Overview and Motivation

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)

Questions

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)? What are the primary motivations (including victims) that fuel terrorism?

Data

For this project, we used 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.

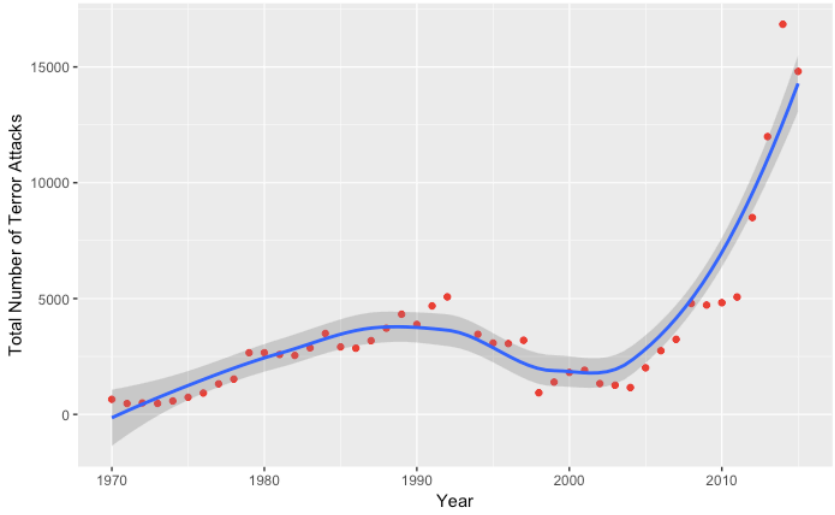
The GTD dataset also includes text summaries of each incident, including the when, where, who, what, how, and why of each incident. We plan to use this data to study common trends in terrorist attacks and how these evolve over time.

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).

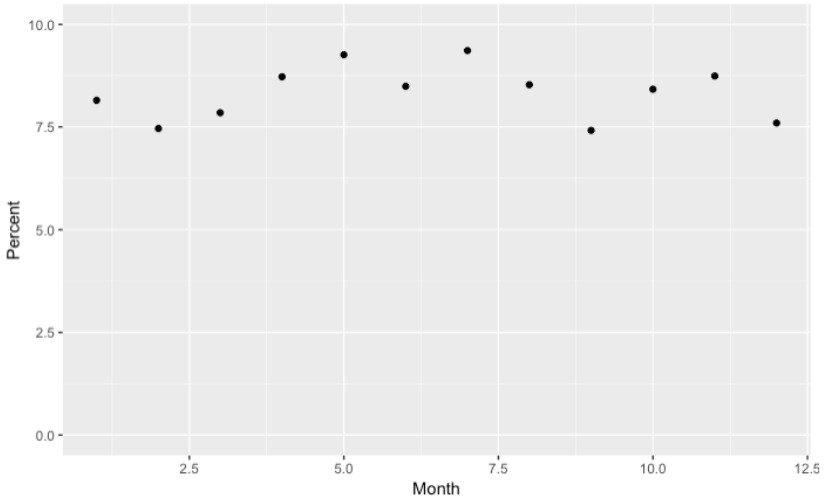
Exploratory Data Analysis

The first step in exploring our data was to import the data file into R, which was actually not a trivial task. Given the size of the GTD, we had to first trim out the data we did not need directly in Excel in order to avoid using up our local memory unnecessarily in R. Once this was done, we first started to explore the data over time, which was the most logical step given the format of the original Global Terrorism Database. We quickly found that there were significant changes in the data over time, so we decided to make a full section of our website related to this.

To begin with, we first looked at how many terrorist attacks happened each year. We could find that terrorism has been on the rise since 1970, wish some drops in mid 1990's and early 2000's. However, terrorism is increasing at a faster past since the 2005, so we had done some further work to see whether there were some patterns regarding terrorist attacks, such as seasonality, attack types, and target types.



We did not find any association between terrorism and seasons, and I think it is because of the fact that if terrorist attacks were to happen in specific times of each year, the police and military would stay alert which would prevent attacks from happening.

