1.INTRODUCTION

Living in 21st century and the word 'Terrorism' hasn't cross your mind is inevitable. We have either heard of these inhumanitarian act or have been victims soem or the other way. Inspite of terrorism being age-old activity, it is beacause of the significant advancements of technology in recent years the one has been able to access deeper insights about the terrorist activities around the globe, however remote be it. The shocking photos and videos captured in the aftermeth of any attack reveals about the damage done to man-kind.

The counter-part of the coin is that this feeling of insecurity and fear has boosted the technological advancements to help prevent it or reduce the after effects.

So why this dataset and data about terrorism? What are the audience that being catered through this analysis? From a Data Analysis perspective, these terrorist attacks can be viewed as data point with time (years) being the timescale. And this analysis can answer various questions in our constant struggle to stablize this world.

This Jupyter Notebook is a small approach to answer various questions or if not answer, atleast clear the blurry image related to terrorism.

1.1 Questions useful for performing EXPLORATORY ANALYSIS

The questions below will be of interest for various groups of people like governments, military, intelligence agencies and normal public.

```
1) How much has the number of terrorist attacks increased over the past 22 years?
2) How are these attacks spread over the globe?
3) What is the success and failure rates?
4) What types of weapons are most popular?
```

The final question is that since US has been able to restrict the amount of attacks after the 09/11, which served as the threshold of being reselient, can the other countries facing issue of terrorism replicate the US methods of counter-terrorism?

1.2 Predictive Questions

1) Can the number of casulties be predicted based on the characteristics of the attack i.e. weapon used, no of attackers, site attacked ?

2) Can we look and predict how tourism in a country will be affected because of terrorism?

2. DATA PREPARARTION

2.1 Importing Libraries

In [2]:
```
import numpy as np
import pandas as pd


# Plotting
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from seaborn import countplot

# Prediction
from sklearn.cross_validation import train_test_split
from sklearn.model_selection import cross_val_score
from sklearn.metrics import confusion_matrix
import itertools

# Classifiers
from sklearn import svm
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.dummy import DummyClassifier # Validation

# Data prepping
from sklearn.preprocessing import LabelEncoder

print("The Libraries have been imported.")
```

```
The Libraries have been imported.
```

2.2 Loading the CSV

In [16]:
```
data = pd.read_csv('G:/Rutgers 2018/subjects/Data Science/final project/Dataset.csv', encoding='ISO-8859-1')
print("Data loaded.")
```

```
C:\Users\HP\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:2785: DtypeWarning: Columns (4,6,33,6
1,62,63,76,79,94,96,114,115,121) have mixed types. Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)

Data loaded.
```

2.3 Data Selection and Description.

In [6]:
```
data_columns = [

        'eventid', # Unique ID alloted to a row,used in Plotting functions.

    # Time and Space where the Act of Terrorism committed is defined by the fields above.
        'iyear', 'imonth', 'iday', 'latitude', 'longitude',

    # Binary Variables (1 -> Yes or 0 -> no)
        'extended', # Did the duration of the Incident extend more than 24 hours?
        'vicinity', # Did the Incident occur in the immediate vicinity of the city? 0 is for IN city.
        'crit1', 'crit2', 'crit3', # The Incident meets the criterion (1, 2, 3), described in the introduction.
        'doubt', # Is there any doubt to whether the attack is an act of terrorism?
        'multiple', # Is this incident connected to other incident(s)?
        'success', # Has the attack reached its goal? Depends on type of attack.
        'suicide', # Did the perpetrator intend to escape alive?
        'claimed', # Was the attack claimed by any organised group?
        'property', # Is there any evidence of property damage from the incident?
        'ishostkid', # Were there victims taken as hostage or kidnapped?

    # Continuous Variables
        'nkill', # Amount of confirmed kills.
        'nwound', # Amount of confirmed wounded.

    # Categorical Variables
        'country_txt', # Name of country.
        'region', # Region id.
        'region_txt', # Name of region.
        'attacktype1_txt', # Describing the type of the attack I.e. Assassination, Bombing or Kidnapping.
        'targtype1_txt', # What type of target did the attack have? I.e. business, government or police.
        'natlty1_txt', # Nationality of the target.
        'weaptype1_txt', # What weapon was used?

    # Descriptive Variables
        'target1', # Description of specific target, if applicable.
        'gname', # Name of the organized group, if applicable.
        'summary', # Summary of the attack.

]
```

