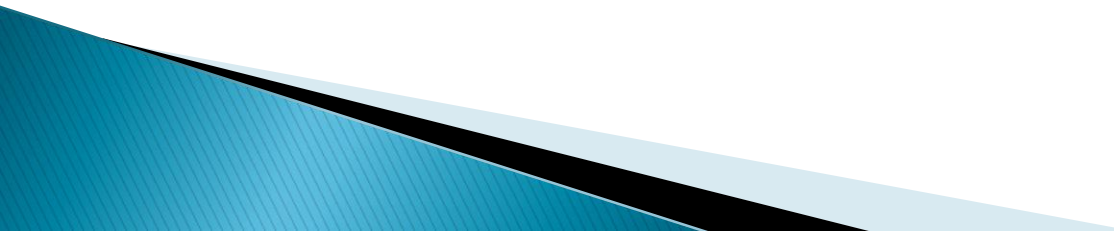# HELP International

## Identifying countries in need of Humanitarian Aid

Problem Statement

- The internationally renowned NGO "HELP International" needs suggestion about the list of countries in the direst need of aid, so that they can make a decision about how to make the most effective use of the $10 Million funding they have raised.

Approach for Data Analysis

▸ Analyze the "Countries" dataset which contains the socio-economic & health information of 167 countries. The objective is to segment these countries into clusters so that we can find the least performing countries based on the given factors, and decide the countries which are in dire need of financial & humanitarian aid.

