TerrorismAnalysis

June 3, 2023

Data Analysis on Global Terrorism Dataset

1 About Data

The Data is available at https://www.start.umd.edu/gtd/ anyone can obtain it by just filling a small form , The timeframe of the whole dataset is 1970 to June 2021, There are 135 columns/features of every associated incident uniquely identified by eventid, We will go through this data using python and try to find some insights about the terrorists attacks over the decades in the whole world.

Explanation of selected columns:

iyear - Year in which attack was done
success - Success of a terrorist strike
suicide - 1 = "Yes" The incident was a suicide attack. 0 = "No" There is no indication that the incident was a suicide
attacktype1 - The general method of attack
attacktype1_txt - The general method of attack and broad class of tactics used.
weaptype1_txt - General type of weapon used in the incident
weapsubtype1_txt - More specific value for most of the Weapon Types
nkill - The number of total confirmed fatalities for the incident

Data is divided in two excel sheets, one for data from 1970 to 2020 and the other sheet contains data from Jan 2021 to June 2021, Going forward to easily plot graphs from the dataframe I will concatenate the two files in a single dataframe.

There are total of 214666 attacks over 40 years, which means **5366.65** attacks every year on average. A whooping **14** Terrorist attacks per day around the globe.

```
Dataframe Info
<class 'pandas.core.frame.DataFrame'>
Int64Index: 214666 entries, 0 to 4959
Columns: 135 entries, eventid to related
dtypes: datetime64[ns](1), float64(54), int64(23), object(57)
memory usage: 222.7+ MB
```

There are null values present in multiple rows for 94 columns, to select a subset of this dataset I will drop the columns that have null value more than 5% of the rows.

```
 These are the columns that have null value less than 5%
['eventid', 'iyear', 'imonth', 'iday', 'extended', 'country', 'country_txt',
```

```
'region', 'region_txt', 'provstate', 'city', 'latitude', 'longitude',
'specificity', 'vicinity', 'crit1', 'crit2', 'crit3', 'doubtterr', 'multiple',
'success', 'suicide', 'attacktype1', 'attacktype1_txt', 'targtype1',
'targtype1_txt', 'target1', 'natlty1', 'natlty1_txt', 'gname', 'guncertain1',
'individual', 'weaptype1', 'weaptype1_txt', 'property', 'ishostkid', 'dbsource',
'INT_LOG', 'INT_IDEO', 'INT_MISC', 'INT_ANY']
```

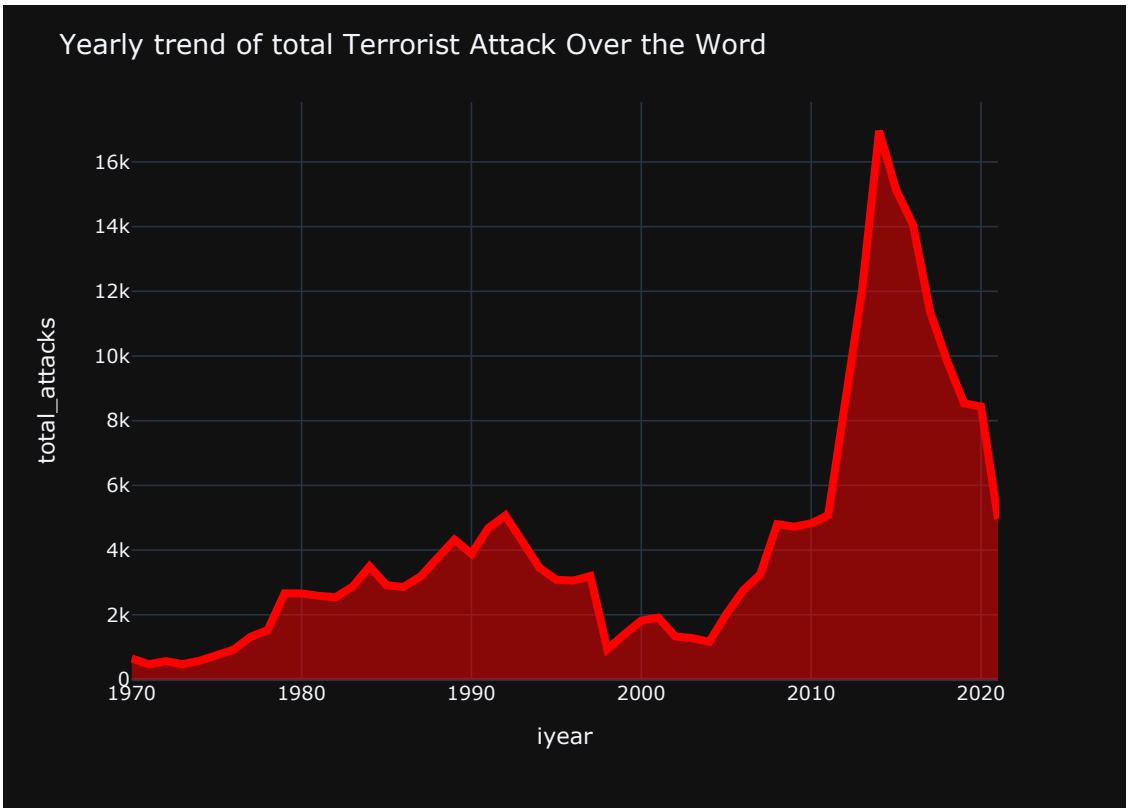
