Report on Global terrorism (1970-2017)

Subject: Practical aspects of data preparation
Master’s Degree in Management, Data Science Specialization
Student: Rahul Rahul
Teacher: Dr. Karol Jędrasiak
WSB University, January 2020

Introduction

This dataset is based on the number of terrorist attacks occurred around the world between the period 1970 – 2017 except 1993. It contains information of more than 180,000 terrorist attack all around the world. The main motive of working on this database is to analyze the terrorism incidents and find some interesting results and conclusions.

The first task is to select the relevant variables from the dataset as there are 135 columns in the dataset but I will work on only 18 variables. The next task is to clean this data as the data is highly unstructured.

Then I will make some visualization on data around the world and also specifically on a particular country (Iraq).

Finally, I will write the conclusion that I draw from the visualization.

Structure of dataset

The original dataset was in Excel format(xlsx). I converted this data into comma separated value(csv) format as it's easy to work if you are working on Anaconda jupyter interface.

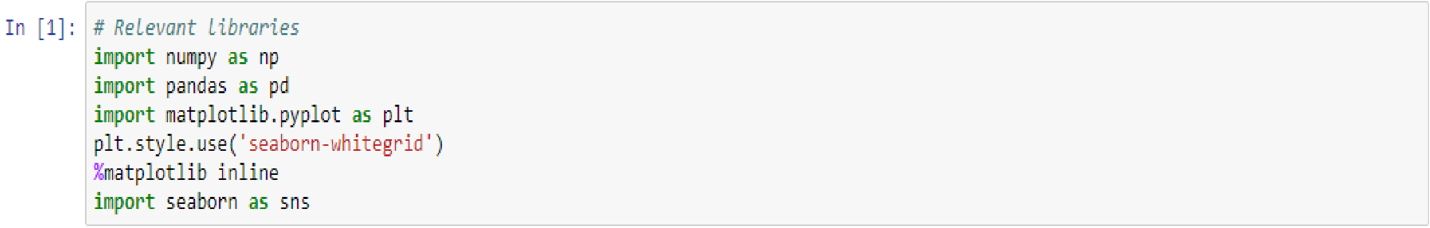
This dataset contains around 135 columns but I worked on only 18 columns/variables which are relevant to my project.

The columns are:

1. Event ID
2. Year
3. Month
4. Country
5. Region
6. State
7. City
8. latitude
9. longitude
10. success
11. Attack type
12. Target
13. Nationality
14. Terrorist Group
15. Weapon type
16. Weapon subtype
17. Killed
18. Wounded

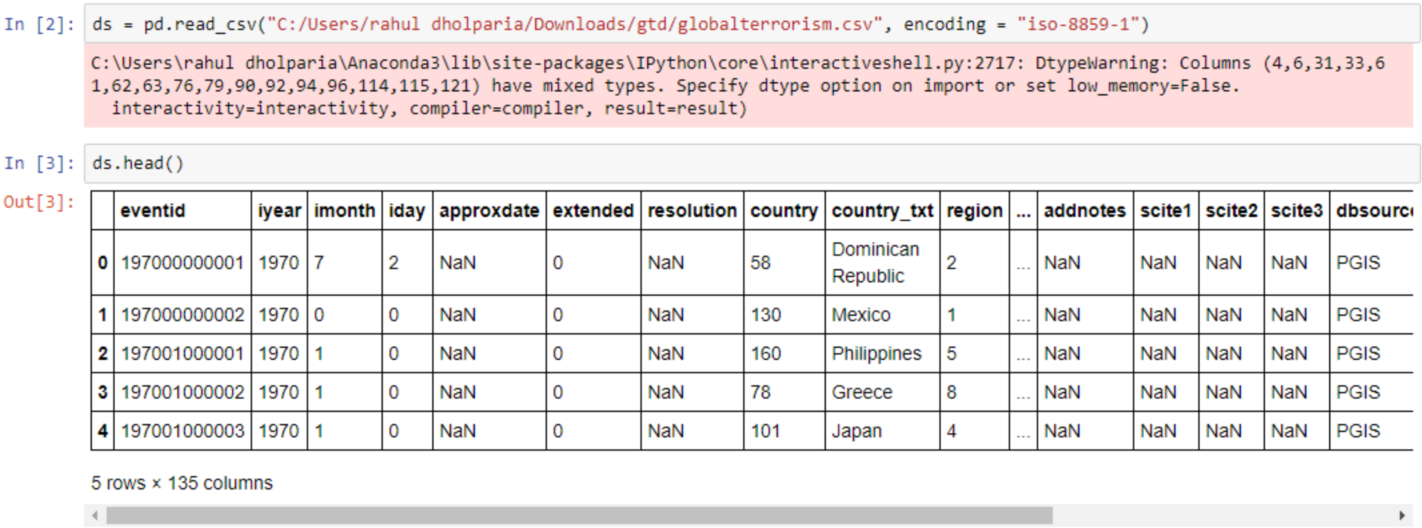
Index

1. Importing Libraries



```
In [1]: # Relevant libraries
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        plt.style.use('seaborn-whitegrid')
        %matplotlib inline
        import seaborn as sns
```

2. Importing the dataset



```
In [2]: ds = pd.read_csv("C:/Users/rahul dholparia/Downloads/gtd/globalterrorism.csv", encoding = "iso-8859-1")

        C:\Users\rahul dholparia\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:2717: DtypeWarning: Columns (4,6,31,33,6
        1,62,63,76,79,90,92,94,96,114,115,121) have mixed types. Specify dtype option on import or set low_memory=False.
          interactivity=interactivity, compiler=compiler, result=result)
```

```
In [3]: ds.head()
```

Out[3]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1	scite2	scite3	dbsourc
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN	NaN	NaN	PGIS
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN	NaN	NaN	PGIS
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN	NaN	NaN	PGIS
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN	NaN	NaN	PGIS
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN	NaN	NaN	PGIS

5 rows × 135 columns

As you can see, we have lots of cell with Null value and not all variables are important for our project. So, first we remove the irrelevant variables and then we clean our dataset.

