**A black and yellow logo

Description automatically generated with medium confidence**

**DSC651 (DATA REPRESENTATION & REPORTING TECHNIQUES)**

**Prepared by:**

|  |  |
| --- | --- |
| **Name** | **Student ID** |
| Mohamad Aizuddin bin Mohamad Rafee | 2019542553 |
| Zikri Zuhairi bin Shamsulbahri | 2019328725 |
| Muhammad Zuhaily bin Ramzi | 2019705303 |
| Mohd Shamil bin Mohd Shukri | 2019707617 |
| Farahiyah Najla binti Azhari | 2019705175 |

**Group Name:** Sleeping AI

**Prepared for:** SIR MOHAMMAD BAKRI CHE HARON

**Submission Date:** 28 January 2021

**1. Problem Statement**

* Difficulty of identifying terrorist group between various countries.
* Difficulty of identifying the type of terrorist assaults in the most affected country.
* Difficulty of predicting the motive of terrorist attacks.
* Difficulty of discovering the type of victims in each country.
* Difficulty of identifying the number of successful attacks in each country.
* Difficulty of discovering the number of fatalities throughout the year.

**2. Dataset Information**

|  |  |  |
| --- | --- | --- |
| **Columns** | **DataType** | **Description** |
| year | date | List of records from the year 1970 until 2017. |
| month | date | Month. |
| country | string | Country that has been attacked by terrorists. |
| region | string | Region that has been attacked by terrorists. |
| provstate | string | The country’s state. |
| city | string | The country’s city. |
| latitude | decimal | Latitude location of the attacked. |
| longitude | decimal | Longitude location of the attack. |
| location | string | Location name of the attacked. |
| success | boolean | Success attack is determined with value 1-success, 0-unsuccessful. |
| suicide | boolean | Suicide attack is determined with value 1-suicide, 0- no suicide. |
| attactype | integer | The type of attack that the terrorist used. |
| attacktype\_txt | string | The type of attack that the terrorist used. |
| target\_type | integer | The type of victim targeted by terrorists. For example, as the military, police, and etc. |
| target\_type\_txt | string | Type of victim. |
| nationality | string | The nationality of the victim. |
| gname | string | The terrorist group name. |
| motive | string | Motive of the attack. |
| weapon\_type | string | The type of weapon used for the attack. |
| n\_kill | integer | The number of casualties or fatalities for an attack. |

**3. Objective**

* To identify the terrorist group between various countries
* To identify the type of terrorist assaults in the most affected country
* To predict the motive of terrorist attacks
* To discover the type of victims in each country
* To identify the number of successful attacks in each country
* To discover the number of fatalities throughout the year.

**4. a. Data pre-processing**

Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format. Data preprocessing is divided into four stages: data cleaning, data integration, data reduction, and data transformation. For this project, data cleaning, data reduction, and data transformation were implemented.

**Data cleaning:**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. The table below shows the attributes that have been cleaned.

|  |  |
| --- | --- |
| **Attributes** | **Description** |
| year | Exclude “null” value |
| month | Exclude “null” value |
| success | Exclude “null” value |
| latitude | Exclude “null” value |
| longitude | Exclude “null” value |
| nationality | Exclude “null” value |
| weapon\_type | Exclude “null” value |
| nkill | Exclude “null” value |

**Data reduction:**

Data reduction is a process that reduces the volume of original data and represents it in a much smaller volume. Data reduction techniques ensure the integrity of data while reducing the data. After deciding the questions there are several attributes that are not needed in the dataset.The table below shows the attributes that have been removed.

|  |  |
| --- | --- |
| **Attributes** | **Description** |
| location | Remove Field |
| city | Remove Field |
| provstate | Remove Field |

**Data transformation:**

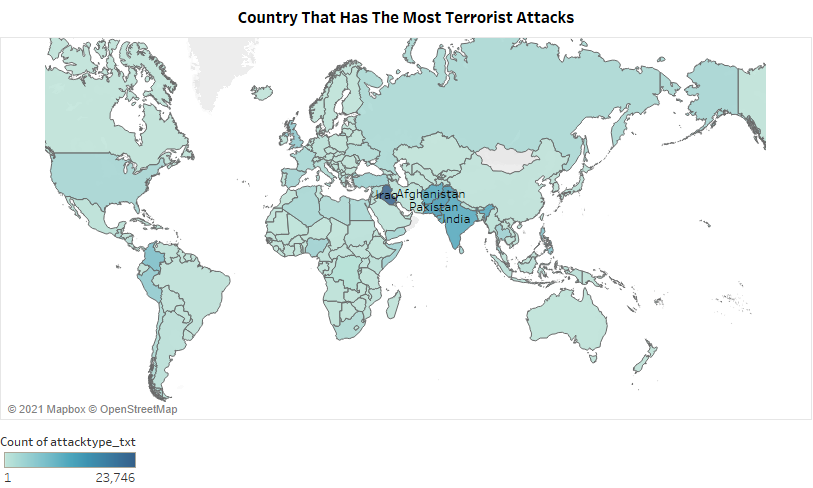
Data transformation is the process of changing the format, structure, or values of data. Because the data contains too many null values in one field. It is not the best choice to remove all the null values to perform data analytic. It will lose too much data. The table below shows the attribute that has been replaced with its null value.

|  |  |
| --- | --- |
| **Attributes** | **Description** |
| motive | Null value replaced with “unknown” |

After implementing all the phases above, the dataset is ready for data analytic processes.

**4. b. Data Visualizations**

1.0) Which country has the most terrorist attacks?

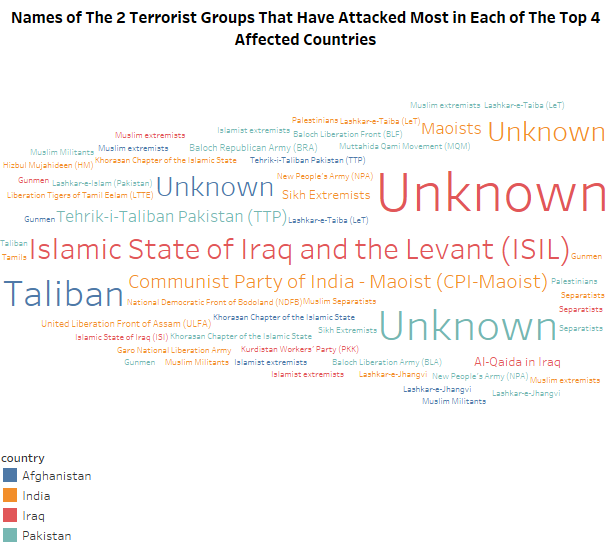


Based on the choropleth map above, the country that has the most terrorist attack is Iraq. Iraq has 23746 attacks. This is probably due to the fact that Iraq was a state sponsor of terrorism under the control of the deceased Saddam Hussein. Saddam Hussein was an Iraqi politician who served as the fifth President of Iraq from 16 July 1979 until 9 April 2003. Under Saddam, Iraq was a major state sponsor of international terrorism. Saddam used foreign terrorist groups as an instrument of foreign policy. The terrorist groups hosted by Saddam were denied protection if he wanted to improve relations with a neighboring country and encouraged to attack those Saddam wanted to pressure.

