

## **STATISTICAL ANALYSIS ON FACTORS INFLUENCING LIFE EXPECTANCY**

### **I. Introduction**

The dataset used for this project is **Life Expectancy (WHO)**. The Global Health Observatory (GHO) data repository under World Health Organization (WHO) keeps track of the health status as well as many other related factors for all countries. The dataset related to life expectancy, health factors for 193 countries has been collected from the same WHO data repository website and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It has been observed that in the past 15 years, there has been a huge development in health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years. Therefore, in this project we have considered data from year 2000-2015 for 193 countries for further analysis. The individual data files have been merged together into a single dataset. On initial visual inspection of the data showed some missing values. As the datasets were from WHO, we found no evident errors. Missing data was handled in R software by using Missmap command. The result indicated that most of the missing data was for population, Hepatitis B and GDP. The missing data were from less known countries like Vanuatu, Tonga, Togo, Cabo Verde etc. Finding all data for these countries was difficult and hence, it was decided that we exclude these countries from the final model dataset. The final merged file(final dataset) consists of 22 Columns and 2938 rows which meant 20 predicting variables. All predicting variables was then divided into several broad categories: Immunisation related factors, Mortality factors, Economical factors and Social factors.

There have been lot of studies undertaken in the past on factors affecting life expectancy of different countries considering factors such as demographics, income composition, mortality rates, pathologies of various diseases. Important immunisation like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunisation factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

### **II. Problem Domain**

The data-set aims to answer the following key questions:

1. Does various predicting factors which has been chosen initially really affect the Life expectancy? What are the predicting variables actually affecting the life expectancy?
2. Should a country having a lower life expectancy value(<65) increase its healthcare expenditure in order to improve its average lifespan?
3. How does Infant and Adult mortality rates affect life expectancy?
4. Does Life Expectancy has positive or negative correlation with eating habits, lifestyle, exercise, smoking, drinking alcohol etc.
5. What is the impact of schooling on the lifespan of humans
6. Does Life Expectancy have positive or negative relationship with drinking alcohol?
7. Do densely populated countries tend to have lower life expectancy?
8. What is the impact of Immunisation coverage on life Expectancy?

### **III. Dataset**

Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

World Health Organization(WHO) and United Nations. The intended use of data : Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that effect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well.

**Attributes :**

Country

Year

Status : Developed or Developing status

Life expectancy : Life Expectancy in age

Adult Mortality : Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)

Infant deaths : Number of Infant Deaths per 1000 population

Alcohol : Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)

Percentage expenditure : Expenditure on health as a percentage of Gross Domestic Product per capita(%)

Hepatitis B : Hepatitis B (HepB) immunization coverage among 1-year-olds

Measles : Number of reported cases per 1000 population

BMI Average : Body Mass Index of entire population

Under five deaths : Number of under-five deaths per 1000 population

Polio : Polio (Pol3) immunization coverage among 1-year-olds (%)

Total expenditure : General government expenditure on health as a percentage of total government expenditure (%)

Diphtheria : Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)

HIV/AIDS : Deaths per 1 000 live births HIV/AIDS (0-4 years)

GDP : Gross Domestic Product per capita (in USD)

Population : Population of the country

Thinness 1-19 years : Prevalence of thinness among children and adolescents for Age 10 to 19 (%)

Thinness 5-9 yrs : Prevalence of thinness among children for Age 5 to 9(%)

Income composition of resources : Human Development Index in terms of income composition of resources (index ranging from 0 to 1)

Schooling : Number of years of Schooling(years)

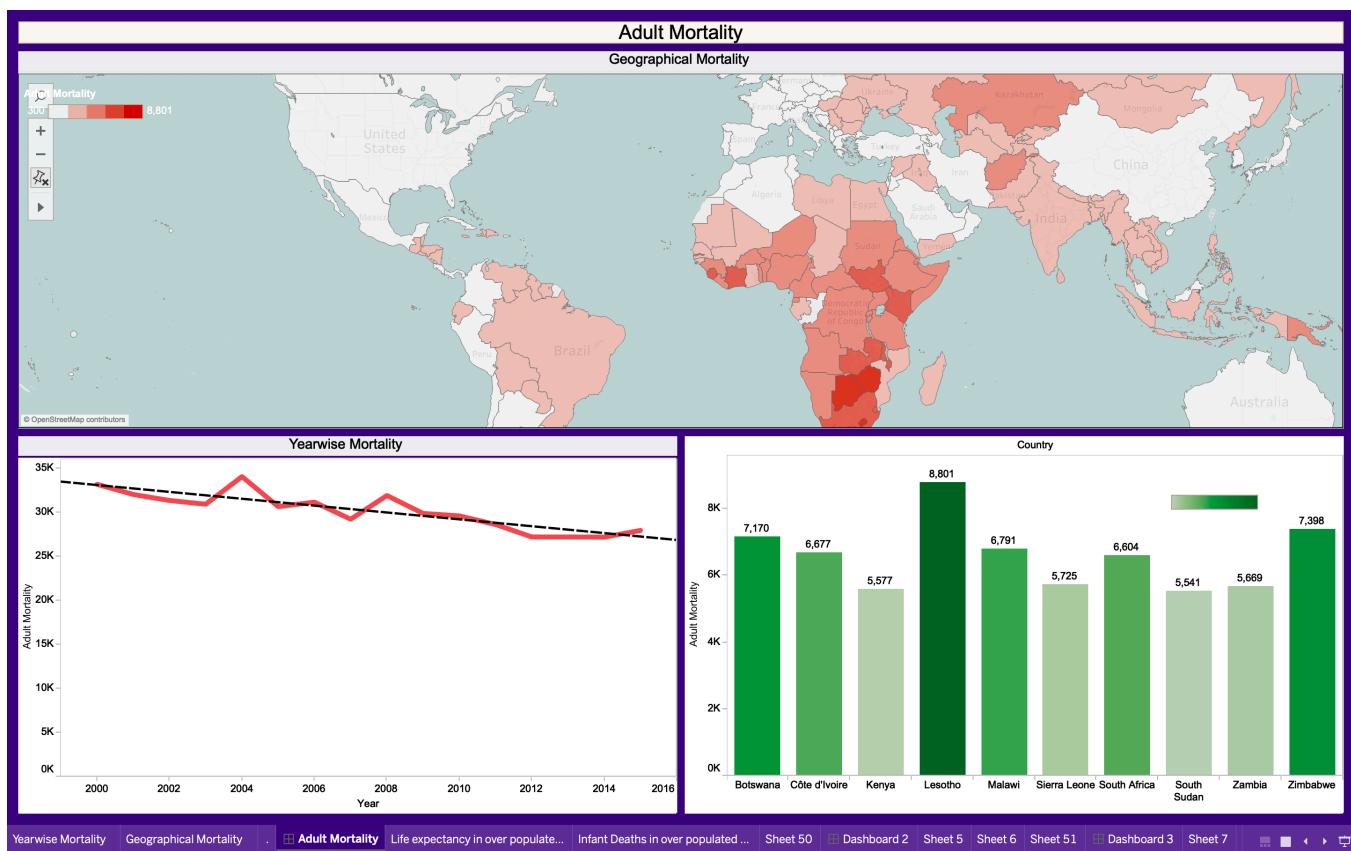
## IV. Analysis

### Dashboard 1 : Adult Mortality

Three visualizations were designed to build this dashboard on Adult Mortality. The attributes used were Country , Year ,and Adult mortality rates .

**Geographical Mortality** : Adult mortality rates plotted against various countries available in the dataset to create an interactive map structure , highlighting that particular region when hovered over by the user. The intensity of the colour ( red ) denotes the level of adult mortality in that country.

**Yearwise Mortality** : The adult mortality rates of all the countries in the chronological order is represented in the form of a line graph. There is a trend line drawn along the graph denoting the regression of adult mortality rates.



---

**Mortality Bar chart :** The countries with the most adult mortality rates are represented in the form of a bar graph labelling their respective mortality rates. The intensity of the colour denotes the level of adult mortality in that country.

