# MT18052_Assignment 1

September 3, 2019

In [4]:
```
import pandas as pd
import numpy as np
from sklearn.utils import shuffle
import numpy as np
from wordcloud import WordCloud as wcl
from wordcloud import STOPWORDS
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

## 1 Tweets

In [80]:
```
import sys
print(sys.executable)
```

```
/home/ashish/anaconda3/bin/python
```

In [81]:
```
df = pd.read_csv(r"1377884570_tweet_global_warming.csv", encoding ="latin-1")
```

In [82]:
```
df.tail()
```

Out[82]:
```
                                                  tweet existence confidence  \
6085  @bloodless_coup "The phrase 'global warming' s...         Y          1
6086  Virginia to Investigate Global Warming Scienti...       NaN          1
6087  Global warming you tube parody you will enjoy ...         N     0.6411
6088  One-Eyed Golfer: Don't dare tell me about glob...         N          1
6089  man made global warming a hair brained theory ...         N          1

     other1 other2  other3
6085    NaN    NaN     NaN
6086    NaN    NaN     NaN
6087    NaN    NaN     NaN
```

```
6088     NaN     NaN      NaN
6089     NaN     NaN      NaN
```

In [83]:
```
df=df.replace(to_replace ="Y", value ="Yes")
```

In [84]:
```
df=df.replace(to_replace ="N", value ="No")
```

In [85]:
```
y1 = df[df.existence == 'Yes']
```

In [86]:
```
n1 = df[df.existence =='No']
```

In [87]:
```
n2 = df[df.existence =='NA']
```

In [88]:
```
y1.tail()
```

Out[88]:
```
                                                  tweet existence confidence  \
6080  Bats, Birds and Lizards Can Fight Climate Chan...       Yes     0.6751
6081  Bats, Birds and Lizards Can Fight Climate Chan...       Yes          1
6082  Global warming: The fossil fuel dilemma: Ameri...       Yes          1
6084  It's 83ï£¡_ï£¡ and climbing in NYC. August wea...       Yes          1
6085  @bloodless_coup "The phrase 'global warming' s...       Yes          1

     other1 other2  other3
6080    NaN     NaN      NaN
6081    NaN     NaN      NaN
6082    NaN     NaN      NaN
6084    NaN     NaN      NaN
6085    NaN     NaN      NaN
```

In [89]:
```
d1=pd.concat([y1,n1])
df=pd.concat([d1,n2])
```

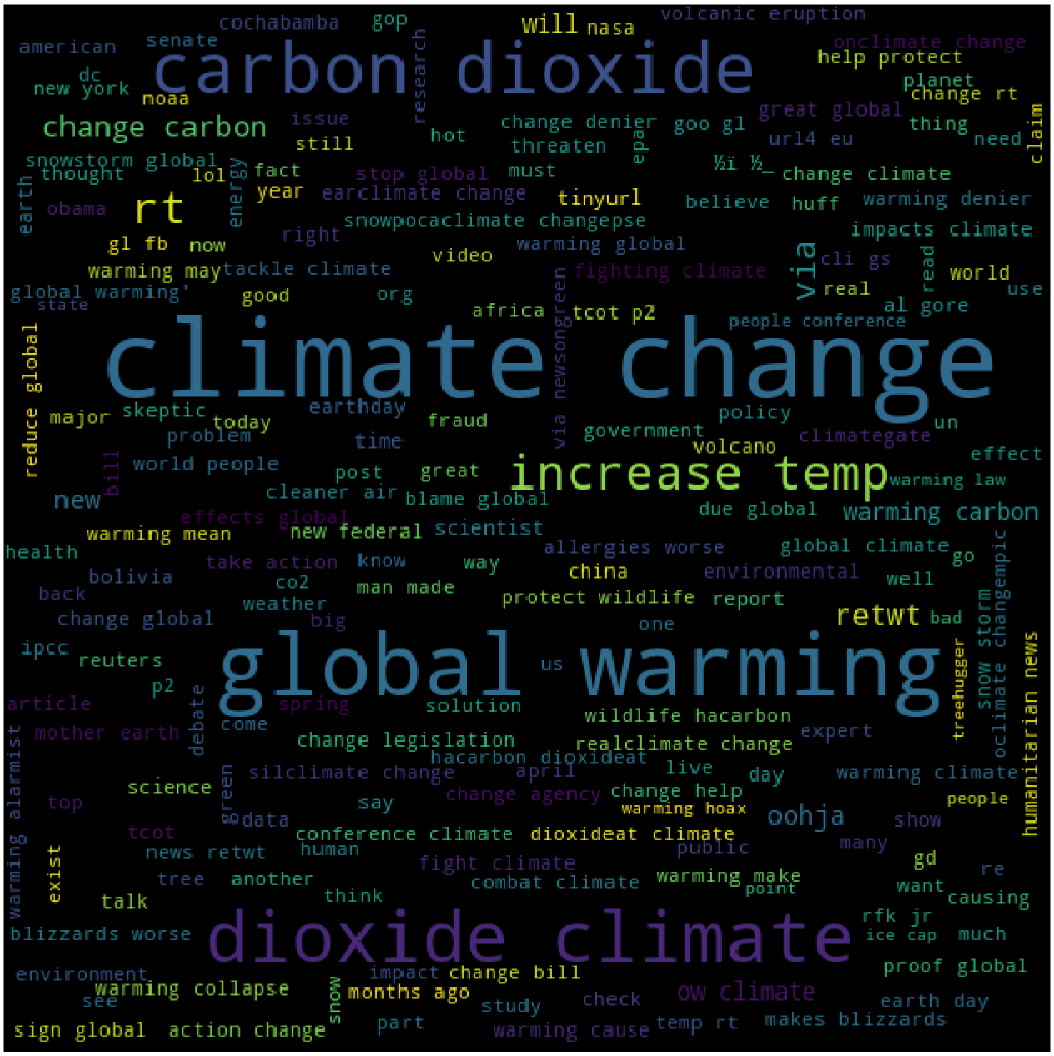
In [90]:
```
def word_cloud(words):
    stopwords = set(STOPWORDS)
    wordcloud = wcl(max_words=204,width = 603, height = 603,background_color ='black'
                min_font_size = 10).generate(words)

    plt.figure(figsize = (9, 9), facecolor = None)
    plt.imshow(wordcloud)
    plt.axis("off")
    plt.tight_layout(pad = 0)
    plt.show()
```

In [91]:
```
def extract_words(tweets):
    distinct_words = ' '
    for tweet in tweets:
        tweet = str(tweet)
        tokens = tweet.split()
        for i in range(len(tokens)):
```

