

Let's Grow More

Name:Rashmi Ranjan Nayak

Task #1

Exploratory Data Analysis on Dataset - Terrorism:

In [1]:

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

Loading the Terrorism dataset:

In [2]:

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

C:\ProgramData\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3063: DtypeWarning: Columns (4,6,31,33,61,62,63,76,79,90,92,94,96,114,115,121) have mixed types.Specify dtype option on import or set low_memory=False.
interactivity=interactivity, compiler=compiler, result=result)

In [3]:

```
df.head()
```

Out[3]:

| | eventid | iyear | imonth | iday | approxdate | extended | resolution | country | country_txt | region | ... | addnotes | scite1 | scite2 | scite3 |
|---|---------------|-------|--------|------|------------|----------|------------|---------|--------------------|--------|-----|----------|--------|--------|--------|
| 0 | 1970000000001 | 1970 | 7 | 2 | NaN | 0 | NaN | 58 | Dominican Republic | 2 | ... | NaN | NaN | NaN | NaN |
| 1 | 1970000000002 | 1970 | 0 | 0 | NaN | 0 | NaN | 130 | Mexico | 1 | ... | NaN | NaN | NaN | NaN |
| 2 | 1970010000001 | 1970 | 1 | 0 | NaN | 0 | NaN | 160 | Philippines | 5 | ... | NaN | NaN | NaN | NaN |
| 3 | 1970010000002 | 1970 | 1 | 0 | NaN | 0 | NaN | 78 | Greece | 8 | ... | NaN | NaN | NaN | NaN |
| 4 | 1970010000003 | 1970 | 1 | 0 | NaN | 0 | NaN | 101 | Japan | 4 | ... | NaN | NaN | NaN | NaN |

5 rows x 135 columns

In [4]:

```
df.shape
```

Out[4]:

(181691, 135)

Rename columns according to user friendly:

In [5]:

```
df.rename(columns={'iyear': 'Year', 'imonth': 'Month', 'iday': 'Day', 'city': 'City', 'country_txt': 'Country', 'region_txt': 'Region', 'attacktype1_txt': 'AttackType', 'targettype1_txt': 'TargetType', 'weaptype1_txt': 'WeaponType', 'gname': 'Group', 'dbsource': 'Source'}, inplace = True)
```

In [6]:

```
df['Casualties'] = df.nkill + df.nwound ## Add a extra column as Casualties
```

In [7]:

```
df.head()
```

Out[7]:

| | eventid | Year | Month | Day | approxdate | extended | resolution | country | Country | region | ... | scite1 | scite2 | scite3 | Source | IN1 |
|---|---------------|------|-------|-----|------------|----------|------------|---------|--------------------|--------|-----|--------|--------|--------|--------|-----|
| 0 | 1970000000001 | 1970 | 7 | 2 | NaN | 0 | NaN | 58 | Dominican Republic | 2 | ... | NaN | NaN | NaN | PGIS | |
| 1 | 1970000000002 | 1970 | 0 | 0 | NaN | 0 | NaN | 130 | Mexico | 1 | ... | NaN | NaN | NaN | PGIS | |
| 2 | 1970010000001 | 1970 | 1 | 0 | NaN | 0 | NaN | 160 | Philippines | 5 | ... | NaN | NaN | NaN | PGIS | |
| 3 | 1970010000002 | 1970 | 1 | 0 | NaN | 0 | NaN | 78 | Greece | 8 | ... | NaN | NaN | NaN | PGIS | |
| 4 | 1970010000003 | 1970 | 1 | 0 | NaN | 0 | NaN | 101 | Japan | 4 | ... | NaN | NaN | NaN | PGIS | |

5 rows × 136 columns

Selecting the required features:

In [8]:

```
df = df[['Year', 'Month', 'Day', 'Country', 'Region', 'City', 'latitude', 'longitude', 'AttackType', 'TargetType', 'WeaponType', 'nkill', 'nwound', 'Source', 'Group', 'Casualties']]
```

In [9]:

```
df.head()
```

Out[9]:

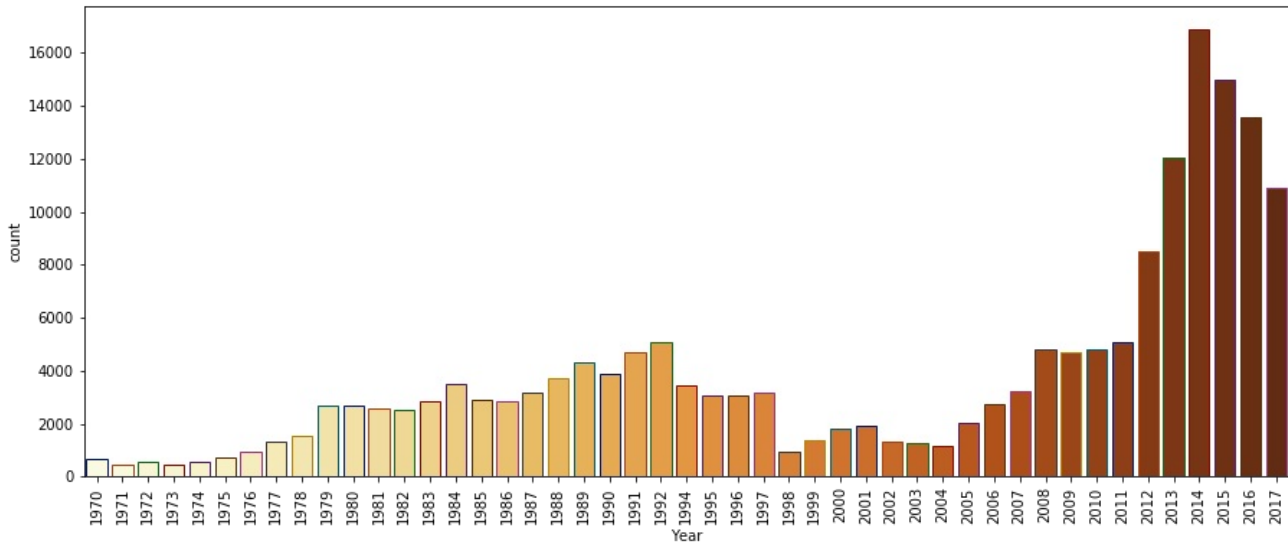
| | Year | Month | Day | Country | Region | City | latitude | longitude | AttackType | TargetType | WeaponType | nkill | nwound |
|---|------|-------|-----|--------------------|-----------------------------|---------------|-----------|------------|--------------------------------|-----------------------------|------------|-------|--------|
| 0 | 1970 | 7 | 2 | Dominican Republic | Central America & Caribbean | Santo Domingo | 18.456792 | -69.951164 | Assassination | Private Citizens & Property | Unknown | 1.0 | |
| 1 | 1970 | 0 | 0 | Mexico | North America | Mexico city | 19.371887 | -99.086624 | Hostage Taking (Kidnapping) | Government (Diplomatic) | Unknown | 0.0 | |
| 2 | 1970 | 1 | 0 | Philippines | Southeast Asia | Unknown | 15.478598 | 120.599741 | Assassination | Journalists & Media | Unknown | 1.0 | |
| 3 | 1970 | 1 | 0 | Greece | Western Europe | Athens | 37.997490 | 23.762728 | Bombing/Explosion | Government (Diplomatic) | Explosives | NaN | |
| 4 | 1970 | 1 | 0 | Japan | East Asia | Fukouka | 33.580412 | 130.396361 | Facility/Infrastructure Attack | Government (Diplomatic) | Incendiary | NaN | |

Count of terrorist attacks year wise:

In [10]:

```
plt.subplots(figsize=(15,6))
sns.countplot('Year',data=df,palette='YlOrBr',edgecolor=sns.color_palette('dark',10))
plt.xticks(rotation=90)
plt.title('Number of Terrorist Activities Each Year')
plt.show()
```

