Child Mortality: An exploratory study of trend around the world over the last two decades and the impact of wealth on this trend

1. Motivation, data, research questions

1.1 Inspiration

Child mortality is defined as the death of children under the age of fourteen. It is measured as the number of deaths per 1000 live births and can be classified into:

a. Neonatal Mortality Rate (NMR) - death under 28 days
b. Infant Mortality Rate (IMR) - death under one year
c. Under Five Mortality Rate (U5MR) - death under five years of age
d. Under 14 Mortality Rate (U14MR) - death under 14 years of age

Millions of children die around the world every day and the leading causes of death are pre-term birth complications, disease and birth abnormalities according to World Health Organisation (WHO) [2]. NMR deaths account for 38% of the deaths followed by IMR (25%), U5MR (22%) and lastly U14MR (15%). In actual numbers 2.6 million children died around the world within the first month in the year 2016, 1.7 million under 1 year and 1.5 million under five years of age. This study is focused on NMR, IMR and U5MR trends around the world.

1.2 About the data

Historical NMR, IMR and U5MR data was obtained from United Nations International Children's Emergency Fund (UNICEF) [1], an organisation that collects data to track, assess and monitor progress of women and children around the world using Multiple Indicator Cluster Surveys (MICS).

1.3 Data Wrangling and Dimension Reduction

Data contains CMR from the year 1956 till 2016, however, this study focuses on the last twenty one years i.e. from 1996 to 2016, where the data is dense and does not need null value imputations. The estimates are further qualified with uncertainty bounds of lower, median and upper margins.

1.4 Additional Data

Gross Domestic Product (GDP) of the countries was sourced from The Organisation for Economic Co-operation and Development (OECD) [3]. GDP is the measure of export minus import in US dollars at current prices and is an indicator of economic well-being of a country. The data is limited to the last twenty one years so that it corresponds to the NMR/IMR/U5MR data timeframe. Missing values were imputed by time series interpolation.

Latitude and longitude of countries was obtained from GitHub. This was merged with NMR/IMR/U5MR data and GDP data to help with visualisation on the world map.

1.5 Data Derivatives

Additional attributes were created to help summarise the data over several years for a country and across countries for a year. This data helped with plotting trends for the world as whole.

1.6 Programming Language and Tools

Python was the programming language of choice because of its rich libraries for both statistical analysis and visualisation. Numpy and Pandas was used for data wrangling, StatsModel and SciPy were used extensively for statistical analysis. Plotly and MatplotlIb were the primary packages used for visualisation.

1.3 What is Being Explored?

Historical data about NMR/IMR/U5MR provides many insights in to the child mortality rates (CMR) around the world. Specifically, this study explores:

1. Study the trend of CMR over the last twenty one years around the world and understand if any spikes can be explained with war, disease or natural disasters.
2. Can CMR of a country be explained by GDP of the country and if so to what degree?

2. Tasks and Approach

2.1 Exploring the Data

This study starts off with some simple bar charts to understand the shape of the data and to gain knowledge of the CMR distribution.

2.1.1. Study of the uncertainty bounds in each category

Worst affected countries:

A grouped bar graph of the top 10 countries with the most mortality rates for NMR, IMR and U5MR are shown in Fig. 1. Three separate graphs show the worst affected countries for each category.

NMR: *Pakistan* has the worst mortality rates in lower and median bounds and *South Sudan* for upper bounds.

IMR: *Sierra Leone* has the most child deaths in lower and median bounds and *Somalia* for upper bounds.

U5MR: Once again *Sierra Leone* is the most affected in lower and median bounds and *Somalia* for upper bounds of CMR.

Least affected countries:

Fig. 2 shows another set of three grouped bar graphs with the countries that have the least mortality rates.

NMR: *San Marino* has the least mortality rates in lower and median bounds and *Japan* has the least rate for upper bounds.

IMR: *Iceland* has the least mortality rates in lower and median bounds and *Singapore* has the least rate for upper bounds.

U5MR: *San Marino* has the least mortality rate in lower bound and *Singapore* has the least in median and upper bounds.

2.1.2 Study of the trend over last two decades:

The cumulative CMR of all the countries over the last two decades is mapped as a time series plot in Fig. 3. There has been a dramatic fall in NMR, IMR and U5MR over time. The arrows show the swing of the upper and lower bounds from the median. The circle marks the noticeable bump in the year 2009.

2.1.2 Study of trend of individual countries over last two decades:

The data as a whole may show strong trends, but we may miss the subtler trends as shown with Gapminder data by Hans Gosling [8]. It is easy to miss these spikes when seen as numbers, so we need to plot the countries individually to visually inspect abnormality in trends.

Time series trend line plot for every country for median bound over the years is shown in Fig. 4. Each line represents a country. Over the years the rate can be seen to be going down, but the plot clearly shows some abnormalities. There are some spikes in NMR, IMR and U5MR. Isolating these spikes will show us the countries which buck the downward trend. Some countries even have an upward trend. Further investigation with GDP might reveal the reasons for the abnormality.

Haiti had a spike in 2010. When we dig deeper, Haiti had an earthquake of the magnitude of 7.0 causing upwards of 160,000 deaths. A further study by Chen and colleagues [11] of the effects of displacement on CMR concluded that there is definitely an increase in CMR in temporary camps.

Cumulative data does not always make irregularities obvious. This leads us to check if the spikes exist in lower and upper uncertainty bounds and so we plot all three bounds as shown in Fig. 5. The spikes appear in all uncertainty bounds.

2.1.3 Compare NMR/IMR/U5MR rates for the worst affected countries:

As per UNICEF [1] literature, NMR rates account for a much larger portion of the CMR than IMR and U5MR. In this study, a separate count of deaths from 0 to 28 days, 28 days to one year, and one year to five years is plotted to see if we can see a trend in these counts for the worst affected countries. Fig. 6 shows this grouped bar chart. Only two countries follow the literature, the rest do not.

2.1.4 Study of the overlap of uncertainty bounds for all categories:

Uncertainty bounds are calculated based on the MICS survey confidence interval set at 95%. Fig. 7 shows the overlap of the uncertainty bounds for NMR / IMR and U5MR. The true value might be between either of these bounds and is determined by the random sample picked for the survey. We note that the upper bound has a larger residual value than the lower bound and the overlaps might be significant.

2.1.5 Choropleth map of the world with CMR:

Choropleth map of the world (Fig. 8) showing CMR reveals that countries in *Africa* are most affected followed by South Asian countries (*India*, *Pakistan*, *Afghanistan*, *Cambodia* and *Indonesia*) and *Bolivia* in South America. The patterns are similar for NMR, IMR and U5MR.

3. Analytical steps

The irregularities in the plots help us to focus on the countries that are affected. We now begin to explore the relationship between GDP and CMR.

3.1 Relation between GDP of a country and CMR:

A scatter plot of NMR, IMR and U5MR against GDP is shown in Fig. 9. The graph line shows an inverse relationship – higher the GDP lower the CMR and the converse. The head of the curve (1) is *Luxembourg* with a GDP of 1.5 million dollars and has 42 as the cumulative CMR over 20 years. On the other hand *India* is at the tail of the curve (2) with a mere 150,000 dollars as GDP but over 1500 as the cumulative CMR. This is closely followed by *South Africa* with 216,000 as GDP but with 1275 as the cumulative CMR. Countries with high GDP (3) have low CMR and this includes *United States, Sweden, New Zealand and Japan*. Countries with low GDP have increasingly high CMR (4). There are a few countries which have low CMR in spite of the low GDP (5) – these include *Bulgaria, Costa Rica, Russian Federation, Romania, Poland* and many other European countries. When the GDP falls below 2 million the curve flattens out and CMR varies from 245 to 783. *Saudi Arabia* (6) stands out as an outlier – in spite of the high GDP, the CMR is high – as much as 396. For other countries with comparable GDP the rate is much lower, in the range of 100 to 125. An investigation into the cause of high CMR in *Saudi Arabia* strongly points to poor parental education (Al-Mazrou and colleagues [14]).