3. Data cleaning

```
In [4]: # The total columns(variables) we have in dataset
        ds.columns.values
```

```
Out[4]: array(['eventid', 'iyear', 'imonth', 'iday', 'approxdate', 'extended',
               'resolution', 'country', 'country_txt', 'region', 'region_txt',
               'provstate', 'city', 'latitude', 'longitude', 'specificity',
               '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_txt', 'corp3', 'target3', 'natlty3',
               'natlty3_txt', 'gname', 'gsubname', 'gname2', 'gsubname2', 'gname3',
               'gsubname3', 'motive', 'guncertain1', 'guncertain2', 'guncertain3',
               'individual', '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',
```

```
In [5]: Motive = ds["motive"]
        # Selecting the relevant variables for our project.
        ds = ds[['eventid','iyear', 'imonth', 'country_txt', 'region_txt', 'provstate', 'city', 'latitude', 'longitude', 'success', 'atta
```

```
In [6]: ds.columns.values
```

```
Out[6]: array(['eventid', 'iyear', 'imonth', 'country_txt', 'region_txt',
               'provstate', 'city', 'latitude', 'longitude', 'success',
               'attacktype1_txt', 'targtype1_txt', 'natlty1_txt', 'gname',
               'weaptype1_txt', 'weapsubtype1_txt', 'nkill', 'nwound'], dtype=object)
```

```
In [7]: # Renaming the columns
        ds.rename(columns={'imonth': 'Month',
         'iyear': 'Year',
         'eventid': 'Event ID',
         'country_txt': 'Country',
         'region_txt': 'Region',
         'provstate': 'State',
         'city': 'City',
         'attacktype1_txt': 'Attack type',
         'targtype1_txt': 'Target',
         'natlty1_txt': 'Nationality',
         'gname': 'Terrorist Group',
         'weaptype1_txt': 'Weapon type',
         'weapsubtype1_txt': 'Weapon subtype',
         'nkill': 'Killed',
         'nwound': 'Wounded'},inplace=True)
```

```
In [8]: ds.columns.values
```

```
Out[8]: array(['Event ID', 'Year', 'Month', 'Country', 'Region', 'State', 'City',
               'latitude', 'longitude', 'success', 'Attack type', 'Target',
               'Nationality', 'Terrorist Group', 'Weapon type', 'Weapon subtype',
               'Killed', 'Wounded'], dtype=object)
```

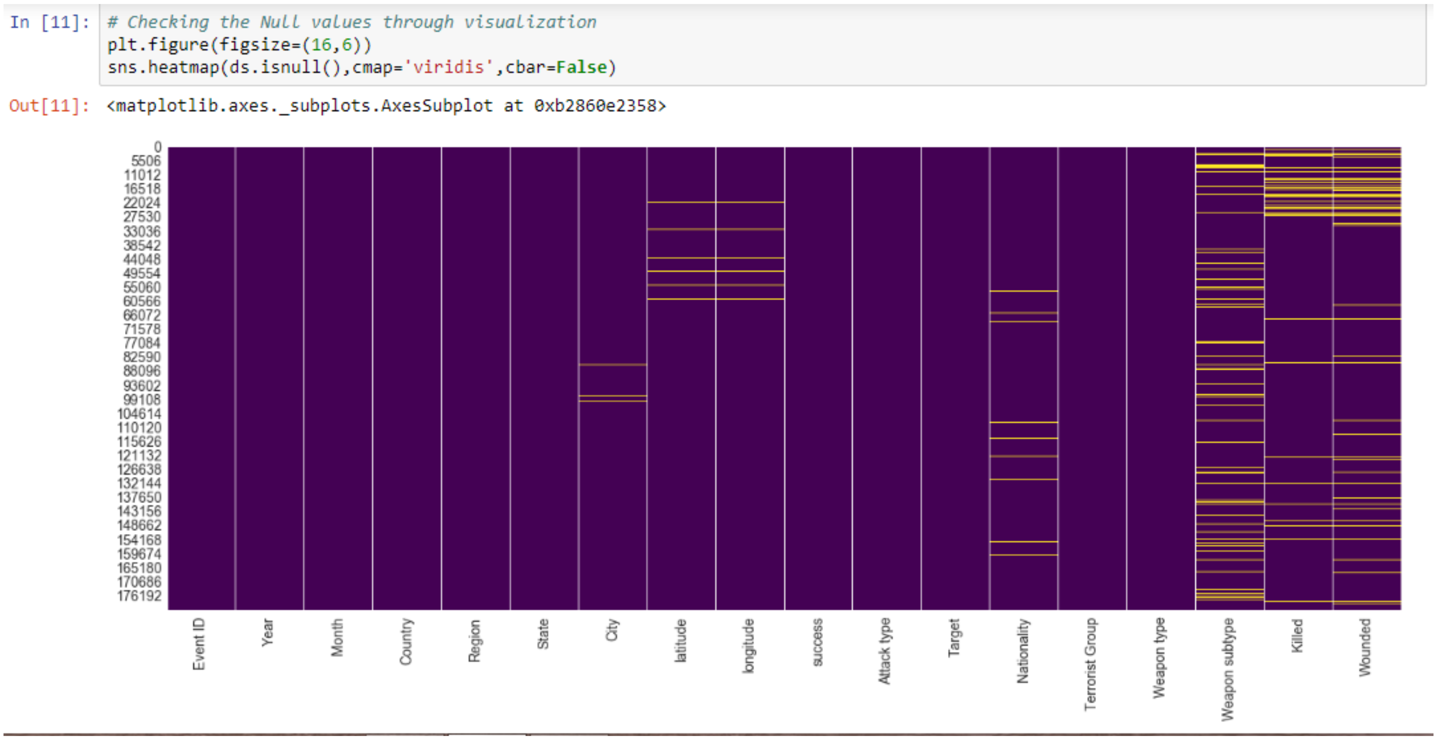
In [9]:
```
ds.head()
```

Out[9]:

