PROCESS BOOK

PART 1 :- PROPOSAL

Basic Info

Title - Global Terrorism Visualization

Name	Email Address
Akshay Khatwani	Akshay.khatwani@utah.edu
Tejas Hirekatur Sreedhar	tejashs@cs.utah.edu
Giorgi Kvernadze	u0950369@utah.edu

Link https://github.com/tejashs/GlobalTerrorismVisualization

Background and Motivation

16 years after the horrifying attack on the twin towers, the fight against terrorism continues and this global menace lingers to plague the world. Many questions still remain unanswered. We try to answer some of these questions through our visualizations using some interesting correlations and design choices.

Project Objectives:

1. To find if US involvement in the Middle East impacted Global Terrorism
2. To find if increase in population of a specific religion in a country increased terrorism in that country
3. To find if US involvement in the Middle East has caused increase in terrorism in that country
4. Correlating factors like GDP growth, literacy etc. with increase in terrorism and pinpointing what factors might have caused this.

Data

The Global Terrorism Database (GTD). The Dataset is available on Kaggle. This dataset contains information about worldwide terrorism events back to 1970, including location, circumstances, number of casualties, weapons used, etc.

The link to the dataset - **https://www.kaggle.com/START-UMD/gtd/data**

Data Processing

We plan to use Pandas to manipulate data to extract relevant features. This data is complex in terms of the number of missing values. We plan to deal with this by either predicting the missing values or eliminating data points with substantial missing information.

Visualization Design

This is described in Design.pdf

Must-Have Features

It includes the following,

- The World map and US map showing terror attacks.
- A line Graph with nodes for events representing US' involvement in Middle East showing increase or decrease in number of terror attacks after a major event.
- Selecting a node (e.g. US invades Iraq 2003) provides drill down of Heat map of attacks before and after 2003.
- A bar chart showing Global terror attacks based on the religion of perpetrator.
- Word cloud based on Weapons used, attack type etc.

Optional Features

It includes the following,

- Charts for correlating metrices like GDP, literacy rates, etc. with terrorism. This largely depends on different datasets related to our main dataset.
- An Alexa voice based interface to interact with the visualization using the Alexa API.

Project Schedule

We have 6 weeks to project submission, this is what we plan to do

Week1 - Data cleaning and feature extraction

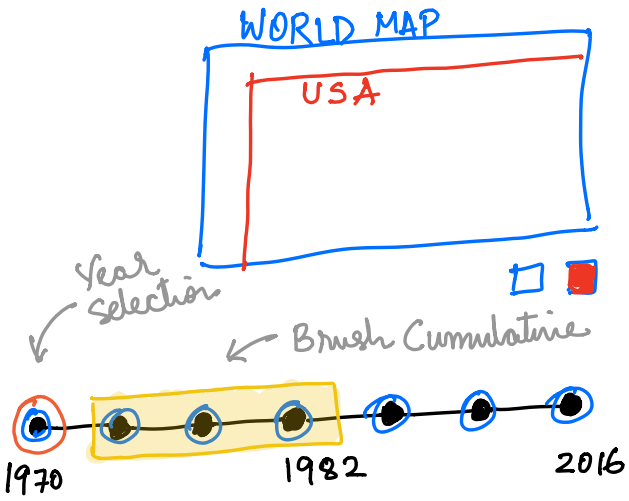
Week2 – Layout(Tejas), World Map(Giorgi) and US map(Akshay) along with brush(Tejas)

Week 3,4,5 - Line Chart with Nodes(Akshay), Heat Maps(Giorgi), Bar Charts(Tejas) with their respective interactions

Week 5,6 – Making All Interactions work together and taking care of styling, positioning to tie the whole thing together

