CS526: Global Terrorism Analysis

Keya Desai
Rutgers University
New Brunswick, NJ, USA
Email: kd706@scarletmail.rutgers.edu

Prakruti Joshi
Rutgers University
New Brunswick, NJ, USA
Email: phj15@scarletmail.rutgers.edu

Shang-Hao Huang
Rutgers University
New Brunswick, NJ, USA
Email: sh1384@scarletmail.rutgers.edu

***Abstract*— The Global Terrorism Database (GTD) [1] is an open-source database with information of ∼180K terrorist events across the world in the years of 1970-2017. Using the information of the events, as part of the course *CS526: Data Interaction and Visualisation*, we have designed and developed an interactive dashboard which explores the dataset in depth. In addition to summarising the terrorism trends over the years, the dashboard provides means to view important statistics of an event and that of a country. Such understanding of the data can be used to form global and local anti-terrorism policies to prevent such events.**

Index terms - Global terrorism, data visualization, data interaction, statistical analysis

I. Project Description

A detailed project design is discussed in the following sections using the project design methodology.

A. *DATA*

a) Data Description: The Global Terrorism Dataset (GTD) [1] is an open-source database maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland (UMD). This dataset recorded information on 181,693 terrorist attacks all over the world from 1970 to 2017. The START organization defines *terrorism* as follows:
"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."
Thus, following the definition, the GTD database covers events such as assasination, kidnapping, armed assault in rallies, attacks on religious group etc, apart from all the global major terror events.

b) Selected Data set:

1. Size: 181,693 rows, 160 MB
   Original Format: CSV
2. Is the data at rest? Is it Streaming?
   Data is at rest.
3. Is the data Time Variant data?
   Yes.
4. Is Time an essential component of the desired data processing?
   Yes, time is an essential component of the data processing. Identifying trends in terrorism is central to facilitating data exploration of GTD. In the final model, a page is dedicated for this purpose which can be explored in the 'Trends' section. In this section, the trends in number of attacks and deaths worldwide and for each individual country over the years can be observed. This helps in understanding of the political, social or economic standing of a country at a given period of time and how the policies of countries is helping in either curbing or fostering terrorist attacks. As an example, from the visualisation in Figure 1, it can be seen that there was a sudden peak in terrorism in the year 2014, due to the large number of terror events occurring in Iraq (Figure 2). The visualisation over time has a potential to be useful for experts to chain events together through time and form policies based on historical events.

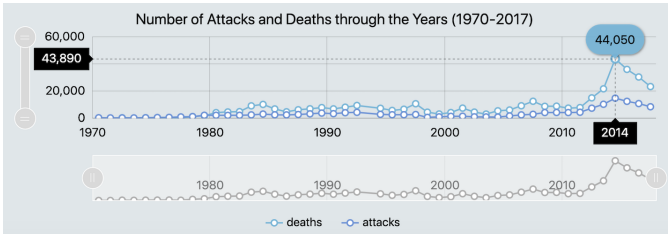

Fig. 1: Trends in terrorism (Worldwide)

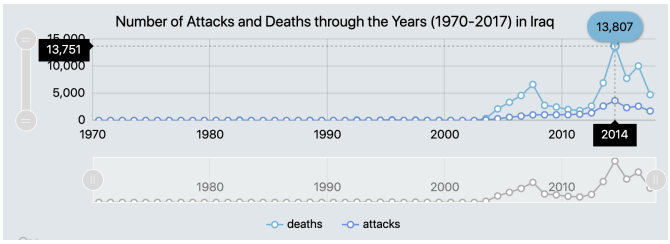

Fig. 2: Trends in terrorism (Iraq)

c) Identify central entities and relationships
The central entities in GTD are:

1. Incident
2. Location
3. Time of Occurrence
4. Target/Victim
5. Perpetrator
6. Casualties and Consequences

Relationships:

- Incident occurred at Location.

- Incident occurred at time (year, month, day) and lasted for how long.
- Perpetrator carried out the incident.
- Target/Victim casualties or injuries by perpetrator.
- Incident caused casualties.
- Nationality (Location) of the Target/Victim

d) Identify basic Stats that that will be useful for processing: record size, distributions, etc.

1. No. of Attributes = 135
2. No. of Records = 181,693
3. No. of Terrorist groups = 3,537
4. No. of related events: 25,032
   The dataset logs list of related events. For each event, an attribute contains list of eventids of events which are related to the particular event. One example of related events is the terror events which took done the World Trade Center in New York City, United States on 9/11. A event logs the details of the first plane crashing into the towers and a separate event entry logs the second plane crashing into the second tower, minutes after the first one. Both events are marked as related in the database.
5. Relation amongst the following parameters:
   - Location, Attack type, Target type, Weapon type (Political/Religious): Identify the political, social scenario of a country.
   - Country, success of attack: How successful a country is in curbing a threat.
   - Victim type (Tourists/Govt/ Police), casualties
   - Terrorist group, attack type, country attacked: Identify characteristics of a terrorist group.
   - Terrorist group, motive: Identify the ideology or motive of a terrorist group.
   - Event, casualty: Related events and their damage and extent
6. ER Diagram: The ER Diagram is shown in Fig. 3. The central entity in the database is the event which is identified by a unique EventID. The event has attributes and entities of location, time, terrorist group (gname), targets involved and damage associated to it.

B. *QUESTIONS*

a) Motivation:
As the terrorism news are becoming more frequent and affecting people's view toward countries, we wanted to build a visually informative web interface providing an interactive and vivid way of displaying the data in an attempt to not only make the public aware of the issue and understand the context easier, but try to help researchers facilitate exploration of the data.

b) What are the fundamental questions you want to answer about the data?

- Global distribution of terrorism
  - Number of terrorist events around the world and through the years
  - Identify high-conflicted countries/regions
- Understand the dataset in terms of:
  - Countries most affected by terrorism
  - Region wise (Continent wise) distribution of terror activities through the years
  - Activities of top terrorist groups.
  - Worst terror events (in terms of number of deaths) in the years of 1970-2017
  - Motive behind majority of the terror acts
  - Identify most common targets and choice of weapon
  - Visually display the geographic reach of terrorist groups by showing related events on the global map
- To visually represent the year, location and scale of each event and the corresponding statistics such as the terror group responsible for it, motive behind the attack and a short summary instead of a lookup in the database.
- Country-wise Statistics : Identify for each country, most affected cities, terrorist groups active, attack type, target type and weapon type distribution and statistics such as total number of terror events, terror related deaths and injuries, percentage of successful events, contribution to number of global events, worst affected year and event.

c) What is the fundamental data representation that will allow the system and users to query extract information about the data?
As the Global Terrorism Data is available in a csv file, it is has been stored in a SQL database of Postgres for data processing. The system can query the database via SQL queries which fetches relevant data from the data store. On the dashboard the users can interact with the data and the visualisation via sliders, drop downs and

