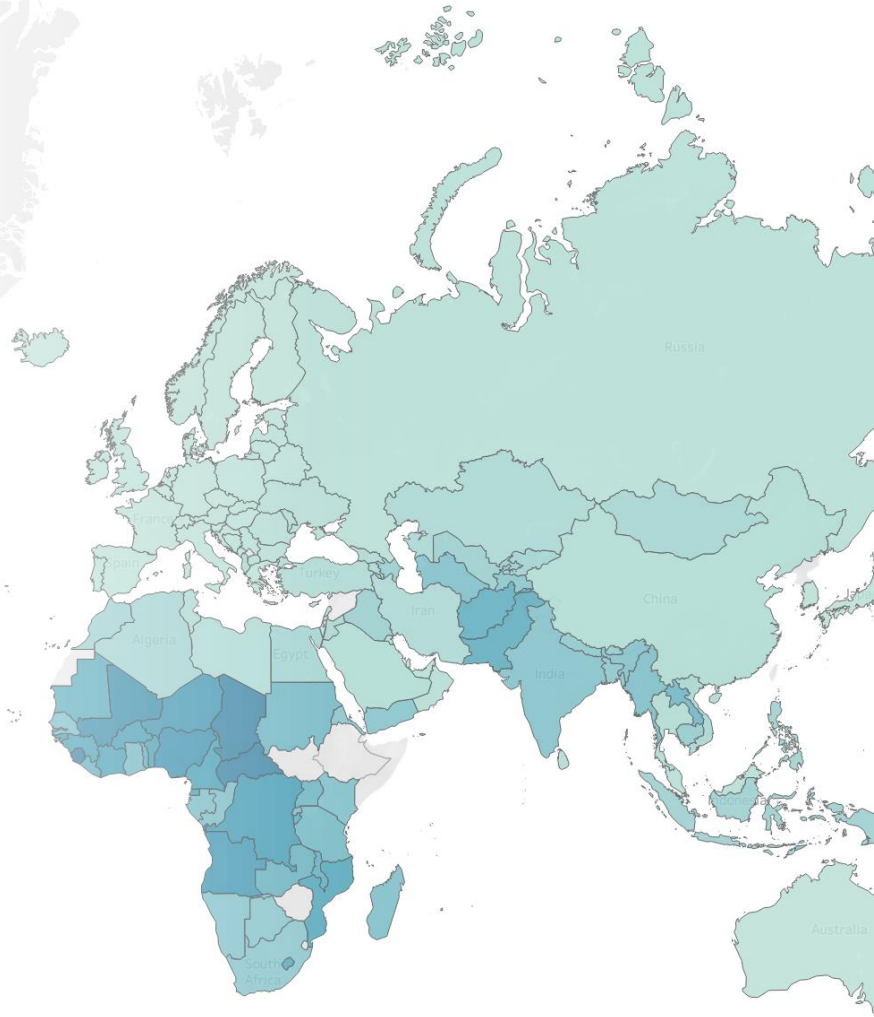
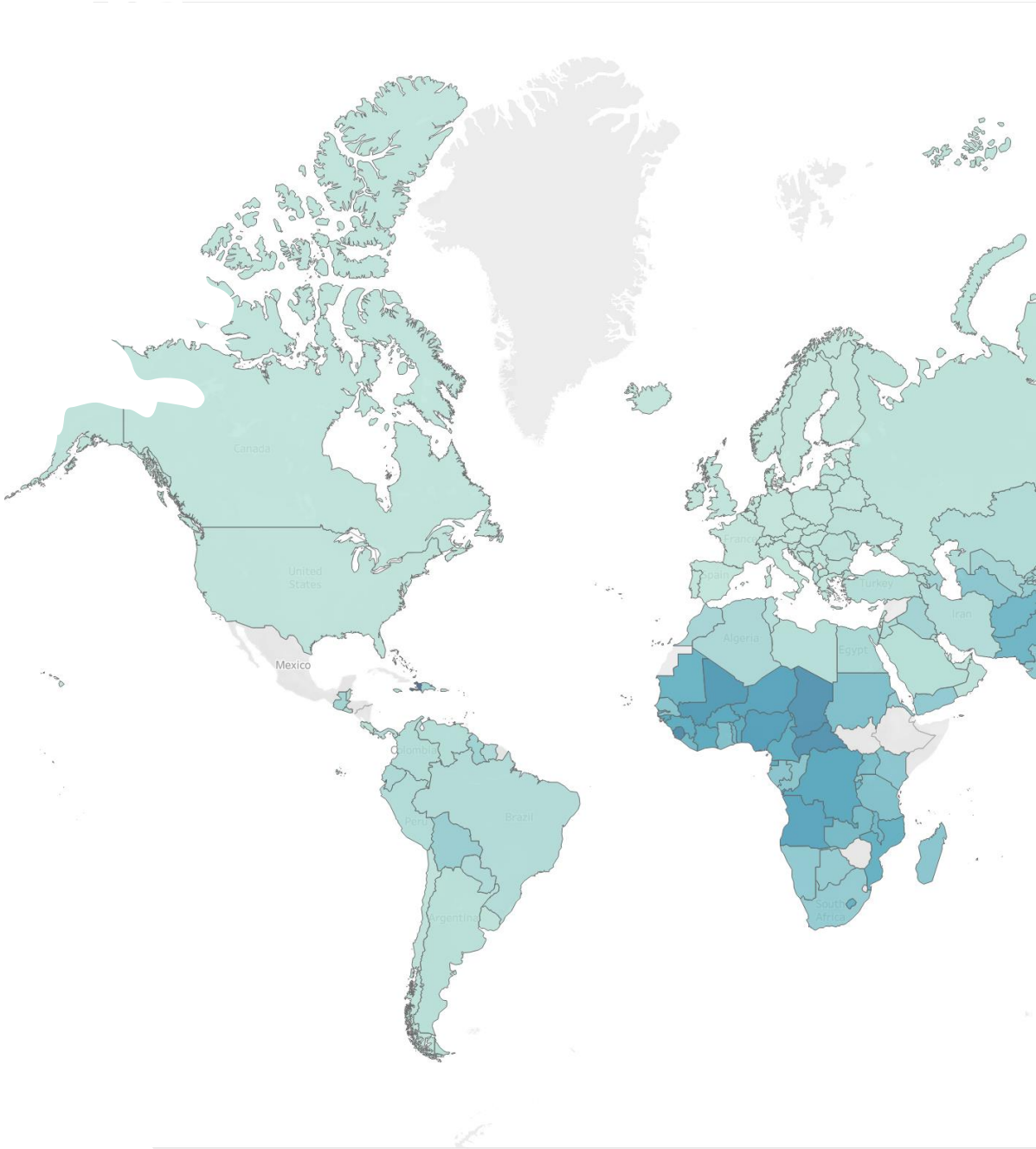
Clustering Analysis

for countries in direst need of aid



By Rahul Negi

This analysis presents the group of countries or categories the countries which will be in the 'direst need of aid' during the time of disaster or natural calamities, this is done through a rigorous analysis (clustering with the help of KMeans and Hierarchical clustering) on the basis of the socioeconomic and health factors that determine the overall development of country.



(generated). Colour shows sum of Child Mort. Details are shown for Country.

