Alfredo Espana
Federico Cestau
Catherine Diagama-Pedroza

EXPLORATORY ANALYSIS ON SUICIDE DATA

INTRODUCTION

It was decided to study a data set of suicide ranges between the years of 1985 to 2016 in order to find trends and understand the suicide rates around the world.

The data set contains data about all the countries of the world.

The data was retrieved from the website:https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016

It was chosen five countries to analyze

DATA CLEAN

A verification was done to confirm the type of variables that were used and some of the variables were transform

Null values were dropped from the data set.

Two columns, the spaces and comas were removed.

The Data Frame was retrieved

DATA CLEAN

	country	year	sex	age	suicides_no	population	suicides/100k pop	country-year	HDI for year	gdp_for_year ($)	gdp_per_capita ($)	generation
0	Albania	1987	male	15-24 years	21	312900	6.71	Albania1987	NaN	2,156,624,900	796	Generation X
1	Albania	1987	male	35-54 years	16	308000	5.19	Albania1987	NaN	2,156,624,900	796	Silent
2	Albania	1987	female	15-24 years	14	289700	4.83	Albania1987	NaN	2,156,624,900	796	Generation X

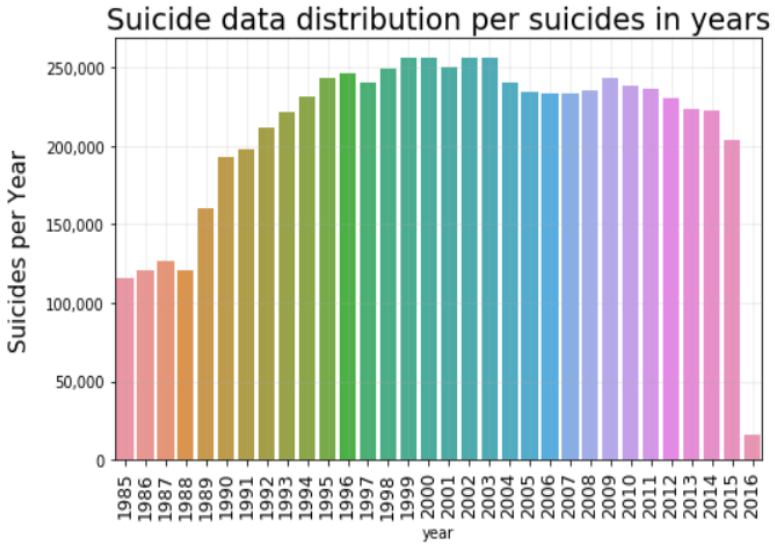
:

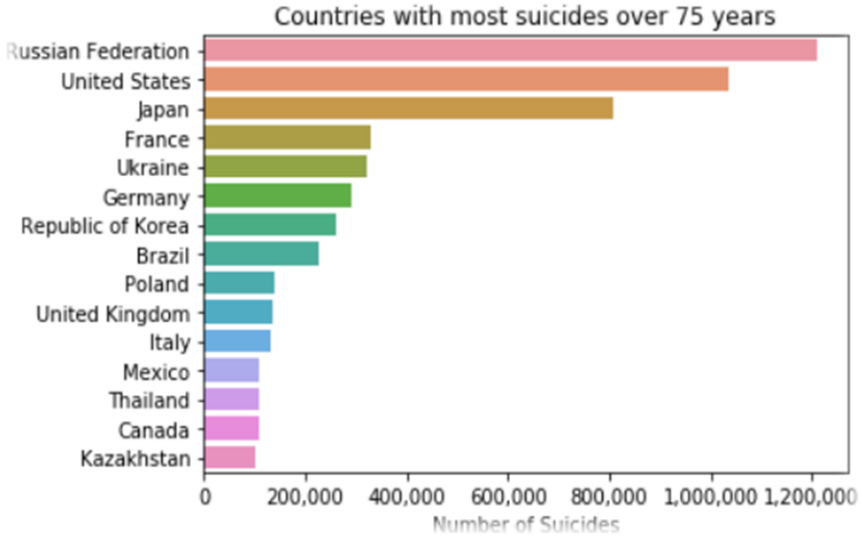
	country	year	sex	age	suicides_no	population	suicides/100k pop	gdp_for_year	gdp_per_capita	generation
0	Albania	1987	male	15-24 years	21	312900	6.71	2.156625e+09	796	Generation X
1	Albania	1987	male	35-54 years	16	308000	5.19	2.156625e+09	796	Silent

