Project 1: Global Terrorism Analysis

Yandi Farinango, Patrick Furman, Matt Hyatt, Corinne Steuk

Introduction

Global terrorism by radical groups has become a point of concern to the public in recent years. With events in recent history to spark unification and uprising, certain terrorist groups have found the opportunity to attack their targets. It is important to become aware of the tactics and tendencies of groups to equip countries to prevent future incidents. Our project utilizes a data set from Kaggle that has entries for documented terrorist attacks from 1970 through 2017. We will use HBase and Phoenix applications to query answers to a variety of questions involving terrorist attack hotspots, targets, time of year, and method of attack. With this information, we hope to find particularly dangerous groups and trends of their attacks for the safety of people moving forward.

Data Description

The original data set from Kaggle has 135 features and over one hundred and eighty thousand examples. We kept 22 of the most descriptive features including date, location (country, state, and city), terrorist group information, weapon details, target details, number of people wounded, and killed. The columns removed had mostly null values and did not pertain to our study parameters. Each example is one terrorist attack denoted with a unique12-digit number but, we reindexed the values (starting with 1) for simplicity. This data set includes all necessary information for our project, so additional data aggregation was not required. We were then able to map this into a HBase table to perform Phoenix SQL queries to address our questions.

Team Tasks

Each team member was assisted in different parts of the data pipeline process. Corinne cleaned and moved the data into the HDFS as well as creating the table in the HBase system and Phoenix. Additionally, she wrote the “Introduction” and “Data Description” portions of the report. All members wrote queries to answer particular questions of interest and Corinne was tasked with executing them and exporting the output CSV files to the computer file system from the local host root directory. Matt took care of creating visuals for half of the queries and some analyzations. Patrick created involved queries, their corresponding visualizations and analysis, and the “Conclusions & Future Work.” Finally, Yandi created many visualizations with their results for both his and Corinne’s queries.

Scripts & Queries

Moving File to HDFS

```
>scp -P 2222 GlobalTerrorism.csv root@127.0.0.1:~/proj.csv
>ssh root@127.0.0.1 -p 2222
>sed -I ‘1d’ proj.csv
>hadoop fs -put ~/proj.csv /tmp
Creating HBase Table
>su hbase
>hbase shell
>create ‘p’, ‘t’
>exit
>su root
```

Populating HBase Table

```
>hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.separator=, -
Dimporttsv.columns="\
>HBASE_ROW_KEY, t:id, t:iyear, t:imonth, t:iday, t:country, t:provstate, t:city, t:success, t:suicide,
t:attacktype, t:targ_type, t:target, t:targ_nat, t:group, t:weap, t:weap2, t:nkill, t:nkillus, t:nkillter,
t:nwound, t:nwoundus, t:nwoundte" p /tmp/proj.csv
```

Creating Phoenix Table View

```
>cd /usr/hdp/current/phoenix-client/bin
>.sqlline.py localhost
>create view "p" ("row" VARCHAR primary key, "t"."id" VARCHAR, "t"."iyear" VARCHAR,
"t"."imonth" VARCHAR, "t"."iday" VARCHAR, "t"."country" VARCHAR, "t"."provstate"
VARCHAR, "t"."city" VARCHAR, "t"."success" VARCHAR, "t"."suicide " VARCHAR,
"t"."attacktype" VARCHAR, "t"."targ_type" VARCHAR, "t"."target" VARCHAR, "t"."targ_nat"
VARCHAR, "t"."group" VARCHAR, "t"."weap" VARCHAR, "t"."weap2" VARCHAR, "t"."nkill"
VARCHAR, "t"."nkillus" VARCHAR, "t"."nkillter " VARCHAR, "t"."nwound" VARCHAR,
"t"."nwoundus" VARCHAR);
```

Moving Output File from Root to Computer File System

```
>scp -P 2222 root@127.0.0.1:~/corinne1.csv
```