The curve in Fig. 9 shows a non- linear relationship called rectangular hyperbole. This can be shown mathematically as:

$$y = \frac{k}{x}$$

Where x is inversely proportional to y and k is the constant of proportionality.

To see the relationship from another point of view, we attempt to straighten the curve with Fig. 10 which shows the plot between GDP and inverse of CMR. The line of fit shows a clear pattern of relationship. Circles 1 to 6 are marked as in Fig. 9. Linear regression lines shown in red, blue and green correspond to NMR, IMR and U5MR.

3.2 How do individual countries fare?

In order to see if increase in GDP results in decrease in CMR, a country from each of the areas is picked to study in detail. Fig. 11 shows the line chart of *Luxembourg* (1), *India* (2), *USA* (3), *Indonesia* (4), *Poland* (5) and *Saudi Arabia* (6). Reading the chart from right to left, as the GDP

increases CMR decreases. A jump in GDP can be seen in India and Indonesia as shown in the circled areas but there is no corresponding drop in CMR.

3.3 GDP, Correlation and Regression

This study will now apply statistical analysis to understand and verify the patterns seen so far.

We begin with measuring GDP and showing it on a choropleth map (Fig. 12). Then the correlation of GDP to CMR is shown and from this we can see the consistent colour all over the world, except for the countries with insignificant GDP which are blanked out. Lastly we look at the R-squared values to see how regression works for each country. Many countries show very high regression values.

3.4 Clustering

Finally we try to group the countries using an unsupervised machine-learning model from Scikit-Learn using the simplest algorithm, K-Means. After experimenting with several cluster sizes from three to eigth, we settle on cluster size of 4. Fig. 13 shows the result of clustering with four centres. We apply this algorithm to all three categories to study similarities and differences.

Luxembourg stands out as the centre of the first cluster. It is the country with the least CMR of 42 with a GDP of 1.5 million. The next cluster is the set of countries with high GDP and low CMR. The next cluster is the set of countries with low GDP yet low CMR. The last cluster is of countries with very low GDP and relatively high CMR. The findings correspond roughly to Fig. 9, where the clustering was done manually.

Increasing the cluster size to eight still did not isolate *India* and *South Africa* which are in the tail end of this inverse curve nor was *Saudi Arabia* shown as an outlier with its own cluster but when looking at this visually, these groups are obvious. The algorithm does not assign centres as you might assign it with your eye.

4. Findings

Our efforts to answer the research questions are as follows:

4.1 Research Question 1*: Study the trend of CMR over the last twenty one years around the world and understand if any spikes can be explained with war, disease or natural disasters.*

CMR in all the categories are in decline throughout the world as can be seen from Fig. 3.

The spikes for *Haiti* in Fig. 4 for the year 2010 is because of the earth quake it suffered. The upward trends in *Somalia* and *Cambodia* are due to the civil war, famine and AIDS epidemic.

Although babies between 0-28 days have a lower chance of survival than babies between 28 days and one year who in turn have higher chances that a child between one and five years, Fig. 6 shows that in the worst affected countries, this pattern is not prevalent. In *Mali*,

Somalia, *Nigeria* and *South Sudan* the third category is noticeably more that the first and second.

Choropleth map of the world shown in Fig. 8 distinctly shows that the countries towards the southern hemisphere such as African and South Asian countries are severely affected by CMR, *Australia* being the exception. As we move towards the northern hemisphere, the rates drop dramatically.

4.2 Research Question 2: *Can CMR of a country be explained by GDP of the country and if so to what degree?*

Plot of CMR against GDP shows a rectangular hyperbolic relationship (Fig. 9). GDP does have a significant effect on low CMR, but only up to a certain point. *Saudi Arabia* stands out as an outlier - it has a high CMR in spite of high GDP and the reason is found to be poor parental education. A further plot of one country from each of the areas (Fig.11) shows that in all cases there has been a steady decrease of CMR over years. *India* and *Indonesia* show a jump in GDP, but this is not significant enough to cause a corresponding drop in NMR/IMR or U5MR.

Correlation between CMR and GDP ranges from very high negative correlation (-9.97) for Slovakia to positive correlation (0.57) for Argentina. Fig. 12 shows how the countries look, the GDP map is almost the opposite of Fig. 8 showing CMR. Correlation and R-squared values show that GDP does have a high relation to the CMR in the countries. Some countries have very low GDP and they have almost dropped out of the map like most African countries and *Afghanistan*.

Clustering the countries into 4 groups as in Fig. 13 show the very high, mid and low GDP countries all have similar CMR, while the very low GDP countries do vary a lot with their CMR. This confirms the findings from Fig. 10.

5. Critical Reflection

5.1 Conclusion

GDP is a measure of well-being of a country and it might directly correlate to good pre-natal health of the mother, quality neo-natal care, better education, higher living conditions, less displacement, easy access to medical facilities and much more. But GDP is not the only measure of well-being of a country. However, this study only deals with the effect of GDP on the CMR.

There may be one or more confounding factors which may lead us to erroneous correlation between CMR and GDP.

1. War and terrorism: Civil and international is one of them. *Pakistan* has been severely affected by the war on terror since 2001 killing over 60,000 people. *Afghanistan* has had severe casualties too. Bhutta and colleagues [10] have noted in their study that

children have paid a disproportionate price due to malnutrition, disease and death. *Sierra Leone* has endured untold suffering due to the civil war lasting over 10 years. Kiros and colleagues [11] have said that the CMR is also affected by parent's education. Fig. 1 features these countries as most affected.

2. Disease: Countries with extreme HIV population also show high paediatric AIDS numbers – *India* and *South Africa* are among the top five countries as per the UN report [12] and co-incidentally they have been identified in Fig. 9.

3. Natural disasters: Famine and earthquakes affect CMR adversely as can be expected. *Africa* (famine) and *Haiti* (earthquake) have suffered increased CMR. Incidentally, *Haiti* is the spike identified in Fig. 4.

4. Refugee immigration: *Pakistan* has had a huge influx of refugees from *Afghanistan* which can increase CMR in Pakistan. According to Essen and colleagues [13], perinatal mortality is high in children of women from sub-Saharan immigrants compared to women of Swedish origin. This might mean, without the immigrants *Sweden* might have even lesser CMR.

Instead of doom and gloom, we view CMR armed with real facts. In general, CMR is decreasing around the world. Developed countries with high GDP fare much better than countries with low GDP. Looking at the cross-sectional view of CMR and GDP data, when we compare 1996 data to 2016 data (Fig. 14), we see a marked drop in CMR and a steady increase in GDP for most countries. This study concludes that CMR will continue to drop and we can look forward to a better future should this trend continue.