```
                tokens[i] = tokens[i].lower()
            for words in tokens:
                distinct_words = distinct_words + words + ' '
        return distinct_words
```

```
In [92]: unique_wrds = extract_words(df.tweet)
```

```
In [53]: word_cloud(unique_wrds)
```




From Tweets we can infer that increase in carbon dioxide leads to increase in temperature which is responsible for awareness among people about global warming.

## 2 Co2 Emmision

```
In [5]: co2 = pd.read_csv("./data/CO2 emissions per capita per country.csv")
```

Unit metric

```
In [6]: co2.head()
```

```
Out[6]:   Country Name Country Code      1960      1961      1962      1963      1964  \
       0         Aruba          ABW       NaN       NaN       NaN       NaN       NaN
       1   Afghanistan          AFG  0.046060  0.053604  0.073765  0.074233  0.086292
       2        Angola          AGO  0.097472  0.079038  0.201289  0.192535  0.201003
       3       Albania          ALB  1.258195  1.374186  1.439956  1.181681  1.111742
       4       Andorra          AND       NaN       NaN       NaN       NaN       NaN

              1965      1966      1967  ...       2009       2010       2011  \
       0       NaN       NaN       NaN  ...  25.915833  24.670529  24.505835
       1  0.101467  0.107637  0.123734  ...   0.241723   0.293837   0.412017
       2  0.191528  0.246413  0.154912  ...   1.232495   1.243406   1.252789
       3  1.166099  1.333055  1.363746  ...   1.495600   1.578574   1.803715
       4       NaN       NaN       NaN  ...   6.121652   6.122595   5.867130

               2012      2013      2014  2015  2016  2017  2018
       0  13.155542  8.351294  8.408363   NaN   NaN   NaN   NaN
       1   0.350371  0.315602  0.299445   NaN   NaN   NaN   NaN
       2   1.330843  1.254617  1.291328   NaN   NaN   NaN   NaN
       3   1.692908  1.749211  1.978763   NaN   NaN   NaN   NaN
       4   5.916597  5.900753  5.832170   NaN   NaN   NaN   NaN

       [5 rows x 61 columns]
```

```
In [64]: co2 = co2.drop(['2015','2016','2017','2018'],axis=1)
```

```
In [65]: y = co2.columns[12:]
```

```
In [66]: y = y.to_list()
```

```
In [190]: b = []
          for i in y:
              b.append(int(i))
```

```
In [191]: values = []
          for key in y:
              val = np.sum(co2[key])
              values.append(val)
```

