ANALYSIS OF GLOBAL TERRORISM DATA

Project Title

Analysis of Global Terrorism Data

Group Members:

Abhishek Kumar Singh, GIET University, 1701210128

Mr Janmejoy Das Adhikari, GIET University, 1701210017

Ashok Kumar, GIET University, 1701210320

MD Afzal Alam, Chandigarh University, 11717629

Requirement Specification

- Acknowledgement
- Project Objective
- Data Description
- Data Loading
- Interpreting the Data
- Distribution Analysis
- Data Cleaning and Munging
- Analysing the Data based on various parameters
- Drawing Inferences based on Data Analysis
- Future Scope of Improvements
- Project Certificate

Acknowledgement

I take this opportunity to express my profound gratitude and deep regards to my faculty Mr. Kaushik Ghosh for his exemplary guidance, monitoring and constant encouragement throughout the course of this project. The blessing, help and guidance given by him time to time shall carry me a long way in the journey of life on which I am about to embark.

I am obliged to my project team members for the valuable information provided by them in their respective fields. I am grateful for their cooperation during the period of my assignment.

Abhishek Kumar Singh

Project Objective

This project is a comprehensive study of analysis of the impact of terrorism for 163 countries and which covers 99.7 per cent of the world's population.

Given the significant resources committed to counter terrorism by governments across the world, it is important to analyse and aggregate the available data to better understand its various properties.

Examples of the information contained in this study are:

- The long-term trends and how terrorism changes over time.
- The types of strategies deployed by terrorists, their tactical targets and how these have evolved over time.

This project produces a composite score to provide a ranking of countries on the impact of terrorism. The Global Terrorism Data is unique is the resource consisting of systematically and comprehensively coded data for 180,000+ terrorist incidents.

Project Scope

Defining terrorism is not a straightforward matter. There is no single internationally accepted definition of what constitutes terrorism and the terrorism literature abounds with competing definitions and typologies. So, we define terrorism as ‘*The threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.*’

The project will feature analysis of tens of thousands of records to understand and answer various questions, like:

- Why terrorism is more prevalent in some regions?
- How the terrorism shifted over a period?
- What are the tactics preferred by terrorist groups?

Data Description

Our main resource used for the compilation of this report remains Global Terrorism Database (GTD). The GTD is a database of incidents of terrorism from 1970 onwards. The GTD describes itself as the “Most comprehensive unclassified database on terrorist events in the world” and includes over 180,000+ terrorist attacks. The entire database (about 150 MB excel file) is open source and can be used for non-commercial purpose.

The database includes some 80+ columns describing each record on various parameters like attack-date, resolution, weapons used, motive, organisation name, success, damage, civilians killed, etc. The data itself need to be inferred properly in order to get as accurate a result as possible.

Data Loading

```python
import pandas as pd

df=pd.read_csv("terror_data.csv")
```

Data loading is the process of loading the dataset for further use of it. By using the Pandas module, we can load the dataset. In this project we will use 4 modules, matplotlib and seaborn for graphical analysis, pandas for Series, DataFrame object and NumPy for nD array.

```python
print(df)
```

```
In [6]: df
```

[6]:

	year	month	day	country_id	country_name	resolution	region_id	region_name	state	city	...	motive	weap_id	weap_name	kill	wo
0	1970	7	2	58	Dominican Republic	Unknown	2	Central America & Caribbean	Unknown	Santo Domingo	...	Motive unknown	13	Unknown	1	
1	1970	1	0	130	Mexico	Unknown	1	North America	Federal	Mexico city	...	Motive unknown	13	Unknown	0	
2	1970	1	0	160	Philippines	Unknown	5	Southeast Asia	Tarlac	Unknown	...	Motive unknown	13	Unknown	1	
3	1970	1	0	78	Greece	Unknown	8	Western Europe	Attica	Athens	...	Motive unknown	6	Explosives	0	
4	1970	1	0	101	Japan	Unknown	4	East Asia	Fukouka	Fukouka	...	Motive unknown	8	Incendiary	0	
5	1970	1	1	217	United States	Unknown	1	North America	Illinois	Cairo	...	Protest	5	Firearms	0	

Output:

```
(181691, 24)
```

Interpreting the Data

In this process we will identify the dependent and independent variables for our analysis process. We are going to identify the dependent variables according to the dependency on the dataset. We can also say that the dependent variables are the most important attributes of our dataset in order to better understand and interpret the data for some definite conclusion.

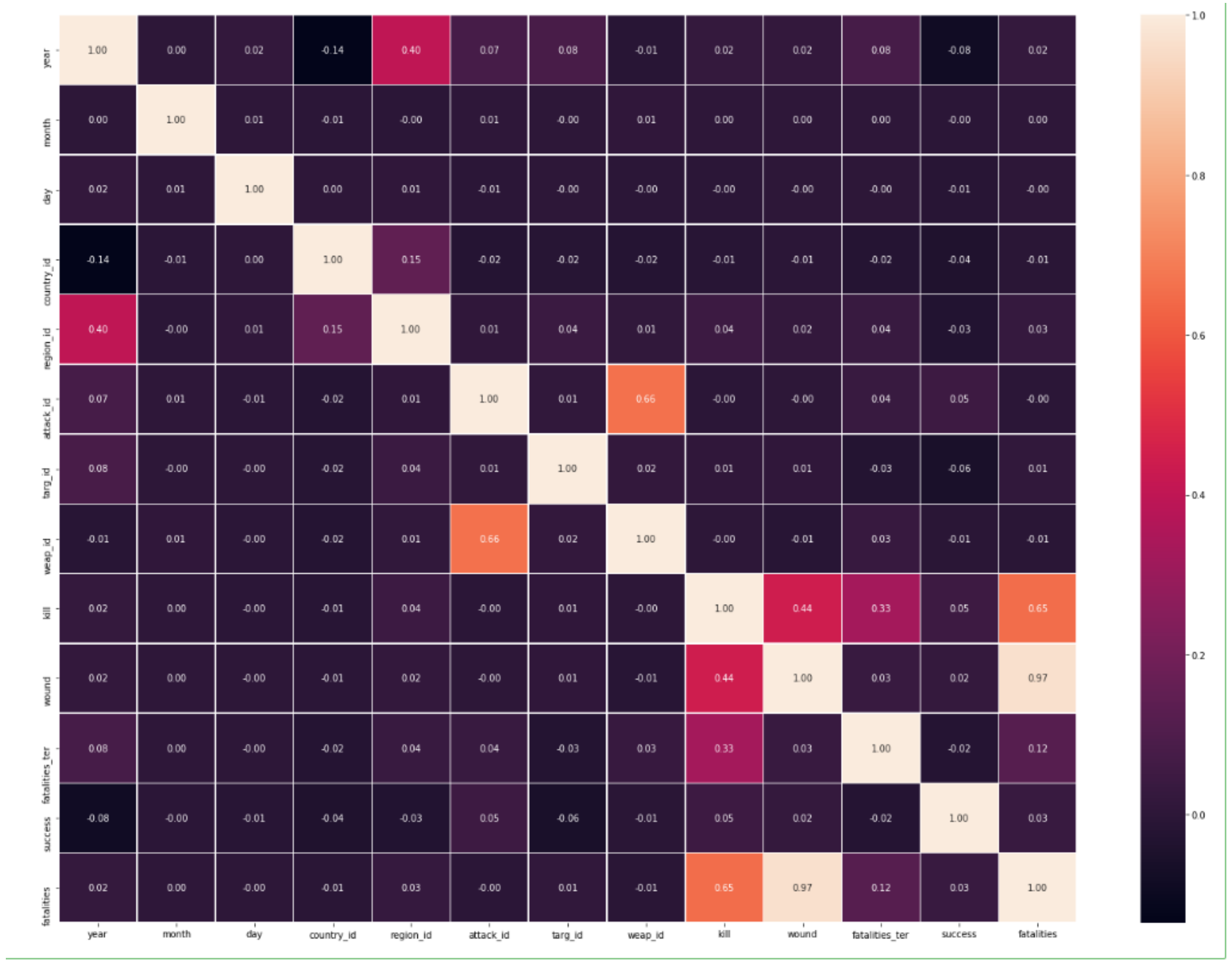
Below are our following data attributes/columns:

```
Iyear, imonth, iday, country, country_txt, resolution
, region, region_txt, provstate, city, success, attac
ktype1, attacktype1_txt, attacktype2, attacktype2_txt
, targtype1, targtype1_txt, gname, motive, weaptype1,
weaptype1_txt, nkill, nwound, fatalities, fatalities_
ter, propextent_txt
```

These above are the attributes of our dataset. By going through, dataset we can recognize that success and fatalities are the dependent variables. Because by using this we can analyse the other variable which will be useful for the future analysis.

Correlation heatmap

```
ax = plt.subplots(figsize=(25, 18))
sns.heatmap(df.corr(), annot=True, linewidths=.5,
fmt= '.2f', ax=ax)
plt.show()
```



This correlation heatmap helps us identify the effect of independent variable on dependent one. The correlation value can range from

-1 < correlation < +1.

Data cleaning and munging

In this process we will remove the dirty value or NAN values from our corresponding dataset, and we will drop the unnecessary columns or the columns which will not use in our dataset. And after this we will fill the columns with appropriate values, which is nothing but called data filling.

- df.info # by this we will get the information about every column containing the non-null values in it.
- df.isnull().sum() # by this we can get the info about every column containing the null values in it

```
before cleaning
iyear                 0, imonth              0
iday                  0, country             0
country_txt           0, resolution          0
region                0, region_txt          0
provstate           421, city              434
success               0, attacktype1         0
attacktype1_txt       0, attacktype2    175377
attacktype2_txt  175377, targtype1           0
targtype1_txt         0, gname               0
motive           131130, weaptype1           0
weaptype1_txt         0, nkill               0
nwound                0, fatalities          0
fatalities_ter        0, propextent_txt 117626
ransom           104310, ransomamt      180341
dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 28 columns):
iyear              181691 non-null int64, imonth            181691 non-null int64
iday               181691 non-null int64, country           181691 non-null int64
country_txt        181691 non-null object, resolution        181691 non-null object
region             181691 non-null int64, region_txt        181691 non-null object
provstate          181270 non-null object, city              181257 non-null object
success            181691 non-null int64, attacktype1       181691 non-null int64
attacktype1_txt    181691 non-null object, attacktype2        6314 non-null float64
attacktype2_txt    6314 non-null object, targtype1         181691 non-null int64
targtype1_txt      181691 non-null object, gname             181691 non-null object
motive             50561 non-null object, weaptype1         181691 non-null int64
weaptype1_txt      181691 non-null object, nkill             181691 non-null int64
nwound             181691 non-null int64, fatalities        181691 non-null int64
fatalities_ter     181691 non-null int64, propextent_txt    64065 non-null object
ransom             77381 non-null float64, ransomamt         1350 non-null float64
dtypes: float64(3), int64(13), object(12), memory usage: 38.8+ MB
```

```
df=df.drop(['attacktype2','attacktype2_txt','ransom','ransomamt'],axis=1)
```

By this we will drop or delete the columns which have NaN values.

Replacing the null values from the columns

```
print(df[('provetext').fillna('unknown'))
print(df['provstate'].fillna('unknown'))
print(df['motive'].mode())
df['motive'].fillna(str(df['motive'].mode().values[0]),inplace=True)
print(df['motive'])
```

After cleaning and dropping the column we have now the new columns which have no null values in it:

Rename columns: In this process we will rename the column for our

further use of the dataset.

```
df.rename(columns={'iyear': 'year', 'imonth': 'month', 'iday': 'day',
'country': 'country_id', 'country_txt': 'country_name', 'region':
'region_id',                'region_txt': 'region_name', 'provstate':
'state', 'city': 'city', 'success': 'success', 'attacktype1': 'attcakid',
'attacktype1_txt': 'attack_type', 'targtype1': 'targid',
'targtype1_txt': 'targ_name', 'gname': 'group_name', 'motive': 'motive',
'weaptype1': 'weap_id',                'weaptype1_txt': 'weap_name',
'nkill': 'kill', 'nwound': 'wound', 'fatalities': 'fatalities',
'fatalities_ter': 'fatalities_ter', ‘propextent_txt’: ‘propextent_txt’
}, inplace=True)
```

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690, Data columns (total 24 columns):
year              181691 non-null int64, month            181691 non-null int64
day               181691 non-null int64, country_id       181691 non-null int64
 intry_name       181691 non-null object, resolution       181691 non-null object
 gion_id          181691 non-null int64, region_name      181691 non-null object
 ite              181691 non-null object, city             181691 non-null object
 tack_id          181691 non-null int64, attack_type      181691 non-null object
 g_id             181691 non-null int64, targ_name        181691 non-null object
 oup_name         181691 non-null object, motive           181691 non-null object
 ip_id            181691 non-null int64, weap_name        181691 non-null object
 l                181691 non-null int64, wound            181691 non-null int64
 alities_ter      181691 non-null int64, damage           181691 non-null object
success           181691 non-null int64, fatalities       181691 non-null int64
dtypes: int64(13), object(11)
memory usage: 33.3+ MB
```

Analysing the Data

We have analysed the data based on various parameters like year, weapons used, attack pattern, weapons. The content of analysis is as below:

- *Change in number of terror attacks since 1970*
- *Death and wound toll over a period*
- *Five deadliest terrorist groups*
- *Failed and Successful terrorist attack, 1998-2017*
- *Civilian's fatalities Continent wise*
- *Institutions most frequently targeted by terrorists*
- *Ten countries most frequently targeted, 1970-2000 to 2001-2017*
- *Countries where military base is most vulnerable*
- *Countries most often targeted by Explosives*
- *Countries where airline industry is vulnerable*

Analysing the Data

- *Countries where government institute is most often targeted*
- *Top 3 most frequently targeted Indian states, 1997-2017*
- *Rate of change in terrorism in India, 2002-2017*
- *Analysing civilian's and terrorist's death pattern, 1998-2017*
- *Largest decreases in death from terrorism, 2016-2017*
- *Most favoured weapons types used by terrorists*
- *Declining phase of terrorism*
- *Maximum people targeted over entire period*
- *Frequency of attack in North America*
- *Attacks based on weapon: 'Firearm'*

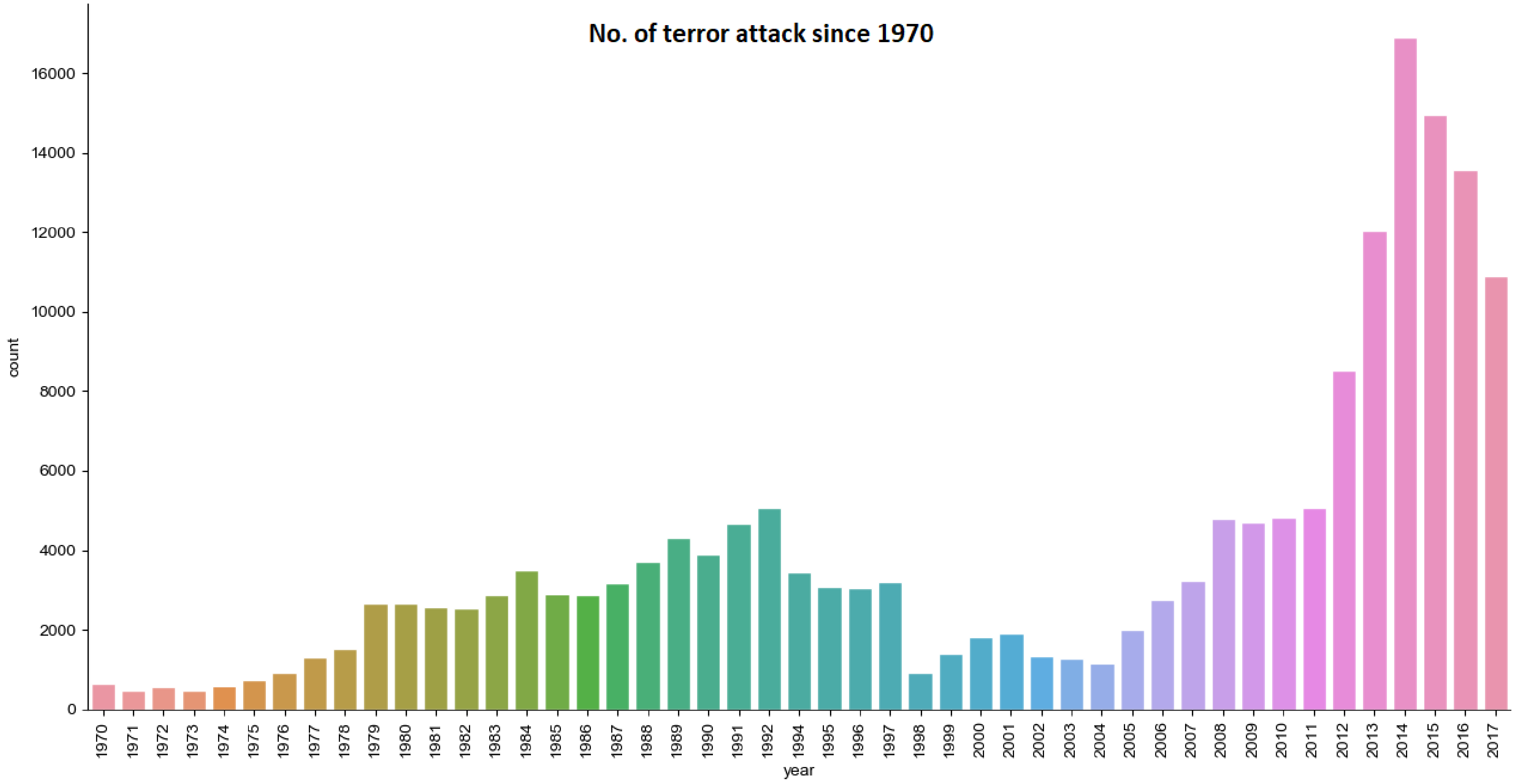
Change in number of terror attacks since 1970

Inference: *21st Century marked itself as the terror-era.*

- The total terror attacks reported in between 2000-2017 is nearly twice to that occurred in time period of 1970-1999.

- Nearly 69,836 incidents were reported before 2000, which spiked up to 1,11,855 in first two decades of 21st century.