Table 1: Queries

File Name	*Description*	*Query*
Corinne1.csv	Top 10 countries with the most successful terrorist attacks.	select "country" as Country , SUM(TO_NUMBER("success")) as TotalSuccess from "p" group by "country" order by TotalSuccess desc limit 10;
Corinne2.csv	All the Countries included in this dataset (204).	select distinct "country" as Country from "p";
Corinne3.csv	Which attacks performed by the Taliban with the country in which the attack occurred, the year, and if it was a success.	select "iyear" as Year, "country" as Country, TO_NUMBER("success") as Success from "p" WHERE "group" = "Taliban" group by Country, Year;
Corinne4.csv	The different attack types and how they changed through the years.	select "iyear" as Year, "attacktype" as AttackMethod, "country" as Country from "p" group by Year, AttackMethod, Country;
Yandi1.csv	Top 10 terrorist organizations with most successful attacks and which nationality they were targeting.	select "targ_nat", "target", SUM(TO_NUMBER("success")) as Success from "p" group by "targ_nat", "target" order by Success desc limit 10;
Yandi2.csv	Distinct weapons used throughout entire data set.	select distinct "weap" from "p";
Yandi3.csv	Top 10 weapons used in successful terrorist attacks	select "weap", SUM(TO_NUMBER("success")), as SuccessfulWeap from "p" group by "weap"

		order by SuccessfulWeap desc limit 10;
Yandi4.csv	Top 10 deadliest weapons used	select "weap", SUM(TO_NUMBER(nkill)) as DeadliestWeap from "p" group by "weap" order by tot desc limit 10;
Yandi5.csv	Top 10 type of groups successfully targeted by terrorist attacks	select "attacktype", SUM(TO_NUMBER("success")) as SuccessfulAttack from "p" group by "attacktype order by desc limit 10;
Matt1.csv	Total number of deaths per year.	select "iyear" as Year, COUNT(TO_NUMBER("nkill")) as Total from "p" group by Year;
Patrick1.csv	Top 10 successful terrorist groups and their average kills	select "tar_nat", TO_NUMBER("success") as Kill, avg(TO_NUMBER("nkillter")) as KillTer from "p" group by "targ_nat" order by Kill desc limit 10;
Patrick2.csv	Top 10 terrorist groups and weapons with their total number of successful attacks	select "targ_nat", "weap", SUM(TO_NUMBER("success")) as S from "p" group by "targ_nat", "weap" order by S desc limit 10;
Patrick3.csv	The average number of suicides for both successful and unsuccessful terrorist attacks	select AVG(TO_NUMBER("suicide")) as Suic, TO_SUMBER("Success") from "p" group by Suic;

Results

Query	**Analysis/Commentary**
Total Number of Attacks by Group: The 10 terrorist groups with the highest number of successful attacks as well as their most commonly used weapon type	The most surprising finding from this query was that the majority (8 out of 10) of the most active terrorist organizations are actually unknown which indicates a significant gap in either the intelligence gathering surrounding these attacks or in reporting to the data source we used. Also noteworthy is that the groups that have the most successful attacks seem to favor explosive weapons. Guns are only slightly less popular with these groups, but for planning purposes, it's valuable to know what types of weapons the most successful groups rely on.
Average Rate of Suicides by Attack Success: The proportion of successful and unsuccessful attacks that result in the perpetrator committing suicide.	We found that, at least in our data, there were no occurrences of the perpetrator committing suicide when the attack was unsuccessful and that in successful attacks the rate was also fairly low, just 6.67%. One possible explanation for the difference between suicide rates in successful vs unsuccessful attacks is that unsuccessful attacks are probably also more likely to end in arrest rather than violence leaving both less opportunity and motivation for suicide.
Average Casualties: The average total number of casualties (perpetrator and victims) in successful attacks for the 10 most lethal terrorist	In most successful attacks, we found that the average number of reported casualties was still quite low for almost every group. There was however one major outlier with the "Abdullah Azzam Brigades" which had an average of 159 casualties per attack which is more than 11 times the average of the next most lethal group.

