Data Exploration on African's Food Production and Supply (2004-2013)

Name: Toyese Williams.

Introduction

The food production and food supply systems of a country depend on many factors including government policy, ecological potential, and the level of technology, inputs and skills of the food producers. Food self-sufficiency has been a food policy objective of many African governments. However, while this objective may have been politically desirable, it has not always been economically and ecologically feasible. In some countries food self-sufficiency has been taken to mean adequacy of food supplies of locally produced staple cereals (e.g. maize, sorghum, rice) to meet the population's per caput energy requirements, but more often it has been seen as adequacy to meet market demand. This project is to learn the trend in the food production and supply in African (2004 to 2013) and gain better insights on inferring solution to food shortage in Africa.

Methodology

The food production dataset covers 45 countries, different food items (rice, beans, oats etc) and years ranging from 2004 to 2013.

```
production.head()
```

	Country	Item	Year	Value
0	Algeria	Wheat and products	2004	2731
1	Algeria	Wheat and products	2005	2415
2	Algeria	Wheat and products	2006	2688
3	Algeria	Wheat and products	2007	2319
4	Algeria	Wheat and products	2008	1111

```
supply.head()
```

	Country	Year	Value
0	Algeria	2004	2987
1	Algeria	2005	2958
2	Algeria	2006	3047
3	Algeria	2007	3041
4	Algeria	2008	3048

Data Visualization of the Food production and supply: Both datasets are presented in charts and graphs to show trends in the food production and supply.

Fig1

Food Production in Africa (2004-2013)
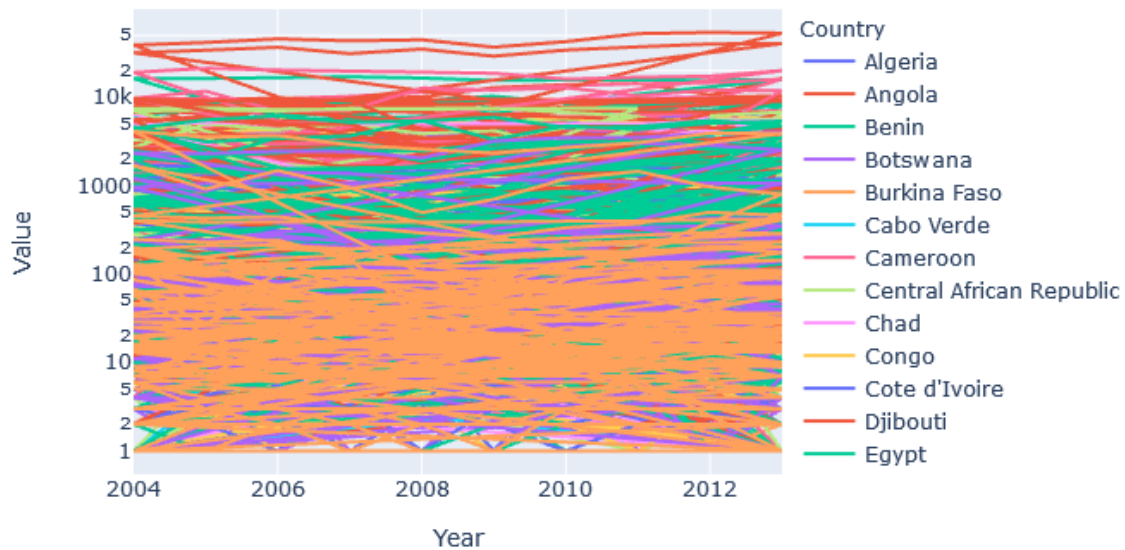
The fig1 plot shows that Nigeria is the highest food producer in African followed by Egypt. Therefore other African country can learn from the techniques and policies used by Nigeria to boost their food production.

Fig2

Food Supply in Africa

The fig2 food supply chart shows that Egypt, Tunisia and Morocco top the chart of highest food supplier in African and Zambia as the least supplier from 2004 to 2013.

The trend in the two charts (food production and supply production) shows that the two datasets are different, although Nigeria was the highest food producer, the case was different in food supplying as Egypt top the chart.

Outlier: Outliers are values that are significantly higher or lower than the rest in a dataset. The two dataset will explore more through visualization to further gain insight and detect outlier. The outlier should be removed before analysis for better results.