I will add 'nkill' to the dataset as well to analyse some columns relationship to the number of deaths due to these attacks.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 214666 entries, 0 to 4959
Data columns (total 42 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   eventid          214666 non-null  int64
 1   iyear            214666 non-null  int64
 2   imonth           214666 non-null  int64
 3   iday             214666 non-null  int64
 4   extended         214666 non-null  int64
 5   country          214666 non-null  int64
 6   country_txt      214666 non-null  object
 7   region           214666 non-null  int64
 8   region_txt       214666 non-null  object
 9   provstate        214666 non-null  object
 10  city             214240 non-null  object
 11  latitude         209940 non-null  float64
 12  longitude        209939 non-null  float64
 13  specificity      214665 non-null  float64
 14  vicinity         214666 non-null  int64
 15  crit1            214666 non-null  int64
 16  crit2            214666 non-null  int64
 17  crit3            214666 non-null  int64
 18  doubtterr        214666 non-null  int64
 19  multiple         214663 non-null  float64
 20  success          214666 non-null  int64
 21  suicide          214666 non-null  int64
 22  attacktype1      214666 non-null  int64
 23  attacktype1_txt  214666 non-null  object
 24  targtype1        214666 non-null  int64
 25  targtype1_txt    214666 non-null  object
 26  target1          214031 non-null  object
 27  natlty1          212592 non-null  float64
 28  natlty1_txt      212592 non-null  object
 29  gname            214666 non-null  object
 30  guncertain1      214286 non-null  float64
 31  individual       214666 non-null  int64
 32  weaptype1        214666 non-null  int64
 33  weaptype1_txt    214666 non-null  object
```

```
 34  property          214666 non-null  int64
 35  ishostkid         214488 non-null  float64
 36  dbsource          214666 non-null  object
 37  INT_LOG           214666 non-null  int64
 38  INT_IDEO          214666 non-null  int64
 39  INT_MISC          214666 non-null  int64
 40  INT_ANY           214666 non-null  int64
 41  nkill             201715 non-null  float64
dtypes: float64(8), int64(23), object(11)
memory usage: 70.4+ MB
```

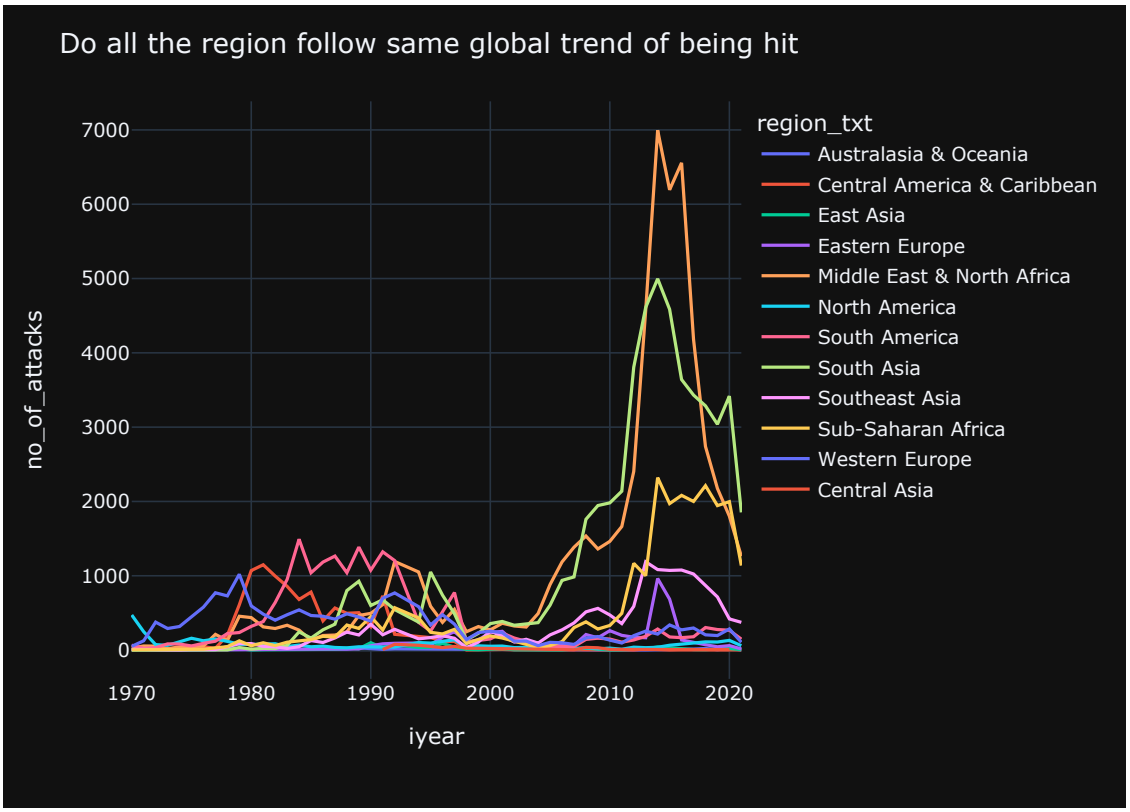
Ahh! A little relief for RAM and CPU . Now we are left with 41 columns and the column year of all the events can help us identify various trends of the terrorists activities around the globe every year.

