Data Analysis

On

Global Terrorism Attack

Using

Python

Under The Guidance of

Mr. KAUSHIK GHOSH

Group Members

MANISH SHARMA, BUIE

IMANKALYAN SARKAR, RCCIIT

SAYANTIKA BISWAS, RCCIIT

SUBHAM SAMANTA, IEM

SAYAK LAHIRI, IEM

RUDRADEEP CHAUDHURI, UEM

BISWADEEP NANDI, UEM

ACKNOWLEDGEMENT

I take this opportunity to express my profound gratitude and deep regards to my

faculty Mr. KAUSHIK GHOSH for his exemplary guidance, monitoring and constant

encouragement throughout the course of this project. The blessing, help and guidance

given by him time to time shall carry me a long way in the journey of life on which I

am about to embark.

I am obliged to my project team members for the valuable information provided by

them in their respective fields. I am grateful for their cooperation during the period of

my assignment.

Abstract

Recently, there has been a dramatic increase in both the amount of terrorism

related literature and public interest towards the matter. Even so, it is difficult to

gain access to relevant data. This is because much of the available data is not well

organized nor of high quality, and available data is not very well presented. The

GTD (Global Terrorism Database) is an open-source database that has information

on 75,000 terrorist events around the world since 1970 through 2004. It is managed

by the National Consortium for the Study of Terrorism and Responses to

Terrorism (START) at the University of Maryland. This report will introduce the

GTD Explorer, a web-based interactive visual exploratory tool that deals with this

data. It counts the number of incidents grouped over a certain criteria, and stack

charts on top of each other to see both individual and accumulated patterns of

incidents over time. This tool provides insight to experts while making the data

approachable and informative. Making the visualization light-weight and

universally accessible was one of the main concerns. Since this is a web-based

tool, there are certain challenges and limitations to the approach. Details of the

implementation will be covered. Interesting facts that were found using the tool

will be discussed.

**Table of Contents:**

I. Introduction

* Context
* Aims and Objectives
* Basic Project Working Data

II. Pre-Processing

* Describing the Data
* Cleaning the Data
* Categorizing Perpetrators
* Categorizing Targets
* Filling Missing Data

III. Exploratory Data Analysis

* Visual Exploration
  + Casualties by Year
  + Casualties by Region
  + Casualties and Attacks by Country
  + Top Perpetrators
  + Terror Attacks by Weapon Type
  + Terror Tactics throught the Years
* Feature Analysis
  + Feature Selection
  + Target Selection

**Introduction**

The past few decades have been painted with growing geo-political instability across the world and terrorism has been one of the main ways through which this global decay manifested itself.

In recent years, increased access to technology has allowed the average individual to get a deeper insight in what terror acts occurring around the world. As such, terrorism attacks have been increasingly mediatized. I thought it would be interesting to explore what is all over the news from another approach, a more methodic one so to say.

Dealing with terrorism has become the quintessential social problem for civilized societies. The world has changed

in many ways during the last decade in response to this increased awareness. Yet, terrorism has always been

existent.

Recently, there has been a dramatic increase in both the amount of terrorism related literature and public interest

towards the matter. Even so, it is difficult to gain access to relevant data. This is because much of the available

data is not well organized nor of high quality, and available data is not very well presented.

The Global Terrorism Database is an open-source database that has information of terrorist events around

the world since 2013 through 2016. It is managed by the National Consortium for the Study of Terrorism and

Responses to Terrorism (START) at the University of Maryland, which claims to be “the most comprehensive

open source data set on terrorism”. The START intends to make this data more accessible to the public and this

tool is one way of doing so.

The Global Terrorism Database (GTD) records terror attacks around the world from 1970 through 2015, it includes systematic data on domestic and international terrorist incidents durint this time period.

1. Contains information on over 5,000 terrorist attacks
2. Currently the most comprehensive unclassified data base on terrorist events in the world
3. Includes information on more than 2,000 bombings, 1,500 assassinations since 2013
4. Includes information on at least 45 variables for each case, with more recent incidents including information on more than 120 variables.

**Aims and Objectives**

* Visually exploring the extremely rich data on the 5,000 terrorist attacks, trying to answer questions such as:
  + Which countries/region are the most targeted?
  + Where are there the most casualties?
  + How have casualties evolved throughout the years?
  + What are the casualties by weapon type?
  + Are some countries better at defending themselves against terrorist attacks?

**Basic Project Working Data**

Project Plan:

Use Python Code to do the following:​

* Suggest multiple suitable hypothesis.​
* Clean the Data​
* Do distribution analysis​
* Represent our data in different graphical representation and with detailed analysis​

Working hours:

We expect a time of 35-40 manhours for the completion of the project

Distribution analysis:

Description of the data:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Year | Month | Day | Nkill | Nwound | Ransomamt |
| Count | 43915 | 43915 | 43915 | 41290 | 39313. | 2.730000e+02 |
| Mean | 2014.688079 | 6.442537 | 15.664716 | 3.110463 | 3.504006 | 4.066460e+06 |
| Std | 1.229350 | 3.390650 | 8.778536 | 13.343306 | 13.988988 | 2.180638e+07 |
| Min | 2012.000000 | 1.000000 | 1.000000 | 0.000000 | 0.000000 | -9.900000e+01 |
| 25% | 2014.000000 | 4.000000 | 8.000000 | 0.000000 | 0.000000 | -9.900000e+01 |
| 50% | 2015.000000 | 6.000000 | 15.000000 | 1.000000 | 0.000000 | 1.58604e+04 |
| 75% | 2016.000000 | 9.000000 | 23.000000 | 2.000000 | 3.000000 | 2.189562e+05 |
| Max | 2016.000000 | 12.000000 | 31.000000 | 1500.000000 | 1500.000000 | 2.000000e+08 |

**Correlation among the values**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Year | Month | Day | Nkill | Nwound | Ransomamt |
| Year | 1.000000 | -0.032845 | -0.002996 | -0.019340 | -0.025865 | -0.165084 |
| Month | -0.032845 | 1.000000 | 0.002112 | 0.005468 | -0.001070 | 0.077339 |
| Day | -0.002996 | 0.002112 | 1.000000 | -0.008516 | -0.001725 | 0.041875 |
| Nkill | -0.019340 | 0.005468 | -0.008516 | 1.0000000 | 0.364357 | 0.029284 |
| Nwound | -0.025865 | -0.001070 | -0.001725 | 0.364357 | 1.000000 | -0.038720 |
| Ransomamt | -0.165084 | 0.077339 | 0.041875 | 0.029284 | -0.038720 | 1.000000 |

**Content:**

1. Data Information:
   1. Data Info
   2. Data Describe
   3. Data Correlation
   4. Data Correlation Map
   5. Features of Data
2. Data Visualize & Analyze:
   1. US Terror Attacks Death and Injuries
   2. Death and Injuries at All Time
   3. Frequency of Terrorism Actions
   4. Frequency of Terrorist Actions in Customized Region
   5. Comparing People Who Died in Terrorist Attacks in The World and in The Region
3. Graphics and Analysis on Customized Data:
   1. Terrorist Attacks of a Particular Year and Their Locatios
   2. Total Number of People Killed in Terror Attack
   3. Types of Attacks That Cause Deaths
   4. Number of Killed in Terrorist Attacks by Countries

Relevent data:

The dataset has the following fields​

* iyear, imonth and iday: data of event​
* country and city​
* attacktype1\_txt: attack type​
* targtype1\_txt: target institution type​
* corp1: target institution​
* target1: target population type​
* natlty1\_txt: nationality of attacker​
* gname: name of attacking organization/terrorist group​
* weaptype1\_txt, weapsubtype1\_txt and weapdetail: describes the weapon of attack​
* nkill and nwound: number of people killed and wounded​
* propextent\_txt: extent of property damage​
* ransomamt: ransom amount (e.g. in case of hostage situations): values like 0 or negative indicate that either this was not a ransom case or the amount was unknown.