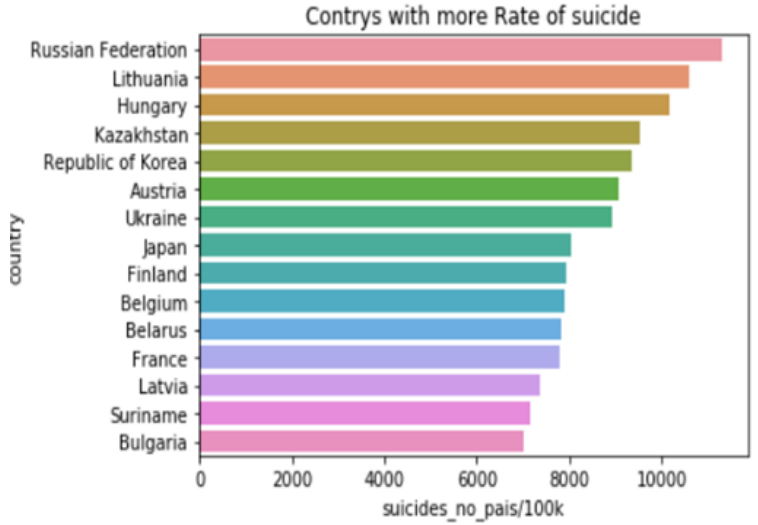
EXPLORATION

- Over the years , have the number of suicides increased?
- Which are the countries that commit more suicides?
- Do the distribution of suicides vary when we change countries?

```
suicides_no_year = []

for y in Suicide_data['year'].unique():
    suicides_no_year.append(sum([Suicide_data['year'] == y]['suicides_no']))

n_suicides_year = pd.DataFrame(suicides_no_year, columns=['suicides_no_year'])
n_suicides_year['year'] = Suicide_data['year'].unique()

top_year = n_suicides_year.sort_values('suicides_no_year', ascending=False)['year']
top_suicides = n_suicides_year.sort_values('suicides_no_year', ascending=False)['suicides_no_year']
```



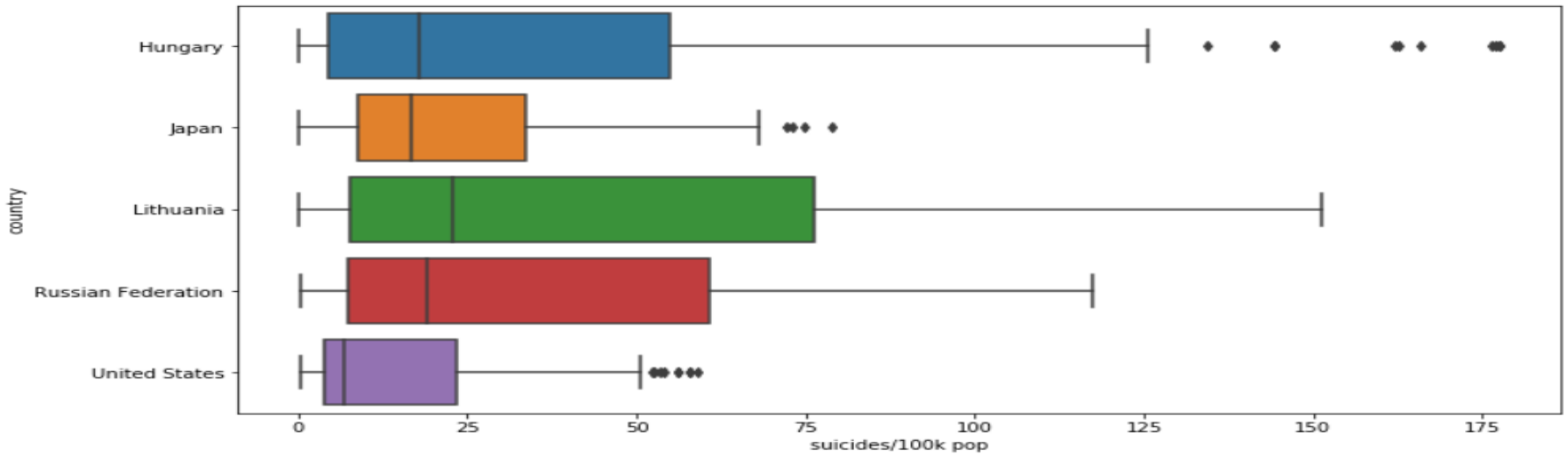




EXPLORATION

HOW THE SUICIDE'S RANGE VARIES WHEN IS COMPARED BETWEEN THE 5 COUNTRIES SELECTED ?

