Abstract

In this project, I explored Global terrorism dataset which contains information about global terrorism acts since 1970. I attempted to make the findings as visual as possible using D3. Many visualizations like bar chart, line plot, geo map, Sankey diagram, histogram etc. Ire created in the process for purpose of understanding and analyzing data better.

I then developed some hypotheses from visualization and implemented a model for prediction purpose. Depending on the model which uses complete data set I will predict the terrorist group name which may be behind the attack

*Keywords:* Sankey, Bar plot, ggplot, Histogram

**Terrorist Groups activities analysis**

Terrorist attacks over the world are increasing day by day. They result in large number of casualties and are considered as global threat. Mostly the people affected are the general public. Considering this need of observing and discovering relations and rules of behavior based on terrorism incidents becomes very important.

The dataset used for this project is global terrorism data and can be found at [Kaggle](https://www.kaggle.com/anshumanyp/rossman/data). The dataset is an open smyce database which contain information on terrorist attacks around the world from 1970 – 2016. Data for year 2013 is missing from the dataset.

It contains 170,350 terrorist incidents worldwide, with 135 different attributes for each incident. In order to simplify and make the analysis much more efficient, I chose to discard over 100 attributes, and only retain a few of the key variables such as date, attack type, location, target type, casualties, group name and Iapon type.

**Data Understanding**

* *Description of the Data.*

The dataset contains 135 variables and most of them are not important for my analysis therefore I are going to focus on a subset of variables for the sake of keeping my analysis clear.

Variable of importance with their descriptions can be found below:

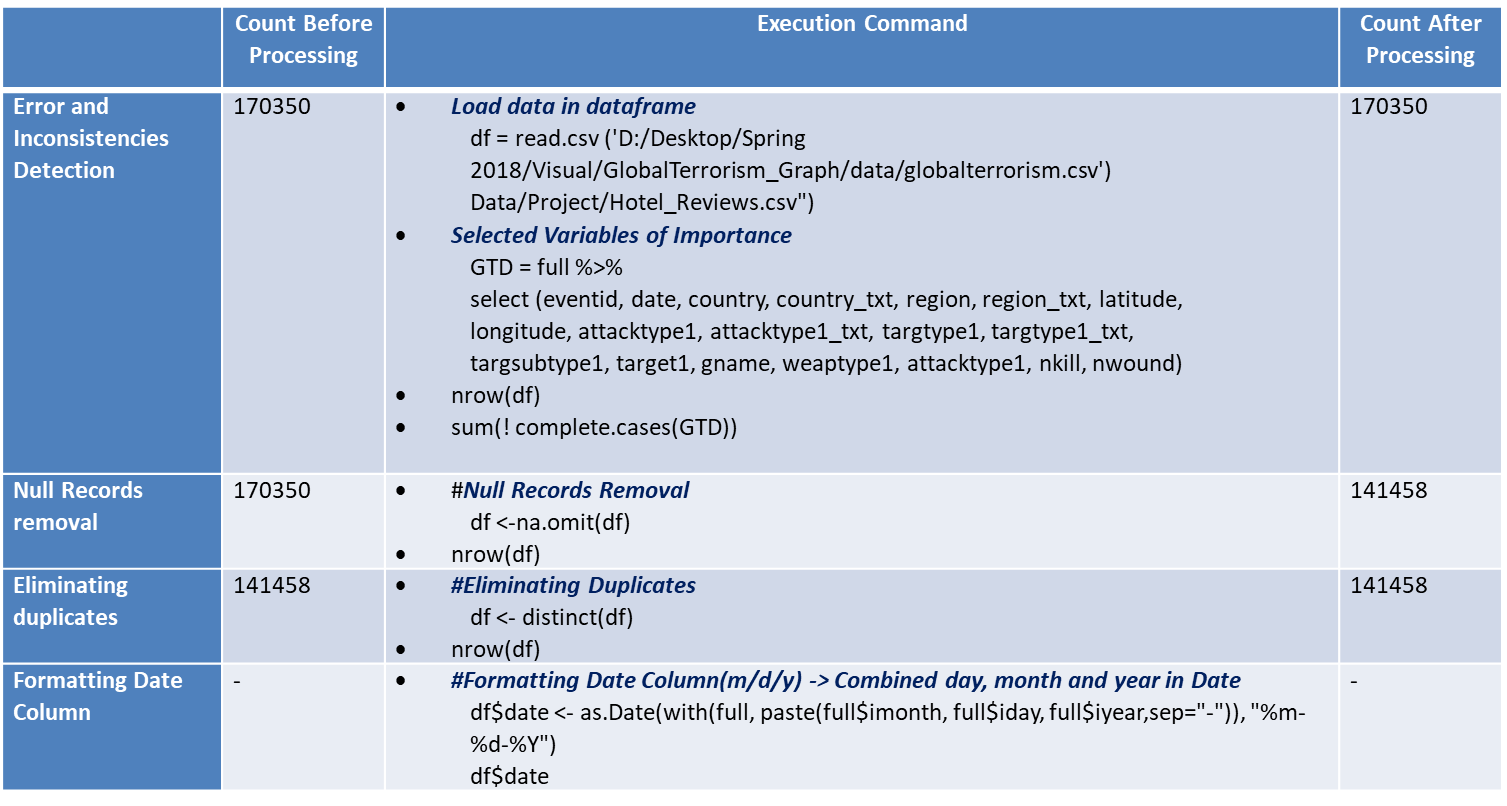
* **Iyear, imonth, Iday** => Three attributes combined in one called “Date”.
* **Country\_txt** => Contains name of country where attack took place.
* **Region\_txt** => Contains Region name.
* **Lat, lon** => Gives the exact location of attack on map.
* **Attacktype1\_txt** => Type of attack like Bombing, Armed Assault etc.
* **Targtype1\_txt** => Targeted type like Government, Citizen, Military etc.
* **Gname** => Name of the terrorist group responsible
* **Iaptype1\_txt** => Type of Iapon used for attack
* **Nkill** => number of people killed
* **Nwound** => number of people wounded