2 Yearly trend of attacks across the Globe



The maximum number of attacks was in the year 2014 with total of 16.96 thousands.
The least number of attacks was in the year 1970 with total of 651 hundred.
The number of attacks have decreased after 2014 by 75.88% till 2021

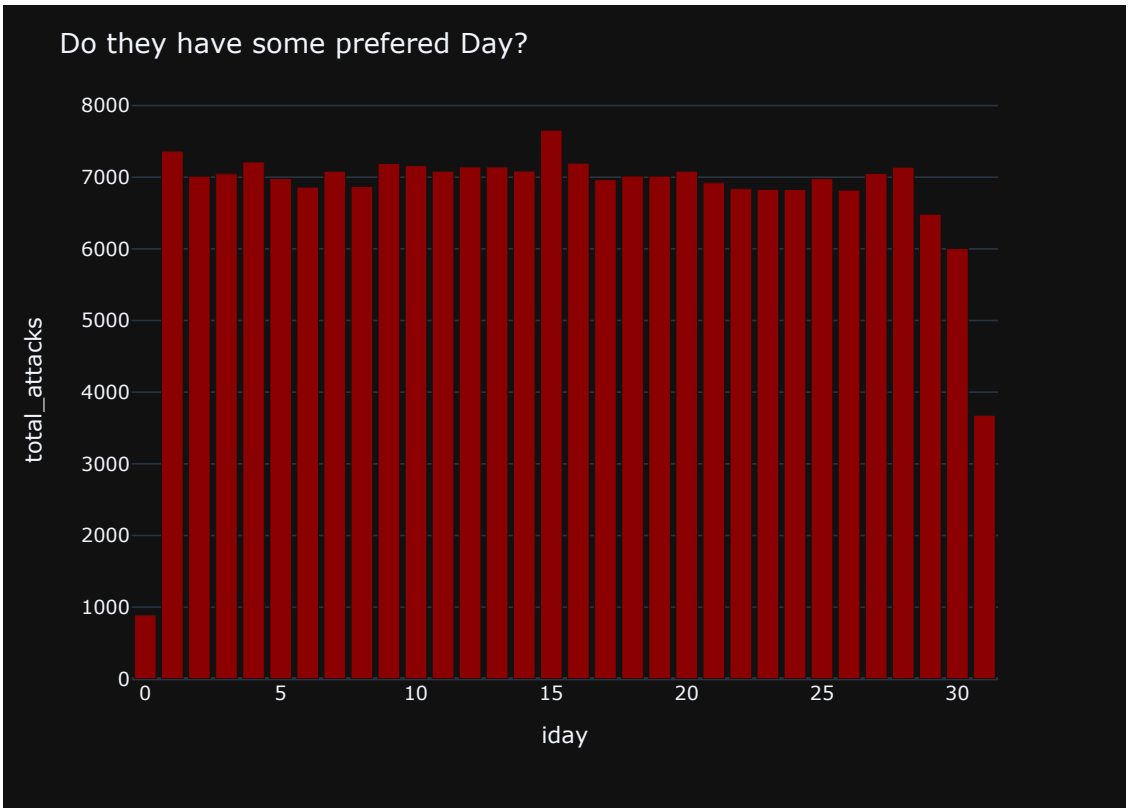
2.1 How the yearly trend differs in different region for the number of attacks



These regions don't follow the usual global trend

```
