Data Visualisation : Assignment 3

Ayush Arya Kashyap
IMT2022129
IIIT - Bangalore
Bangalore, India
ayusharya.kashyap@iiitb.ac.in

Uttam Hamsaraj
IMT2022524
IIIT - Bangalore
Bangalore, India
uttam.hamsaraj@iiitb.ac.in

Pranav Laddhad
IMT2022074
IIIT - Bangalore
Bangalore, India
pranav.laddhad@iiitb.ac.in

I. VISUAL ANALYTICS WORKFLOW

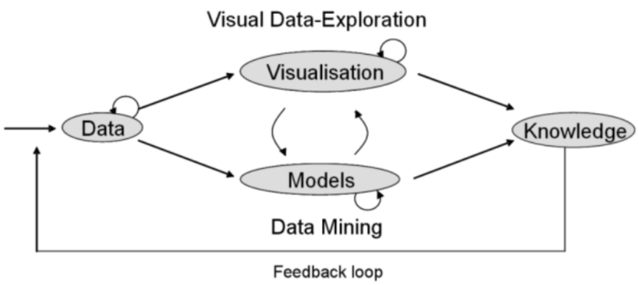


Fig 1. Kiem et al. Visual Analytics Workflow, Image Courtesy: [1]

II. TASK 1 : AN ADVANCED ANALYSIS ON THE GDP OF THE COUNTRIES AND THEIR SHIFTS

We start by summarizing Task 1 from the DAS732 Assignment 1, detailing the visualization methodology and the inferences derived. This is considered the initial iteration (run) of our workflow. Based on the insights and observations gained, we identify areas in our methods that could be improved and work on integrating these refinements into subsequent iterations of the workflow. The workflow illustrated in Figure 1 by Kiem et al. [1] serves as a source of inspiration.

A. *Prior Visualizations and Inferences*

Please Refer to DAS732 ssignment 1 Report [2]. This section summarizes my task for DAS732 Assignment 1 (Task 1), where I aimed to give an analysis of trends and impacts of major economic events across the globe over the years. It serves as an appendix for Assignment 3, and may initially be skipped by the reader.

Visualization Methodology Assignment 1

- We began by visualizing the GDP per capita of major global economies, including the USA, Japan, China, India, and Russia, from 1960 to 2022.
- Next, we created stacked bar charts and traditional bar charts to visualize the tax revenues of these key countries over the same period.
- An in-depth analysis was conducted on Luxembourg, the country with the highest GDP per capita, to explore its economic performance in greater detail.
- Additionally, we generated several visualizations to highlight different aspects of economic trends:
  - A line and bar chart illustrating Ukraine's GDP trends over the years.
  - A combined bar and line chart comparing the central government debt and interest rates of the USA.
  - A circular chart showcasing the countries with the highest annual inflation rates.
  - An area graph depicting the fluctuating inflation rates in Sri Lanka over time.
  - A treemap representing Argentina's GDP per capita over the past 25 years.
  - Finally, a heatmap was created to compare the GDP per capita of countries across the world, offering a comprehensive view of global economic disparities.

Inferences from Assignment 1 Please note that the figure numbers in this subsection indicate the figure numbers as mentioned in the report for Assignment 1 [2].

- **Global Economic Trends:** The report analyzes various economic indicators across countries, comparing leaders like the USA, Luxembourg, and China, as well as countries facing instability like Ukraine and Sri Lanka. The focus is on understanding how government policies, geopolitical events, and other factors shape economic growth, resilience, and vulnerabilities.
- **GDP per Capita:** Examining trends in GDP per capita reveals how economies such as Luxembourg, the USA, and China have grown, while countries like Afghanistan show stagnation (Fig. 1).
- **Tax Revenue:** High tax-to-GDP ratios in countries like Lesotho and New Zealand show the role of taxation in sustaining public services, while developing nations face challenges in tax collection (Fig. 2).
- **Luxembourg's Growth:** Luxembourg stands out with its high GDP per capita, sustained by its status as a financial hub. Despite global economic crises, Luxembourg has maintained resilience (Fig. 3, 4).
- **Ukraine's Struggles:** The economic difficulties in Ukraine, especially post-2014, underline the negative impacts of geopolitical conflict on economic performance, including severe GDP contraction (Fig. 5).
- **USA's Financial Crisis:** The USA's government

debt and real interest rate trends from 2005 to 2010 illustrate the economic responses to the 2008 financial crisis, where stimulus measures resulted in rising debt levels despite declining interest rates (Fig. 6).

- **Luxembourg's Model:** Luxembourg's economic strategies, including embracing technology and focusing on renewable energy and services, offer a model for small, high-income economies (Fig. 3, 4).
- **China's Growth:** China's rapid GDP per capita growth from the late 1990s reflects the success of its economic reforms, industrialization, and integration into the global market (Fig. 9).
- **USA's Debt and Interest Rates:** The USA's experience shows the delicate balance between using debt for economic recovery and the potential long-term consequences on fiscal policy and inflation (Fig. 10).

New ideas and improvements The following points are considered as feedback from the visualizations generated for Assignment 1:

- **Technological Innovation:**
  - **Luxembourg**: Leverage digital transformation and renewable energy for sustainable growth.
  - **China**: Transition to a consumption-driven economy while maintaining industrial strength.
  - **Germany**: Continue investing in green technologies to maintain economic leadership.
- **Economic Inequality:**
  - **USA**: Address rising inequality and its effects on economic mobility.
  - **Brazil**: Focus on poverty reduction and social inclusion to boost growth.
  - **India**: Enhance access to education and healthcare to reduce disparity.
- **Government Debt & Fiscal Policy:**
  - **USA**: Reevaluate fiscal policies, especially post-crisis stimulus measures.
  - **Ukraine**: Strengthen economic resilience through international cooperation and internal reforms.
  - **Sri Lanka**: Improve fiscal management to avoid further debt crises.
- **Trade & Global Integration:**
  - **China**: Expand its global market presence while mitigating trade risks.
  - **Germany**: Enhance trade partnerships in a post-Brexit Europe.
  - **Brazil**: Diversify trade to reduce dependence on a few partners.
- **Natural Disasters & Climate Change:**
  - **Japan**: Invest in climate resilience and disaster preparedness.
  - **Australia**: Adapt to climate challenges with sustainable resource management.
  - **Luxembourg**: Develop green technologies to meet sustainability goals.
- **Economic Growth & Development:**
  - **Luxembourg**: Continue driving growth through financial services and innovation.
  - **Afghanistan**: Focus on rebuilding infrastructure and attracting international investment.
  - **Germany**: Maintain stability through industrial innovation and social welfare systems.
- **Smaller economies** like Luxembourg should focus on technological adoption and high-value industries.
- **Larger economies** like the USA and China need balanced strategies addressing inequality and global trade risks.
- **Emerging economies** like India and Brazil should improve social systems and diversify trade for sustainable growth.

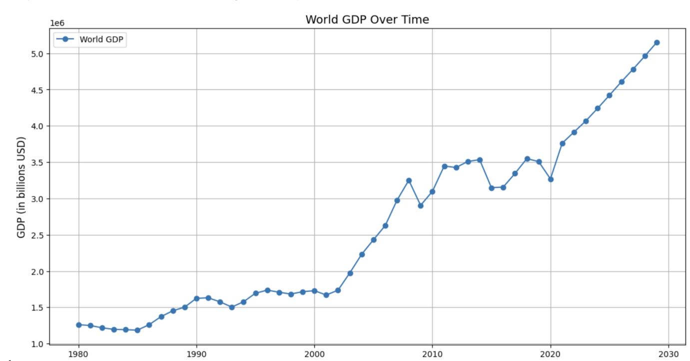
Thus, from among these ideas (based on feedback) we will be focusing on a few and incorporating them in the workflow.

B. Data Preprocessing

Preprocessing using Python and Pandas

- Remove rows or columns with missing values (e.g., drop rows where the "Year" column is NaN).
- Replace missing values with a statistical measure, such as the mean or median of the respective column.
- Ensure correct data types for each column, such as converting the "Year" column to integers.
- Convert all values to numeric where applicable, using coercion for non-numeric data.
- Normalize the data (scaling values to a range, like 0-1) or standardize it (converting to have zero mean and unit variance).
- Remove columns with no useful data (e.g., columns filled entirely with NaN values).
- Identify and treat any outliers that could skew results or analysis.

C. Visualisations and Observations



],

Fig 2. World GDP growth over time

We initially plot a line chart (Fig. 2) that plots the global GDP (sum of all countries' GDP) over time. The trend shows how the global economy has grown or contracted across the years. There might be certain periods where a sharp decline or growth is observed, which corresponds to global economic events like recessions or booms. Increases in global GDP correlate with industrial revolutions, globalization, and technological advancements, while decreases indicate financial crises or global slowdowns. The chart gives us the predicted valued too till 2029.

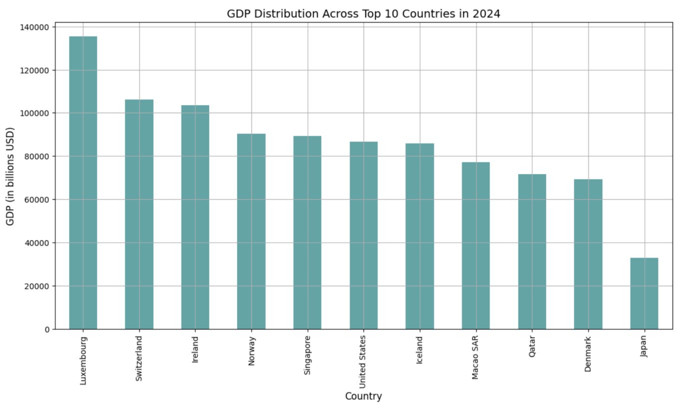


Fig 3. GDP per capita distribution across countries in the most recent year

We plot a bar chart (Fig. 3) that displays the GDP distribution among all countries in the most recent year. This chart would clearly show the disparity in GDP between countries, where a small number of countries account for the majority of the GDP. Many countries likely have much smaller GDPs compared to major economies like the U.S. and China, highlighting global economic inequality.

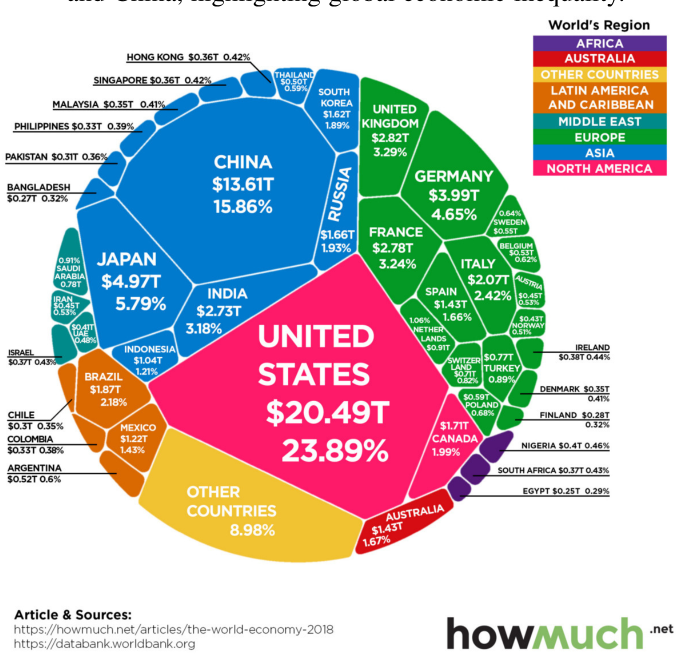


Fig 4. Treemap depicting the global GDP distribution, Image Courtesy: [3, 4]

The treemap depicting the global GDP distribution in 2018 highlights the dominance of the United States and China, which together account for nearly 40% of the world's GDP, with the U.S. contributing 23.89% and China 15.86%. Other significant economies include Japan, Germany, and India, with North America and Europe collectively making up over 40% of the global economy. Meanwhile, regions like Africa and Latin America have smaller contributions, with Nigeria and Brazil being the largest economies in their respective regions. The data suggests that emerging markets such as India, China, and Brazil show substantial growth potential, while there is a noticeable disparity between the economic outputs of high-income countries and smaller economies. Overall, the U.S. and China remain the largest players in the global economy, but emerging markets and developing regions are gradually increasing their economic influence.

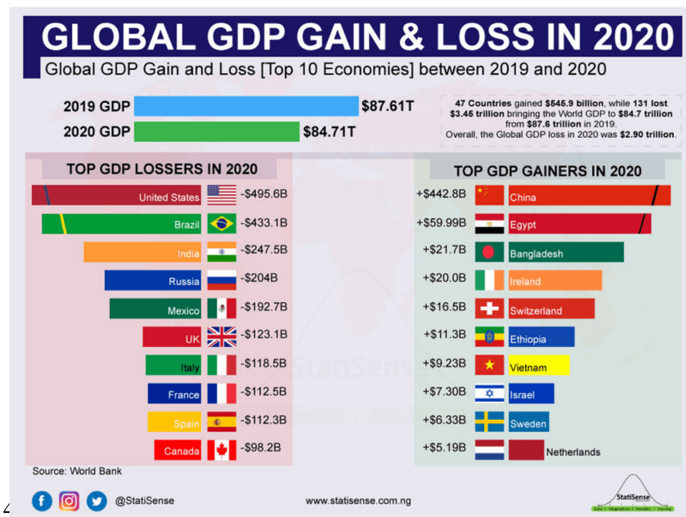
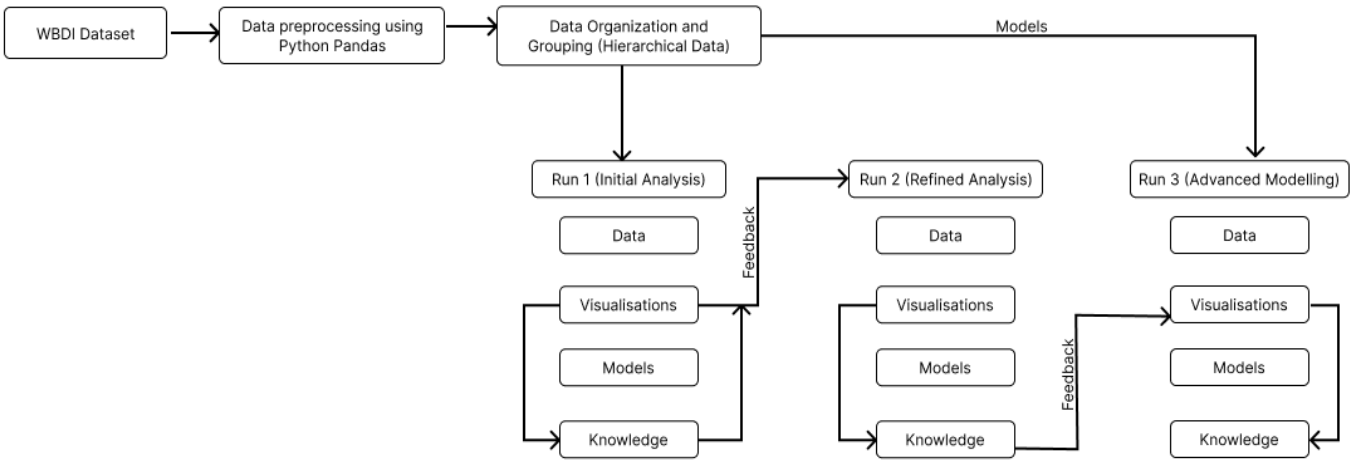


Fig. 5 Global GDP Gain and Loss [Top 10 Economies] between 2019 and 2020, Image Courtesy: [4]

The infographic highlights the global GDP impact of the COVID-19 pandemic, with a $2.90 trillion loss globally as GDP fell from $87.61 trillion in 2019 to $84.71 trillion in 2020. The United States suffered the largest loss at $495.6 billion, followed by Brazil ($433.1 billion) and India ($247.5 billion), while countries like Russia, Mexico, and major European economies also faced significant declines. On the other hand, 47 countries recorded gains, led by China with $442.8 billion, followed by Egypt and Bangladesh. Emerging economies like Vietnam, Ethiopia, and Ireland also posted growth. This reflects advanced economies facing steeper declines, while some developing and emerging markets showed resilience and even growth, underlining varying recovery capacities.

D. Describing The Visual Analytics Workflow

1) First Run
   - **Data :** Minimal data processing was done, null values were appropriately imputed based on variables. New column called `gdp_per_capita` was added to the dataset for advanced visualisations. This was calculated by dividing the `GDP_current_US` by the `Population` column which gives us the GDP per capita of a country. Data over for all the years was handled efficiently.
   - **Visualisations :** Refer to A1 report [2] for the visualizations. Figure numbers are as per A1.
     * **GDP per Capita Trends for Selected Countries:** A stacked bar chart displays the GDP per capita for Afghanistan, Brazil, China, Germany, India, and the USA from 1960 to 2022.
     * **Tax Revenue as a Percentage of GDP:** A bar chart illustrates the top 10 countries with the highest tax revenues as a percentage of GDP from 1960 to 2022.
     * **Luxembourg's Economic Analysis:** Line graphs show Luxembourg's GDP and GDP per capita trends (1974–2022), complemented by additional charts on internet adoption, renewable energy consumption, and trade in services.
     * **Ukraine's Economic Decline:** A line graph depicts Ukraine's GDP trends from 1989 to 2023, highlighting major fluctuations during periods of geopolitical conflict.
     * **USA Government Debt and Real Interest Rates:** A mixed line and bar chart illustrates real interest rates and government debt trends in the USA from 2005 to 2010.
     * **Inflation Trends in High-Inflation Nations:** A circle chart highlights countries with the highest annual inflation percentages during key periods.
     * **Inflation in Sri Lanka:** An area graph illustrates the annual inflation rates in Sri Lanka from 1962 to 2022, emphasizing major spikes.
     * **Global GDP Growth in 2002:** A heatmap visualizes GDP growth percentages across countries in 2002.
     * **Argentina's GDP Per Capita:** A treemap shows Argentina's GDP per capita trends from 1998 to 2005.
   - **Models :** No models were used when going through the workflow for the first run. Simple statistical measures such as mean and total sum alongwith historical and noticable events were used to choose which countries to focus on.
   - **Knowledge :** Refer to A1 report [2] for the visualizations. Figure numbers are as per A1.
     * **GDP per Capita Trends for Selected Countries (Fig. 1):** The United States and Germany show consistent growth, while China and India exhibit rapid development since the 1990s and 2000s, respectively. Afghanistan remains stagnant due to political instability.
     * **Tax Revenue as a Percentage of GDP (Fig. 2):** Lesotho and New Zealand have high tax-to-GDP ratios, indicating robust tax systems, while developing nations face challenges in tax collection.
     * **Luxembourg's Economic Analysis (Fig. 3, 4):** Luxembourg demonstrates resilience with high GDP per capita, increasing internet adoption, and rising renewable energy consumption, despite minor dips during crises.
     * **Ukraine's Economic Decline (Fig. 5):** Severe GDP declines during 2014–2017 were caused by geopolitical conflict, trade loss with Russia, and inflation.
     * **USA Government Debt and Real Interest Rates (Fig. 6):** Government debt spiked during the 2008 financial crisis, while real interest rates fell to stimulate recovery.
     * **Inflation Trends in High-Inflation Nations (Fig. 7):** Countries like Brazil, Peru, and Angola experienced severe inflation due to economic mismanagement and external shocks.
     * **Inflation in Sri Lanka (Fig. 8):** Inflation surged above 50% in 2022, driven by economic distress, following periods of stability in earlier years.
     * **Global GDP Growth in 2002 (Fig. 9, 10):** Many countries saw negative growth post-dot-com bubble, while China and India maintained moderate positive growth.
     * **Argentina's GDP Per Capita (Fig. 11):** Argentina experienced a sharp decline during the 2001–2002 crisis, followed by slow recovery from 2003 onwards.
   - **Feedback :**
     * **Focus on the USA and Japan:** The USA and Japan are among the largest economies, significantly shaping global trade, finance, and technology. The USA's consistent GDP growth and Japan's resilience despite demographic challenges make them benchmarks for economic analysis. Highlighting sectoral contributions (e.g., technology in the USA, manufacturing in Japan) can provide deeper insights.
     * **Global Perspective:** A global analysis reveals disparities between developed and developing nations, emphasizing growth potential and



Workflow 1. A sketch of the Visual Analytics Workflow used for the assignment. The rectangles depict a particular sub-task component, pill-boxes depict final outcome or result, lines depict the general flow and feedback (as labelled).

challenges. Heatmaps or clustering countries based on GDP, inflation, and tax revenues could improve understanding of regional differences and global trends.

