Analysis World Bank Socio Economic Data Analysis

A. Introduction

This is a personal project using the World Bank dataset (from Kaggle) to explore key socio-economic indicators and their relationship to life expectancy and development outcomes.

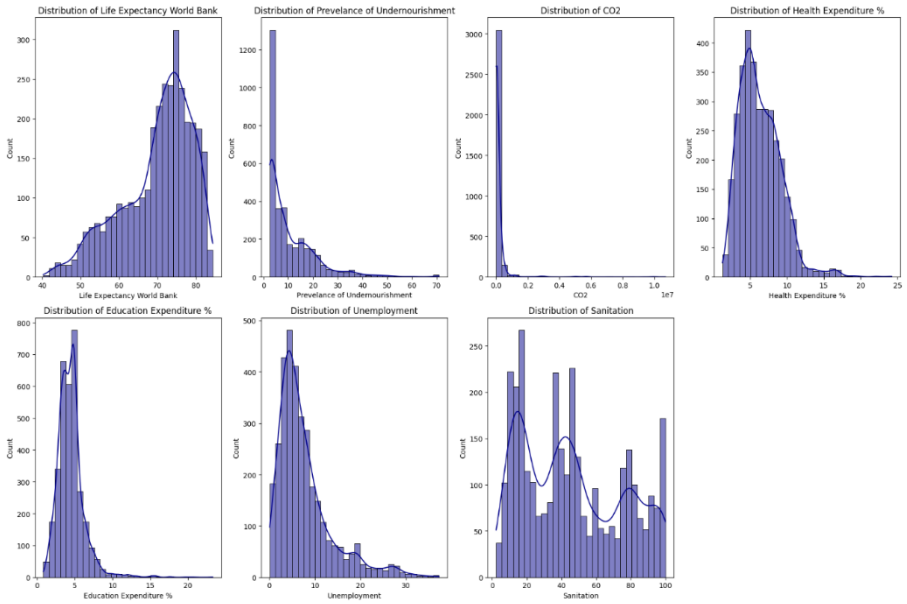
B. Data:

- **Source:** https://www.kaggle.com/datasets/mjshri23/life-expectancy-and-socio-economic-world-bank
- **Overview**
  1. Health Expenditure (% of GDP): Current health expenditure as a percentage of GDP. Includes healthcare goods and services consumed each year but excludes capital expenditures such as buildings, equipment, and vaccine stocks.
  2. Education Expenditure (% of GDP): Government expenditure on education (current, capital, and transfers) as a percentage of GDP, including transfers from international sources.
  3. Unemployment (% of labor force): The share of the labor force without work but available for and seeking employment.
  4. Prevalence of Undernourishment (% of population): Share of the population whose habitual food consumption is insufficient for a healthy, active life.
  5. Carbon Dioxide Emissions (kilotons): Emissions from burning fossil fuels and manufacturing cement, including gas flaring.
  6. Sanitation Access: Percentage of the population with access to improved sanitation facilities.

C. Method

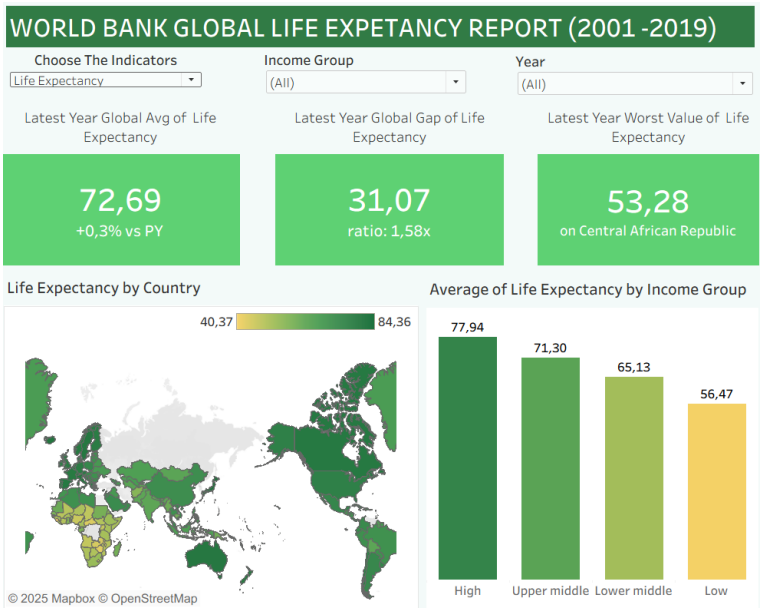
1. Data cleaning using Python
   - Columns with <10% missing values (e.g., Life Expectancy, CO_2, Health Expenditure, Unemployment) were filled using country-year interpolation.
   - Columns with >10% missing values were imputed using the median of region-year groups.
2. EDA and visualization using python and tableau.

