CS-697 (INDEPENDENT STUDY IN C S)

Global Terrorist Attacks

Submitted By-Swati Mishra

UIN- 01212676

December 12th, 2022

Summary:

Terrorist attacks have an impact on the global economy and security, terrorist attacks have increased in recent years, blocking economic development. As a result, it is necessary to pinpoint the locations of the attacks. It is also necessary to understand the global number of terrorist attacks, as well as the most common types of attacks, in order to predict what type of terrorist attack will occur and in which areas of the world.

The objective of the project is to visualize the dataset which consist of incidents that have occurred during the time period (1970-2020) and includes more than 200,000 cases. It is an open-source database that contains information on over 200,000 terrorist events worldwide from 1970 to 2020.

System Architecture:

Execution Task for the Project:

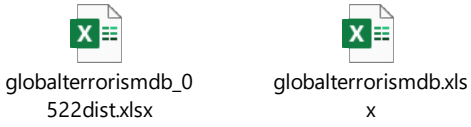
- Data Gathering
- Data Cleaning
- Descriptive Analysis
- Clustering Analysis
- Data Visualization
- Designing and Publishing Tableau Dashboards

Tools:

- Tableau
- Open Refine
- IBM SPSS

Dataset:

Below is the dataset for your reference:



Dataset Overview:

The dataset, which contains data on global terrorism from 1970 to 2020, is derived from the University of Maryland Repository. Both qualitative and quantitative data are included in the dataset. Over 200,000 records are included in the dataset. The dataset contains 1355 variables in total. When analyzing the data, we can determine that various attack types have taken place in various regions with other related factors.

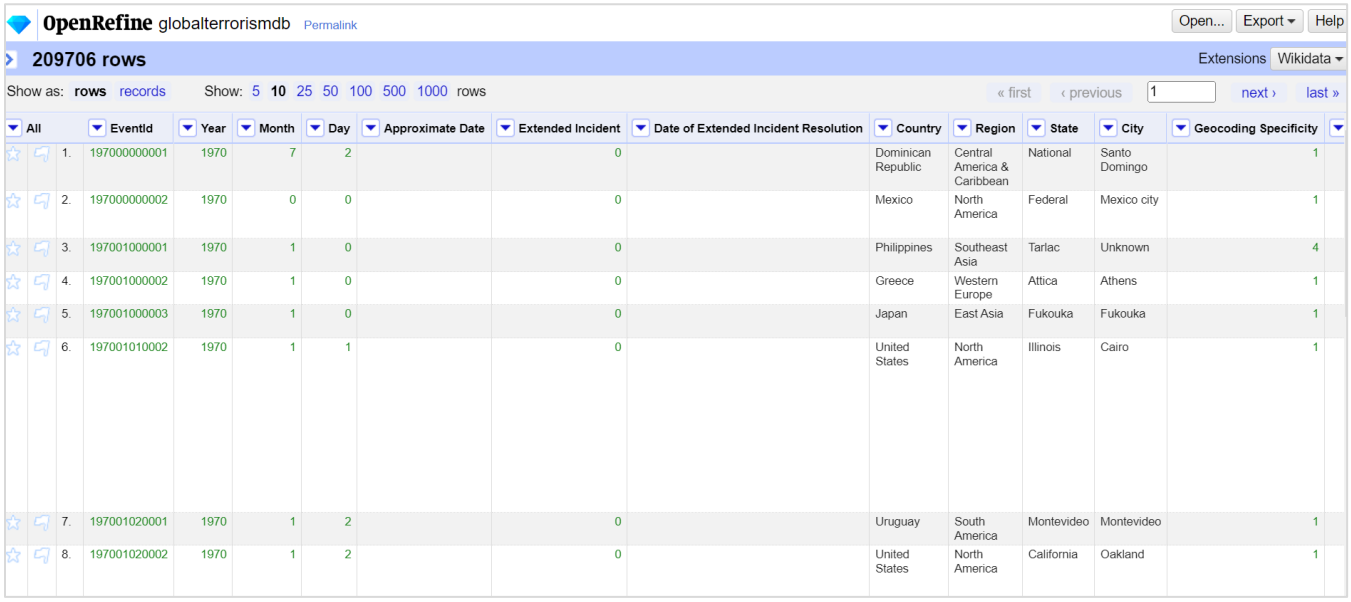
Data Cleaning:

Data cleaning is the process of removing inaccurate, duplicate, or other wrong data from a dataset. These errors, which frequently occur when two or more datasets are joined, can include poorly structured data, redundant entries, mislabeled data, and other problems. Data cleansing enhances the quality of your data and any business decisions you make based on it.

To clean the dataset for my project, I used Open Refine (i.e., Change the column name, removed the unwanted column). Open Refine (was previously known as Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

To clean the dataset, firstly upload the dataset in Open Refine and then perform the desired operation. All the carried operation can be exported then in the Json format.

Below is the snapshot of cleaned dataset and steps carried out are listed in the json file.



OpenRefine globalterrorismdb Permalink — Open... Export Help

209706 rows — Extensions: Wikidata

Show as: rows records — Show: 5 10 25 50 100 500 1000 rows — « first ‹ previous 1 next › last »

All		EventId	Year	Month	Day	Approximate Date	Extended Incident	Date of Extended Incident Resolution	Country	Region	State	City	Geocoding Specificity
1.	197000000001	1970	7	2		0		Dominican Republic	Central America & Caribbean	National	Santo Domingo	1
2.	197000000002	1970	0	0		0		Mexico	North America	Federal	Mexico city	1
3.	197001000001	1970	1	0		0		Philippines	Southeast Asia	Tarlac	Unknown	4
4.	197001000002	1970	1	0		0		Greece	Western Europe	Attica	Athens	1
5.	197001000003	1970	1	0		0		Japan	East Asia	Fukouka	Fukouka	1
6.	197001010002	1970	1	1		0		United States	North America	Illinois	Cairo	1
7.	197001020001	1970	1	2		0		Uruguay	South America	Montevideo	Montevideo	1
8.	197001020002	1970	1	2		0		United States	North America	California	Oakland	1

history.json

Descriptive Analysis:

Descriptive analysis is a sort of data analysis that helps in accurately describing, displaying, or summarizing data points so that patterns may appear that satisfy all the data's requirements.

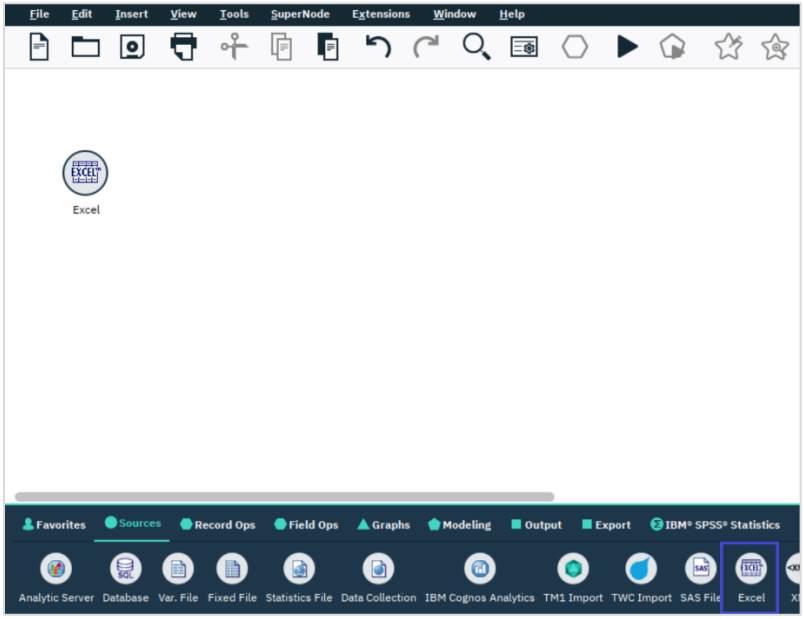
- Statistical and Graphical Analysis:

  The process of gathering and analyzing data with the purpose of identifying patterns and trends and informing decision-making is known as statistical analysis and on the other hand The data are visualized through graphical analysis, which helps in understanding patterns and the relationships between process parameters.

The above analysis is performed using IBM SPSS. IBM SPSS Statistics is a fast and powerful solution that propels research analysis in numerous industries. SPSS Statistics is used in education, market research, healthcare, government, and retail throughout the entire analytics process, from planning and data collection to analysis, reporting and deployment.

Steps for IBM SPSS:

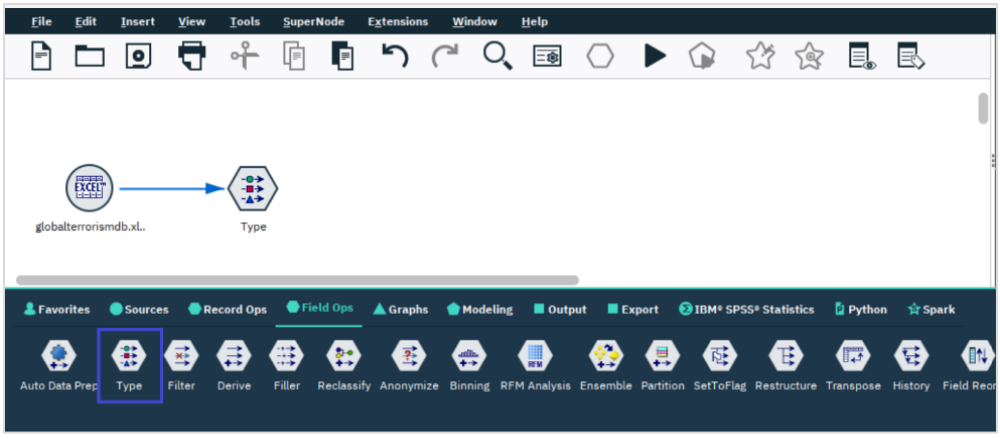
- The dataset was uploaded to IBM SPSS using the excel node located under Sources to begin the statistical analysis, as seen below:

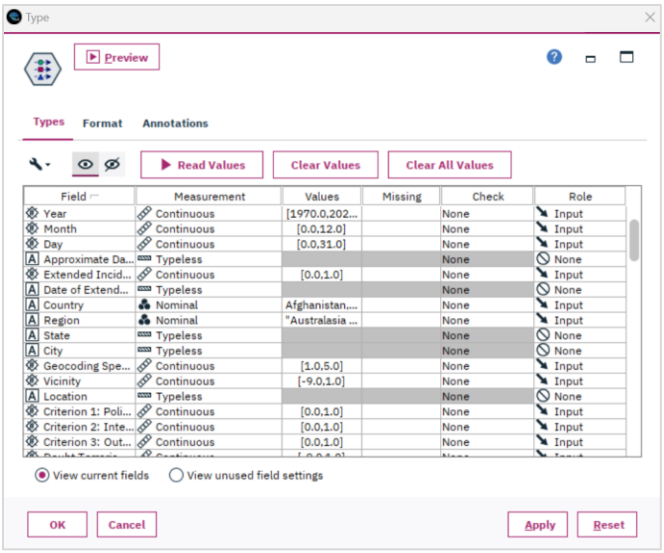


- Once the dataset is uploaded successfully, preview the same to validate as shown below:

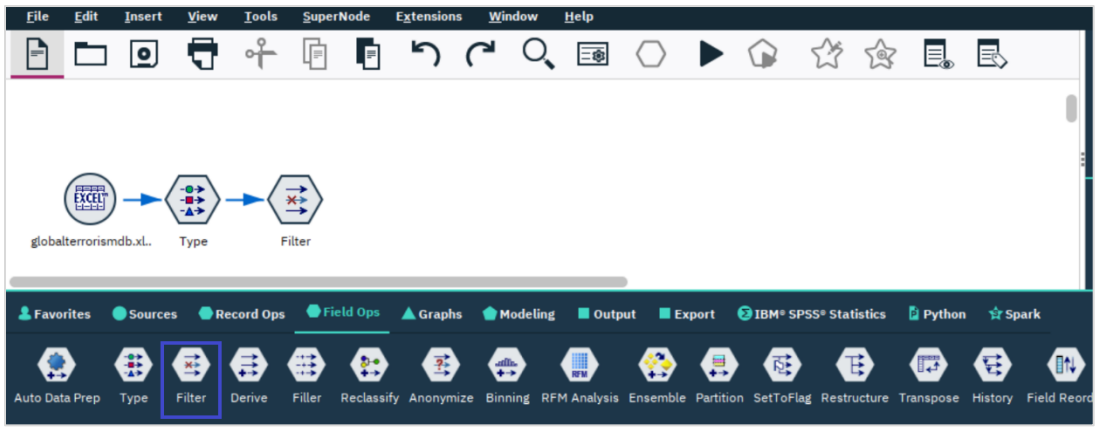
	Year	Month	Day	Approximate Date	Extended Incident	Date of Extended Incident Resolution	Country	Region	State	City	Geocod
1	1970....	7.000	2....		0.000		Dominican Republic	Central America & Caribbean	National	Santo Domingo	1.000
2	1970....	0.000	0....		0.000		Mexico	North America	Federal	Mexico city	1.000
3	1970....	1.000	0....		0.000		Philippines	Southeast Asia	Tarlac	Unknown	4.000
4	1970....	1.000	0....		0.000		Greece	Western Europe	Attica	Athens	1.000
5	1970....	1.000	0....		0.000		Japan	East Asia	Fukouka	Fukouka	1.000
6	1970....	1.000	1....		0.000		United States	North America	Illinois	Cairo	1.000
7	1970....	1.000	2....		0.000		Uruguay	South America	Montevideo	Montevideo	1.000
8	1970....	1.000	2....		0.000		United States	North America	California	Oakland	1.000
9	1970....	1.000	2....		0.000		United States	North America	Wisconsin	Madison	1.000
10	1970....	1.000	3....		0.000		United States	North America	Wisconsin	Madison	1.000

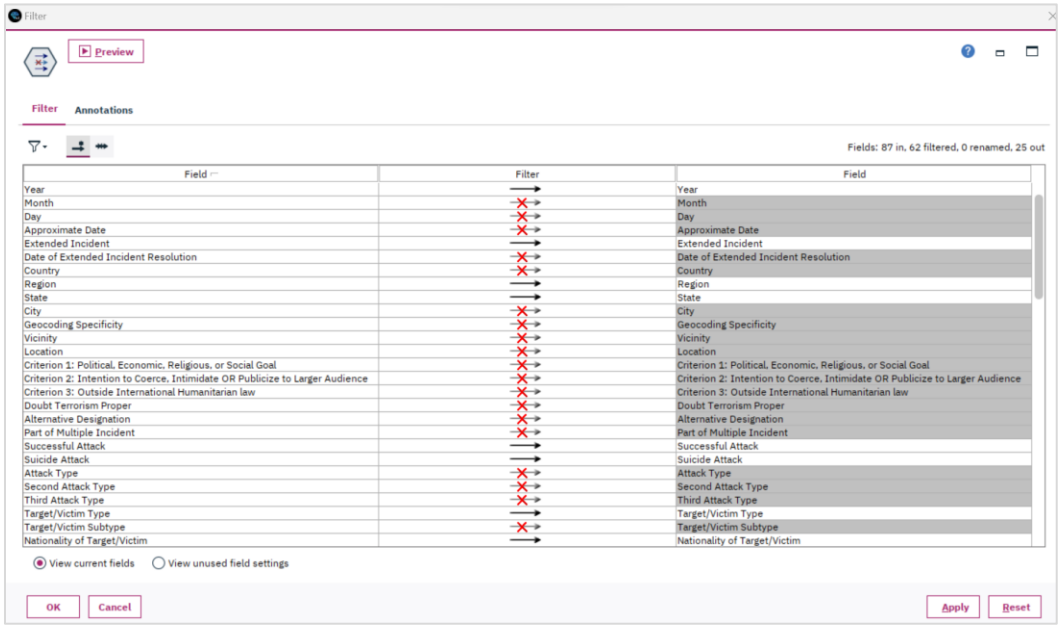
- Following completion of the validation, I select the Type node from Field Ops to read the values as shown below:



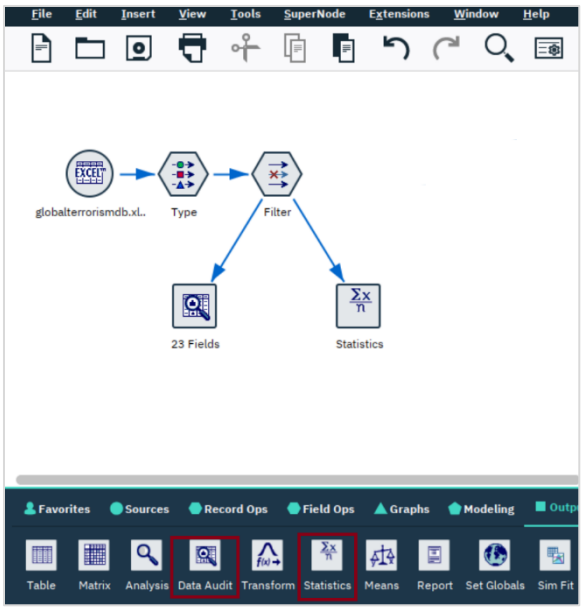


- Once the above operation is done choose Filter node from Field Ops to filter out the unwanted variables shown below:





- Now, to perform statistical analysis, I choose data audit and statistics from output menu as shown below:



Statistical Output:

Total Number of Fatalities

Statistics

Count	197179
Mean	2.431
Min	0.000
Max	1700.000
Range	1700.000
Variance	128.616
Standard Deviation	11.341
Standard Error of Mean	0.026

According to the statistical analysis the average number of fatalities is 2.431 and Minimum number of fatalities is 0 and Maximum number of fatalities is 1700.

Number of US Fatalities

Statistics

Count	145269
Mean	0.039
Min	0.000
Max	1361.000
Range	1361.000
Variance	26.107
Standard Deviation	5.110
Standard Error of Mean	0.013

According to the statistical analysis the average number of US fatalities is 0.039 and Minimum number of US fatalities is 0 and Maximum number of US fatalities is 1361.

Total Number of Injured

Statistics

Count	189770
Mean	3.086
Min	0.000
Max	10878.000
Range	10878.000
Variance	1674.133
Standard Deviation	40.916
Standard Error of Mean	0.094

According to the statistical analysis the average number of injured is 3.086 and Minimum number of injured is 0 and Maximum number of injured is 10878.

Number of US Injured

Statistics

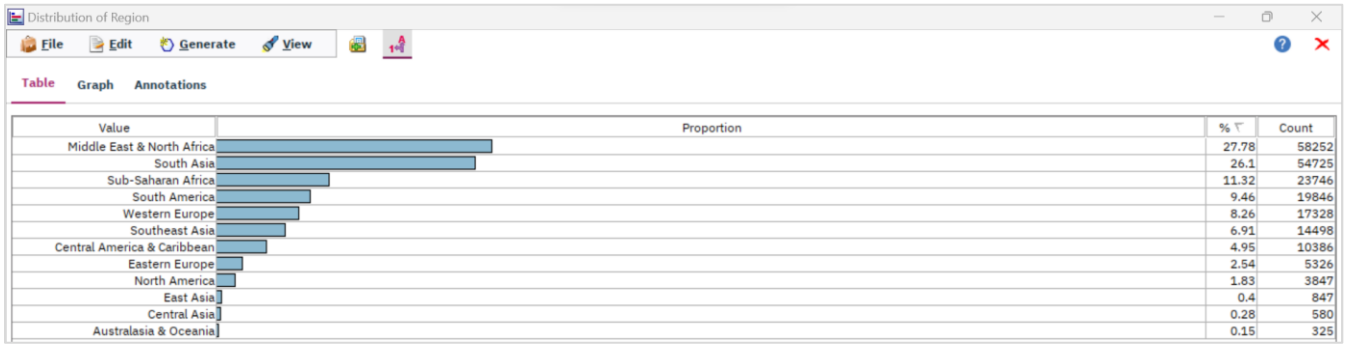
Count	145009
Mean	0.034
Min	0.000
Max	751.000
Range	751.000
Variance	7.620
Standard Deviation	2.760
Standard Error of Mean	0.007

According to the statistical analysis the average number of US injured is 0.034 and Minimum number of US injured is 0 and Maximum number of US injured is 751.

Graphical Output:

Distribution of Region:

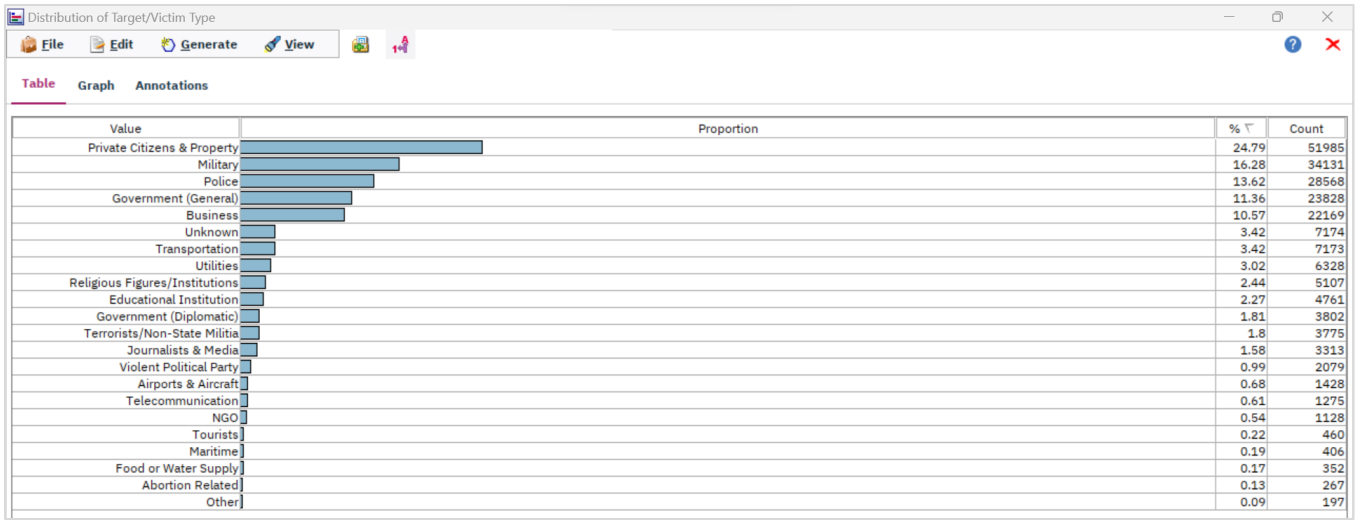
The distribution below demonstrates that the Middle East and North Africa make up most of the region affected by global terrorism w.r.t other regions.