```
import pandas as pd
import seaborn as sns

data = pd.read_csv('dataset/terror_data.csv')
axis = plt.subplot()
sns.set(color_codes='purple')
g = sns.countplot(x='year', data=data, ax=axis)
g.set_xticklabels(g.get_xticklabels(), rotation=90)
g.tick_params(axis='x', labelsize=10)
sns.despine()
plt.show()
```



Death and wound toll over a period

In the beginning of 1970, the number of deaths and wounded was noticeably small with deaths about 8,000 and wounded about 400. However, due to the increase in activity of terror groups like Boko Haram, the death and wounded toll increased by 6-fold in 1980-1989.

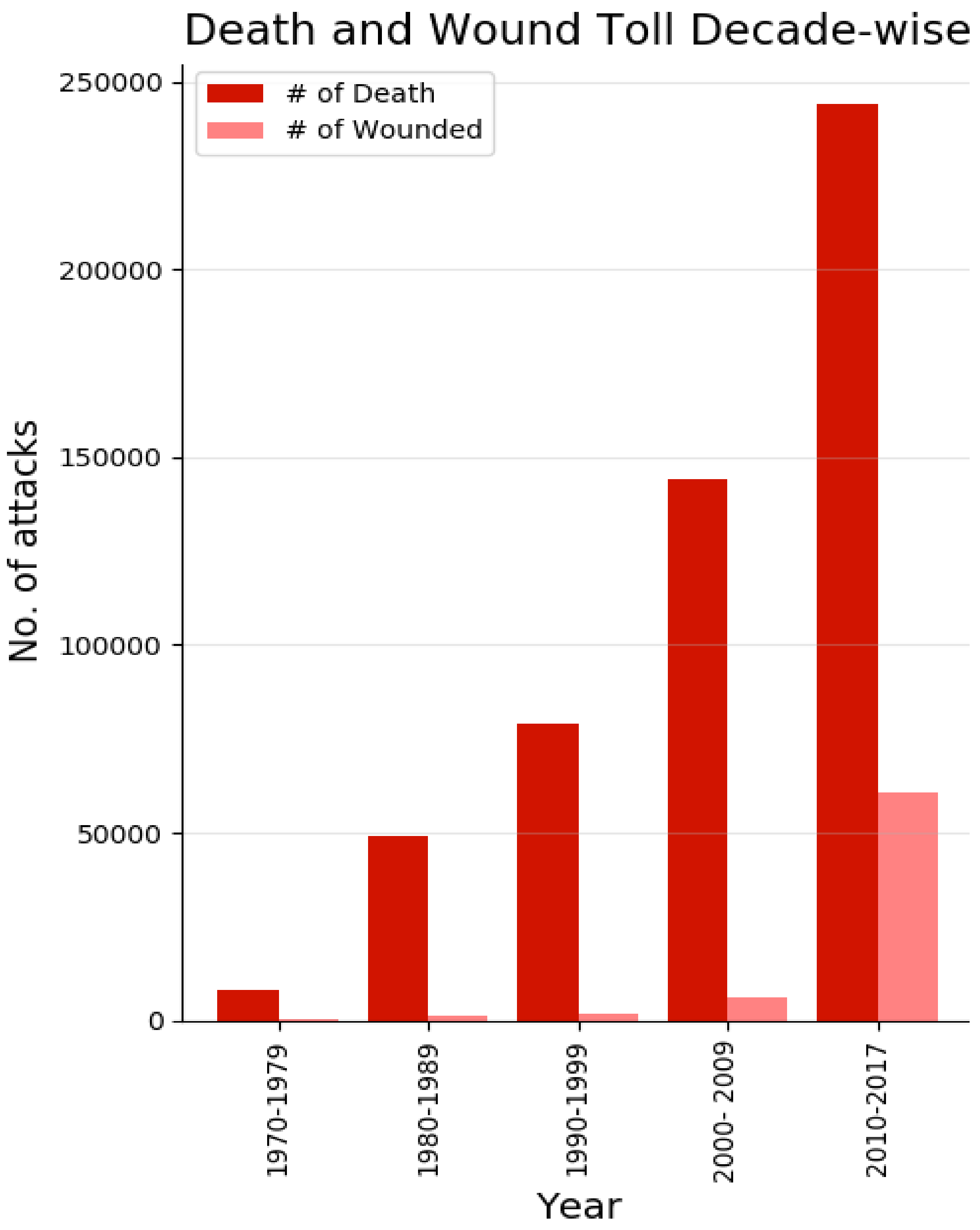
In decade 1990-1999 the trend remains same with increase in fatality of death and wound toll by 160%. This suggests that wound incurred by civilians were fatal enough to cause death. The increase in the intensity of conflict in the Middle East fuelled the fire thus increasing overall fatality of the decade by another 200%. The terror attacks further intensified by another 180% in decade 2010-present. This was mainly because increase in terror activity by ISIL and Middle East terror groups.

```python
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('dataset/terror_data.csv')
li = [['1970-1979', 0, 0], ['1980-1989', 0, 0], ['1990-1999',
0, 0], ['2000- 2009', 0, 0], ['2010-2017', 0, 0]]
for i in range(len(data)):
    if 1970 <= data.iloc[i, 0] < 1980:
        li[0][1]+=data.iloc[i,19]; li[0][2]+=data.iloc[i,20]
    elif 1980 <= data.iloc[i, 0] < 1990:
        li[1][1]+=data.iloc[i,19]; li[1][2]+=data.iloc[i,20]
    elif 1990 <= data.iloc[i, 0] < 2000:
        li[2][1]+=data.iloc[i,19]; li[2][2]+=data.iloc[i,20]
    elif 2000 <= data.iloc[i, 0] < 2010:
        li[3][1]+=data.iloc[i,19];li[3][2]+=data.iloc[i,20]
    else:
        li[4][1]+=data.iloc[i,19]; li[4][2]+=data.iloc[i,20]
df = pd.DataFrame(li); xpos = np.arange(len(df))
plt.xticks(xpos, df[0], rotation='vertical')
plt.xlabel('Year', fontdict={'size':14})
plt.ylabel('No. of attacks', fontdict={'size':14})
plt.title('Death and Wound Toll Decade-wise',
fontdict={'size':16})
plt.bar(xpos - 0.2, df[1], width=0.4, label='# of Death',
color='#d11400')
plt.bar(xpos + 0.2, df[2], width=0.4, label='# of Wounded',
color='#ff8282'); plt.legend()
plt.grid(axis='y', alpha=.3)
plt.show()
```

Contd...

Inference: *Terror attacks are more prevalent and fatal in the later decade.*

- Terror groups of Middle East flourished well, leading to significant increase in property and life loss.
- Losses of lives increased by 600% from decade 1970 to that of 1980.
- Losses of lives again increased by 160% in decade 1990.
- Decade 2000 did not relieved either since death toll increased by another 200%.
- The death toll peaked in decade 2010 claiming as many as 2,43,837 lives and 60,768 wounded thus increasing by 180%.
- The overall increase in death toll from 1970 to 2010 is as high as 1000% i.e. 10 times claiming almost about 5.5 lakhs lives and leaving 1 lakh wounded.



Five deadliest terrorist groups

Determining which terrorist groups are the most active and responsible for the most deaths can be difficult, as many groups have regional affiliates and other groups working in partnership or partially under the same command.

For the purposes of this section, we did not include affiliates in its definition of a terrorist group. For example, ISIL refers only to the Islamic State of Iraq and the Levant, and does not include the Khorasan chapter or Sinai Province of the Islamic State, despite the strong connections between the two groups. Similarly, Al-Shabaab is counted as a single group, rather than an affiliate of Al-Qa'ida.

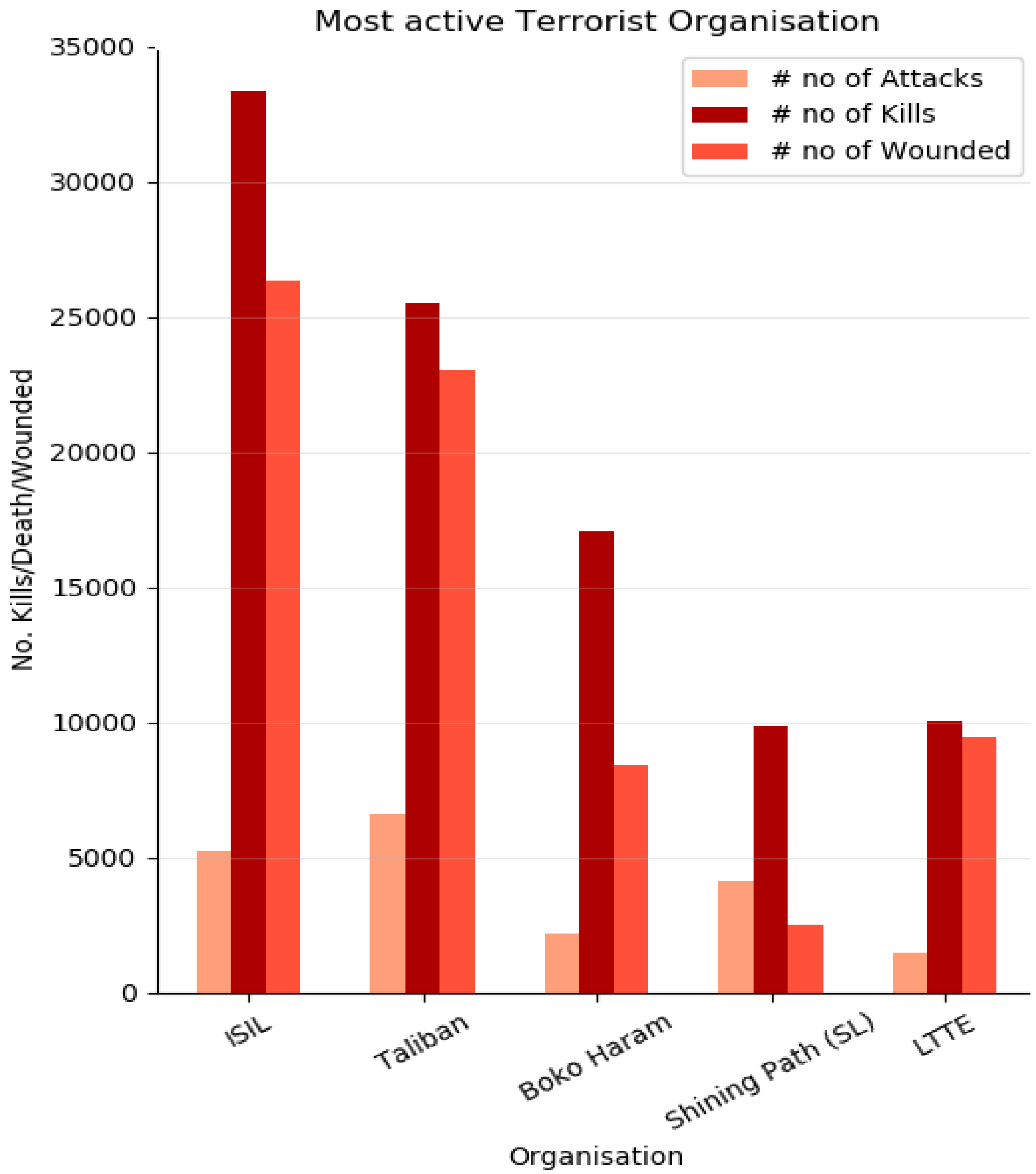
```
import pandas as pd
import matplotlib.pyplot as plt
from collections import defaultdict