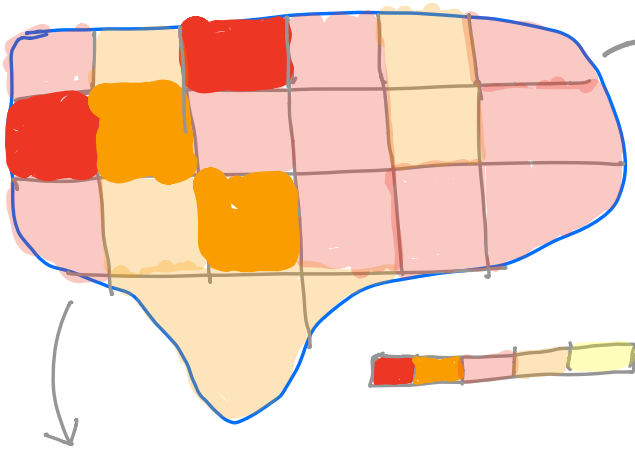
PART 2 :- PROJECT DESIGN

1970 - 2016



INTERACTION

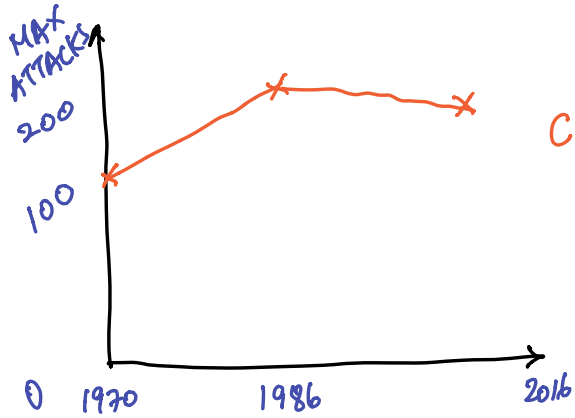
* Clicking on legend buttons Changes Map from World Map to USA.
* Clicking on Countries in World Map and clicking on States in USA Map offers drill down



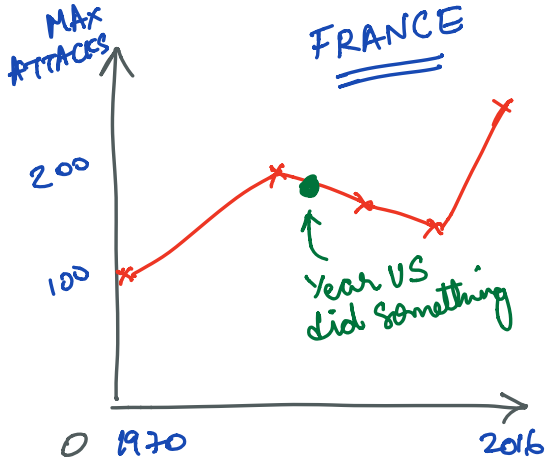
WILL BE GRADIENT (DIDNT HAVE SO MANY COLORS)

