Project Report

Kriti Faujdar, Eric Hung, Kshiti Mehta, JD Rao, Tessa Patapoutian

1 Introduction

1.1 Data and problem selection

The dataset that we have selected is the Global Terrorism Database [1] created by the University of Maryland, College Park. We will create visualizations to facilitate exploration and analysis of this data to solve the following problems:

1. How the different regions compare with respect to the number of recorded terrorist incidents?
2. Is there a correlation between time of year and attacks? Are there recurring attacks?
3. What is the relationship between terrorist organisations and number of attacks/successful attacks/coordinated attacks/etc.?
4. Is there a correlation between weapon type and attack type?

1.2 Data abstraction

Dataset type: Table, Network
Number of items: 191465
Number of total fields: 152
Number of relevant fields to be analyzed: 20

1. Field A
   Name: Event ID
   Meaning: A unique number used to identify an incident
   Type: Quantitative
   Cardinality/Range: 191465 (1 per item)
   Further notes: The first 8 numbers are date recorded and last 4 are sequential case number for a given day
2. Field B
   Name: Year
   Meaning: The year the incident occurred

Type: Quantitative
Cardinality/Range: 1970 - 2018

3. Field C
   Name: Month
   Meaning: The month the incident occurred
   Type: Quantitative
   Cardinality/Range: 1 - 12
   Further notes: If unknown, it's encoded as 0

4. Field D
   Name: Day
   Meaning: The day the incident occurred
   Type: Quantitative
   Cardinality/Range: 1 - 31
   Further notes: If unknown, it's encoded as 0

5. Field I
   Name: Country text
   Meaning: The country or location where the incident occurred
   Type: Categorical
   Cardinality/Range: 205
   Further notes: Separatist regions are coded as part of the "home" country; in the case where a country that represented the location of terrorist attacks no longer exists, the country name for the year the event occurred is recorded

6. Field L
   Name: Province/state
   Meaning: The name (at time of event) of the 1st order subnational administrative region in which the incident occurred
   Type: Categorical
   Cardinality/Range: 2565

7. Field M
   Name: City
   Meaning: The name of the city, village, or town in which the incident occurred
   Type: Categorical
   Cardinality/Range: 39115
   Further notes: If unknown, then the smallest administrative area below province/state which can be found

8. Field T
   Name: Criterion 1
   Meaning: Indicates if the incident was aimed at attaining a political,

economic, religious, or social goal, or attaining a pure profit or idiosyncratic personal goal
Type: Categorical
Cardinality/Range: 2 (0 for no, 1 for yes)

9. Field U
   Name: Criterion 2
   Meaning: Indicates if there was any evidence of an intention to coerce, intimidate, or convey some other message to a larger audience than the immediate victims
   Type: Categorical
   Cardinality/Range: 2 (0 for no, 1 for yes)
   Further notes: Evidence can include (but is not limited to) pre- or post-attack statements by the perpetrator(s), past behaviour by the perpetrators, or the particular nature of the target/victim, weapon, or attack type

10. Field V
    Name: Criterion 3
    Meaning: Indicates if the event is outside the international humanitarian law
    Type: Categorical
    Cardinality/Range: 2 (0 for no, 1 for yes)
    Further Notes: If the action is outside the context of legitimate warfare activities, insofar as it targets non-combatants

11. Field W
    Name: Doubt terrorism
    Meaning: Indicates if there is doubt as to whether the incident is an act of terrorism
    Type: Categorical
    Cardinality/Range: 2 (0 for no, 1 for yes)
    Further Notes: In ambiguous cases, where there is a strong possibility but not a certainty that an incident represents an act of terrorism, the incident is coded as "1"; this field is only systematically available with incidents occurring after 1997, and is coded as "-9" otherwise

12. Field Z
    Name: Multiple
    Meaning: Indicates if a particular attack was part of a coordinated multi-part incident
    Type: Categorical
    Cardinality/Range: 2 (0 for no, 1 for yes)
    Further Notes: In cases where several attacks are connected but where the various actions do not constitute a single incident (either the time

of occurrence of incidents or their locations are discontinuous), then "1" is selected to denote that the particular attack was part of a "multiple" incident; this field is only systematically available with incidents occurring after 1997

13. Field AA
    Name: Success
    Meaning: Indicates the success of a terrorist event
    Type: Categorical
    Cardinality/Range: 2 (0 for no, 1 for yes)
    Further notes: Success is defined according to the tangible effects of the attack and not judged in terms of the larger goals of the perpetrators

14. Field AD
    Name: Attack type 1 text
    Meaning: Indicates the general method of attack
    Type: Categorical
    Cardinality/Range: 9

15. Field AJ
    Name: Target type 1 text
    Meaning: Indicates the type of target/victim
    Type: Categorical
    Cardinality/Range: 22