### Insights :

The dashboard overviews the mortality rates of various countries around the globe from the years 2000 - 2015. Considering the whole world , the mortality rate has declined significantly over the years. African countries like Botswana , Cote d'ivoire, Kenya , Malawi , South Sudan rank the most in terms of mortality , with Lesotho being the highest ( mortality rate of 8801 ) . A linear trend model is computed for sum of Adult Mortality given Year. The model may be significant at  $p \leq 0.05$ . There had been many ups and downs over the years but , on the whole , mortality rates have decreased. With the advancement of technology and digitization , people and the society have improved notably over the years . The overall adult mortality has gone down from 33,210 to 27,974 within a period of 15 years. With the invention of new medicines and improvements in the healthcare domain pandemics and deadly diseases can be cured and eventually mortality rates would reduce. Africa , being a less developed continent , holds the most number of countries with high mortality rates. It is evident that many countries have developed in the past decade. Already technology and new techniques are ruling the world . Within the next 10 years , it is certain that many more countries would emerge developed , improving the lives of people and gradually reducing the death rate.

### Dashboard 2 : Comparing infant deaths with life expectancy rates of overpopulated countries.

Three visualizations were designed to build this dashboard comparing the life expectancy rate and infants death rate of various overpopulated countries ( above 3 billion ) . The attributes used were Country , Year , Life expectancy and infant death rates.

**Life expectancy in over populated countries :** The dataset was filtered in terms of population to return overpopulated countries around the world. ( 3 billion here ) Sum of Life expectancy for each Country was plotted against the respective countries. Color shows details about Country. The data is filtered on sum of Population, which ranges from 300,000,000 to 6,743,483,055.

### Infant Deaths :

The sum of Infant Deaths marked against the highly populated countries. The marks are labeled by Country. The data is filtered on sum of Population, which includes values greater than or equal to 300,000,000.

**Sum of Life expectancy vs. sum of Infant Deaths :** Color shows details about Country. The view is filtered on Country, which keeps 10 of 193 members.The mark type is Circle. The marks are labeled by Country.

**Filters:** Population

**Text:** Country

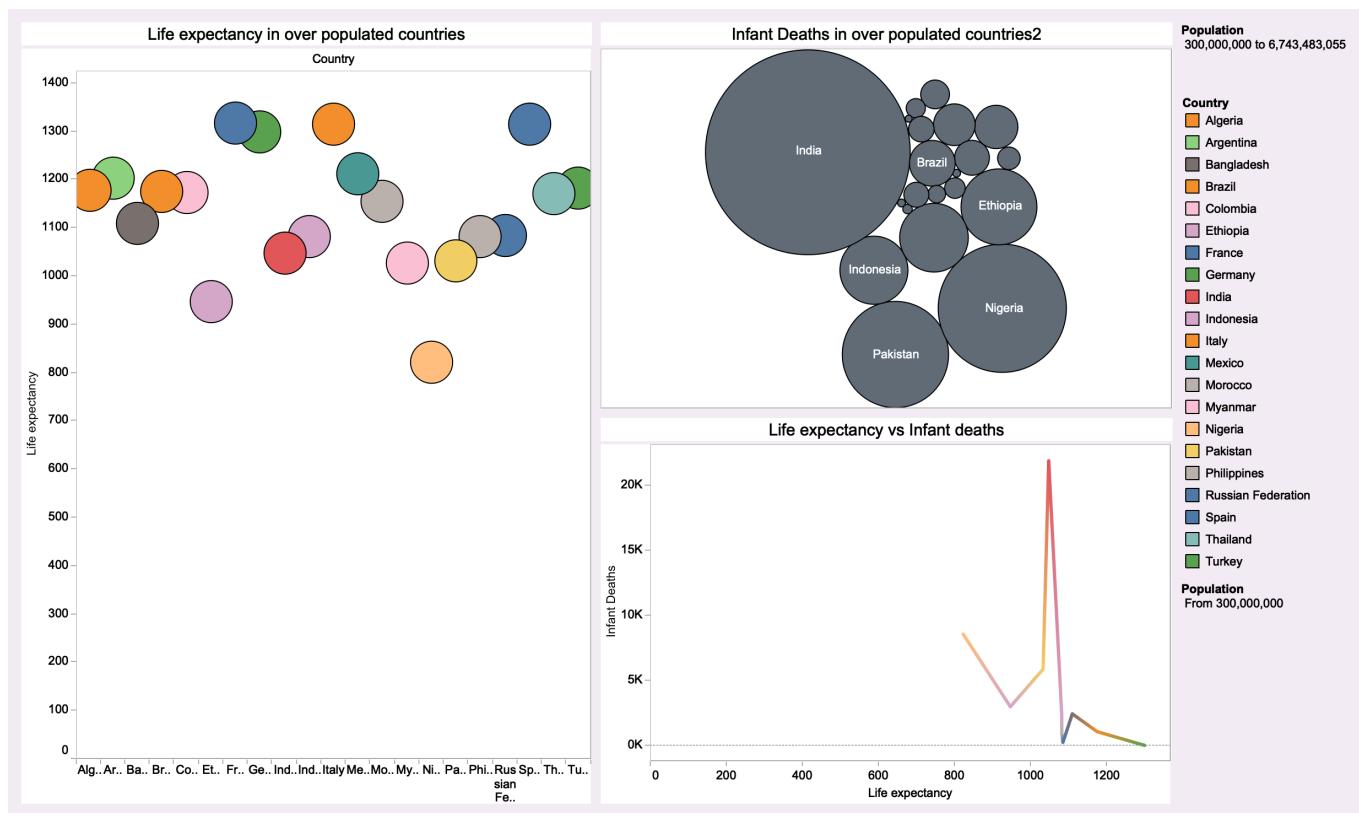
**Size:** Sum of Infant Deaths

Country has 21 members on this sheet

**Members:** Argentina; Bangladesh; Brazil; France; India; Indonesia ; Italy

**Measures:**

Sum of Population has the value 300,000,000 on this sheet. The filter associated with this field includes values greater than or equal to 300,000,000. Sum of Infant Deaths ranges from 26 to 21,867 on this sheet.



### **Life expectancy vs Infant deaths :**

A coloured line graph is drawn to determine the relationship between Sum of Life expectancy vs. sum of Infant Deaths. Color shows details about Country. The view is filtered on Country, which keeps 10 of 193 members. The mark type is Line.

#### **Dimensions**

**Country has 10 members on this sheet**

**Members:** Brazil; India; Indonesia; Nigeria; Pakistan; ...

#### **Measures**

Sum of Life expectancy ranges from 821.7 to 1,298.8 on this sheet.

Sum of Infant Deaths ranges from 40 to 21,867 on this sheet.

#### **Insights:**

Life expectancy equals the average number of years a person born in a given country is expected to live if mortality rates at each age were to remain steady in the future. The life expectancies of France , Italy and Spain are 1315.5 , 1315 and 1311 respectively and they're decent enough. India has both high life expectancy rates as well as high infant deaths. Germany has low infant death rate 40 and high expectancy rate 1298 whereas Nigeria has a poor expectancy rate of 821.7 but too high infant deaths 8571. From the line graph , we can clearly see that India has both high expectancy rate and high infant death rate. Germany is found to be having decent values for both. Overall , we can see that Infant deaths and life expectancy rates of many overpopulated nations are somewhat inversely related. More the life expectancy , lesser the infant deaths. As we move across the curve , we can see how infant death rates and life expectancies are related. This implies that the infant death rate is more when expectancy is high and mortality is high as in India. Population surely impacts these variables in the course of time. A nation is said to be good when life expectancy is high and death rate is low.

### **Dashboard 3 : Forecast - Adult Mortality & Hepatitis-B ( Brazil )**

**Hepatitis - B :** The trend of Hepatitis B for Year broken down by Country. Color shows details about Hepatitis B and Forecast indicator. The view is filtered on Country, which keeps Brazil. The mark type is Line.

**Country, Hepatitis B , Year**

**Filters:** Country

**Color:** Measure Names and Forecast indicator

### **Trend Lines Model**

A linear trend model is computed for sum of Hepatitis B (actual & forecast) given Year.