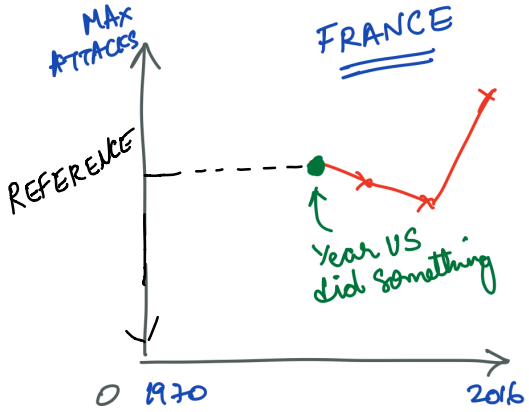
World Map

INTERACTABLE

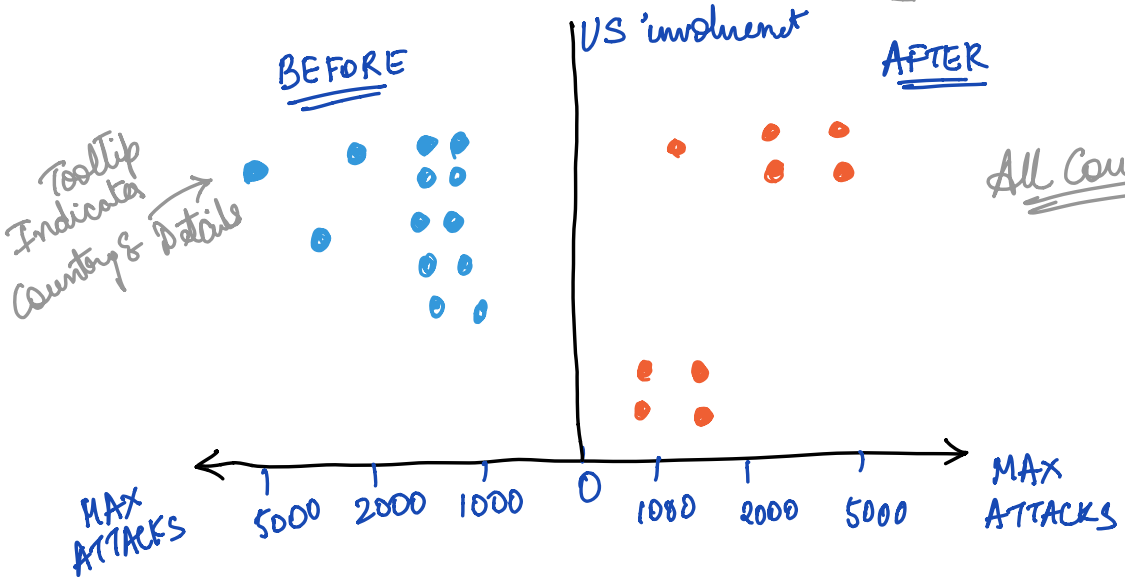


CALIFORNIA



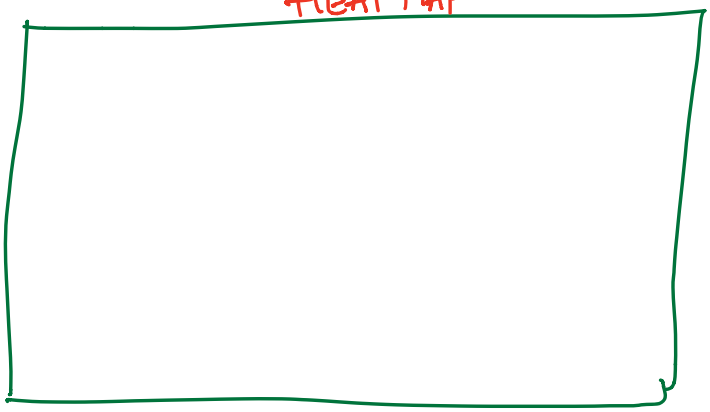


⇒ Did US involvement affect the number of terror attacks in the Country.



WORLD MAP (BEFORE 2003) → US Invade IRAQ

HEAT MAP



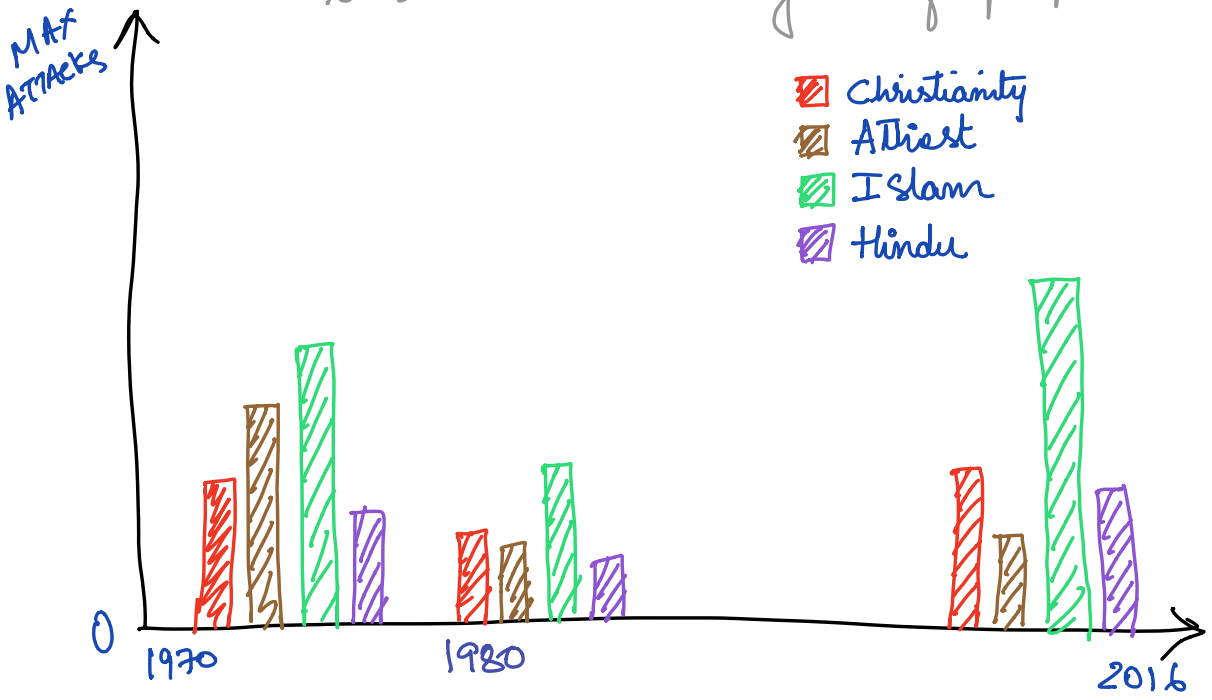
Shows terror attacks globally before 2003

WORLD MAP (AFTER 2003)