Value	Proportion	%	Count
Middle East & North Africa		27.78	58252
South Asia		26.1	54725
Sub-Saharan Africa		11.32	23746
South America		9.46	19846
Western Europe		8.26	17328
Southeast Asia		6.91	14498
Central America & Caribbean		4.95	10386
Eastern Europe		2.54	5326
North America		1.83	3847
East Asia		0.4	847
Central Asia		0.28	580
Australasia & Oceania		0.15	325

Proportion of Target/Victim Type:

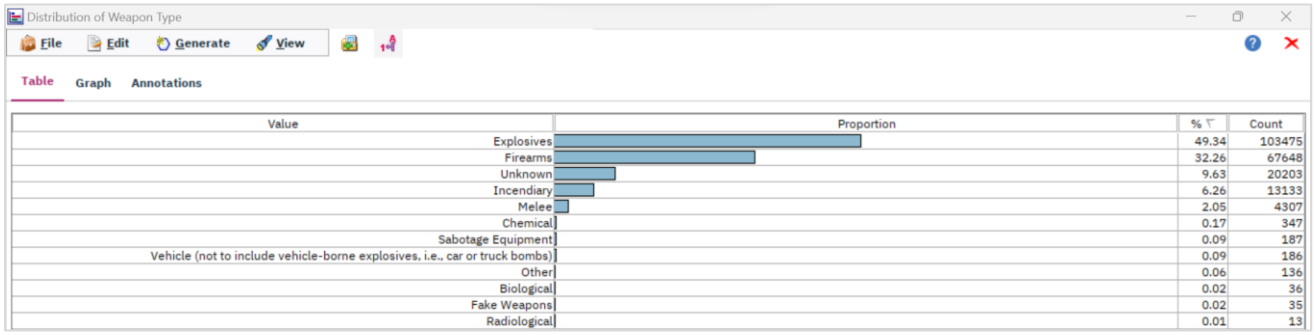
Below graph represents the distribution of Target/Victim Type in which it can be clearly seen that most of them belongs to Private Citizens & Property (i.e., 24.79%).



Value	Proportion	%	Count
Private Citizens & Property		24.79	51985
Military		16.28	34131
Police		13.62	28568
Government (General)		11.36	23828
Business		10.57	22169
Unknown		3.42	7174
Transportation		3.42	7173
Utilities		3.02	6328
Religious Figures/Institutions		2.44	5107
Educational Institution		2.27	4761
Government (Diplomatic)		1.81	3802
Terrorists/Non-State Militia		1.8	3775
Journalists & Media		1.58	3313
Violent Political Party		0.99	2079
Airports & Aircraft		0.68	1428
Telecommunication		0.61	1275
NGO		0.54	1128
Tourists		0.22	460
Maritime		0.19	406
Food or Water Supply		0.17	352
Abortion Related		0.13	267
Other		0.09	197

Distribution of different Weapon Types:

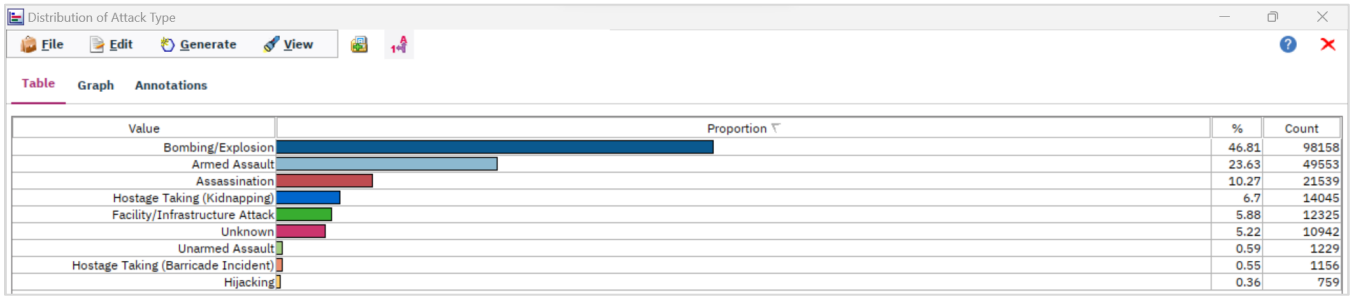
The following graph displays the number of different weapon types used in worldwide terrorist attacks:



Value	Proportion	%	Count
Explosives		49.34	103475
Firearms		32.26	67648
Unknown		9.63	20203
Incendiary		6.26	13133
Melee		2.05	4307
Chemical		0.17	347
Sabotage Equipment		0.09	187
Vehicle (not to include vehicle-borne explosives, i.e., car or truck bombs)		0.09	186
Other		0.06	136
Biological		0.02	36
Fake Weapons		0.02	35
Radiological		0.01	13

Distribution of Attack Types:

The distribution of attack types is seen in the graph below, where it is evident that bombing and explosion attacks represent the majority of attacks.



Value	Proportion	%	Count
Bombing/Explosion		46.81	98158
Armed Assault		23.63	49553
Assassination		10.27	21539
Hostage Taking (Kidnapping)		6.7	14045
Facility/Infrastructure Attack		5.88	12325
Unknown		5.22	10942
Unarmed Assault		0.59	1229
Hostage Taking (Barricade Incident)		0.55	1156
Hijacking		0.36	759

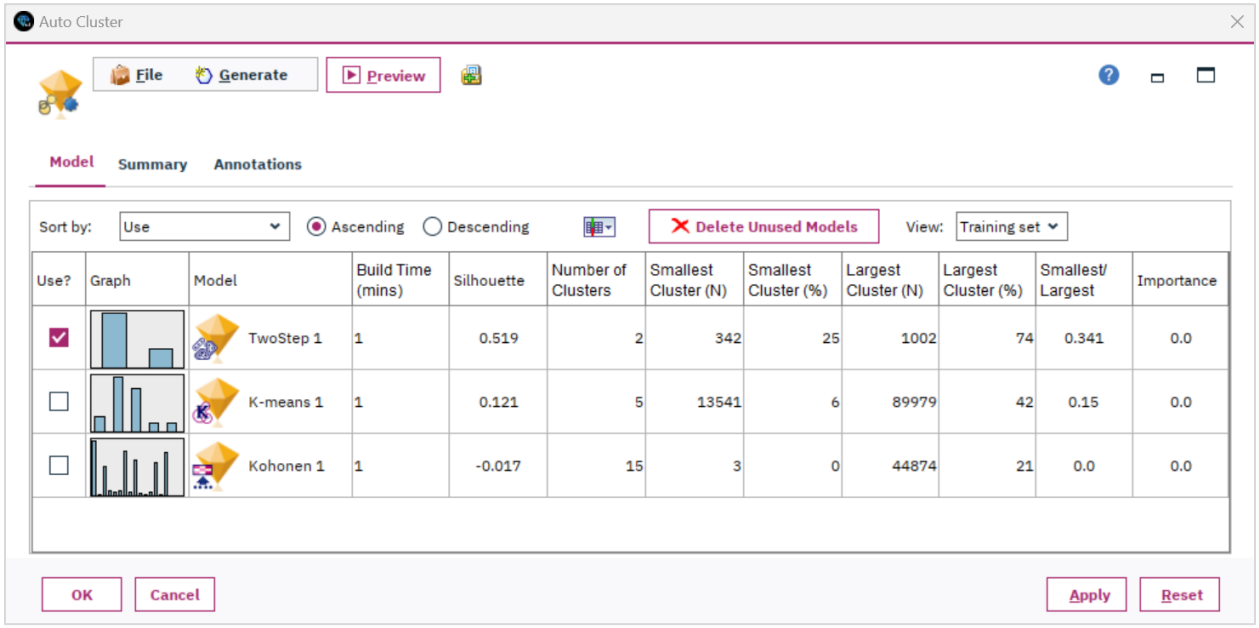
Clustering Analysis:

Cluster analysis is a type of exploratory analysis that looks for patterns in the data. If the grouping was not previously known, it attempts to locate homogenous groupings of cases. It does not distinguish between dependent and independent variables because it is exploratory. Binary, nominal, ordinal, and scale (interval or ratio) data can all be handled by the many cluster analysis techniques that SPSS provides.

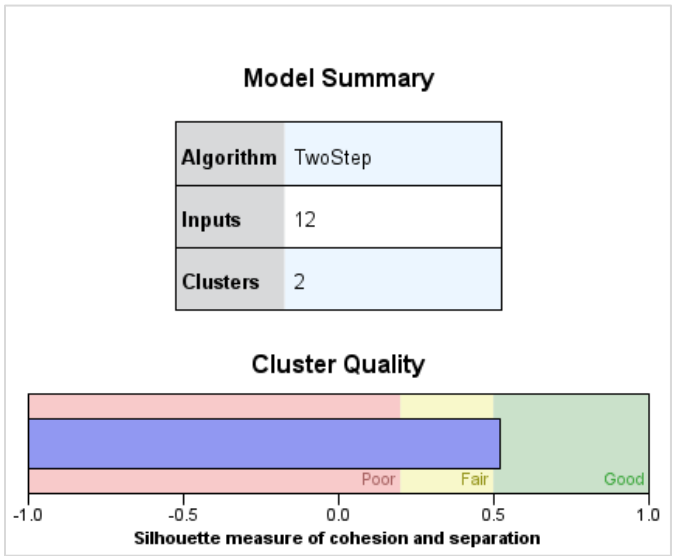
To do the cluster analysis in IBM SPSS, the dataset and its type must first be loaded and then we can choose Auto Cluster option from Modeling and then choose the variables and execute the cluster.

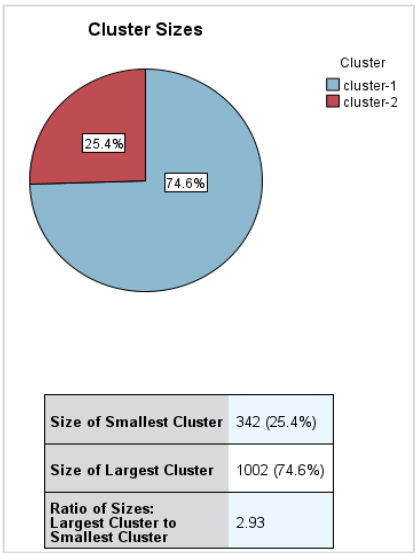
Initial Model:

The initial model consists of 12 variables in it. After running cluster analysis, we can see that only three models returned results, by which we can see that the value of Silhouette is greater than 50% (i.e,51.9 % in our scenario), which is considered as the good clustering scenario as shown below:



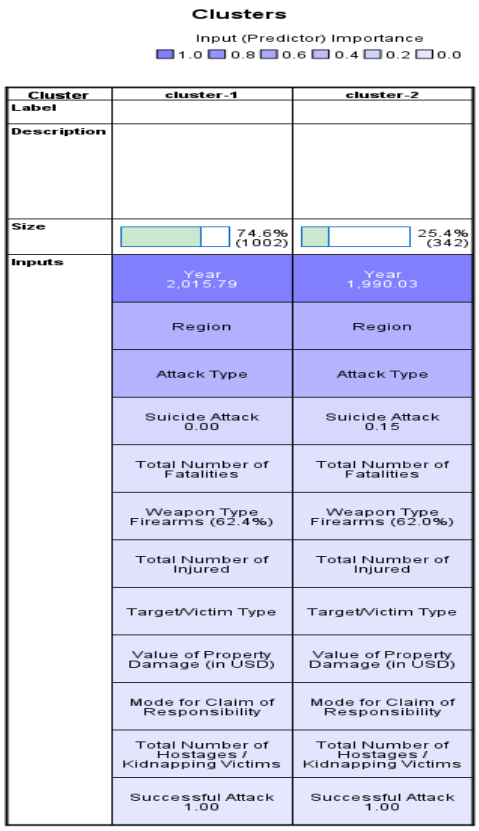
While validating the TwoStep 1 Model it was discovered that two clusters are obtained from the 12 variables as shown below:



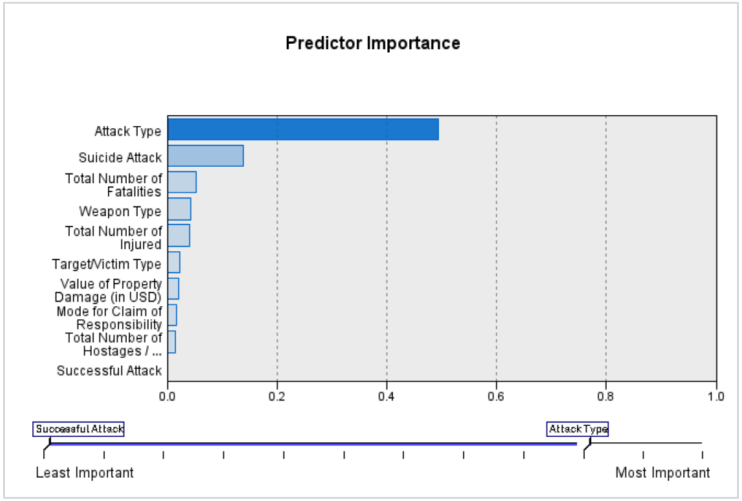


The above graphic demonstrates how our input was separated into 2 clusters, with the smallest cluster size being 342 and the largest cluster size being 1002, respectively. The ratio of the biggest cluster to the smallest cluster is 2.93.