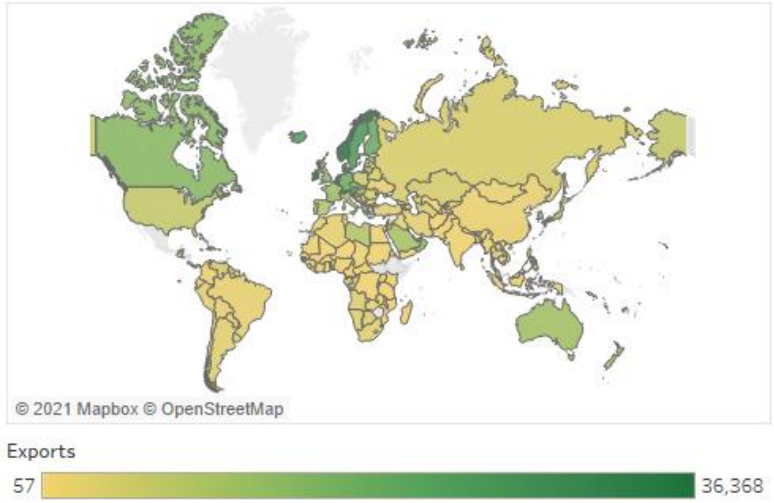
Data available for the list of countries

Afghanistan	Central African Republic	Guatemala	Madagascar	Romania	Uganda
Albania	Chad	Guinea	Malawi	Russia	Ukraine
Algeria	Chile	Guinea-Bissau	Malaysia	Rwanda	United Arab Emirates
Angola	China	Guyana	Maldives	Samoa	United Kingdom
Antigua and Barbuda	Colombia	Haiti	Mali	Saudi Arabia	United States
Argentina	Comoros	Hungary	Malta	Senegal	Uruguay
Armenia	Congo, Dem. Rep.	Iceland	Mauritania	Serbia	Uzbekistan
Australia	Congo, Rep.	India	Mauritius	Seychelles	Vanuatu
Austria	Costa Rica	Indonesia	Micronesia, Fed. Sts.	Sierra Leone	Venezuela
Azerbaijan	Cote d'Ivoire	Iran	Moldova	Singapore	Vietnam
Bahamas	Croatia	Iraq	Mongolia	Slovak Republic	Yemen
Bahrain	Cyprus	Ireland	Montenegro	Slovenia	Zambia
Bangladesh	Czech Republic	Israel	Morocco	Solomon Islands	
Barbados	Denmark	Italy	Mozambique	South Africa	
Belarus	Dominican Republic	Jamaica	Myanmar	South Korea	
Belgium	Ecuador	Japan	Namibia	Spain	
Belize	Egypt	Jordan	Nepal	Sri Lanka	
Benin	El Salvador	Kazakhstan	Netherlands	St. Vincent and the Grenadines	
Bhutan	Equatorial Guinea	Kenya	New Zealand	Sudan	
Bolivia	Eritrea	Kiribati	Niger	Suriname	
Bosnia and Herzegovina	Estonia	Kuwait	Nigeria	Sweden	
Botswana	Fiji	Kyrgyz Republic	Norway	Switzerland	
Brazil	Finland	Lao	Oman	Tajikistan	
Brunei	France	Latvia	Pakistan	Tanzania	
Bulgaria	Gabon	Lebanon	Panama	Thailand	
Burkina Faso	Gambia	Lesotho	Paraguay	Timor-Leste	
Burundi	Georgia	Liberia	Peru	Togo	
Cambodia	Germany	Libya	Philippines	Tonga	
Cameroon	Ghana	Lithuania	Poland	Tunisia	
Canada	Greece	Luxembourg	Portugal	Turkey	
Cape Verde	Grenada	Macedonia, FYR	Qatar	Turkmenistan	

Various parameters on which the countries are analyzed are listed down below.

- Child mortality
- Exports
- Health
- Imports
- Income
- Inflation
- Life Expectancy
- Total fertility
- GDPP

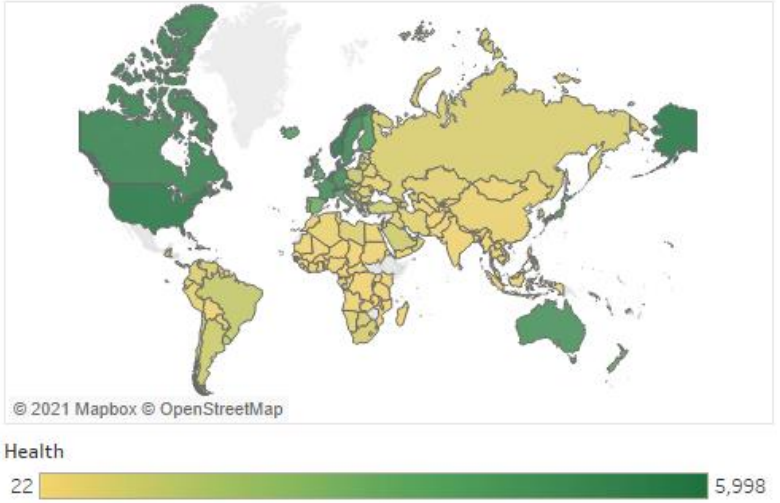
Exports



Top 10 countries with low exports:

- Central African Republic
- Myanmar
- Eritrea
- Burundi
- Afghanistan
- Nepal
- Liberia
- Sierra Leone
- Rwanda
- Niger

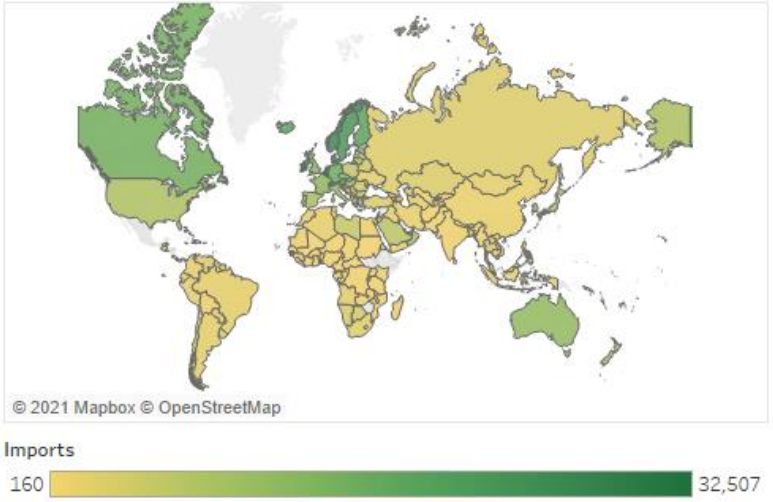
Health



Top 10 countries with low spending on health:

- Central African Republic
- Niger
- Myanmar
- Madagascar
- Eritrea
- Mozambique
- Pakistan
- Congo, Dem. Rep.
- Bangladesh
- Burundi

