Hi Hallie,

I weaved data visualization and inferential statistics write-up together to make a more coherent data story.

For the inferential statistics part, I did a hypothesis test on correlation, a one-way ANOVA test and a post hoc Tukey's Range Test. This part of writing is on page 5-7.

P.S. The formatting of this document is not ideal. I don't want to spend too much time on formatting since we're still adding new materials to this project. I will format the final report nicely.

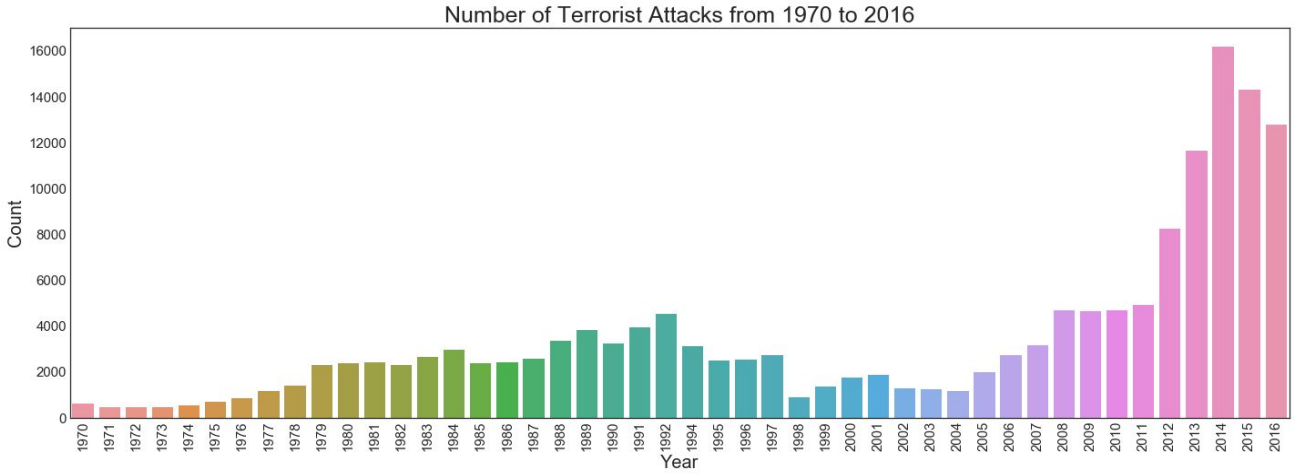
Arianna Lang Wang
Springboard Jan 2nd, 2018 Cohort

Global Terrorism Exploratory Data Analysis Write-up

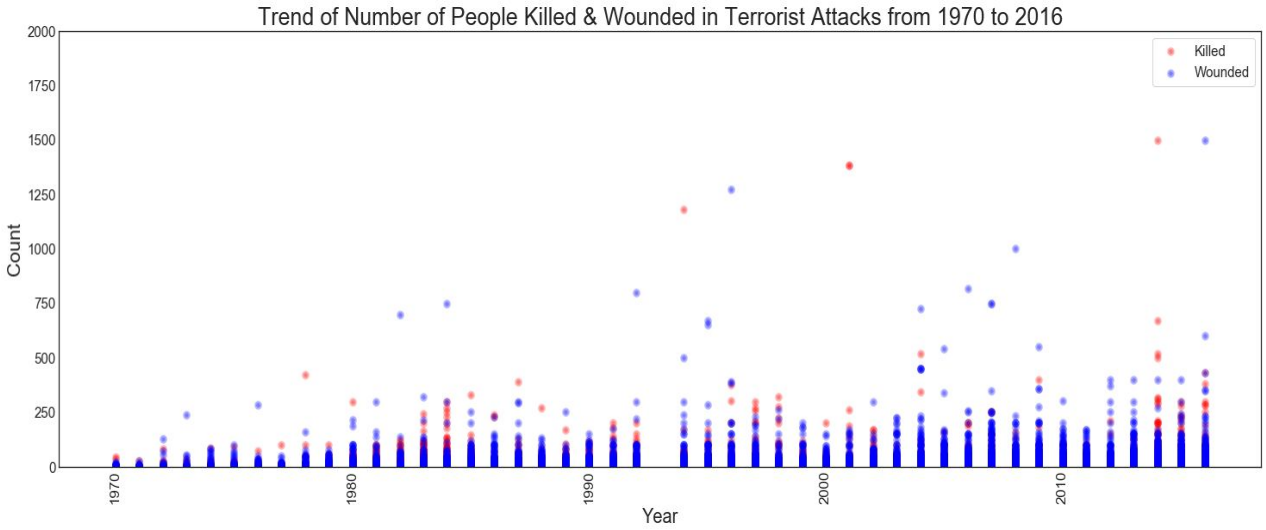
Feb 22nd, 2018

Before analyzing this dataset, I had no prior knowledge of global terrorism except for hearing about it occasionally in the news. It has been an interesting project to work on because it allowed me to gain some much-needed understanding of this important issue in today's society.