D. Analysis



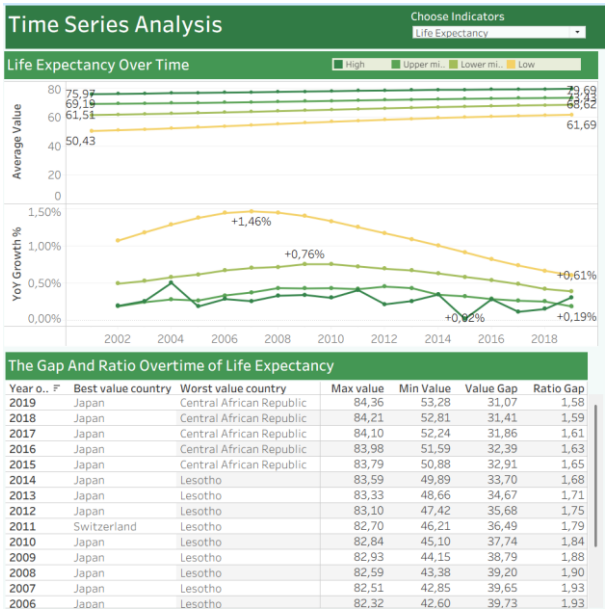
- Life Expectancy: Skewed to the left (the tail), indicating that the majority of countries already have moderate to high life expectancy, although a few countries have very low values.
- Prevalence of Undernourishment: Skewed to the right, indicating that most countries have very low levels of undernourishment, but some still experience very high cases.
- CO_2 Emissions: The distribution is skewed to the right with outliers, meaning a few countries have significantly higher levels of CO_2 emissions compared to the majority.
- Health Expenditure: The tail stretches to the right, showing that most countries spend a low to moderate budget (below 15%) on health, while a few (developed countries) spend a moderate to high share.
- Education Expenditure: Similar to health, the distribution is skewed to the right, indicating that most countries spend a low to moderate share on education compared to others.
- Unemployment: Some countries experience a high unemployment rate, while the majority have low to moderate rates.
- Sanitation: The distribution is multimodal, indicating that groups of countries fall into different levels of sanitation access.

1. Life Expectancy



Life expectancy at birth indicates the number of years a newborn infant is expected to live if prevailing patterns of mortality at the time of its birth remain the same throughout its life. It is a key metric for assessing population health. Overall, the most recent life expectancy value was 72.69 years, an increase of 0.3 years compared with the previous year.

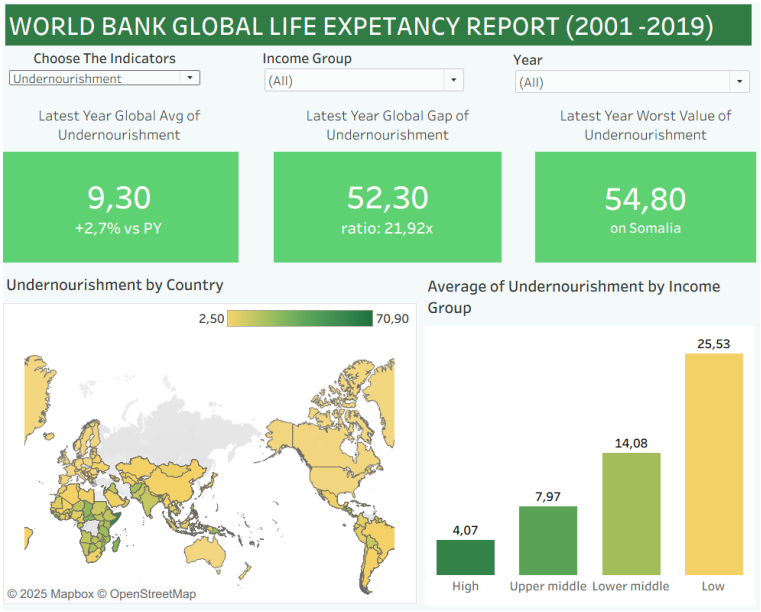
However, when looking at the details, this latest growth was largely driven by countries in the high-income group, while other groups declined—most notably the low-income group. The gap between countries with the highest and lowest life expectancy was 31 years, specifically between Japan and the Central African Republic. The positive trend is that this gap has been gradually decreasing over the years, narrowing inequality.