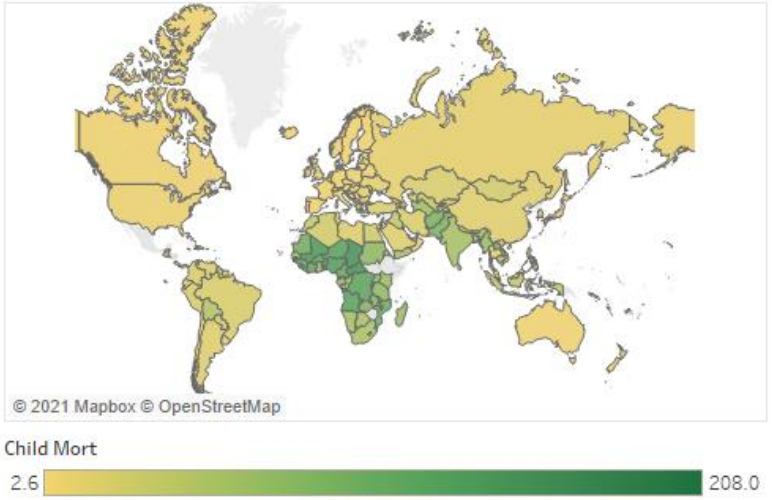
Imports



Top 10 countries with high imports:

- Belgium
- Singapore
- Ireland
- Switzerland
- Luxembourg
- Malta
- Netherlands
- Denmark
- Norway
- Austria

Child mortality



Top 10 countries with high mortality rate:

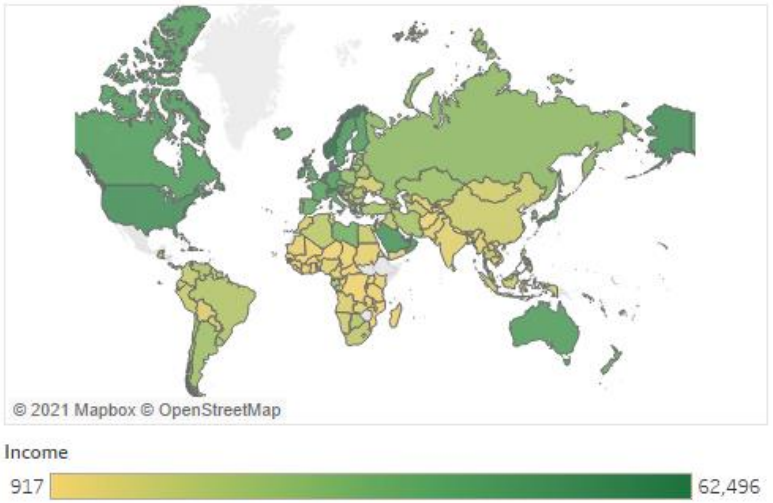
- Haiti
- Sierra Leone
- Chad
- Central African Republic
- Mali
- Nigeria
- Niger
- Angola
- Burkina Faso
- Congo, Dem. Rep

Life_Expectancy

Top 10 countries with low life expectancy:

- Central African Republic
- Myanmar
- Eritrea
- Burundi
- Afghanistan
- Nepal
- Liberia
- Sierra Leone
- Rwanda
- Niger

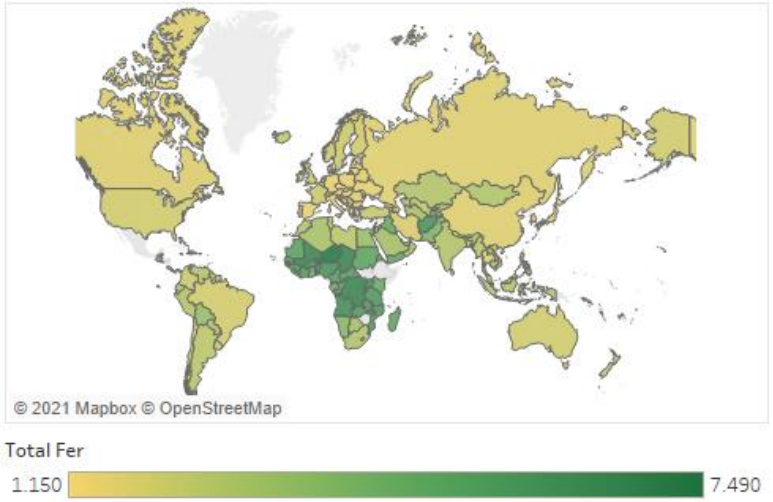
Income



Top 10 countries with low income:

- Liberia
- Niger
- Congo, Dem. Rep.
- Central African Republic
- Burundi
- Mozambique
- Malawi
- Guinea
- Togo
- Sierra Leone

Total_Fertilation



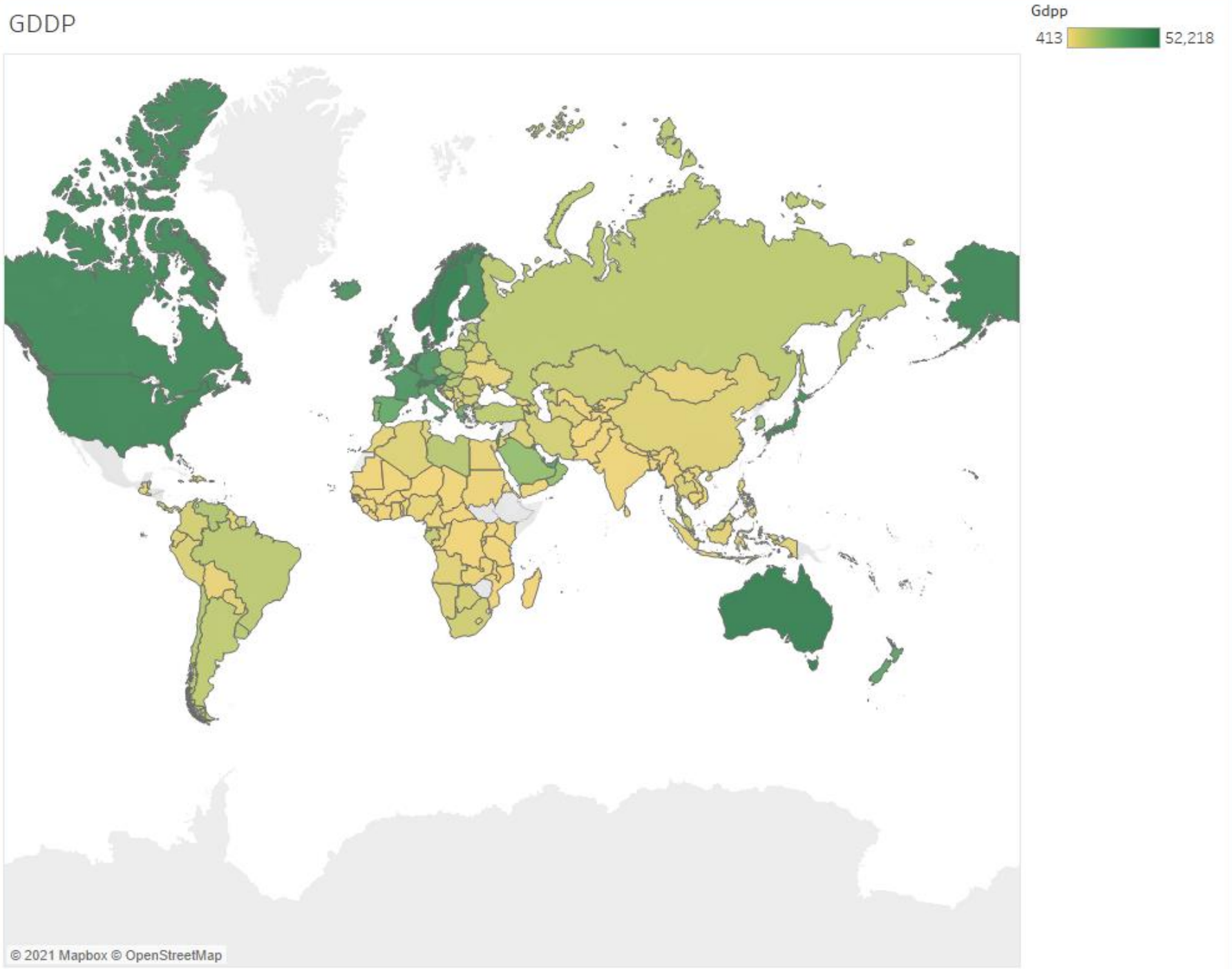
Top 10 countries with high fertility rate:

- Niger
- Chad
- Mali
- Congo, Dem. Rep.
- Burundi
- TimorLeste
- Angola
- Uganda
- Burkina Faso
- Nigeria

Inflation

Top 10 countries with high Inflation:

- Nigeria
- Venezuela
- Mongolia
- TimorLeste
- Equatorial Guinea
- Yemen
- Sri Lanka
- Angola
- Argentina
- Congo, Dem. Rep.

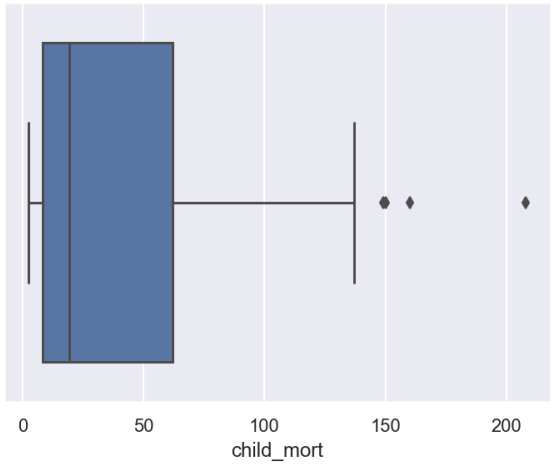


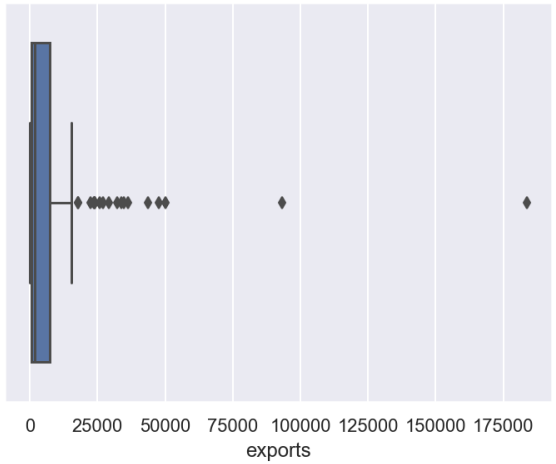
Top 10 countries with low GDPP:

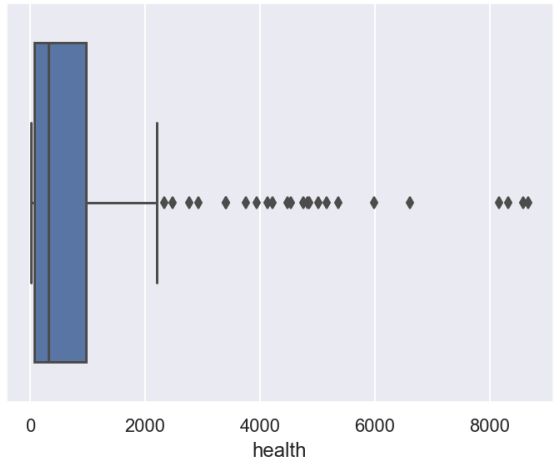
- Congo, Dem. Rep.
- Sierra Leone
- Burundi
- Liberia
- Niger
- Madagascar
- Mozambique
- Central African Republic
- Malawi
- Eritrea

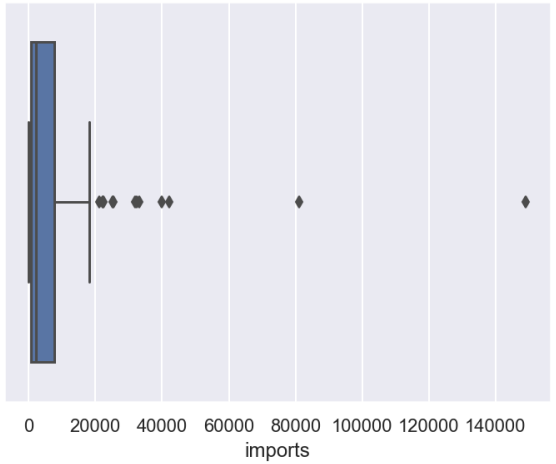
Boxplots

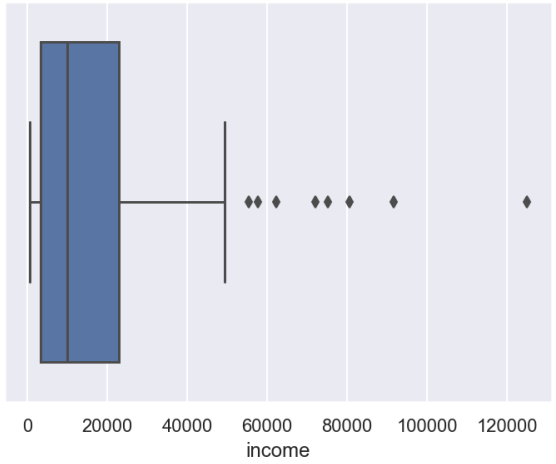
By looking at the plots for various parameters it can be seen that the some of the values are outside the whiskers (outliers) which are dealt while doing analysis.