Australasia & Ocenia
Central America and Caribbean
East Asia
North America
South America
Western Europe
Central Asia
```

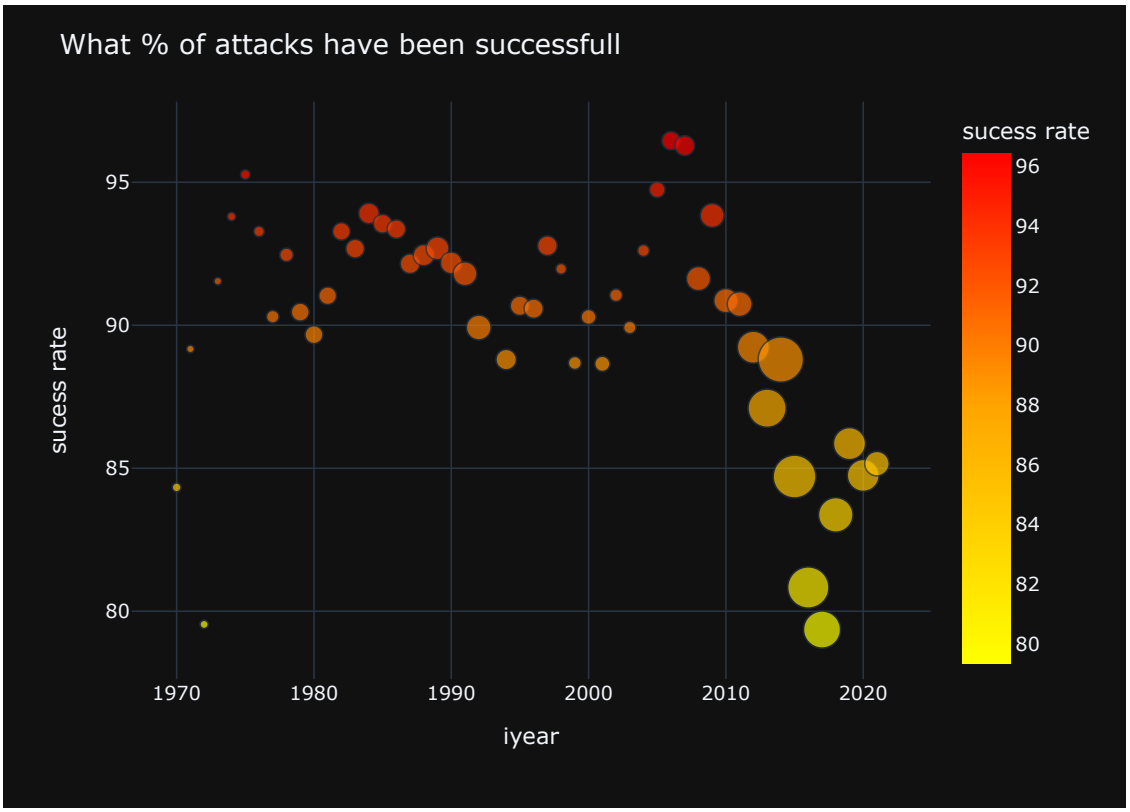
3 Any trends by Months or Days?



There is no prefrences of day or months shown in terrorist attacks.

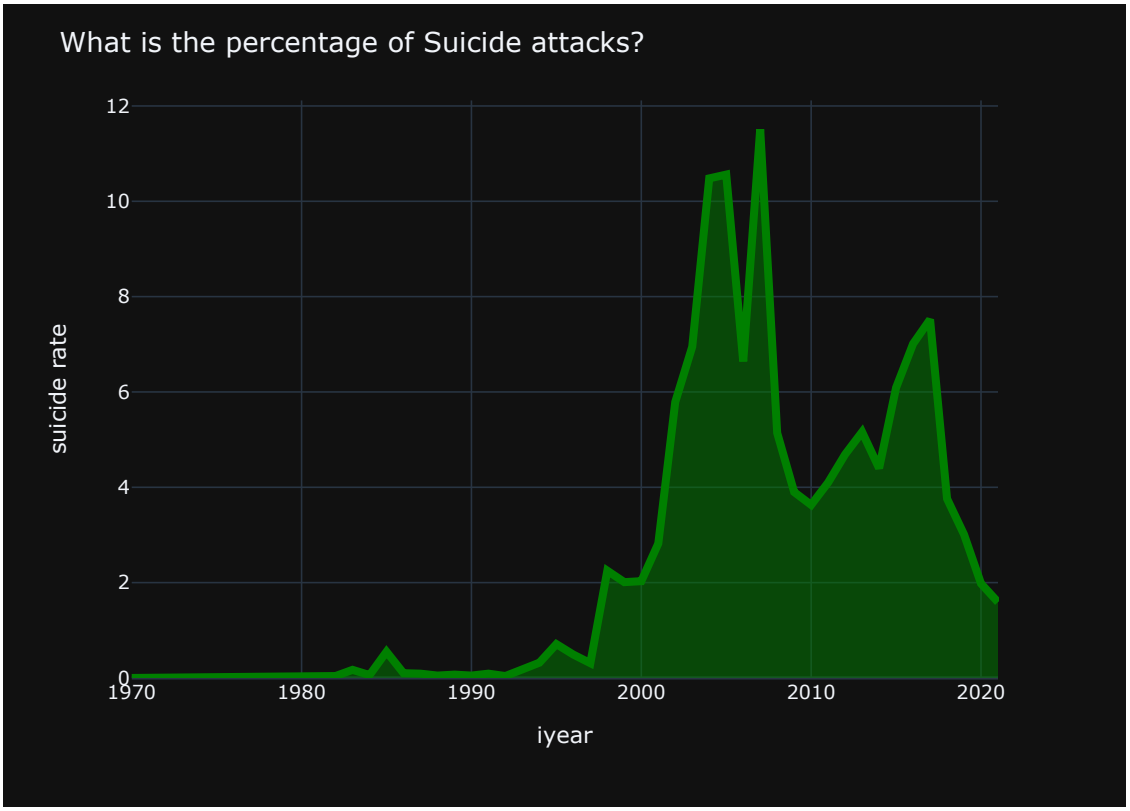
The Reason why we have 0 as day : `For attacks that took place between 1970 and 2011, if the exact day of the event is unknown, this is recorded as "0."`

4 Have we been able to fights these attacks?



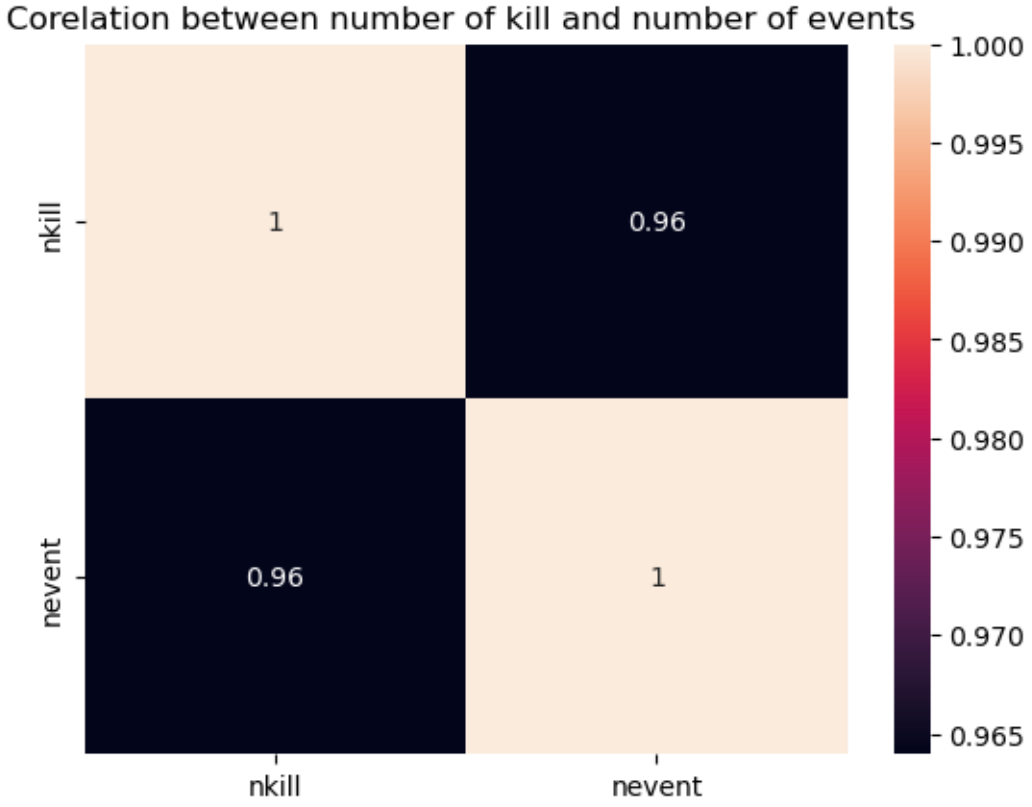
With the decrease in success rate over time the number of attacks have increased.
All time lowest percentage of success rate was in year 2017 with 79.36% of attacks being successfull.
When compared the number of attacks from 2017 to 1972 it has increased by 95%.

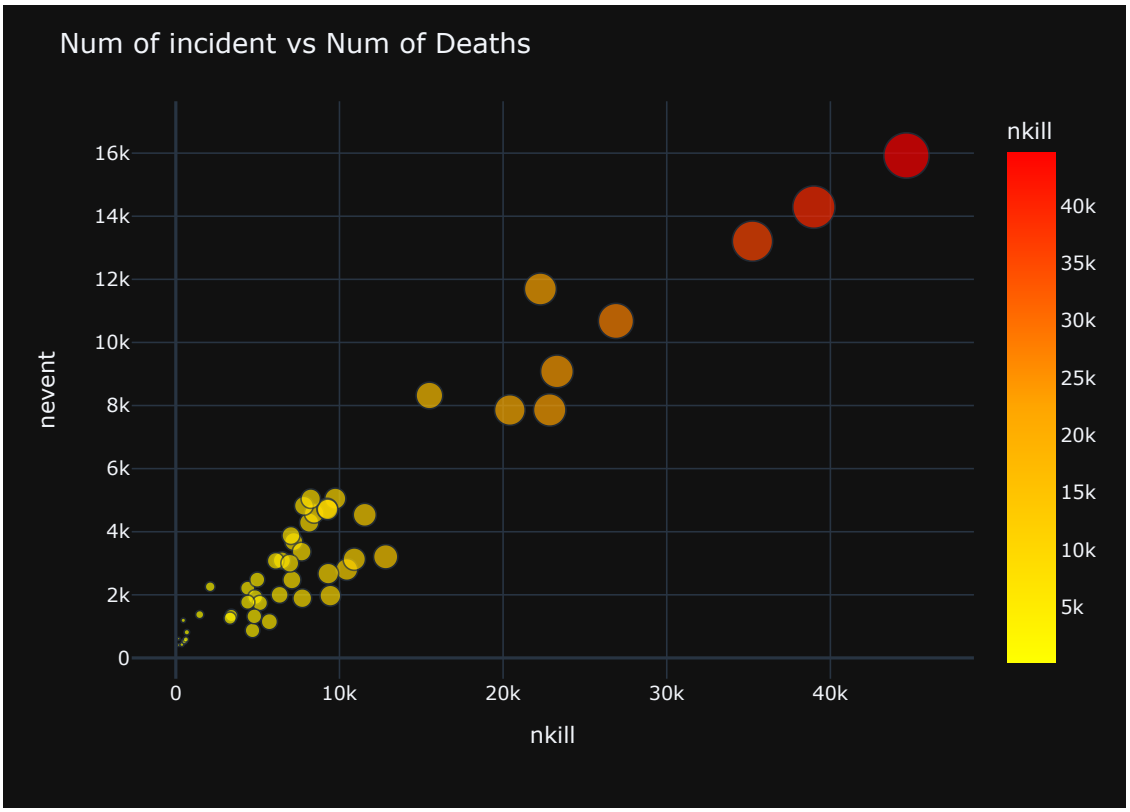
5 Suicidal Attacks



By 1980 there were no Suicidal attacks and has increased rapidly since 2000.