2) Second Run
   - **Data :** We have extracted the data only for USA and Japan as those are the only 2 countries we will be focusing on for further advanced analysis. We took into account different factors for these countries which mostly impact their GDP.
   - **Models :** We have not used any models in this run, it is mostly just visualisations for Japan and USA.
   - **Visualisations :**

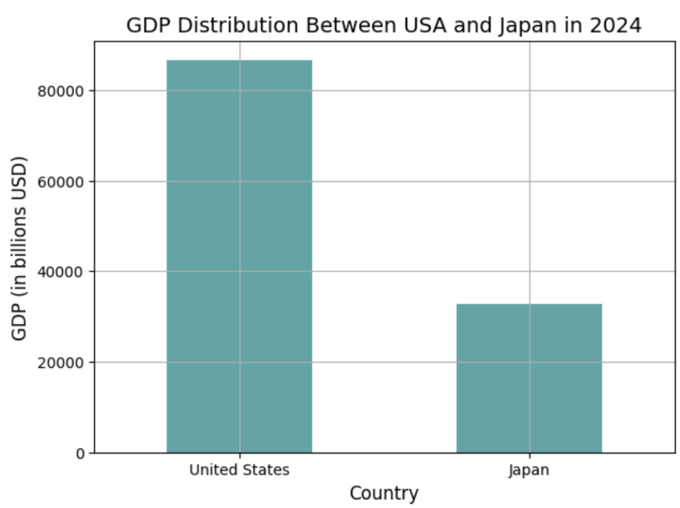


Fig. 6 GDP of USA and Japan (In billions USD)

GDP of USA and Japan : Plotted a detailed bar chart illustrating the GDP of Japan and the USA in billions of USD over several years for comparative analysis.

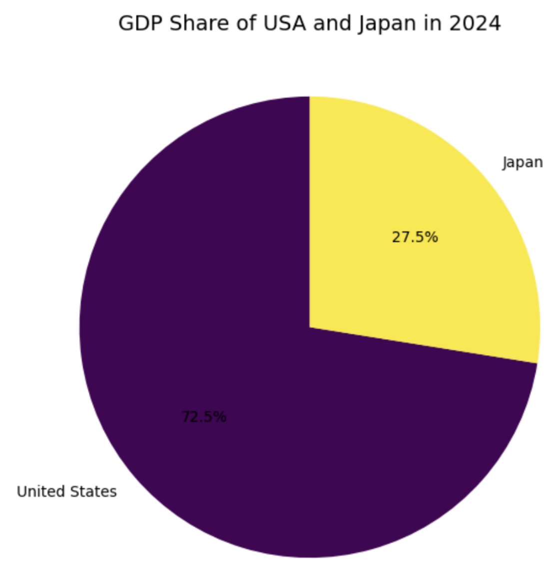


Fig. 7 GDP comparison of USA and Japan in terms of percentage

Relative comparison of GDP of USA and Japan : Plotted a pie chart showing the comparison in terms of GDP for Japan and USA.

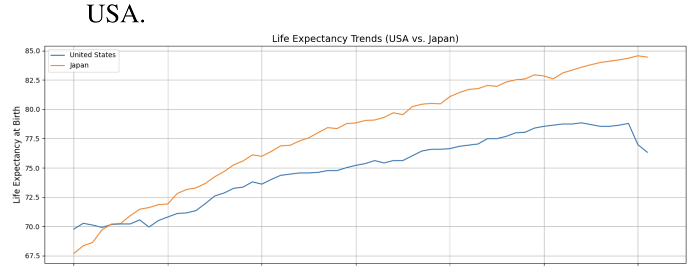


Fig. 8 Population trend of USA and Japan over the years

∗ **Trends in the life expectancy of USA and**

Japan : Plotted a comprehensive line graph showing the trends in the life expectancy of the USA and Japan over the years, highlighting key differences and patterns.

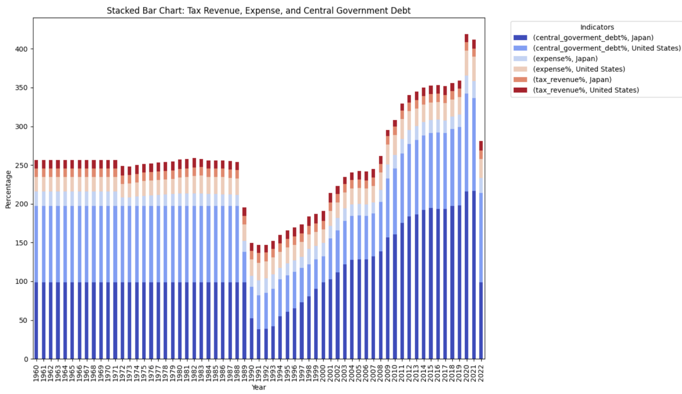


Fig. 9 Stacked bar chart for the tax revenue, expense and central government debt for Japan and USA over the years

* **Trends in the different factors affecting GDP for USA and Japan :** Plotted a stacked bar chart showing the comparison in the tax revenue, expense and central government debt of Japan and USA.

– **Knowledge :**
  * It is evident that the United States has a significantly higher GDP compared to Japan, with its economy being more than twice the size of Japan's. This disparity highlights the dominant position of the United States in global economic output, driven by stronger economic productivity, larger consumer markets, and more extensive resource utilization. Japan's comparatively lower GDP underscores its smaller economic scale in relation to the United States.
  * Japan's population is consistently higher and the gap widens. The USA experiences a sharp decline, indicating possible health or socioeconomic issues, while Japan's growth stabilizes. Higher life expectancy typically supports GDP through a healthier workforce, but the USA's decline may increase healthcare costs and reduce productivity. Japan's aging population and plateau in life expectancy could strain GDP due to rising healthcare and pension demands.
  * We see a rising central government debt in both Japan and the U.S., with Japan's debt-to-GDP ratio significantly higher in recent years. Government expenses consistently exceed tax revenue, contributing to this growth. A notable spike in debt occurs post-2008, reflecting borrowing to counter the global financial crisis. While high debt levels may initially stimulate GDP through economic support, they can hinder long-term growth due to increased interest payments and reduced fiscal flexibility. Insufficient tax revenue further challenges GDP growth by limiting funds for essential programs and infrastructure.

– **Feedback :** The visualizations successfully highlight key trends and contrasts, such as the widening life expectancy gap between Japan and the U.S., the sharp post-2008 spike in government debt, and the disparity in GDP distribution between the two countries. However, they fall short in providing a holistic narrative or actionable insights. For example, the life expectancy visualization does not delve into the socioeconomic, healthcare, or demographic drivers behind the trends, nor does it account for the impact of specific policies or crises. The government debt and expense visualization effectively shows rising debt and its mismatch with tax revenue but fails to connect these trends to macroeconomic factors, such as fiscal policies or global economic events. Finally, the GDP comparison offers a simplistic view without contextualizing the role of factors like population size, economic structure, or productivity. Incorporating contextual data or annotations explaining these trends and their implications would make the visualizations more informative and impactful.

3) Third Run

– **Data :** No changes from the second run
– **Models :** We selected Linear Regression, Polynomial Regression, ARIMA, and Random Forest for GDP prediction due to their complementary strengths. Linear Regression provides a simple, interpretable baseline for modeling the relationship between year and GDP, assuming linear growth. Polynomial Regression extends this by capturing non-linear trends, which are often observed in economic data. ARIMA (Auto-Regressive Integrated Moving Average) is a time series model well-suited for capturing temporal dependencies and trends, allowing for forecasting based on past GDP values. Lastly, Random Forest, an ensemble method, is robust against overfitting and effective in modeling complex, non-linear relationships. Together, these models enable a comprehensive analysis, balancing interpretability, flexibility, and predictive power to address the intricacies of GDP data. The following graph shows the model comparisons for both the countries.

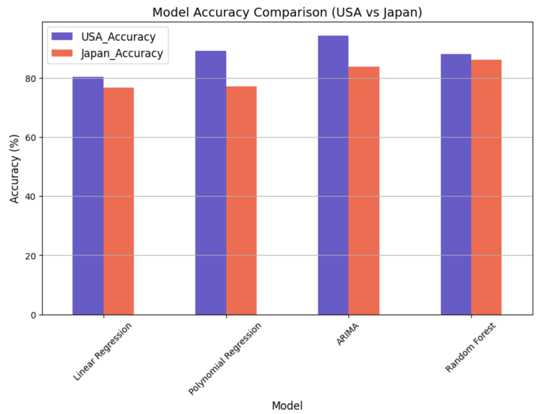


Fig. 10 Bar chart comparing the performance of each model for predicting the GDP of Japan and USA

The best accuracy was given by ARIMA for USA and Random Forest for Japan. We had taken different train and test data for Japan and USA after extracting their information from the dataset. The following line chart shows us the continuation of our actual data till 2024 and the predicted values from 2024 to 2029.

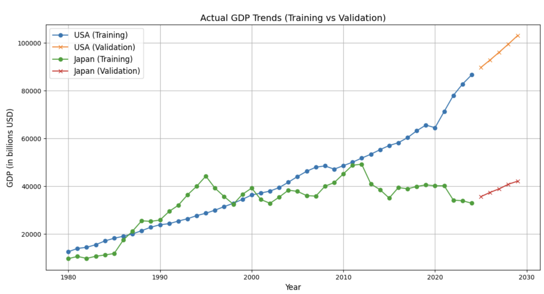


Fig. 11 Line chart representing the continuation in actual and predicted values for GDP of USA and Japan

Accuracy Metrics (MAPE and Percentage Accuracy):

Model	USA_MAPE	USA_Accuracy	Japan_MAPE	Japan_Accuracy
Linear Regression	19.6146	80.3854	23.3513	76.6487
Polynomial Regression	10.749	89.251	22.7787	77.2213
ARIMA	5.59674	94.4033	16.2375	83.7625
Random Forest	11.9101	88.0899	13.7524	86.2476

Table 1 shows us the accuracies of the different models used

– **Visualisations :**

∗ We analyzed and visualized the predicted GDP values for the next five years, specifically from 2024 to 2029, to gain a deeper understanding of the economic trends during this period. As part of our analysis, we created a pie chart that highlights the GDP comparison between the United States and Japan. This chart effectively illustrates the relative proportions of GDP contributed by each country in terms of percentage, providing a clear and concise visual representation of their economic standing.

∗ The radar chart illustrates the GDP proportions of the USA and Japan as percentages for the years 2024 to 2029. Each vertex of the chart represents a specific year, while the blue and orange polygons denote the proportions for the USA and Japan, respectively. By comparing the size and shape of these polygons, the chart highlights the relative economic standing of the two countries over time. If the blue polygon (USA) consistently covers a larger area than the orange polygon (Japan), it signifies that the USA maintains a higher GDP proportion throughout the observed period. Variations in the size or structure of the polygons across the years reflect changes in the relative economic growth or contraction of each country.

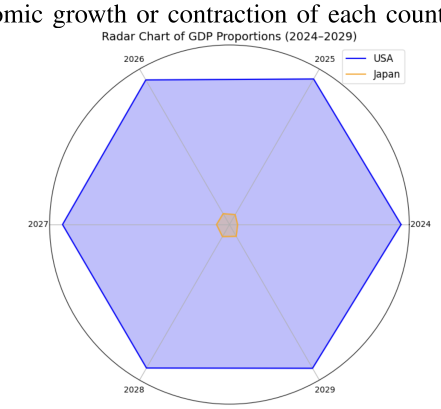


Fig. 12 Radar Chart of GDP Proportions (2024–2029): Comparison of USA and Japan's GDP as Percentages Over Time

∗ In addition to the pie chart, we also generated a bar chart to further compare the GDP per capita of Japan and the United States. This bar chart serves as a valuable tool for performing a comparative study, showcasing the differences in GDP per capita over the specified period.

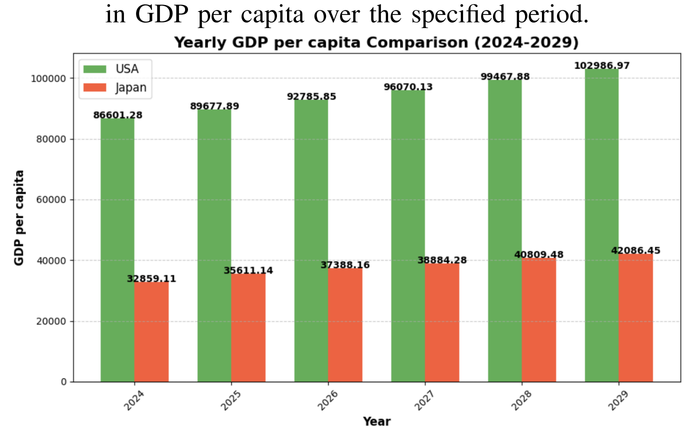


Fig. 14 Bar chart showing the GDP per capita of Japan and USA from 2024 to 2029

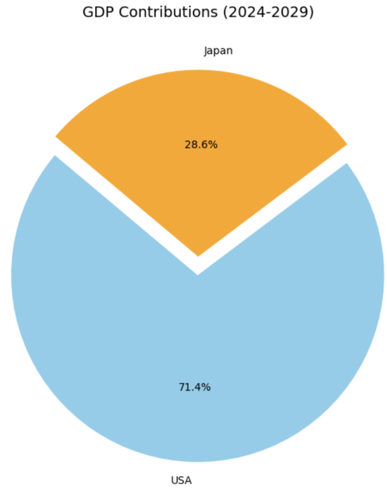


Fig. 13 Pie chart showing GDP comparison as percentage for USA and Japan from 2024 to 2029

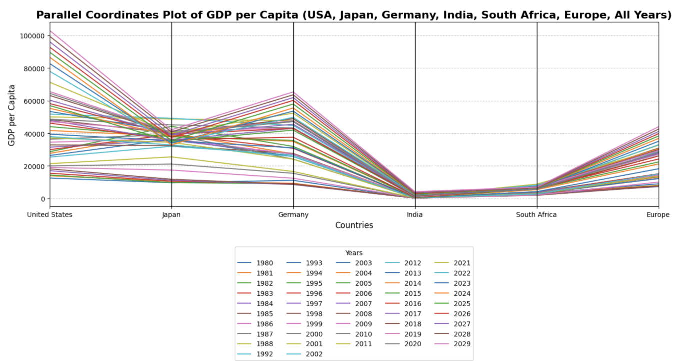


Fig. 15 Parallel coordinates plot showing GDP of Japan, USA, Europe, South Africa and India

* We plotted a SPLOM which shows the comparison between Japan and USA in a consolidated way.

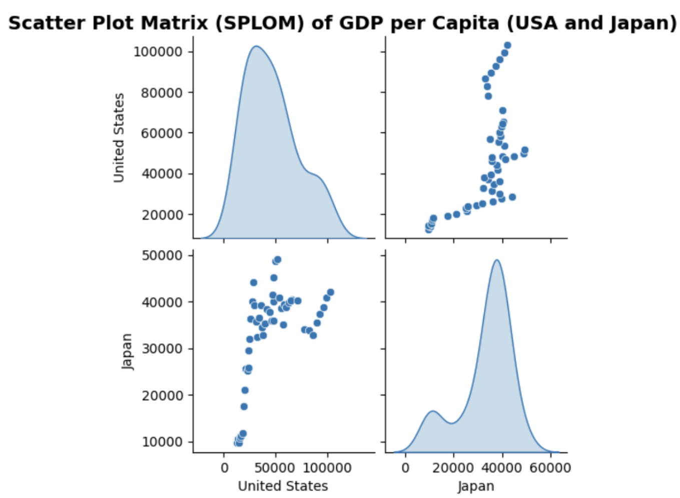


Fig. 16 SPLOM showing the GDP comparison of USA and Japan

* We plotted a Parallel Coordinates Plot which shows us the trends in the GDP for USA, Japan, Germany, India, South Africa and Europe. We have included more countries and continents other than USA and Japan to show a comparison of the GDP trends in those countries as predicted from 2024 to 2029.

– **Knowledge :**
  * From 2024 to 2029, the USA contributes 71.4% of the total GDP, while Japan accounts for 28.6%. This disparity likely reflects the USA's larger population, diversified economy, and global dominance in high-revenue industries such as technology, finance, and energy, compared to Japan's export-driven and aging economy.
  * We see a consistently higher GDP per capita of the USA, growing from \$86,601.28 in 2024 to \$102,986.97 in 2029, while Japan increases from \$32,859.11 to \$42,086.45. The USA's lead may be attributed to higher productivity, stronger innovation, and a more robust service sector, whereas Japan faces structural challenges such as a declining workforce and slower economic diversification.
  * We see a stark contrast in GDP per capita between the USA, Japan, and other regions. The USA consistently leads with the highest GDP per capita, followed by Japan, showcasing their advanced economies. In comparison, developing regions like India and South Africa lag significantly, reflecting economic disparities. Europe and Germany show moderate GDP per capita levels, bridging the gap between the extremes. This visualization emphasizes the dominance of the USA and Japan in individual economic output compared to other countries and continents.
  * We can observe the distribution of GDP values for each country, where the United States has a wider range and significantly higher GDP per capita compared to Japan, which has a more concentrated distribution. The off-diagonal scatter plots highlight the correlation between the two countries' GDP per capita. A slight positive trend can be seen, indicating that as the GDP per capita of one country increases, there tends to be a corresponding increase in the other, although Japan consistently lags behind the United States in absolute values.