This type of visualization is used to provide an easy way to visualize how a variable varies across a geographic area or show the level of variability within a region. In this case, the above visualization shows the country that used longitude and latitude which are necessary to the generate the geographical area.

This visualization is achieved by using longitude, latitude, attacktype\_txt and country attributes. The country attribute is filtered with the top 4 countries. The intensity of the colour shows the count of attacktype\_txt. The details about country are shown in the choropleth map while the marks are labelled by country. The longitude and latitude attributes are involved in the data pre-processing which is data cleaning as it contains missing values.

1.1) What are the names of the 2 terrorist groups that have attacked most in each of the top 4 affected countries?

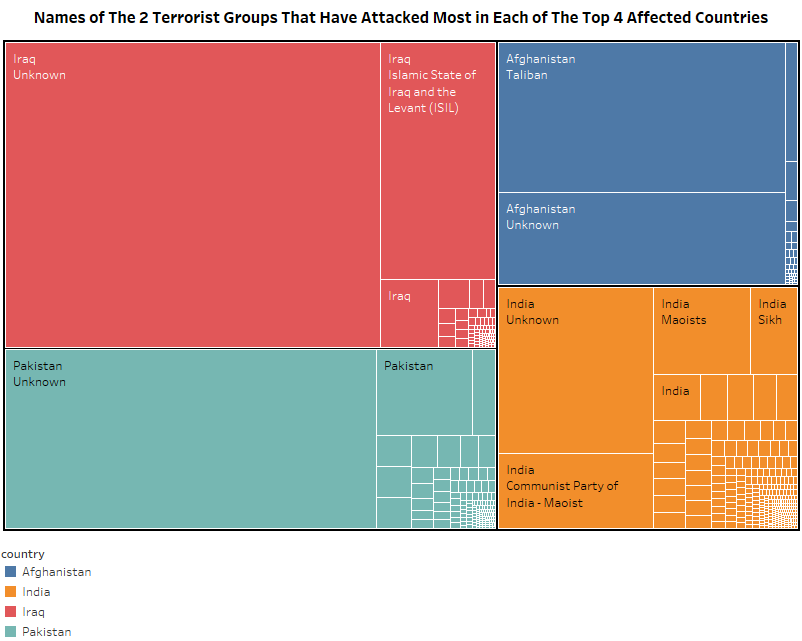


The text data visualization above shows the names of the terrorist groups that have attacked most in each of the top 4 affected countries. From the highest affected country, the top 2 terrorist groups are the unknown and Islamic State of Iraq and the Levant (ISIL) group have attacked Iraq, the unknown and Tehrik-i-Taliban Pakistan (TTP) have attacked Pakistan, the Taliban and unknown groups have attacked Afghanistan, and the unknown and Communist Party of India-Maoist (CPI-Maoist) groups have attacked India. The same terrorist organization which is the unknown group often chooses the same area where terrorist incidents occur.

Iraq is the most attacked by the terrorists is probably because of the country’s mountainous and desert areas. ISIL has 4335 terrorist groups that have attacked Iraq. In the case of TTP, the action of entering or gaining access to an organization or place secretively has both allowed the TTP to launch terrorist attacks on targets in Pakistan's cities and to engage in armed crime across Pakistan. In India, CPI-Maoist group have been at war with the government of India. Therefore, CPI-Maoist group has frequently launched attacks against their political and economic enemies.

This type of visualization is used in this question to visualize numerical and categorical features of data. The numerical value in this data is represented by the count of terrorist groups, while the categorical value in the data is represented by the countries.

This visualization is achieved by using country and gname attributes. The country attribute is filtered with the top 4 countries. The intensity of colour shows the details about country while the size of the text shows the count of gname.

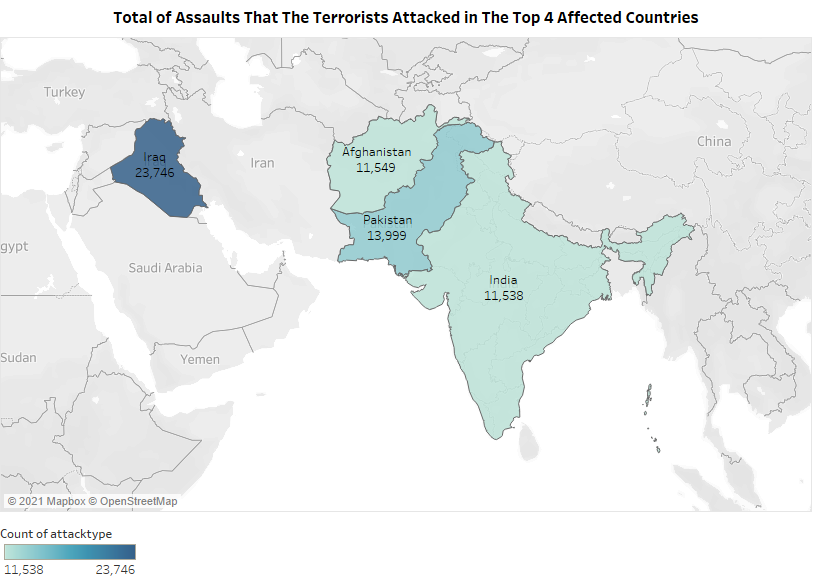


The treemap visualization above shows the names of the terrorist groups that have attacked most in each of the top 4 affected countries. From the highest affected country, the top 2 terrorist groups are the unknown and Islamic State of Iraq and the Levant (ISIL) group have attacked Iraq, the unknown and Tehrik-i-Taliban Pakistan (TTP) have attacked Pakistan, the Taliban and unknown groups have attacked Afghanistan, and the unknown and Communist Party of India-Maoist (CPI-Maoist) groups have attacked India.

This type of visualization, which is treemap, is used to visualize part-to-whole relationship between a large number of categories. It can easily distinguish between categories and data values with just one glance. The categorized data in this visualization is country, which is classified by colours. Meanwhile, the size of the data values is shown by the terrorist group.

The treemap visualization above is achieved by using country and gname attributes. The country attribute is filtered with the top 4 countries. The details about the countries are shown based on the intensity of colours. The size of each treemap is shown by the count of gname.

1.2) How many assaults did the terrorists attack in the top 4 affected countries?