data = pd.read_csv('dataset/terror_data.csv')
dic = defaultdict(int)
for i in range(len(data)):
    dic[data.iloc[i, 14]] += data.iloc[i, 18]
org = [i[0] for I in sorted(dic.items(), reverse=True,
key=lambda x: x[1])[1:6]]
cols = ['count', 'kills', 'wounded']
df = pd.DataFrame(0, index=org, columns=cols)
for i in range(len(data)):
    if data.iloc[i, 14] in org and data.iloc[i, 21] in cols:
        df.loc[data.iloc[i, 14], 'count'] += 1
        df.loc[data.iloc[i, 14], 'kills'] += data.iloc[i, 18]
        df.loc[data.iloc[i,14], 'wounded'] += data.iloc[i, 19]
        df.loc[data.iloc[i, 14], data.iloc[i, 21]] += 1
cols = ['count', 'kills', 'wounded']
org = ['ISIL', 'Taliban', 'Boko Haram', 'Shining Path
(SL)','LTTE']
xpos = np.arange(len(df)); plt.xticks(xpos, org, rotation=30)
plt.ylabel('No. Kills/Death/Wounded')
plt.xlabel('Organisation')
plt.title('Most active Terrorist Organisation')
plt.bar(xpos-0.2, df['count'], width=0.2, label='# no of
Attacks', color='#ff9f7a')
plt.bar(xpos, df['kills'], width=0.2, label='# no of Kills',
color='#ad0000')
plt.bar(xpos+0.2, df['wounded'], width=0.2, label='# no of
Wounded', color='#ff513a')
plt.legend(); plt.grid(axis='y', alpha=.3)
plt.show()
```

Contd...

Inference: *ISIL, Taliban, Boko Haram remains the deadliest terror group*

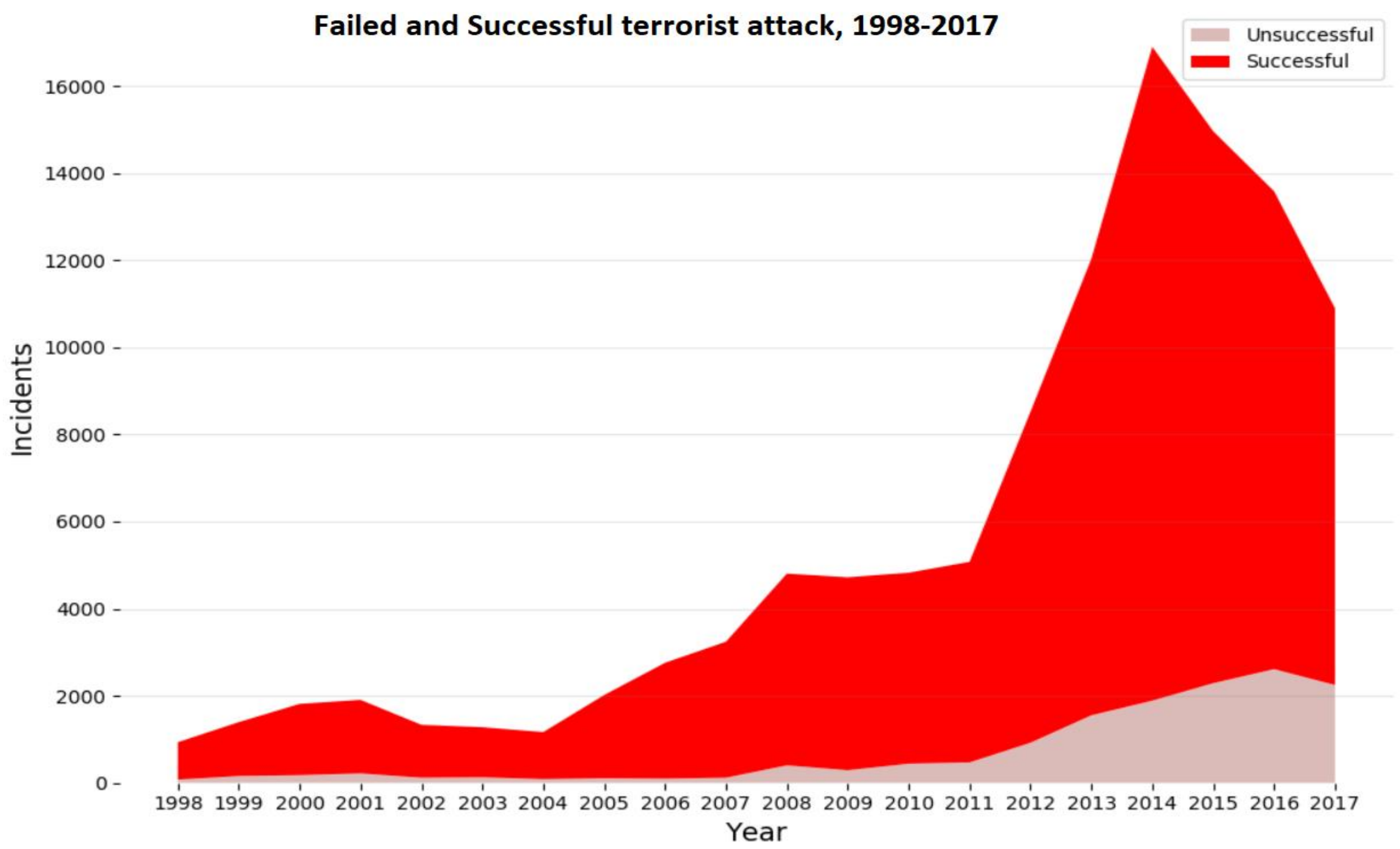
s

- ISIL, Taliban, Boko Haram, Shining Path, LTTE combined caused About 63% of overall fatalities.
- Top three organisation is most active in Middle East and is fuelled by Islamic Extremists.
- The remaining two organisation is communist party and liberation Group.
- ISIL is active in ten countries in 2017.
- The deadliest attack committed by the Taliban was from a suicide explosion in Gardez, Paltika, killing 74 people and injuring and additional 236 people.
- Boko Haram has specialised in maximum-impact bombings and explosions since its initial insurgency in 2009.



Failed and Successful terrorist attack, 1998-2017

In 2002, eight per cent of all terrorist attacks were unsuccessful.



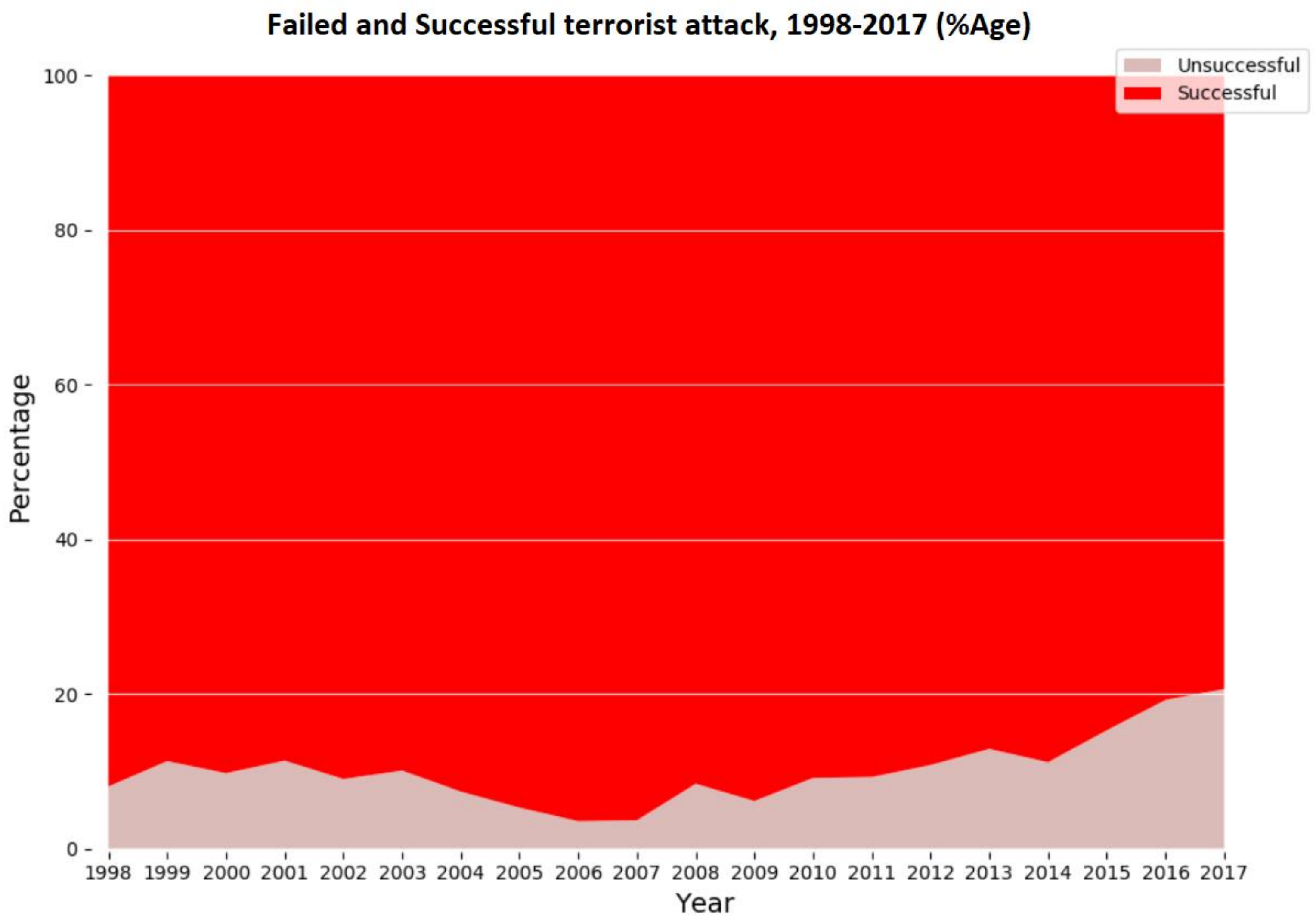

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from collections import defaultdict

data = pd.read_csv('dataset/terror_data.csv')
year = [i for i in range(1998, 2018)]
df = pd.DataFrame(0, index=year, columns=['failed', 'success',
'suc_per', 'fail_per'])
for i in range(len(data)):
    if 1998 <= data.iloc[i, 0] <= 2017:
        if data.iloc[i, 23]:
            df.loc[data.iloc[i, 0], 'success'] += 1
        else:
            df.loc[data.iloc[i, 0], 'failed'] += 1
for i in range(len(df)):
    df.iloc[i,2]=df.iloc[i,1]/(df.iloc[i,0]+df.iloc[i,1])*100
    df.iloc[i, 3] = 100 - df.iloc[i, 2]
xpos = np.arange(1998, 2018)
plt.xticks(xpos, year)
plt.ylabel('Incidents', fontdict={'size': 14})
plt.stackplot(xpos, df['failed'], df['success'],
colors=['#dbbbb8', 'red'], labels=['Unsuccessful',
'Successful'])
plt.xlabel('Year', fontdict={'size': 14})
plt.grid(axis='y', alpha=.2, color='gray')
plt.gca().spines["top"].set_alpha(.0)
plt.gca().spines["bottom"].set_alpha(.0)
plt.gca().spines["right"].set_alpha(.0)
plt.gca().spines["left"].set_alpha(.0)
plt.legend()
plt.show()
```

Contd...

Inference: *Almost 20% of attempted terrorist attacks in between 1998-2017 failed.*

- The percentage of failed attacks remaining under ten percent for every year bar one until 2012.
- In 2017, just over 20 per cent of attacks were unsuccessful, rising from just over 12 per cent in 2014, the year in which the highest number of total attacks were recorded.

```
plt.xticks(xpos, year)
plt.ylabel('Percentage', fontdict={'size': 14})
plt.xlabel('Year', fontdict={'size': 14})
plt.stackplot(xpos, df['fail_per'], df['suc_per'],
colors=['#dbbbb8', 'red'], labels=['Unsuccessful',
'Successful'])
plt.grid(axis='y', alpha=.8, color='white')
plt.gca().spines["top"].set_alpha(.0)
plt.gca().spines["bottom"].set_alpha(.0)
plt.gca().spines["right"].set_alpha(.0)
plt.gca().spines["left"].set_alpha(.0)
plt.legend()
plt.show()
```



Attack pattern Continent-wise

The trend of terror related deaths and wounded remains proportional to the size of the Continent and its population, with an exception of Asia where deaths and and wounded contributes about 40% and 50% respectively.

This effect can be understood by the rising effect of terror groups of Middle East. Those terror group like ISIL, Taliban combinedly contributed about 62% of total fatalities.

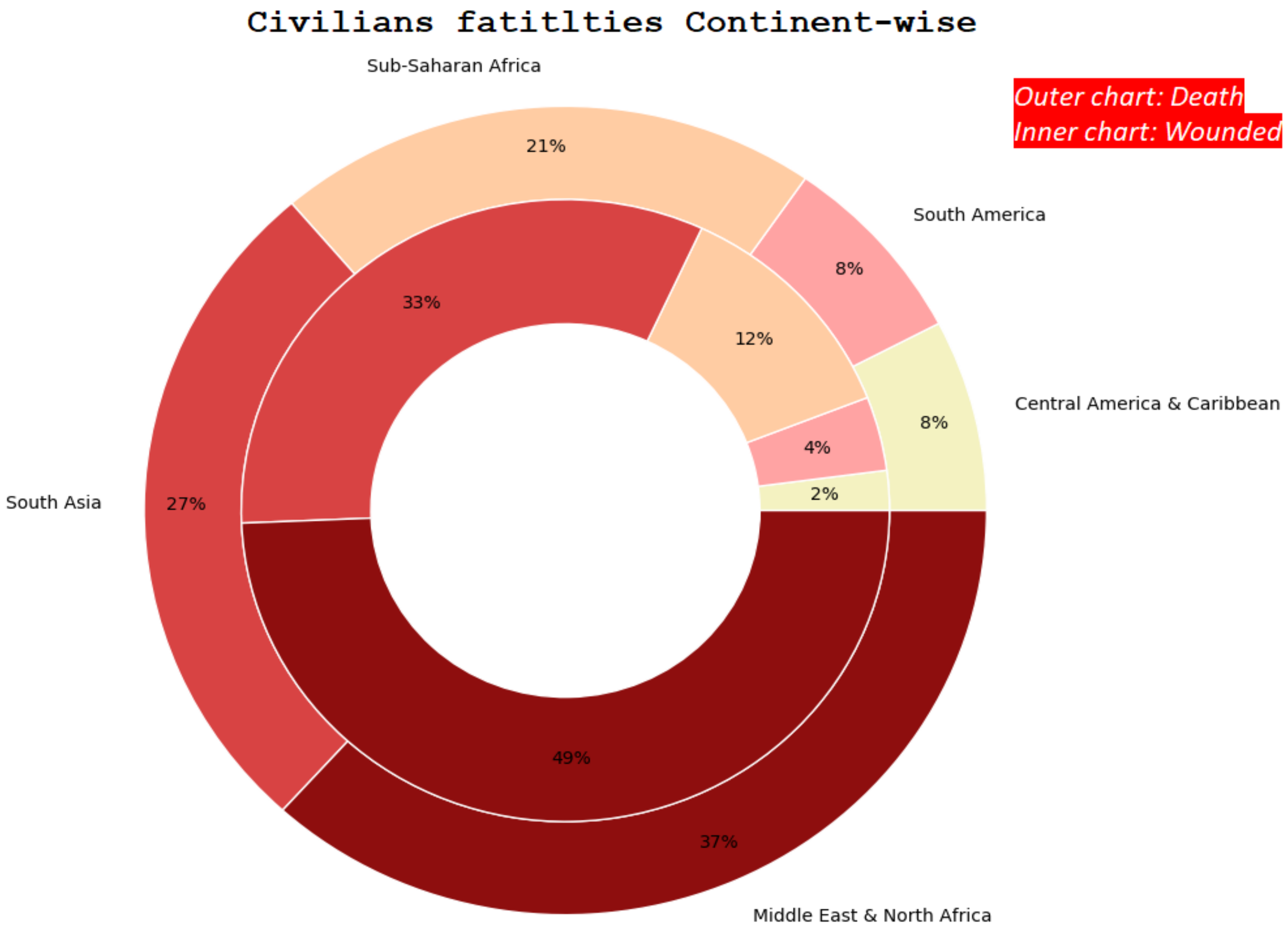
North America has had a higher proportion of attacks directed at infrastructure than other regions, owing to attacks by ecoterrorist organisations. However, these attacks were not responsible for any deaths.

In the past decade there has been very little activity from eco-terrorist groups, with a concurrent change in the predominant type of terrorism

```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import pyplot

empdf = pd.read_csv('terror_data.csv')
continent = ['Central America & Caribbean', 'North America',
'Southeast Asia', 'Western Europe', 'East Asia', 'South
America', 'Eastern Europe', 'Sub-Saharan Africa', 'Middle East
& North Africa', 'Australasia & Oceania', 'South Asia',
'Central Asia']
df=pd.DataFrame(0,index=continent,columns=['kill','wound'])
for i in range(len(empdf)):
    if empdf.iloc[i,7] in continent:
        df.loc[empdf.iloc[i,7], 'kill']+=empdf.iloc[i, 18]
        df.loc[empdf.iloc[i, 7], 'wound'] += empdf.iloc[i, 19]
df1=df.sort_values(by=['kill'])[-5:]
fig, ax = plt.subplots(); ax.axis('equal')
mypie, _, juck = ax.pie(df1['kill'], radius=1.3,
labels=df1.index.values, autopct='%1.0f%%',colors=['#f4f2c1',
'#ffa3a3', '#ffcca3', '#d84343', '#8e0e0e'])
plt.setp(mypie, width=0.3, edgecolor='white')
mypie2, _, juck = ax.pie(df1['wound'], radius=1.3 - 0.3,
autopct='%1.0f%%', colors=['#f4f2c1', '#ffa3a3', '#ffcca3',
'#d84343', '#8e0e0e'])
plt.setp(mypie2, width=0.4, edgecolor='white')
plt.show()
```

Contd...