```
In [192]: print (len(values))
```

```
45
```

In [193]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [194]:
```
def plot_CO2_emission(feature,years,col = 'orange'):
    sns.regplot(x=years, y=feature,color=col, lowess=True,label="Co2 emission")
    plt.title("CO2 Emissions: 1970-2013 ", size=18)
    plt.ylabel("CO2 emission ", size=15)
    plt.xlabel("Year", size=15)
    plt.xticks(size=12)
    plt.yticks(size=12)
    plt.legend()
  # plt.show()
```

In [195]:
```
India = co2.iloc[107,]
india_values = []
for key in y:
    india_values.append(India[key])
```

In [218]:
```
ireland = co2.iloc[109,]
ireland_values = []
for key in y:
    ireland_values.append(ireland[key])
```

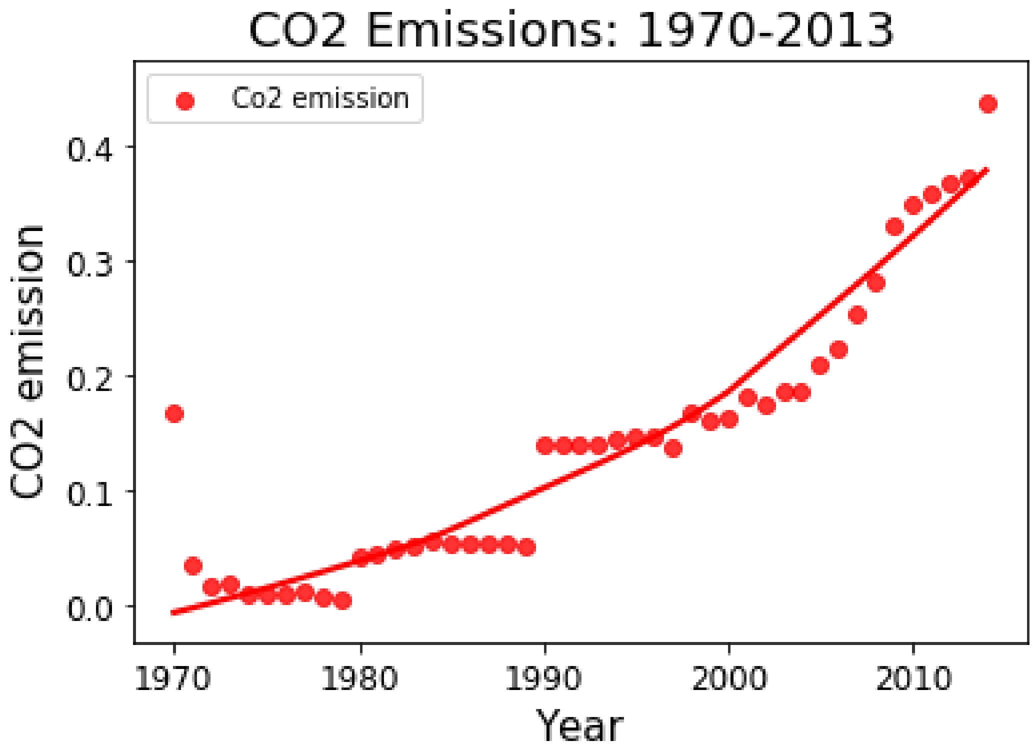
In [228]:
```
cambodia = co2.iloc[121,]
cambodia_values = []
for key in y:
    cambodia_values.append(cambodia[key])
```

In [214]:
```
len(india_values)
```

Out[214]: 45

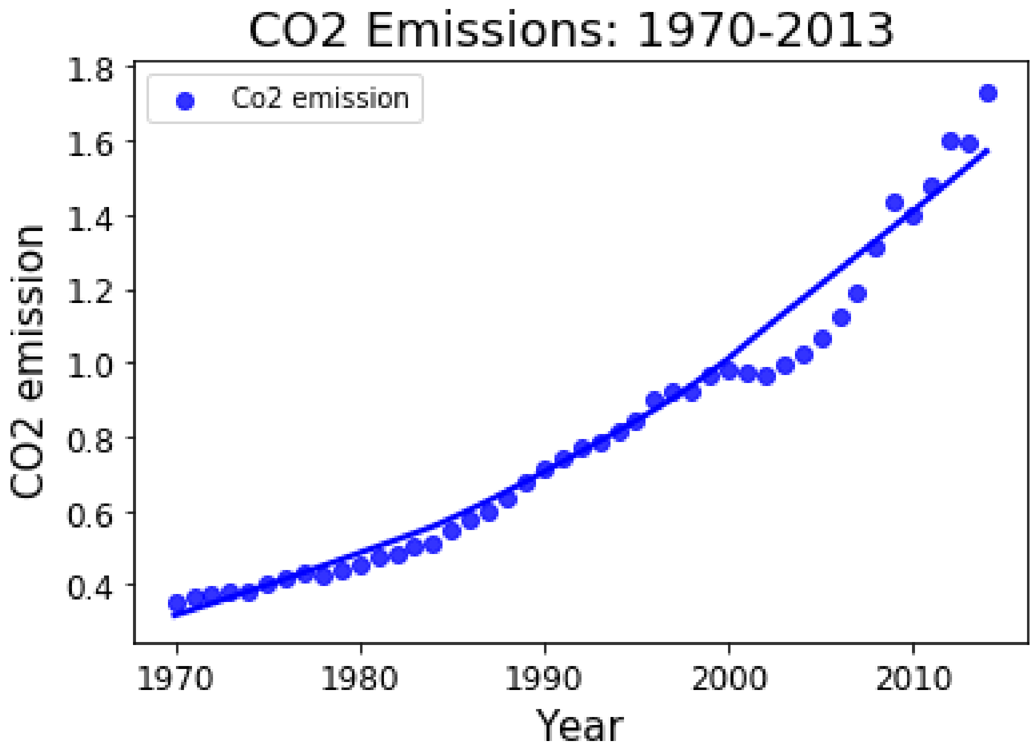
In [215]:
```
india_values = list(india_values)
```

In [230]:
```
# india_values
plot_CO2_emission(feature=np.array(cambodia_values),years=np.array(b),col = 'red')
```



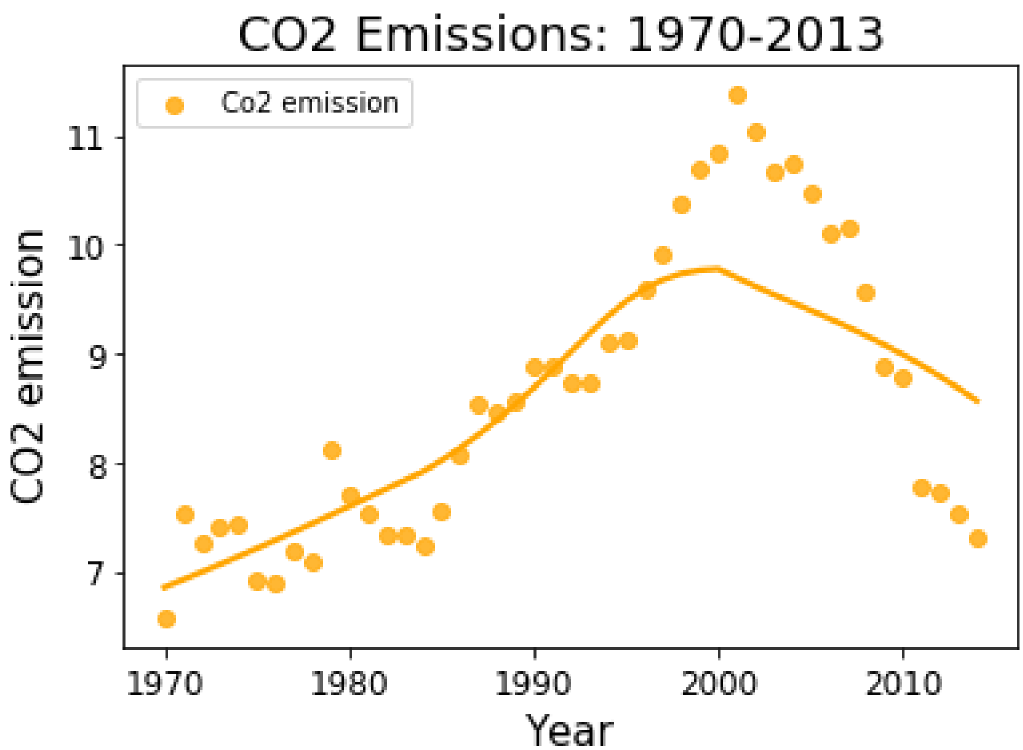

Emmision increases and then gets stable for few years and then again on rise

