Data Science

suicide analysis and model predictions

By: Eitan Tgania & Aviv Musai

Research question:

can we predict the suicide rate in certain population?

Data acquisition

Resources:

World Health Organization

First ideas:

- First idea - predict the car accidents rate in the next year- not enough database.

- second idea- predict volcanic eruptions. - good idea but not enough database

Database before cleaning

	country	year	sex	age	suicides_no	population
0	Albania	1985	female	15-24 years	NaN	277900.0
1	Albania	1985	female	25-34 years	NaN	246800.0
2	Albania	1985	female	35-54 years	NaN	267500.0
3	Albania	1985	female	5-14 years	NaN	298300.0
4	Albania	1985	female	55-74 years	NaN	138700.0
5	Albania	1985	female	75+ years	NaN	34200.0
6	Albania	1985	male	15-24 years	NaN	301400.0
7	Albania	1985	male	25-34 years	NaN	264200.0
8	Albania	1985	male	35-54 years	NaN	296700.0
9	Albania	1985	male	5-14 years	NaN	325800.0

```
country         43776
year            43776
sex             43776
age             43776
suicides_no     41520
population      38316
dtype: int64
```

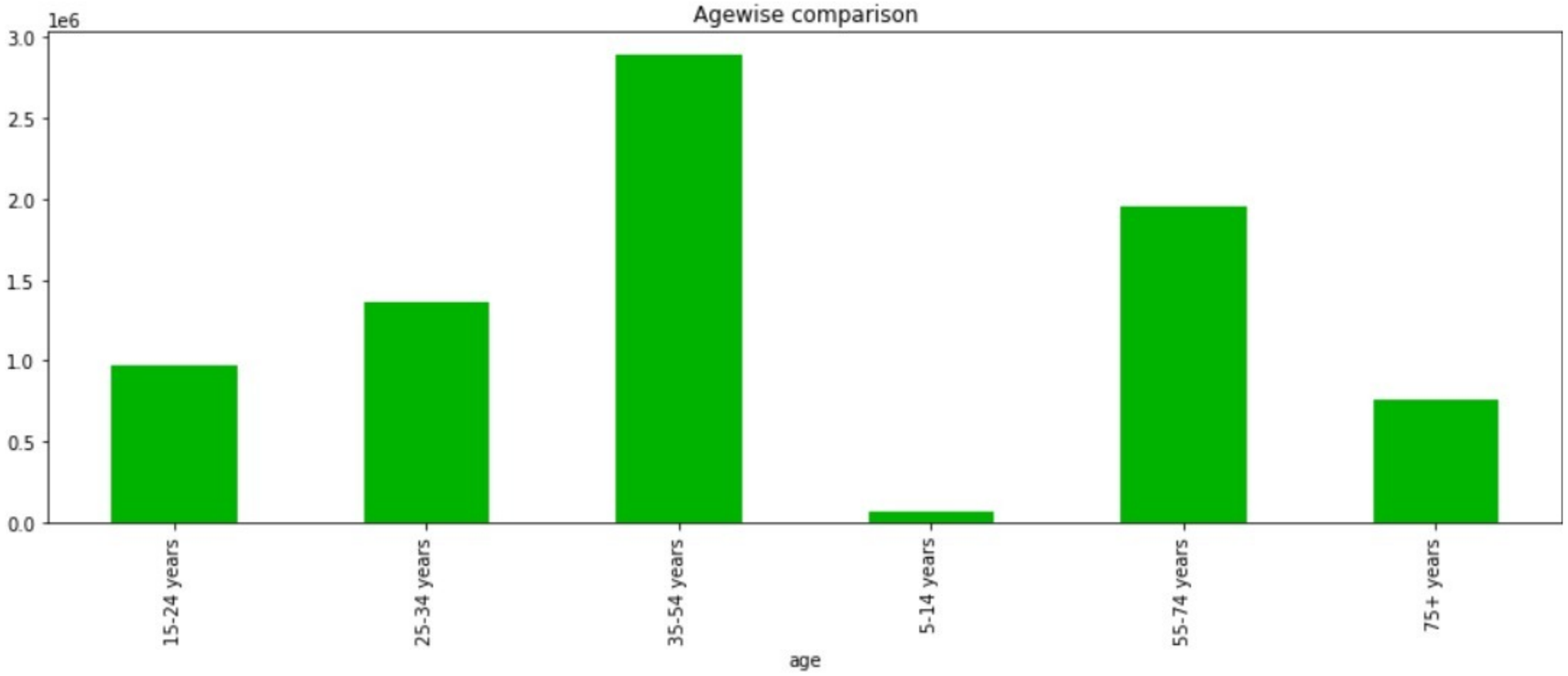
Database after cleaning

	country	year	sex	age	suicides_no	population
24	Albania	1987	female	15-24 years	14.0	289700.0
25	Albania	1987	female	25-34 years	4.0	257200.0
26	Albania	1987	female	35-54 years	6.0	278800.0
27	Albania	1987	female	5-14 years	0.0	311000.0
28	Albania	1987	female	55-74 years	0.0	144600.0
29	Albania	1987	female	75+ years	1.0	35600.0
30	Albania	1987	male	15-24 years	21.0	312900.0
31	Albania	1987	male	25-34 years	9.0	274300.0
32	Albania	1987	male	35-54 years	16.0	308000.0
33	Albania	1987	male	5-14 years	0.0	338200.0

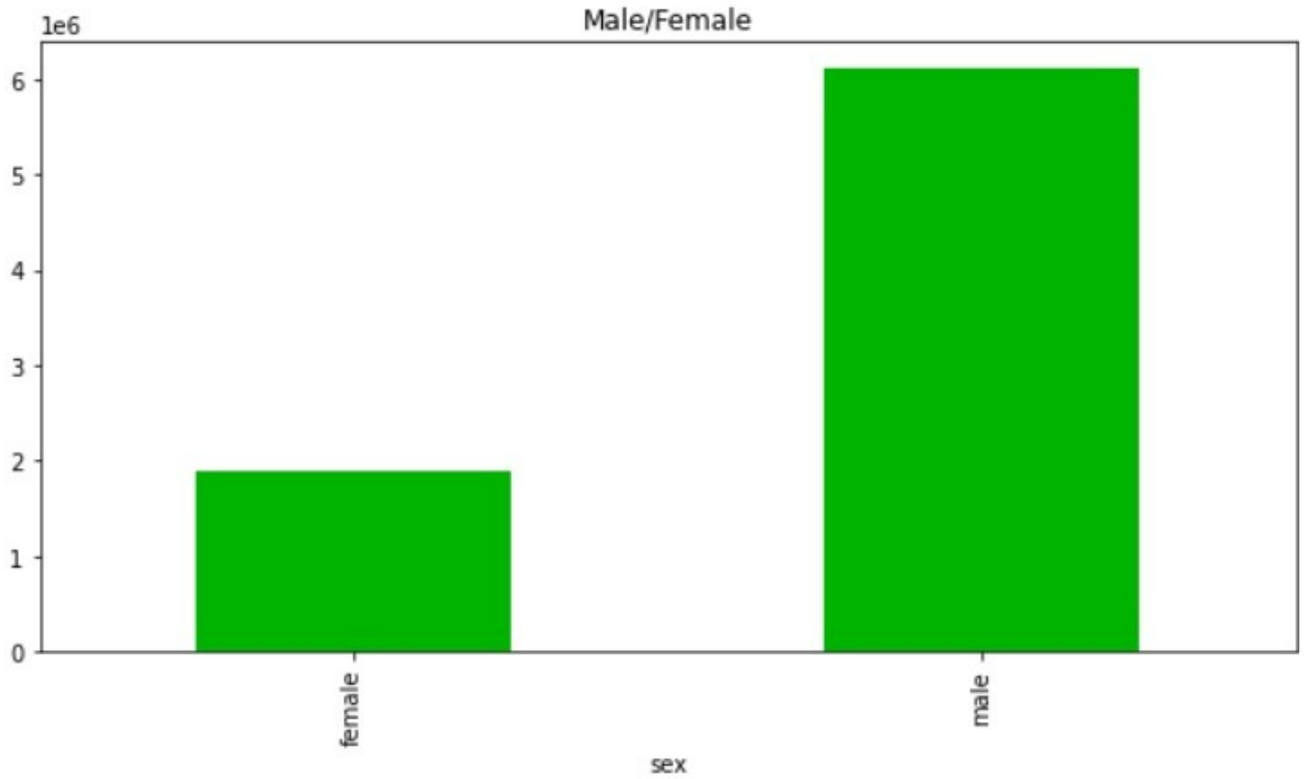
```
country        36060
year           36060
sex            36060
age            36060
suicides_no    36060
population     36060
dtype: int64
```