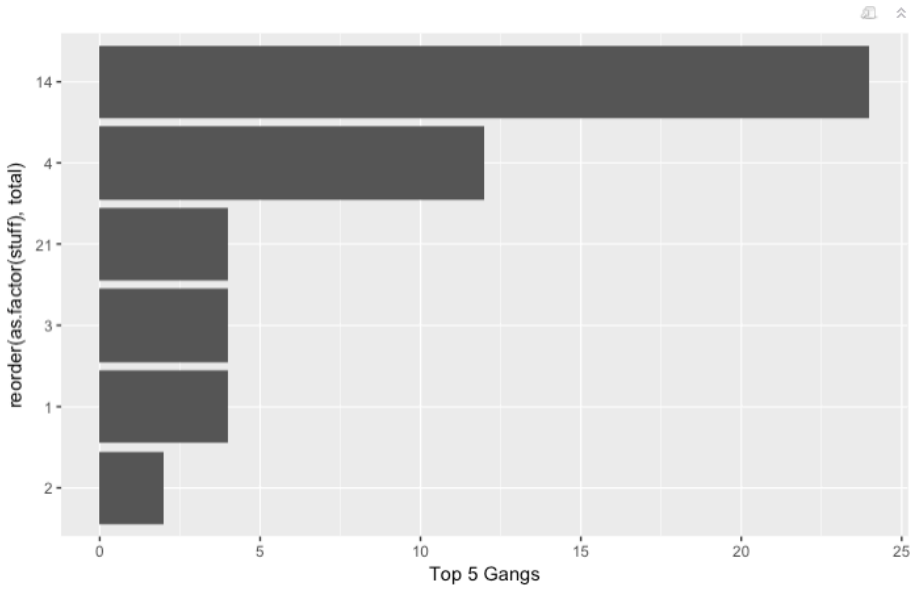


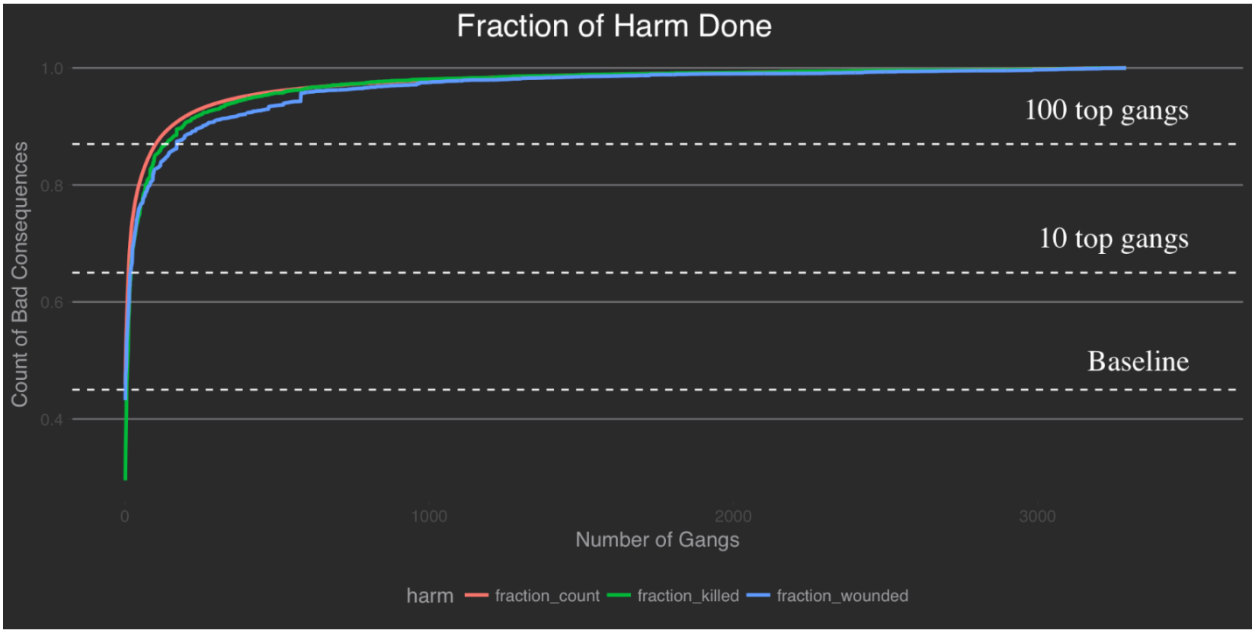
We also found that some big gangs tend to do more terrorist attacks and each attack's consequence was larger than those committed by small gangs. Further work on this is presented on our website.

gname <chr>	stuff <int>
Taliban	5502
Shining Path (SL)	4548
Farabundo Marti National Liberation Front (FMLN)	3351
Islamic State of Iraq and the Levant (ISIL)	2833
Irish Republican Army (IRA)	2670
Revolutionary Armed Forces of Colombia (FARC)	2474
New People's Army (NPA)	2241
Al-Shabaab	2127
Basque Fatherland and Freedom (ETA)	2024
Boko Haram	1839

1–10 of 50 rows

Previous 1 2 3 4 5 Next





When it comes to terrorism activities one always want to know more about the terrorist organizations. We arranged the terrorist organizations according to how many activities they conducted over all the years, and plotted the accumulative percentage as above. The baseline represents the activities that no organization has claimed to be responsible yet. But we find that top 10 gangs by activity frequency has already accounted for 20% of total terrorist attacks, fatality, and injuries. This indicates the skewedness in the performance of the gangs, in other words, we have big terrorist organizations which have done a great deal of harms and we also have many more less organized small gangs.