2.4 DATA CLEANING

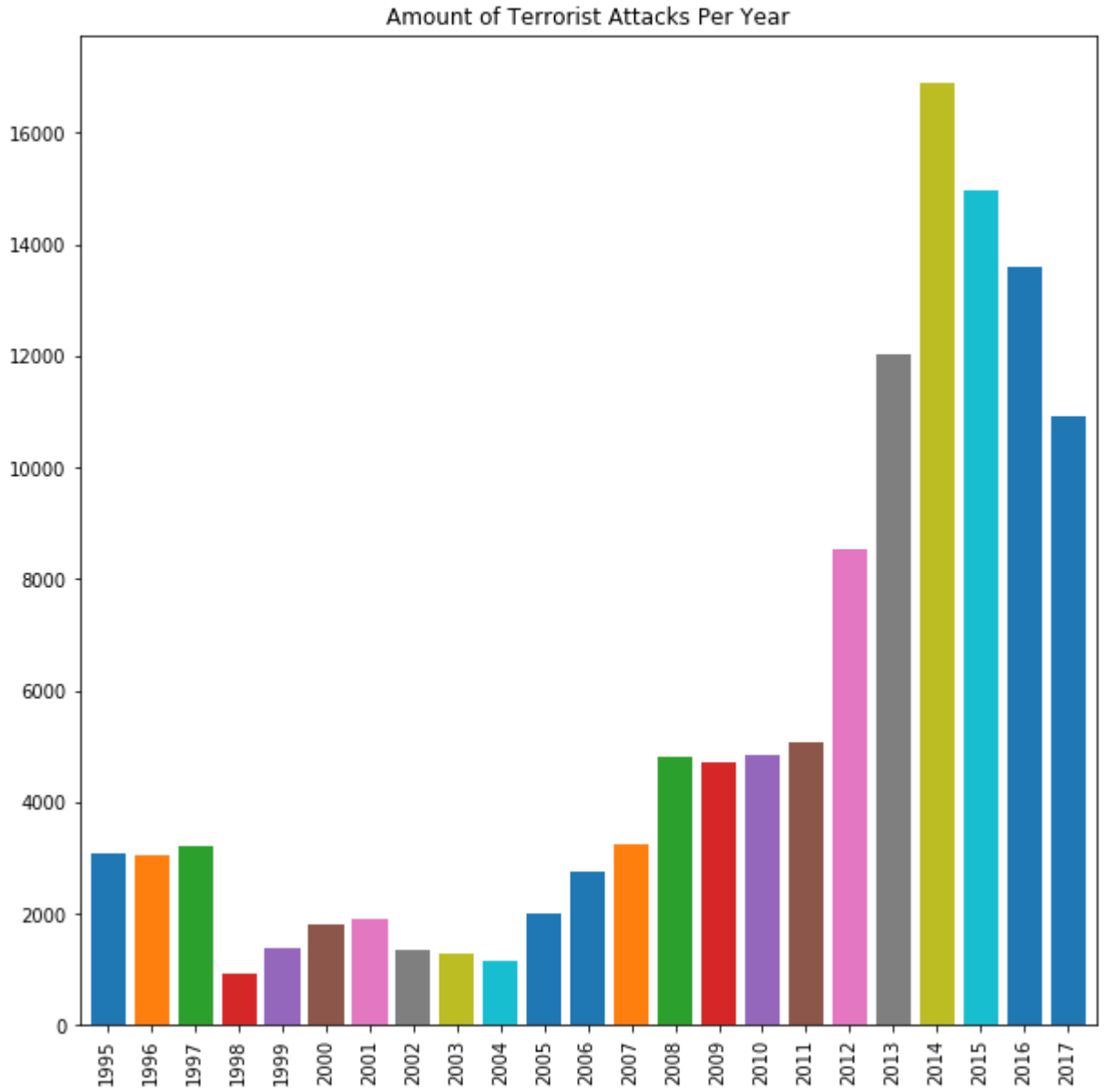
In [15]:
```
data = data.loc[:, data_columns] #Only keeps all the described columns.

# Random acts of violence and other outliers should not be part of the data.
# Thus, restrict the set the only attacks where the terrorism motive is certain.

#Restricts the set where terrorism motive is ceratin since all the random acts of violence and other
# outliers shouldnt be part of it.
data = data[(data.crit1 == 1) & (data.crit2 == 1) & (data.crit3 == 1) & (data.doubt == 0)]

#Here the column for vehicle property has a very long name ie. Weapontype so we are shortening it.
data.weaptype1_txt.replace(
    'Vehicle (not to include vehicle-borne explosives, i.e., car or truck bombs)',
    'Vehicle', inplace = True)

# We are replacing unknowing values such as -9 with 0. -9 values are more likely to be false than true.
data.iloc[:,[6, 15, 16, 17]] = data.iloc[:,[6, 15, 16, 17]].replace(-9,0)


# Claimed Category has few values '2' (here is should be 0 or 1),now ASSUMING it as input mistakes we are setting
data.claimed.replace(2,1, inplace = True)


# Converting everything to lowercase and ensuring consistent values.
data.target1 = data.target1.str.lower()
data.gname = data.gname.str.lower()
data.summary = data.summary.str.lower()
data.target1 = data.target1.fillna('unknown').replace('unk','unknown')

# Some nwound and nkill are NaN so we are replacing them with median.
data.nkill = np.round(data.nkill.fillna(data.nkill.median())).astype(int)
data.nwound = np.round(data.nwound.fillna(data.nwound.median())).astype(int)



#Since database only reports victim as nkill and nwound so combining them into ncasualties column
#We are also adding has_casualties column.
data['ncasualties'] = data['nkill'] + data['nwound']
data['has_casualties'] = data['ncasualties'].apply(lambda x: 0 if x == 0 else 1)

print("The data has been cleaned and prepared.")
```

```
The data has been cleaned and prepared.
```