16. Field AL
    Name: Target subtype 1 text
    Meaning: Captures the more specific target category
    Type: Categorical
    Cardinality/Range: 113
    Further notes: There are multiple target subtypes for each target type

17. Field BG
    Name: Group name
    Meaning: Indicates the name of the group that conducted the attack
    Type: Categorical
    Cardinality/Range: 3617
    Further notes: If information is unavailable, this field is coded as "Unknown"

18. Field CE
    Name: Weapon type 1 text
    Meaning: Indicates the weapon types used for each incident
    Type: Categorical
    Cardinality/Range: 13

19. Field CG
    Name: Weapon subtype 1 text
    Meaning: Indicates the subtypes of weapon types
    Type: Categorical
    Cardinality/Range: 30

20. Field EE
    Name: Related
    Meaning: Indicates related incident IDs
    Type: Categorical
    Cardinality/Range: up to 191465 (specific to each incident)
    Further notes: This field is only systematically available with incidents occurring after 1997

1.3 Users identification and tasks abstraction

- Intelligence agent persona
  - Current occupation: Government intelligence officer
  - Education: Bachelor's degree
  - Technical skills: Background in statistical science, awareness of global/international affairs
  - Work experience: Has worked in law enforcement for 10 years
- Political Science researcher persona
  - Current occupation: Political Science Professor
  - Education: PhD, Post Doctorate
  - Technical skills: Highly skilled in Political Science, background in statistical science, very familiar with global/international affairs
  - Work experience: Tenured professor

To reiterate, our tasks are:

1. *How the different regions compare with respect to the number of recorded terrorist incidents?*
   The second question's action and target pairs are to (summarize distribution) between location and frequency of attacks, and to (derive value) of risk.

2. *Is there a correlation between time of year and attacks? Are there recurring attacks?*
   Using the Munzner Taxonomy, the the first question's action and target pairs are first to (identify correlations) between time of year and

frequency of terrorist attacks, and second to (compare values) of number of attacks per time period.

3. *What is the relationship between terrorist organisations and number of attacks/successful attacks/coordinated attacks/etc.?*
   The fourth question's action and target pair is to (discover correlations) between terrorist organizations and a chosen user feature.

4. *Is there a correlation between weapon type and attack type?*
   The third question's action and target pair is to (identify correlation) between weapon and attack type.

A chart is needed to answer all of our questions. To answer question 1, fields B, C, and D are needed. Fields I, L, and M are needed to answer question 2, fields AD, CE, and CG for question 3, and fields Z, AA, AD, AJ, AL, BG, CE, CG, and EE for question 4. We will not need to transform the data to answer any of our questions, and the data is sufficient to answer all of them.

2 Related work

We have focused this section specifically on prior work using the GTD.

Many visualizations created from this database primarily use maps, like this one [2] which shows terrorist attacks from 2011 - 2014 and encodes place of attack, attack type, and number of attacks. This visualization is perhaps useful for seeing where attacks take place, but there are too many colors for the attack type information to be easily understood. It is also hard to get information about frequency of attacks per place because the circles often overlap and it is difficult to see differences in size unless the circles are very close to each other on the map.

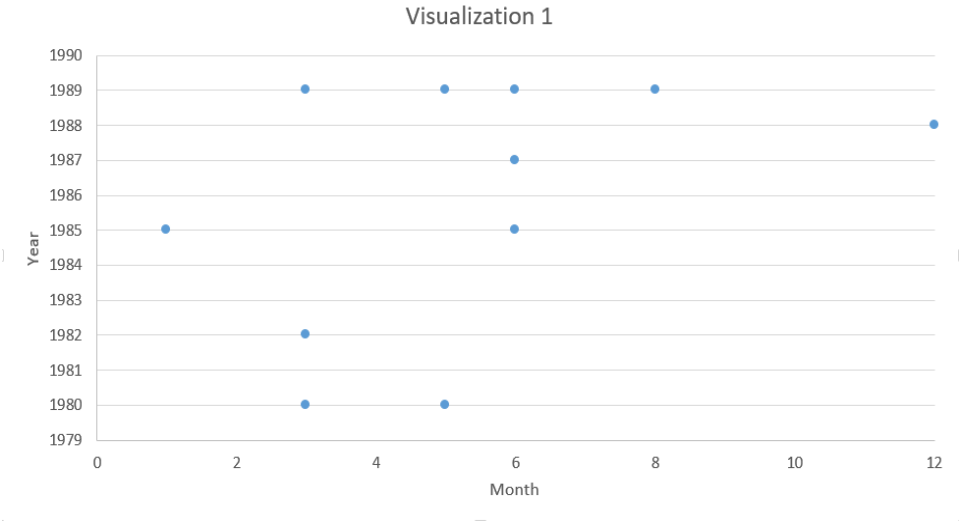
Other visualizations, such as [3], display the data over time using a line plot and encode lots of information with multiple faceted views. This visualization also uses too many colors to usefully encode information, as well as often unappealing color schemes, and the stacked aspect of the line plot means that there isn't alignment and thus the information is difficult to understand.