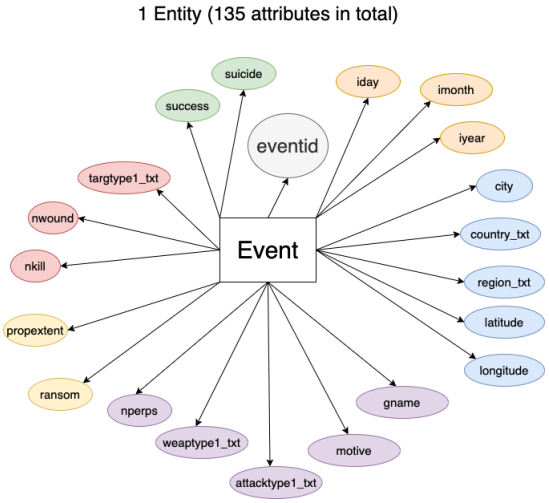

Fig. 3: ERD of data displaying the central entity as the event which is identified by an unique eventID. The other entities and their relationships.

check boxes. The user can manipulate the year of which the data is displayed via a slider. Country can be selected from a dropbox to view that country's terrorism statistics. The input from the user is used to fetch data from the database and update the plots.

d) Create a name that faithfully describe the purpose of your Interactive System.
Global Terrorism Analysis. The purpose of the system is to enable its users to explore the GTD and draw insights about global terrorism using interactive visualisation. Hence, the system has been named 'Global Terrorism Analysis'.

e) Types of user:
The following are the types of users which can be benefited from the interface.
- General Public: To have a big-picture about global terrorism and country specific information without going through the csv data.
- Policy-maker: To better analyze the global terror situation and make suitable policies.
- Anti-crime Agency/Police department: To better identify/predict the terrorist events and mitigate it
- Researchers: To facilitate exploration of the data for researchers and academics

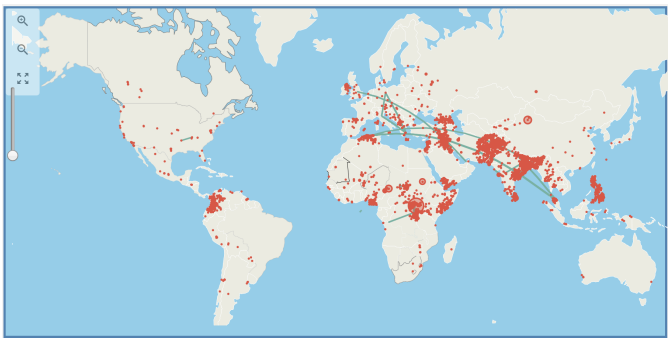

Fig. 4: World map showing related events in 2009

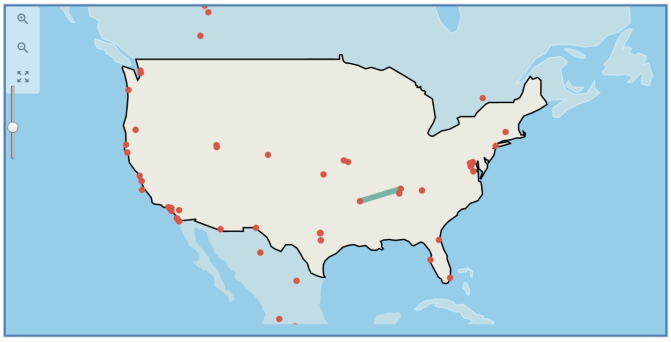

Fig. 5: Related events in United States in 2009

C. *MODE OF PROCESSING*

a) How is the data going to be represented and manipulated internally?
The GTD data is available as tabular formatted (csv file) and structured. Hence, it is efficient to store and manipulate structured data using SQL. If it were unstructured data, NoSQL would have been more efficient. Our application requires running complex queries to create interactive visualisations and users can dynamically query the data to get information. SQL allows for fast search and retrieval of data for complex dynamic queries and hence using SQL is the better for our application.
Hadoop is a software framework which provides processing of large amounts of data. However, due to its high capacity design, the Hadoop Distributed File System, lacks the ability to efficiently support the random reading of small files. Hence, given that the size of the GTD data is only 160MB, using Hadoop will not be efficient. [2].

b) Data Store: Data Base + Network Representation?
Data Base:
Two of the widely used open source SQL databases are PostgreSQL and MySQL. While both databases are ACID compliant and support python, PostgreSQL is faster for executing complex queries while MySQL performs faster reads. For this application, we need to execute complex queries and thus PostgreSQL will be a better choice.

Network Representation:
As mentioned in section I-A, the dataset contains information regarding related events. This information is represented as a network on world map. Figure 4 shows scatter plot of events and network of related events (two events are related if there is line between them) in the year of 2009. Some cross country related events can be observed in the map such as Algeria-Thailand. Intra country connected events can be observed in the United States (Figure 5). On hovering over the two events, the information of the events is displayed, which indicates that both the armed assaults were carried out by 'Jihadi-inspired extremists' a day apart. Such connections can be studied in depth by researchers and experts using this network representation, without going through the large csv file.

c) What questions can be answered via the type of Data Base Store being used?
Using PostrgreSQL to store the data, we can query the data to answer the questions listed in I-B.

d) How is data input into the system after the system is set up initially?
Since the data is static and not streaming, after the initial set up, there is no need of data input into the system after the system is set up. The csv file is imported in postgres database store and SQLALchemy is used to connect the database to python.

II. Dashboard

In this section, the working of the dashboard is described. The video of the demo can be found **here**. The github repository for the project can be found: **here**.

A. *DASHBOARD PAGES*

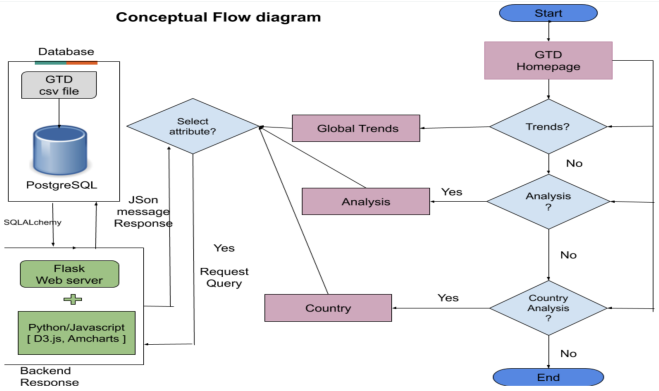

Fig. 6: Flow diagram of the system

a) **Flow Diagram:** The flow diagram of the dashboard is shown in figure 6. The navigation between views and the details of each view are explained in the following sections.

b) **Web Pages in the application:**
A user can navigate to different pages using a floating sidebar on the left. The different plots in each page view are described in detail in section II-B. The basic functioning of each page view is as follows:
Home: This is the first page which renders on opening the application, shown in figure 7. The home page describes GTD and has a button which on clicking opens the GTD homepage. A second button takes the user to the dashboard.

Fig. 7: Home page

Trends: The trends page (figure 8) explores the time component of the dataset. The user can view the trends in terrorism over time worldwide, for each region and country. The user can further set a year to view the terrorist events that occurred in that particular year on an interactive world map, along with the details of the events. The most affected country or region in each year and the corresponding information on events is combined in this view.