5.2 Limitations

UNICEF [1] data is an estimate and not exact, however the lower and upper limits give us a range for uncertainty. Numerous factors affect CMR as identified in the conclusion like war, terrorism, disease, natural disasters and immigration. However this study just focusses on GDP and its effect on CMR. Nevertheless, each of these confounding factors also have a correlation to GDP – war causes GDP to dive down, so do pestilence, earthquakes, famine and refugee immigration. While foreign aid may flood in, the effect of displacement and living in camps far outweigh these aids on CMR. We have taken a time capsule of 21 years to study, only because this means we have reliable dense data in this period and it does not involve intense missing value imputation. A study of longer period might reveal more insight into this data. Missing values were imputed with forward and backward time series interpolation which may have affected the statistics. CMR is calculated as an average for the whole country, but separating this out into provinces might show huge disparities within the country. While a country might show low CMR, it may have distinct difference in rates between say rural and urban population.

5.3 Applications to other domains

The study of this time series data of CMR can be applied to many other domains with time series data. Poverty, education, morbidity, pollution, road accidents, economic freedom of women and happiness are some of the domains which can have time series data over the

years and we will be able to study the effect of GDP and trends around the world. The visualisation used in this study range from bar graphs to choropleth maps and these can be easily reused to study other domains. Analytical tools used include correlation, regression and clustering all of which can be readily used for other domains to understand collinearity. GDP itself may be replaced by other independent variables such as peace index or measure of natural resources and this study will still be reusable across multiple domains.

6.0 Graphs and Maps

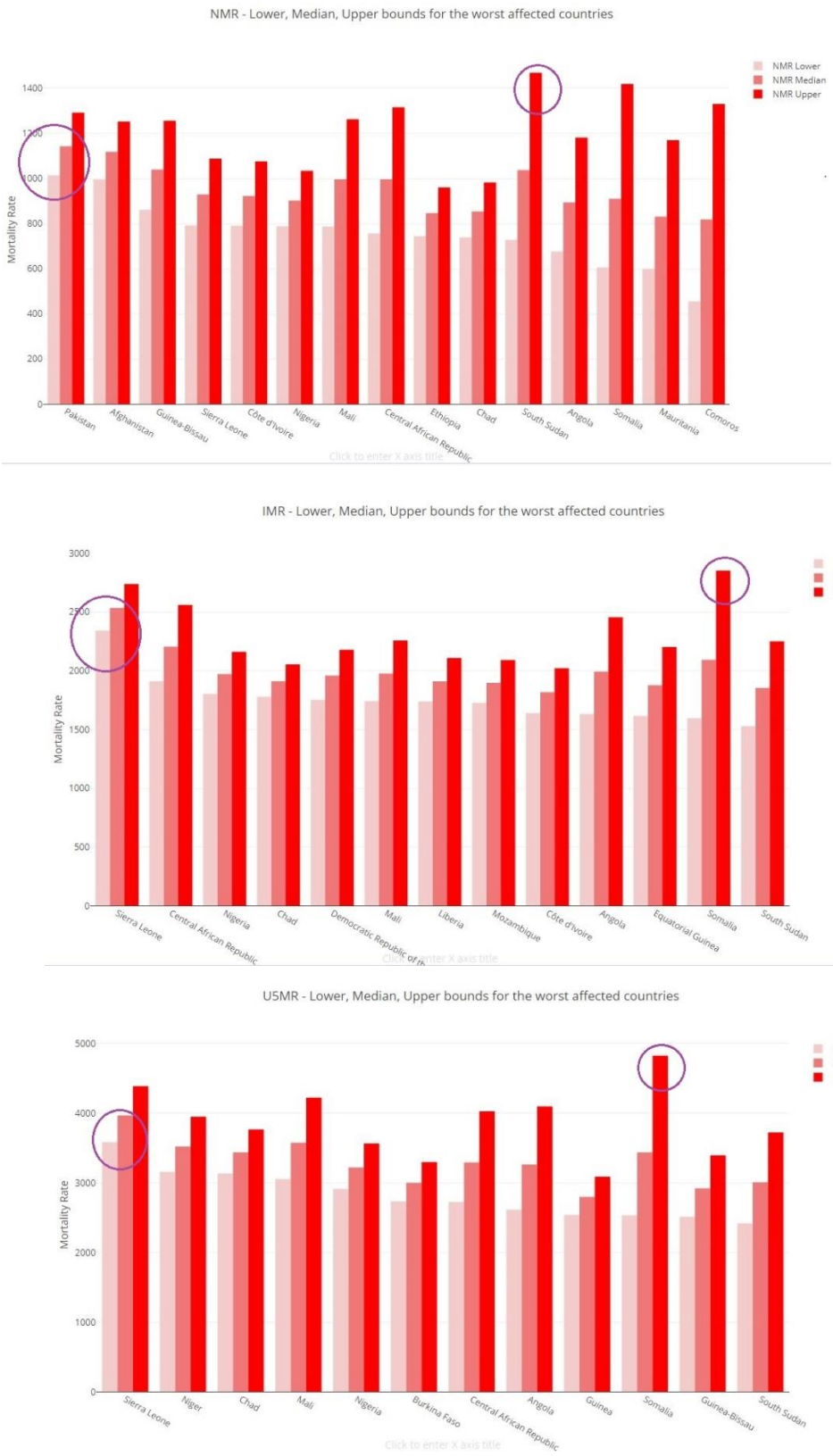


Fig. 1: Lower, Median and Upper uncertainty bounds of NMR/IMR/U5MR of the worst affected countries are shown. The circles indicate the worst countries for each of the categories for its uncertainty bounds.

Fig. 2: Lower, Median and Upper uncertainty bounds of NMR/IMR/U5MR of the least affected countries are shown. The circles indicate the least affected countries for each of the categories for its uncertainty bounds.

Fig. 3: Time series trend line showing cumulative CMR of all the countries over the last two decades. The swing of the lower and upper bounds are shown with arrows. The circle denotes the bump in CMR in the year 2009.

Fig. 4: Time series trend line showing line per country. Although the trend line is generally going down the circles show the spikes in the trend. Arrow shows that one of the countries has increasing CMR.