Show 3 entries Search:

	Winners	Gang_Name	Frequency	Fatality_Rate	Injury_Rate
1	Most frequent activities	Taliban	5502	3.6	4
2	Top fatality rate	Dissident Military Mmbrs of Northern Tribal Group	1	150	0
3	Top injury rate	Aum Shinri Kyo	8	2.6	750

Showing 1 to 3 of 3 entries Previous 1 Next

We are motivated to see which are the biggest offenders in by these statistics. The finding is that each terrorist organizations has quite different style of attacks. Taliban has the most frequent activities, by committing 5502 terror attacks till 2015; Dissident Military Members of Northern Tribal Group has committed only one attack by killed 150 people on the attack; Aum Shinri Kyo has the highest injury rate by hurting approximately 750 persons on every attack, while killing a relatively small number of them, which is less than 3 on average.
With the intension both to cut done the data size and focus more on the organizations which our attention would bring more payoff in understanding and reducing its harms, we made a few assumptions to filter out a few organizations by relative importance. The assumptions we made towards defining an important organization are:

1. It should be active in its terrorist attacks throughout the years, which indicates high activity frequency;

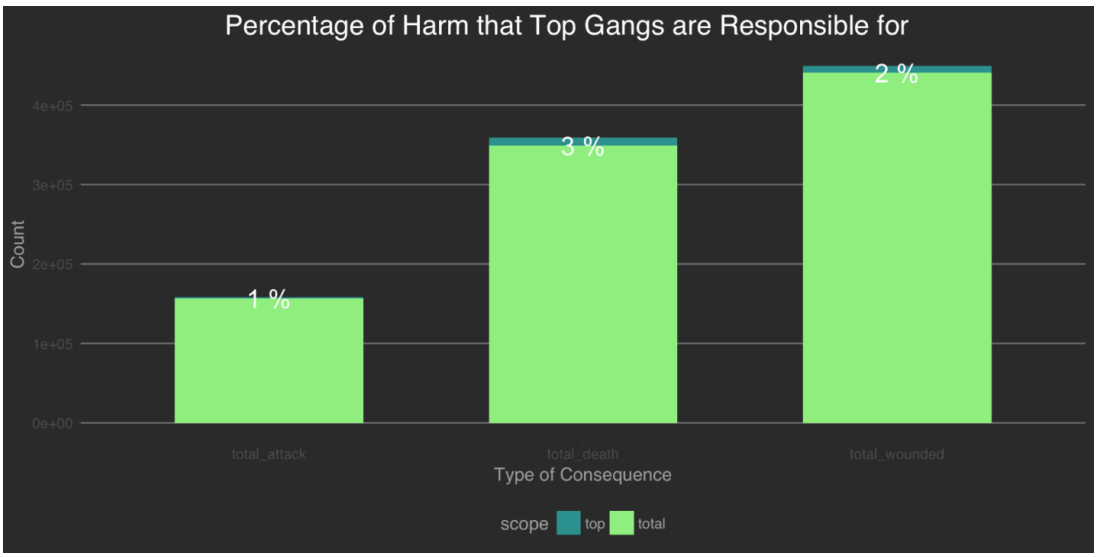
2. With each attack it would bring severe consequences, which indicates high fatality rates and high injury rates;

Show 9 entries Search:

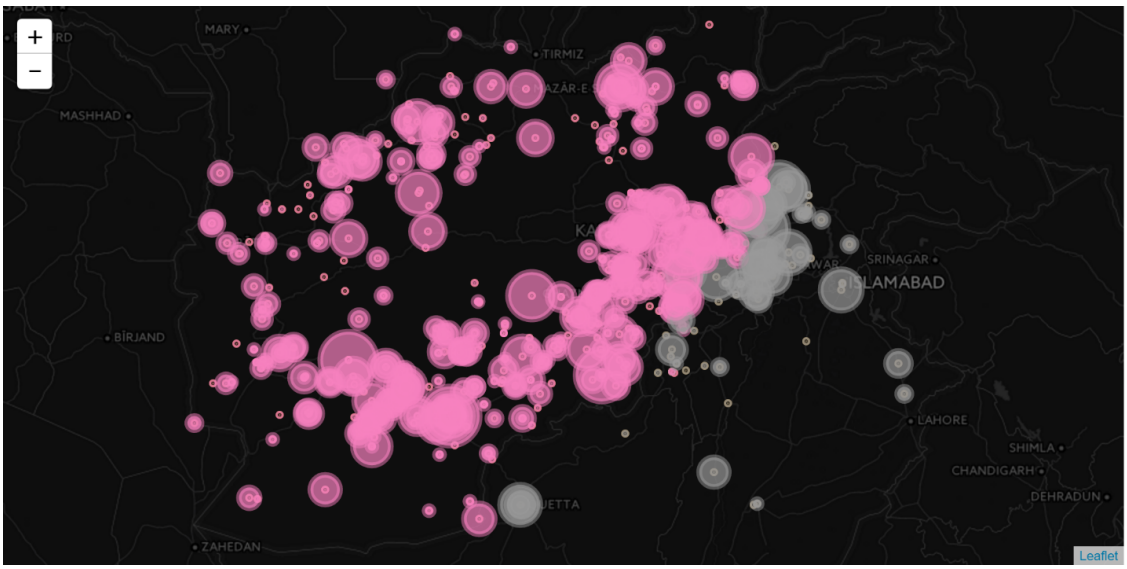
	gname	total_num_attacks	death_ratio	wounded_ratio
1	Taliban	5502	3.6	4
2	Islamic State of Iraq and the Levant (ISIL)	2833	7.5	7
3	Boko Haram	1839	10.1	5
4	Liberation Tigers of Tamil Eelam (LTTE)	1606	6.9	7
5	Tehrik-i-Taliban Pakistan (TTP)	1153	4.7	7
6	Al-Qaida in Iraq	635	6.9	17
7	Huthi Extremists	554	3.8	5
8	National Union for the Total Independence of Angola (UNITA)	431	6.5	5
9	Chechen Rebels	327	5.2	9