Number of Terrorist Activities Each Year

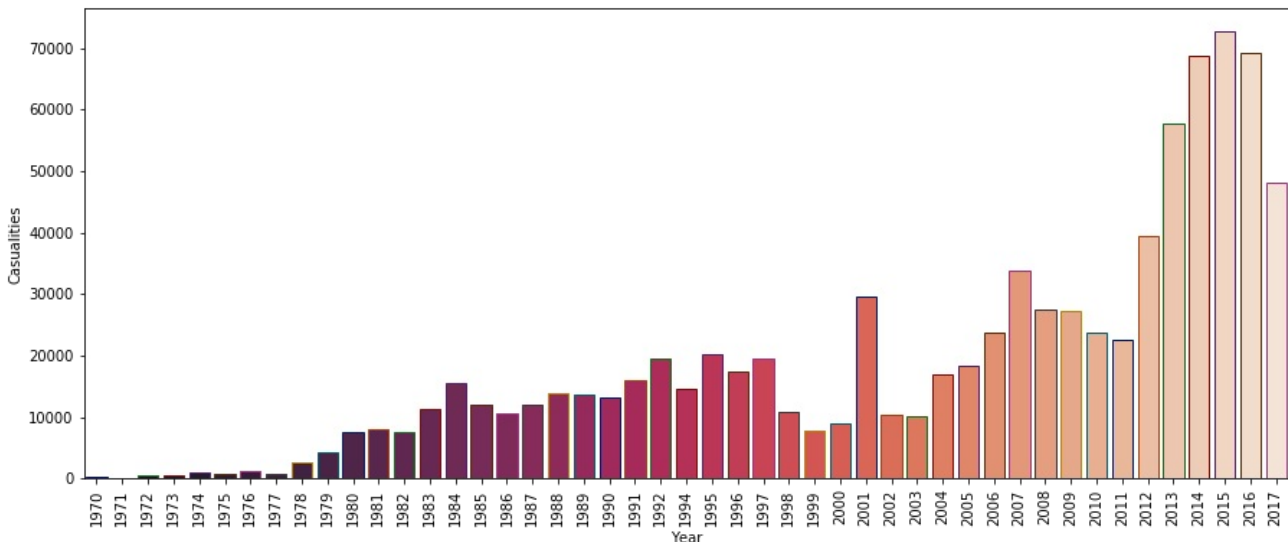


Casualties :: Years :

In [11]:

```
plt.subplots(figsize=(15,6))
year_cas = df.groupby('Year').Casualties.sum().to_frame().reset_index()
year_cas.columns = ['Year','Casualties']
sns.barplot(x=year_cas.Year, y=year_cas.Casualties, palette='rocket',edgecolor=sns.color_palette('dark',10))
plt.xticks(rotation=90)
plt.title('Number of Casualties Each Year')
plt.show()
```