**Adult Mortality :** The trend of sum of Adult Mortality (actual & forecast) for Year broken down by Country. Color shows details about Country and Forecast indicator. The view is filtered on Country, which keeps Brazil.

The mark type is Line (Automatic).

**Rows:** Country, Adult Mortality

**Columns:** Year

**Filters:** Country

**Color:** Country and Forecast indicator

### **Trend Lines Model**

A linear trend model is computed for sum of Adult Mortality (actual & forecast) given Year.

**Adult mortality Vs Hepatitis-B :** The trends of sum of Adult Mortality (actual & forecast) and sum of Hepatitis B (actual & forecast) for Year. Color shows details about Year. Details are shown for Forecast indicator. The data is filtered on Country, which keeps Brazil. The view is filtered on Year, which ranges from 2000 to 2015.

### **Sum of Adult Mortality (actual & forecast) Properties**

The mark type is Bar.

**Rows:** Adult Mortality, Hepatitis B

**Columns:** Year

**Filters:** Country, Year

**Level of detail:** Forecast indicator

**Color:** Year

### **Sum of Hepatitis B (actual & forecast) Properties**

The mark type is Line (Automatic).

**Rows:** Adult Mortality, Hepatitis B

**Columns:** Year

**Filters:** Country, Year

**Level of detail:** Forecast indicator

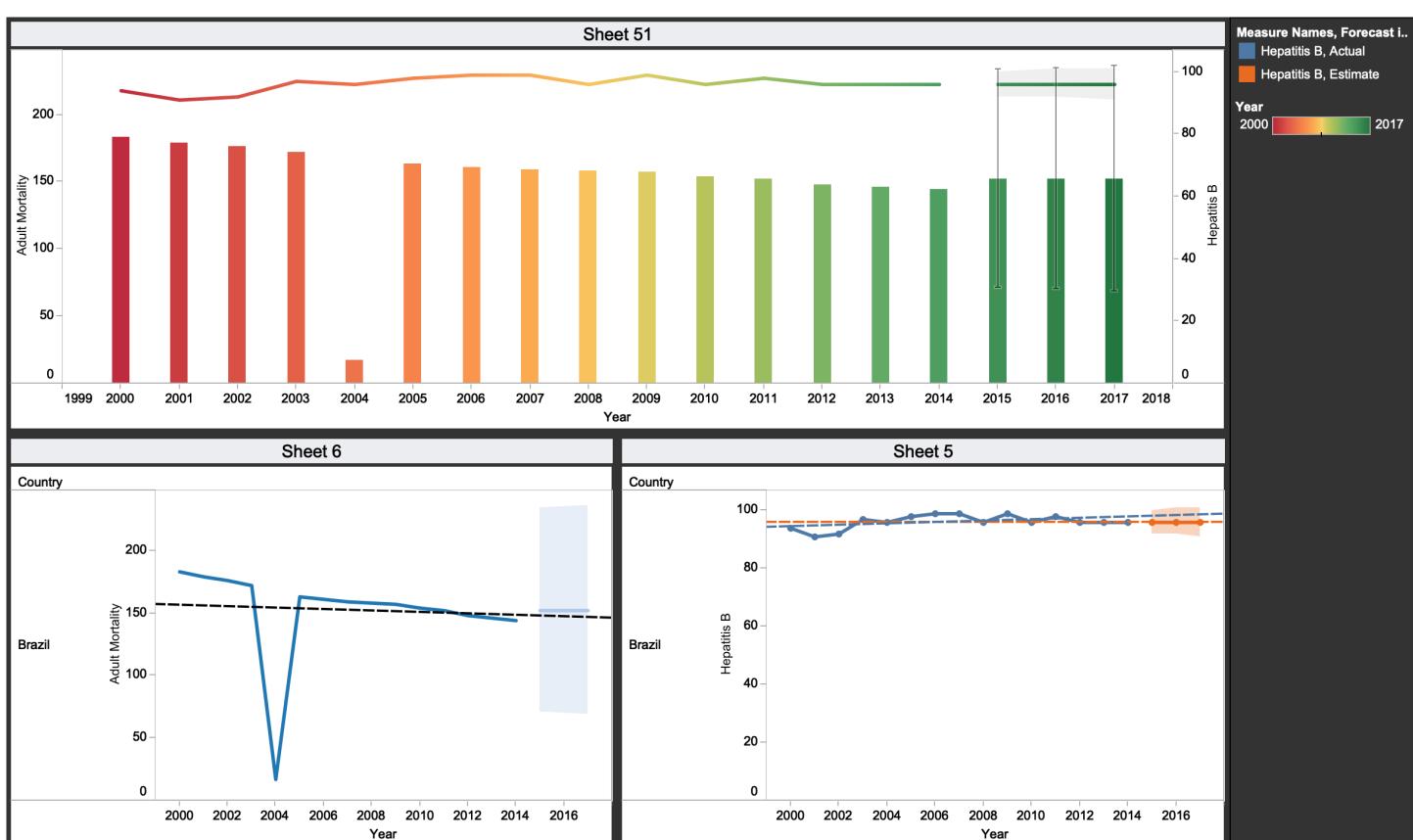
**Color:** Year

Sum of Adult Mortality (actual & forecast) ranges from 17 to 183 on this sheet.

Sum of Hepatitis B (actual & forecast) ranges from 91 to 99 on this sheet.

### **Insights:**

With respect to Brazil , adult mortality has shown a decreasing trend over the years ( 2000 - 2015 ) In 2000 , the value was 183 and in 2014 , it had gone down to 144. The percentage of people affected by Hepatitis-B has ranged around 94 - 96 in the past decade. It hasn't changed much , except for some marginal fluctuations. In the dual combination chart portraying mortality and Hepatitis-B values of Brazil , we can elucidate that there is no strong relationship between the two. Forecast shows that the mortality would rise upto 152 and Hepatitis-B value would be around 96 in the next 5 years. There is no actual cure for Hepatitis - B , sometimes it gets cured on its own but rarely , people never get a complete cure for the disease , in a way becoming carriers . Vaccines are available to prevent the disease . As years pass , if proper medications are found , the rate of victims would come down , gradually decreasing the mortality rate in Brazil.



## **Dashboard 4 : Gross Domestic Products of Various Countries**

### **GDP ( Developed nations )**

Sum of GDP for each Country. The data is filtered on Status, which keeps Developed. The view is filtered on sum of GDP, which keeps non-Null values only.

The mark type is Bar (Automatic).

**Rows:** Sum of GDP

**Columns:** Country

**Filters :** Status, Sum of GDP

**Dimensions :** Country has 28 members on this sheet

**Members:** Croatia; Denmark; Japan; Netherlands; Sweden; ...

**Members:** Developed

### **Measures**

Sum of GDP ranges from 68,143 to 917,806 on this sheet.

### **GDP ( Developing countries )**

Sum of GDP for each Country. Color shows details about GDP. The data is filtered on Status, which keeps Developing. The view is filtered on Country, which keeps 54 of 193 members.

The mark type is Bar (Automatic).