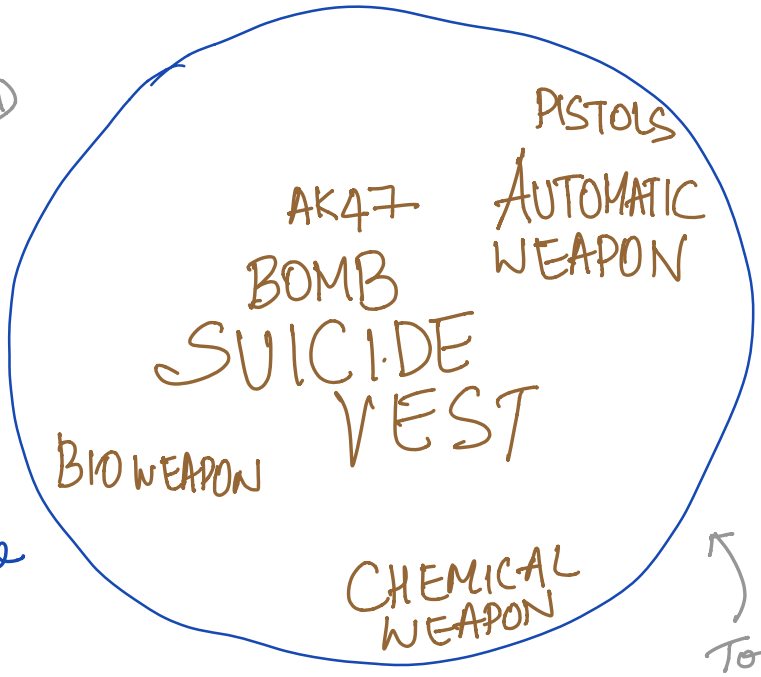
HEAT MAP

Shows terror attack globally after 2003

Global terror attacks per year based on the religion of perpetrator



↳ Use similar bar graphs / line charts for correlating other metrics like GDP, literacy rates of countries etc. (Atleast another 6-8 charts) i.e CASUALTIES etc.



WORD CLOUD

↓

Based on Weapons used, attack type etc.

↑

Tooltip for Countries with years

2-3 WORD CLOUDS

1] We are planning to do a lot of interaction for all the charts, maps and graphs.

2] Allow drill down for most of the Countries, States (US) for all the charts.

3] Brushing, filters, for charts wherever applicable.

PART 3 :- PEER REVIEW

- Differentiate terrorist attacks from mass shootings? (Mohammed)

Answer: We are currently planning to only visualize terrorist attacks. But if time permits we would be correlating data of terrorist attacks in the USA with the data of mass shootings in the USA.
The data set we could be using for mass shootings would be :

https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years

- Estimating religious affliation of victims using the country's population data would be inaccurate. An obvious way to address this would be to get specific details religious affliation of victims for various incidents/attacks (Ram)

Answer: The correlation between religious affiliation vs country's population is done so that we can find trends between the the religious affiliations of perpetrators and the religious affiliations of affected population.
We can draw trends corresponding to religious groups being targeted in a certain country for instance.

- In the "before" and "after" chart showing scatter plots of attacks, if we need to compare a particular country "before" and "after" US involvement, it doesn't seem to be convenient to do so. First, I need to hover over a point on the "before" side to see what country it represents and in order to check out the same country on the "after" side, based on the design shown, we would need to hover over each point to find that particular country we are looking for - this inconvenience is a side affect of using tooltips. You could link the points across so that hovering on one country in the "before" part highlights the associated point for the country in the "after" part (Mohammed)

Answer: The before and After US involvement in scatter plots is to distinguish between the number of attacks a country has had before and after a certain event. Tool tips are to indicate the number and details of the attack. The scatter plot graph is per country and not overall. The behavior indicated in the comment seems to be misunderstood.

- The color codes shown on the map seem to be based on the number of attacks, how about also considering the outcome of attacks (i.e. casualty count) for deciding the color code? (Ram)
Answer: Yes. The graphs will be color coded based on the number of attacks. It would be a gradient.
However there are graphs like heat map that would be based on other attributes as well.

- For a selected Country or state, depict the ranking of attackers based on appropriate metrics (Mohammed)
Answer: There is not enough data or appropriate metrics to "rank" attackers. The metrics used would be misleading.

- Details of specific (or major) attacks based on user selection of country/state on world/US map (Ram)
Answer: There would be summary of the attacks shown as tooltips. This would provide details and specifics of the attacks.