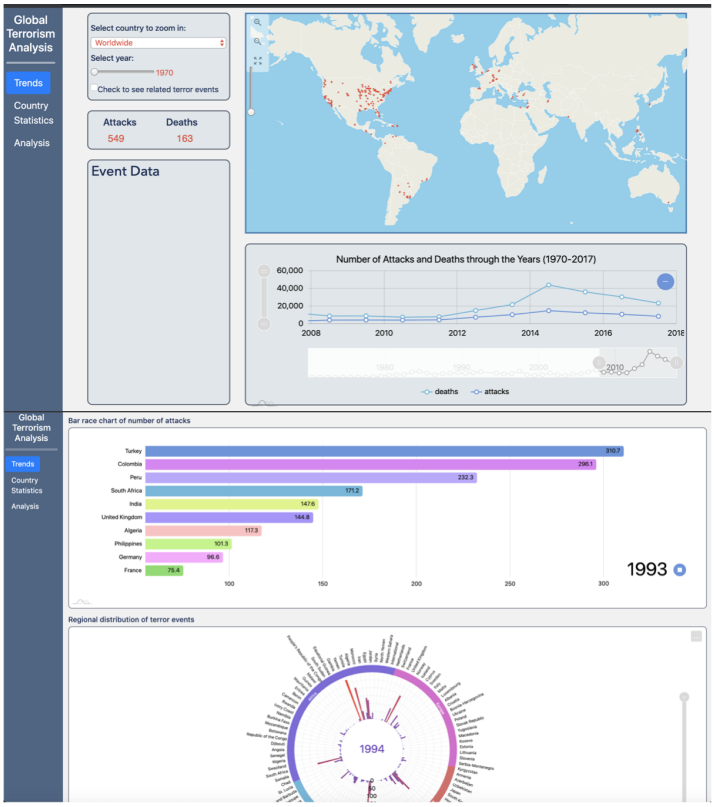

Fig. 8: Trends page

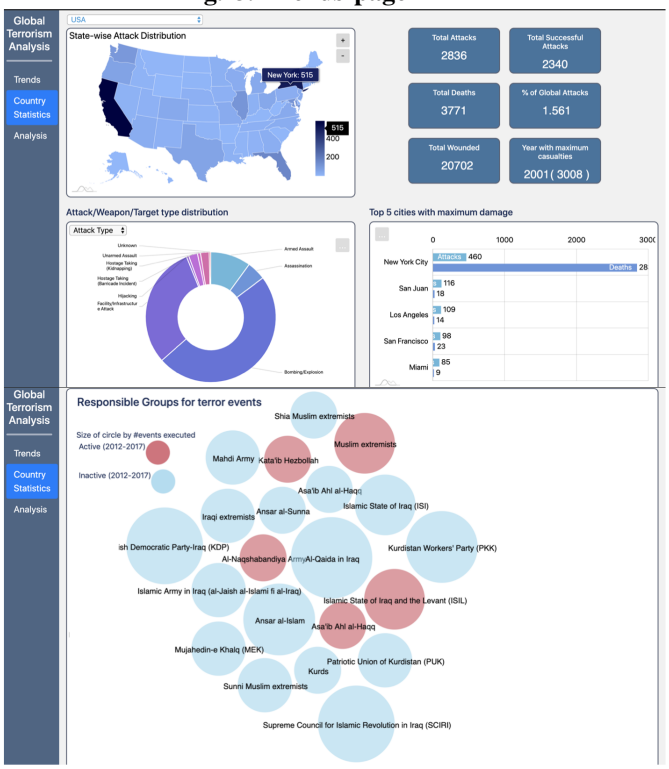

Fig. 9: Country Statistics page

Country Statistics: On this page (figure 9) , terrorism statistics of all the countries can be explored. A country can be selected from the dropdown menu and the plots are updated accordingly. The page describes the state-wise distribution of terrorism, distribution of attack, target and weapon type, top cities in a country affected by terrorism and the terrorist organisations responsible for the attacks.
Analysis: This is the page (figure 10) which provides the aggregate analysis of the entire dataset. Information of activities of top terrorist groups, deadliest events, attack, target, weapon type distribution, countries most affected by terrorism and motive can be explored on this page.

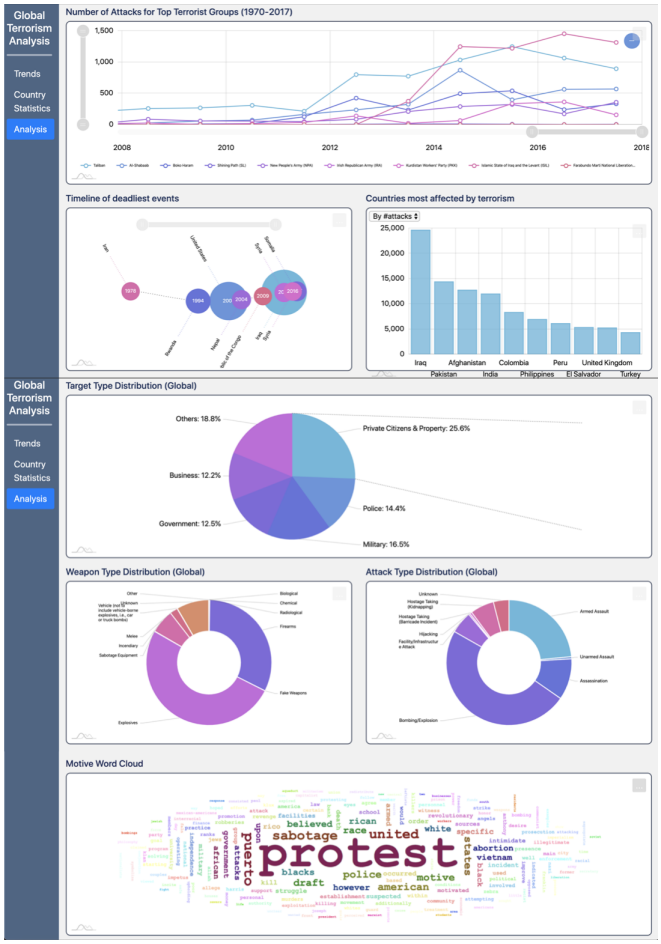

Fig. 10: Analysis page

B. *VISUAL REPRESENTATION*

a) **Plots**

Line Plot: Trends of number of attacks and terrorism related deaths worldwide and for each country is represented using a line plot (Figure 1). Further, trends in terrorist activities by top 10 terrorist groups (in terms of number of attacks carried out), can be observed by a line plot.

Pie Plot: Pie plots are used to represent the distribution of attack, target and weapon type. Pie plots offer a relatively easier method to visually summarise large data. One example of a pie plot in the dashboard is given in figure 12. The sub-type distribution of a particular distribution is visualized best by using pie charts. We have visualized the sub-type of target type using this hierarchical distribution.

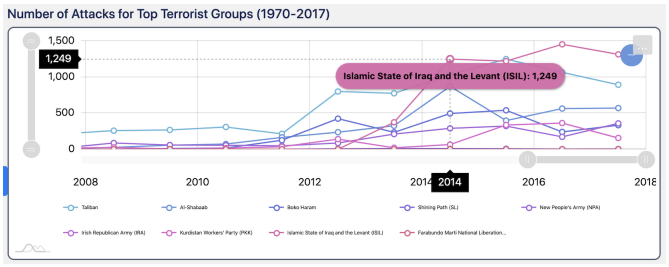

Fig. 11: Line plot showing activity of terrorist organisations

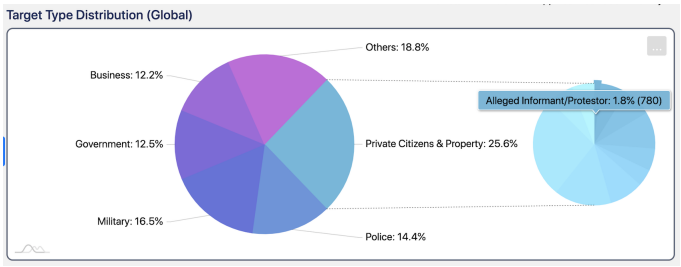

Fig. 12: Pie plot of target type and subtype distribution

Bar Chart: Comparison between the number of attacks and deaths among different countries and cities is given using bar chart. Figure 13 represents one such plot which represents the number of attacks occurring in countries between the years 1970-2017 in a sorted order. It can be seen that Iraq, Pakistan and Afghanistan are the countries most affected by terrorism. The dropdown on the top left allows the user to view the most affected countries according to terrorism related deaths. To make it visually attractive and avoid redundancy, we have also used clustered column bar charts and horizontal bar chart.

