Visual Analytics on Global Terrorism

CSE 6242 | Progress Report | Team INTERPolo | December 5, 2017

Cooper Colglazier | Eugene Kang | Huaiwei Sun | Minghan Xu | Vineet Pasupulety | Wonhee Lee

---

Introduction and Motivation

In lieu of increased terrorist activity in recent years, it is meaningful to investigate terrorism and its causes over time. Drawing conclusions about motives of terrorists and finding links between socio-economic aspects and terrorist activities is tough yet necessary for governmental bodies to respond amicably in the long run. Our aim is to combine key information from these datasets and discover important trends in attacks in terms of time, place and circumstance and provide efficient visualizations. We not only attempt to draw correlations with external factors such as education, poverty, and religion employing machine learning techniques, but also create an interface for users to perform their own analysis using the data attributes. We hope that this analysis can assist counter-terrorism efforts and raise awareness amongst the general public, intelligence agencies and governments.

Problem Definition

Current analysis and visualizations of terrorism only indicate what is happening in terms of location, attack type, and the number of victims with limited interactivity. In general, why these events occur is not well understood. To effectively counteract terrorism, understanding why and what factors contribute to fostering terrorism are critical.

Survey

Terrorism

Terrorist activity has increased recently and each incident is 17% more likely to result in death or injury compared to the pre-Cold war era [H]. Terrorist attacks in the U.S. have varied with respect to lethality, location, type of weapons, targets since 2000 [I][J]. However, finding a general explanation of the sources of terrorism is not a simple task [L]. Many researchers have tried to identify the causes and backgrounds for terrorism such as education, poverty and terrorism [K]. Their analysis showed that there was little conclusive evidence to suggest direct connections between these factors.

Analytics

Previous analysis done with respect to terrorism has employed different regression and clustering techniques. Many studies attempt to extract patterns or inter-dependencies from within the dataset using clustering techniques such as modified K-means partitional algorithm [M] or regression techniques such as hierarchical regression analysis to draw correlations between culture and terrorism [N]. Survival analysis could be exploited to estimate which kinds of activities alter the frequency of terrorist events and which kinds of activities reduce the probability of future terrorist events [Q]. Improved visualizations [R] and computationally efficient social network analysis tools [S] for large scale criminal networks provide us motivation to implement them for terrorism datasets. Another motivating work uses HADOOP to mine relations between terrorist group ideology and terrorist attack type from Google News data [T] but doesn't integrate vital socio-economic backgrounds of victims and terrorists.

Visualization

Several visual analytic systems have been proposed to explore terrorism data. GTD Explorer is a web-based visualization using GTD (Global Terrorism Database) dataset. It supports interaction and is easily extendable to support different types of data [D]. However, this traditional visualization might cause distraction when visualizing geospatially uncertain events within large datasets. To avoid this, using a squarified treemaps that maintains the strength of traditional map but incorporates some measure of data verity [E]. To make a visual analytical system insightful, it has been suggested to focus on depicting the five W's(who, what, where, when and why). This approach allows us to think both tactically and strategically [F]. Also, the unified visualization environment, which could present various types of patterns and thus to facilitate explorations of the incident data from different perspectives, has been introduced [G].

Proposed Method

Intuition and Hypothesis

We believe by providing a tool that allows for visual and analytic exploration of a large terrorism dataset, conjoined with data that indicates levels of development of a country and region, we will allow the global community to better understand why terrorism happens, and in turn influence policy that aims to curb terrorism.

