

GRIP @ THE SPARKS FOUNDATION

Name- Dipranjan Parida

Batch- JUNE 2022

TASK-4 Exploratory Data Analysis - Terrorism

(LEVEL - INTEMEDIATE)

Perform Exploratory Data Analysis on Dataset "Global Terrorism". Try to find the hot zone of the Terrorism.

Here the aim is to derive insights from the EDA for security/defence analysis..

```
In [ ]: #Importing necessary libraries
```

```
In [22]: import warnings
warnings.filterwarnings('ignore')
```

```
In [23]: import numpy as np

import pandas as pd
from pandas import Series,DataFrame

import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [ ]: #Importing the Dataset
```

```
In [24]: df = pd.read_csv('globalterrorismdb_0718dist.csv',encoding='latin1')
df.head()
```

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	8
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	4

5 rows × 135 columns

In [25]:	df.shape																																																																							
Out[25]:	(181691, 135)																																																																							
In [26]:	df.columns																																																																							
Out[26]:	Index(['eventid', 'iyear', 'imonth', 'iday', 'approxdate', 'extended', 'resolution', 'country', 'country_txt', 'region', '', 'addnotes', 'scite1', 'scite2', 'scite3', 'dbsource', 'INT_LOG', 'INT_IDEO', 'INT_MISC', 'INT_ANY', 'related'], dtype='object', length=135)																																																																							
In [27]:	#Renaming the columns df.rename(columns={'iyear':'Year','imonth':'Month','iday':'Day','country_txt':'Country','provstate':'State', 'region_txt':'Region','city':'City','latitude':'Latitude','longitude':'Longitude', 'attacktype1_txt':'AttackType','targtype1_txt':'TargetType','targsubtype1_txt':'TargetSubType', 'weaptype1_txt':'WeaponType','gname':'Group','nkill':'Killed','nwound':'Wounded','summary':'Summary','m eleeve': 'Motive'})																																																																							
In [28]:	#Extracting Necessary Columns df=df[['eventid','Year','Month','Day','Country','State','Region','City','Latitude','Longitude','AttackType','TargetType','Targ eSubType','Suicide','Killed','Wounded','Summary','Group','WeaponType','Motive']]																																																																							
In [29]:	#Seeing First Five Entries df.head()																																																																							
Out[29]:	<table border="1"> <thead> <tr> <th></th><th>eventid</th><th>Year</th><th>Month</th><th>Day</th><th>Country</th><th>State</th><th>Region</th><th>City</th><th>Latitude</th><th>Longitude</th></tr> </thead> <tbody> <tr> <td>0</td><td>197000000001</td><td>1970</td><td></td><td>7</td><td>2</td><td>Dominican Republic</td><td>NaN</td><td>Central America & Caribbean</td><td>Santo Domingo</td><td>18.456792</td><td>-69.951164</td></tr> <tr> <td>1</td><td>197000000002</td><td>1970</td><td></td><td>0</td><td>0</td><td>Mexico</td><td>Federal</td><td>North America</td><td>Mexico city</td><td>19.371887</td><td>-99.086624</td></tr> <tr> <td>2</td><td>197001000001</td><td>1970</td><td></td><td>1</td><td>0</td><td>Philippines</td><td>Tarlac</td><td>Southeast Asia</td><td>Unknown</td><td>15.478598</td><td>120.599741</td></tr> <tr> <td>3</td><td>197001000002</td><td>1970</td><td></td><td>1</td><td>0</td><td>Greece</td><td>Attica</td><td>Western Europe</td><td>Athens</td><td>37.997490</td><td>23.762728</td></tr> <tr> <td>4</td><td>197001000003</td><td>1970</td><td></td><td>1</td><td>0</td><td>Japan</td><td>Fukouka</td><td>East Asia</td><td>Fukouka</td><td>33.580412</td><td>130.396361</td></tr> </tbody> </table>		eventid	Year	Month	Day	Country	State	Region	City	Latitude	Longitude	0	197000000001	1970		7	2	Dominican Republic	NaN	Central America & Caribbean	Santo Domingo	18.456792	-69.951164	1	197000000002	1970		0	0	Mexico	Federal	North America	Mexico city	19.371887	-99.086624	2	197001000001	1970		1	0	Philippines	Tarlac	Southeast Asia	Unknown	15.478598	120.599741	3	197001000002	1970		1	0	Greece	Attica	Western Europe	Athens	37.997490	23.762728	4	197001000003	1970		1	0	Japan	Fukouka	East Asia	Fukouka	33.580412	130.396361
	eventid	Year	Month	Day	Country	State	Region	City	Latitude	Longitude																																																														
0	197000000001	1970		7	2	Dominican Republic	NaN	Central America & Caribbean	Santo Domingo	18.456792	-69.951164																																																													
1	197000000002	1970		0	0	Mexico	Federal	North America	Mexico city	19.371887	-99.086624																																																													
2	197001000001	1970		1	0	Philippines	Tarlac	Southeast Asia	Unknown	15.478598	120.599741																																																													
3	197001000002	1970		1	0	Greece	Attica	Western Europe	Athens	37.997490	23.762728																																																													
4	197001000003	1970		1	0	Japan	Fukouka	East Asia	Fukouka	33.580412	130.396361																																																													
In [30]:	df.describe()																																																																							