– **Feedback :** The visualizations provide a clear and comprehensive depiction of GDP per capita trends across various countries, effectively illustrating the stark contrasts between economically

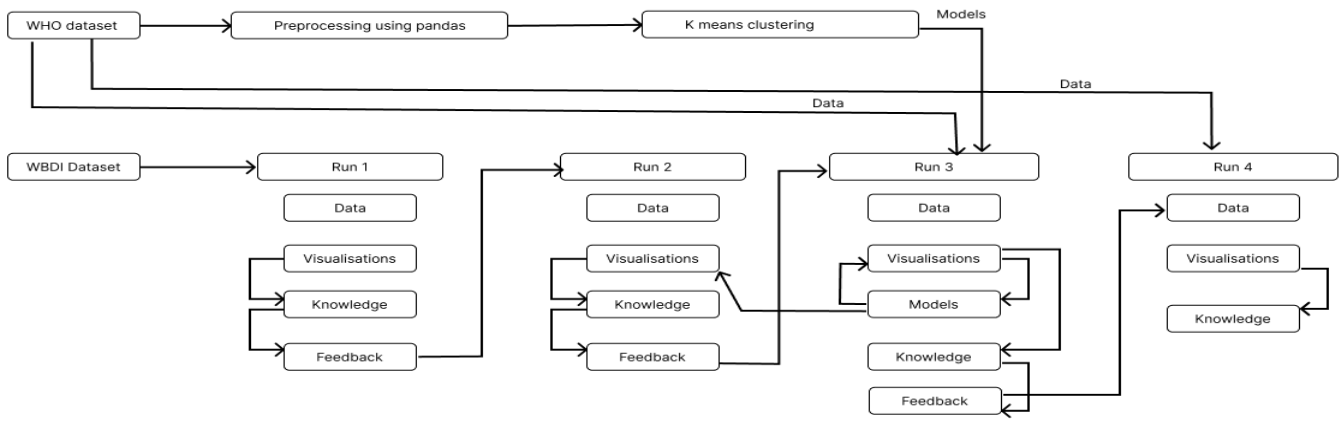


Figure 17. A sketch of the Visual Analytics Workflow used for the TASK-2. The rectangles depict a particular sub-task component, pill-boxes depict final outcome or result, lines depict the general flow and feedback (as labelled).

advanced nations like the USA and Japan and other regions. They highlight temporal and regional patterns in a visually engaging manner, enabling quick identification of global economic disparities. These insights can serve as a valuable foundation for further analysis, helping policymakers and researchers focus on specific regions or periods to understand and address economic differences more effectively.

- **Summary :** The three runs in the document reflect progressive iterations of the experiment or model setup. The first run served as a baseline, using a standard configuration or default parameters to establish a foundational performance benchmark. Building on this, the second run involved modifications to optimize specific variables or test hypotheses derived from the first run, aiming for improved outcomes or deeper insights. Finally, the third run incorporated further refinements or a distinct experimental approach to validate previous findings and achieve a more robust or enhanced performance. Each run represents a step toward fine-tuning and understanding the underlying system or process.

III. TASK 2 : AN ADVANCED ANALYSIS ON LIFE EXPECTANCY

A. *Visual Analytics Workflow*

In this workflow, we focus on analyzing life expectancy in detail, a key social indicator. Our goal is to gain deeper insights into life expectancy trends and patterns. In Assignment 1, we explored various social indicators and examined how they impact countries. In this visual workflow, we focus specifically on life expectancy, building on the insights gained from Assignment 1. The workflow is shown in Fig. 17. We first discuss about the datasets used in this workflow, and then we will look at the detailed visual analytics description.

B. *Datasets used in the workflow*

In this workflow, we have used the World Bank's World Development Indicators dataset, which was used in A1. In addition, we have included the gender-wise distribution of life expectancy dataset and the number of deaths due to HIV/AIDS by country. These additional datasets help us analyze life expectancy in greater depth.

C. *Describing The Visual Analytics Workflow*

We now go through each of the runs in detail.

1) FIRST RUN

In the first run, we revisit and go through the analysis from Assignment 1 in more detail with respect to life expectancy. The same can be found under Task 2 - Hypothesis 2 in the A1 assignment.

DATA: We used the World Bank's World Development Indicators dataset. Additionally, we categorized countries by continent to analyze life expectancies for each continent. The average life expectancy for all countries was calculated for the years 2000 to 2015.

Visualization:
We first plot the heatmap of life expectancies across the countries for the years 2000-2015. A sequential colormap is used to show the trend of life expectancy. The range of life expectancy is from 40 years to 85 years. This shows the life expectancy of all the countries
.
This can be seen in **Fig. 18**.

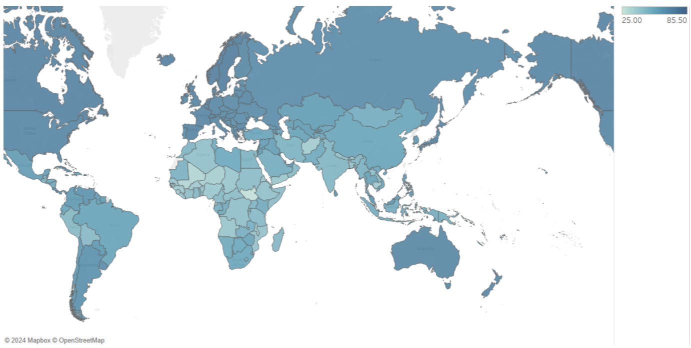


Fig. 18 Heatmap of Life Expectancy Across Countries (2000-2015)

We created a box plot to show the range of life expectancies for each continent: Asia, Europe, Africa, Oceania, North America, and South America. The plot uses boxes to represent the middle 50% of the data (the interquartile range), with a line inside the box marking the median life expectancy for each continent. The whiskers extend to show the range of most of the data, while any points outside this range are marked as outliers. This helps us see the spread, the central values, and any unusual life expectancy data for each continent. This can be seen in **Fig. 19**.

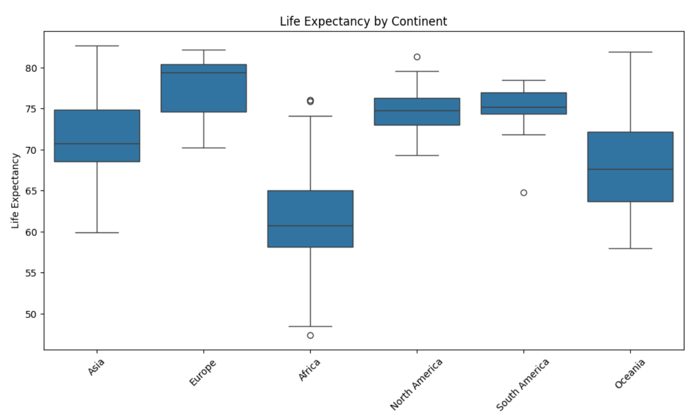


Fig. 19 Box Plot showing variations in Life Expectencies in different continents

The horizontal bar chart displays the top 10 countries with the highest life expectancies and the 10 countries with the lowest life expectancies worldwide. Each bar represents a country, with the length of the bar indicating its life expectancy.
The bars are color-coded to show the continent each country belongs to, making it easy to identify regional patterns.
The chart allows for a visual comparison between countries with the highest and lowest life expectancies, highlighting the disparities in global health outcomes. Can be seen in **FIG 20**

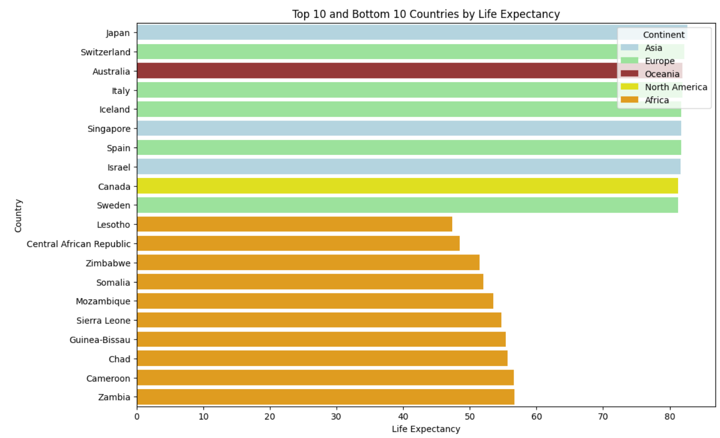


Fig. 20 Bar chart of Life Expectancy Across Countries

Models: We would not be using any models in this round.

Knowledge:
From the heatmap, we can clearly see the trend in life expectancy across different countries. European countries generally have higher life expectancies compared to most African and Asian countries. The countries in Europe are shown in darker shades, indicating higher life expectancies, while many African and Asian countries are in lighter shades, reflecting lower life expectancies

The box plot also shows that African countries have the lowest life expectancies. The highest life expectancy in Africa is almost the same as the lowest life expectancy in other continents. This clearly highlights the challenges faced by many countries in Africa in terms of health and life expectancy.

In the horizontal bar chart, interestingly, we can see that the countries with the lowest life expectancies are all in Africa. There are no African countries in the top 10 for the highest life expectancies. This shows a clear gap between Africa and other regions in terms of life expectancy, with African countries at the bottom and no representation in the higher ranks.

Feedback:
The visualizations above clearly show that African countries have much lower life expectancies compared to the rest of the world. Moving forward, we would like to analyze the life expectancies of countries specific to Africa in detail.

2) SECOND RUN

In the second run , we will be focusing on the life expectencies in Africa

DATA: We used the World Bank's World Development Indicators dataset for the life expectancies of countries. We used gender-wise life expectancy distribution dataset. We calculated life expectancies of different regions of africa such as northern africa , southern africa , middle , eastern and western africa by taking average of life expectancies of all the countries of respective regions.

Visualization:

We first plot the heatmap of life expectancies across the countries for the years 2005-2015 inside Africa. A sequential colormap is used to show the trend of life expectancy. The range of life expectancy is from 46 years to 75 years. This can be seen in **Fig. 21**.

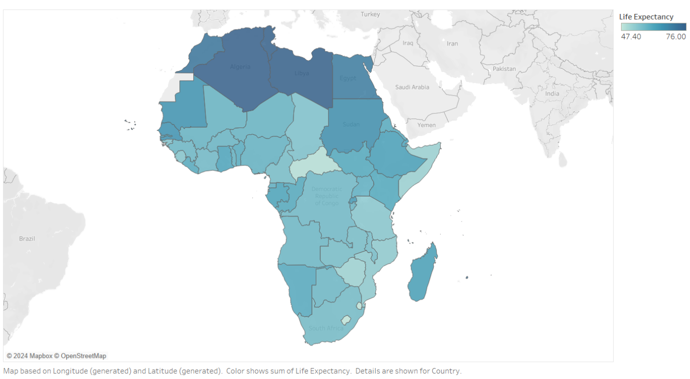


Fig. 21 Heatmap of Life Expectancy Across Countries in Africa (2000-2015)

.

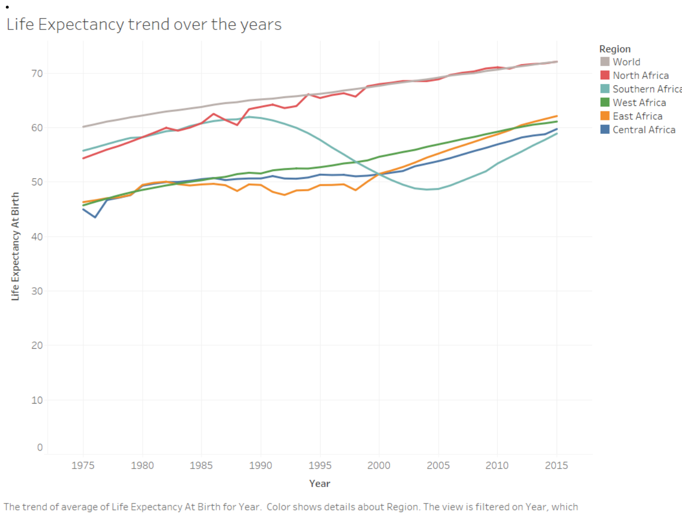


Fig. 22 Line chart showing life expectancy trend of different African regions

We created a line chart to display the life expectancy trends of five different regions in Africa from 1975 to 2015. This is show in **FIG 22** .The chart shows the life expectancy data for each region, with lines representing each one. Additionally, we included the global average life expectancy across the same years for comparison. . The lines are color-coded to differentiate between the regions, making it easier to track the trends for each one.

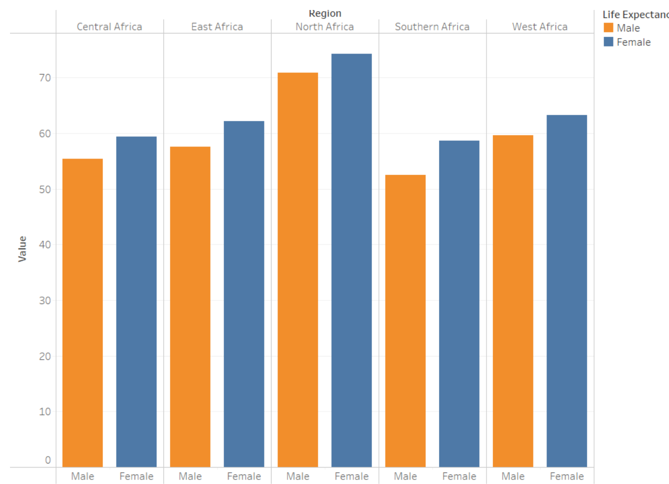


Fig. 23 Double Bar chart showing gender-wise life expectancy

We created a double bar chart (**FIG 23**) to compare life expectancy between males and females across five different regions of Africa. In this chart, the blue bars represent life expectancy for females, and the yellow bars represent life expectancy for males. The y-axis shows the range of life expectancy values, while the x-axis represents the five regions of Africa. For each region, there are two bars—one for females and one for males.

Models: We would not be using any models in this round.

Knowledge:

From the heatmap, we can clearly see the trend in life expectancy across different countries in Africa. We can see that life expectancies vary widely across different countries , while many countries have relatively low life expectancy, there are a few exceptions where life expectancy is much higher. These countries at the top of the map have life expectancies that are significantly better compared to those at the bottom, The line chart supports the previous point. It shows that Northern African countries have higher life expectancies compared to the rest of Africa. On the other hand, other regions have much lower life expectancies. We also see a sharp drop in life expectancy in Southern Africa around 2000-2005. This decline in Southern Africa is particularly significant, as it stands out compared to the relatively stable trends in other regions. Map of Africa showing it's main regions is shown in **FIG 24**

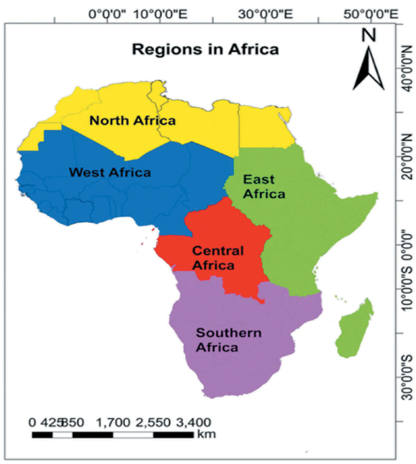


Fig. 24 Five Regions of Africa

The double bar chart highlights a similar pattern across different regions, where females generally have a lower life expectancy compared to Northern Africa and the world average. This trend is also observed for males in most regions. However, it is noticeable that, overall, females tend to have a higher life expectancy than males. Southern Africa, in particular, has the lowest life expectancy for both males and females, showing a significant gap compared to other regions.

Feedback:
Looking at all the visualizations from this round, one thing is clear: while all regions of Africa have lower life expectancy, Northern Africa has the highest, and the rest of Africa has much lower life expectancy. Specifically, Southern Africa has the lowest life expectancy. We now want to focus on understanding the factors behind this, especially why Southern Africa has such a low life expectancy.

3) THIRD RUN
In third run, we will focus on Southern Africa and the factors contributing to its lower life expectancy.

DATA: We used the World Bank's World Development Indicators dataset for the life expectancies , health expenditure percentage , intentional homicides of African countries. We calculated rural population percentage for African countries from the dataset. We used the HIV/AIDS deaths country wise dataset by WHO. We filtered out African countries in HIV/AIDS deaths dataset , which we used in our model.

Visualization :
We first tried to find the correlation between life expectancy and other factors that is mentioned in the data section using correlation matrix. We have used a diverging colormap. Dark red represents a strong positive correlation and dark blue represents strong negative correlation. White represents no correlation, We have considered African countries for this matrix (**FIG 25**). We have considered 3 countries from each region of the country (Regions of the countries that was discussed in the previous round) over the years 2000-2015.

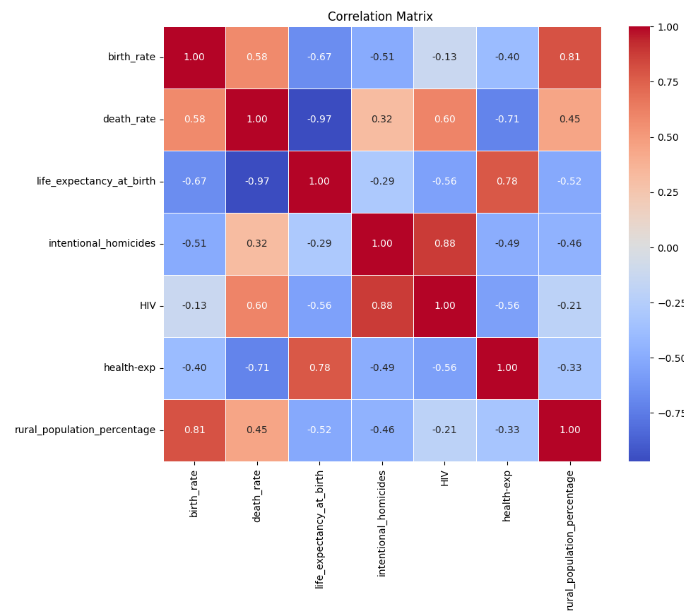


Fig. 25 Correlation matrix between Life expectancy and other factors

Knowledge :
From the correlation matrix above, the relationship between various factors and life expectancy has been determined. These factors include intentional homicides, HIV deaths, health expenditure, and rural population percentage. The correlations are summarized in the table below.

TABLE I
CORRELATION BETWEEN FACTORS AND LIFE EXPECTANCY

Factor	Correlation
Intentional homicides	-0.29
HIV Deaths	-0.56
Health expenditure	0.78
Rural population percentage	-0.52

There is a strong negative correlation between Life expectancy and Number of deaths due to HIV/AIDS. This is also true. This is because around 1990-2000 , African countries were hit very hard by HIV/AIDS and millions of people lost their lives. Poor health and poor health facilities are one among the major factors for lesser life expectancy.

Model:
We have used K-means clustering . We have taken data of deaths due to HIV/AIDS in African countries.

