**PROBLEM**

**Unavailability of Data**
Due to political bias, a lot of attacks go undocumented and unreported. Even when the data is available, it is not synthesized.

**Fake and Unreliable News**
There’s no proper mechanism to tackle the problem of fake news propagating disinformation.

**Lack of an Analytical Model**
Data comes in many forms and is unformatted. It usually doesn’t naturally answer a specific question about a particular event.
WHAT DOES A SOLUTION LOOK LIKE?

Modular
Has a model that accepts multiple forms of data.

Multiple Sources of Data with Confidence Levels
Data consists of news articles, live news feeds, google searches, tweets, and direct reports from individuals.

Each unique source of data answers 4 Questions
WE BUILT IT

Databases
Historical data was collected via two different sources.

Confidence
Each data source is given a weight. For each attack/event, a confidence level is calculated.

Current News
A web-crawler, google search, Twitter Streaming API and a witness program were used to fetch current news.

Location-Based
Our dashboard is location-based and gives in-depth analysis of events in different regions.

Triangulation
We created a predictive model defining the red-zones around a specific geographical region.

Adaptable
In future, new data sources could be integrated into our model.
LET'S TAKE A LOOK.
Our simple yet effective interactive map displays historic events across different time periods and regions.
EXAMPLE OF A CURRENT EVENT

Clicking a specific geographic region opens up a dialogue box displaying information about an attack occurring in the region.

Clicking on the link takes you to another page with more information.
CATEGORIZATION OF RESULTS

Each event/attack is categorized based on the types of common attacks on education.
Geographical Profile: Syria in 2016

Region Attacks

- Aleppo
  - August 2016
- Al-Raqqa
  - August 2016
- Idlib
  - September 2016
- Al-Bab
  - October 2016

When did people start searching for each attack in each city?

- Direct attacks on Education
- Indirect attacks on Education
- Educational Personnel Killed/Injured
- Military Use of Schools
- Schools Damaged or Closed
- Gender Based Violence
- Teachers Under Attack
- Use of Explosive Weapons

Methodology: First minute searched is defined as the first minute where a related search showed up.
METRICS PER CATEGORY OF EVENT

Each category can further be broken down in terms of statistics such as but not limited to the number of casualties in an attack, stakeholders...etc.

Clicking on a region takes you to another page with specific Tweets, Google Trends, Live News, and Testimonials relative to the event.
Each event has a confidence level describing the possibility of the occurrence of an event. The confidence level is calculated by the weighted sum of the accuracy of the different data sources used to describe the event.
For each event, our dashboard displays the key search terms, relevant tweets and news articles along with the testimonials from witnesses that submit personal accounts of the event.
We then run a classifier which predicts if a certain region is a ‘red-zone’ i.e. has a high probability of a future attack. We then use triangulation to display those regions. We hypothesize that a region is possibly under attack if it is surrounded by attacks.
Each event is then linked to a closest organization one could contact to get details about the event.
TECHNOLOGIES USED

d3.js
Google Trends
django
CARTO
The GDELT Project
BeautifulSoup
JavaScript
Natural Language Analyses with NLTK
THANKS!

For further information, you can contact us at:

» isoltan@cmu.edu
» shahan@cmu.edu
» smkhan@cmu.edu
» smehdi@andrew.cmu.edu