Showing 1 to 9 of 9 entries Previous 1 Next

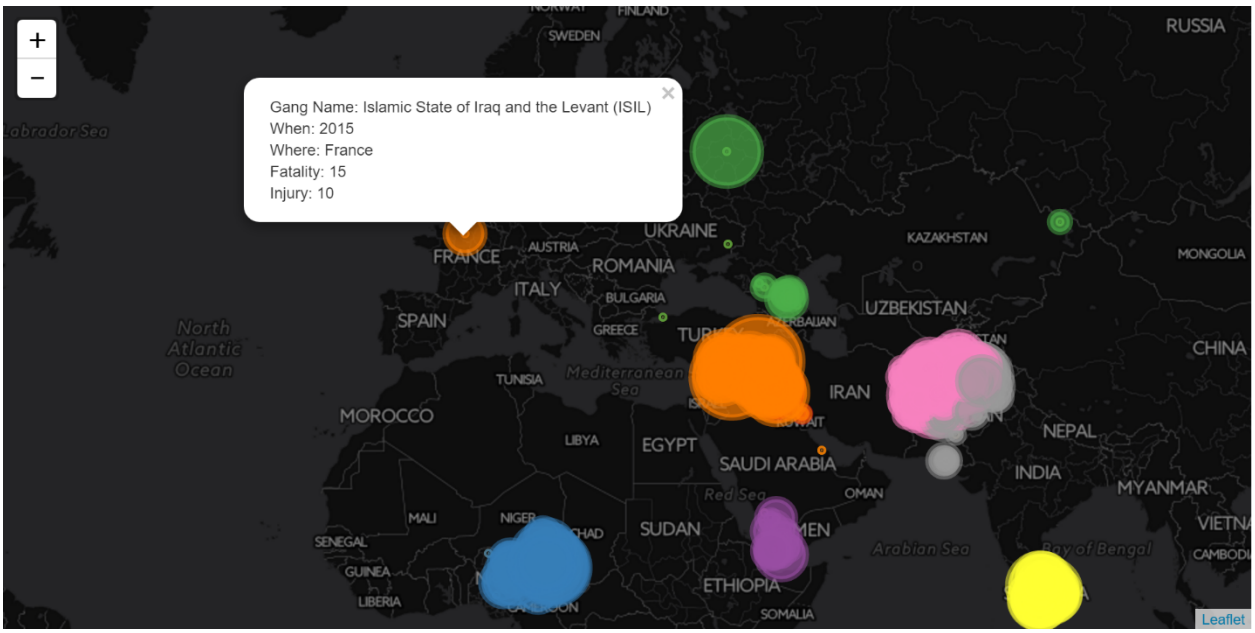
By these assumptions we filtered out 9 gangs. We see familiar names in them. With these 9 gangs in mind, we continue to study their geographical patterns and across time.



We can see that the 9 gangs that we filtered out has already accounted for 1% of total attacks, and 3% of total people killed in the attacks and 2% total injuries.