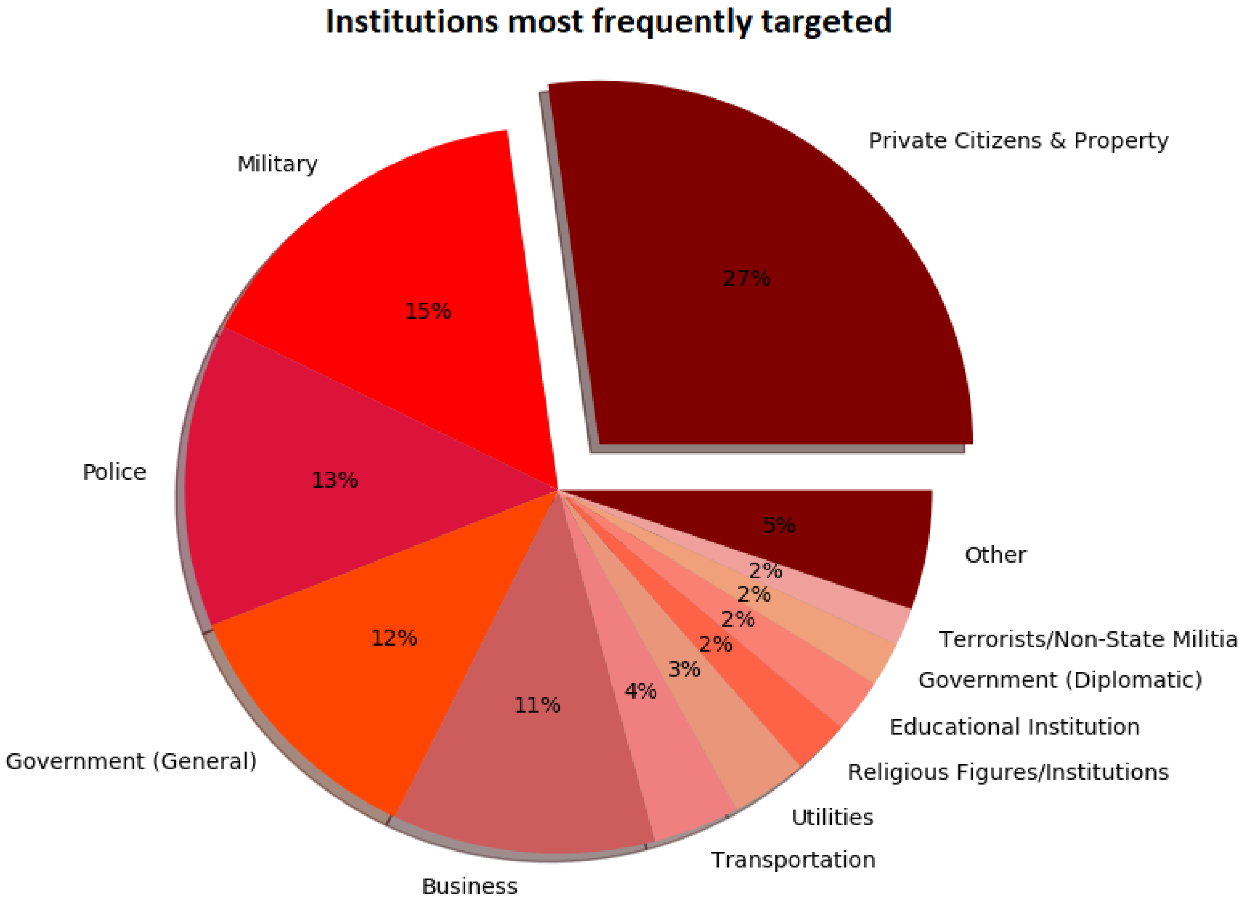
Inference: *Fatality trend Continent-wise remains proportional to its population, with Pacific Asia contributing the most.*



Institutions most frequently targeted

Inference: *Private Citizens & Property remains the primary targets of terrorists to disrupt peace.*

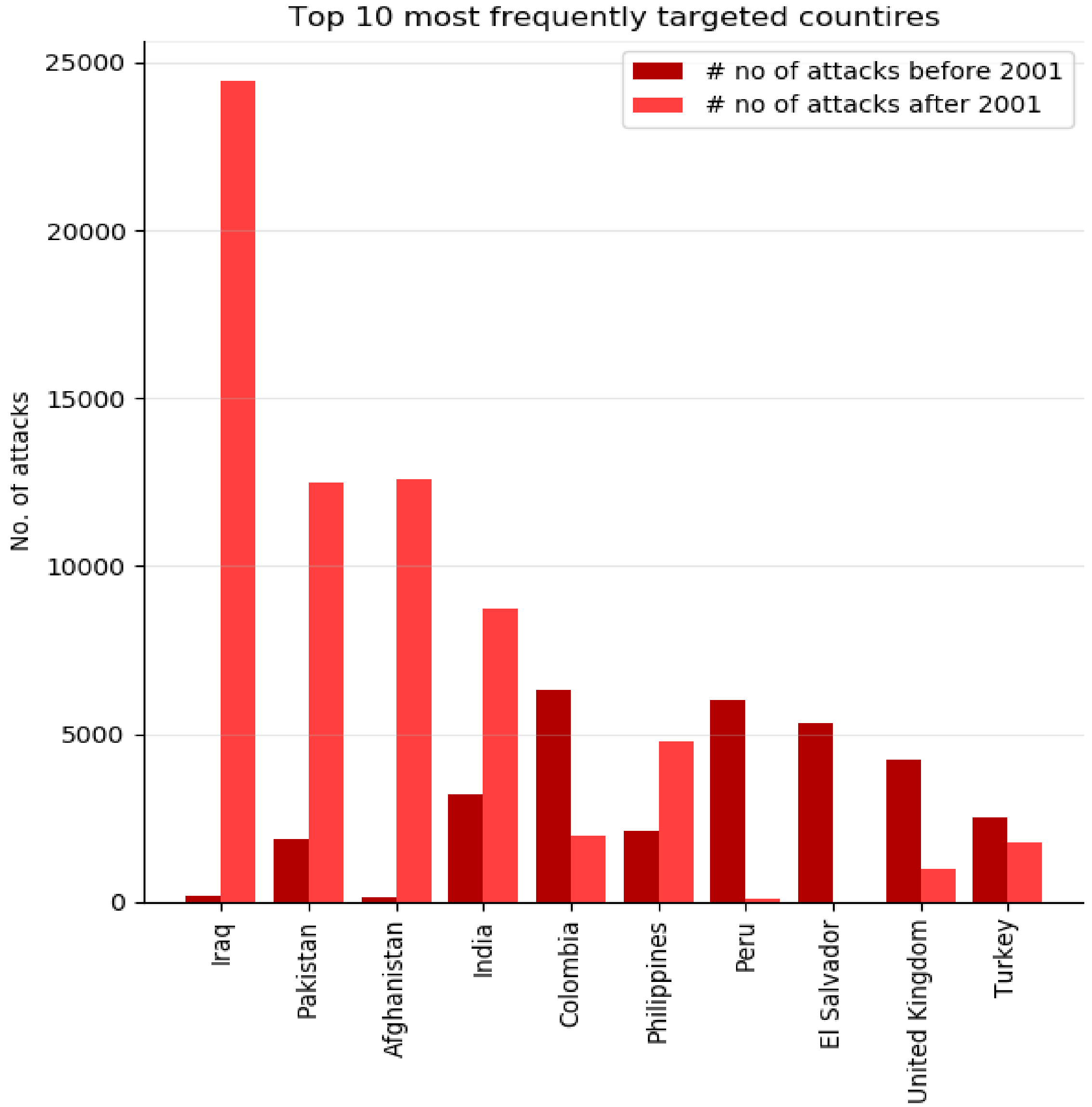
```
import pandas as pd
import matplotlib.pyplot as plt
from collections import defaultdict

dic = defaultdict(int)
data = pd.read_csv('dataset/terror_data.csv')
for i in data['targ_name']:
    dic[i] += 1
tup =
df = pd.DataFrame(tup)
for i in range(len(df) - 1, 11, -1):
    df.iloc[i-1,0]=df.iloc[i,0]; df.iloc[i-1,1]+=df.iloc[i,1]
df.drop([12, 13, 14, 15, 16, 17, 18, 19, 20], inplace=True)
plt.axis('equal')
plt.pie(df[1], labels=df[0], radius=1.2, autopct='%0.0f%%',
explode=[0.2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], shadow=True,
colors=['#800000', '#FF0000', '#DC143C', '#FF4500','#CD5C5C',
'#F08080', '#E9967A', '#FF6347', '#FA8072', '#F0A07A',
'#F0A09B'])
plt.show()
```



Ten countries most frequently targeted, 1970-2000 to 2001-2017

Inference: *Most of the countries witnessed exponential jump in incidents at the beginning of 21st century.*

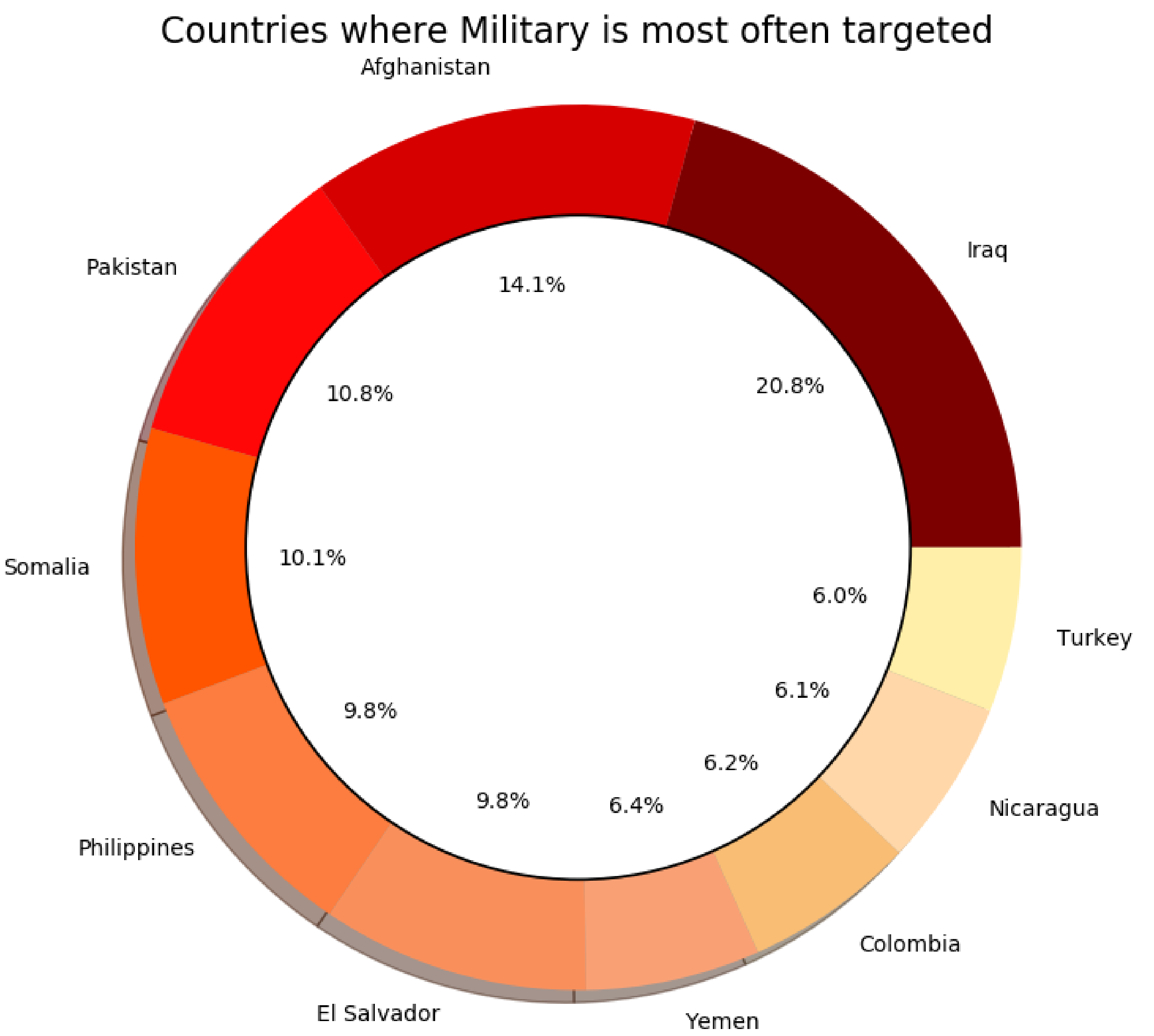
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from collections import defaultdict

data = pd.read_csv('dataset/terror_data.csv')
dic = defaultdict(int); li1 = []; li2 = []
for i in range(len(data)):
    dic[data.iloc[i, 4]] += 1
tup = sorted(dic.items(), reverse=True, key=lambda x: x[1])
df = pd.DataFrame(tup[0: 10]); del df[1]
for i in df[0]:
    sum1 = 0; sum2 = 0
    for j in range(len(data)):
        if data.iloc[j, 4] == i and data.iloc[j, 0] >= 2001:
            sum2 += 1
        elif data.iloc[j, 4] == i:
            sum1 += 1
    li1.append(sum1); li2.append(sum2)
df[1] = li1; df[2] = li2
xpos = np.arange(len(df))
plt.xticks(xpos, df[0], rotation='vertical')
plt.ylabel('No. of attacks')
plt.title('Top 10 most frequently targeted countires')
plt.bar(xpos - 0.2, df[1], width=0.4, label='# no of attacks before 2001', color='#b20000')
plt.bar(xpos + 0.2, df[2], width=0.4, label='# no of attacks after 2001', color='#ff3f3f')
plt.legend(); plt.grid(axis='y', alpha=.3)
plt.show()
```



Countries where military bases are most prone

Inference: *These 10 countries combinedly contributed to about 73% of total attacks on Military.*