Possible approaches for data wrangling:

According to the dataset given we see that some of the fields in the given dataset are provided with NaN. So we need to take care of those fields and we must remove some of the unneccesary columns which may be irrelevant for our analysis.

1. In the column targertype1\_txt we see that some of the values are given as NaN. We can fill then with the most frequent target type (Private citizen) as per the given data set.

|  |  |
| --- | --- |
| Private citizens & property | 11780 |
| Military | 9419 |
| Police | 6691 |
| Government(general) | 3510 |
| Business | 3124 |
| Unknown | 2412 |

2. In the column ‘natlty1\_txt’ we see some values are missing . we can fill it up by the most frequent nationality of that country. For example

Afghanistan

|  |  |
| --- | --- |
| Nationality | Count |
| Afganistan | 5316 |

3. In the column ‘corp1’ we see that some samples have data as ‘Not Applicable’ we will replace them with NaN for our better computational efficiency.

Not Applicable count count in corp1: 737

4. In the column ‘gname’ we see that the maximum frequency of data lies in the unknown category . so we find an outlier . but we cannot remove it as it is difficult to predict the group name.

|  |  |
| --- | --- |
| Unknown | 12210 |
| Taliban | 4900 |
| Islamic State of Iraq and the Levant(ISIL) | 4273 |
| Al-Shabaab | 2350 |
| Boko Haram | 1921 |
| Maoists | 1124 |
| New People’s Army(NPA) | 1076 |

5: in the column weap\_detail is unneccesary as it is a mere decription of firearm type.

6.in the column propextent\_txt it will be useful to leave the NaN variables alone .

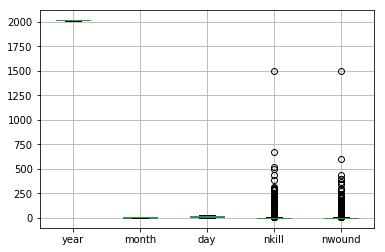
7: the column ransomamt is unneccesary as we see from the dataset that most of the terrorism didnt result to ransom amt.

no of nan values in ransomant : 43642

8: Now we will rename the columns as follows and drop ransomamt column

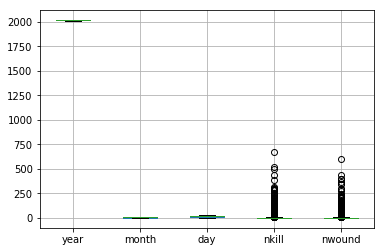
|  |  |
| --- | --- |
| From | To |
| iyear' | ‘year', |
| imonth' | 'month' |
| 'iday' | 'day' |
| 'country\_txt | 'country' |
| ‘attacktype1\_txt’ | 'attack1' |
| 'targtype1\_txt' | 'target0',' |
| natlty1\_txt' | 'nationality' |
| 'weaptype1\_txt' | ‘weapon' |
| 'weapsubtype1\_txt' | 'weapon\_sub' |
| 'propextent\_txt' | 'damage' |

9: From the box plot given below we see the following outliers



We see that nkill and n wound has values greater than or equal to 1500. So we remove the values.

After removing we see that the data is uniformly spread..



10: we find that in nkill column we find 2625 nan values. So we replace them by their mean.

11: we find that in the nwound column there are 4602 nan values . so we replace them by the mean nwound.

Parameters:

* change in terrorism frequency in the years for a country

.

* count of severe attacks and non severe attacks on acountry for a year
* Distribution of weapons
* To see distribution of global terrorism through out years
* to display the top 10 countries affected by terrorism.
* to see how each country was attacked
* most vernerable terrorist group
* top 10 terrorist group with highest kill to wound ratio
* top 10 vernerable places to terrorism
* favoured way of attack of each terrorist group
* distribution of attack places according to the year
* attack places distribution by each terrorist group
* categories of damage per year
* top 10 attack targets
* Distribution of terrorism within cities
* no of domestic and international casualties
* which weapon type was the most vurnerable
* distribution of terrorism within months of year
* attack prob of each country
* military and religious attacxks in years
* distribution of global terrorism in the world map

.

# Parameter 1: Change in terrorism frequency in the years for a country

tempdf=df.groupby(df.country)

with PdfPages ('year\_wise\_terrorism.pdf') as pdf:

for name,group in tempdf:

print("\n",name,"\n")

t=group.groupby(group.year)

k=t.year.agg(np.count\_nonzero)

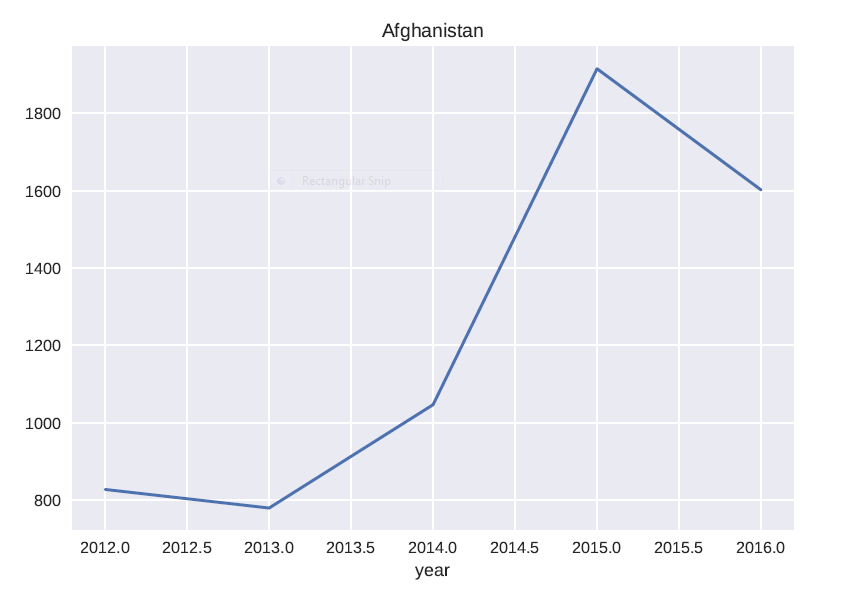
k2=k.diff()

k.plot(title=name)

pdf.savefig()

plt.close()

Output:



# Parameter 2: Count of severe attacks and non severe attacks on a country for a year

severe = df[df.nkill>df.nwound]

non\_severe= df[df.nkill<df.nwound]

print(severe,"\n\n",non\_severe)

with PdfPages ('severity\_test.pdf') as pdf:

tempdf1=severe.groupby(severe.country)

tempdf2=non\_severe.groupby(non\_severe.country)

for name,group in tempdf1:

for name1,group1 in tempdf2:

g1=group.groupby(group.year)

g2=group1.groupby(group1.year)

if name==name1:

k1=g1.year.agg(np.count\_nonzero)

k2=g2.year.agg(np.count\_nonzero)

k1.plot(kind="bar",title=name+' severe')