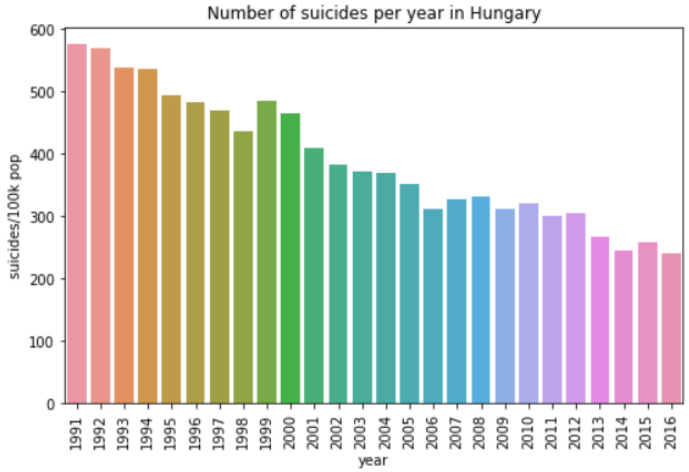
```
1 countries = ['Russian Federation', 'Lithuania', "Hungary", 'United States','Japan']
2 df_filtred = Suicide_data[[Suicide_data['country'][i] in countries for i in range(len(Suicide_data))]]  #
```

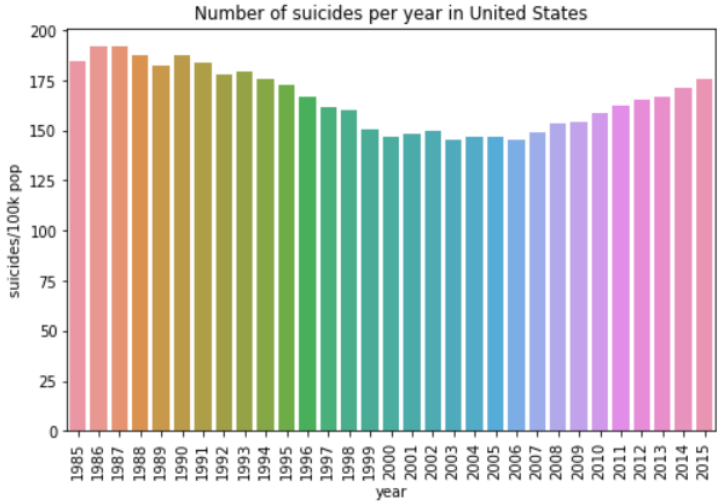


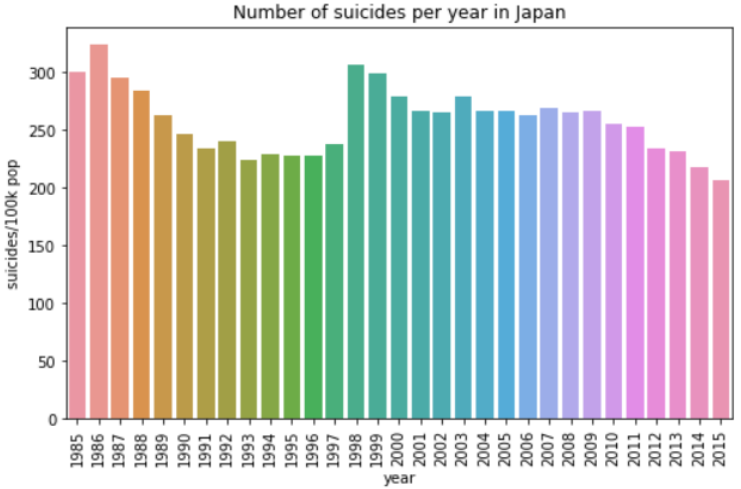
ANALYSIS

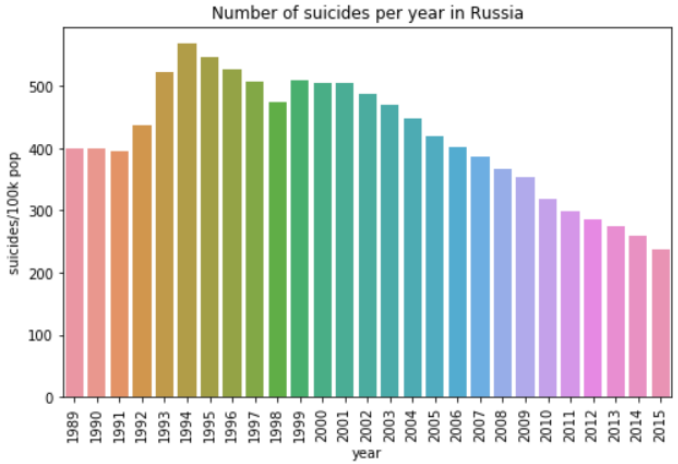
HOW DO THE NUMBER OF SUICIDES VARY OVER TIME?