```
In [226]: plot_CO2_emission(feature=np.array(india_values),years=np.array(b),col = 'blue')
```



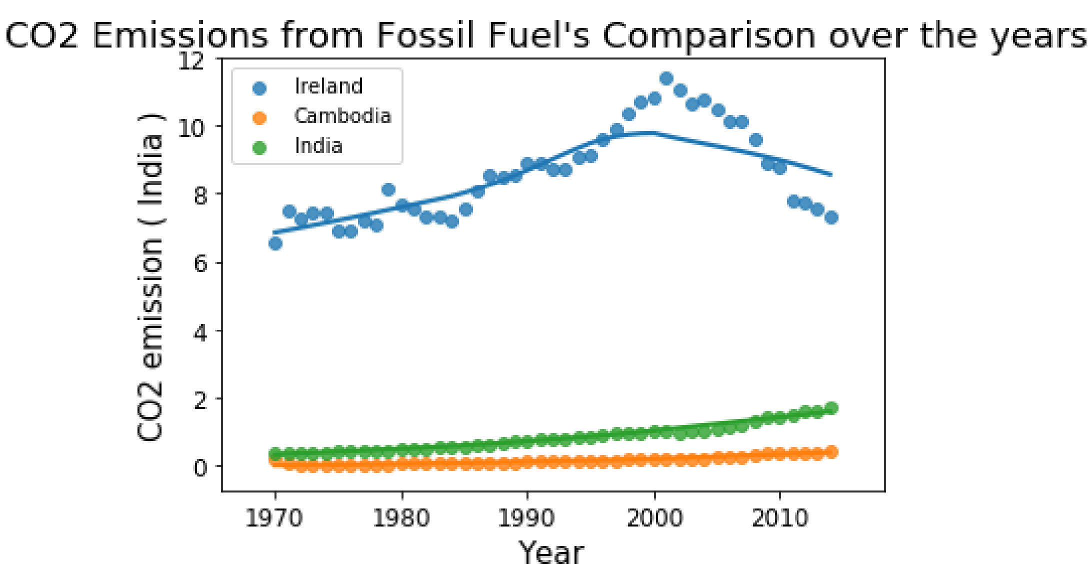

Trend in Increase in Emission level in India

```
In [227]: plot_CO2_emission(feature=np.array(ireland_values),years=np.array(b),col = 'orange')
```




There has been decrease in Co2 emmision level in Ireland

```
In [231]: plot_CO2_emission_all(np.array(b),ireland_values,'Ireland')
          plot_CO2_emission_all(np.array(b),cambodia_values,'Cambodia')
          plot_CO2_emission_all(np.array(b),india_values,'India')
```



Emmision of Co2 by different countries in years india emission is low as comparison to ireland but it is increasing whereas in ireland it is decreasing

```
In [167]: countries = np.array(co2.iloc[:,1])
```

```
In [168]: len(countries)
```

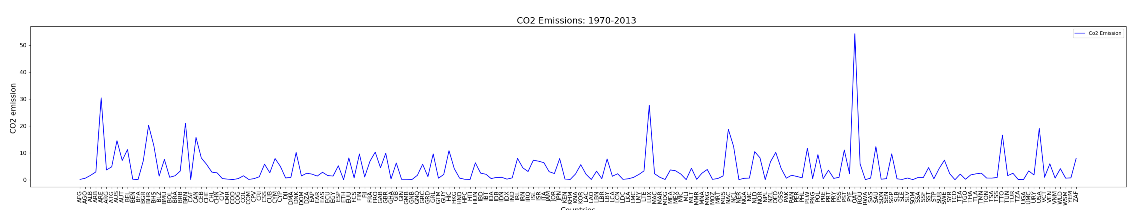
```
Out[168]: 264
```

```
In [169]: countries_emmision = []
          for i in range(len(countries)):
              tmp = np.array(co2.iloc[i,2:])
              countries_emmision.append(tmp.mean())
```

```
In [170]: def plot_CO2_emission_2(feature,countries_,col='blue'):
              plt.figure(num=None, figsize=(40,6),dpi = 100, facecolor='w', edgecolor='k')
          #     sns.regplot(x=countries_, y=feature,color='orange', lowess=True,label="Co2 emi
              plt.plot(countries_,feature,color = col)
          #     plt.annotate(feature)
              plt.title("CO2 Emissions: 1970-2013 ", size=18)
              plt.ylabel("CO2 emission ", size=15)
              plt.xlabel("Countries", size=15)
              plt.xticks(size=12, rotation='vertical')
              plt.yticks(size=12)
              plt.legend(['Co2 Emission'])
```

```
In [171]: countries_clear_2 = []
          countries_emmision_clear_2 = []
          for i in range(len(countries_emmision)):
              if(np.isnan(countries_emmision[i]) == False):
                  countries_clear_2.append(countries[i])
                  countries_emmision_clear_2.append(countries_emmision[i])
```

```
In [172]: plot_CO2_emission_2(feature=np.array(countries_emmision_clear_2),countries_=np.array
```




Co2 Emmision sum of QATAR(QAT) is maximum among all others this can be inferred by the graph plotted above

```
In [108]: data = np.array(countries_emmision_clear_2)
          print ("Observations of CO2 Emmision among different Countries")
          print ("Mean of CO2 " + str(data.mean()))
          print ("Variance of CO2 " + str(data.var()))
          print ("Standard Deviation of CO2 " + str(data.std()))
```

```
Observations of CO2 Emmision among different Countries
Mean of CO2 4.135514541776172
Variance of CO2 38.58275043767678
Standard Deviation of CO2 6.211501464032411
```

## 3 Fossil Fuels

Unit million metric tons of carbon

```
In [115]: df = pd.read_csv('Fossil-fuel co2 Global Estimates (1751-2013).csv')
```

