Letsgrowmore Task 2-Exploratory Data Analysis on Dataset - Terrorism
Done by:Samyuktha Rajkumaran

Understanding our data

In [18]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy as sp
import warnings
warnings.filterwarnings("ignore")
```

In [19]:

```
data=pd.read_csv(r"C:\Users\Rajkumar\Downloads\terrorism.csv",encoding='ISO-8859-1')
data.round(5)
```

Out[19]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes
0	197000000000.000	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN
1	197000000000.000	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN
2	197000000000.000	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN
3	197000000000.000	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN
4	197000000000.000	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
181686	202000000000.000	2017	12	31	NaN	0	NaN	182	Somalia	11	...	NaN
181687	202000000000.000	2017	12	31	NaN	0	NaN	200	Syria	10	...	NaN
181688	202000000000.000	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN
181689	202000000000.000	2017	12	31	NaN	0	NaN	92	India	6	...	NaN
181690	202000000000.000	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN

181691 rows × 135 columns

In [20]:

```
data
```

Out[20]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes
0	197000000000.000	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN
1	197000000000.000	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN
2	197000000000.000	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN
3	197000000000.000	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN
4	197000000000.000	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
181686	202000000000.000	2017	12	31	NaN	0	NaN	182	Somalia	11	...	NaN
181687	202000000000.000	2017	12	31	NaN	0	NaN	200	Syria	10	...	NaN
181688	202000000000.000	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN
181689	202000000000.000	2017	12	31	NaN	0	NaN	92	India	6	...	NaN
181690	202000000000.000	2017	12	31	NaN	0	NaN	160	Philippines	5	...	NaN

181691 rows × 135 columns

In [21]:

```
data.shape
```

Out[21]:

```
(181691, 135)
```

In [22]:

```
data.describe()
```

Out[22]:

	eventid	iyear	imonth	iday	extended	country	region	latitude	longitude
count	181691.000	181691.000	181691.000	181691.000	181691.000	181691.000	181691.000	177135.000	177134.000
mean	200323758469.049	2002.639	6.467	15.506	0.045	131.969	7.161	23.498	-458.696
std	1383522764.661	13.259	3.388	8.814	0.208	112.415	2.933	18.569	204778.989
min	197000000000.000	1970.000	0.000	0.000	0.000	4.000	1.000	-53.155	-86185896.000

	eventid	iyear	imonth	iday	extended	country	region	latitude	longitude
25%	199000000000.000	1991.000	4.000	8.000	0.000	78.000	5.000	11.510	4.546
50%	201000000000.000	2009.000	6.000	15.000	0.000	98.000	6.000	31.467	43.247
75%	201000000000.000	2014.000	9.000	23.000	0.000	160.000	10.000	34.685	68.710
max	202000000000.000	2017.000	12.000	31.000	1.000	1004.000	12.000	74.634	179.367

8 rows × 77 columns

In [23]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 135 entries, eventid to related
dtypes: float64(56), int64(21), object(58)
memory usage: 187.1+ MB
```

In [24]:

```
data.value_counts
```

Out[24]:

```
<bound method DataFrame.value_counts of               eventid  iyear  imonth  iday appr
oxdate  extended resolution  \
0      197000000000.000   1970       7     2         NaN         0         NaN
1      197000000000.000   1970       0     0         NaN         0         NaN
2      197000000000.000   1970       1     0         NaN         0         NaN
3      197000000000.000   1970       1     0         NaN         0         NaN
4      197000000000.000   1970       1     0         NaN         0         NaN
...                 ...    ...     ...   ...         ...       ...         ...
181686 202000000000.000   2017      12    31         NaN         0         NaN
181687 202000000000.000   2017      12    31         NaN         0         NaN
181688 202000000000.000   2017      12    31         NaN         0         NaN
181689 202000000000.000   2017      12    31         NaN         0         NaN
181690 202000000000.000   2017      12    31         NaN         0         NaN

        country        country_txt  region  ... addnotes  \
0            58  Dominican Republic       2  ...      NaN
1           130              Mexico       1  ...      NaN
2           160         Philippines       5  ...      NaN
3            78              Greece       8  ...      NaN
4           101               Japan       4  ...      NaN
...         ...                 ...     ...  ...      ...
181686      182             Somalia      11  ...      NaN
181687      200               Syria      10  ...      NaN
181688      160         Philippines       5  ...      NaN
181689       92               India       6  ...      NaN
181690      160         Philippines       5  ...      NaN

                                                   scite1  \
0                                                     NaN
1                                                     NaN
2                                                     NaN
3                                                     NaN
4                                                     NaN
...                                                   ...
181686  "Somalia: Al-Shabaab Militants Attack Army Che...
181687  "Putin's 'victory' in Syria has turned into a ...
181688  "Maguindanao clashes trap tribe members," Phil...
181689  "Trader escapes grenade attack in Imphal," Bus...
181690  "Security tightened in Cotabato following IED ...

                                                   scite2  \
0                                                     NaN
1                                                     NaN
2                                                     NaN
3                                                     NaN
4                                                     NaN
```

```
...                                                            ...
181686  "Highlights: Somalia Daily Media Highlights 2 ...
181687  "Two Russian soldiers killed at Hmeymim base i...
181688                                                         NaN
181689                                                         NaN
181690  "Security tightened in Cotabato City," Manila ...

                                                    scite3  \
0                                                      NaN
1                                                      NaN
2                                                      NaN
3                                                      NaN
4                                                      NaN
...                                                    ...
181686  "Highlights: Somalia Daily Media Highlights 1 ...
181687  "Two Russian servicemen killed in Syria mortar...
181688                                                 NaN
181689                                                 NaN
181690                                                 NaN

                        dbsource  INT_LOG  INT_IDEO INT_MISC INT_ANY  related
0                           PGIS        0         0        0       0      NaN
1                           PGIS        0         1        1       1      NaN
2                           PGIS       -9        -9        1       1      NaN
3                           PGIS       -9        -9        1       1      NaN
4                           PGIS       -9        -9        1       1      NaN
...                          ...      ...       ...      ...     ...      ...
181686  START Primary Collection        0         0        0       0      NaN
181687  START Primary Collection       -9        -9        1       1      NaN
181688  START Primary Collection        0         0        0       0      NaN
181689  START Primary Collection       -9        -9        0      -9      NaN
181690  START Primary Collection       -9        -9        0      -9      NaN

[181691 rows x 135 columns]>
```

In [25]:

```
data.columns
```

Out[25]:

```
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 [26]:

```
data.isnull().sum()
```

Out[26]:

```
eventid             0
iyear               0
imonth              0
iday                0
approxdate     172452
                ...
INT_LOG             0
INT_IDEO            0
INT_MISC            0
INT_ANY             0
related        156653
Length: 135, dtype: int64
```

In [27]:

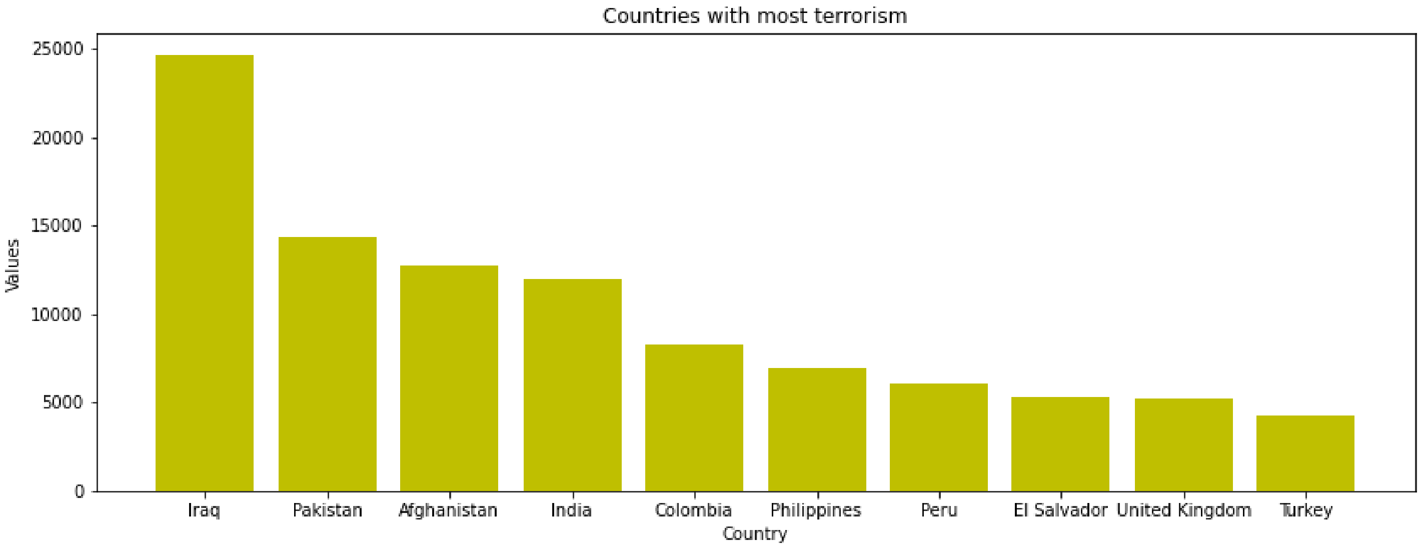
```
countries_with_most_terrorism = data.country_txt.value_counts().head(10)
countries = list(countries_with_most_terrorism.index)
```

Visualizing the data

Most Affected Country

In [34]:

```
fig, ax = plt.subplots(figsize=(14,5))
ax.bar(countries_with_most_terrorism.index,countries_with_most_terrorism.values,color='y'
)
plt.title('Countries with most terrorism')
plt.xlabel('Country')
plt.ylabel('Values')
plt.show()
```



In [29]:

```
print(countries)
```

```
['Iraq', 'Pakistan', 'Afghanistan', 'India', 'Colombia', 'Philippines', 'Peru', 'El Salva
dor', 'United Kingdom', 'Turkey']
```

Therefore,it's evident that Iraq is the most affected country.

In [30]:

```
data.iyear.value_counts().head()
```

Out[30]:

```
2014    16903
2015    14965
2016    13587
2013    12036
2017    10900
Name: iyear, dtype: int64
```

In [31]:

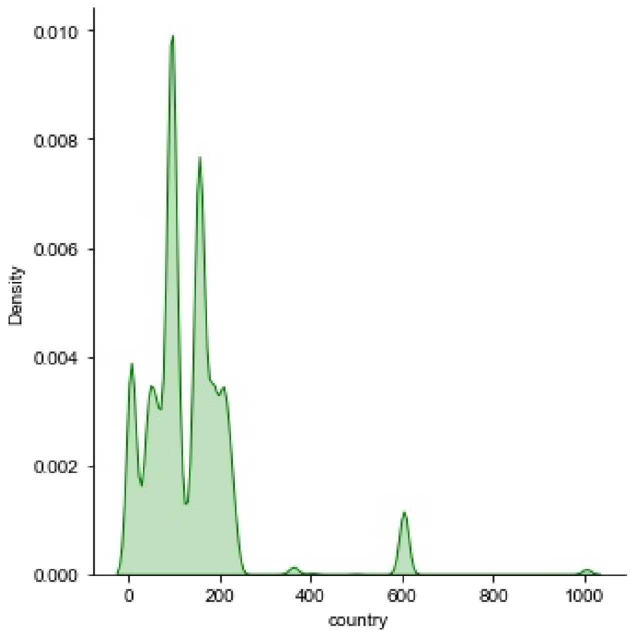
```
correlation=data.corr()
print(correlation)
```

```
           eventid  iyear  imonth   iday  extended  country  region  latitude  \
eventid      1.000  0.974  -0.003  0.018     0.092   -0.135   0.391     0.160
iyear        0.974  1.000   0.000  0.018     0.092   -0.135   0.401     0.167
imonth      -0.003  0.000   1.000  0.005    -0.000   -0.006  -0.003    -0.016
iday         0.018  0.018   0.005  1.000    -0.005    0.003   0.010     0.003
extended     0.092  0.092  -0.000 -0.005     1.000   -0.020   0.038    -0.025
...            ...    ...     ...    ...       ...      ...     ...       ...
nreleased   -0.179 -0.182  -0.012  0.002    -0.192   -0.044  -0.150     0.003
INT_LOG     -0.126 -0.144  -0.002 -0.002     0.072    0.070  -0.083    -0.100
INT_IDEO    -0.116 -0.133  -0.002 -0.002     0.075    0.068  -0.072    -0.094
INT_MISC    -0.082 -0.078  -0.003 -0.002     0.027    0.207   0.043     0.098
```

```
INT_ANY      -0.159 -0.176  -0.006 -0.001     0.081    0.153  -0.048    -0.042

          longitude  specificity  ...  ransomamt  ransomamtus  ransompaid  \
eventid       0.004        0.032  ...     -0.013       -0.036      -0.007
iyear         0.004        0.031  ...     -0.010       -0.018      -0.014
imonth       -0.004        0.004  ...     -0.001        0.047       0.059
iday         -0.002       -0.007  ...      0.013       -0.011       0.003
extended      0.001        0.058  ...     -0.008        0.028       0.002
...             ...          ...  ...        ...          ...         ...
nreleased    -0.018       -0.031  ...      0.055        0.035       0.049
INT_LOG       0.002        0.073  ...      0.036        0.031       0.007
INT_IDEO      0.002        0.071  ...      0.039        0.042       0.013
INT_MISC      0.000       -0.019  ...      0.024        0.125       0.037
INT_ANY       0.002        0.061  ...      0.028        0.053       0.007

          ransompaidus  hostkidoutcome  nreleased  INT_LOG  INT_IDEO  \
eventid         -0.112           0.252     -0.179   -0.126    -0.116
iyear           -0.165           0.256     -0.182   -0.144    -0.133
imonth          -0.017           0.011     -0.012   -0.002    -0.002
iday            -0.007          -0.007      0.002   -0.002    -0.002
extended         0.009           0.233     -0.192    0.072     0.075
...                ...             ...        ...      ...       ...
nreleased        0.017          -0.555      1.000    0.039     0.041
INT_LOG         -0.046          -0.015      0.039    1.000     0.996
INT_IDEO        -0.040          -0.016      0.041    0.996     1.000
INT_MISC         0.129          -0.120      0.085    0.053     0.082
INT_ANY          0.056          -0.062      0.065    0.891     0.894

          INT_MISC  INT_ANY
eventid     -0.082   -0.159
iyear       -0.078   -0.176
imonth      -0.003   -0.006
iday        -0.002   -0.001
extended     0.027    0.081
...            ...      ...
nreleased    0.085    0.065
INT_LOG      0.053    0.891
INT_IDEO     0.082    0.894
INT_MISC     1.000    0.252
INT_ANY      0.252    1.000

[77 rows x 77 columns]
```

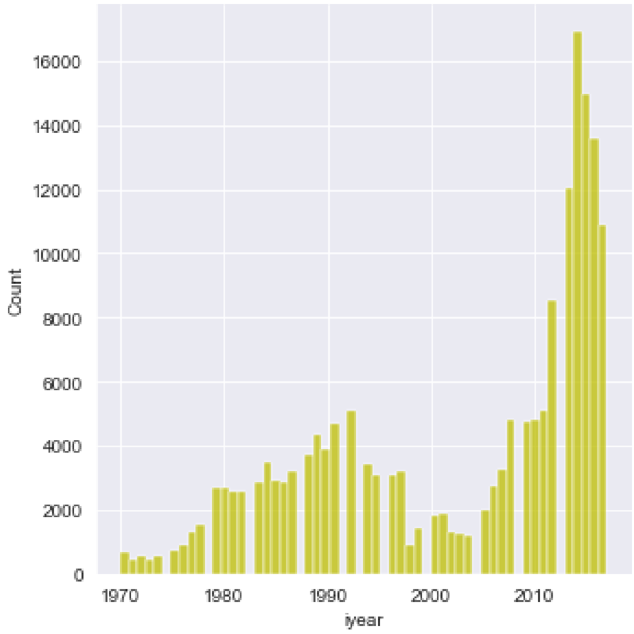
In [36]:

```
sns.displot(data,x='country',kind='kde',fill='tree',palette='colorblind',color='g')
sns.set_style("darkgrid")
```



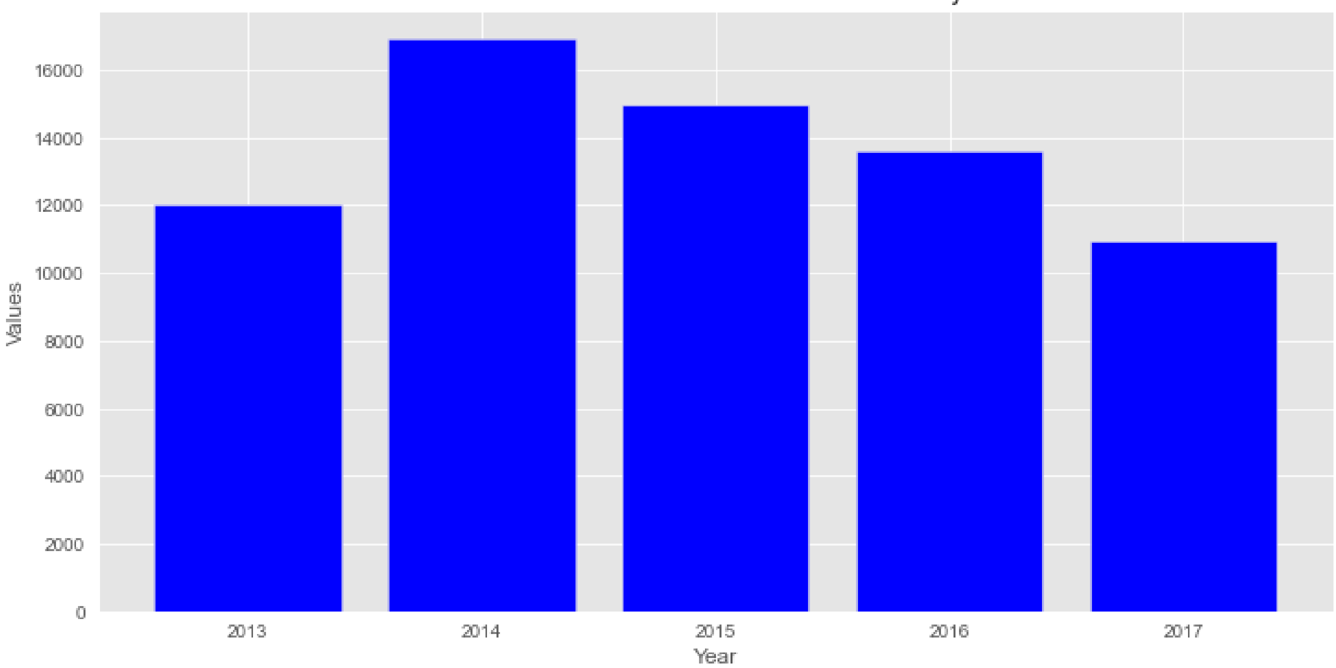
In [37]:

```
sns.displot(data,x='iyear',kind='hist',palette='colorblind',color='y')
plt.style.use("ggplot")
```



In [38]:

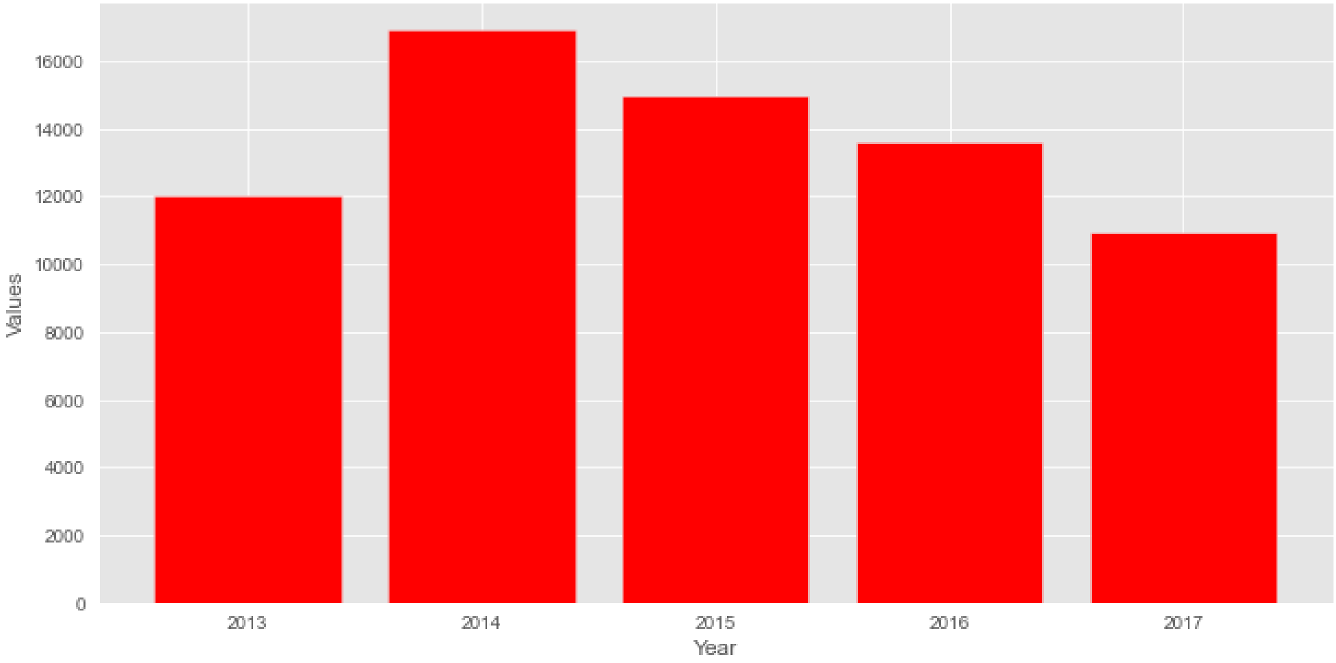
```
year = data.iyear.value_counts().head(5)
plt.figure(figsize=(12,6))
plt.bar(year.index,year.values,color='b')
plt.title("Years with most number of terrorist activity")
plt.xlabel("Year")
plt.ylabel("Values")
plt.show()
plt.style.use("ggplot")
```



In [39]:

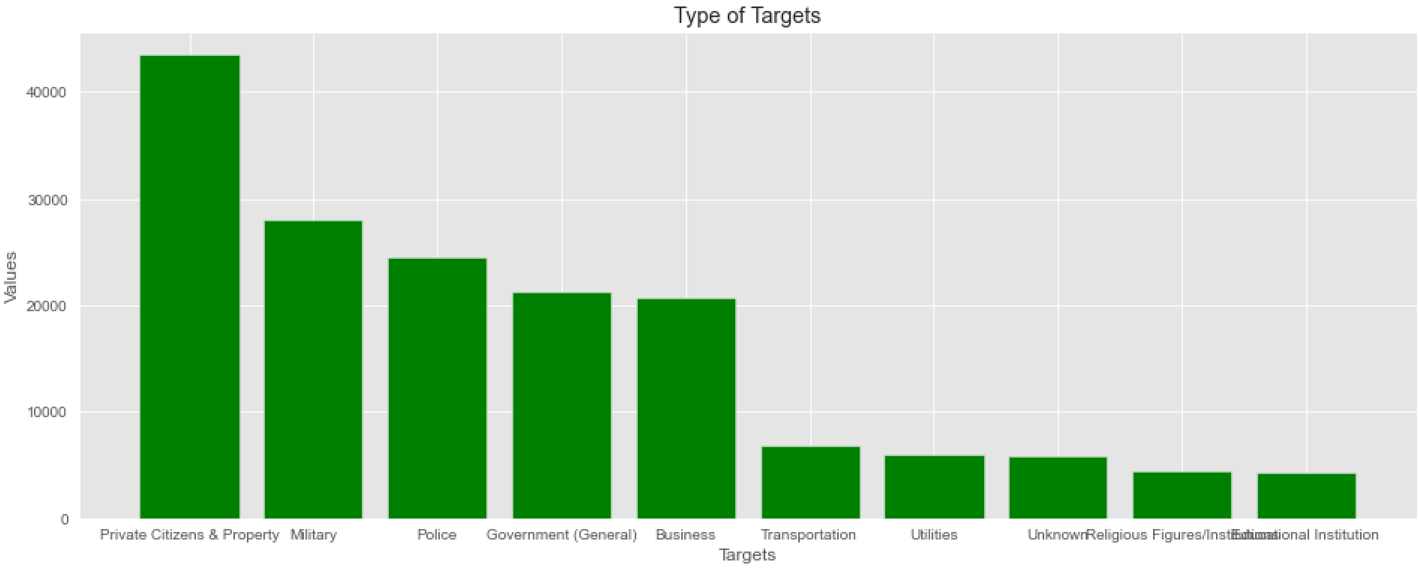
```
year = data.iyear.value_counts().head(5)
plt.figure(figsize=(12,6))
plt.bar(year.index,year.values,color='r')
plt.title("Years with most number of terrorist activity")
plt.xlabel("Year")
plt.ylabel("Values")
plt.show()
plt.style.use("ggplot")
```

Years with most number of terrorist activity



In [40]:

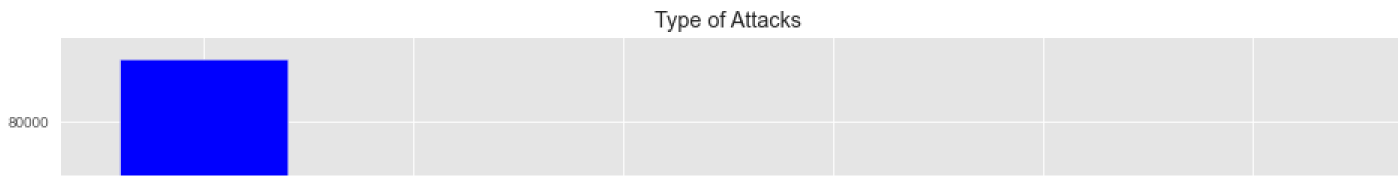
```
target = data['targtype1_txt'].value_counts().head(10)
fig,ax = plt.subplots(figsize=(16,6))
ax.bar(target.index,target.values,color='g')
plt.title('Type of Targets')
plt.xlabel("Targets")
plt.ylabel("Values")
plt.show()
```

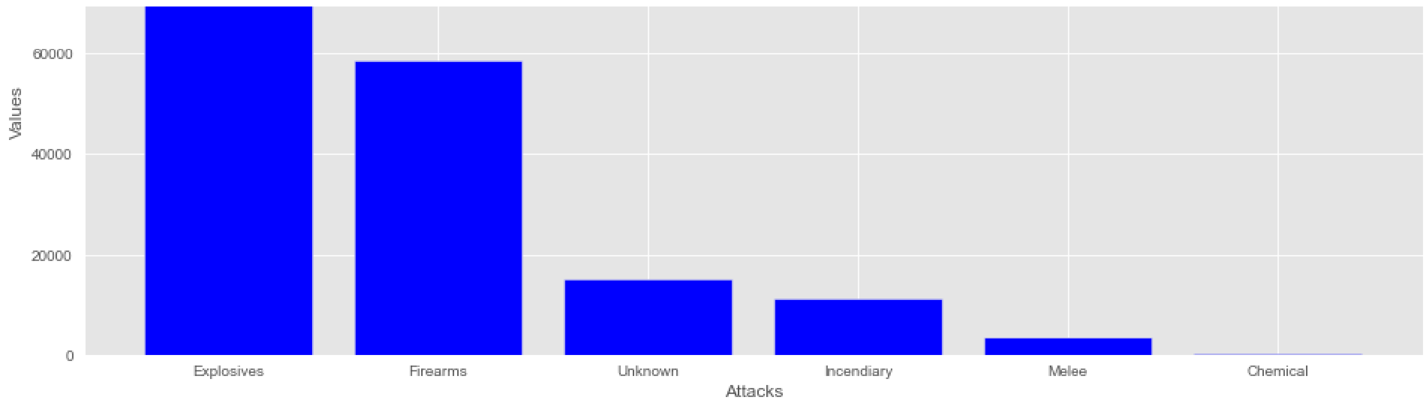


We could see from the above graphs that the most number of attacks took place in 2014 and private citizens were attacked the most

In [41]:

```
weapon_type = data['weaptype1_txt'].value_counts().head(6)
fig,ax = plt.subplots(figsize=(16,6))
ax.bar(weapon_type.index,weapon_type.values,color='b')
plt.title('Type of Attacks')
plt.xlabel("Attacks")
plt.ylabel("Values")
plt.show()
```



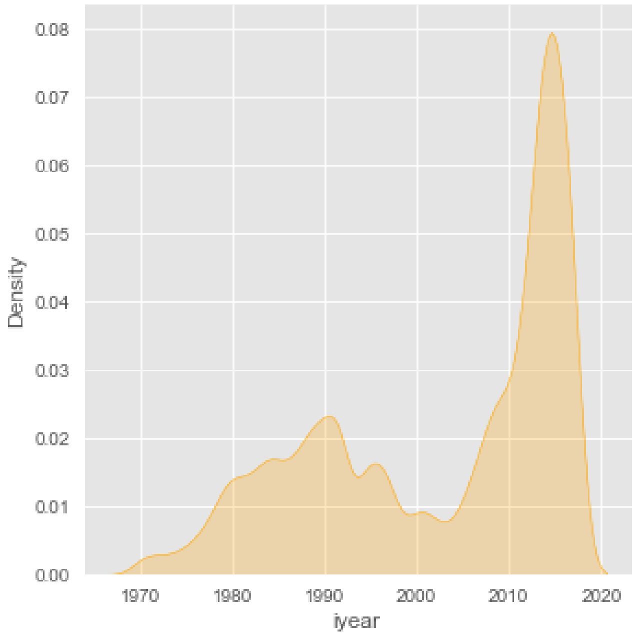


In [42]:

```
sns.displot(data,x='iyear',kind='kde',fill='tree',palette='colorblind',color='orange')
```

Out[42]:

```
<seaborn.axisgrid.FacetGrid at 0x222d8b245e0>
```

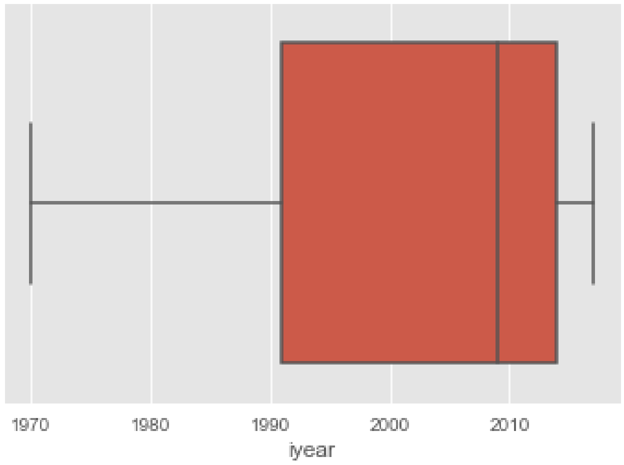


In [43]:

```
sns.boxplot(x=data['iyear'])
```

Out[43]:

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



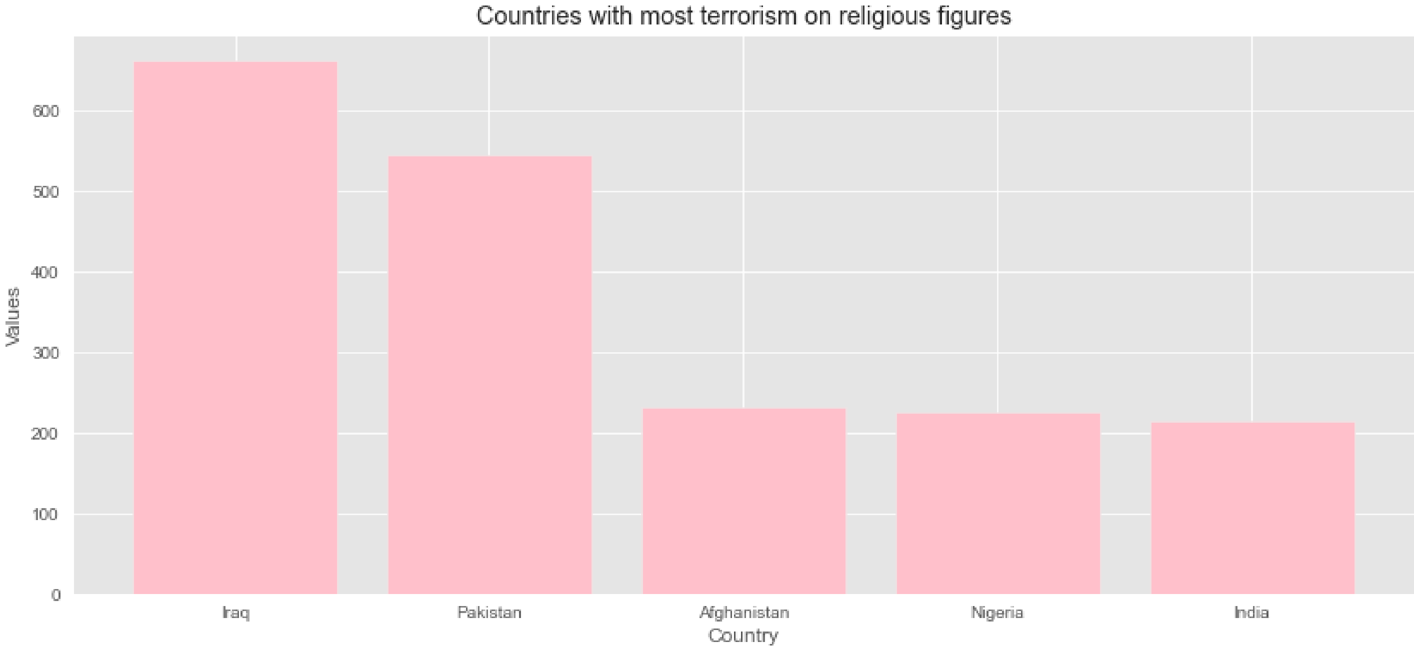
In [44]:

```
religious_target = data[data['targtype1_txt']=='Religious Figures/Institutions']
plt.figure(figsize=(14,6))
```

```
plt.bar(religious_target['country_txt'].value_counts().head().index,religious_target['country_txt'].value_counts().head().values,color='pink')
plt.title("Countries with most terrorism on religious figures")
plt.xlabel("Country")
plt.ylabel("Values")
plt.show
```

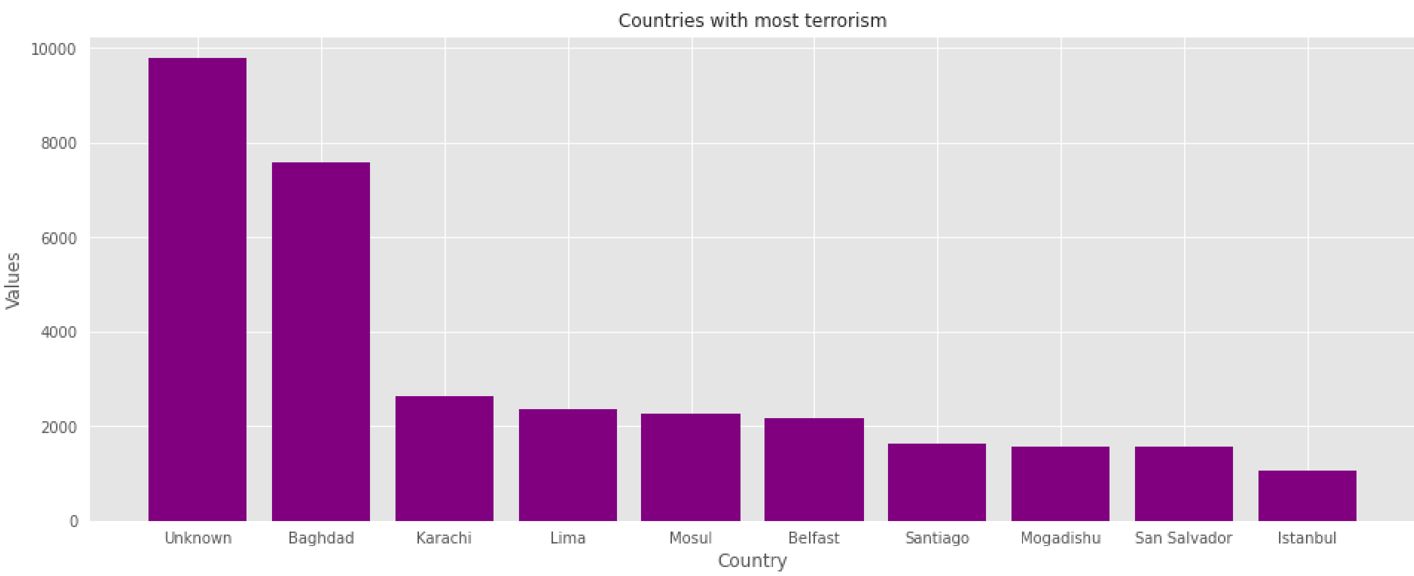
Out[44]:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [45]:

```
cities_with_most_terrorism = data.city.value_counts().head(10)
cities = list(cities_with_most_terrorism.index)
cities_with_most_terrorism
fig,ax = plt.subplots(figsize=(16,6))
plt.style.use('default')
ax.bar(cities_with_most_terrorism.index,cities_with_most_terrorism.values,color='purple')
plt.title('Countries with most terrorism')
plt.xlabel('Country')
plt.ylabel('Values')
plt.show()
```

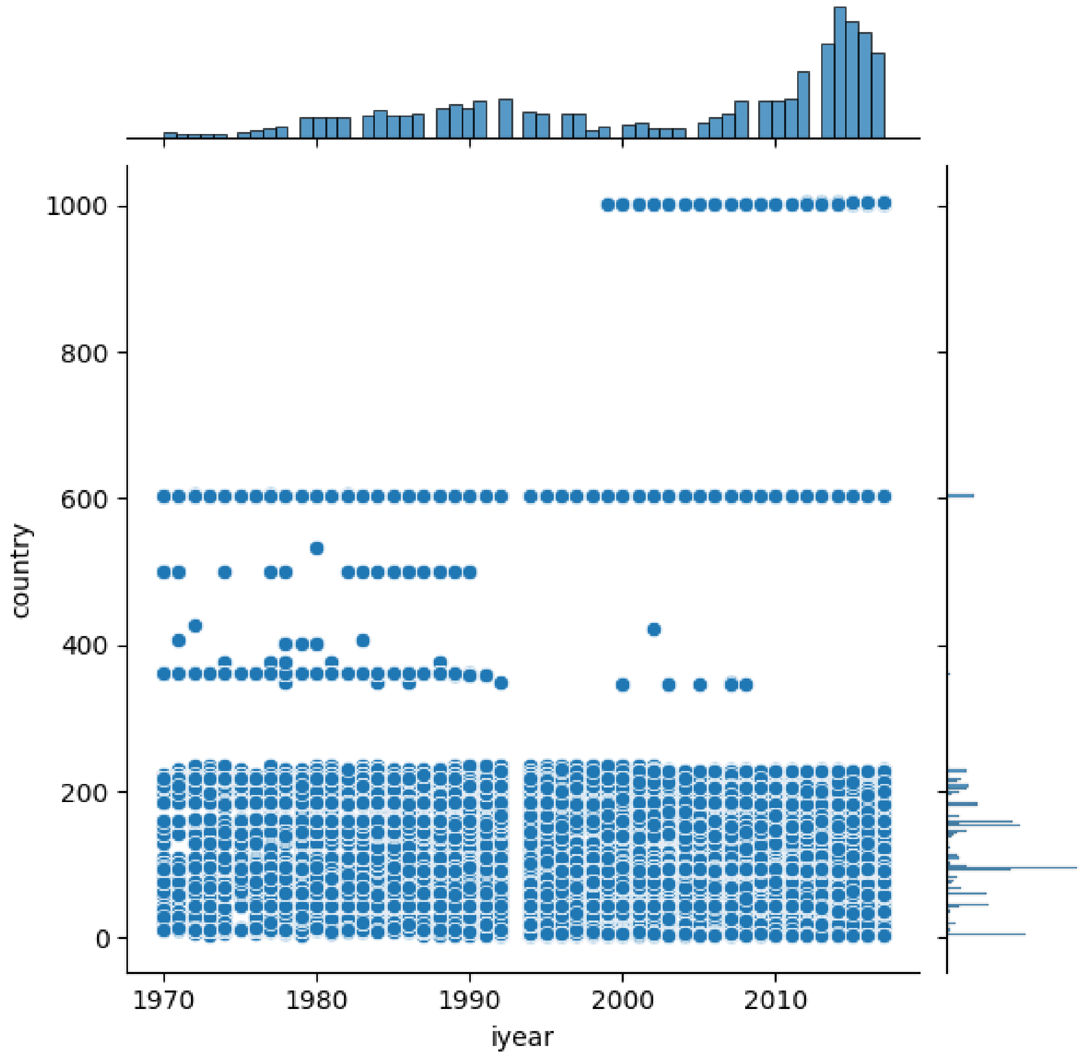


In [46]:

```
sns.jointplot(x=data['iyear'],y=data['country'])
plt.figure(figsize=(2,2))
```

Out[46]:

<Figure size 200x200 with 0 Axes>



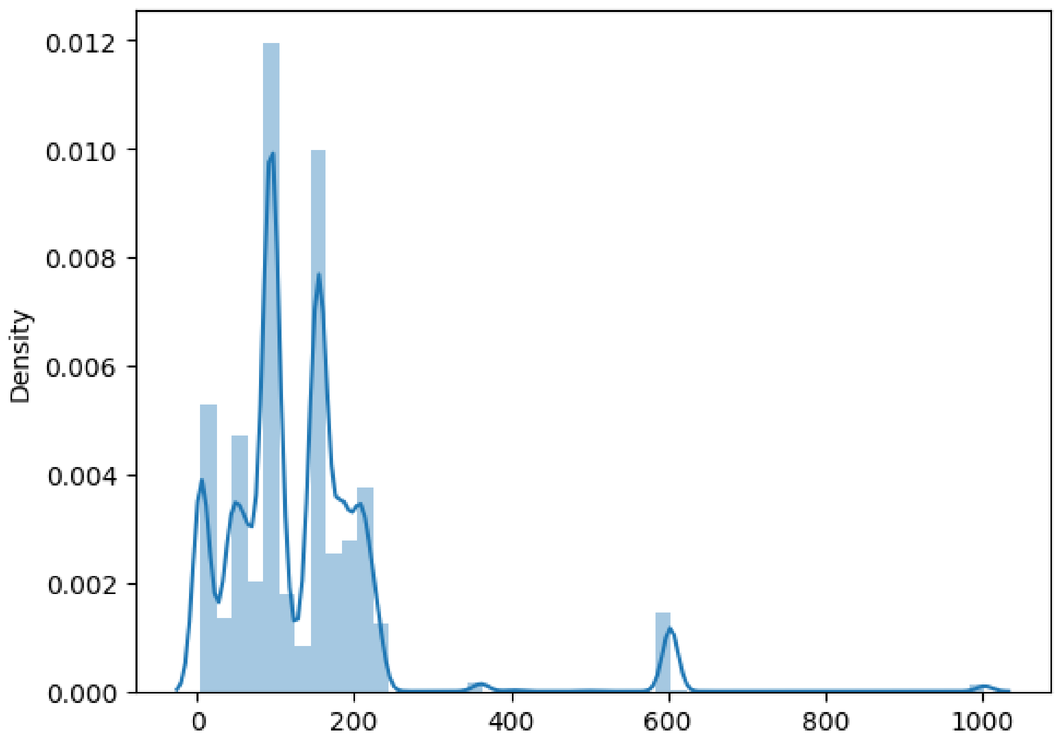
<Figure size 200x200 with 0 Axes>

In [47]:

```
sns.distplot(a=data['country'])
```

Out[47]:

<AxesSubplot:xlabel='country', ylabel='Density'>



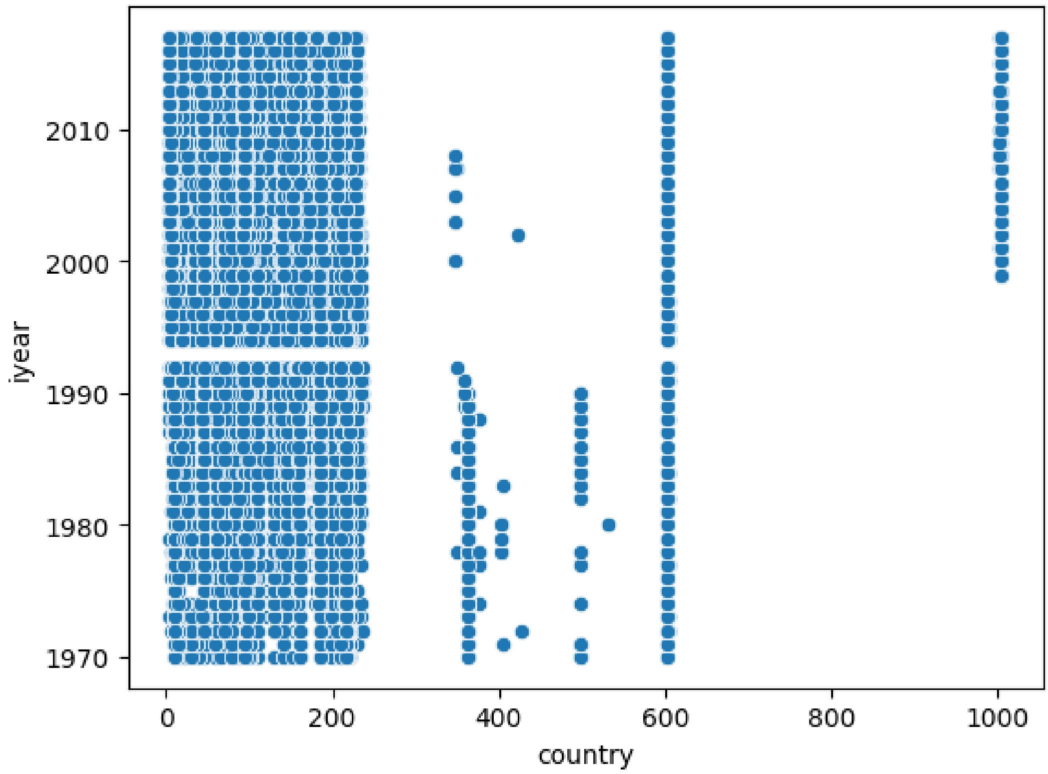
country

In [48]:

```
sns.scatterplot(x=data['country'],y=data['iyear'])
```

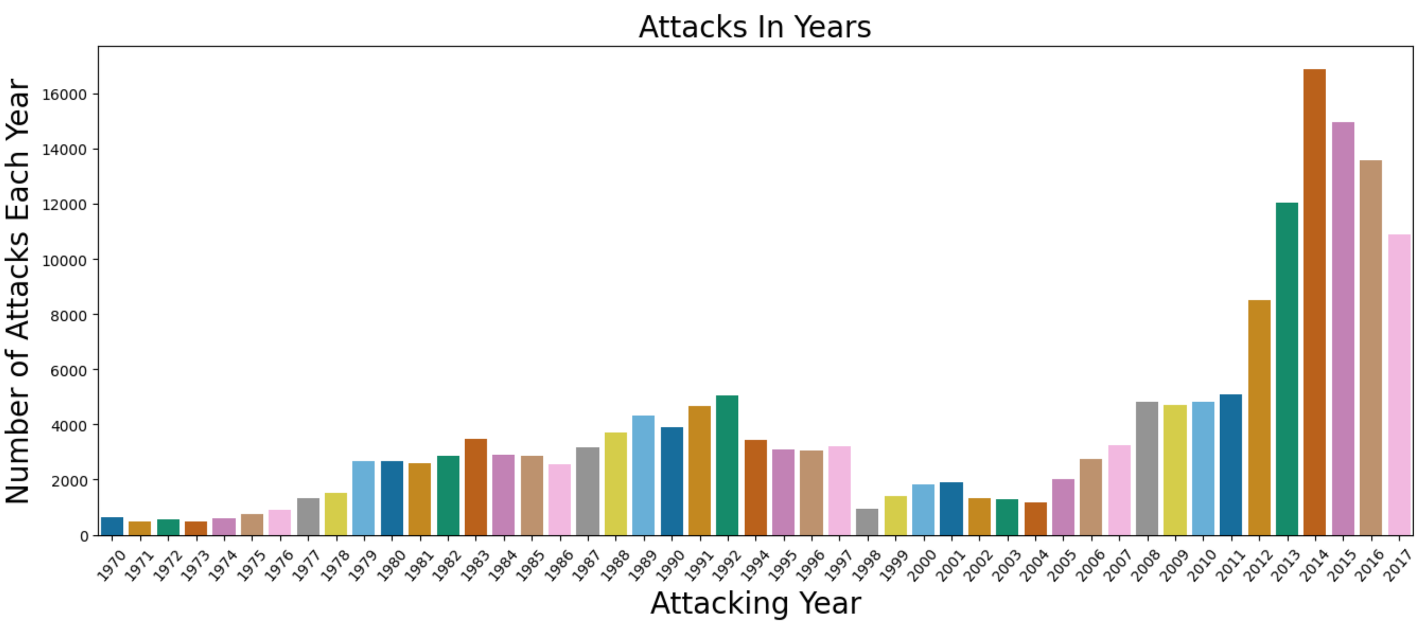
Out[48]:

```
<AxesSubplot:xlabel='country', ylabel='iyear'>
```



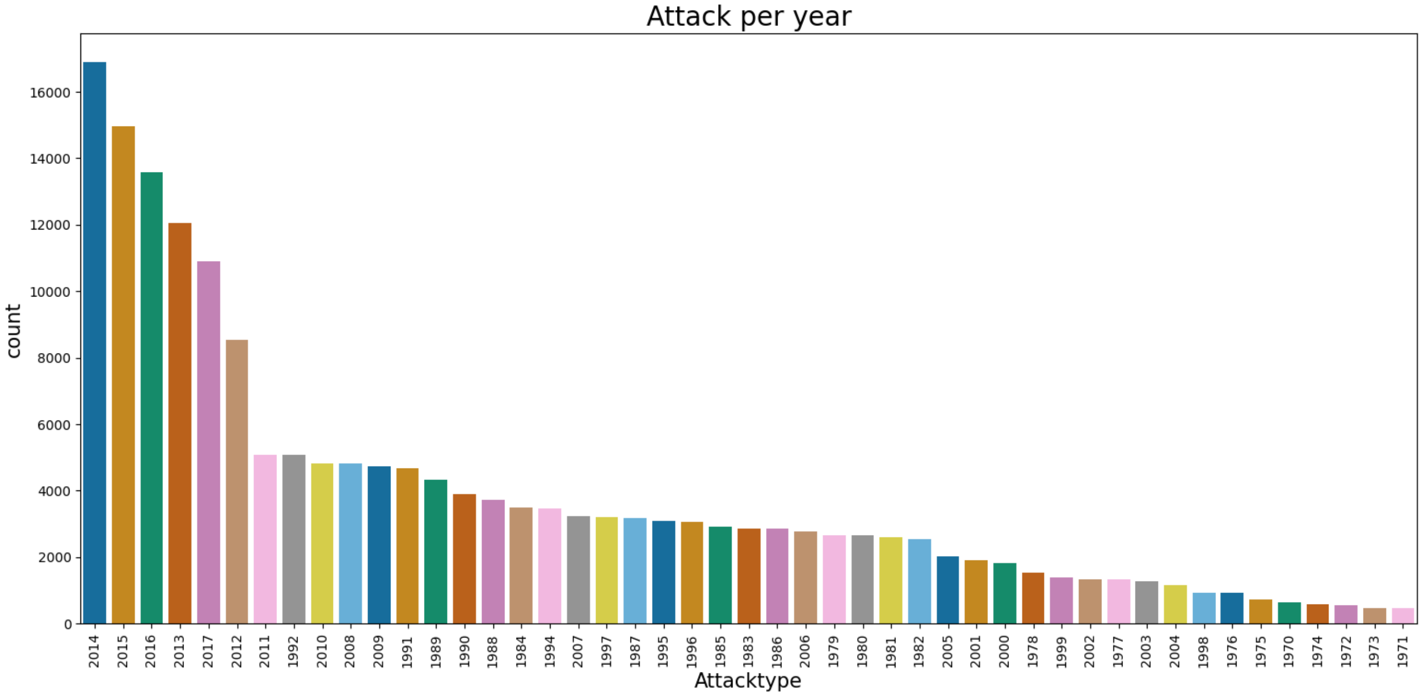
In [49]:

```
year = data['iyear'].unique()
years_count = data['iyear'].value_counts(dropna = False).sort_index()
plt.figure(figsize = (16,6))
sns.barplot(x = year,
            y = years_count,
            palette = "colorblind")
plt.xticks(rotation = 50)
plt.xlabel('Attacking Year',fontsize=20)
plt.ylabel('Number of Attacks Each Year',fontsize=20)
plt.title('Attacks In Years',fontsize=20)
plt.show()
```



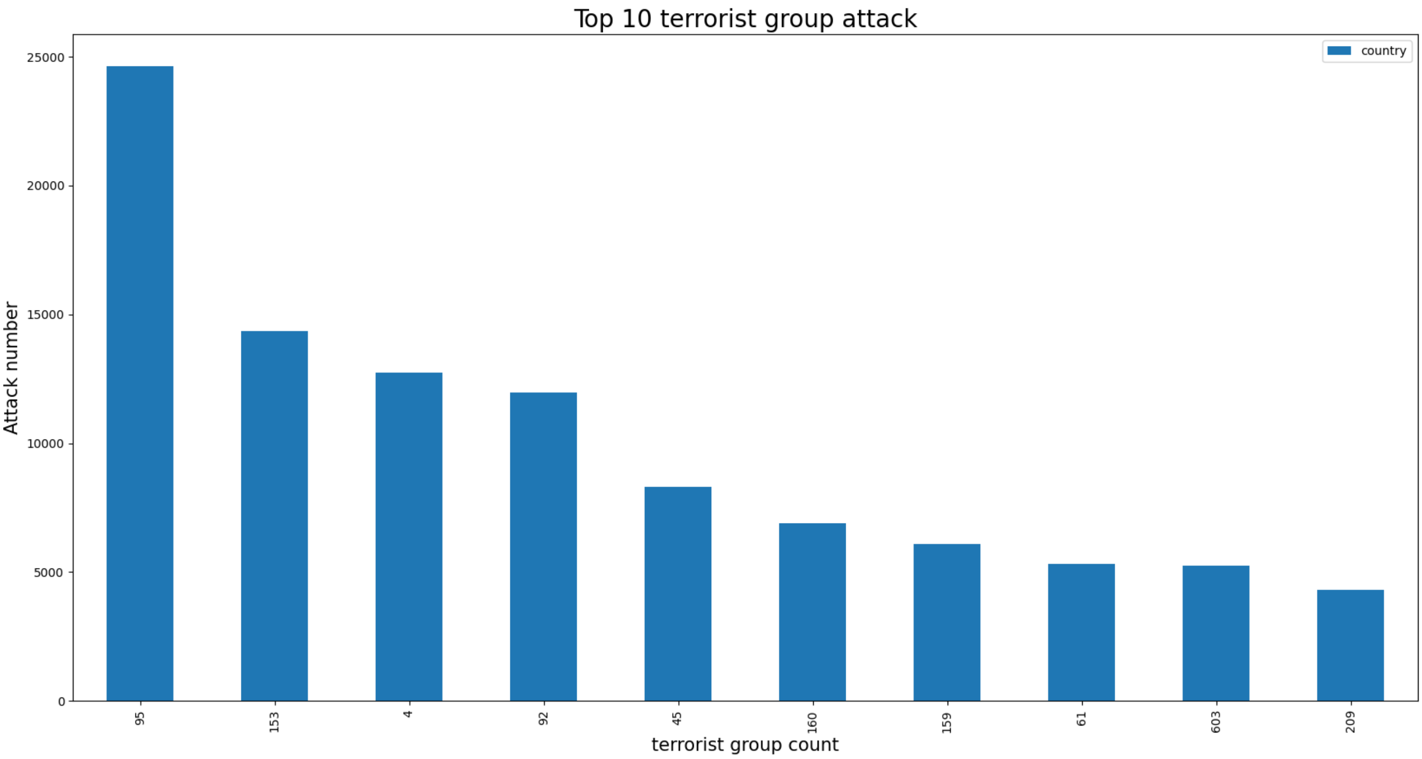
In [50]:

```
plt.subplots(figsize=(18,8))
sns.countplot(data["iyear"],order=data['iyear'].value_counts().index,palette="colorblind
")
plt.xticks(rotation=90)
plt.xlabel("Attacktype",fontsize=15)
plt.ylabel("count",fontsize=15)
plt.title("Attack per year",fontsize=20)
plt.show()
```



In [52]:

```
data['country'].value_counts().to_frame().head(10).plot(kind='bar',figsize=(20,10))
plt.title("Top 10 terrorist group attack",fontsize=20)
plt.xlabel("terrorist group count",fontsize=15)
plt.ylabel("Attack number",fontsize=15)
plt.show()
```



Conclusion

The death toll has been increasing every decade since 1970.The middle east is most prone to terror attacks

The death toll has been increasing every decade since 1970.The middle east is most prone to terror attacks followed by North Africa and South Asia.Most of the attacks were perpetrated by unknown terrorist groups followed by Taliban.Baghdhad is the most affected city.Civilians were the most targetted victims.

In [ ]:

```

```