3. DATA EXPLORATION

3.1 Is there any increase in the amount of terrorist attacks during the recent years?

In [18]:
```
barplot = pd.value_counts(data['iyear'])\
.sort_index().plot.bar(width=0.8, figsize=(10, 10), title="Amount of Terrorist Attacks Per Year")
```



By looking at the graph, we have come to the shocking conclusion that the Amount of Terrorist attacks have been drastically increasing

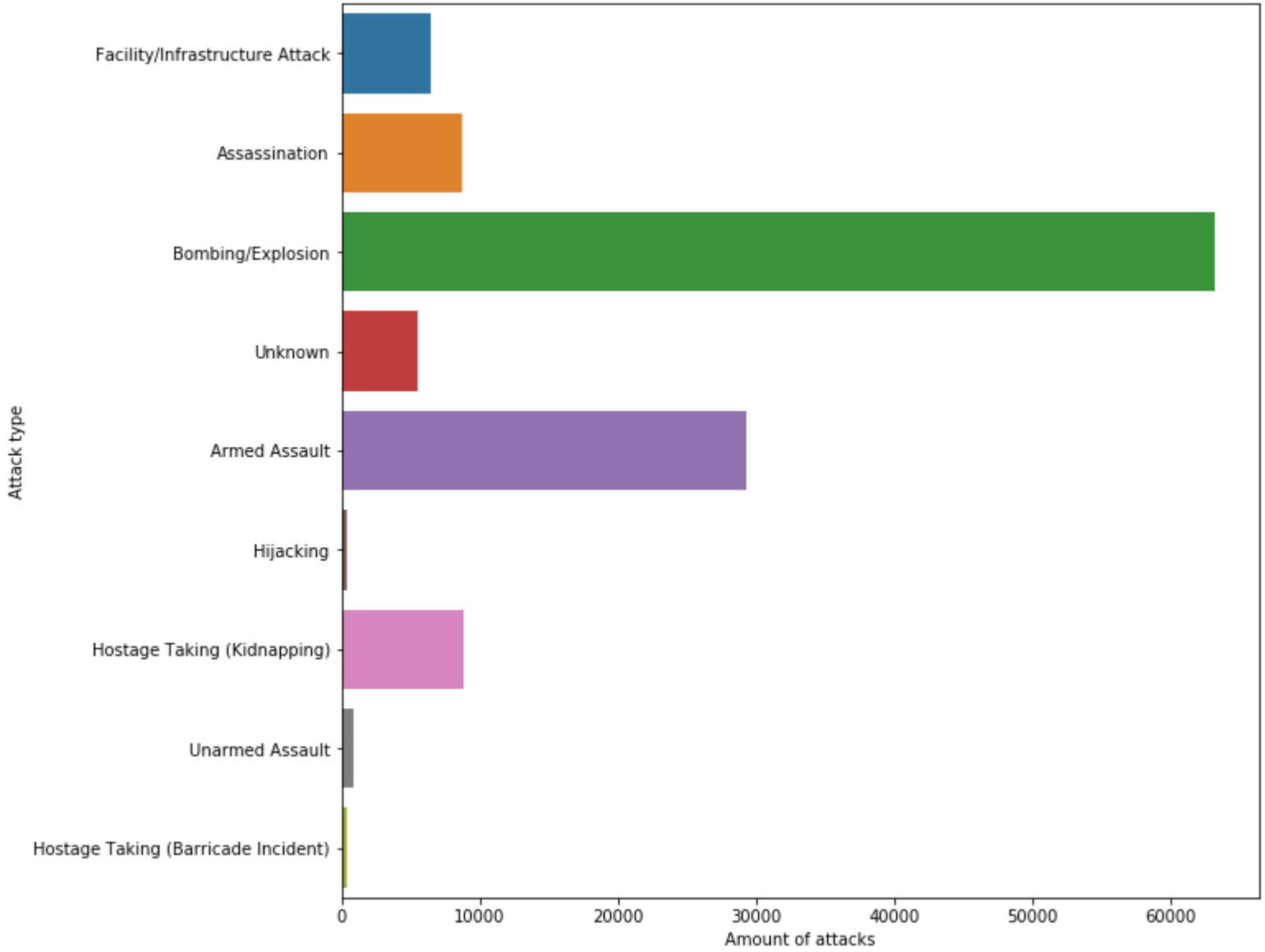
since the past five years. However, it is important to take into account the effictiveness of data collection since 2012. This can be interpreted with relating it with the increase in the media coverage over the last decade. The advancements in technology sector has made the world more connected and even the attacks in remote locations can be reported or covered.