```python
import pandas as pd
import matplotlib.pyplot as plt
from collections import defaultdict

empdf = pd.read_csv('terror_data.csv')
country=defaultdict(int)
for i in range(len(empdf)):
    if empdf.iloc[i,13] =='Military':
        country[empdf.iloc[i,4]]+=1
tup=sorted(country.items(),reverse=True,key=lambda x:x[1])
df1=pd.DataFrame(tup[0:10])
plt.pie(df1[1],labels=df1[0],autopct='%1.1f%%',shadow=True, colors=['#7c0000',
'#d60000', '#ff0707', '#ff5400', '#fc7d3f', '#f9905c', '#f9a175', '#f9bd74',
'#ffd7a8', '#ffefa8'])
centre_circle=plt.Circle((0,0),0.75,color='black',fc='white',linewidth=1.25)
fig=plt.gcf()
plt.title("Countries where Military is most often targeted",
fontdict={'size':16})
fig.gca().add_artist(centre_circle)
plt.axis('equal')
plt.show()
```



Five countries most often targeted, 2002-2017

This analysis highlights the top five countries most impacted by terrorism according to the in year 2002-2017 and how they have ranked on the Global Terrorism Index since its inception in 2002.

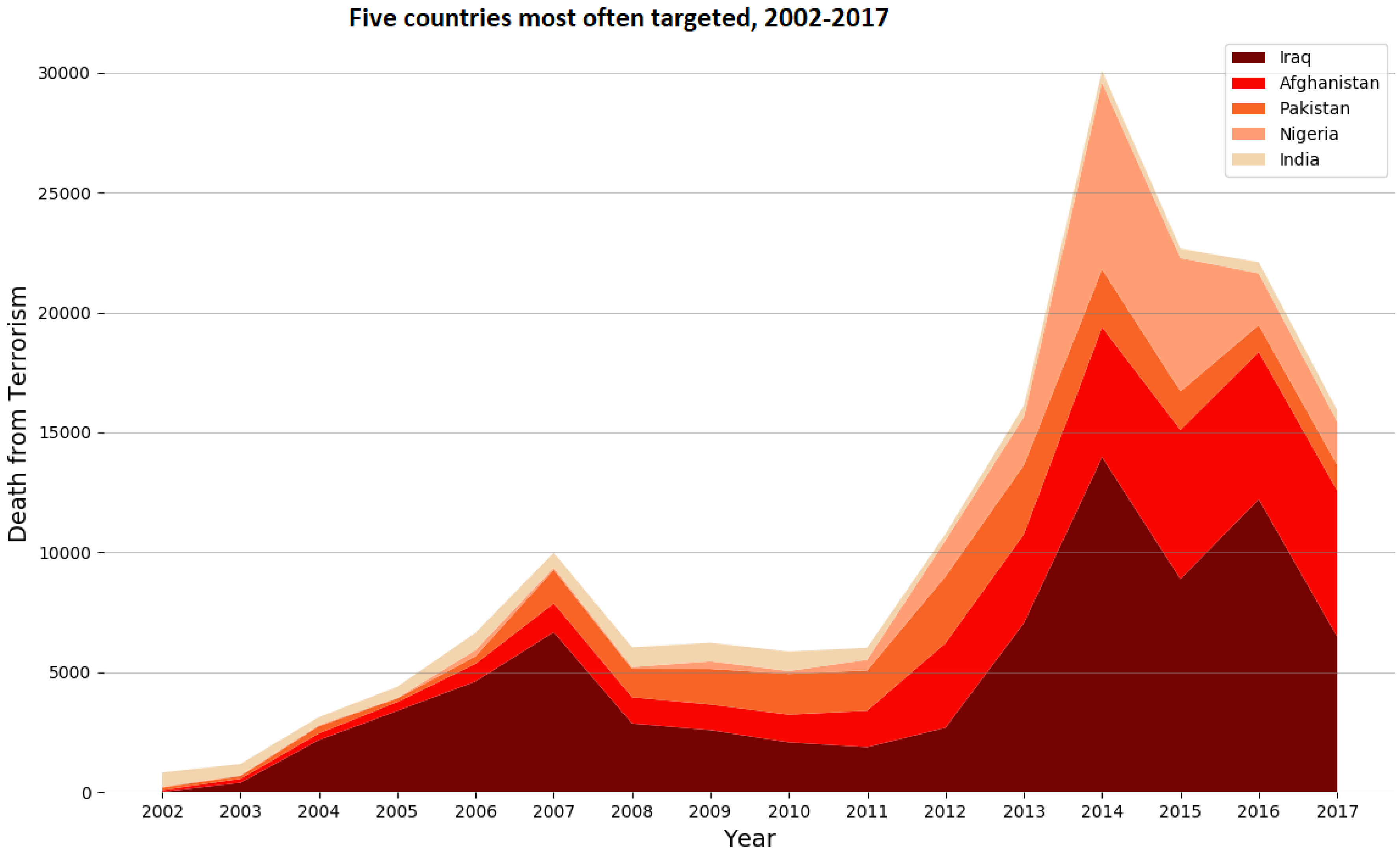
Conflict continued to be the primary driver of terrorist activity for the countries most impacted by terrorism in this period. Four out of five countries were classified as being in the state of war whereas India involved in multiple minor conflicts.

```
import pandas as pd
import numpy as np
from collections import defaultdict
import matplotlib.pyplot as plt

df=pd.read_csv('terror_data.csv')
dict = defaultdict(int)
for i in range(len(df)):
    dict[df.iloc[i, 4]] += df.iloc[i, 18]
tup = sorted(dict.items(), reverse=True, key=lambda x:x[1])[0:5]
year = [i for i in range(2002, 2018)]
country = [i[0] for i in tup[0:5]]; df = pd.DataFrame(0, index=year, columns=country)
for i in range(len(df)):
    if 2002 <= df.iloc[i, 0] <=2017 and df.iloc[i, 4] in country:
        df.loc[df.iloc[i, 0], df.iloc[i, 4]] += df.iloc[i, 18]
xpos=np.arange(len(df))
plt.xticks(xpos, year); plt.ylabel('Death from Terrorism', fontdict={'size': 14})
plt.xlabel('Year', fontdict={'size': 14}); plt.stackplot(xpos, df['Iraq'],
df['Afghanistan'],df['Pakistan'],df['Nigeria'],df['India'],colors=['#750301','#f70602','#f76425','
#ff9e75','#f2d4ae'], labels=['Iraq', 'Afghanistan','Pakistan','Nigeria','India'])
plt.grid(axis='y', alpha=.5, color='gray')
plt.gca().spines["top"].set_alpha(.0); plt.gca().spines["bottom"].set_alpha(.0);
plt.gca().spines["right"].set_alpha(.0); plt.gca().spines["left"].set_alpha(.0); plt.legend()
plt.show()
```

Contd...

Inference: *Iraq and Afghanistan maintained the top position in death incurred since 2007. Developing and under-developed country remains catalyst of terrorism.*

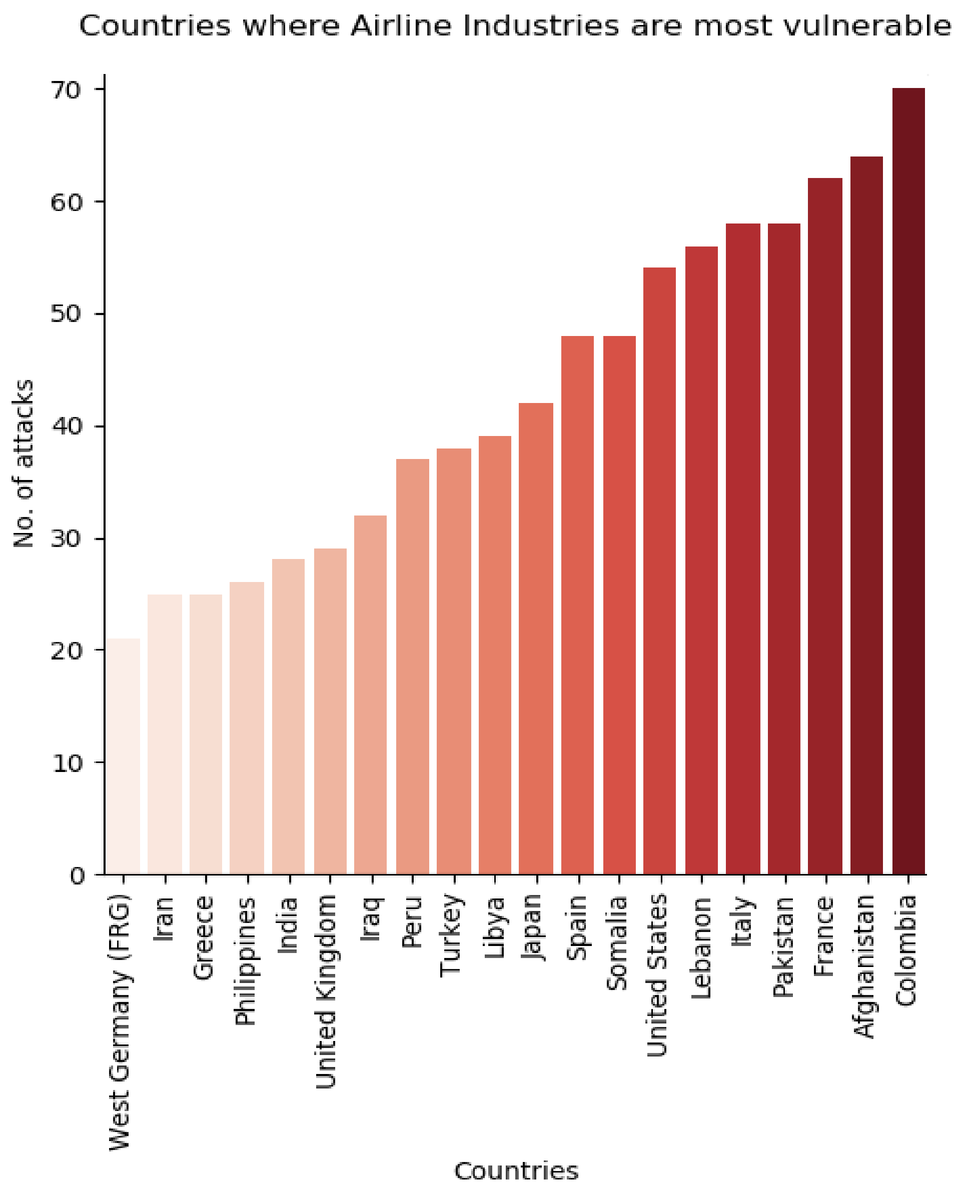
- Iraq remains the country most impacted by terrorism, a position it has held since 2014.
- In 2017, Afghanistan was the country that recorded the highest number of deaths from terrorism, replacing Iraq which had held the position since 2013.
- Terrorism increased substantially in the FATA and Sindh regions in Pakistan, with deaths increasing by 117 and 104 per cent respectively in recent years.



Countries where Airline Industry is highly vulnerable

Inference: *Most of the countries already struggling with terrorism has it's Airline Industry vulnerable.*

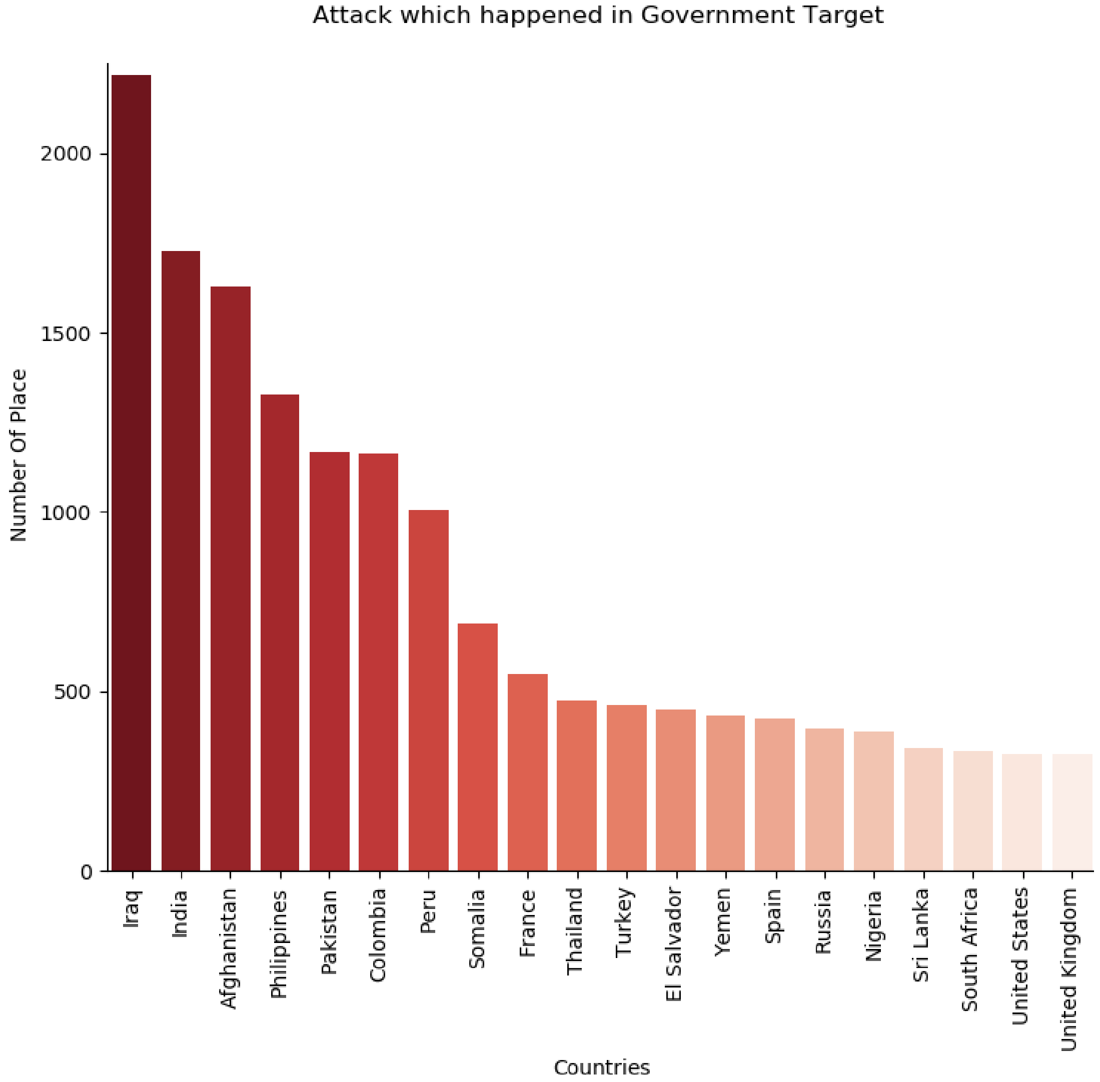
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from collections import defaultdict

df1 = pd.read_csv('dataset/terror_data.csv')
dict=defaultdict(int)
for i in range(len(df1)):
    if df1.iloc[i,13] == 'Airports & Aircraft':
        dict[df1.iloc[i,4]]+=1
df = pd.DataFrame(sorted(dict.items(),reverse=False,key=lambda x:x[1]) [-20:])
ax = sns.barplot(x=df[0], y=df[1], data=df, palette='Reds')
plt.xticks(rotation='vertical')
plt.title('Countries where Airline Industries are most vulnerable')
ax.set(xlabel='Countries', ylabel='No. of attacks')
plt.show()
```



Countries where government institute is most often targeted

Inference: *Developing and under-developed countries yet again remains the main cause. Including with already sky-rocketing increase terror related incidents in 21st century.*

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from collections import defaultdict

empdf = pd.read_csv('terror_data.csv')
dict = defaultdict(int)
for i in range(len(empdf)):
    if empdf.iloc[i,13] == 'Government (General)':
        dict[empdf.iloc[i,4]]+=1
df=pd.DataFrame(sorted(dict.items(),reverse=True,key=lambda x:x[1])[0:20])
ax = sns.barplot(x=df[0], y=df[1], data=df, palette='Reds_r')
plt.xticks(rotation='vertical')  # change to vertical
plt.title('Attack which happened in Government Target ')
ax.set(xlabel='Countries', ylabel='Number Of Place')
plt.show()
```



Three most frequently targeted Indian states, 1997-2017

Terrorism poses a significant threat to people of India.

Terrorism found in India includes ethno-nationalist terrorism,

Religious terrorism, left wing terrorism and narco terrorism.

South Asia Terrorism Portal has listed 180 terrorist groups

That have operated within India over the last 20 years,

Many of them co-listed as transnational terror networks

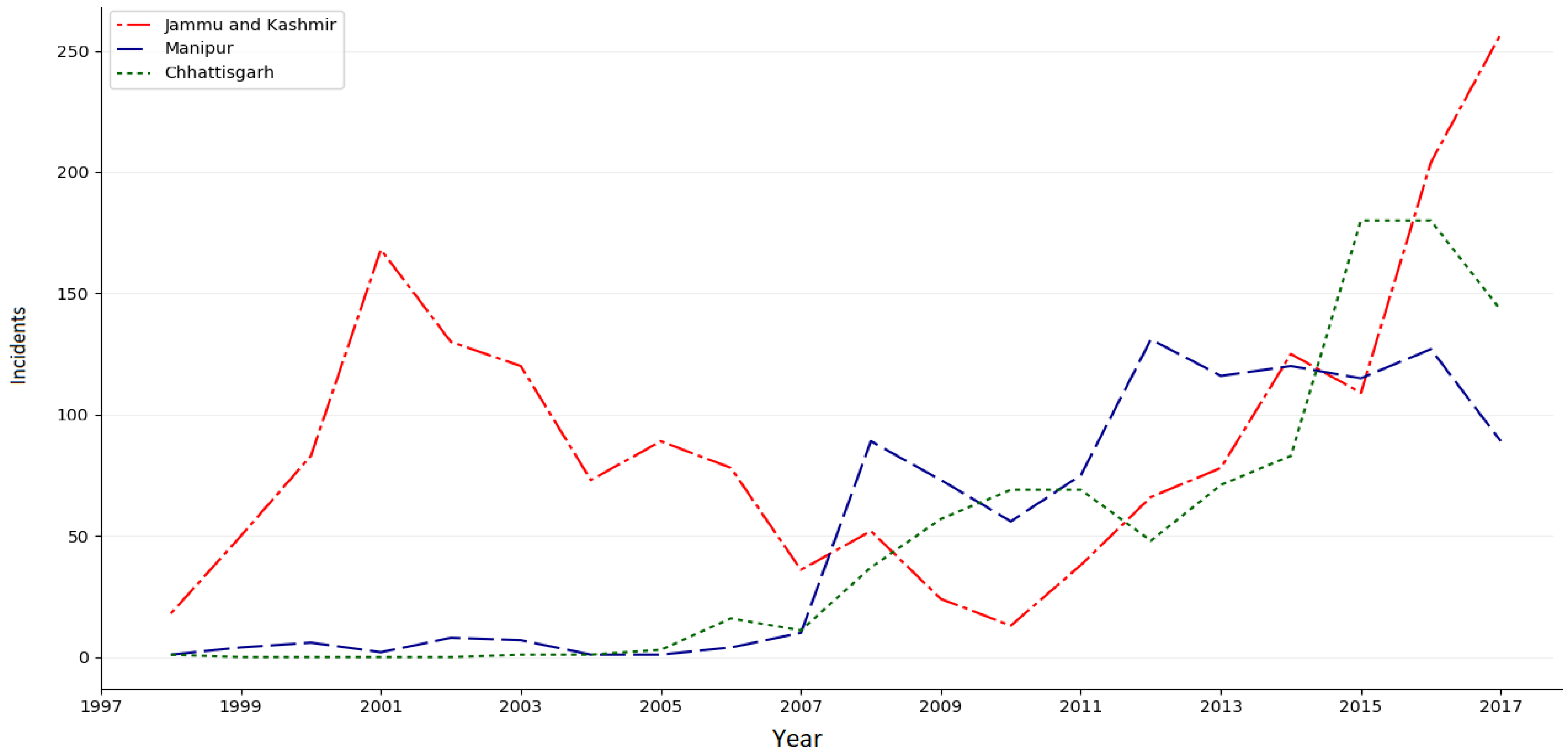
From neighbouring countries like Pakistan, Bangladesh, etc.

```
import pandas as pd
import matplotlib.pyplot as plt
from collections import defaultdict

empdf = pd.read_csv('dataset/terror_data.csv')
dict = defaultdict(int)
for i in range(len(empdf)):
    if empdf.iloc[i,4] == 'India' and empdf.iloc[i,0]>1996:
        dict[empdf.iloc[i,8]]+=1
df=pd.DataFrame(sorted(dict.items(),reverse=True,key=lambda x:x[1])[0:3])
states = [i for i in df.loc[0:3, 0]]
df1 = pd.DataFrame(0, index=np.arange(1998, 2018), columns=states)
for i in range(len(empdf)):
    if empdf.iloc[i, 0] > 1997 and empdf.iloc[i, 8] in states:
        df1.loc[empdf.iloc[i, 0], empdf.iloc[i, 8]] += 1
li = [int(i) for i in range(1998, 2018)]; fig, ax = plt.subplots()
plt.title('Top 3 Indian states most frequently targeted since 1997')
l1, = ax.plot(df1['Jammu and Kashmir'], dashes=[2, 2, 10, 2], label='Jammu and
Kashmir', color='red')
l2, = ax.plot(df1['Manipur'], dashes=[10, 3], label='Manipur', color='darkblue')
l3, = ax.plot(df1['Chhattisgarh'], dashes=[2, 2], label='Chhattisgarh',
color='darkgreen')
plt.xticks(range(1997, 2018, 2))
plt.grid(axis='y', alpha=.2); ax.legend()
plt.show()
```

Contd...

Inference: *States opposing government's policies tend to pick up arms to disrupt peace, like separatists, Maoist.*

Top 3 Indian States most frequently targeted since 1997



Rate of change in terrorism India, 2002-2017

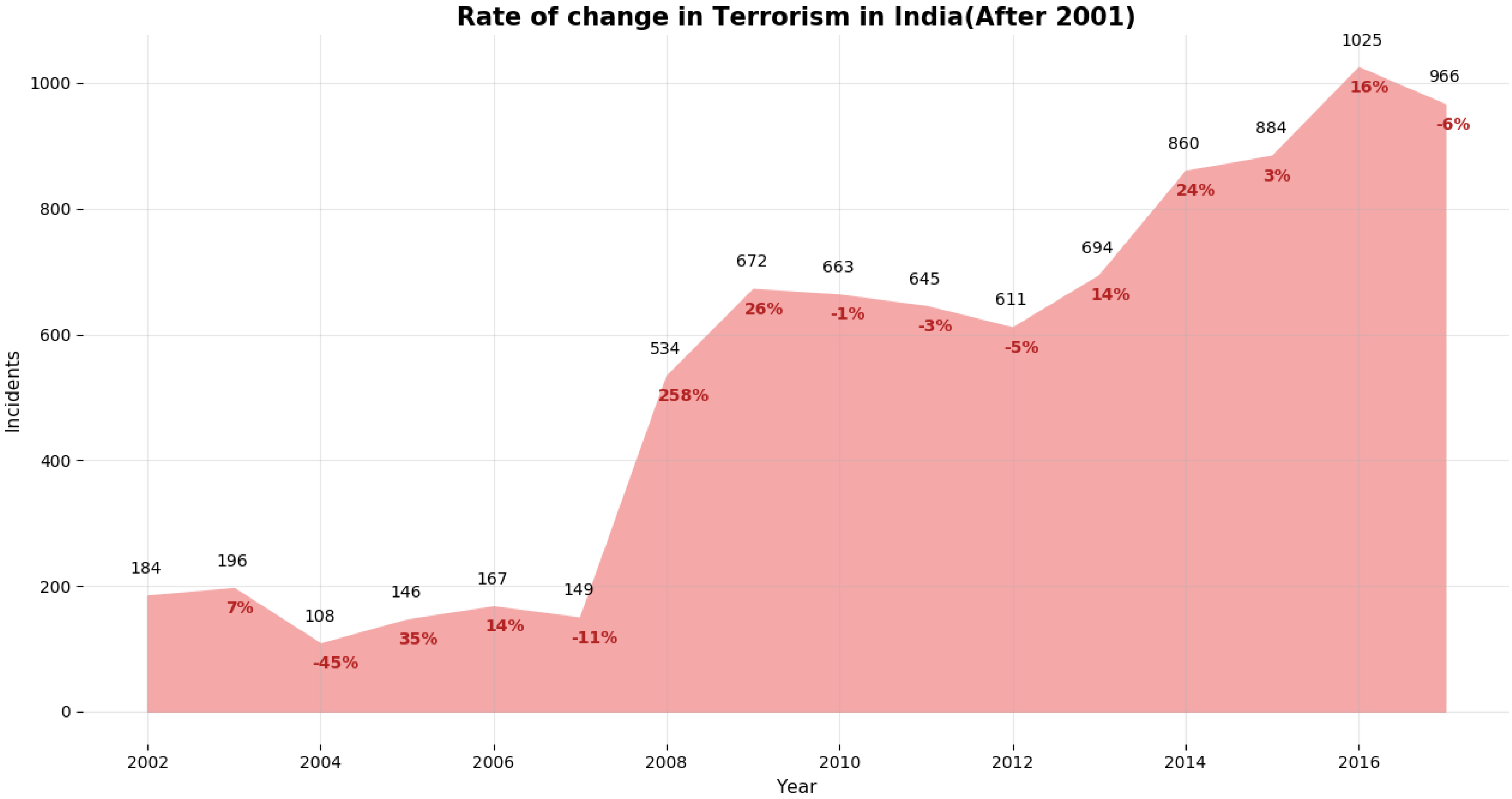
Deaths from terrorism in India rose to 384 in 2017, a 12 per cent increase. India is now ranked seventh on the Global Terrorism Index.

The scope of terrorism and violent conflict in India is particularly broad, with 51 different terrorist's groups being responsible for at least one terrorist attack in 2017 and 25 groups being responsible for at least one terrorism death.

```python
import pandas as pd
from collections import defaultdict
import matplotlib.pyplot as plt
from matplotlib import pyplot

data = pd.read_csv('dataset/terror_data.csv')
dic = defaultdict(int)
for i in range(len(data)):
    if data.iloc[i, 0] > 2001 and data.iloc[i, 4] == 'India':
        dic[data.iloc[i, 0]] += 1
li = [[i, dic[i]] for i in dic]
li1 = [str(round(((li[i+1][1]-li[i][1])/li[i][1])*100))+'%' for i in range(len(li)-1)]
li1.insert(0, '')
df = pd.DataFrame(li)
plt.fill_between(df[0], df[1], color='#f4a8a8')
plt.title("Rate of change in Terrorism in India(After 2001)", fontdict={'size': 15,
'weight': 'bold'}); a = 0
for i, j in zip(df[0], df[1]):
    pyplot.text(i-0.105, j-40, li1[a],fontdict={'color': '#B22222', 'weight': 'bold'})
    pyplot.text(i-0.2, j+35, str(j)); a += 1
plt.xlabel('Year',fontdict={'size':11}); plt.ylabel('Incidents',fontdict={'size': 11})
plt.gca().spines["top"].set_alpha(.0); plt.gca().spines["bottom"].set_alpha(.0)
plt.gca().spines["right"].set_alpha(.0); plt.gca().spines["left"].set_alpha(.0)
plt.grid(axis='both', alpha=.3)
plt.show()
```

Contd...

Inference: *Total incidents saw a significant rise of 425% in year as compared to 2002, indicating higher spread of terrorists in country.*



Civilian's and terrorist's death pattern, 1998-2017

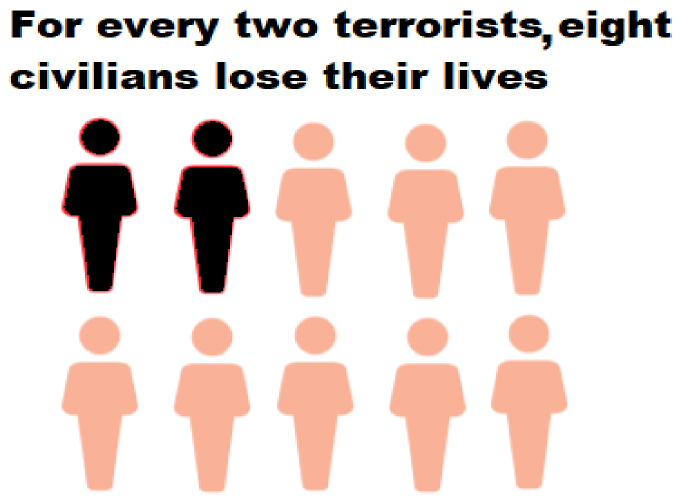
Civilian's and terrorist's death indicate us how strict the government has taken a measure to neutralize the terrorists in order to prevent any further terrorist attacks by implementing a vigilant anti-terrorist squad to neutralize them.

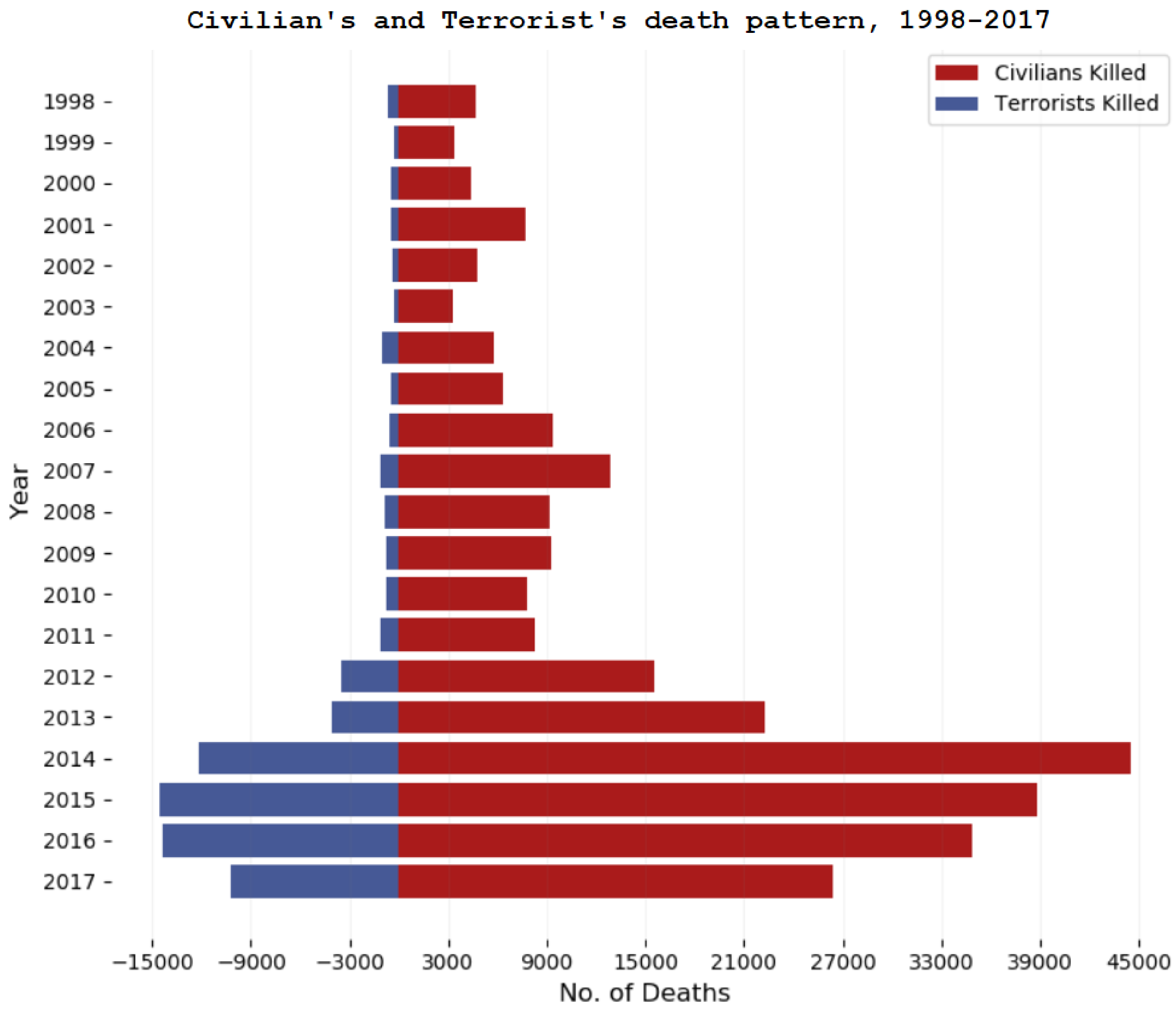
It also indicates how well a country has recovered from any incident and have taken a precautionary measures to prevent multiple such incidents.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

data = pd.read_csv('dataset/terror_data.csv')
year = [i for i in range(2017, 1997, -1)]
df = pd.DataFrame(0, index=year, columns=['kill', 'ter_kill'])
for i in range(len(data)):
    if 1998 <= data.iloc[i, 0] < 2018:
        df.loc[data.iloc[i, 0], 'kill'] += data.iloc[i, 18]
        df.loc[data.iloc[i, 0], 'ter_kill'] += data.iloc[i, 20]
xpos = np.arange(len(df)); plt.yticks(np.arange(len(df)), year)
plt.xticks(np.arange(-15000, 45001, 6000)); plt.grid(axis='x', alpha=.1,
color='gray')
plt.title('Type of Terrorist attack total trend, 2002-2017',fontdict={'size':16})
plt.xlabel('No. of Deaths',fontdict={'size':12});
plt.ylabel('Year',fontdict={'size':12})
plt.barh(xpos, df['kill'], color='#aa1b1b', label='Civilians Killed')
plt.barh(xpos, -df['ter_kill'], color='#465996', label='Terrorists Killed')
plt.gca().spines["top"].set_alpha(.0); plt.gca().spines["bottom"].set_alpha(.0)
plt.gca().spines["right"].set_alpha(.0); plt.gca().spines["left"].set_alpha(.0)
plt.legend()
```

Contd...

