Exploratory Data Analysis using EBOLA dataset

by Anooraag Basu

ABOUT EBOLA

- **What is Ebola?**

Ebola virus disease (EVD), formerly known as Ebola haemorrhagic fever, is a rare but severe, often fatal illness in humans.

- **How does it spread?**

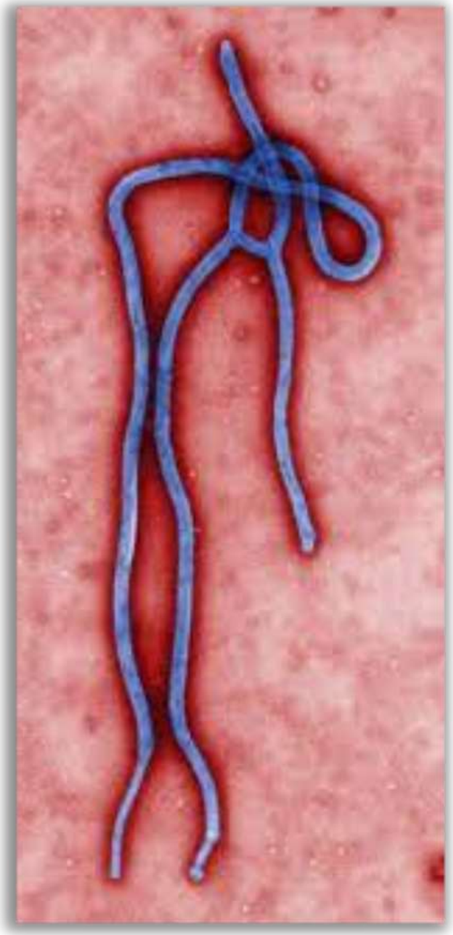
The virus is transmitted to people from wild animals and spreads in the human population through human-to-human transmission.
Common example - Fruit Bats

- **What are the symptoms?**

The Ebola virus causes an acute, serious illness that is often fatal if untreated. Other symptoms include fever, fatigue, muscle pain, headache and sore throat.

- **How can it be treated?**

A range of potential treatments including blood products, immune therapies and drug therapies are currently being evaluated.



DATASET USED

	A	B	C	D
1	**Country**	**Date**	**Cumulative no. of confirmed, probable and suspected cases**	**Cumulative no. of confirmed, probable and suspected deaths**
2	Guinea	29-08-2014	648	430
3	Nigeria	29-08-2014	19	7
4	Sierra Leone	29-08-2014	1026	422
5	Liberia	29-08-2014	1378	694
6	Sierra Leone	05-09-2014	1261	491
7	Nigeria	05-09-2014	22	8
8	Liberia	05-09-2014	1871	1089
9	Guinea	05-09-2014	812	517
10	Senegal	05-09-2014	1	0
11	Senegal	08-09-2014	3	0
12	Guinea	08-09-2014	862	555
13	Sierra Leone	08-09-2014	1361	509
14	Liberia	08-09-2014	2046	1224
15	Nigeria	08-09-2014	21	8
16	Guinea	12-09-2014	861	557
17	Sierra Leone	12-09-2014	1424	524
18	Nigeria	12-09-2014	21	8
19	Liberia	12-09-2014	2081	1137
20	Senegal	12-09-2014	3	0
21	Senegal	16-09-2014	1	0

INSTALLING AND IMPORTING LIBRARIES

- **Installing the 3 libraries**

```
In [1]: pip install pandas
```

```
In [2]: pip install matplotlib
```

```
In [3]: pip install seaborn
```

- **Importing the 3 libraries**

```
In [4]: import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sn
```

USING PANDAS

In [5]:
```
ebola = pd.read_csv(r"C:\Users\Anoorag\OneDrive\Desktop\eboladatatset.csv")
ebola.head()
```

Out[5]:

	Country	Date	Cumulative no. of confirmed, probable and suspected cases	Cumulative no. of confirmed, probable and suspected deaths
0	Guinea	29-08-2014	648.0	430
1	Nigeria	29-08-2014	19.0	7
2	Sierra Leone	29-08-2014	1026.0	422
3	Liberia	29-08-2014	1378.0	694
4	Sierra Leone	05-09-2014	1261.0	491

In [6]:
```
ebola = pd.read_csv(r"C:\Users\Anoorag\OneDrive\Desktop\eboladatatset.csv")
ebola.head(7)
```

Out[6]:

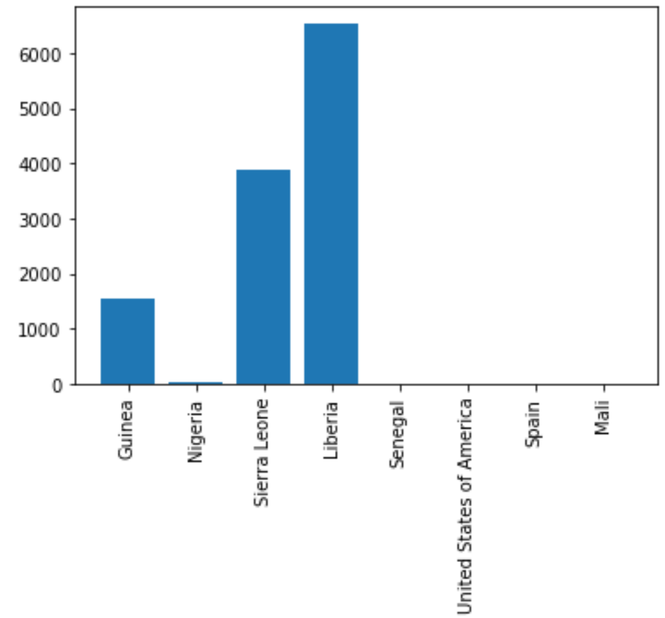
	Country	Date	Cumulative no. of confirmed, probable and suspected cases	Cumulative no. of confirmed, probable and suspected deaths
0	Guinea	29-08-2014	648.0	430
1	Nigeria	29-08-2014	19.0	7
2	Sierra Leone	29-08-2014	1026.0	422
3	Liberia	29-08-2014	1378.0	694
4	Sierra Leone	05-09-2014	1261.0	491
5	Nigeria	05-09-2014	22.0	8
6	Liberia	05-09-2014	1871.0	1089