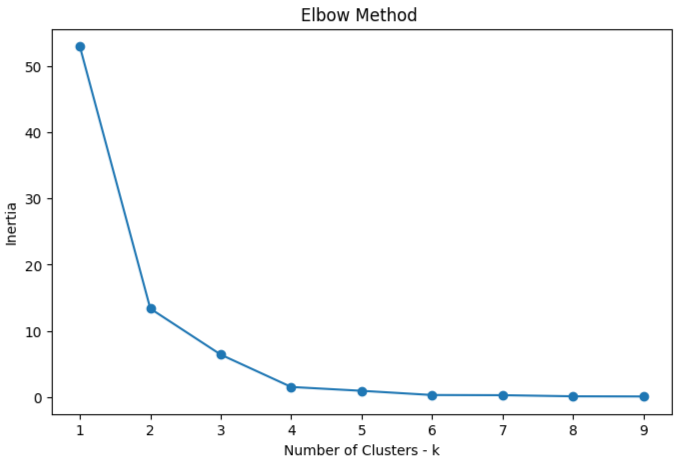


Fig. 26 Elbow diagram to find optimal K

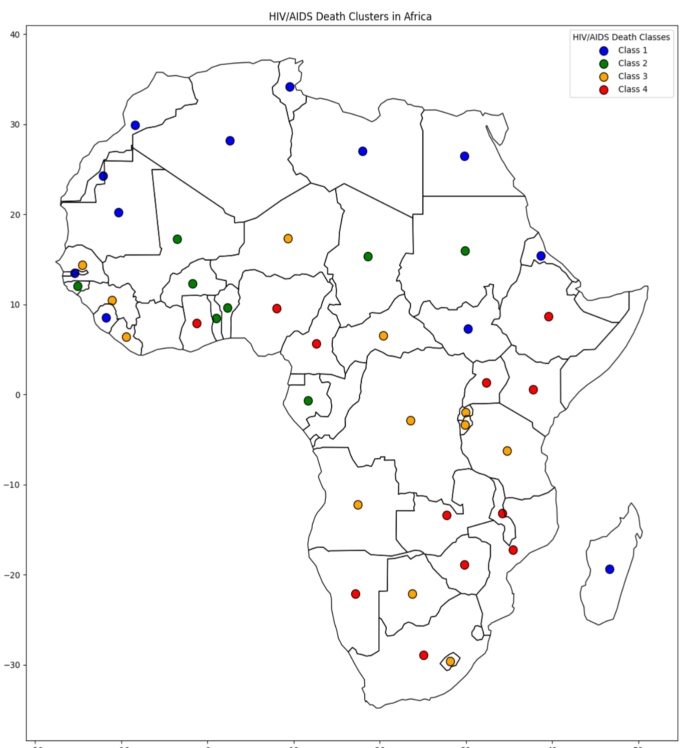


Fig. 27 Output of the model. 4 Clusters being formed

We used K-means clustering from the sklearn library to group African countries into clusters based on the total number of deaths due to HIV/AIDS. In this approach, countries with similar numbers of deaths are grouped together, allowing us to identify patterns and categorize regions with comparable impacts. To find the optimal number of clusters, we applied the elbow method as shown in **FIG 26**, which suggested **k = 4**. This resulted in four distinct groups of countries, each reflecting a unique range of HIV/AIDS-related deaths.
We can look at the output of the model in **Fig. 27**. We can see that the model has grouped countries into clusters, so the countries in a cluster have around the same number of deaths. We can clearly see that the countries with lesser life expectancy (as per previous run knowledge) are clustered into the same group, which has a high number of cases. We will use this inference to improve our understanding using further visualizations.

Feedback:
We now understand that HIV/AIDS is one of the major reason for the drop in life expectancy in African countries , especially Southern African countries. In Assignment 1, we studied other factors like rural population percentage and health spending in detail. In next run, we want to look more closely at HIV/AIDS trends in Southern Africa.

4) FOURTH RUN
In this run , we will try to analyze the deaths due to HIV/AIDS in Southern Africa in detail.

DATA: We used the HIV/AIDS deaths by country dataset. We calculated the HIV/AIDS deaths percentage in few countries out of the total worlds' HIV/AIDS deaths.

Visualisations:
The model showed us that the Southern African countries have higher HIV/AIDS deaths. We will try to refine our visualizations using this output.

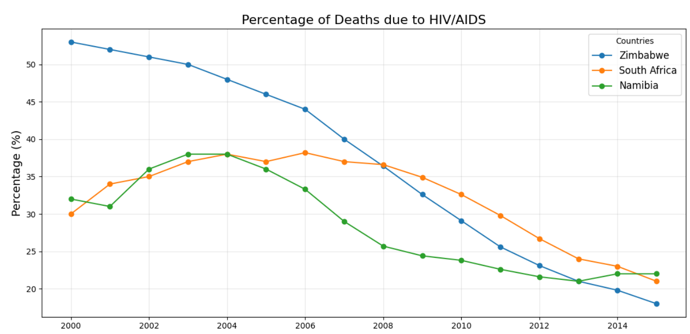


Fig. 28 Line graph showing percentage deaths due to HIV/AIDS

The model showed us that the Southern African countries have higher HIV/AIDS deaths. We will try to refine our visualizations using this output. We plotted a line graph (**FIG 28**) to display the percentage of deaths caused by HIV/AIDS in three specific Southern African countries: Namibia, Zimbabwe, and South Africa. The graph covers the time period from 2000 to 2015, with the x-axis representing the years and the y-axis showing the percentage of deaths. Each country is represented by a distinct colored line to make comparisons. .

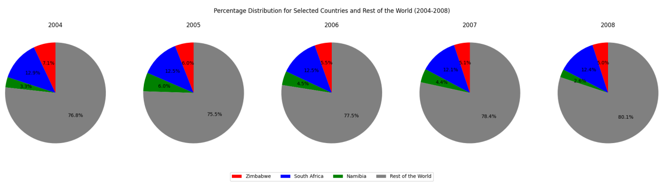


Fig. 29. Percentage HIV/AIDs deaths

We created a set of pie charts (**FIG 29**) to show the distribution percentages of HIV/AIDS deaths worldwide from 2004 to 2009. Each pie chart represents one year, and it compares the number of deaths in three Southern African countries with the rest of the world. The charts visually break down the percentage of deaths in these countries versus other regions.

Knowledge:
The line chart shows that the major source of death in Southern African countries like South Africa , Zimbabwe, Namibia was HIV/AIDS in the timeline shown in graph. More than 50 percent of the deaths in Zimbabwe was because of HIV/AIDS. We can clearly infer the impact of HIV/AIDS on life expectancy of these countries.

This pie chart shows that nearly 25 percent of the world's HIV/AIDS deaths came from just three Southern African countries—Namibia, Zimbabwe, and South Africa—between 2004 and 2009. This shows how serious the HIV/AIDS problem is in this region, where the disease has caused a major health crisis. These countries account for a large portion of the world's deaths, pointing to issues like limited healthcare, poverty, and a lack of education on how to prevent the disease. The chart makes it clear that these nations have been at the center of the HIV/AIDS epidemic, contributing a significant number of deaths during these years

Conclusion:
In this workflow, we analyzed the life expectancy trends of African countries, identifying key trends and factors that influence life expectancy. By examining data over time, we were able to highlight the main reasons behind the differences in life expectancy across the continent.

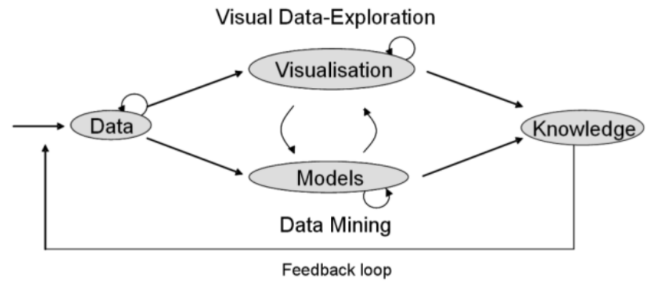


Figure 18. Kiem et al. Visual Analytics Workflow, Image Courtesy: [1]

IV. Visual Analytics Workflow

V. Task 3: An Advanced Analysis on Greenhouse Gas Emissions

A. *Visual Analytics Workflow*

In this workflow, we focus on analyzing different greenhouse gas emissions in detail, a key environmental indicator. Our goal is to gain deeper insights into the emissions, such as causes, sources, sectors, distribution across regions, and patterns. In Assignment 1, we explored various environmental indicators and examined how they impact countries or groups of countries based on income levels. This workflow builds on those insights, focusing specifically on greenhouse gas emissions.

B. *Datasets Used in the Workflow*

We used the World Bank World Development Indicators dataset, previously used in Assignment 1. Additionally, we included datasets detailing state-wise and sector-wise distributions of greenhouse gas emissions and related factors, particularly for India. These additional datasets allow for a more in-depth analysis.

C. *Describing the Visual Analytics Workflow*

We detail each analysis run below:
First Run
Data :
Visualisations : Refer to A1 report [2] for the visualizations. Figure numbers are as per A1.

- **Fig 27**. Double bar chart representing the per capita CO_2 emissions for individuals across different income. groups (1990 and 2007)
- **Fig 29 and 30** Maps representing global trends for CO_2 emissions and overall greenhouse gas emissions across countries for the year 2010.
- **Fig 31**. Top 10 largest contributing countries to CO_2 emissions (2010).
- **Fig 32 and 33** Show the share of global greenhouse gas emissions as by the top 10 CO_2 emitting countries compared to the rest of the world (2010).

Models : No models were used when going through the workflow for the first run. Simple statistical measures such as mean and total sum along with historical and noticable events were used to choose which countries to focus on.

Knowledge : Refer to A1 report [2] for the visualizations. Figure numbers are as per A1.

- **Fig 27 :** Here, the chart highlights a significant disparity in CO_2 emissions per person among different income groups worldwide. It illustrates that individuals in high-income countries utilize significantly more resources, leading to much higher carbon footprints compared to those in low-income countries. The graph clearly shows that people in high-income countries emit 12 times more CO_2 per person than those in low-income countries, emphasizing the stark contrast in resource usage and environmental impact between these groups.
- **Fig 28, 29, 31 :** This visual representation highlights the dominant contributors to global carbon dioxide emissions, providing a clear comparison among the top emitters. The symbol map depicting CO_2 emissions across countries for the year 2010 and the bar chart showing the top 10 largest contributing countries to CO_2 emissions provide a com- plementary view of global emission patterns. The top 4 darkest countries in the map which are also represented by top 4 longest bars in the horizontal bar chart in their decreasing order of emission for the year 2010 are : China, United States of America, India and Russian Federation.
- **Fig 32 and 33:** It can be concluded that the 10 largest contributors to CO_2 emissions account for approximately 70% of the global total, underscoring the significant concentration of emissions within a relatively small number of countries. This disproportionate share of global emissions highlights the crucial role these nations play in shaping the trajectory of global climate change. Also, it can be seen that the same top 10 CO_2 emitting countries are also responsible for about 64% of total greenhouse gas emissions of the world.

Feedback :

- **Focus on one of the high emitting countries – *India* in this case.** India was ranked 3 rd in 2010 according to global CO_2 emissions and with total CO_2 emissions close to 1650 kilo tonnes of CO_2 equivalent for that particular year.
- **Sector-wise and state-wise :** See the distribution across different states and sectors to understand the patterns and main sources responsible for emissions.
- **Different gas-wise distributions :** Study the percentage share of different kinds of greenhouse gases as a part of total emissions.

1) Second Run: **Data:** The GHG Emissions Estimates Economy-Wide (2005-2015) for India dataset was used.
Visualization:

- A bar chart of the top 10 CO2-emitting states in India for 2010 (Fig. 30).
- A pie chart showing the composition of greenhouse gases (Fig. 31).
- A pie chart displaying the distribution of emissions by categories (Fig. 32).

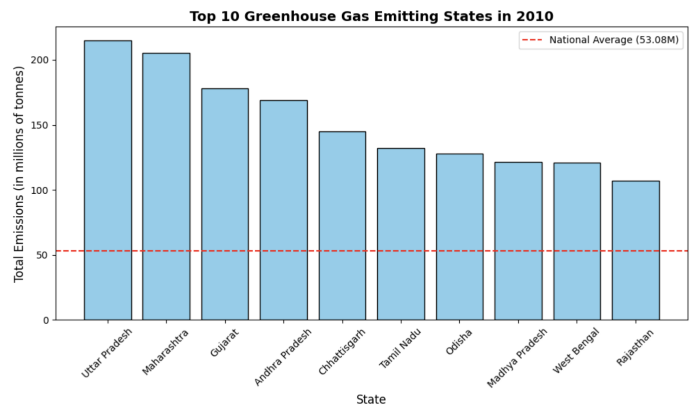


Figure 30. Top 10 CO2 Emitting States in India (2010)

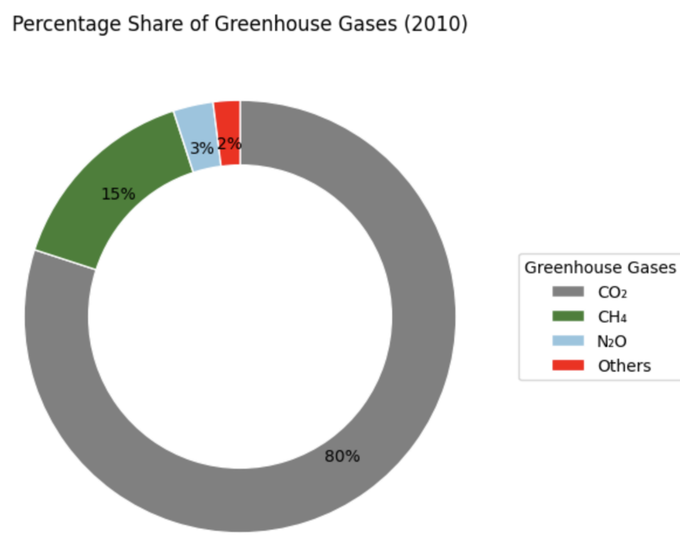


Figure 31. Composition of Greenhouse Gases in Emissions

Knowledge:

- The bar chart identifies the states contributing the most to CO2 emissions, with industrialized states topping the list.
- The gas composition pie chart highlights which greenhouse gases dominate emissions.
- The category pie chart provides clarity on sectoral contributions to emissions.

Feedback: Understand and dive deeper into each of the categories to understand trend over the years from 2005 to 2015 and also particularly for the year 2010.

2) Third Run: Here, we see the sector wise and sub sector wise distributions in detail.

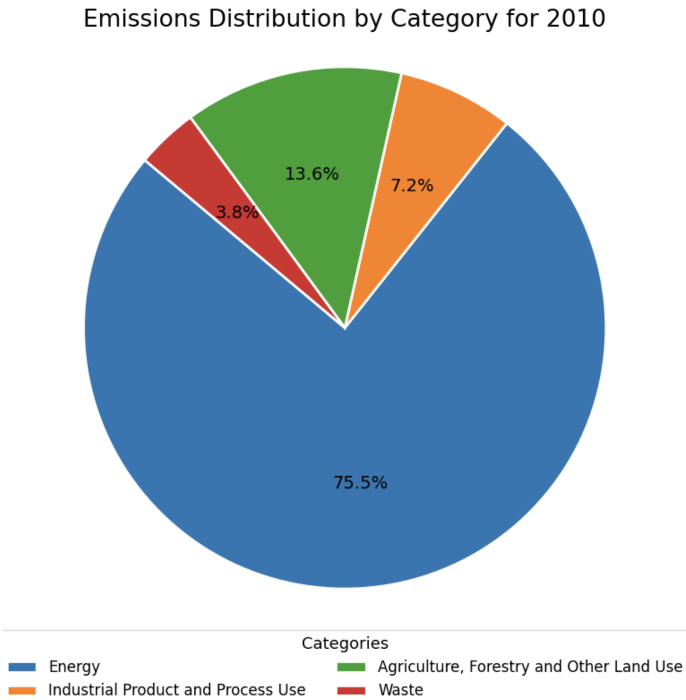


Figure 32. Emission Distribution by Categories in 2010

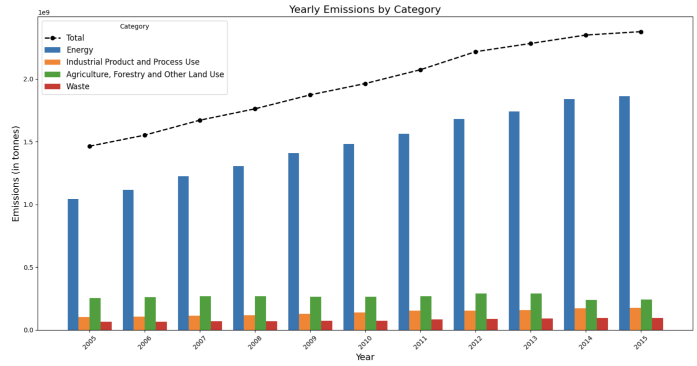


Figure 33. Top 10 CO2 Emitting States in India (2010)

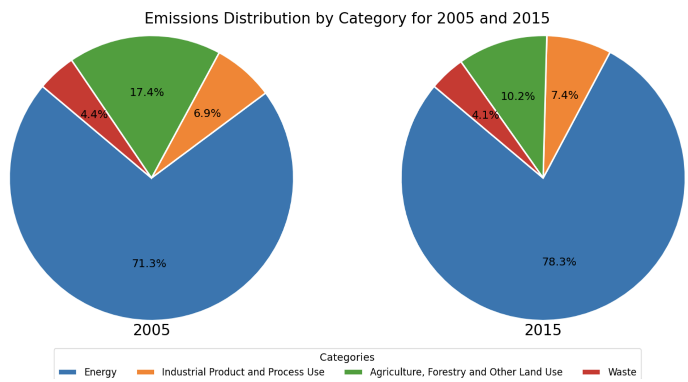


Figure 34. Comparison across categories for 2005 and 2015

Data: The GHG Emissions Estimates Economy-Wide (2005-2015) for India dataset was used.
Visualization:
(1) A bar chart showing the yearly GHG Emissions Estimates of India (2005 to 2018) across different categories. (Fig. 33).

The various categories of greenhouse gas (GHG) emissions include:

- **Energy:** Covers emissions from the production and consumption of energy.
- **Agriculture, Forestry, and Other Land Use (AFOLU):** Refers to emissions from agriculture, forestry, and other land uses.
- **Industrial Processes and Product Use (IPPU):** Includes emissions from industrial processes and product use.
- **Waste:** Encompasses emissions from the treatment and disposal of waste materials.

.

(2) Two pie charts showing the comparison of emissions across different categories for the years 2005 and 2015. (Fig. 34).

Since, we got an idea of the emissions across different categories, now we dive into each of the categories.
(3) The clustered bar chart below represents the breakdown of emissions in the **Energy Category** from 2005 to 2015. The bars display the **Fuel Combustion Emissions** and **Fugitive Emissions** subcategories, while the line shows the **Total Emissions**.

- **Fuel Combustion Emissions**: These emissions result from the burning of fuels during energy production and consumption.
- **Fugitive Emissions**: These are emissions that occur from the intentional or unintentional release of gases, primarily methane, during energy production processes, especially in the oil and gas sector.
- **Total Emissions**: The line shows the cumulative emissions from both fuel combustion and fugitive emissions.

This chart visually represents how these subcategories of energy-related emissions have compared and evolved over time, from 2005 to 2015. (Fig. 35).

(4) The stacked bar graph below represents the emissions in various subcategories (Level 3) of the **Energy Category**. The length of each bar corresponds to the value of emissions for each subcategory, while the bar colors are indicative of the different subcategories (according to Level 3). (Fig. 36).

(5) The clustered bar chart below represents the breakdown of emissions in the **IPPU Category** from 2005 to 2015. The bars display the emissions from various

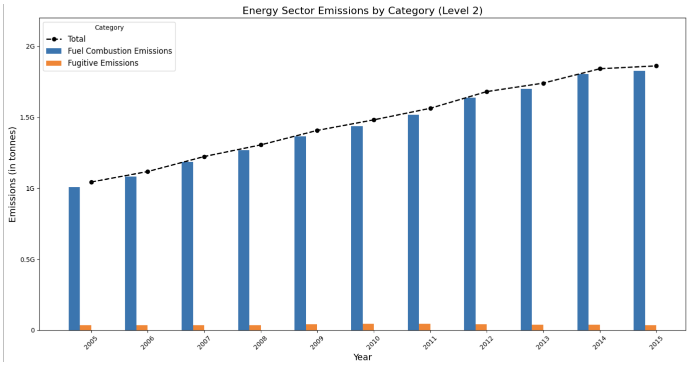


Figure 35. Emission Distribution in the Energy Category (Level 2) from 2005 to 2015

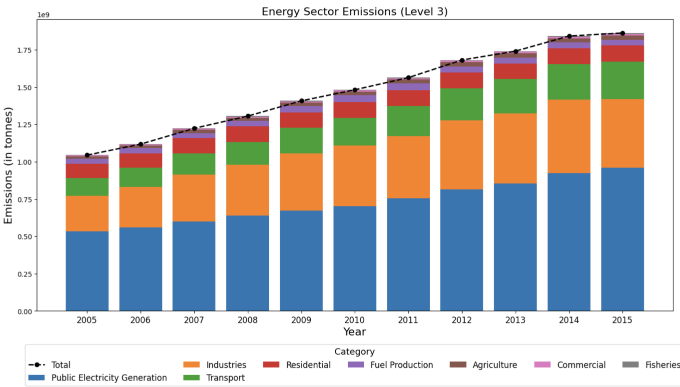


Figure 36. Emission Distribution in the Energy Category (Level 3) from 2005 to 2015

subcategories, such as **Chemical Industry**, **Metal Industry**, **Mineral Industry**, and **Non-Energy Products from Fuels and Solvent Use**, while the line shows the **Total Emissions**.