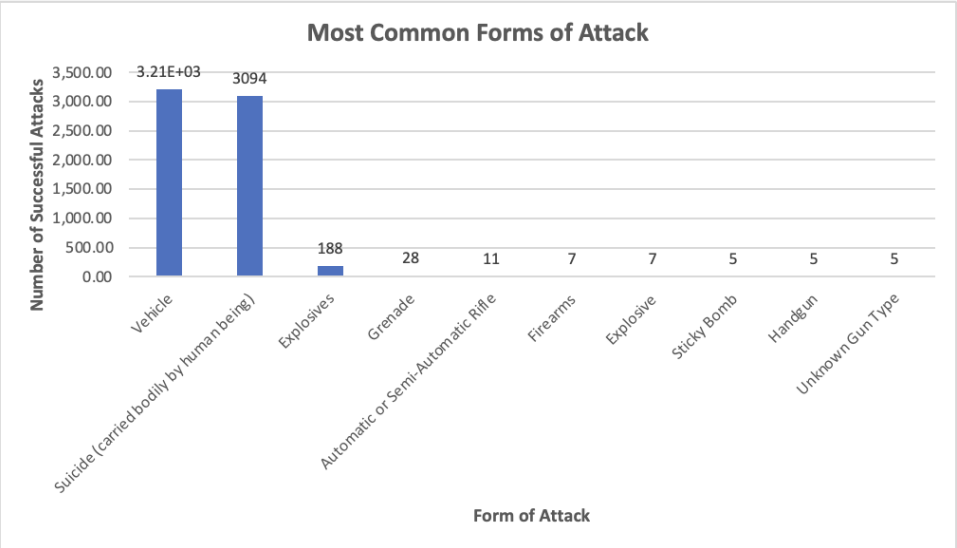
groups	This is likely symptomatic of a small sample size but would be worthy of further investigation.
Most Targeted Countries & Terrorist Organizations Claiming Responsibility for Attacks The top 10 most successful terrorist attacks organized by country and organization responsible for attacks.	We found the countries most affected by terrorist attacks to be Middle Eastern countries. Countries including Iraq, Afghanistan, and Nigeria were most affected by terrorist attacks – by number of successfully carried out attacks. The groups responsible for these attacks include Islamic State of Iraq and the Levant (ISIL), The Taliban, and Boko Haram.
Weapons Used List of distinct weapons used to carry out attacks	We found that there have been over 30 distinct types of weapons used by terrorist organizations for carrying out attacks.
Weapons Used to Execute Successful Attacks 10 distinct weapons most commonly used to carry out successful terrorist attacks	We found 10 distinct weapons most used during successfully carried out terrorist attacks. We found most successful attacks were carried out using vehicles, suicide by carried explosives, and explosives.
Deadliest Weapons List of weapons used ordered by resulting casualties	We found the 10 most deadly weapons used during successfully carried out attacks. Firearms were the deadliest weapons used claiming the lives of 10,253 people.
Most Targeted Organizations Organizations most targeted by terrorist organizations	We found the organizations that were most targeted by terrorist attacks. We found most attacks were against Military, Police, and Private Citizens.