Principal Component Analysis

- The features in the dataset are scaled using Standardization and then transformed into what is called Principal Components (PCs). A total of 9 PCs are obtained, but we can infer that 5 PCs explain 96% data variance. So we can proceed analysis with 5 components.
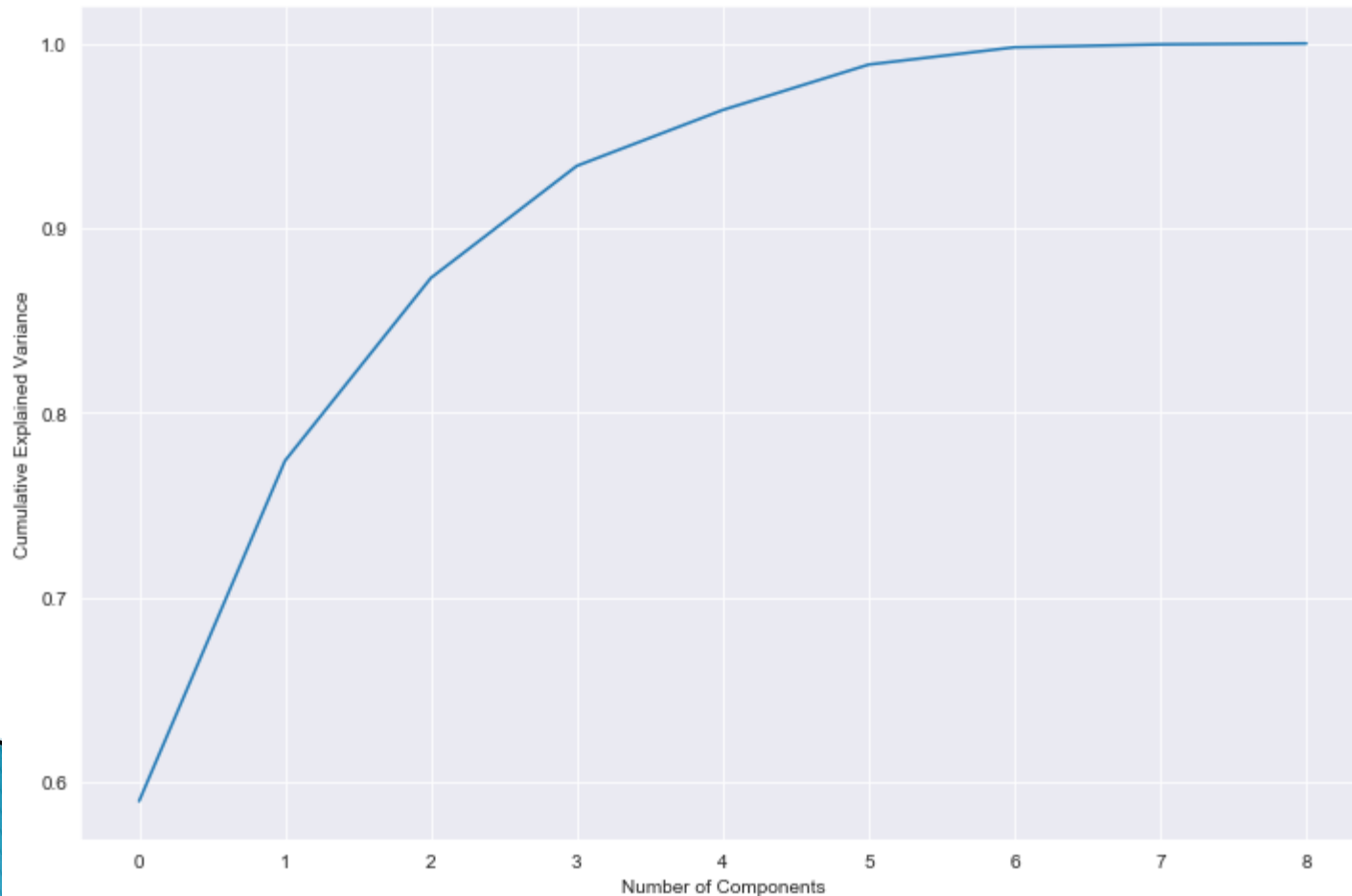