	Event ID	Year	Month	Country	Region	State	City	latitude	longitude	success	Attack type	Target	Nation
0	197000000001	1970	7	Dominican Republic	Central America & Caribbean	NaN	Santo Domingo	18.456792	-69.951164	1	Assassination	Private Citizens & Property	Domin Repub
1	197000000002	1970	0	Mexico	North America	Federal	Mexico city	19.371887	-99.086624	1	Hostage Taking (Kidnapping)	Government (Diplomatic)	Belgiu
2	197001000001	1970	1	Philippines	Southeast Asia	Tarlac	Unknown	15.478598	120.599741	1	Assassination	Journalists & Media	United States
3	197001000002	1970	1	Greece	Western Europe	Attica	Athens	37.997490	23.762728	1	Bombing/Explosion	Government (Diplomatic)	United States
4	197001000003	1970	1	Japan	East Asia	Fukouka	Fukouka	33.580412	130.396361	1	Facility/Infrastructure Attack	Government (Diplomatic)	United States

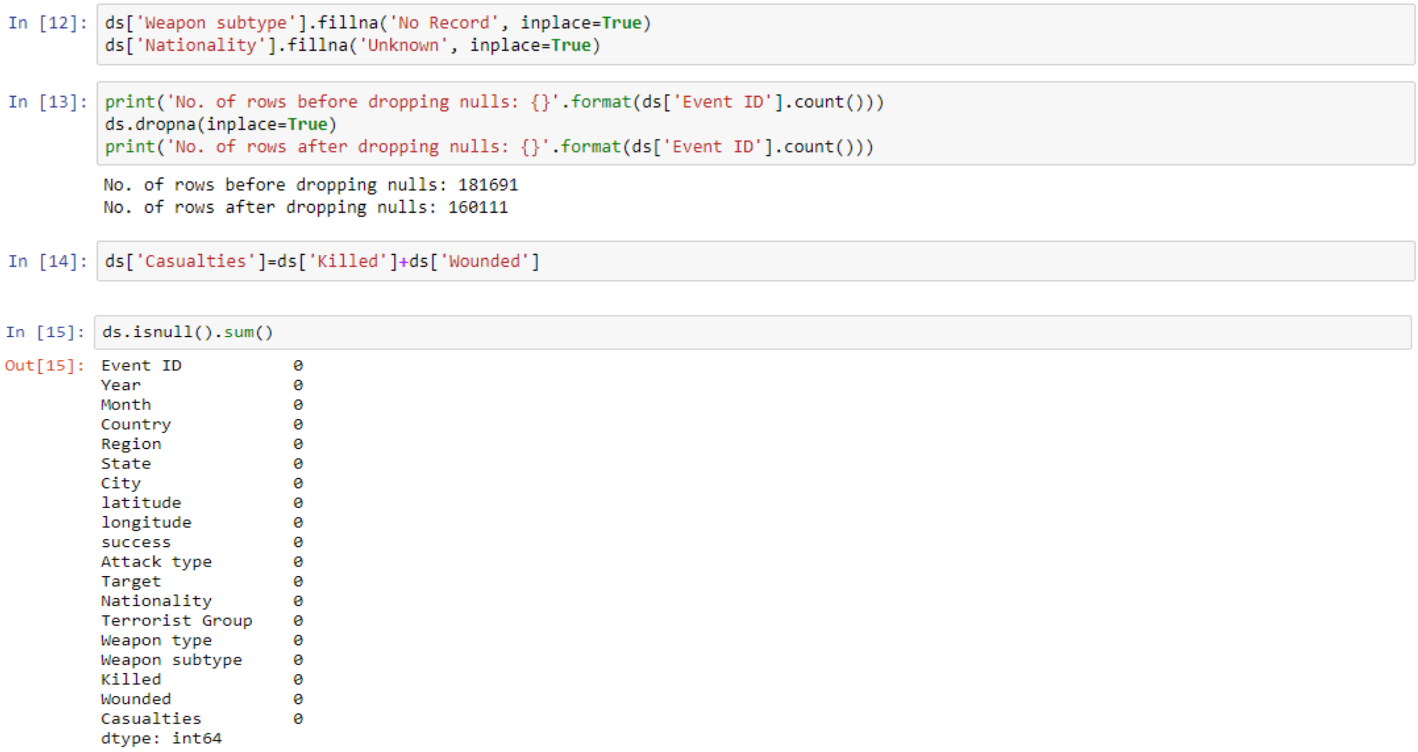
In [10]:
```
# Total no. of Null values in columns
ds.isnull().sum()
```

Out[10]:
```
Event ID               0
Year                   0
Month                  0
Country                0
Region                 0
State                421
City                 434
latitude            4556
longitude           4557
success                0
Attack type            0
Target                 0
Nationality         1559
Terrorist Group        0
Weapon type            0
Weapon subtype     20768
Killed             10313
Wounded            16311
dtype: int64
```