Conclusions & Future Work

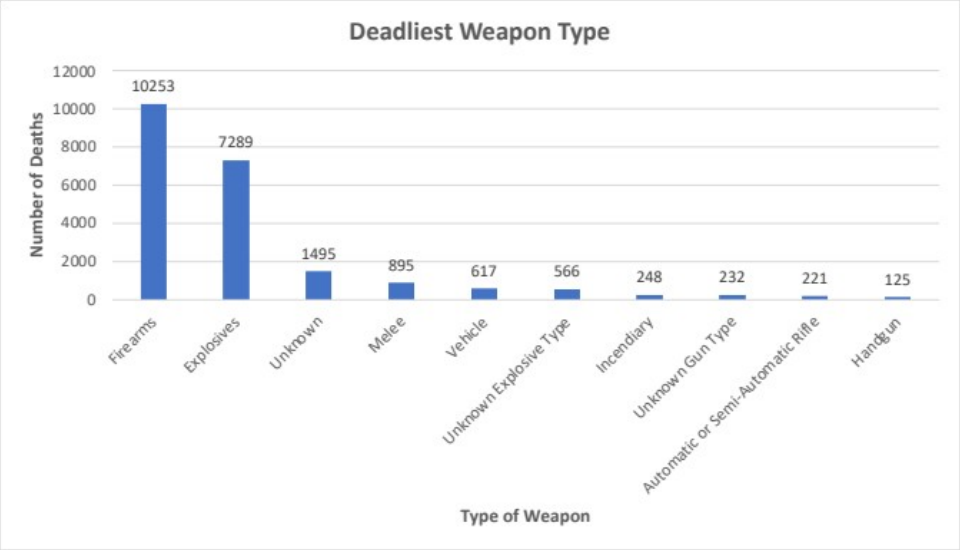
Our analysis primarily focuses on the terrorist organizations, their methods and targets of attack, and how often they succeeded. The goal of this analysis to provide a general idea of how terrorist attacks are evolving over time and how they might trend in the near future. This provides a good overview of how terrorism has looked on a global scale in the past 50 years but to provide more meaningful insights to any individual group, further analysis could be done that focuses on either an individual terrorist organization or target country of interest. For instance, in the case of the United States, it may be valuable to focus more closely on terrorist events surrounding 9/11 to study the effectiveness of efforts such as the War on Terror. Additionally, work could be done to create predictive models to assist in contingency preparations for terrorist attacks. One example of this could be a model predicting the potential damage and casualties based on factors such as location, type of weapon used, the type of target, and recent trends using the historical data provided in the dataset. Such models would allow for law enforcement to more efficiently allocate resources in order to better respond to potential future terrorist attacks in the event that they cannot be prevented.

Dataset: https://www.kaggle.com/START-UMD/gtd
Graph scripts: https://github.com/mhyatt000/comp358/blob/main/proj1/main.py