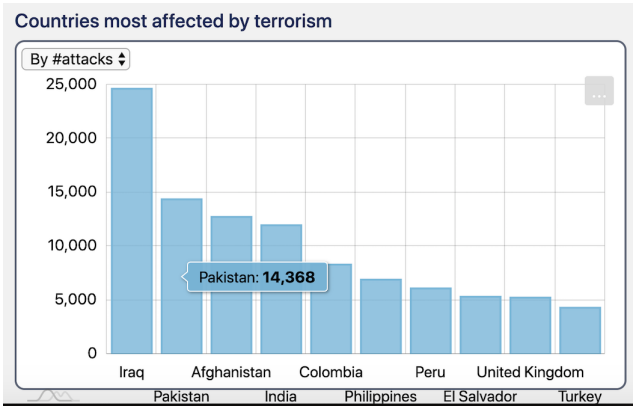

Fig. 13: Bar chart showing most affected countries

Fishbone timeline: The worst 10 events in terms of number of deaths it incurred is represented in figure 14. The plot gives information of the year and the country in which the event took place and on hovering the user can see the terrorist group responsible for the attack and the number of deaths. The bubble size is proportional to the deaths.

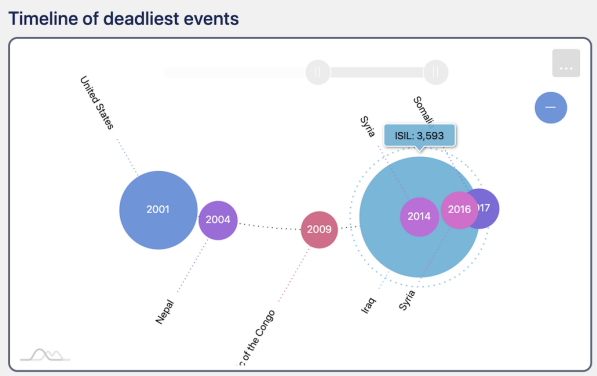

Fig. 14: Time line of worst terror events represented by a Fishbone diagram

Map: The geographic information in the dataset of longitude and latitude about the event is used to represent the event on a map, as shown in 4. This method of visual representation of the data, helps the user in understanding the exact geographic location of the event. Further, a map of a country is used to represent the state-wise distribution of terror events. Figure 15 shows the map of United States. Shades of blue are used to represent the intensity of the attacks in a state. Using the scale, it is evident from the plot that the states of California and New York are the most targeted for terrorist attacks. Both the maps can be zoomed in or out to focus on the details. In the world map, a country can be clicked or a country can be selected from a dropdown menu to zoom into it. Buttons on the top left of the map allows to zoom in or out over a specific region and 'reset' button can be used to zoom out to the world map.

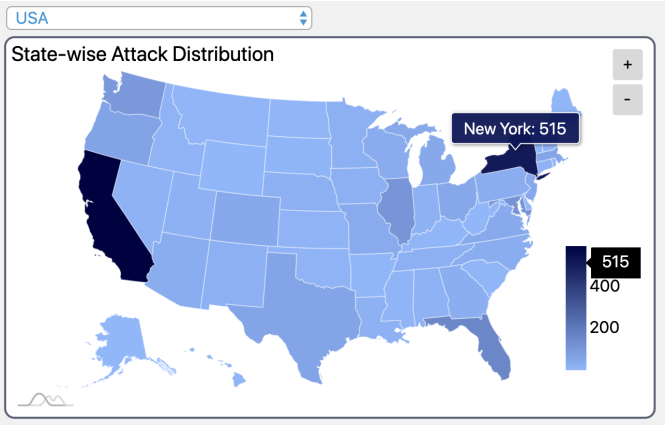

Fig. 15: Unites States choropleth

Bubble Plot: Representation of the terrorist groups operating in a country is done with a bubble plot. Figure 16 shows the groups operating in Iraq. The size of the bubbles is proportional to the number of terror events for which that group is responsible for. The exact number for each can be viewed on hovering. In this plot, colour has been used to convey the information of the activity of each group. Red colour indicates that the group was active in the last 5 years (2012-2017), whereas blue indicates that the group was active before that. The importance of this information lies in the fact that a country has potentially more danger from the terrorist gangs active in recent times. A bubble plot is the best choice to represent multi level information of large number of terrorist groups.

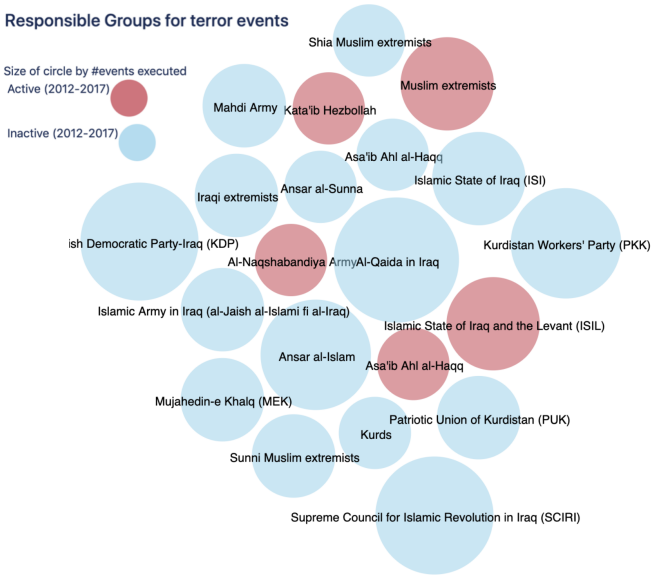

Fig. 16: Bubble plot of terrorist groups active in Iraq

Word cloud: One of the attributes in the database is the motive behind a terror event. This information has been used as a whole to construct a word cloud as show in figure 17. The size of a word is based on the number of times it appears in motive. The plot captures the main reasons behind any terror attack. 'protest', 'sabotage', 'race', 'abortion' are some of the most frequently occurring words in the motive.