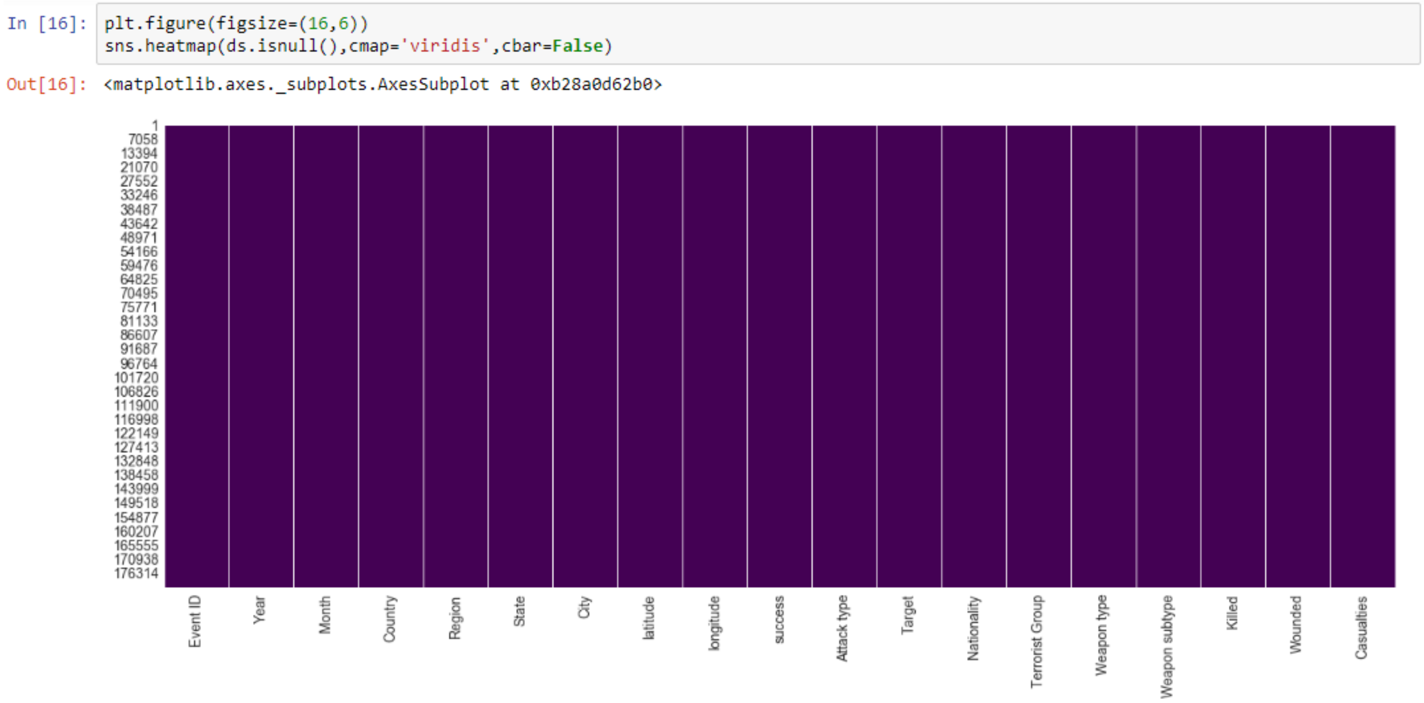


```
In [11]: # Checking the Null values through visualization
         plt.figure(figsize=(16,6))
         sns.heatmap(ds.isnull(),cmap='viridis',cbar=False)

Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0xb2860e2358>
```

Here the yellow lines representing Null values.



```
In [12]: ds['Weapon subtype'].fillna('No Record', inplace=True)
         ds['Nationality'].fillna('Unknown', inplace=True)

In [13]: print('No. of rows before dropping nulls: {}'.format(ds['Event ID'].count()))
         ds.dropna(inplace=True)
         print('No. of rows after dropping nulls: {}'.format(ds['Event ID'].count()))

         No. of rows before dropping nulls: 181691
         No. of rows after dropping nulls: 160111

In [14]: ds['Casualties']=ds['Killed']+ds['Wounded']

In [15]: ds.isnull().sum()

Out[15]: Event ID           0
         Year               0
         Month              0
         Country            0
         Region             0
         State              0
         City               0
         latitude           0
         longitude          0
         success            0
         Attack type        0
         Target             0
         Nationality        0
         Terrorist Group    0
         Weapon type        0
         Weapon subtype     0
         Killed             0
         Wounded            0
         Casualties         0
         dtype: int64
```

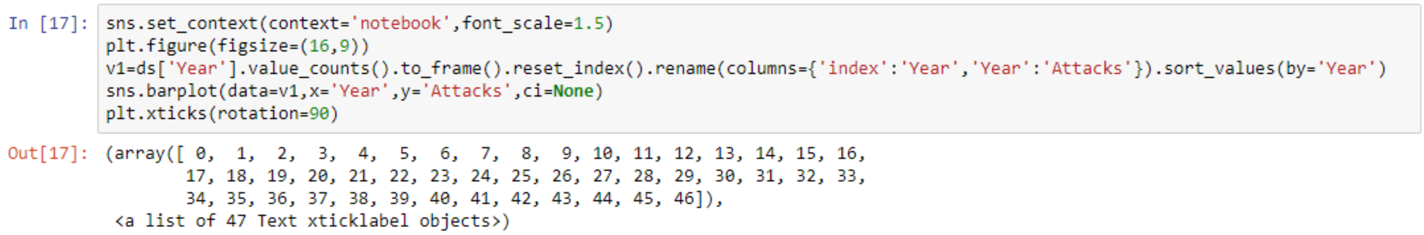


```
In [16]: plt.figure(figsize=(16,6))
         sns.heatmap(ds.isnull(),cmap='viridis',cbar=False)

Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0xb28a0d62b0>
```