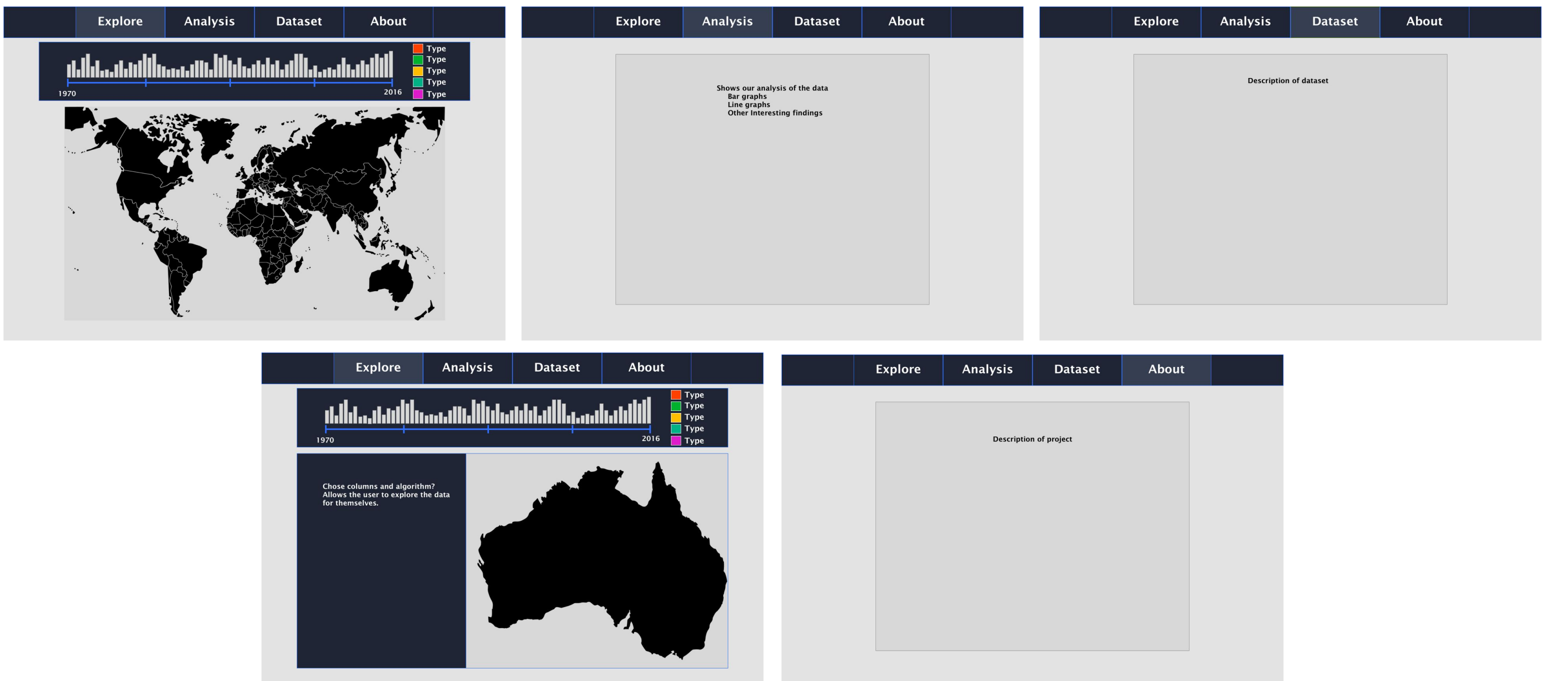
Innovations

1. Incorporate key socio-economic factors from external data to identify relationships beyond the Global Terrorism Database (GTD)
2. Interactive machine learning algorithm and visualization on GTD
3. User exploratory driven visualization design

Approaches

User Interfaces

We first created two interactive prototypes of the web page interface. Ability to explore terrorism events spatially, filter by time, and filter by specific features of our dataset were key interactions. One prototype is shown below.



After finishing the prototype, we discussed which functions and interactions should be included in the final website. We walked through each feature in the prototype and perform evaluation in terms of complexity feasibility, and resources. We then implemented the following two features and subfeatures in D3.js:

Explore Data Through Interactive Map

We adopted an “Overview to Details” design model and implemented an interactive visualization system for users to complete the sense-making loop including searching, filtering, schematizing, case-building and telling stories. The system allows users to firstly explore the data through an interactive terrorist activities map. The occurrence of terrorist attacks are visually encoded by the density of the points, higher density indicates a higher number of attacks. With zoom-in and zoom-out functionality of the map, users are able to view data points at different geo-granularity and synthesize their observations. To reveal more interesting insights, we build range slider for the user to dissect data in the time domain which helps them perform the visual analysis between different time windows in the history. Once users acquire high-level information from the above interactions and thinking processes, they will proceed to the next function for refined analysis.

