

Global Terrorism Data Analysis and Visualization

```
In [3]: import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
import seaborn as sns
import numpy as np
plt.style.use('fivethirtyeight')
import plotly.offline as py
py.init_notebook_mode(connected=True)
import plotly.graph_objs as go
import plotly.tools as tls
from mpl_toolkits.basemap import Basemap
import folium
import folium.plugins
from matplotlib import animation,rc
from scipy.misc import imread
import codecs
from subprocess import check_output
print(check_output(["ls", "../input"]).decode("utf8"))
```

globalterrorismdb_0718dist.csv

Reading the data

```
In [4]: terror=pd.read_csv('../input/globalterrorismdb_0718dist.csv',encoding='ISO-8859-1')
terror.rename(columns={'iyear':'Year','imonth':'Month','iday':'Day','country_txt':'Country',
terror=terror[['Year','Month','Day','Country','Region','city','latitude','longitude','AttackType','Killed','Wounded']]
terror['casualties']=terror['Killed']+terror['Wounded']
terror.head(3)
```

Out[4]:

	Year	Month	Day	Country	Region	city	latitude	longitude	AttackType	Killed	Wounded
0	1970	7	2	Dominican Republic	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	Assassination	1.0	0.0
1	1970	0	0	Mexico	North America	Mexico city	19.371887	-99.086624	Hostage Taking (Kidnapping)	0.0	0.0
2	1970	1	0	Philippines	Southeast Asia	Unknown	15.478598	120.599741	Assassination	1.0	0.0

```
In [5]: terror.isnull().sum()
```

```
out[5]: Year          0  
Month         0  
Day           0  
Country        0  
Region         0  
city          434  
latitude       4556  
longitude      4557  
AttackType     0  
Killed        10313  
Wounded        16311  
Target         634  
Summary        66129  
Group          0  
Target_type    0  
Weapon_type    0  
Motive         131130  
casualties     16874  
dtype: int64
```

Basic EDA

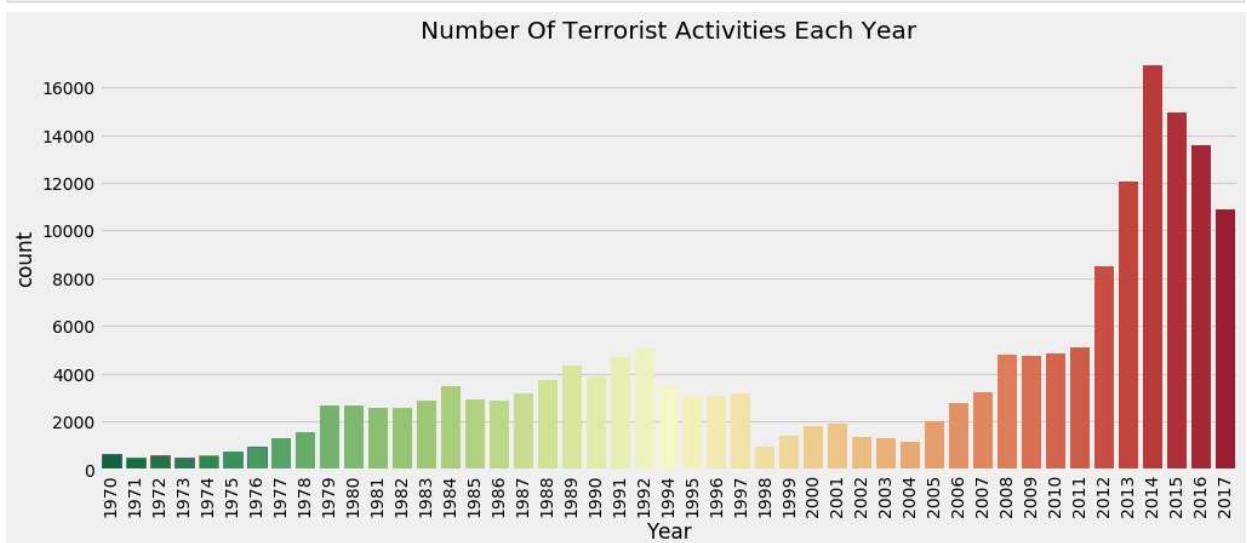
```
In [6]: print('Country with Highest Terrorist Attacks:',terror['Country'].value_counts().index[0])  
print('Regions with Highest Terrorist Attacks:',terror['Region'].value_counts().index[0])  
print('Maximum people killed in an attack are:',terror['Killed'].max(),'that took place in',terror['Location'].value_counts().index[0])
```