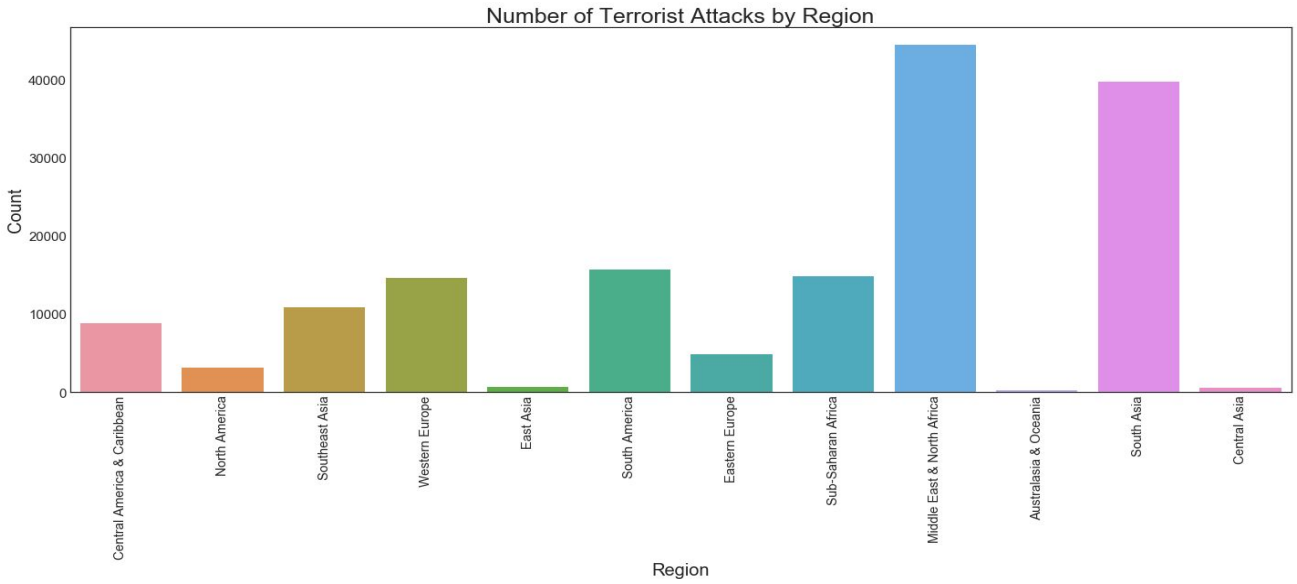
The first important question I asked was: "Has the world become a safer place today?" The answer seems to be "No". The number of global terrorist attacks has been increasing and it has increased drastically since 2012.



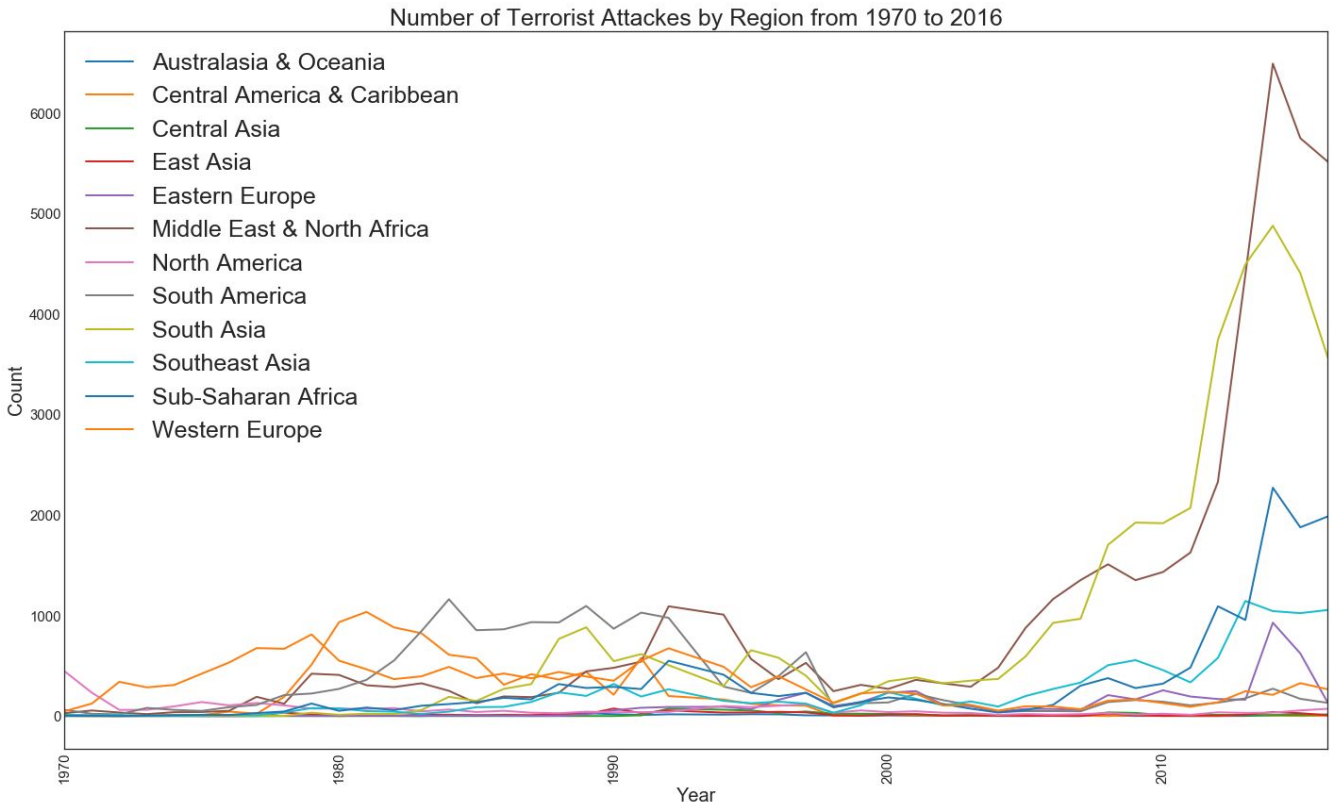
Not surprisingly, the number of people killed or wounded in these attacks are also increasing over the decades.