Out[30] :

	eventid	Year	Month	Day	Latitude	Longitude	
count	1.816910e+05	181691.000000	181691.000000	181691.000000	177135.000000	1.771340e+05	181691.000000
mean	2.002705e+11	2002.638997	6.467277	15.505644	23.498343	-4.586957e+02	
std	1.325957e+09	13.259430	3.388303	8.814045	18.569242	2.047790e+05	
min	1.970000e+11	1970.000000	0.000000	0.000000	-53.154613	-8.618590e+07	
25%	1.991021e+11	1991.000000	4.000000	8.000000	11.510046	4.545640e+00	
50%	2.009022e+11	2009.000000	6.000000	15.000000	31.467463	4.324651e+01	
75%	2.014081e+11	2014.000000	9.000000	23.000000	34.685087	6.871033e+01	
max	2.017123e+11	2017.000000	12.000000	31.000000	74.633553	1.793667e+02	

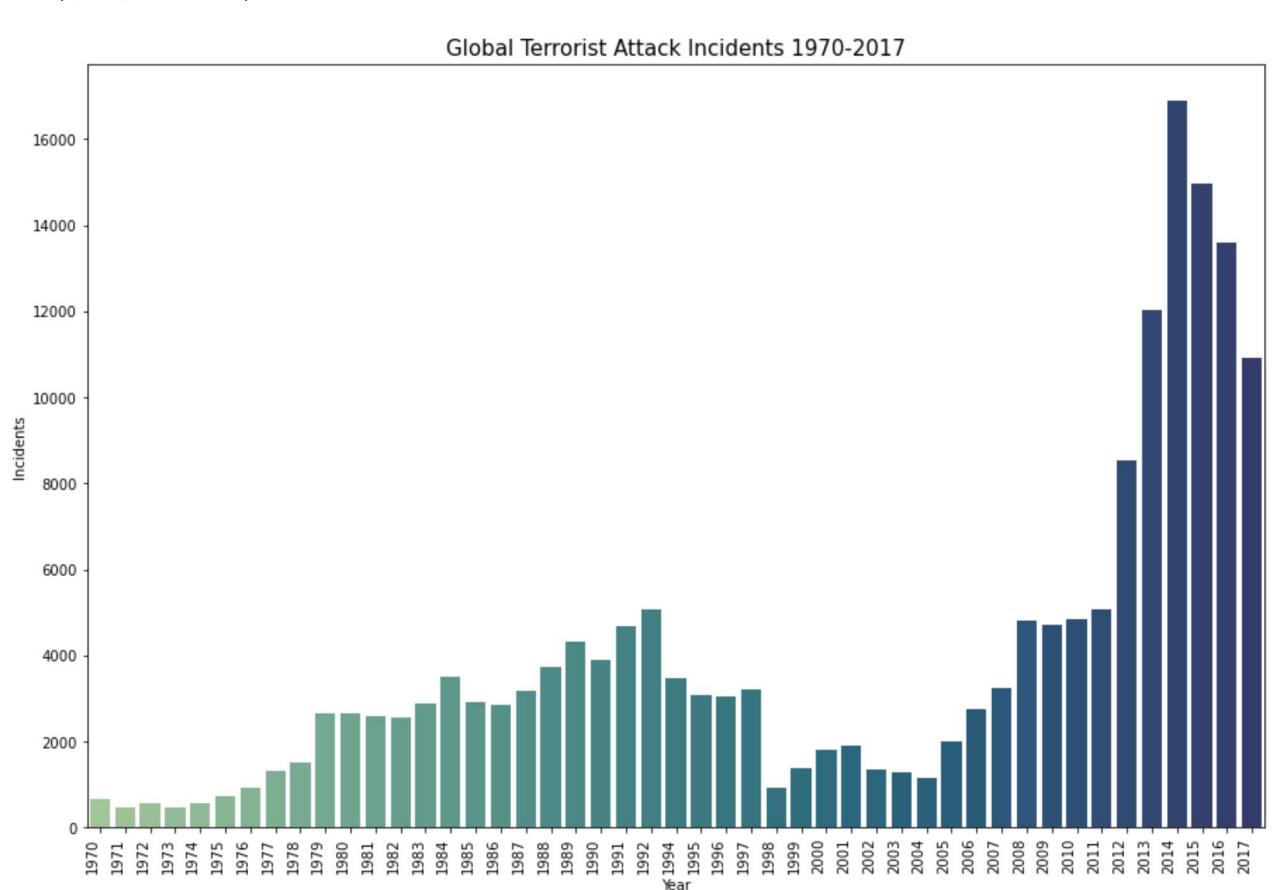


Exploratory Data Analysis Number of Global Attacks per year

In [31] :

```
plt.figure(figsize=(15,10))
ax = sns.countplot(x='Year',data=df,palette='crest')
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha="right")
plt.title('Global Terrorist Attack Incidents 1970-2017',size=15)
plt.ylabel('Incidents')
```

Out[31] :



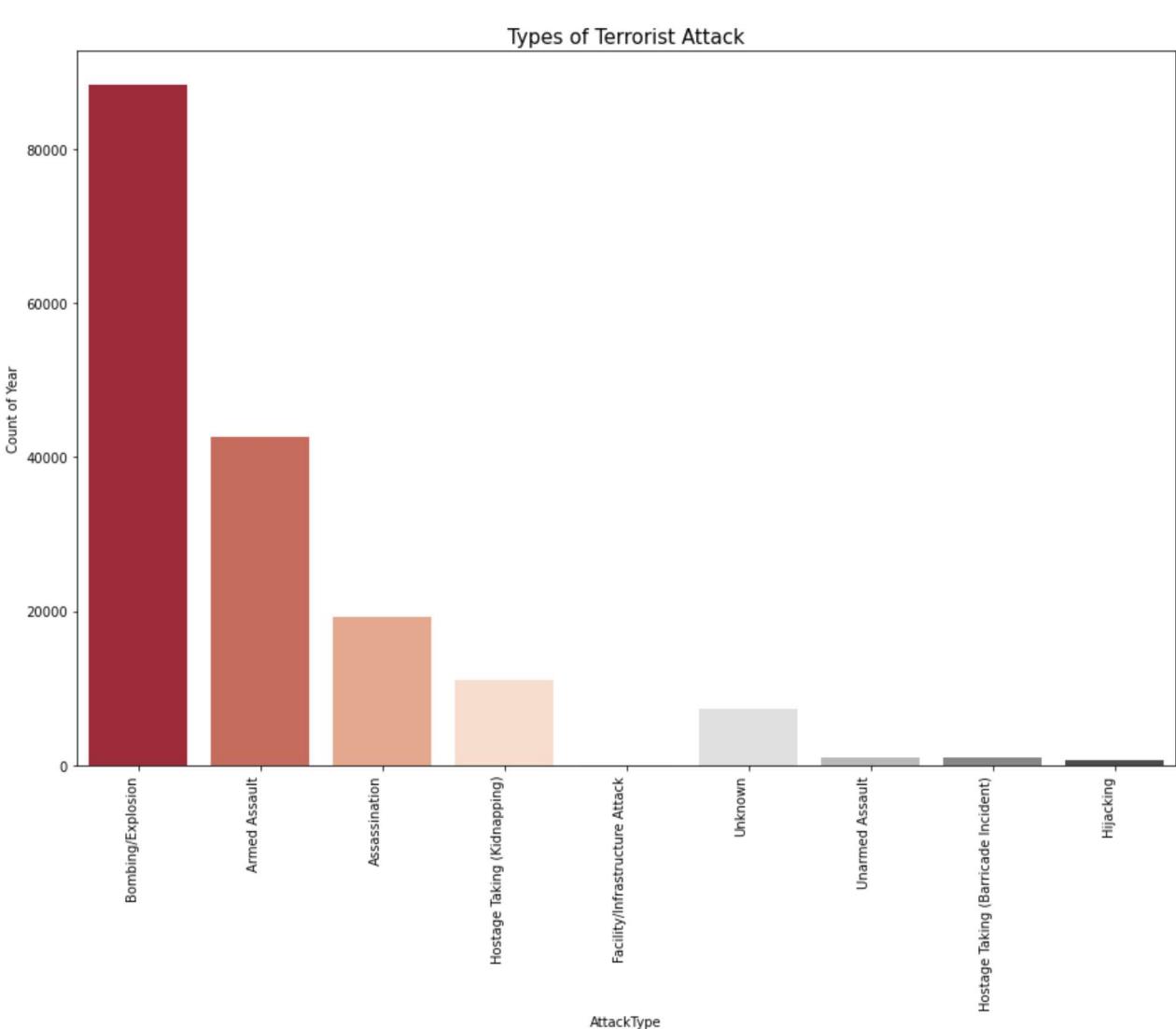
We can see here that attacks slowly start to increase from 1970 until about 1993 where until 2003 we see a downward global trend in Terrorist incidents. Then there is a sharp rise in the number of attacks/incidents. Since 2014 and 2015 has a large number of Terrorist activities.

Types of Terror Attack

In [32]:

```
plt.figure(figsize=(15,10))
ax = sns.countplot(x='AttackType',data=df,palette='RdGy',order = df['AttackType'].value_counts().index)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha="right")
plt.title('Types of Terrorist Attack',size=15)
plt.ylabel('Count of Year')
```

Out[32]:



Here we can see that in the past years the overwhelming favored form of attack has been bombings with it being the choice of terrorists more than twice that of Armed Assaults. Interestingly, hijacking is in last despite its prevalence.