Country with Highest Terrorist Attacks: Iraq

Regions with Highest Terrorist Attacks: Middle East & North Africa

Maximum people killed in an attack are: 1570.0 that took place in Iraq

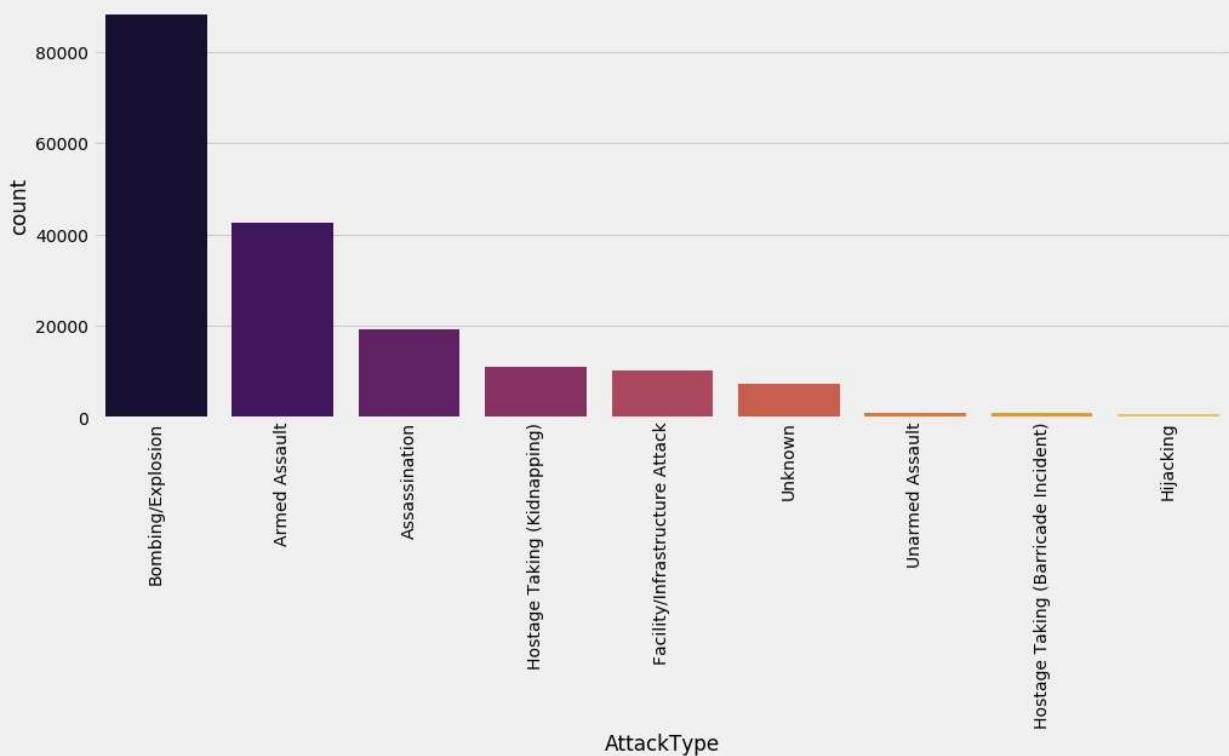
```
In [7]: plt.subplots(figsize=(15,6))  
sns.countplot('Year',data=terror,palette='RdYlGn_r',edgecolor=sns.color_palette('dark')  
plt.xticks(rotation=90)  
plt.title('Number Of Terrorist Activities Each Year')  
plt.show()
```



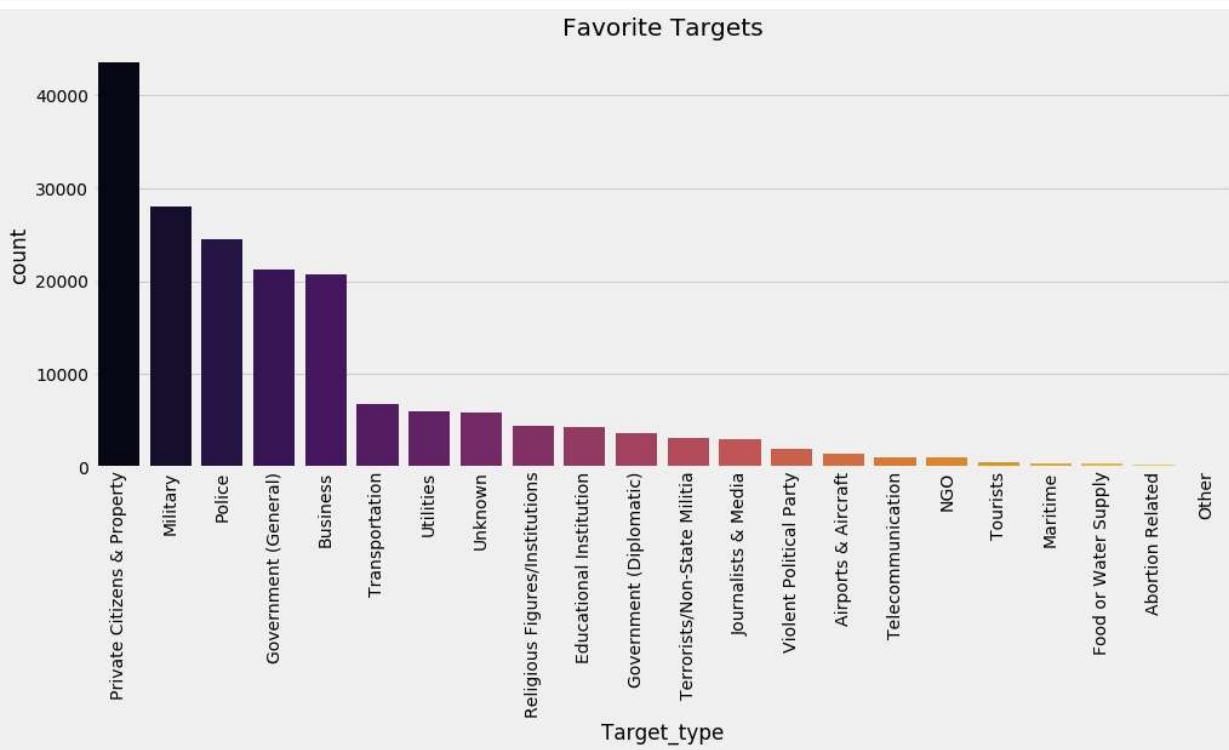
We can find that the number of terrorist activities have gone up sharply after the year 2011.

```
In [8]: plt.subplots(figsize=(15,6))  
sns.countplot('AttackType',data=terror,palette='inferno',order=terror['AttackType'].value_counts().index)  
plt.xticks(rotation=90)  
plt.title('Attacking Methods by Terrorists')  
plt.show()
```