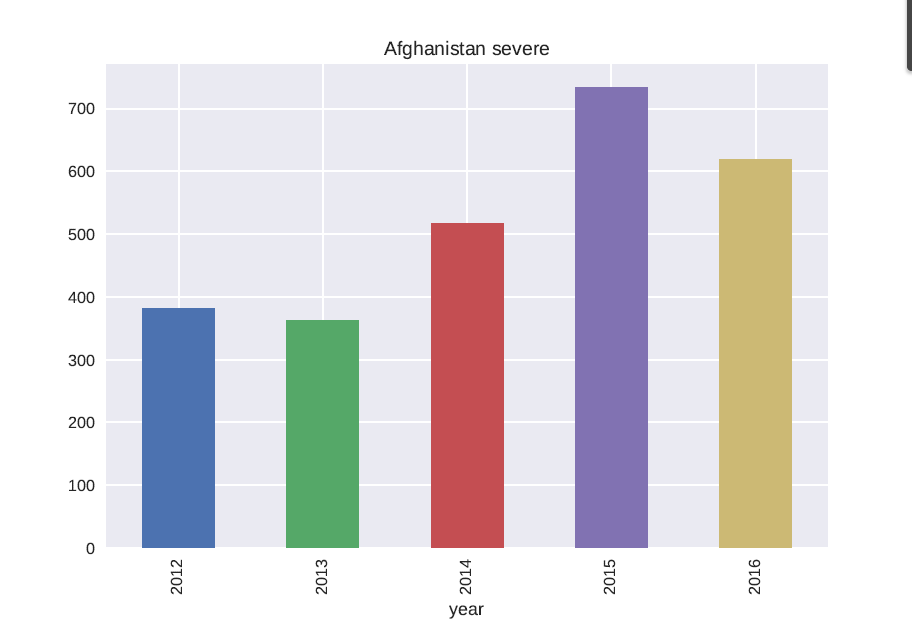
pdf.savefig()

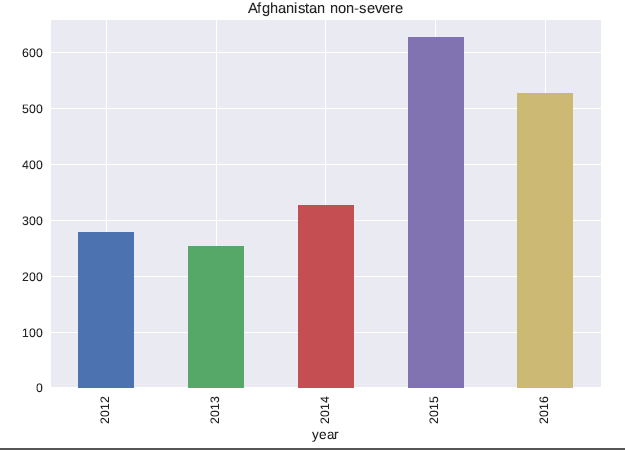
plt.close()

k2.plot(kind="bar",title=name1+' non-severe')

pdf.savefig()

plt.close()





# Parameter 3: Distribution of weapons

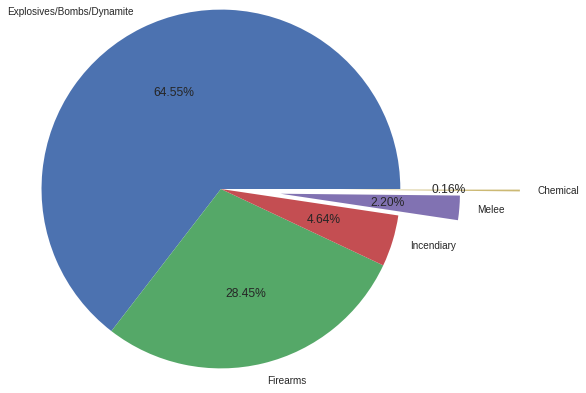
tempdf=df.groupby(df.weapon)

df1=tempdf.weapon.agg(np.count\_nonzero).sort\_values(ascending=False).head(5)

df1.plot(kind="pie",subplots=True,autopct="%.2f%%",explode=[0,0,0,0.5,1],radius=1.5)

# print(df1)

Output:

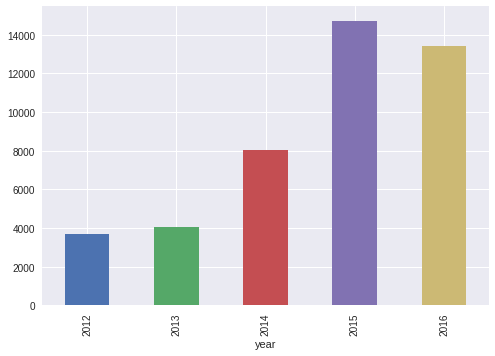


# Parameter 4 : To see distribution of global terrorirm through out years

tempdf=df.groupby(df.year)

df1=tempdf.year.agg(np.count\_nonzero)

df1.plot(kind="bar")



# Parameter 5: To display the top 10 countries affected by terrorism

tempdf=df.groupby(df.country)

df1=tempdf.country.agg(np.count\_nonzero).sort\_values(ascending=False).head(10)

print (df1)

Output:

country

Iraq 7977

Afghanistan 6172

India 3277

Pakistan 3271

Nigeria 2457

Philippines 2251

Somalia 2233

Yemen 2099

Syria 1314

Turkey 1204

# Parameter 6 : To see how each country was attacked

tempdf=df.groupby(df.country)

with PdfPages ('param6.pdf') as pdf:

for name,group in tempdf:

print("\n",name,"\n")

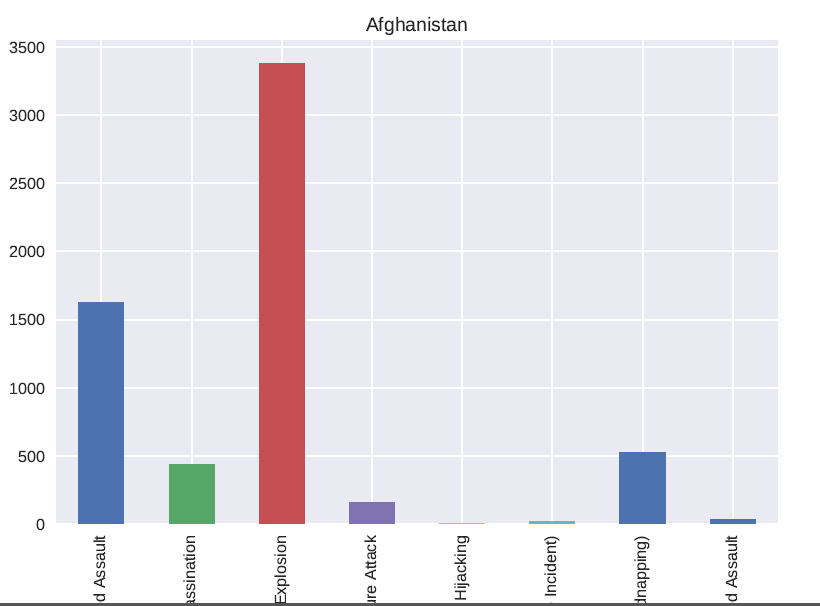
t=group.groupby(group.attack1)

k=t.attack1.agg(np.count\_nonzero)

fig=k.plot(kind="bar",title=name).get\_figure()

pdf.savefig(fig)

