In [0]:

```
from google.colab import files
files.upload()
```

In [0]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from scipy.stats import norm
from sklearn.preprocessing import StandardScaler
from scipy import stats
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
```

In [10]:

```
!ls
```

```
check4.csv  open?id=1f-AQkQmw4CNgOS_A8qKTSn3Vf94zhzTm  sample_data
data.csv    open?id=1f-AQkQmw4CNgOS_A8qKTSn3Vf94zhzTm.1
```

In [0]:

```
data = pd.read_csv('check4.csv')
```

In [12]:

```
data.head()
```

Out[12]:

	Unnamed: 0	Country	Symbol	Size	Population	Users	Penetration	date
0	0	Afghanistan	AF	645807	3.637318e+07	5700905.0	15.7 %	Dec/2017
1	1	Africa	--	30221532	1.287914e+09	453329534.0	35.2 %	Dec/2017
2	3	Albania	AL	28748	2.911428e+06	1916233.0	65.8 %	Mar/2017
3	4	Algeria	DZ	2381741	4.200805e+07	18580000.0	44.2 %	Mar/2017
4	5	American Samoa	AS	197	5.565300e+04	22000.0	39.5 %	Mar/2017

Most important criteria to analyse country is its population. Let's work on it first.

In [17]:

```
data['Population'].describe()
```

Out[17]:

```
count    2.520000e+02
mean     6.202275e+07
std      3.099332e+08
min      1.200000e+02
25%      3.801508e+05
50%      5.189607e+06
75%      2.374042e+07
max      4.207588e+09
Name: Population, dtype: float64
```

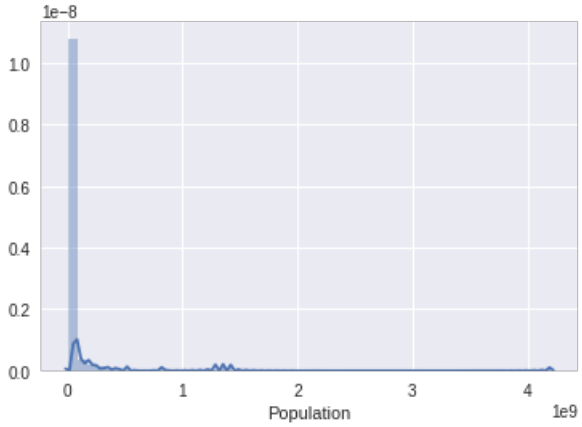
In [26]:

```
 max_pop = max(data['Population'])
min_pop = min(data['Population'])
print('Area with max population: ', data['Country'][data['Population'] == max_pop].values[0])
print('Area with min population: ', data['Country'][data['Population'] == min_pop].values[0])
```

```
Area with max population:  Asia
Area with min population:  French Southern Terr.
```

In [18]:

```
sns.distplot(data['Population']);
```



It can be observed that there lies some outliers and graph is wery peaked. Let's check its skewness and kurtness.

In [19]:

```
print("Skewness: %f" % data['Population'].skew())
print("Kurtosis: %f" % data['Population'].kurt())
```

```
Skewness: 10.568004
Kurtosis: 131.471672
```

Data is highly skewed and showing very high kurtness also.

Let's check other variables.

Another important deciding factor is coutry's Size. Let's check it out.

Size

In [42]:

```
from io import StringIO
cont_size = data[data.Size.apply(lambda x: str(x).isnumeric())]['Size']
cont_pop = data[data.Size.apply(lambda x: str(x).isnumeric())]['Population']
cont_size.head(3)
```

Out[42]:

```
0      645807
1    30221532
2       28748
Name: Size, dtype: object
```

In [36]:

```
cont_size.describe()
```

Out[36]:

```
count       250
unique      242
top       14604
freq          2
Name: Size, dtype: object
```

It's time to see relation between population and size of countries.