Now that we've seen a time trend, my second question was: "Is there a geographical trend?"

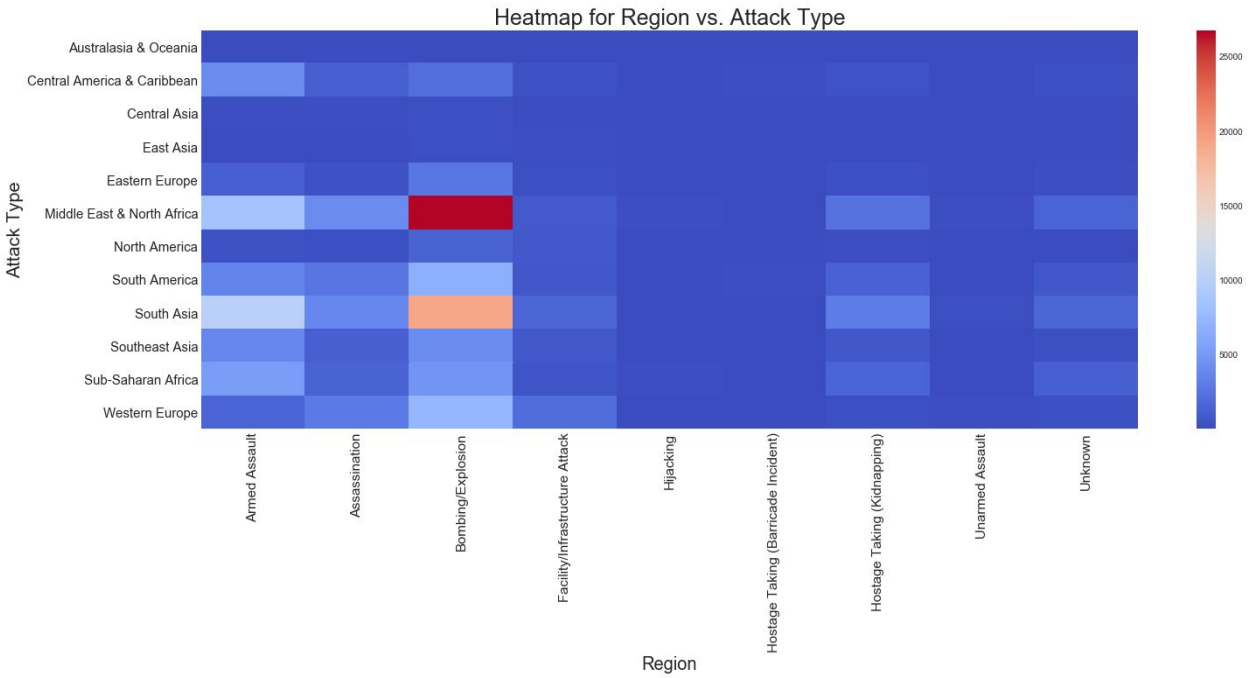


The answer is "Yes". We see that regions such as Middle East/North Africa and South Asia are most vulnerable to terrorist attacks. East Asia, Central Asia and North America are relatively safe regions. However, tying to the time-trend discussed before, a lot of regions suffer from increasing terrorist attack occurences in the recent decade.