```
In [116]: df.tail()
```

Out[116]:

| | Year | Total | Gas | Liquids | Solids | Production | Flaring | Capita |
|---|---|---|---|---|---|---|---|---|
| 258 | 2009 | 8641 | 1580 | 3065 | 3517 | 415 | 64 | 1.26 |
| 259 | 2010 | 9137 | 1700 | 3129 | 3795 | 448 | 66 | 1.32 |
| 260 | 2011 | 9508 | 1762 | 3158 | 4027 | 496 | 64 | 1.36 |
| 261 | 2012 | 9671 | 1787 | 3214 | 4086 | 520 | 65 | 1.36 |
| 262 | 2013 | 9776 | 1806 | 3216 | 4131 | 554 | 68 | 1.36 |

Data snapshot of dataset used

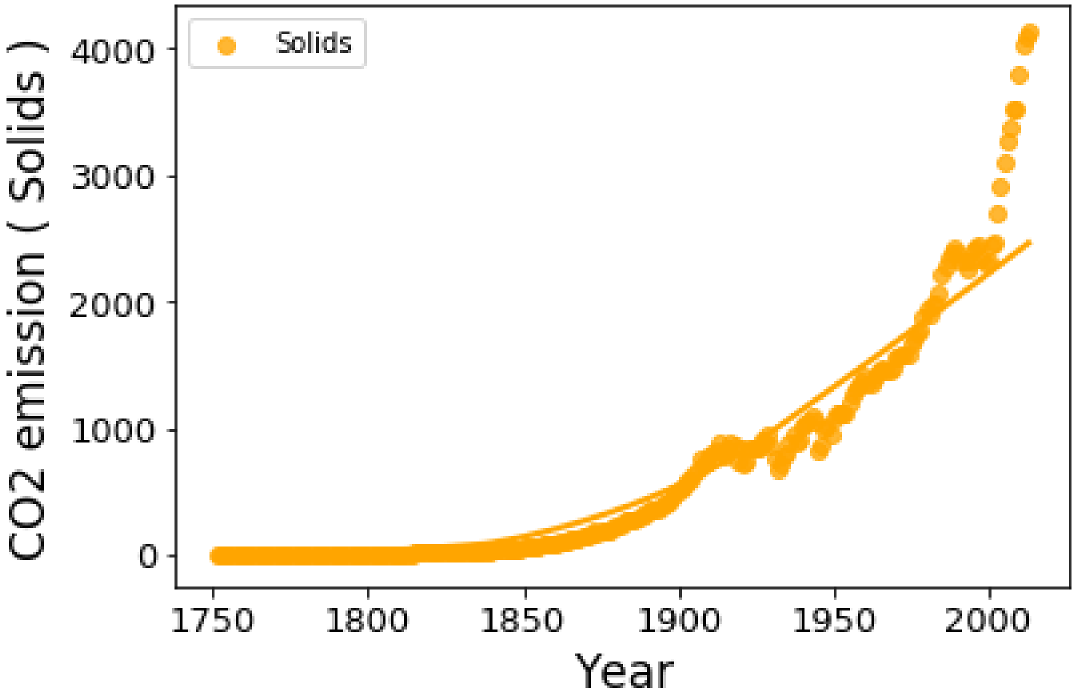
```
In [159]: def plot_CO2_emission(feature1,feature2,name,col='blue'):
              sns.regplot(x=feature1, y=feature2,color=col ,lowess=True,label=name)
          #     plt.title("CO2 Emissions from Fossil Fuel's via Cement Manufacture, and Gas Fl
              plt.ylabel("CO2 emission ( "+name+" )", size=17)
              plt.xlabel("Year", size=17)
              plt.xticks(size=13)
              plt.yticks(size=13)
              print ("Mean of CO2 " + str(feature2.mean()))
              print ("Variance of CO2 " + str(feature2.var()))
              print ("Standard Deviation of CO2 " + str(feature2.std()))
              plt.legend()
              plt.show()
```

```
In [160]: def plot_CO2_emission_all(feature1,feature2,name):
              sns.regplot(x=feature1, y=feature2,lowess=True,label=name)
              plt.title("CO2 Emissions from Fossil Fuel's Comparison over the years ", size=18
              plt.ylabel("CO2 emission ( "+name+" )", size=15)
              plt.xlabel("Year", size=15)
              plt.xticks(size=12)
```

```
plt.yticks(size=12)
plt.legend()
```

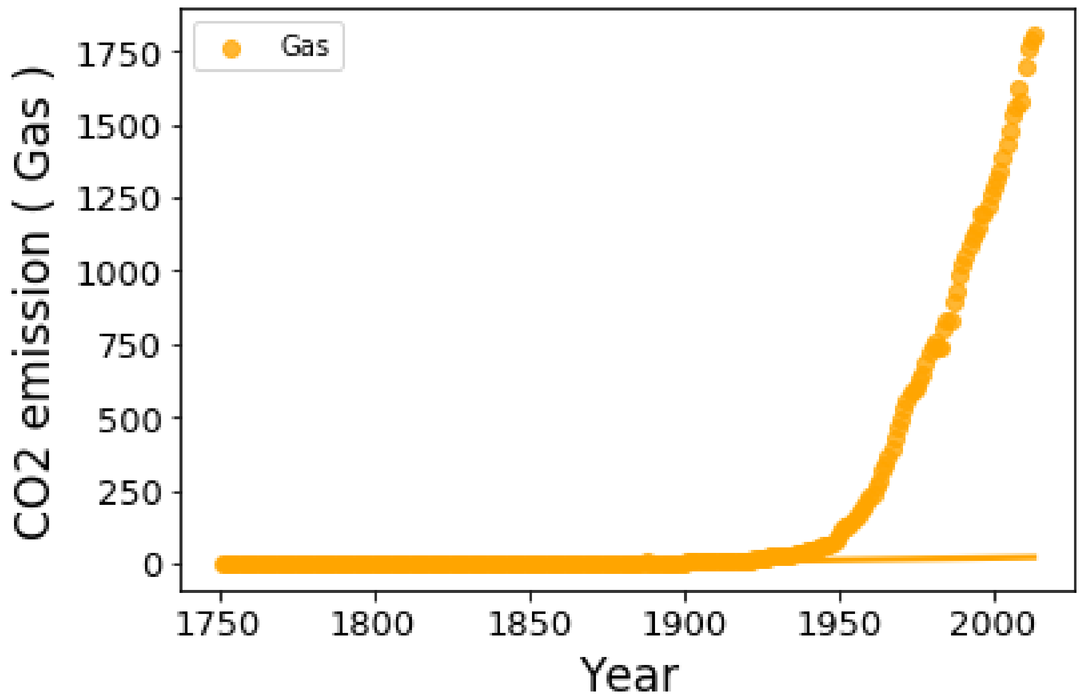
```
In [161]: plot_CO2_emission(df.Year,df.Solids,'Solids','orange')
```