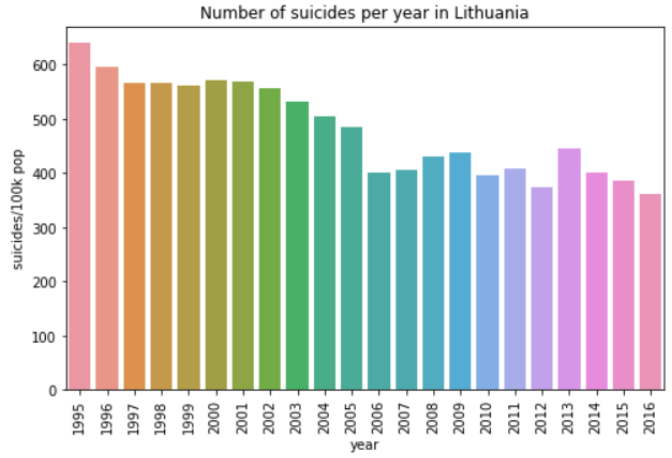
```
for y in df_united_states['year'].unique():
    suicides_no_year.append(sum(df_united_states[df_united_states['year'] == y]["suicides/100k pop"]))

n_suicides_year = pd.DataFrame(suicides_no_year, columns=["suicides/100k pop"])
n_suicides_year['year'] = df_united_states['year'].unique()

top_year = n_suicides_year.sort_values('suicides/100k pop', ascending=False)['year']
top_suicides = n_suicides_year.sort_values('suicides/100k pop', ascending=False)['suicides/100k pop']
```







