Exploratory Data Analysis - Terrorism

Task:

- Perform "Exploratory Data Analysis" on dataset 'Global Terrorism'
- As a security/defense analyst, try to find out the hot zone of terrorism.
- Derive the security issues and insights.

In [1]:
```python
#importing libraries.
import pandas as pd
import plotly.express as px
import plotly.io as pio
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

In [2]:
```python
#importing the data
csv=pd.read_csv("globalterrorismdb_0718dist.csv",encoding='ISO-8859-1')
csv.head(3)
```

Out[2]:

	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
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS	0	0	0	0	NaN
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS	0	1	1	1	NaN
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS	-9	-9	1	1	NaN

3 rows × 135 columns

In [3]:
```python
#getting the column names
columns=csv.columns.to_list()
print(f"shape of the dataset:  {csv.shape[0]} rows x {csv.shape[1]} columns\n\n columns:\n{columns}")
```

```
shape of the dataset:  181691 rows x 135 columns

 columns:
['eventid', 'iyear', 'imonth', 'iday', 'approxdate', 'extended', 'resolution', 'country', 'country_txt', 'region', 'region_txt', 'provstate', 'city', 'latitude', 'longitude', 'spec
ificity', 'vicinity', 'location', 'summary', 'crit1', 'crit2', 'crit3', 'doubtterr', 'alternative', 'alternative_txt', 'multiple', 'success', 'suicide', 'attacktype1', 'attacktype1
_txt', 'attacktype2', 'attacktype2_txt', 'attacktype3', 'attacktype3_txt', 'targtype1', 'targtype1_txt', 'targsubtype1', 'targsubtype1_txt', 'corp1', 'target1', 'natlty1', 'natlty1
_txt', 'targtype2', 'targtype2_txt', 'targsubtype2', 'targsubtype2_txt', 'corp2', 'target2', 'natlty2', 'natlty2_txt', 'targtype3', 'targtype3_txt', 'targsubtype3', 'targsubtype3_t
xt', 'corp3', 'target3', 'natlty3', 'natlty3_txt', 'gname', 'gsubname', 'gname2', 'gsubname2', 'gname3', 'gsubname3', 'motive', 'guncertain1', 'guncertain2', 'guncertain3', 'indivi
dual', 'nperps', 'nperpcap', 'claimed', 'claimmode', 'claimmode_txt', 'claim2', 'claimmode2', 'claimmode2_txt', 'claim3', 'claimmode3', 'claimmode3_txt', 'compclaim', 'weaptype1',
'weaptype1_txt', 'weapsubtype1', 'weapsubtype1_txt', 'weaptype2', 'weaptype2_txt', 'weapsubtype2', 'weapsubtype2_txt', 'weaptype3', 'weaptype3_txt', 'weapsubtype3', 'weapsubtype3_t
xt', 'weaptype4', 'weaptype4_txt', 'weapsubtype4', 'weapsubtype4_txt', 'weapdetail', 'nkill', 'nkillus', 'nkillter', 'nwound', 'nwoundus', 'nwoundte', 'property', 'propextent', 'pr
opextent_txt', 'propvalue', 'propcomment', 'ishostkid', 'nhostkid', 'nhostkidus', 'nhours', 'ndays', 'divert', 'kidhijcountry', 'ransom', 'ransomamt', 'ransomamtus', 'ransompaid',
'ransompaidus', 'ransomnote', 'hostkidoutcome', 'hostkidoutcome_txt', 'nreleased', 'addnotes', 'scite1', 'scite2', 'scite3', 'dbsource', 'INT_LOG', 'INT_IDEO', 'INT_MISC', 'INT_AN
Y', 'related']
```

Extracting usefull columns.

In [4]:
```python
df=csv[['iyear', 'imonth', 'iday','country_txt','region_txt', 'provstate','latitude', 'longitude', 'city','location', 'summary','success','suicide',
        'targtype1_txt','gname','motive','weapdetail','nkill','nwound']]