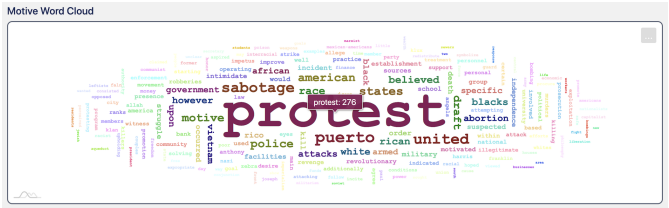

Fig. 17: Word cloud of motive

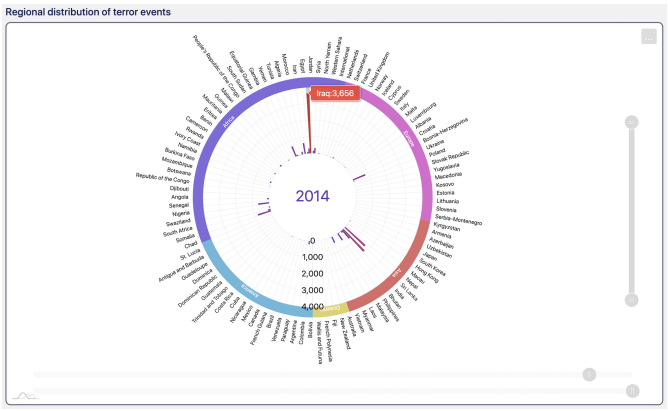

Fig. 18: Radar timeline

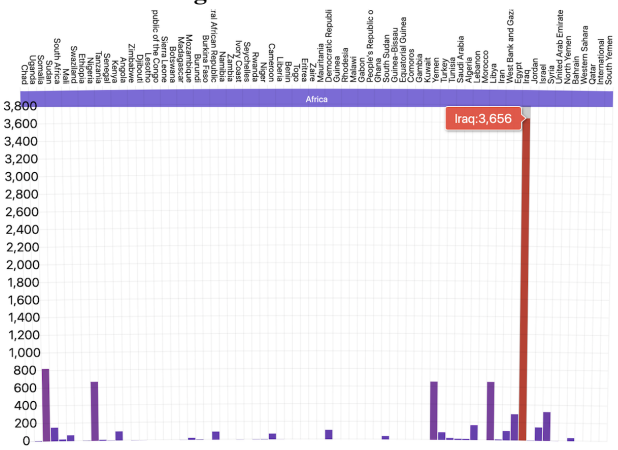

Fig. 19: Histogram of terror events in Africa in 2014

Radar timeline: This plot provides a concise region wise analysis of terrorism. For a given year, the chart shows the number of attacks for every country in each region. The scale is determined by the height of the bar and the number of attacks is displayed on hovering. In figure 18, it can be seen that Iraq had the most attacks in the year of 2014. The user can zoom into a region and convert the plot into a histogram using the bottom most slider for a better comparison (figure 19).

b) **Linking of different views** Linking of views is done in two sections of the dashboard, *Trends* and *Country Statistics*. All the views that are linked in the *Trends* page are shown in figure 20. The world map shows the events in a year which is set using the slider. The size of the bubble is proportional to the number of deaths caused by an event. On hovering over the event, its data is displayed in the 'Event Data' text box. On selecting the related events checkbox, the related events for the selected year are displayed on the world map.
Figure 20 shows event data of an attack which took place in Las Vegas, United States in 2017, which killed 59 people. The text box of 'Attacks' and 'Deaths' shows the total attacks and terror related deaths, worldwide in the year 2017. The line plot on the bottom right is the trends in number of attacks and deaths worldwide. On selecting a country from the drop down, all the views change,as shown in 21.
In the section of the *Country Statistics*, linking of views is shown in figure 22 and 23. On selecting a country from the dropdown menu on the top left, all the 5 views are updated. The 5th plot is shown in figure 16.
All the page views are connected to each other using the floating side navigation menu bar. Switching between the pages and back to the dashboard is easily navigable through the side menu.

C. *INTERACTIVITY*

The interactivity is less than 1s for all the plots accept the map on the Trends page (figure 4). This figure takes 2s to render because of the huge data which is changed dynamically.

a) **Interface Layout:** The layout of the system is shown in section II-Ab). It is a dashboard with navigation bar on the left side.

b) **Interface Interaction mechanisms and answers representation for further user interaction**
The interface interaction mechanisms are both textual and graphical.
Mouse Hovering, Clicking and selection: These features are present in every plot described in section II-Ba). On hovering or clicking the maps, the value of the hovered item is displayed.
Zooming and Panning: These are employed in the plots of 4, 15, 18, 11 and 14. The maps can be drilled down using Panning. Also, each map and some of the plots have buttons or sliders to zoom in and zoom out and reset to original view.
Summary: Text boxes are display on the *Country Statistics* page to display the statistics for a country as shown in figure 24.

Fig. 24: Text boxes

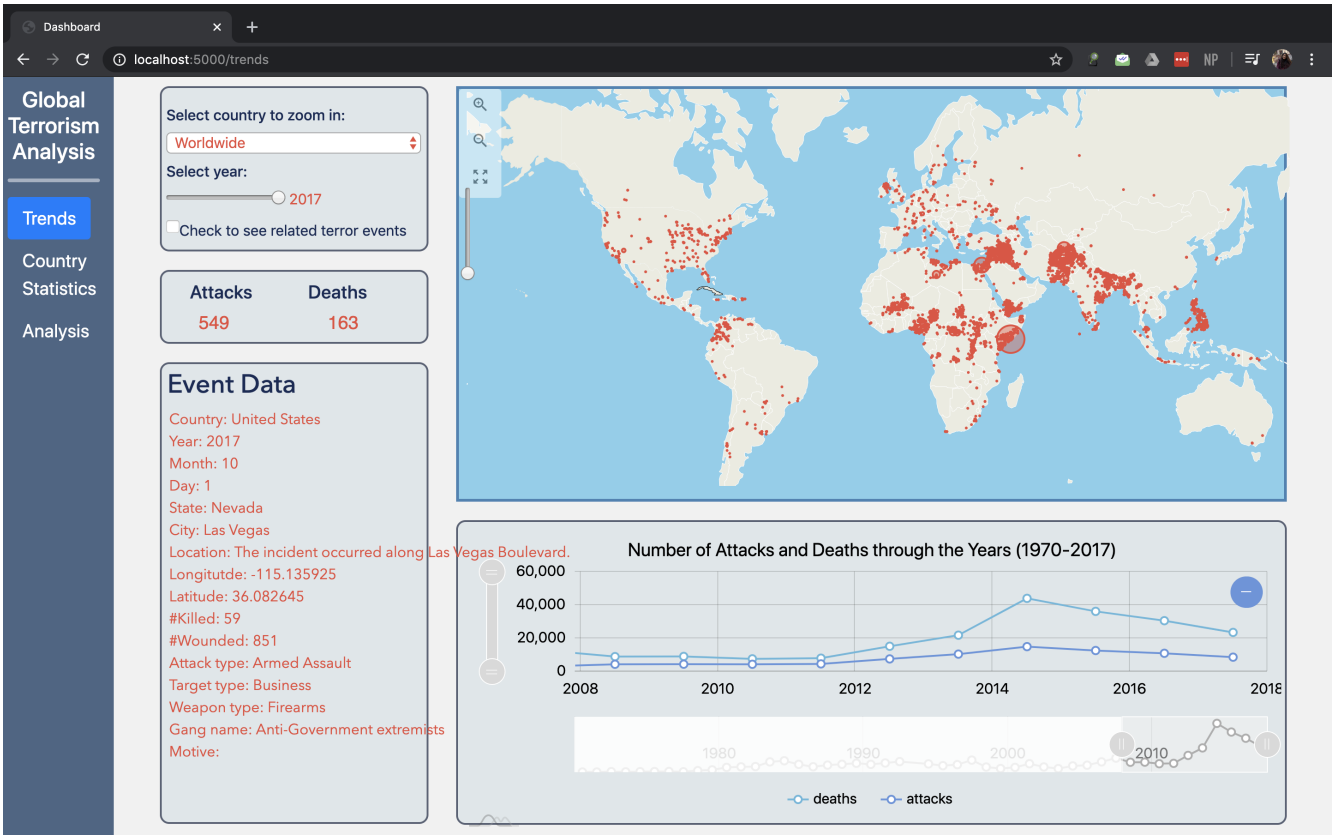

Fig. 20: Trends page - Selection: Worldwide

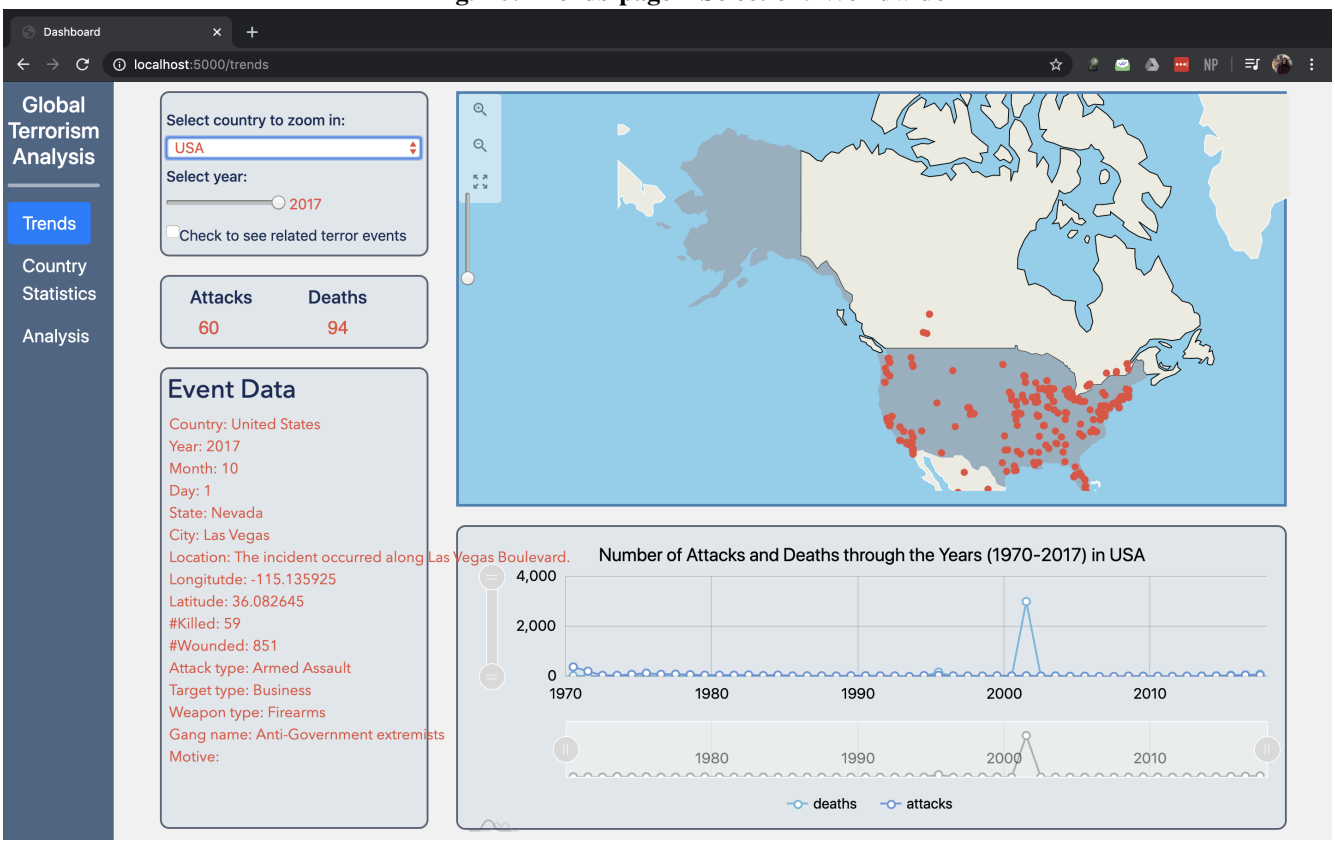

Fig. 21: Trends page - Selection: United States

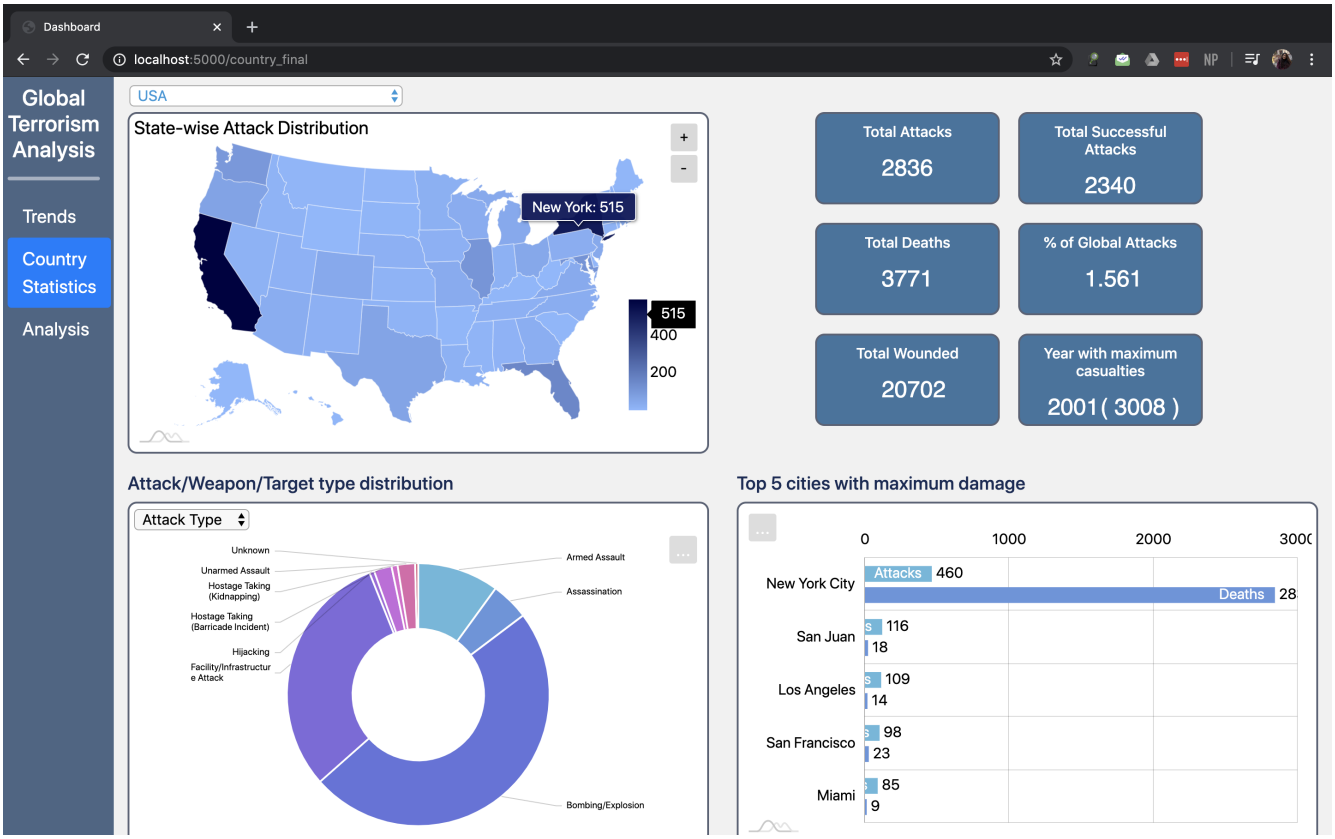

Fig. 22: Country Statistics page - Selection: United States

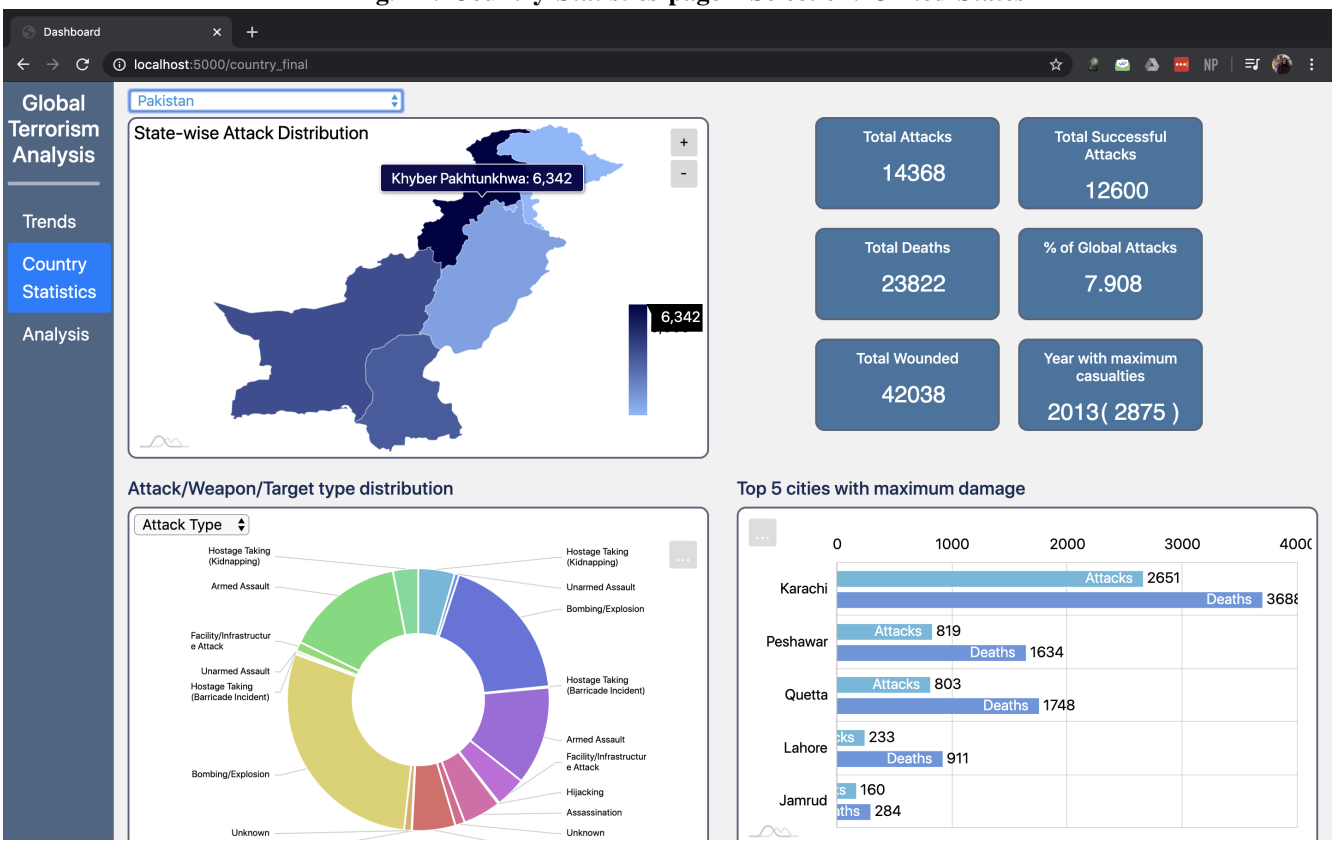

Fig. 23: Country Statistics page - Selection: Pakistan

Menu Driven Buttons and Tabs: Button are present on the home page (7). On the *Trends* page(8), there is a dropdown to zoom into a country and a checkbox to show related events on the map. Similarly, there is a dropdown to select country on *Country Statistics* page(9) and a drop down to select the pie plot to display of attack/weapon/target distribution. On the *Analysis* page (10), dropdown is present on the bar plot (figure 13) to select number of attacks or deaths. Another import menu button in each plot is the option to download the plot in different formats or download the data shown in the plots in different formats. The menu for this feature can be seen in Fig. 25

Sliders: A bi-directional Slider is present in line plots 11 and 20 at the bottom to select a range of year to zoom in or out. A separate time slider is also present on 8 to select a year for which the data is displayed on the map 4.

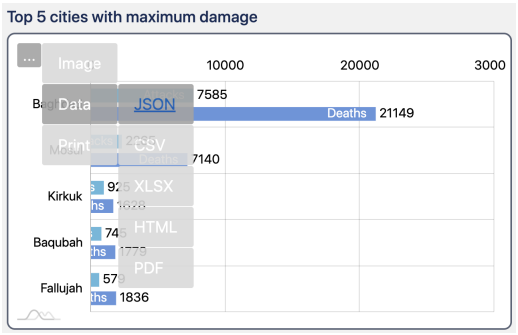

Fig. 25: The feature of downloading the plot in different formats like png, jpeg etc; downloading the data in plot and printing the plot.

D. *Analytics and Evaluation*

a) Level of interactivity achieved based on statistics collected by the system from User Data Exploration Tasks Performance: In II-C, we list a set of different types of interaction we′ve achieved in our system. What′s more important, in order to maintain a proper balance between the interactivity and the amount of information displayed in a page, we′ve tried to link multiple views together and each of them involves different types of interaction mechanism. For example, in our dashboard *Trends* page, upon selection of "Worldwide", users can use the slider to choose which year they′d like to explore. Then, the world map shows each event in the selected year with a bubble, and when hovering on it, the event information is displayed in the "Event Data" box. Also, there′s a line plot showing number of attacks and deaths caused by terrorist events through the years. This helps us to present the data in an organized and interactive way while ensuring the level of interactivity around 1s per action.

b) Data Insights and Findings:

I) Trends in Terrorist Events: We can see that in Fig.[26], in the 1980s, Central and South America are at the center of global terrorism, followed subsequently by a down period around 1995-2002 with average events per year less than 2500. The global terrorism started to increase after 2003, and rapidly after 2009. It reached the peak in 2013, where more than 15000 events were recorded. This peak coincides with the sharp increase in terrorist violence carried out by the ISIL (Islamic state of Iraq and Levant, also known as ISIS), particularly in Iraq, and by the Taliban in Afghanistan. The Highest number ever recorded for a year in a region is 6939 in Middle East, 2013.

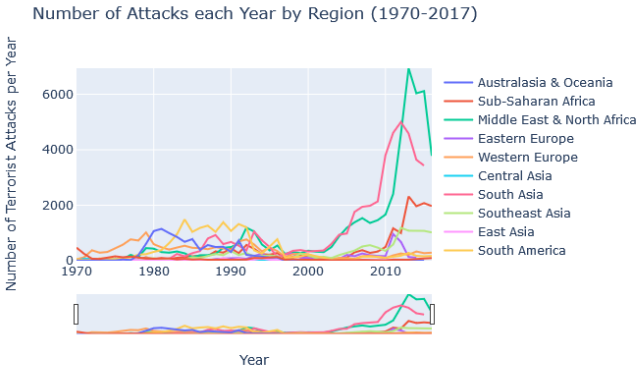

Fig. 26: Number of Attacks per Year by Region

II) Region-wise Statistics:

i) Top 3 types of attacks across all regions: Bombing/Explosion, Armed/Unarmed Assault, and Facility/Infrastructure Attack

ii) Top 3 most frequently used weapons: Explosives, Firearms, and Incendiary

iii) Most affected types of targets

- North America and Western Europe: Business, Private Citizen & Property
- South Asia: Private Citizen & Property, Police
- East Asia: Government, Business
- Middle East and North Africa: Private Citizen & Property, Military

III) General Findings:

i) Amongst the top 15 worst events of all time, 50% of the attacks are of type Hijacking or Hostage situation and the rest 50% are Bombing and Armed Assault.

ii) Percent of successful attacks : 88.96%

iii) Number of related events: 14k

iv) Iraq, Afghanistan, and Pakistan are the worst affected countries.

v) ISIL has carried out terror events in 50 countries resulting in mass casualties over the years,

becoming one of the largest and deadliest terror groups.

c) What are the Data Characteristics that will allow the REUSE of your visual system interface. Give concrete example.
The GTD data set is spatial-temporal, and it has spatial information for different scales, from region, country, to city. This allows us to explore and visualize the data on each level of the hierarchies.
Suppose we want to visualize and analyze current situation of COVID-19 and even make prediction about when will this global pandemic start to decline and to become stable. We can utilize the similar interface and visualization techniques. For example, a world map with a time slider that shows how the spatial distribution of all confirmed cases and deaths from the beginning till now, similarly to what we′ve done in the *Trends* page. In this way, we can observe how the center of this pandemic shifts from one region to another over time, then we maybe can predict which country and when it′ll be impacted based on the observation. Moreover, we can apply the country-wise analysis used in *Country Statistics* page to help policy-makers better and more easily analyze the situation on which they can base their decisions. Last, we can also calculate some statistics summarizing the whole COVID-19 data.

d) What type of Data is not suitable for the mode of exploration adopted by your system interface?
Our interface is more suitable for the type of data that is time-variant but static. We also expect the data with a certain extent of hierarchy in spatial and temporal information to fit our system interface much more smoothly.

e) What are possible enhancements that can be added to your system interface and/or processing?
Our current system focuses more on visualizing various patterns found in the data and facilitating the exploratory analysis. We′d like to incorporate the following parts to augment system′s functionality as well as usability.

I) User Query Interface: With a user query interface, researchers from different disciplines can apply their own analysis methods to explore the data. The data can be accessed by querying some attributes like the terrorist group, country, year etc.

II) Prediction: We′d expect to employ techniques like linear regression, neural networks or time series analysis on the data to help us make prediction on terrorist events in the future.

III) Feedback: Through a feedback interface we can gather valuable information from different types of users, which can help us to constantly improve user experience, interface layout, and the findings they are interested in.

E. *Development Documentation*

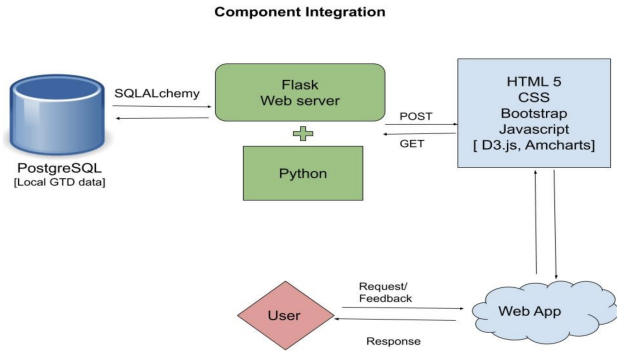

Fig. 27: Component Integration diagram

The components used in frontend and backend and the flow of data and query between them can be seen in Fig. 27.

a) Language(s) used
- Backend: Python
- Frontend: HTML 5, CSS, JavaScript
- Query: SQL in PostgreSQL

b) Basic Hardware - Software platform required (Backend, Front End)
- Database: PostgreSQL
- Web server: Flask, and we use SQLALchemy to handle the communication between the local PostgreSQL database and web server
- Front-end framework: Bootstrap

c) Software Libraries Used
- Plotting and visualization libraries: D3.js [3] and amCharts [4] which are Javascript based libraries.

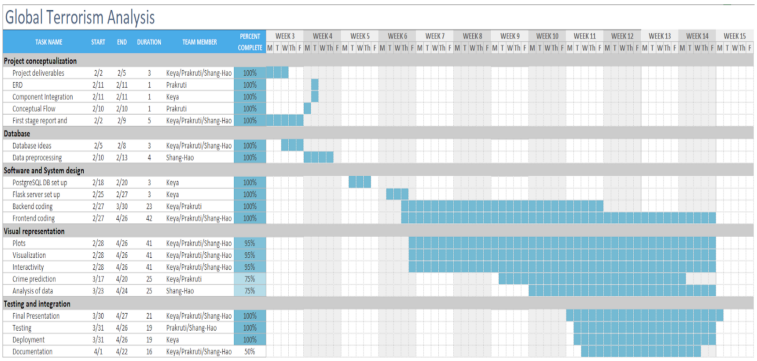

Fig. 28: Gantt chart

d) Data Sets Sources: The data set we use for this project is called Global Terrorism Database [5], which is maintained by START of the University of Maryland, and can be downloaded from Kaggle [6]. The data is available as CSV file.

e) Gantt chant: The chart which systematically documents our progress is given in Fig. 28

f) References from the literature: The references used in the development of the project are mentioned below in the references section.

III. Acknowledgements

We would like to thank Professor James Abello who was given excellent guidelines and has constantly motivated us throughout the project. We would also like to thank Fangda Han and Nishant Kumar for giving us insightful feedback on the system conceptualization, interface design, and implementation.

References

[1] G. LaFree, "The global terrorism database: Accomplishments and challenges," *Perspectives on Terrorism*, vol. 4, no. 1, 2010. [Online]. Available: http://www.terrorismanalysts.com/pt/index.php/pot/article/view/89

[2] Stevepaul. (2019, May 20) Hadoop advantages and disadvantages. Accessed: 10-Feb-2020. [Online]. Available: https://www.mindsmapped.com/hadoop-advantages-and-disadvantages/.

[3] M. Bostock, V. Ogievetsky, and J. Heer, "D3 data-driven documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, p. 2301â2309, Dec. 2011. [Online]. Available: https://doi.org/10.1109/TVCG.2011.185

[4] amCharts. Javascript charts maps - amcharts. [Online]. Available: https://www.amcharts.com/

[5] START. (2020, May 20) Gtd — global terrorism database. Accessed: 10-Feb-2020. [Online]. Available: https://www.start.umd.edu/gtd/

[6] START Consortium. (2018, July) Global terrorism database (version 3). Accessed: 3-Feb-2020. [Online]. Available: https://www.kaggle.com/START-UMD/gtd