Visualisation of Original features vs two PCs
PC1 and PC2 factor loading for all 9 original features from the data set
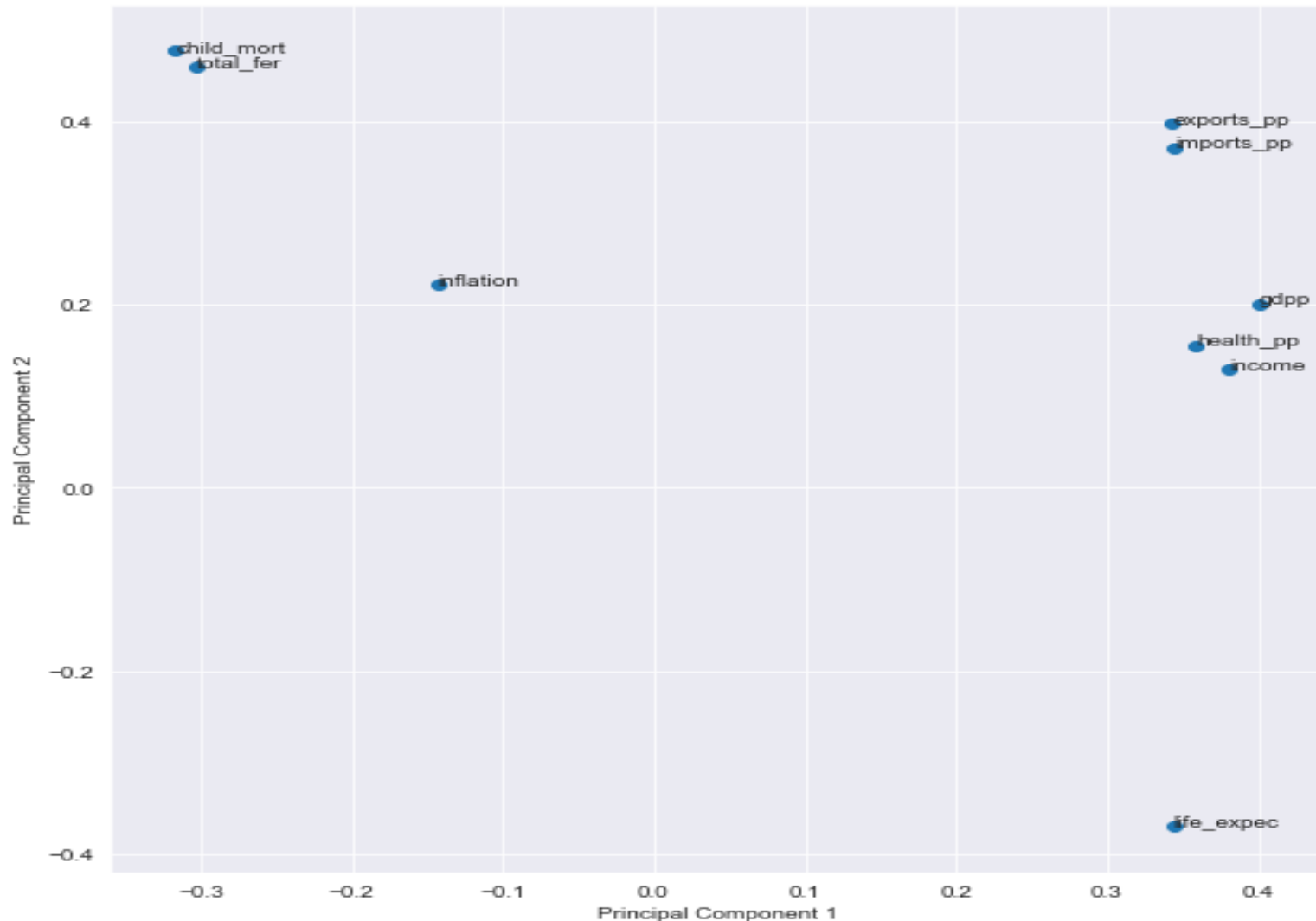
Clustering – is the assignment of a set of observations into subsets (called **clusters**) so that observations in the same **cluster**are similar in some sense.