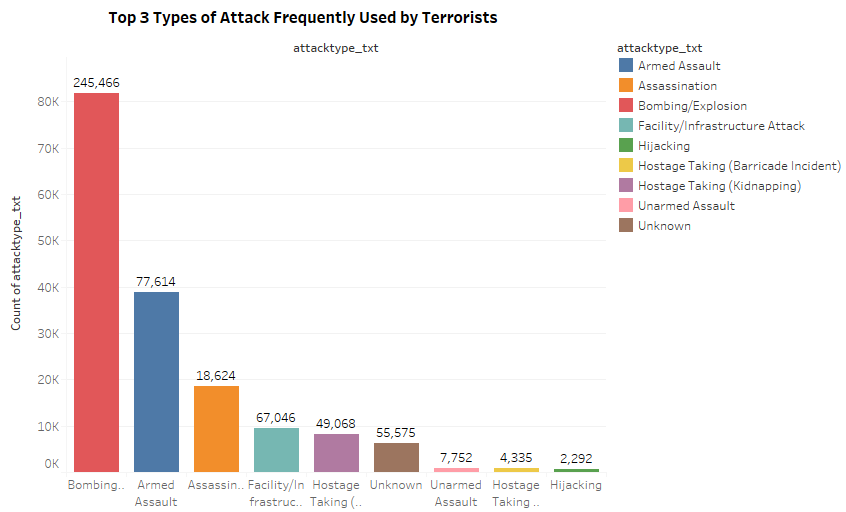


The choropleth map above shows the total of assaults that the terrorists attack in the top 4 affected countries. From the highest affected country, Iraq has 23746 terrorist attacks, Pakistan has 13999 attacks, Afghanistan has 11549 attacks and India has 11538 attacks. Iraq has the most terrorist attacks is probably due to the fact that Iraq was a state sponsor of terrorism under the control of the deceased Saddam Hussein. Saddam Hussein was an Iraqi politician who served as the fifth President of Iraq from 16 July 1979 until 9 April 2003. Under Saddam, Iraq was a major state sponsor of international terrorism. Saddam used foreign terrorist groups as an instrument of foreign policy. The terrorist groups hosted by Saddam were denied protection if he wanted to improve relations with a neighboring country and encouraged to attack those Saddam wanted to pressure.

This type of visualization is used to provide an easy way to visualize how a variable varies across a geographic area or show the level of variability within a region. In this case, the above visualization shows the country that used longitude and latitude which are necessary to the generate the geographical area.

This visualization is achieved by using longitude, latitude, attacktype\_txt and country attributes. The country attribute is filtered with the top 4 countries. The intensity of the colour shows the count of attacktype\_txt. The details about country are shown in the choropleth map while the marks are labelled by country. The longitude and latitude attributes are involved in the data pre-processing which is data cleaning as it contains missing values.

2.0) What are the top 3 types of attack that the terrorists have frequently used?

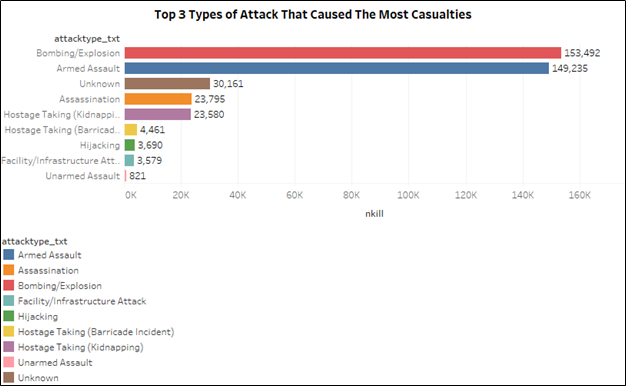


The bar chart above shows the top 3 types of attack that the terrorists have frequently used. From the highest type of attacks, the terrorists used 245466 bombings or explosions, 77614 armed assaults, and 18624 assassinations.

This type of visualization is used to display and compare the number, frequency or other measure for different discrete categories of data. Bar chart is also simple to create and very easy to interpret.

This visualization is achieved by using attacktype\_txt attribute. The type of colours is differentiated to show each attacktype\_txt. The marks at the top of each bar chart are labeled by sum of attacktype.

2.1) What are the top 3 types of attack that caused the most casualties?



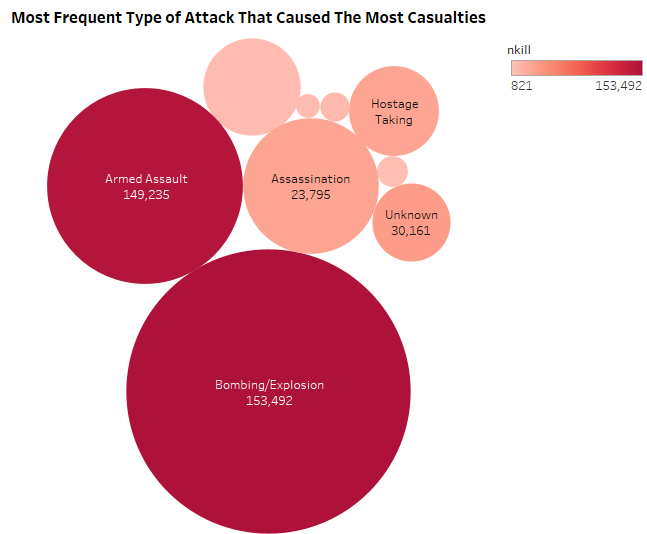
The top three types of attack that caused the most casualties are bombing/explosion, armed assault and unknown attacks. Different types of attacks will directly affect the degree of harm of the incident. Bombing or explosion, armed assault and assassination are generally political and can cause bad effects. Bombing or explosion and armed assault and explosion always result in casualties. These methods are more harmful. Unarmed attacks and other means involve small casualties, resulting in a relatively low degree of harm. Therefore, the type of attack that have used by the terrorists is the important criteria for the terrorists to evaluate the degree of damage to launch attacks.

This type of visualization is used in this question to compare items between different groups. This bar chart compares the data among the top three types of attack that caused the most casualties.

This visualization is achieved by using each nkill, gname and year attributes. The type of colours is differentiated to distinguish between each gname. The marks on each bar chart are labeled by the sum of nkill. The year is filtered by using the 2014 year only. The nkill and year attributes are involved in the data pre-processing which is data cleaning as they contain missing values.

This visualization is achieved by using attacktype\_txt and nkill attributes. The type of colours is differentiated to distinguish between each attacktype\_txt. The marks on each bar chart are labeled by the sum of nkill. The year is filtered by using the top three types of attack. The nkill attribute is involved in the data pre-processing which is data cleaning to remove the missing values.

2.2) Did the most frequent type of attack cause the most casualties?



Based on the bubble chart above, the most frequent type of attack, which is bombing or explosion attack has caused the most casualties. The number of casualties achieved by this type of attack is 153492. Bombing or explosion attacks are used by the terrorist is likely due to the effect of explosion that can be resulted in immediate structural collapse. The structural damage can be also occurred within a confined space or occurred in open air. As a consequence, most terrorists would choose bombing or explosion attacks as their preference to attack the victims.