Cluster Distribution:



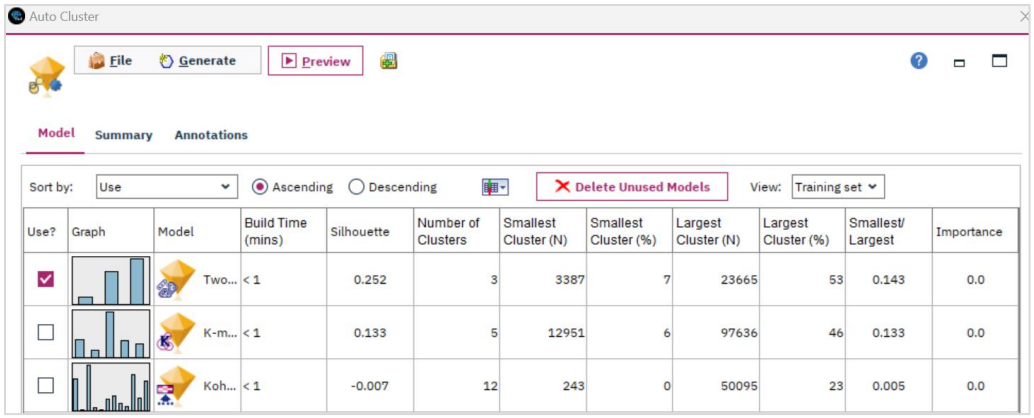
Cluster	cluster-1	cluster-2
Label		
Description		
Size	74.6% (1002)	25.4% (342)
Inputs	Year 2,015.79	Year 1,990.03
	Region	Region
	Attack Type	Attack Type
	Suicide Attack 0.00	Suicide Attack 0.15
	Total Number of Fatalities	Total Number of Fatalities
	Weapon Type Firearms (62.4%)	Weapon Type Firearms (62.0%)
	Total Number of Injured	Total Number of Injured
	Target/Victim Type	Target/Victim Type
	Value of Property Damage (in USD)	Value of Property Damage (in USD)
	Mode for Claim of Responsibility	Mode for Claim of Responsibility
	Total Number of Hostages / Kidnapping Victims	Total Number of Hostages / Kidnapping Victims
	Successful Attack 1.00	Successful Attack 1.00



The least important variable, which is shown in the above Predictor Importance chart and graph, can be removed in order to enhance the clustering model and clustering score. The variables can be eliminated from the filtering node and then the clustering model will be executed.

Refined Model:

After making the above modifications to the model and running it again, it was discovered that the value of Silhouette had decreased and was now 82.5%, which was lower than the initial model. Hence, it's not a good clustering scenario, as shown below:



Use?	Graph	Model	Build Time (mins)	Silhouette	Number of Clusters	Smallest Cluster (N)	Smallest Cluster (%)	Largest Cluster (N)	Largest Cluster (%)	Smallest/ Largest	Importance
☑		Two...	< 1	0.252	3	3387	7	23665	53	0.143	0.0
☐		K-m...	< 1	0.133	5	12951	6	97636	46	0.133	0.0
☐		Koh...	< 1	-0.007	12	243	0	50095	23	0.005	0.0

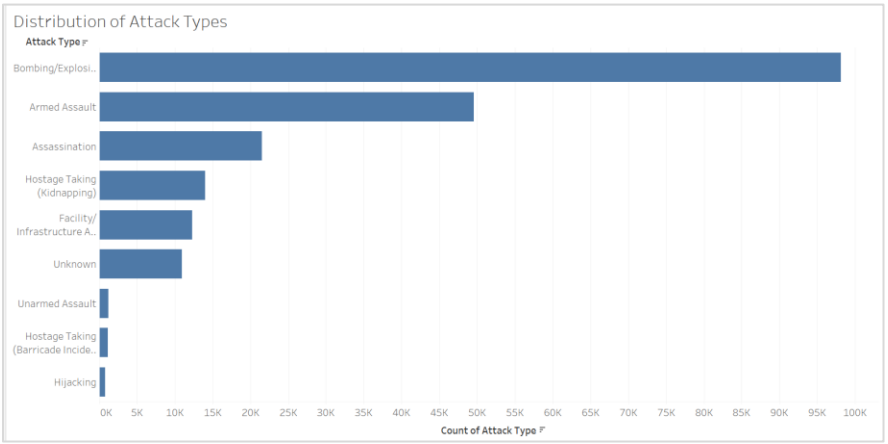
Additionally, while looking at the above results it can be concluded that initial model with 12 variables in it and having 51.9% Silhouette score is the good clustering scenario.

Data Visualization:

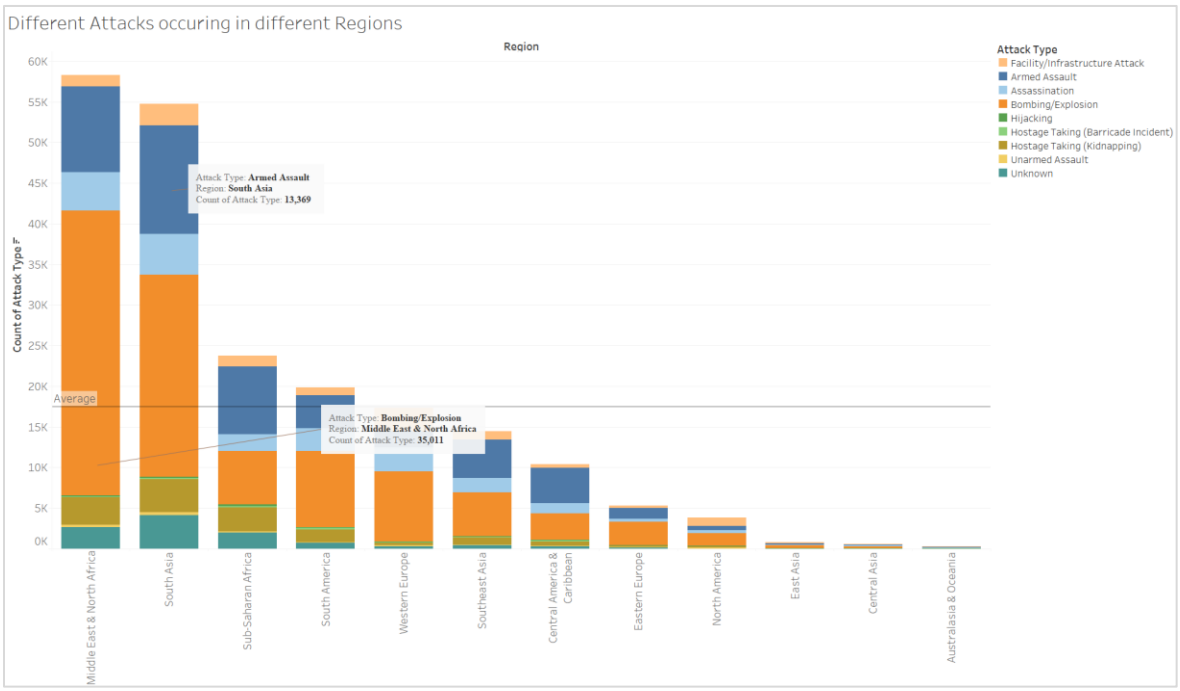
Data visualization is the graphic depiction of information and data. Data visualization tools offer a simple approach to spot and comprehend trends, outliers, and patterns in data by utilizing visual elements like charts, graphs, and maps. These informational visual displays provide complicated data relationships and data-driven insights in a way that is simple to comprehend. Tools and technology for data visualization are crucial in the world of big data to analyze vast volumes of information and make data-driven decisions.

The data visualization has been caried out using Tableau. Tableau enables individuals and businesses to become more data driven. The analytics platform, the market-leading option for contemporary business intelligence, makes it simpler for users to explore and manage data as well as faster to find and share insights that have the potential to transform industries and the global economy.

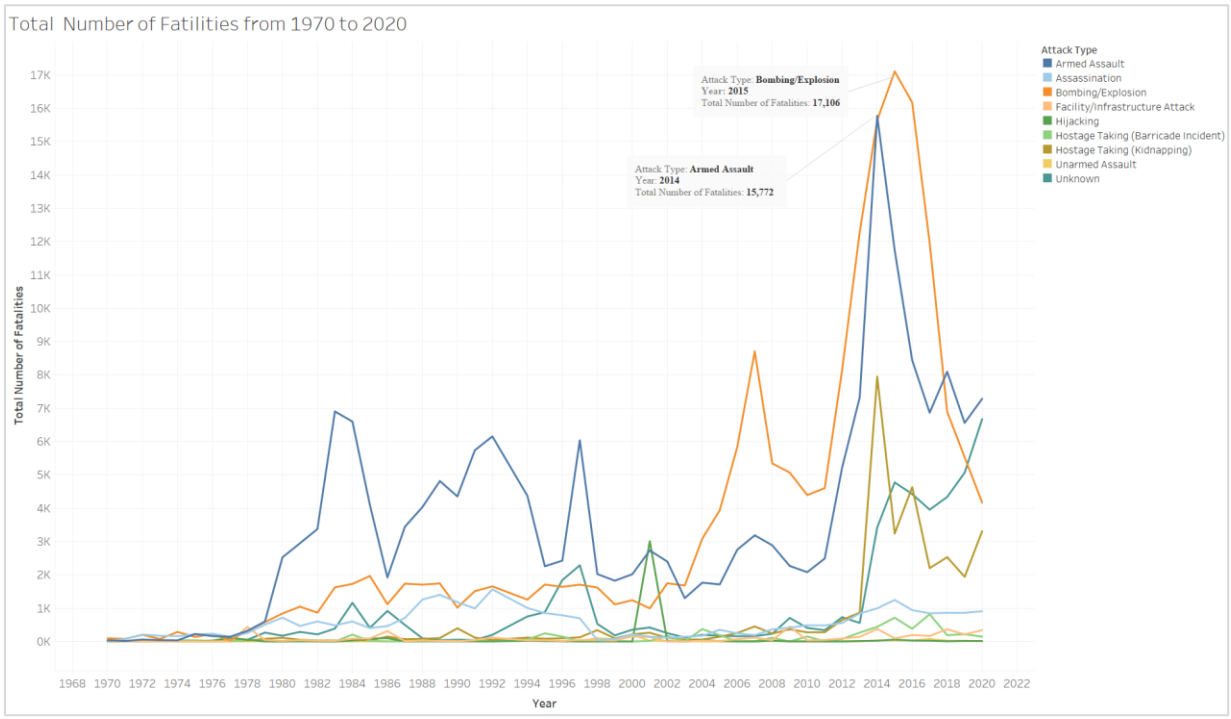
Below are the different data visualization which was been carried out using Tableau to gain insight from the dataset:



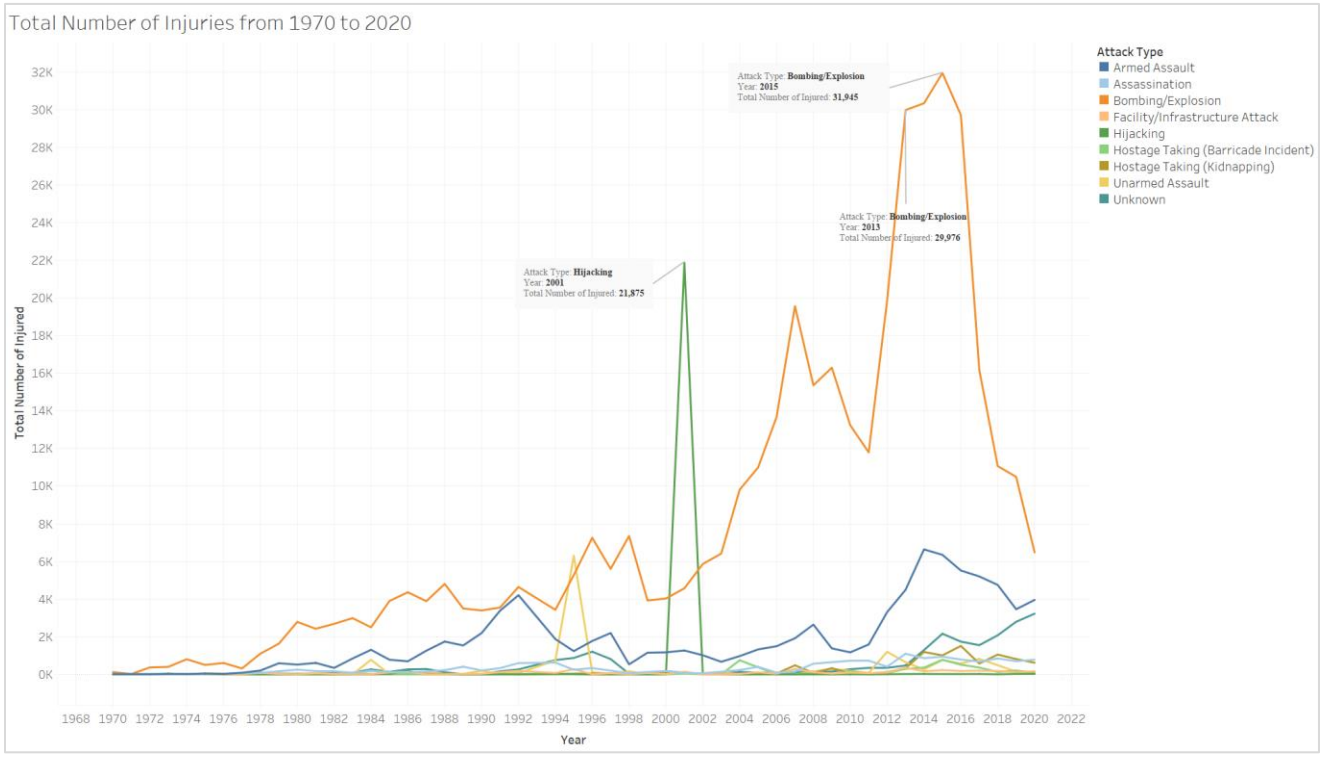
From the above chart it can be seen that Bombing/Explosion is the major cause of attacks occurring globally.



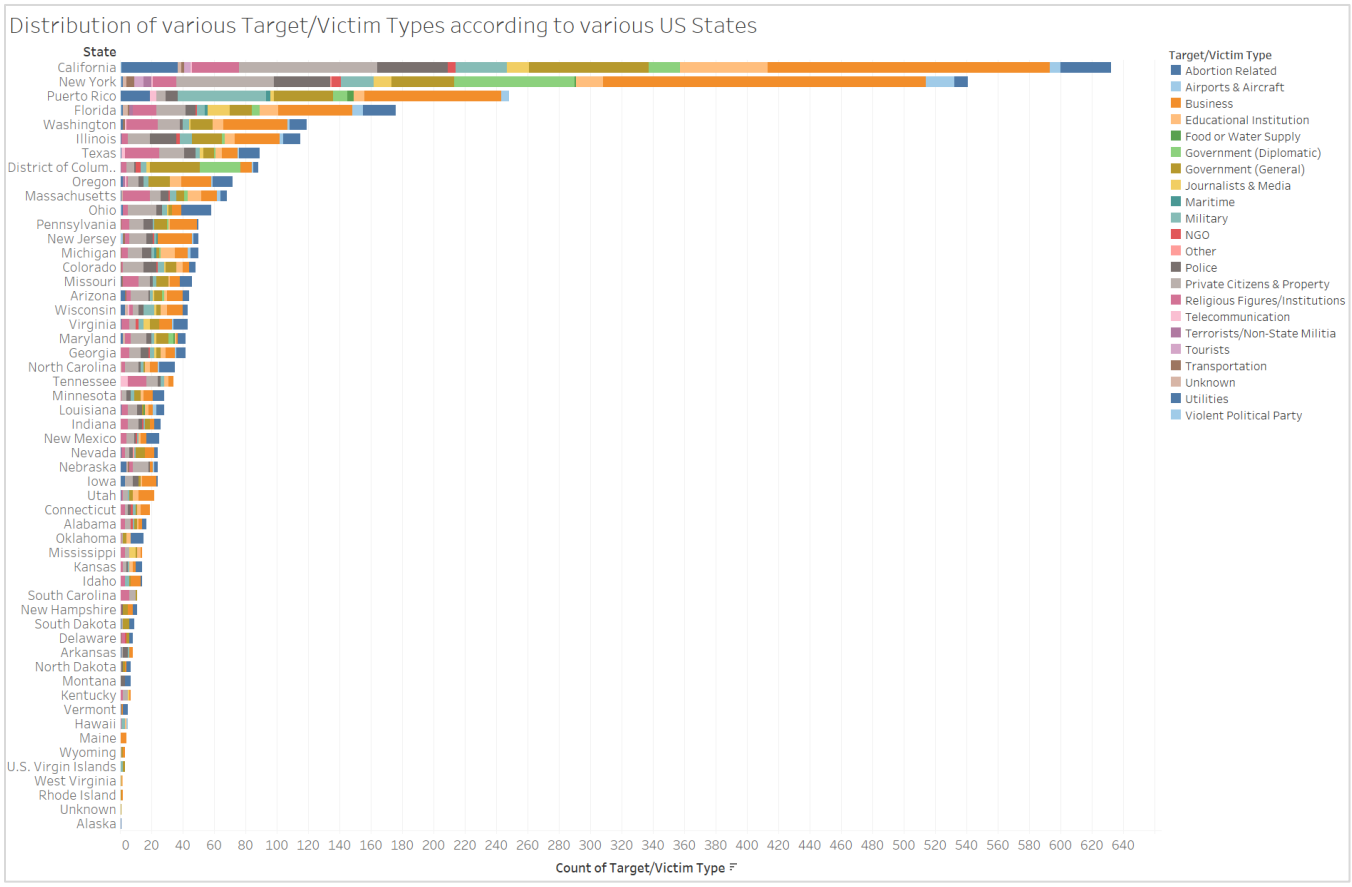
The chart above illustrates various attacks that have taken place in different regions of the world. Furthermore, the majority of attacks in the Middle East and North Africa were caused by bombing or explosions, whereas the majority of attacks in South Asia were caused by armed assault and bombing or explosions.



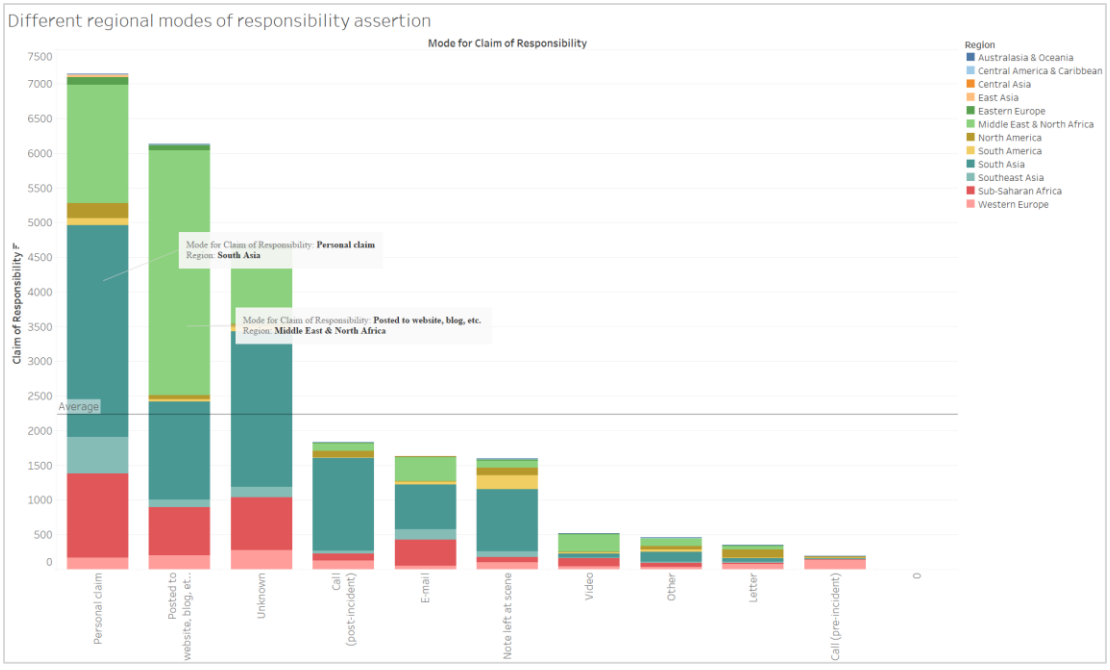
The above graph shows the total number of deaths occurred from 1970-2020 based on different attack types. The graph above shows a high number of fatalities occurring in the year 2015 due to Bombing/Explosion and on the other hand in the year 2014 too there's a high peak in the fatalities caused due to Armed assault. Furthermore, it can be also seen that after 2015 there's a drop in the fatalities number.