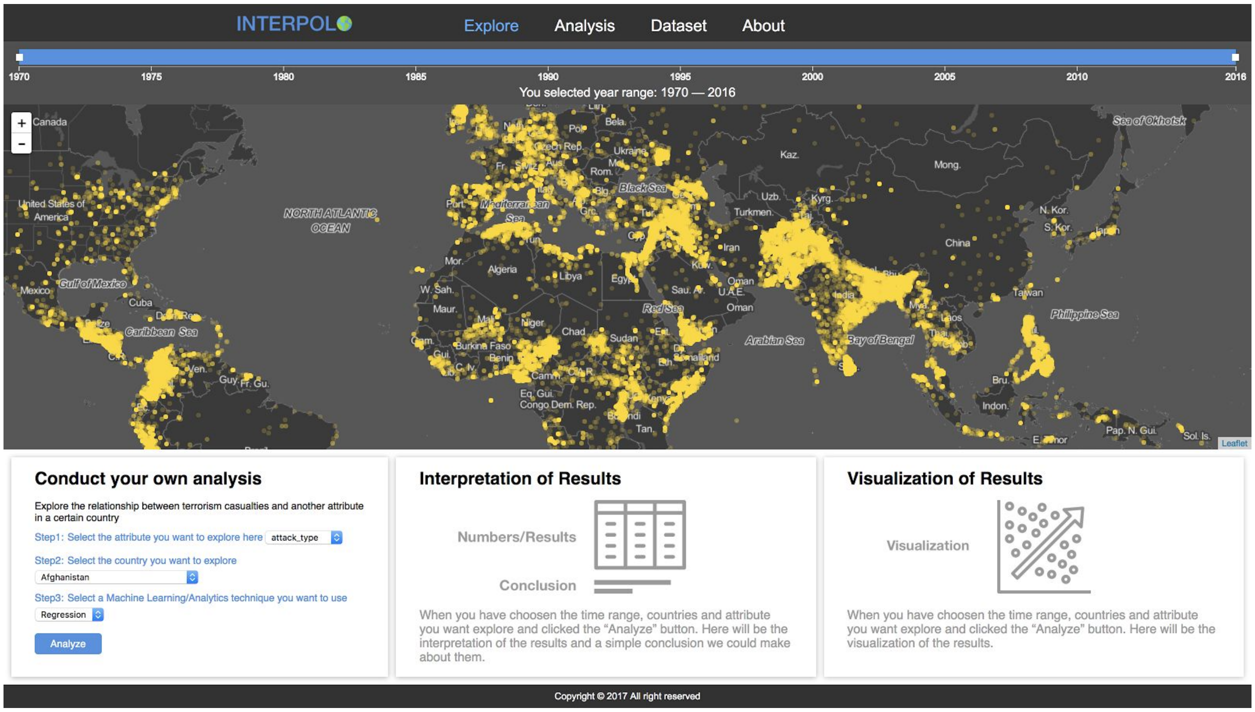


Figure: Overview of the "Explore" page

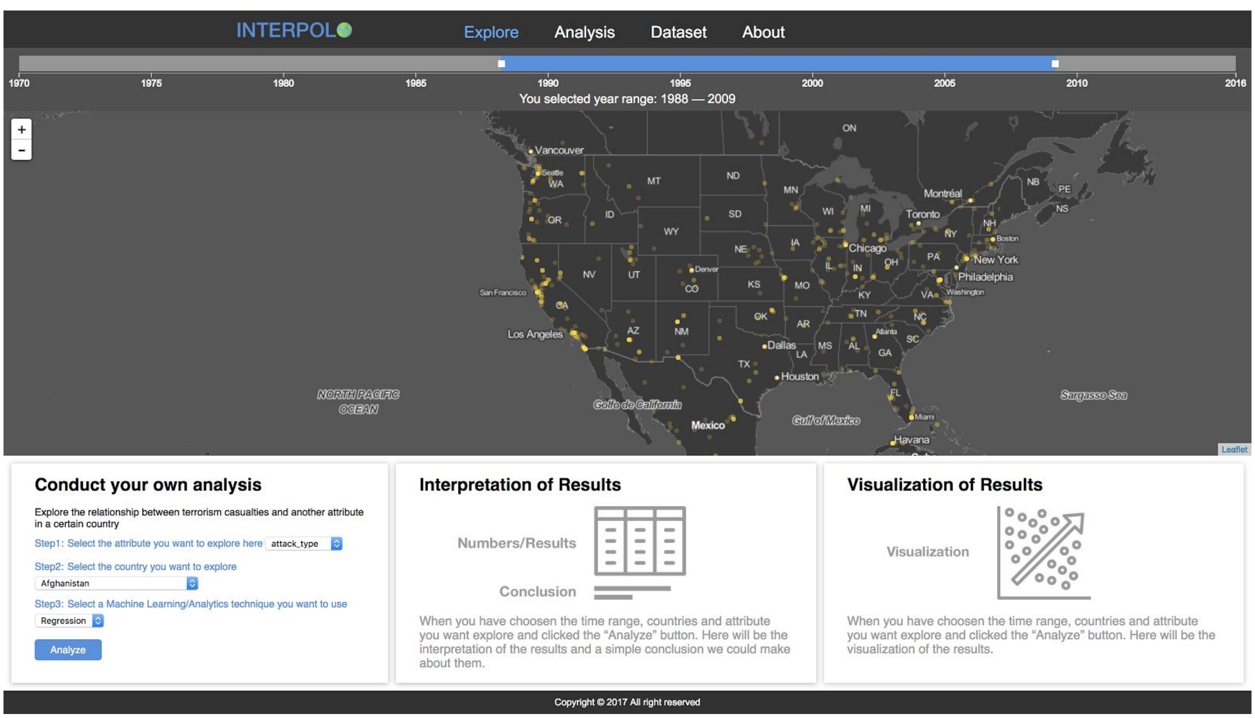


Figure: Use slider to select time range and zoom in/out for focused study

Analyze Data Using Linear Regression

The user can choose the attribute and country in his/her interest to perform the analysis. Once the user selects inputs, Interpolo builds a linear regression model to learn the relationship between casualties and user-specified predictors. The system returns the model coefficients back to the interface for statistical interpretation. If certain predictor causes more casualties in the terrorist attacks, its coefficient will be higher. Such implementation provides a unique model for the user to derive personalized insights.

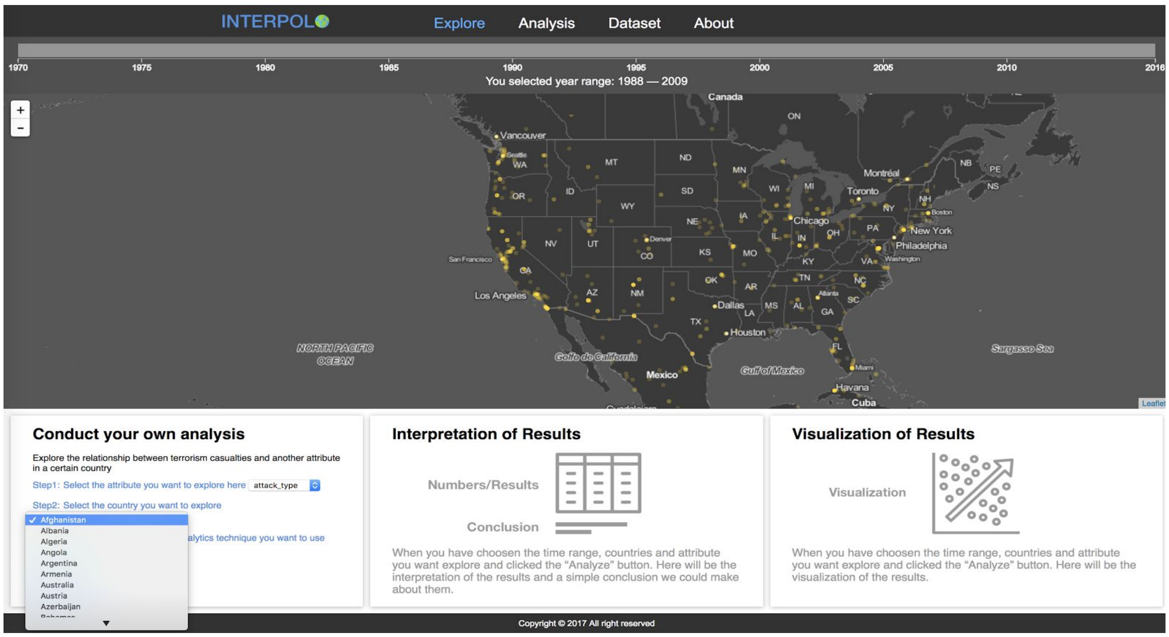


Figure: Select one country users want to study.