This type of visualization is used to compare and show relationships between numerical and categorical variables. Different bubble sizes are useful to visually emphasize specific values in this question that specifies the measurement of number of casualties or sum of kills.

This visualization is achieved by using attacktype\_txt and nkill attributes. The intensity of the colours is shown to measure the sum of nkill. The size of each bubble chart shows attacktype\_txt. The marks in the bubble chart are also labeled with attacktype\_txt. The nkill attribute is involved in the data pre-processing which is data cleaning to remove the missing values.

3.0) What are the top 10 motives of terrorist attacks?

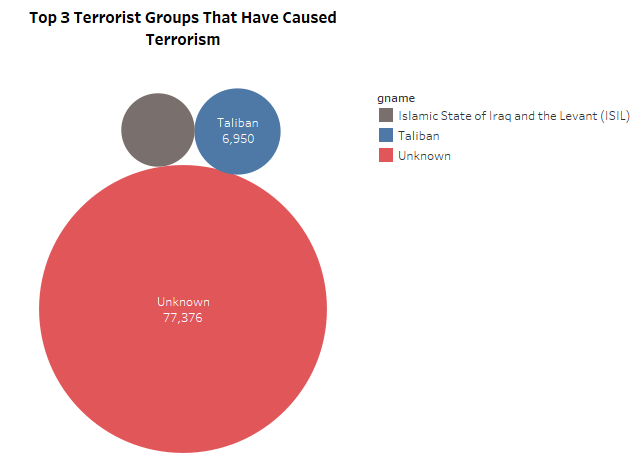


Based on the text data visualization above, the top 10 motives of terrorist attacks are mostly unknown. However, some of the unknown motives include other motives, such as the attack may have been part of a larger trend of violence related to Bangladesh's nationwide hartal which began on 6 January 2015 and sources suspected that the attack, which targeted members of the Sunni community, may have been part of a larger trend of sectarian violence between Iraq's minority Sunni and majority Shiite communities. Meanwhile, the specified motives are part of a campaign by Islamic extremists to destabilize Algeria by weakening security forces protecting the 'apostate' Algerian government, for political purpose, the attack was carried out because the victim was accused of being a police informer. The motive of terrorists, which are for political purposes, are likely to be based on terrorist actions that are generally carried out in a manner that achieves maximum publicity.

This type of visualization is used in this question to summarize large amounts of text. The text visualization also automatically highlights the key terms in a series of texts, and categorize text by topic, sentiment and saving hours of reading time. The motives of the terrorist attacks are varied. Therefore, the text data visualization

This visualization is achieved by using attacktype\_txt attribute. The type of colours is differentiated to show each attacktype\_txt. The marks at the top of each bar chart are labeled by sum of attacktype. The motive attribute is involved in the data pre-processing which is data transformation as it contains null values. Therefore, the null values are replaced with “unknown”.

3.1) What are the top 3 terrorist groups that have caused terrorism?



Based on the bubble chart above, the top 3 terrorist groups that have caused terrorism are unknown, Taliban and u Islamic State of Iraq and the Levant (ISIL) groups. The Taliban who refers to themselves as the Islamic Emirate of Afghanistan (IEA), are a Sunni Islamic fundamentalist political movement and military organization in Afghanistan currently waging war within that country. The resurgence of the Taliban in a war that has increased in Afghanistan up until now.

This type of visualization is used in this question to compare and show relationships between numerical and categorical variables. Different bubble sizes are useful to visually emphasize specific values.

This visualization is achieved by using gname and count of gname attribute. The type of colours is differentiated to shows the details about gname. The size of each bubble chart shows count of gname. The marks in the bubble chart are labeled by gname and count of gname.

3.2) What are the top 5 motives of the top 3 terrorist groups?

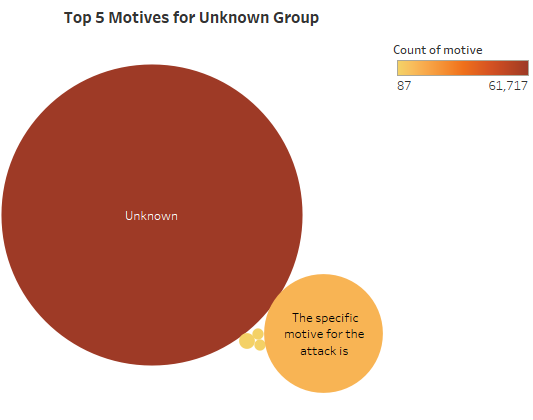


The text data visualization above shows the top 5 motives of the top 3 terrorist groups. However, some of the unknown motives include other motives, such as the attack may have been part of a larger trend of violence related to Bangladesh's nationwide hartal which began on 6 January 2015 and the source was not reported.

This type of visualization is used in this question to summarize large amounts of text. The text visualization also automatically highlights the key terms in a series of texts, and categorize text by topic, sentiment and saving hours of reading time. The motives of the terrorist attacks are varied. Therefore, the text visualization makes the text data easy to understand as the human brain loves visual data.

This visualization is achieved by using gname and motive attributes. The type of colours is differentiated to represent each gname. The size of the text shows the measurement of count of motive. The motive attribute is involved in the data pre-processing which is data transformation as it contains null values. Therefore, the null values are replaced with “unknown”.

Unknown group:

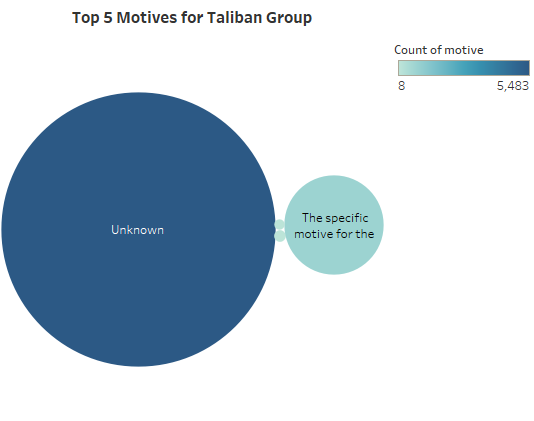


The bubble chart visualization above shows the top 5 motives for unknown group. The motives for this data are quite vague as the data mostly consist of unknown motives.

This type of visualization is used in this question to compare and show relationships between numerical and categorical variables. Different bubble sizes are useful to visually emphasize specific values.

This visualization is achieved by using gname, motive, and count of gname attributes. The intensity of the colours is used to represent the measurement of count of motive. The size of the bubble chart shows the measurement of count of motive. The motive attribute is involved in the data pre-processing which is data transformation as it contains null values. Therefore, the null values are replaced with “unknown”.