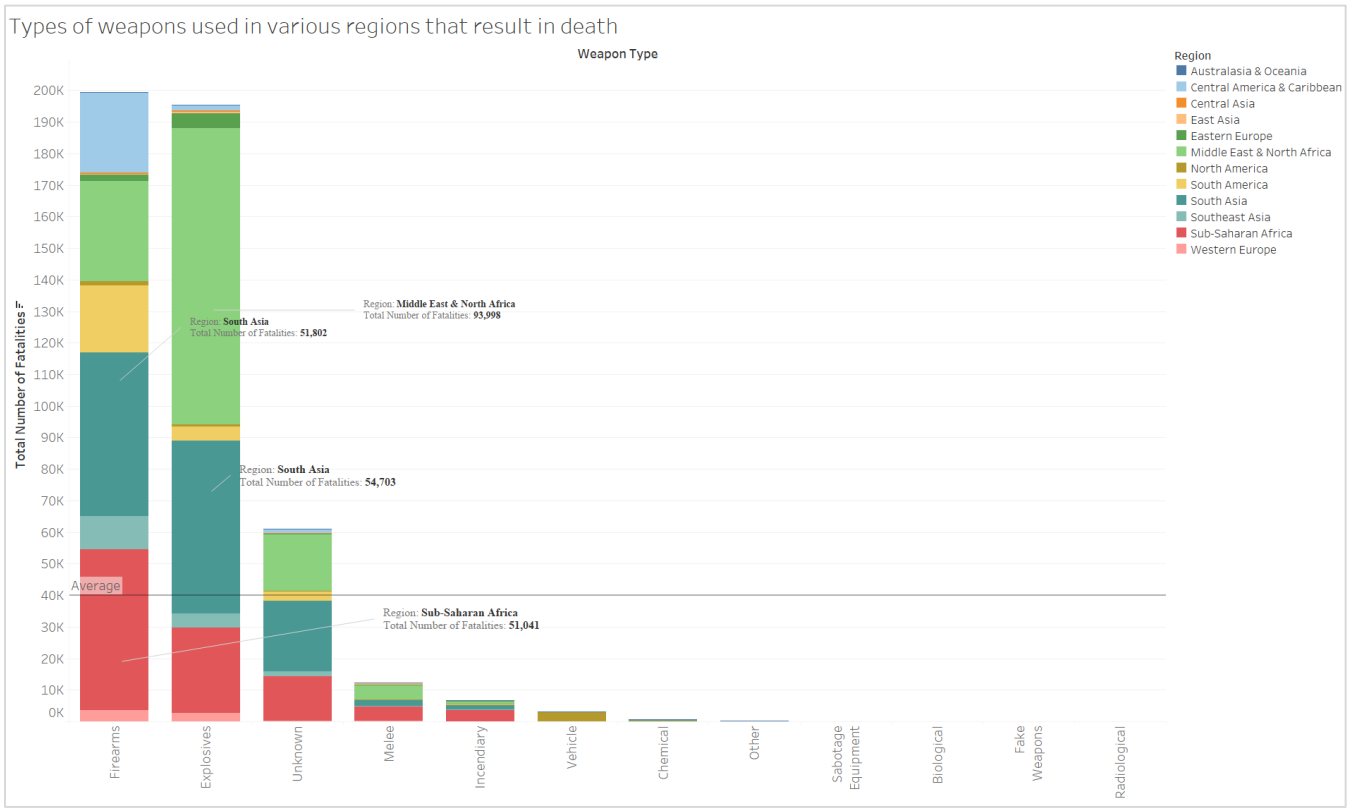
The above graph shows the total number of injuries occurred from 1970-2020 based on different attack types. The graph above shows a high number of injuries occurring in the year 2015 due to Bombing/Explosion and on the other hand in the year 2001 Hijacking was the major cause of injuries which injured about 21,875 people.



The distribution of distinct Target/Victim Types across different US states is seen in the stacked bar graph above. Additionally, it has been noted that California is the largest state with the greatest number of Target/Victim types.



The stacked bar graph above displays various regional claims of responsibility based on regions. From the graph above, it may be inferred that Personal claims are among the most common types of responsibility claims in South Asia, and posting to websites, blogs, and other online publications is among the most common types in the Middle East and North Africa.

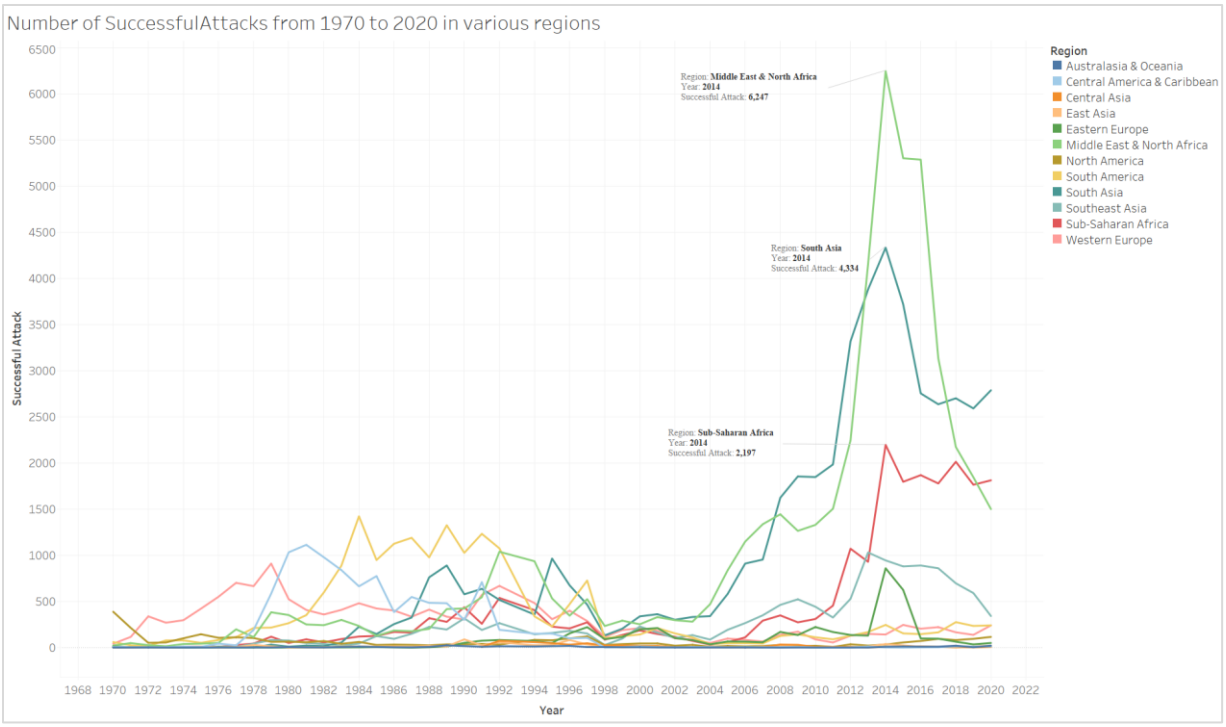


The above chart displays the types of weapons being used in various regions that result in death. From the graph above it can be seen that Explosives is one of the major weapons being used in Middle East and North Africa and another common type of weapon used is Firearms in South Asia and Sub-Saharan Africa region.

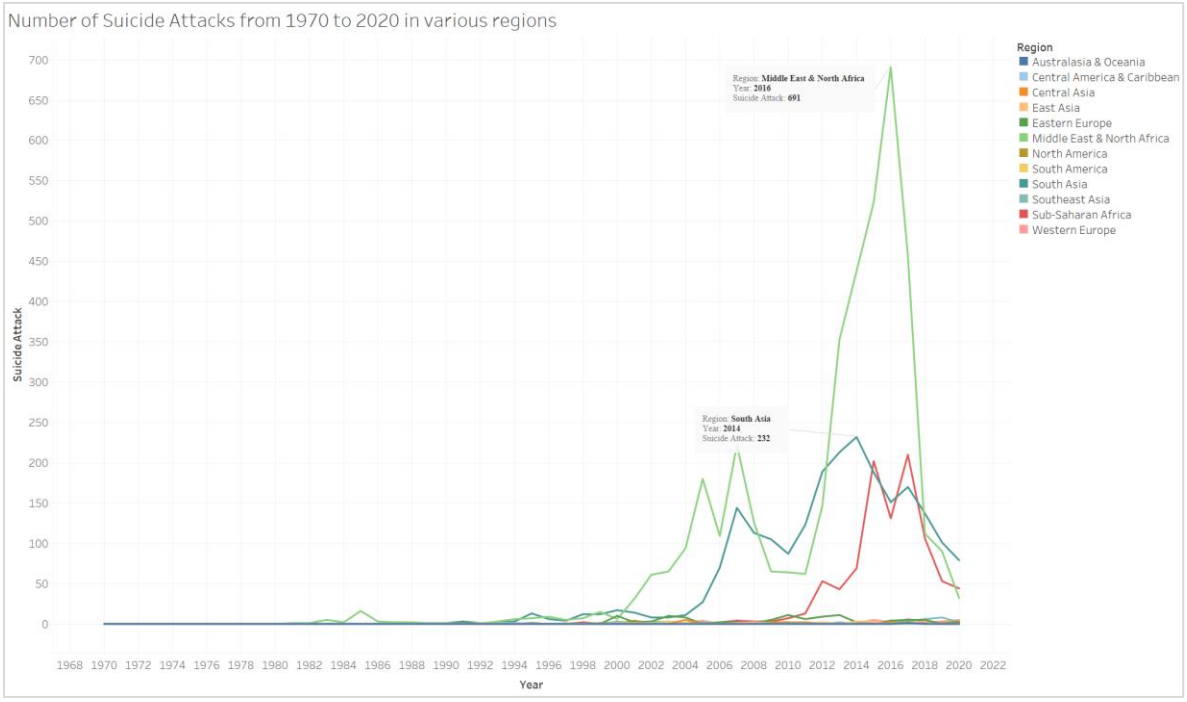
Damage done in different region and the cost of such property damage (in USD)

Region	Property Damage	Value of Property Damage (in USD)
South America	10,328	1,272,970,344
Western Europe	9,361	4,640,552,785
Central America ..	8,187	277,557,707
North America	1,315	1,070,669,428
East Asia	250	12,345,372
Australasia & Oc..	204	2,403,175
Central Asia	-34	69,194
Eastern Europe	-7,041	8,480,831
Southeast Asia	-9,932	79,646,860
Sub-Saharan Afri..	-23,382	69,580,510
South Asia	-51,648	482,925,265
Middle East & No..	-87,113	548,507,159

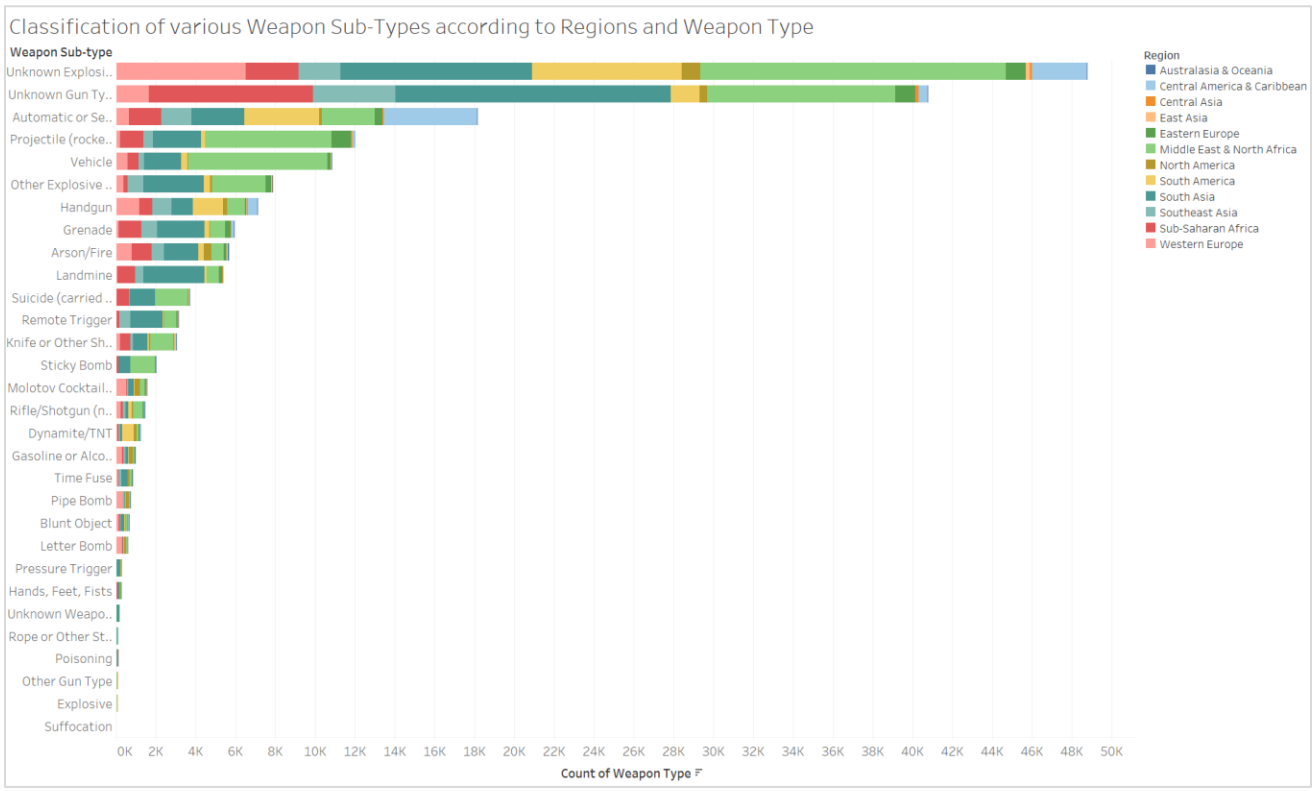
The tabular data above shows the quantity of property damaged throughout several regions, along with the cost associated with it.