**Data Preparation**

For the data preprocessing part, slicing for the variable of importance cleaning of the data by eliminating null values was done. Combined some attributes to create a new variable and rename few attributes.

* Null Records Removal
* Eliminating Duplicates
* Combining day, month, year in new variable date
* Renaming of few attributes

Following are the commands used and result after preprocessing:



**Case Study**

For my project I will be concentrating on two case studies:

1. ***Deeper look into terrorism to determine the trend***

J. Górecki, K. Slaninová and V. Snášel (2011) in their paper states

“Terrorist attacks have been practiced by a wide array of organizations or groups for achieving their objectives. I can include political parties, nationalistic and religious groups, revolutionaries, ruling governments or others. Due to this fact the need of observing and discovering relations and rules of behavior based on terrorism incidents becomes very important.” (pp. 255-260)

It becomes important to analyze data and determine the relation betIen them in order to prevent attacks.

1. ***Prediction of responsible terrorist group***

“Several machine learning algorithms Ire trained on the Indian subset of the Global Terrorism Database to learn to predict the perpetrator of a terrorist attack, given data about the types of attack, target and Iapon in addition to the location, year and other attributes of the event. It was found that Support Vector Machine technique gave accuracy higher than 75% in predicting the perpetrators. This approach has the potential to aid investigating agencies and carries significant implications for national and international security.” (D. Talreja, J. Nagaraj, N. J. Varsha & K. Mahesh, pp. 1723-1726)

**Case Study 1 – Deeper look**

* ***Terrorist Attacks over time***: Using ggplot and histogram I created histogram to show terrorist attacks over time. From the graph we see that over 170000 attacks happening, and they seem to have gone up.

