Source Data: Suicide Rates Overview 1985 to 2016

https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/version/1#_=
(https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/version/1#_=)_

In [43]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

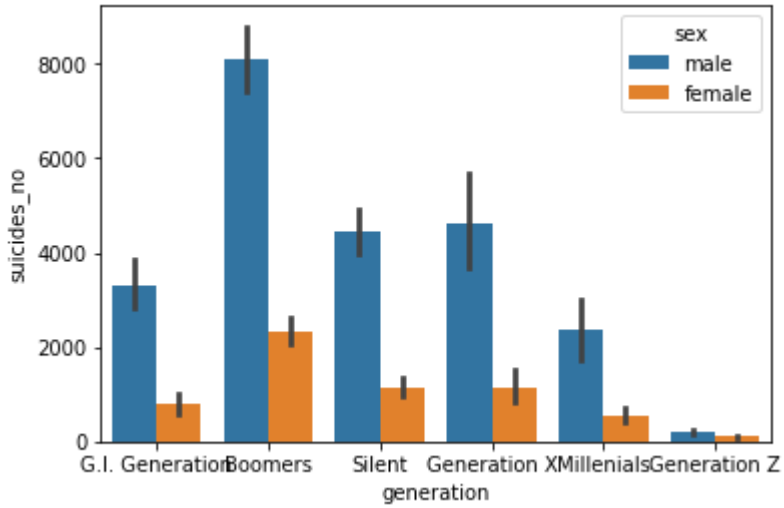
In [44]:
```
#import data from csv file
suicide = pd.read_csv('C:/Users/Louis/Desktop/4170/Seaborn graph practice/suicide-rates-ove
suicide.head()
```

Out[44]:

	country	year	sex	age	suicides_no	population	suicides/100k pop	country-year	HDI for year	gdp_fo
0	United States	1985.0	male	75+ years	2177.0	4064000	53.57	United States1985	0.841	4,346,73
1	United States	1985.0	male	55-74 years	5302.0	17971000	29.50	United States1985	0.841	4,346,73
2	United States	1985.0	male	25-34 years	5134.0	20986000	24.46	United States1985	0.841	4,346,73
3	United States	1985.0	male	35-54 years	6053.0	26589000	22.77	United States1985	0.841	4,346,73
4	United States	1985.0	male	15-24 years	4267.0	19962000	21.38	United States1985	0.841	4,346,73

In [97]:
```
import seaborn as sns
# use barplot to show mean and 95 CI of suicide rate across different generation between se
sns.barplot(x='generation',y = 'suicides_no', hue ='sex',data = suicide)
```

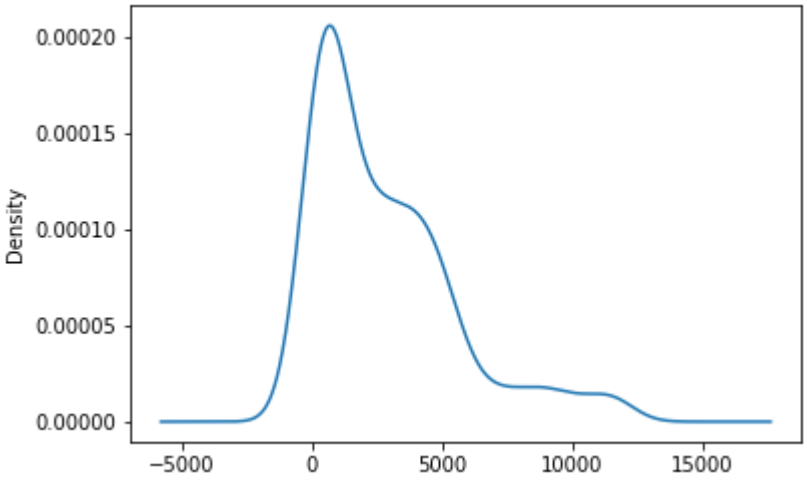
Out[97]: <matplotlib.axes._subplots.AxesSubplot at 0x1560d088780>



In [47]:
```
#show the distribution of number of suicides using KDE plot
suicide['suicides_no'].plot.density()

# it seems to be a right-skewed distribution leaning on 1000.
```

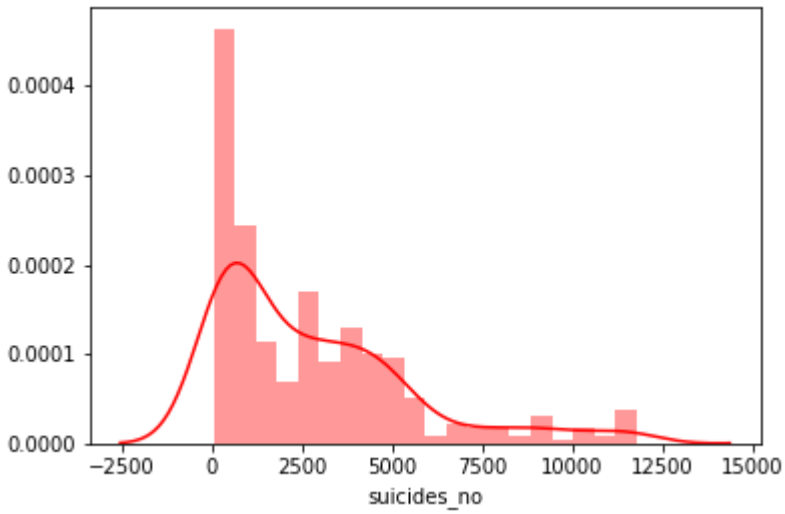
Out[47]: <matplotlib.axes._subplots.AxesSubplot at 0x1567575fa58>



In [61]:
```
#also a displot is vital to see how the suicides number are distributed, I want to know how
sns.distplot(suicide['suicides_no'], bins=20, color='r')
```

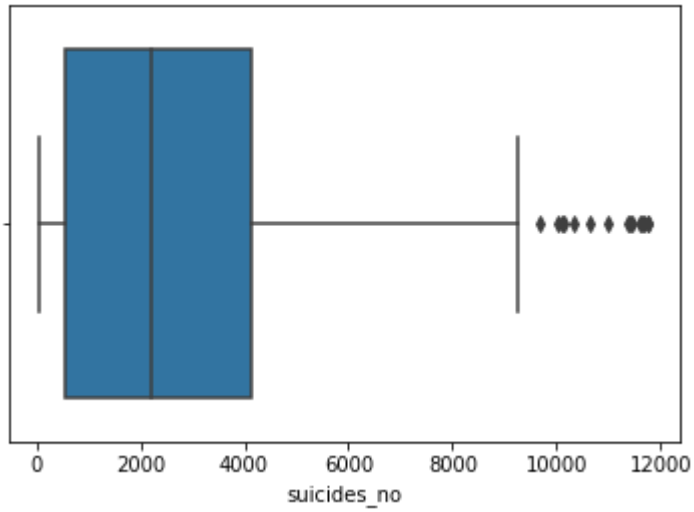
```
C:\Users\Louis\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning:
The 'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```

Out[61]: <matplotlib.axes._subplots.AxesSubplot at 0x156752e0da0>

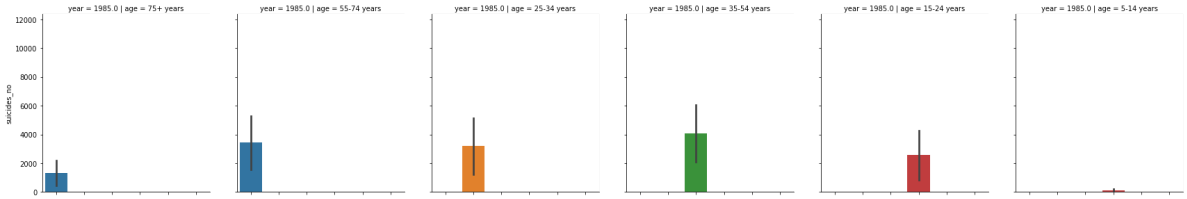


In [64]:
```
#since the number seems a bit sketchy, I try to see if there is outliers by looking at the
sns.boxplot(x=suicide['suicides_no'])
```

Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x156740ac5c0>

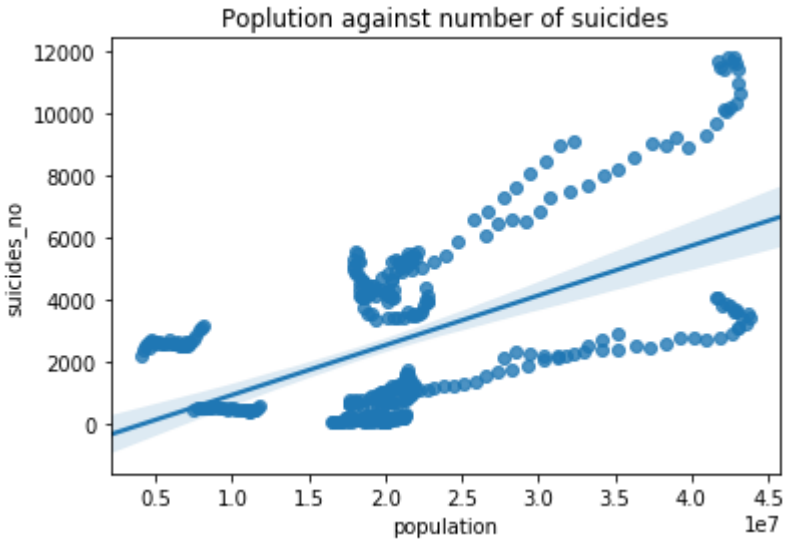


In [76]:
```
#I try to grasp the details of this dataset by visualizing them one by one
sns.factorplot(x='generation',y='suicides_no', row = 'year',col ='age', kind ='bar',data =
```



In [92]:
```
sns.regplot('population','suicides_no', data=suicide)
plt.title('Poplution against number of suicides')
```

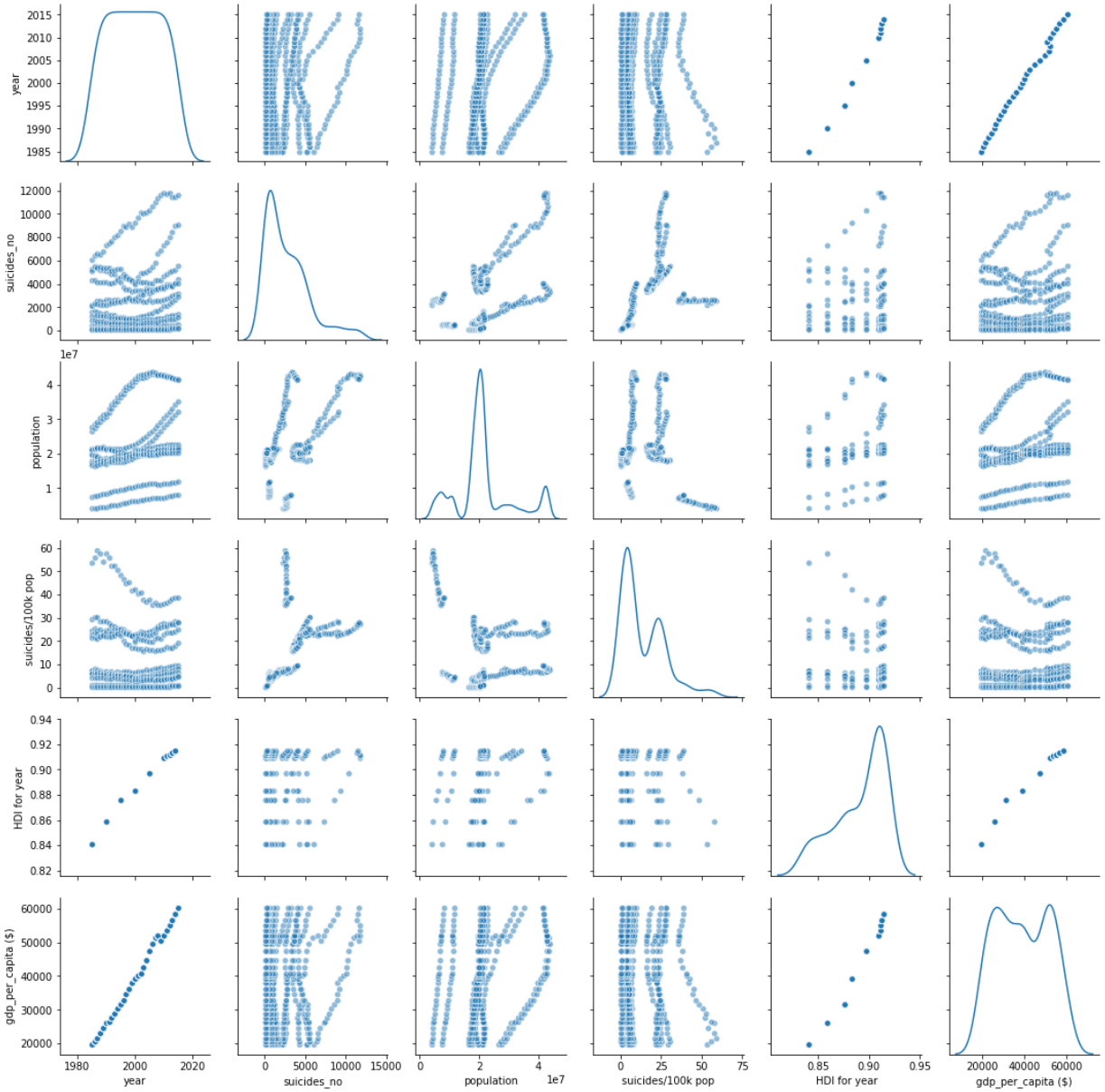
Out[92]: Text(0.5,1,'Poplution against number of suicides')



In [94]:
```
sns.pairplot(suicide, diag_kind='kde', plot_kws={'alpha': 0.5})
# series of scatter plot plotting against each columns with others and the diagonsis shows
```

```
C:\Users\Louis\Anaconda3\lib\site-packages\statsmodels\nonparametric\kde.py:448: RuntimeWarning: invalid value encountered in greater
  X = X[np.logical_and(X > clip[0], X < clip[1])] # won't work for two columns.
C:\Users\Louis\Anaconda3\lib\site-packages\statsmodels\nonparametric\kde.py:448: RuntimeWarning: invalid value encountered in less
  X = X[np.logical_and(X > clip[0], X < clip[1])] # won't work for two columns.
```

Out[94]: <seaborn.axisgrid.PairGrid at 0x1560aa17b00>



__There are lots of stuff may lead to suicidal mind but apprently a few in this dataset can reflect a certain relationships amongst number of suicides GPD, Sex and Generation.One stood out to me in the barplot is that male tends to commit suicide especially babyboomers.__

In [ ]:
```
```