The above chart shows the number of Successful Attacks being carried out from 1970-2020 in different regions. It can be seen that Middle East and North Africa region has the maximum number of successful attacks in the year 2014 which is about 6,247 attacks. Moreover, there were 4,334 successful attacks in South Asia region in the year 2014 and about 2,197 attacks in the region Sub-Saharan Africa in the year 2014.Furthemore, it can be concluded that most of the attacks which happened in the year 2014 were successful.



From the above graph it can be illustrated that maximum number of Suicide attack which happened in the Middle East and North Africa region was in the year 2016 which is about 691.Additonally it can be seen that South Asia also had lots of suicide attacks in the year 2014 which is about 232.



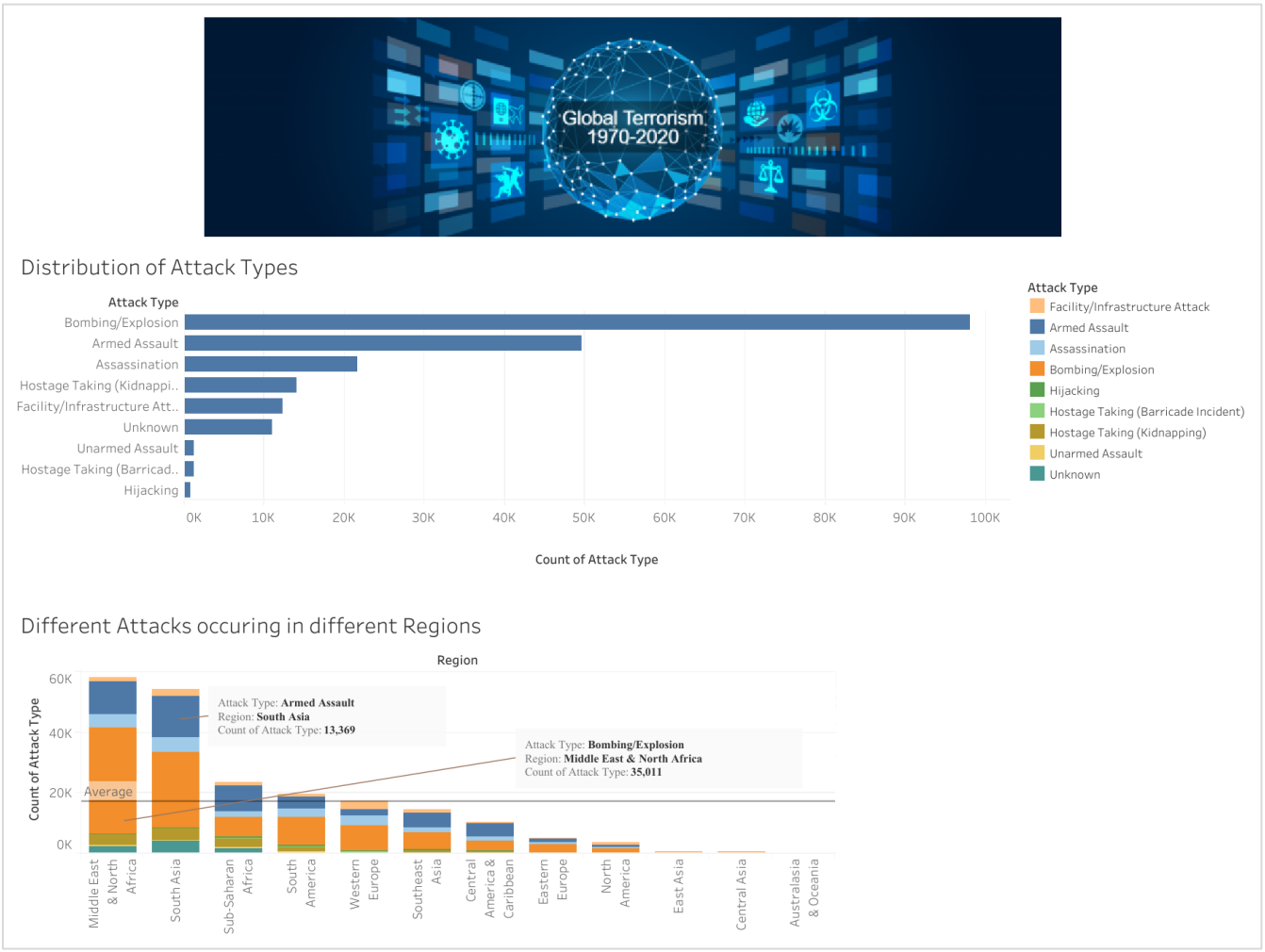
The above stack bar chart shows different sub-types of weapons been involved in global terrorism based on different regions and their weapon type.

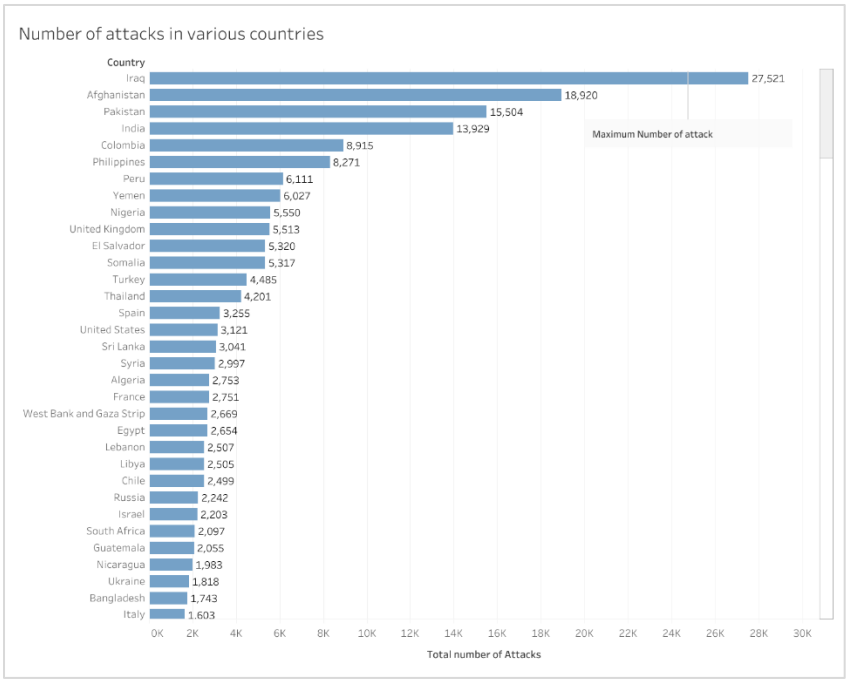
It can be seen from the above graph that there are some Unknown Explosion Type as a weapon under Explosion weapon type and its usage is majorly in the Middle East and North Africa region w.r.t other regions.

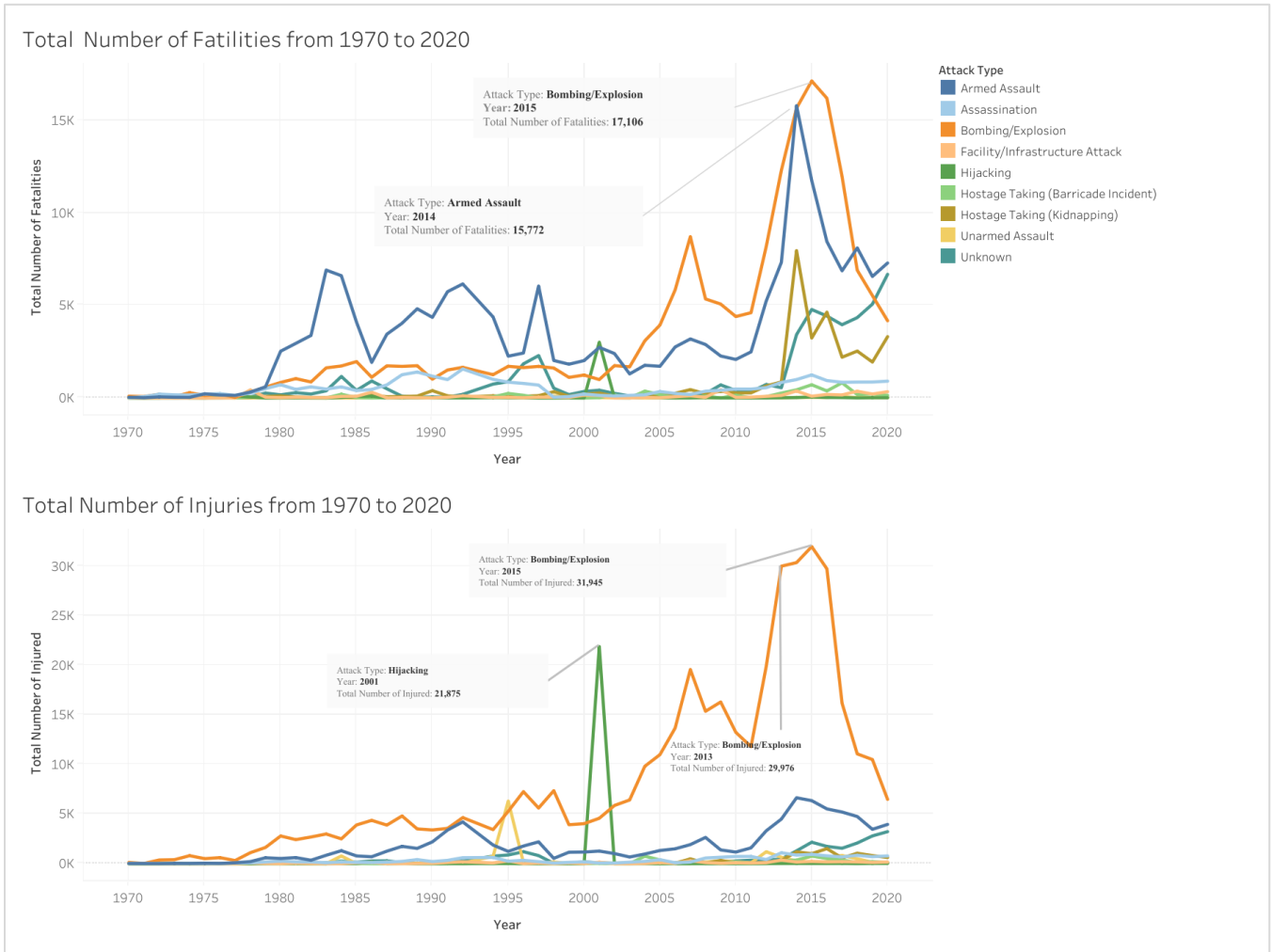
Designing and Publishing Tableau Dashboards:

A dashboard is a tool for condensing different kinds of visual data. Typically, a dashboard's purpose is to present various, linked facts in an easy-to-understand style. It is a collection of many views that enables you to compare a variety of data simultaneously.