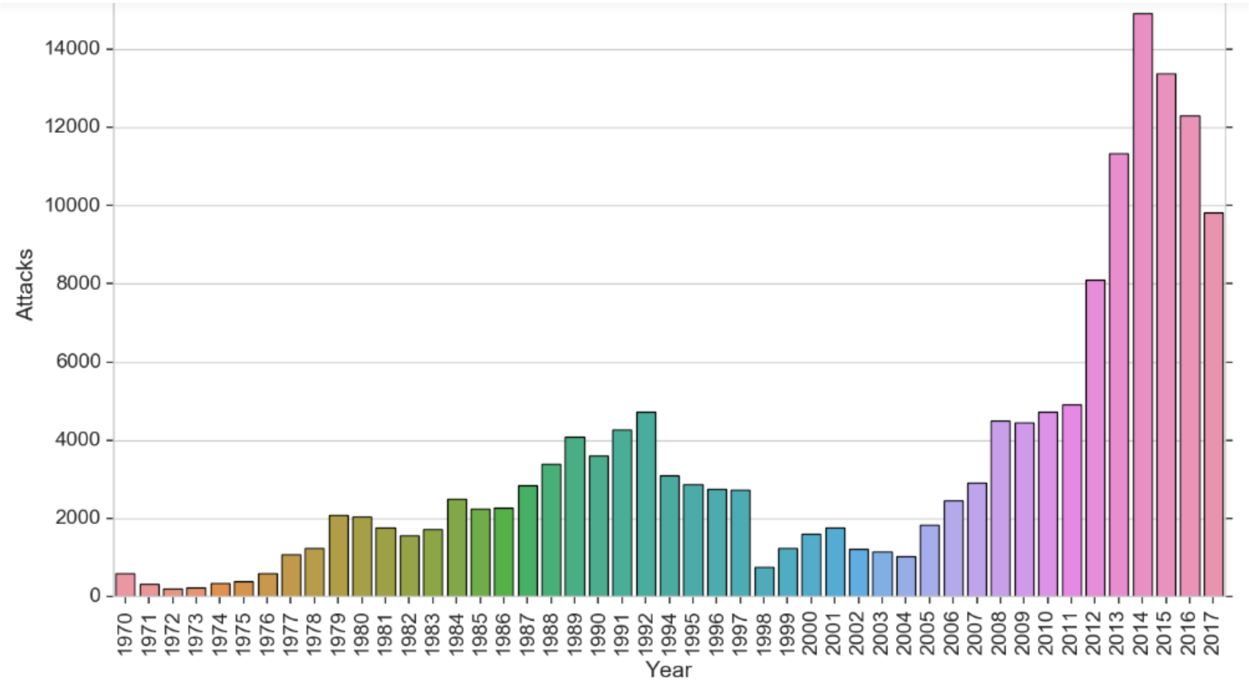
Now our data is clean since there are no yellow lines.

4. Visualizing the Dataset

4.1. Number of attacks worldwide (1970 - 2017)



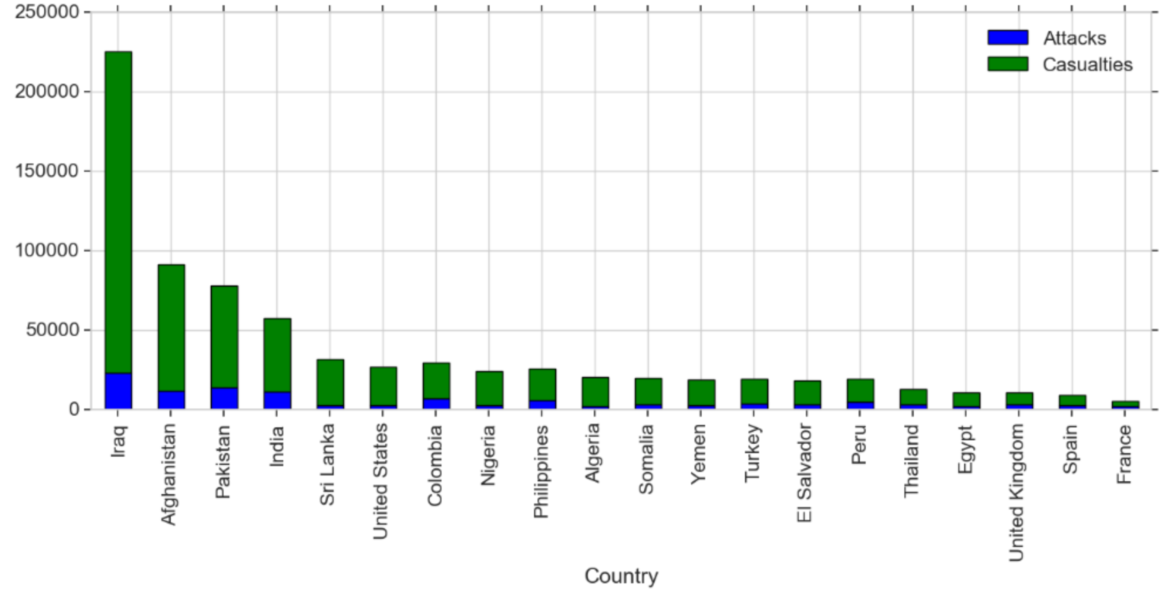
```
In [17]: sns.set_context(context='notebook',font_scale=1.5)
         plt.figure(figsize=(16,9))
         v1=ds['Year'].value_counts().to_frame().reset_index().rename(columns={'index':'Year','Year':'Attacks'}).sort_values(by='Year')
         sns.barplot(data=v1,x='Year',y='Attacks',ci=None)
         plt.xticks(rotation=90)

Out[17]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
                17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
                34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46]),
          <a list of 47 Text xticklabel objects>)
```