Number of Casualties Each Year



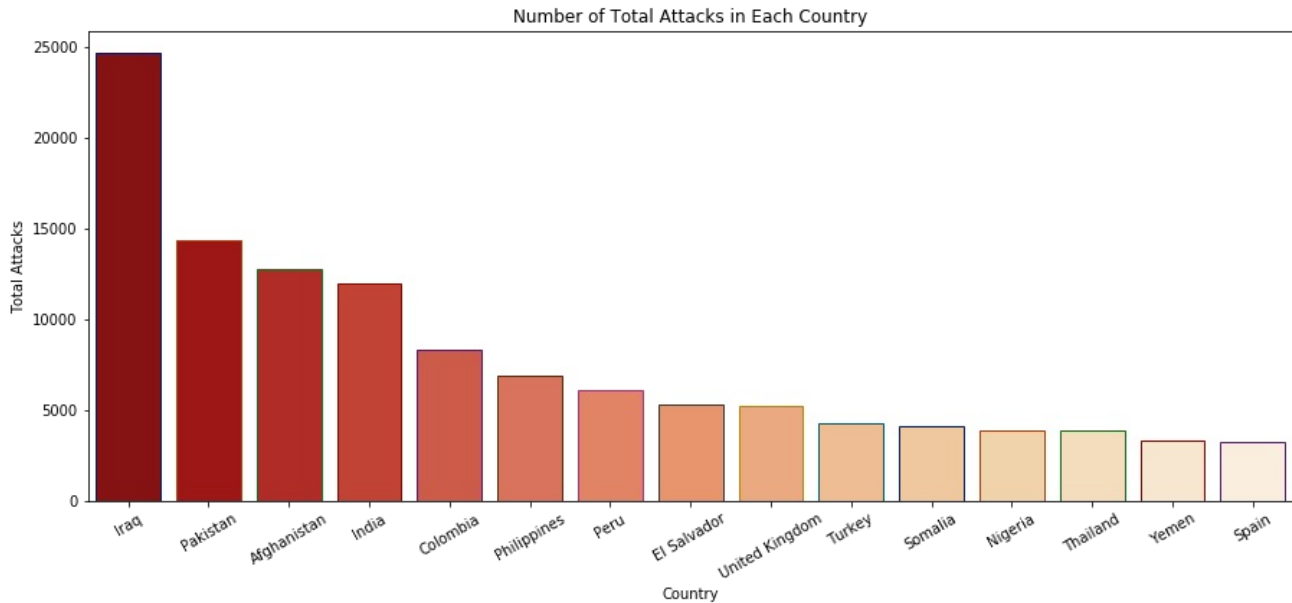
Attacks :: Countries

As you can see #Iraq has maximum number of attacks so, Iraq has to implement more on deffence system

Iraq >> Afganistan >> Pak [TOP #3]

In [12]:

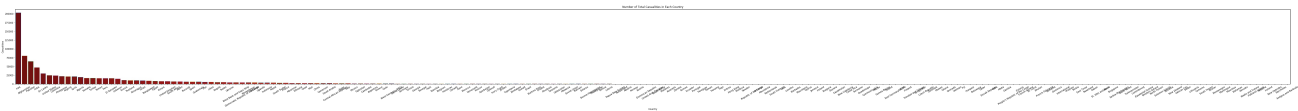
```
plt.subplots(figsize=(15,6))
country_attacks = df.Country.value_counts()[15].reset_index()
country_attacks.columns = ['Country','Total Attacks']
sns.barplot(x=country_attacks.Country, y=country_attacks['Total Attacks'], palette='OrRd_r',edgecolor=sns.color_palette('dark',10))
plt.xticks(rotation=30)
plt.title('Number of Total Attacks in Each Country')
plt.show()
```



Iraq has most attacks so Number of Casualities is also high

In [13]:

```
plt.subplots(figsize=(100,6))
count_cas = df.groupby('Country').Casualities.sum().to_frame().reset_index().sort_values(by='Casualities',ascending=False)
sns.barplot(x=count_cas.Country, y=count_cas.Casualities, palette='OrRd_r',edgecolor=sns.color_palette('dark',10))
plt.xticks(rotation=30)
plt.title('Number of Total Casualities in Each Country')
plt.show()
```