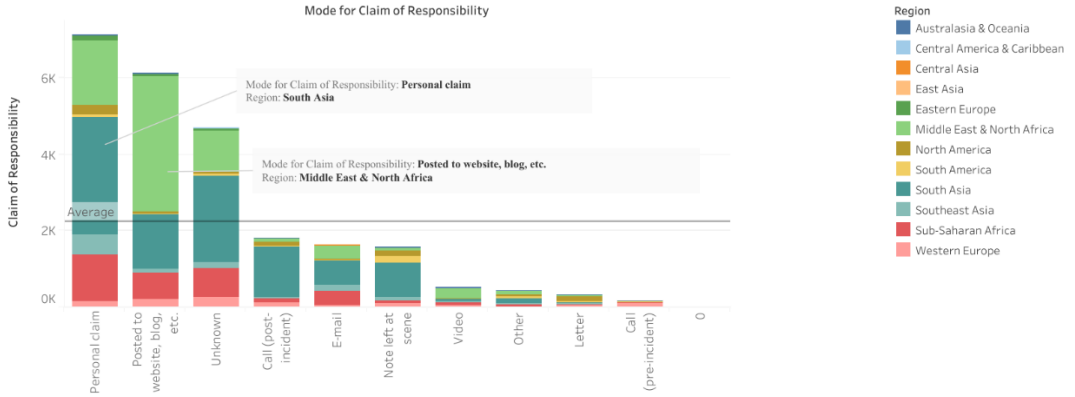
Below is the list of dashboards being designed for the project:




Number of attacks in various countries
Country
Iraq 27,521
Afghanistan 18,920
Pakistan 15,504
India 13,929
Colombia 8,915
Philippines 8,271
Peru 6,111
Yemen 6,027
Nigeria 5,550
United Kingdom 5,513
El Salvador 5,320
Somalia 5,317
Turkey 4,485
Thailand 4,201
Spain 3,255
United States 3,121
Sri Lanka 3,041
Syria 2,997
Algeria 2,753
France 2,751
West Bank and Gaza Strip 2,669
Egypt 2,654
Lebanon 2,507
Libya 2,505
Chile 2,499
Russia 2,242
Israel 2,203
South Africa 2,097
Guatemala 2,055
Nicaragua 1,983
Ukraine 1,818
Bangladesh 1,743
Italy 1,603
Maximum Number of attack
0K 2K 4K 6K 8K 10K 12K 14K 16K 18K 20K 22K 24K 26K 28K 30K
Total number of Attacks


Total Number of Fatilities from 1970 to 2020
Attack Type: Bombing/Explosion
Year: 2015
Total Number of Fatalities: 17,106
Attack Type: Armed Assault
Year: 2014
Total Number of Fatalities: 15,772
Total Number of Fatalities
15K
10K
5K
0K
1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020
Year
Attack Type
Armed Assault
Assassination
Bombing/Explosion
Facility/Infrastructure Attack
Hijacking
Hostage Taking (Barricade Incident)
Hostage Taking (Kidnapping)
Unarmed Assault
Unknown
Total Number of Injuries from 1970 to 2020
Attack Type: Bombing/Explosion
Year: 2015
Total Number of Injured: 31,945
Attack Type: Hijacking
Year: 2001
Total Number of Injured: 21,875
Attack Type: Bombing/Explosion
Year: 2013
Total Number of Injured: 29,976
Total Number of Injured
30K
25K
20K
15K
10K
5K
0K
1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020
Year

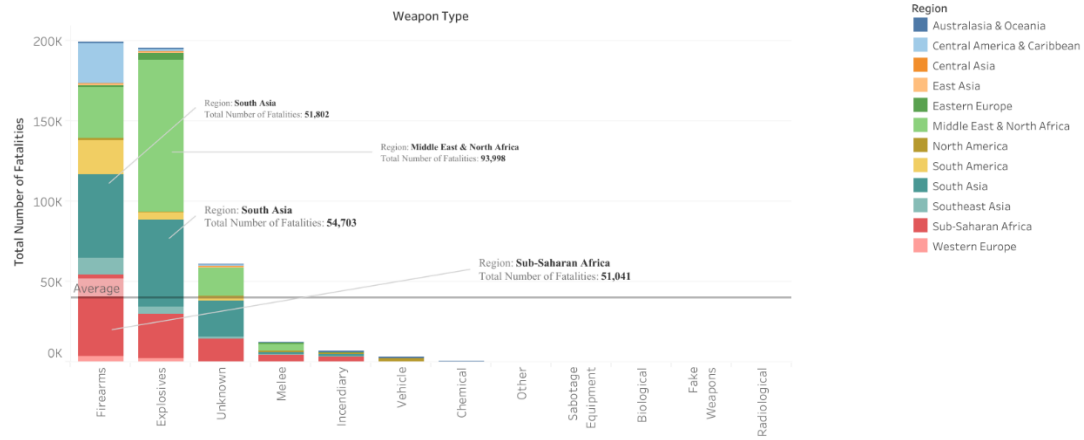
Different regional modes of responsibility assertion



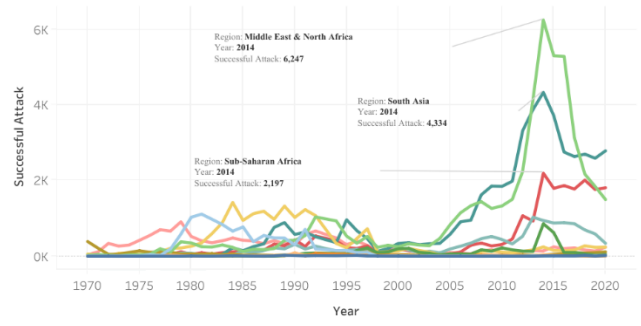
Damage done in different region and the cost of such property damage (in USD)

Region	Property Damage	Value of Property Damage (in USD)
South America	10,328	1,272,970,344
Western Europe	9,361	4,640,552,785
Central America & Caribbe..	8,187	277,557,707
North America	1,315	1,070,669,428
East Asia	250	12,345,372
Australasia & Oceania	204	2,403,175
Central Asia	-34	69,194
Eastern Europe	-7,041	8,480,831
Southeast Asia	-9,932	79,646,860
Sub-Saharan Africa	-23,382	69,580,510
South Asia	-51,648	482,925,265
Middle East & North Africa	-87,113	548,507,159

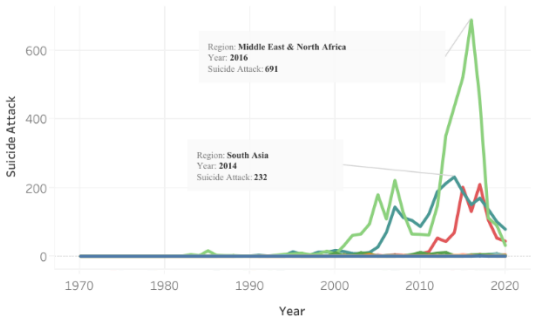
Types of weapons used in various regions that result in death

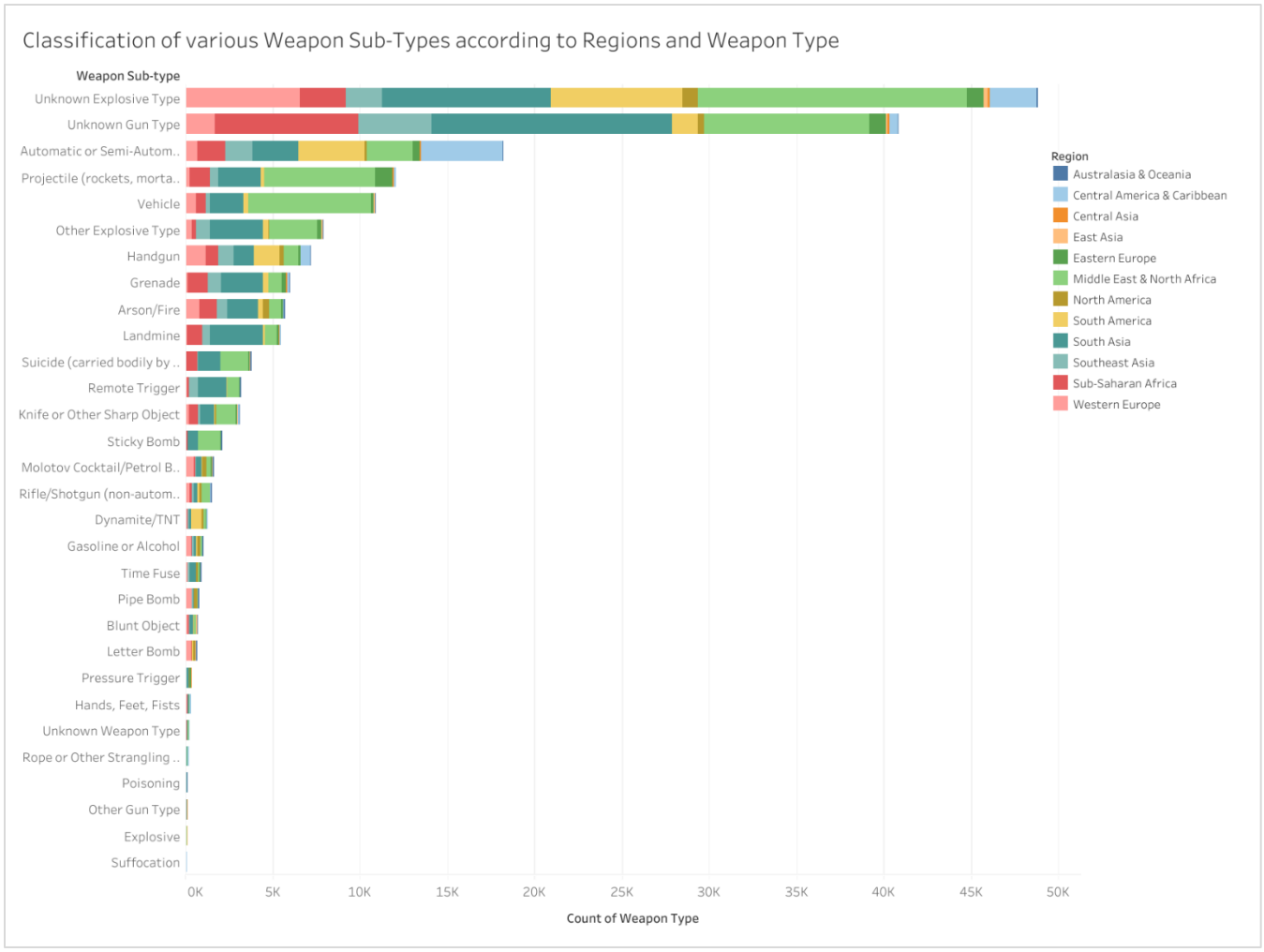


Number of SuccessfulAttacks from 1970 to 2020 in various regions



Number of Suicide Attacks from 1970 to 2020 in various regions





References:

https://www.start.umd.edu/gtd/

https://help.tableau.com/

https://www.tutorialspoint.com/tableau/index.htm

https://365datascience.com/