This poster presentation [4] offers three visualizations, two of which we have just touched on. Their map plot and stacked bar charts suffer similar flaws. The third visualization is a parallel coordinates plot, which has so many overlapping lines and presents information in such an abstract way that it does not seem to offer much intuition about the data.

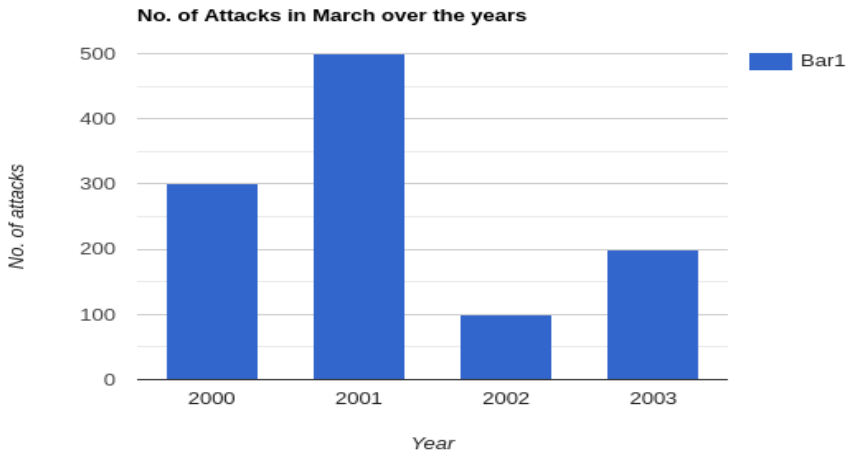
3 Solution

We propose four visualizations, one for each of our tasks.

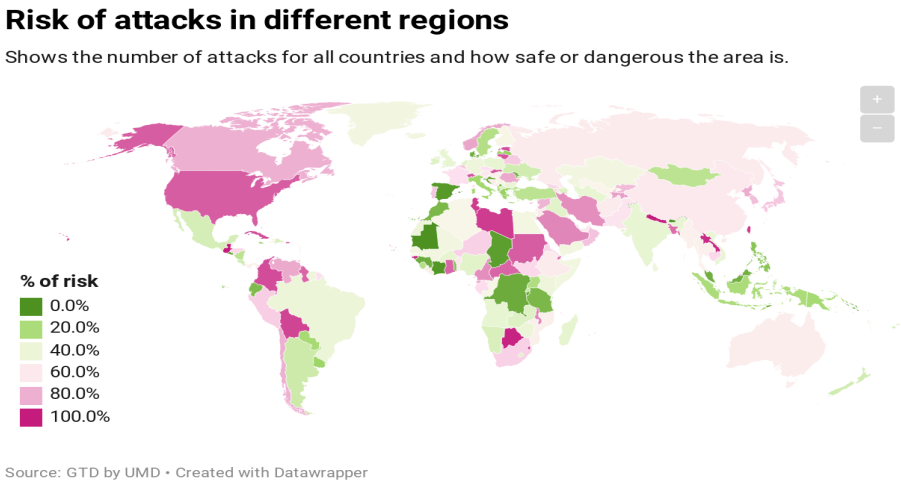
1. A scatter plot of month/date on the x-axis vs year on the y-axis.



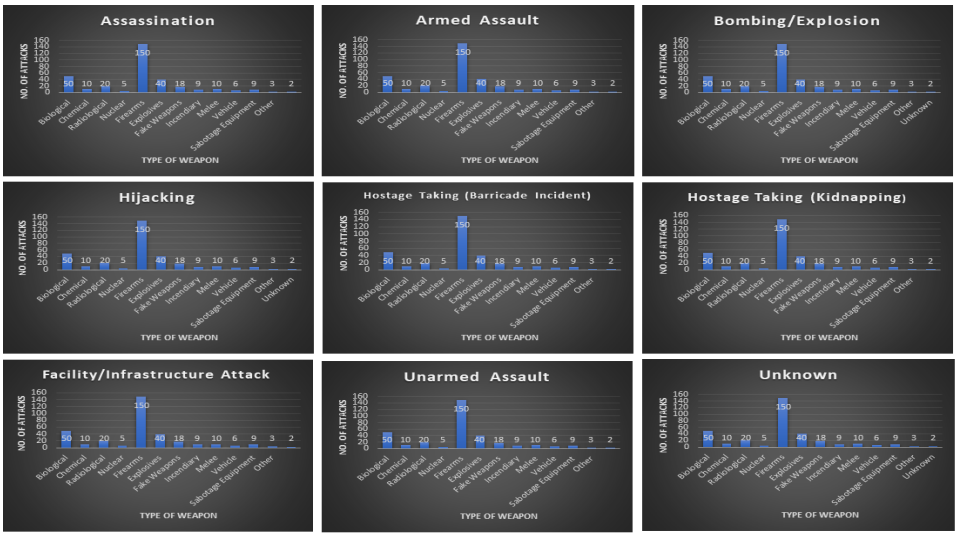
   The user can click on a particular month or year to view a bar chart representing that time period.