The Gap And Ratio Overtime of Life Expectancy

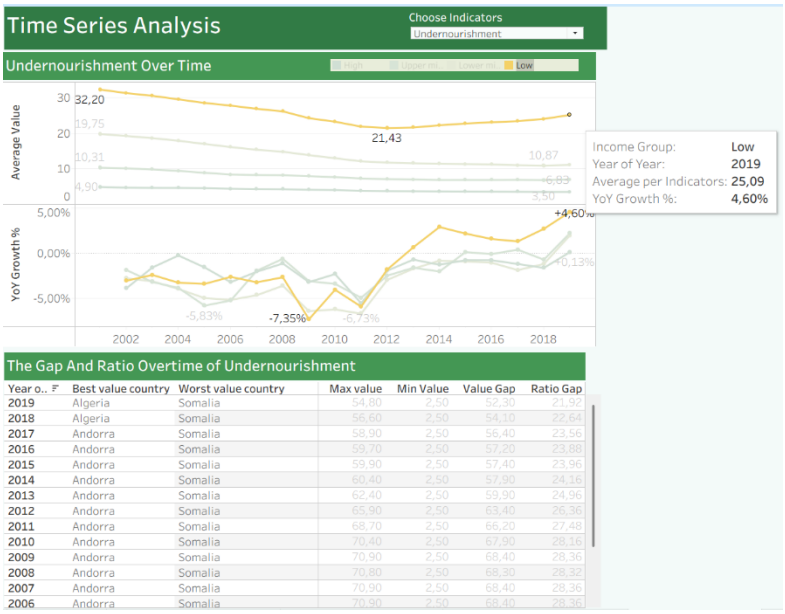
Year o..	Best value country	Worst value country	Max value	Min Value	Value Gap	Ratio Gap
2019	Japan	Central African Republic	84,36	53,28	31,07	1,58
2018	Japan	Central African Republic	84,21	52,81	31,41	1,59
2017	Japan	Central African Republic	84,10	52,24	31,86	1,61
2016	Japan	Central African Republic	83,98	51,59	32,39	1,63
2015	Japan	Central African Republic	83,79	50,88	32,91	1,65
2014	Japan	Lesotho	83,59	49,89	33,70	1,68
2013	Japan	Lesotho	83,33	48,66	34,67	1,71
2012	Japan	Lesotho	83,10	47,42	35,68	1,75
2011	Switzerland	Lesotho	82,70	46,21	36,49	1,79
2010	Japan	Lesotho	82,84	45,10	37,74	1,84
2009	Japan	Lesotho	82,93	44,15	38,79	1,88
2008	Japan	Lesotho	82,59	43,38	39,20	1,90
2007	Japan	Lesotho	82,51	42,85	39,65	1,93
2006	Japan	Lesotho	82,32	42,60	39,73	1,93