After Milestone Changes

Design Evolution

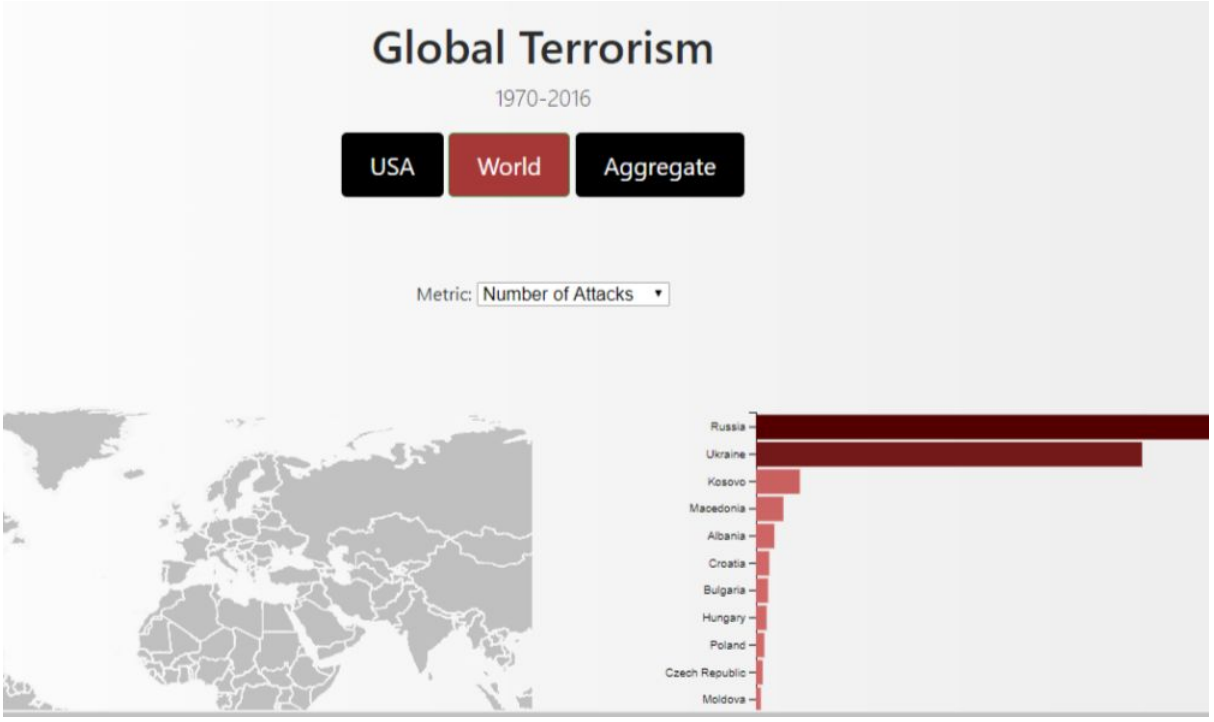
We considered An Event Based Design initially to be the core of our design.
The Data Processing turned out to be a really difficult task. The data was over 150mb with all kinds of data. We had a lot of things that we had to reject due to the non viability of the complexity and storytelling.
We deviated quite a bit from our proposal but in the end, what we show answers the more important questions like what is going on in a region. What may have caused it, who were the victims. Interestingly for US, 1970 was the peak year and surprisingly the chief perpetrators were anti- student scholarship group. Many such trends were picked and for each country a snapshot was displayed.
Over 125,00 incidents were present and one of the interesting things that was done was take a tf-idf of the summary of all the incidents and plot a word cloud.

All must have features were implemented except for the scatter plot but the place we changed the most was the questions that we were trying to answer.

Implementation

The GTD has information for over 125,000 total incidents. The amount of data that is there is immense and finding structure that would answer the questions we laid out was a hard task.