Inference: *No mercy policy has been adopted towards terrorism. As the number of civilian's deaths increased so did that of terrorist's, keeping the numbers proportional.*

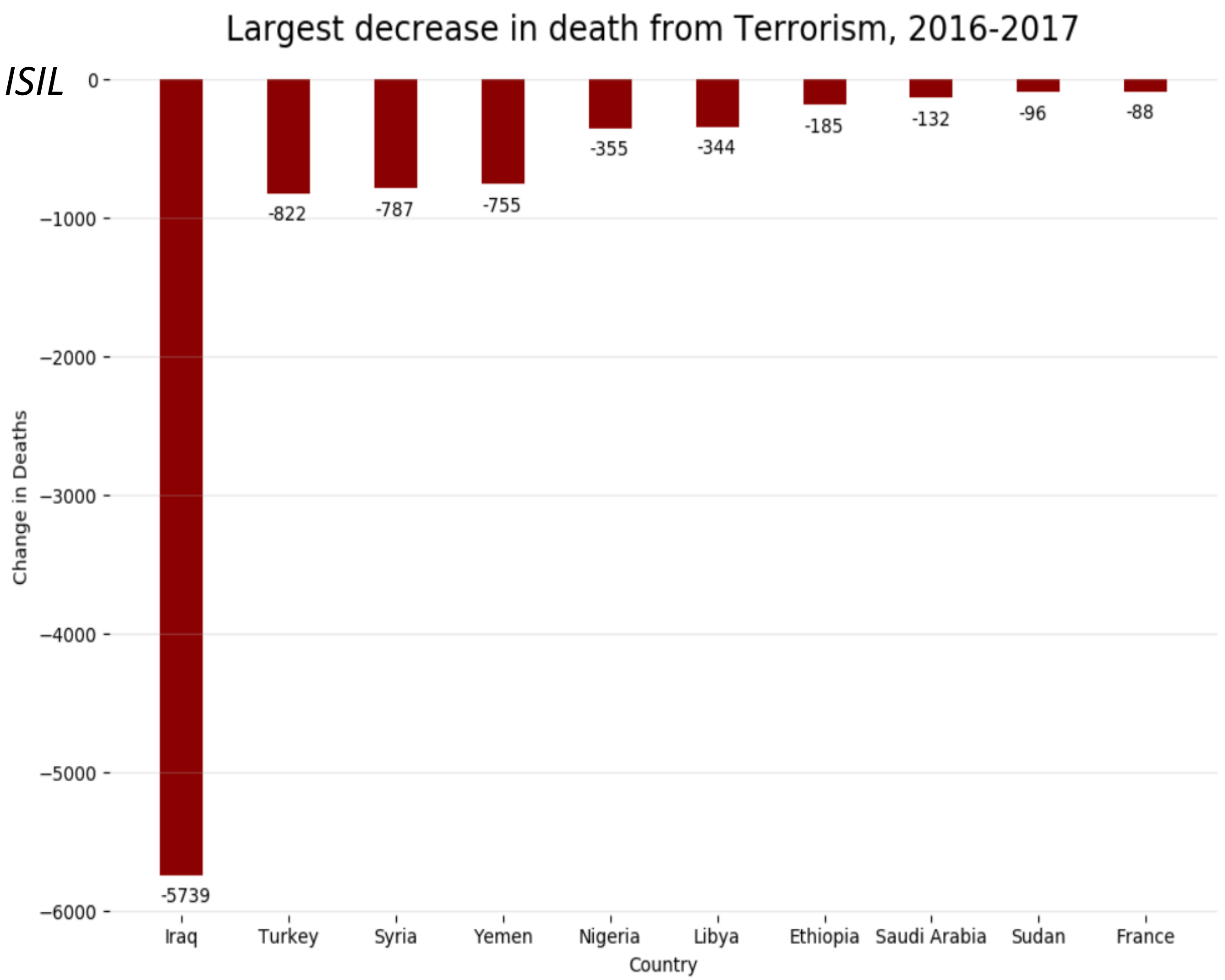




Largest decrease in death from terrorism, 2016-2017

Inference: *Iraq witnessed the largest decrease because of weakening of ISIL*

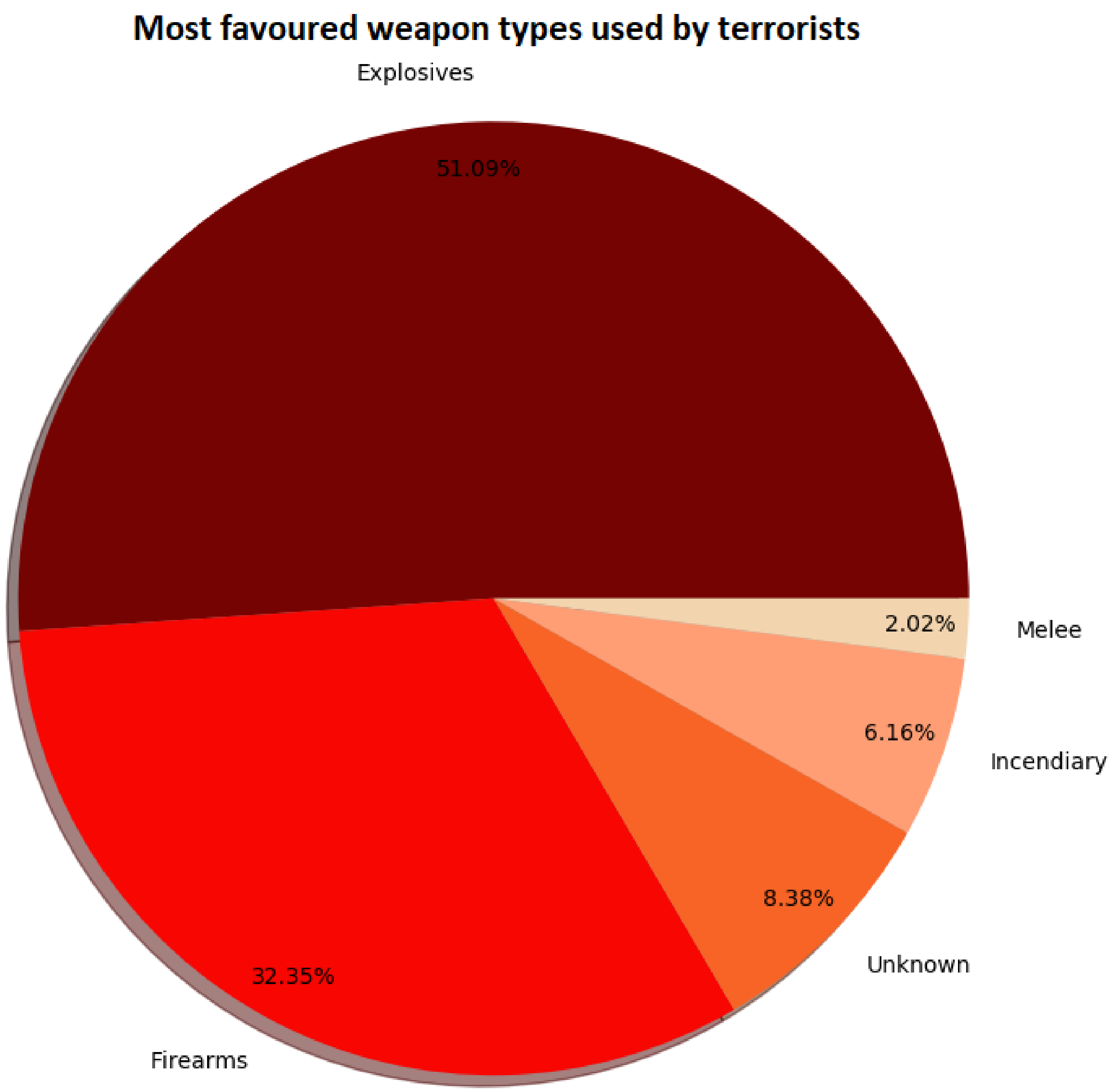
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from collections import  defaultdict
from matplotlib import pyplot

df = pd.read_csv('dataset/terror_data.csv')
year = [int(i) for i in range(1970, 2018)]
dic = defaultdict(int); dic1 = defaultdict(int); dic2 = defaultdict(int)
for i in range(len(df)):
    if df.iloc[i, 0] == 2016:
        dic1[df.iloc[i, 4]] += df.iloc[i, 18]
    elif df.iloc[i, 0] == 2017:
        dic2[df.iloc[i, 4]] += df.iloc[i, 18]
for i in dic1:
    if i in dic1 and i in dic2:
        dic[i] = dic2[i] - dic1[i]
df = pd.DataFrame(sorted(dic.items(), reverse=False, key=lambda x: x[1])[:10])
xpos = np.arange(len(df)); plt.xticks(xpos, df[0]); plt.ylabel('Change in Deaths')
plt.title('Largest decrease in death from Terrorism, 2016-2017',fontdict={'size': 18})
plt.bar(xpos, df[1], width=0.4, color='darkred'); plt.xlabel('Country'); a = -0.2
for j in df[1]:
    pyplot.text(a, j-200, str(j)); a += 1.001
plt.gca().spines["top"].set_alpha(.0); plt.gca().spines["bottom"].set_alpha(.0)
plt.gca().spines["right"].set_alpha(.0); plt.gca().spines["left"].set_alpha(.0);
plt.grid(axis='y', alpha=0.3)
plt.show()
```



Most favoured weapon types used by terrorists

Inference: *Explosives like grenades, bombs are the most favoured weapon because of their high burst capacity and easy to hurl in the public or military places causing significant damage.*

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('terror_data.csv')
val = df['weap_name'].value_counts().head(5)
head=['Explosives','Firearms','Unknown','Incendiary','Melee']
plt.pie(val,labels=head,radius=1.2,autopct='%0.2f%%',
pctdistance=0.9,shadow=True,colors=['#750301','#f70602','#f76425','#ff9e75
','#f2d4ae']); plt.axis("equal")
plt.show()
```



Declining phase of terrorism

Through the number of terrorist attacks did kept increasing in 21st century but the death toll reduced by a noticeable difference, this can be understood by the fact that terrorist’s attacks recently are less fatal as compared to few decades earlier.

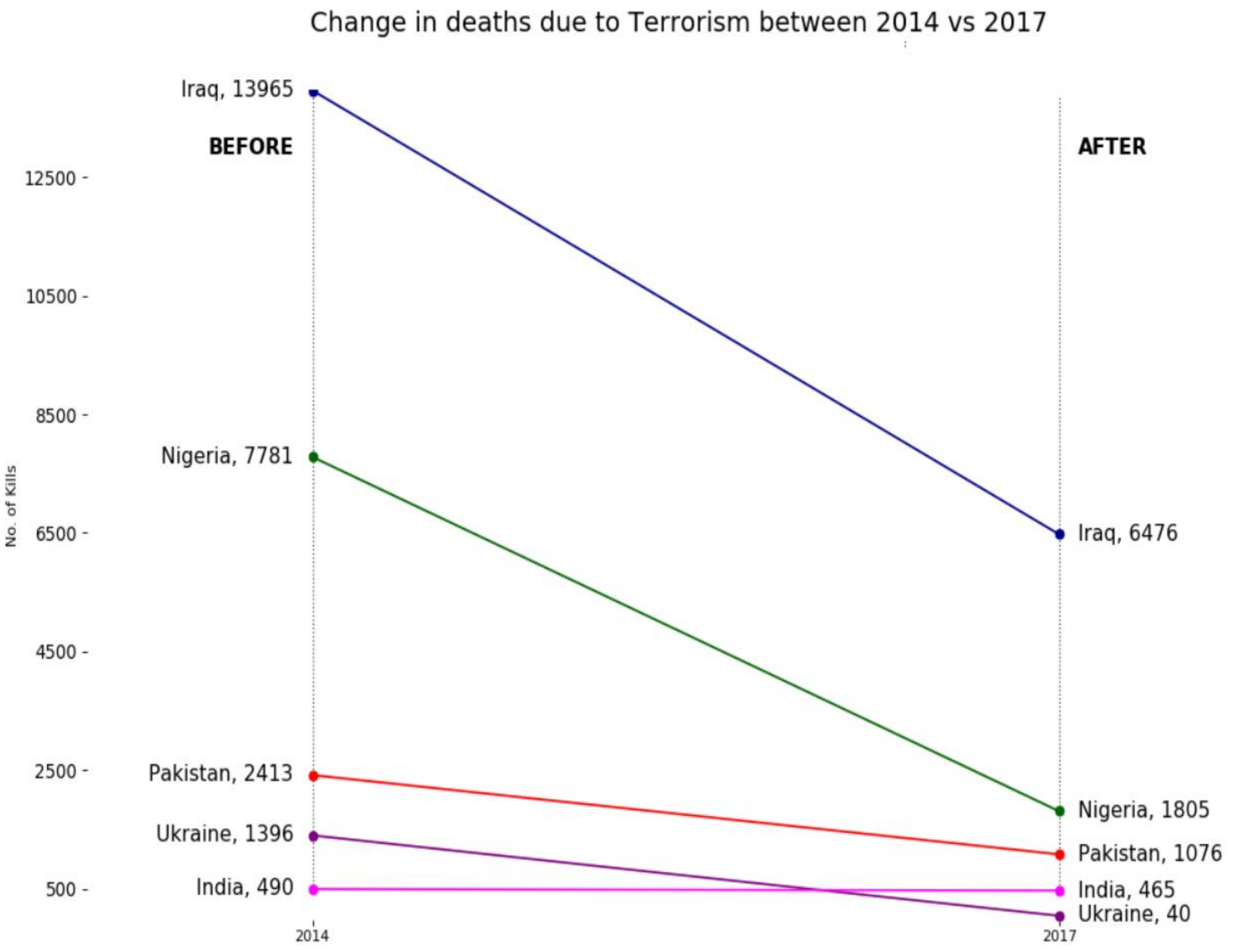
Many terrorist groups like ISIL, Al Qaida has witnessed a huge fall in funding, which caused the weakening of the organisations.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.lines as mlines