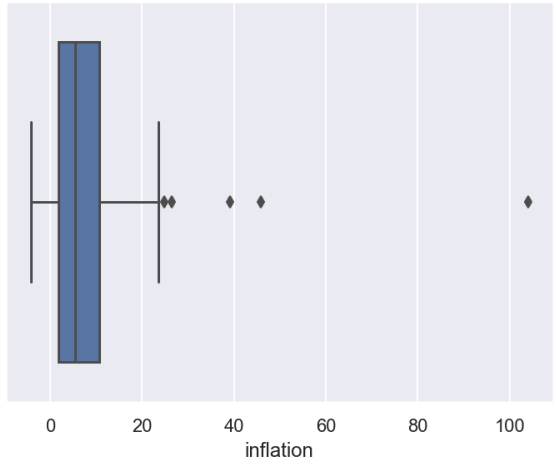


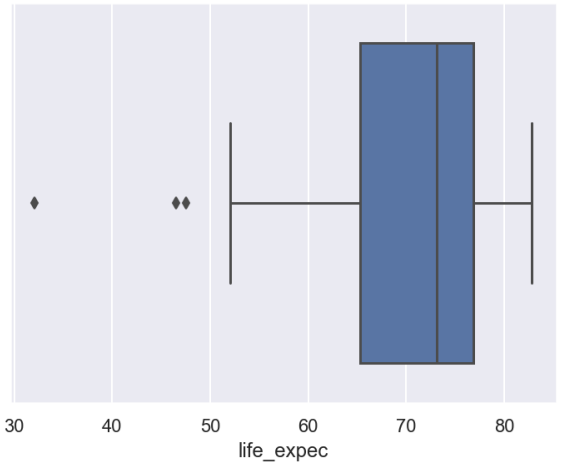


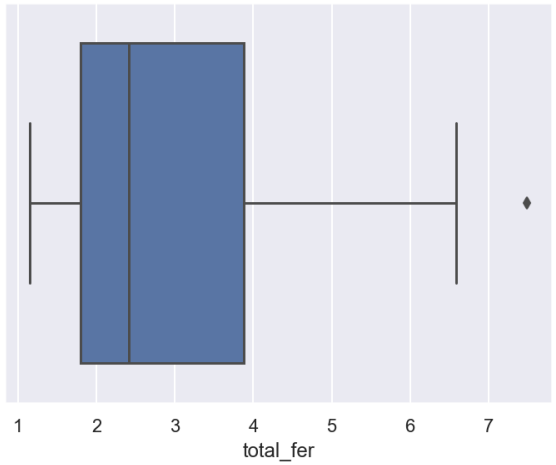


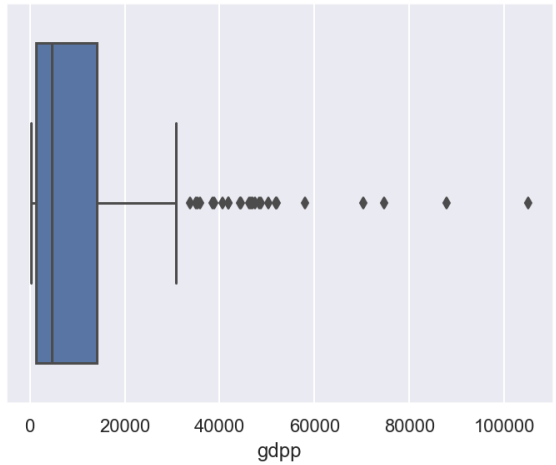








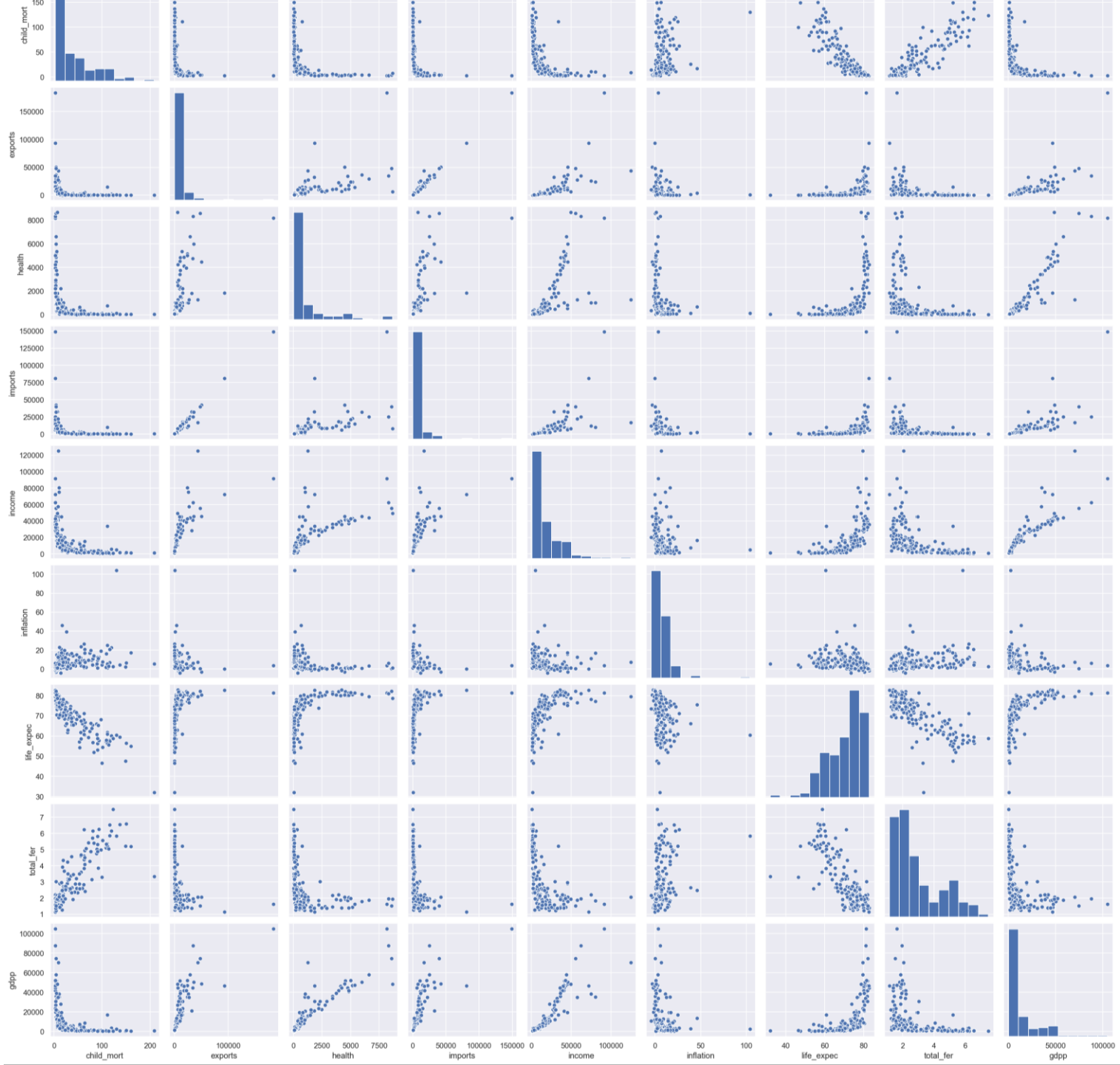




Pair-Plots for checking the variability of the variables with each other.