Now let's look at the rates....

Agewise comparison

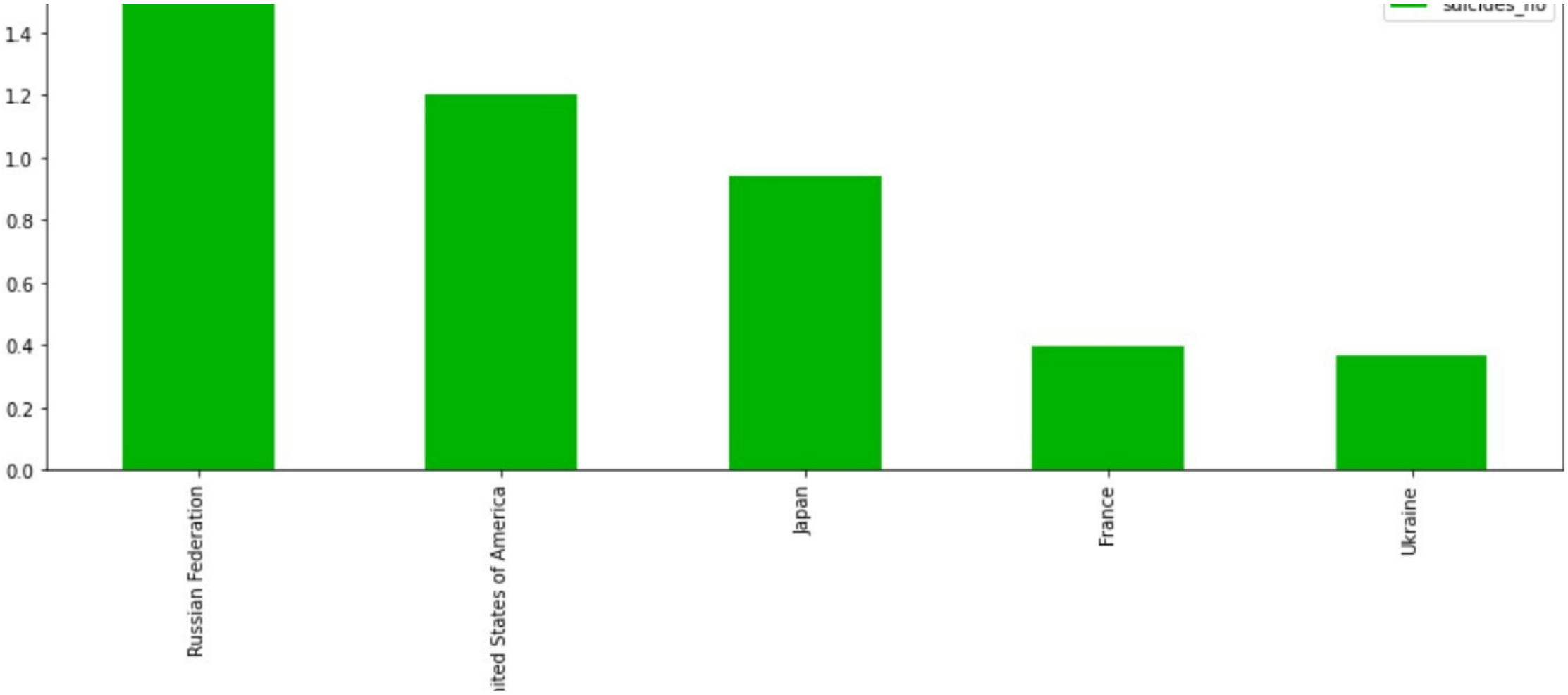