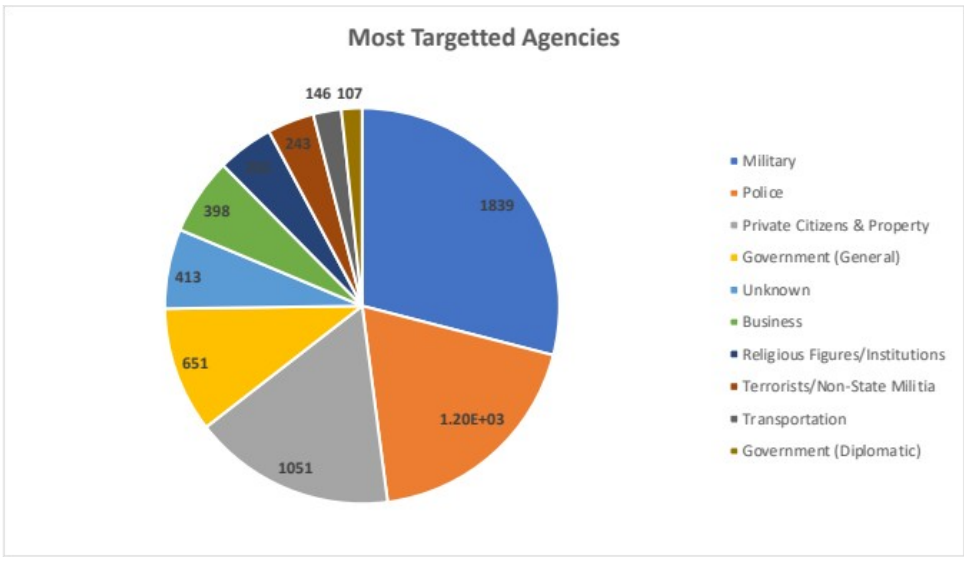
Appendix – Results



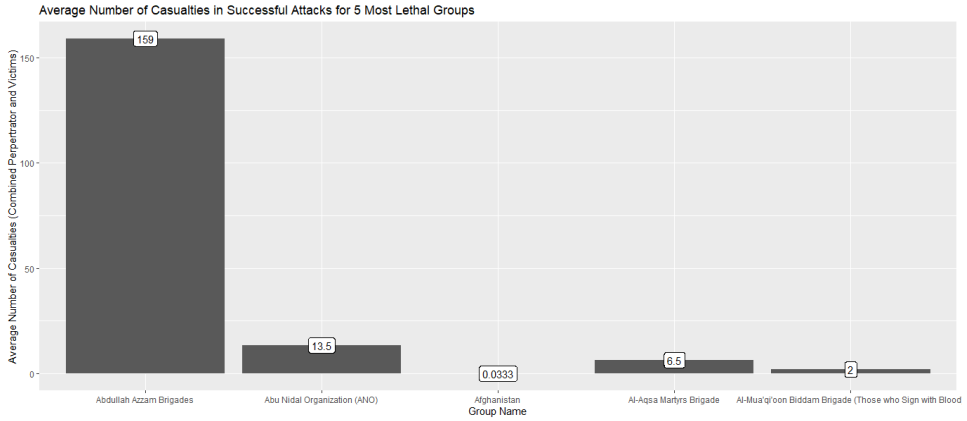
Most Common Form of Attack (yandi3.csv)



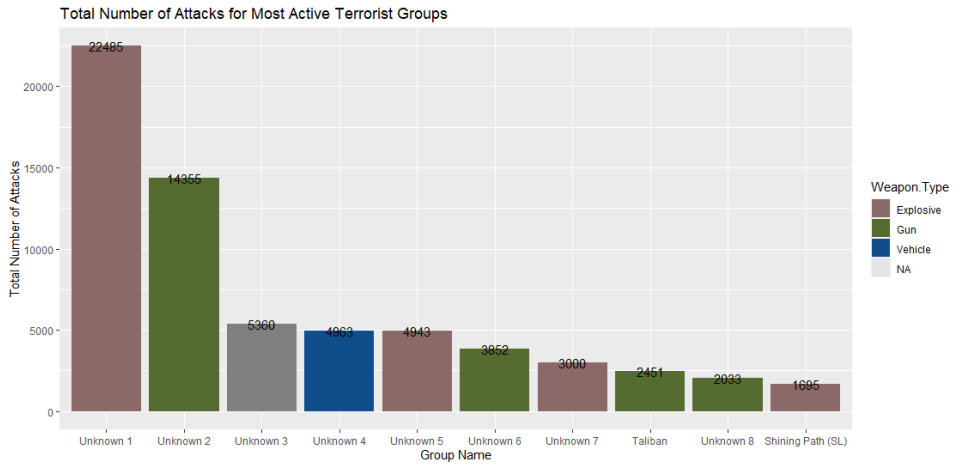
Deadliest Weapon (yandi4.csv)