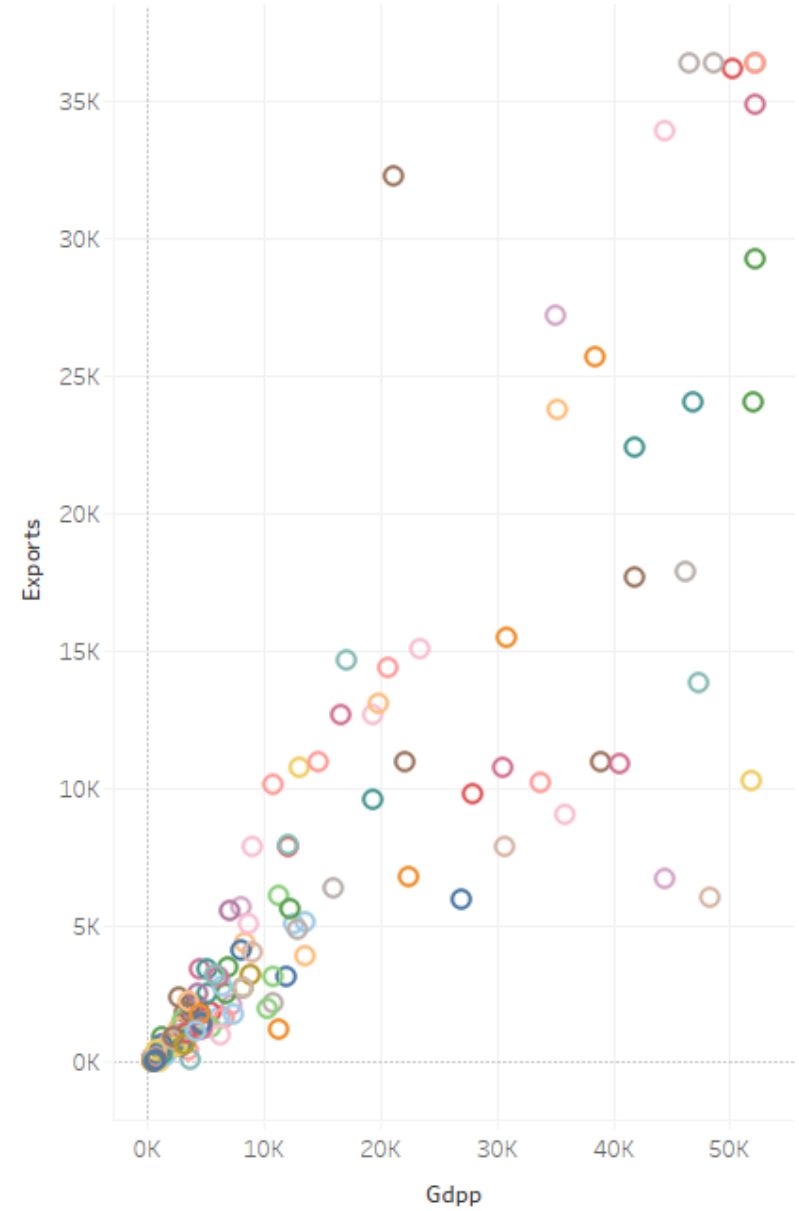
From the plots above it an been seen that some of the variables which are linearly varying with each other are listed down below:

- child_mort and life_expec
- child_mort and total_fer
- exports and imports
- exports and income
- exports and gddp
- health and gddp
- health and income
- imports and exports
- imports and gddp
- child_mort and total_fer



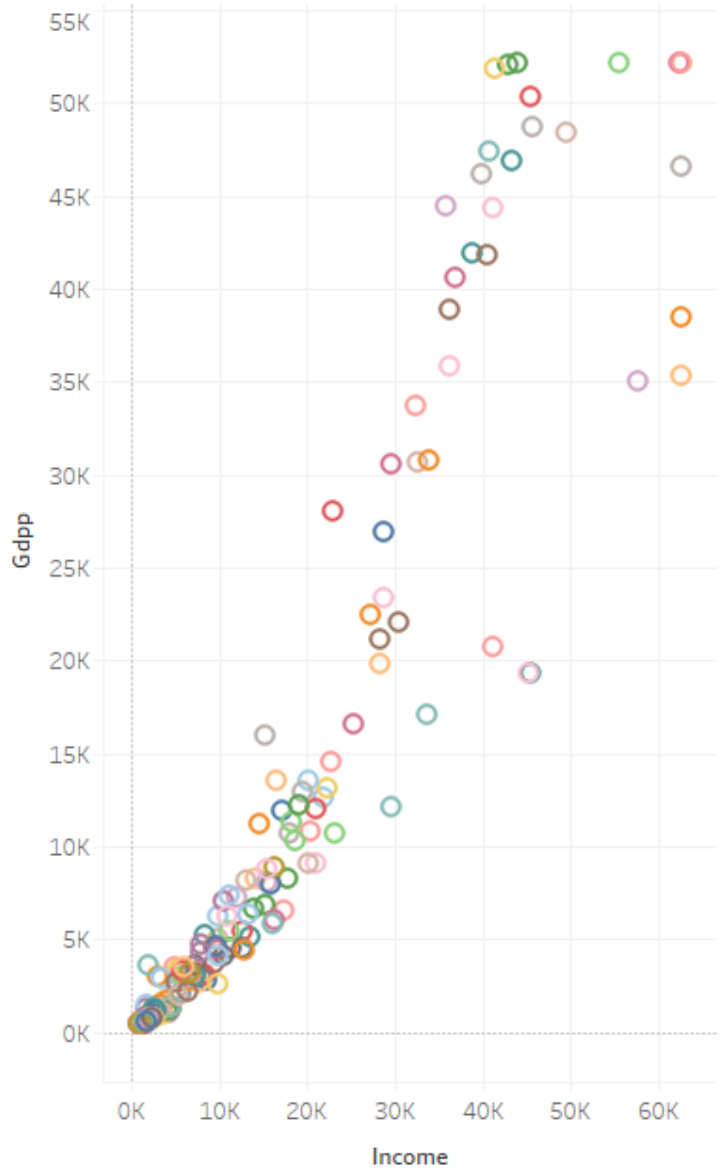
Exports vs GDDP

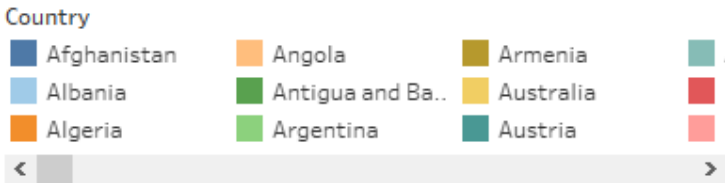
From the scatterplot it can be seen that ‘Exports’ and ‘GDDP’ are linearly varying and the bubble chart with larger circles shows higher Exports and the darker shade tells the higher GDDP.



Income vs Income

From the scatterplot it can be seen that ‘Income’ and ‘GDDP’ are linearly varying and the bubble chart with larger circles shows higher Income and the darker shade tells the higher GDDP.