2. Prevalence of undernourishment.

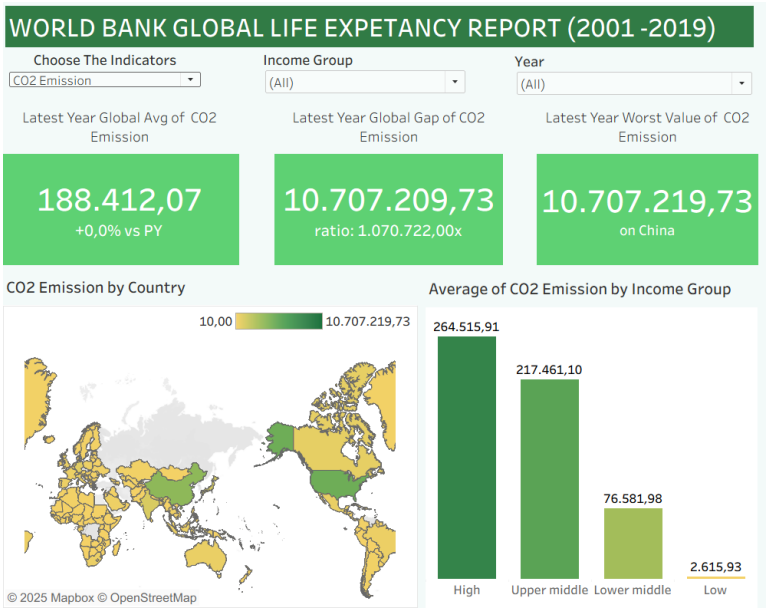


The global prevalence of undernourishment was 9.3%, an increase of 2.7% points from the previous year. Somalia recorded the highest level, with nearly 55% of its population undernourished, which was about 22 times the lowest country value. Trends show that undernourishment declined from 2002 to 2011 but has risen again since 2012, especially in low-income countries.

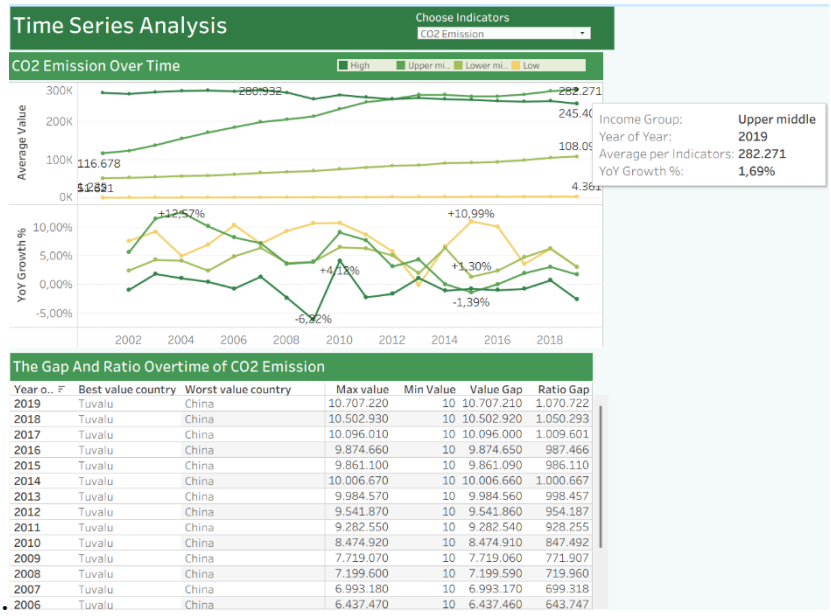
In detail, we can see a notable increase among low-income countries (yellow line). The year-on-year growth was negative from 2002 to 2011, meaning the prevalence of undernourishment declined during that period, but it began to rise again from 2012 to 2019. Given this large gap compared to other income groups, along with the increasing trend, low-income countries require urgent nutrition interventions.



3. Carbon Emission

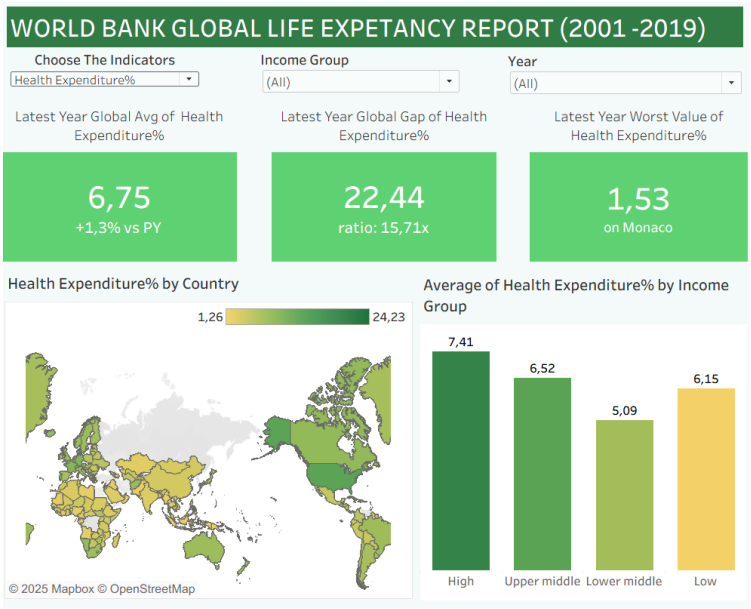


Carbon dioxide emissions come from the burning of fossil fuels and the manufacture of cement. They also include emissions produced during the consumption of solid, liquid, and gas fuels, as well as gas flaring. The latest global average of carbon emissions was 188 million tons (188 thousand kilotons).In detail, most income groups show a declining trend, while the upper-middle-income group (medium green line) experienced an increase and even surpassed the high-income group (darkest green line). This notable shift was largely the result of China's rapid industrialization, which accounts for nearly 27% of global carbon emissions and about 74% of emissions from upper-middle-income countries.



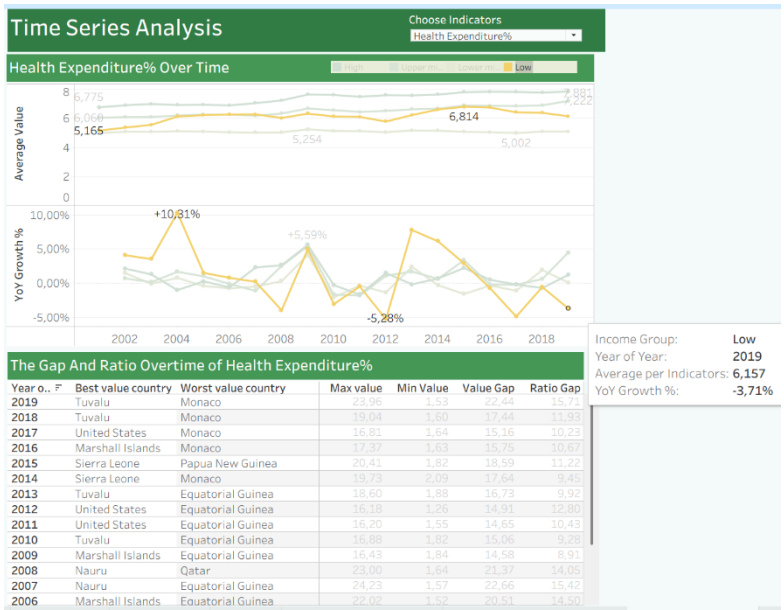
Year o..	Best value country	Worst value country	Max value	Min Value	Value Gap	Ratio Gap
2019	Tuvalu	China	10.707.220	10	10.707.210	1.070.722
2018	Tuvalu	China	10.502.930	10	10.502.920	1.050.293
2017	Tuvalu	China	10.096.010	10	10.096.000	1.009.601
2016	Tuvalu	China	9.874.660	10	9.874.650	987.466
2015	Tuvalu	China	9.861.100	10	9.861.090	986.110
2014	Tuvalu	China	10.006.670	10	10.006.660	1.000.667
2013	Tuvalu	China	9.984.570	10	9.984.560	998.457
2012	Tuvalu	China	9.541.870	10	9.541.860	954.187
2011	Tuvalu	China	9.282.550	10	9.282.540	928.255
2010	Tuvalu	China	8.474.920	10	8.474.910	847.492
2009	Tuvalu	China	7.719.070	10	7.719.060	771.907
2008	Tuvalu	China	7.199.600	10	7.199.590	719.960
2007	Tuvalu	China	6.993.180	10	6.993.170	699.318
2006	Tuvalu	China	6.437.470	10	6.437.460	643.747

4. Health Expenditure



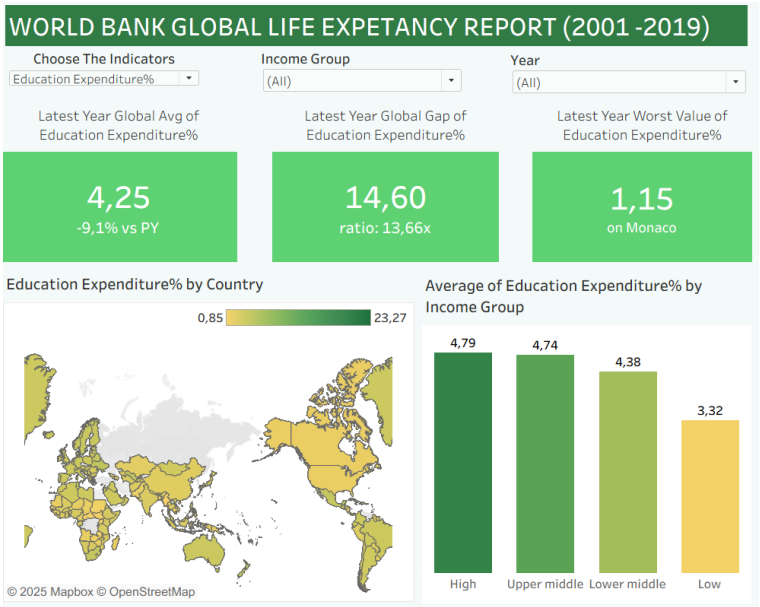
According to the dataset, Health Expenditure (% of GDP) represents the level of current health expenditure expressed as a percentage of GDP. This indicator reflects how much countries allocate to health services relative to the size of their economy. The latest global average was 6.1% of GDP, an increase of 1.3 percentage points compared to the previous year. This suggests a gradual global shift toward prioritizing health spending.

Over time, high-income countries have shown slightly increasing health expenditure, while many low-income and lower-middle-income countries have experienced a decline in recent years. This trend highlights a persistent inequality in healthcare funding, as the regions with the greatest need continue to spend the least relative to their GDP.



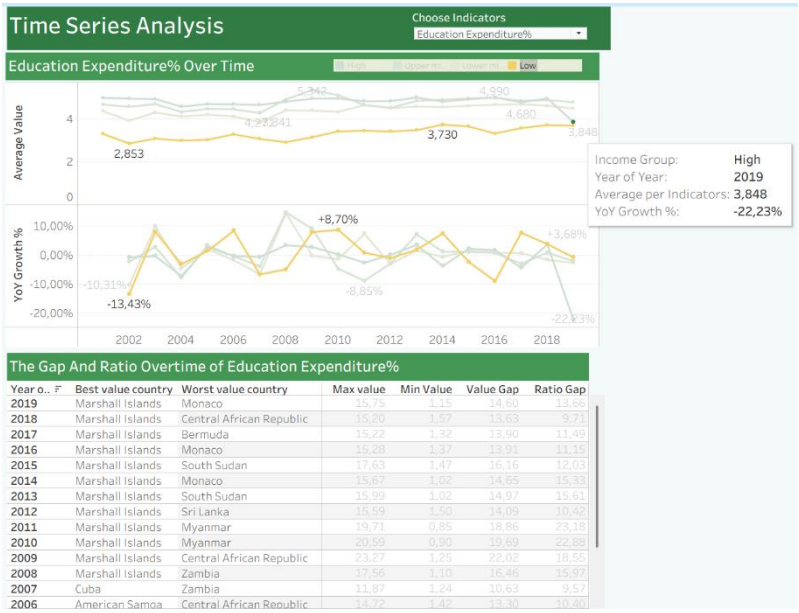
Year o..	Best value country	Worst value country	Max value	Min Value	Value Gap	Ratio Gap
2019	Tuvalu	Monaco	23,96	1,53	22,44	15,71
2018	Tuvalu	Monaco	19,04	1,60	17,44	11,93
2017	United States	Monaco	16,81	1,64	15,16	10,23
2016	Marshall Islands	Monaco	17,37	1,63	15,75	10,67
2015	Sierra Leone	Papua New Guinea	20,41	1,82	18,59	11,22
2014	Sierra Leone	Monaco	19,73	2,09	17,64	9,45
2013	Tuvalu	Equatorial Guinea	18,60	1,88	16,73	9,92
2012	United States	Equatorial Guinea	16,18	1,26	14,91	12,80
2011	United States	Equatorial Guinea	16,20	1,55	14,65	10,43
2010	Tuvalu	Equatorial Guinea	16,88	1,82	15,06	9,28
2009	Marshall Islands	Equatorial Guinea	16,43	1,84	14,58	8,91
2008	Nauru	Qatar	23,00	1,64	21,37	14,05
2007	Nauru	Equatorial Guinea	24,23	1,57	22,66	15,42
2006	Marshall Islands	Equatorial Guinea	22,02	1,52	20,51	14,50

5. Education Expenditure

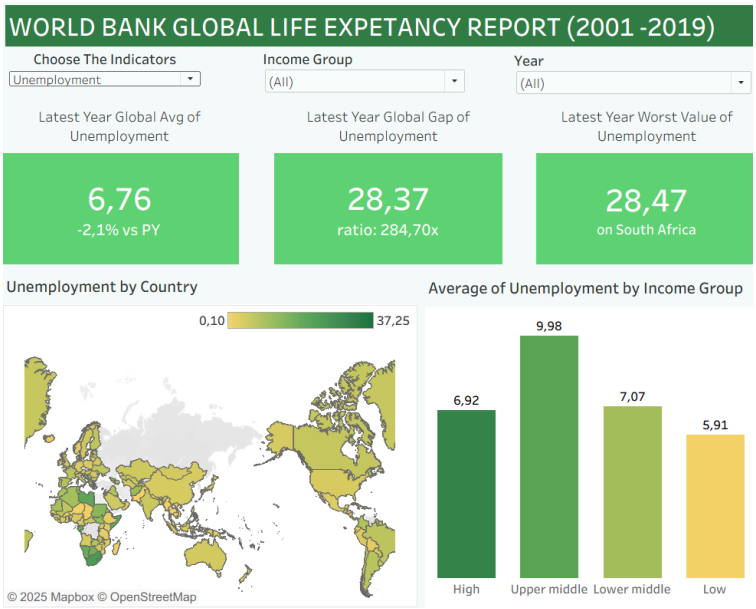


Education Expenditure (% of GDP) refers to general government spending on education, including current, capital, and transfer expenditures, expressed as a percentage of GDP. It also includes expenditures funded by transfers from international sources to the government. Globally, government expenditure on education averaged about 4.5% of GDP.

In detail, low-income countries consistently spent less than 4% of GDP on education, with figures fluctuating but never surpassing that threshold. In contrast, other income groups generally allocated between 5% and 7%. Similar to health expenditure, the gap in education investment reinforces inequality, as lower-income countries remain less able to finance long-term human capital development.