```
Mean of CO2 708.4600760456274
Variance of CO2 856818.7302702229
Standard Deviation of CO2 925.6450347029486
```



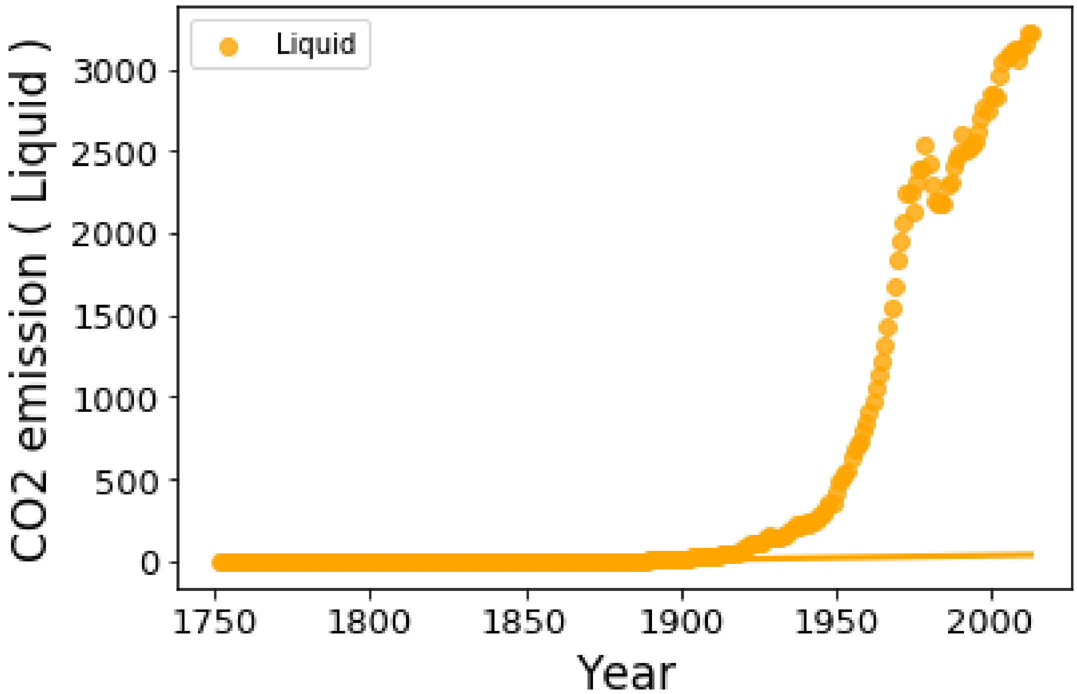

```
In [162]: plot_CO2_emission(df.Year,df.Gas,'Gas','orange')
```

```
Mean of CO2 202.92015209125475
Variance of CO2 183397.89818012947
Standard Deviation of CO2 428.2498081495536
```



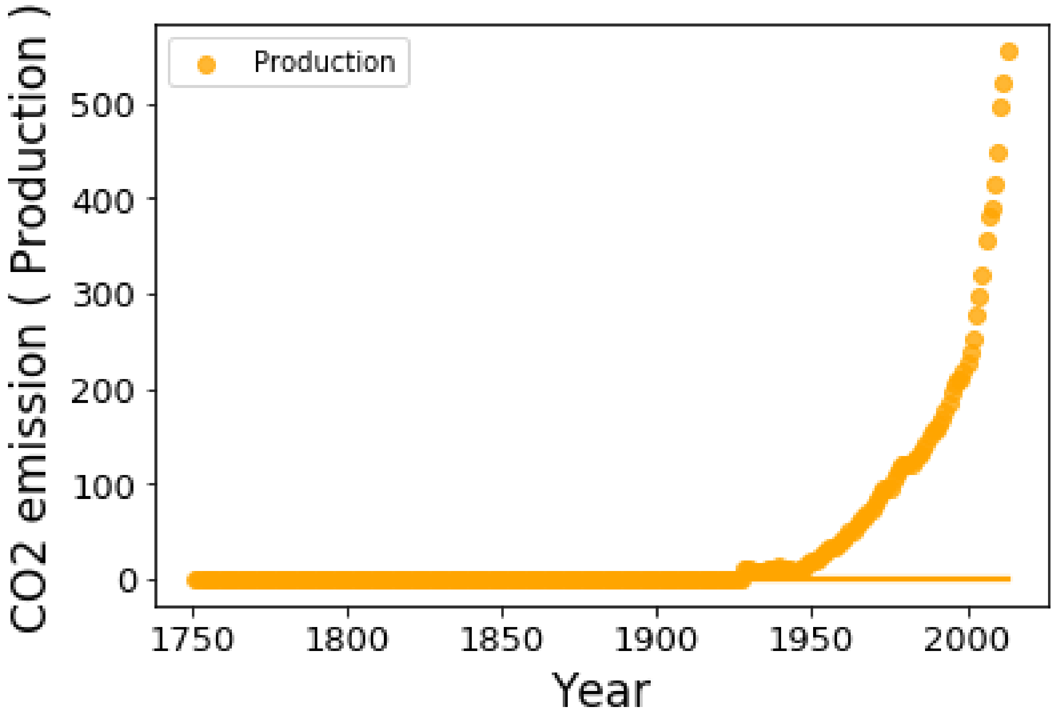