Male/ Female

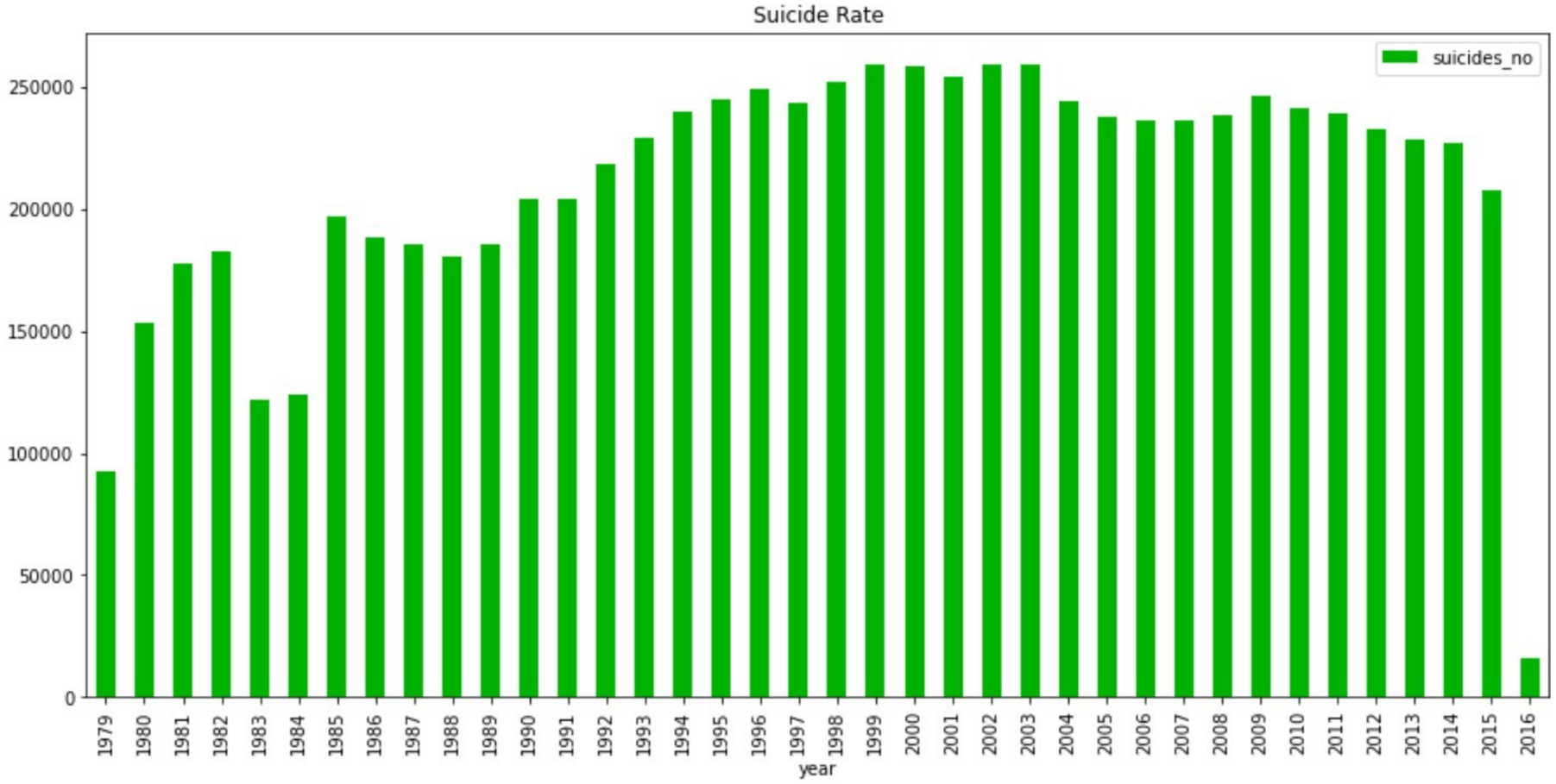


Countreies comparison