4.2. Most affected countries

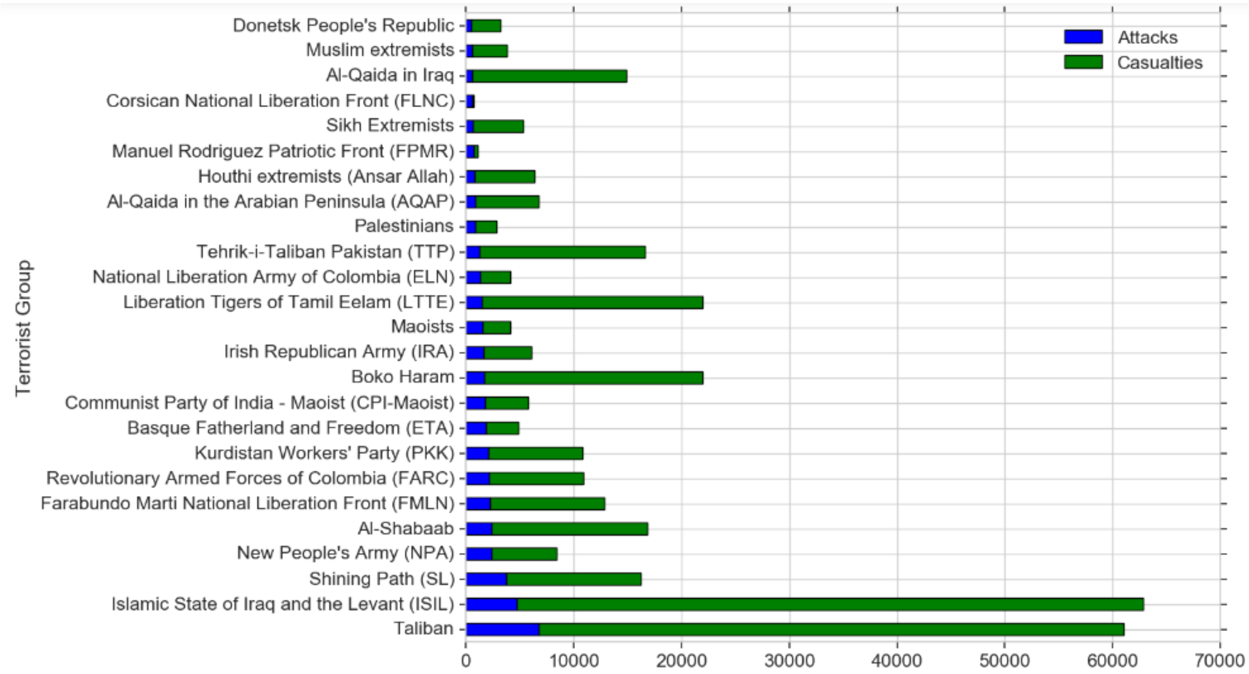
```
In [18]: ds[ds['Country'].isin(ds['Country'].value_counts()[0:20].index)][['Casualties','Country']].groupby('Country').sum().reset_index()
```

```
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0xb289b8dda0>
```



4.3. Most active terrorist groups

```
In [19]: ds[ds['Terrorist Group'].isin(ds['Terrorist Group'].value_counts()[1:26].index)][['Casualties','Terrorist Group']].groupby('Terro
```



5. Analysis of specific country (Iraq)

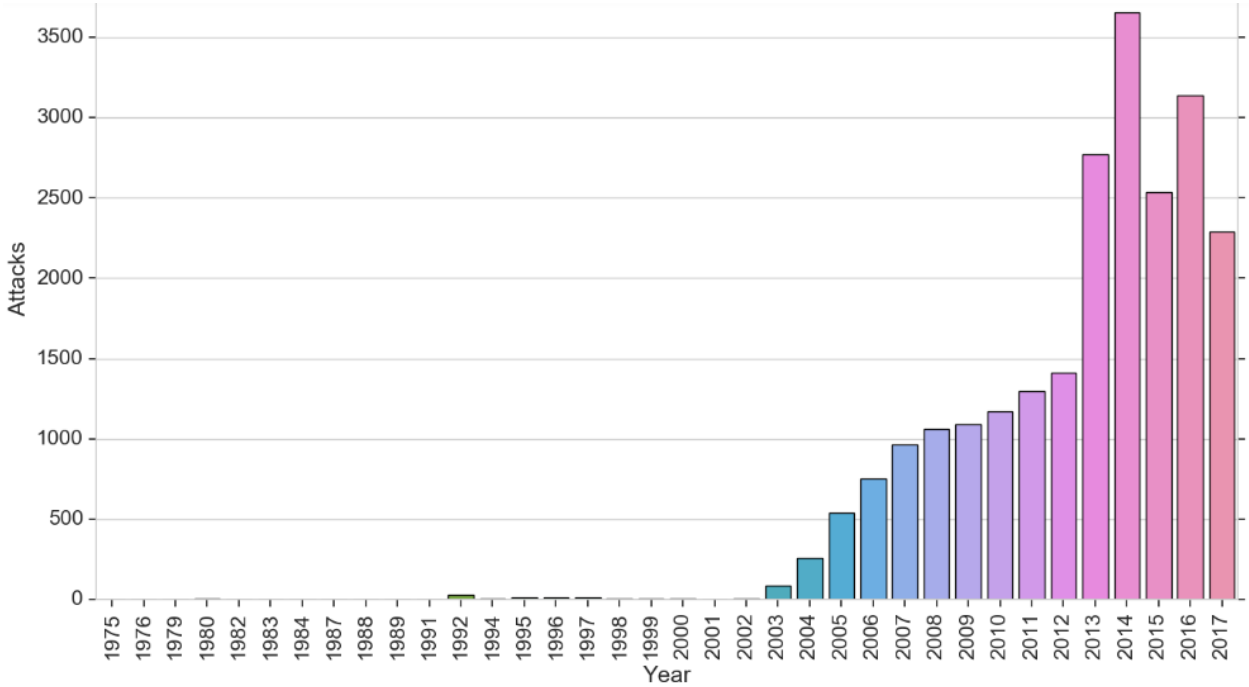
In [20]:

```
Iraq = ds[ds['Country']=='Iraq']
```