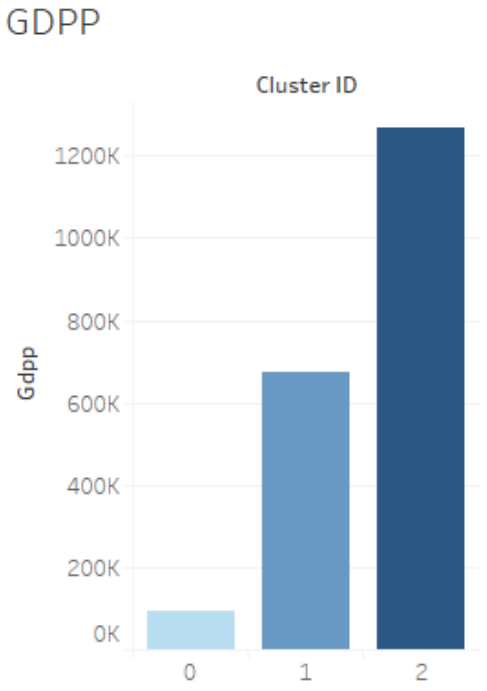


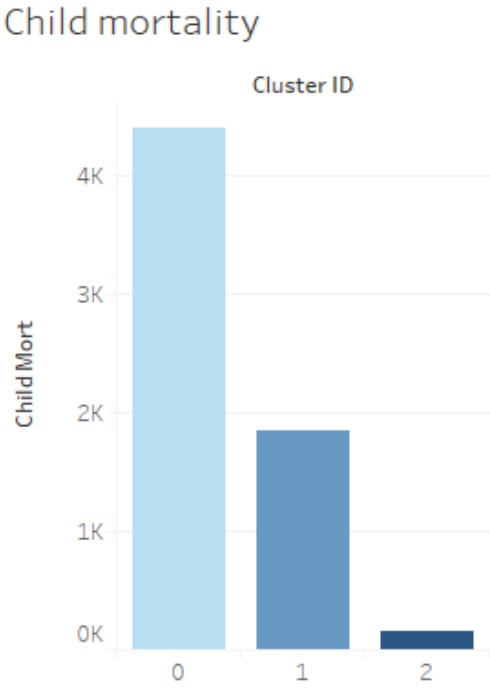
From the plots some of the inferences has been listed below:

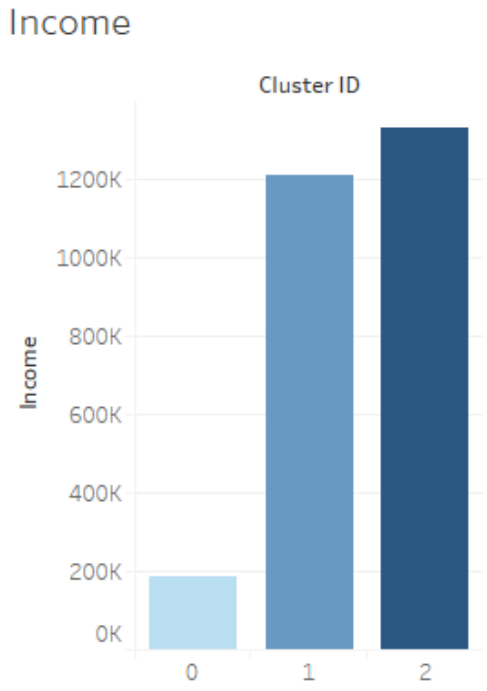
- Group-0: Low GDDP and Income whereas the Child mortality is high
- Group-1: Low GDDP and Child mortality but relatively high Income
- Group-2: High GDDP and Income whereas the Child mortality is low

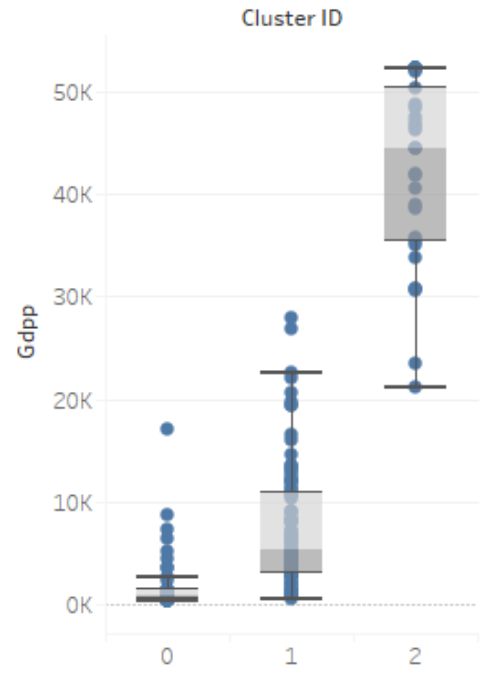
It can be seen that there has been a very clear distinction between three of the groups on the basis of the parameters chosen from the dataset, however these three type of countries can be divided in three groups,

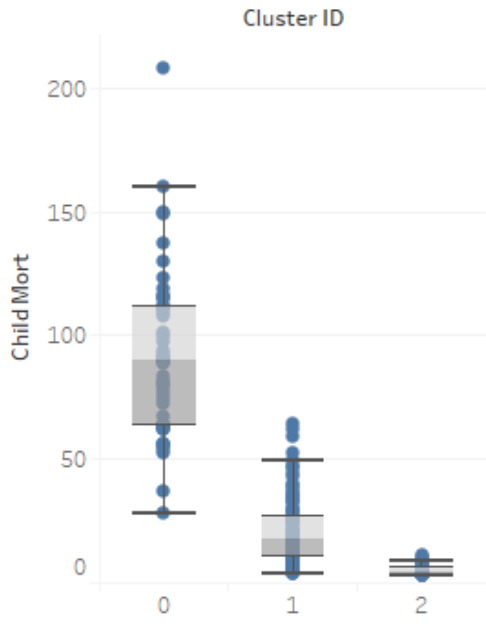
- Group-0: Developing countries
- Group-1: Emerging countries
- Group-2: High-income or Developed countries