- **Chemical Industry**: Emissions resulting from chemical manufacturing processes, including the production of fertilizers, plastics, and chemicals.
- **Metal Industry**: Emissions from the production and processing of metals, including steel, aluminum, and other metals.
- **Mineral Industry**: Emissions originating from the production of minerals such as cement, glass, and ceramics.
- **Non-Energy Products from Fuels and Solvent Use**: Emissions related to the use of fuels and solvents in industrial applications, not related to energy production.
- **Total Emissions**: The line represents the cumulative emissions from all the subcategories within the IPPU sector. (Fig. 37).

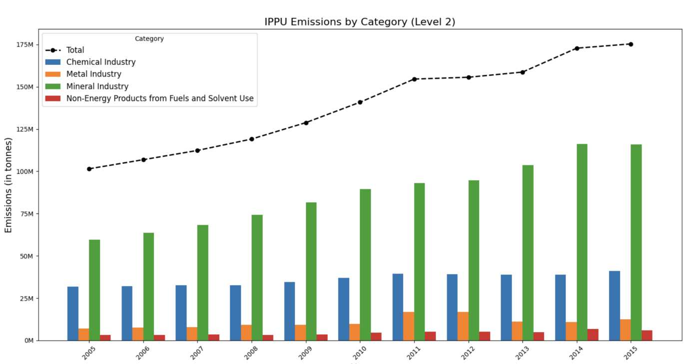


Figure 37. Emission Distribution in the IPPU Category (Level 2) from 2005 to 2015

(6) The stacked bar graph below represents the emissions of top 3 sub categories (Level 3) of the **IPPU Category**. The length of each bar corresponds to the value of emissions for each subcategory, while the bar colors are indicative of the different subcategories (according to Level 3). (Fig. 38).

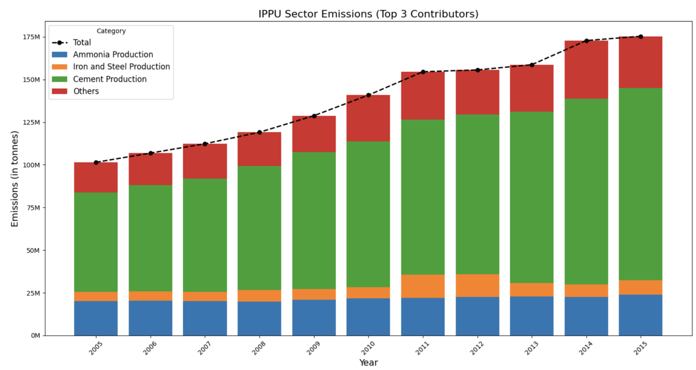


Figure 38. Top 3 Contributors in Subcategories of the IPPU Category (Level 3) from 2005 to 2015

We now examine the Level 2 categories, such as the Chemical Industry, Metal Industry, and Mineral Industry, to determine which of these categories the top three productions, as mentioned in **Fig. 38**, belong to.

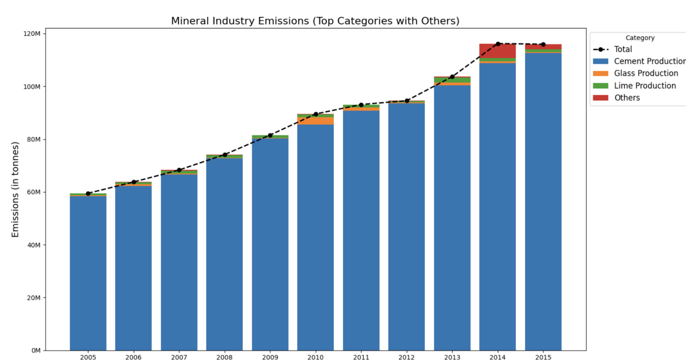


Figure 39. Top 3 Contributors in Mineral Industry (IPPU SubCategory) from 2005 to 2015

(7) The stacked bar graph below represents the emissions of top 3 sub categories (Level 3) in the Mineral Industry of the IPPU Category. The length of each bar corresponds to the value of emissions for each subcategory, while the bar colors are indicative of the different subcategories (according to Level 3). (Fig. 39).

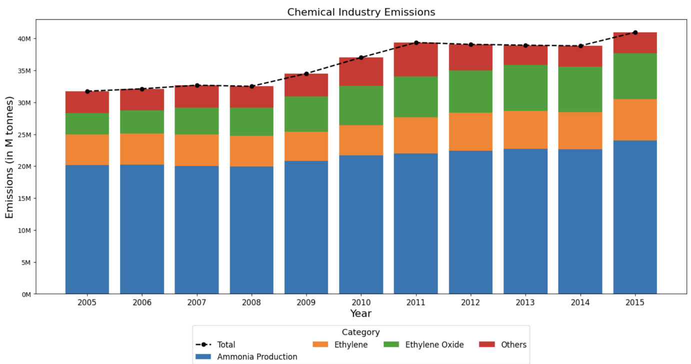


Figure 40. Top 3 Contributors in Chemical Industry (IPPU SubCategory) from 2005 to 2015

(8) The stacked bar graph below represents the emissions of top 3 sub categories (Level 3) in the Chemical Industry of the IPPU Category. The length of each bar corresponds to the value of emissions for each subcategory, while the bar colors are indicative of the different subcategories (according to Level 3). (Fig. 40).

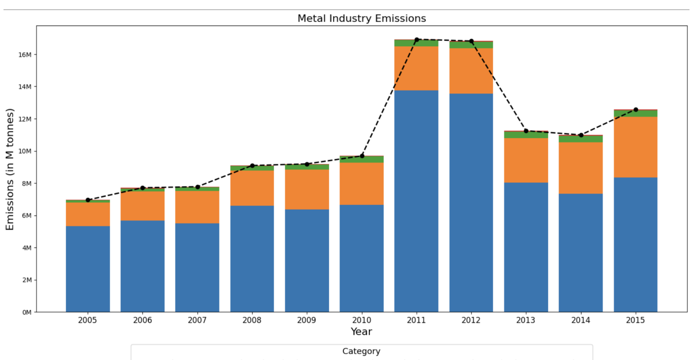


Figure 41. Top 3 Contributors in Metal Industry (IPPU SubCategory) from 2005 to 2015

(9) The stacked bar graph below represents the emissions of top 3 sub categories (Level 3) in the Metal Industry of the IPPU Category. The length of each bar corresponds to the value of emissions for each subcategory, while the bar colors are indicative of the different subcategories (according to Level 3). (Fig. 41).

(10) The clustered bar chart below represents the breakdown of emissions in the **AFOLU Category** from 2005 to 2015. The bars display the emissions from various subcategories, such as **Livestock**, **Land**, and **Aggregate**

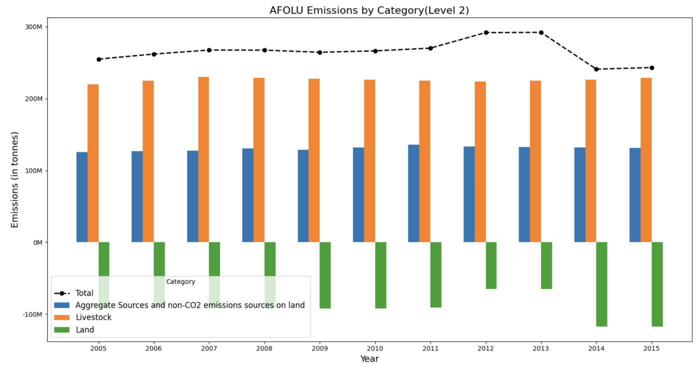


Figure 42. Emission Distribution in the AFOLU Category (Level 2) from 2005 to 2015

Sources and Non-CO2 Emissions Sources on Land, while the line shows the **Total Emissions**.

- **Livestock**: Emissions resulting from agricultural activities related to the raising of livestock, including enteric fermentation and manure management.
- **Land**: Emissions associated with land use changes, such as deforestation, land degradation, and the conversion of natural habitats.
- **Aggregate Sources and Non-CO2 Emissions Sources on Land**: Emissions from various non-CO2 greenhouse gases and other sources that occur on land, such as methane and nitrous oxide from agricultural soils.
- **Total Emissions**: The line represents the cumulative emissions from all the subcategories within the AFOLU sector. (Fig. 42).

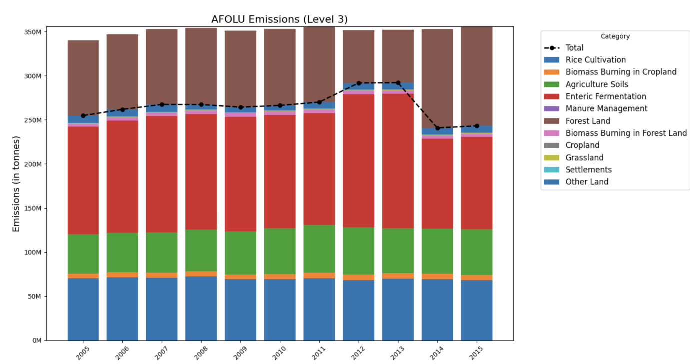


Figure 43. Top 3 Contributors in Subcategories of the AFOLU Category (Level 3) from 2005 to 2015

(11) The stacked bar graph below represents the emissions in various subcategories (Level 3) of the AFOLU Category. The length of each bar corresponds to the value of emissions for each subcategory, while the bar colors are indicative of the different subcategories (according to Level 3). (Fig. 43)

(12) The donut chart below illustrates the percentage-wise distribution of emissions within the AFOLU Category (Level 3) for the year 2010. The chart highlights the

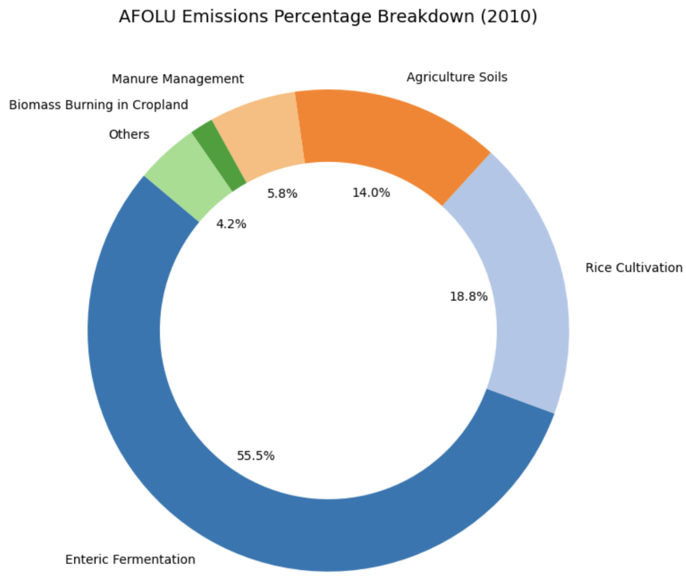


Figure 44. AFOLU sector emissions breakdown for 2010

relative contributions of various subcategories, providing insights into the distribution of emissions from different sources within the AFOLU sector during that year (Fig. 44)

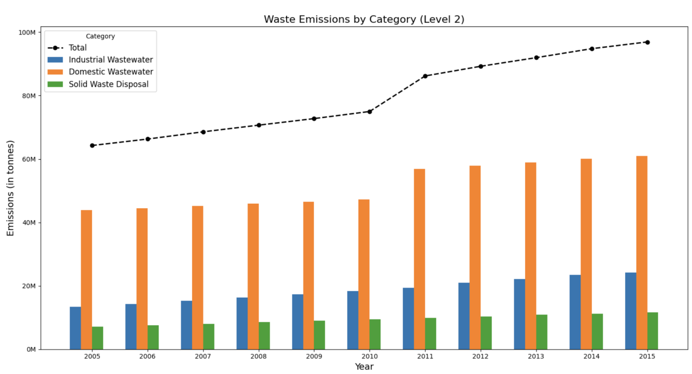


Figure 45. Emission Distribution in the Waste Category (Level 2) from 2005 to 2015

(13) The clustered bar chart below represents the breakdown of emissions in the **Waste Category** from 2005 to 2015. The bars display the emissions from various subcategories, such as **Industrial Wastewater**, **Domestic Wastewater**, and **Solid Waste Disposal**, while the line shows the **Total Emissions**.

- **Industrial Wastewater**: Emissions resulting from the treatment and disposal of wastewater generated by industrial processes.
- **Domestic Wastewater**: Emissions from the treatment and disposal of wastewater from domestic sources, such as households and municipalities.
- **Solid Waste Disposal**: Emissions from the management and disposal of solid waste, including landfills and waste incineration.
- **Total Emissions**: The line represents the cumulative emissions from all the subcategories within the Waste sector.(Fig. 45)

(14) The donut chart below illustrates the percentage-wise distribution of emissions within the **Waste Category** (Level 3) for the year 2010. The chart highlights the relative contributions of various subcategories, providing insights into the distribution of emissions from different sources within the Waste sector during that year. (Fig. 46)

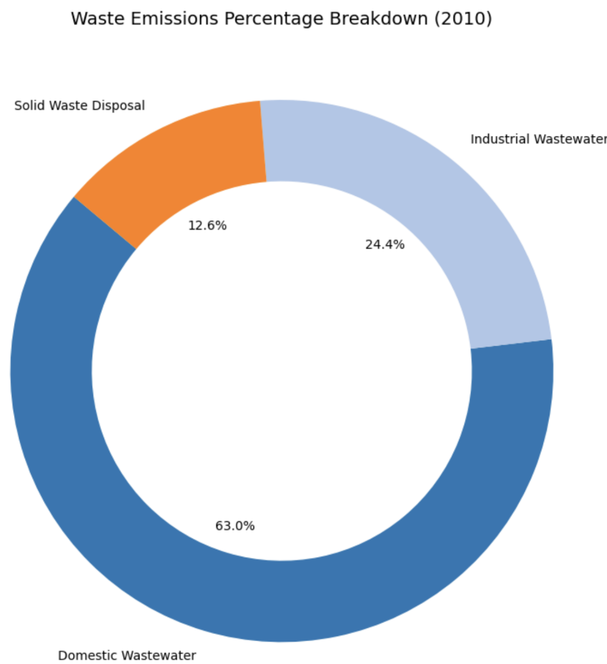


Figure 46. Waste Sector emissions breakdown for 2010

Knowledge:

- (1) tells that most of the emissions are because of the energy category, followed by AFOLU category, followed by IPPU category and then waste category.
- (2) Shows that the overall trend in percentages remains almost the same but emissions due to AFOLU and waste categories decreased in the year 2015 as compared to 2005.
- (3) Shows that in the energy category, most of the emissions are due to fuel combustion and only a few are due to fugitive emissions.
- (4) Shows that in the energy sub category (level 3), most of the emissions are due to public electricity generation, industries, transport, residential and fuel production.
- (5) Shows that in the IPPU category, most of the emissions are due to mineral industry, then chemical industry, followed by metal industry and then non-energy products.

– (6) Shows that in the IPPU category (level 3), the top 3 contributors to emissions are cement production, ammonia production and iron and steel production.
– (7), (8), (9) make it evident that the three contributors as seen in the point (6) account for the highest factors in three different industries (i.e sub categories (level 2)). So cement production is the major contributor in Mineral Industry as seen in figure of point (7), ammonia production is the major contributor in Chemical Industry as seen in figure of point (8) and iron and steel production is the major contributor in Metal Industry as seen in figure of point (9).
– (10) Show that for the AFOLU category, livestock is responsible for major emissions followed by other aggregate sources on land.
– (12) Shows that for the year 2010 the waste category emissions comprised of enteric fermentation (about 55 percent), rice cultivation (about 19 percent), agricultural soils (about 14 percent).
– (13) Shows that in the waste category, domestic wastewater tops the emission game with industrial wastewater and solid waste disposal following it respectively.
– (14) Shows that for the year 2010 the waste category emissions comprised of domestic wastewater (about 63 percent), industrial wastewate (about 24 percent), solid waste disposal(about 13 percent).

VI. AUTHORS' CONTRIBUTIONS

Ayush Arya Kashyap: (TASK 1 : AN ADVANCED ANALYSIS ON THE GDP OF THE COUNTRIES AND THEIR SHIFTS)

– Conducted data analysis using Jupyter Notebook, Python, and libraries like pandas, datetime, os, among others to identify trends in economic activities in and around the world over the years especially in the GDP.
– Visualizations Used : bar chart, stacked bar chart, PCP, SPLOM, radar chart, pie chart, treemaps, line chart, and bubble diagram to represent the data effectively.
– used 3 runs to to do the analysis
– shown the visual analytics workflow used in the three runs
– Interpreted the visualizations and predicted values using machine learning models to draw conclusions about the overall trends in economic activities.
– Wrote the narrative and descriptions for Task 1, providing context and insights.

Uttam Hamsaraj: (TASK 2: AN ADVANCED ANALYSIS ON LIFE EXPECTENCY)

– Preprocessed the data using pandas. Merged data from 3 different datasets and played around with it.
– Plotted few of the visualizations with tableau , and most of them using pandas.
– Visualizations Used : Box Plot , Dual Bar chart , Horizontal Bar chart , Heat map, Correlation Matrix , Line plot , pie chart .
– ML model used is K-means clustering. Used k means clustering using sklearn.

Pranav Laddhad: (TASK 3: AN ADVANCED ANALYSIS ON GREENHOUSE GAS EMISSIONS)

– Preprocessed and analyzed data on environmental indicators using python libraries and tabeau prep.
– Analyzed and visualized 3 hypotheses with inferences.
– Visualizations Used : lines charts, dual combinations chart, double bar charts, horizontal bar chart and pie charts to represent the data effectively.
– Interpreted the visualizations and drew conclusions
– Wrote the narrative and descriptions for Task 3, providing context and insights.

VII. REFERENCES

1) Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Gorg, Jorn Kohlhammer, et al.. Visual Analytics: Definition, Process and Challenges. Andreas Kerren and John T. Stasko and Jean-Daniel Fekete and Chris North. Information Visualization - Human-Centered Issues and Perspectives, 4950, Springer, pp.154-175, 2008, LNCS. fflirmm-00272779
2) https://github.com/ayusharyakashyap/world-bank-development-indicators-data-visualisation-project
3) https://www.weforum.org/stories/2019/09/fifteen-countries-represent-three-quarters-total-gdp/
4) https://databank.worldbank.org
5) https://documents.worldbank.org/en/publication/documents-reports
6) https://apps.who.int/gho/data/view.main.SDG2016LEXREGv?lang=en

World Bank World Development Indicators

Ayush Arya Kashyap
IMT2022129
IIIT - Bangalore
Bangalore, India
ayusharya.kashyap@iiitb.ac.in

Uttam Hamsaraj
IMT2022524
IIIT - Bangalore
Bangalore, India
uttam.hamsaraj@iiitb.ac.in

Pranav Laddhad
IMT2022074
IIIT - Bangalore
Bangalore, India
pranav.laddhad@iiitb.ac.in

I. Introduction

In this document, we have explored the economic, social and sustainability performance of various countries by analyzing and visualizing data from the "World Bank World Development Indicators" dataset. The given dataset has data from 1960 to 2022.