Taliban group:

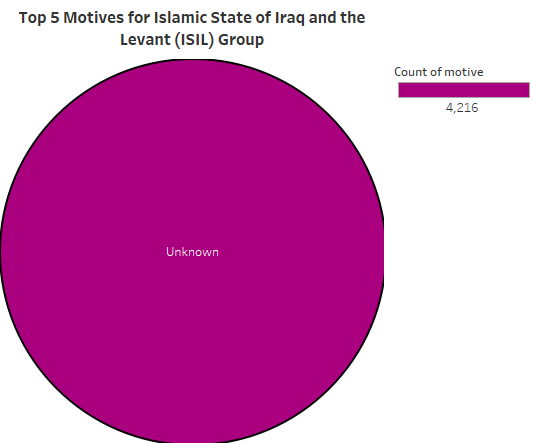


The bubble chart visualization above shows the top 5 motives for the Taliban group. The motives for this data are quite vague as the data mostly consist of unknown motives.

This type of visualization is used in this question to compare and show relationships between numerical and categorical variables. Different bubble sizes are useful to visually emphasize specific values.

This visualization is achieved by using gname and motive attributes. The intensity of the colours is used to represent the measurement of count of motive. The size of the bubble chart shows the measurement of count of motive. The motive attribute is involved in the data pre-processing which is data transformation as it contains null values. Therefore, the null values are replaced with “unknown”.

Islamic State of Iraq and the Levant (ISIL) group:

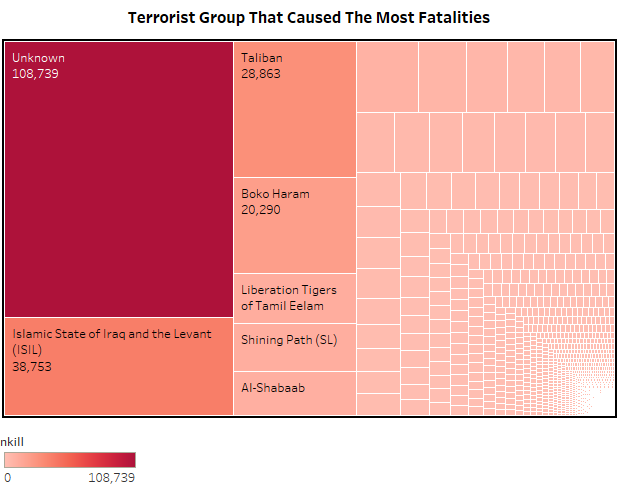


The bubble chart visualization above shows the top 5 motives for the Islamic State of Iraq and the Levant (ISIL) group. The motives for this data are quite vague as the data mostly consist of unknown motives.

This type of visualization is used in this question to compare and show relationships between numerical and categorical variables. Different bubble sizes are useful to visually emphasize specific values.

This visualization is achieved by using gname and motive attributes. The intensity of the colours is used to represent the measurement of count of motive. The size of the bubble chart shows the measurement of count of motive. The motive attribute is involved in the data pre-processing which is data transformation as it contains null values. Therefore, the null values are replaced with “unknown”.

4.0) Which terrorist group caused the most fatalities?

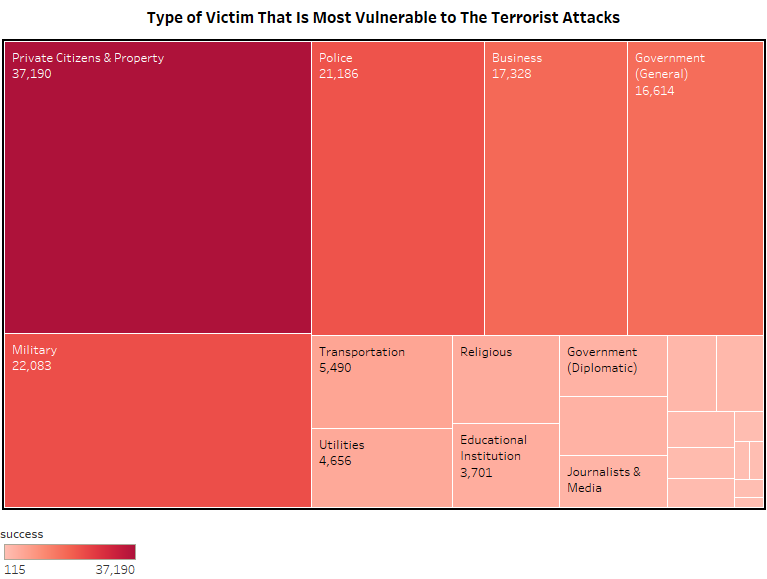


Based on the treemap visualization above, the terrorist group that caused the most fatalities is unknown group with 108739 fatalities.

In this question, this type of visualization is used to visualize the hierarchical structure of the tree diagram while also displaying the quantity by area size for each category. A rectangle area with its sub-category rectangles nested within it is assigned to each category. Treemap is also very good at comparing the proportions between categories by their size of area.

This visualization is achieved by using gname and nkill attributes. The intensity of the colours is used to represent the measurement of sum of nkill. The size of each treemap shows the measurement of sum of nkill. The marks in each treemap rectangle are labeled by gname and sum of nkill. The nkill attribute is involved in the data pre-processing which is data cleaning as it contains missing values.

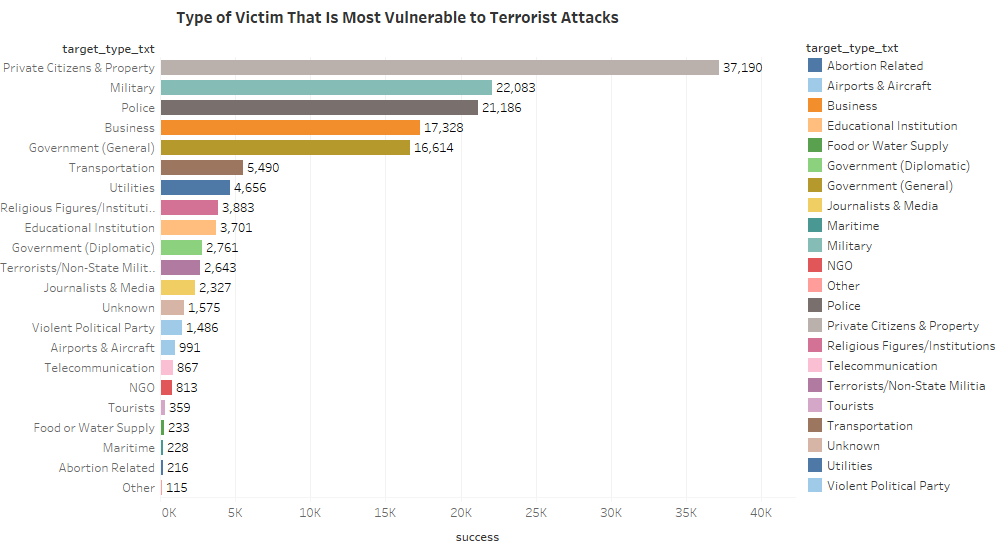
4.1) What type of victim is most vulnerable to the terrorist attacks?