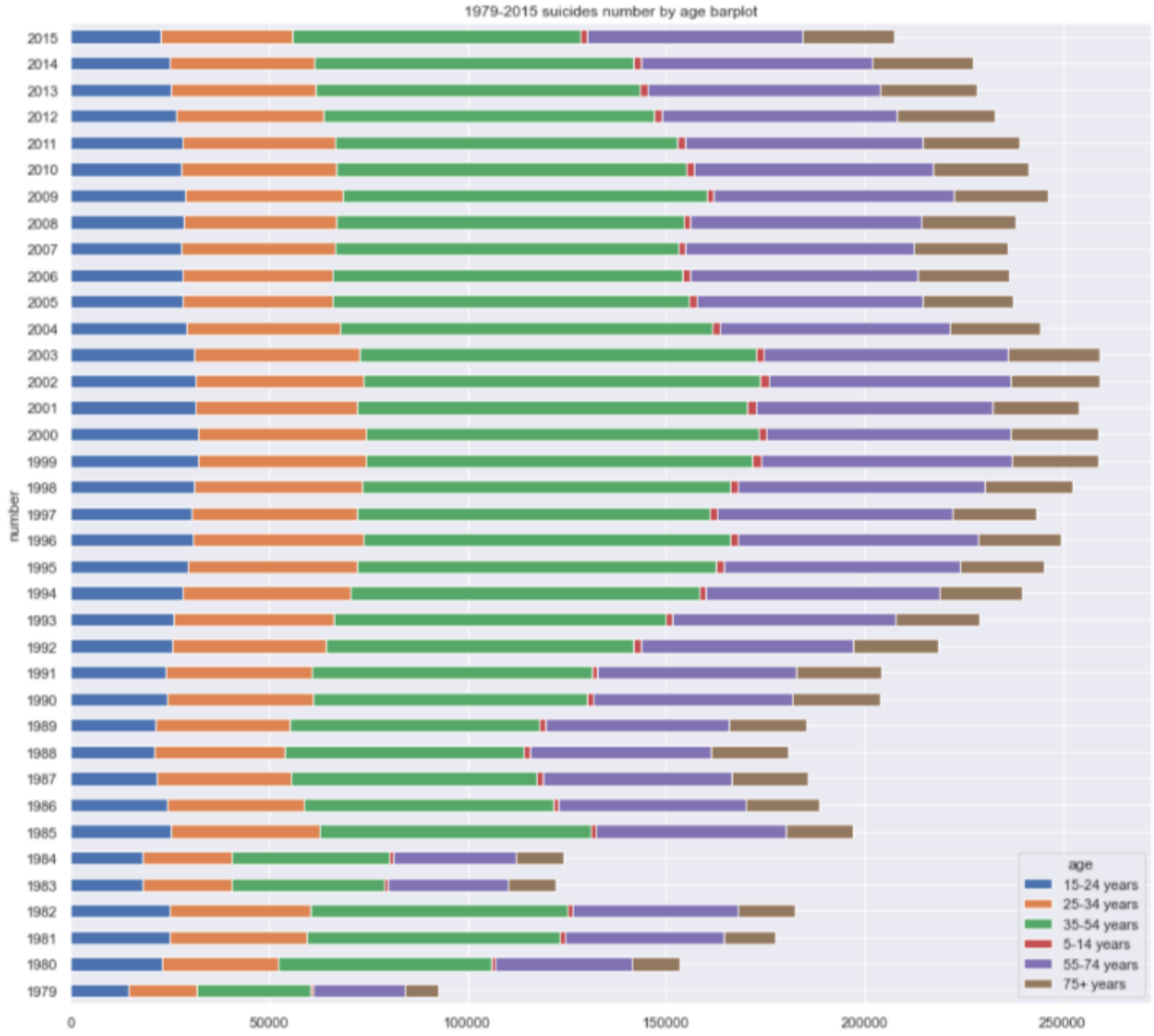
Top 5 countries with highest suicide rate



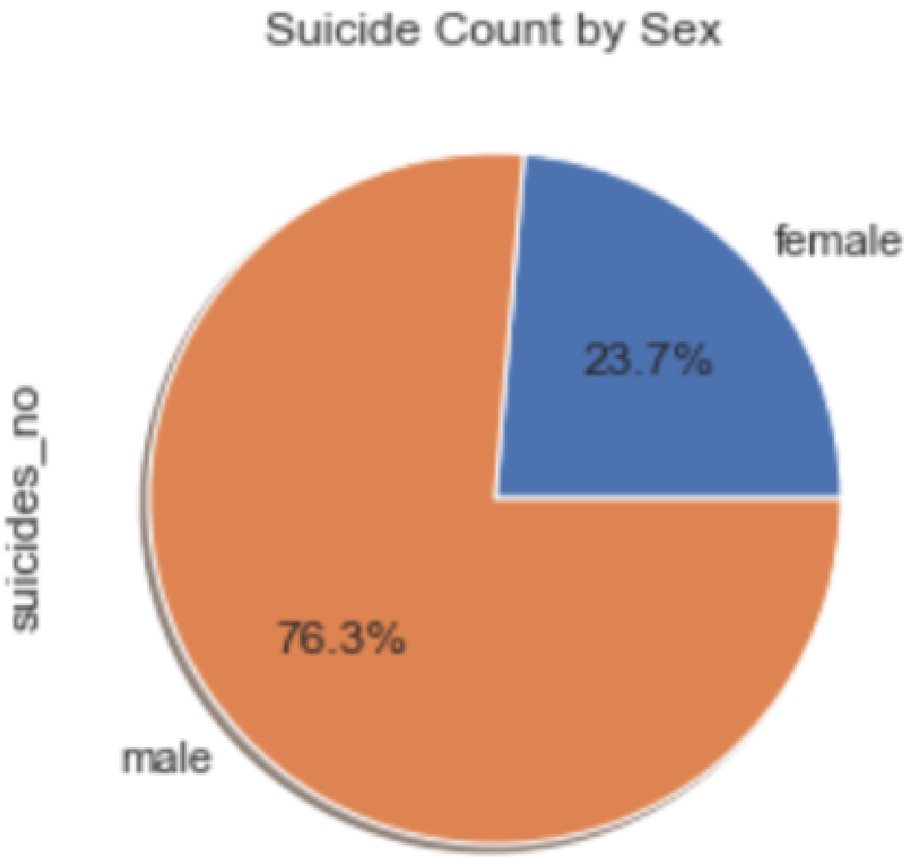
Suicide rate over the years 1979-2016



suicide numbers by age:



Suicide count by sex



Machine learning

machine learning

We decided to build models in order to try and see if we can predict suicide rates. We examined two models : **logistic regression** and **decision tree**. We trained the models and got almost accurate result.

```
Logistic Regression : 0.999

Decision Tree  0.9924570160843039
```

conclusions

In conclusion, as we all know we can't predict suicide rates, but the results of the models show we can get close to the original rates.

The End