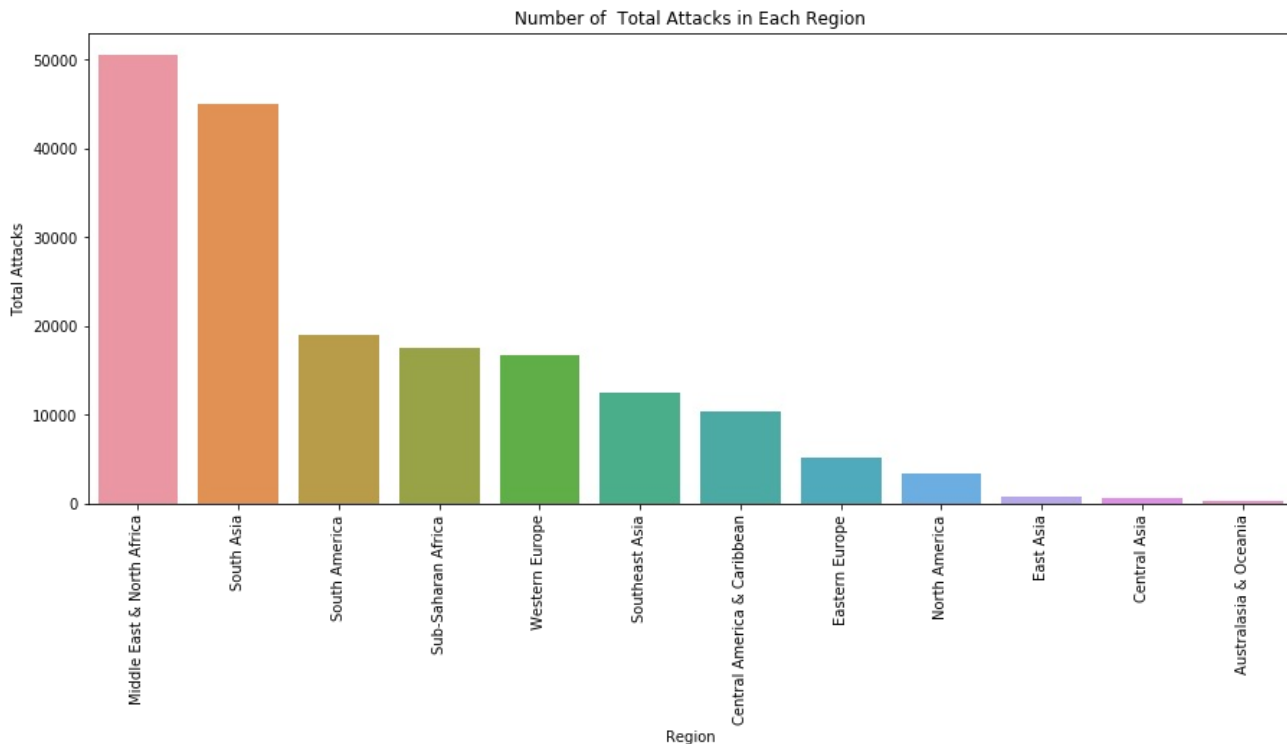


Region wise attack graph:

Red-zone :- Middle-East and North Africa

In [14]:

```
region_attacks = df.Region.value_counts().to_frame().reset_index()
region_attacks.columns = ['Region', 'Total Attacks']
plt.subplots(figsize=(15,6))
sns.barplot(x=region_attacks.Region, y=region_attacks['Total Attacks'])
plt.xticks(rotation=90)
plt.title('Number of Total Attacks in Each Region')
plt.show()
```



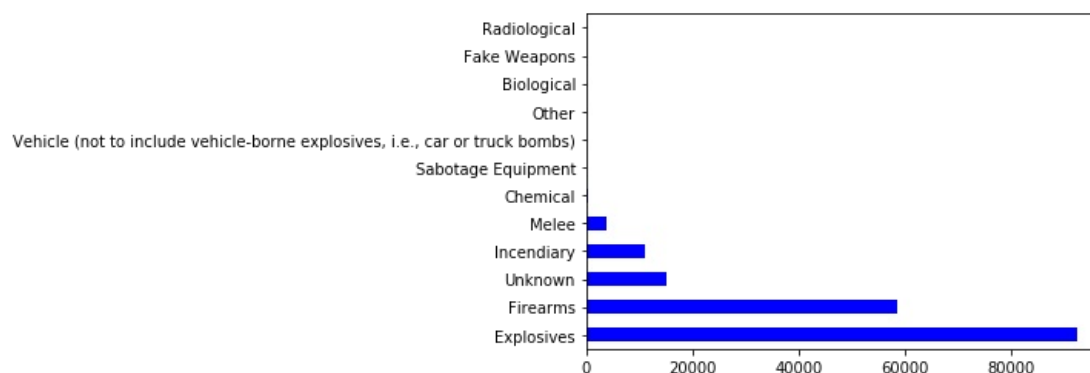
Explosives are widely used for terrorist attacks: As we can see from the bellow graph:

In [15]:

```
df['WeaponType'].value_counts().plot(kind='barh',color="blue")
```

Out[15]:

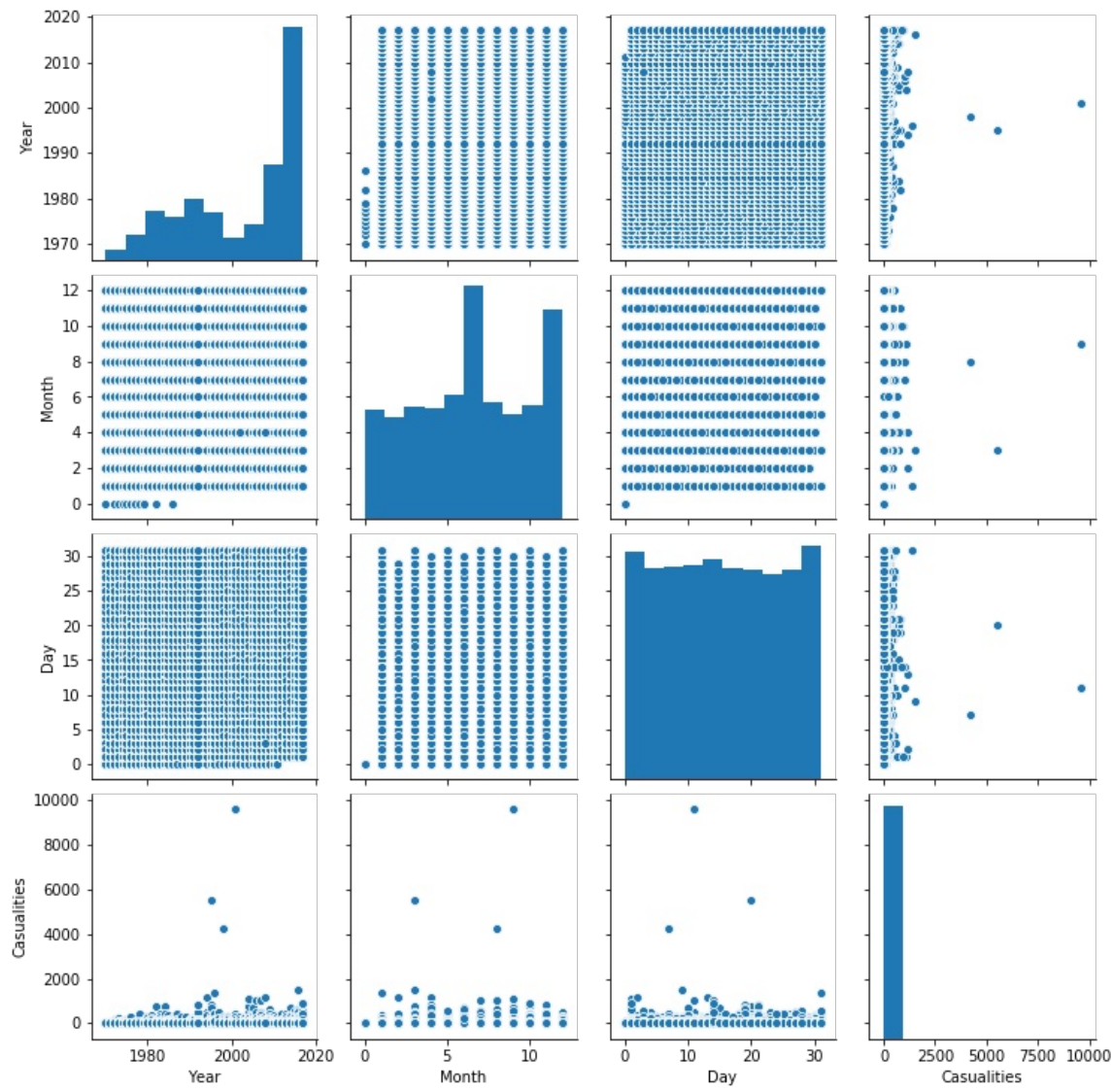
<matplotlib.axes._subplots.AxesSubplot at 0x26f22ac7748>