The counter-part of the argument can be to debate upon the advancements in technology that has led to more evolved weaponary and thus more successful attacks. The question that arises from the above question is the comparison of the rate of development of advanced weapons and attack technique versus the rate of the development of defence mechanism for those advanced weapons and technique.

3.2 Which type of attacks are popular?

In [36]:
```
plt.figure(figsize=(10,10))
attack = sns.countplot(y="attacktype1_txt", data=data)
attack.set_xlabel("Amount of attacks")
attack.set_ylabel("Attack type")
```

```
Bombing/Explosion                      63188
Armed Assault                          29252
Hostage Taking (Kidnapping)             8781
Assassination                           8659
Facility/Infrastructure Attack          6494
Unknown                                 5538
Unarmed Assault                          816
Hostage Taking (Barricade Incident)      400
Hijacking                                392
Name: attacktype1_txt, dtype: int64 ['Facility/Infrastructure Attack' 'Assassination' 'Bombing/Explosion'
 'Unknown' 'Armed Assault' 'Hijacking' 'Hostage Taking (Kidnapping)'
 'Unarmed Assault' 'Hostage Taking (Barricade Incident)']
```



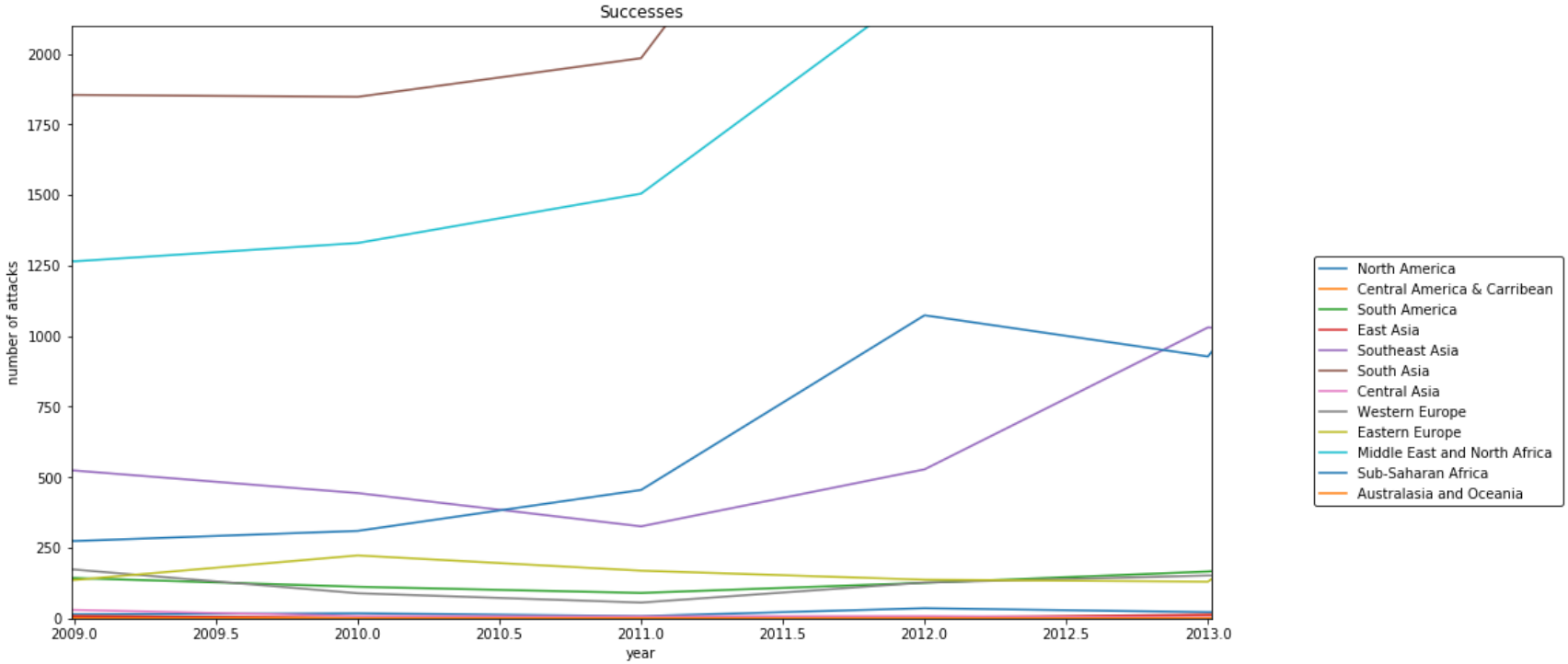
What are the number of attacks per region?

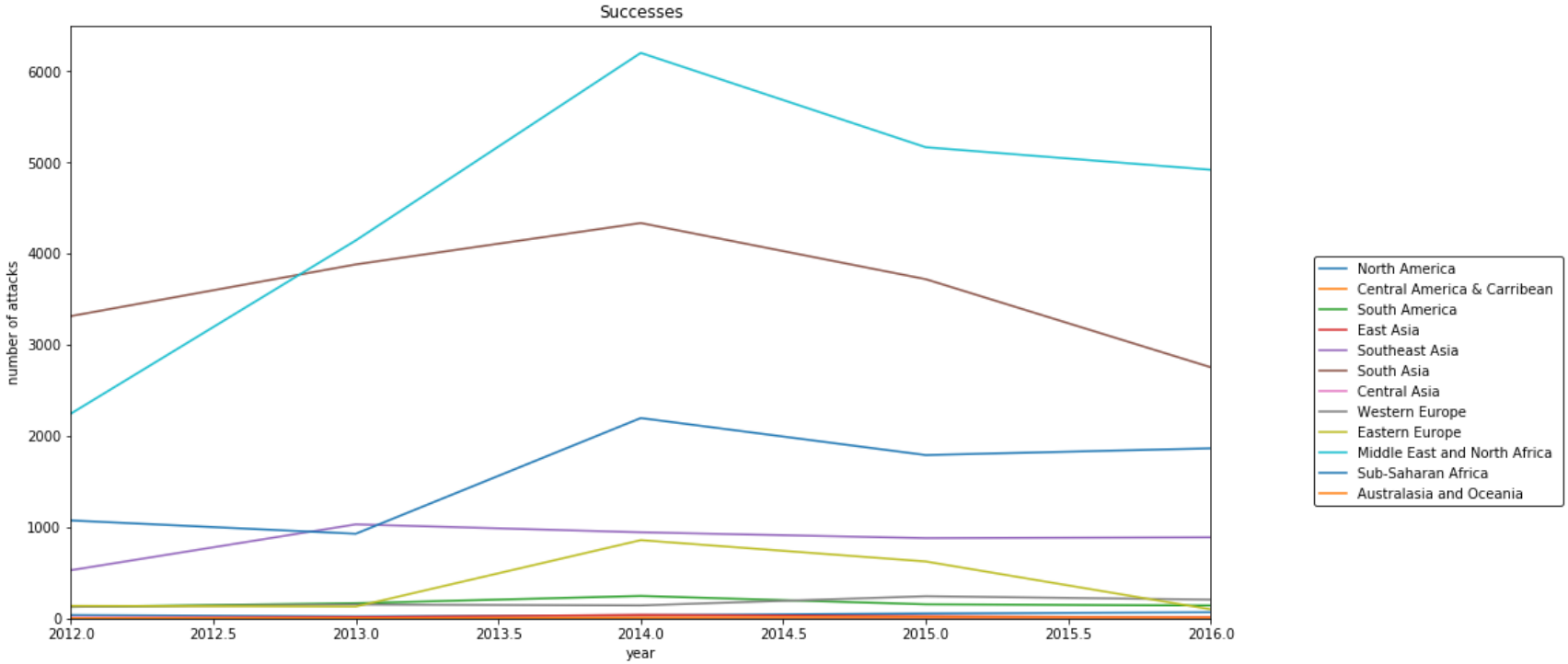
Thsi seems to give us a fair idea of how the amount of terrorist activity is spread across the globe. We will divide the world into 12 major region and then try and plot the number of successful attacks and failed attacks.

