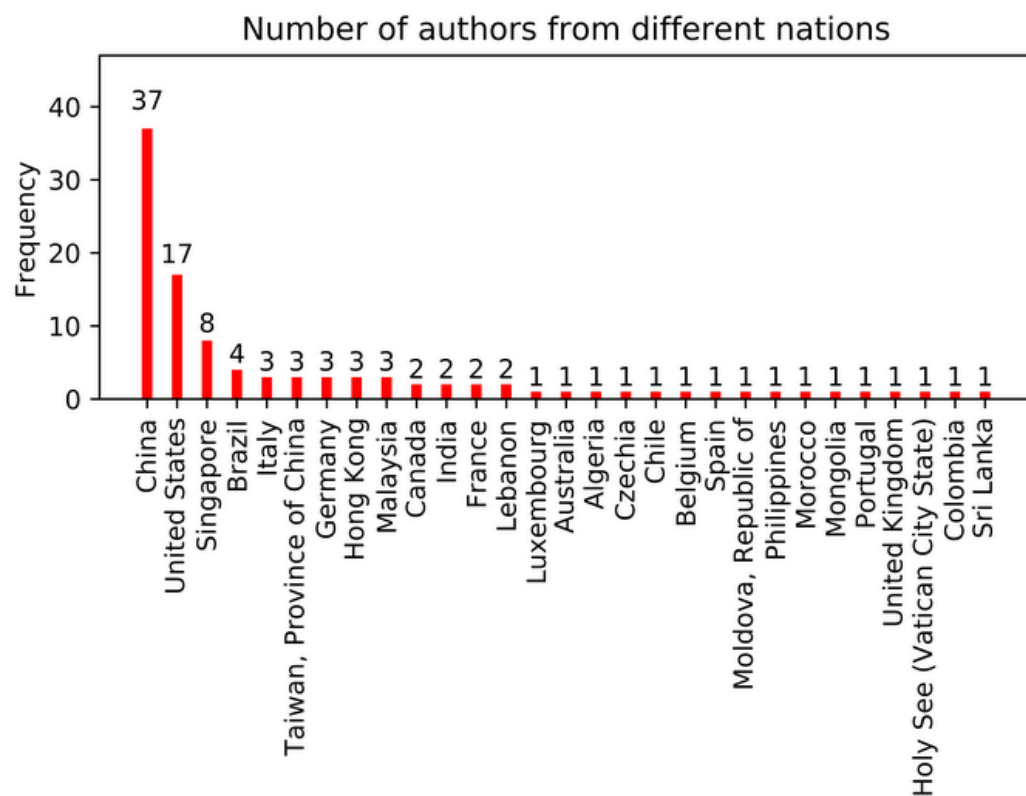## READ ME

This project takes a dataset which has the list of paper published for a certain topic. The universities/institution name of the author is also given in the data. The program inputs this data through an spreadsheet and then extracts out the country name from the universities name. In the end, it plots a graph of the frequency of the paper published to the nation of the author.

The code makes use of pycountry and geograpy packages in python, in order to extract country name out of the given data.

Number of authors from different nations

Downloading and Installing Packages:

pip install xlrd

pip install numpy

pip install matplotlib

pip install re

pip install pandas

pip install geograpy

pip install pycountry

Working And Algo

We can use each of Geograpy's modules on their own. For example:

```
from geograpy import extraction

e = extraction.Extractor(url='http://www.bbc.com/news/world-europe-26919928')
e.find_entities()

# You can now access all of the places found by the Extractor
print e.places
```

If country still not extracted, splitting the string using space as delimiter.

```
    country2 = text2.split()
```

If country still not extracted, splitting the string using ',' as delimiter.

```
    country3 = [x.strip() for x in text2.split(',')]
```

So, three lists we have: e.places , country2 and country3; which has names extracted out, now we will check if these names exist in pycountry module or not, if yes we will mark that country count as 1 and hence, in the end, plot a graph.