```
In [163]: plot_CO2_emission(df.Year,df.Liquids,'Liquid','orange')
```

```
Mean of CO2 526.6045627376426
Variance of CO2 945568.5300554379
Standard Deviation of CO2 972.4034811000205
```

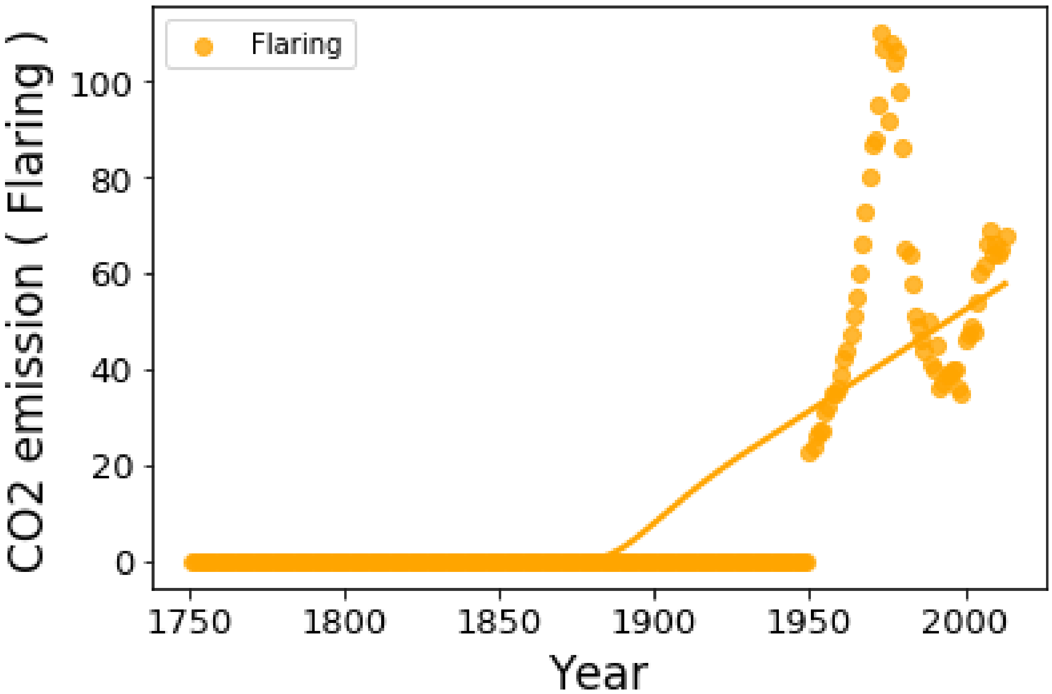



```
In [164]: plot_CO2_emission(df.Year,df.Production,'Production','orange')
```

```
Mean of CO2 39.741444866920155
Variance of CO2 8868.513046759357
Standard Deviation of CO2 94.1727829405044
```

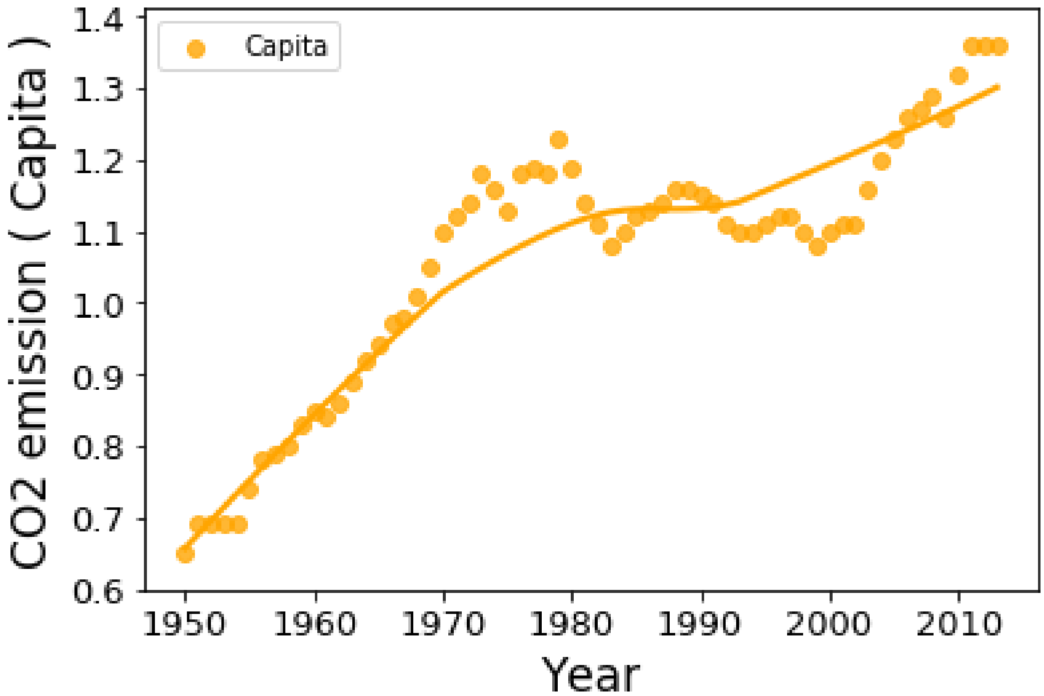



```
In [165]: plot_CO2_emission(df.Year,df.Flaring,'Flaring','orange')

