*Team participants*

**Suer Hu** (sh3589@columbia.edu)

**Tiantian Yin** (ty2349@columbia.edu)

**Donggwan Kim** (dk2930@colubia.edu)

**Pedro Perez** (pjp2129@columbia.edu)

*Purpose of the Project*

The title of our thesis is ***Patterns in Global Terrorism***. In this study, we seek to answer these general questions:

1. Where is terrorism most prevalent?
2. How have global terrorism patterns evolved over time (types and location of attacks)?
3. What are the primary motivations (including victims) that fuel terrorism?

*Abtract*

Terrorism is clearly a very hot topic, as evidenced by the elevated media coverage of terrorist attacks. The focus on terrorism by both parties during the most recent U.S. presidential debate shows that it is a growing policy concern around the world, with the likes of ISIS and Boko Haram making headlines on a daily basis.

Our goal is to take a broad look at terrorism around the globe and determine for ourselves if the soundbites that are echoed by different media outlets are backed by the data in the GTD dataset. Our analysis will contain three major pieces:

* Static plots of numerical and categorical variables (e.g. types of weapons)
* Geospatial data maps (e.g. global terrorism hotspots/density)
* Incident summary text analysis (e.g. evolutions in primary motivations)

*Overview of the data*

For this project, we will be using the Global Terrorism Database (GTD), which is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START). The dataset provides multi-dimensional data on global terrorism attacks from 1970 to 2015. Some variables of interest include the specific location of an attack (country, city, long/lat), attack type (e.g. assassination, bombing, hijacking, etc.), weapon information, target/victim information, and perpetrator information.

The GTD dataset also includes text summaries of each incident, including the when, where, who, what, how, and why of each incident. We plan to use this data to study common trends in terrorist attacks and how these evolve over time.

The biggest limitation we see with this data is that some of the filtering rules change over time since the GTD started. For example, the categorization of weapons is slightly different from 1970 to 2015, which could lead to irregular results when we compare across time. In order to solve for this, we will focus any temporal analysis on specific time periods during which the filtering rules and definitions of the variables are constant (e.g. 2012 – 2015).

The data is readily available in *.csv* format, so most of the analysis can easily be done in R.