- Two type of Clustering can be done in Machine Learning.
- – KMeans Clustering
- – Hierarchical Clustering (Single Linkage and Complete Linkage)
- Here, we will cluster the data using 3 clusters. We have arrived at this number based on what is called Elbow Curve and Silhouette score analysis.
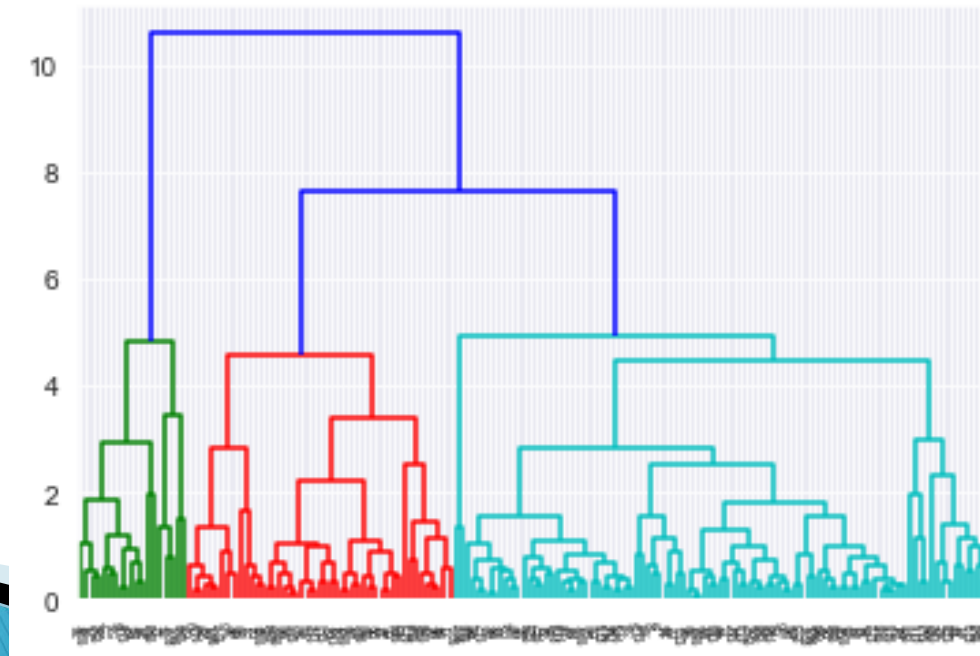- **ELBOW CURVE–**   **Silhouette Scores–**
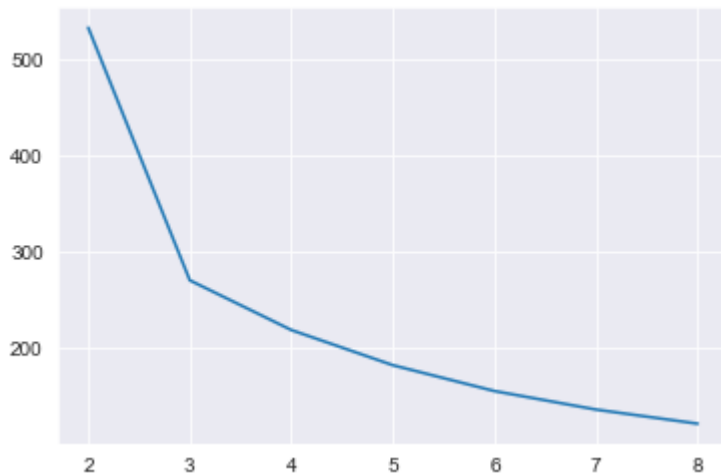

```
For n_clusters=2, the silhouette score is 0.40029890857536693
For n_clusters=3, the silhouette score is 0.4853499284355971
For n_clusters=4, the silhouette score is 0.39461875288487164
For n_clusters=5, the silhouette score is 0.36823953537189236
For n_clusters=6, the silhouette score is 0.36985483322443935
For n_clusters=7, the silhouette score is 0.32764900859260454
For n_clusters=8, the silhouette score is 0.36544359618358063
```
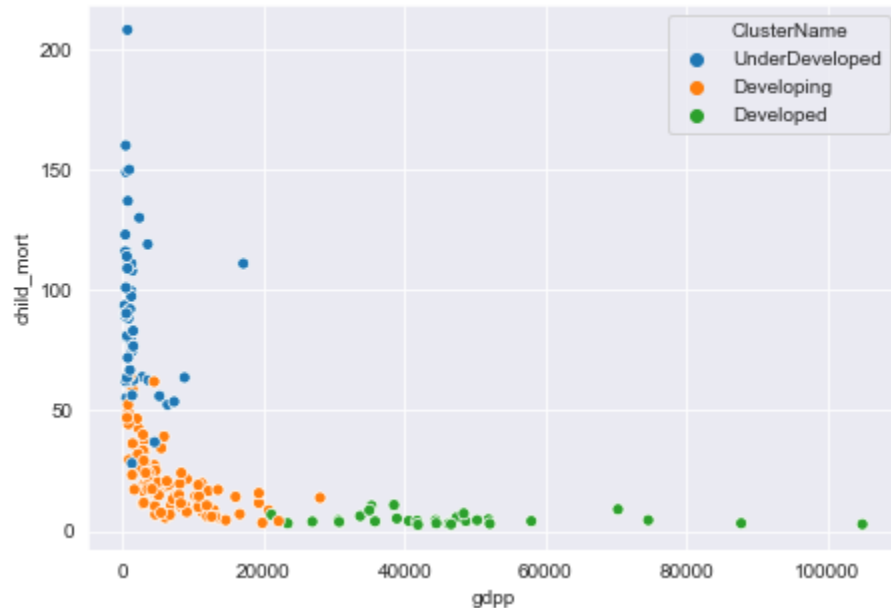
Dendrogram formed through Hierarchical Clusterin – Complete Linkage.