Mean of CO2 13.752851711026617
Variance of CO2 721.0341044321245
Standard Deviation of CO2 26.85207821439757
```

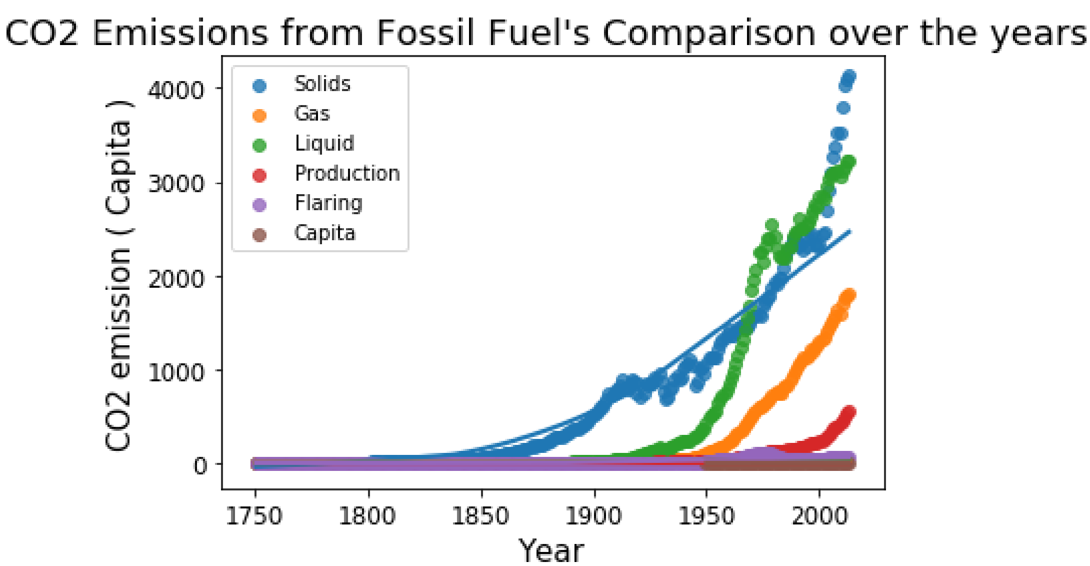



```
In [166]: plot_CO2_emission(df.Year,df.Capita,'Capita','orange')
```

```
Mean of CO2 1.0643749999999998
Variance of CO2 0.03319007936507937
Standard Deviation of CO2 0.18218144627013852
```

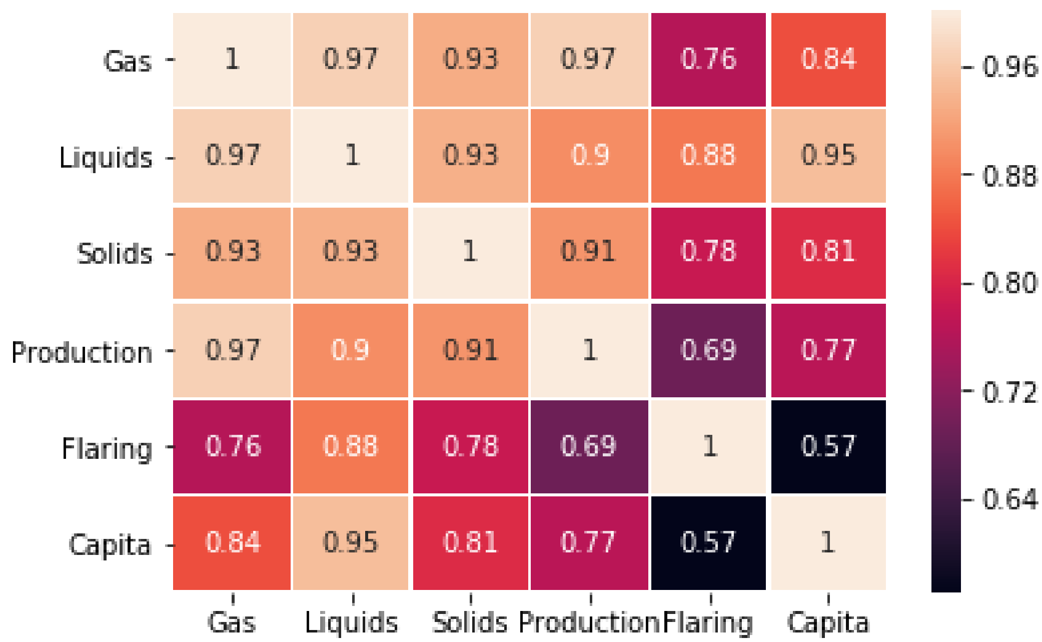



```
In [158]: plot_CO2_emission_all(df.Year,df.Solids,'Solids')
          plot_CO2_emission_all(df.Year,df.Gas,'Gas')
          plot_CO2_emission_all(df.Year,df.Liquids,'Liquid')
          plot_CO2_emission_all(df.Year,df.Production,'Production')
          plot_CO2_emission_all(df.Year,df.Flaring,'Flaring')
          plot_CO2_emission_all(df.Year,df.Capita,'Capita')
```



As per the comparison the rate of emission of Solids is more than any other followed by Liquid.

```
In [100]: df.keys()

Out[100]: Index(['Year', 'Total', 'Gas', 'Liquids', 'Solids', 'Production', 'Flaring',
                 'Capita'],
                dtype='object')

In [174]: sns.heatmap(df.drop(['Year','Total'],axis=1).corr(),linecolor='white',linewidths=0.4

Out[174]: <matplotlib.axes._subplots.AxesSubplot at 0x7f3ee98ca4a8>
```




Increase in levels of emmision of Gas and Liquids are highly coorelated with each other in comparison to others, Production and Flaring are least coorelated to each other

As we can see from multiple plots the trends in climate change and increase in level's of carbon dioxide in our surrounding over the years span.

# 4 Story

As per the plots created above we can see that the level of carbon dioxide gases is increasing day by day leading to increase in temperature. This increase in emission is made from multiple sources including solid, fossil fuels, liquids and gases. As per the deforestation have affected the control of carbondioxide leading to its increase in atmosplhere among different countries as we can see from the plots. These increase leads to increase in global temperature and people are spreading awareness about these issues on twitter as we can see from the plot made on frequency of most common words plotted.

```
In [ ]:
```