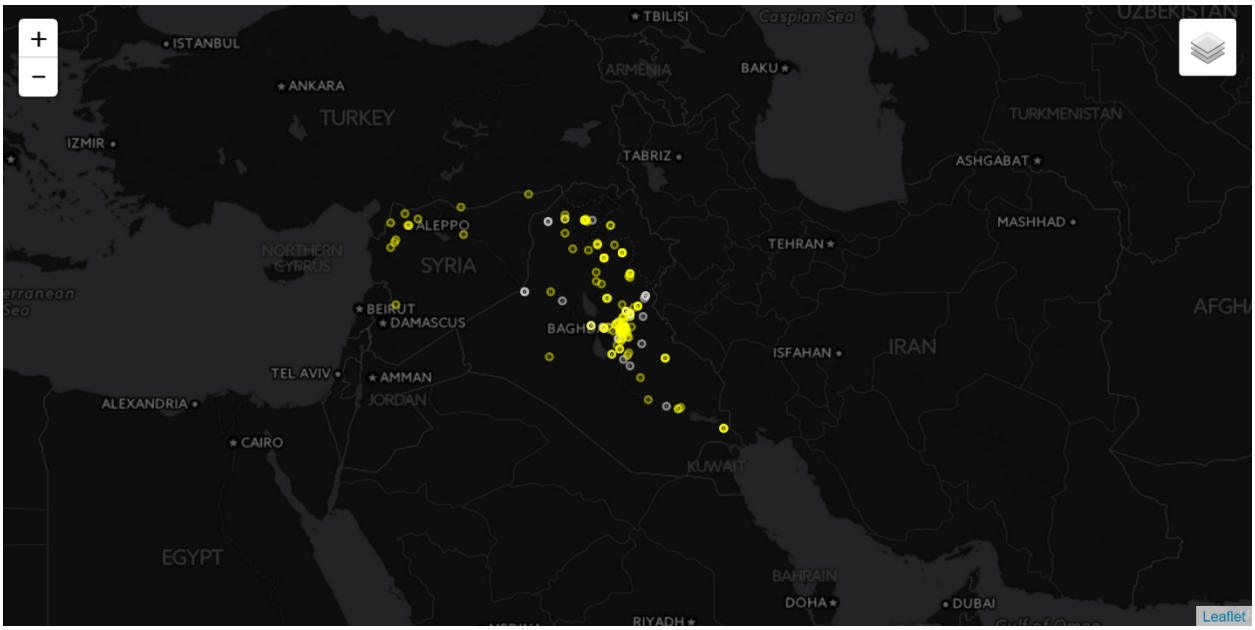
We find out that every one of the nine gangs has an active area to its own, indicated by different colors of the circles. We also find out that the biggest attacks (The severity of an attack is indicated by the size of the circle around the dot.) they committed were always within their own territories, which make sense because it takes high risks to perform attacks outside their territory, with high alerts of police of other countries. The one outlier in this sense is the attack of ISIL, who committed one attack in France in 2015 that killed 15 persons.



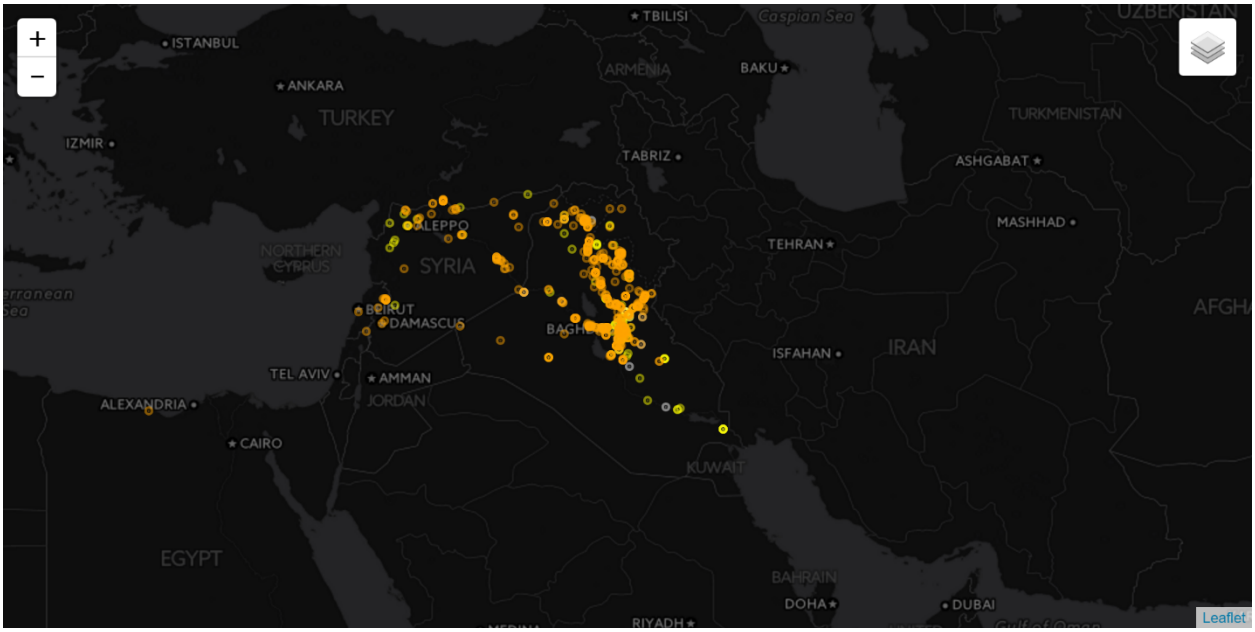
We also looked at the trend across time for a particular gang, ISIL. We can see that it was originally active only in Iraq before 2012 (ISI), but later moved to Syria in 2013 (and took another name of ISIL). We can see a clear expansion towards north Africa and Europe.