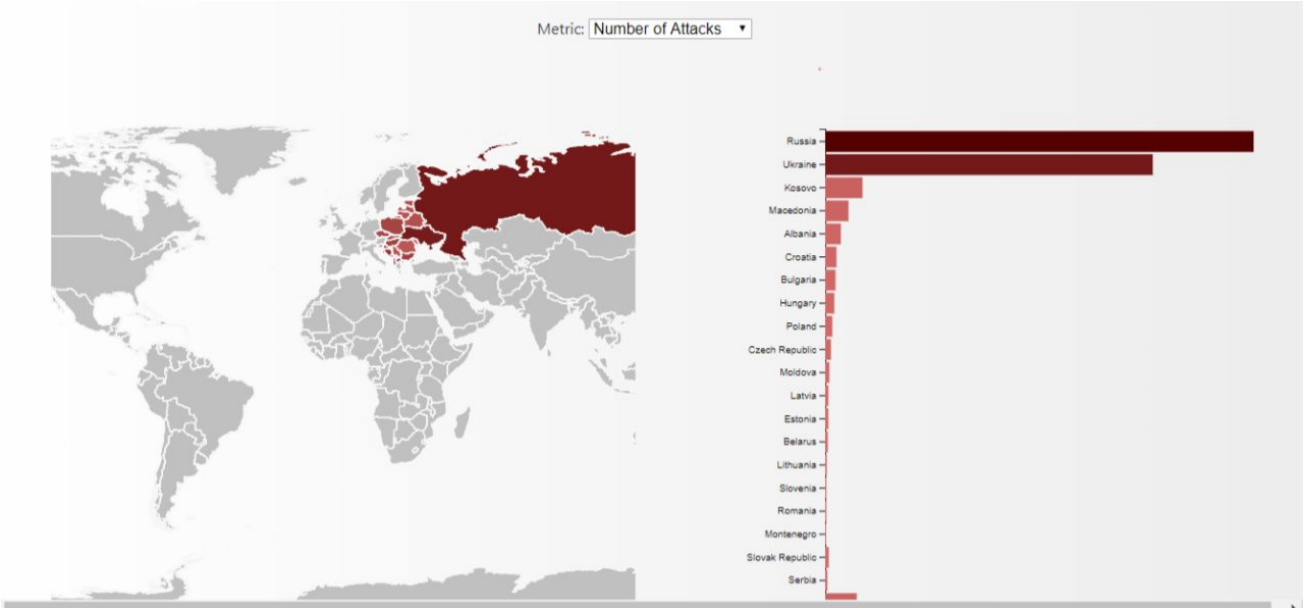
There are 2 tabs that we can toggle, USA and world.



First we focus on the world Tab.

The first thing that we did was to segregate the world map into regions that would be affected by similar attacks. For example, West Europe and East Europe were separated. In Asia,we aggregated over SouthEast Asia, Middle East, East Asia because of how diverse these regions are. This is an important point because of how the story unfolds

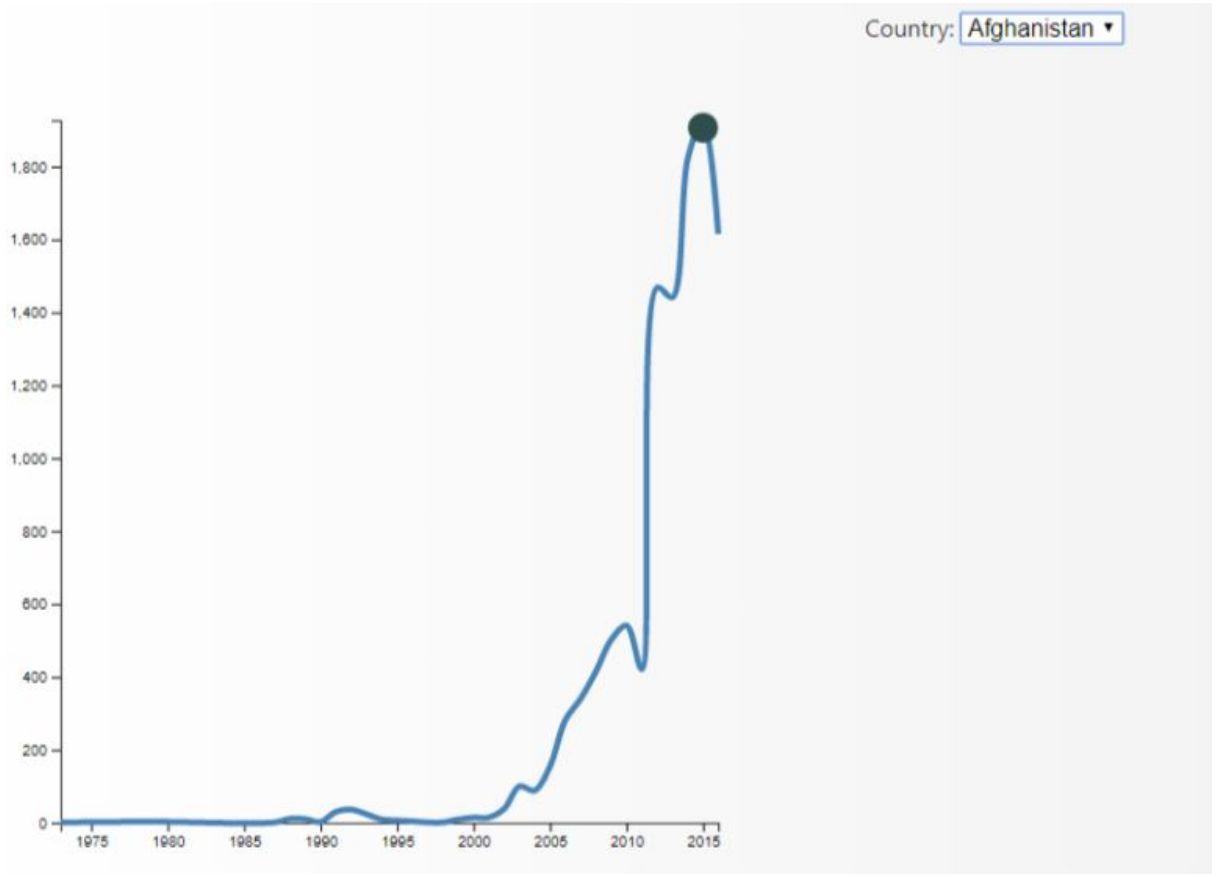
Hovering over a region, highlights the countries in that region by gradient of number of attacks.