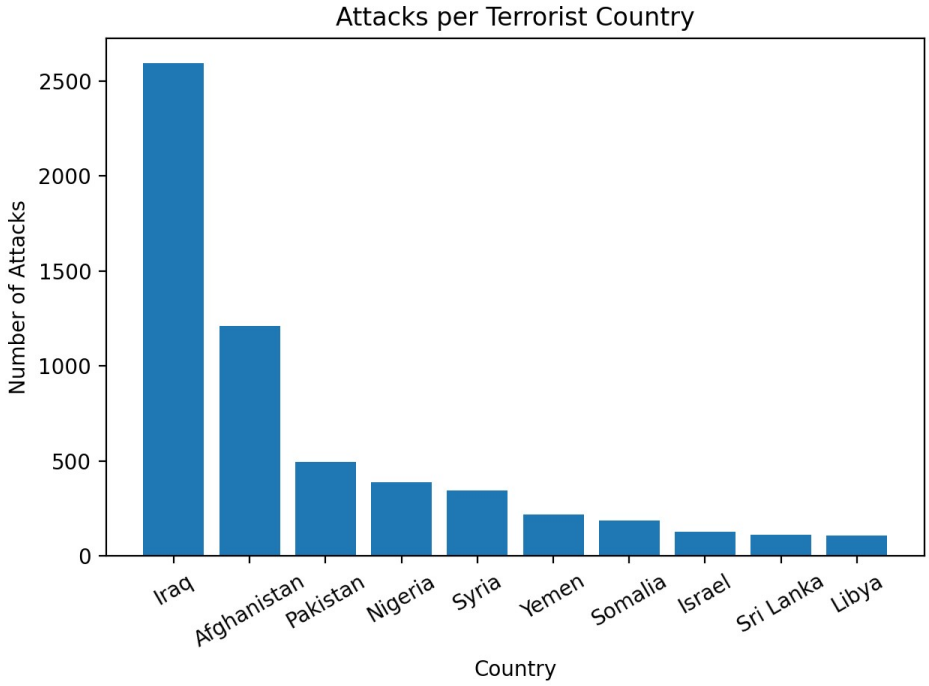
Most Targeted Agency (yandi5.csv)