▸ Dendrogram is used to visualise the cluster formed as a part of Hierarchical Clustering.

▸ We can also decide the number of clusters we need and cut the dendrogram such we can precisely get the clusters as desired.

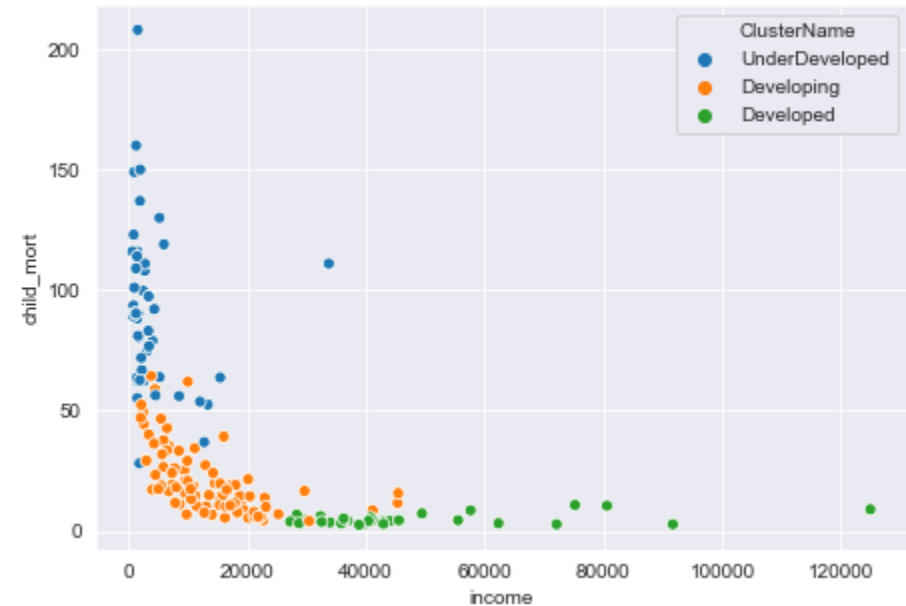▸ In terms of Clusters, KMeans and Hierarchical have given almost identical results for our data.

Visualising the data distribution in Clusters formed through our analysis

- ▸ **From the above scatter plot, it is clearly visible that:**
- ▸ The countries with high Child Mortality Rate and very low income/GDPP are clustered as "UnderDeveloped"
- ▸ The countries with relatively lower Child Mortality Rate and medium income/GDPP are clustered as "Developing"
- ▸ The countries with very low Child Mortality Rate and high income/GDPP are clustered as "Developed"

Visualising the data distribution in Clusters formed through our analysis – Principal Components


PC1 vs. PC2 - Cluster Visualisation
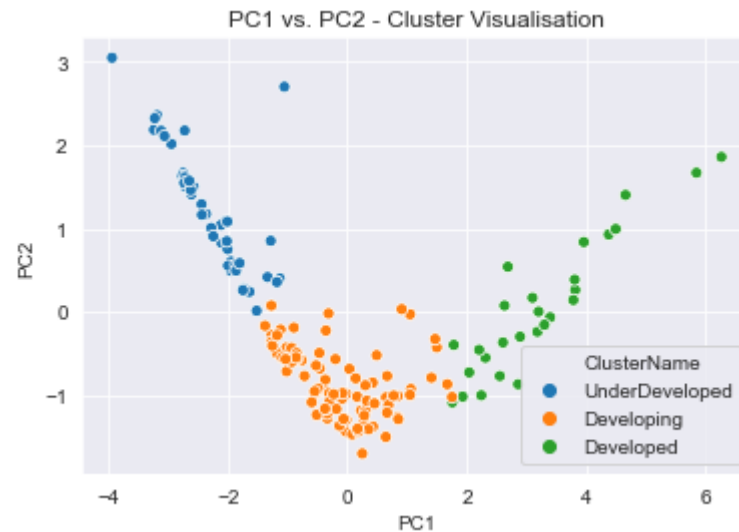
▸ The two Principal Components have done very clear representation of the Clusters to be formed. "UnderDeveloped" cluster has high PC2 and low PC1. "Developing" cluster has low PC2 and PC1 values. "Developed" cluster has high PC1 and PC2 values.

Factor Distribution across Clusters

From the box plots seen below it is clearly evident that the UnderDeveloped cluster has the highest average Child Mortality Rate, lowest GDPP and INCOME among all 3 clusters.

