**Project Proposal**

Ho Huang, Yuhan Cui, Hsin-Jung Yu

**Introduction:**

Terror attacks recently have been a serious issue. Each country is dedicated to preventing terrorism and also aware of any possible attacks. With the development of scientific and technological progress, there are a variety of ways of terror attacks occurring. In response to this situation, the government from all over the world are urging to use new intelligence efforts to protect the nation’s borders, defend against threats, minimize infrastructure vulnerabilities, and improve emergency responses. In order to understand how active and the pattern of the terror attacks, we want to analyze the data from the Global Terrorism Database(GTD), which includes information on terrorist attacks around the world from 1970 through 2017.

**The detailed description of the datasets:**

The GTD is maintained by the National Consortium for the Study of Terrorism (START) at the University of Maryland, College Park. The data set records terrorisms from 1970 to 2017, including systematic data on domestic as well as transnational and international terrorist incidents that have occurred during this time period. There are 182,000 rows and 135 columns in the data set, where each row indicates an incident of terrorism.

The dataset contains a wide range of entries(columns) related to the terrorist attack, including the GTD ID and Date, Incident Information, Location(Region, Country), Attack Type, Weapon Type, Victim Information, Perpetrator Information, Casualties, and Consequences, etc.

Addition dataset: World GDP from 1960 to 2017, by current US$. Data source: The World Bank Group

**Describe how you plan to acquire the data:**

* Download from Kaggle (<https://www.kaggle.com/START-UMD/gtd>)
* Download from The World Bank Group (<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?page=>)
* Download from the United Nations Office on Drugs and Crime

(<https://dataunodc.un.org/drugs/prevalence_time_series>)

**Processing tasks you will need to accomplish to prepare the data for analysis:**

* Drop columns which have a large number of nulls (number of nulls exceeds 80% of data).

For example: columns like attacktype2\_txt/attacktype3/ransomnote etc.

* Text processing several columns which include summary text.

For example: capture the feature words from the Additional Notes(addnotes).

* Merge multiple datasets to do further analysis.

For example: merge the two datasets mentioned above to find the relationship between GDP and terrorism.

* Generate new columns that will be used in further analysis.

For example: create a new column ‘MultipleWeapon’, which is a binary variable. If weapsubtype2~4 contains values, MultipleWeapon is 1, else 0.

**Description of what you hope to find in the data:**

* What is the frequency of attacks happened in each region and countries?
* What is the killing/wounded rate in each attacked country and year?
* What is the causation of terrorism? (Using drugs to earn money or GDP?)
* How many American citizens were killed or wounded in each year/states?
* Which type of target is the most frequently under attack by terrorists?
* What are the most active terrorist group?
* What is the most frequently used weapon used by terrorists?
* What proportion of terrorism is the suicide attack? And what is the difference between suicide and a normal terrorist attack?
* How many kidnapping occurs among terrorism? What are the descriptive statistics of kidnapping incidents?
* What is the relationship between GDP and terrorism?
* How to predict a terrorist will be a success or not in order to prevent the attack?
* What are the factors that can lead to a successful terror attack?

The report should be developed as a Jupyter Notebook (and submitted as an .html file), in order to demonstrate your progress to date. The progress report should begin with a restatement of the overall project objectives, a description of the data, and your proposed (or revised) direction of analysis. In addition, you should summarize your progress to date, and demonstrate your progress in code. Finally, you should conclude your progress report with a description of the pending tasks that remain before you will complete your projects. Your grade will not only be based on the quality of your progress report (5 points), but also an assessment of your progress itself (5 points).