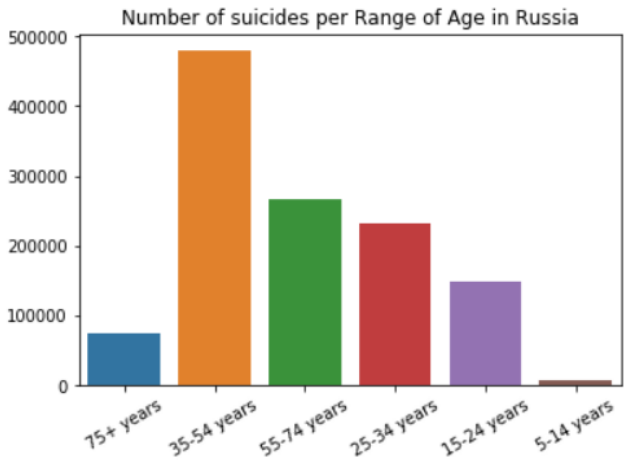


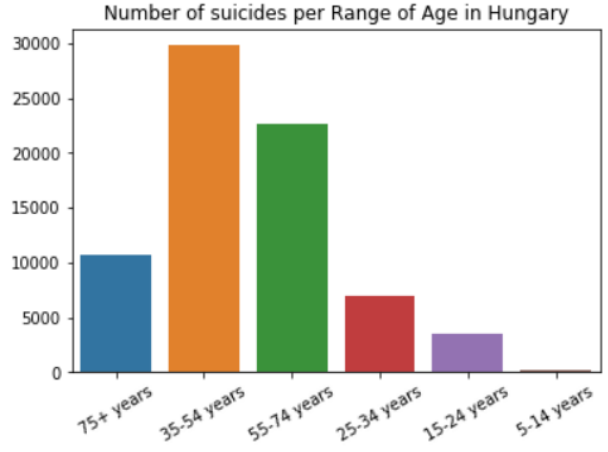


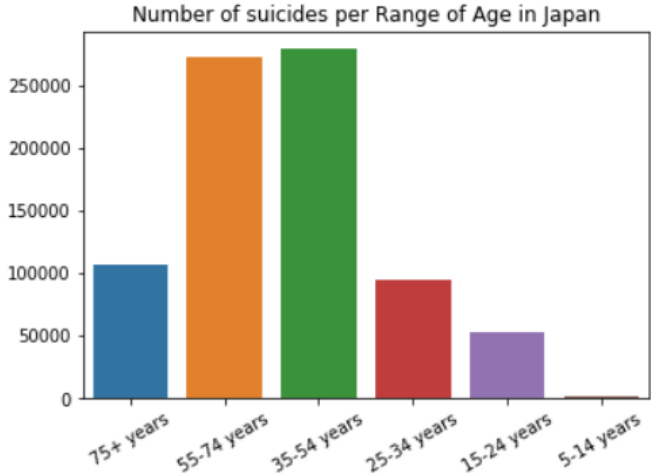
ANALYSIS

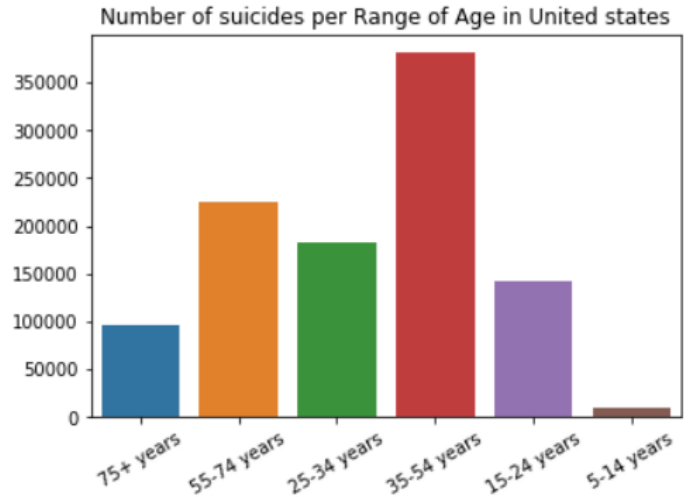
WHO TENDS TO COMMIT MORE SUICIDE? TEENS? ADULTS? ELDERS?

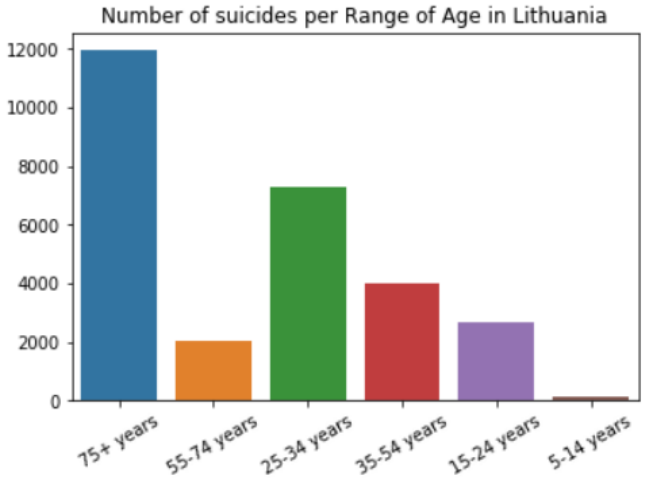
```
1  suicides_no_age = []
2
3  for a in df_united_states['age'].unique():
4      suicides_no_age.append(sum(df_united_states[df_united_states['age'] == a]['suicides_no']))
5
```







