# Task: Exploratory Data Analysis - Terrorism

**Our Aim :**

- Perform 'Exploratory Data Analysis' on dataset 'Global Terrorism'
- As a security/defense analyst, try to find out the hot zone of terrorism.
- What all security issues and insights you can derive by EDA?

## DATA PREPROCESSING

In [1]:

```
import math
import warnings
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
import plotly.offline as py
import plotly.graph_objs as go
import matplotlib.pyplot as plt
warnings.filterwarnings('ignore')
```

In [2]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px


import warnings
warnings.filterwarnings('ignore')
```

In [3]:

```
# Let's import to our data and check the basics.
data = pd.read_csv(r"globalterrorismdb_0718dist[1].csv",encoding='latin1')
```

In [4]:

```
data.head()
```

Out[4]:

| | eventid | iyear | imonth | iday | approxdate | extended | resolution | country | country_txt |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 197000000001 | 1970 | 7 | 2 | NaN | 0 | NaN | 58 | Dominican Republic |
| **1** | 197000000002 | 1970 | 0 | 0 | NaN | 0 | NaN | 130 | Mexico |
| **2** | 197001000001 | 1970 | 1 | 0 | NaN | 0 | NaN | 160 | Philippines |
| **3** | 197001000002 | 1970 | 1 | 0 | NaN | 0 | NaN | 78 | Greece |
| **4** | 197001000003 | 1970 | 1 | 0 | NaN | 0 | NaN | 101 | Japan |

5 rows × 135 columns

In [5]:

```
data.shape
```

Out[5]:

```
(181691, 135)
```

In [6]:

```
nul=[]
for i in data.columns:
    if data[i].isna().sum() > 0:
        nul.append(i)
len(nul)
```

Out[6]:

```
106
```

In [7]:

```
for i in nul:
    data.drop(i,axis=1,inplace=True)

data.shape
```

Out[7]:

```
(181691, 29)
```

In [8]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 29 columns):
 #   Column           Non-Null Count   Dtype 
---  ------           --------------   ----- 
 0   eventid          181691 non-null  int64 
 1   iyear            181691 non-null  int64 
 2   imonth           181691 non-null  int64 
 3   iday             181691 non-null  int64 
 4   extended         181691 non-null  int64 
 5   country          181691 non-null  int64 
 6   country_txt      181691 non-null  object
 7   region           181691 non-null  int64 
 8   region_txt       181691 non-null  object
 9   vicinity         181691 non-null  int64 
 10  crit1            181691 non-null  int64 
 11  crit2            181691 non-null  int64 
 12  crit3            181691 non-null  int64 
 13  success          181691 non-null  int64 
 14  suicide          181691 non-null  int64 
 15  attacktype1      181691 non-null  int64 
 16  attacktype1_txt  181691 non-null  object
 17  targtype1        181691 non-null  int64 
 18  targtype1_txt    181691 non-null  object
 19  gname            181691 non-null  object
 20  individual       181691 non-null  int64 
 21  weaptype1        181691 non-null  int64 
 22  weaptype1_txt    181691 non-null  object
 23  property         181691 non-null  int64 
 24  dbsource         181691 non-null  object
 25  INT_LOG          181691 non-null  int64 
 26  INT_IDEO         181691 non-null  int64 
 27  INT_MISC         181691 non-null  int64 
 28  INT_ANY          181691 non-null  int64 
dtypes: int64(22), object(7)
memory usage: 40.2+ MB
```

In [9]:

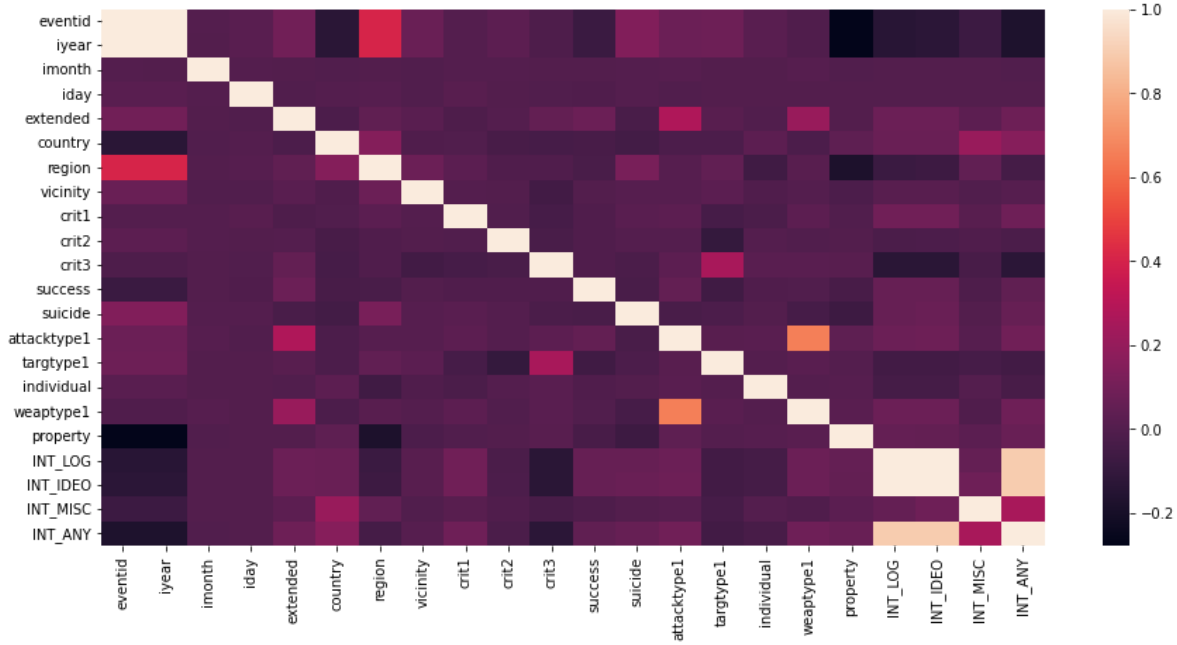
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Data columns (total 29 columns):
 #   Column           Non-Null Count   Dtype 
---  ------           --------------   ----- 
 0   eventid          181691 non-null  int64 
 1   iyear            181691 non-null  int64 
 2   imonth           181691 non-null  int64 
 3   iday             181691 non-null  int64 
 4   extended         181691 non-null  int64 
 5   country          181691 non-null  int64 
 6   country_txt      181691 non-null  object
 7   region           181691 non-null  int64 
 8   region_txt       181691 non-null  object
 9   vicinity         181691 non-null  int64 
 10  crit1            181691 non-null  int64 
 11  crit2            181691 non-null  int64 
 12  crit3            181691 non-null  int64 
 13  success          181691 non-null  int64 
 14  suicide          181691 non-null  int64 
 15  attacktype1      181691 non-null  int64 
 16  attacktype1_txt  181691 non-null  object
 17  targtype1        181691 non-null  int64 
 18  targtype1_txt    181691 non-null  object
 19  gname            181691 non-null  object
 20  individual       181691 non-null  int64 
 21  weaptype1        181691 non-null  int64 
 22  weaptype1_txt    181691 non-null  object
 23  property         181691 non-null  int64 
 24  dbsource         181691 non-null  object
 25  INT_LOG          181691 non-null  int64 
 26  INT_IDEO         181691 non-null  int64 
 27  INT_MISC         181691 non-null  int64 
 28  INT_ANY          181691 non-null  int64 
dtypes: int64(22), object(7)
memory usage: 40.2+ MB
```

# Data Visualization

In [10]:

```
fig,ax = plt.subplots(1,1,figsize=(15,7))
sns.heatmap(data.corr())
plt.show()
```

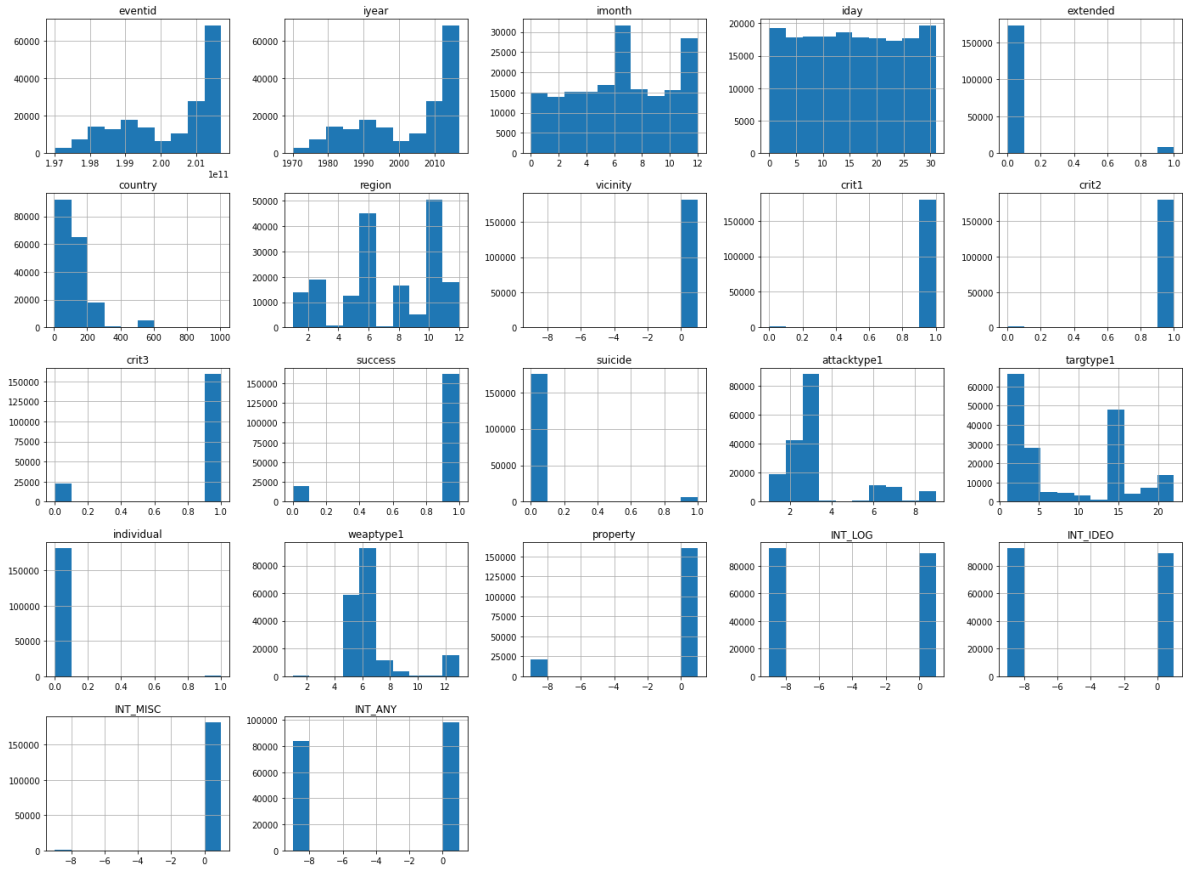


In [11]:

```
data.hist(figsize=(24,18))
```

Out[11]:

```
array([[<AxesSubplot:title={'center':'eventid'}>,
        <AxesSubplot:title={'center':'iyear'}>,
        <AxesSubplot:title={'center':'imonth'}>,
        <AxesSubplot:title={'center':'iday'}>,
        <AxesSubplot:title={'center':'extended'}>],
       [<AxesSubplot:title={'center':'country'}>,
        <AxesSubplot:title={'center':'region'}>,
        <AxesSubplot:title={'center':'vicinity'}>,
        <AxesSubplot:title={'center':'crit1'}>,
        <AxesSubplot:title={'center':'crit2'}>],
       [<AxesSubplot:title={'center':'crit3'}>,
        <AxesSubplot:title={'center':'success'}>,
        <AxesSubplot:title={'center':'suicide'}>,
        <AxesSubplot:title={'center':'attacktype1'}>,
        <AxesSubplot:title={'center':'targtype1'}>],
       [<AxesSubplot:title={'center':'individual'}>,
        <AxesSubplot:title={'center':'weaptype1'}>,
        <AxesSubplot:title={'center':'property'}>,
        <AxesSubplot:title={'center':'INT_LOG'}>,
        <AxesSubplot:title={'center':'INT_IDEO'}>],
       [<AxesSubplot:title={'center':'INT_MISC'}>,
        <AxesSubplot:title={'center':'INT_ANY'}>, <AxesSubplot:>,
        <AxesSubplot:>, <AxesSubplot:>]], dtype=object)
```

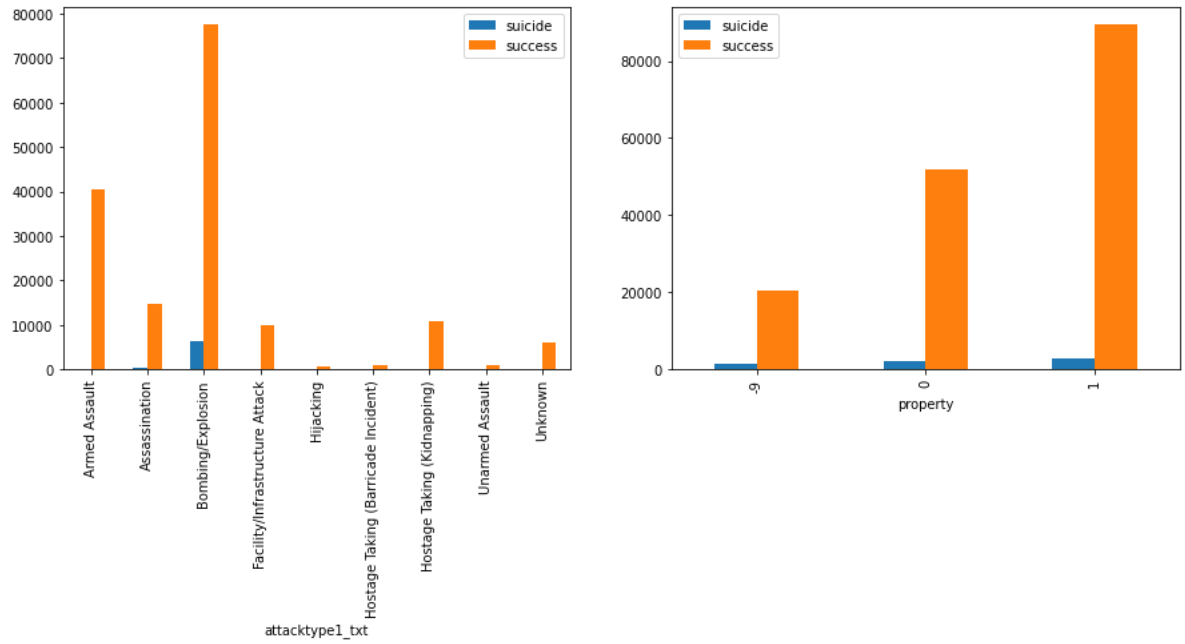


In [12]:

```
fig,axes = plt.subplots(1,2,figsize=(15,5))
data.groupby('attacktype1_txt')['suicide','success'].agg(sum).plot(kind='bar',ax=axes[0])
data.groupby('property')['suicide','success'].agg(sum).plot(kind='bar',ax=axes[1])
```

Out[12]:

```
<AxesSubplot:xlabel='property'>
```

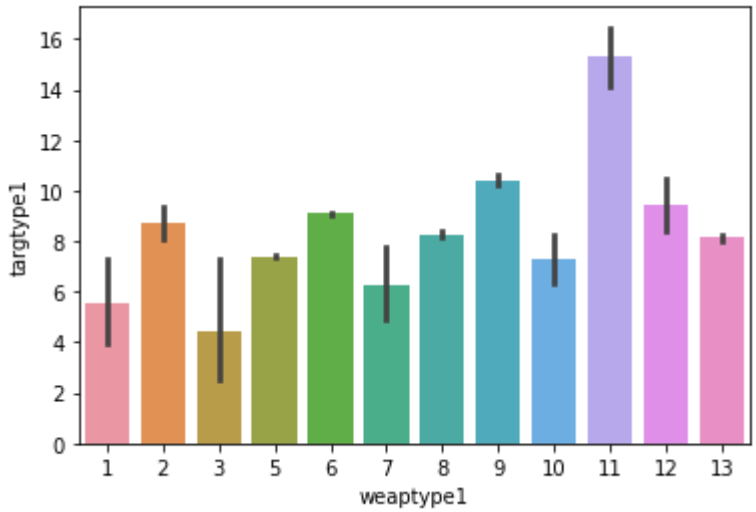


In [13]:

```
sns.barplot(y=data['targtype1'], x=data['weaptype1'])
```

Out[13]:

```
<AxesSubplot:xlabel='weaptype1', ylabel='targtype1'>
```

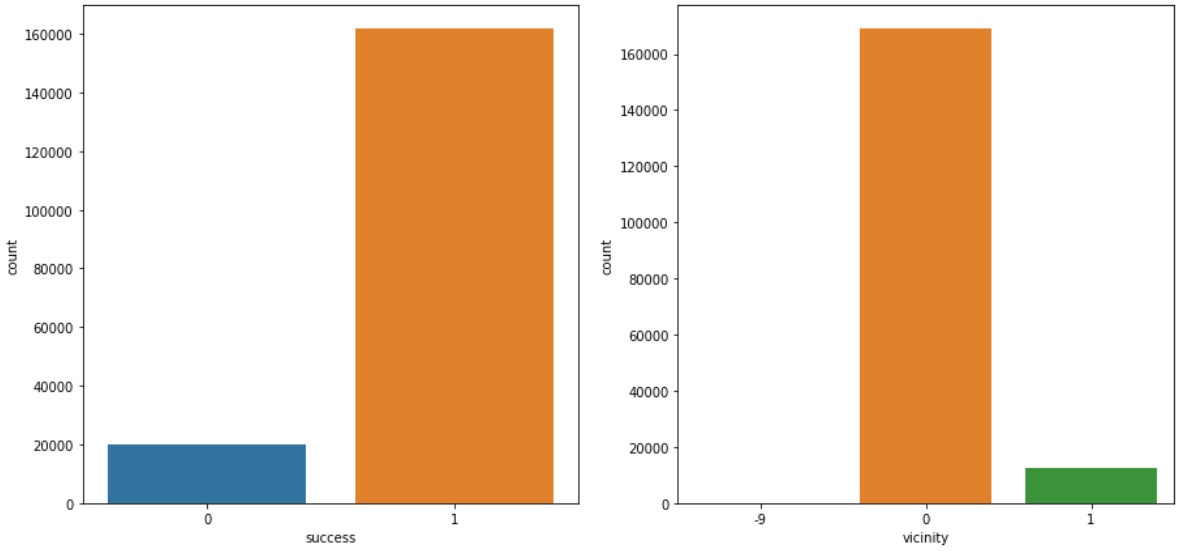



In [14]:

```
fig,ax = plt.subplots(1,2,figsize=(15,7))
sns.countplot(data.success,ax=ax[0])
sns.countplot(data.vicinity,ax=ax[1])
```

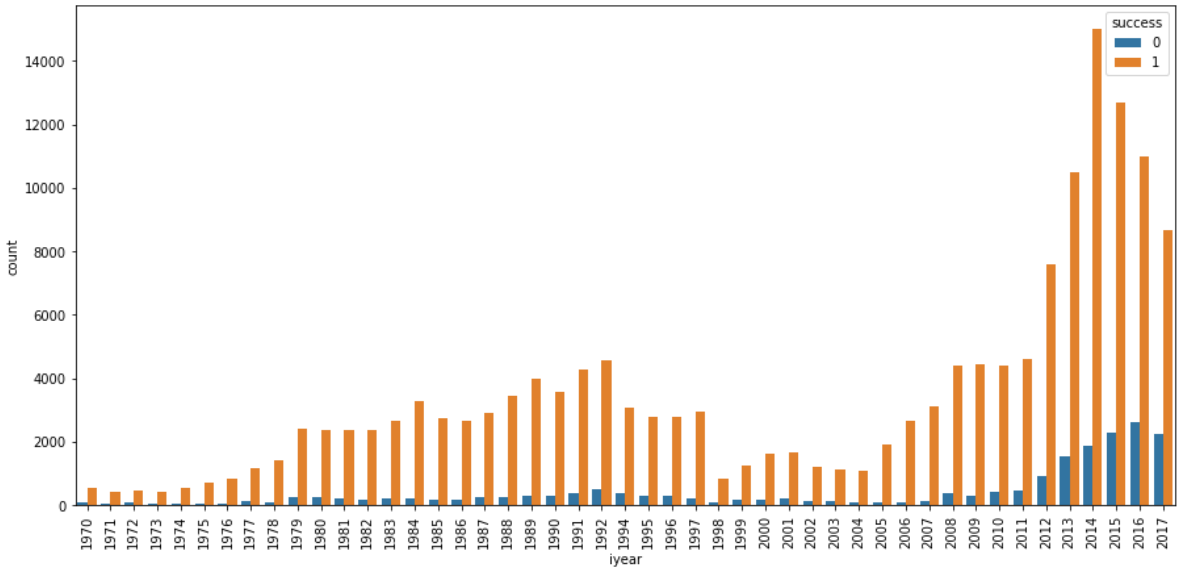
Out[14]:

```
<AxesSubplot:xlabel='vicinity', ylabel='count'>
```



In [15]:

```
fig,ax = plt.subplots(1,1,figsize=(15,7))
sns.countplot(data['iyear'],hue=data['success'])
plt.xticks(rotation=90)
plt.show()
```



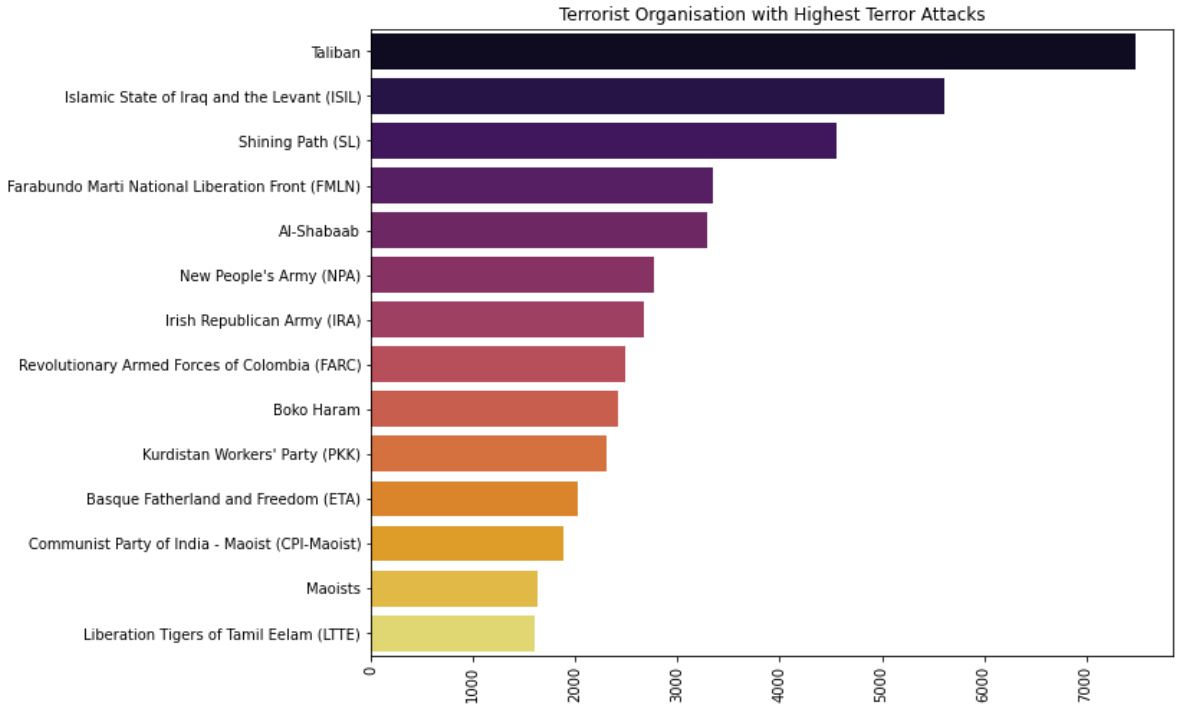

In [16]:

```
dat = data.groupby(['country_txt'],as_index=False).count()
```

# Terrorist Organisation with Highest Terror Attacks

In [17]:

```
sns.barplot(data['gname'].value_counts()[1:15].values,data['gname'].value_counts()[1:15].in
plt.xticks(rotation=90)
fig=plt.gcf()
fig.set_size_inches(10,8)
plt.title('Terrorist Organisation with Highest Terror Attacks')
plt.show()
```




# Using of Pyplot to show this visualization

In [18]:

```
fig = px.choropleth(dat,locations='country_txt',locationmode='country names',
                    color='success', hover_name='country', projection ='orthographic',
                    title='Total number of attack', labels={'Year':'Attacks'})
fig.show()
```

## CONCLUSION :

> So after different type of analyzation ,Overall terrorism is suddenly increased from 2010. I have ranked the Hot zone in terms of terrorist activities in all regions.

### Middle East & North Africa

1.Iraq

2.Libya

3.Yemen

## South Asia

1.Pakistan

2.Afghanistan

3.India

## Sub-Saharan Africa

1.Nigeria

2.Somalia

3.Sudan