#Renaming the columns for better understanding.
df.rename(columns={'iyear':'YEAR', 'imonth':'MONTH', 'iday':'DAY','country_txt':'COUNTRY','region_txt':'REGION', 'provstate':'PROVINCE/STATE','latitude':'LATITUDE','longitude':'LON
        'targtype1_txt':'TARGET','gname':'GROUP/ORG','motive':'MOTIVE','weapdetail':'WEAPON_TYPE','nkill':'KILLED','nwound':'WOUNDED'},inplace=True)
```

Checking for null(nan) values and filling them with 0

In [5]:
```python
print(df.isnull().sum())
#Have to replace the 'NaN' values with 0 to make our data compatible with Mapbox.
df[['LONGITUDE','LATITUDE','KILLED','WOUNDED','LOCATION','CITY']]=df[['LONGITUDE','LATITUDE','KILLED','WOUNDED','LOCATION','CITY']].fillna(0)
df.head(3)
```

```
YEAR                 0
MONTH                0
DAY                  0
COUNTRY              0
REGION               0
PROVINCE/STATE     421
LATITUDE          4556
LONGITUDE         4557
CITY               434
LOCATION        126196
SUMMARY          66129
SUCCESS              0
SUICIDE              0
TARGET               0
GROUP/ORG            0
MOTIVE          131130
WEAPON_TYPE      67670
KILLED           10313
WOUNDED          16311
dtype: int64
```

Out[5]:

	YEAR	MONTH	DAY	COUNTRY	REGION	PROVINCE/STATE	LATITUDE	LONGITUDE	CITY	LOCATION	SUMMARY	SUCCESS	SUICIDE	TARGET	GROUP/ORG	MOTIVE	WEAPON_TYPE	KILLED	WOUNDED
0	1970	7	2	Dominican Republic	Central America & Caribbean	NaN	18.456792	-69.951164	Santo Domingo	0	NaN	1	0	Private Citizens & Property	MANO-D	NaN	NaN	1.0	0.0