Suicidal attacks have decreased in percentage since 2017.
In 2021 June the Suicidal attacks are 1.79% of the total attacks which is a decrease since 201

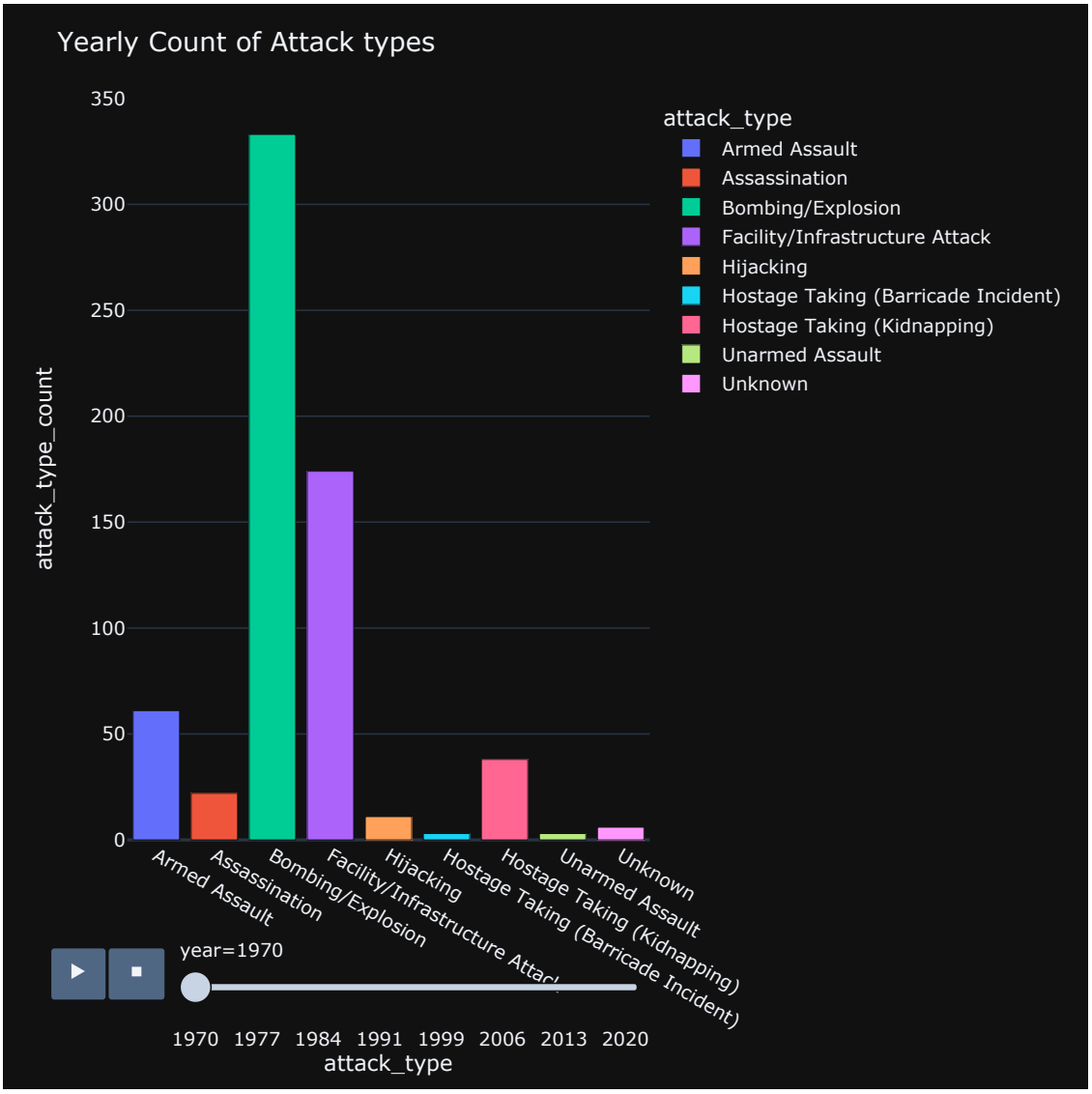
6 Relation between no. of events and no. of deaths



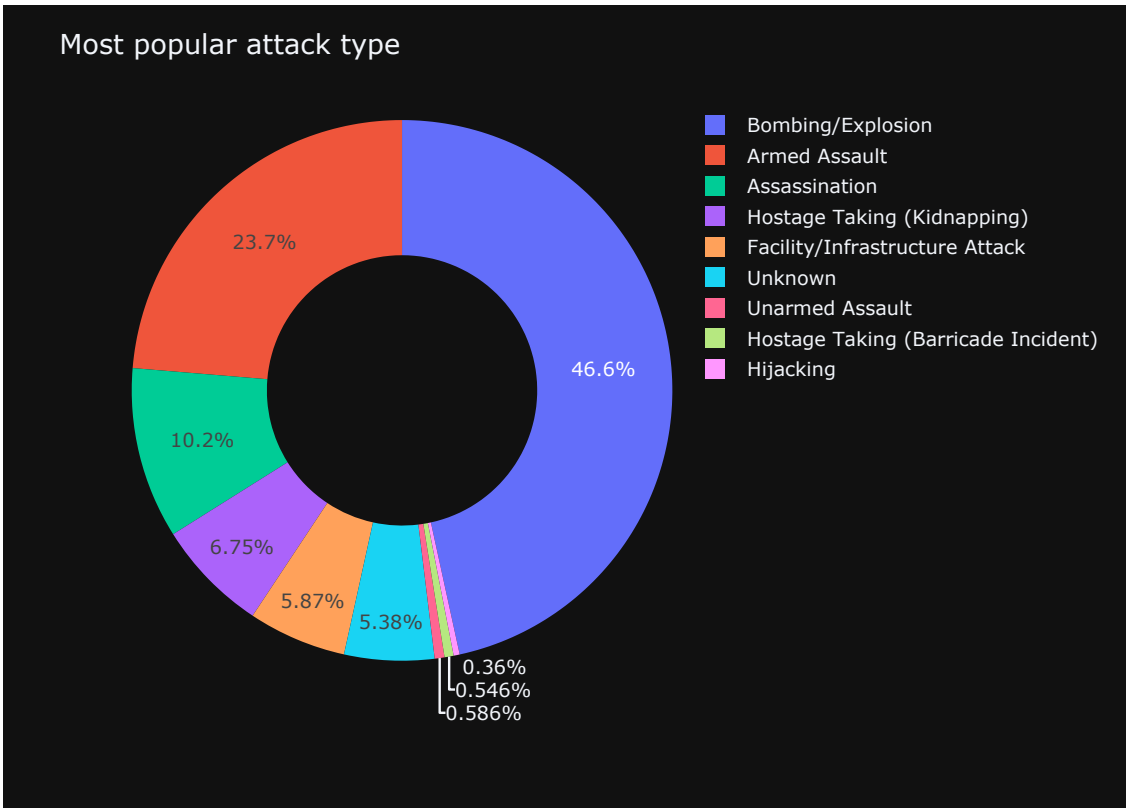


The number of kills is highly correlated with the number of attacks.
This two variables show a nearly perfect linear relationship.