Output:



# Parameter 7 : Most vurnerable terrorist group

df1=df.copy()

df1['total']=np.nan

df1['total']=df['nkill']+df['nwound']

tempdf=df1.groupby(df1.gname)

dict1={'name':[],'total\_cas':[]}

for name,group in tempdf:

g1=group.total.agg(np.sum)

dict1['name'].append(name)

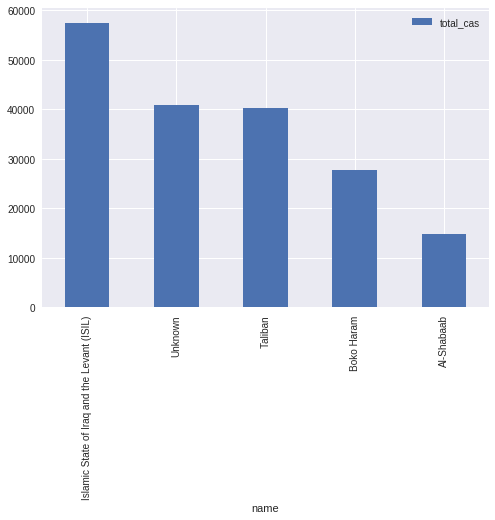
dict1['total\_cas'].append(g1)

df2=pd.DataFrame.from\_dict(dict1)

res=df2.sort\_values(by=['total\_cas'],ascending=[False]).head(5)

res.plot(x=res.name,kind='bar')

Output:



# Parameter 8 : Top 10 terrorist group with highest kill to wound ratio

df1=df.copy()

tempdf=df1.groupby(df1.gname)

dict1={'name':[],'ratio':[]}

for name,group in tempdf:

kill=group.nkill.agg(np.sum)

wound=group.nwound.agg(np.sum)

if wound!=0:

res1=kill/wound

else:

res1=0

dict1['name'].append(name)

dict1['ratio'].append(res1)

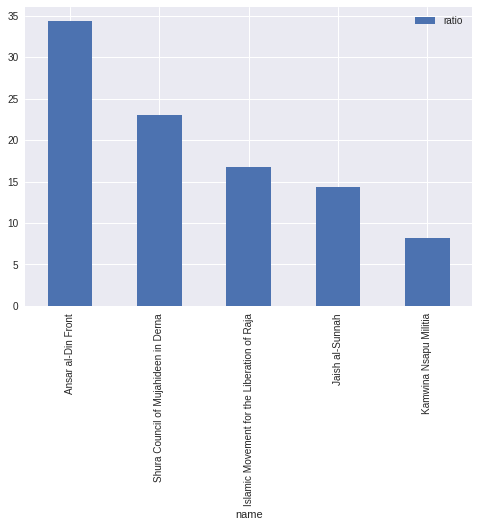
df2=pd.DataFrame.from\_dict(dict1)

res=df2.sort\_values(by=['ratio'],ascending=[False]).head(5)

print(res)

res.plot(x=res.name,kind='bar')

Output:



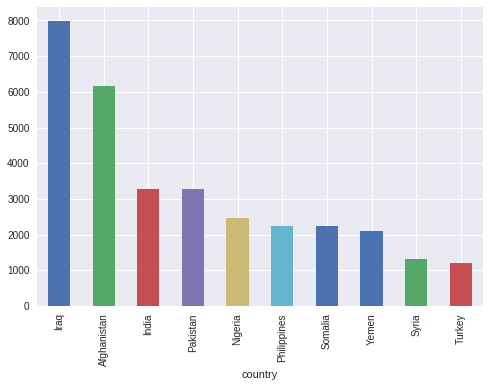
# Parameter 9 : Pop 10 vurnarable places to terrorism

tempdf=df.groupby(df.country)

res=tempdf.country.agg(np.count\_nonzero).sort\_values(ascending=False).head(10)

res.plot(kind="bar")

Output:



# Parameter 10 : Favoured way of attack of each terrorist group

tempdf=df.groupby(df.gname)

for name,group in tempdf:

g1=group.groupby(group.weapon)

res=g1.weapon.agg(np.count\_nonzero).sort\_values(ascending=False).head(1)

print("name: ",name,end="\t")

res1=res.to\_dict()

for key in res1:

print("way: ",key)

Output:

name: 313 Brigade (Syria) way: Explosives/Bombs/Dynamite

name: A'chik Matgrik Elite Force (AMEF) way: Explosives/Bombs/Dynamite

name: Aba Cheali Group way: Explosives/Bombs/Dynamite

name: Abbala extremists way: Firearms

name: Abdul Ghani Kikli Militia way: Firearms

name: Abdul Qader Husseini Battalions of the Free Palestine movement way: Explosives/Bombs/Dynamite