But as we already have 48 countries in this cluster, we will have to identify the list of countries that are performing poorly in all 3 factors, not just 1 or 2 of them. We will be doing further analysis in the next part and arrive at the final list countries to be recommended.
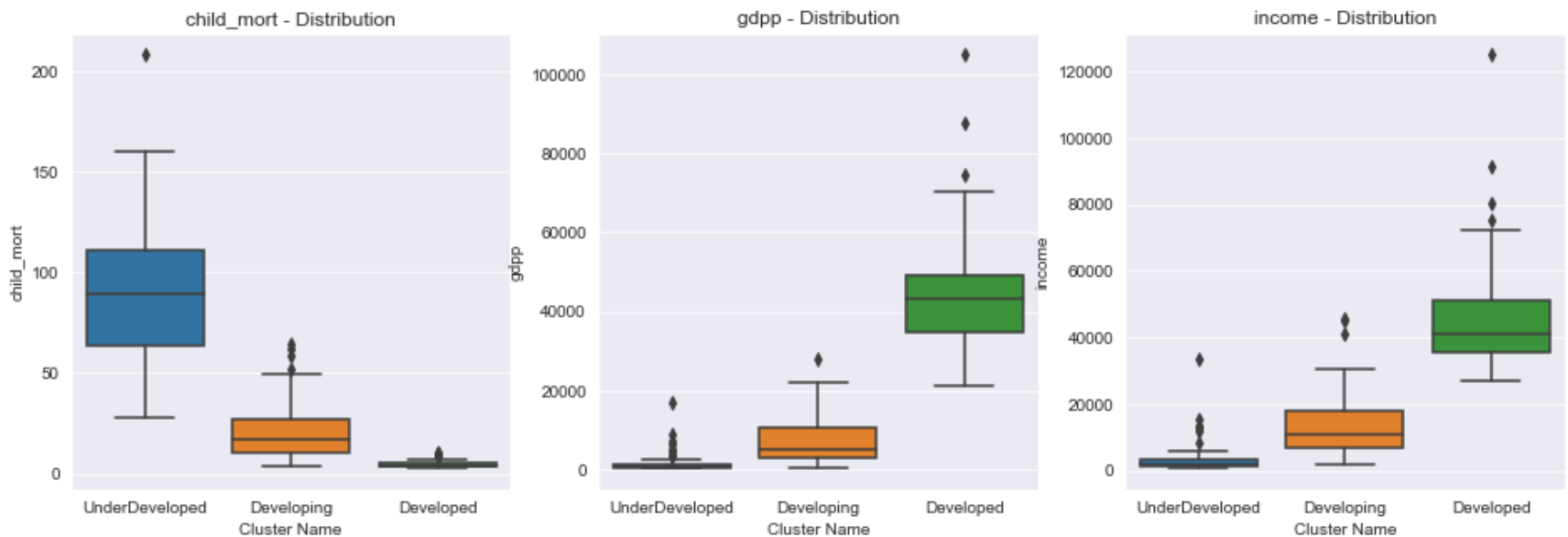
Identify the UnderDeveloped countries with inferior GDPP & Income, and higher Child Mortality rate - when compared to the mean values of the "UnderDeveloped" Cluster countries.

INCOME vs. Child Mortality Rate - Plot of all countries needing aid
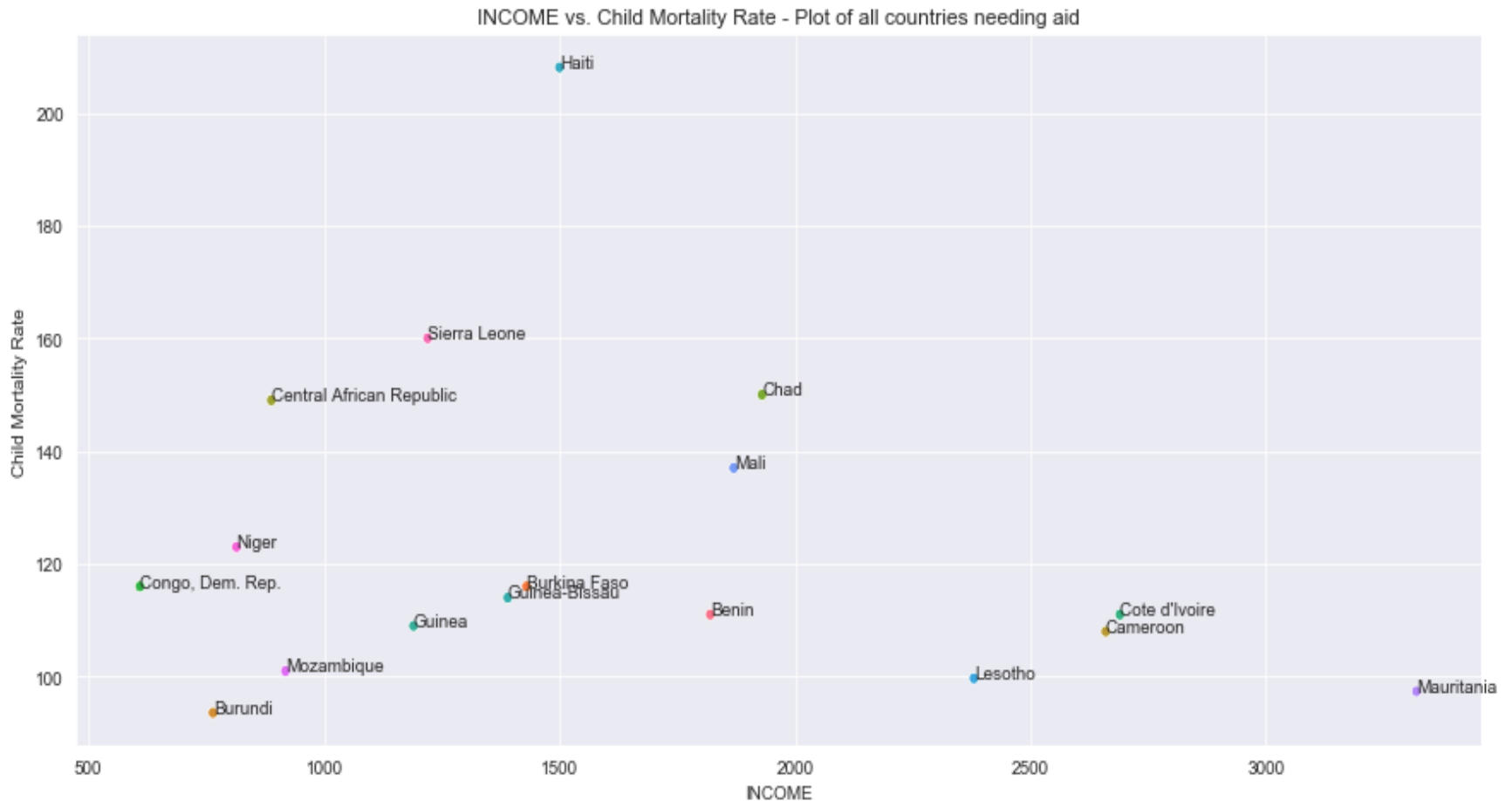
Conclusion: The 17 countries listed below are in dire need to Financial & humanitarian aid to battle high Child Mortality Rate & low Income/GDPP

- Haiti, Sierra Leone, Chad, Central African Republic, Mali, Niger, Congo Dem. Rep., Burkina Faso, Guinea-Bissau, Benin, Cote d'Ivoire, Guinea, Cameroon, Mozambique, Lesotho, Mauritania, Burundi

country	gdpp	income	child_mort
Haiti	662	1500	208.0
Sierra Leone	399	1220	160.0
Chad	897	1930	150.0
Central African Republic	446	888	149.0
Mali	708	1870	137.0
Niger	348	814	123.0
Congo, Dem. Rep.	334	609	116.0
Burkina Faso	575	1430	116.0
Guinea-Bissau	547	1390	114.0
Benin	758	1820	111.0
Cote d'Ivoire	1220	2690	111.0
Guinea	648	1190	109.0
Cameroon	1310	2660	108.0
Mozambique	419	918	101.0
Lesotho	1170	2380	99.7
Mauritania	1200	3320	97.4
Burundi	231	764	93.6