5.1. Attacks in iraq (1975 - 2017)

In [21]:

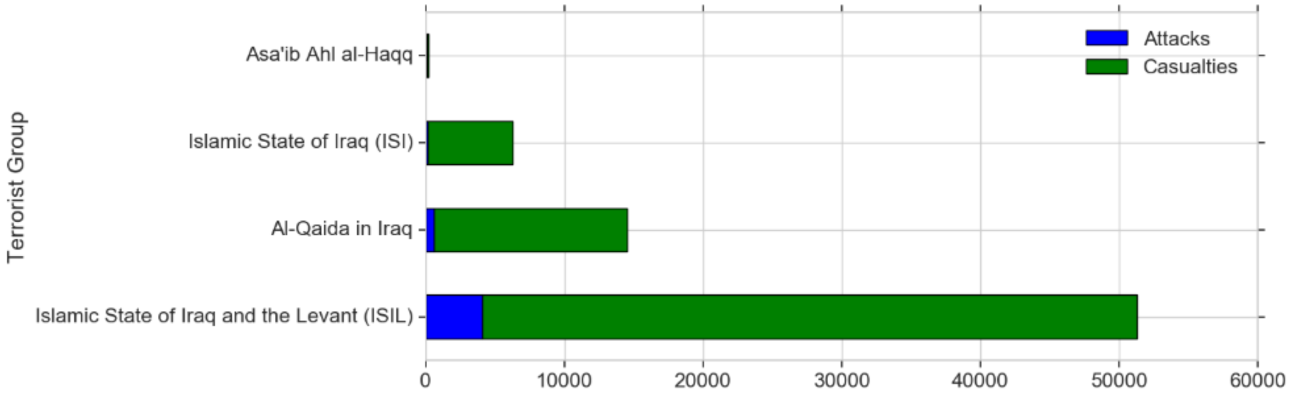
```
sns.set_context(context='notebook',font_scale=1.5)
plt.figure(figsize=(16,9))
v1 = Iraq['Year'].value_counts().to_frame().reset_index().rename(columns={'index':'Year','Year':'Attacks'}).sort_values(by='Year'
sns.barplot(data=v1,x='Year',y='Attacks',ci=None)
plt.xticks(rotation=90)
```



5.2. Most active terror groups in Iraq

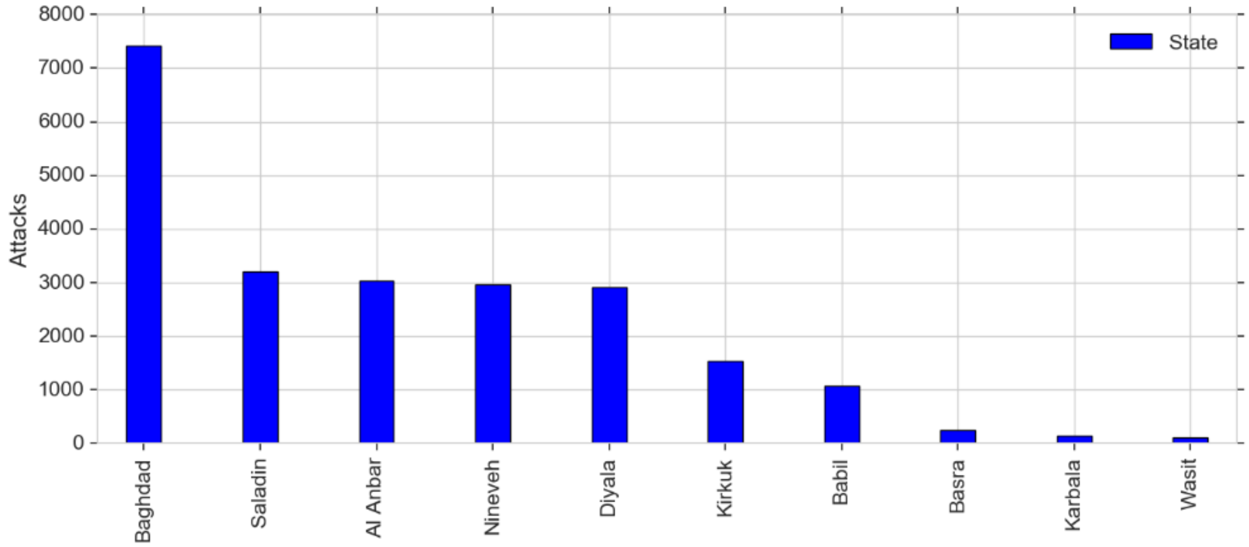
In [22]:
```
Iraq[Iraq['Terrorist Group'].isin(Iraq['Terrorist Group'].value_counts()[1:5].index)][['Casualties','Terrorist Group']].groupby('
```