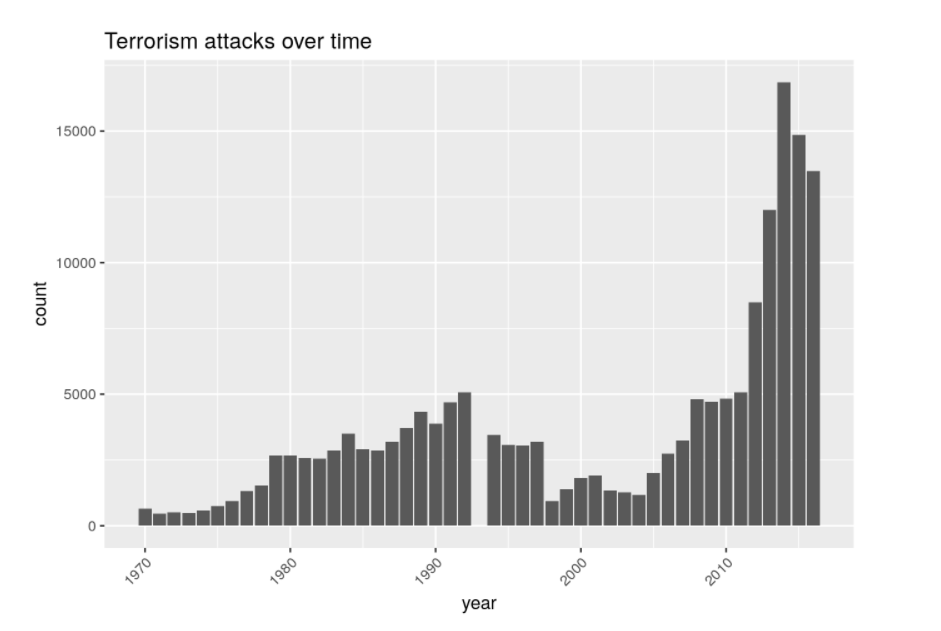


Figure 1- Terrorism Over Time

* ***Locations of terrorism***: Middle East and N Africa (27% of total), South Asia (24%) and S America (11%) are the top three regions in terms of number of attacks. Iraq (12.9% of total), Pakistan, (8%), Afghanistan (6.6%), India (6.4%) and Colombia (4.7%) are the top five countries in terms of number of attacks.

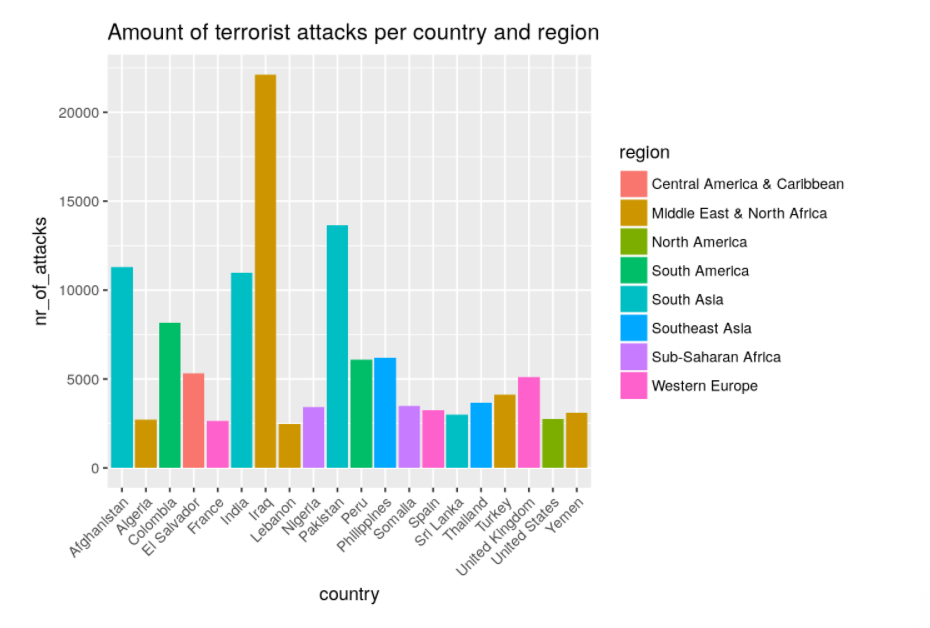


Figure 2-Terrorism Location

* ***Activity of groups over time:*** As we can see from plot, terrorist activities are rapidly increasing over time.

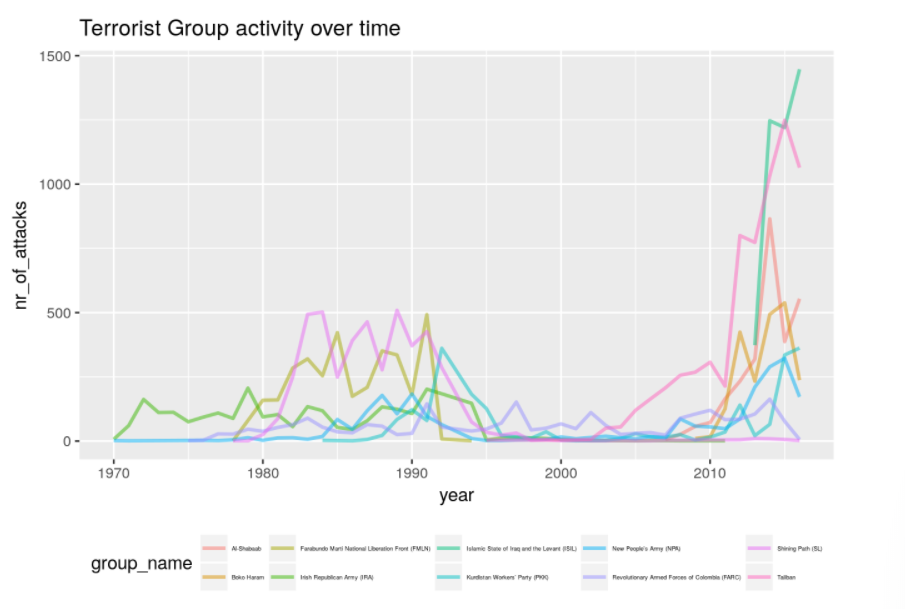


Figure 3- Activity Over Time

* ***Iapon choice over time***: As I can see from visualization, Bomb Attacks are the deadliest and have caused half of all casualties. The next deadliest Iapon grouping is "Firearms" responsible for 32% of all Terror attack deaths.

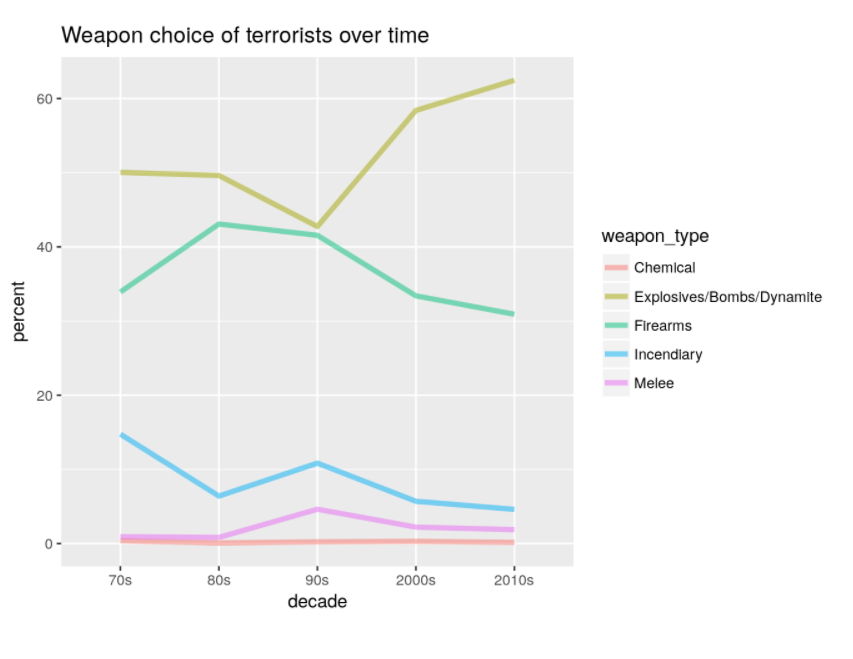


Figure 4- Iapon Choice

* ***Target choice over time***: Private citizens have become the first target. It seems that violence has escalated to this innocent group. Besides that the Military has also become a bigger target over the decades.

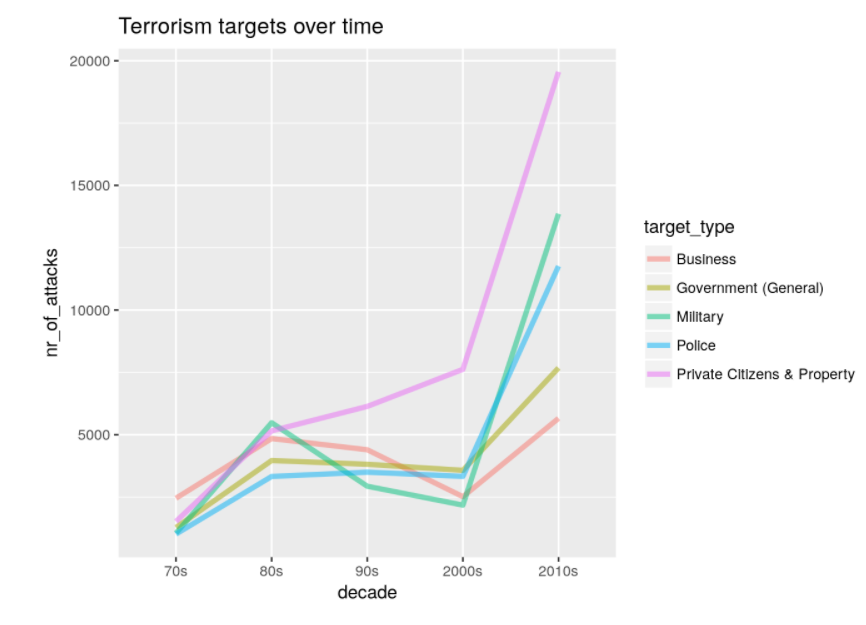


Figure 5- Target Choice

**Case Study 2 – Prediction**

Predicting and visualizing about the responsible group after a terrorism incident can be very useful in order to device a reactive strategy.

In order to perform prediction, I first gathered all the information about terrorist groups like which are the top 10 groups responsible for most casualties. Location where they have attacked and number of people killed and wounded. I created some visualization for that:

* ***Top 10 terrorist groups:*** The spike in Terror Activity has been maintained primarily by 5 Main Groups:

1. Taliban
2. Shining Path (SL)
3. Islamic State of Iraq and the Levant (ISIL)
4. Farabundo Marti National Liberation Front (FMLN)
5. Al-Shabab

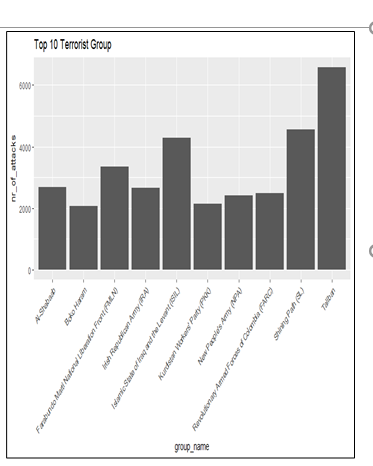
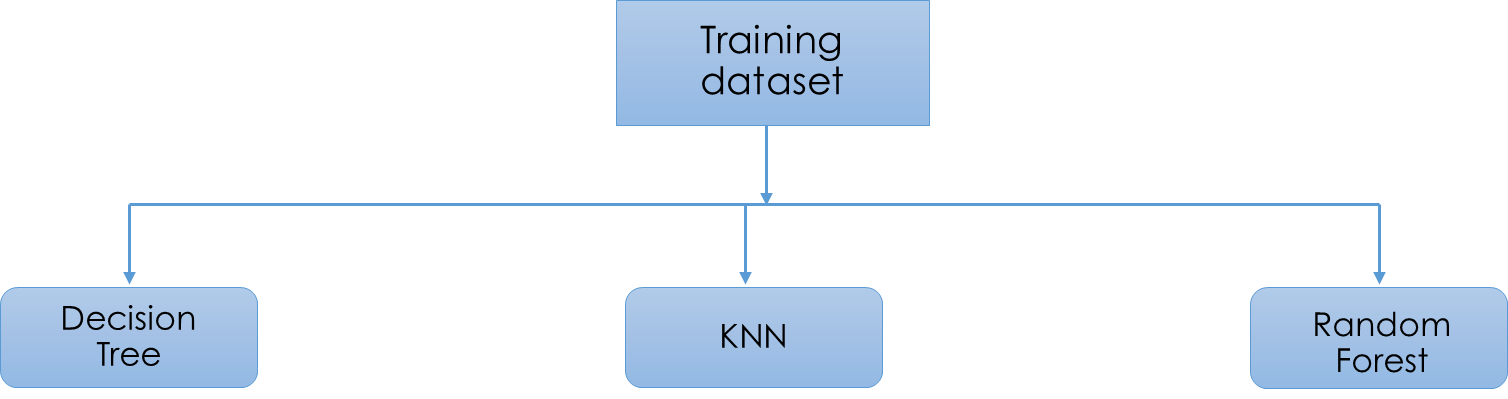


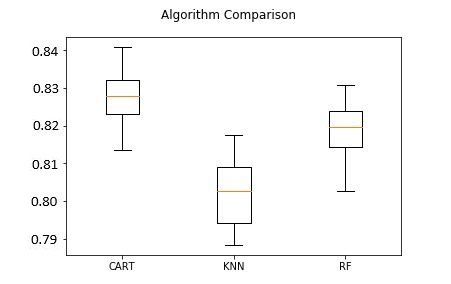
Figure 6- Top Terrorist Groups

**Predictive Model:** For the model following steps were followed:

1. *Data Partitioning*: First of all, data was partitioned into training data to train the model and test data to evaluate the model
   * Training Dataset
   * Testing Dataset
2. *Data Training*: I trained the dataset using decision tree, KNN and random forest model.