Attacking Methods by Terrorists

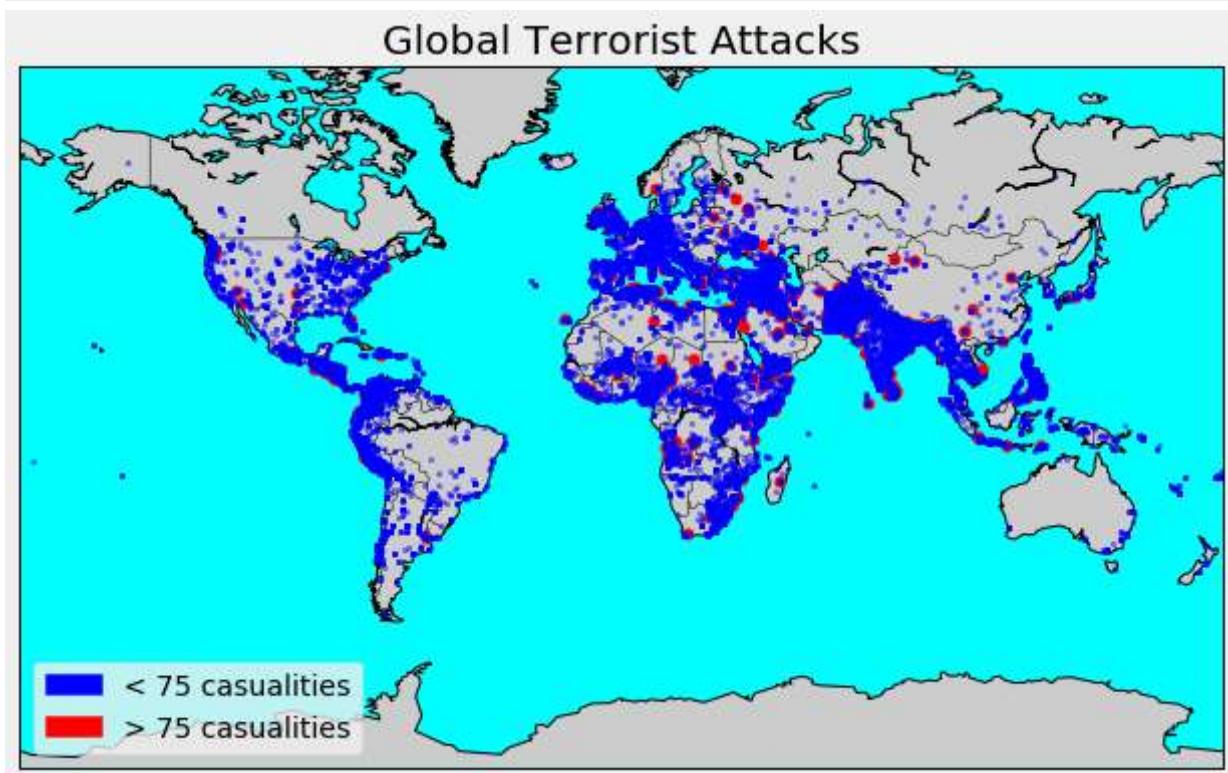


```
In [9]: plt.subplots(figsize=(15,6))
sns.countplot(terror['Target_type'], palette='inferno', order=terror['Target_type'].value_counts().index)
plt.xticks(rotation=90)
plt.title('Favorite Targets')
plt.show()
```



Global Terrorism Attacks

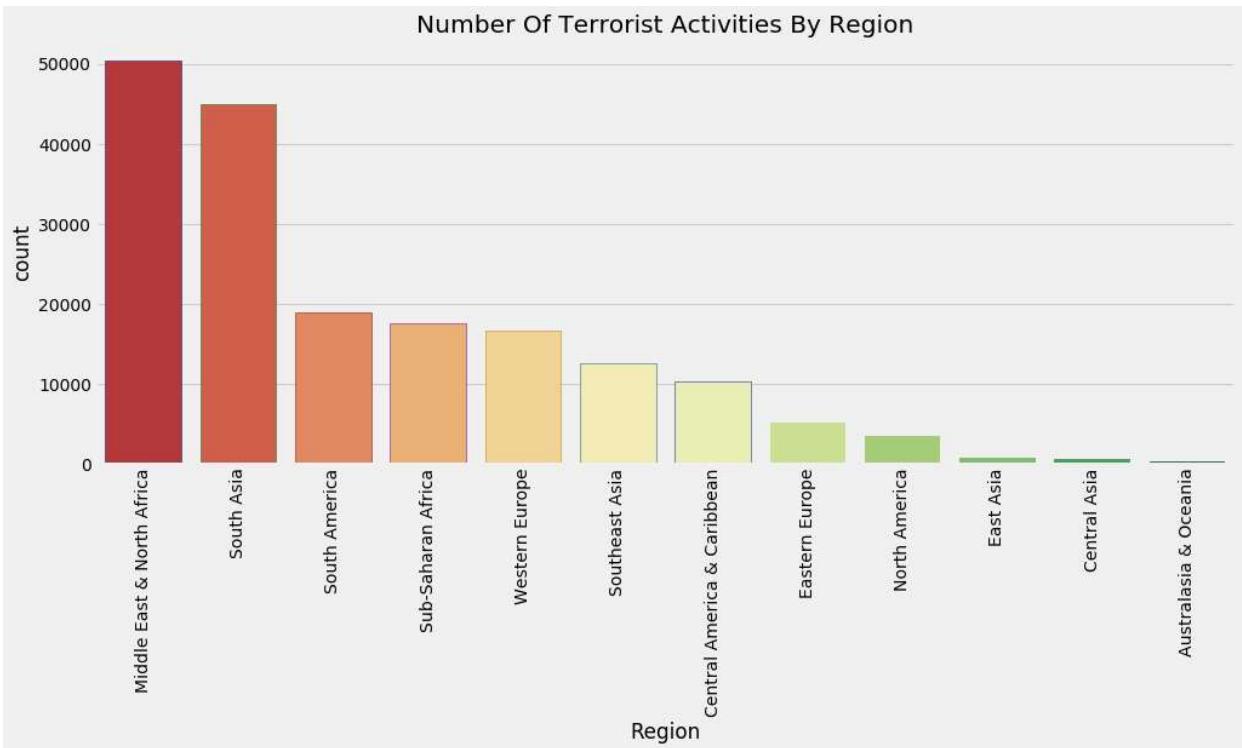
```
In [10]: m3 = Basemap(projection='mill',llcrnrlat=-80,urcrnrlat=80, llcrnrlon=-180,urcrnrlon=180)
lat_100=list(terror[terror['casualties']>=75].latitude)
long_100=list(terror[terror['casualties']>=75].longitude)
x_100,y_100=m3(long_100,lat_100)
m3.plot(x_100, y_100,'go',markersize=5,color = 'r')
lat_=list(terror[terror['casualties']<75].latitude)
long_=list(terror[terror['casualties']<75].longitude)
x_,y_=m3(long_,lat_)
m3.plot(x_, y_,'go',markersize=2,color = 'b',alpha=0.4)
m3.drawcoastlines()
m3.drawcountries()
m3.fillcontinents(lake_color='aqua')
m3.drawmapboundary(fill_color='aqua')
fig=plt.gcf()
fig.set_size_inches(10,6)
plt.title('Global Terrorist Attacks')
plt.legend(loc='lower left',handles=[mpatches.Patch(color='b', label = "< 75 casualties"),
                                     mpatches.Patch(color='red',label='> 75 casualties')])
plt.show()
```