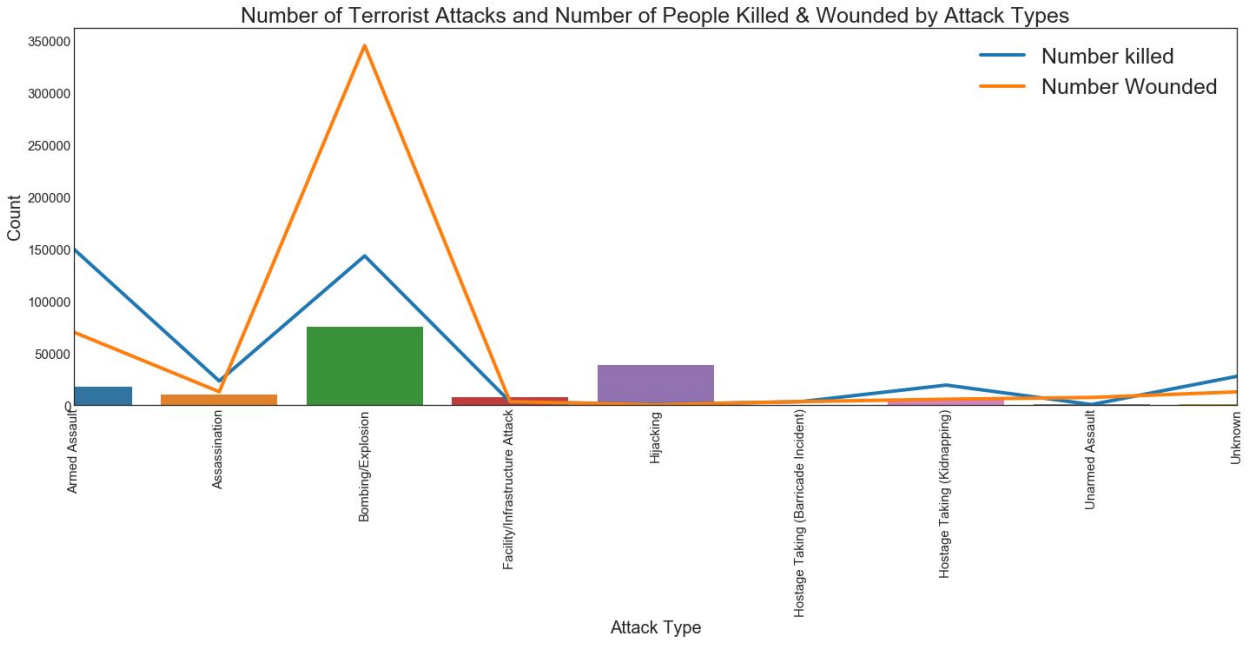
The third question I wanted to ask was around terrorist attack types.

We see previously Middle East/North Africa and South Asia are the most vulnerable regions to terrorist attacks. Are these regions prone to a particular type of terrorist attack?”



The answer is “yes”. We see from the heatmap that Middle East/North Africa and South Asia have the most number of bombing/explosion type of terrorist attacks.

Bombing/explosion is also the most used and the most lethal means by terrorists since it kills and wounds most people.

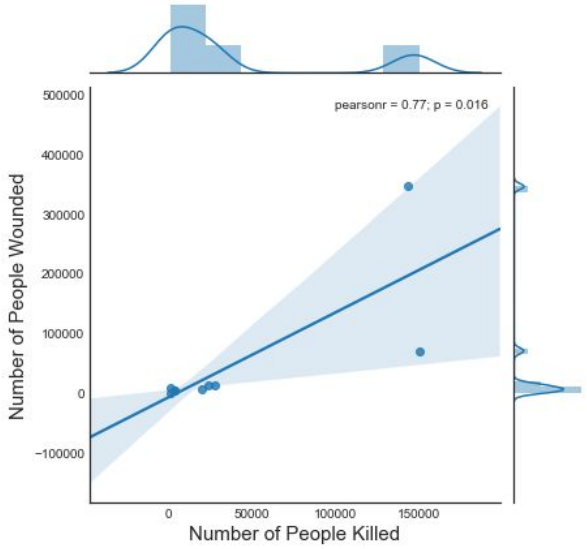


From the graph above, there seems to be a correlation between number of people killed and number of people wounded by attack types. To quantify my observation, I set up a hypothesis test to test this correlation.

Ho: under different attack types, number of people killed and number of people wounded are independent of each other.

Ha: under different attack types, number of people killed and number of people wounded are not independent of each other.