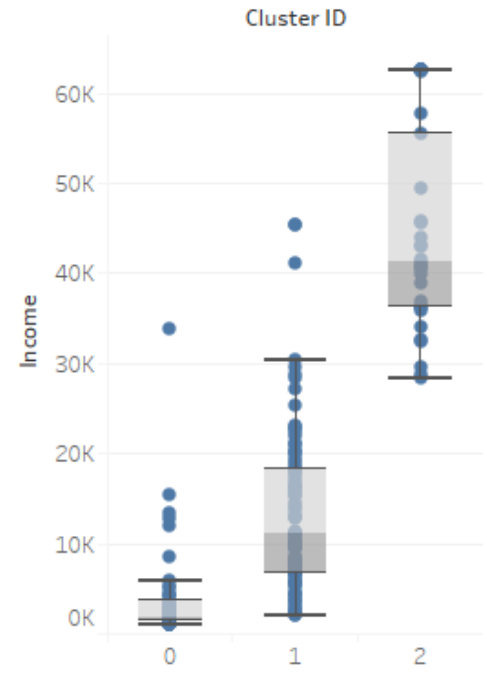








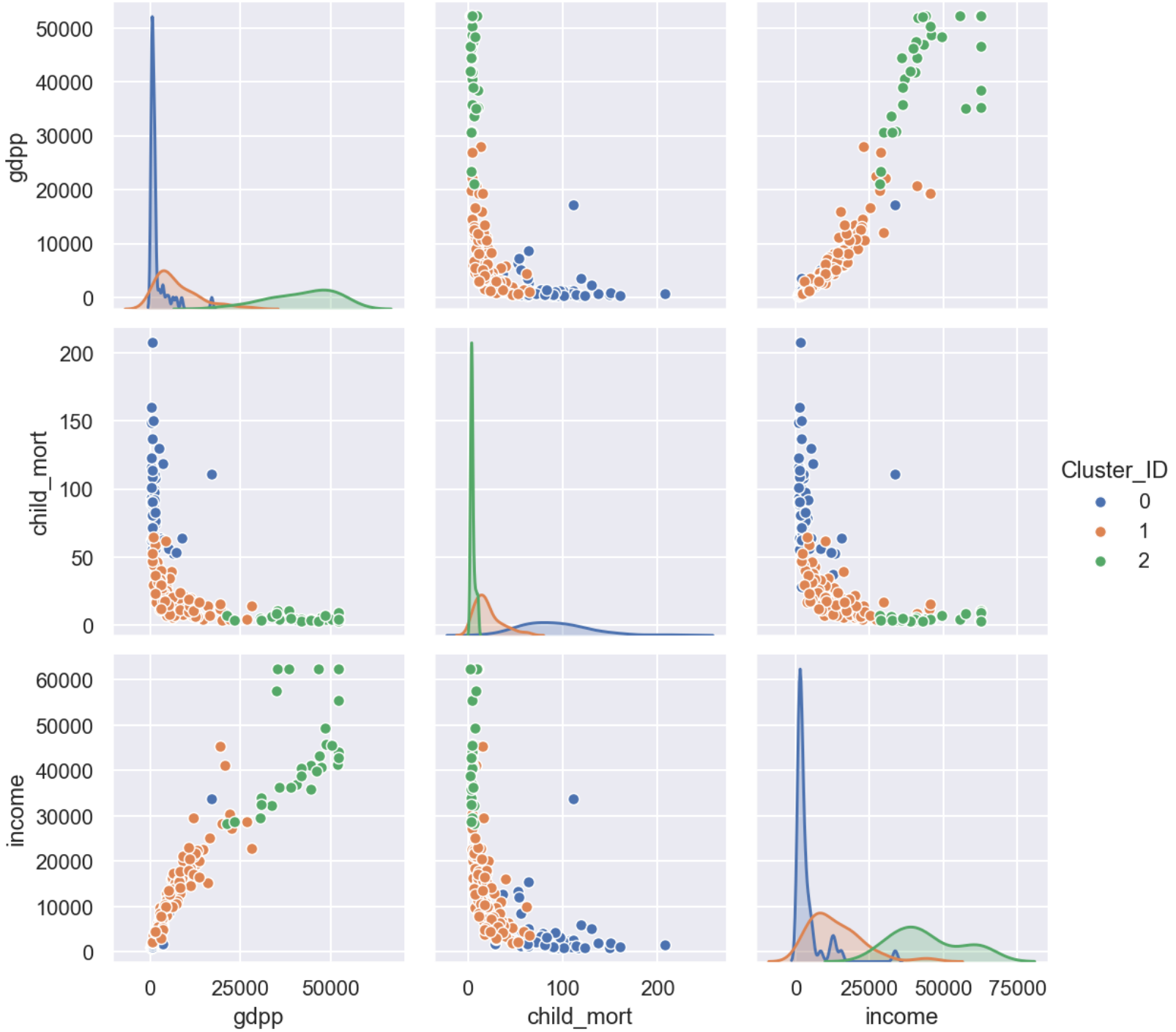




Some inference from the above pair plots:

- GDDP and Income have a liner relationship in which 'Group-2 or Developed is at the top' and 'Group-1' in the middle where as the 'Group-0' at bottom near the origin.

- GDDP and Child mortality shows the expected behaviour wherein the 'Group-0 or Developing countries' lies near the X-axis which mean that the GDP of the country is low and child mortality rate is high on contrary to that 'Group-2 or Developed countries' near the Y-axis and 'Group-1 or Emerging countries' in between



List of countries in Group-0 (Developing countries)

- Congo, Dem. Rep.
- Sierra Leone
- Burundi
- Liberia
- Niger
- Madagascar
- Mozambique
- Central African Republic
- Malawi
- Eritrea
- Togo
- Guinea-Bissau
- Afghanistan
- Gambia
- Rwanda
- Burkina Faso
- Uganda
- Guinea
- Haiti
- Tanzania
- Mali
- Benin
- Comoros
- Chad
- Kenya
- Senegal
- Pakistan
- Lao
- Lesotho
- Mauritania
- Cote d'Ivoire
- Solomon Islands
- Yemen
- Ghana
- Cameroon
- Zambia
- Sudan
- Kiribati
- Nigeria
- Congo, Rep.
- Angola
- Timor-Leste
- Iraq
- Namibia
- Botswana
- South Africa
- Gabon
- Equatorial Guinea

Map based on Longitude (generated) and Latitude (generated). Colour shows details about Country. The data is filtered on Cluster ID, which keeps 0.

List of countries in Group-1 (Emerging countries)

- Nepal
- Tajikistan
- Bangladesh
- Cambodia
- Kyrgyz Republic
- Myanmar
- Vietnam
- India
- Uzbekistan
- Moldova
- Bolivia
- Philippines
- Bhutan
- Egypt
- Mongolia
- Sri Lanka
- Guatemala
- Morocco
- Micronesia, Fed. Sts.
- Georgia
- Ukraine
- Vanuatu
- El Salvador
- Guyana
- Indonesia
- Armenia
- Paraguay
- Cape Verde
- Samoa
- Tonga
- Fiji
- Jordan
- Albania
- Tunisia
- Belize
- Turkmenistan
- Algeria
- Macedonia, FYR
- China
- Bosnia and Herzegovina
- Ecuador
- Jamaica
- Peru
- Thailand
- Serbia
- Dominican Republic
- Azerbaijan
- Belarus
- St. Vincent and the Grenadines
- Colombia
- Iran
- Montenegro
- Bulgaria
- Bulgaria
- Maldives
- Grenada
- Mauritius
- Panama
- Costa Rica
- Romania
- Suriname
- Lebanon
- Kazakhstan
- Malaysia
- Argentina
- Russia
- Turkey
- Seychelles
- Brazil
- Latvia
- Uruguay
- Lithuania
- Libya
- Antigua and Barbuda
- Poland
- Chile
- Hungary
- Venezuela
- Croatia
- Estonia
- Barbados
- Slovak Republic
- Saudi Arabia
- Oman
- Czech Republic
- Bahrain
- South Korea
- Portugal
- Greece
- Bahamas

Map based on Longitude (generated) and Latitude (generated). Colour shows details about Country. The data is filtered on Cluster ID, which keeps 1.

List of countries in Group-2 (Developed countries)

- Malta
- Slovenia
- Israel
- Spain
- Cyprus
- New Zealand
- United Arab Emirates
- Brunei
- Italy
- Kuwait
- United Kingdom
- France
- Germany
- Iceland
- Belgium
- Japan
- Finland
- Singapore
- Austria
- Canada
- United States
- Ireland
- Netherlands
- Australia
- Sweden
- Qatar
- Norway
- Luxembourg
- Denmark
- Switzerland

Map based on Longitude (generated) and Latitude (generated). Colour shows details about Country. The data is filtered on Cluster ID, which keeps 2.

Summary

In the above analysis for HELP international we have used the K-Means and Hierarchical clustering analysis for segregating out the countries and also finding out the countries that are in the direst need of the aid. From the analysis we have came to a conclusion that these countries can be divided into three groups

- Developed countries
- Emerging countries
- Developing countries

Even though the analysis is done though two types of clustering methods and the number of clusters are different for both the analysis (KMeans-3 clusters and Hierarchical clustering-4 clusters) but ultimately we have been getting the same countries form both the clustering methods which are listed below:

- Haiti, Sierra Leone, Chad, Central African Republic, Mali, Nigeria, Niger, Angola, Burkina Faso, and Congo, Dem. Rep.