Columns given in the dataset :

- **country**: The country or geographic region.
- **date**: Date of the measurement. This column along with country can be used as index.
- **agricultural_land%**: Agricultural land as a % of land area of the country/region.
- **forest_land%**: Forest area as the % of land area of the country/region.
- **land_area**: Land area, measured in km^2.
- **avg_precipitation**: Average precipitation in depth, measured in mm per year.
- **trade_in_services%**: Trade in services as a % of GDP.
- **control_of_corruption_estimate**: Index that makes an estimate of the control of corruption.
- **control_of_corruption_std**: Standard error of the estimate of control of corruption.
- **access_to_electricity%**: Percentage of the population that has access to electricity.
- **renewvable_energy_consumption%**: Renewable energy consumption as a % of total final energy consumption.
- **electric_power_consumption**: Electric power consumption, measured in kWh per capita.
- **CO2_emisions**: CO2 emissions measured in kt.
- **other_greenhouse_emisions**: Total greenhouse gas emissions, measured in kt of CO2 equivalent.
- **population_density**: Population density, measured in people per km^2 of land area.
- **inflation_annual%**: Inflation, consumer prices, as annual %.
- **real_interest_rate**: Real interest rate (%).
- **risk_premium_on_lending**: Risk premium on lending (lending rate minus treasury bill rate, %).
- **research_and_development_expenditure%**: Research and development expenditure, as a percentage of GDP.
- **central_goverment_debt%**: Central government debt, total, as a % of GDP.
- **tax_revenue%**: Tax revenue as a % of GDP.
- **expense%**: Expense as a % of GDP.
- **goverment_effectiveness_estimate**: Index that makes an estimate of Government Effectiveness.
- **goverment_effectiveness_std**: Standard error of the estimate of Government Effectiveness.
- **human_capital_index**: Human Capital Index (HCI) (scale 0-1).
- **doing_business**: Ease of doing business score (0 = lowest performance to 100 = best performance).
- **time_to_get_operation_license**: Days required to obtain an operating license.
- **statistical_performance_indicators**: Statistical performance indicators (SPI): Overall score (scale 0-100).
- **individuals_using_internet%**: Percentage of population using the internet.
- **logistic_performance_index**: Logistics performance index: Overall (1=low to 5=high).
- **military_expenditure%**: Military expenditure as a % of GDP.
- **GDP_current_US**: GDP (current US$).
- **political_stability_estimate**: Index that makes an estimate of Political Stability and Absence of Violence/Terrorism.
- **political_stability_std**: Standard error of the estimate of Political Stability and Absence of Violence/Terrorism.
- **rule_of_law_estimate**: Index that makes an estimate of the Rule of Law.
- **rule_of_law_std**: Standard error of the estimate of Rule of Law.
- **regulatory_quality_estimate**: Index that makes an estimate of Regulatory Quality.
- **regulatory_quality_std**: Standard error of the estimate of Regulatory Quality.
- **government_expenditure_on_education%**: Government expenditure on education, total, as a % of GDP.
- **government_health_expenditure%**: Domestic general government health expenditure as a % of GDP.
- **multidimensional_poverty_headcount_ratio%**: Multidimensional poverty headcount ratio (% of total population).
- **gini_index**: Gini index.
- **birth_rate**: Birth rate, crude (per 1,000 people).

- **death_rate**: Death rate, crude (per 1,000 people).
- **life_expectancy_at_birth**: Life expectancy at birth, total (years).
- **population**: Total population.
- **rural_population**: Rural population.
- **voice_and_accountability_estimate**: Index that makes an estimate of Voice and Accountability.
- **voice_and_accountability_std**: Standard error of the estimate of Voice and Accountability.
- **intentional_homicides**: Intentional homicides (per 100,000 people).

Added columns to the dataset for visualisation

- **gdp_per_capita**: The GDP per capita, measured in current US dollars.It is calculated by dividing the total GDP by the population of the country/region.
- **debt_to_gdp%**: Debt to GDP ratio, expressed as a percentage, shows the country's total government debt compared to its GDP.
- **gdp_growth%**: GDP growth percentage measures the rate at which a country's GDP is growing or shrinking compared to the previous year.
- **tax_revenue_gdp%**: Tax revenue as a percentage of GDP shows the portion of a country's economic output that is collected through taxes.
- **rural_population_percentage**: It is calculated by dividing the rural population of the country/region by total population.
- **co_2_per_capita**: It is calculated by dividing the total CO_2 emissions of the country/region by total population (in kt).

We have broken down our analysis into the following three tasks :

- T1: Analyzing Trends and Impacts of major economic events across the globe over the years
- T2: Analyzing key social indicators across countries to uncover trends and insights into global social development
- T3: Analyzing trends and relations of environmental and sustainability indicators to evaluate global progress in addressing climate change.

II. Tasks

A. T1: Analyzing Trends and Impacts of major economic events across the globe over the years

In this comprehensive analysis, we aim to delve into a wide range of critical economic indicators across a diverse set of countries, spanning different time periods, economic landscapes, and stages of development. By examining key metrics such as GDP per capita, inflation rates, government debt levels, trade performance, and energy consumption, we can uncover significant patterns that reveal both growth trajectories and underlying vulnerabilities in national economies.

Our approach highlights the economic progression of global leaders like Luxembourg, the USA, and China, alongside countries experiencing notable economic transformations, such as Ukraine and Sri Lanka. Through this exploration, we aim to gain a deeper understanding of how countries with vastly different economic conditions and government policies have managed growth, handled crises, and responded to external pressures like global recessions, financial instability, and geopolitical events.

Luxembourg, with the highest GDP per capita, provides a case study of how smaller economies can leverage high productivity and innovation to maintain sustained growth. The USA and China serve as examples of economic giants that demonstrate different approaches to development—one characterized by long-term stability and high-income status, and the other by rapid industrialization and expansion into global markets.

In contrast, the inclusion of countries like Ukraine and Sri Lanka in the analysis helps illustrate how socio-political instability, conflict, or economic mismanagement can severely hinder a country's ability to achieve sustainable growth. These outliers also provide a deeper context for understanding inflation patterns, government debt dynamics, and economic recovery efforts.

By correlating various economic factors across these countries, we aim to draw inferences about economic resilience, the impact of government policies on growth and stability, and the broader socio-political influences that shape national economies. Ultimately, this analysis provides a holistic view of the interplay between economic indicators and how different economies respond to both challenges and opportunities over time.

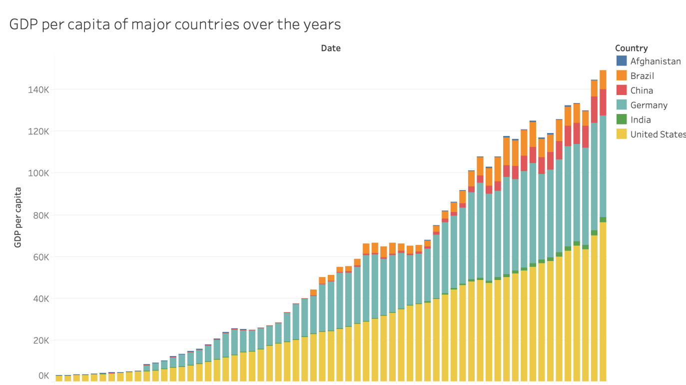


Fig 1. Stacked bar chart for countries such as Afghanistan, Brazil, China, Germany, India and USA representing the trend in GDP per capita over the years from 1960 to 2022

- **United States**: In (Fig 1.) the country shows a consistent and significant increase in GDP per capita, especially from the mid-1980s onwards, reaching its highest point by 2022. This reflects its dominance as one of the world's largest and most developed economies, driven by a stable economy, high productivity, and technological advancements, leading to high living standards.
- **Germany**: In (Fig 1.) the country exhibits a similar trend to the United States with robust economic growth and a high-income status throughout the period from 1960 to 2022, maintaining its position as a leading global economy.
- **China**: In (Fig 1.) the country demonstrates an extraordinary rise in GDP per capita starting in the late 1990s and early 2000s, corresponding to its economic reforms, industrialization, and rapid expansion in global trade. By 2022, China's GDP per capita sees substantial growth, but the gap between China and more developed economies like the US and Germany persists.
- **Brazil**: In (Fig 1.) the country experiences an upward trend in GDP per capita, particularly from the 2000s, benefiting from periods of economic growth in Latin America. However, its progress has been slower compared to China, and it remains behind more developed economies.
- **India**: In (Fig 1.) the country starts seeing significant GDP per capita growth from the early 2000s, driven by rapid industrialization and economic development. India still lags behind major developed countries.
- **Afghanistan**: In (Fig 1.) the country shows little to no significant rise in GDP per capita over the years from 1960 to 2022, primarily due to ongoing political instability, conflict, and lack of substantial economic development or industrialization.

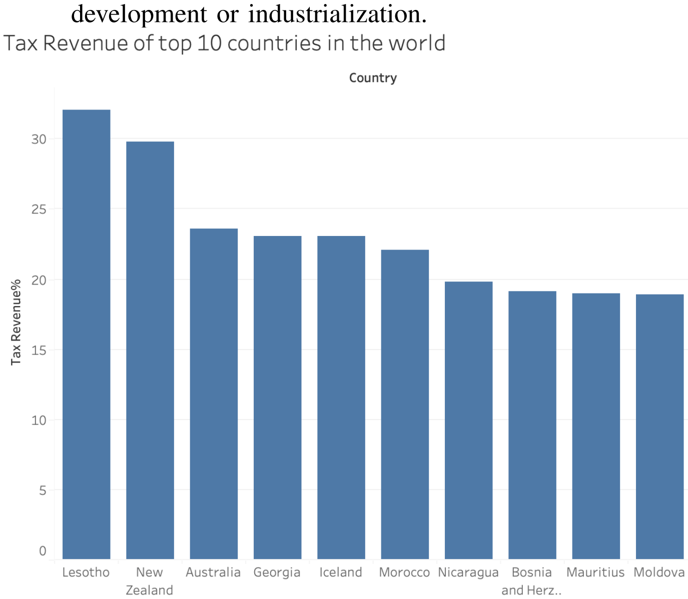


Fig 2. Bar chart representing top 10 countries with the maximum tax revenue from 1960 to 2022

- In (Fig 2.) The bar chart illustrates tax revenues as a percentage of GDP for several countries from 1960 to 2022, emphasizing those with the highest tax collections.
- In (Fig 2.) **Lesotho** stands out with over 30% of its GDP generated through tax revenue, suggesting a strong reliance on taxes due to limited other economic resources.
- In (Fig 2.) **New Zealand** follows closely, just under 30%, reflecting its efficient tax system and reliance on public funding for comprehensive services such as healthcare and education.
- In (Fig 2.) **Australia**, **Georgia**, and **Iceland** demonstrate high tax-to-GDP ratios, ranging between 22% and 25%, indicating robust tax collection frameworks that support government functions and infrastructure.
- In (Fig 2.) **Morocco**'s slightly lower tax revenue, around 21%, suggests a balanced economy with both tax and other income sources contributing to GDP.
- In (Fig 2.) **Nicaragua**, **Bosnia and Herzegovina**, **Mauritius**, and **Moldova**, with tax revenues between 18% and 20%, represent economies where tax collection may be more challenging due to factors such as weaker administrative systems or a larger informal sector.
- The presence of higher tax-to-GDP ratios in developed countries highlights their ability to maintain robust tax systems, which is critical for sustaining public services and infrastructure development.
- Lower tax revenue in some countries can also reflect a higher reliance on foreign aid or non-tax revenue sources, which may not be sustainable for long-term economic growth.
- The data reveals significant variations in tax collection across these countries, underscoring the importance of efficient tax systems in funding public services. It also highlights the struggles of developing nations in collecting substantial tax revenues.

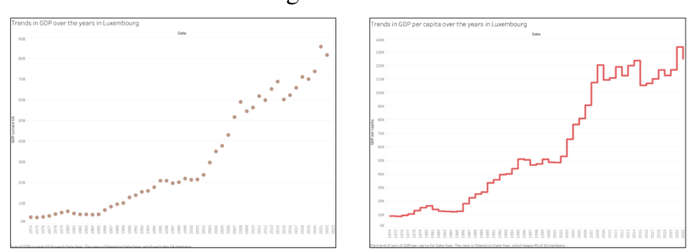


Fig 4.1 Scatter Plot representing changes in the annual GDP from 1974 to 2023

Fig 3.2 Line graph representing changes in the GDP per capita from 1974 to 2023

Fig 3. Graphs representing annual GDP and GDP per capita of Luxembourg

- In (Fig 3.) the analysis of Luxembourg highlights several key trends that provide insights into the country's economic and technological evolution.
- In (Fig 3.) Luxembourg, known for having the highest GDP per capita in the world, shows significant progress across various dimensions from 1974 to

2022.

- In (Fig 3.) the GDP, both in current US dollars and per capita, displays a steady increase over the years, reflecting the nation's economic growth, resilience, and its prominent position as a global financial hub.
- In (Fig 3.) the graphs reveal that during major global economic crises, such as the stock market crash and the dot-com bubble burst, Luxembourg experienced a slight dip in its economic performance. Luxembourg's economy also dipped during the 2015-16 period mainly due to changes in EU tax regulations aimed at curbing tax avoidance. These stricter rules led many multinational corporations, which Luxembourg's economy heavily relies on, to restructure their tax arrangements or relocate. The end of favorable tax rulings further contributed to the decline, as it affected corporate tax strategies and reduced economic activity in the country.
- Despite these global crises, Luxembourg's economy quickly rebounded, demonstrating its ability to recover and maintain its growth trajectory.

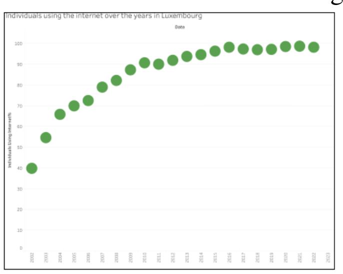


Fig 3.1 Scatter Plot representing Individuals using internet from 2002 to 2023

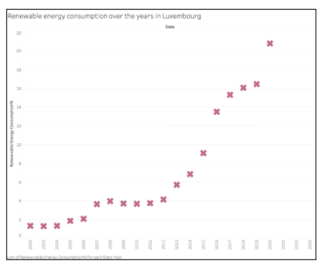


Fig 3.2 Scatter Plot representing Renewable energy consumption from 2002 to 2023

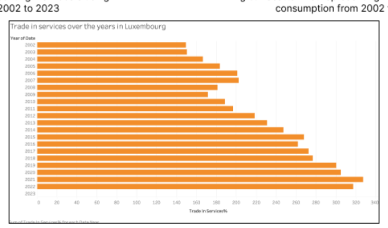


Fig 3.3 Scatter Plot representing Individuals using internet from 2002 to 2023

Fig 4. Graphs representing different data for Luxembourg

- In (Fig 4.1) The adoption of the internet is a clear indicator of Luxembourg's modernization, with a marked increase in individuals using the internet from 2002 to 2022.
- This growth parallels the global technological boom and shows Luxembourg's commitment to becoming a tech-forward nation.
- In (Fig 4.2) The rising consumption of renewable energy over the same period reflects Luxembourg's emphasis on sustainability, aligning with global efforts to transition to cleaner energy sources.
- In (Fig 4.3) Additionally, trade in services has shown continuous growth, crucial for Luxembourg's status as an international financial and services hub.
- This suggests that Luxembourg has effectively capitalized on globalization and international trade, further boosting its economic prosperity.
- In sum, Luxembourg's economic, technological, and environmental strategies have positioned it as a high-income, progressive nation with sustained growth in key areas over the past few decades.
- In (Fig 4.1) This growth parallels the global technological boom and shows Luxembourg's commitment to becoming a tech-forward nation.
- In (Fig 4.2) The rising consumption of renewable energy over the same period reflects Luxembourg's emphasis on sustainability, aligning with global efforts to transition to cleaner energy sources.
- In (Fig 4.3) Additionally, trade in services has shown continuous growth, crucial for Luxembourg's status as an international financial and services hub.

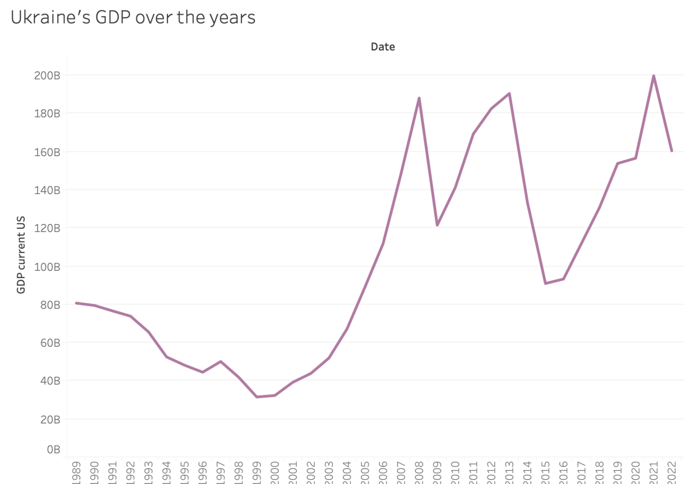


Fig 5. Line graph representing annual GDP for Ukraine from 1989 to 2023

- In (Fig 5.) Ukraine's GDP drop in 2016 and 2017 was largely a consequence of the ongoing geopolitical crisis and its economic fallout.
- The annexation of Crimea by Russia in 2014 and the conflict in the Donbas region severely disrupted key industries and infrastructure, particularly in eastern Ukraine, leading to a sharp decline in industrial output.
- The conflict triggered a severe recession in 2014-2015, which carried over into 2016-2017, slowing recovery efforts.
- The loss of trade with Russia, Ukraine's largest trading partner, further exacerbated the economic downturn.
- The sharp depreciation of the Ukrainian hryvnia led to high inflation and reduced purchasing power.
- Ukraine's structural economic challenges, such as corruption and reliance on heavy industries, worsened the situation.
- Bailout packages from the IMF helped stabilize the economy, but accompanying austerity measures added to short-term economic pain, delaying recov-

ery.

- These factors combined to significantly impact Ukraine's GDP during this period.
- In (Fig 3.) In contrast, Luxembourg's GDP growth displays consistent growth over the past two decades, emphasizing the nation's economic stability and effective governance.
- In (Fig 5.) International sanctions imposed on Russia following the annexation of Crimea had a spillover effect on Ukraine's economy, disrupting regional trade flows.