Out[22]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xb28cd7ca90>
```



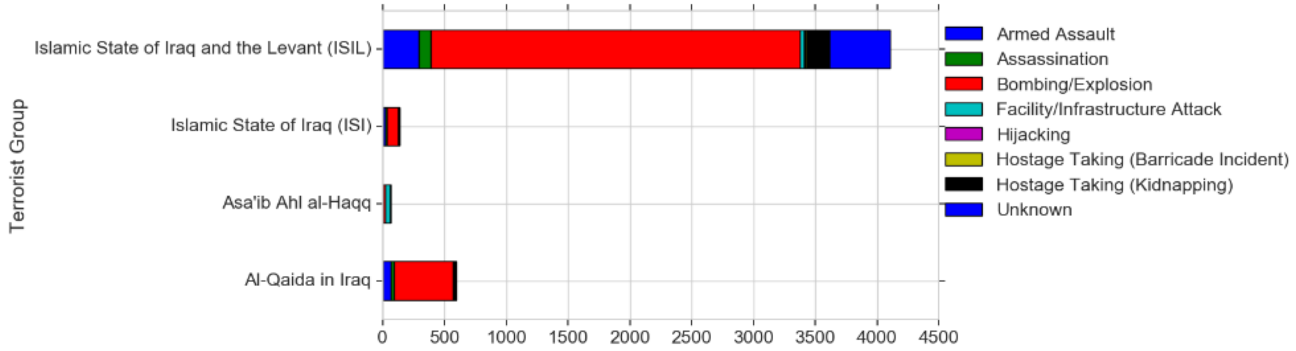
5.3. Most affected states

In [23]:
```
Iraq[Iraq['State'].isin(Iraq['State'].value_counts()[0:10].index)]['State'].value_counts().to_frame().plot.bar(figsize=(16,6),wid
plt.ylabel('Attacks')
```



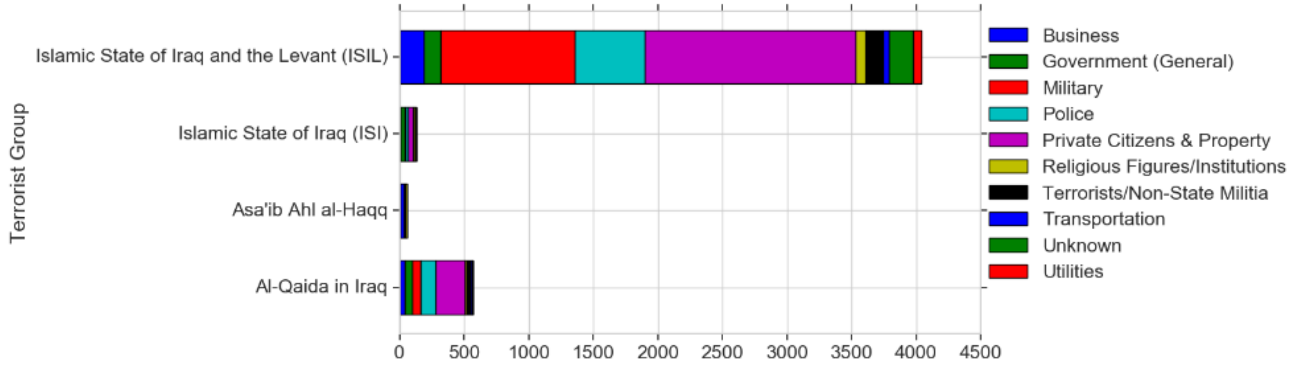
5.4. Attack types

```
In [24]: v1=Iraq[Iraq['Terrorist Group'].isin(Iraq['Terrorist Group'].value_counts()[1:5].index)]
         pd.crosstab(v1['Terrorist Group'],v1['Attack type']).plot.barh(stacked=True,figsize=(9,5),width=0.5)
         plt.legend(loc=9,bbox_to_anchor=(1.32,1))

Out[24]: <matplotlib.legend.Legend at 0xb289a82080>
```



5.5. Favourite targets of different terror groups

```
In [25]: v1=v1[v1['Target'].isin(v1['Target'].value_counts()[0:10].index)]
         pd.crosstab(v1['Terrorist Group'],v1['Target']).plot.barh(stacked=True,figsize=(9,5),width=0.7)
         plt.legend(loc=9,bbox_to_anchor=(1.27,1))

Out[25]: <matplotlib.legend.Legend at 0xb28f672da0>
```



5. Conclusions

1. 2014 is the deadliest year in terrorism with more than 14000 attacks happen all around the world.

2. Iraq is the most affected country followed by Afghanistan, Pakistan, India, Sri Lanka and United States. I am quite surprised that India and United States are in top 5 countries in term of casualties.

3. In term of casualties, the ISIL is the most deadliest terror group followed by Taliban. But in no. of attacks Taliban is leading.
   Boko haram, LTTE, TTP, Al-Qaida and Al-Shabab are others major terrorist group.

4. Until 2002 there were only few terrorist incidents happened in Iraq. But after 2002 it's grows exponentially reaching an all time high in 2014. The possible reason can be since USA invaded Iraq in 2003.

5. ISIL is the most deadliest terror group in Iraq followed by Al-Qaida and ISIS.

6. Bagdad the capital of Iraq is the most affected state with almost twice more attacks happened than in Saladin which is on 2nd place.

7. Bombing is the most used attack type by ISIL in Iraq.

8. Private citizens & property followed by military and police are the highly targeted group in Iraq.