Global Terrorism Exploration

Since the Sept 11 attack, the terrorism has been intensified worldwide, and there have been more and more diversified trends such as the targets of attacks, modus operandi and arms tactics, and hot topics are emerging in an endless stream. Terrorist attacks have become important issues that all of us need to pay attention on.

1. Data Collecting & Preprocessing

Our data sources include Global Terrorism Database[1], Wikipedia[2], New York Times[3] and World Bank[4] .

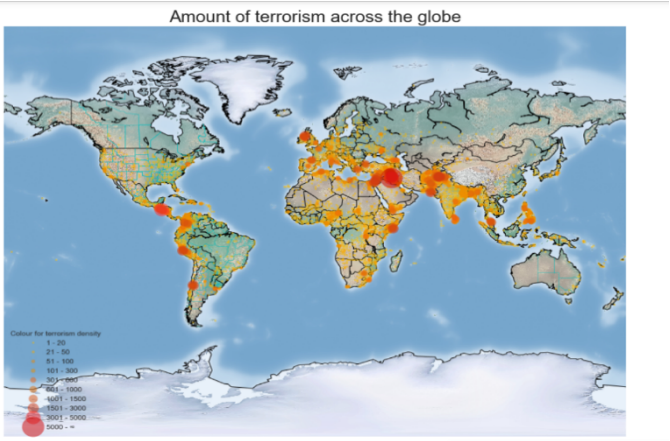
After getting total five data files, we do data cleaning to eliminate useless information and add new columns for future analysis. Next we use SQL to create a database called Terrorism, including five data tables (Events, Gun, Religion, Education & GDP).

2. Dataset Analysis

2.1 Geographic Feature

The geographical plot reveals some typical characteristics of terrorism from 1970 to 2016. First, terrorism attack mainly happened in middle and southeast Asia, Europe, middle and south Africa, northwest of America and also United States. Iraq, Pakistan and Afghanistan are the top 3 countries that suffered the most number of terrorism attack. The lregional ine chart shows that since 2004, Middle East and North Africa as well as Asia have a significant rising in the regional terrorism events. Since 2004, terrorism happened in the world each year showed an increasing trend. Relevant researches show that the number of attacks by left-wing extremists increased 80 percent in the 2000s. These attacks were nearly all motivated by environmentalism and carried out by perpetrators affiliated with ALF and ELF. All of these attacks were intended to cause property damage and intimidation; none were lethal. In comparison to the 2000s, there was a sharp decline in the proportion of terrorist attacks carried out by left-wing, environmentalist extremists during the first seven years of the 2010s, which should be paid more attention to by the government. The top 3 organizations are Taliban, Shining Path and ISIL. All of these organizations organized over 3500 of attacks. These organizations are mainly conflict with the government in religion or political ideology, which are the main conflict reasons.

To analyze the relationship between GDP and terrorism,



we chose the time period 2000-2005 since the world-shocked 911 event happened in this period and since then, the public began to pay attention to terrorism. The graph shows that countries have frequent terrorism are mainly developed countries or less developed countries.

To analyze the relationship between nationality of terrorists and their targets, we choose the most severe 1000 events and set up the geographical network. The lines connect the capital of the country that terrorist came from as "origin" and the place where event happened as "target". The network shows that the terrorist mainly came from Europe, Middle West of Asia and Northwest of South America. The United States is also an outstanding origin of many terrorists. In addition, terrorists from the US always make US and Europe as their target. What's more attracting is that West Africa, North Africa and Europe form a triangle, and the first two are the origins. So these countries should enhance their cooperation in security.