The above basemap shows the locations of attacks. The red circles represent those with more than 75 casualties (wounded + killed).

Terrorism By Regions

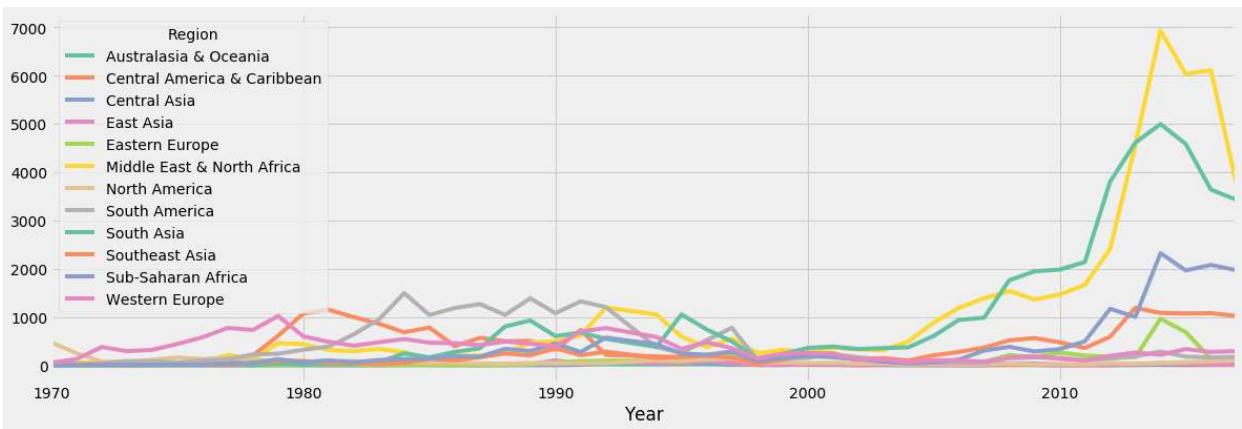
```
In [12]: plt.subplots(figsize=(15,6))
sns.countplot('Region',data=terror,palette='RdYlGn',edgecolor=sns.color_palette('dark')
plt.xticks(rotation=90)
plt.title('Number Of Terrorist Activities By Region')
plt.show()
```



The Middle East and North Africa are the most terrorism-prone regions, followed by South Asia. The Australian region has experienced very few terrorist events. Collectively, we can say that the African and Asian continents experience the highest number of terrorist attacks.

Trend in Terrorist Activities

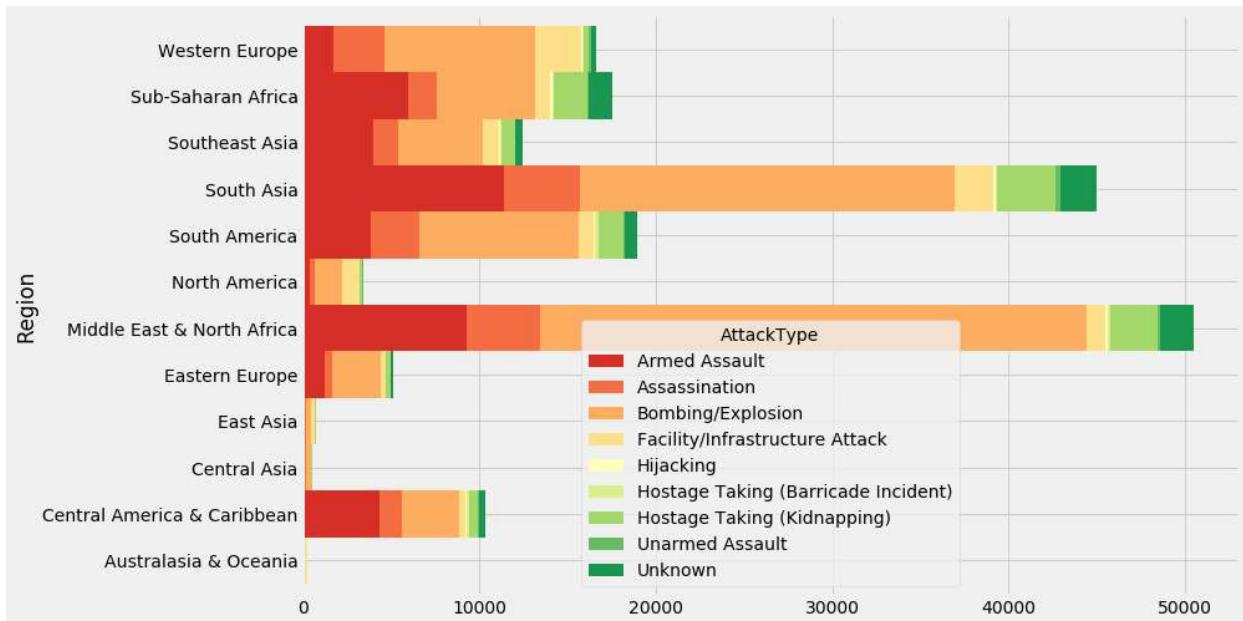
```
In [13]: terror_region=pd.crosstab(terror.Year,terror.Region)
terror_region.plot(color=sns.color_palette('Set2',12))
fig=plt.gcf()
fig.set_size_inches(18,6)
plt.show()
```