As we can see from the (Year-Casualities) graph that most of the yearly attacks has under 2000 casualties

```
In [16]:
sns.pairplot(df, vars=['Year', 'Month', 'Day', 'Casualties'])
```

Out[16]:
<seaborn.axisgrid.PairGrid at 0x26f22bb2088>

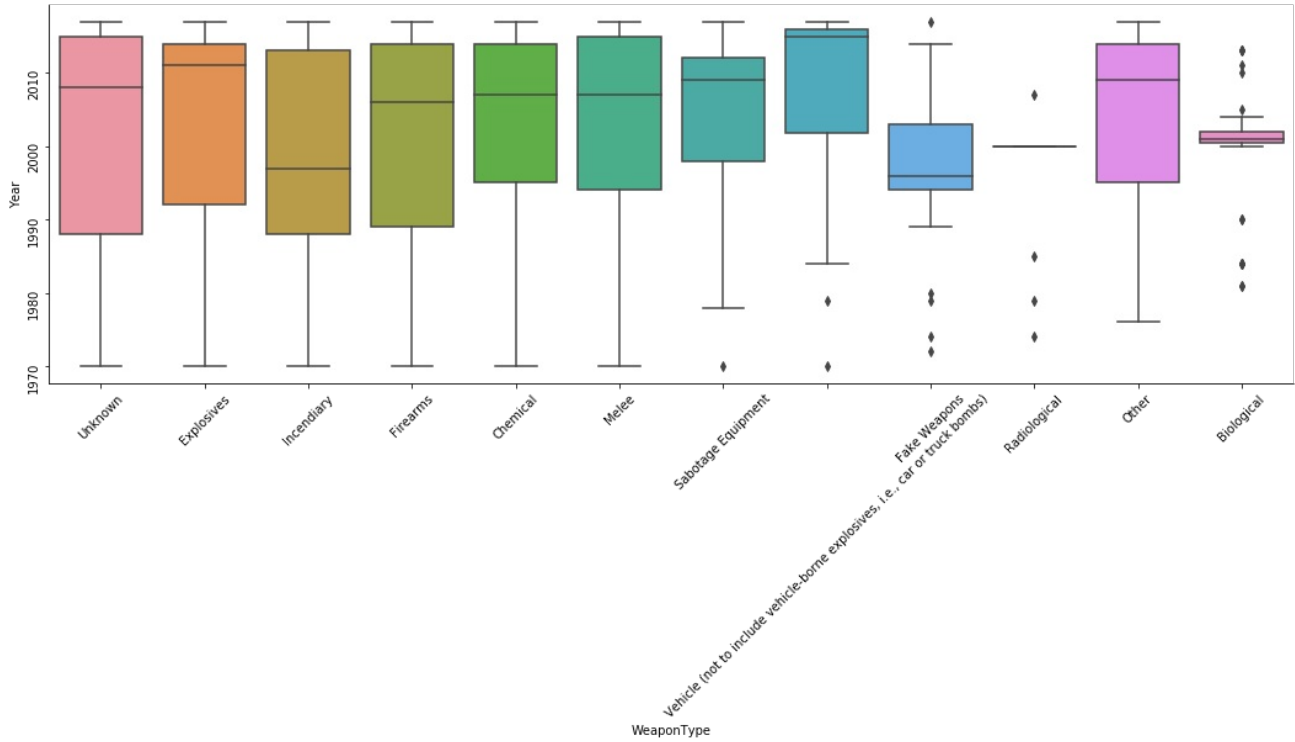


In [17]:

```
chart = sns.catplot(x='WeaponType',y='Year', kind='box', height=5, aspect=3, data=df,orient='v')
chart.set_xticklabels(rotation=45)
chart.set_yticklabels(rotation=90)
```

Out[17]:

<seaborn.axisgrid.FacetGrid at 0x26f22a76fc8>



In []: