# Introduction

Using the Global health and nutrition data, which includes 345 indicators across 263 countries around the world, the below report explains various visualizations used to do an “**Exploratory Analysis**” on this dataset. Data was collected on a yearly basis from 1960-2016 and consist of themes like population dynamics, nutrition, reproductive health, health financing, medical resources and usage, immunization, infectious diseases, HIV/AIDS, DALY, population projections and lending.

Since there is no evident conclusion when examining raw data, due to factors like missing values, scattered fields, the variety and the volume, these visualizations play an important role in analysing the information given and to make conclusion, as well as comparison, on such a vast data. The conclusions made from these different graphs can in turn be related to build the knowledge required to make decisions regarding our health systems. Here are few important links to this project:

1. Dataset: <https://data.world/data-society/global-health-nutrition-data>
2. Visual Artefact: <https://akashverma5.wixsite.com/mysite>
3. Presentation Video: <https://vimeo.com/409497181>

# Description

The dataset has a lot of information to offer and can be explored in multiple ways. However, looking at indicators while trying to summarize it into a progressive relatable thread of information to build a knowledge, I have divided data into following set of visualizations:

**2.1. Mortality:**

To understand a health system, it makes a logical sense to look at the morality rate and observe what’s the rate of death along with factors causing it. Fortunately, this dataset offers mortality rate and has categorised it into gender and age group. This might be enough to build a visualization. But to make it relatable to our next visualization, there is an extra category of income group that will help understand this data better as we further explore this analysis.

|  |  |  |
| --- | --- | --- |
| **IDIOM** | **Sunburst Graph** | Visualization |
| Data | It represents Two-dimensional table with Multiple categorical key attributes (Sex, Age, Income group) and one quantitative value (Mortality rate per 1000). | Figure 1 |
| Task | Discover Trends/Represent Data/Categorize Values. |
| Visual Encoding | Each circle encodes different category. Arc encodes mortality rate for respective category. |
| Approach | Data pre-processing, filtering, data transformation and visualize using Python’s Plotly library. |

There is another visualization of bar graph showing population division for each income group to make sense of data seen in sunburst graph.

|  |  |  |
| --- | --- | --- |
| **IDIOM** | **Stacked Bar Graph** | **Visualization** |
| Data | It represents Two-dimensional table with two categorical key attributes (Sex & Income group) and one quantitative value (Percentage of Population). | Figure 2 |
| Task | Compare different categories based on numerical value. |
| Visual Encoding | Size of bar encode population size. Color encodes sex. |
| Approach | Data pre-processing, filtering, data transformation and visualize using Python’s Plotly library. |

**2.2. Life Expectancy:**

After knowing mortality of different income group, we now look at timeline data from 1995 to 2014 to understand how health expenditure has affected life expectancy. This will allow us to see the correlation of health with economy.

|  |  |  |
| --- | --- | --- |
| **IDIOM** | **Scatter Plot** | **Visualization** |
| Data | It represents Two-dimensional table with two categorical key attributes (Year & Income group) and two quantitative value (Life Expectancy at Birth and Health expenditure). | Figure 3 |
| Task | Discover Trends/ Find Outliers, distribution and correlation. |
| Visual Encoding | Point marks. Express with horizontal and vertical position. Size represents population. Motion represents Year. |
| Approach | Data pre-processing, filtering, data transformation and aggregation. Visualized using Python’s Plotly library. |

**2.3. Diseases:**

Its time now to explore the actual cause of these deaths and see what effects the health system most.

|  |  |  |
| --- | --- | --- |
| **IDIOM** | **Geospatial Map** | **Visualization** |
| Data | Geographic geometry data. Table with multiple categorical key attributes and their respective quantitative value per region. | Figure 4 |
| Task | Analyse geographically distributed values, trends; comparison, search, analyse. |
| Visual Encoding | Dots: countries geo location on map. Colour: Different categorical key defining disease. Size: Respective quantitative values. |
| Approach | Data pre-processing, filtering, data transformation and aggregation. Visualized using Python’s Plotly library. |

**2.4. Hospitals:**

This visualization will show how capable we are in tackling these diseases?

|  |  |  |
| --- | --- | --- |
| **IDIOM** | **Pictograph** | **Visualization** |
| Data | Two-dimensional table with multiple categorical key attributes and their respective quantitative value per 1000. | Figure 5 |
| Task | Compare multiple categories using numerical values. |
| Visual Encoding | Image encode category. Repetition encode value. |
| Approach | Get Data, Find image, Loop repetition. Visualized using Processing. |

**2.5. HIV:**

Finally, we look at an example disease of HIV.

|  |  |  |
| --- | --- | --- |
| **IDIOM** | **Infographic** | **Visualization** |
| Data | Two-dimensional table with multiple categorical key attributes and their respective quantitative. | Figure 6 |
| Task | Discover/present/rank/categories |
| Visual Encoding | Icons encode category. Jigsaw design encode separation of category. Background image encodes data being talked about. |
| Approach | Get Data, find image background, icons and boundary separator, Edit the whole using image editor Paint3D. |

# Results

This series of visualization starts exploring the world health data by looking at mortality rate where it finds the value to be lesser for higher economies whereas the division of death among gender remains same for all. Moving to Life expectancy chart, it shows how average life has increased over time with some level of correlation with the amount of expenditure. However, there are extreme outliers on both ends of expenditure defying the general norm. Further analysis of diseases on world map shows how differently diseases impact different regions of world. It’s no surprise to see that majority of population has malnutrition of some kind rather than disease. The next visualization of hospital is important as it simply shows how different factors of life, death and hospitals are related. It clearly shows that our health system is ill equipped to handle any major crisis. Finally, the HIV example states some numbers that could be related to understand why this disease is still spreading?

Here are some strength and weakness of this visualization:

* 1. **Strength**

1. Dynamic: Visualizations have motions to represent data with extra information when you hover over data point.
2. Concise: Multidimensional keys have been aggregated and compared in simple ways that are simply too complex to read through raw data.
3. Interactive: It allows data to be zoomed, filtered on different axis and have interactive legends.
   1. **Weakness**
4. Missing Data: In some cases, there was no raw data for same year, hence data some previous years had to be taken to complete visualization.
5. Exactness: Although we can see multi dimensions and their correlations. It can sometimes be hard to understand exact numbers behind the visuals.
6. Missing knowledge: To keep the project concise not all data could be visualized and has many more unexplored information.

# Novelty

Novelty of this visualization lies in the way this data has been visualized. Even though there exists multiple Kaggle projects for this dataset, none of them have visualized the dataset as its been done here. Some of the key factor that lies in this exploratory analysis is how each visualization is related and represents a continues stream of knowledge that can be summed to make logical decision or understanding of our health system.

# Citing third party resources

Data was collected from data.world [1]. While most libraries are default in python, some third-party libraries used included Plotly [2] and Matplotlib [3]. Figure 5 and Figure 6 were created using Processing [4] and Paint3D [5] respectively.

# References

|  |  |
| --- | --- |
| [1] | "Data.World",[Online]. Available: https://data.world/data-society/global-health-nutrition-data. |
| [2] | "Plotly",[Online]. Available: <https://plotly.com/python/>. |
| [3] | " Matplotlib ",[Online]. Available: <https://matplotlib.org/>. |
| [4] | " Processing ",[Online]. Available: <https://processing.org/>. |
| [5] | " Paint3D ",[Online]. Available: https://www.microsoft.com/en-us/p/paint-3d/9nblggh5fv99. |