets generated, along with a caption and a label so that it can be referred to automatically as “Figure 1” (or whatever) in the writeup.

For a demonstration of a line plot on a polar axis, see Figure 1.

```
import numpy as np
import matplotlib.pyplot as plt

r = np.arange(0, 2, 0.01)
theta = 2 * np.pi * r
fig, ax = plt.subplots(
    subplot_kw = {'projection': 'polar'}
)
ax.plot(theta, r)
ax.set_rticks([0.5, 1, 1.5, 2])
ax.grid(True)
plt.show()
```

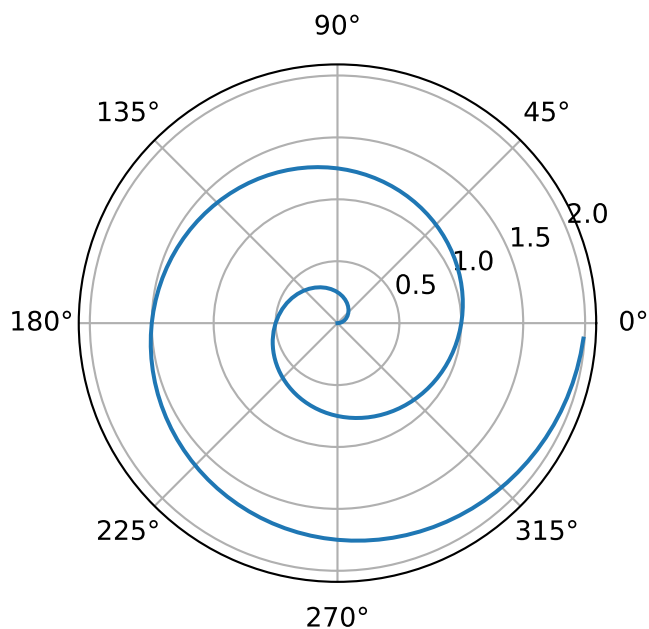


Figure 1: A line plot on a polar axis

Here's an example of citing a source (see Phillips 1999, 33–35). Be sure the source information is entered in “BibTeX” form in the `references.bib` file.

## 6 References

[The bibliography will automatically get generated. Any sources you cite in the document will be included. Other entries in the `.bib` file will not be included.]

Phillips, T. P. 1999. “Possible Influence of the Magnetosphere on American History.” *J. Oddball Res.* 98: 1000–1003.