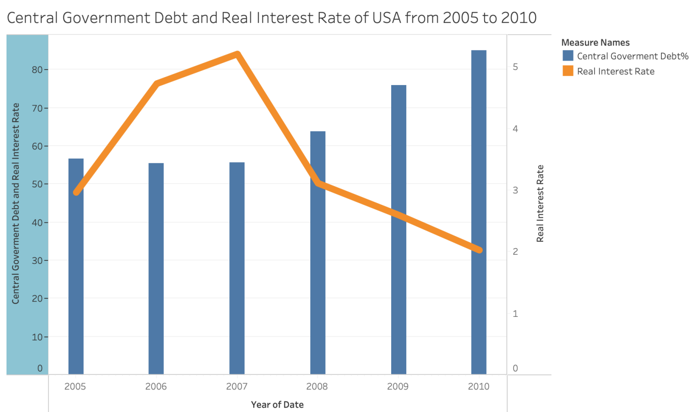


Fig 6. Mixed Line and Bar chart representing the real interest rate and central government debt of USA from 2005 to 2010

- In (Fig 6.) The graph depicts the USA's central government debt percentage alongside the real interest rate from 2005 to 2010, showing key trends during this period.
- Government debt remains relatively steady until 2007, after which there is a sharp rise in 2008, likely due to the global financial crisis.
- The spike in debt reflects increased government spending aimed at stabilizing the economy, including stimulus measures and bailouts.
- Meanwhile, the real interest rate peaks in 2007 before starting a steady decline, a typical response during financial crises.
- Central banks, such as the Federal Reserve, lower interest rates to encourage borrowing, stimulate the economy, and counteract the recession.
- By 2010, government debt remains high while the real interest rate continues to decrease, signaling continued economic strain.
- This divergence between rising debt and falling interest rates underscores the government's reliance on borrowing to fund stimulus measures and infrastructure investments.
- The declining real interest rate reflects the Federal Reserve's efforts to make borrowing cheaper, spurring consumer spending and business investment.
- Despite these measures, the persistently high levels of debt indicate long-term fiscal challenges, which could have implications for future monetary policy and economic growth.
- The sustained rise in government debt post-2008 suggests that recovery efforts, although necessary, may have come at the cost of increased fiscal burden for future generations.
- The continued decrease in real interest rates raises concerns about the potential long-term effects on savings rates and inflation management in the post-crisis recovery phase.
- As the government focused on stimulus, the trade-off between short-term economic recovery and long-term fiscal sustainability became increasingly apparent.

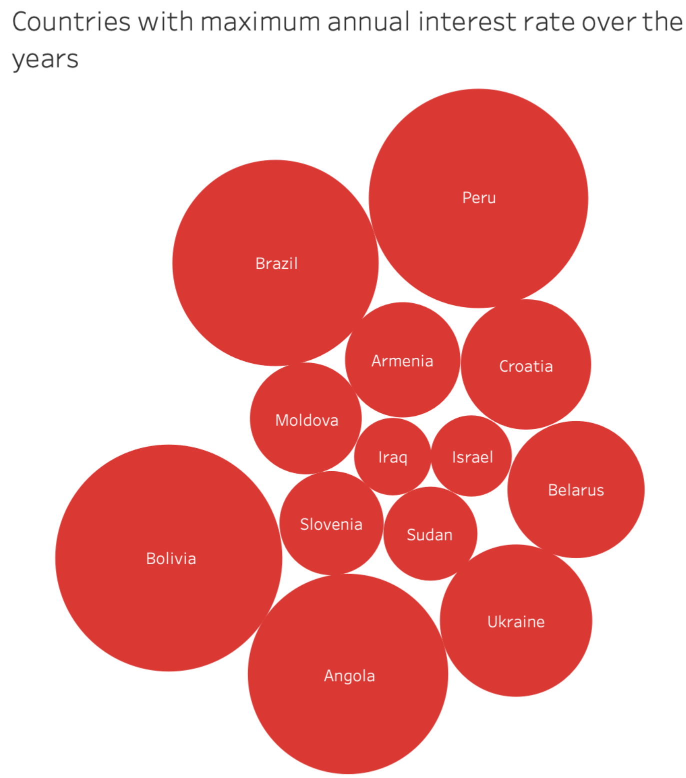


Fig 7.Circle chart to show the countries with maximum annual inflation percentage over the years

- In (Fig 7.) The Bubble chart highlights countries with the highest annual inflation percentages over the years.
- Peru, Bolivia, and Brazil have the largest circles, indicating severe inflationary pressures during this period.
- This likely resulted from economic mismanagement or crises, leading to significant currency devaluation and rising prices.
- Countries like Angola, Ukraine, and Belarus also experienced notable inflation, potentially due to external shocks or economic transitions.

- The inflation levels in these countries disrupted economic stability, affecting overall economic performance and reducing living standards.
- High inflation typically erodes purchasing power, making it harder for populations in these nations to maintain consistent standards of living.
- Inflationary spikes also likely led to economic instability, forcing governments to implement corrective measures, such as monetary tightening or fiscal adjustments.
- Sustained high inflation often triggers a loss of investor confidence, potentially leading to reduced foreign direct investment and slower economic growth.
- Countries with high inflation may face challenges in maintaining stable exchange rates, further contributing to economic volatility.

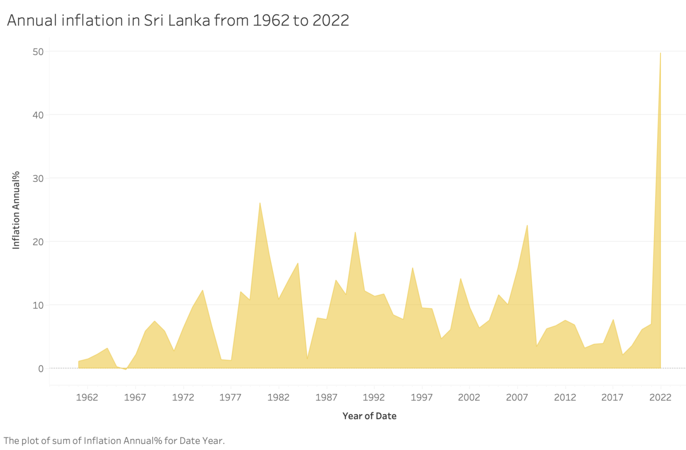


Fig 8. Area graph showing the annual inflation rate in Sri Lanka from 1962 to 2022

- **Inflation Spikes**: In (Fig 8.) These fluctuations have mainly been caused due to the civil war among other factors which ended in 2008. The country heavilty relies on imports, tourism and foreign exchanges as well which have not been very lucrative over the years.
- **Periodicity**: Inflation shows a cyclical behavior with sharp increases and decreases approximately every 10-15 years.
- **Recent Inflation Surge**: The inflation rate in 2022 exceeds 50%, the highest point in the country's modern economic history.
- **Stable Periods**: Between 2000-2007 and 2010-2019, inflation remained relatively stable, ranging from 5% to 10%.
- **Global Events Impact**: Certain peaks, such as the one observed in 2008, align with global economic events like the financial crisis, suggesting external factors may influence domestic inflation.
- **Government Policies and Monetary Decisions**: Changes in government policies and monetary decisions, such as interest rate adjustments or fiscal stimulus measures, have impacted inflation. For example : Aggressive monetary easing during downturns can lead to higher inflation in later years.

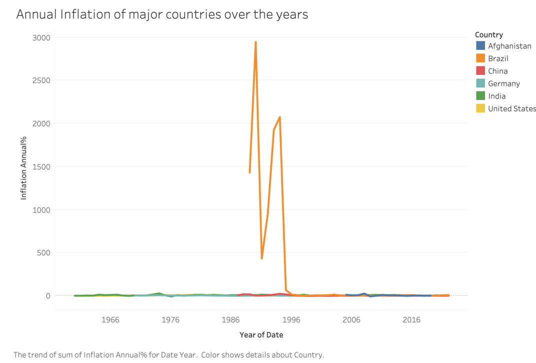


Fig 9. Continuous line graphs for countries such as Afghanistan, Brazil, China, Germany, India and USA

- In (Fig 9.) Brazil exhibits more volatile inflation, influenced by political instability and external market shocks.
- Afghanistan does not show any significant spikes in inflation up to 2020 due to the lack of available data for the year 2021.
- After the Taliban takeover in 2021, Afghanistan's inflation rate spiked, but this is not reflected in the graph because of null values.
- To address the missing data, the average inflation rate was imputed to handle the discrepancies effectively.
- China, despite rapid economic growth, maintains controlled inflation, highlighting a focus on economic stability.
- The USA and Germany show moderate but stable inflation trends, reflecting their mature economies and well-developed fiscal controls.
- This comparison underscores the economic stability of developed nations like the USA and Germany, whereas countries like Afghanistan and Brazil remain vulnerable to inflation spikes due to external and internal pressures.

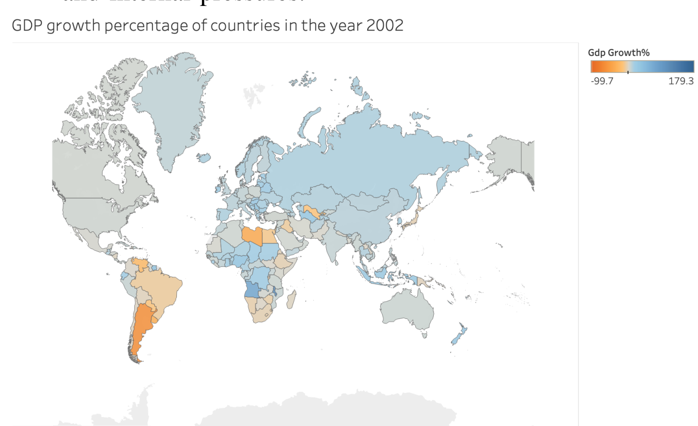


Fig 10. Heatmap for the GDP growth percentage of all the countries in the year 2002

- **Global Economic Impact:** In (Fig 10.) We chose this particular year so that the heatmap could depict the widespread economic impact of the 2000-2002 recession, following the burst of the dot-com bubble. Many countries in light blue exhibit low or negative GDP growth, reflecting the global slowdown.

- **Severe Decline in Argentina:** In (Fig 10.) Argentina shows a significantly negative GDP growth rate (dark orange), indicating the country's major economic crisis during this period, including its debt default in 2001-2002.
- **Mixed Performance in Africa:** In (Fig 10.) Africa displays varying GDP growth percentages, with some countries experiencing growth (e.g., blue regions in East Africa) while others, such as in Central Africa, have negative growth (orange regions). This suggests the region's diverse economic conditions in response to global events.
- **Stable Growth in China and India:** In (Fig 10.) China and India (light blue) show positive but moderate GDP growth, reflecting their growing economies at the time, somewhat insulated from the dot-com bubble's direct effects.
- **Recovery in North America:** In (Fig 10.) The United States and Canada appear with slight positive growth after recovering from the initial shock of the dot-com bubble burst. The U.S. experienced a mild recession, but recovery was underway by 2002.
- **Resilient Growth in Eastern Europe:** In (Fig 10.) Countries in Eastern Europe show moderate GDP growth rates, potentially reflecting their transition economies benefitting from market reforms and foreign investment during this time.

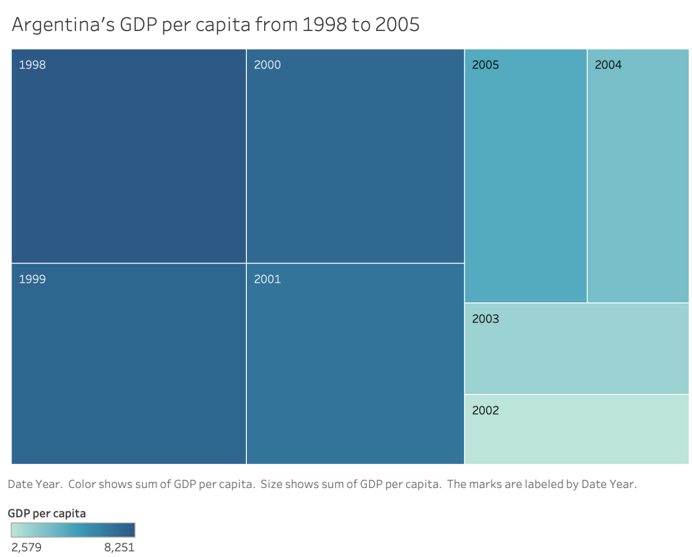


Fig 11. Treemap for the GDP per capita of Argentina from 1998 to 2005

- **Economic Stability Pre-2001:** In (Fig 11.) From 1998 to 2000, Argentina's GDP per capita remained relatively stable, as indicated by the darker blue shades, representing higher GDP per capita values. This suggests relative economic stability during these years.
- **Sharp Decline in 2002:** In (Fig 11.) The year 2002 shows a notable dip in GDP per capita, as indicated by the lighter color in the treemap. This corresponds to Argentina's severe economic crisis and debt default in 2001-2002, which led to a sharp drop in the country's economic output.
- **Gradual Recovery Post-2002:** In (Fig 11.) From 2003 onwards, the GDP per capita shows a slow recovery, with the color transitioning back to a darker shade by 2005. This suggests that Argentina's economy started to rebound after the crisis, although the recovery was gradual.
- **Impact of Argentina's Crisis:** In (Fig 11.) The treemap visually depicts the significant impact of Argentina's financial crisis during this period, with 2002 standing out as a year of severe economic decline.
- **Sustained Growth Post-2005:** In (Fig 11.) From 2006 onwards, Argentina's GDP per capita consistently increased, indicated by the darker and more consistent shades, suggesting a period of sustained economic growth following the recovery phase.

B. T2: Analyzing key social indicators across countries to uncover trends and insights into global social development

In this task , we will analyze a range of social indicators across different countries, focusing on factors that directly or indirectly affect the quality of life for their citizens. By examining social indicators , we aim to identify trends and patterns that highlight the social development and well-being of populations. This analysis will help us understand how these indicators vary across nations and the extent of their impact on people's lives.

Hypothesis 1: Access to Internet and Electricity

Our hypothesis examines the relationship between access to electricity and internet usage, leveraging two key attributes from the dataset: Access to Electricity and Individuals Using the Internet . We analyzed these trends to better understand how internet and electricity accessibility has evolved across countries

We have used the 2 main labels given in the dataset , Access to electricity precentage and INdividuals using internet

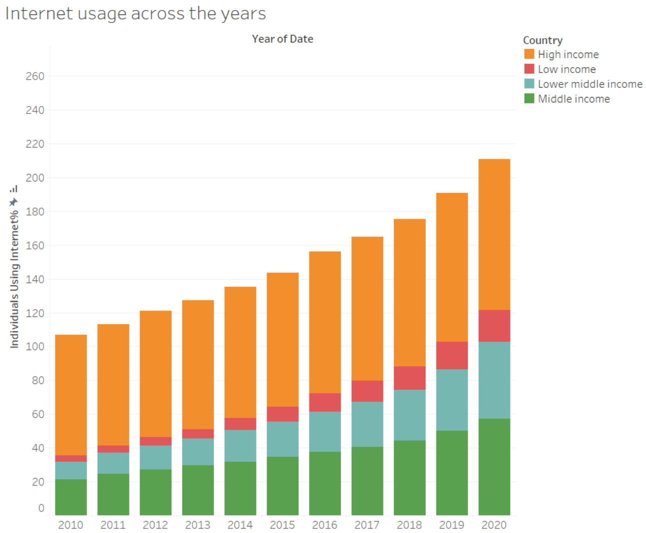


Fig 12. Stack Bar showing Access to electricity across years for different countries

Figure 12 presents a stacked bar chart depicting internet usage percentages for country categories based on income levels: high income, middle income, lower middle income, and low income. The data spans from 2010 to 2020, a period chosen because internet access prior to 2010 was either limited or non-existent in many countries.

The chart shows that by 2010, high-income countries already had widespread internet access, which grew steadily throughout the decade. In contrast, low-income countries had minimal access in 2010 and made gradual progress, though a significant gap remained by 2020. This disparity is due to differences in infrastructure, economic resources, and technological investment.

High-income countries have well-established digital networks, while low-income countries face challenges like limited infrastructure and resources, contributing to the persistent digital divide. Additionally, the availability of affordable internet and technology plays a crucial role in determining access, particularly in low-income regions. Countries in the middle-income group made noticeable strides, bridging the gap slightly but still lagging behind higher-income nations.

The role of government policies, such as investments in digital infrastructure and the promotion of internet access, also shaped the expansion of usage across income groups. By 2020, even lower middle-income countries experienced a significant increase in internet access, signaling progress but still highlighting the persistent inequities. These ongoing disparities have implications not only for digital connectivity but also for education, healthcare, and economic opportunities in the 21st century..

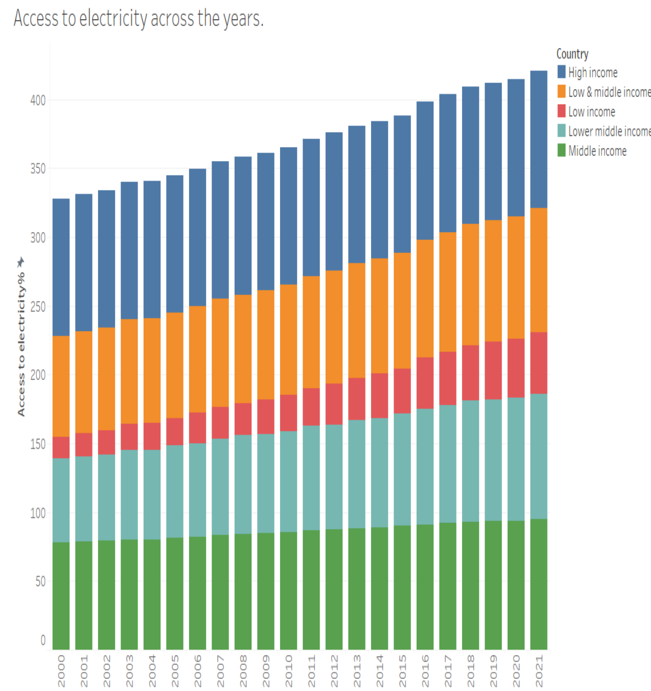


Fig 13. Stack Bar showing Access to electricity across years for different countries

Figure 13 shows the trends in Access to Electricity across high, middle, lower middle, and low-income countries from 2000 to 2020 using a stacked bar chart. High-income countries had near-universal access to electricity by 2000, and this remained consistent over the two decades.

On the other hand, low-income countries, which started with significantly lower access in 2000, showed gradual improvement over time. However, by 2020, the gap between low-income and high-income countries remained significant. This disparity can be attributed to differences in infrastructure investment, economic resources, and policy focus.

High-income countries benefit from longstanding investments in their national electrical grids, while low-income nations face challenges like remote populations and limited resources, slowing their progress in achieving universal electricity access. Additionally, political instability and inadequate government support have also hampered efforts in certain low-income regions. In middle-income countries, access to electricity improved significantly, with several nations on the path to achieving near-universal access.

Despite the progress, achieving full electrification for low-income countries will require sustained international support, increased funding, and technology transfer to overcome persistent barriers.
High-income countries benefit from longstanding investments in their national electrical grids, while low-income nations face challenges like remote populations and limited resources, slowing their progress in achieving universal electricity access

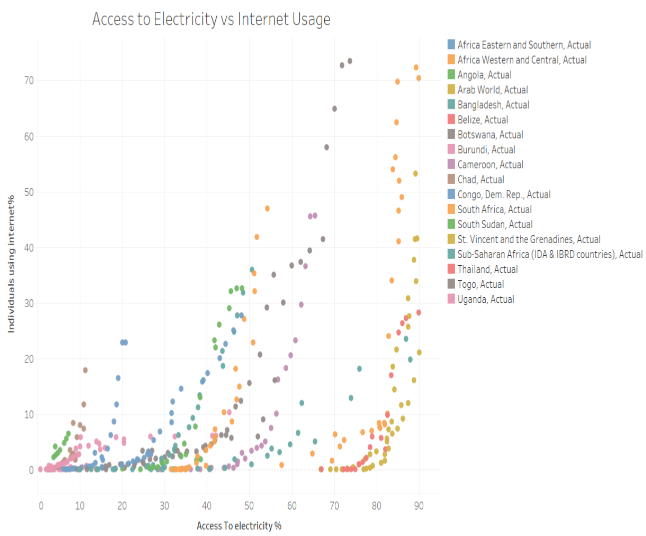


Fig 14. Scatter plot showing relation between Access to electricity and internet usage across the year

Figure 14 shows the correlation between Access to Electricity and Individuals Using the Internet for 20 countries from 1990 to 2020. The scatter plot reveals a clear trend: as electricity access increases, internet usage rises. Countries with widespread electricity in the 1990s saw rapid internet growth, while those with lower access experienced slower progress, improving as they expanded their electrical infrastructure.