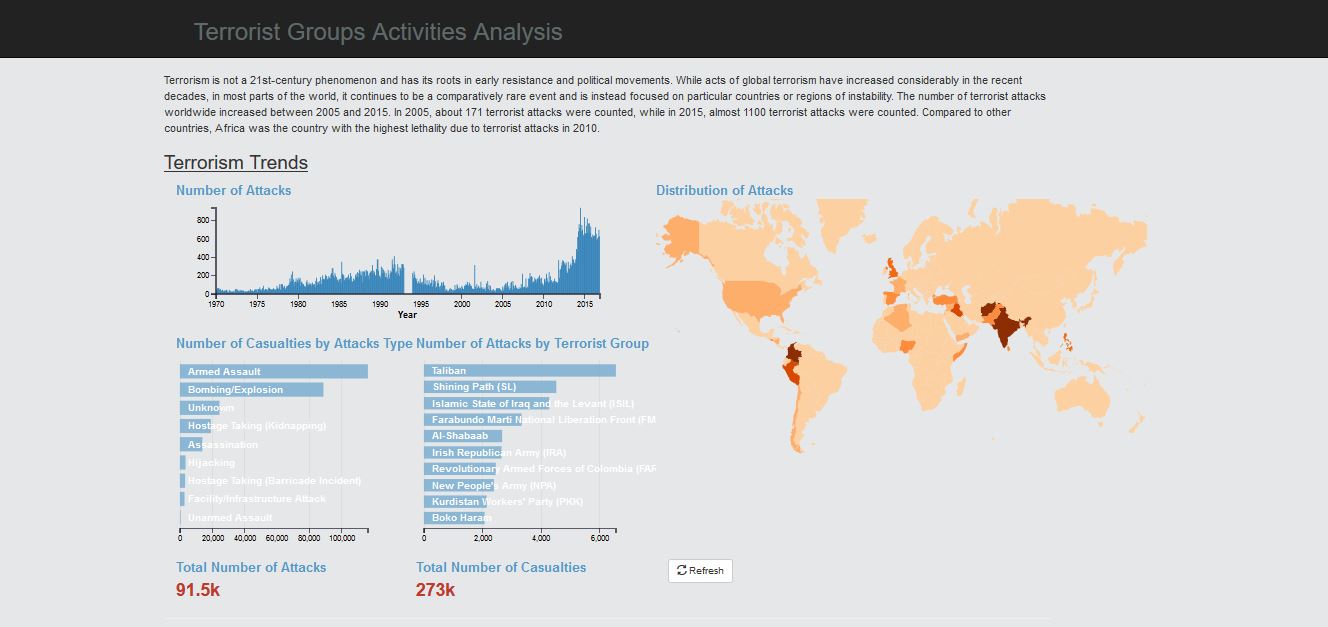
1. *Algorithm Comparison:* Compared all three models. Apparently, Decision tree algorithm has highest accuracy of 83% among all three algorithms.

**

**Final Visualization Snapshots**

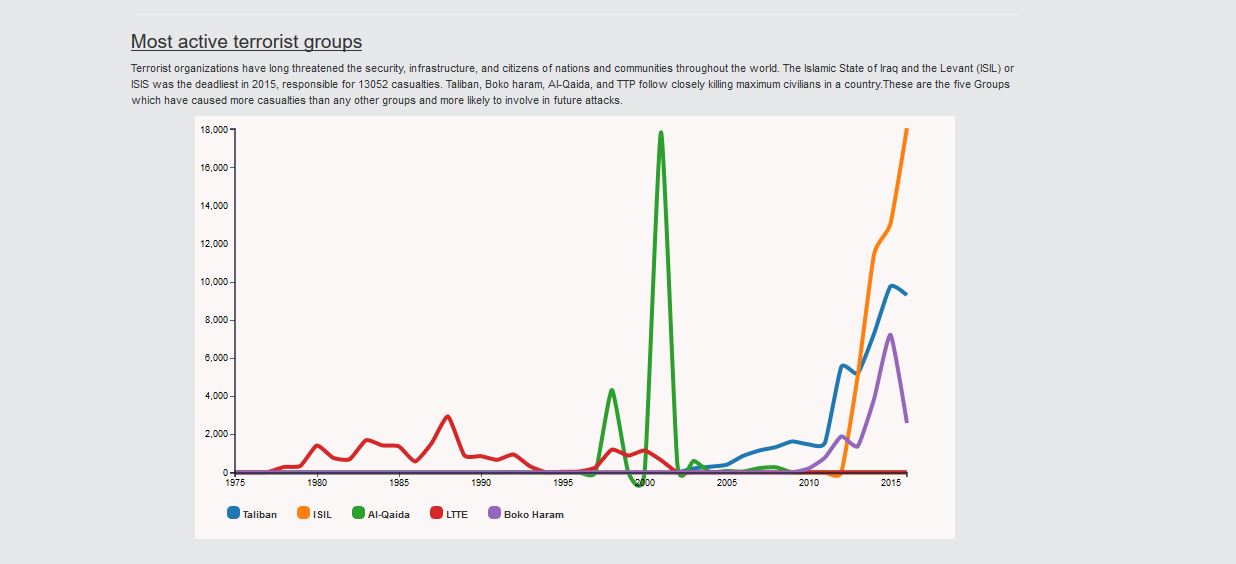
* My project follows Story telling approach which would have broader perspective about how terrorism and terrorist groups have evolved.
* Overview of the terrorism trends over the years followed by terrorist groups activities visualization which gives summary about horror of terrorist groups over the past decade and lastly predicting group behind attacks.