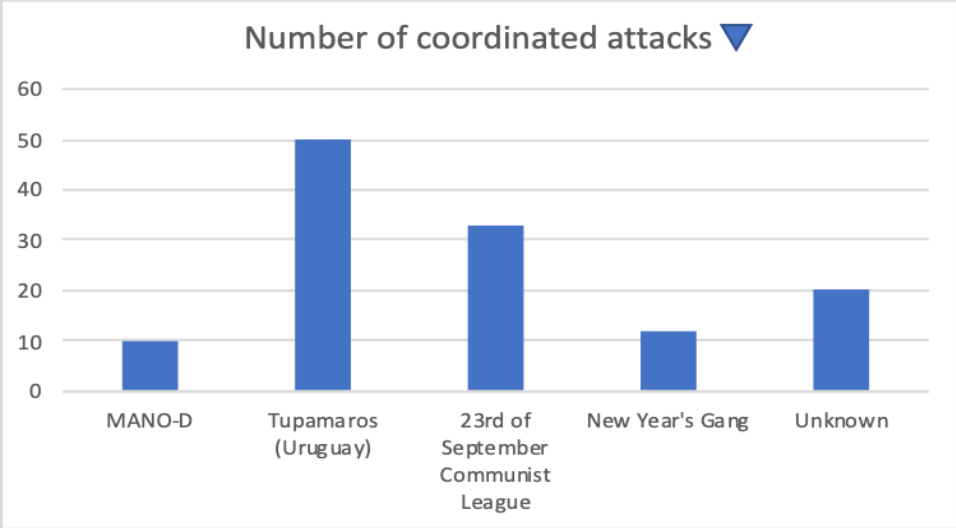
2. A choropleth map (diverging) according to number of attacks. The user can zoom or choose a specific country to focus on.



3. A grid of nine bar charts, where each chart represents an attack type and the bars represent the number of weapon types used.



4. A bar chart with terrorist organizations on the x-axis and a user-selected field (e.g. number of coordinated attacks, number of successful attacks, weapon type, attack type, target type) on the y-axis. User can use side bars and interactive tools to filter the details.



Users can choose a field by clicking the filter to see details.

3.1 Argument for our solution

We selected a scatter plot because it is the best visualization for identifying clusters and extreme points, which is what our first question is seeking to find. A common visualization chosen by others is a line chart, but this design would be too cluttered and a user wouldn't be able to identify a particular point or discern trends in the data. The data is also fairly sparse, so a scatter plot is much better suited than a line chart.

There is no prior work addressing the second question and is inherently spatial data, so a map is the best visualization for identifying geographic regions with specific attributes. It is also useful when comparing regions that are close to each other geographically, which is doubtless relevant when considering terrorist attacks.

That majority of our visualizations use bar charts because they're the most efficient visualization when looking at the relationship between two attributes. The user can specify exactly which features to compare and make it easy to identify any correlations between selected fields.

4 Implementation Details

4.1 Diverging Interactive Choropleth Map

Tools used:

- CSS used for borders, caption and tool tip.
- In D3 we used Map (geoPath and geoNaturalEarth) and Projection. For transformation purposes we use scale and translation for the map,

legend, colour scales and labels.

- For loading data we used d3.csv and d3.json.
- We use mouse-over and mouse-out to show tool tips and click to enable zoom in and zoom out features for countries and states/provinces.
- We use animation like transition to enable smooth transition of the zoom in and zoom out feature in our interactive map. We also use duration to give enough time for the user to experience the transition that takes place.
- We use d3.mouse to pick up the location at which the user is pointing/hovering and return a tool tip that gives more information about that specific location on the map.
- We use attribute fill to give colours to the regions of the map according to the colour scale for the number of attacks that have occurred in that region.

4.2 Scatter Plot

Tools used:

- For loading data, we used d3.csv.
- In D3, we used translation for graph and label positioning and scales for axes and dot size.
- We used the fill attribute for coloring the points.
- We used filter to only include entries with the necessary data.

4.3 Linked Bar Charts

Tools used:

- CSS flex used for layout and borders caption, d3-tip for tooltips.
- We use nested/grouped d3 bar charts to show sub categories such as weapon types, attack types and target types corresponding to specific organization.
- For loading data we used d3.csv.
- We use mouse-over and mouse-out to show tooltips and click to enable change the nested bar charts corresponding to the chosen organization.
- We use two filters to control the item on x axis and y axis respectively. For x axis, we can change the top N organizations that conducted the most attacks or with the most success decided by the other filter.