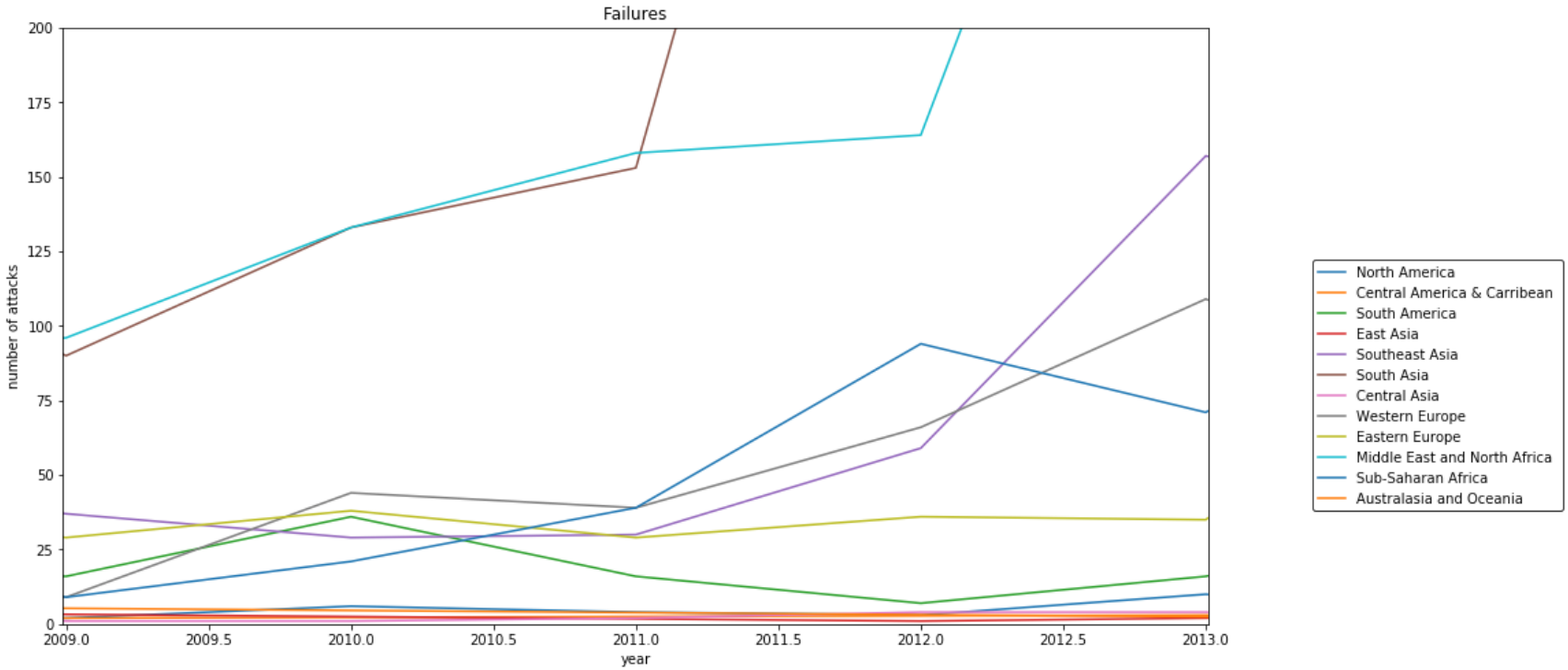
In [46]:
```
region_dictionary = {1: 'North America', 2: 'Central America & Carribean', 3: 'South America',
                     4: 'East Asia', 5: 'Southeast Asia', 6: 'South Asia', 7: 'Central Asia',
                     8: 'Western Europe', 9: 'Eastern Europe', 10: 'Middle East and North Africa',
                     11: 'Sub-Saharan Africa', 12: 'Australasia and Oceania'}

def multi_graph(result,result_list, xmin, xmax, ymin, ymax):
    fig2, ax2 = plt.subplots(figsize = (15,8))
    number = 1 #the for-loop in append_list processes the regions in order from 1 to 12
    for j in result_list:
        ax2.plot(j.index, j.eventid, label = '%s ' % region_dictionary[number] )
        number += 1

    plt.xlim([xmin,xmax])
    plt.ylim([ymin,ymax])
    plt.xlabel('year')
    plt.ylabel('number of attacks')
    plt.title(result)
    ax2.legend(loc = 'center', frameon = True, edgecolor = 'black',bbox_to_anchor =(1.2,0.4))


success_list = []
failure_list = []

for i in region_dictionary:
    region_data = data[(data.region == i)]
    region_data_success = region_data[(region_data.success == 1)]
    region_data_failure = region_data[(region_data.success == 0)]
    region_grouped_success = region_data_success.groupby('iyear').count()
    region_grouped_failure = region_data_failure.groupby('iyear').count()


    success_list.append(region_grouped_success)
    failure_list.append(region_grouped_failure)

multi_graph('Successes',success_list, 2011, 2011, 0, 2100)
multi_graph('Successes',success_list, 2012, 2016, 0, 6500)
multi_graph('Failures',failure_list, 2011, 2011, 0, 200)