data = pd.read_csv('dataset/terror_data.csv')
data1 = pd.read_csv('top10.csv')
li = list(data1.iloc[0:5, 0])
data1 = pd.DataFrame(0, index=li, columns=[2014, 2017])
for i in range(len(data)):
    if data.iloc[i, 4] in li:
        if data.iloc[i, 0] == 2014:
            data1.loc[data.iloc[i, 4], 2014] += data.iloc[i, 18]
        elif data.iloc[i, 0] == 2017:
            data1.loc[data.iloc[i, 4], 2017] += data.iloc[i, 18]
left_label = [str(c) + ', ' + str(round(y)) for c, y in zip(li, data1[2014])]
right_label = [str(c) + ', ' + str(round(y)) for c, y in zip(li, data1[2017])]
klass = ['darkblue', 'darkgreen', 'purple', 'red', 'magenta']; z = 0
def newline(p1, p2, z, color='black'):
    l = mlines.Line2D([p1[0], p2[0]], [p1[1], p2[1]], color=klass[z], marker='o',
markersize=6)
    ax = plt.gca(); ax.add_line(l); return l
fig, ax = plt.subplots(1, 1, figsize=(14, 14), dpi=80)
ax.vlines(x=1,ymin=500,ymax=15000,color='black',alpha=0.7,linewidth=1,linestyles='dotted')
ax.vlines(x=3,ymin=500,ymax=15000,color='black',alpha=0.7,linewidth=1,linestyles='dotted')
ax.scatter(y=data1[2014], x=np.repeat(1, data1.shape[0]), s=10, color='black', alpha=0.7)
ax.scatter(y=data1[2017], x=np.repeat(3, data1.shape[0]), s=10, color='black', alpha=0.7)
```

Contd...

Inference: *The lethality of attacks has decreased as the major terrorist groups weaken since 2014. Most of the country topping the list has shown a significant decrease in fatality.*

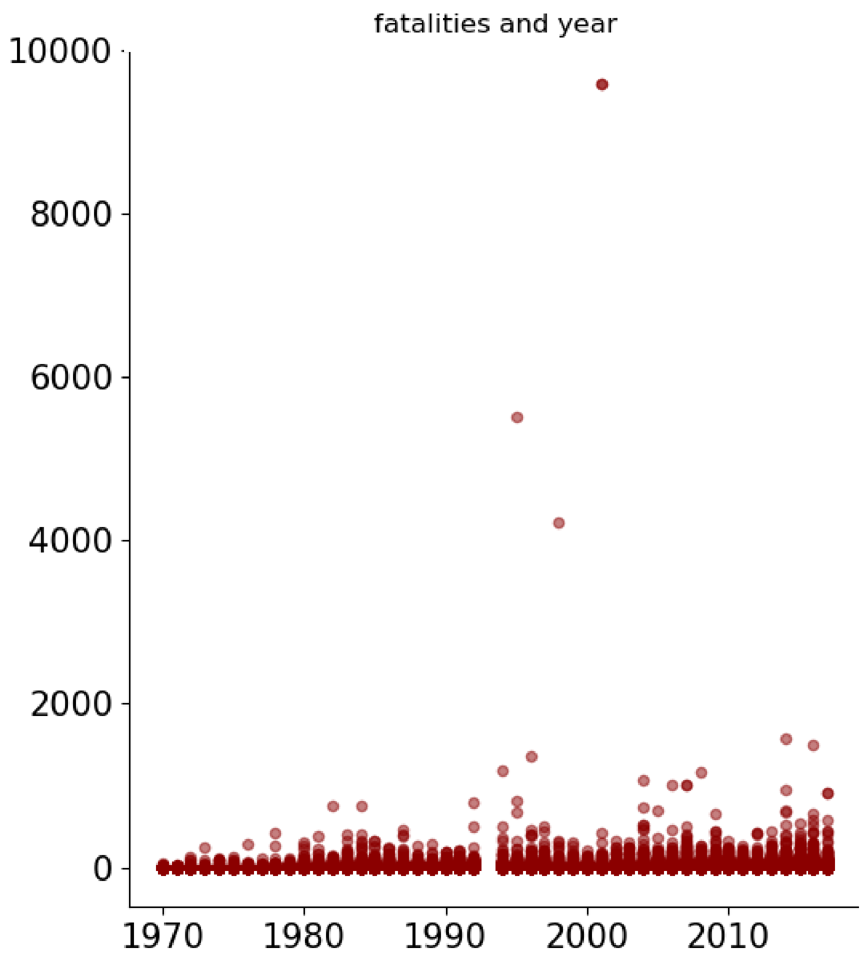
```python
for p1, p2, c in zip(data1[2014], data1[2017], li):
    newline([1, p1], [3, p2], z); z += 1
    ax.text(1-0.05, p1, c + ', ' + str(round(p1)), horizontalalignment='right',
verticalalignment='center', fontdict={'size': 14})
    ax.text(3+0.05, p2, c + ', ' + str(round(p2)), horizontalalignment='left',
verticalalignment='center', fontdict={'size': 14})
ax.text(1-0.05, 13000, 'BEFORE', horizontalalignment='right',

verticalalignment='center', fontdict={'size': 14, 'weight': 700})
ax.text(3+0.05, 13000, 'AFTER', horizontalalignment='left',

verticalalignment='center', fontdict={'size':14, 'weight': 700})
ax.set_title("Change in deaths between 2014 vs 2017",fontdict={'size':18})
ax.set(xlim=(0, 4), ylim=(0, 14000), ylabel='No. of Kills')
ax.set_xticks([1, 3]); ax.set_xticklabels([2014, 2017])
plt.yticks(np.arange(500, 13000, 2000), fontsize=12)
plt.gca().spines["top"].set_alpha(.0); plt.gca().spines["bottom"].set_alpha(.0)
plt.gca().spines["right"].set_alpha(.0); plt.gca().spines["left"].set_alpha(.0)
plt.show()
```



Maximum people targeted over entire period

Inference: *Year 2014 witnessed the most number of kills, reaching the the astounding number of 9825.*

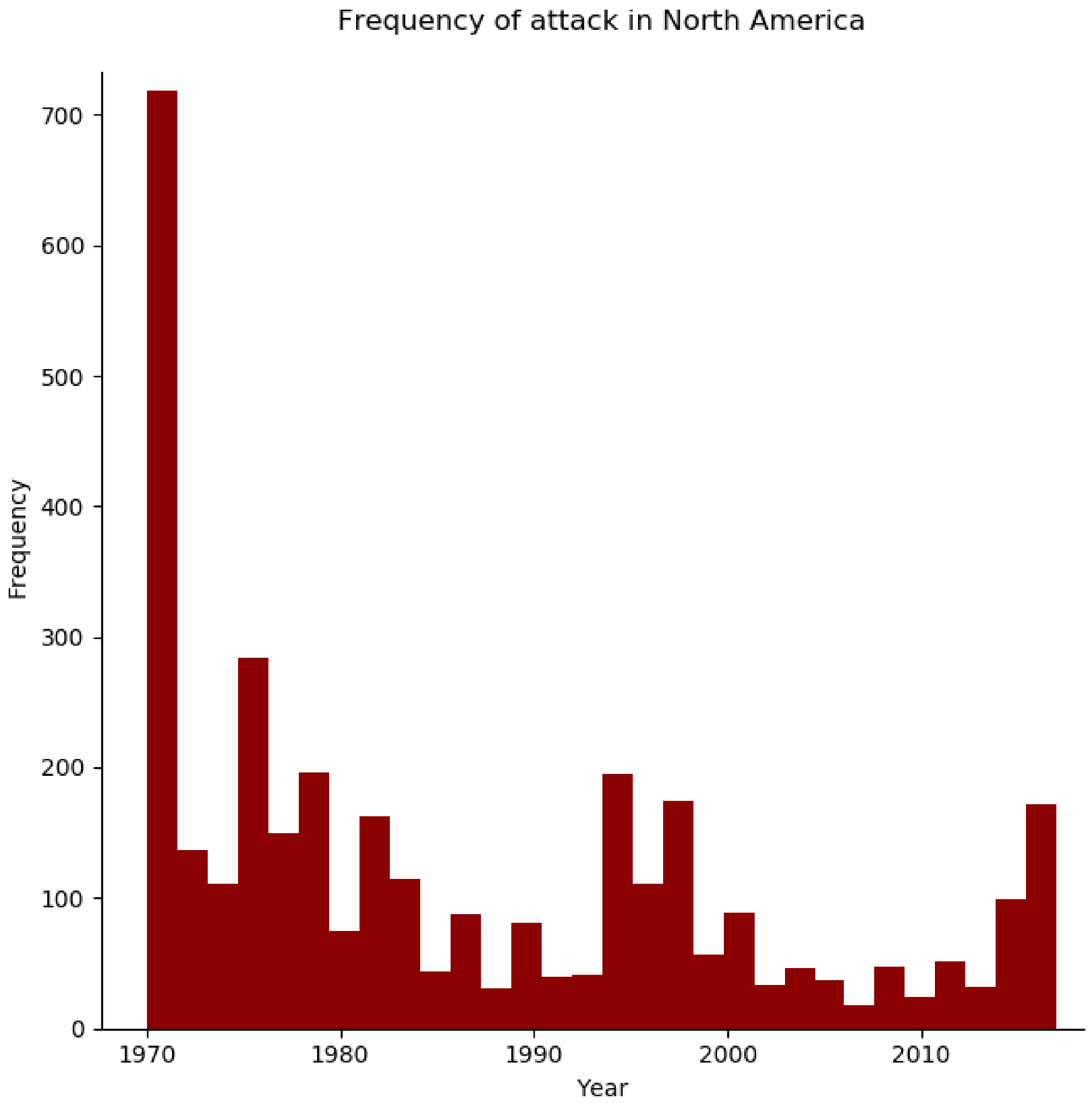
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('dataset/terror_data.csv')
print('most  kill in attacks:
',df['fatalities'].max(),df.loc[df['year'].idxmax()].year)
df.plot(kind = 'scatter', y = 'fatalities', x = 'year', alpha = 0.5, color
= 'darkred', fontsize=15)
plt.xlabel('fatalities', fontsize=15)
plt.ylabel('year', fontsize=15)
plt.title('fatalities and year')
plt.show()
```



Frequency of attack in North America

Inference: *North America remains a much peaceful continent and is successful in handling terrorists well by their policies.*

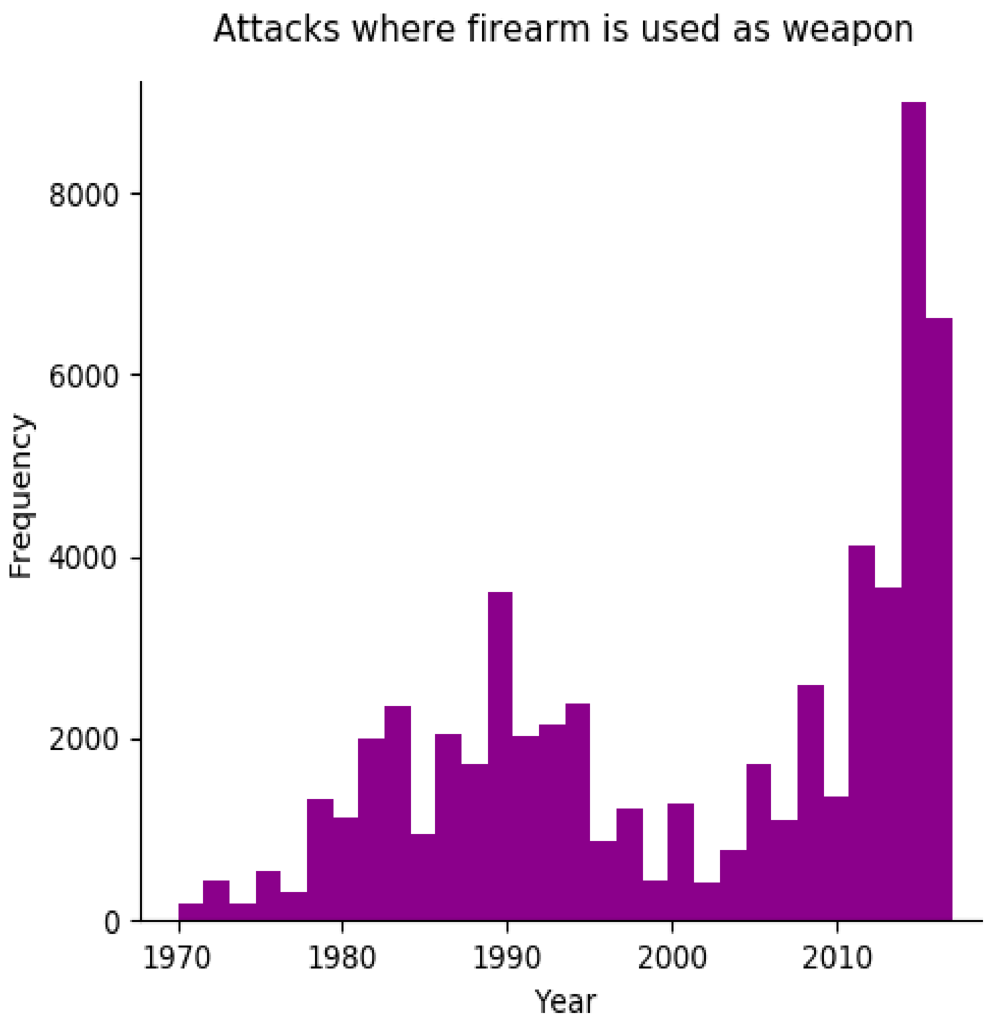
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('dataset/terror_data.csv')
mf = df[df['region_id'] == 1]
mf.year.plot(kind = 'hist', bins = 30, color = 'darkred', fontsize=10)
plt.xlabel('Year', fontsize=10)
plt.ylabel('Frequency', fontsize=10)
plt.title('Frequency of attack in North America')
plt.show()
```



Attacks based on weapon: ‘Firearm’

Inference: *Attacks with firearm has increased in recent years because of higher availability of modern weapons and being handy to use.*

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('dataset/terror_data.csv')
arr=["Firearms"]
df3=df.loc[df["weap_name"].isin(arr)]
df3.year.plot(kind = 'hist', bins = 30, color = 'darkmagenta')
plt.xlabel('Year')
plt.ylabel('Frequency')
plt.title('Attacks where firearm is used as weapon')
plt.show()
```



Drawing Inferences based on various parameters

By all the above analysis we can summarize the following inferences:

- Deaths from terrorism fell for the third consecutive year, after peaking in 2014.
- The total number of deaths fell by 27 per cent between 2016 and 2017, with the largest falls occurring in Iraq and Syria.
- 94 countries improved its GTI, compared to 46 that deteriorated. This is the highest number of countries to record a year on year improvement since 2004.
- The large falls in the number of deaths in Middle East is mainly the result of ISIL's continuing decline.
- Despite its reduced capacity, ISIL remained the deadliest terrorist group globally in 2017.
- Europe was the region with the biggest improvement from the impact of terrorism and recorded a marked fall in terrorist activity.
- Despite the fall in deaths, the number of terrorist incidents increased to 282 in Europe, up from 253 in the prior year.
- Conflict remains the primary driver of terrorism in most countries throughout the world. The ten countries with the highest impact of terrorism are all engaged in at least one conflict.

- In countries with high levels of economic development, factors other than conflict and human rights abuses are more strongly correlated with the impact of terrorism.
- Social alienation, lack of economic opportunity, and involvement in an external conflict are the major factors associated with terrorist activity in Western Europe, North America, and other highly economically developed regions.
- Despite the fall in deaths, the number of incidents rose in Western Europe. Increased counter-terrorism spending and security measures have reduced the lethality of attacks.
- Every region in the world recorded a higher average impact of terrorism in 2017 than in 2002.
- Bombings and armed assaults have been the most common form of terrorist attack every year for the past twenty years.
- Over 99 per cent of all deaths from terrorism have occurred in countries involved in a violent conflict or with high levels of political terror.
- Terrorist attacks have been more lethal on average in conflict-affected countries than countries not in conflict for every year bar one since 2002.
- Civilian’s property, military bases and government organisation are the popular targets of terrorists.
- Terrorism in India is not stagnant but is rising at noticeable pace in states where there are conflicts between separatists and government policies.

Certificate

This is to certify that Mr Abhishek Kumar Singh of GIET University, Gunupur, registration number: 1701210128, has successfully completed a project on Analysis of Global Terrorism Data using Data Science with Python under the guidance of Mr Kaushik Ghosh.

Mr Kaushik Ghosh

Globsyn Finishing School

Certificate

This is to certify that Mr Janmejoy Das Adhikari of GIET University, Gunupur, registration number: 1701210017, has successfully completed a project on Analysis of Global Terrorism Data using Data Science with Python under the guidance of Mr Kaushik Ghosh.

Mr Kaushik Ghosh

Globsyn Finishing School

Certificate

This is to certify that Mr Ashok Kumar of GIET University, Gunupur, registration number: 1701210320, has successfully completed a project on Analysis of Global Terrorism Data using Data Science with Python under the guidance of Mr Kaushik Ghosh.

Mr Kaushik Ghosh

Globsyn Finishing School

Certificate

This is to certify that Mr Afzal Alam of Chandigarh University, registration number: 11717629, has successfully completed a project on Analysis of Global Terrorism Data using Data Science using Python under the guidance of Mr Kaushik Ghosh.

Mr Kaushik Ghosh

Globsyn Finishing School