After running the test, the Pearson-correlation statistics r = 0.77 with a p-value = 0.016. Since p-value is less than 0.05, we reject the null hypothesis and conclude that under different attack types, number of people killed and number of people wounded are not independent of each other. We can also say that there is a strong positive correlation between number of people killed and number of people wounded because the correlation coefficient r is 0.77.



From the “Number of Terrorist Attacks and Number of People Killed & Wounded by Attack Types” graph above, it is also easy to see that there are certain attacks types that kill more people than other types. For instance, Bombing/explosion incidences kill most people compared to other attack types. However, Bombing/explosion was also the most frequent atack type used by the terrorists. Doing a one-way ANOVA test allows us to compare the average number of people killed for each attack type on a per incident basis.

Ho: per incident, all the average number of people killed for each attack type are equal.

Ha: per incident, not all the average number of people killed for each attack type are equal.

After running the ANOVA test, the F- statistics is 416.76 and the p-value is zero. Hence, we reject the null hypothesis and conclude that per incident, not all the means of number of people killed for each attack type are equal.

However, we still do not know which one(s) are different from the others. This is why ANOVA test is often followed by a post hoc analysis. Tukey's range test, named after the American mathematician John Tukey, is a common method used as post hoc analysis after one-way ANOVA. This test compares all possible pairs and we can use it to precisely identify difference between two means that's greater than the expected standard error.

For each pair of mean values:

Ho: the means are equal.

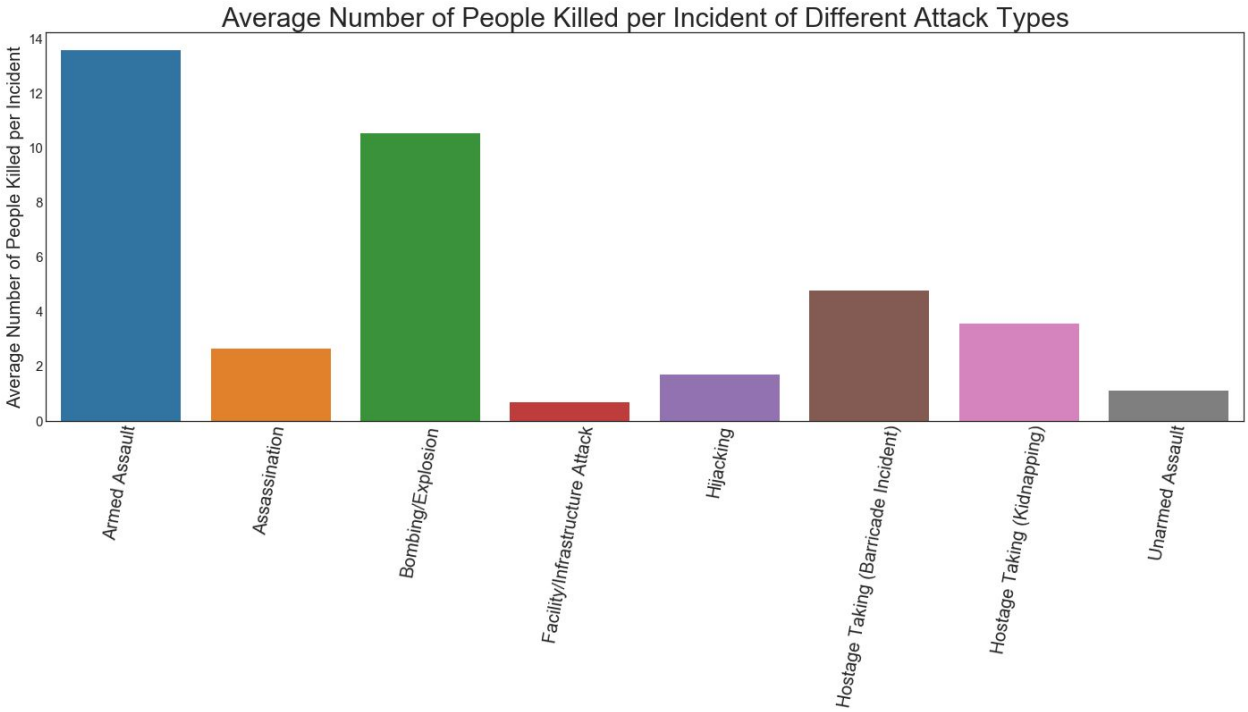
Ha: the means are not equal.

The result below shows each pairs' mean difference. If the pair's mean values are statistical-significantly different, then we reject the null hypothesis and conclude that that pairs' mean values are not equal. (The 'reject' column will have a true value.)