Global bombings

In [33]:

```
df_bombing = df[df['AttackType'] == 'Bombing/Explosion']
```

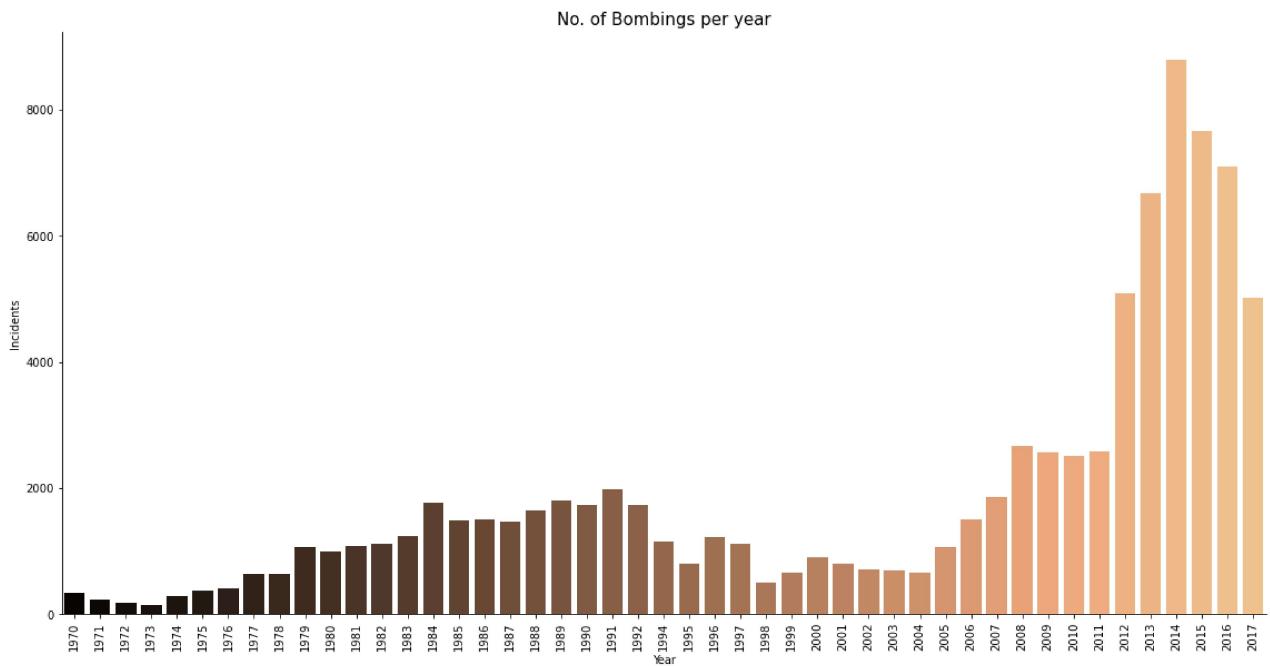
In [34]:

```

sns.catplot(x='Year',data=df_bombing,kind='count',height=8, aspect=2,palette='copper')
plt.xticks(rotation=90)
plt.title('No. of Bombings per year',size=15)
plt.ylabel('Incidents')

```

Out[34]: Text(-2.699999999999815, 0.5, 'Incidents')



We can see that the number of bombings follows roughly the same trend as global attack, but that it has a sharper decline of occurrences in the early 90s.

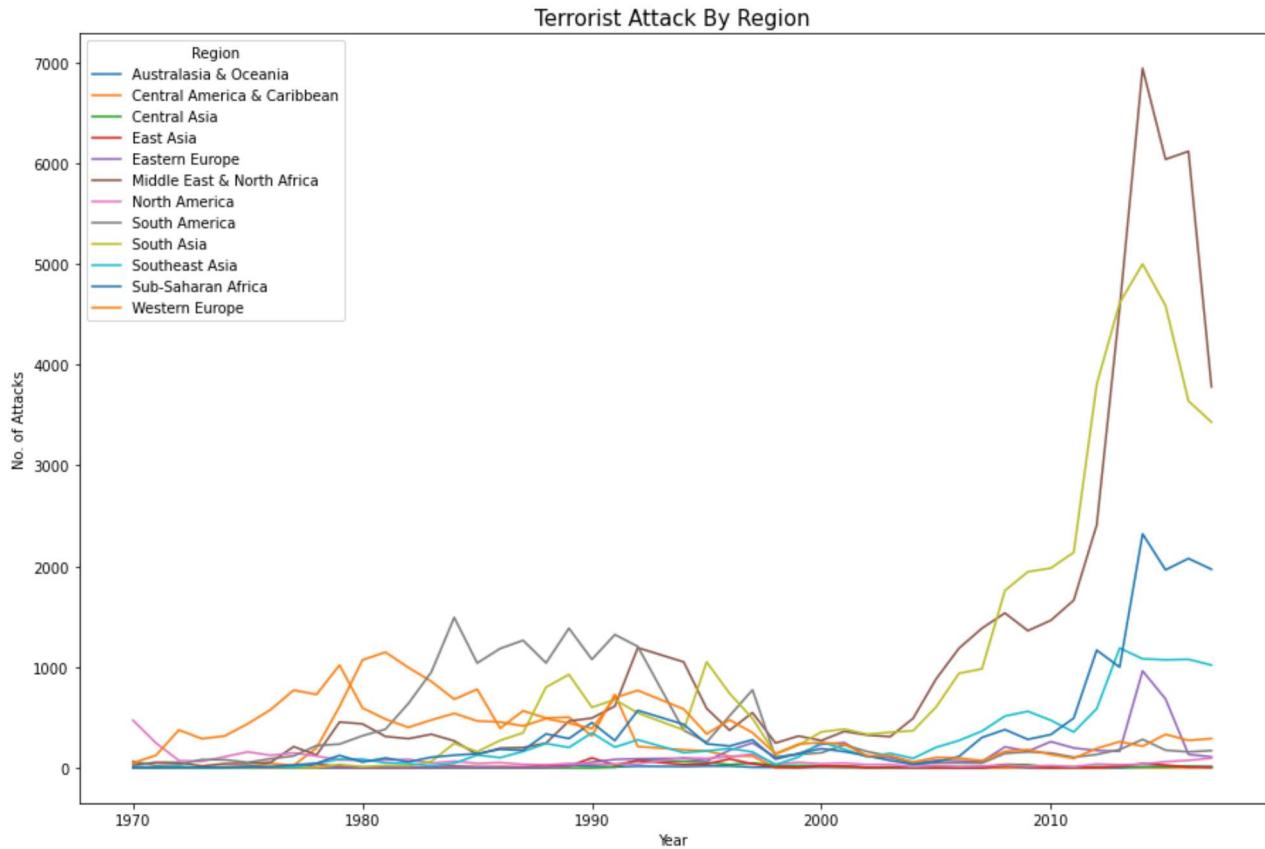
Terrorist Attack by Region

```

In [35]: pd.crosstab(df.Year,df.Region).plot(figsize=(15,10))
plt.title('Terrorist Attack By Region',size=15)
plt.ylabel('No. of Attacks')