As seen in the Image, On hovering over different regions, The region gets highlighted and the red gradient represents the no of attacks in that region.

A drop down menu to select “No of attacks”, “No of Fatalities” and “No of injuries” changes the bar chart showing a quick overview of what’s going on in the region.

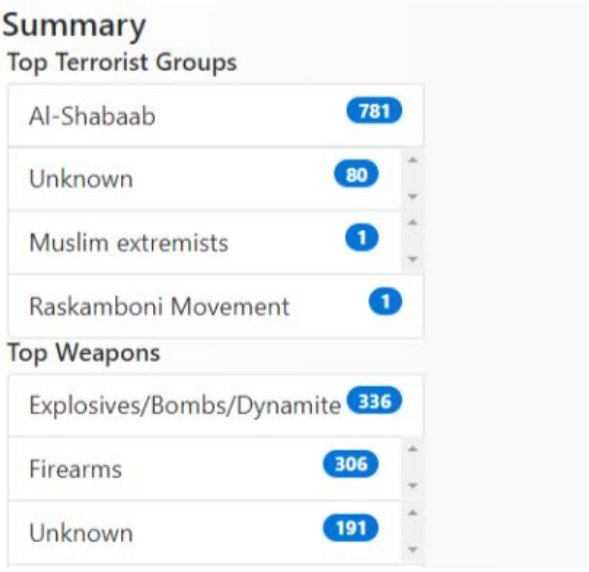
On Clicking on a region, A Line Chart with a drop down menu to select different countries in the region is updated.

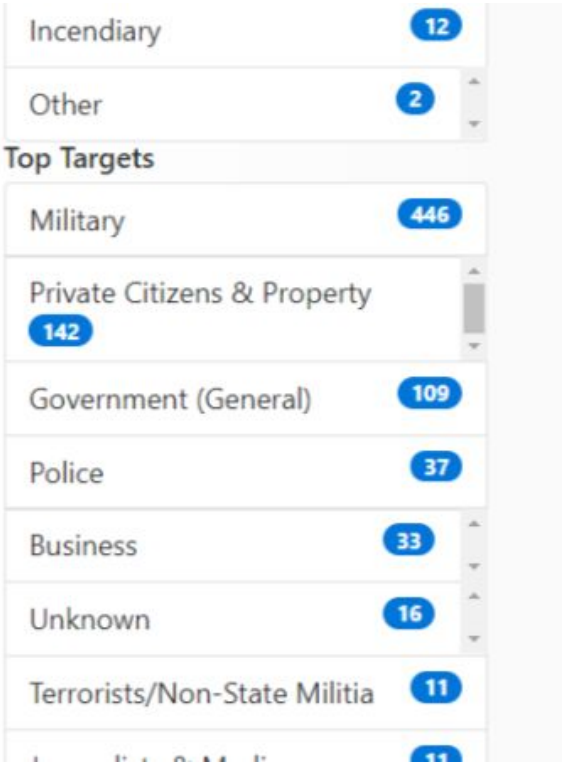


This gives us a trend about what has happened over the years in a particular country. There is a node on the line chart that is clickable to show what might have caused this increase in number of attacks.

On clicking the node the following are updated.

A word cloud for Somalia's Top year is shown above. This populates for every country's line graph on clicking the node signifying the year with the top attacks.