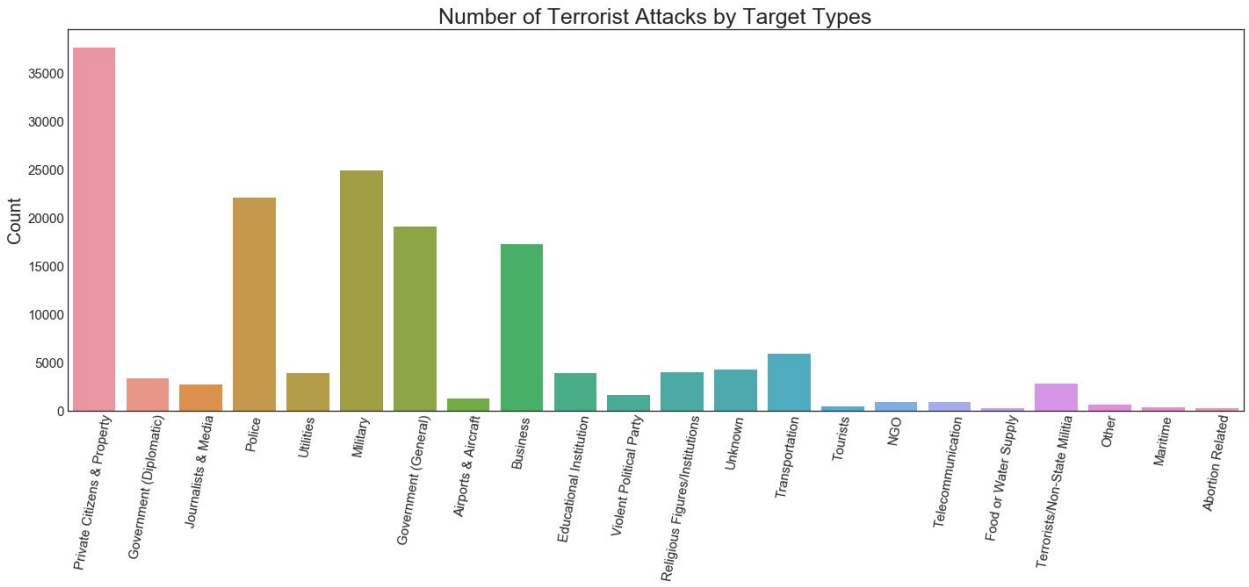
Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
Armed Assault	Assassination	-2.563	-2.816	-2.31	True
Armed Assault	Bombing/Explosion	-1.9753	-2.15	-1.8006	True
Armed Assault	Facility/Infrastructure Attack	-3.4607	-3.8004	-3.1209	True
Armed Assault	Hijacking	-2.3124	-3.473	-1.1519	True
Armed Assault	Hostage Taking (Barricade Incident)	0.3962	-0.5853	1.3778	False
Armed Assault	Hostage Taking (Kidnapping)	-1.9285	-2.2421	-1.6148	True
Armed Assault	Unarmed Assault	-2.9285	-3.9159	-1.941	True
Armed Assault	Unknown	0.7486	0.3616	1.1355	True
Assassination	Bombing/Explosion	0.5877	0.3549	0.8205	True
Assassination	Facility/Infrastructure Attack	-0.8977	-1.2707	-0.5247	True
Assassination	Hijacking	0.2506	-0.9201	1.4213	False
Assassination	Hostage Taking (Barricade Incident)	2.9592	1.9657	3.9528	True
Assassination	Hostage Taking (Kidnapping)	0.6345	0.2852	0.9839	True
Assassination	Unarmed Assault	-0.3655	-1.3648	0.6339	False
Assassination	Unknown	3.3115	2.8952	3.7279	True
Bombing/Explosion	Facility/Infrastructure Attack	-1.4854	-1.8103	-1.1604	True
Bombing/Explosion	Hijacking	-0.3371	-1.4934	0.8192	False
Bombing/Explosion	Hostage Taking (Barricade Incident)	2.3715	1.395	3.3481	True
Bombing/Explosion	Hostage Taking (Kidnapping)	0.0468	-0.2507	0.3444	False
Bombing/Explosion	Unarmed Assault	-0.9532	-1.9356	0.0293	False
Bombing/Explosion	Unknown	2.7238	2.3499	3.0978	True
Facility/Infrastructure Attack	Hijacking	1.1483	-0.0442	2.3407	False
Facility/Infrastructure Attack	Hostage Taking (Barricade Incident)	3.8569	2.8378	4.876	True
Facility/Infrastructure Attack	Hostage Taking (Kidnapping)	1.5322	1.1158	1.9487	True
Facility/Infrastructure Attack	Unarmed Assault	0.5322	-0.4925	1.557	False
Facility/Infrastructure Attack	Unknown	4.2092	3.7351	4.6833	True
Hijacking	Hostage Taking (Barricade Incident)	2.7087	1.202	4.2153	True
Hijacking	Hostage Taking (Kidnapping)	0.384	-0.8013	1.5692	False
Hijacking	Unarmed Assault	-0.616	-2.1265	0.8944	False
Hijacking	Unknown	3.061	1.8542	4.2677	True
Hostage Taking (Barricade Incident)	Hostage Taking (Kidnapping)	-2.3247	-3.3354	-1.314	True
Hostage Taking (Barricade Incident)	Unarmed Assault	-3.3247	-4.7024	-1.947	True
Hostage Taking (Barricade Incident)	Unknown	0.3523	-0.6834	1.3881	False
Hostage Taking (Kidnapping)	Unarmed Assault	-1.0	-2.0164	0.0164	False
Hostage Taking (Kidnapping)	Unknown	2.677	2.2213	3.1328	True
Unarmed Assault	Unknown	3.677	2.6357	4.7183	True