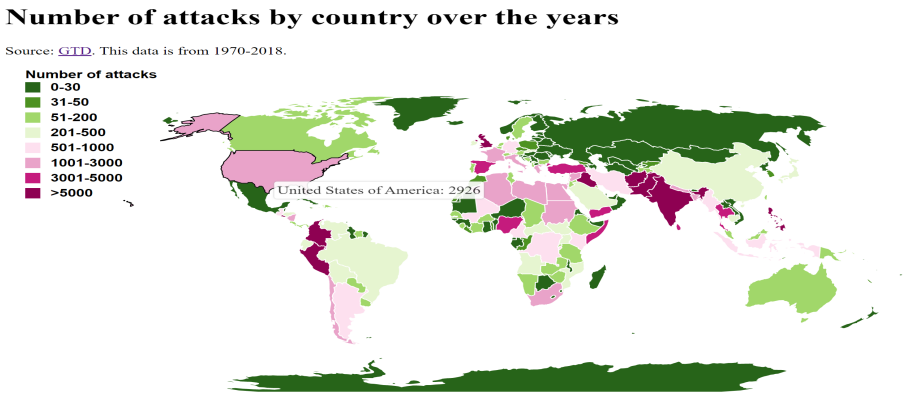
4.4 Other Bar charts

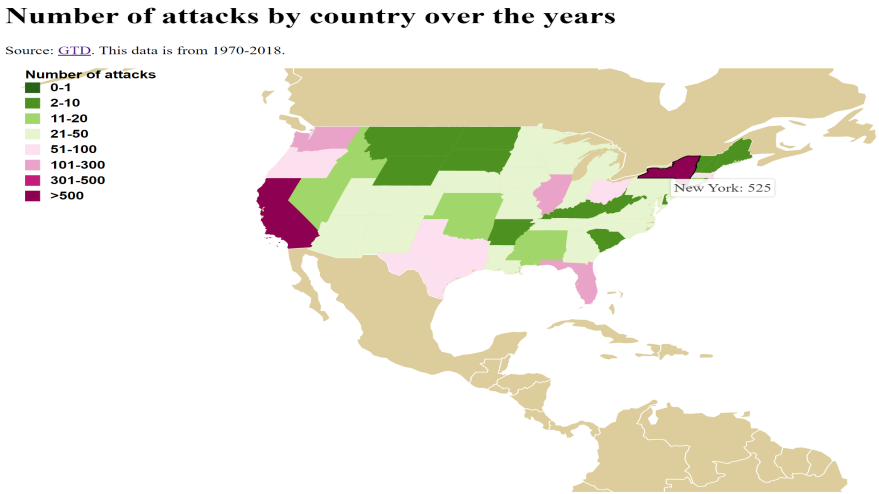
Tools used:

- CSS for styling and d3-tip for tooltips.
- For loading data we used d3.csv.
- We use mouse-over and mouse-out to show tooltips and data values on bars
- To show multiple bar charts on a web page, we used css grid.

5 Final Product Sample

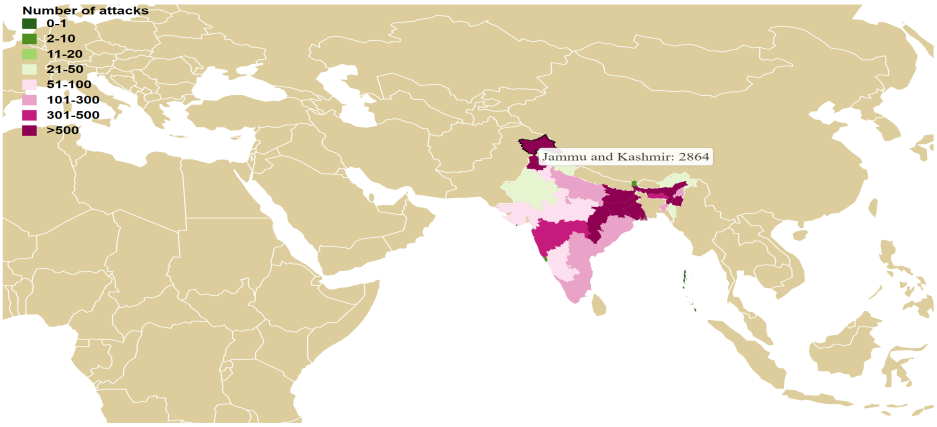
5.1 Diverging Interactive Choropleth Map





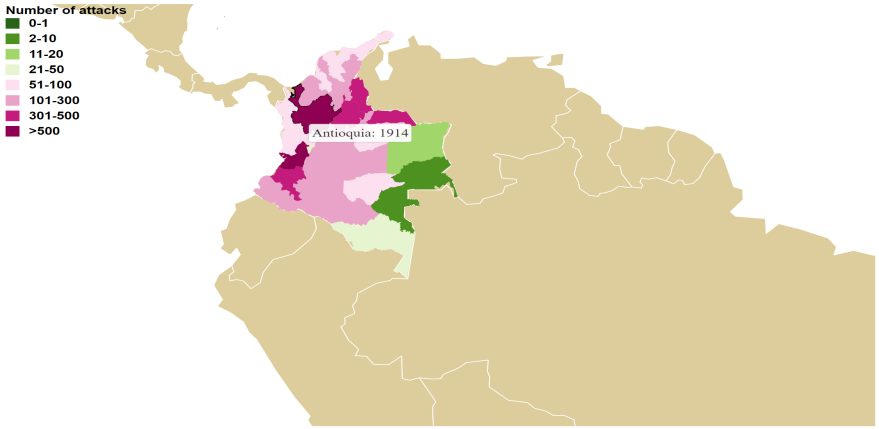
Number of attacks by country over the years

Source: GTD. This data is from 1970-2018.

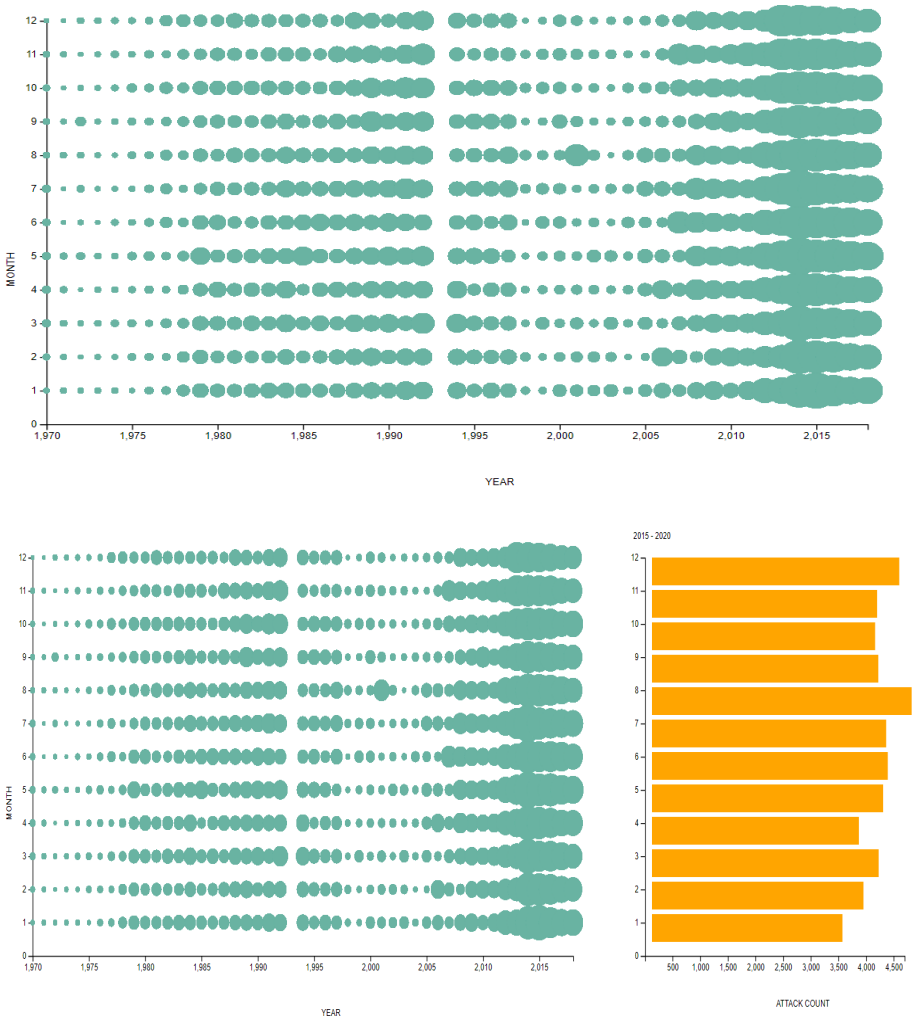


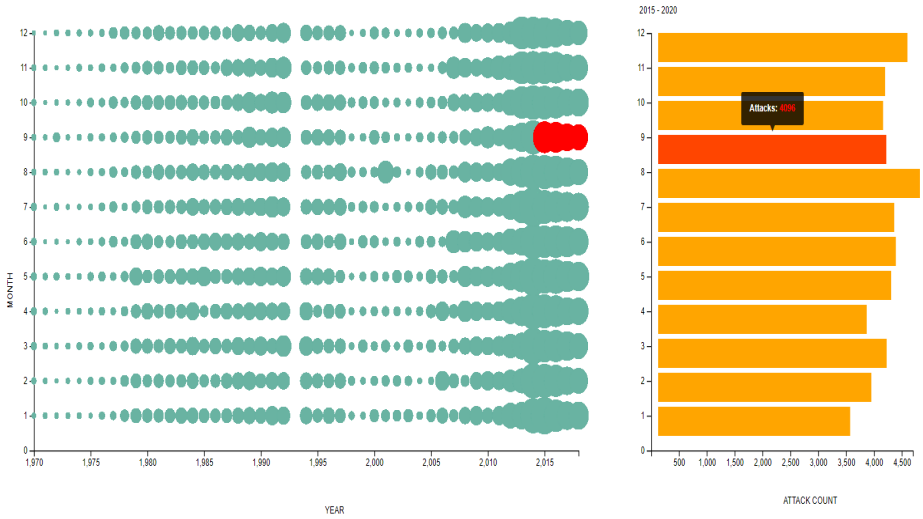
Number of attacks by country over the years

Source: GTD. This data is from 1970-2018.

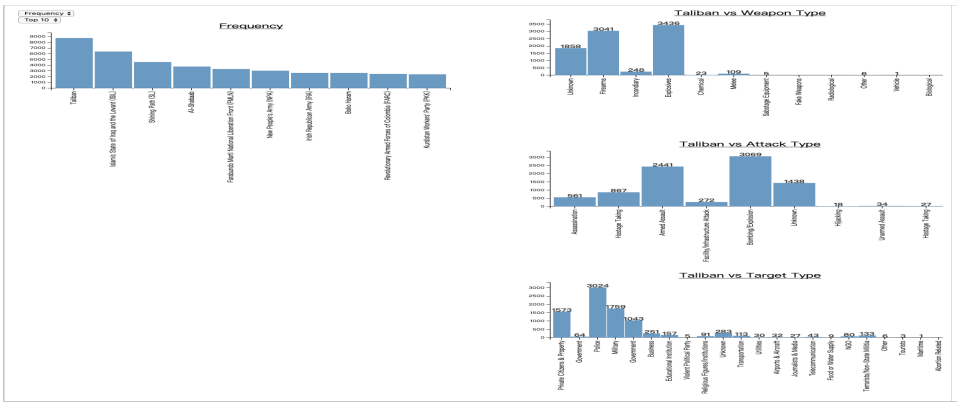


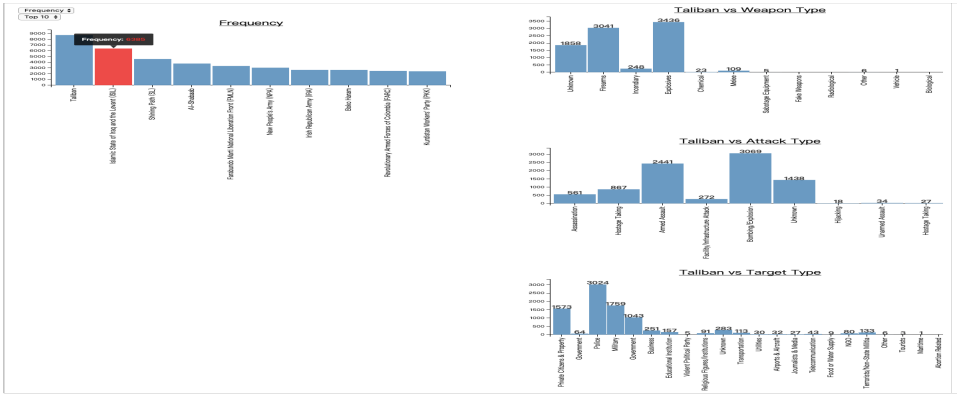
5.2 Interactive Scatter Plot and Bar Chart

