The past few decades have been painted with growing geo-political instability across the world and terrorism has been one of the main ways through which this global decay manifested itself.

In recent years, increased access to technology has allowed the average individual to get a deeper insight in what terror acts occurring around the world. As such, terrorism attacks have been increasingly mediatized. I thought it would be interesting to explore what is all over the news from another approach, a more methodic one so to say.

The Global Terrorism Database (GTD) records terror attacks around the world from 1970 through 2015, it includes systematic data on domestic and international terrorist incidents durint this time period.

* Contains information on over 150,000 terrorist attacks
* Currently the most comprehensive unclassified data base on terrorist events in the world
* Includes information on more than 75,000 bombings, 17,000 assassinations, and 9,000 kidnappings since 1970
* Includes information on at least 45 variables for each case, with more recent incidents including information on more than 120 variables
* Over 4,000,000 news articles and 25,000 news sources were reviewed to collect incident data from 1998 to 2015 alone

For the purpose of this study, the data analyzed will be strictly restricted to attacks that satisfy the three criteria per the Codebook guidelines.

**CRITERIA 1 : POLITICAL, ECONOMIC, RELIGIOUS, OR SOCIAL GOAL**

The violent act must be aimed at attaining a political, economic, religious, or social goal. This criterion is not satisfied in those cases where the perpetrator(s) acted out of a pure profit motive or from an idiosyncratic personal motive unconnected with broader societal change

**CRITERIA 2 : INTENTION TO COERCE, INTIMIDATE OR PUBLICIZE TO LARGER AUDIENCE(S)**

To satisfy this criterion there must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims. Such evidence can include (but is not limited to) the following: pre‐ or post‐attack statements by the perpetrator(s), past behavior by the perpetrators, or the particular nature of the target/victim, weapon, or attack type.

**CRITERIA 3 : OUTSIDE INTERNATIONAL HUMANITARIAN LAW**

The action is outside the context of legitimate warfare activities, insofar as it targets non‐combatants (i.e. the act must be outside the parameters permitted by international humanitarian law as reflected in the Additional Protocol to the Geneva Conventions of 12 August 1949 and elsewhere).

**Aims and Objectives**

[Table of Contents](http://localhost:8889/notebooks/EDA_PredictingCasualties.ipynb#Table-of-Contents:)

This notebook aims to explore the global terrorism database in two ways:

* Visually exploring the extremely rich data on the 150,000 terrorist attacks, trying to answer questions such as:
  + Which countries/region are the most targeted?
  + Where are there the most casualties?
  + How have casualties evolved throughout the years?
  + What are the casualties by weapon type?
  + Are certain nationalities more targeted?
  + Are some countries better at defending themselves against terrorist attacks?
* Analyze the data using scikit learn's classification library:
  + Can we predict if a terrorist attacks will result in casualties or yield no death and wounded.
  + If so, which model will be the most appropriate, should we use linear models, ensemble, support vector machines or neural networks?
  + How accurate will they be? How do we determine our scoring metrics?
  + Can we delve deeper by predicting thresholds of casualties? Will an attack be devastating or benign?
  + What about predicting the future? Can we get descent results running an ARIMA model?

Now let's look at this database!

**Data Dictionary**

[Table of Contents](http://localhost:8889/notebooks/EDA_PredictingCasualties.ipynb#Table-of-Contents:)

Here is a quick look at the data that will be used in this notebook, the complete descriptions can be found in the [codebook.](https://www.start.umd.edu/gtd/downloads/Codebook.pdf)

## Spatio-Temporal Variables:

'iyear' : year of the incident

'imonth' : month of the incident

'iday' : day of the incident

'latitude' : latitude of the incident

'longitude' : longitude of the incident

## Continous variables:

'nkill' : number of dead

'nwound' : number of wounded

## Binary Variables:

'crit1' : was the attack aimed at attaining a political, economic, religious, or social goal?

'crit2' : was there intent to coerce or intimidate a larger audience than the victims?

'crit3' : was the incident outside legitimate warfare activities (i.e. target non-combattants)?

'doubtter' : aws there doubt as to whether or not the incident is a terrorist attack

'extended' : has the incident lasted for more than 24 hours?

'multiple' : is the incident connected to other attacks?

'success' : did the terrorist attack achieve its goal (i.e. assassination, etc.)?

'suicide' : did the incident involve a suicide attack?

'guncertain1' : was the terrorist group confirmed?

'claimed' : was the incident claimed by a particular group?

'property' : was property damaged during the attack?

'ishostkid' : were victims taken hostages or kidnapped?

## Categorical Variables:

'country\_txt' : country in which the incident occured

'region\_txt' : region in which the incident occured

'alternative\_txt' : type of attack if it was not terrorist for certain

'attacktype1\_txt' : general method of attack used (i.e. assassination, hijacking, bombing/explosion, etc.)

'targtype1\_txt' : general type of target/victim (i.e. business, government, police, military, etc.)

'natlty1\_txt' : nationality of the target/victim

'weaptype1\_txt' : general type of weapon used in the incident (i.e. biological, chemical, firearms, etc.)

## Descriptive Variables:

'target1' : specific person, building, installation, etc. that was targeted

'gname' : terrorist group responsible for the attack

'summary' : summary of the incident, when avaialble