```

Out[35]: Text(0, 0.5, 'No. of Attacks')



Here it becomes very clear how the overall trend over attacks has changed over the years in different regions across the globe. We can also see that come we experience the same jump in numbers and see that the Middle East and North Africa along with South Asia are the hardest hit by terrorist attacks.

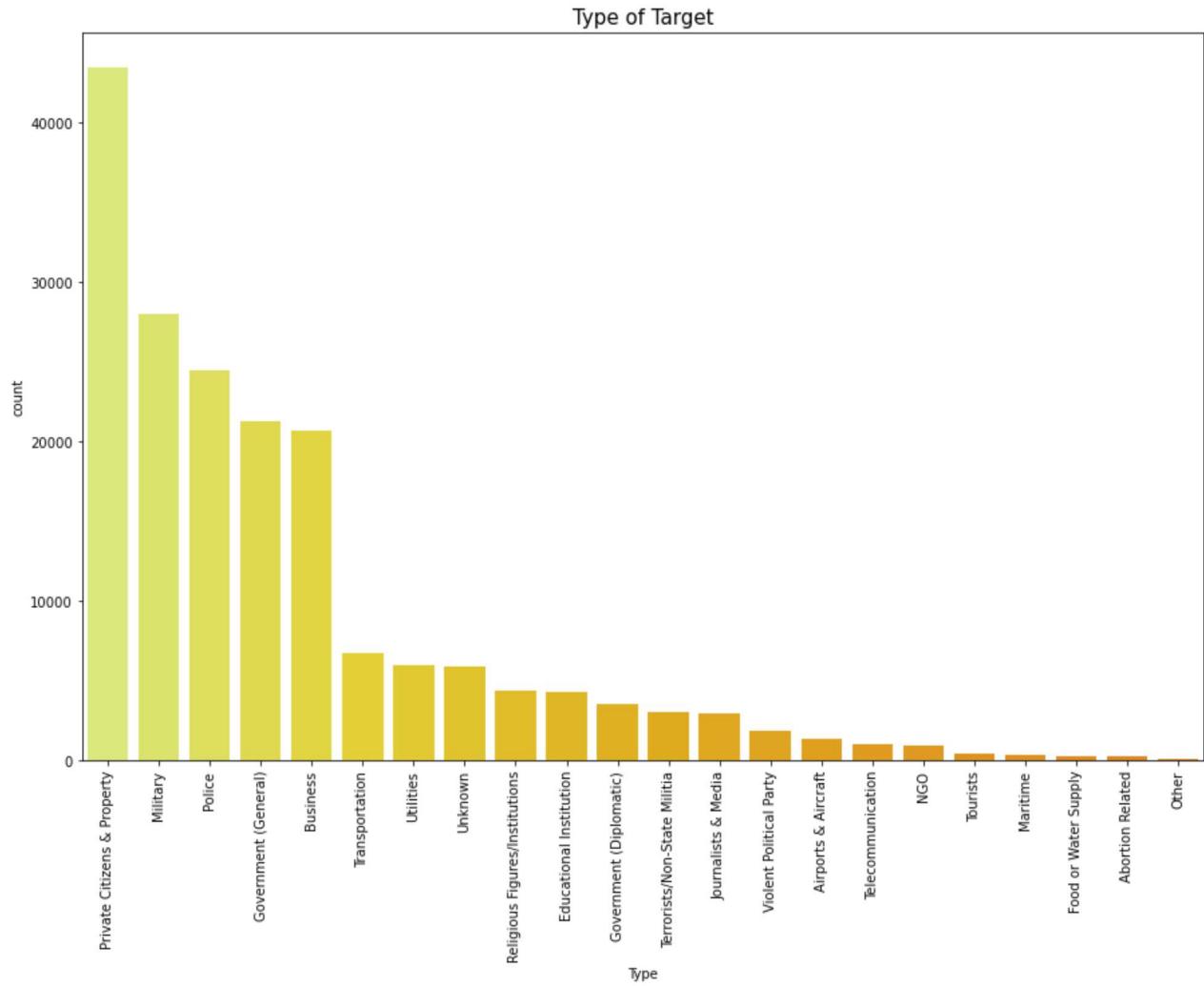
Target Type Plot

In [36]:

```
plt.figure(figsize=(15,10))
sns.countplot(df['TargetType'],order=df['TargetType'].value_counts().index,palette='Wistia')
plt.xticks(rotation=90)
plt.xlabel('Type')
plt.title('Type of Target',size=15)
```

Out[36]:

Text(0.5, 1.0, 'Type of Target')



We can see overwhelmingly that the most frequent target is civilians who as a result also suffer the most at the hands of terrorists.