USING MATPLOTLIB

- **Bar Graph**

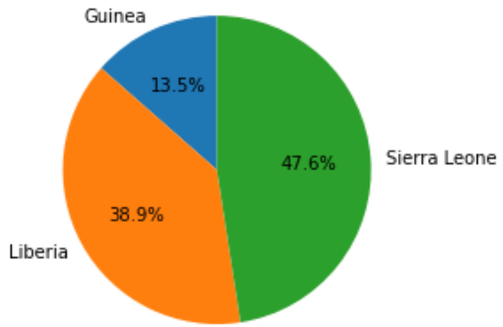
```
In [7]: c=ebola.head(100)
        plt.xticks(rotation='vertical')
        plt.bar(c["Country"],c["Cumulative no. of confirmed, probable and suspected cases"])
```

```
Out[7]: <BarContainer object of 100 artists>
```



- **Pie Chart**

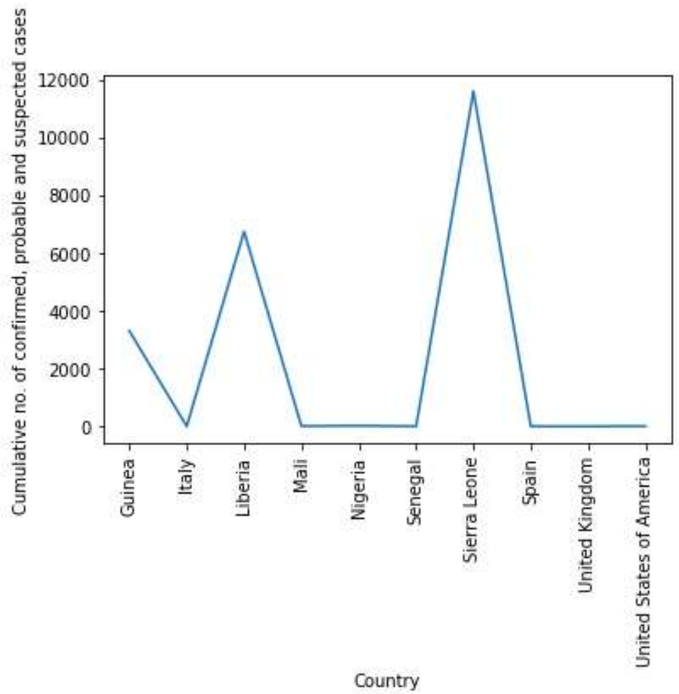
In [8]:
```
from matplotlib.pyplot import pie, axis, show
%matplotlib inline
df = pd.read_csv(r"C:\Users\Anoorag\OneDrive\Desktop\eboladatatset.csv")
sums = (df.loc[df["Cumulative no. of confirmed, probable and suspected cases"]>=500]).groupby(df["Country"])["Cumulative no.
axis('equal');
pie(sums, labels=sums.index, startangle=90, autopct='%.1f%%');
show()
```



USING SEABORN

- **Line Graph 1**

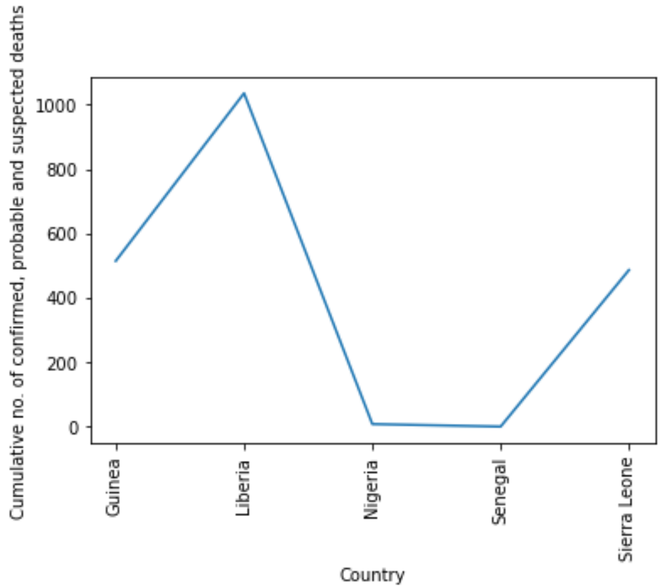
```
In [9]: plt.xticks(rotation=90)
        ax = sn.lineplot(x="Country", y="Cumulative no. of confirmed, probable and suspected cases", ci=None, data = ebola)
```



- **Line Graph 2**

In [10]:

```
c=ebola.head(20)
plt.xticks(rotation=90)
ax=sn.lineplot(x="Country", y="Cumulative no. of confirmed, probable and suspected deaths", ci=None, data=c)
```



In [11]:
```
print("THANK YOU")
```

```
THANK YOU
```