Before 2012:

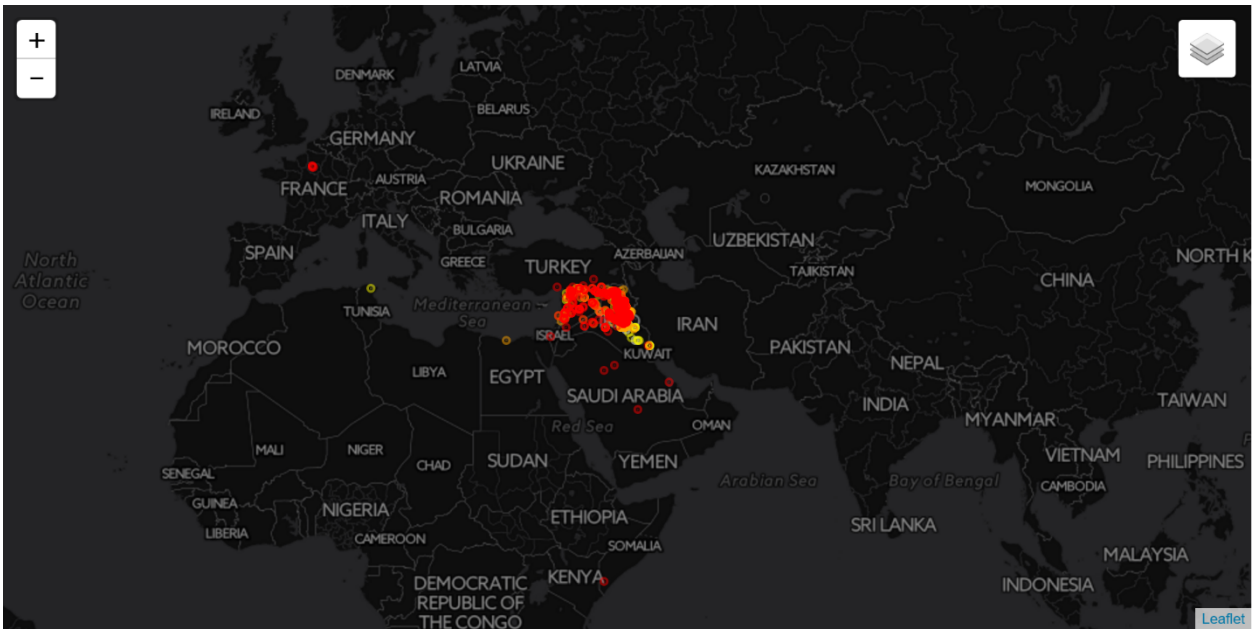
2013:


TBILISI
Caspian Sea
ISTANBUL
ARMENIA
BAKU
ANKARA
TURKEY
TURKMENISTAN
IZMIR
TABRIZ
ASHGABAT
ALEPPO
MASHHAD
NORTHERN CYPRUS
SYRIA
TEHRAN
BEIRUT
DAMASCUS
BAGHDAD
ISFAHAN
IRAN
TEL AVIV
AMMAN
JORDAN
ALEXANDRIA
CAIRO
KUWAIT
EGYPT
BAHRAIN
DOHA
DUBAI
RIYADH
Leaflet

2014:


TBILISI
Caspian Sea
ISTANBUL
ARMENIA
BAKU
ANKARA
TURKEY
TURKMENISTAN
IZMIR
TABRIZ
ASHGABAT
ALEPPO
MASHHAD
NORTHERN CYPRUS
SYRIA
TEHRAN
BEIRUT
DAMASCUS
BAGHDAD
ISFAHAN
IRAN
TEL AVIV
AMMAN
JORDAN
ALEXANDRIA
CAIRO
KUWAIT
EGYPT
BAHRAIN
DOHA
DUBAI
RIYADH
Leaflet

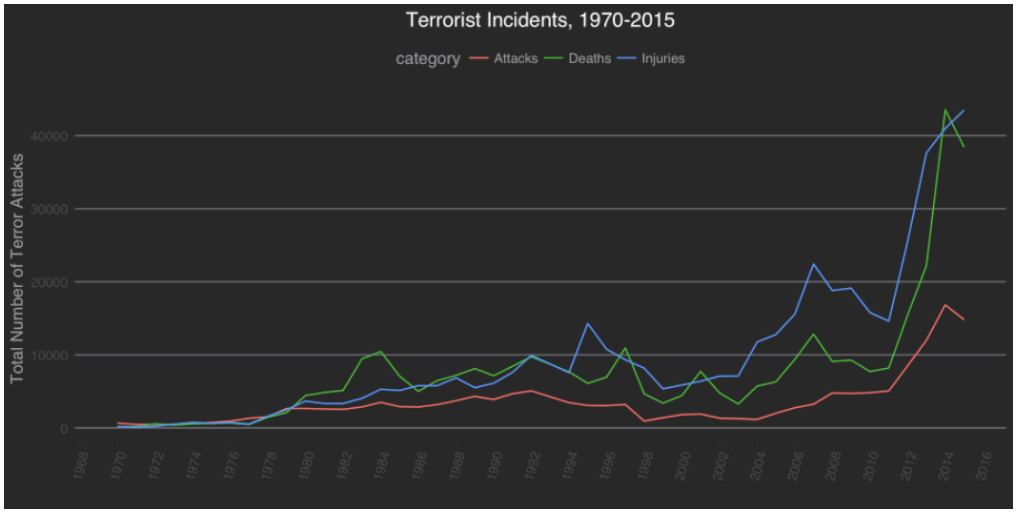
2015:

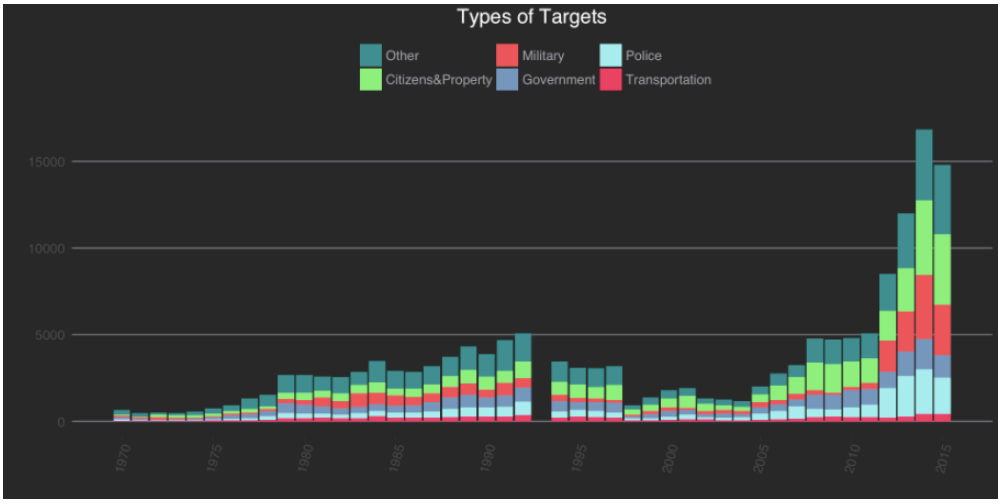


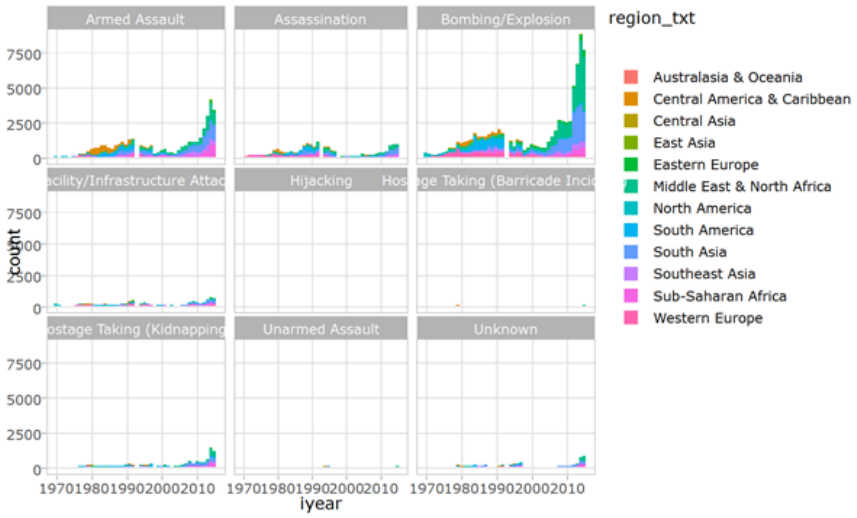
For the text mining section of our analysis, we also had to take the extra step of removing incident reports where there the motive was missing. In many of the earlier years (1970-1990) a lot of the motives were missing, which made our analysis a bit challenging.

Design Evolution

When we were deciding what colors schemes to use in our project, we tried to pick out a dark theme, so it would give a feeling that our project is about some serious societal issues. We had to tilt x-axis tick values in order to show them in other year. More importantly, we tried to put enough information in each plot without violating visualization principles.