1	1970	0	0	Mexico	North America	Federal	19.371887	-99.086624	Mexico city	0	NaN	1	0	Government (Diplomatic)	23rd of September Communist League	NaN	NaN	0.0	0.0
2	1970	1	0	Philippines	Southeast Asia	Tarlac	15.478598	120.599741	Unknown	0	NaN	1	0	Journalists & Media	Unknown	NaN	NaN	1.0	0.0

In [6]:
```python
df.describe()
```

Out[6]:

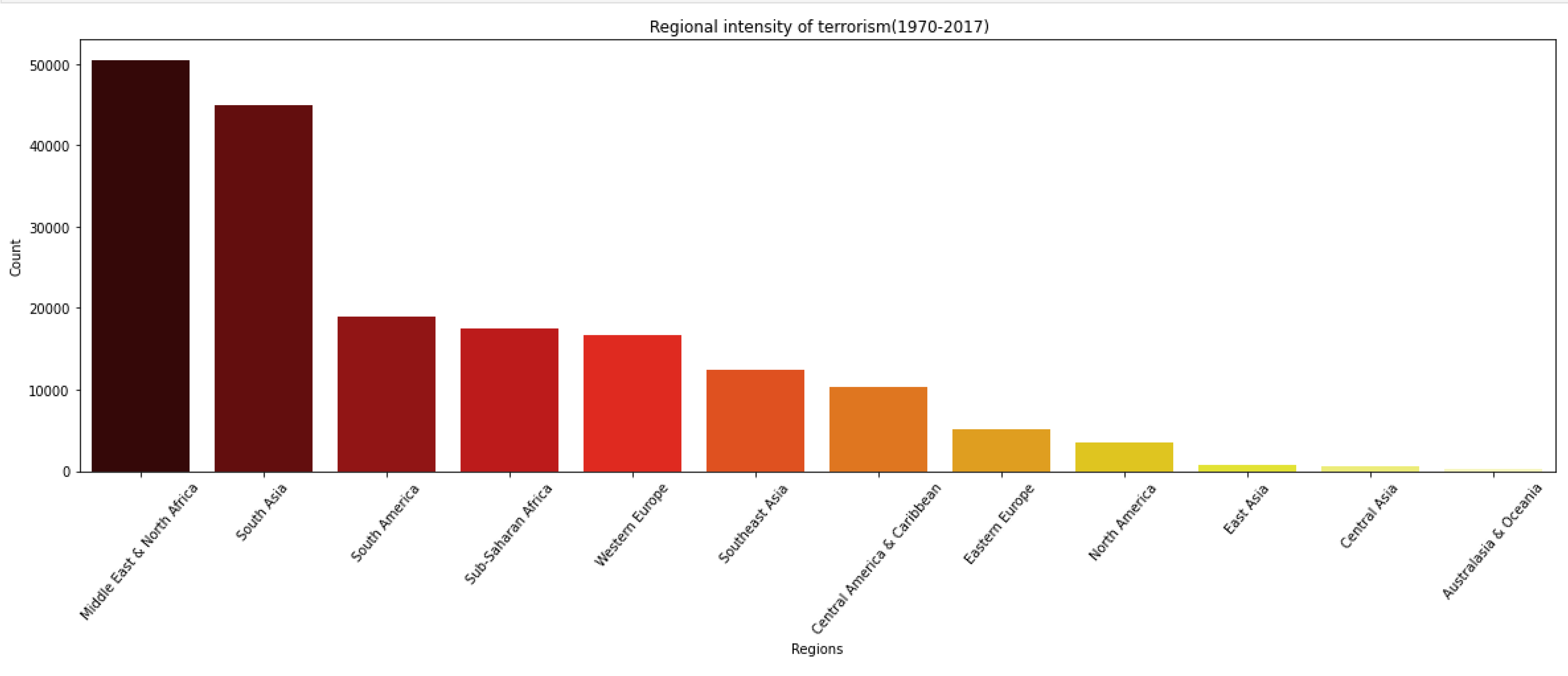
	YEAR	MONTH	DAY	LATITUDE	LONGITUDE	SUCCESS	SUICIDE	KILLED	WOUNDED
count	181691.000000	181691.000000	181691.000000	181691.000000	1.816910e+05	181691.000000	181691.000000	181691.000000	181691.000000
mean	2002.638997	6.467277	15.505644	22.909109	-4.471911e+02	0.889598	0.036507	2.266860	2.883296
std	13.259430	3.388303	8.814045	18.699442	2.021946e+05	0.313391	0.187549	11.227057	34.309747
min	1970.000000	0.000000	0.000000	-53.154613	-8.618590e+07	0.000000	0.000000	0.000000	0.000000
25%	1991.000000	4.000000	8.000000	9.518645	1.231572e+00	1.000000	0.000000	0.000000	0.000000
50%	2009.000000	6.000000	15.000000	31.126646	4.314357e+01	1.000000	0.000000	0.000000	0.000000
75%	2014.000000	9.000000	23.000000	34.538561	6.835734e+01	1.000000	0.000000	2.000000	2.000000
max	2017.000000	12.000000	31.000000	74.633553	1.793667e+02	1.000000	1.000000	1570.000000	8191.000000

In [7]:
```python
MxRegion=df['REGION'].value_counts().idxmax()
MxTarget=df['TARGET'].value_counts().idxmax()
MxCountry=df['COUNTRY'].value_counts().idxmax()
MxCity=df['CITY'].value_counts().index[1]
MxGroup=df['GROUP/ORG'].value_counts().index[1]
MxYear=df['YEAR'].value_counts().idxmax()