The Middle East, North Africa, and South Asia have seen a rise in the number of terrorist activities over the years.

AttackType vs Region

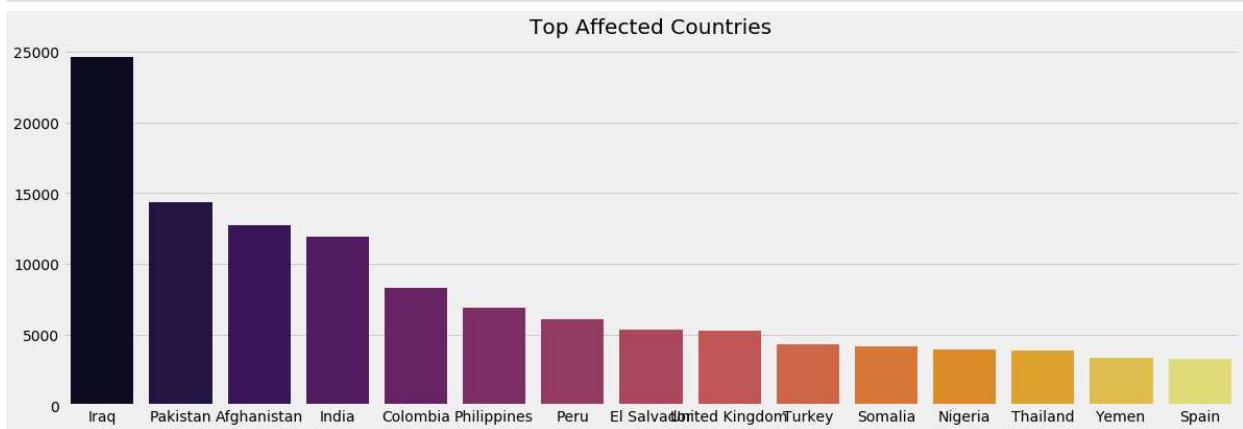
```
In [14]: pd.crosstab(terror.Region,terror.AttackType).plot.barh(stacked=True,width=1,color=sns.color_palette('Set2',12))
fig=plt.gcf()
fig.set_size_inches(12,8)
plt.show()
```



Bombings and armed assaults are the most prominent types of attacks irrespective of regions.

Terrorism By Country

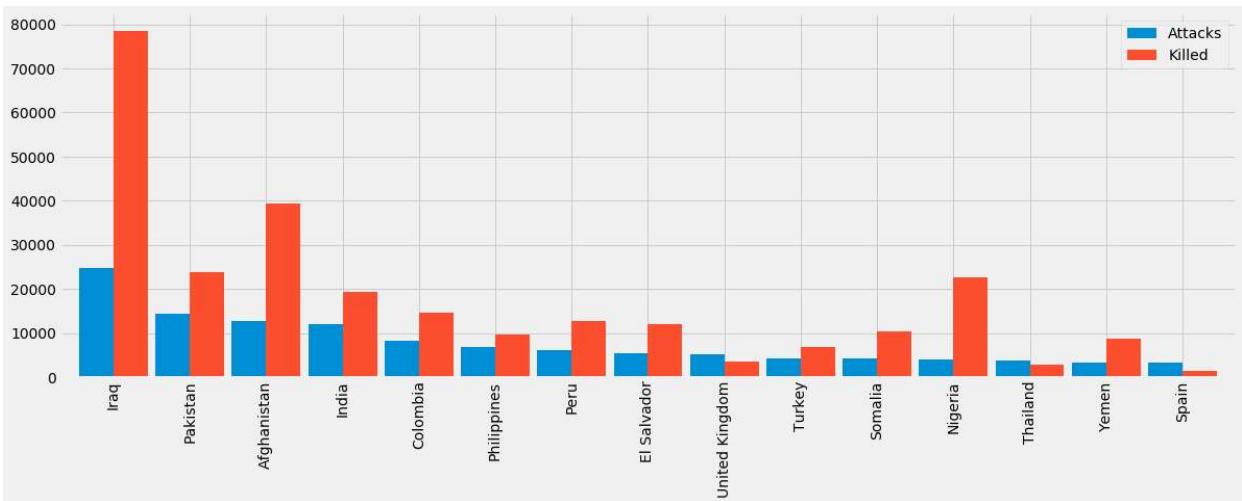
```
In [15]: plt.subplots(figsize=(18,6))
sns.barplot(terror['Country'].value_counts()[:15].index,terror['Country'].value_counts())
plt.title('Top Affected Countries')
plt.show()
```



Iraq has witnessed a very large number of terrorist activities, followed by Pakistan. One thing to note is that the countries with the highest number of attacks are mostly densely populated, which means they will ultimately claim many lives.