2.2 Multi-dimensional Analysis

We have also explored deeper into our dataset for more specific analysis about the terrorism. Firstly, we try to find out whether the amount of terrorism has increased over the years. It is observed that the amount has been notably increasing from 1970 to 2016, especially after 2004. There may be multiple reasons for this happening and the result should be treated with care as the data collection becomes more and more effective. Secondly, we want to investigate about the attack type, target type and weapon type involved in each terrorist attack. After plotting the amount shares of each type, we conclude that Bombing/Explosion, Private Citizens & Properties and Explosives are the most frequent attack, target and weapon type, respectively. We also get to know the success rate and average kills for each type. It turns out that success rate maintains a similar level among attack

[1] http://www.start.umd.edu
[2] https://www.wikipedia.org
[3] http://www.nytimes.com
[4] http://www.worldbank.org

types while hijacking possesses the highest average kills. For weapon types, it is surprising that vehicles have the highest success rate and average kills, and the values for radiological attacks are 0. The property damages caused by each attack type, target type and weapon type has also been approached. These outcomes can be reference for the government to interfere and prevent the specific kind of terrorism attack.

It is also important to know the motives of terrorists and how these motives have changed across the years. After conducting a simple sentiment analysis on the words in terrorists' motives, we find that negative words count far more than positive ones, which is also expected. It is also found that both positive words and negative words become more and more in recent years, which may be also ascribed to the effective collection of data recently.

2.3 News Report Analysis

From 1998 to 2017 (total 20 years), we get 50 articles from New York Times each year querying "Terrorist Attack" and combine them into a single doc on a yearly basis.

Firstly, we do the sentiment analysis using NRC data based on the whole news reports. In the 20 years period, the emotions (fear, trust, negative, positive, joy, disgust, anticipation, sadness, surprise) conveyed by news reports are relatively stable.

Secondly, we research on the complexity of the news articles, trying to find if their writing styles change from 1998 to 2017. We find that, besides 2006, the variety of words using is stable and continuous, range from 0.118 to 0.149. What's more, the average length of a single sentence has an increasing trend in the 20 years period. From 1998 to 2006, the average sentences length is around 24. From 2007 to 2017, this number comes to 28.

Thirdly, we conduct the topic analysis toward a single doc containing all the news from 1998 to 2017 to identify the major concepts underlying it. Topics that have highest probability 45.45% is ('0.018*"program"+ 0.010*"surveillance"+0.010*"million"+0.010*"Taliban" + 0.008*"pentagon"').

Next, we set each single year's doc as new text and see which topics it best matches. We can see from the results that each year's topic with highest probability keep change. For example, from Russia to Palestinian.

Finally, we draw the word cloud to illustrate the changes of topics in each year in a straighter forward way.

2.4 Past & Future

We make use of machine learning to get a summary on the past and make our own forecasts on future terrorism events. First, we grab data from joined tables using SQL query. Second, use Futures Importance Figure to analyze 8 factors of terrorism, we conclude that the most important factor underlying terrorism is GDP, followed by religion and gun, and education is the least important factor among four.

Because the place with higher GDP growth will have larger population density, terrorists would like to launch an attack in such a place to get much damage and more kills. What's more, terrorism is always related to religion, so a place which has more religious people will be more likely to be attacked. Based on this, we deeply analyze four indexes to draw scatter for these four factors. We find that countries with moderate GDP growth have more attacks, that is because most developed countries like USA, England which are targets of terrorists may no longer have fast economic development. And the most interesting thing is that the more religious people a country has, the more it will be attacked. Besides, the education index has no influence on terrorism actually.

For four types, the attack type is the most important, followed by weapon type. Because terrorists always like to choose a wider range of attack to have more severe influence.

The conclusion us that government and public should take more care of developed areas and pay more attention to assembly occupancies which may be chosen by terrorists.

Third, in order to help government and organizations to better prevent terrorism, we build decision tree to predict the future terrorism. We divide the damage types into two kinds, major damage and minor damage. Different indexes will lead to different damage kind. Hence, we could predict future influence of terrorism in each country. This figure could offer advice for government management on terrorism.

3. Conclusion

Based on multi-sourced data support and our further analysis, this project explores several dimensions of terrorist attacks, including geographical distribution, types of attacks, analysis of relevant news reports, summary and forecast for the future. The data tells more of the hidden truth and we hope this study will shed more light on the details of terrorist attacks hiding under the iceberg so that people will have a fuller understanding of terrorism.