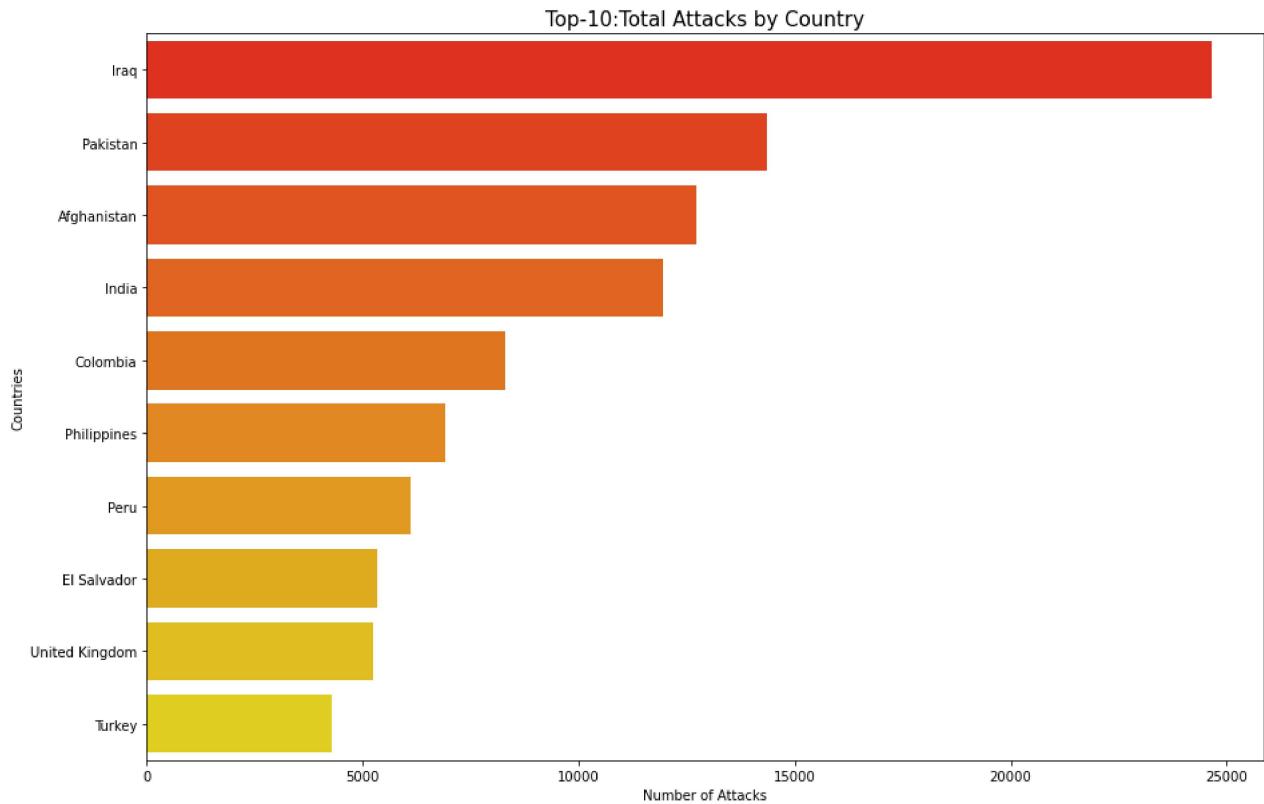
Total Number of Terrorist Attack in each country

In [37]:

```
plt.figure(figsize=(15,10))
sns.barplot(x = df['Country'].value_counts().values[:10], y = df['Country'].value_counts()[:10].index, palette = 'autumn')
plt.xlabel('Number of Attacks')
plt.ylabel('Countries')
plt.title('Top-10:Total Attacks by Country',size=15)
```

Out[37]:

Text(0.5, 1.0, 'Top-10:Total Attacks by Country')



With the plot above we can see that Iraq followed Pakistan are the countries with most terrorism occurences in 2017 (and also in all period).