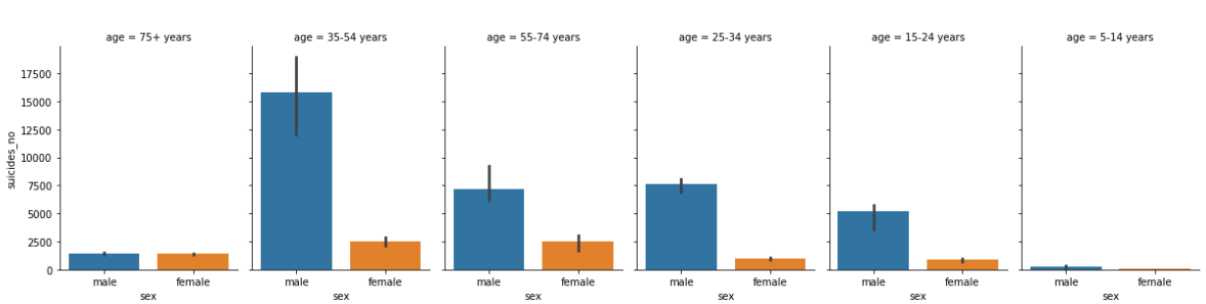


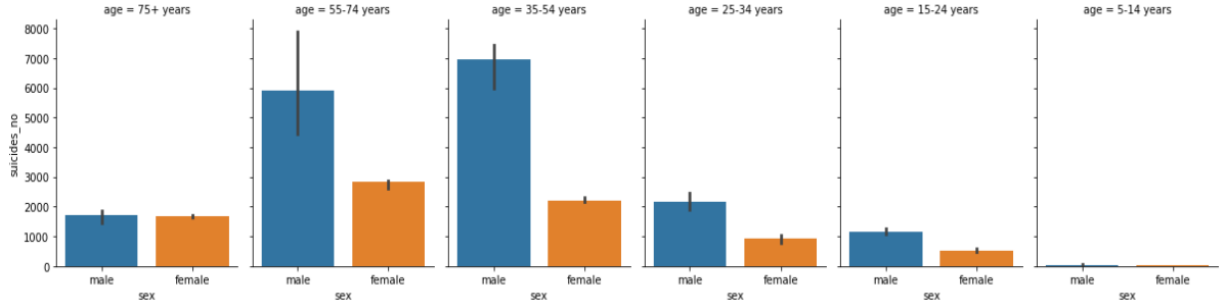


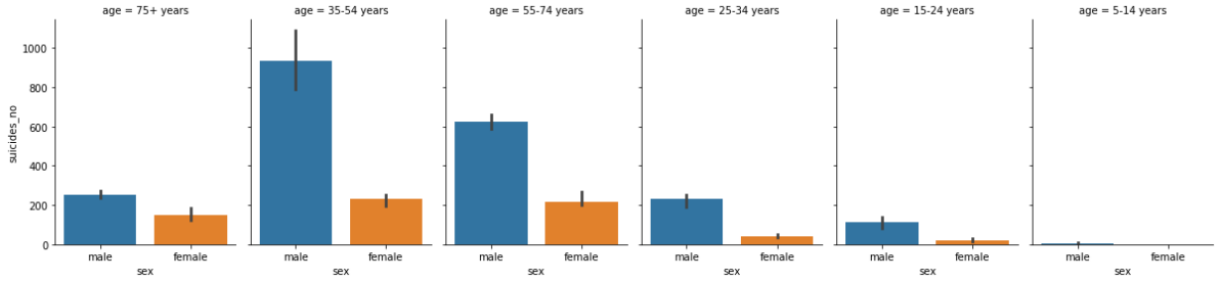
ANALYSIS

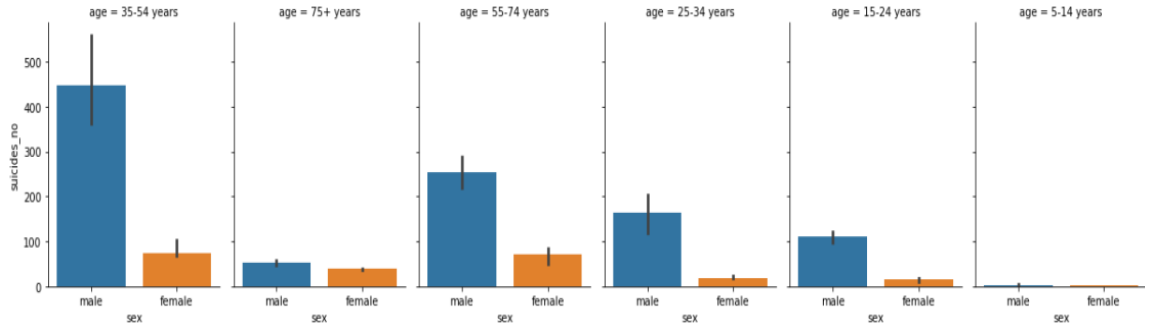
WHAT ABOUT SEX? WHO COMMITS MORE SUICIDE, MEN OR WOMEN?

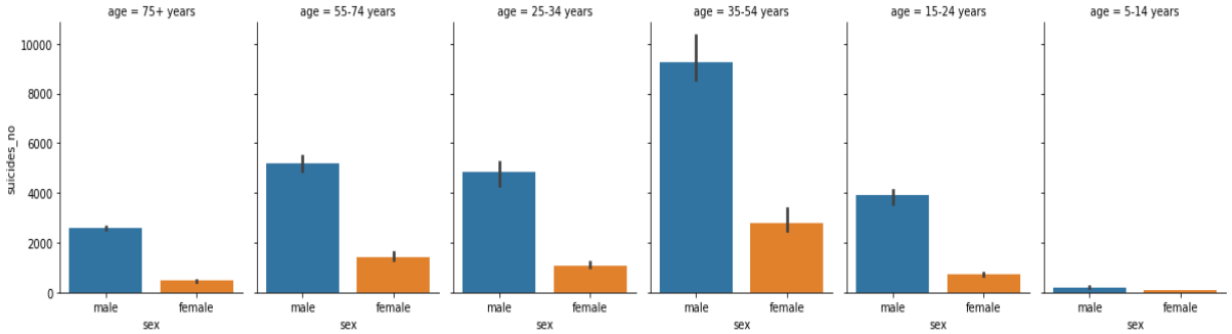
```
1 sns.catplot(x='sex', y='suicides_no',col='age', data=df_united_states, estimator=np.median,height=4, aspect=.7,kind='bar
```









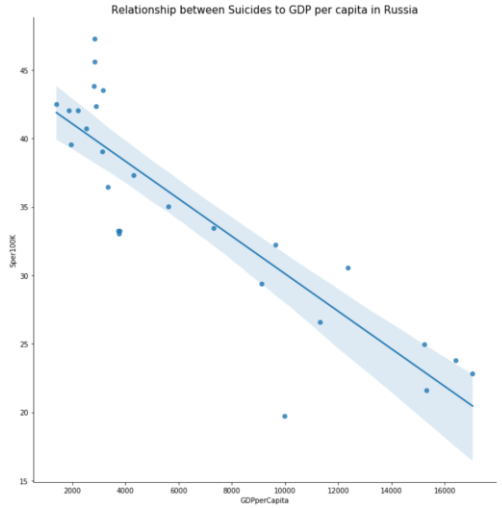


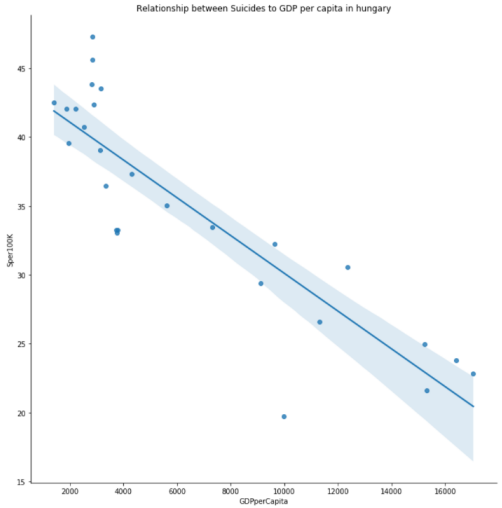
ANALYSIS

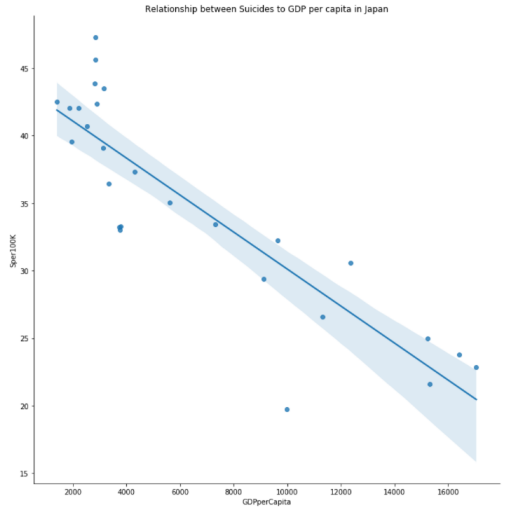
AS FOR THE GDP PER CAPITA, IS THERE ANY INFLUENCE?

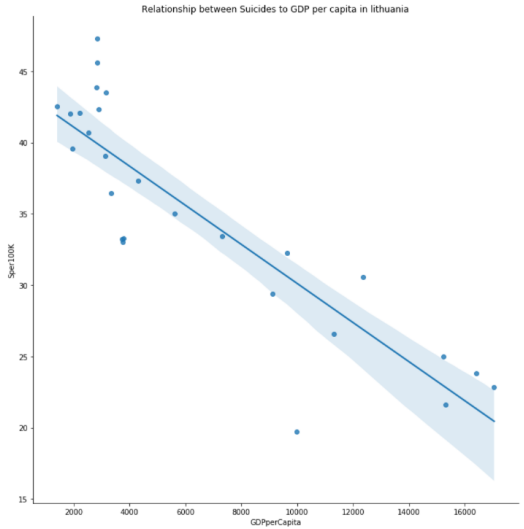
```
1 df_compare=pd.DataFrame({'Year':gdp_per_capita.index,'GDPperCapita':gdp_per_capita,'Sper100K':suicides_100k_pop})
```

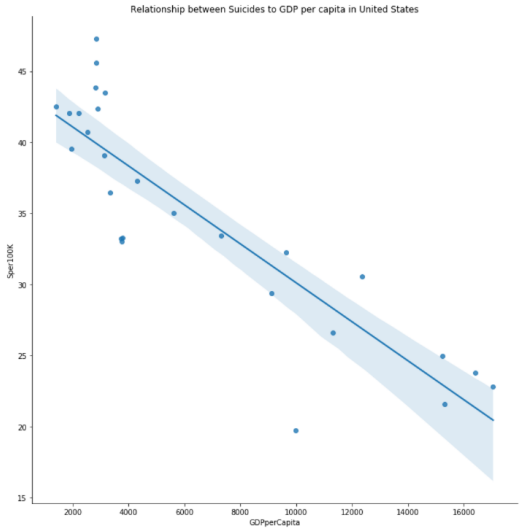
```
1 sns.lmplot(x='GDPperCapita',y='Sper100K',fit_reg=True,data=df_compare,height=10)
```











ANALYSIS

IS THE SUICIDE A POPULAR TREND IN USA?

- It was observed that the suicide rate doesn't decrease in the United States, a New York times API was used to confirm if this was an important topic in the news.

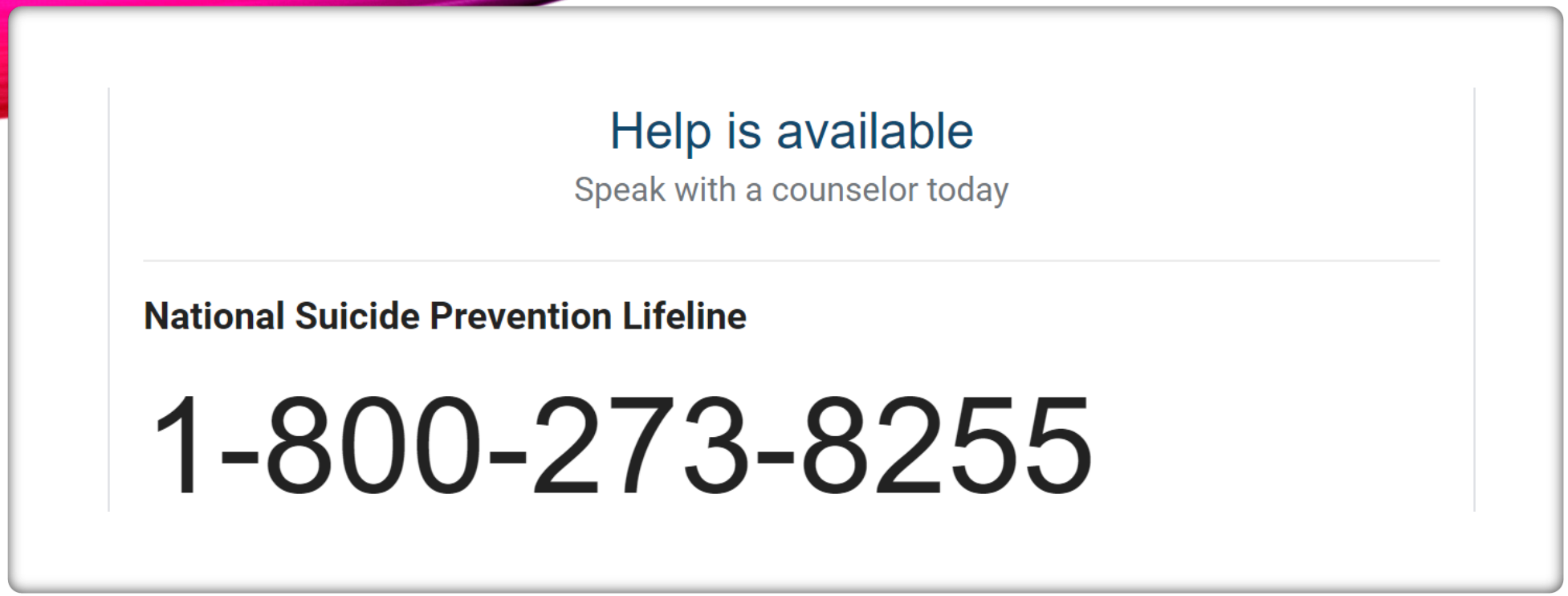
.

```
In [97]:
apik='YfhSSAXAIv21DhZA4W5LjxuaWbKVokWN'

url2= f'https://api.nytimes.com/svc/search/v2/articlesearch.json?q=suicide&api-key={apik}'

In [98]:
requests.get(url2).url

Out[98]: 'https://api.nytimes.com/svc/search/v2/articlesearch.json?q=suicide&api-key=YfhSSAXAIv21DhZA4W5LjxuaWbKVokWN'

In [99]:
data={
    'title':[],
    'detail':[]
}

In [100]:
response=requests.get(url2).json()

In [101]:
txt=requests.get(url2).text

In [102]:
len(txt.split('Suicide')) + len(txt.split('suicide'))

Out[102]: 35
```

CONCLUSIONS

The trends that were found
The peak of the rate was found in the range of 1987- 1995
The highest suicide rate was during the range of 35-54 years old
The gender with the highest rate was male
When the GDP is lower, there are more prone to negative biases.
If the country has a lower temperature the rate of suicides increases
It was found 35 articles related to suicide in less than 3 months.


QUESTIONS??



HTTPS://SUICIDEPREVENTIONLIFELINE.ORG/