1. Terrorism trends and casualties:
   * Interactive Map
   * Line chart to show number of attacks per year
   * Bar plot to show number of casualties by attack type and another to show number of attacks by terrorist groups

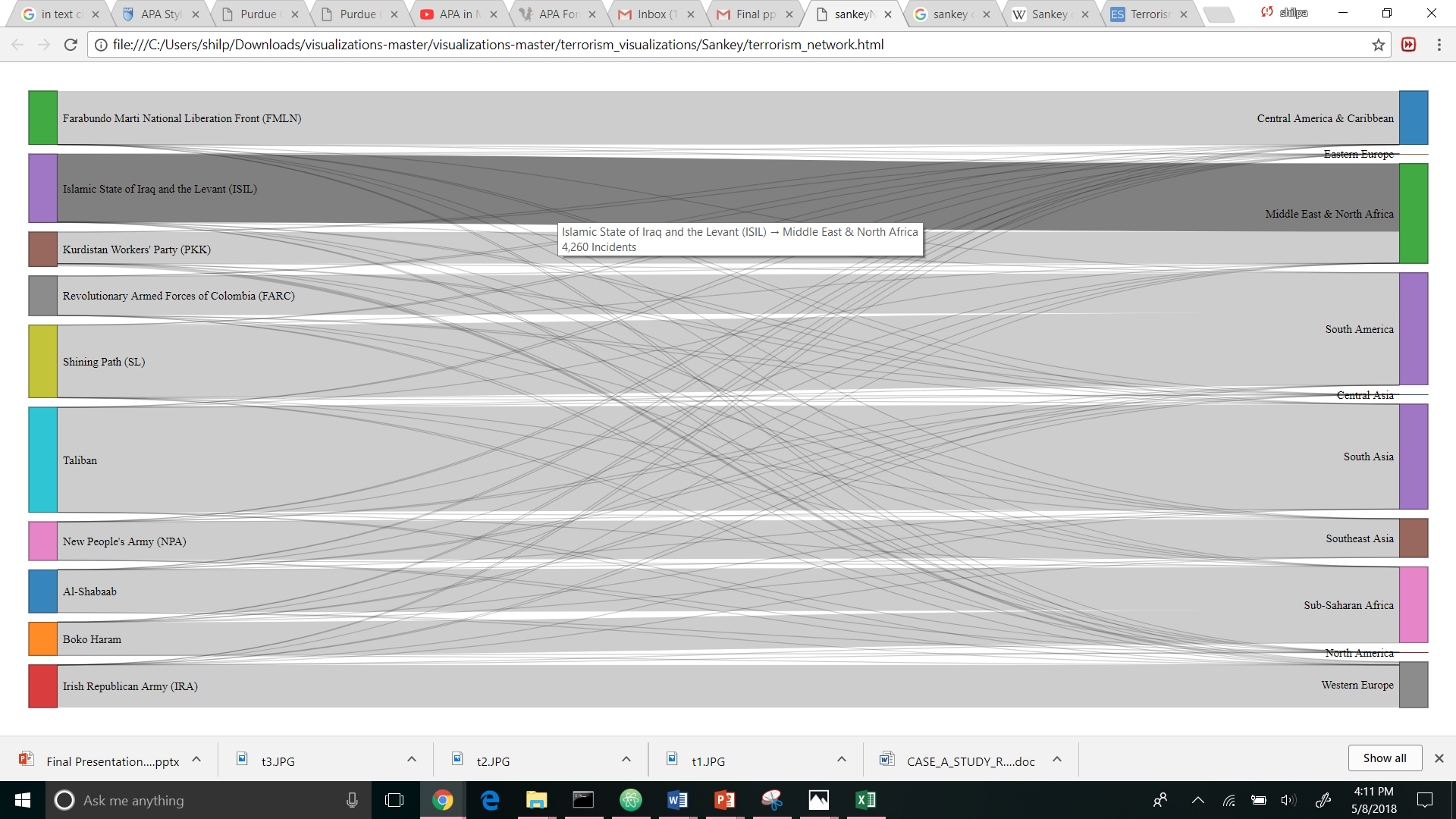


1. Terrorist group activities

Line chart showing top 5 most dangerous group attack over past five years

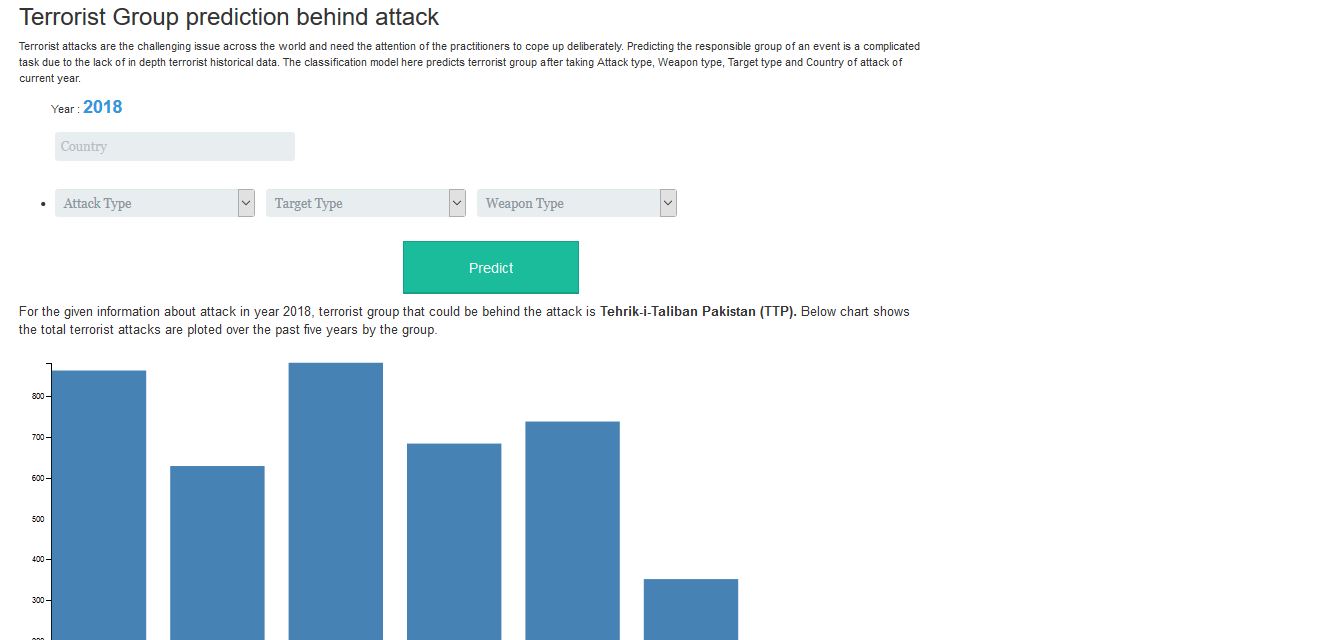


1. Sankey graph showing regions impacted by the terrorist groups.



1. Predicting terrorist group

Interactive bar chart



**References**

1. D. Talreja, J. Nagaraj, N. J. Varsha and K. Mahesh, "Terrorism analytics: Learning to predict the perpetrator," *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, 2017, pp. 1723-1726.
2. J. Górecki, K. Slaninová and V. Snášel, "Visual investigation of similarities in Global Terrorism Database by means of synthetic social networks," 2011 International Conference on Computational Aspects of Social Networks (CASoN), Salamanca, 2011, pp. 255-260.
3. <http://square.github.io/crossfilter/>
4. <https://www.start.umd.edu/gtd/>