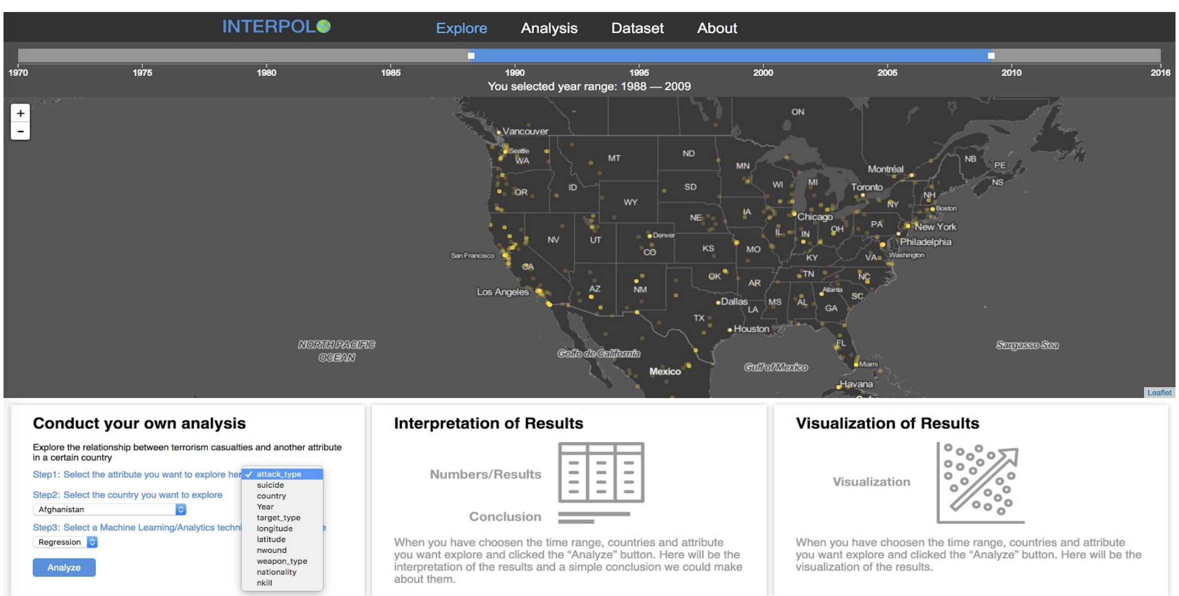


Figure: Select the attribute interested in, e.g., attack_type or weapon_type.

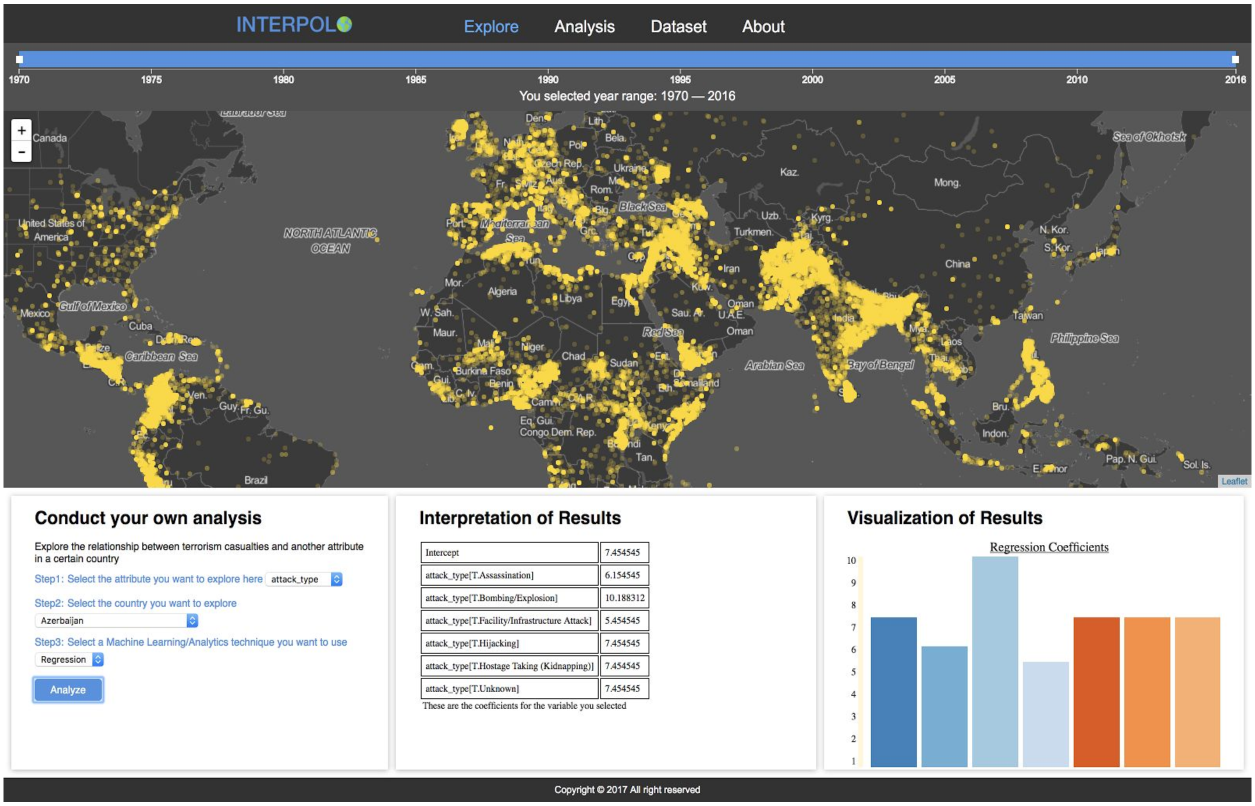


Figure: Press the “Analyze” button and get the results of the analysis.

Gain additional insights from Tableau Dashboard

We embedded our interactive visualization of basic Tableau analysis of the terrorism in our website and helped users learn more about weapons and targets involved in the terrorist events over time.

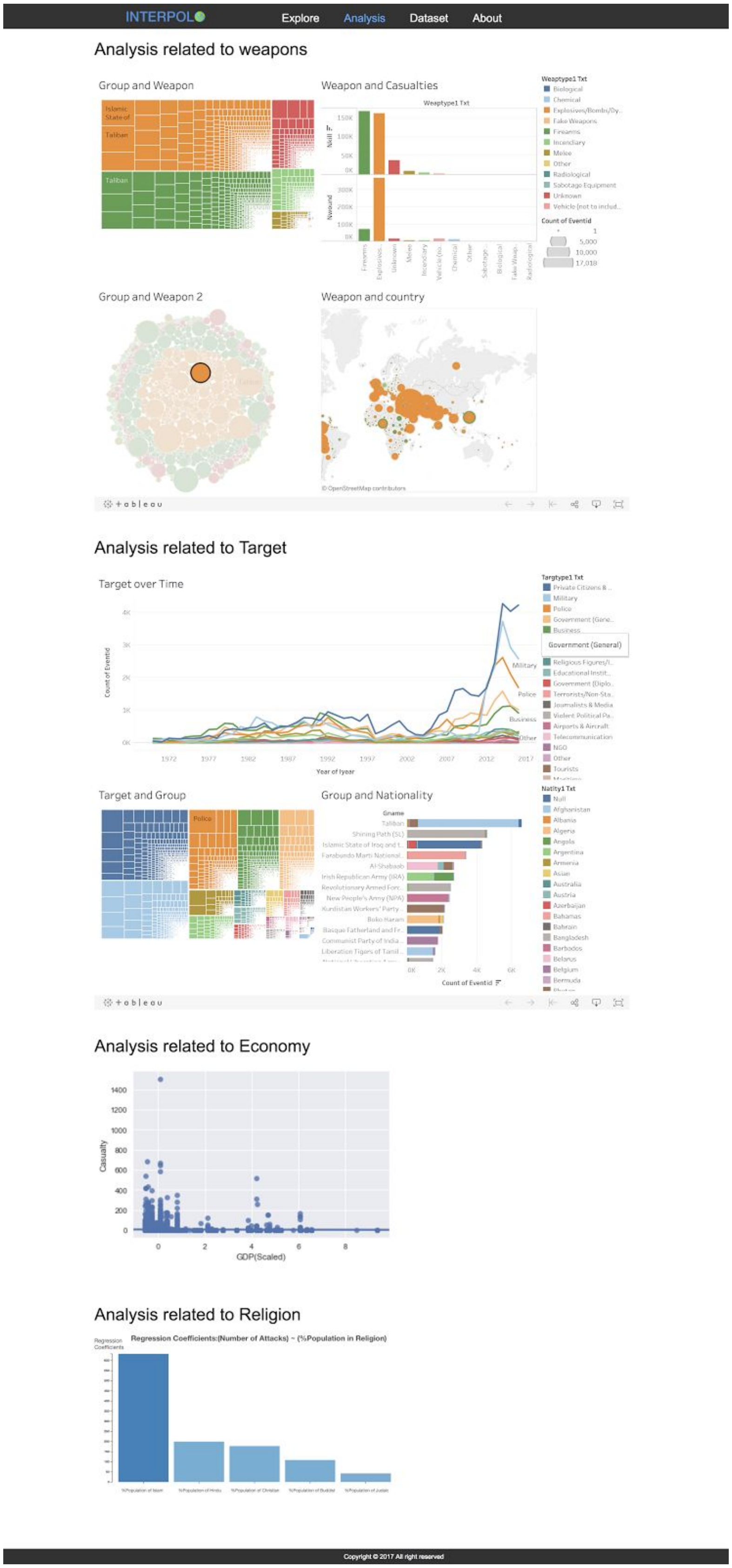


Figure: “Analysis” page

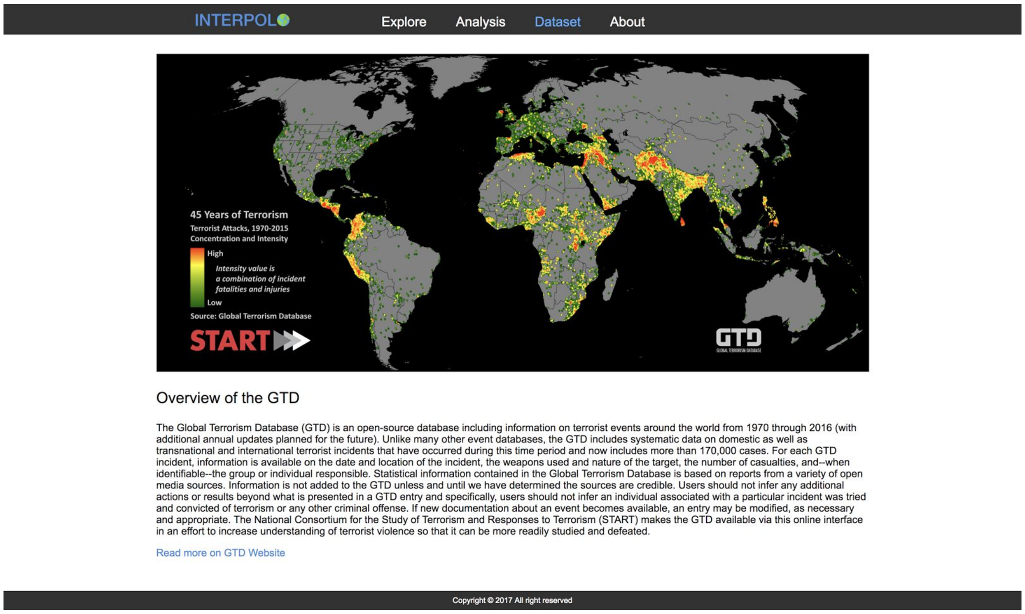


Figure: “Dataset” page: Overview of Global Terrorism Dataset

Figure: “About” page: Introduction to our project and team.

Analytics

Regression

Linear regression analysis attempts to measure the linear relationships between explanatory variables and a response variable. By comparing how the response varies with respect to predictors, we are able to identify trends and correlations. Furthermore, we are able to measure the magnitudes of such effects by comparing the coefficients of the regression model.

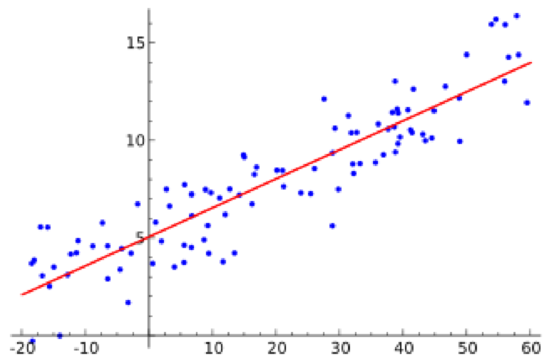


Figure: Illustration of Linear Regression

We first implemented the regression algorithm to draw relationships between the number of casualties and different variables in the Global Terrorism Database, such as attack type and, target type, and weapon type. By allowing the user to choose the predictor, we enable the user to compare the magnitudes of the effects. For instance, bombing as an attack type may have a higher regression coefficient than firearms; the user is able to compare all the different factors of a categorical variable in this manner.

Furthermore, we explored which external factors affect or correlate to terrorist attacks around the globe. Based on various research we conducted concerning the motives of terrorist attacks, we made an educated guess on which factors could potentially be correlated to terrorist incidents. For example, GDP of the country in which the incident occurred may be a significant factor to analyze the terrorist attack. In order to incorporate such external factors, we cleaned and merged other datasets with the Global Terrorism Database. This approach enabled us to widen the scope of our analysis.

For instance, we tested whether economic indicators of a country were correlated to the number of terrorist activities in that country and the number of casualties. To assess this, we employed a regression model testing the relationship between various economic indicators and outputs referred to above. The resulting R^2 values were low, illustrating no particular correlation between the predictors and the response.

Furthermore, we tried to gauge whether the prevalence of different religions in the country in which the activities occurred was a major factor. Again, taking the World Religion Dataset from the Association of Religion Data Archives, we ran a regression model with percentages of the population that followed a particular religion as predictors and the number of incidents in that country as the response. Again, the R^2 value was low, even when we limited the number of predictors to major religions. However, one notable observation was that no matter what variables we chose, Islam seemed to be a significant predictor based on p-value.

Throughout our research and survey, we had read about the difficulty of identifying trends in terrorist activity; the nature of such activities is that often times, individuals or small groups engage in attacks based on a wide variety of motives, making them difficult to generalize. Given such circumstances, it did not come as a surprise that the coefficients of regression were small and the correlations between the various predictors and the response were weak. Due to the nature of terrorist activities being largely independent from each other, discovering relationships that satisfy the majority of the data may be challenging.

K-Means Algorithm

In exploratory data analysis, we implemented K-means algorithm with sklearn and matplot libraries. We are able to identify clusters of geolocation (longitude-latitude coordinates) where terrorist attacks occured. Due to the complexity of integration, we recommend to incorporate this feature in the future work.

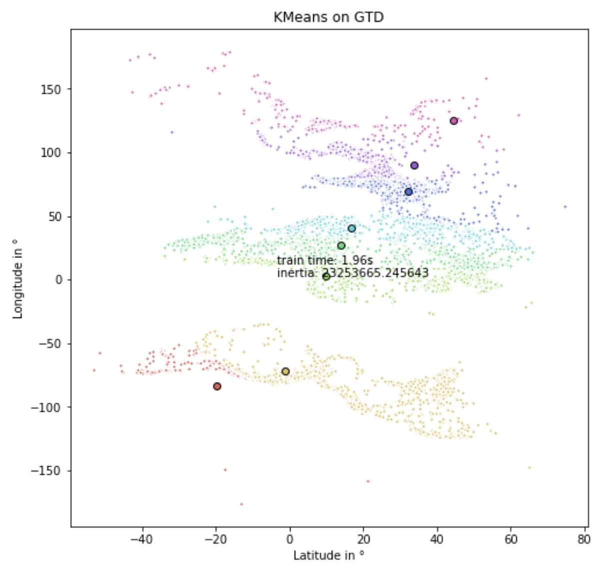


Figure: PCA results on Longitude-Latitude

:

PCA & Correlation

Principal components analysis (PCA) Correlation matrix are useful techniques to find associations between predicting variables. We have reported our results in the progress report and these two techniques helped the team understand the dataset and facilitate the design and implementation of regression analysis.

Text Analytics

In an attempt to mine deeper insights from events, we extracted the summaries of news articles describing the event. We chose to simultaneously extract news headlines from the 1 Million Headlines Dataset [U]. Topic Modelling in the original article is done by creating a Latent Dirichlet Allocation (LDA) model and correlations between these topics and those in the corpora were measured. The results were unpromising and we hope to improve the methodology in the future.

Visualization

Interactive Map

The GTD dataset contains attribute columns for latitude and longitude for each terrorism activity. Two main strategies have been identified for visualizing events on map:

1. Creating static world map using d3-geo library
2. Using JavaScript-powered map APIs

The strength of the first approach is its general simplicity in implementation, but it comes with the limitation of zooming and panning. Understanding user interaction is a key component in visual analytics, we executed the second approach.

The team originally built the visualization using d3 with Google Map API. The challenge we faced was its performance. Plotting over 10,000 data points using Google Map API and d3 would slow down the browser and make it unresponsive.

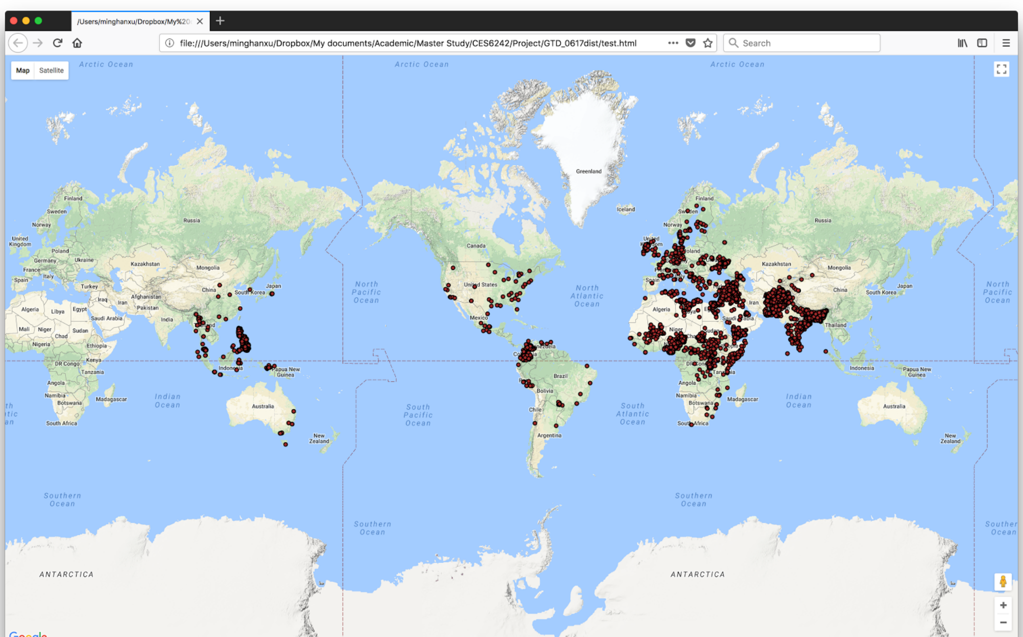


Figure: 1st map implementation using d3 and Google Map API

Hence the team switched to leaflet, an open-source JavaScript library for mobile-friendly interactive maps. Leaflet is a light library to build visualization and we were able to load a much larger dataset on a single interactive map. To further enhance system performance, we adopted dynamic plotting strategy where optimized level of abstraction is applied and the program only plots data points that are visible to users. When the user performs panning and zooming on our interactive map, the program will adjust the plotting window and re-plot the data points within the dynamic plotting region. By using leaflet library, data abstraction and dynamic plotting, we are able to visualize ~170,000 terrorism records on the interactive map.