MnRegion=df['REGION'].value_counts().idxmin()
MnCountry=df['COUNTRY'].value_counts().idxmin()
MnGroup=df['GROUP/ORG'].value_counts().index[-1]
MnYear=df['YEAR'].value_counts().idxmin()

df['VICTIMS']=df['KILLED']+df['WOUNDED']
Total_Victims=df['VICTIMS'].sum()
```

In [8]:
```python
print("Observation:")
print(f"\nFrom 1970 to 2017, there were total 181691 attacks.\nThis attacks resulted in {int(Total_Victims)} casualties.\n{MxTarget} were primary target.\nThe success rate of those
print(f"\nMost Attacked Region: {MxRegion}\nMost Attacked Country: {MxCountry}\nMost Attacked City: {MxCity}\nPrimary Target: {MxTarget}\nMost Attacks Carried Out By(Group/Organisa
print(f"\nLeast Attacked Region: {MnRegion}\nLeast Attacked Country: {MnCountry}\nLeast Attacks Carried Out By(Group/Organisation): {MnGroup}\nLeast Attacked Year: {MnYear}\n")
```

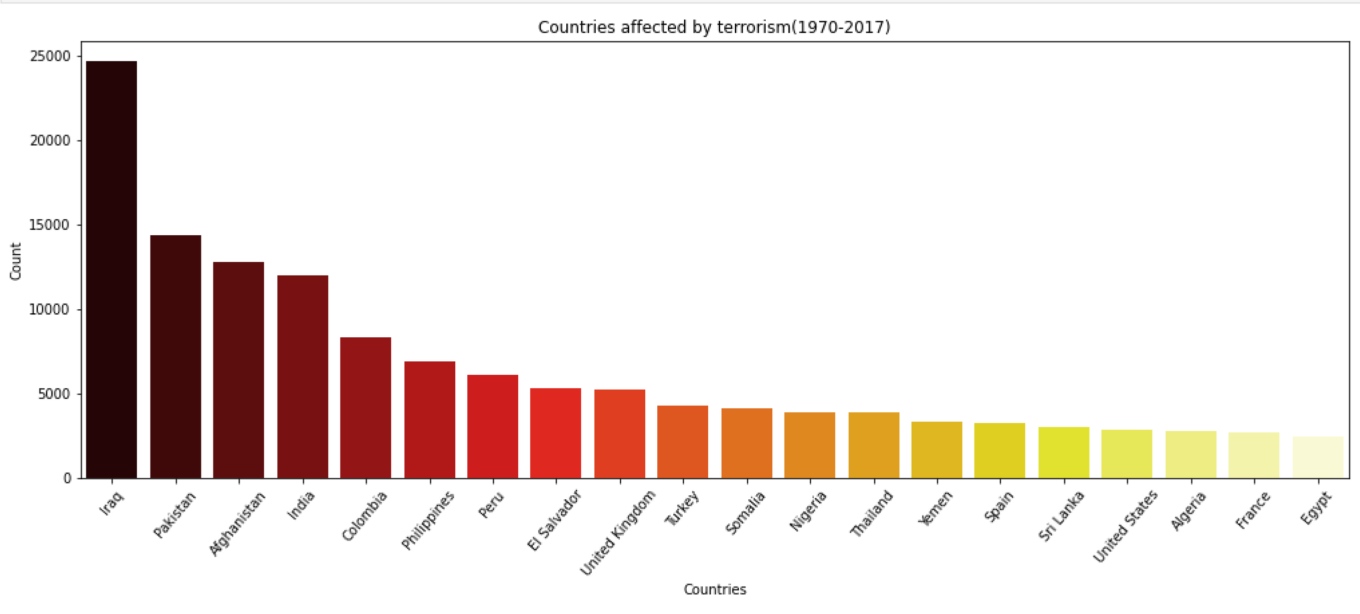
```
Observation:

From 1970 to 2017, there were total 181691 attacks.
This attacks resulted in 935737 casualties.
Private Citizens & Property were primary target.
The success rate of those attacks was approx 89 %.The highest death toll caused by an attack was 1570.