7 What various types of attacks are being used over the years.

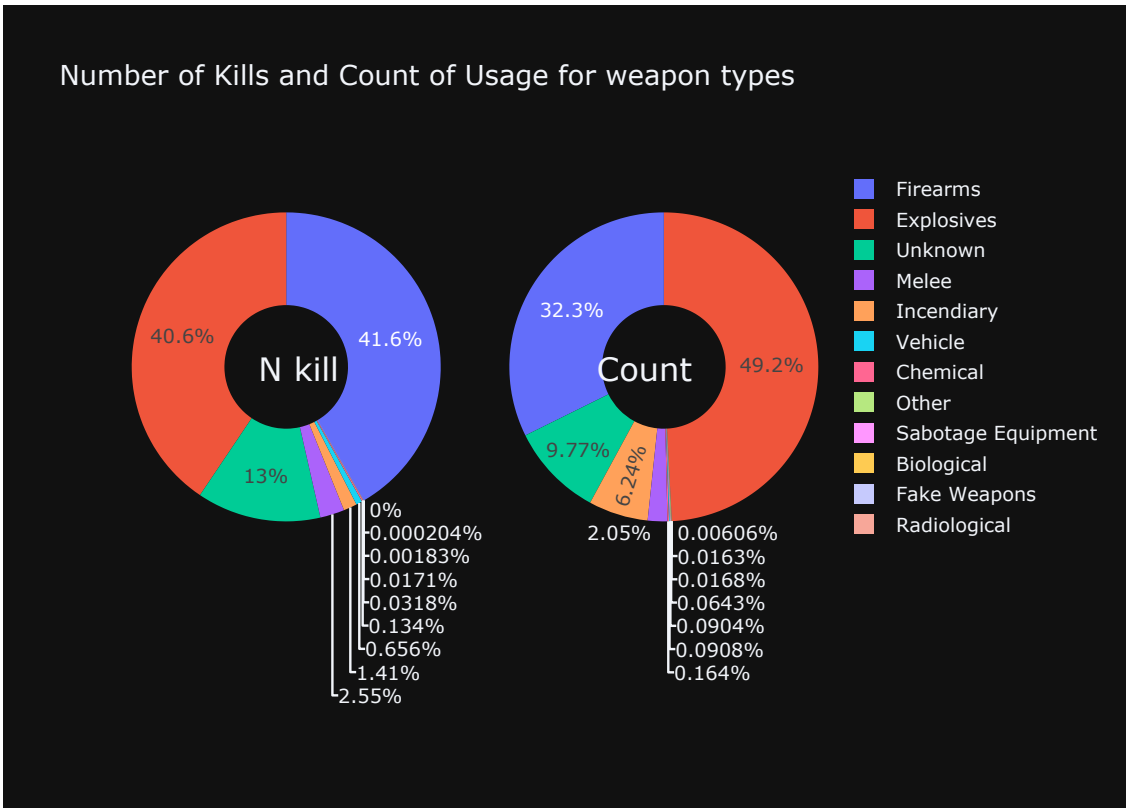


8 What is the most used attack type



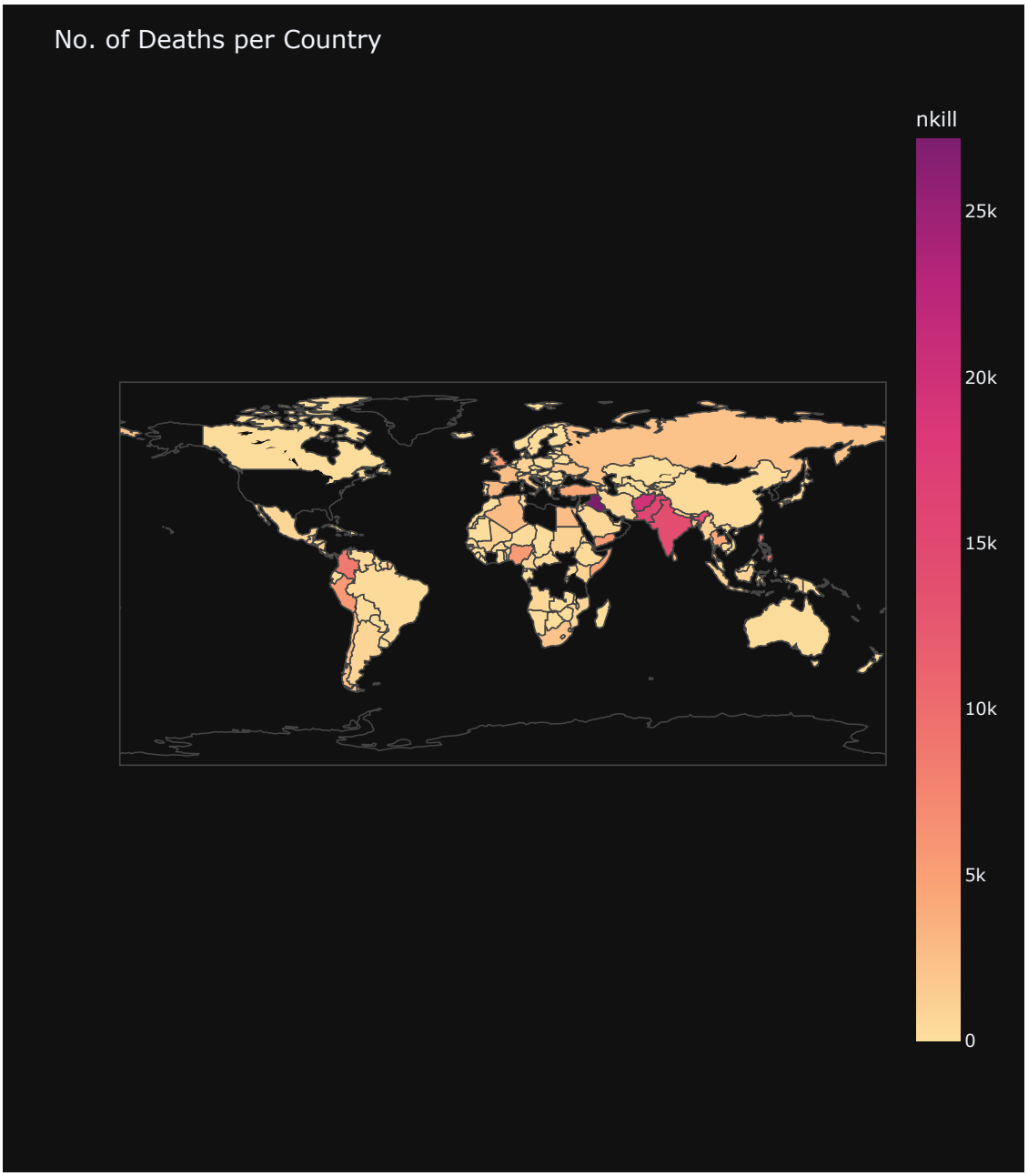
Bomb Explosions is the most used attack type.
23.7% of the attacks are of Armed Assault type.
Least Used method is Hijacking 0.36% of all the attacks.

9 What kind of Weapons are used for these attacks.



Firearms are used in 34.6% of the attacks and still has caused 42.6% deaths of the total casaulties.
Friearms are more lethal than Explosives.
Radiological have the least percentage of deaths and as also been used the least amount of time.

10 Map time



The most casualties of terrorsim are in Iraq, followed by Afghanistan,Pakistan and India. Columbia has the highest casualties from the American Continents followed by Peru. Turkmenistan has the least number of deaths only 2.