To visualize what we've been doing with the one-way ANOVA test and the Tukey's range test, I made a graph below showing the average number of people killed per incident of different attack types. This graph clearly indicates that not all of the average number of people killed for each attack type are equal. From the first line of the Tukey's test result table, we reject the null hypothesis that the average number of people killed from an armed assault terrorist attack equals to the average number of people killed from an assassination terrorist attack. Again, this graph below clearly shows that these two means are very different.

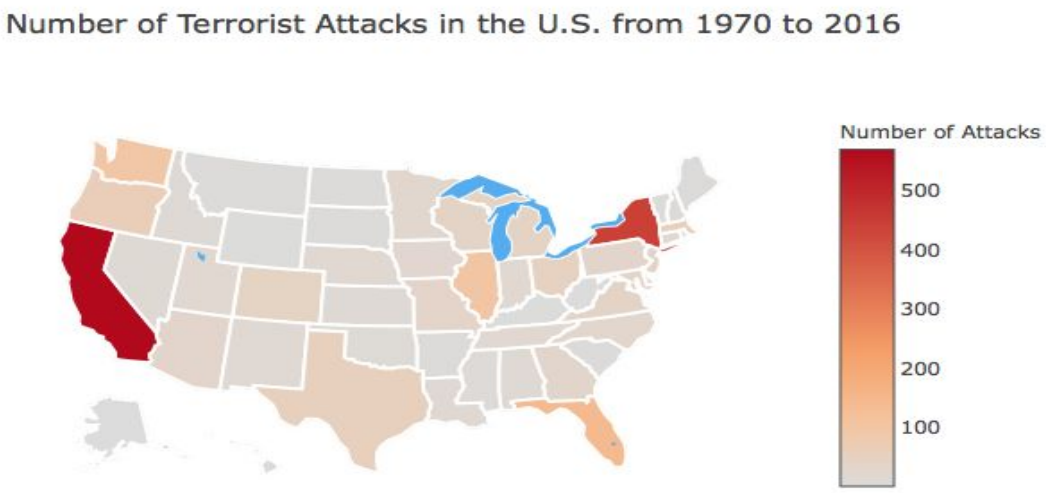


Now that we have some understanding on terrorist attack types, a natural question to ask next is on target types. Are there any particular target groups that are more prone to terrorist attacks? The bar chart below shows that private citizens/properties and military are the most frequent targets.