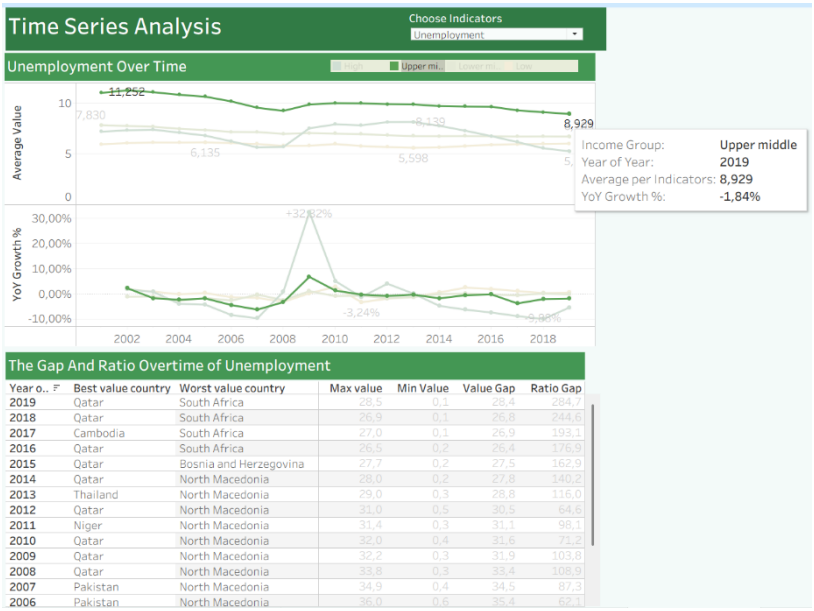


6. Unemployment

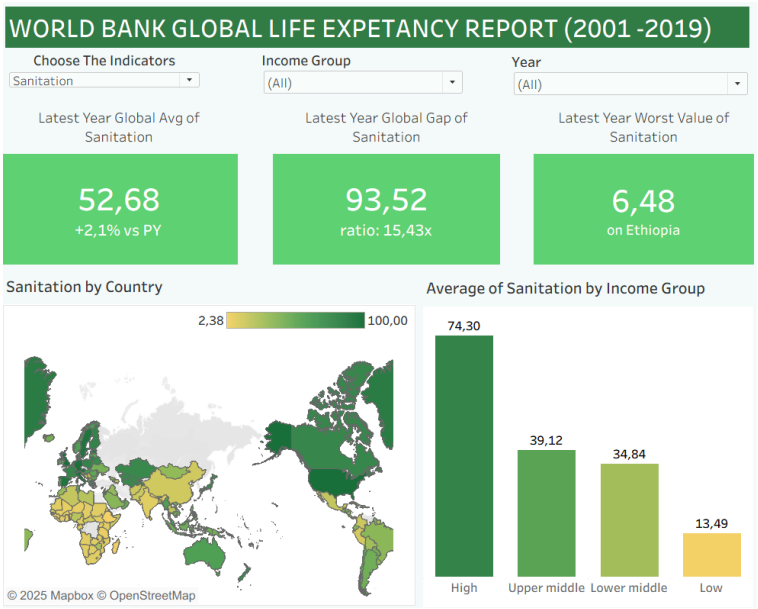


The latest global unemployment rate was 5.8%, although regional disparities remain very large. For example, South Africa recorded an unemployment rate of 28%, while Qatar reported only 0.1%—a difference of nearly 280 times. Interestingly, the income group that consistently showed the highest unemployment rates was the upper-middle-income group.

According to World Bank data, many middle-income countries (MICs) face unfinished development agendas and risk being “trapped” in middle-income status if they do not accelerate economic, social, and structural transformation. These so-called “second-generation” reforms are crucial for addressing challenges associated with a more advanced stage of development, and the lack of progress in this area is particularly evident in upper-middle-income countries (UMICs).

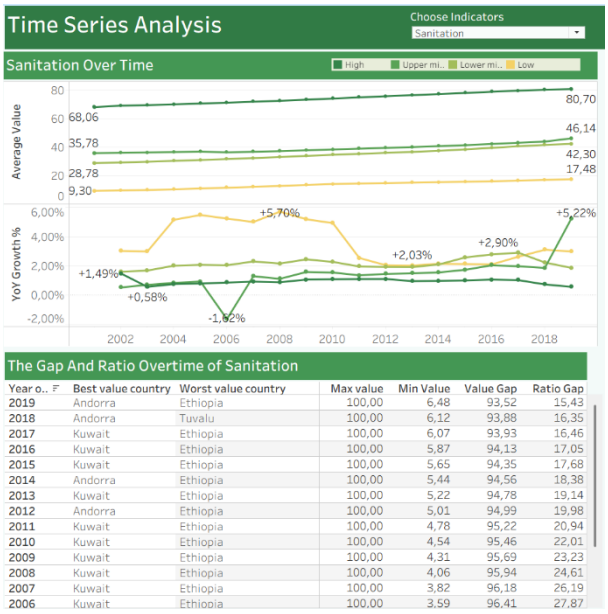


7. Sanitation



Access to improved sanitation showed a multimodal distribution, reflecting stark differences between income groups. High-income countries had nearly universal access, while low-income countries lagged far behind, with some reporting coverage below 40%. Progress has been gradual but uneven. Between 2000 and 2020, the global average improved significantly, yet the gap between low-income and high-income countries remained wide.

In detail, sanitation access in lower-middle, upper-middle, and high-income groups grew steadily, with values gradually rising over time. In contrast, low-income countries showed volatile trends.



Year o..	Best value country	Worst value country	Max value	Min Value	Value Gap	Ratio Gap
2019	Andorra	Ethiopia	100,00	6,48	93,52	15,43
2018	Andorra	Tuvalu	100,00	6,12	93,88	16,35
2017	Kuwait	Ethiopia	100,00	6,07	93,93	16,46
2016	Kuwait	Ethiopia	100,00	5,87	94,13	17,05
2015	Kuwait	Ethiopia	100,00	5,65	94,35	17,68
2014	Andorra	Ethiopia	100,00	5,44	94,56	18,38
2013	Kuwait	Ethiopia	100,00	5,22	94,78	19,14
2012	Andorra	Ethiopia	100,00	5,01	94,99	19,98
2011	Kuwait	Ethiopia	100,00	4,78	95,22	20,94
2010	Kuwait	Ethiopia	100,00	4,54	95,46	22,01
2009	Kuwait	Ethiopia	100,00	4,31	95,69	23,23
2008	Kuwait	Ethiopia	100,00	4,06	95,94	24,61
2007	Kuwait	Ethiopia	100,00	3,82	96,18	26,19
2006	Kuwait	Ethiopia	100,00	3,59	96,41	27,87