Attacks vs Killed

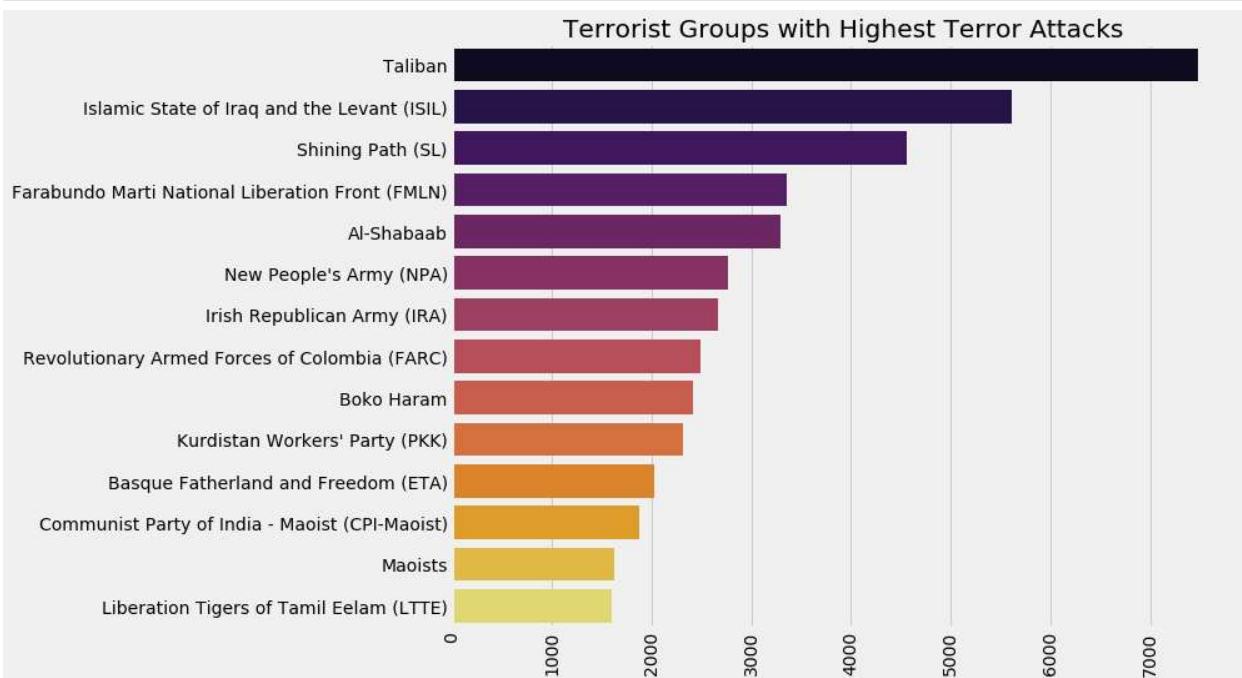
```
In [16]: coun_terror=terror['Country'].value_counts()[:15].to_frame()
coun_terror.columns=['Attacks']
coun_kill=terror.groupby('Country')['Killed'].sum().to_frame()
coun_terror.merge(coun_kill, left_index=True, right_index=True, how='left').plot.bar(wid
fig=plt.gcf()
fig.set_size_inches(18,6)
plt.show()
```



Look at the killed bar for Iraq. The number of killed is almost three times more than the number of attacks in Iraq. Thus, the densely populated theory holds true.

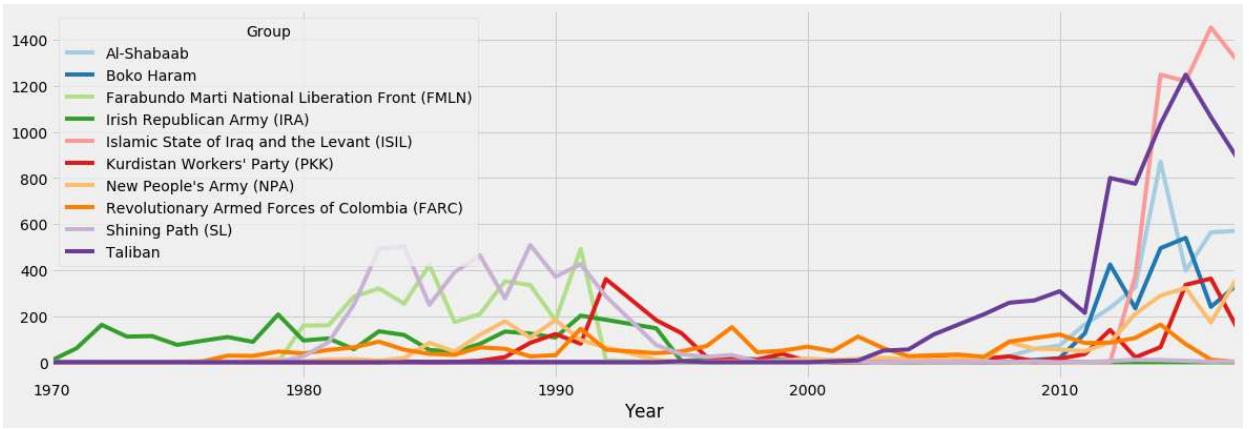
Look at the most terrorist groups

```
In [18]: sns.barplot(terror['Group'].value_counts()[1:15].values,terror['Group'].value_counts()
plt.xticks(rotation=90)
fig=plt.gcf()
fig.set_size_inches(10,8)
plt.title('Terrorist Groups with Highest Terror Attacks')
plt.show()
```



Activity of Top Terrorist Groups

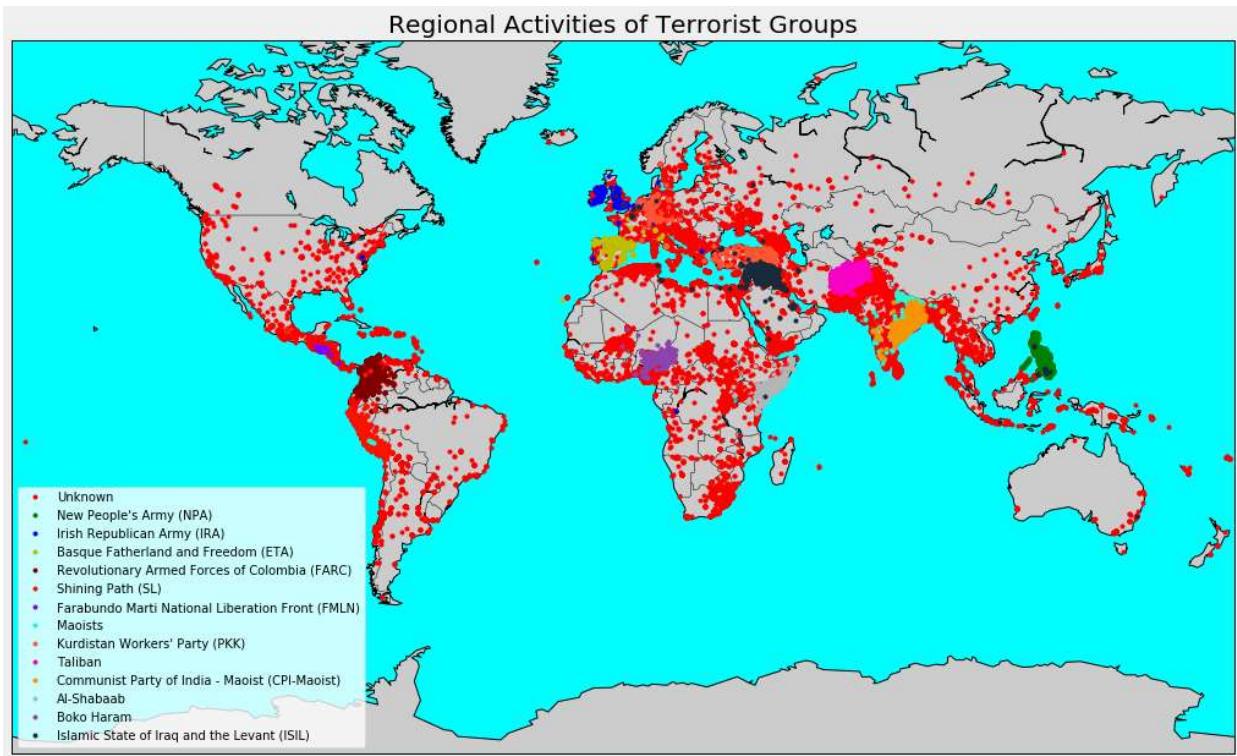
```
In [19]: top_groups10=terror[terror['Group'].isin(terror['Group'].value_counts()[1:11].index)]
pd.crosstab(top_groups10.Year,top_groups10.Group).plot(color=sns.color_palette('Paired'))
fig=plt.gcf()
fig.set_size_inches(18,6)
plt.show()
```



The Irish Republican Army (IRA) is the oldest terrorist group, having started in the 1960s or 1970s, possibly as a response to the mass killings during or after World War II. However, it likely ceased its activities in the late 1990s. Some of the groups that emerged more recently in the 2010s, such as ISIL and the Taliban, have shown a significant increase in the number of attacks in recent years.

Regions Attacked By Terrorist Groups

```
In [20]: top_groups=terror[terror['Group'].isin(terror['Group'].value_counts()[:14].index)]
m4 = Basemap(projection='mill',llcrnrlat=-80,urcrnrlat=80, llcrnrlon=-180,urcrnrlon=180)
m4.drawcoastlines()
m4.drawcountries()
m4.fillcontinents(lake_color='aqua')
m4.drawmapboundary(fill_color='aqua')
fig=plt.gcf()
fig.set_size_inches(22,10)
colors=['r','g','b','y','#800000','#ff1100','#8202fa','#20fad9','#ff5733','#fa02c6','#ff9800']
group=list(top_groups['Group'].unique())
def group_point(group,color,label):
    lat_group=list(top_groups[top_groups['Group']==group].latitude)
    long_group=list(top_groups[top_groups['Group']==group].longitude)
    x_group,y_group=m4(long_group,lat_group)
    m4.plot(x_group,y_group,'go',markersize=3,color=j,label=i)
for i,j in zip(group,colors):
    group_point(i,j,i)
legend=plt.legend(loc='lower left',frameon=True,prop={'size':10})
frame=legend.get_frame()
frame.set_facecolor('white')
plt.title('Regional Activities of Terrorist Groups')
plt.show()
```



ISIL appears to be the most notorious group in Iran and Iraq, as well as across the broader Middle East. Similarly, the Taliban is concentrated in Afghanistan and Pakistan.

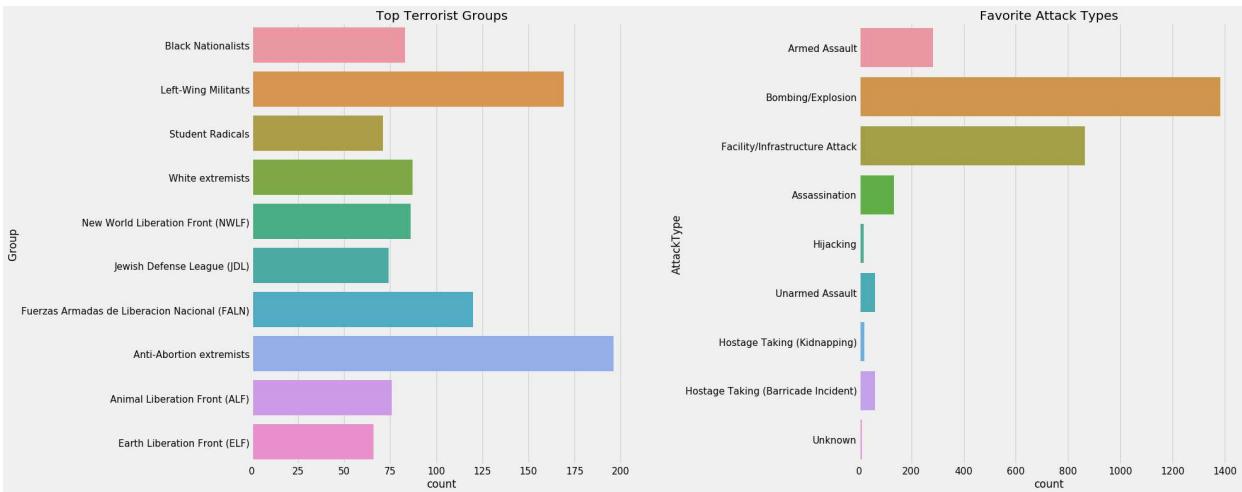
The Unknown markers may indicate individual attacks stemming from resentment, personal grudges, or activities of less well-known groups.