**Rows:** Sum of GDP

**Columns:** Country

**Filters:** Status, Country

**Color:** GDP

**Dimensions :** Country has 28 members on this sheet

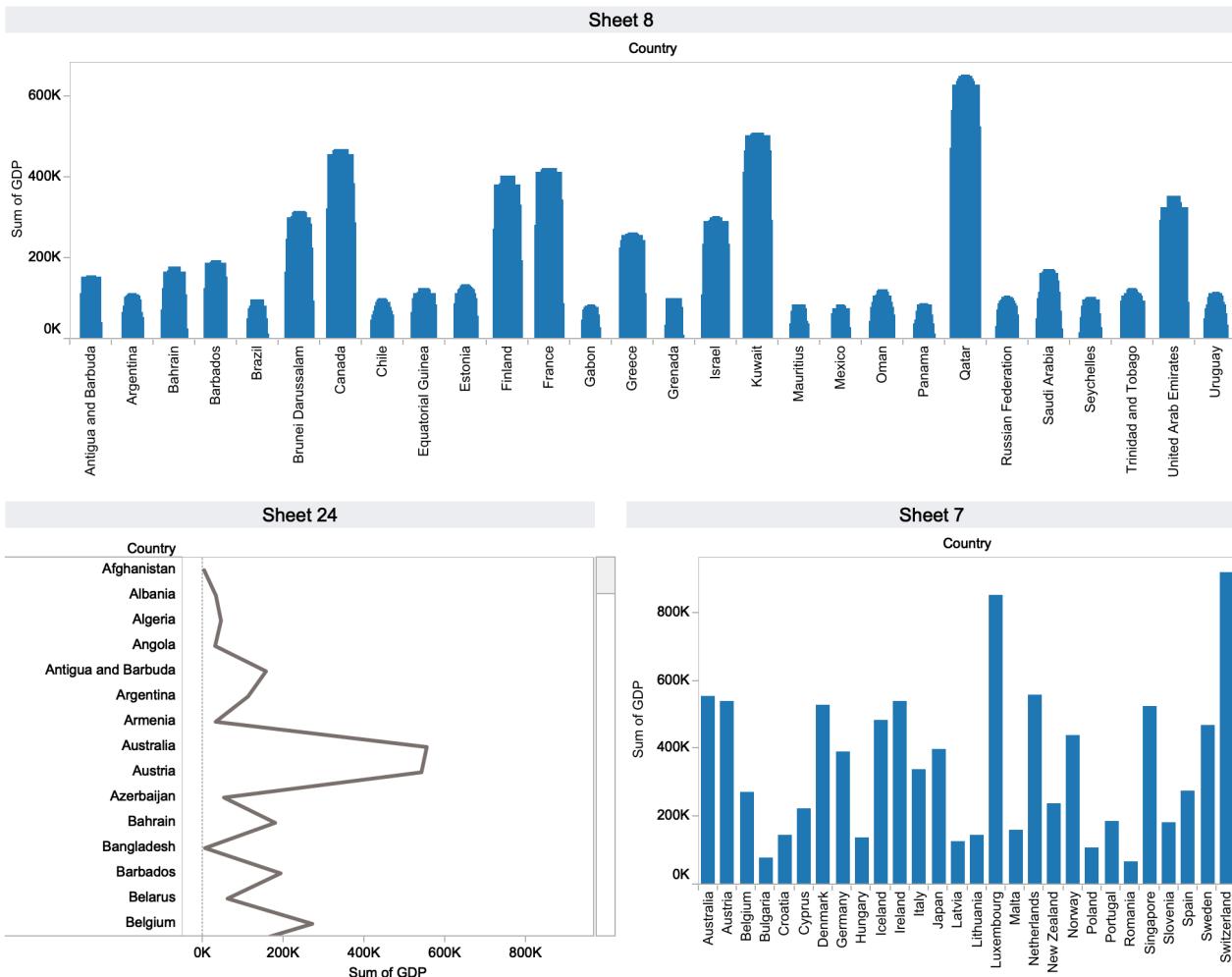
**Members:** Antigua and Barbuda; Canada; Greece; Russian Federation; Uruguay; ...

GDP has 448 members on this sheet

## Members: Developing

### Measures

Sum of GDP ranges from 8 to 88,565 on this sheet.



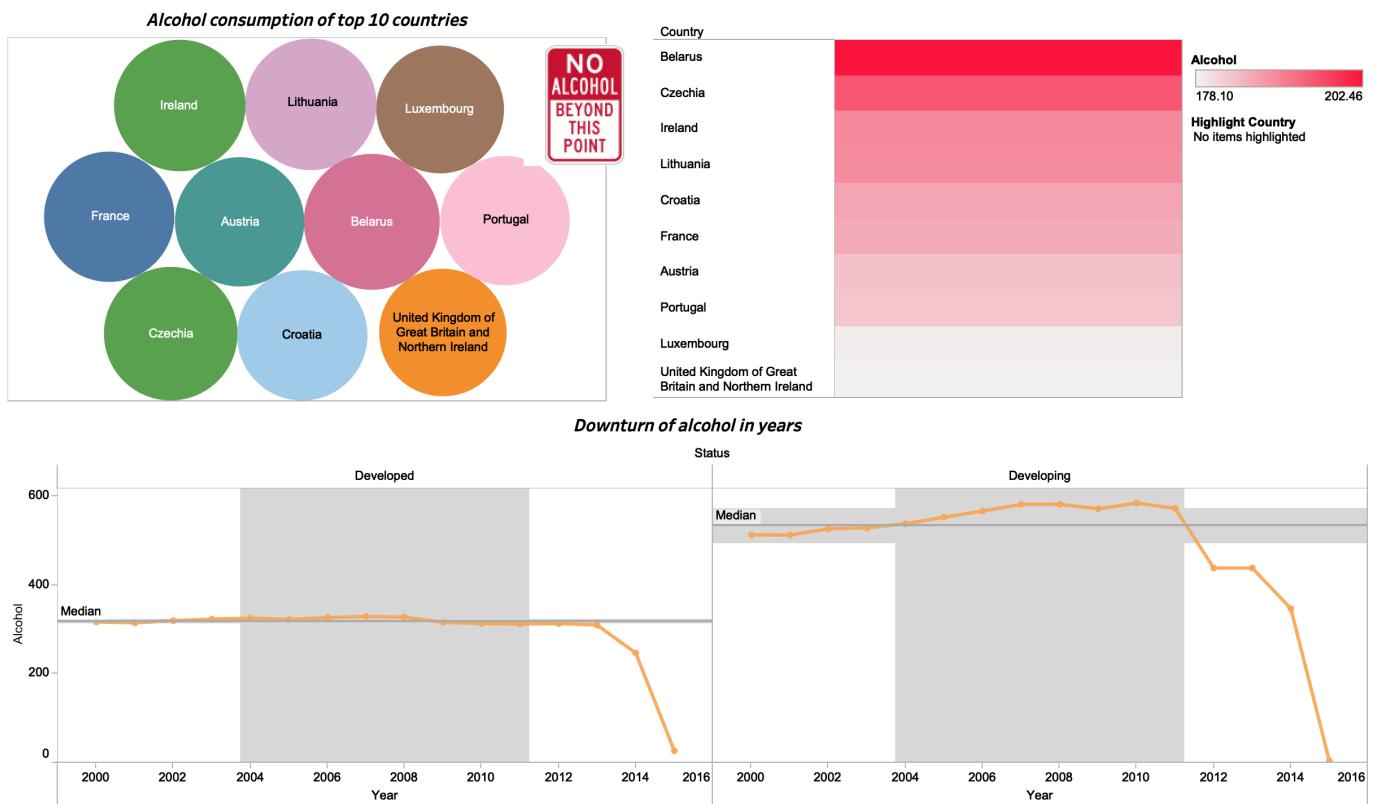
### Insights :

On considering the sum of GDP's of various developed countries selected from the dataset , Switzerland and Luxembourg hold the top places. On the other hand , when developing countries are considered , the middle eastern countries like Qatar , Kuwait , UAE do well. Primarily , the development status of a country is highly influenced towards the GDP of the country. In other words , the more developed a country is , better the GDP , more the budget and ultimately , blessing the nation with a better healthcare and stable society.

## Dashboard 5 : Top 10 Countries with most Alcohol consumption

### Insights :

Belarus has the most alcohol consumption in the whole world. Alcohol consumption in developed and developing countries were plotted separately against the respective years. The line graph shows the downturn of alcohol consumption globally in the past decade. The top ten countries are put in the form of bar graphs where color intensity denotes the alcohol consumption. There is a vast difference in people consuming alcohol taking the data after 2000 time period , which is a positive sign that awareness is being recognised. Hopefully , seeing the trend , we can say alcohol consumption might stay low for the coming years.

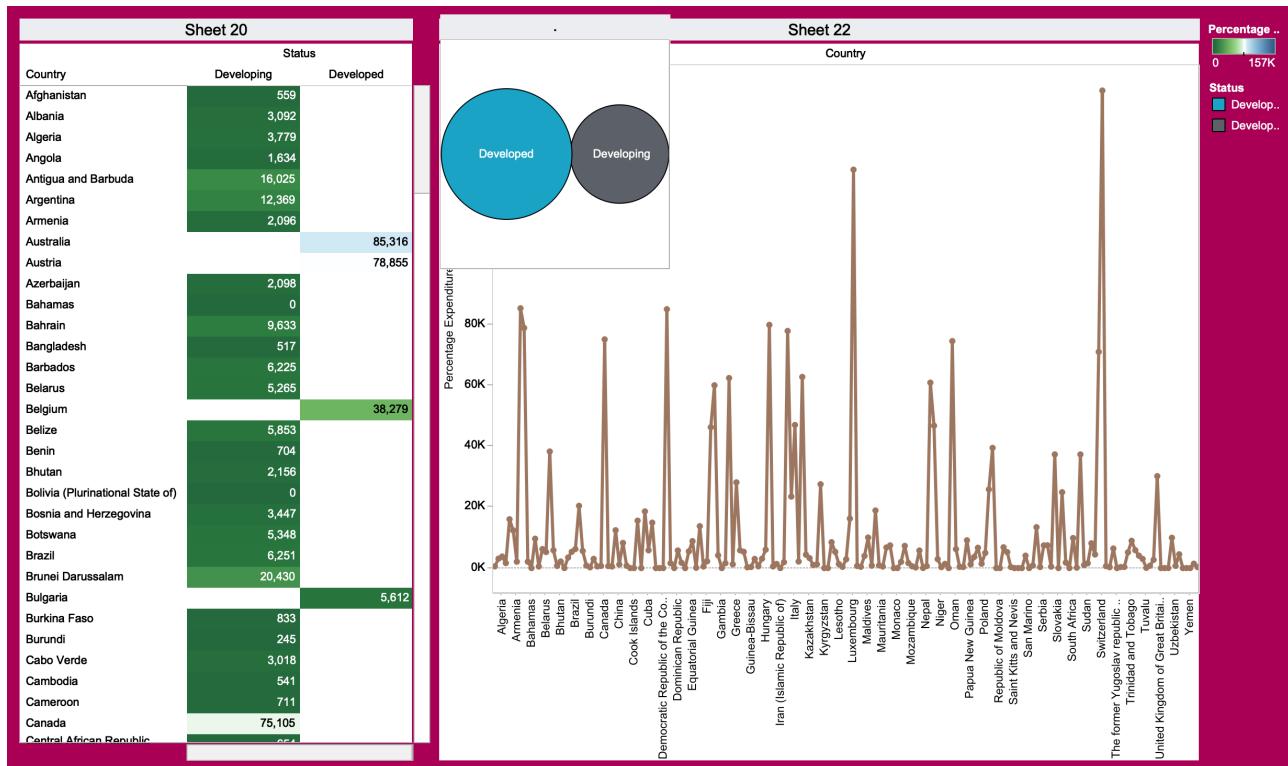


## Dashboard 6 : Country Level Percentage Expenditure Data

### Insights:

Percentage expenditure of various developed and developing countries are visualised in this dashboard. The percentage of developed countries is more than half of developing countries. It is evident that developed countries spend more on

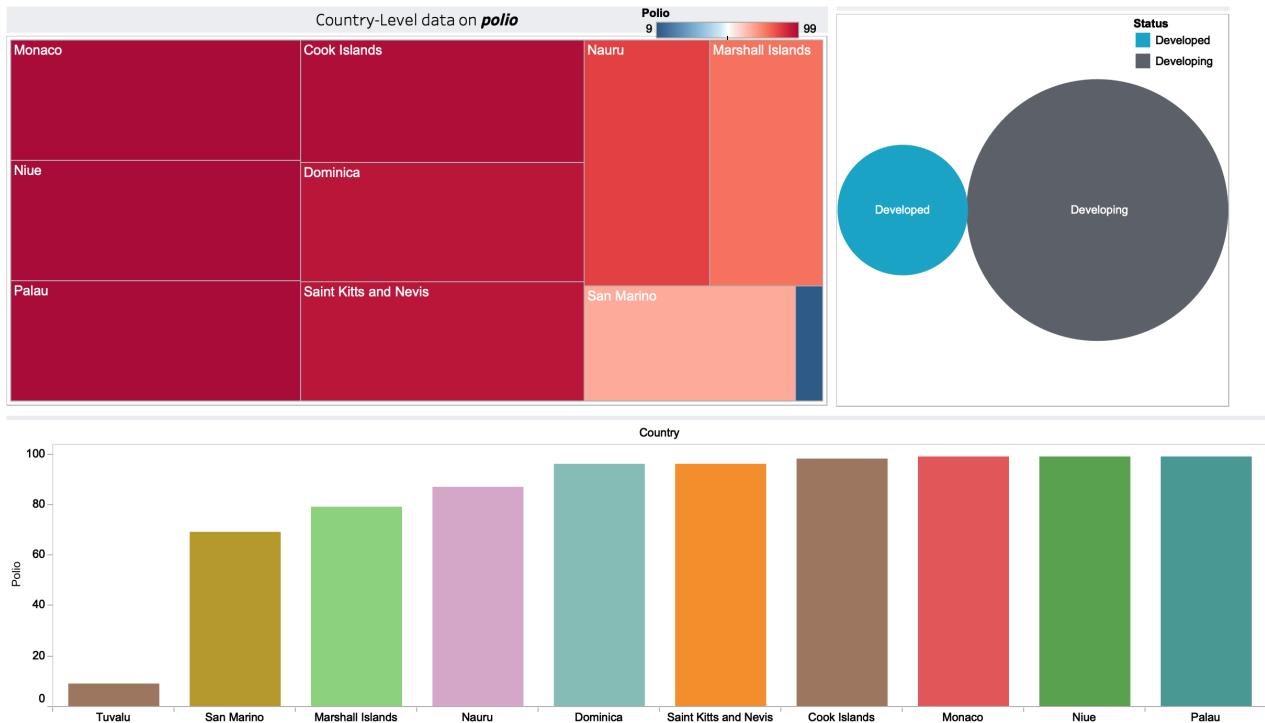
expenses then developing countries. Countries like Luxembourg , Switzerland are very rich and have great economical standards . Some African countries like Sudan , Congo , Namibia have very poor economical standards.



## Dashboard 7 : Country Level Polio Data

### Insights :

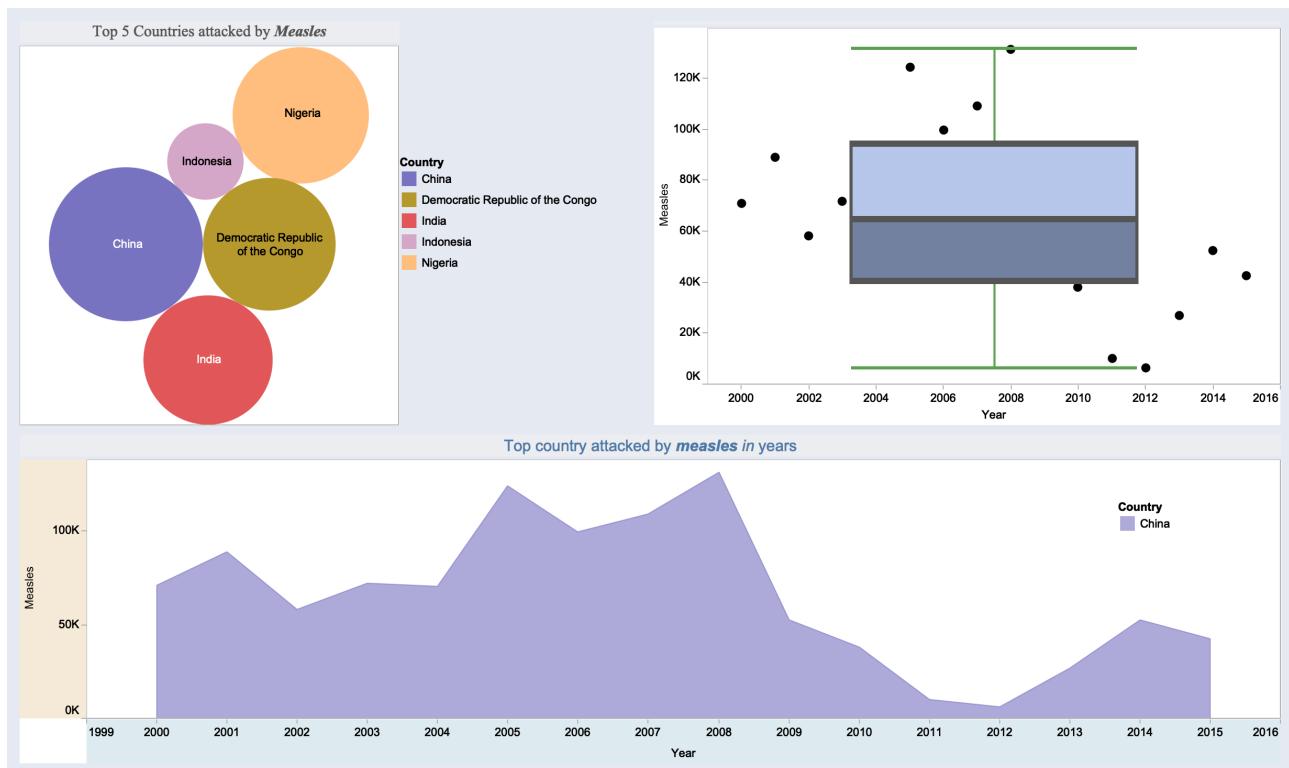
In developing countries , the amount of children affected by polio is far high compared to the developed countries. Reasons might be the percentage of total expenditure on health could be low in slowly developing countries , which in turn reduces the polio vaccination , resulting in more polio attacks. The top ten countries having polio attacked children is mapped using a tree map where colour intensity denotes the number of attacks , Monaco , Dominica , Cook Islands being the highest.



### Dashboard 8 : Measles in different countries.

#### Insights :

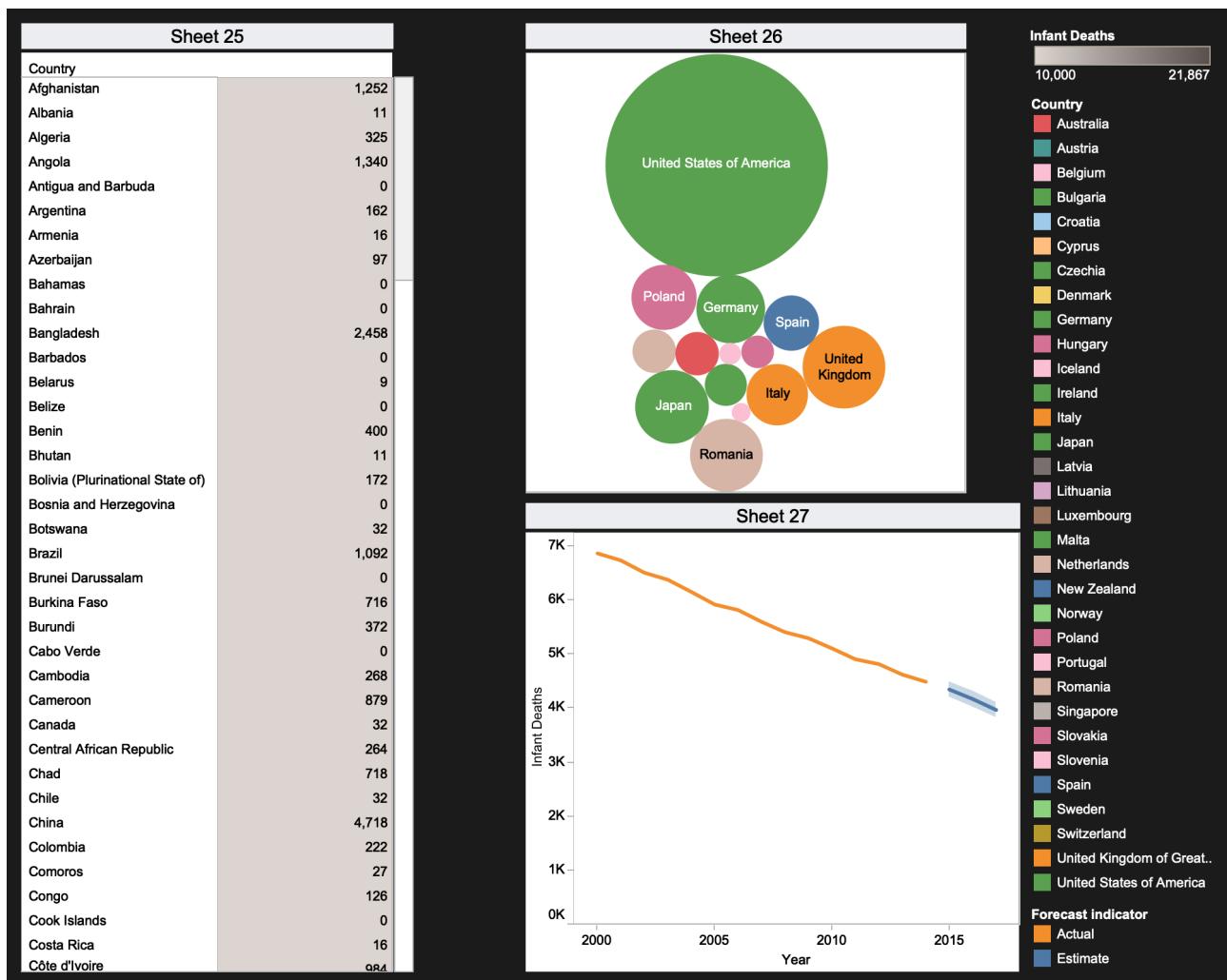
When the number of Measles victims are plotted countrywide and yearwise in the form of box plots and area maps , shows that China , India , Congo , Indonesia and Nigeria are the most affected countries in terms of measles . The average number ranges between 50000 and 80000 . 2005 - 2009 has been the peak period of measles considering the whole world.



## Dashboard 9 : Infant deaths around the Globe.

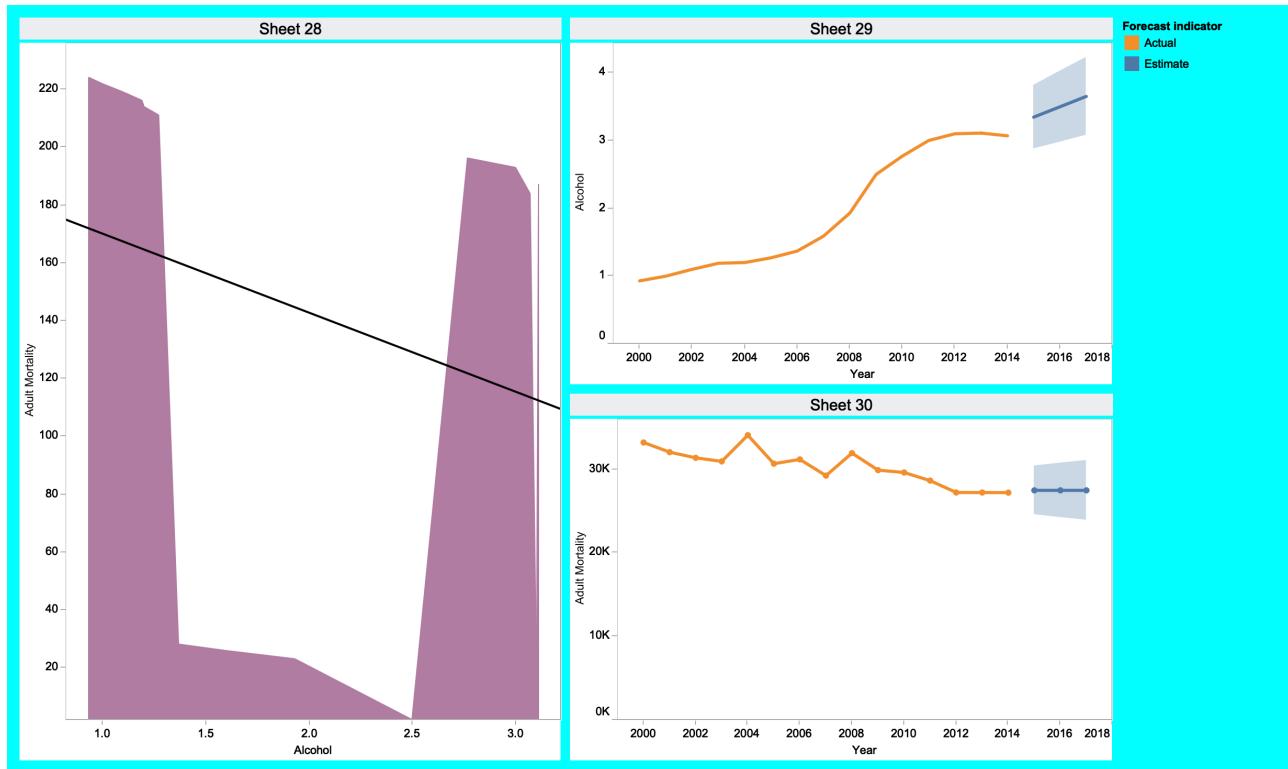
### Insights :

The infant death rates of several countries were fed into a structure to see the trend over the decade. The United States of America stands above all other nations as it has the most number of child death , followed by Uk , Japan and Romania. Forecast shows that the child death rate will drop up to 3000 by the end of 2017. In 2000 , the value was as high as 7000. The child death rate has almost halved in the time period.



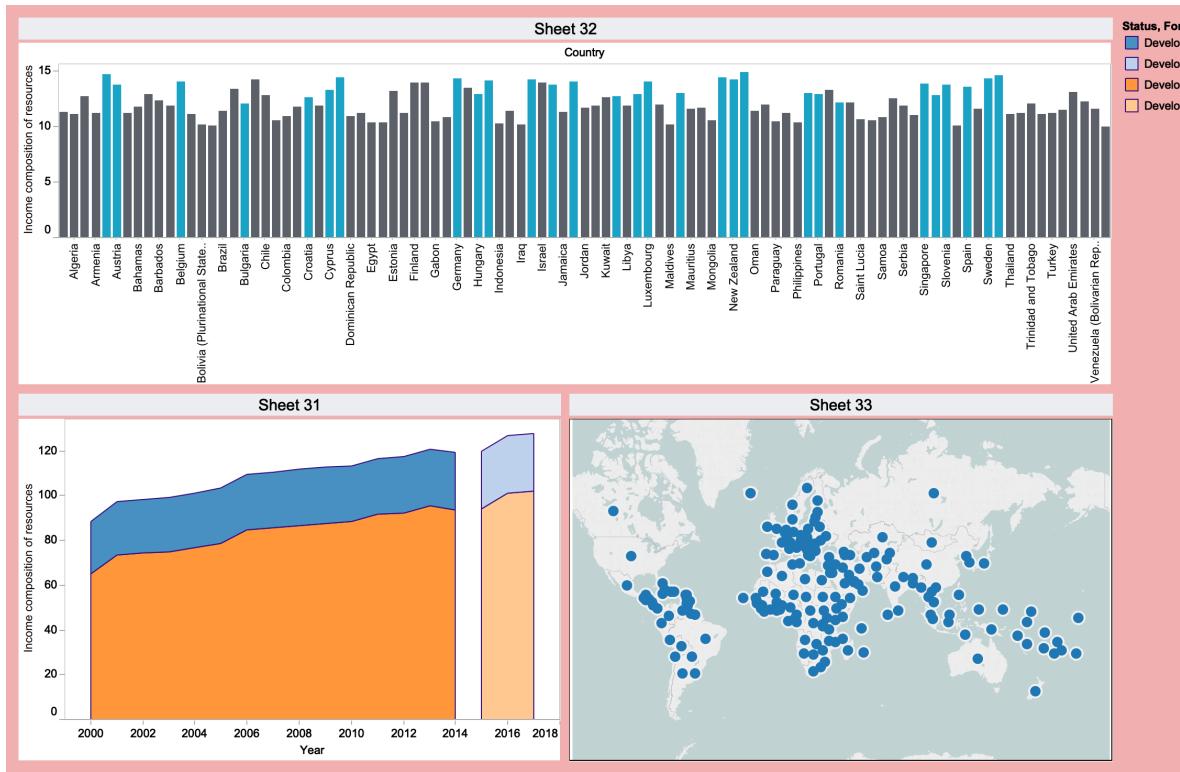
## Dashboard 10 : Adult mortality Vs Alcohol ( India )

Alcohol vs. Adult Mortality. The data is filtered on Country, which keeps India. The view is filtered on Alcohol, which keeps non-Null values only. The alcohol consumption in India has increased dramatically over the past few years. Prediction is that its will move up to almost three times from where it started , from the line curve. Globally , alcohol consumption is being controlled. It has come down a lot comparatively. The actual and estimate lines from both graphs show us the forecast of alcohol consumption.



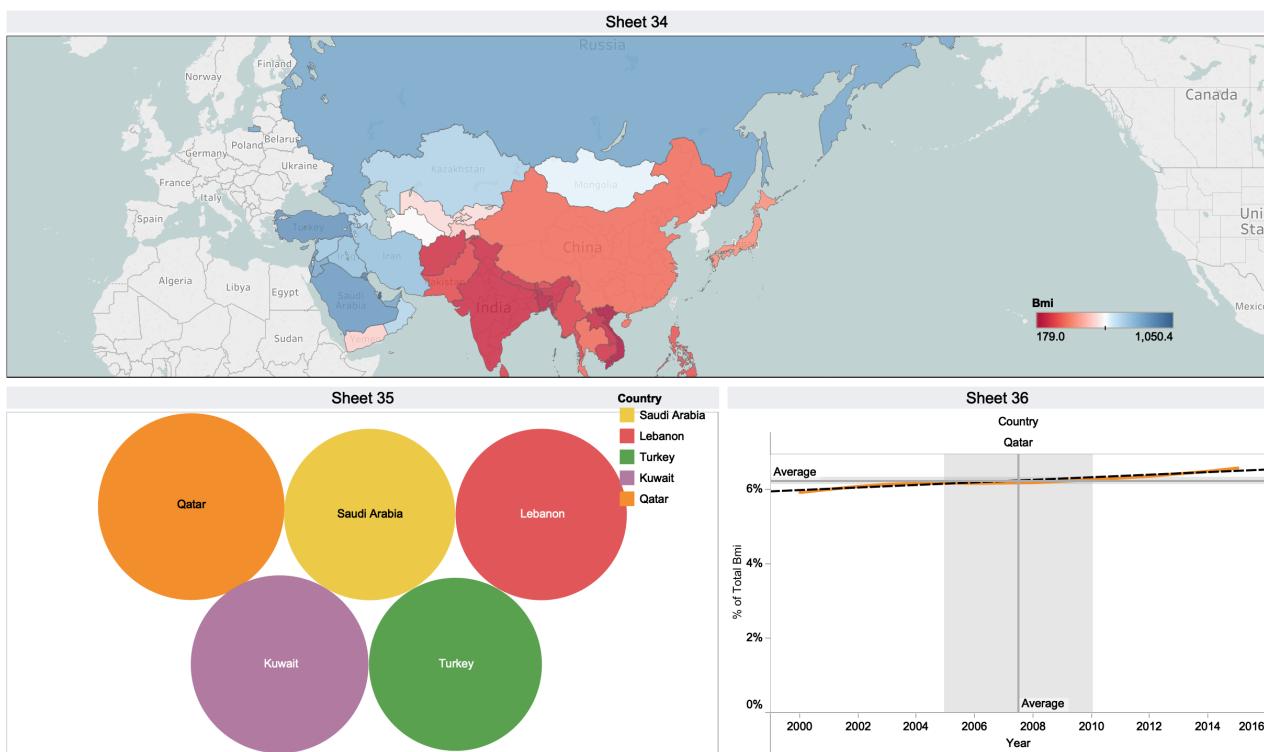
### **Dashboard 11 : Income Composition of Resources ( Human Development Index )**

The income composition of resources also called human development index of both developed and developing countries are visualised to determine relationships. The bar graph represents the income composition of the various nations and those nations are categorised as Developed, Developing. The income composition of resources for various years are analysed and the income composition is gradually increasing in the upcoming years. It is also predicted that there will be an increase in the future years (2015 to 2017). The same is also plotted for various countries across the globe.



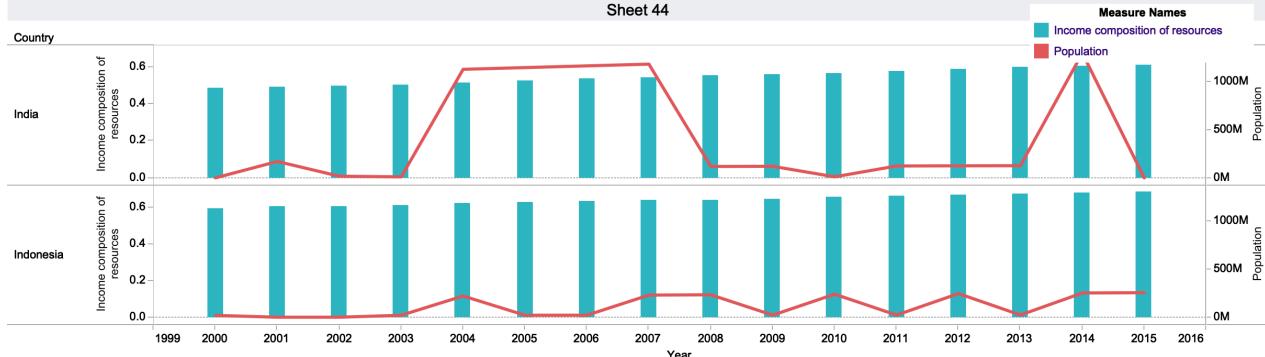
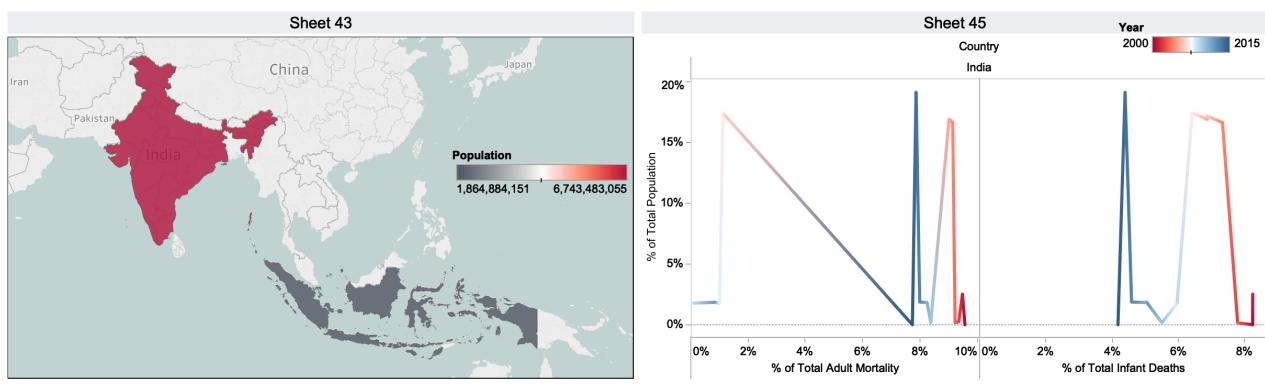
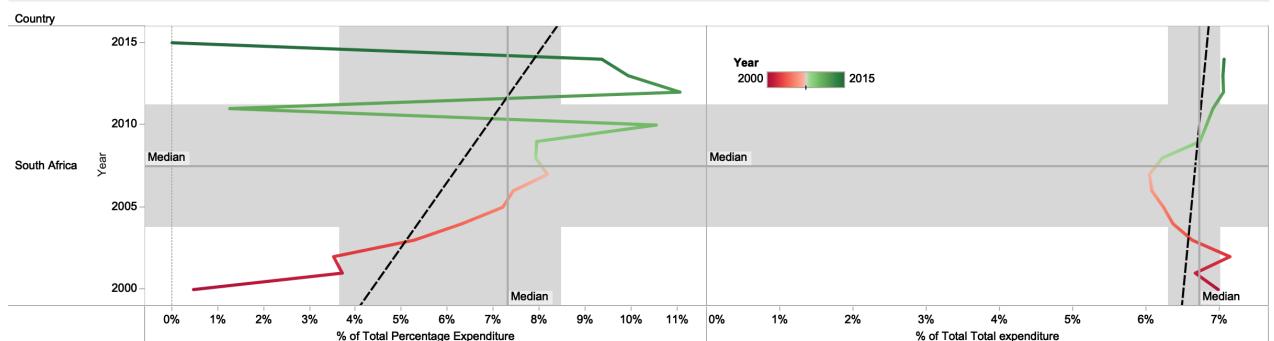
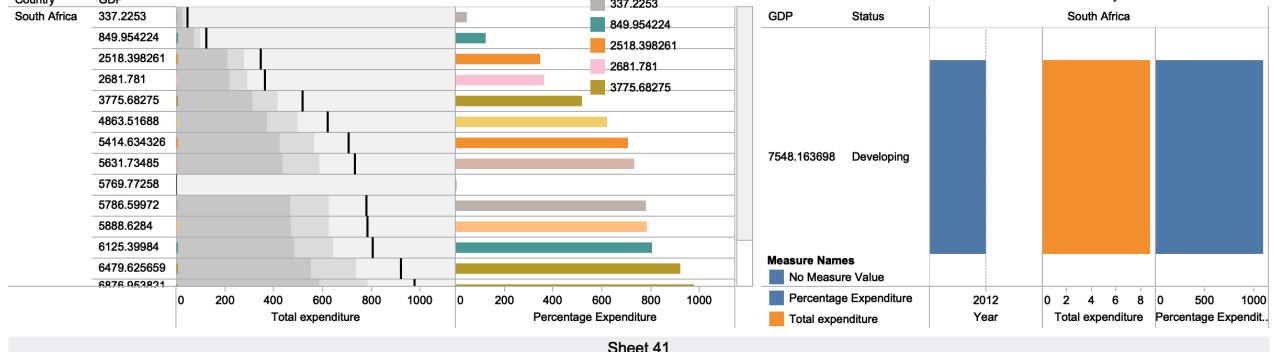
### Dashboard 12 : Body Mass Index of Countries

Body Mass Index of people across the different countries are analysed and visualised in this graph. Qatar is the leading country among all its rivalries. By 2000, the percentage of total bmi was around 6% and till the end of 2016, there was a very slow increase in its BMI growth. By these analysis, Qatar is the most healthiest nation in the world.



### **Dashboard 13 : Schooling - India & Ireland**

The following dashboard represents the visualisation of schooling in India and Ireland. The percentage of schooling is comparatively lesser than in Ireland when compared to India. In later years, the percentage of Schooling in Ireland gradually increased. Schooling in India was always the same and developed a lot in recent times.



## Dashboard 14 : GDP Vs Total Expenditure of South Africa

The visualisation represents the analysis of the difference between the Total expenditure and GDP of South Africa. GDP was comparatively lesser than the Total expenditure in the developing nation. Total expenditure is same as that of the GDP during the mid years and it varies from 2000 to 5000.

## Dashboard 15 : India Vs Indonesia ( A comparison )

India and Indonesia are compared in this visualisation. Adult mortality rate are analysed for two different nations. Adult mortality rate of Indonesia is high compared to India. There was a slight variance in the adult mortality rate for both the countries.