name: Abdullah Azzam Brigades way: Explosives/Bombs/Dynamite

name: Abida Tribe way: Firearms

name: Abu Amarah Battalion way: Explosives/Bombs/Dynamite

name: Abu Bakr Unis Jabr Brigade way: Firearms

name: Abu Jaafar al-Mansur Brigades way: Explosives/Bombs/Dynamite

name: Abu Obaida bin Jarrah Brigade way: Explosives/Bombs/Dynamite

name: Abu Salim Martyr's Brigade way: Explosives/Bombs/Dynamite

name: Abu Sayyaf Group (ASG) way: Firearms

name: Abu Tira (Central Reserve Forces) way: Firearms

name: Aceh Singkil Islamic Care Youth Students Association (PPI) way: Incendiary

name: Achik Matgrik Army (AMA) way: Explosives/Bombs/Dynamite

name: Achik National Cooperative Army (ANCA) way: Explosives/Bombs/Dynamite

name: Achik National Liberation Army (ANLA) way: Explosives/Bombs/Dynamite

name: Achik National Volunteer Council-B (ANVC-B) way: Firearms

name: Achik Songna An'pachakgipa Kotok (ASAK) way: Firearms

name: Achik Tiger Force way: Explosives/Bombs/Dynamite

name: Adan-Abyan Province of the Islamic State way: Explosives/Bombs/Dynamite

name: Adivasi National Liberation Army (ANLA) way: Firearms

name: Adivasi People's Army (APA) way: Explosives/Bombs/Dynamite

name: Afar Revolutionary Democratic Unity Front way: Firearms

name: Agwelek Forces way: Incendiary

name: Ahfad al-Sahaba-Aknaf Bayt al-Maqdis way: Explosives/Bombs/Dynamite

name: Ahle Sunnat Wal Jamaat (ASWJ-Pakistan) way: Firearms

name: Ahlu-sunah Wal-jamea (Somalia) way: Explosives/Bombs/Dynamite

name: Ahmad Luebaesa Group way: Firearms

name: Ahrar al-Sham way: Explosives/Bombs/Dynamite

name: Aisha Umm-al Mouemeneen (Brigades of Aisha) way: Explosives/Bombs/Dynamite

name: Ajnad Misr way: Explosives/Bombs/Dynamite

name: Ajnad al-Sham way: Explosives/Bombs/Dynamite

name: Akhil Terai Mukti Morcha (ATMM) way: Explosives/Bombs/Dynamite

name: Al Bayda Province of the Islamic State way: Melee

name: Al-Aqsa Martyrs Brigade way: Firearms

name: Al-Ashtar Brigades way: Explosives/Bombs/Dynamite

name: Al-Faruq Militia way: Firearms

name: Al-Fateh Al-Jadid way: Explosives/Bombs/Dynamite

name: Al-Furqan Brigades way: Explosives/Bombs/Dynamite

name: Al-Haydariyah Battalion way: Explosives/Bombs/Dynamite

name: Al-Herak Al-Tihami Movement way: Firearms

name: Al-Islah Party way: Explosives/Bombs/Dynamite

name: Al-Jihad (Pakistan) way: Incendiary

name: Al-Khobar way: Explosives/Bombs/Dynamite

name: Al-Madani Brigade way: Firearms

name: Al-Mua'qi'oon Biddam Brigade (Those who Sign with Blood) way: Explosives/Bombs/Dynamite

name: Al-Mus'abi Tribe way: Firearms

# Parameter 11: Distribution of attack places according to the year

tempdf=df.groupby(df.year)

with PdfPages ('target\_places\_year.pdf') as pdf:

for name,group in tempdf:

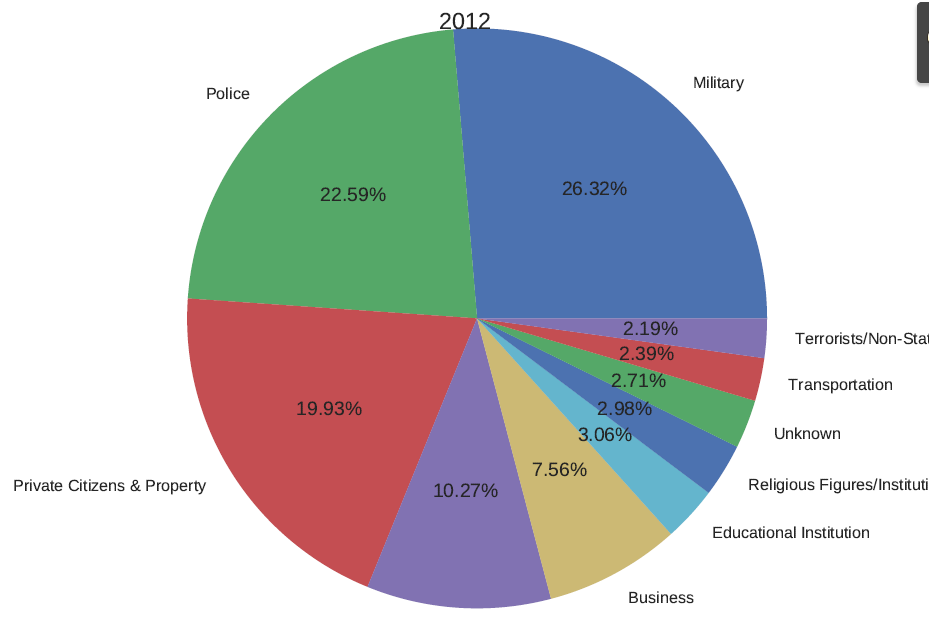
g1=group.groupby(group.target0)

res=g1.target0.agg(np.count\_nonzero).sort\_values(ascending=False).head(10)

res.plot(kind='pie',subplots=True,title=name,autopct="%.02f%%",radius=1.5)

pdf.savefig()

plt.close()



# Parameter 12 : Attack places distribution by each terrorist group

tempdf=df.groupby(['gname'])

with PdfPages ('attack\_places\_gname.pdf') as pdf:

for name,group in tempdf:

print(name,"\n")

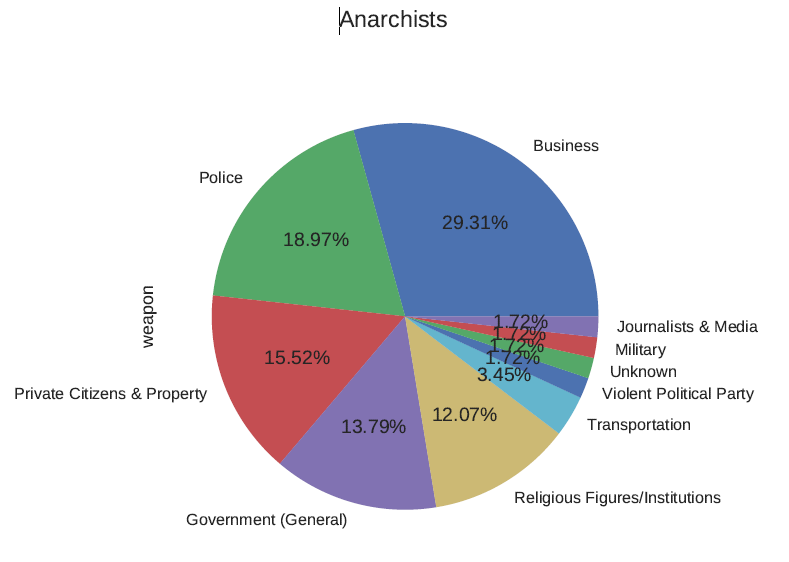
g1=group.groupby(group.target0)

res=g1.weapon.agg(np.count\_nonzero).sort\_values(ascending=False).head(10)

res.plot(kind="pie",subplots=True,title=name,autopct="%.2f%%")

pdf.savefig()

plt.close()



# Parameter 13: Categories of damage per year

tempdf=df.groupby(df.year)

for name,group in tempdf:

print("\n",name,"\n")

g1=group.groupby(group.damage)

print(g1.size())

Output:2012

damage

Major (likely > $1 million but < $1 billion) 1

Minor (likely < $1 million) 1298

Unknown 307

dtype: int64

2013

damage

Major (likely > $1 million but < $1 billion) 2

Minor (likely < $1 million) 1296

Unknown 379

dtype: int64

2014

damage

Major (likely > $1 million but < $1 billion) 3

Minor (likely < $1 million) 2649

Unknown 901

dtype: int64

2015

damage

Minor (likely < $1 million) 3849

Unknown 817

dtype: int64

2016

damage

Major (likely > $1 million but < $1 billion) 12

Minor (likely < $1 million) 2996

Unknown 1104

dtype: int64

# Parameter 14 : Top 10 attack type doing the major number of casualties

df1=df.copy()

df1['total']=np.nan

df1['total']=df1['nkill']+df1['nwound']

tempdf=df1.groupby(df1.attack1)

dict={'name':[],'total\_cas':[]}

for name,group in tempdf:

print("\n",name,"\n")

res=group.total.agg(np.sum)

dict['name'].append(name)

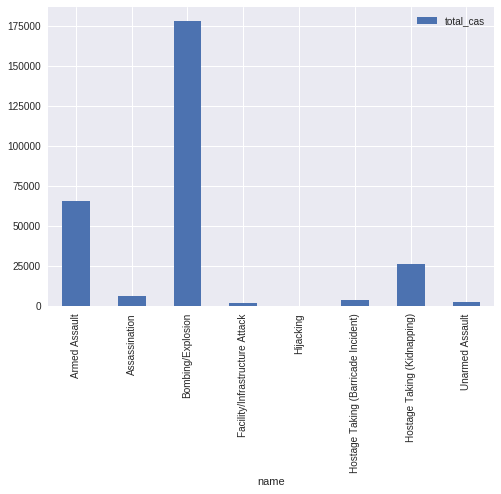
dict['total\_cas'].append(res)

dfr=pd.DataFrame.from\_dict(dict)

#dfr.plot(kind='pie',subplots=True,autopct="%.2f%%",radius=1.5)

dfr.plot(kind="bar",x=dfr.name)

OUTPUT :



# Parameter 15 : Top 10 attack targets

df1=df.copy()

df1['total']=np.nan

df1['total']=df1['nkill']+df1['nwound']

tempdf=df1.groupby(df1.target0)

dict={'name':[],'total\_cas':[]}

for name,group in tempdf:

print("\n",name,"\n")

res=group.total.agg(np.sum)

dict['name'].append(name)

dict['total\_cas'].append(res)

dfr=pd.DataFrame.from\_dict(dict)

#dfr.plot(kind='pie',subplots=True,radius=1.5)

dfr.plot(kind="bar",x=dfr.name)

# Parameter 16: Distribution of terrorism within cities

df1=df.copy()

df1.city.replace('Unknown',np.nan,inplace=True)

tempdf=df1.groupby(df1.country)

with PdfPages ('city.pdf') as pdf:

for name,group in tempdf:

g1=group.groupby(group.city)

res=g1.city.agg(np.count\_nonzero).sort\_values(ascending=False).head(10)

if res.empty:

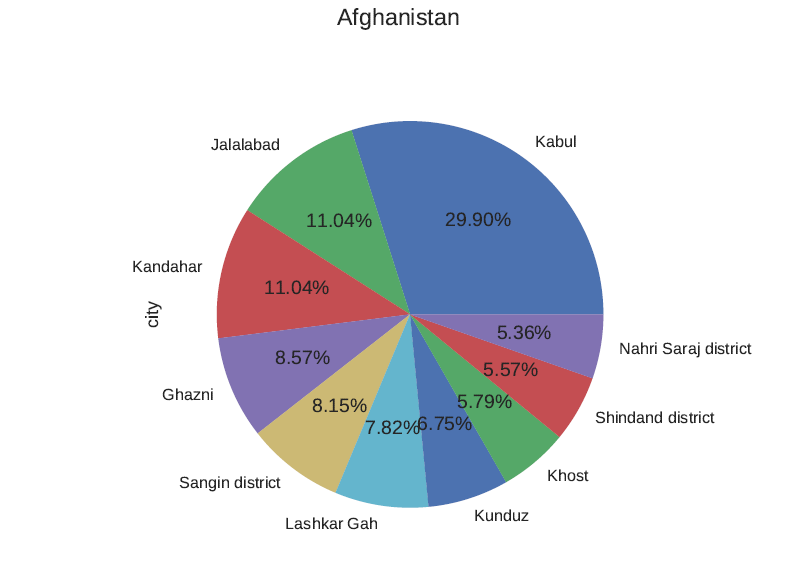
pass

else:

res.plot(kind="pie",title=name,subplots=True,autopct="%.2f%%")

pdf.savefig()

plt.close()



# Parameter 17 : Number of domestic and international casualties

tempdf=df.groupby(df.country)

with PdfPages ('national\_international.pdf') as pdf:

for name,group in tempdf:

ar={'national':0,'international':0}

international=0

print("\n\n",name,"\n")

g1=group.groupby(group.nationality)

res=g1.nationality.agg(np.count\_nonzero)

d1=res.to\_dict()

for key in d1:

if (key==name):

ar['national']=d1[key]

else:

international+=d1[key]

ar['international']=international

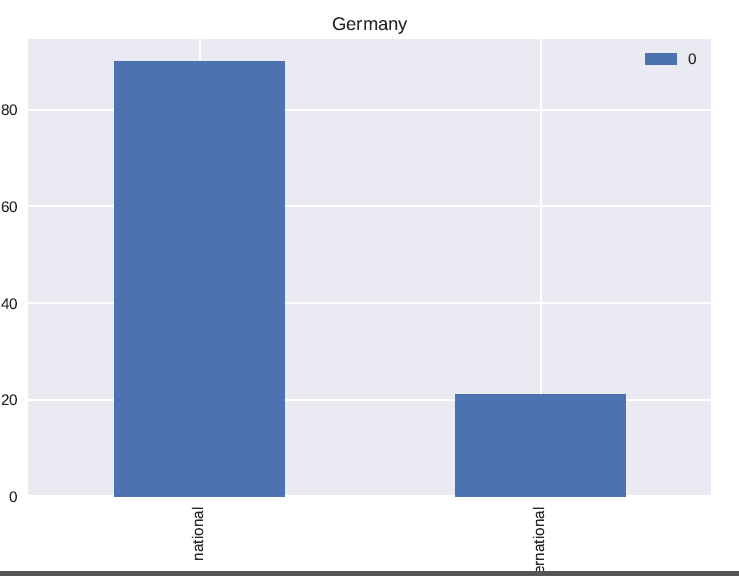
df1=pd.DataFrame.from\_dict(ar,orient='index')

df1.plot(kind='bar',title=name)

pdf.savefig()

plt.close()

Output:



# Parameter 18 : Which weapon type was the most vurnerable

df1=df.copy()

df1['total']=np.nan

df1['total']=df1['nkill']+df1['nwound']

tempdf=df1.groupby(df1.weapon\_sub)

dict={'name':[],'total\_cas':[]}

for name,group in tempdf:

print("\n",name,"\n")

res=group.total.agg(np.sum)

dict['name'].append(name)

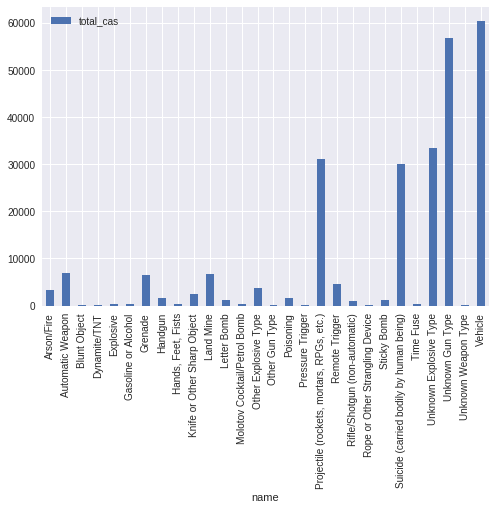
dict['total\_cas'].append(res)

dfr=pd.DataFrame.from\_dict(dict)

#dfr.plot(kind='pie',subplots=True,radius=1.5)

dfr.plot(kind="bar",x=dfr.name)

Output:



# Parameter 19 : Distribution of terrorism within months of year

df1=df.copy()

df1.replace({'month':{1:'jan',2:'feb',3:'mar',4:'apr',5:'may',6:'jun',7:'july',\

8:'aug',9:'sep',10:'oct',11:'nov',12:'dec'}},inplace=True)

tempdf=df1.groupby(df1.year)

with PdfPages ('monthly\_distribution.pdf') as pdf:

for name,group in tempdf:

g1=group.groupby(group.month)

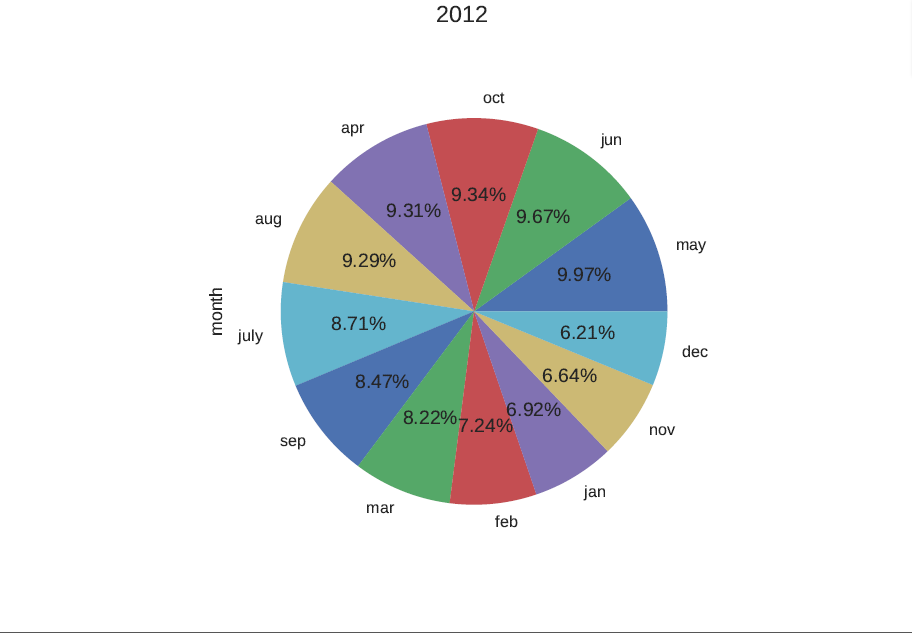
res=g1.month.agg(np.count\_nonzero).sort\_values(ascending=False)

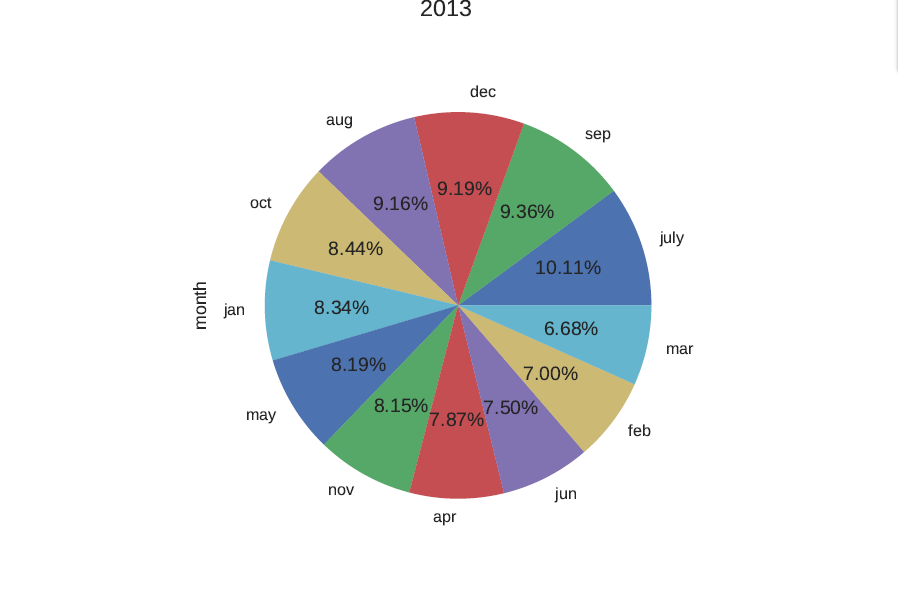
res.plot(kind='pie',subplots=True,title=name,autopct="%.2f%%")

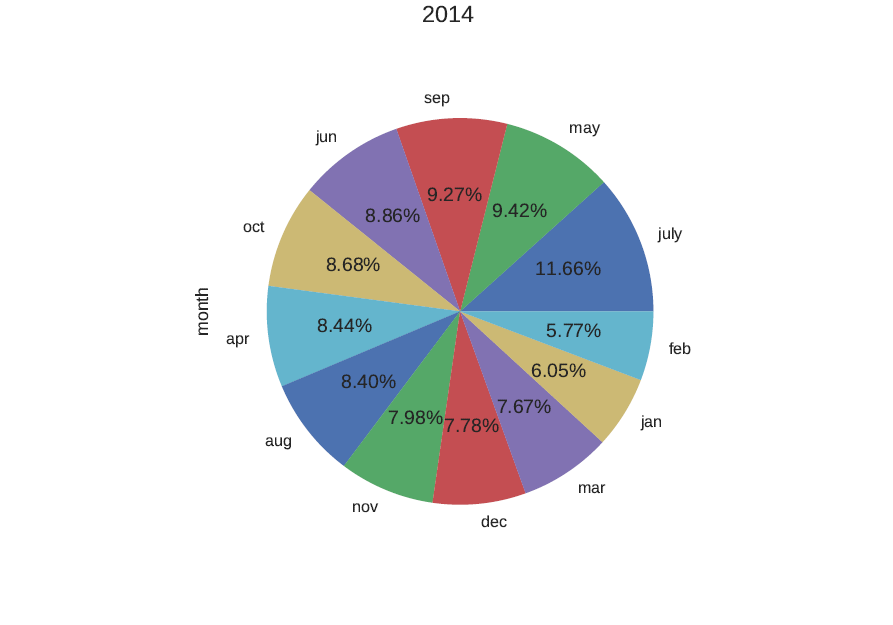
pdf.savefig()

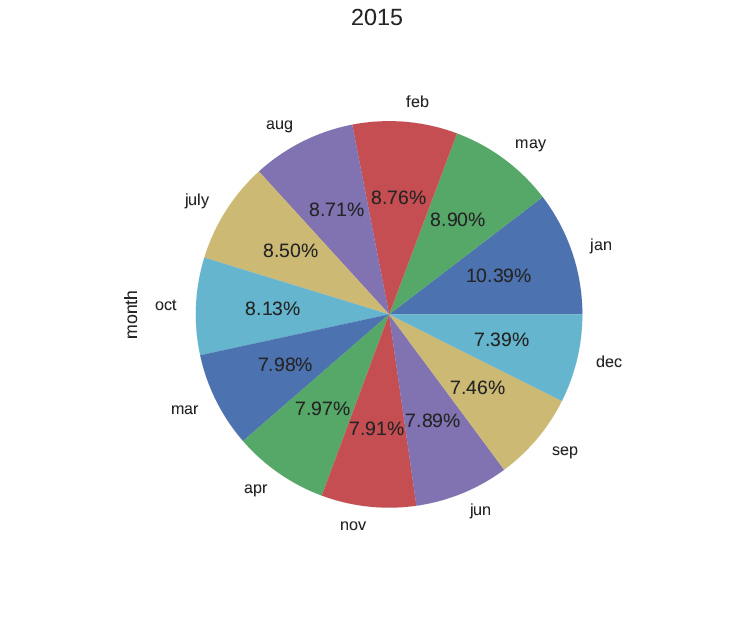
plt.close()