Most Attacked Region: Middle East & North Africa
Most Attacked Country: Iraq
Most Attacked City: Baghdad
Primary Target: Private Citizens & Property
Most Attacks Carried Out By(Group/Organisation): Taliban
Most Attacked Year: 2014


Least Attacked Region: Australasia & Oceania
Least Attacked Country: Vatican City
Least Attacks Carried Out By(Group/Organisation): MANO-D
Least Attacked Year: 1971
```

In [9]:
```python
plt.subplots(figsize=(20,6))
x=df['REGION'].value_counts()[:20].index
y=df['REGION'].value_counts()[:20].values


sns.barplot(x,y,palette='hot')
plt.title(' Regional intensity of terrorism(1970-2017)')
plt.xlabel('Regions')
plt.ylabel('Count')
plt.xticks(rotation= 50)
plt.show()
```



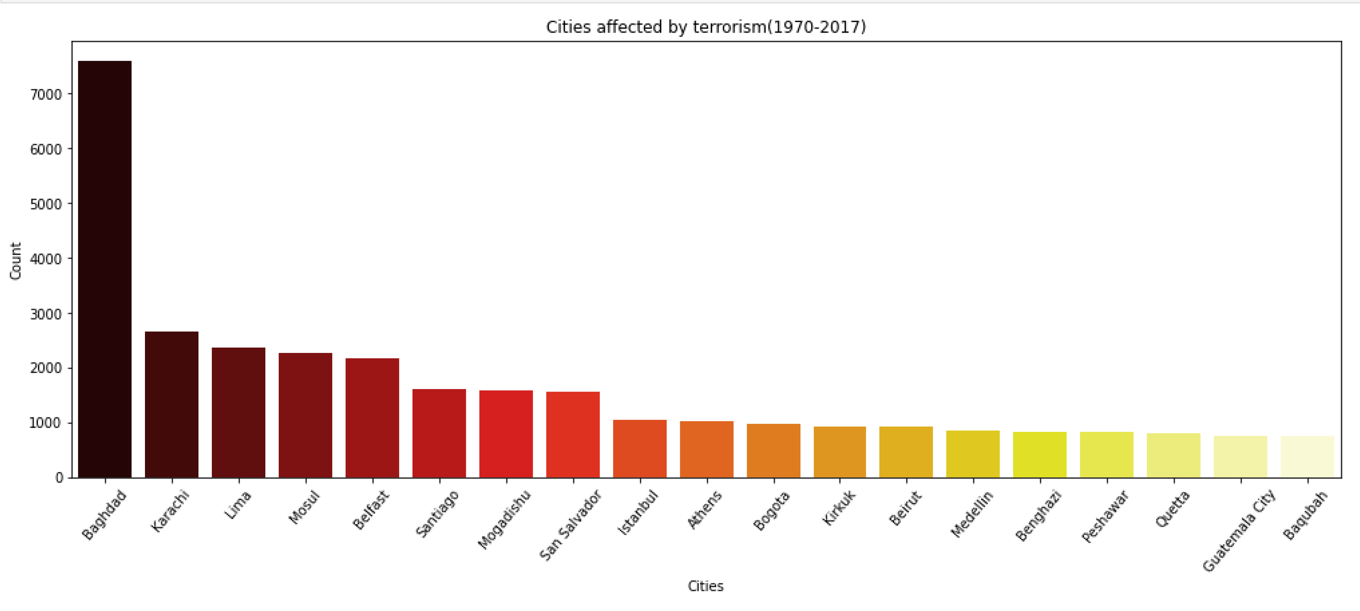
- Middle East & North Africa are the regions with high terrorist activities.

In [10]:
```python
plt.subplots(figsize=(17,6))
x=df['COUNTRY'].value_counts()[:20].index
y=df['COUNTRY'].value_counts()[:20].values


sns.barplot(x,y,palette='hot')
plt.title('Countries affected by terrorism(1970-2017)')

plt.xlabel('Countries')
plt.xticks(rotation=50)
plt.ylabel('Count')
plt.show()
```



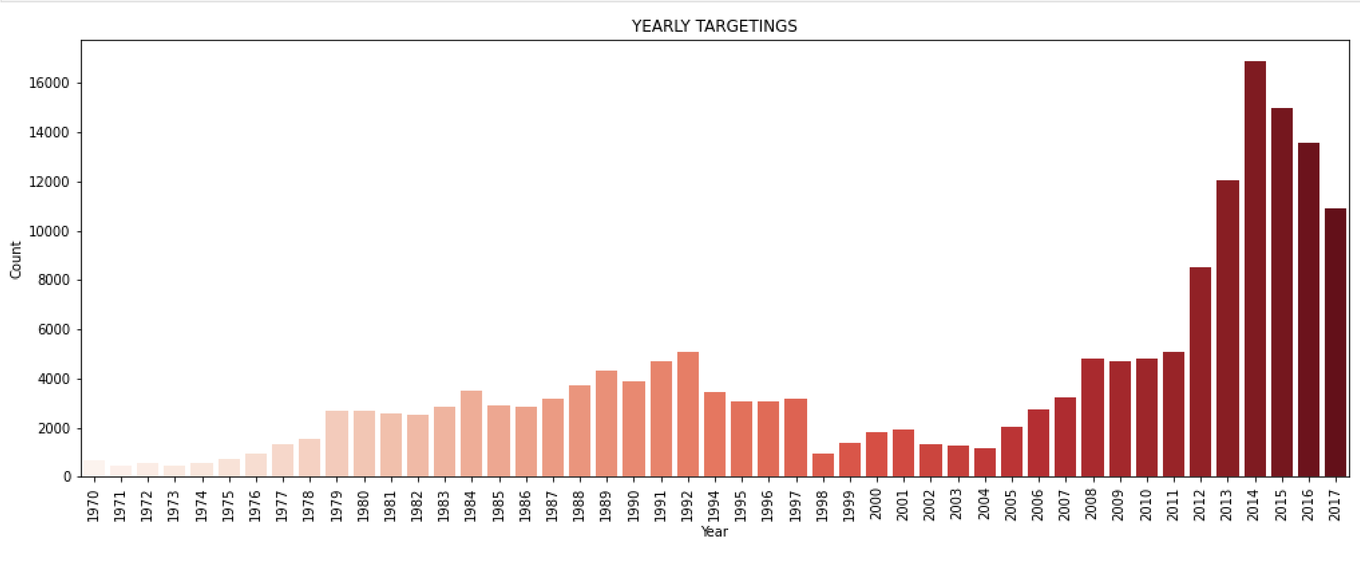
- Iraq is most affected by terrorism.

In [11]:
```python
plt.subplots(figsize=(17,6))
x=df['CITY'].value_counts()[1:20].index
y=df['CITY'].value_counts()[1:20].values


sns.barplot(x,y,palette='hot')
plt.title('Cities affected by terrorism(1970-2017)')

plt.xlabel('Cities')
plt.xticks(rotation=50)
plt.ylabel('Count')
plt.show()
```



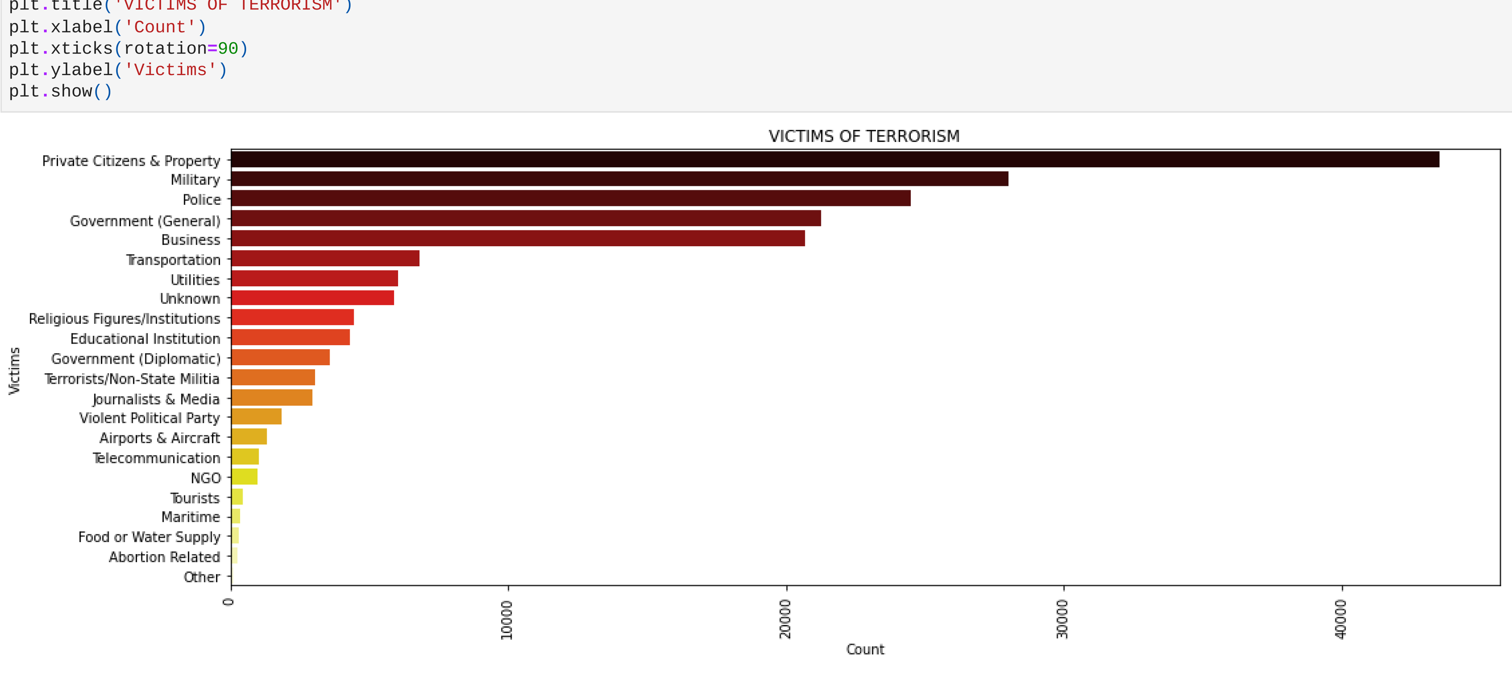
Baghdad & **Karachi** are the most unsafe cities in terms of Terrorism.

In [12]:
```python
plt.subplots(figsize=(17,6))
x=df['YEAR'].value_counts().index
y=df['YEAR'].value_counts().values

sns.barplot(x,y,palette='Reds')
plt.title('YEARLY TARGETINGS')
plt.xlabel('Year')
plt.xticks(rotation=90)
plt.ylabel('Count')
plt.show()
```



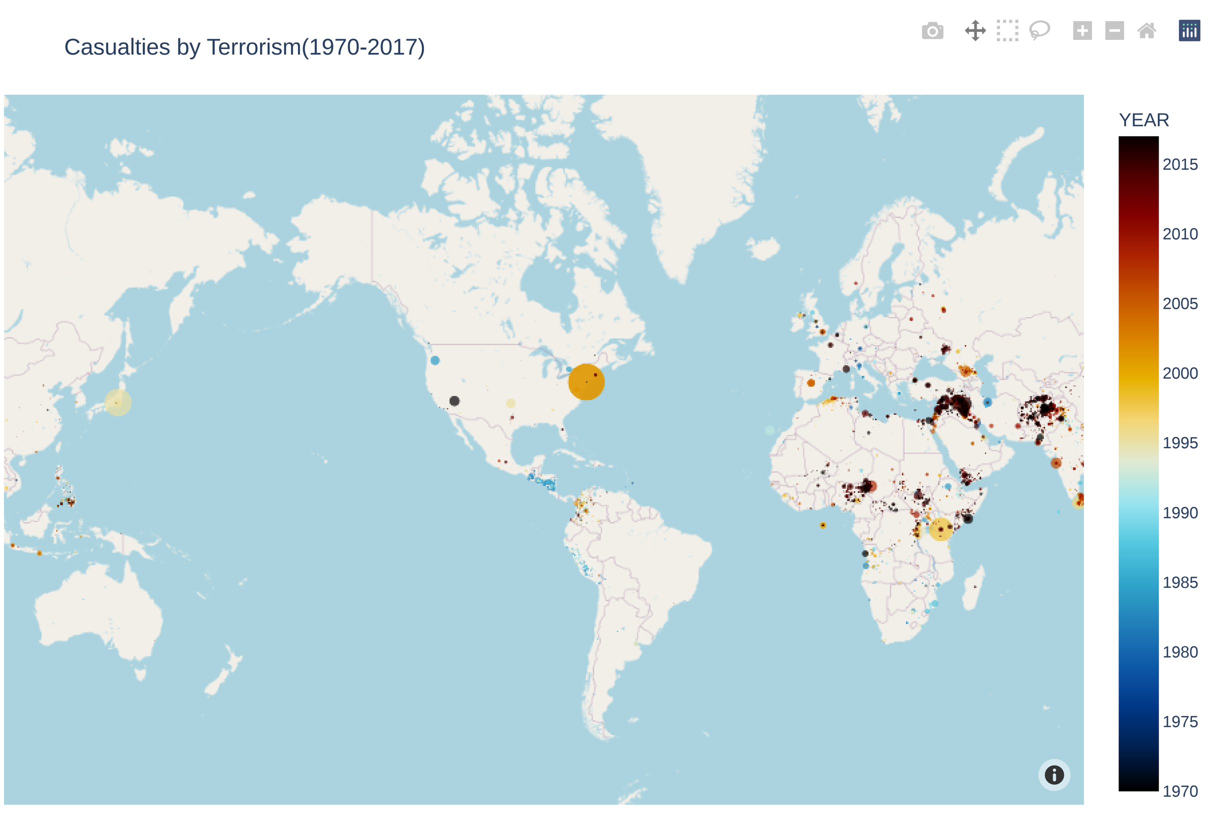
- The terrorist activities have **increased in recent years** and were **highest in 2014**.

In [13]:
```python
plt.subplots(figsize=(17,6))
x=df['TARGET'].value_counts().values
y=df['TARGET'].value_counts().index

sns.barplot(x,y,palette='hot')
plt.title('VICTIMS OF TERRORISM')
plt.xlabel('Count')
plt.xticks(rotation=90)
plt.ylabel('Victims')
plt.show()
```



- Private Citizens, Military ,Police , Government(General), Businessmen are mostly targeted by terrorist organizations.
- Person with those professions should avoid visiting places with high terrorist activities.

In [14]:
```python
fig= px.scatter_mapbox(df, lon=df['LONGITUDE'], lat=df['LATITUDE'], zoom=0.75, color= df['YEAR'], size=df['VICTIMS'], width= 900, height=600, title='Casualties by Terrorism(1970-20
fig.update_layout(mapbox_style="open-street-map")
fig.update_layout(margin={"r":0,"t":60,"l":0,"b":10})
fig.show()
print('The Map above shows the places which are mostly affected by Terrorism. \nPlaces with red marks are high-risk areas for terrorist activity. \nCountries that should be avoided
```




```
The Map above shows the places which are mostly affected by Terrorism.
Places with red marks are high-risk areas for terrorist activity.
Countries that should be avoided include Iraq, Syria, Nigeria, Afghanistan, and Pakistan.
```

... ...

Insights and Takeaways:

- It is advisable to avoid visiting the countries in **Middle Eastern & North African region** specifically **Iraq and Pakistan**, as these of the countries are **more prone to terrorism.**
- If there is an emergency requirement to visit the Iraq or pakistan one should avoid visiting **Baghdad and Karachi** at all costs.
- **Taliban** is most dangerous extrimist organization.
- **Australasia & Oceania** are the **safest region** from the terrorism.
- **Vatican City** is **safest country.**
- **Private Citizens, Military ,Police , Government(General), Businessmen** are **mostly targeted** by terrorist organizations. Person with these professions should avoid visiting places with high terrorist activities.