Range Slider

After we enabled the interaction on the geospace, we extended user interaction to the time domain. Interactive range slider was chosen for implementation as it allows the user to perform range selection in a very natural manner. A standalone year range slider was constructed with left & right handles adjustment, drag and click features. User's current selection will be dynamically displayed below the slider to assist user to make refined adjustment.

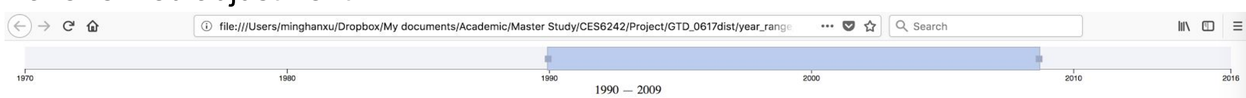


Figure: Interactive Range Slider

With the interactive map and range slider, the team integrated these two important visualization components. To ensure smoothness in interaction, we designed and built our program to respond new range selection of years only after user actions are completed (mouse released). The program will subset the dataset and replot the points based on filtering criteria (time horizon and geolocation). Such integration enables user-focused data visualization and provides natural and intuitive interactions, which outperforms current visualization system on terrorism activities.

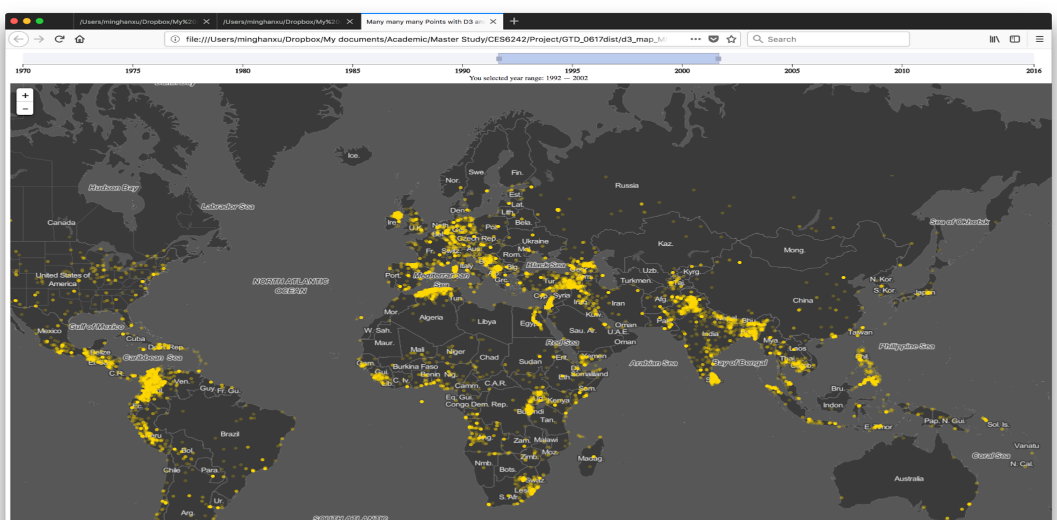


Figure: Prototype of Interactive Map

D3 and leaflet are powerful tools for visualizing geo-information of terrorist attacks. As the team planned to extend the visualization and cover other data attributes, we looked for other effective solutions to build interactive visualizations. We chose Tableau and Tableau Public for fast deployment of data dashboard. We built interactive dashboards on Tableau Desktop and host them on Tableau Public. To integrate them into our system, we applied Tableau Public API and embedded dashboards on the Analysis tab of our website.

Server and Backend

The team built the back-end system to support host-server communication. Hence user inputs on the HTML page were collected and sent to the backend server for fast analysis. The results of data analysis could be returned to the webpage for visualizations and user interpretation. We chose Flask to setup our Web sockets for communication between the backend server and frontend webpage. Our back-end was written in Python due to versatility in data analytics and availability of stable communication protocols. We used Jinja2 templates to allow the frontend HTML page to use data stored on the backend to serve up the options in the dropdown boxes. Flask was built for the communication protocol to ensure data was received from the frontend at the appropriate sequence instances and the correct locations. Flask helped us route server GET and POST requests through the appropriate API routes to ensure efficient and smooth transfer of data, while rendering the right pages on the particular requests. We used AJAX requests to communicate between the back-end analytics functions and the D3 Javascript so that the results of our analysis could be appropriately passed to the visualisation functions.

Experiments and Evaluation

The goal of the experiment is to test if our interactive prototype could help the public gain knowledge about terrorism in an easy and intuitive manner. To evaluate our system, we employed user tests that focused on the participant's preference between an existing terrorism data exploration tool (GTD WebGL Globe) and InterPolo. Our testbeds were in a controlled setting. We conducted in-person moderated usability tests. The participants were our classmates and friends.

We recruited 6 participants in total. We adopted the within-subject procedure and showed the two websites to every participant. They finished the tasks assigned by us. They thought aloud along the process and answered qualitative questions after they finished all the tasks on two websites. These questions let our participants to self-report preferences between the two systems based on six subjective metrics.

The questions are as follows:

- Which website interface do you feel is easier to learn?
- Which website interface do you feel is easier to use?
- Which website interface do you feel is more helpful in learning about terrorism?
- Which website interface do you feel is more enjoyable to use?
- Which website interface are you more likely to use in the future?

We documented their votes and visualized it in a bar chart. The results show that our interface is generally preferred over the existing implementation. Further prototype iteration is needed to improve the interface usability and effectiveness.

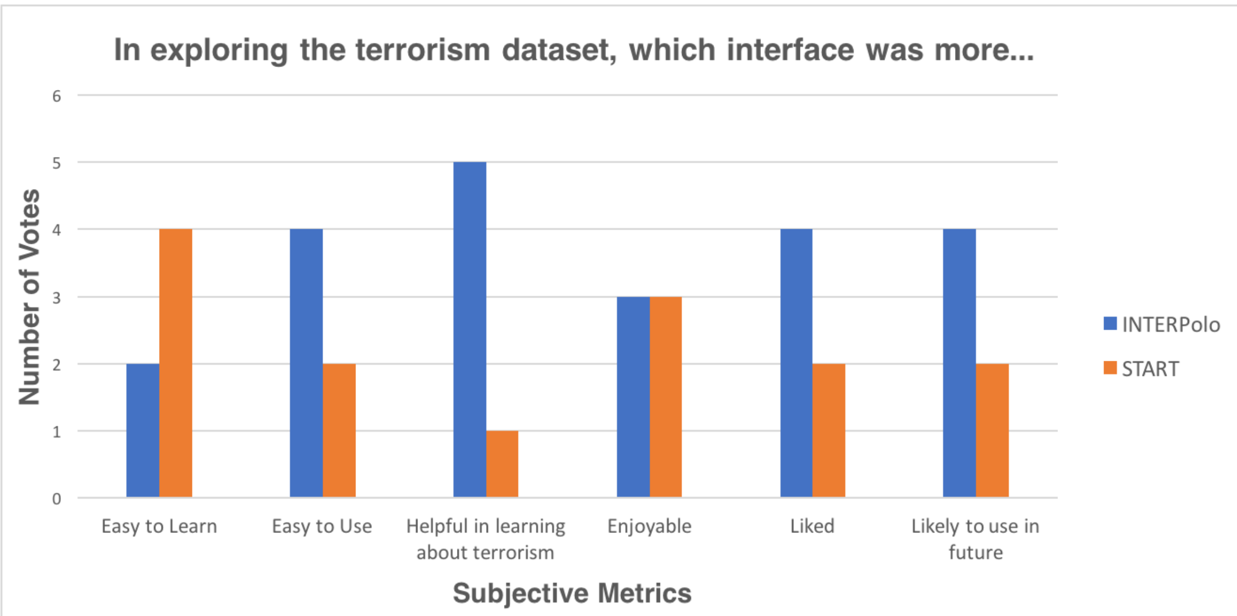


Figure: Experiment and Evaluation Results

Conclusions and Discussion

Throughout the semester, our team collaborated closely, with each member contributing unique technical expertise in similar amounts of effort to successfully develop a visual analytics tool to better understand terrorism. Using our web app, one can easily navigate and visualize the global terrorism database. Moreover, we have provided tools for further analysis so that people can better understand why and how terrorism occurs. This concept of interactively visualizing a dataset and allowing for exploration and discovery is one that could go beyond terrorism. Although it requires a team to construct the tool, the end result is data that is more accessible to people who are interested.

References

A. Yvonne Rogers, Helen Sharp, and Jenny Preece. 2011. Interaction Design: Beyond Human - Computer Interaction (3rd ed.). Wiley Publishing.
B. The World Factbook 2017. Washington, DC: Central Intelligence Agency, 2017.
C. World development indicators. Washington, D.C. :The World Bank.
D. Joonghoon Lee. 2008. Exploring global terrorism data: a web-based visualization of temporal data. Crossroads 15, 2 (December 2008), 7-14.
E. Jones, Josh, et al. "Visualizing uncertainty for geographical information in the global terrorism database." SPIE Defense and Security Symposium. International Society for Optics and Photonics, 2008
F. Wang, X., Miller, E., Smarick, K., Ribarsky, W. and Chang, R. (2008), Investigative Visual Analysis of Global Terrorism. Computer Graphics Forum, 27: 919–926. doi:10.1111/j.1467-8659.2008.01225.x
G. Guo, D., Liao, K., & Morgan, M. (2007). Visualizing Patterns in a Global Terrorism Incident Database. Environment and Planning B: Planning and Design, 34(5), 767-784.
H. Enders, Walter and Sandler, Todd. Is Transnational Terrorism becoming more threatening?
I. Overview: Terrorism in 2016 (University of Maryland, August 2017)
J. Miller, Erin. Patterns of Terrorism in the United States, 1970-2014. College Park, MD: START, 2016.
K. Krueger, Alan and Maleckova, Jitka. Education, Poverty and Terrorism: Is There a Causal Connection? Journal of Economic Perspectives - Volume 17, Number 4 - Fall 2003
L. Crenshaw, Martin. The Causes of Terrorism. Comparative Politics, Vol. 13, No. 4(Jul., 1981)
M. Ben Salem, Semeh & Naouali, Sami & Sallami, Moetez. (2017). A Computational Cost-Effective Clustering Algorithm in Multidimensional Space Using the Manhattan Metric: Application to the Global Terrorism Database.
N. Sofia Pinero Kluch & Alan Vaux (2017) Culture and Terrorism: The Role of Cultural Factors in Worldwide Terrorism (1970–2013), Terrorism and Political Violence, 29:2, 323-341, DOI: 10.1080/09546553.2015.1038523
O. Pattern Recognition Approach in Multidimensional Databases: Application to the Global Terrorism Database
P. Górecki, Jan, Kateřina Slaninová, and Václav Snášel, Visual investigation of similarities in global terrorism database by means of synthetic social networks, Computational Aspects of Social Networks (CASoN), 2011 International Conference on. IEEE, 2011.
Q. Stevenson, John. Statistical Analysis of Event Data Concerning Boko Haram in Nigeria (2009-2013) (January 2015)
R. Ferrara, Emilio, Pasquale De Meo, Salvatore Catanese, and Giacomo Fiumara, Visualizing criminal networks reconstructed from mobile phone records, *arXiv preprint arXiv:1407.2837*(2014).

S. Ferrara, Emilio, Salvatore Catanese, and Giacomo Fiumara. "Uncovering criminal behavior with computational tools." In *Social Phenomena*, pp. 177-207. Springer International Publishing, 2015.
T. Strang, Kenneth David, and Zhaohao Sun. "Analyzing relationships in terrorism big data using hadoop and statistics." *Journal of Computer Information Systems* 57, no. 1 (2017): 67-75.
U. Rohit Kulkarni (2017), A Million News Headlines [CSV Data file], Retrieved from https://www.kaggle.com/therohk/million-headlines