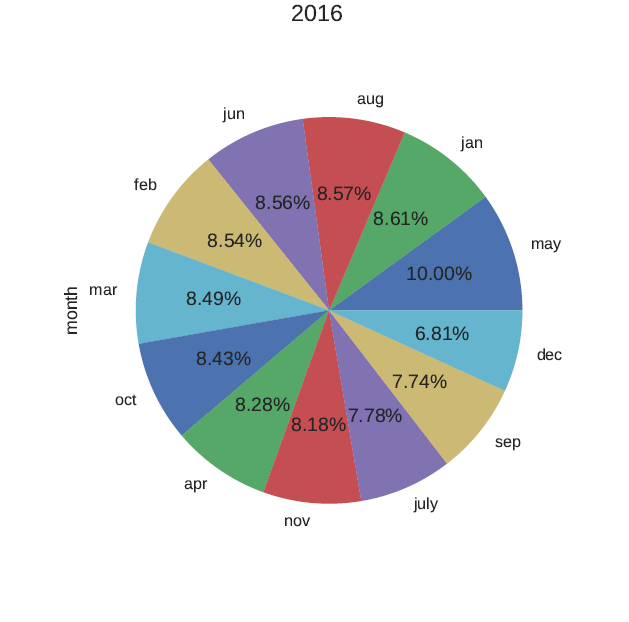
Output:











# Parameter 20: Attack probability of each country

tempdf=df.groupby(df.country)

res=tempdf.country.agg(np.count\_nonzero)

df2=res.to\_frame()

df2['prob']=np.nan

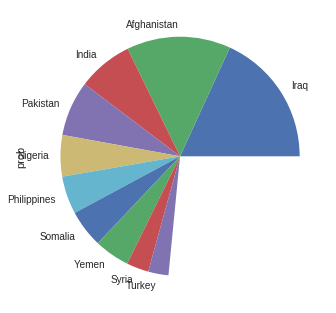
df2['prob']=df2['country']/43914

df2.sort\_values(by=['prob'],ascending=[False],inplace=True)

res=df2.prob.head(10)

res.plot(kind='pie',subplots=True)

Output:



# Parameter 21 : Military and religious attacks in years

tempdf=df.groupby(df.year)

df2=pd.DataFrame(columns=['military','religious'])

for name,group in tempdf:

res1=group[group.target0=='Military']

res2=group[group.target0=='Religious Figures/Institutions']

c1=res1.year.agg(['count'])

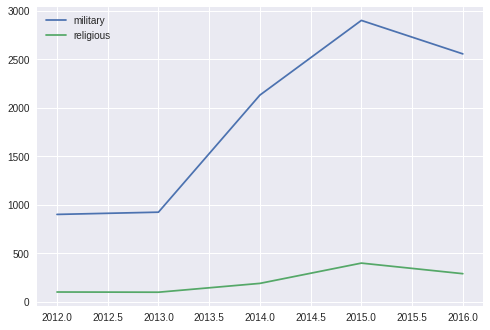
c2=res2.year.agg(['count'])

list1=[c1.loc['count'],c2.loc['count']]

df2.loc[name]=[c1.loc['count'],c2.loc['count']]

df2.plot()

Output:



# Parameter 22 : Distribution of global terrorism in the world map

from pygal\_maps\_world.maps import World

dfc=pd.read\_csv('gg.csv')

df1=df.copy()

d2=dfc.to\_dict()

d2=dfc.set\_index('country').T.to\_dict()

mm =World()

tempdf=df.groupby(df.country)

res=tempdf.country.agg(np.count\_nonzero)

d1=res.to\_dict()

dt={}

for key in d1:

if key in d2.keys():

dt[d2[key]['code']]=d1[key]

d1=dt

dn={}

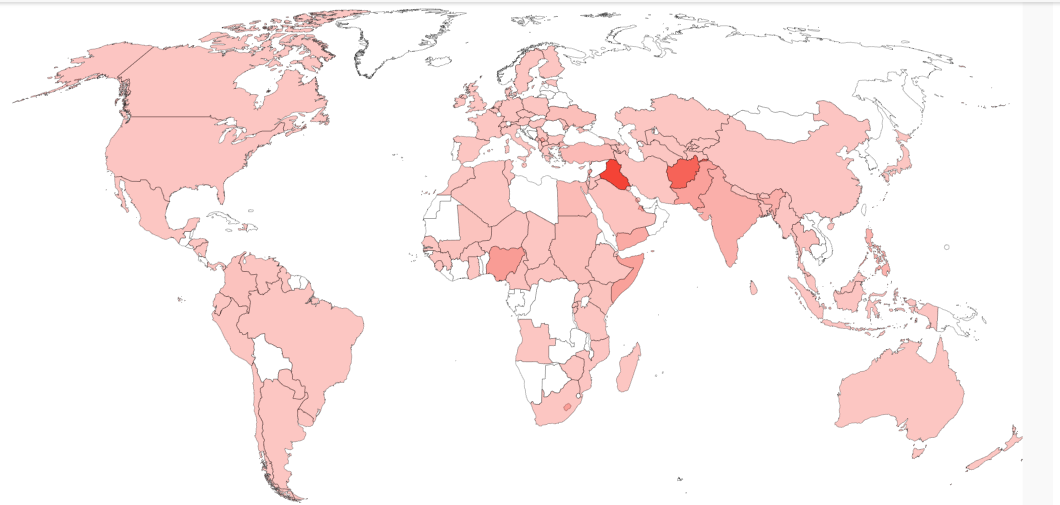
for key in d1:

r=key.rstrip()

dn[r]=d1[key]

mm.add('terrorism distribution in world',dn)

Output:



Additional objectives:

* Try to predict some data from the data model we created.
* Try to find a pattern how the known terrorist attacks are made and make a prediction.
* Try to make a prediction of attack of a given year by generating a pattern of attack from the given data.
* To see which countries will be at peace in future and free from terrorism .

**Challenges faced during the project:**

the dataset supplied was in a old encoding format. So we had to import the dataset in a proper encoding format.

* Second thing where we faced a challenge was when we tried to plot more than one line plots . there we learned how to close a plot and to avoid the unexpected result.
* Thirdly when we tried to show multiple graphs on console we found only one graph could be shown. We avoided the problem by writing the graphs in a pdf format.

HYPOTHESIS

**From all the data we analysed using Python we come to know about various data according to our parameters. They are as follows:**

1. In the count of severe attacks Afganistan is at top
2. Explosive/Bomb/Dynamites are majorly used in Attacks
3. Iraq has the maximum attack probability
4. Vehicle is the weapon type that was the most vurnerable
5. Each terrorist group has used different weapons but Explosive/Bomb/Dynamites are used widely

Also we estimated all the attack types and the damage occurred even we also calculated how many country were most vurnerable to all terrorist attack and also known the attack probability of the countries.

We also analyzed the decrease and increase of attack in the given countries also the distribution of casualities for each terrorism attacks.