multi_graph('Failures',failure_list, 2012, 2016, 0, 1300)
```

```
C:\Users\HP\Anaconda3\lib\site-packages\matplotlib\axes\_base.py:3124: UserWarning: Attempting to set identical
left==right results
in singular transformations; automatically expanding.
left=2011, right=2011
```

```
  'left=%s, right=%s') % (left, right))
C:\Users\HP\Anaconda3\lib\site-packages\matplotlib\axes\_base.py:3124: UserWarning: Attempting to set identical
left==right results
in singular transformations; automatically expanding.
left=2011, right=2011
  'left=%s, right=%s') % (left, right))
```




Successes
number of attacks
year
0
1000
2000
3000
4000
5000
6000
2012.0
2012.5
2013.0
2013.5
2014.0
2014.5
2015.0
2015.5
2016.0
North America
Central America & Carribean
South America
East Asia
Southeast Asia
South Asia
Central Asia
Western Europe
Eastern Europe
Middle East and North Africa
Sub-Saharan Africa
Australasia and Oceania


Failures
number of attacks
year
0
25
50
75
100
125
150
175
200
2009.0
2009.5
2010.0
2010.5
2011.0
2011.5
2012.0
2012.5
2013.0
North America
Central America & Carribean
South America
East Asia
Southeast Asia
South Asia
Central Asia
Western Europe
Eastern Europe
Middle East and North Africa
Sub-Saharan Africa
Australasia and Oceania

Failures

4. PREDICTIVE ANALYSIS

In [ ]:

Work in Progress!

The exploratory data analysis has been done and we are yet molding our question on what to predict and what methods will be appropriate to predict the data to answer the question.

In [ ]:

In [ ]:

4.1 Can the success of an attack be predicted by cross-referencing weather conditions?

In [ ]: