Is world a better place than we think?

Team: Nishant Jain (nishjain@iu.edu)

Abstract:

This project is about getting insights from the World Development Indicator dataset provided by the World Bank and explore other related datasets, and in the process validating the results presented in the book "Factfulness" by Hans Rosling. And finally being able to at least explore if not answer the question: is world a better place than we think?

Introduction:

Motivation:

Going through the book "Factfulness" (Rosling), I was both shocked and amazed by the lack of knowledge I had about the world, and how humanity has progressed in the recent years. The progress has been made slowly over the years and hence most of the people didn't pay attention about what the world has become today. Most of the people still live under the perception of world as it was 30-40 years ago. To test the people around the globe, author had devised a test about some of the world level statistics, and surprisingly people performed worse than the apes would have performed in almost every question. The people we are talking about are the most educated people across the world, members or attendees of conferences at UN and World Bank, and world-renowned journalists who keep themselves updated about the world every day.

After going through the book, I was so excited and motivated to explore more and tell everyone about it. This is something that everyone on the earth should know. Today the world is filled with negativity instinct, major reason being the consumption of increasing negative news through the media. The TED talk by Hans Rosling (TED) and website (www.gapminder.org) provides more info about the study and the test.

Background:

The existing work on this is presented on the Gapminder website where they also provide the aggregated data sourced from various sources like the World Bank, United Nations and other relevant organizations. They have also provided online tools to visualise the data on the web which mainly focuses on interactive line charts and bubble charts with animated timeline to showcase the change over the years. I really liked the idea of the bubble charts, firstly because one can represent three dimensions in 2D using them, and secondly, they provide a comparison between two parameters as they are basically scatter plots with size of markers being the third dimension. In most of the cases they have used the population size as the third dimension, which constantly reminds user about where the most population live, for example when looking at poverty indices, one can see that India and China are the two biggest bubbles, and how China has progressed much faster after 1990s. But in some cases, I feel the bubbles are redundant and interfere with the presentation, especially when there are too many countries represented. Bubbles look good when they represent the world in regional divisions provided by the world bank (Bank). And it is good to

introduce with first the population and economy level bubble charts, and then later just show the scatter plots and line plots incorporating some representative countries and using colours to distinguish between them.

Another relevant and beautiful work is done by the OurWorldinData (https://ourworldindata.org), they have performed similar work done by Gapminder but in a different and more organized way. They have used the similar sources and sometimes used the data directly from Gapminder. I really liked how they have focussed on each topic in detail exploring the history, the dynamics, and other parameters around it. In terms of visualizations, they have focussed more on the geographical maps i.e. choropleths which are interactive, and I also liked how they have used an accurate representation of world map instead of using Mercator. They have provided line charts as well (most of the time interactive) in many cases, but by default the choropleths are shown.

Objectives:

The objective here is to go through the facts given in the book and represent them using visualization. Wherever possible, explore the other things around the ones mentioned in the book and depict what the author has tried to convey, again through visualizations.

This will help in validating the viewpoints and results presented by the author and will help in exploring more around the topic. Also, one has the opportunity to recreate/improve the existing visualisation in visually better way, as well as to create interactive visualizations. And lastly, create a concise document listing all the key points and visualisations at one place so that one can go through it as a summary by looking at the visualisation instead of going through the whole book.

Data & Methods:

The data in focus here is mainly the WDI data from World Bank, because it is the most trusted data and being used along all the other data in the previous works mentioned above. This data is very comprehensive and is available for all the countries in the world, as well as organized in the form of group of countries based on regional and economic criterions. The data is available freely to download from the World Bank website, as well as, it is offered through many APIs in Python, R and other tools.

Apart from this, the data from the Gapmider has been used in few cases where there was a need to have more historical data as well as the data predictions. The predictions in the data are all given by the UN. Hence, it is a trusted source of data, and this has been verified by matching the data with the WDI dataset.

First step was to tidy the data as all the visualization tools and libraries work best with the tidy data, for that purpose pandas library have been used. For data visualization, I started with using matplotlib and seaborn as they are robust and have excellent community support and documentation along with examples. But soon I realized that I needed something more, one option was the altair library which has been used in the course as well, but I wanted to try something different and more intuitive. Then I came across plotly, which offers excellent tools to create vastly different types of visualization. And along with plotly, the cufflinks library provides support to use it directly with pandas as an API, hence making the process easier and quicker. And another huge advantage of plotly is that all the plots created by it are interactive, hence are a better alternative to matplotlib plots when creating and sharing Jupyter notebooks. Plotly helped in creating very nice looking and easy to understand bubble plots. It offers three

themes as well as different color schemes. Wherever cufflinks was not sufficient, plotly's express library has been used.

I have restricted myself to mostly bubble charts, scatter plots, and line charts. The reason being the easy interpretability of these charts. Color labelling has been used across all the plots to distinguish between different series in the same plot. The advantage of bubble charts is that they help in visualizing the countries and geographies in a better way, as the size of the bubble can represent a third dimension which helps in differentiating between them and provides an added perspective to the plot and helps in narrating the story. The disadvantage is that many times there is overlapping due to the size of one bubble being too big. Though this is addressed using transparency in the bubbles, but it still makes the visualization a bit crowded. Another problem is that in case of time series data only one snapshot of the data can be shown. Hence, the line plots come handy in this case, showcasing the change over the time. That's the biggest advantage of line plots, but they can only display limited information as only one factor can be plotted at once, as well as, not many countries can be shown at the same time.

The approach followed is to complement bubble/scatter plots with line plots wherever required.

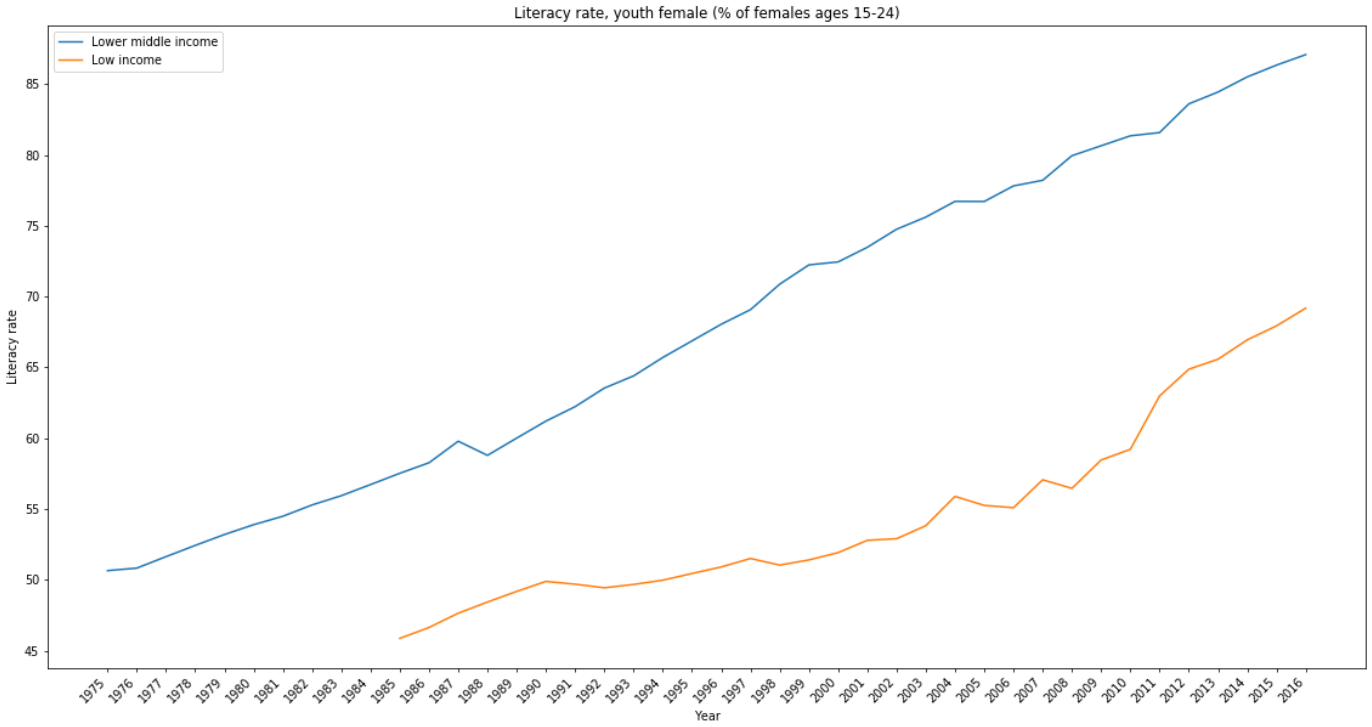
Results and Insights:

Primary education in Females:

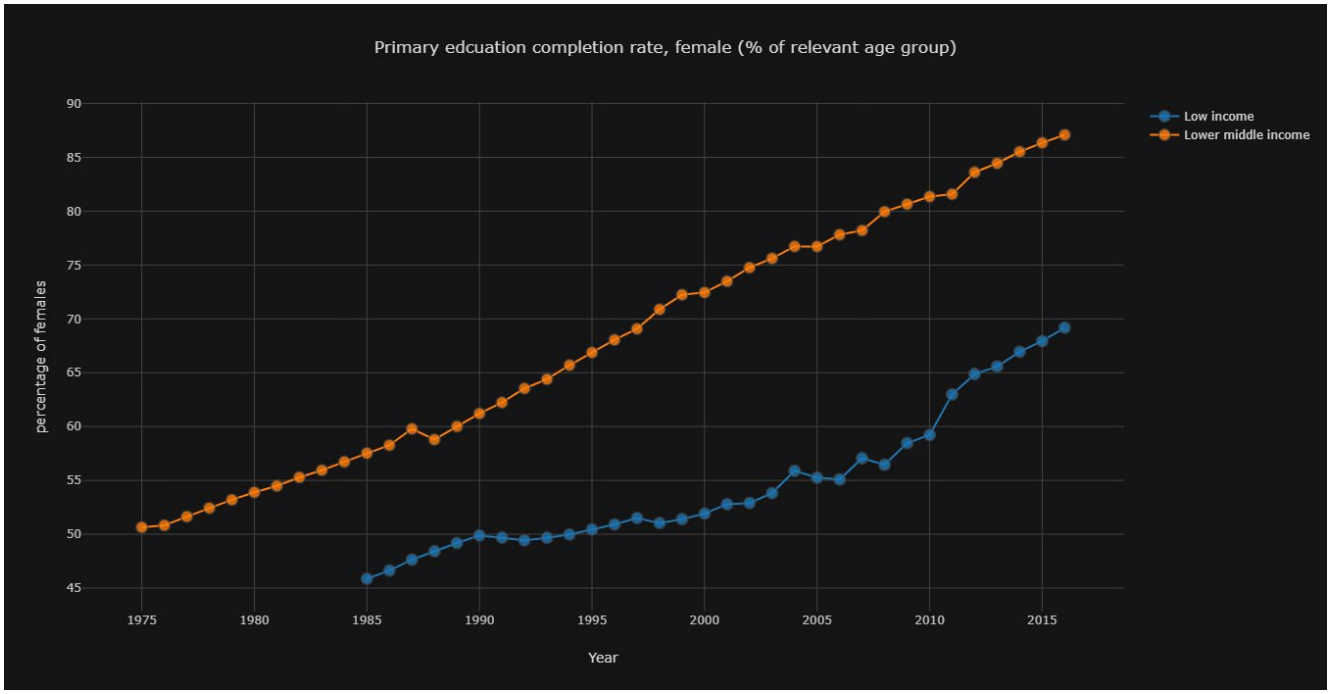
The first question the author asks in the book is:

Q1. In all low-income countries across the world today, how many girls finish primary school?
A: 20 percent
B: 40 percent
C: 60 percent

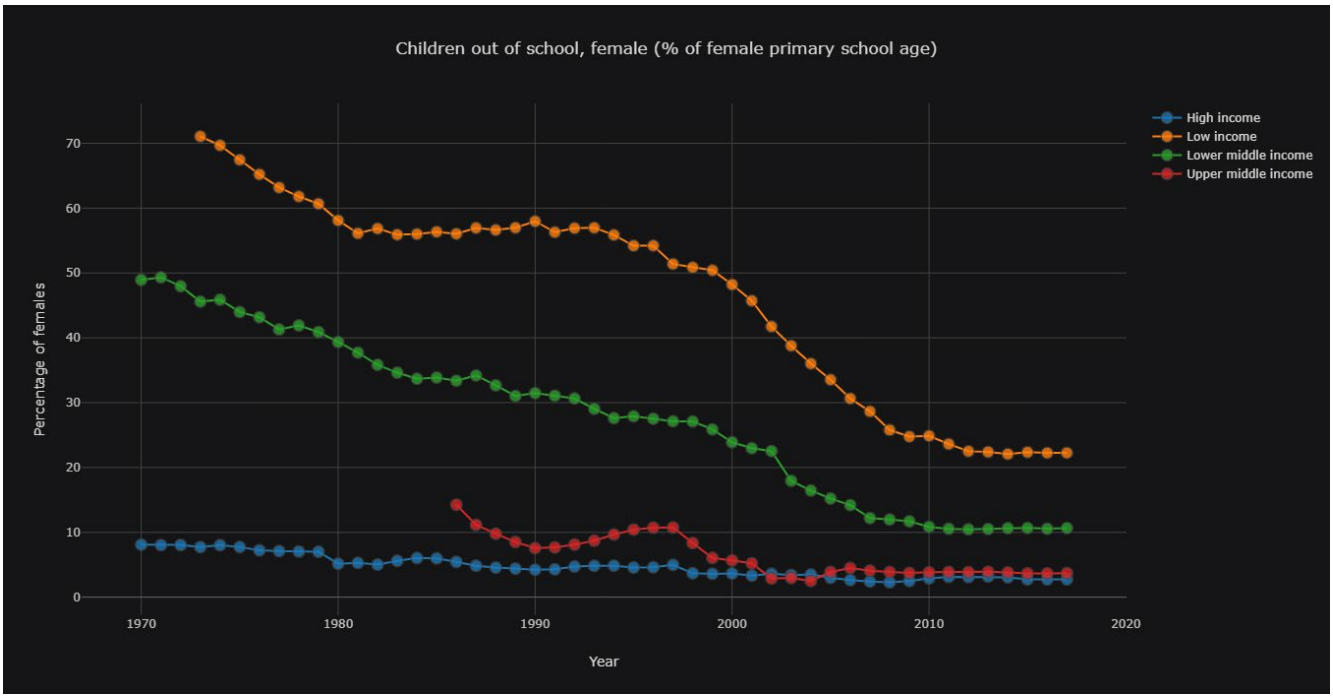
The right answer can be interpreted by the plot below, which is 70% and seems shocking at first. The below plot is created using seaborn and matplotlib. But probably I have a better one.



This is a plot generated by plotly, I have used ‘solar’ theme which makes the background black, and in actual the plot is interactive in nature if viewed in a browser.

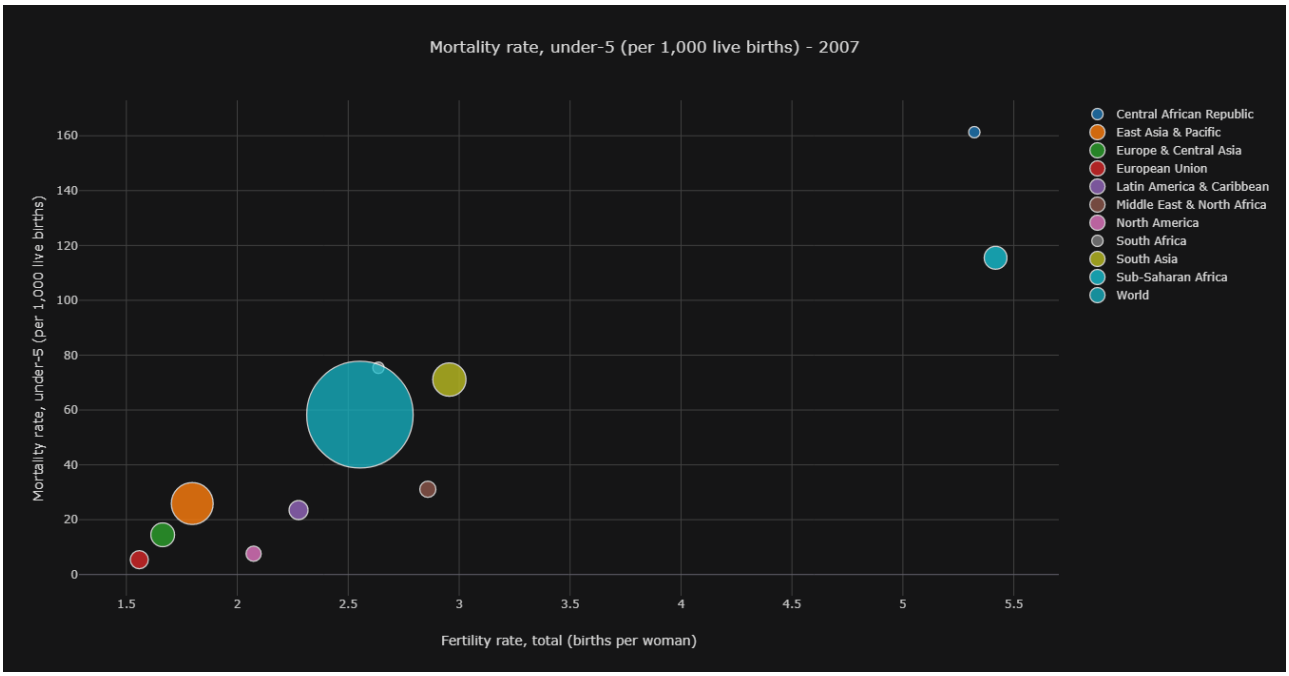


To further explore this, I looked at another parameter, which is school dropout rate. And clearly it has improved a lot specially in low income countries.

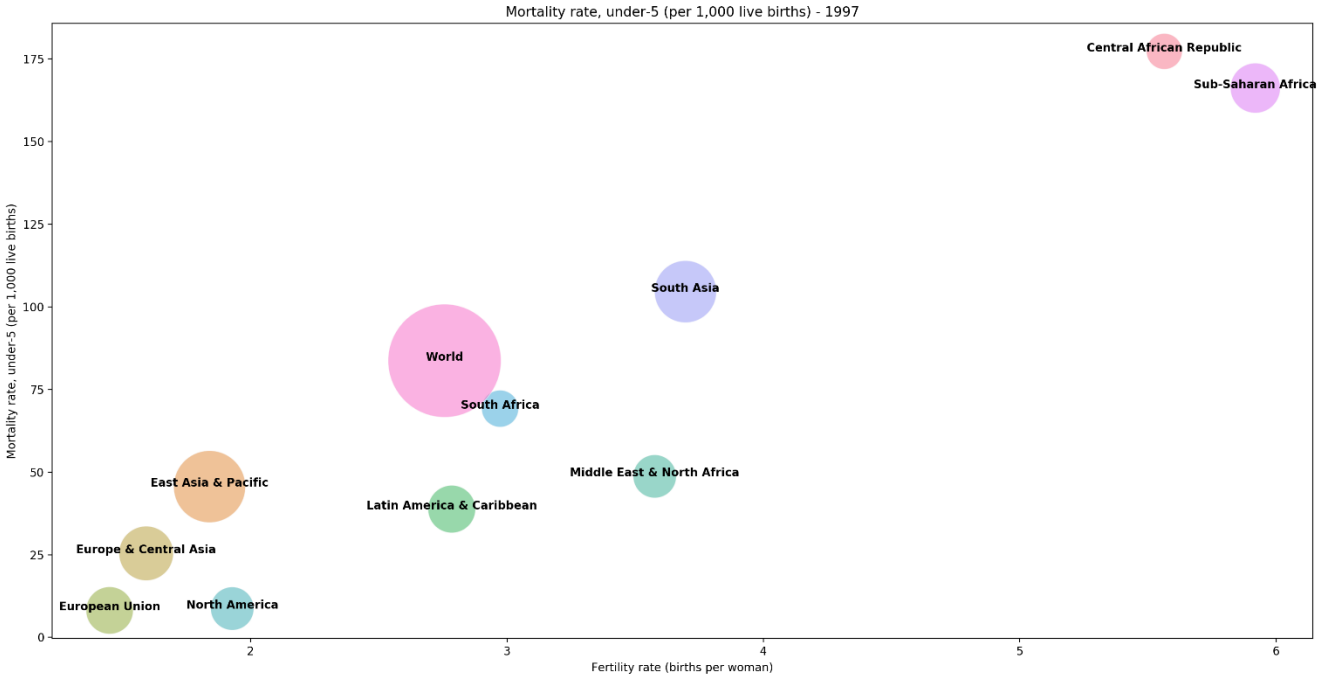


Child Mortality Rate:

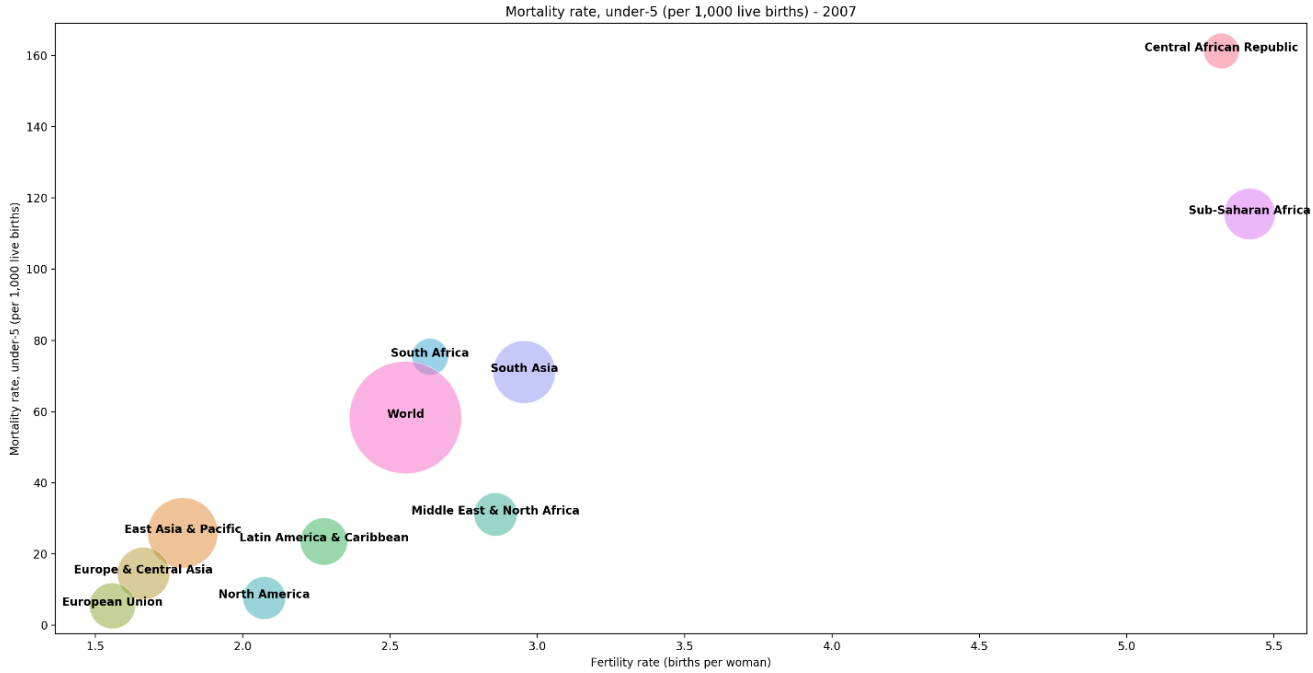
Now we look at child mortality rate across some selected countries from each region in the world to gauge the difference. I have plotted a bubble plot for mortality rate and fertility rate with total population as the bubble size.

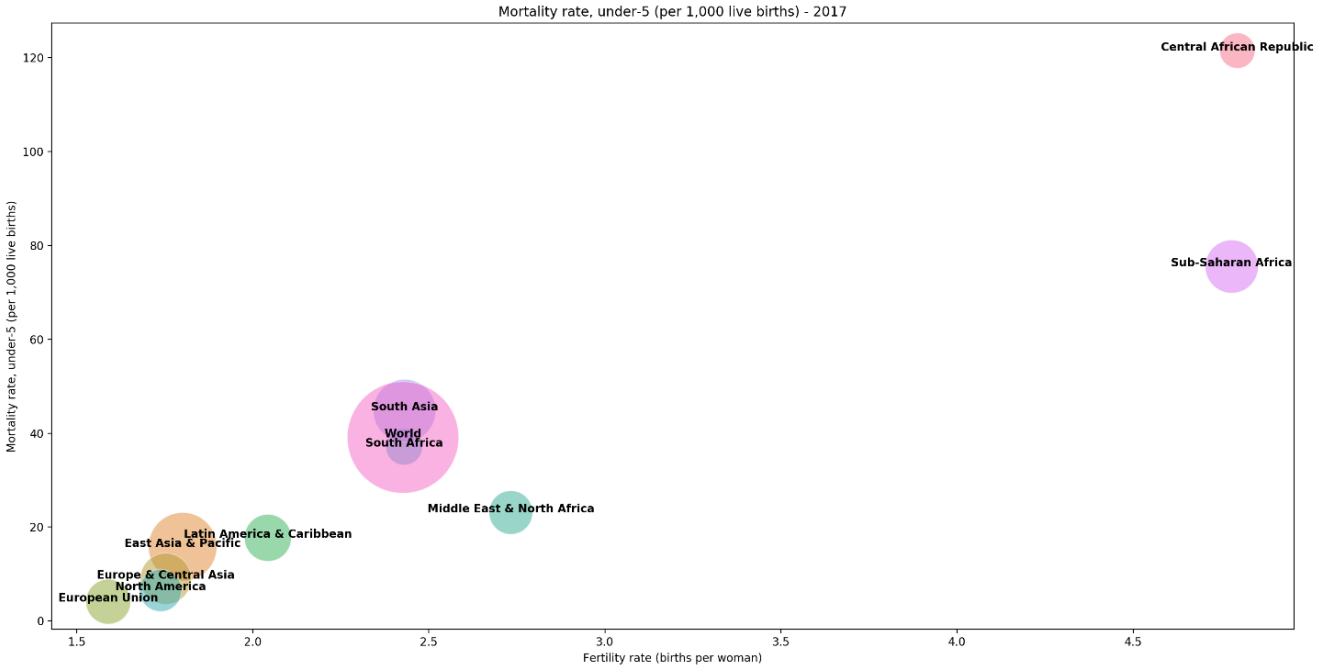


The plot looks nice, but there is a problem, which is that it may be difficult for viewer to immediately identify each bubble. So below is another attempt using seaborn:



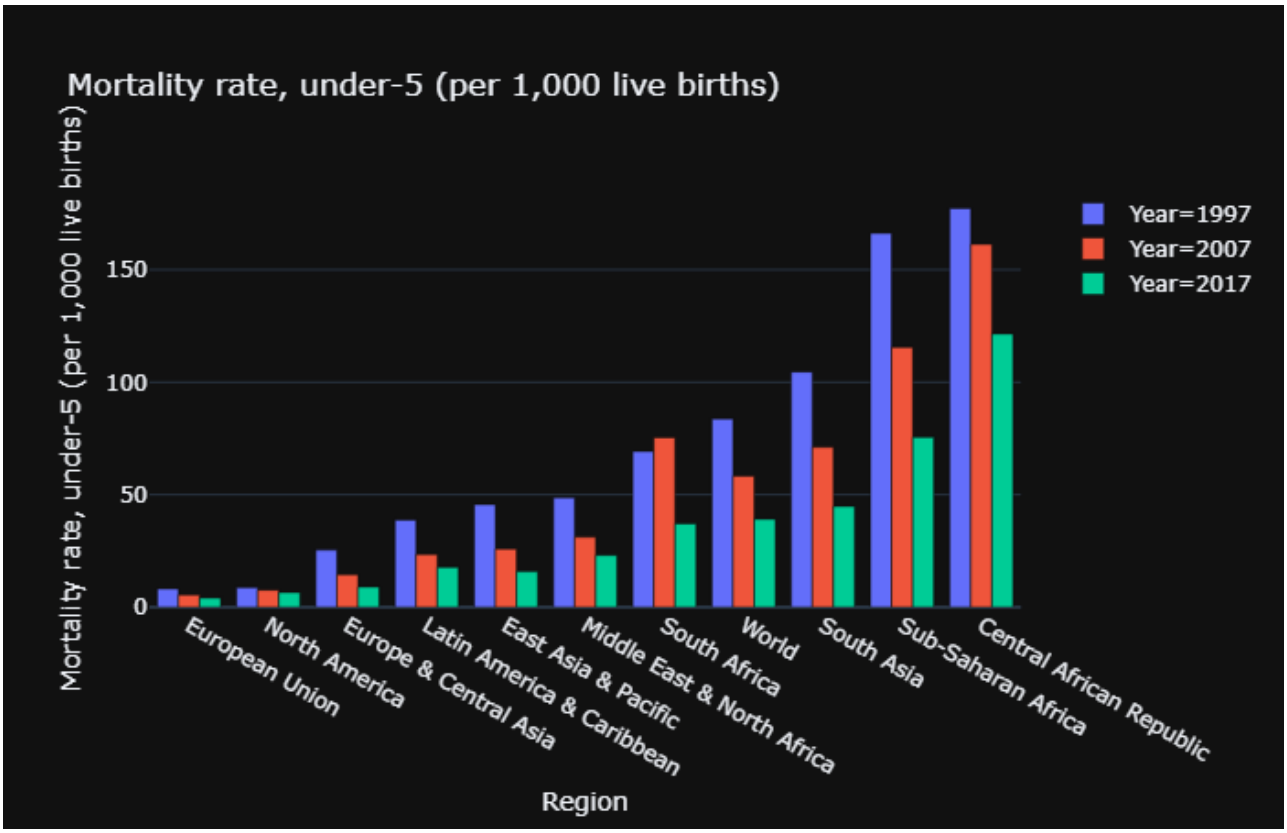
But now we need to compare across the years.





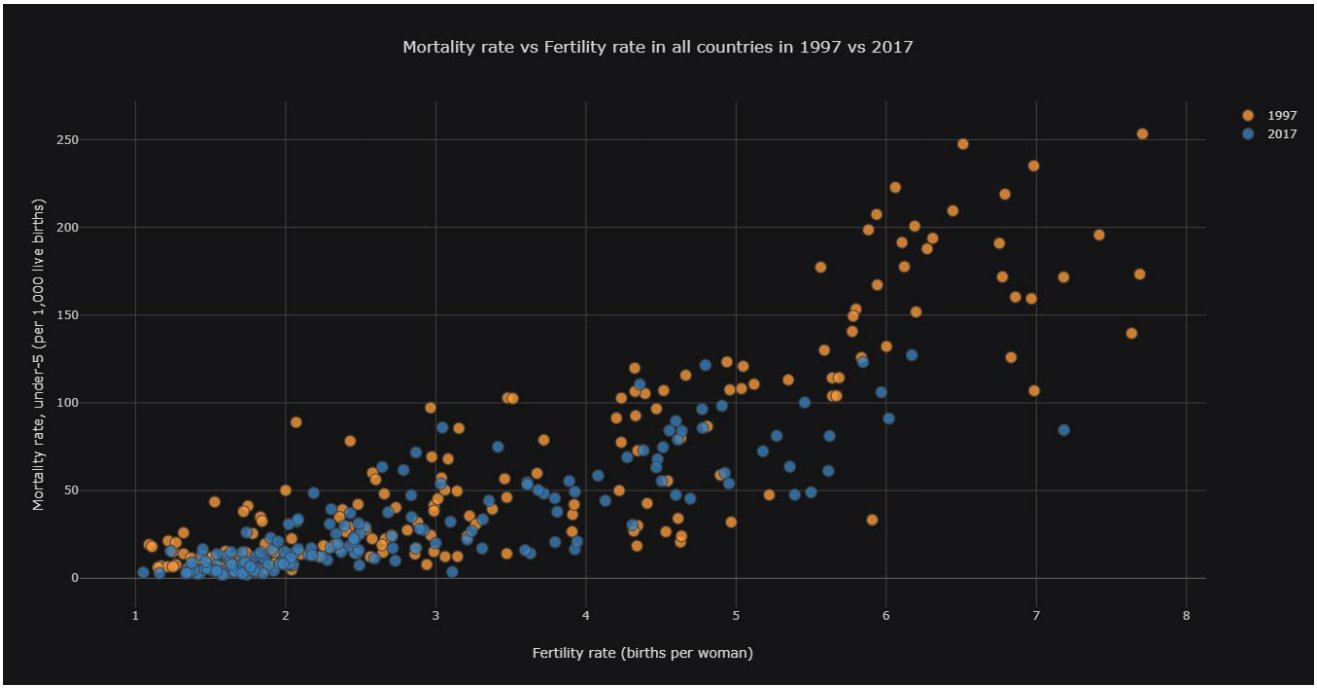
These charts show that both mortality rates and fertility rates have dropped across all regions in the world. This is due to increase in the living standards and education of the people. Poor people have more children because more children means more people working in the family. Another reason is that as many children die before the age of 5, they tend to have more children to ensure at least some of them survive.

Below is a bar chart depicting just only the mortality rate across the years 1997, 2007, and 2017:



This again shows that child mortality rate has gone down.

Another interesting visualization below shows the scatter plot for all the countries drawing the comparison between 1997 and 2017. This looked to me the best plot for this purpose. I have included an animation of the same over the years in the ipython notebook.



Population Division:

Developing vs Developed?

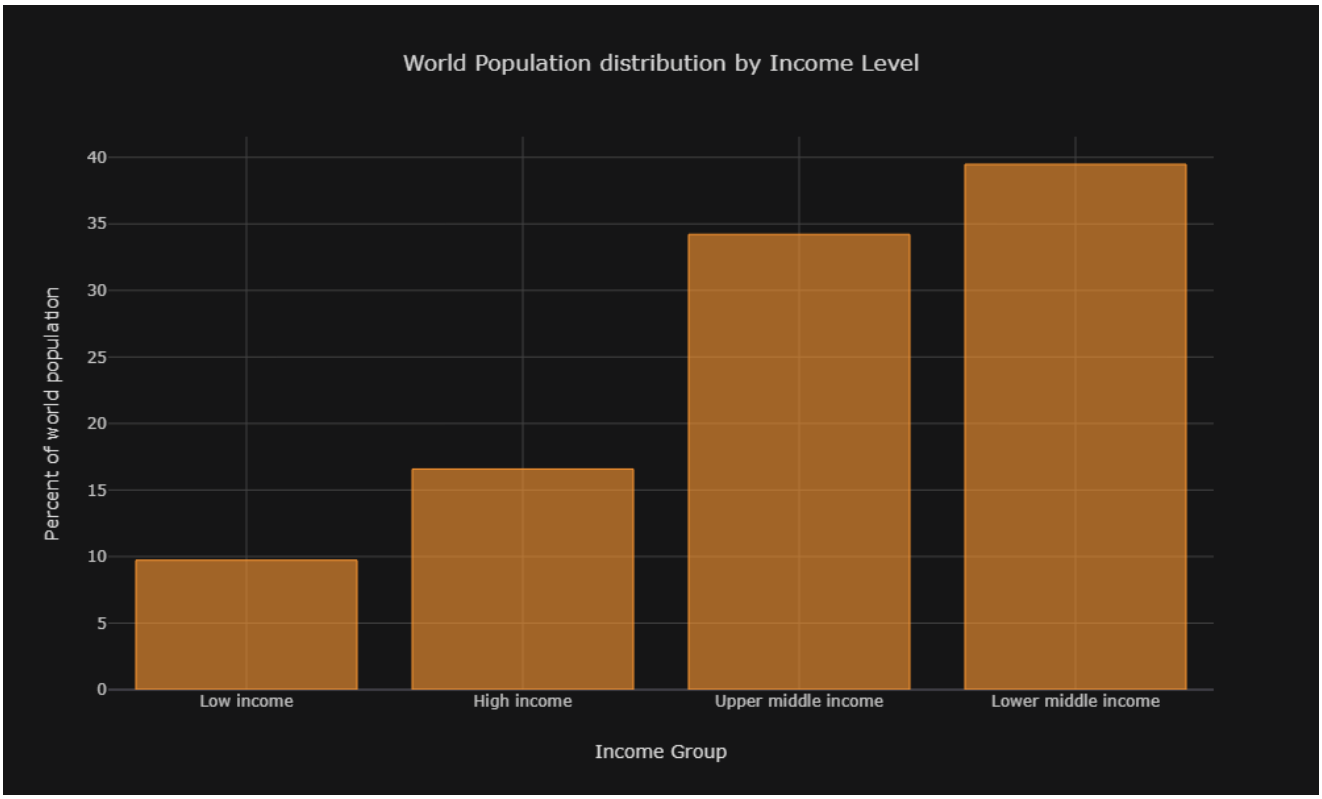
When we divide the world based on the income of people or economic condition, we always have only two distinctions in mind, rich-poor, developed-developing, west-rest etc.

But is this still true in today's world?

The answer is no, this would have been the picture 40-50 years ago, but today there are four different distinctions in the world. We can label them as four levels as per Gapmider, or we can use similar four level scheme by UN and World Bank, just named differently. The below chart shows the percent of world population living in each level at the current stage. (The total world population is 7.5 Billion as per 2017 data)

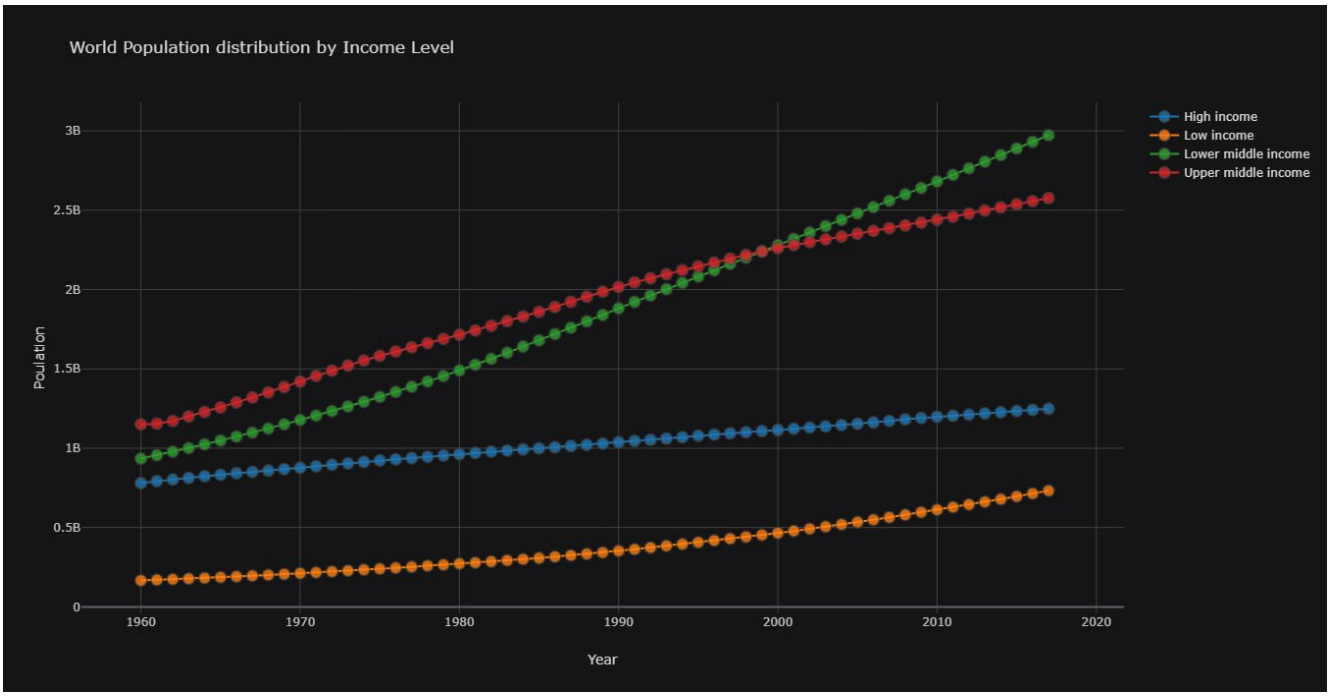
The majority of world now lives middle income levels, the lower income comprises only of 9% of world population.

These categories are based on specific criterion by world bank described in an article by World Bank - Classifying countries by income[1].



Below is a plot showing the actual world population over the time under these categories.

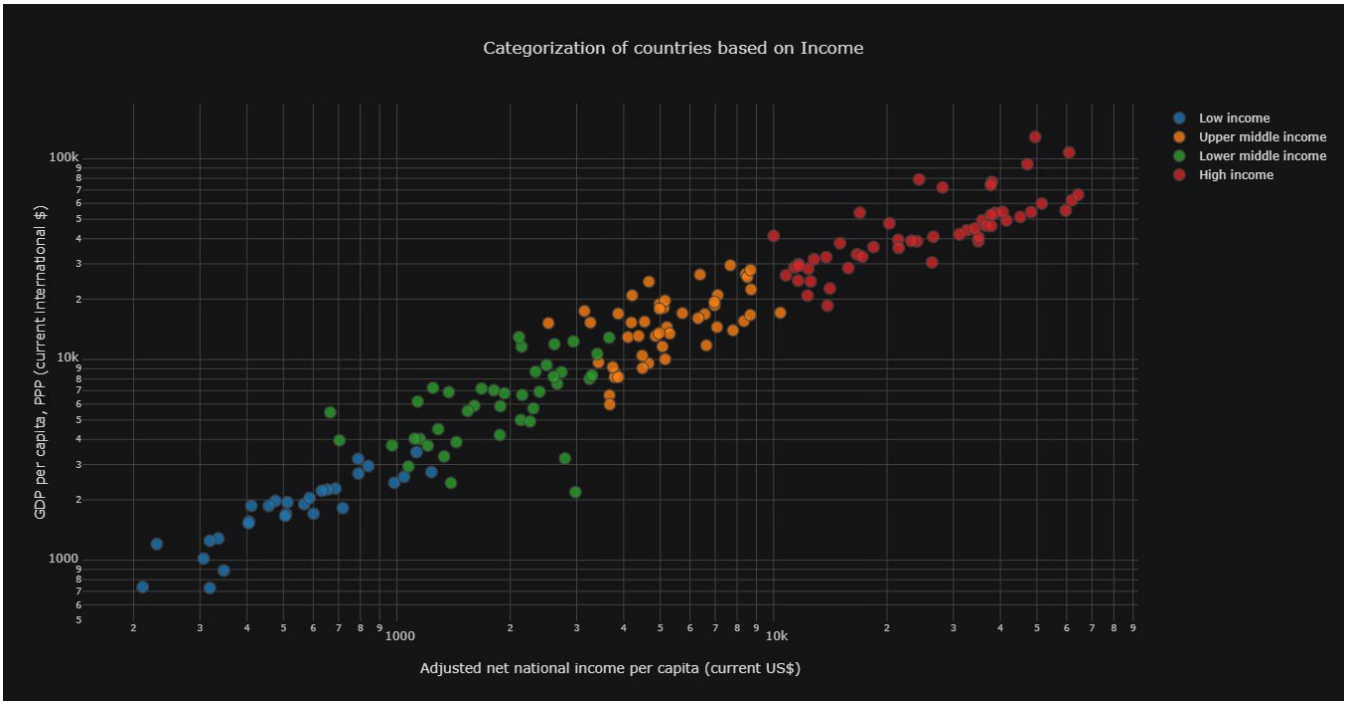
[1] http://datatopics.worldbank.org/world-development-indicators/stories/the-classification-of-countries-by-income.html



Q2. Where does the majority of the world population live?
A: Low-income countries
B: Middle-income countries
C: High-income countries

This is the 2nd question in the book, and now we know the answer.

To see whether this distinction works well, I have plotted a scatter plot for all the countries with GDP and GNI indices both in terms of PPP (purchasing power parity) and per capita. Both the axes are in log scale due to obvious reasons. As per the plot this distinction works great as we can clearly see four separate country groups.



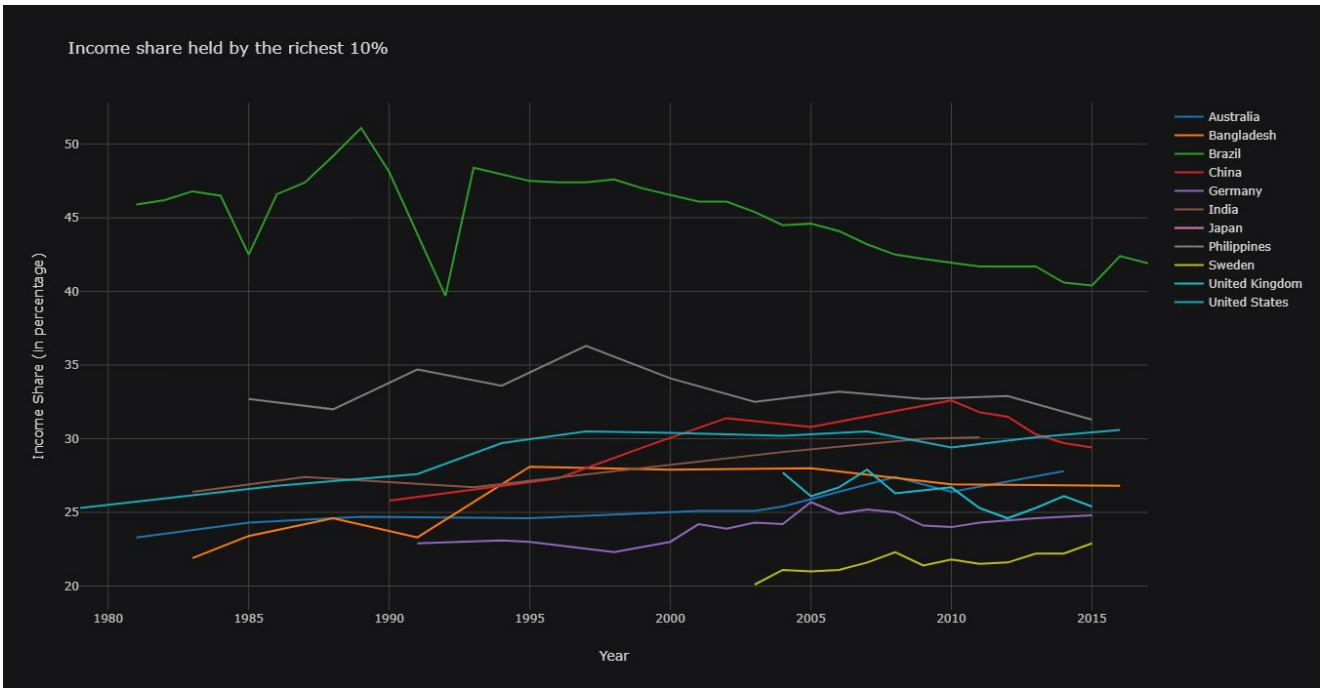
I like the way how the axes are plotted in plotly, as it is clear from the chart that both the axes are in log scale.

Income Disparity:

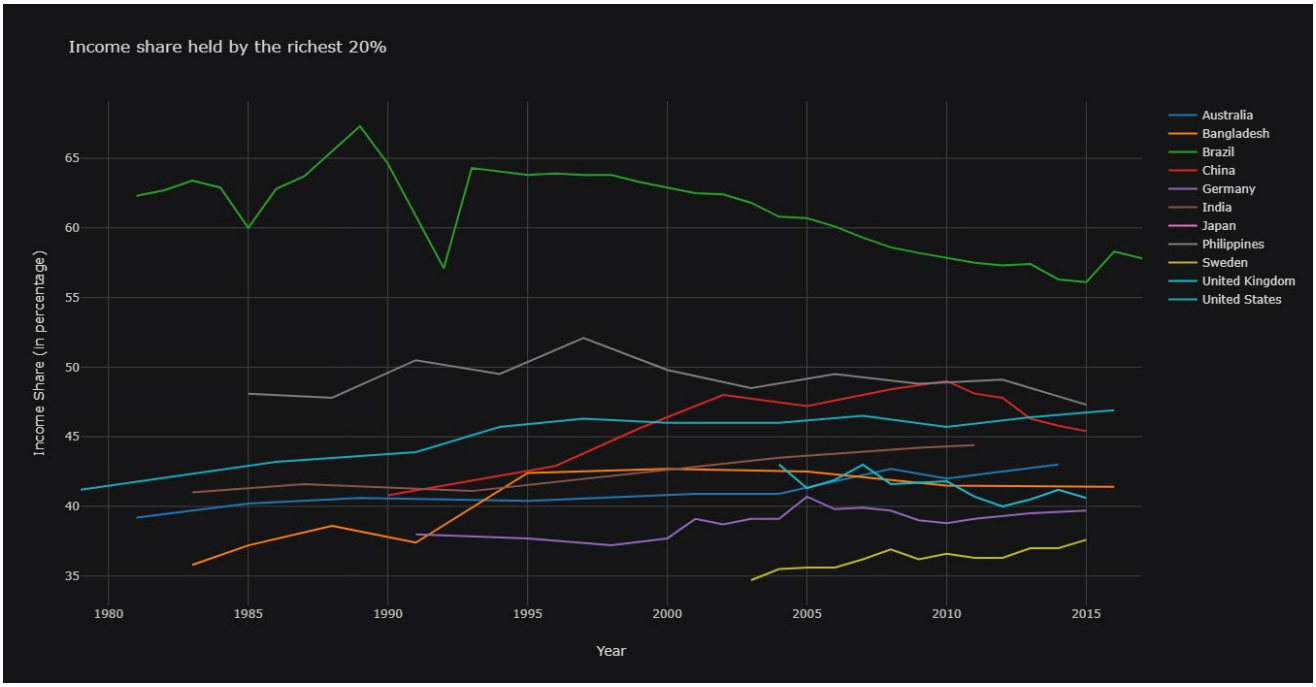
We regularly hear about how the rich are getting rich and poor are getting poorer. The gap is increasing day by day. So, let's explore these claims through data and establish some facts.

Firstly, I have explored the income share held by the richest 10%.

Looking at Brazil, we see that the percentage is huge, but is continually decreasing. Same is the case with China and UK. Although in countries like USA it is on increasing trend. This doesn't imply that income disparity is not increasing, but it is though better is some of the cases, and hence it is important to look at individual countries than the world at average.

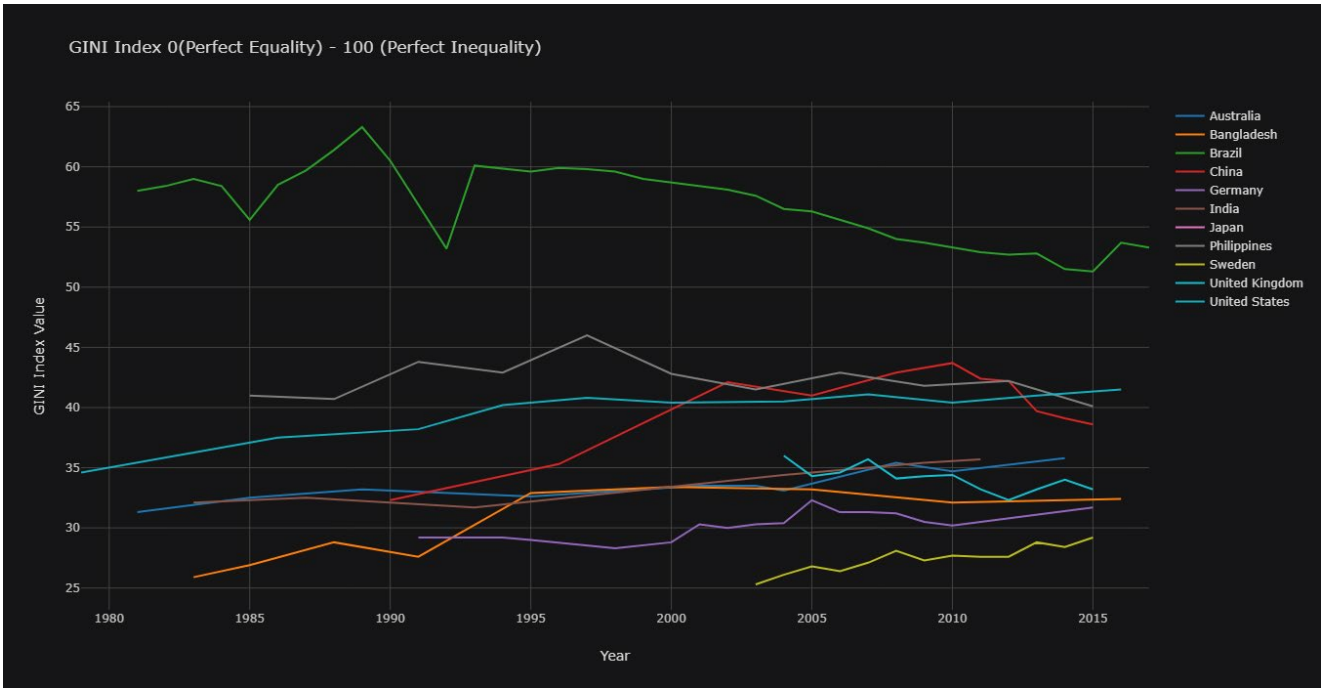


There is an interesting 80:20 rule which is applicable in many cases in the world. One of the examples for it is the income distribution, i.e. 20% of the people hold as much wealth as the other 80%, which is being explore in the below charts. We see the rule being true and over the time on average it is more or less same.



Another way to measure the income disparity in the world is the GINI Index. According to World Bank definition:

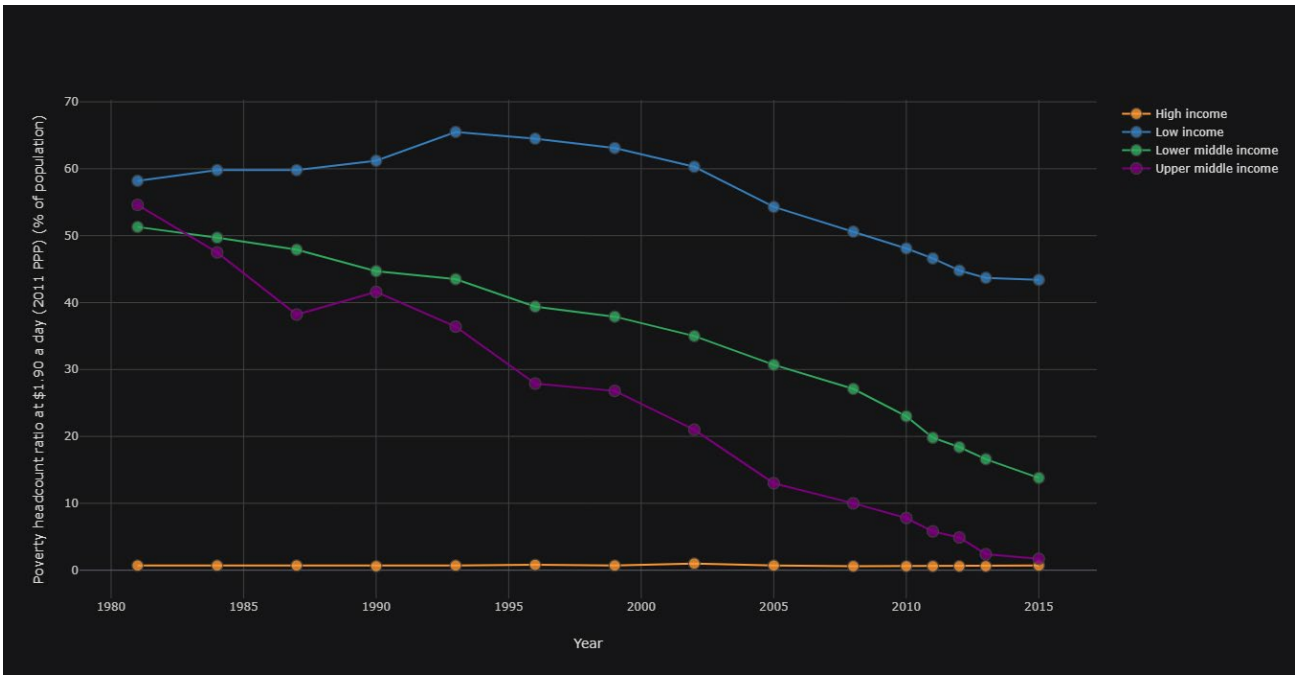
Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus, a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.

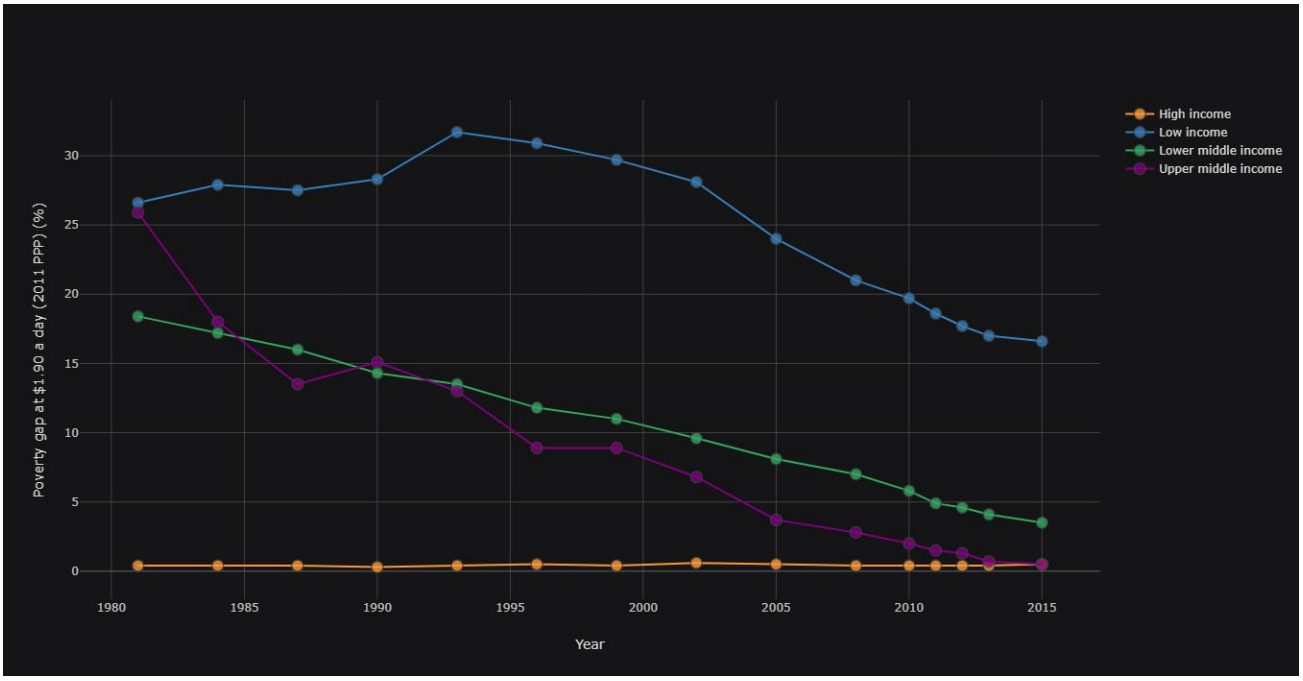


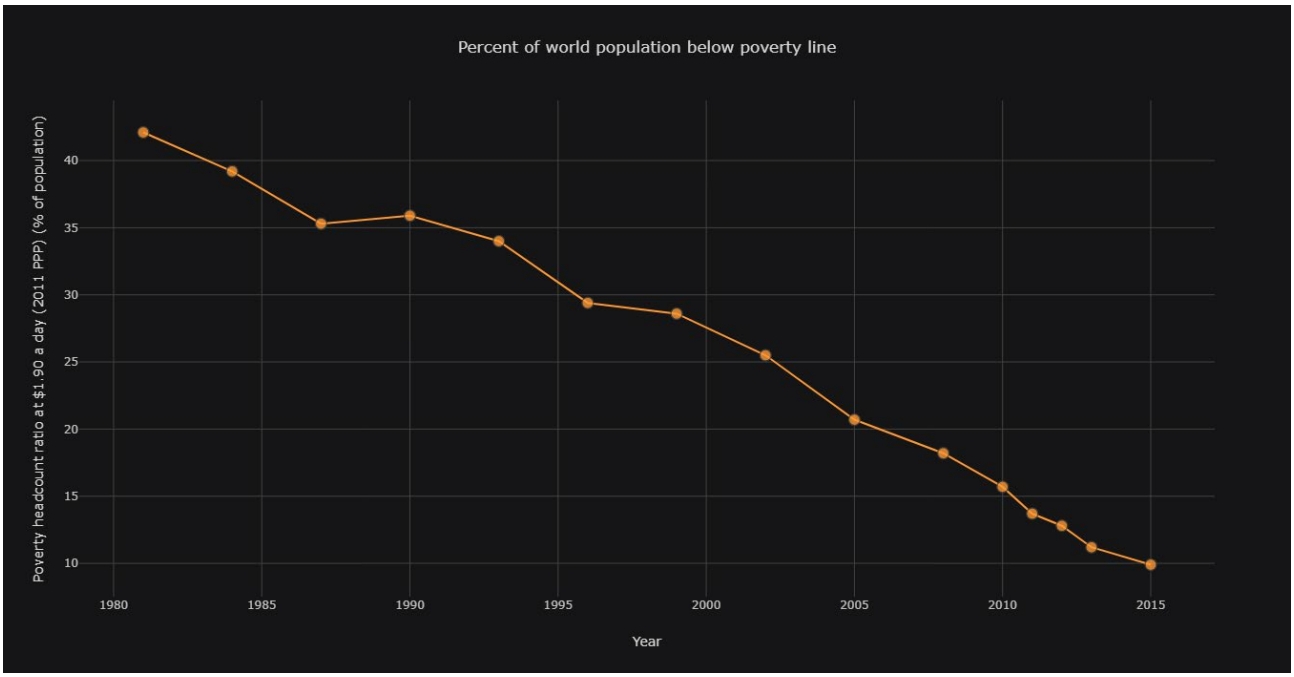
Poverty:

Q3. In the last 20 years, the proportion of the world population living in extreme poverty has …
A: almost doubled
B: remained more or less the same
C: almost halved

So, let's look at some of the poverty indices, and see where it has come in the past 20 years.

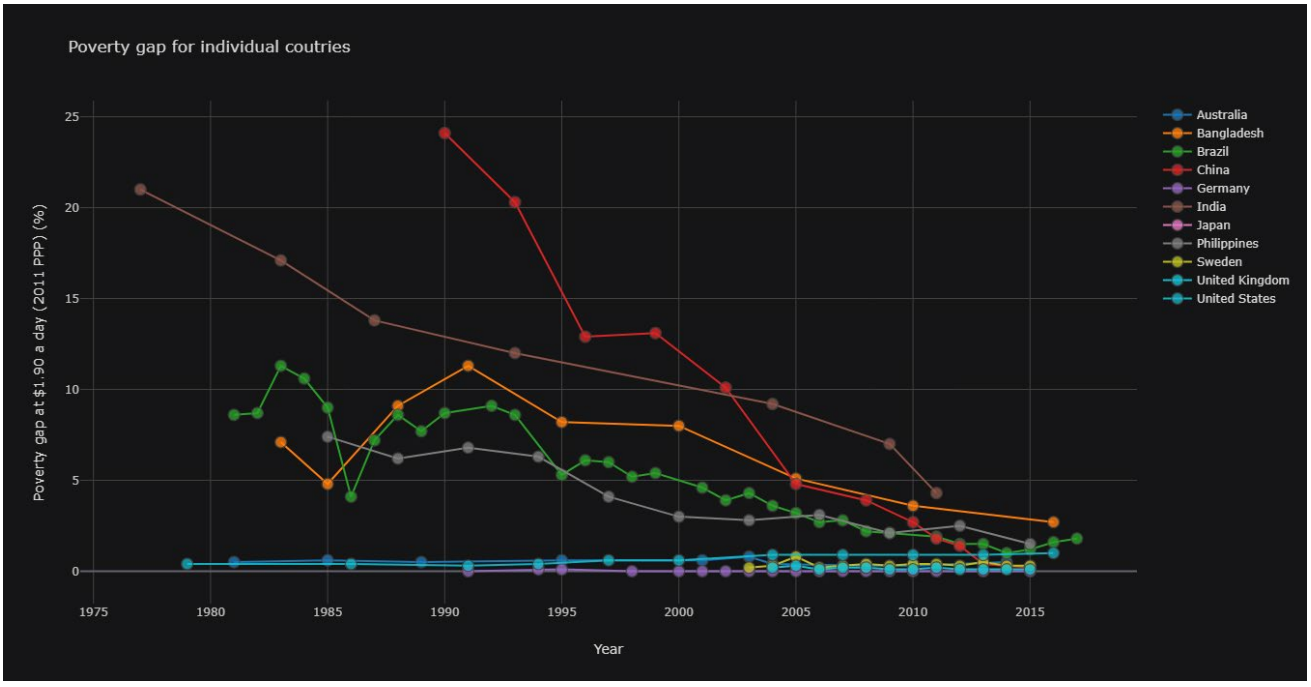

Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population)
70
60
50
40
30
20
10
0
1980
1985
1990
1995
2000
2005
2010
2015
Year
High income
Low income
Lower middle income
Upper middle income


Poverty gap at $1.90 a day (2011 PPP) (%)
30
25
20
15
10
5
0
1980
1985
1990
1995
2000
2005
2010
2015
Year
High income
Low income
Lower middle income
Upper middle income



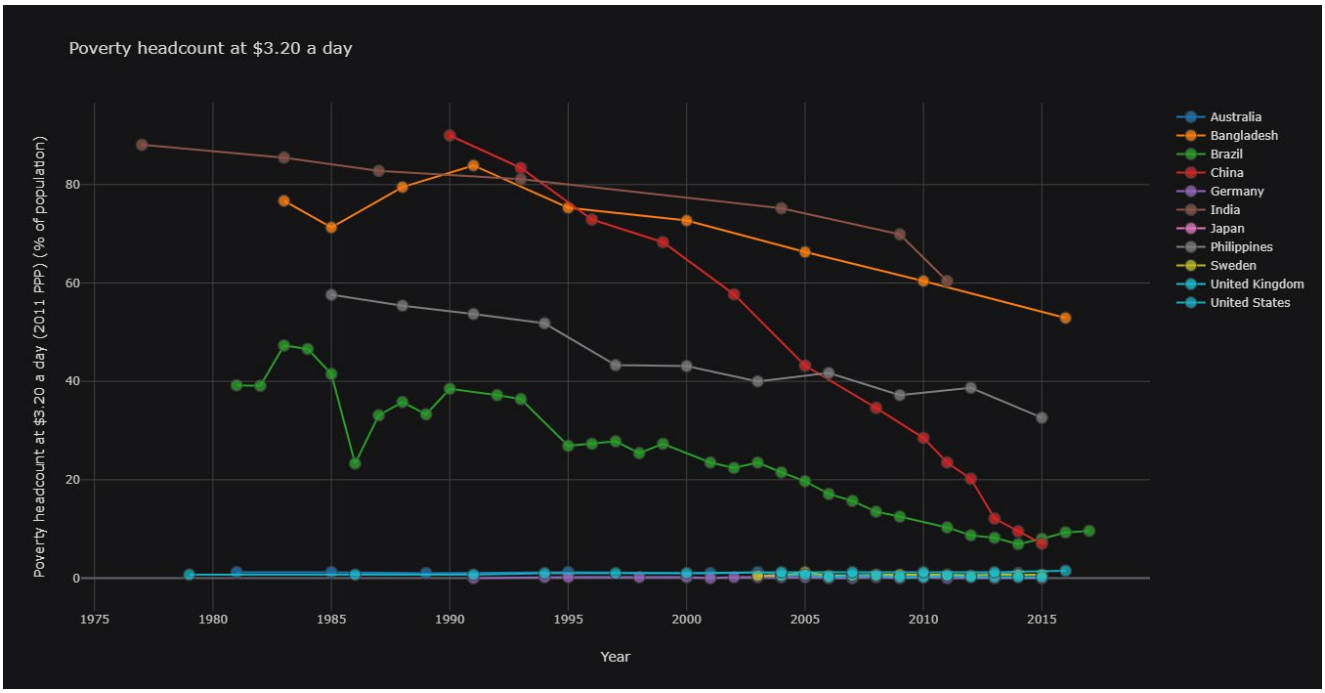
The answer is clear from the above charts that the poverty has halved over the years. Apart from observing that the poverty GAP is decreasing, this also shows that if we look at 1980, there were only two categories rich and poor, but the distinction among 4 poverty levels have become significant over the time.

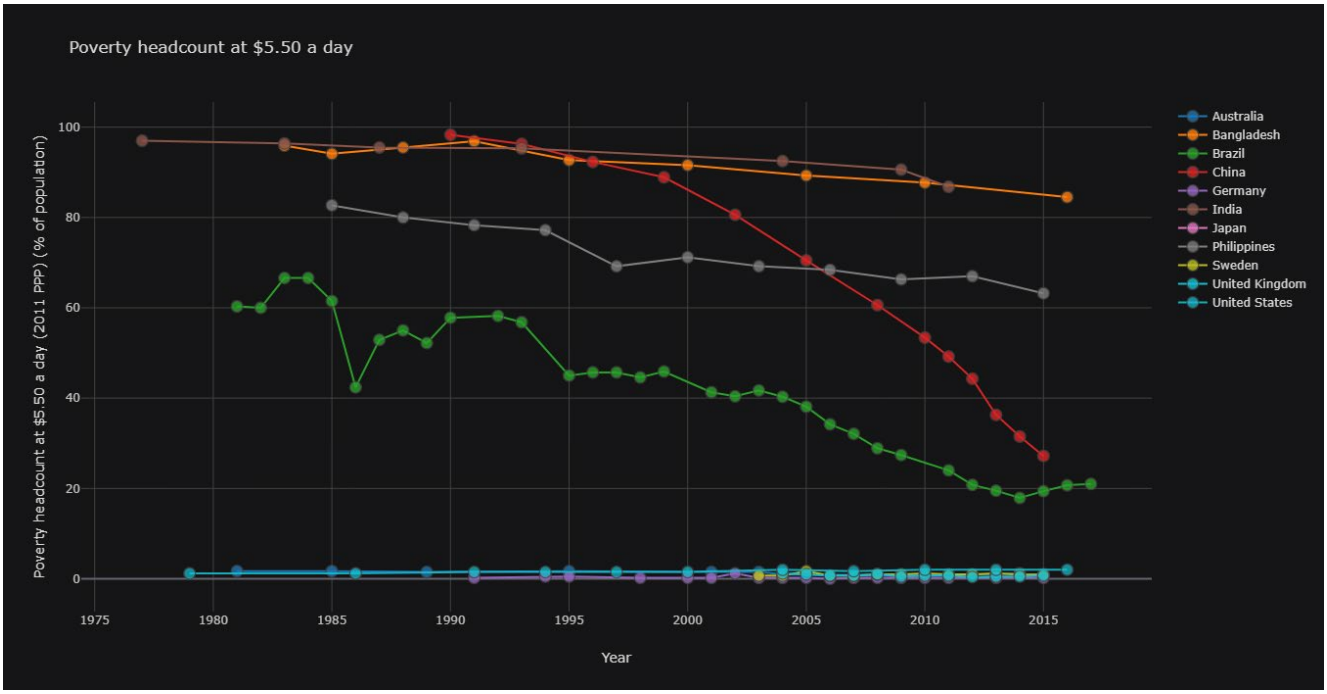
Even the poverty gap statistics for individual countries look good, specially for India and China.

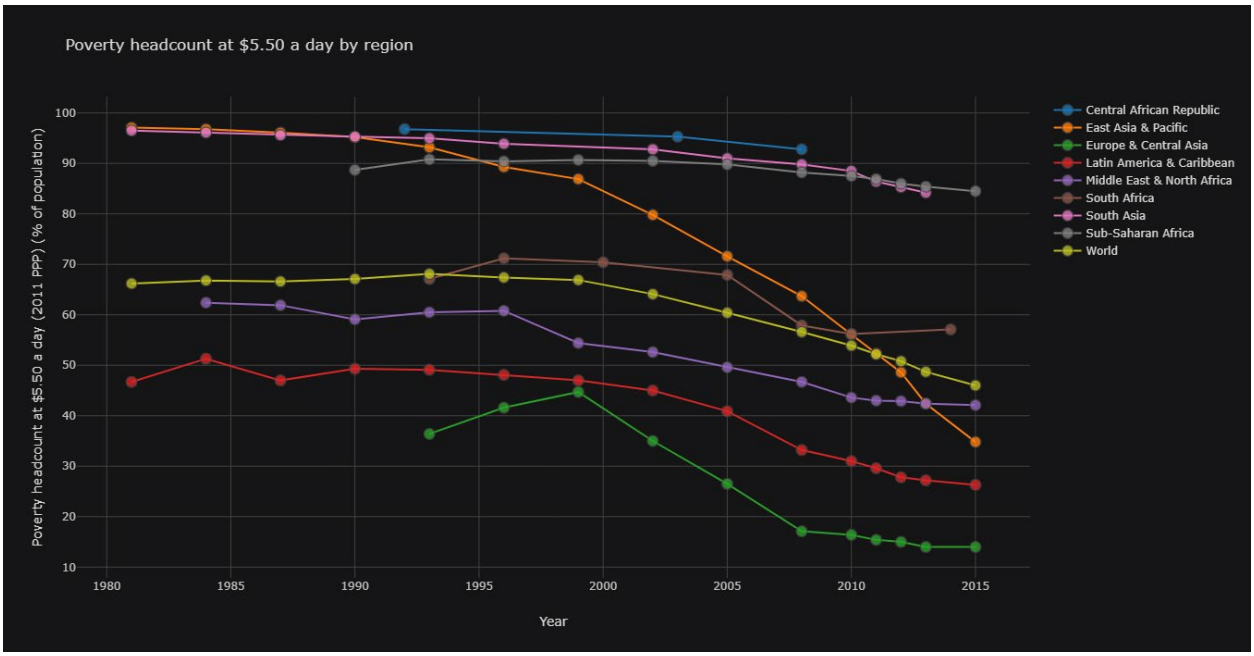


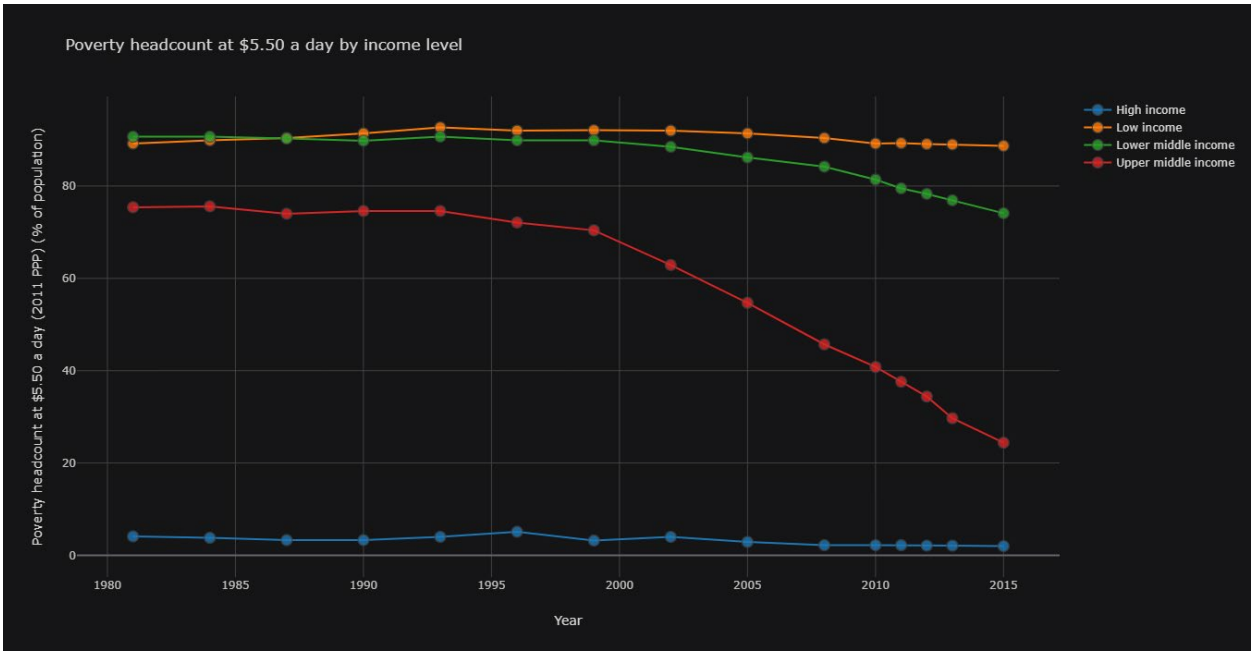
These are all based on $1.90 a day, which is treated as standard by World Bank and other organizations to measure poverty. Though, World Bank provides two more stats based on $3.20 and $5.50 a day. Below

is a look at that which shows there is room for much more improvement, especially in countries like India, Bangladesh, and Philippines to name a few.









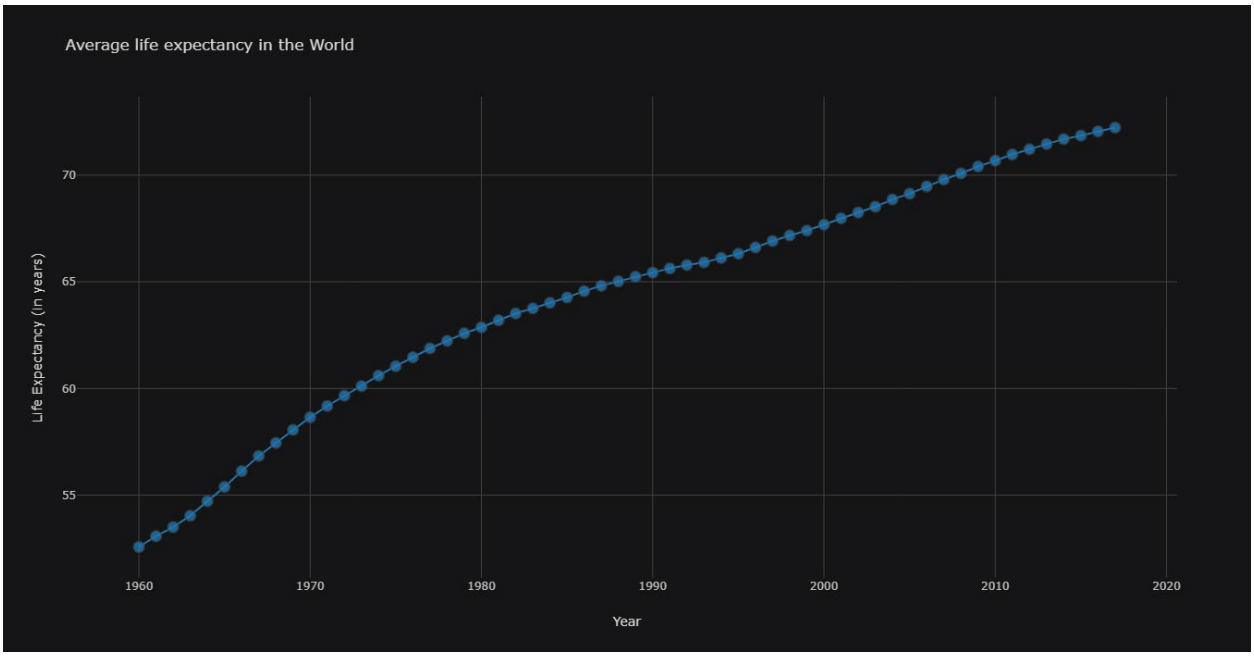
Life Expectancy

Q4. What is the life expectancy of the world today?

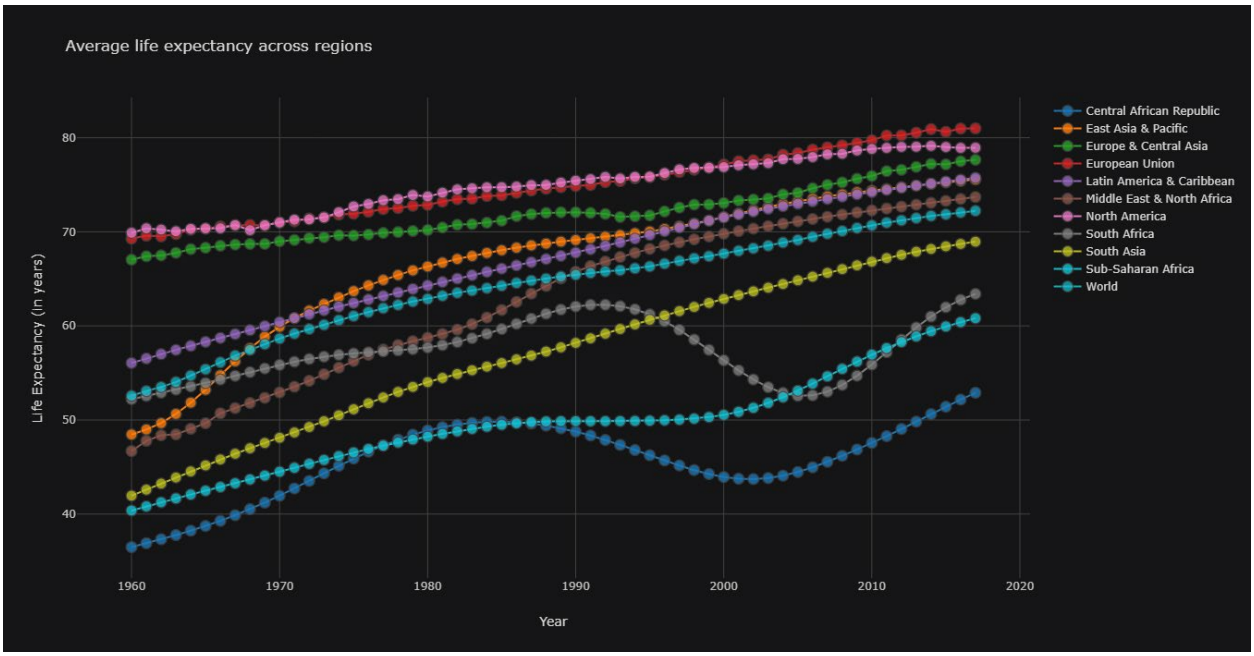
A: 50 years

B: 60 years

C: 70 years

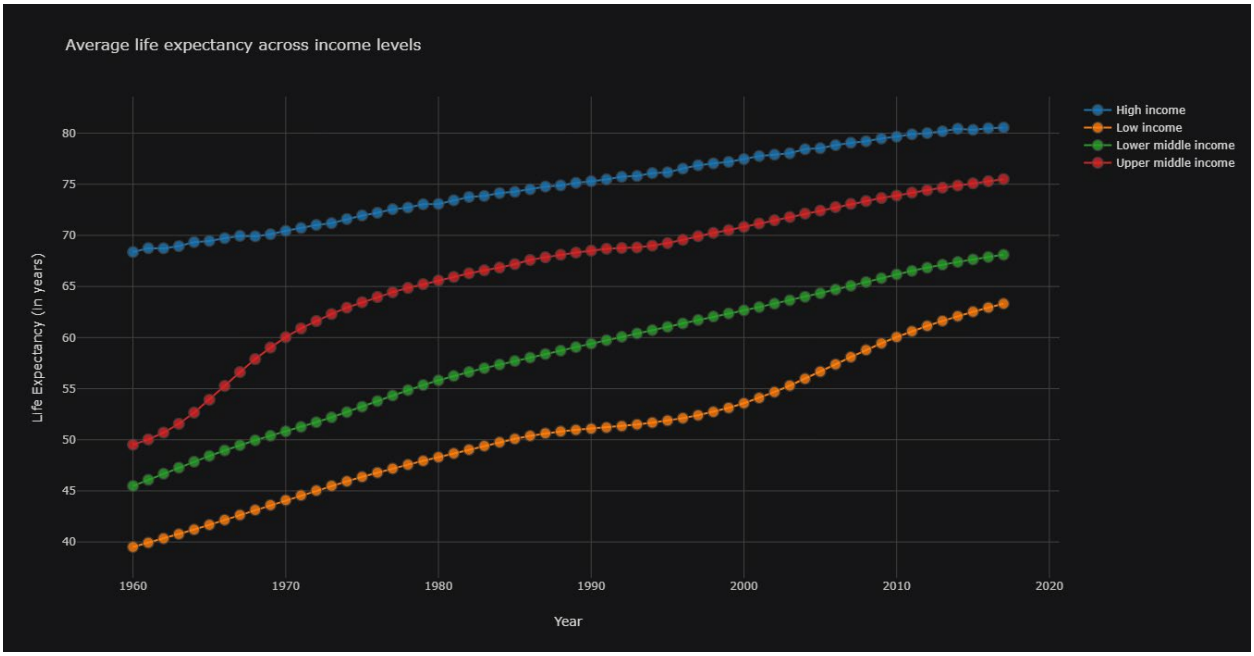


That means the answer is 70 years.



One interesting thing to note in this graph is that in Africa (central and south) there was a period decline as opposed to the whole world. Why did that happen? Well the answer is not Ebola outbreak, that happened later. After searching through web, I stumbled upon an article which stated the reason for this exact observation which is high deaths due to AIDS and other diseases, and secondly high child mortality

rates due to nutritional deficiencies. Both the problems are being tackled now through more awareness about AIDS and availability of medicines, better medical care, and availability of nutritional food.[2]

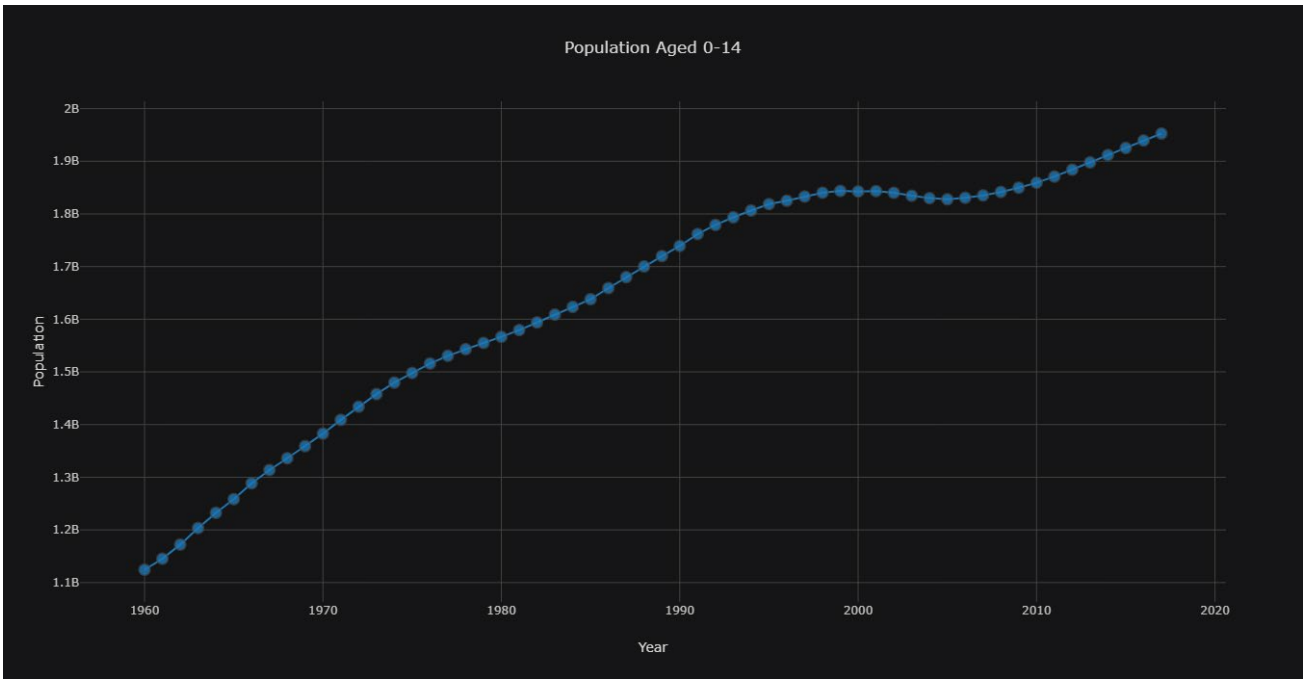


We see that life expectancy has been continuously increasing, that's good news. This means every place on earth has been continually developing and increasing its living standards. More people are moving from Level 1 to Level 2 & Level 3, and are able escape extreme poverty. They are getting better nutrition and health care facilities.

World Population:

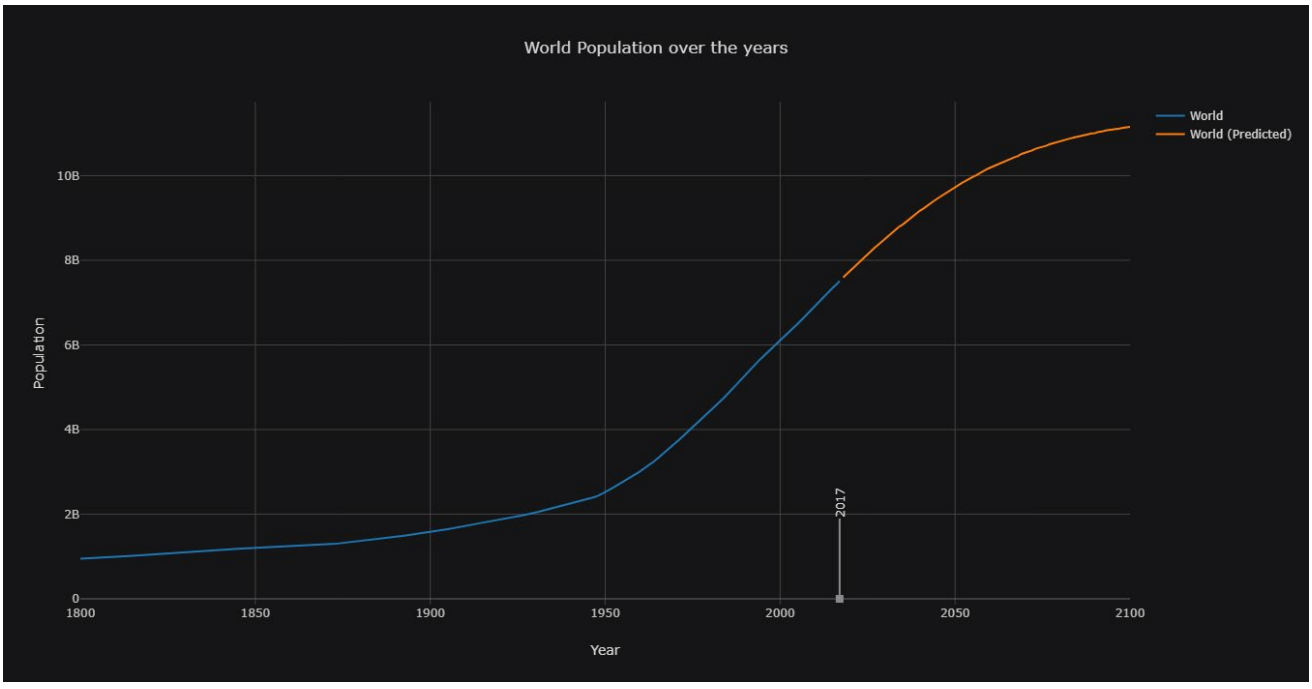
Q5. There are 2 billion children in the world today, aged 0 to 15 years old. How many children will there be in the year 2100, according to the United Nations?
A: 4 billion
B: 3 billion
C: 2 billion

[2] https://bhekisisa.org/article/2014-12-19-sa-life-expectancy-drops/



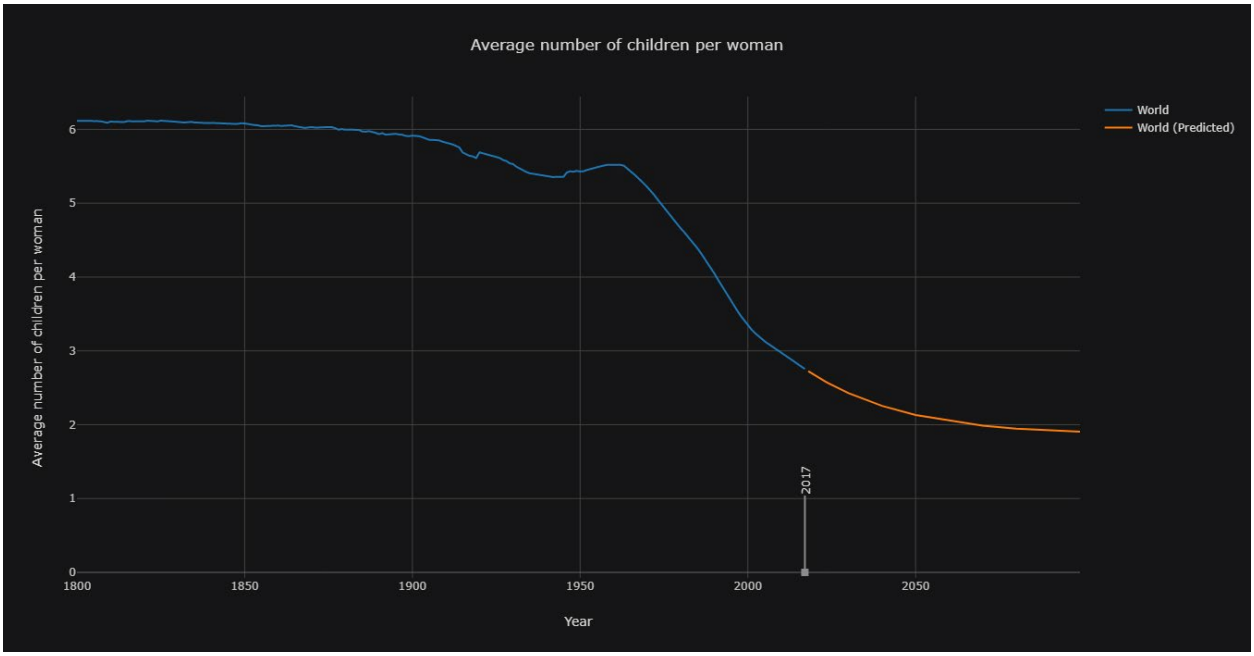
Well if we follow the above chart, we can see it moving towards 2 Billion. And instinctively we extrapolate it in straight line to visualize in our mind how it would be in next 80 years, we see it rising only maybe to around 3 to 4 Billion. But is it true?

Let's first look at historical World Population and future predictions by the UN.

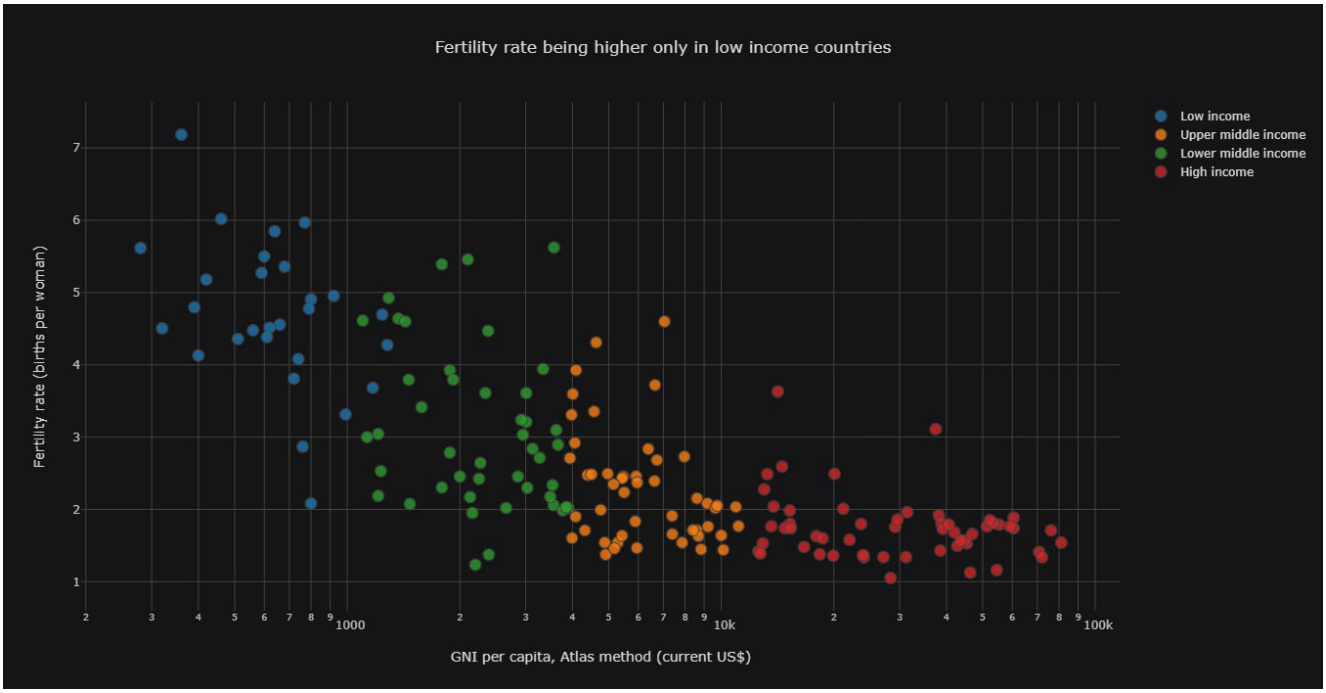


Well the predictions says that the future growth will slowdown and the world population will reach the equilibrium at around 11 billion. How's that possible?

Well the answer to why 0-14 age group population won't rise is in the below chart. The average number of children per woman have been decreasing since a long time, which we saw in the fertility rates charts as well. And it will continue to go down, as a result overall there won't be an increase in child population even if we account more number of female adults in the future.



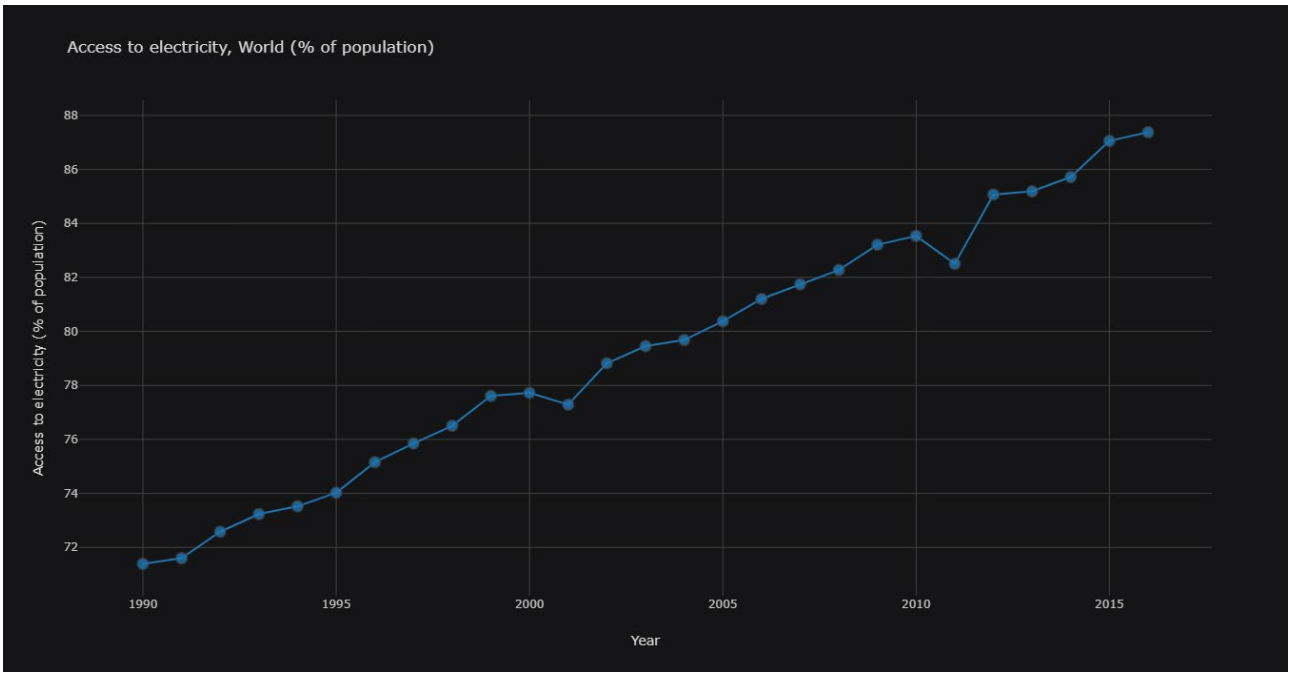
To dive deeper let's look at another chart.

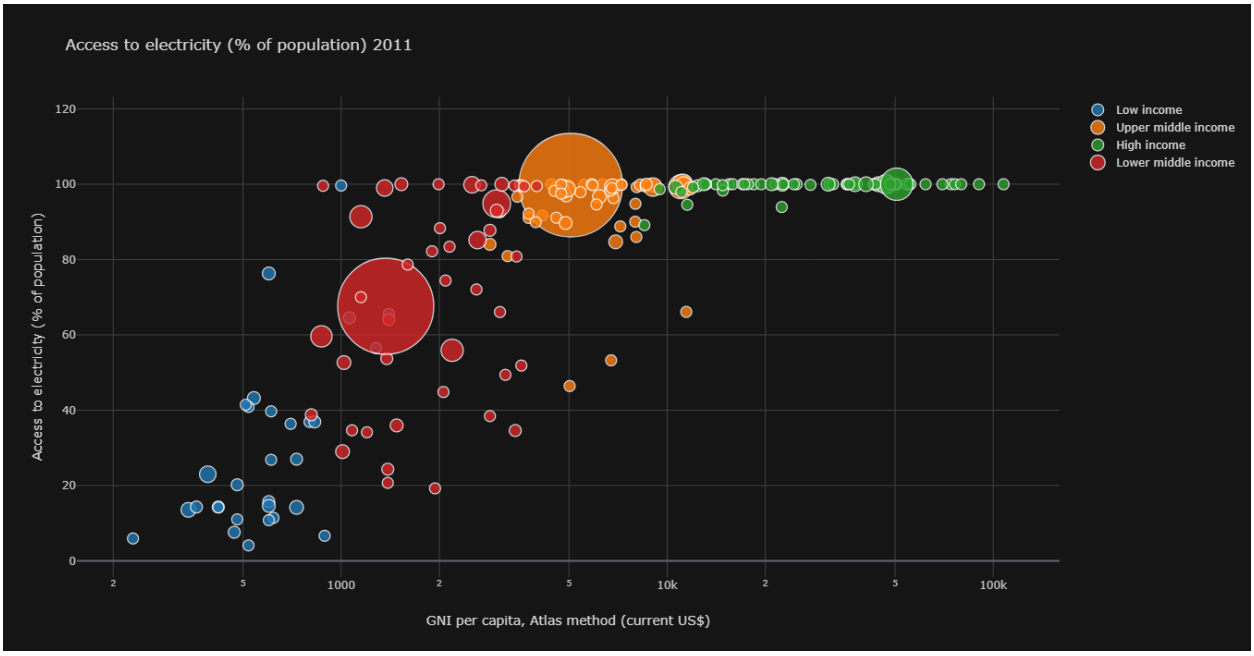


The fertility rate is higher mostly in the low-income countries, and the world is moving towards middle income, hence overall fertility rate is going down.

Access to Electricity:

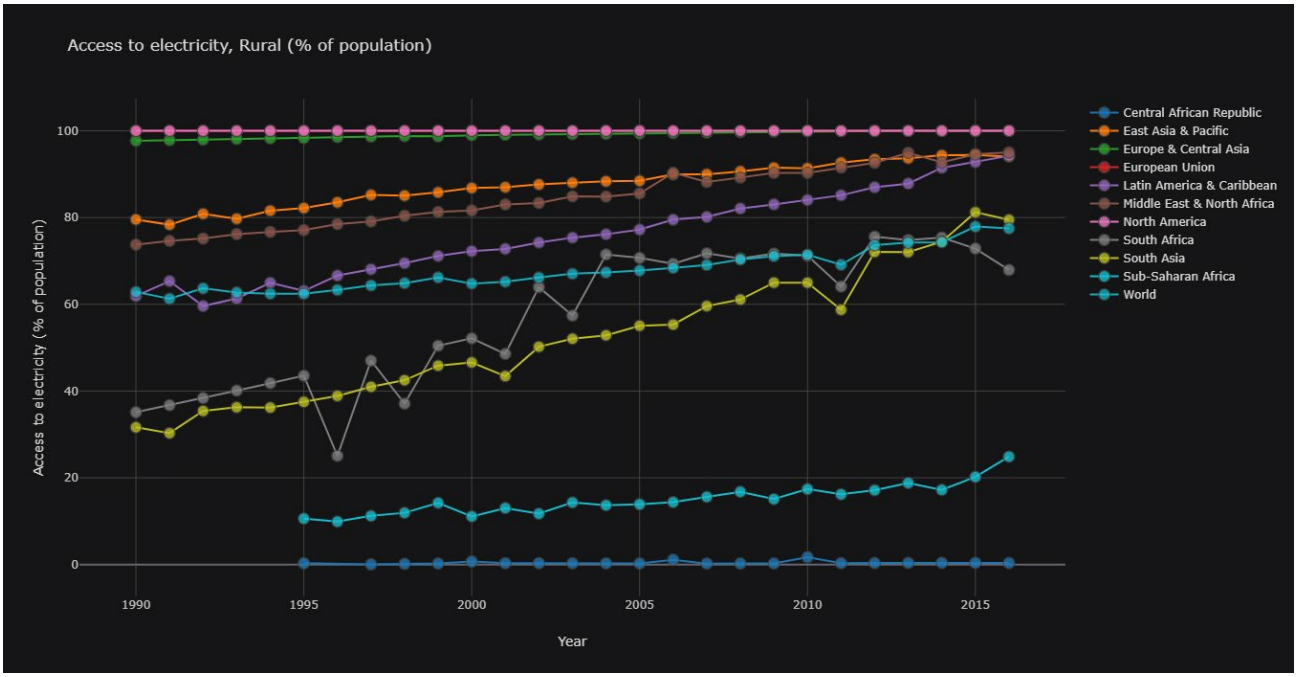
How many people in the world have some access to electricity?
A: 20 percent
B: 50 percent
C: 80 percent

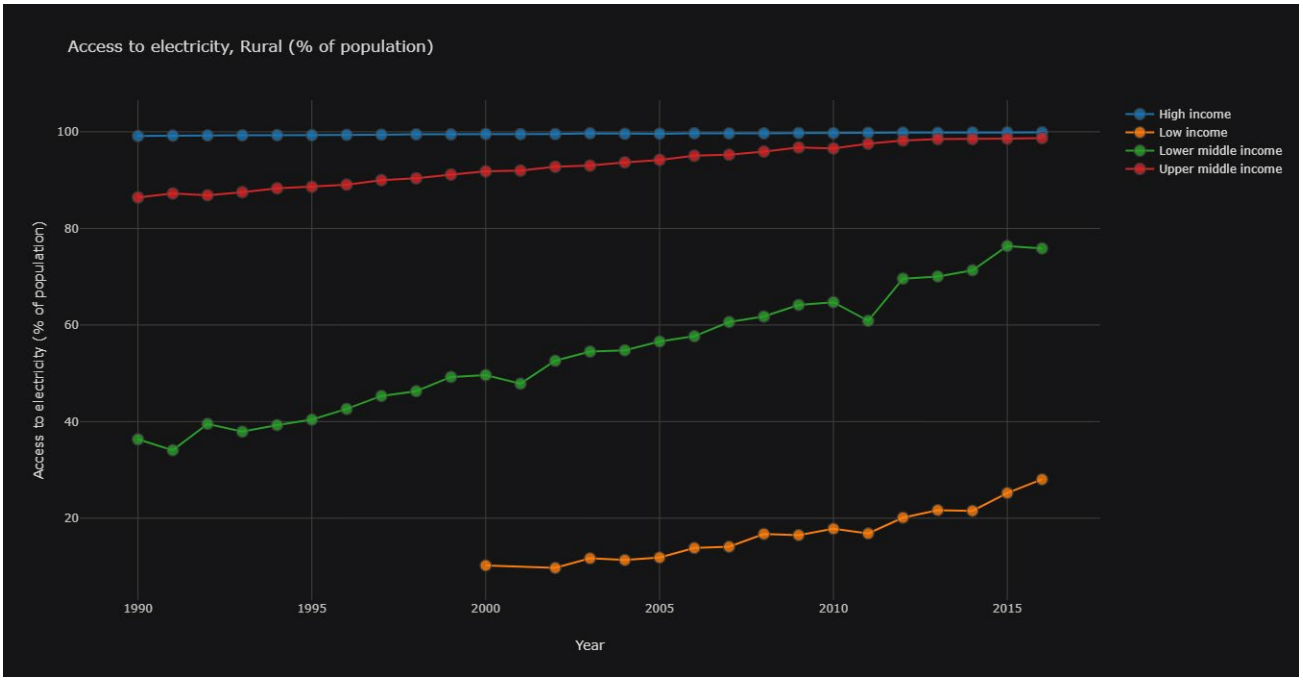




100 % in rich countries due to urbanization. But other countries are catching up.

Let’s take a look at Rural population.



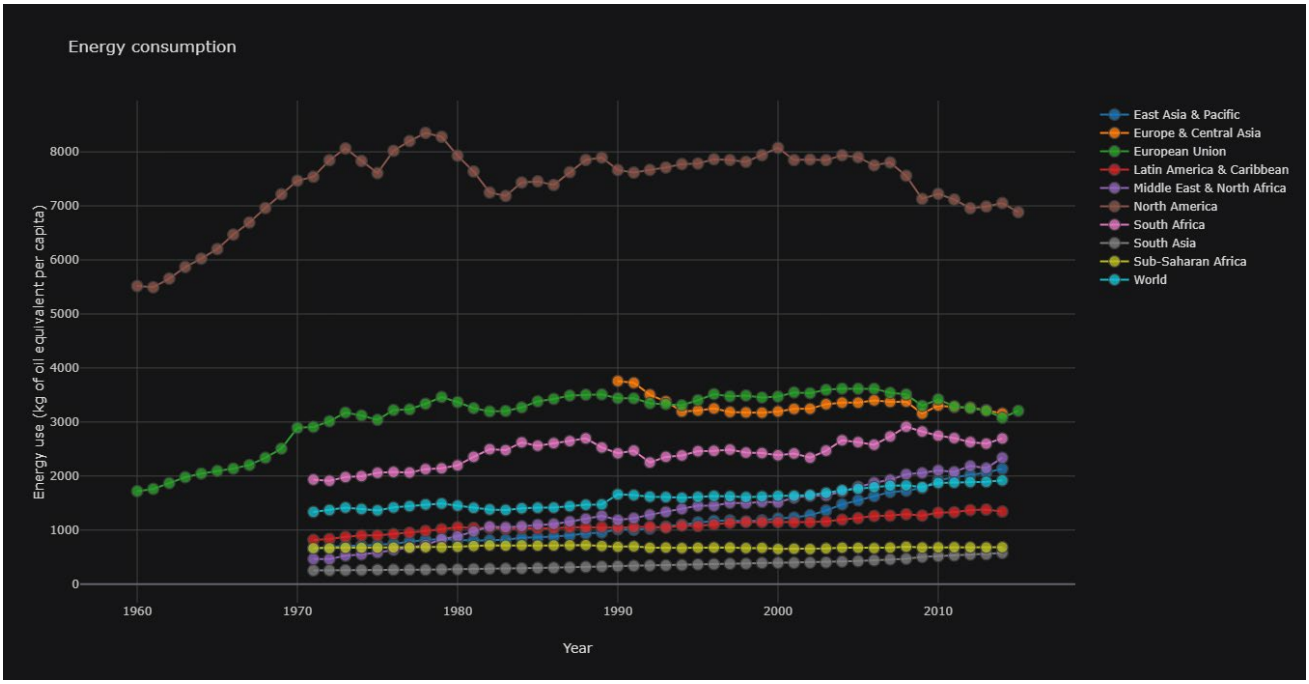


It is clear that a lot has to be done at rural level in low and lower-middle income countries even though 80% of the world have electricity today.

Miscellaneous:

There has been a lot of debate over the climate change and global warming, who and what is causing it over the recent years. And the world now know that it is moving towards hotter climate and global warming is real. To reduce global warming we have to reduce the main cause, which is CO2 emissions, and most CO2 emissions occur from the use of energy (mostly electric and oil). As most of the world population live in east, and as east is progressing, for sure the total energy consumption is going to rise, and hence people (west) may argue that rest cannot live like us because if they do it will create more global warming. But is this the complete truth?

Let's have a look at the below charts.



There is line in brown which is way above the rest of the world. Yes, that's North America with the highest per capita energy consumption, and that too double than the Europe. Hence, one has to look at per capita income and then realize that the rest isn't contributing as much as the west is doing. And everyone has started or must start moving towards renewable and nonpolluting sources of energy. Below is another chart depicting total greenhouse emissions. Somehow that figure was too high erratic for 'Central Africa' and hence have been excluded from the below chart (probably due to some erroneous data, have to be checked further).


Total Greenhouse Emissions
Total Greenhouse Emissions per capita
0.025
0.02
0.015
0.01
0.005
0
1970
1975
1980
1985
1990
1995
2000
2005
2010
Year
Central African Republic
East Asia & Pacific
Europe & Central Asia
European Union
Latin America & Caribbean
Middle East & North Africa
North America
South Africa
South Asia
Sub-Saharan Africa
World

Discussion, Conclusion and Future Work:

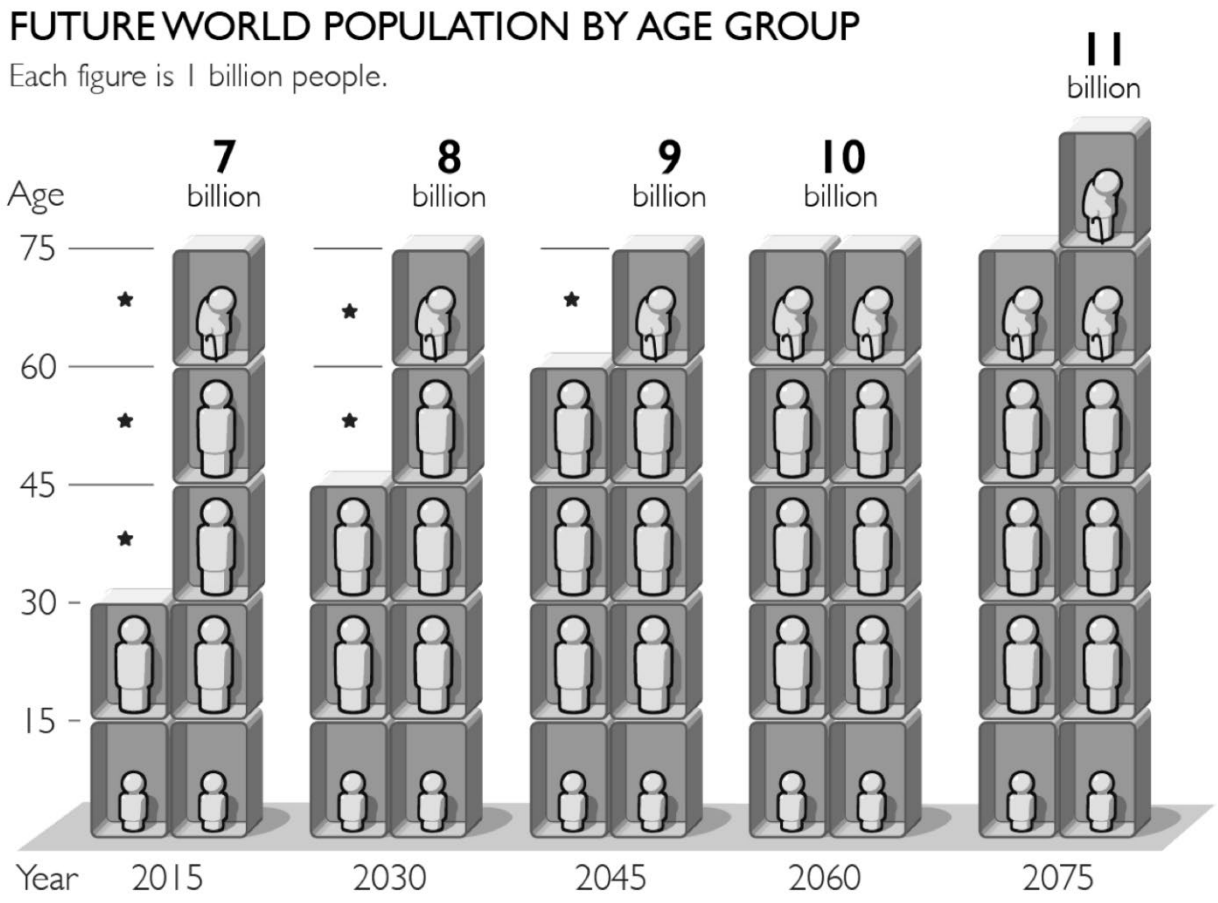
A lot of topics have been explored and discussed on in the previous section which shed some light about how our perspective of world is and, how the world is at this point. We are living in past perceptions about our world which can only be realized if we carefully look at the data. I have been able to explore major portion of the book and work by the author. I have carefully explored the World Development Indices which broadly represents how our world has come to be. There are huge number of indices recorded and provided by World Bank and it is not possible or in the scope of this project to go through them all. One could have approached this data and project in some other way as well, for example, looking through the data without any preconceptions or ideas on exploring it, and just calculating correlations, trends, raw statistics etc.

But my approach to this data and project was different, I had a clear objective and methodology in mind about exploring and making use of this dataset to come up with convincing stories through the use of visualization. And I believe I have done a good job in doing so. But as always there remains much more to be explored further. I have listed some of the limitations and failures I went through in the start of the project, and what other techniques I used to solve those problems in the results section. Hence, I have been able to create effective visualizations while progressing through the project.

For poverty index, more deeper understanding has to be established about the methodology behind the numbers. Per day income categorization, though in PPP, is still vague, there are other factors at play like where the person is living, and what are the living expense in that area locally. Five to four divisions can be made based on the rural – urban setting, and further each country has to perform regular surveys about the living expense in each of the areas. Another important thing is per family instead of per capita, which can provide a different picture than what we have.

For future works, one can start looking at correlations between indices and try to infer more about what is being affected by what and why, and how does it translates to real world. Secondly, in terms of visualization, more use of animated plots depicting the world statistics through time and using choropleth maps along with that could be done.

One of the visualizations I liked the most from book is below which shows how the population of the world be distributed as per age group in future. This plot beautifully handles so many things in one place and that too in simple 2D. Currently it is difficult to produce such kind of plot in python, but I hope to create many more simplistic yet powerful plots in the future.



Finally, to answer the main question, the world has certainly improved a lot in the recent past, and is not as gloomy as we see it. Most of the factors point at more improvements and a balanced world in the future.

References

1. Bank, W. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

2. Rosling, H. *Factfulness.*

3. Gapmider: https://www.gapminder.org/data/

4. WDI data: https://datacatalog.worldbank.org/dataset/world-development-indicators

5. OurWorldinData : https://ourworldindata.org/

6. Use of WDI by World Bank: http://datatopics.worldbank.org/world-development-indicators/

7. List of countries by Income Group :
https://en.wikipedia.org/wiki/List_of_countries_by_GNI_(nominal)_per_capita

8. TED. https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.