Terrorist Activities in USA

The Terrorist Groups in USA and the most favorite attack type

```
In [25]: terror_usa=terror[terror['Country']=='United States']
terror_usa_fol=terror_usa.copy()
terror_usa_fol.dropna(subset=['latitude','longitude'],inplace=True)
location_usa=terror_usa_fol[['latitude','longitude']]
city_usa=terror_usa_fol['city']
killed_usa=terror_usa_fol['Killed']
wound_usa=terror_usa_fol['Wounded']
target_usa=terror_usa_fol['Target_type']

f,ax=plt.subplots(1,2,figsize=(25,12))
usa_groups=terror_usa['Group'].value_counts()[1:11].index
usa_groups=terror_usa[terror_usa['Group'].isin(usa_groups)]
sns.countplot(y='Group',data=usa_groups,ax=ax[0])
sns.countplot(y='AttackType',data=terror_usa,ax=ax[1])
plt.subplots_adjust(hspace=0.3,wspace=0.6)
ax[0].set_title('Top Terrorist Groups')
ax[1].set_title('Favorite Attack Types')
ax[0].tick_params(labelsize=15)
ax[1].tick_params(labelsize=15)
plt.show()
```

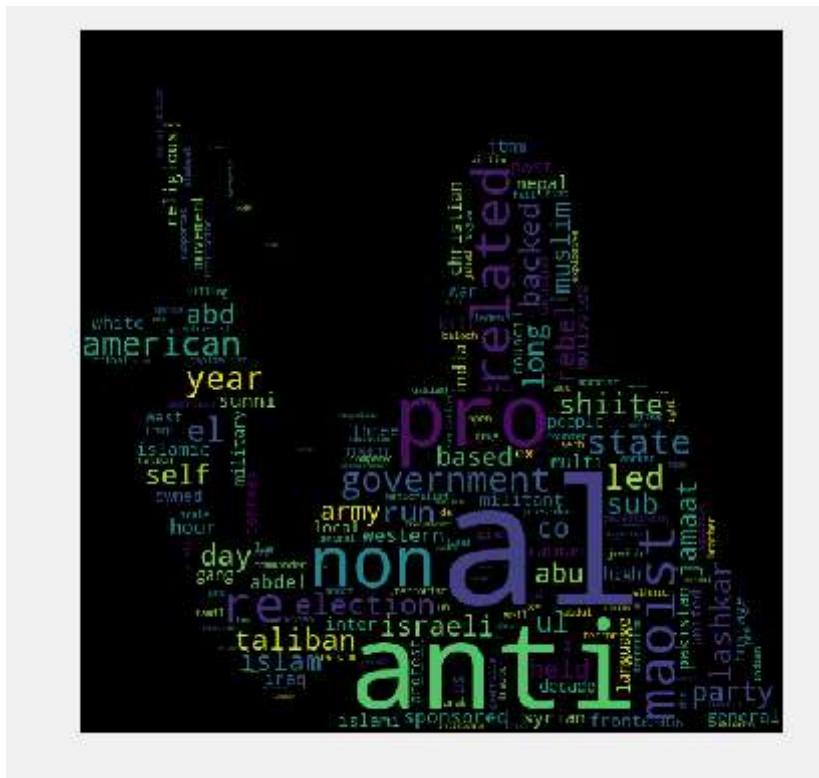


Why Attacks Happened

Analyze the main reasons behind terrorist activities using NLTK for Natural Language Processing. The reason for using NLP is that if we simply count words and create a word cloud, many common words like 'the,' 'and,' 'is,' etc., will have the highest count as they are very common in the English language. Therefore, by using NLTK, we can filter out these words and identify other important ones.

In [28]:

```
import nltk
from wordcloud import WordCloud, STOPWORDS
motive=terror['Motive'].str.lower().str.replace(r'\|', ' ').str.cat(sep=' ')
words=nltk.tokenize.word_tokenize(motive)
word_dist = nltk.FreqDist(words)
stopwords = nltk.corpus.stopwords.words('english')
f1=open("Attacks.png", "wb")
f1.write(codecs.decode(kaggle, 'base64'))
f1.close()
img1 = imread("Attacks.png")
hcmask1 = img1
words_except_stop_dist = nltk.FreqDist(w for w in words if w not in stopwords)
wordcloud = WordCloud(stopwords=STOPWORDS,background_color='black',mask=hcmask1).gener
plt.imshow(wordcloud)
fig=plt.gcf()
fig.set_size_inches(10,6)
plt.axis('off')
plt.show()
```



Religious sentiments or religion appear to be the main causes of terrorism. Some of the most dangerous terrorist group names, like 'al' for 'al-Qaeda,' are prominent. Additionally, words such as 'government' and 'anti' also indicate attacks driven by resentment against the government.