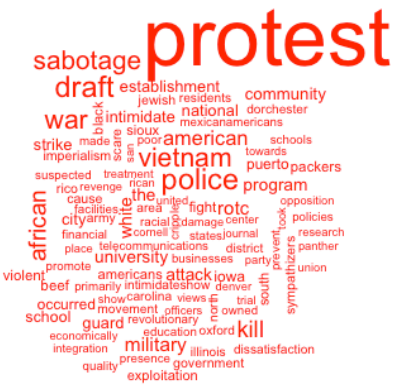




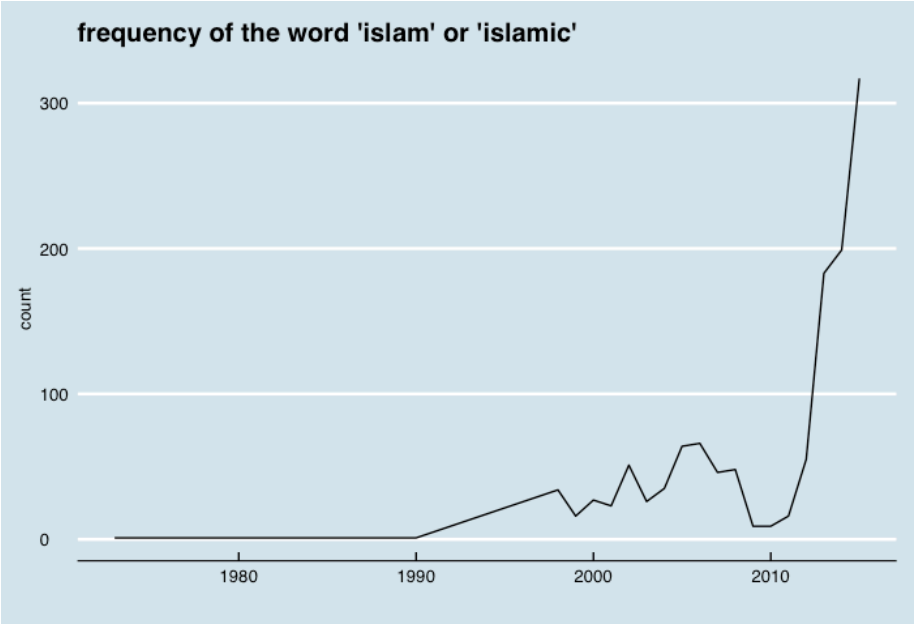
This plot shows that how many incidences happened in each region for different attacks. But due to the few amount of some kind attacks and we have 12 regions in this case, the information conveyed by this plot is hard to tell. Thus, we didn't use this plot in our final project.

The biggest challenge we faced for the text mining part of this project was cleaning the text and coming up with actual words that had some significance. Given the size of our data, this was a very time intensive process since R had to loop through and create all the different cuts of the data that we needed.

When we created word clouds, we first did not use different colors for the different words, which made it hard to clearly distinguish between frequent words and not so frequent words. The examples below show what the word clouds looked before we fixed that issue.



Another challenge we faced when refining our plots was to keep the format and themes consistent throughout. In our first iteration of the website, we were using different themes, but we felt that it would be better for the reader if we kept a consistent style throughout all our plots. Below is an example of what our plots of word popularity over time looked like before we made the styles consistent.



Implementation

When thinking about how we wanted to present the geographical data related to terrorist attacks, we wanted to let the reader explore the data on their own, which is why we decided to make the maps interactive, while also providing enough data on the default view to convey the larger point that we were trying to make, which was that each individual terrorist organization has its own "territory" outside of which it rarely ventured (with the exception of ISIL).

Our first interactive map lets the reader zoom in to particular areas of the world where there are many attacks and click to see the details of each particular attack. We plot all attacks across the entire time period covered by the data, so we cannot exactly extract trends over time from this map. However, our second interactive map features different layers and allows the reader to click through and see how ISIL has evolved and expanded over recent years. Both of these maps combined give the reader a good sense of the global reach of terrorism and how it has been evolving over time.

Evaluation

Some countries in the Middle East, Nigeria, Afghanistan, the Philippines, and Europe appear to suffer the most from terrorism.

From our final work, the first thing we know about is the patterns of terrorism. We find that terrorist incidents saw a huge increase during the period 1970-2015, but with no obvious pattern within each year. For the types of attacks, we find that bombing and explosion is the main one, which we assume is because of its massive killing and injury ability. For the types of target, attacks towards the military and police have been increasing at a fast pace since

Our research question regarding motivation is hard to answer with the analysis we have done here. However, one valuable theme we have uncovered with our analysis is the transition from politically related motivations (think Vietnam War), to more religious motivations (think ISIS). We can see this with the plots that we did over time and also in the word clouds that plot the most popular motivation terms.

Next Steps

Our visualization turned out pretty well given the things we set out to do at the beginning of the project. We covered most of the things we wanted to cover, but one thing that we could have improved upon would be to do some more statistically sound analysis of the data. Given the time constraints we were all working with, and given that the purpose of the class was to learn how to visualize data instead of analyzing it with statistical tools, we decided that was outside the scope of our project and so we focused on simply applying the skills we learned during class.

One idea we had that we could not find out how to implement was to create a sort of network of terrorist organizations where the distance between nodes would be somehow correlated to the common motives and types of attacks shares by terrorist organizations. Our intention was then to see how these common terrorist organizations evolve over time and analyze these patterns independently. This would be our biggest priority if we had more time to work on this.