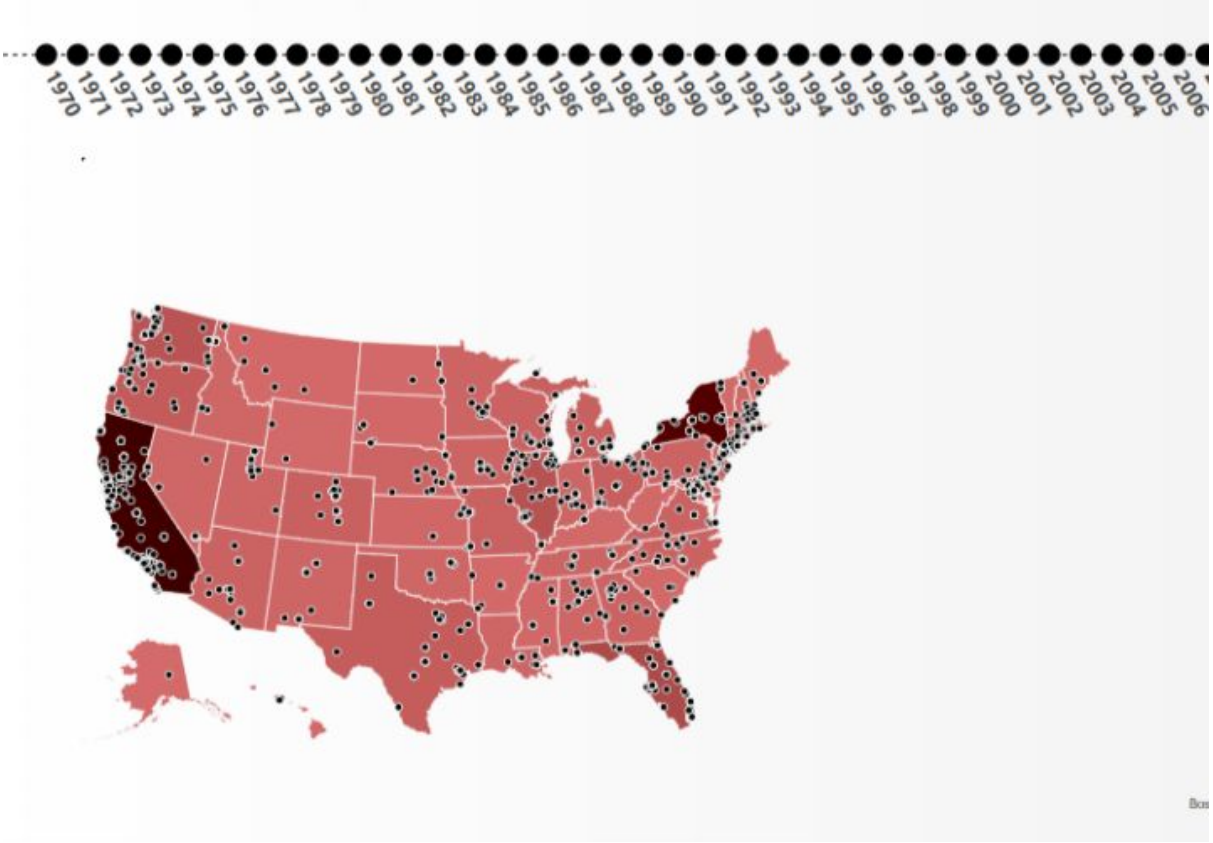


List with Badges also get updated. There are 3 lists, each for Top Organization, weapons and Targets. The badges show the number of attacks in each list item.

How this binds together is we have a region for which we can see overall trends. On selecting a region, We see line graphs for the countries in the region over the years. There is a node for when most attacks have occured. To know what went on in that year, On clicking the node, a word cloud and lists are populated. This quickly gives a snapshot of what happened in a particular year and what were the targets, who were the perpetrators and what kind of attack took place.
Terrorism is somewhat of a subjective term. “Freedom fighters” and “revolutionaries” can easily be put into this category alongside other titles. Our initial plan of answering questions by displaying what happened before and after an attack was very difficult to answer. Also selecting such events brings a bias as we would have to select the events.

What we finally opted for lets the user decide what kind of terrorism took place in a region. We have displayed data using a lot of different aggregations, giving an unbiased snapshot of what occured.

Coming back to the other Tab, USA.



We were initially planning to make this the focus of the project but as we sifted through the data, we learned really interesting things. This displays the US map will all the terror attacks that took place. There is a year chart with a brush for localizing on a particular year.

Future Scope

The problems we faced due to it being a really large dataset and having a lot of preprocessing to do. We wanted to show how a perp organization beginning from the first attack it ever did, spread over the years to different countries. We want to show this over a map with links growing from source to destination in each year. On the Whole, We knew we could have done way more, but the problems we faced really slowed us down. One of them being the map was really slow due to all the data being loaded. We tried our best to make that faster. We fell short in a lot of places as we focussed overly on what data should be used.