The treemap above highlights that private citizens & property is overwhelmingly most vulnerable to the terrorist attacks. The number of successful attacks by terrorists to the private citizens & property is 37190. The terrorists attacked most towards private citizens and property is most likely because they want to reveal private information about the organization. Moreover, the type of victims that are attacked by the terrorists are followed by the military and the police with 22083 and 21186 successful attacks respectively. Polices are started to being targeted by the terrorists since the beginning of the war in Iraq.

This type of visualization is used in this question to visualize the hierarchical structure of a tree diagram while also displaying quantities for each category via area size. Each category is assigned a rectangle area with their subcategory rectangles nested inside of it. Treemap is also great at comparing the proportions between categories through their size of area.

This visualization is achieved by using target\_type\_text and success attributes. The intensity of colour shows the value of the sum of success while the marks are labeled by target\_type\_txt and sum of success. The success attribute is involved in the data pre-processing which is data cleaning as it contains missing values.

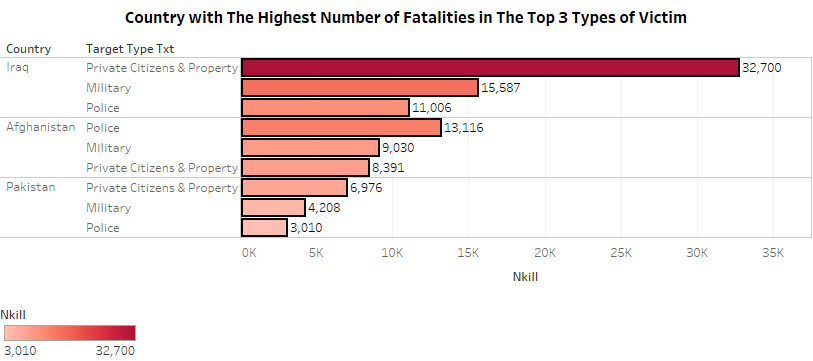


Based on the bar chart above, it is shown that type of victim is most vulnerable to the terrorist attacks which is 37190 successful attacks. The terrorists attacked most towards private citizens and property is most likely because they want to reveal private information about the organization. Furthermore, the type of victims that are attacked by the terrorists are followed by the military and the police with 22083 and 21186 attacks respectively.

This type of visualization is used in this question to compare items between different groups. This bar chart shows a comparison of numbers of the type of victim that is most vulnerable to terrorist attacks.

This visualization is achieved by using target\_type\_text and success attributes. The intensity of colour shows the value of the sum of success while the marks are labeled by target\_type\_txt and sum of success. The success attribute is involved in the data pre-processing which is data cleaning as it contains missing values.

4.2) What is the country with the highest number of fatalities in the top 3 types of victim?



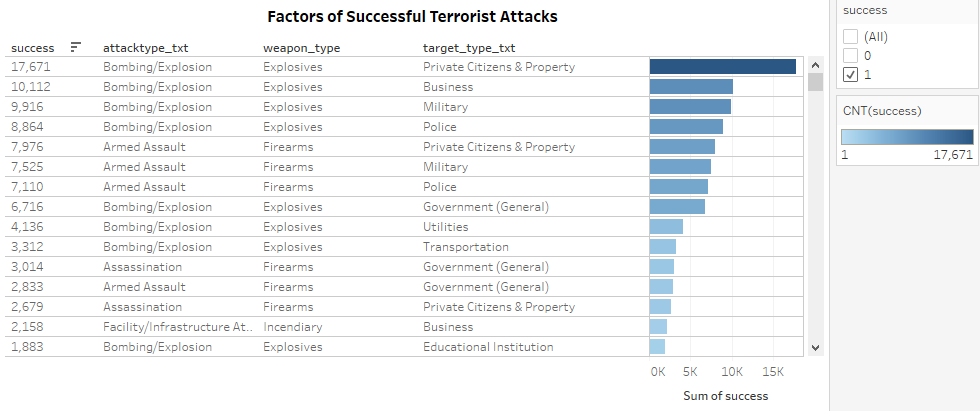
Based on the bar chart above, Iraq has the highest number of fatalities from private citizens and property, and military with the fatalities of 32700 and 15587 respectively. On the other hand, Afghanistan has the highest number of police target type which is 13116 fatalities among them.

Following military invasions in both Afghanistan and Iraq respectively, the instability brought about by conflict allowed an opportunity for a number of terrorist groups to gain a strong position and secure in the region. The vast majority of attacks occur in the countries where the major terrorist organizations such as Al Qaeda and Islamic State are based. The lack of security and order present in places like Iraq and Afghanistan affords more opportunities to terrorist groups to commit more acts of terrorism.

This type of visualization is used in this question to compare items between different groups. This bar chart shows a comparison of numbers of the highest number of fatalities in the top 3 types of victim.

This visualization is achieved by using target\_type\_text, country and nkill attributes. The intensity of colour shows the value of the sum of nkill while the marks are labeled by the sum of nkill. The country is filtered with the top 3 countries that have most fatalities. The target\_type\_text is also filtered with the top 3 target\_type\_text that are resulting into fatalities. The nkill attribute is involved in the data pre-processing which is data cleaning as it contains missing values.

5.0) What are the factors of successful terrorist attacks?

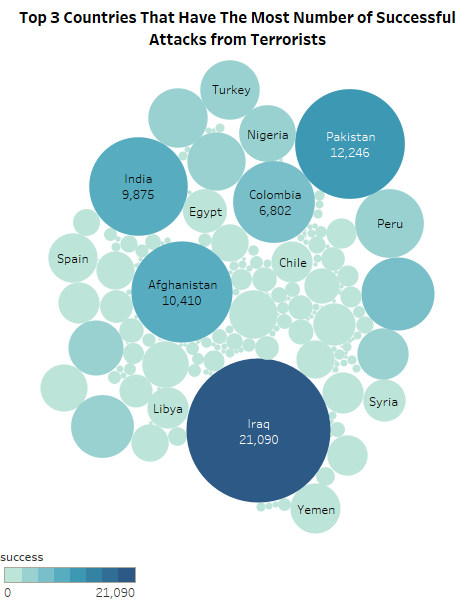


Based on the bar chart above, the factors of successful terrorist attacks are based on the number of successful attacks, type of attacks, type of weapons and the target of type. The highest number of successful attacks is 17671 through the use of explosive weapons that harm private citizens and property targets with bombing or explosion attacks. Terrorist organizations are not static entities. They learn, change their structure, adapt to countermeasures, and continuously look for means to advance their movements and attacks. Sometimes the terrorists' efforts result in successful operations; sometimes they lead to failure. As a result, many methods have been used to achieve successful attacks to harm each type of victim.

This type of visualization is used in this question to compare items between different groups. This bar chart compares the data among categories that shows the factors of successful terrorist attacks.