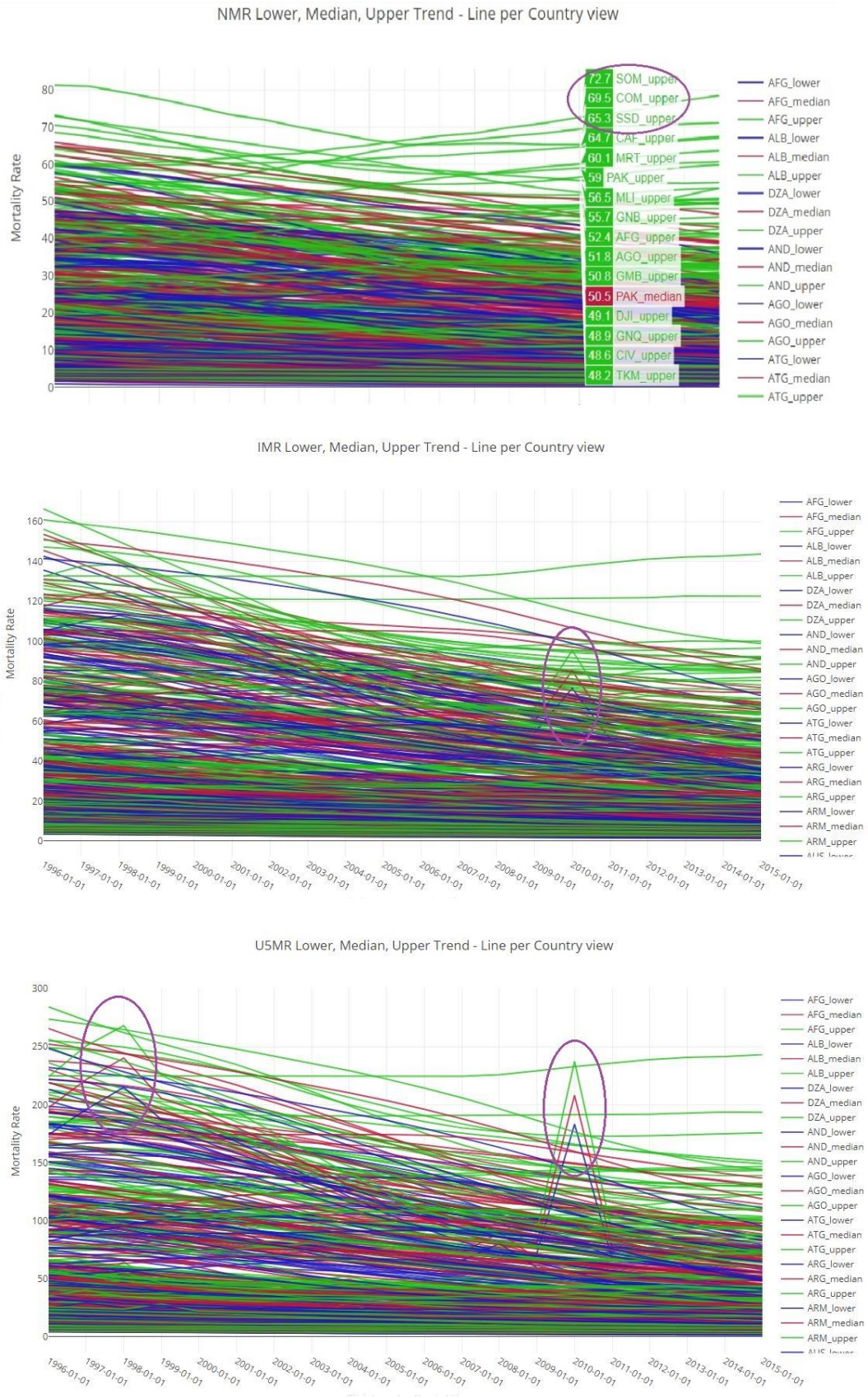


Fig. 5: Time series plot trend line showing lower, median and upper bounds for NMR/IMR /U5MR. Upper bound shows an upward trend as highlighted. Circled portions show the spikes in all three bounds.

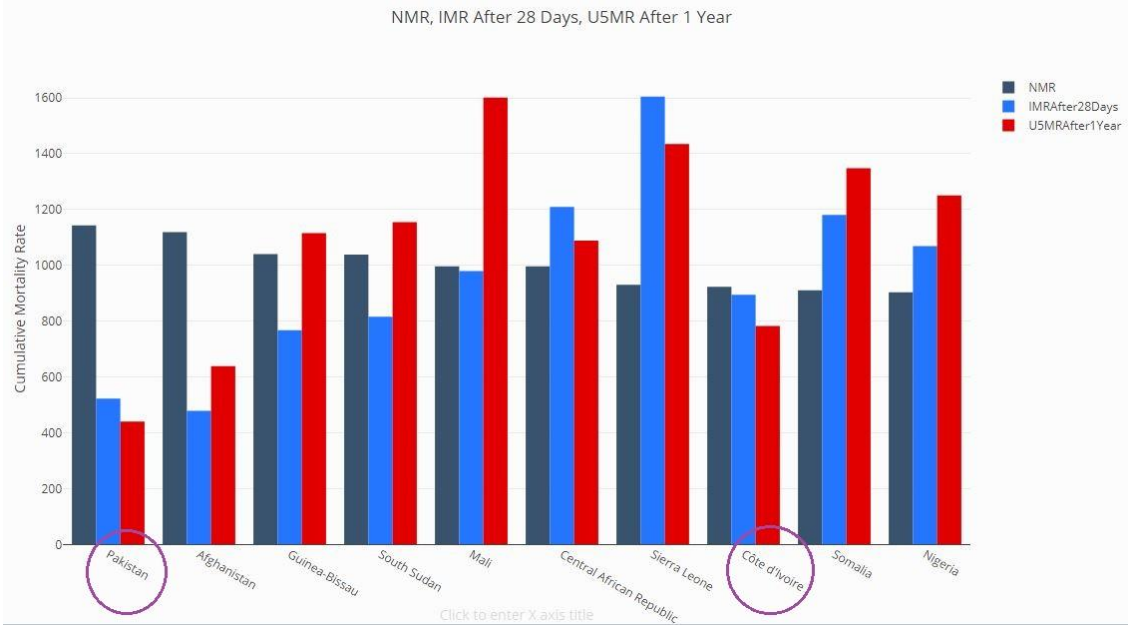


Fig. 6: A grouped bar chart of the top of the worst affected countries, but with IMR excluding NMR, U5MR excluding IMR. When the count is split, only the circled countries follow NMR/IMR/U5MR decreasing rate.

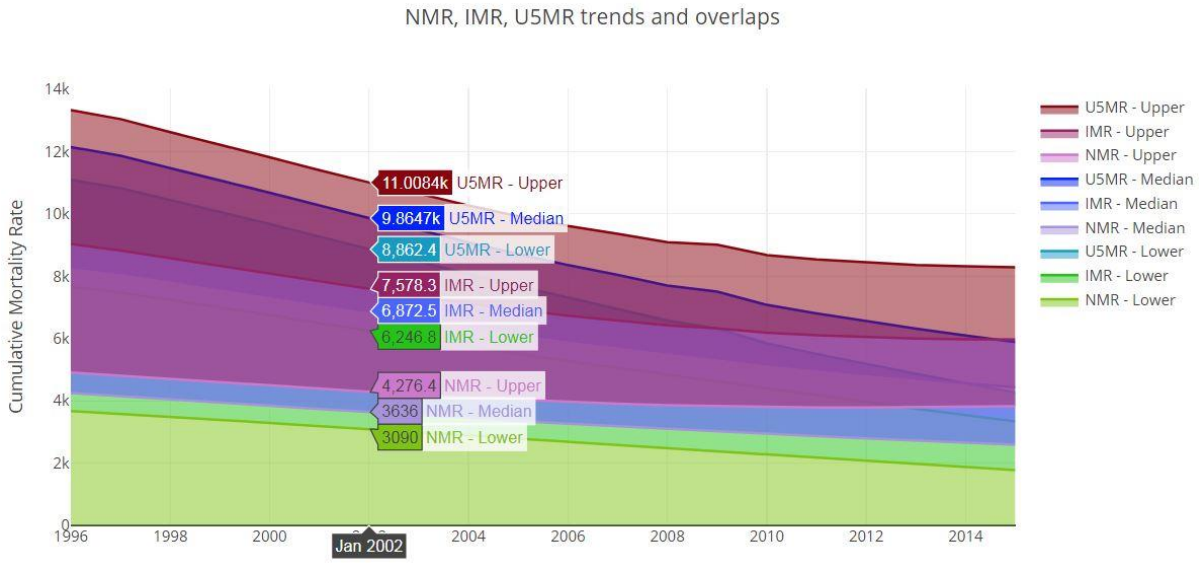


Fig. 7: Stacked area chart showing the overlap of NMR/IMR/U5MR trend over years along with uncertainty bounds.

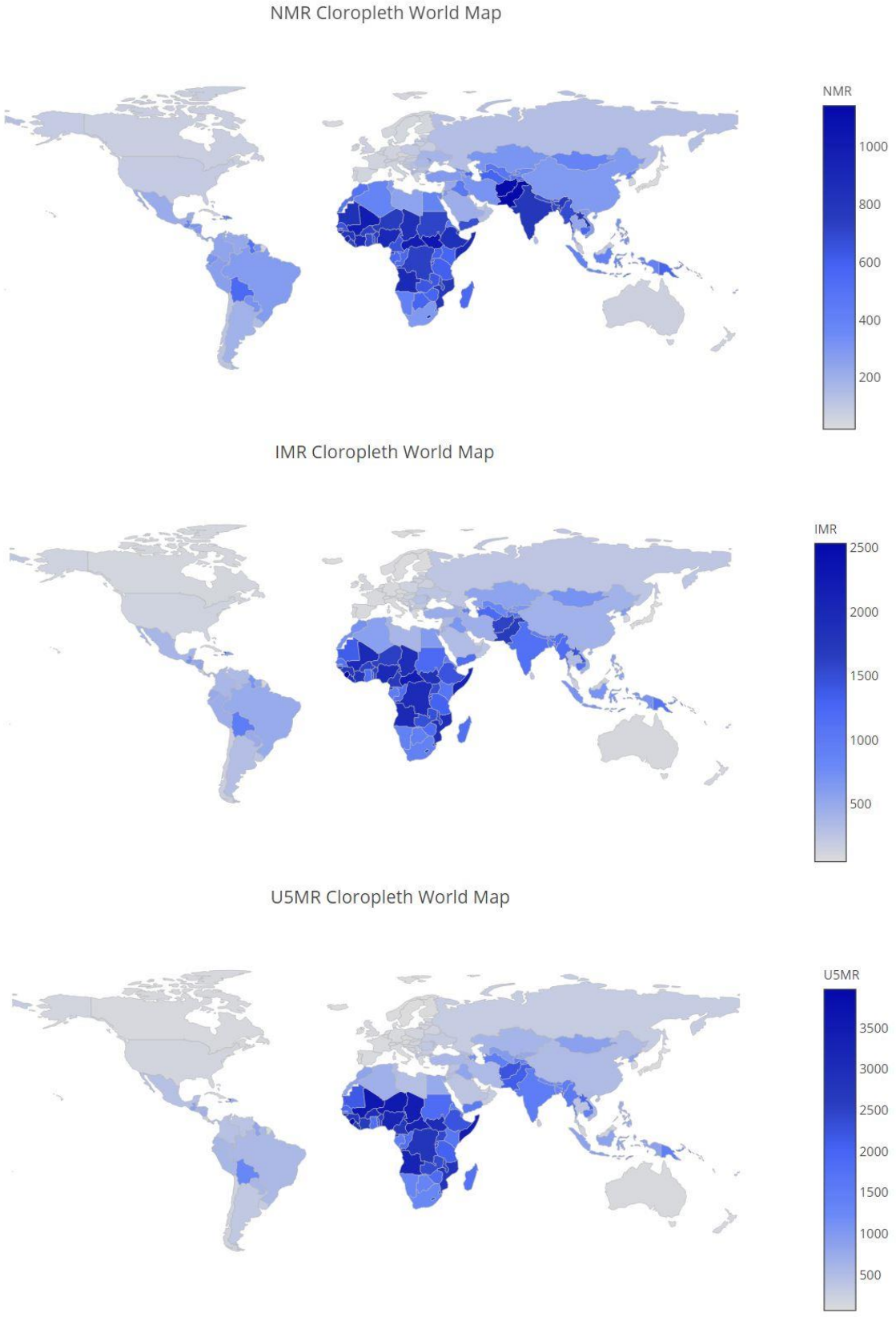


Fig. 8: Choropleth map of the world shows the intensity of NMR/IMR/U5MR in the world. Africa and South Asian countries show up as the most affected.

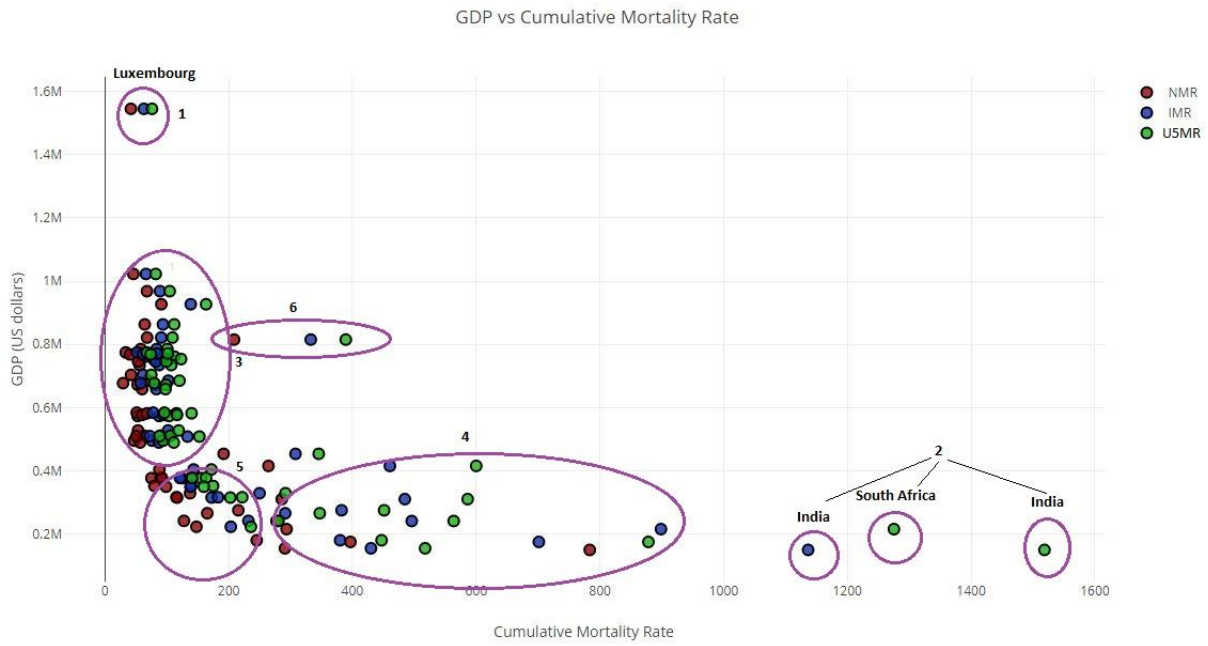


Fig. 9: This scatter plot of GDP vs Cumulative CMR over the last two decades shows the inverse curve of relationship between the two. The circled countries show the head and tail of the inverse curve.

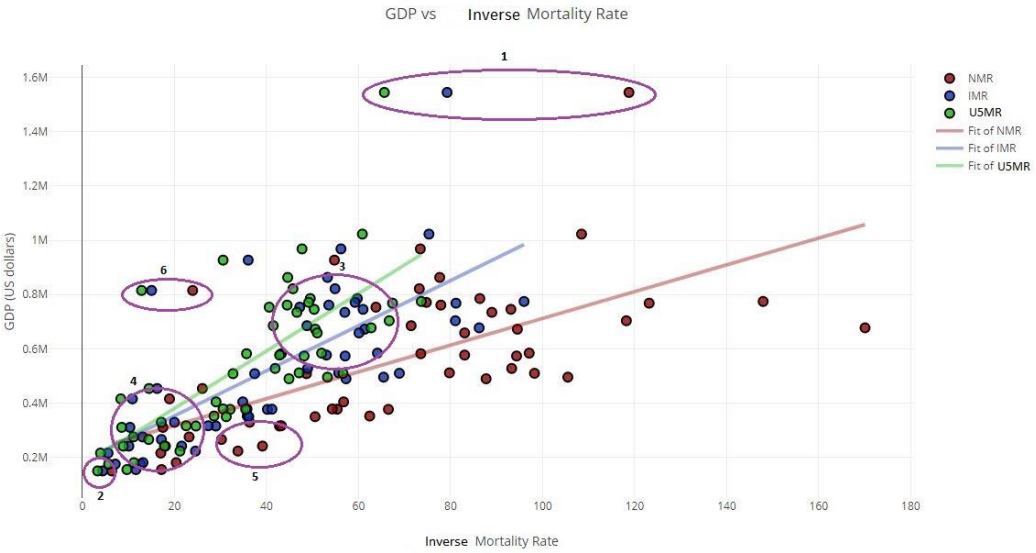
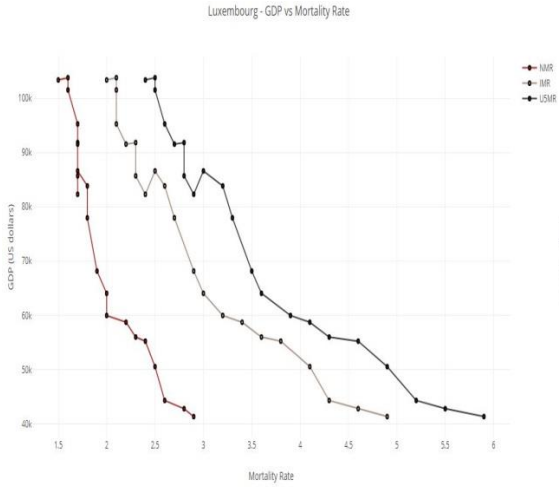
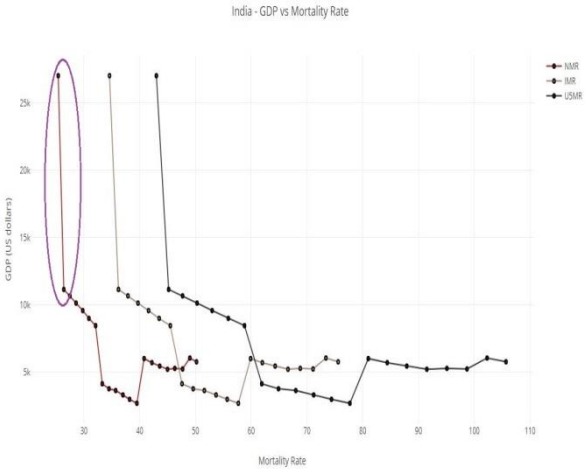


Fig. 10: Scatter plot of GDP with Inverse CMR. The regression lines for all three categories show a linear relationship. The circled sections correspond to Fig. 9.

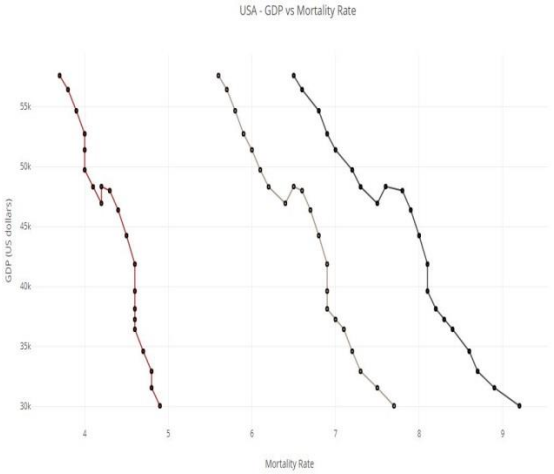
Area 1 – Luxembourg



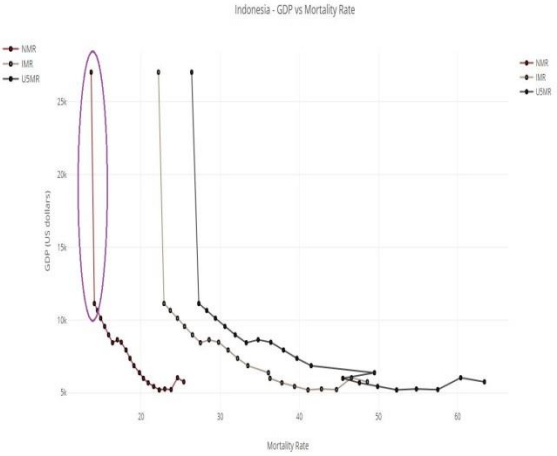
Area 2 - India



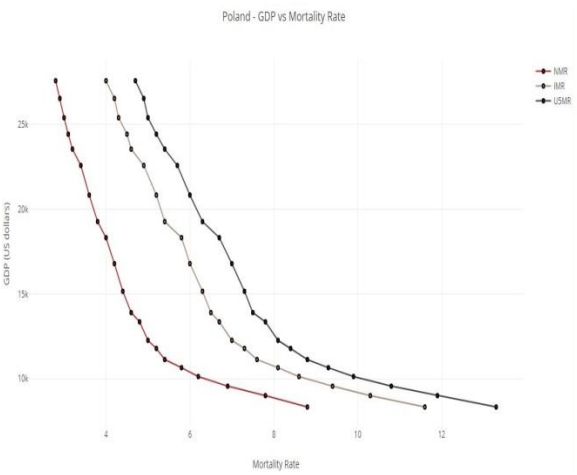
Area 3 – USA



Area 4 - Indonesia



Area 5 – Poland



Area 6 – Saudi Arabia

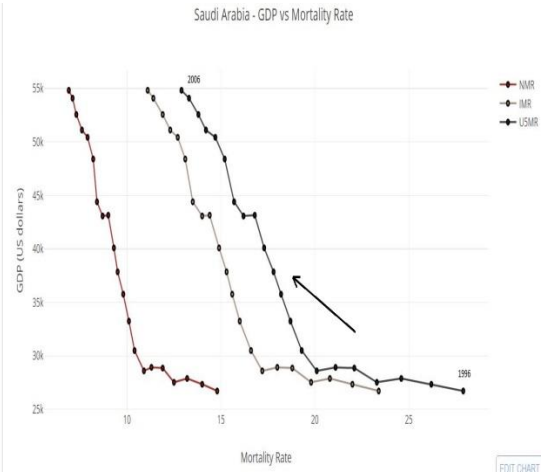


Fig. 11: Line chart of GDP vs CMR is shown for one country from each of the six areas. Reading the map from right to left as GDP increases CMR decreases. Circled areas show the jump in GDP.

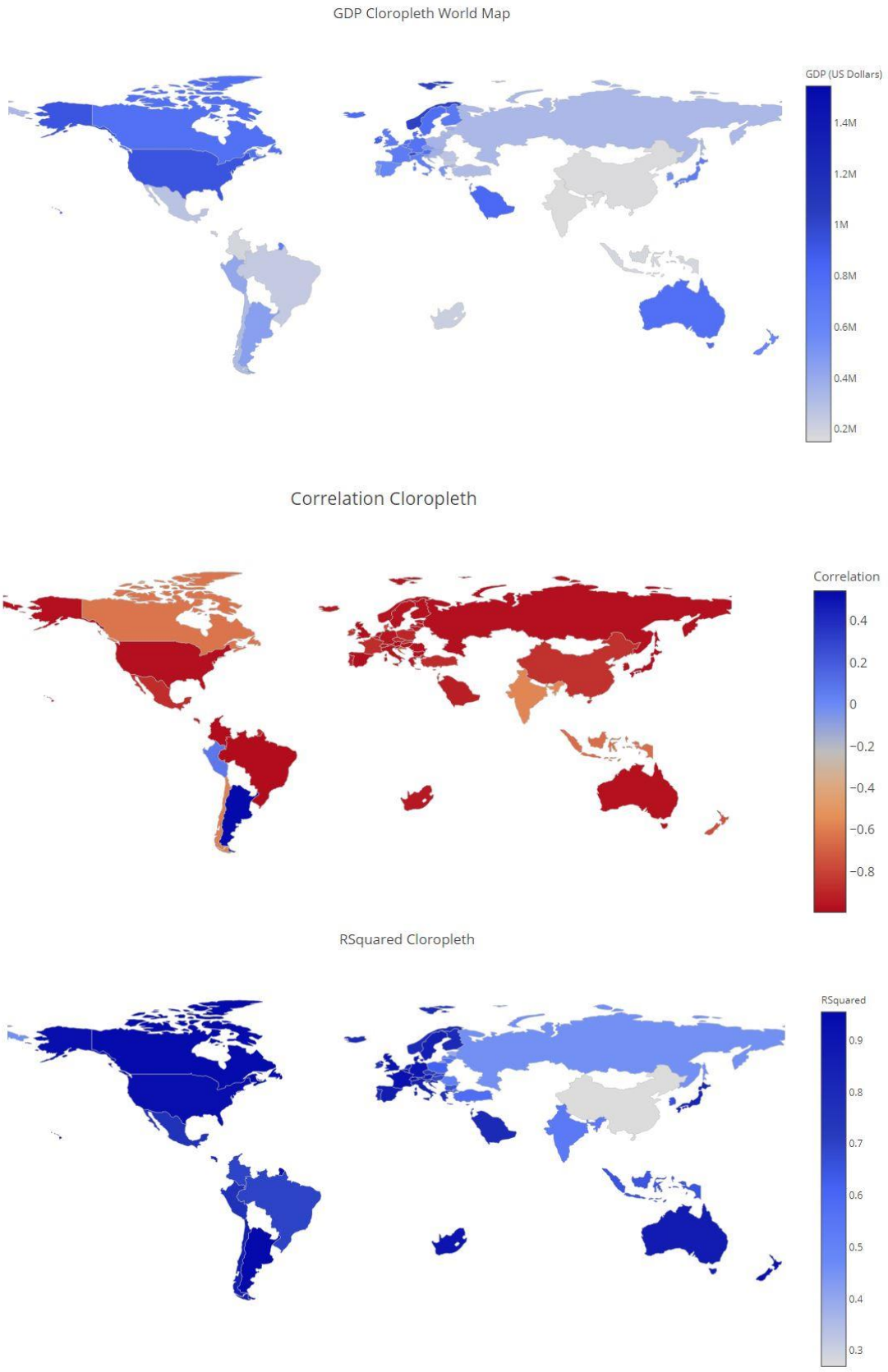
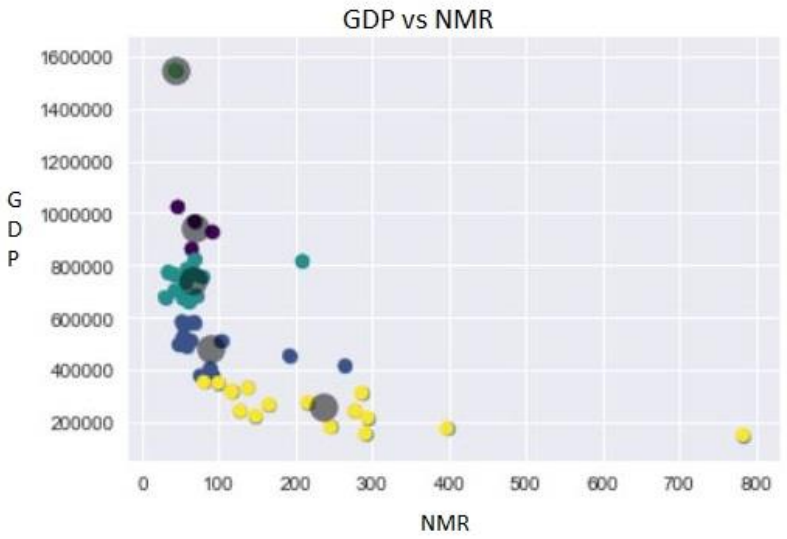
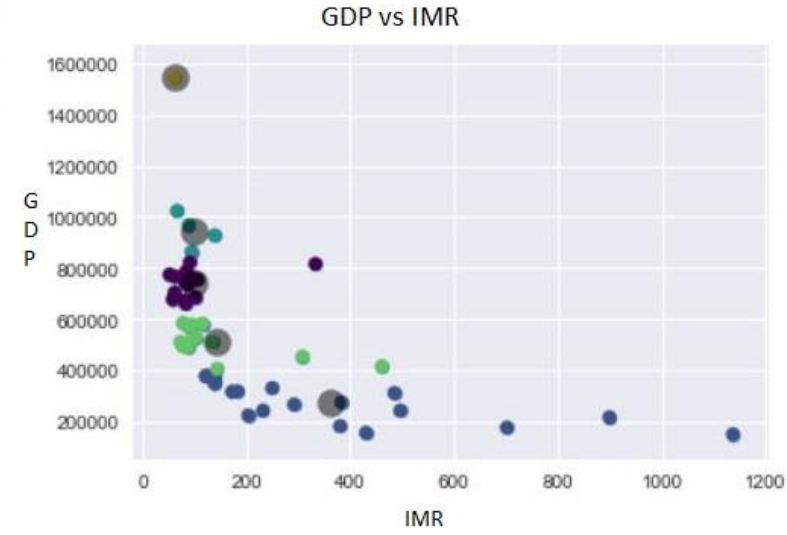


Fig. 12: Choropleth maps show countries GDP, the correlation of GDP to CMR and the R-squared value of GDP and CMR.





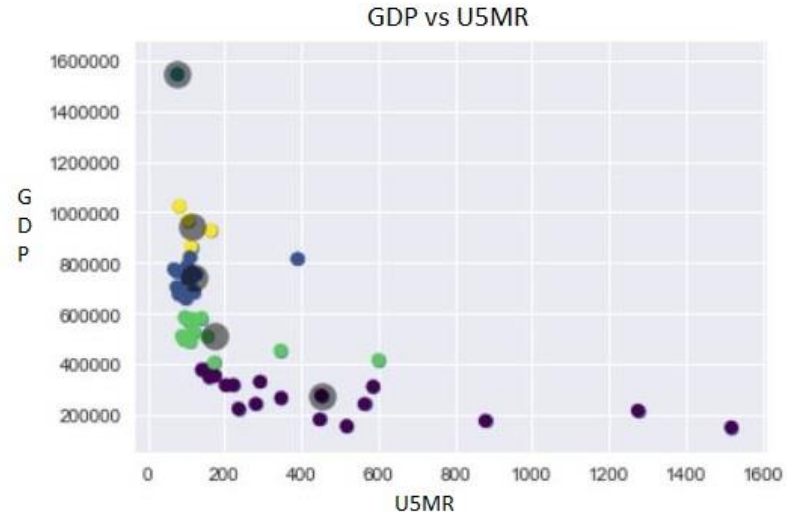


Fig. 13: K-Means cluster shown with NMR / IMR and U5MR. The dots show the centres of the clusters.

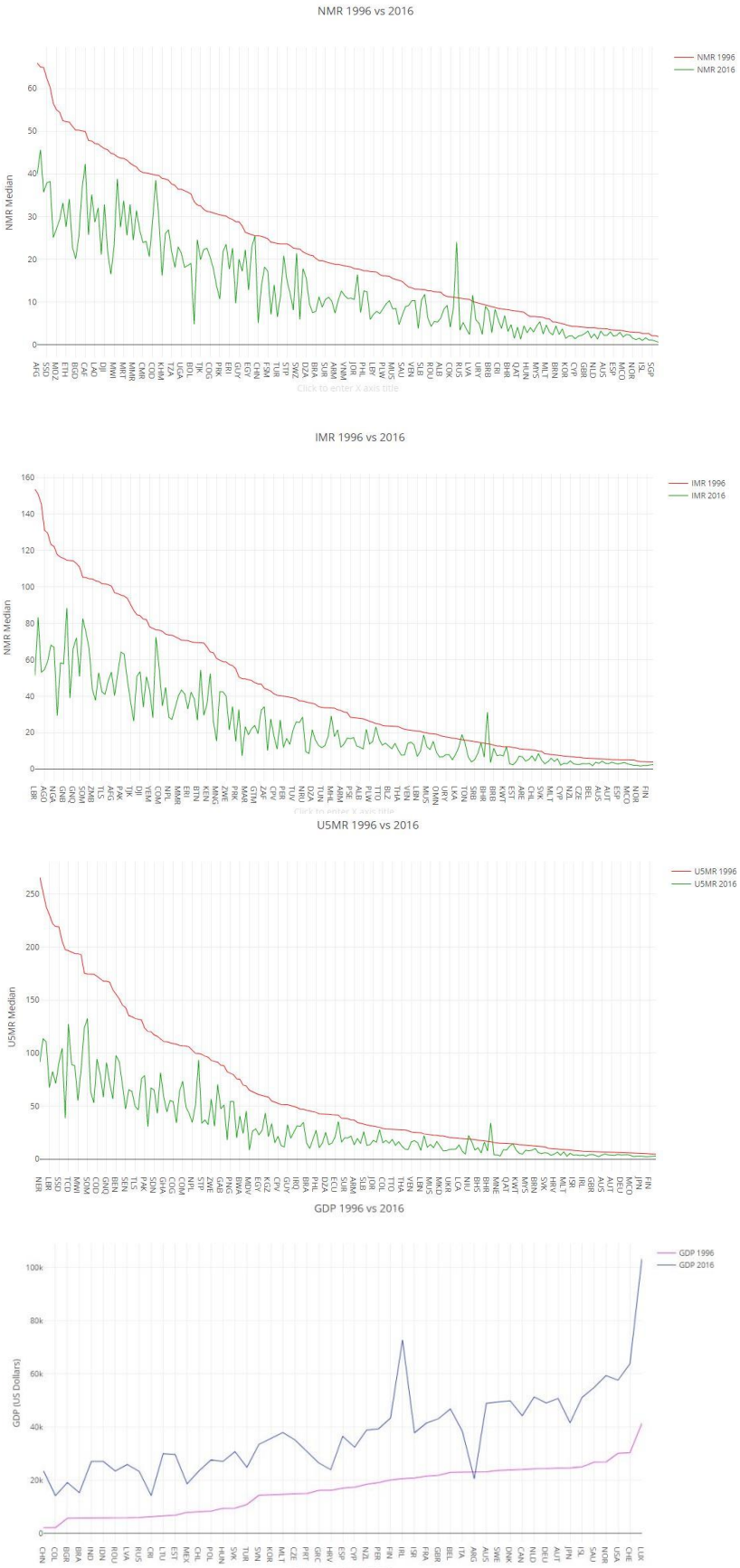


Fig. 14: Line plot of NMR, IMR, U5MR and GDP in 1996 vs 2016. CMR has decreased and GDP as increased.

References

1. Unicef UK. (2017). *Unicef UK - Children's charity - For Every Child in Danger*. [online] Available at: https://www.unicef.org.uk [Accessed 29 Dec. 2017].

2. World Health Organization. (2017). *World Health Organization*. [online] Available at: http://www.who.int [Accessed 29 Dec. 2017].

3. Oecd.org. (2017). *OECD.org - OECD*. [online] Available at: https://www.oecd.org [Accessed 29 Dec. 2017].

4. Author: Plotly Technologies Inc. Title: Collaborative data science Publisher: Plotly Technologies Inc. Place of publication: Montréal, QC Date of publication: 2015 URL: https://plot.ly

5. Wikipedia. (2017). *Main Page*. [online] Available at: https://en.wikipedia.org [Accessed 29 Dec. 2017].

6. GitHub. (2017). *Build software better, together*. [online] Available at: https://github.com/ [Accessed 29 Dec. 2017].

7. Our World in Data. (2017). *Our World In Data*. [online] Available at: https://ourworldindata.org/ [Accessed 29 Dec. 2017].

8. Gapminder.org. (2017). *Gapminder: Unveiling the beauty of statistics for a fact based world view.*. [online] Available at: https://www.gapminder.org/ [Accessed 29 Dec. 2017].

9. Chen, B., Halliday, T.J. and Fan, V.Y., 2016. The impact of internal displacement on child mortality in post-earthquake Haiti: a difference-in-differences analysis. *International journal for equity in health*, *15*(1), p.114.

10. Bhutta, Z.A., 2002. Children of war: the real casualties of the Afghan conflict. *Bmj*, *324*(7333), pp.349-352.

11. Kiros, G.E. and Hogan, D.P., 2001. War, famine and excess child mortality in Africa: the role of parental education. *International journal of epidemiology*, *30*(3), pp.447-455.

12. Un.org. (2017). *COUNTRIES MOST AFFECTED BY HIV/AIDS ARE LEAST ABLE TO PAY FOR PREVENTION AND TREATMENT | Meetings Coverage and Press Releases*. [online] Available at: https://www.un.org/press/en/2001/aids18.doc.htm [Accessed 29 Dec. 2017].

13. Essén, B., Hanson, B.S., ÖSTERGREN, P.O., Lindquist, P.G. and Gudmundsson, S., 2000. Increased perinatal mortality among sub-Saharan immigrants in a city-population in Sweden. *Acta obstetricia et gynecologica Scandinavica*, *79*(9), pp.737-743.

14. Al-Mazrou, Y.Y., Alhamdan, N.A., Alkotobi, A.I., Nour, O.M. and Farag, M.A., 2008. Factors affecting child mortality in Saudi Arabia. *Saudi medical journal*, *29*(1), pp.102-106.