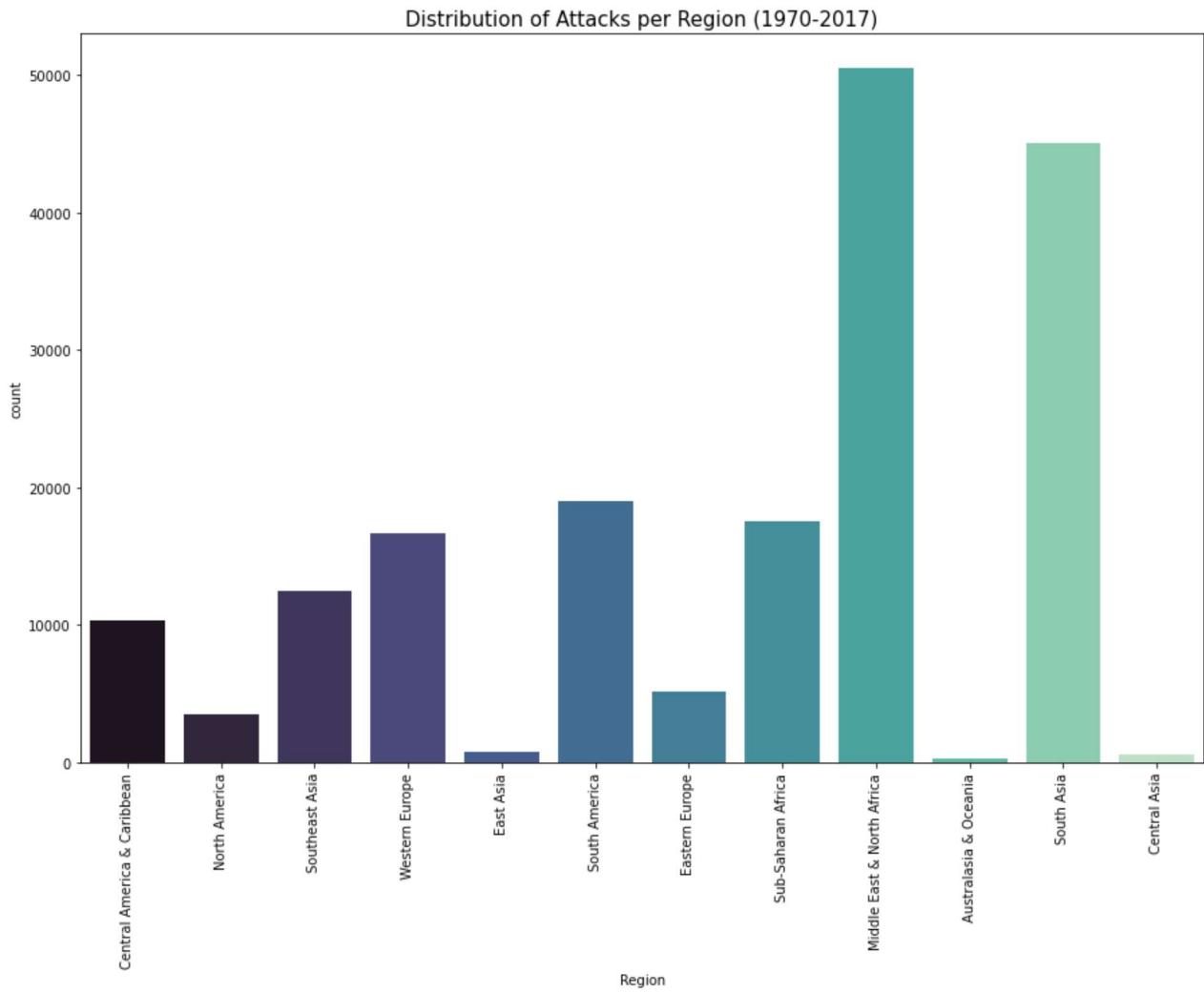
Terrorist Attack By Region

In [38]:

```
fig, ax = plt.subplots(figsize=(15, 10))
sns.countplot('Region', data=df, palette='mako')
plt.xticks(rotation=90)
plt.title('Distribution of Attacks per Region (1970-2017)', size=15)
```

Out[38]:

Text(0.5, 1.0, 'Distribution of Attacks per Region (1970-2017)')



The highest concentration of incidents recorded are from Middle East & South Asia.

Iraq Terrorism Analysis

In [39]:

```
#Preparing the data for analysis
Iraq = df[df.Country == 'Iraq']
Iraq_cities = Iraq.groupby(by='City',as_index=False).count().sort_values(by='eventid',ascending=False).iloc[:5,]

Iraq_suicide_size = Iraq['suicide'].sum() / len(Iraq)
labels = ['Suicide', 'Not Suicide']

Iraq_year = Iraq.groupby(by='Year', as_index=False).sum().loc[:, ['Year', 'Killed']]

Iraq_weapon = Iraq.groupby(by='WeaponType',as_index=False).count().sort_values(by='eventid',ascending=False).iloc[:5,]
```

In [40]:

```
fig, axs = plt.subplots(nrows=2, ncols=2, figsize=(15, 10))

# Plot 1 - Top 5 terrorism cities
sns.barplot(x='eventid', y='City', data=Iraq_cities, ci=None, ax=axs[0, 0], palette='summer')
axs[0, 0].set_title('Top 5 Iraq Cities With Most Terrorism Occurrences')
axs[0, 0].set_ylabel('City')
axs[0, 0].set_xlabel('Victims')
```