This visualization is achieved by using each count of success, attacktype\_txt, weapon\_type and target\_type\_txt attributes. The intensity of colour shows each count of success broken down by attacktype\_txt, weapon\_type and target\_type\_txt. The data is filtered on success, which is 1 represent successful attacks. The weapon\_type and success attributes are involved in the data pre-processing which is data cleaning as they contain missing values.

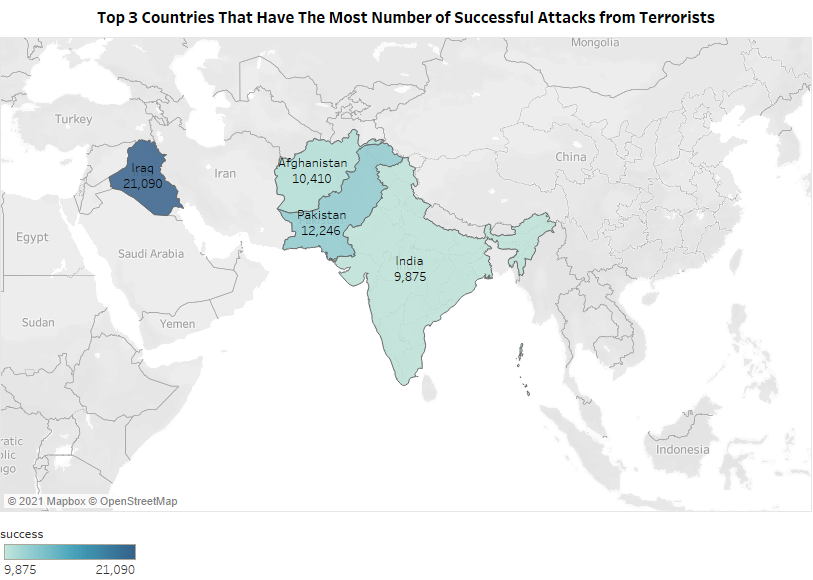
5.1) What are the top 3 countries that have the most number of successful attacks from terrorists?



The bubble chart visualization above shows the top 3 countries that have the most number of successful attacks from terrorists. The top 3 countries that have most number of successful attacks from terrorists are Iraq, Pakistan and Afghanistan with 21090, 12246 and 10410 attacks respectively. Iraq has the highest number of successful attacks because Iraq under Saddam was a major and primary state sponsor of international terrorism.

This type of visualization is used in this question to compare and show relationships between numerical and categorical variables. Different bubble sizes are useful to visually emphasize specific values of which in this case, is the sum of success.

This visualization is achieved by using country and success attributes. The intensity of the colours is used to represent the measurement of sum of success. The size of the bubble chart shows the measurement of sum of success. The marks in each bubble chart are labeled by country and sum of success. The success attribute is involved in the data pre-processing which is data cleaning as it contains missing values.

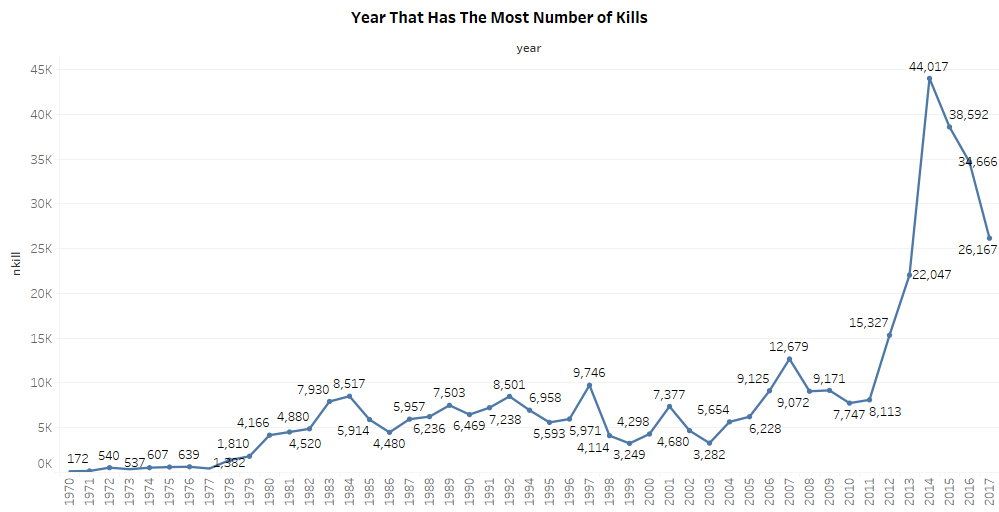


The choropleth map visualization above shows the top 3 countries that have the most number of successful attacks from terrorists. The top 3 countries that have the most number of successful attacks from terrorists are Iraq, Pakistan and Afghanistan with 21090, 12246 and 10410 attacks respectively.

This type of visualization is used in this question to show the clear regional pattern in the data as the data includes the country attribute. In this case, the above visualization shows the country that used longitude and latitude which are necessary to the generate the geographical area.

This visualization is achieved by using longitude, latitude, country and sum of success attributes. The intensity of the colours is used to represent the measurement of sum of success. The size of the bubble chart shows the measurement of sum of success. The marks in each bubble chart are labeled by country and sum of success. The success attribute is involved in the data pre-processing which is data cleaning as it contains missing values.

6.0) Which year has the most number of fatalities?

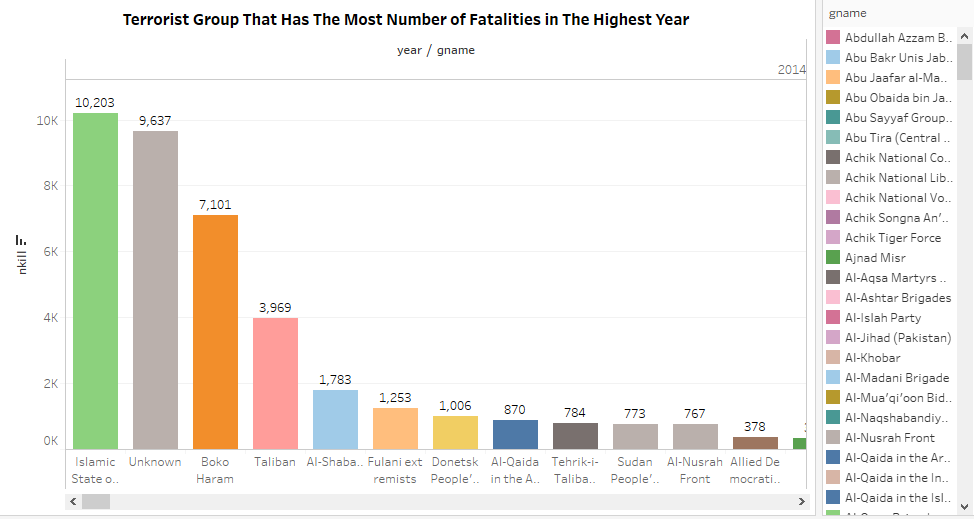


The line chart visualization above shows the year that has the most number of fatalities which is 2014. In 2014, a total of 44017 number of fatalities occurred worldwide. The number of fatalities or deaths are increasing significantly as compared to the previous years. This is probably due to the fact that many of the countries with more terrorism were facing their highest level recorded in 2014. However, there is a decline in deaths of terrorism in the years after 2014. This is likely due to fewer terrorist attacks in certain countries.

This type of visualization is used to show the time-series relationships using continuous data such as year and month. Line chart allows a quick assessment of acceleration when the lines are curving upward, deceleration when the lines are curving downward, and volatility of the up or down frequency. Line chart also helps to track changes over short and long periods of time. In this question, the line chart helps to highlight the year has the most number of fatalities to track the changes over the years.

This visualization is achieved by using year and nkill attributes. The trend line is shown by the sum of nkill for each year. The marks on top of each of the dotted trend line are labeled by the sum of nkill. The nkill and year attributes are involved in the data pre-processing which is data cleaning as they contain missing values.

6.1) Which terrorist group has the most number of fatalities in the highest year?

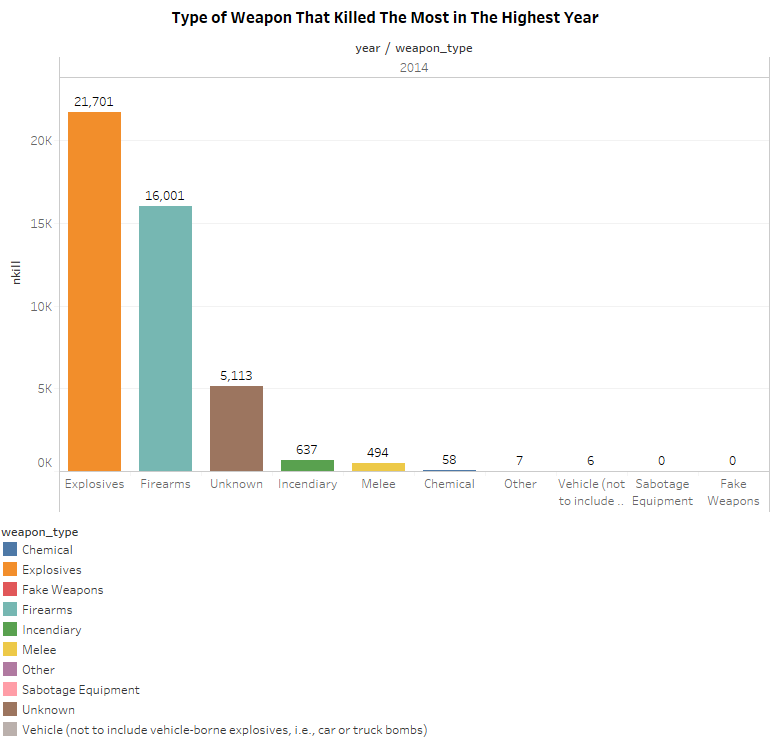


Based on the bar chart above, the terrorist group that records the most total number of fatalities in the highest year, which is 2014, is Islamic State of Iraq and the Levant (ISIL). The total number of people killed in terrorist attacks that peaked in 2015 is likely driven by powerful and lethal attacks carried out by ISIL in Iraq during this year.

This type of visualization is used in this question to compare items between different groups. This bar chart compares the data among the terrorist group has the most number of fatalities in the highest year.

This visualization is achieved by using each nkill, gname and year attributes. The type of colours is differentiated to distinguish between each gname. The marks on each bar chart are labeled by the sum of nkill. The year is filtered by using the 2014 year only. The nkill and year attributes are involved in the data pre-processing which is data cleaning as they contain missing values.

6.2) What type of weapon that caused fatalities the most in the highest year?



Based on the bar chart above, the type of weapon that caused fatalities the most in the highest year, which is 2014, is explosives weapon with 21701 fatalaties. The terrorist attacks are most probably found that explosives type of weapon to be significantly deadlier than any other weapons. Explosives weapon also has wide-area effects that can cause devastating harm to victims and damages or destroys to the infrastructure in the countries.

This type of visualization is used in this question to compare items between different groups. This bar chart compares the data among type of weapon that caused fatalities the most in the highest year.

This visualization is achieved by using each nkill, weapon\_type and year attributes. The type of colours is differentiated to distinguish between each weapon\_type. The marks on each bar chart are labeled by the sum of nkill. The year is filtered by using the 2014 year only. All of the three attributes, which are weapon\_type, nkill and year attributes are involved in the data pre-processing which is data cleaning as they contain missing values.

**5. Summary / Conclusion**

The analysis by using the global terrorism data has sought us to identify factors which are responsible for determining the global terrorism that occurred among various countries. The target type attribute such as private citizens and property, government and police provide the strongest results. These indicate that terrorist organizations are products of their surroundings, which responding the opportunities and constraints provided within the countries. This data analysis is also of benefit because it offers real and meaningful solutions to terrorist violence and targeting. As we interpret the result through the data visualization, it leads to think on how to reduce terrorism through the diminution of the type of victim by influences. Despite the difficulties assessing this data, providing a coherent framework by which it can be studied provides a first step towards providing an effective response. This kind data also provides a promising way to broaden our understanding of terrorism. Hopefully, this project can provide us new knowledge that will allow us to better understand and confront terrorism.

**6. Lesson Learnt from Project**

The first lesson that we have learned from this project is, before conducting the data analysis, we must begin with asking the right questions. Questions should be measurable, clear and concise. The questions are designed to either qualify or disqualify potential solutions to our specific problem. In addition to finding a purpose, consider which metrics to track along the way.

Secondly, decide which dataset we want to collect. The clear measurement principles need to be decided on what and how to measure. We have to consider what kind of data we need to answer our key questions. In answering those key questions, we might need to answer many sub-questions. Finally, in our decision on what to measure, we must be sure to include any reasonable that objections any stakeholders might have.

Next, once data is collected from all the necessary sources, our need to be done with cleaning and sorting through it. Data cleaning is extremely important during the data analysis process, simply because not all data is good data. To generate accurate results, we must identify and purge duplicate data, anomalous data, and other inconsistencies that could skew the analysis.

One of the last lessons that we have learned is choosing the best visualization for our data. We need to start by manipulating our data in a number of different ways, such as plotting it out and finding correlations. As we manipulate our data, we may find that we have the exact data we need, However, there is a likely that we may need to revise our original question or collect more data.

The final lesson is interpreting the results from the data analysis. This part is important because it is how the data analysis will gain actual value from the previous lessons. Interpreting the data analysis should validate why we conducted one in the first place, even if it is not 100 percent conclusive. If our interpretation of the data holds up under all of these questions and considerations, then we likely have come to a productive conclusion. The only remaining step is to use the results of our data analysis process to decide our best course of action.