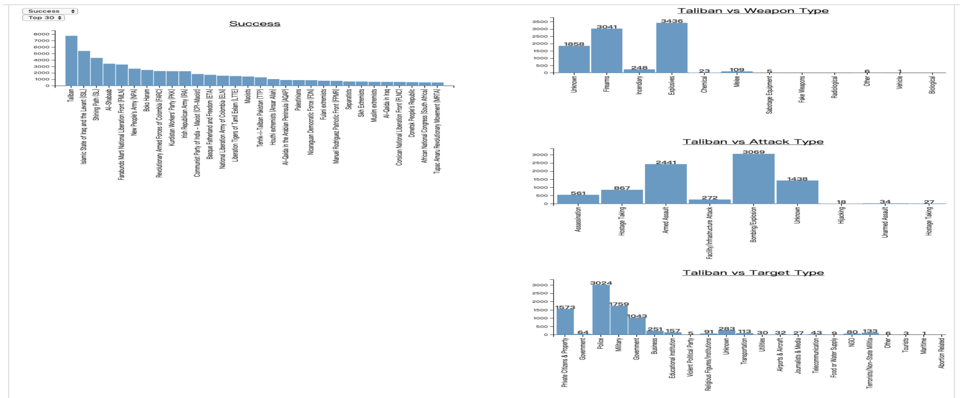


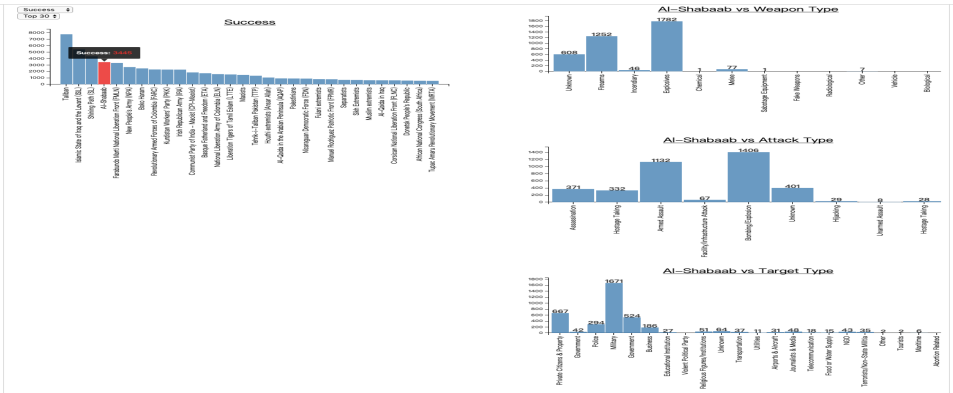

MONTH
YEAR
2015 - 2020
Attacks: 4096
ATTACK COUNT

5.3 Linked Bar Charts

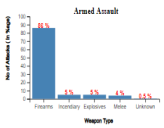


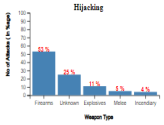


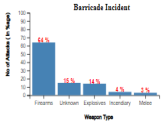

Success
Taliban vs Weapon Type
Taliban vs Attack Type
Taliban vs Target Type

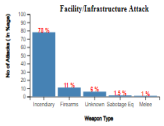

Success
Al–Shabaab vs Weapon Type
Al–Shabaab vs Attack Type
Al–Shabaab vs Target Type

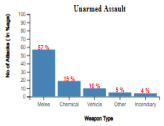
5.4 Bar Charts

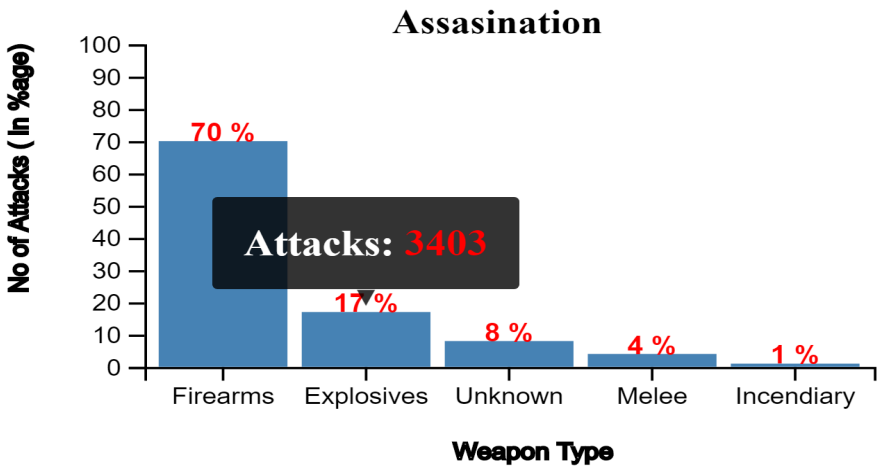












References

[1] G. LaFree and L. Dugan, "Introducing the Global Terrorism Database", *Terrorism and Political Violence* 19: 181-204, 2007.

[2] Mapbox, *Map of all the attacks between 2011 and 2014*, [Online], Accessed at https://public.tableau.com/views/InteractiveVisualization_0/Dashboard 3?:embed=y&:display_count=y&:showTabs=y&:origin=viz_share_link, 28 March 2020.

[3] J. Lee and B. Schneiderman, "GTD Explorer", Human-Computer Interaction Lab at the University of Maryland, [Online], Accessed at http://www.cs.umd.edu/hcil/gtd/gtd/explorer.html, 28 March 2020.

[4] C. Ziemkiewicz, et al., "Global Terrorism Visualization", Southeast Visualization and Analytics Center (SRVAC), University of North Carolina at Charlotte, 2008.