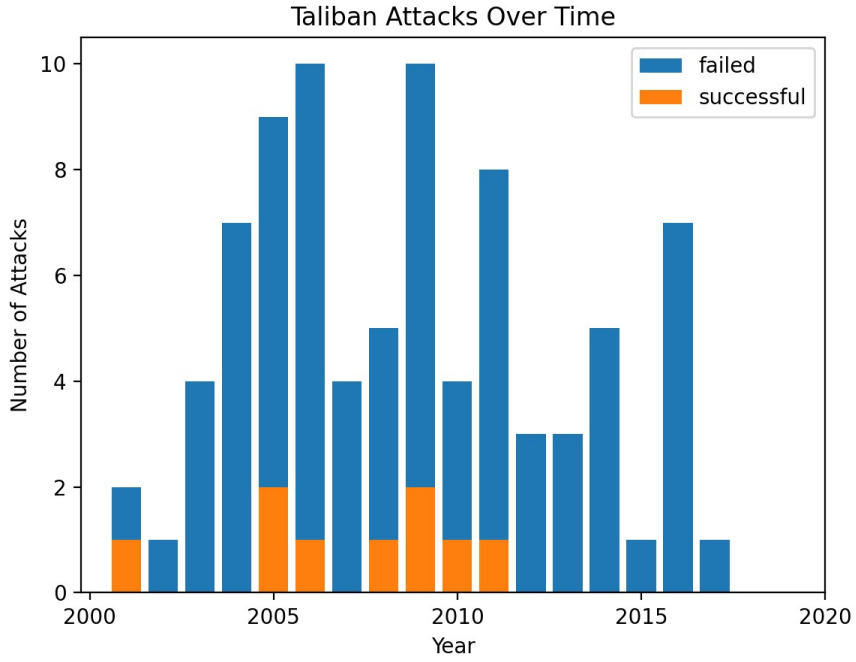


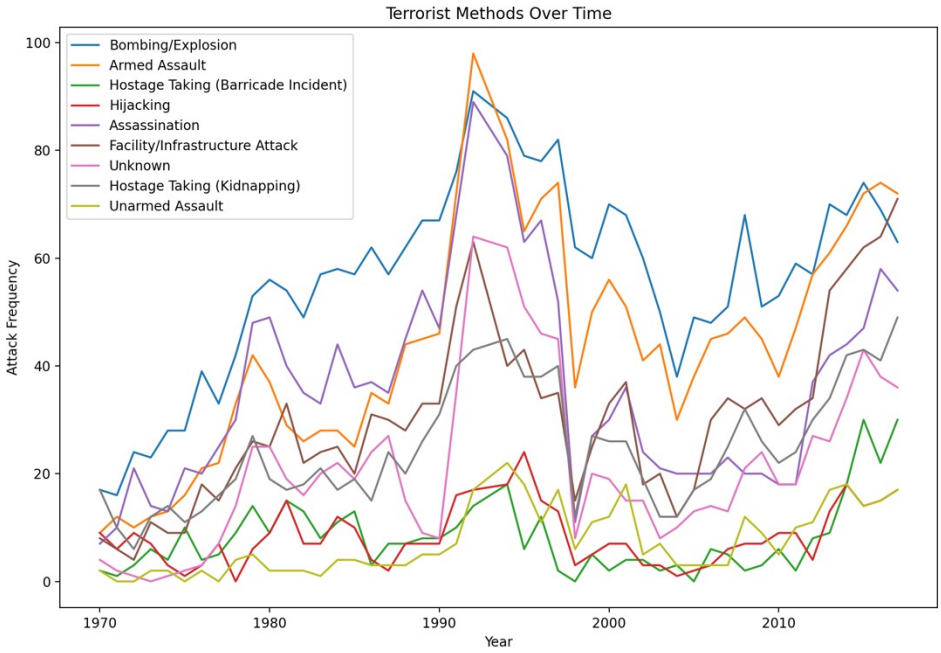
Average Casualties (patrick1.csv)

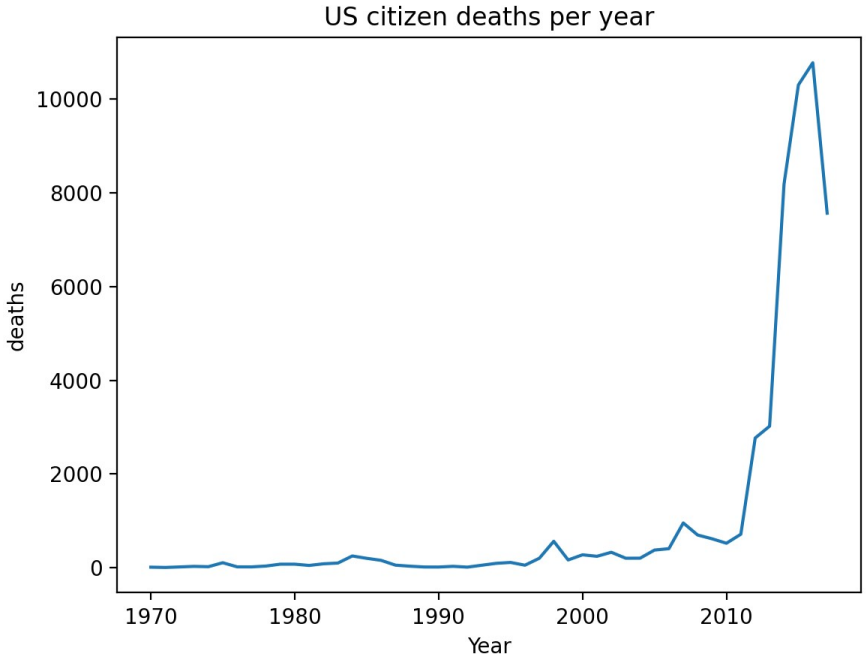


Total Number of Attacks by Group (patrick2.csv)


Attacks per Terrorist Country
Number of Attacks
2500
2000
1500
1000
500
0
Iraq
Afghanistan
Pakistan
Nigeria
Syria
Yemen
Somalia
Israel
Sri Lanka
Libya
Country


Taliban Attacks Over Time
failed
successful
Number of Attacks
10
8
6
4
2
0
2000
2005
2010
2015
2020
Year


Terrorist Methods Over Time
Bombing/Explosion
Armed Assault
Hostage Taking (Barricade Incident)
Hijacking
Assassination
Facility/Infrastructure Attack
Unknown
Hostage Taking (Kidnapping)
Unarmed Assault
Attack Frequency
Year
0
20
40
60
80
100
1970
1980
1990
2000
2010


US citizen deaths per year
deaths
Year
0
2000
4000
6000
8000
10000
1970
1980
1990
2000
2010