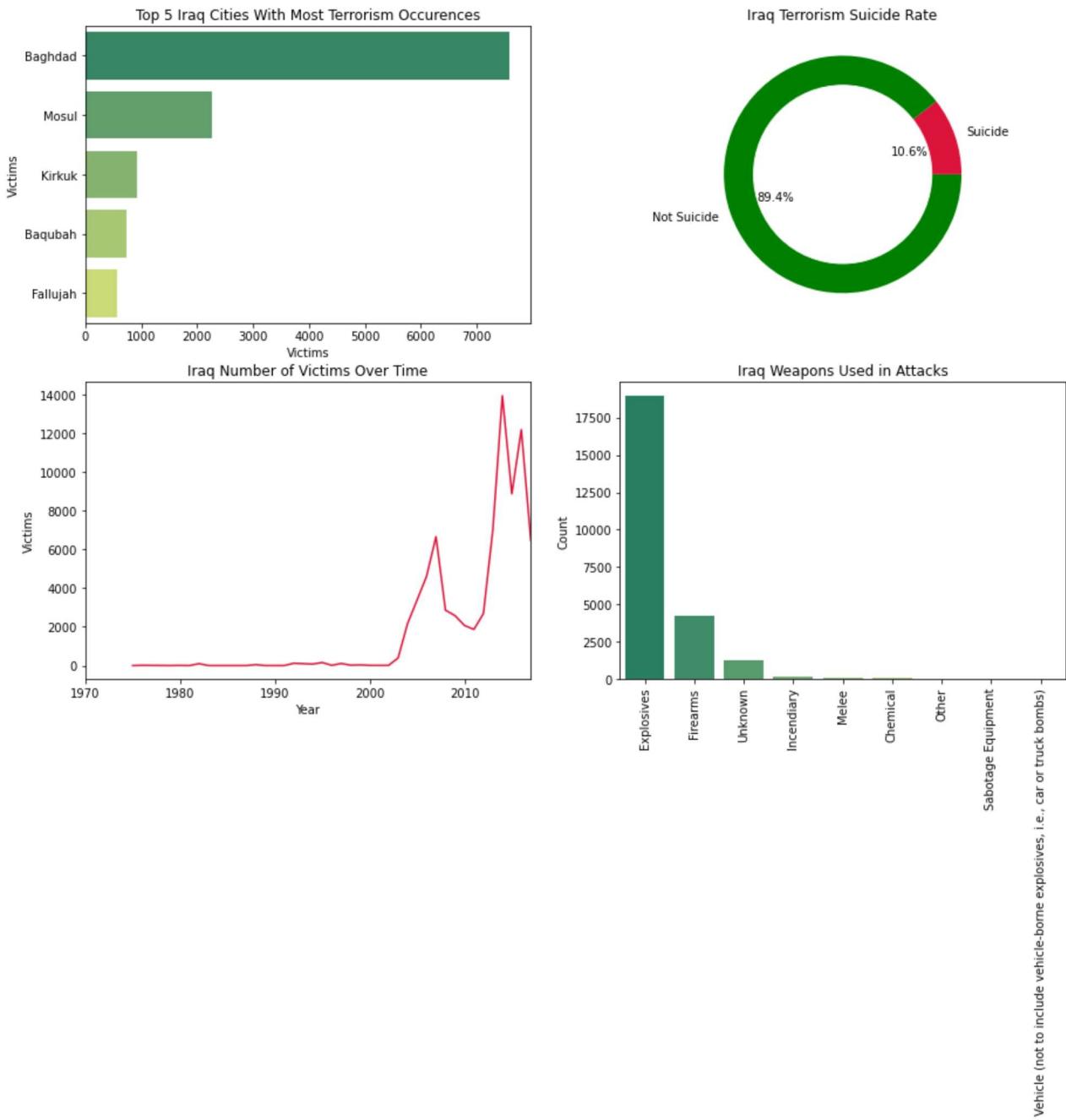
```
# Plot 2 - Suicide Rate
center_circle = plt.Circle((0,0), 0.75, color='white')
axs[0, 1].pie((Iraq_suicide_size, 1-Iraq_suicide_size), labels=labels,colors=['crimson','green'] , autopct='%1.1f%%')
axs[0, 1].add_artist(center_circle)
axs[0, 1].set_title('Iraq Terrorism Suicide Rate')
axs[0, 0].set_ylabel('Victims')

# Plot 3 - Victims through the years
sns.lineplot(x='Year', y='Killed', data=Iraq_year, ax=axs[1, 0],color='crimson')
axs[1, 0].set_xlim([1970, 2017])
axs[1, 0].set_title('Iraq Number of Victims Over Time')
axs[1, 0].set_ylabel('Victims')

# Plot 4 - Terrorism Weapons
sns.barplot(x='WeaponType', y='eventid', data=Iraq_weapon, ci=None, ax=axs[1, 1], palette='summer')
axs[1, 1].set_xticklabels(axs[1, 1].get_xticklabels(), rotation=90)
axs[1, 1].set_xlabel('')
axs[1, 1].set_ylabel('Count')
axs[1, 1].set_title('Iraq Weapons Used in Attacks')

plt.suptitle('Terrorism Analysis in Iraq between 1970 and 2017', size=16)
plt.subplots_adjust(top=0.90)
plt.show()
```

Terrorism Analysis in Iraq between 1970 and 2017



Insights Derived From EDA

1. Most attacks happened in 2014.
2. Bombing was the most favoured form of attack used.
3. The Middle East and South Asia Regions has most targeted.
4. There are maximum number of attacks in Private citizens and Property.
5. Iraq is the country where maximum terrorist attacked happened

Thank You!!!