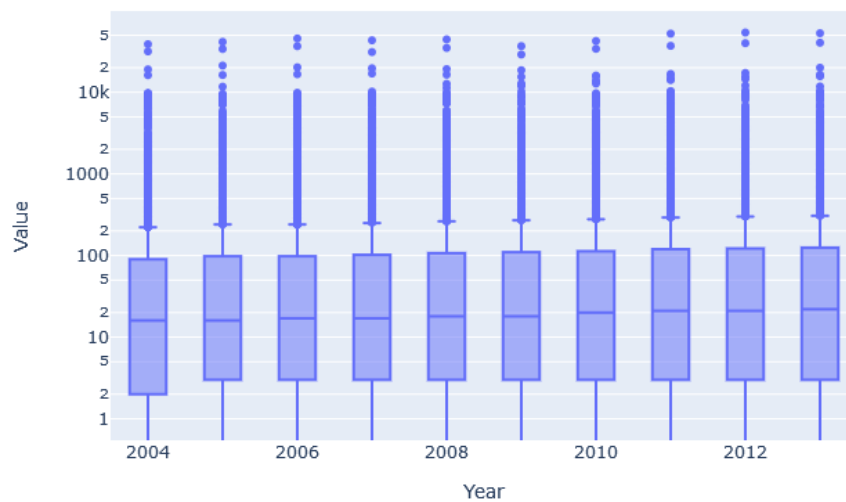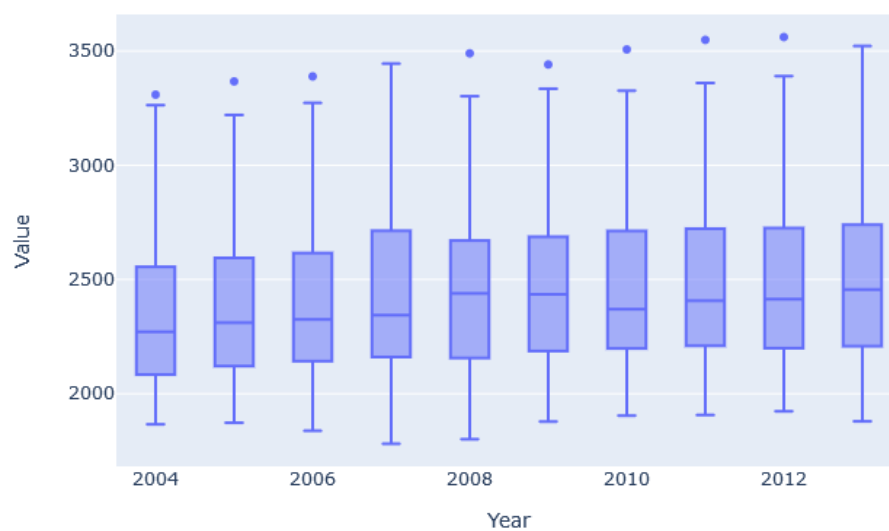Fig3. Outliers in Food production in African countries

Fig4. Outliers in food supply in African countries

The boxplot is a good chart in detecting outlier. Both fig3 and fig4 shows outlier in food production and food supply. The dots above the upper fence are the outlier. The outlier should be removed before analysis for better results.

It is easy to show outliers in each year but not in countries. The better way to show the countries having outlier is through interquartile. The Interquartile Range (IQR) is calculated as

$$Q_3 - Q_1$$

```
B=supply[supply.Value > (q3+1.5*iqr)]
B
```

	Country	Year	Value
123	Egypt	2007	3445
124	Egypt	2008	3490
126	Egypt	2010	3507
127	Egypt	2011	3549
128	Egypt	2012	3561
129	Egypt	2013	3522

Using Inter quartile we were able to detect the country with outlier in each country which is Egypt.

Correlation Test: We can also investigate how the food production and supply production influence the population in order to get insight on how both affect food shortage in Africa. This will be execute by conducting correlation (using PearsonCorrelationTest function) for an item (Rice).

Average value of rice production and population from 2004 to 2013 is as follows.

| Year	rice_prod_value	population_value
2004	326.789474	2.569406e+08
2005	348.763158	2.603340e+08
2006	379.026316	2.637517e+08
2007	361.500000	2.671895e+08
2008	420.289474	2.706754e+08
2009	405.842105	2.741823e+08
2010	457.315789	2.776952e+08
2011	454.157895	2.811942e+08
2012	482.394737	2.858398e+08
2013	489.578947	2.894748e+08

pearson correlation test.

	pearson correlation
p value	0.971281
statistic	0.000003

The correlation coefficient (0.97) is significantly far from 0. I.e., there is a significant correlation between mean rice production in a year, and the mean population of African countries for that year. This means that there's a very high positive correlation between X and Y. And such a small p-value suggests that that it is unlikely that H0 is true.

Fig6

Relationship between Rice production and Population in Africa (2004-2013)
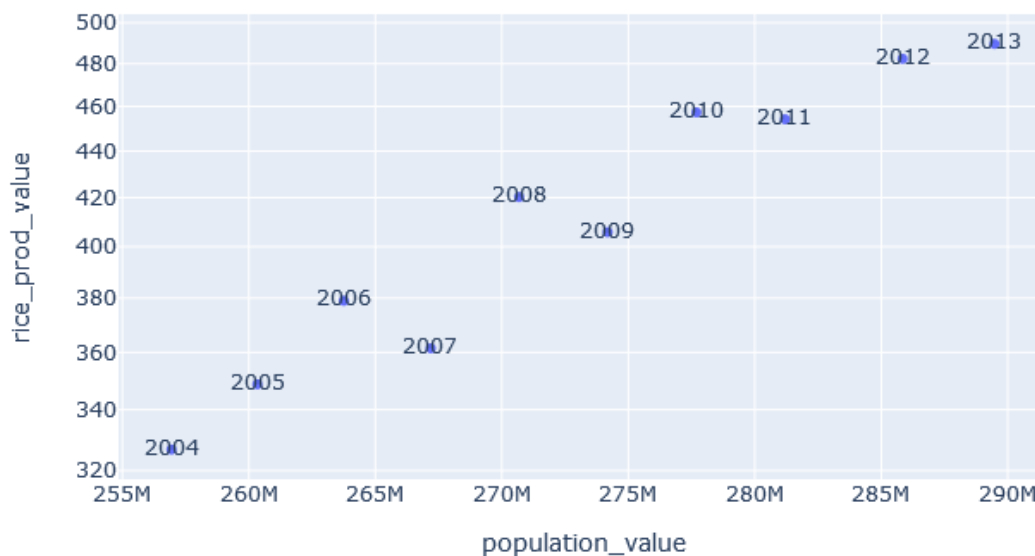
The fig6 chart shows that there is positive correlation between the rice production and the population.

Conclusion

The correlation coefficient (0.97) is significantly far from 0. I.e., there is a significant correlation between mean rice production in a year, and the mean population of African countries for that year. This may mean that the production of rice drives population, or vice versa. Also other African country can learn from the techniques and policies used by Nigeria and Egypt to boost their food production as both countries top the chart for highest food producer in Africa from 2004 to 2013.