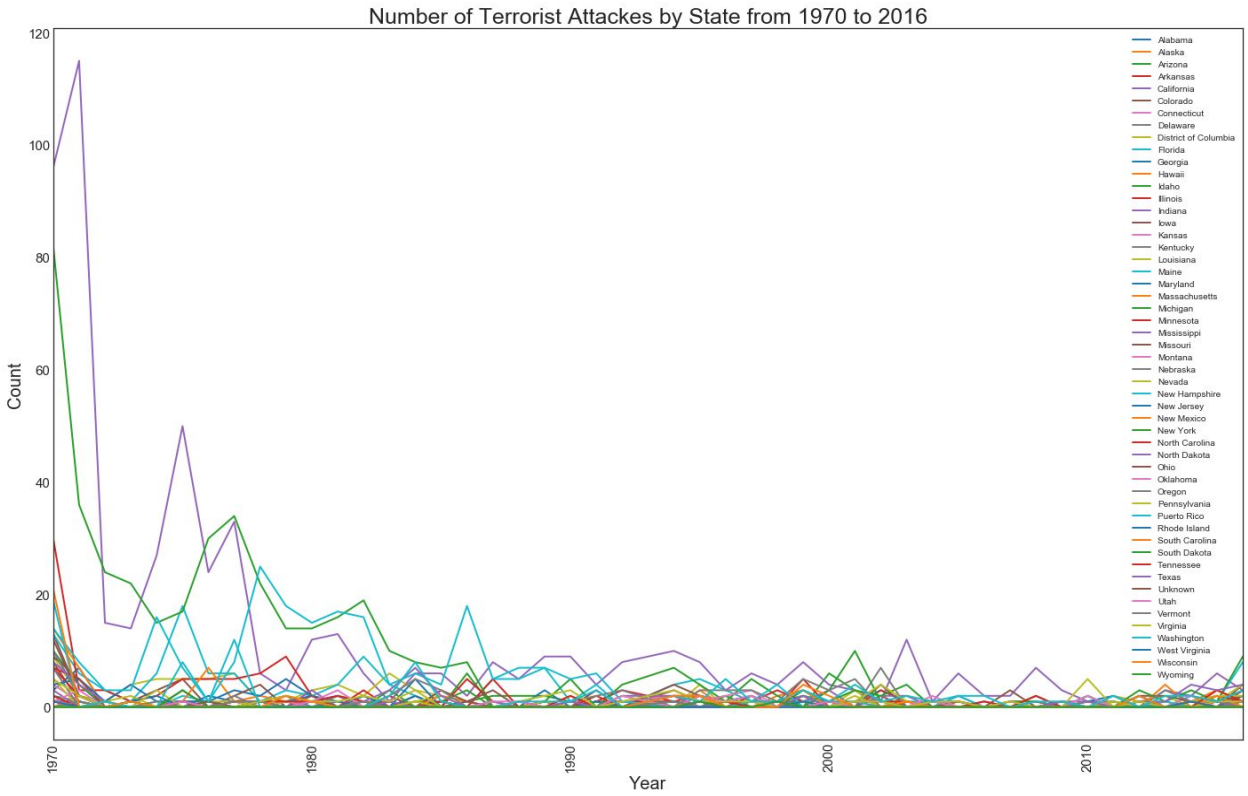


Along with exploring 'global' terrorism trends, I also want to explore a bit closer to home.

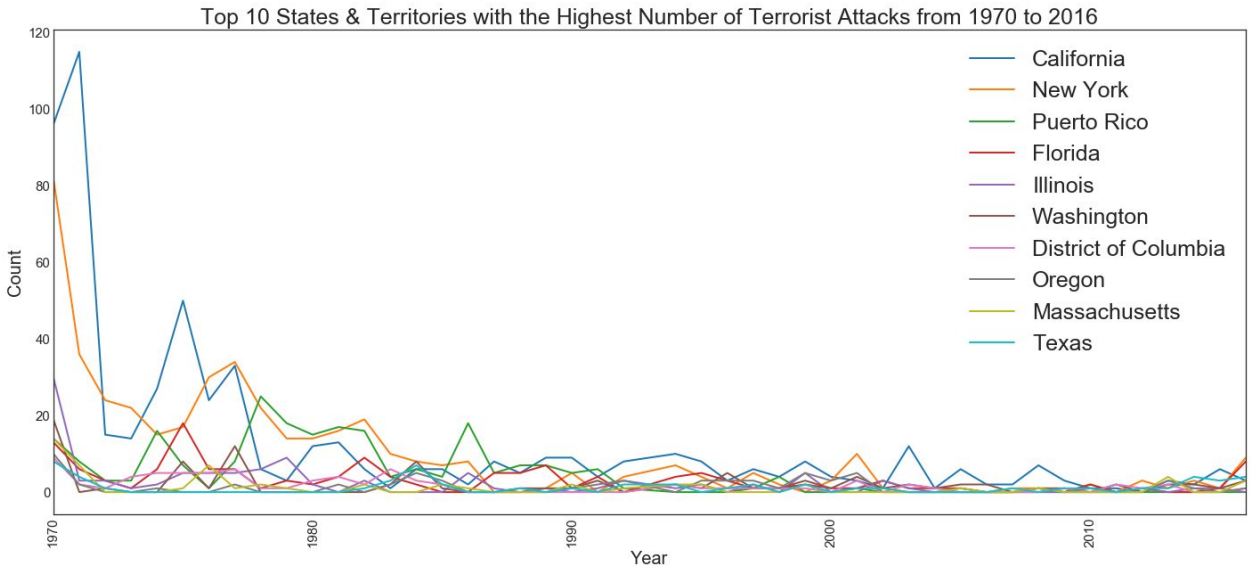
The map below showed that California and New York had the most number of terrorist attacks from 1970 to 2016.



To investigate further, I did a line chart with 'year' on the x-axis and 'number of attacks' on the y-axis. The chart shows that the United States has actually become a lot safer nowadays compared to what it were in the 1970s and 1980s.



If this line chart with all the states and territories are too confusing to look at, I made a new chart below zooming in on the top 10 states and territories that have the highest number of attacks from 1970 to 2016. Note that the legend is in a descending order of the number of attacks occured during this period of time.



Overall, I gained valuable information on terrorism on both a global and a domestic scale after analyzing this dataset. Of course, there are much more I can do with this dataset, but for now, I'm happy with what I've done so far.