The correlation exists because electricity is essential for internet usage. Countries with higher electricity access saw faster growth in internet use, while those with limited access experienced slower progress. Improvements in internet usage closely follow expansions in electricity access. Additionally, countries that invested in both electricity and telecommunications infrastructure simultaneously showed the most rapid advancements in internet connectivity. This highlights the importance of a coordinated approach to development in bridging the digital divide.

Hypothesis 2: Life Expectancy and Birth Rate

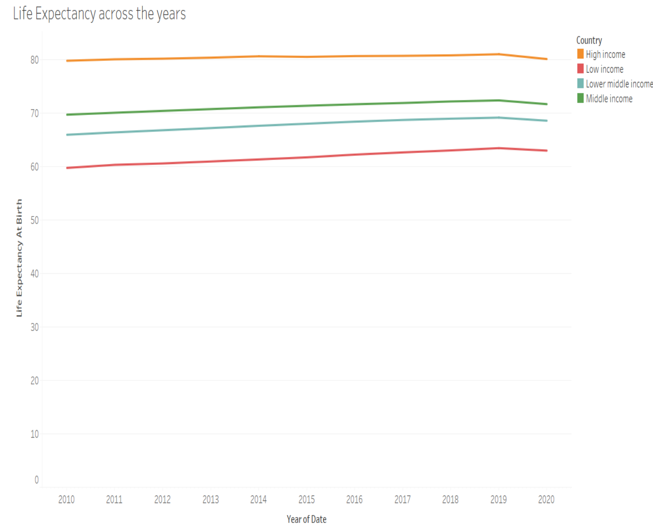


Fig 15. Line graph showing Life expectancy for countries accross the years

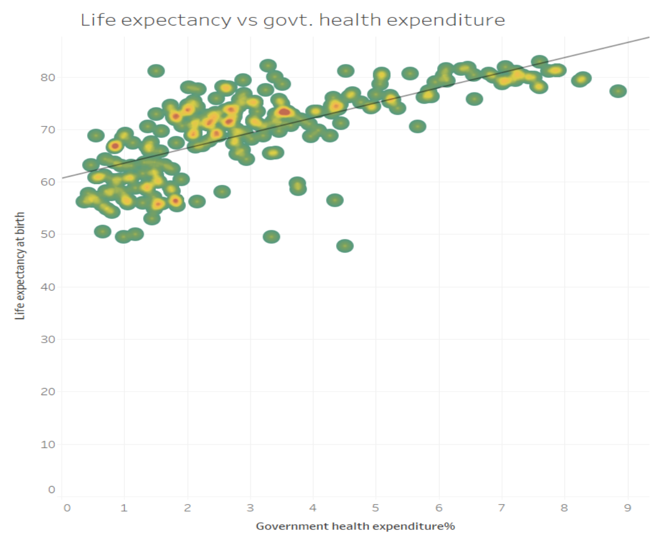


Fig 16. Density plot between Govt Health expenditure and Life expectancy

Figure 15 shows life expectancy trends across income categories: high, upper middle, lower middle, and low income. Higher-income countries consistently have the longest life expectancy due to better healthcare infrastructure and resources.

Figure 16 reinforces this, highlighting a positive correlation between government health expenditure (as a percentage of GDP) and life expectancy. Countries investing more in healthcare tend to see longer lifespans, as increased spending improves medical access and quality of care.

However, while more spending generally leads to longer life expectancy, other factors like healthcare efficiency, lifestyle, and socioeconomic conditions also play significant roles in determining outcomes.

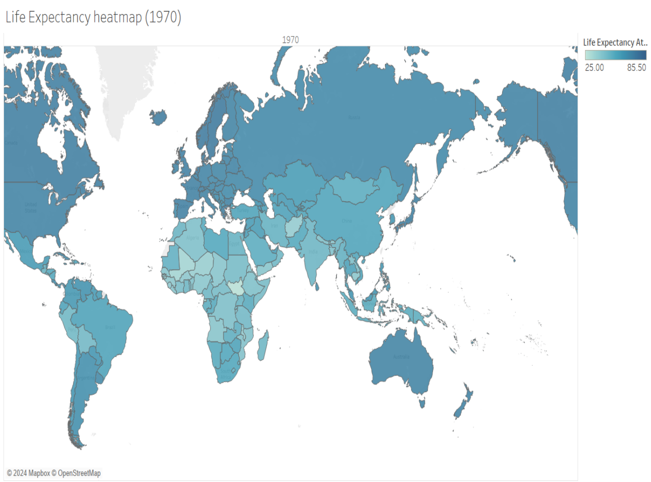


Fig 17.Life Expectancy in 1970

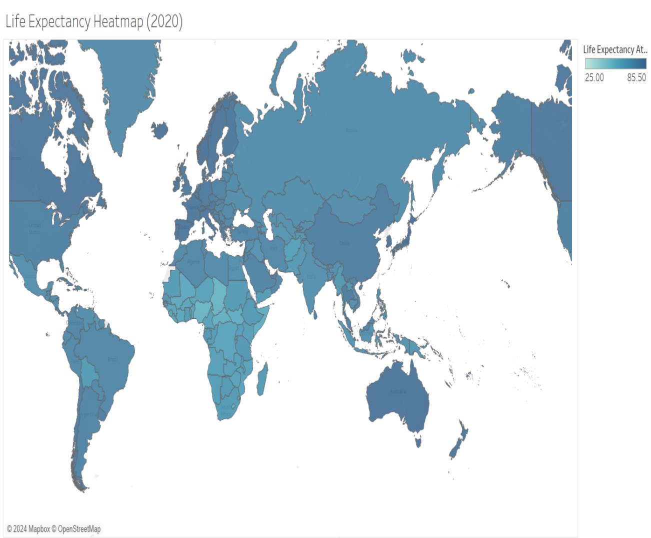


Fig 18. Life Expectancy in 2020

Figure 17 and Figure 18 compares global life expectancy in 1970 and 2020 through heatmaps, showcasing significant improvements over five decades.

Key Trends:

– In 1970, regions like Africa, South Asia, and Latin America had life expectancies below 50-60 years, while North America, Europe, and East Asia were higher. By 2020, most countries, including those in Africa, saw life expectancy rise above 70 years.

– In 1970, developed regions had life expectancies over 70 years, while many other regions lagged far behind. By 2020, global disparities in life expectancy narrowed as healthcare improved worldwide.

.

This progress is largely attributable to advances in healthcare, sanitation, and nutrition, alongside economic growth that enabled greater investments in public health systems. The widespread adoption of vaccines and treatments for diseases played a crucial role in increasing life expectancy in developing countries. Global initiatives such as the eradication of smallpox and efforts to reduce infant mortality have also contributed significantly to the overall rise in life expectancy.
However, despite these advances, some regions continue to face challenges. Issues like political instability, economic inequality, and poor healthcare infrastructure in certain countries have hindered progress, particularly in parts of sub-Saharan Africa and conflict-prone areas. As a result, while the global gap in life expectancy has narrowed, disparities still persist, and ongoing efforts are required to ensure further improvements in health and longevity across all regions.

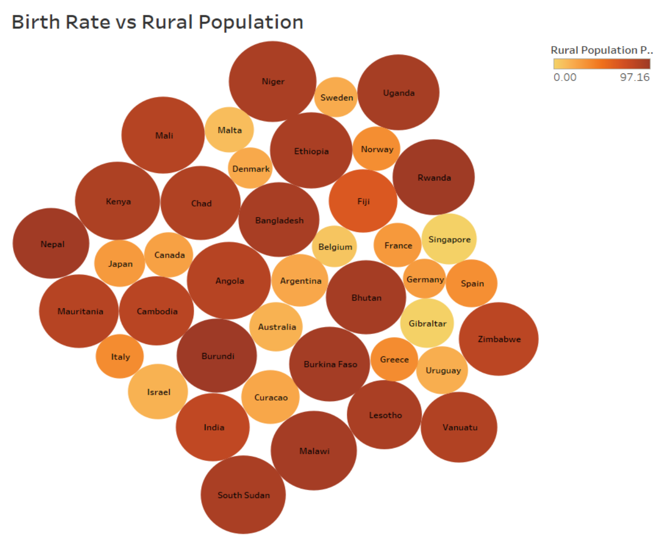


Fig 19. . Bubble chart with color channel highlighting Rural Population Percentage and size channel highlighting Birth rate of different countries

In Fig 19, we have used a bubble chart to visualize birth rates for the top 20 and bottom 20 countries as per the dataset, with bubble size indicating birth rates and color representing rural population percentages.The chart clearly shows that larger bubbles (higher birth rates) are often darker (higher rural populations), highlighting how urban and rural conditions impact demographic trends.

Urbanization, which involves higher living costs and a focus on careers and education, generally leads to lower birth rates. Conversely, rural areas, with less access to family planning and healthcare, tend to have higher birth rates.

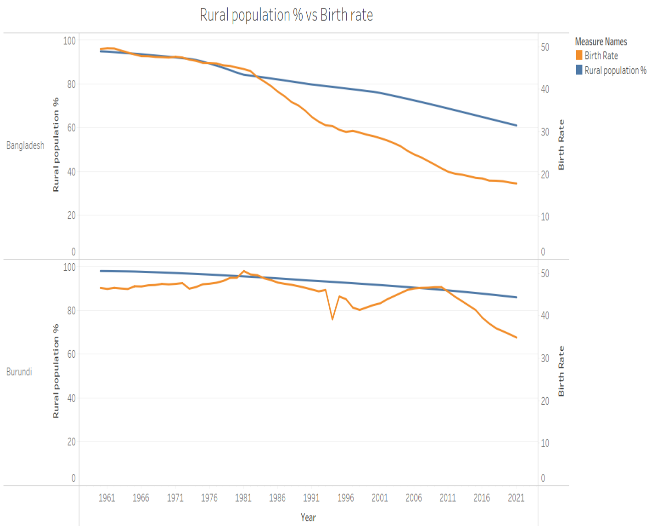


Fig 20. Line chart showing Rural population percentage vs Birth rate across the years for Bangladesh and Burundi

To support this observation, we have used dual axis line plot illustrating the trend with two specific examples: Bangladesh and Burundi , which can be seen in figure 20. In 1960, Bangladesh had a rural population percentage of 94.44 and a birth rate of 49.5 per 1,000 people. Burundi's figures were nearly comparable, with a rural population of 97 and a birth rate of 46.52.

However, over the years, the trends diverged significantly. By 2020, Bangladesh had substantially reduced its rural population to 61 and saw a corresponding sharp decrease in its birth rate, which fell to 18 per 1,000 people. In contrast, Burundi did not achieve a similar reduction in its rural population, which remained high at 85 in 2020. As a result, Burundi's birth rate remained relatively high at 35 per 1,000 people.This clear divergence in the trends between the two countries highlights the strong relationship between rural population reduction and declining birth rates.

Bangladesh's successful reduction in rural population correlates with a dramatic drop in birth rates, while Burundi's persistent high rural population is associated with a higher birth rate.

This demonstrates that countries with lower rural populations tend to experience lower birth rates, reinforcing the observed trend.. While urbanization is a major factor contributing to this trend, it is important to recognize that other factors may also influence birth rates.

Hypothesis 3: Human Capital Index.

In Fig 21, the scatter plot shows a clear positive correlation between GDP per capita and the Human Capital Index (HCI) for the years 2010, 2017, 2018, and 2020

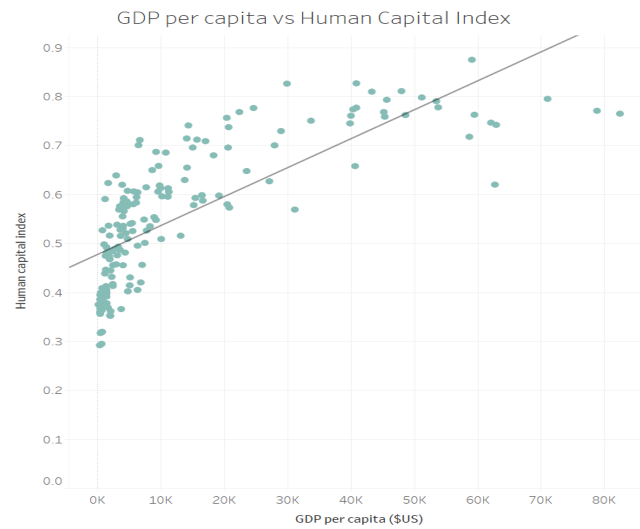


Fig 21. Scatter plot between Human Capital index and GDP per Capita

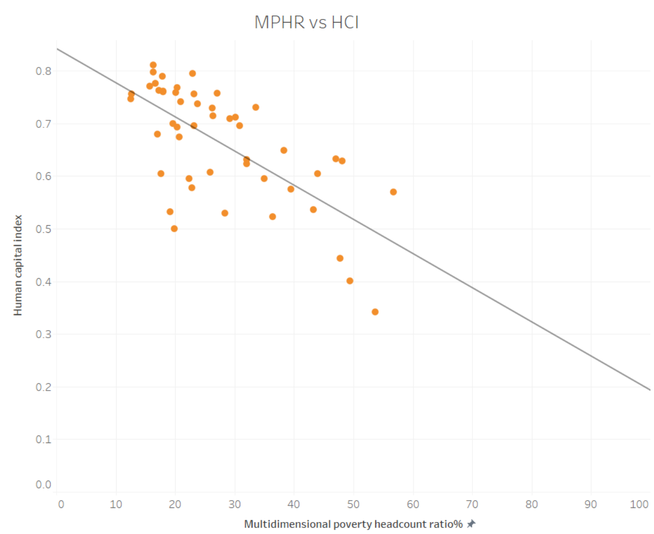


Fig 22. Scatter plot between Human Capital index and Multidimensional poverty headcount ratio

(the years for which we had HCI data). Wealthier countries generally have higher HCI scores due to better education, health, and resources. European countries, with high GDP per capita, benefit from advanced systems, while many African countries, with lower GDP per capita, struggle with lower HCI scores due to limited resources.

In Fig 22, the scatter plot of HCI versus the Multidimensional Poverty Headcount Ratio (MPHR) shows that HCI decreases as MPHR increases. This indicates that higher levels of poverty are associated with lower human capital outcomes.

Hypothesis 4: Intentional Homicides

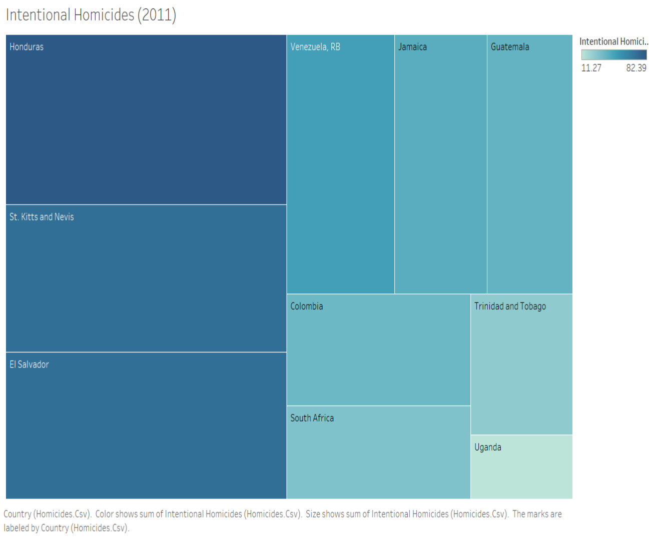


Fig 23. Treemap representing intentional homicides in 2011

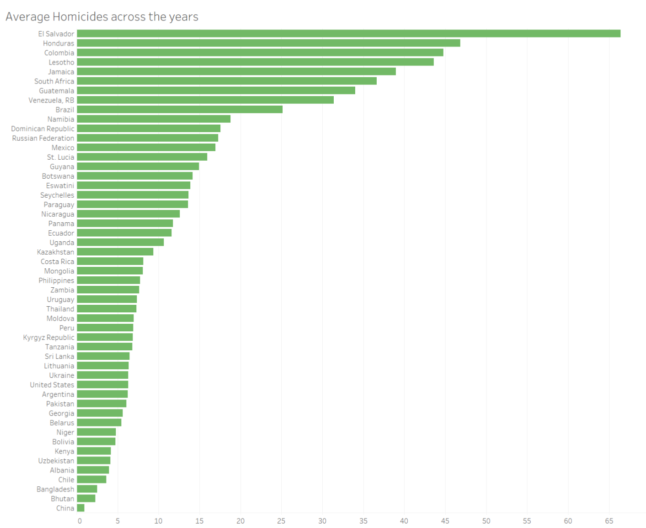


Fig 24. Bar graph representing average intentional homicides for each country

In our analysis on fig 23, we used data from 2011 . The data reveals that homicide rates are particularly high in Latin America and the Caribbean, as well as Sub-Saharan Africa. These regions experience elevated death tolls not only from homicides but also from armed conflicts.

According to the Geneva Declaration on Armed Violence and Development , a quarter of all violent deaths occur in just 15 countries which are shown in treemap (Data of the remaining countries are not available in dataset), averaging over 30 violent deaths per 100,000 people annually. Half of these are in Latin America and the Caribbean. In many of these countries, homicides, rather than armed conflicts, account for the majority of violent deaths.

The relationship between violent death rates and socioeconomic development shows that higher homicide rates are often linked to high levels of income disparity, extreme poverty, and hunger. Conversely, countries that have improved their rule of law have seen a decline in homicide rates, suggesting that stronger legal frameworks and socioeconomic development contribute to reducing violent deaths.

In fig24, we have used bar chart to analyze average intentional homicides for each country across all the years(as per the available data). Again , the same trend was followed.Homicide rates are particularly high in Latin America and the Caribbean, as well as Sub-Saharan Africa.

C. T3: Assessing environmental and sustainability indicators to evaluate global progress in addressing climate change, resource management, and ecological preservation.

In this task, we will analyze a variety of environmental and sustainability indicators across different countries, focusing on the factors that influence ecological balance, resource management, and environmental preservation. By examining key metrics such as carbon emissions, greenhouse emissions, forest and agricultural land area percentage, energy consumption and renewable resource usage, we aim to uncover trends that shed light on global progress towards sustainability. This analysis will provide insights into how countries are managing their environmental responsibilities and the impact of these actions on long-term ecological and human well-being.

Hypothesis 1: Rural Population and Agricultural Land Area

Our hypothesis examines the relationship between rural population percentage and agricultural land area percentage, leveraging two key attributes from the dataset: Rural Population (% of total population) and Agricultural Land

(% of land area). By analyzing these trends, we aim to explore how the proportion of rural populations and agricultural land usage varies across countries, particularly when divided according to income categories. This analysis will help us better understand the dependency of rural populations on agriculture and how these factors are linked across different economic contexts.

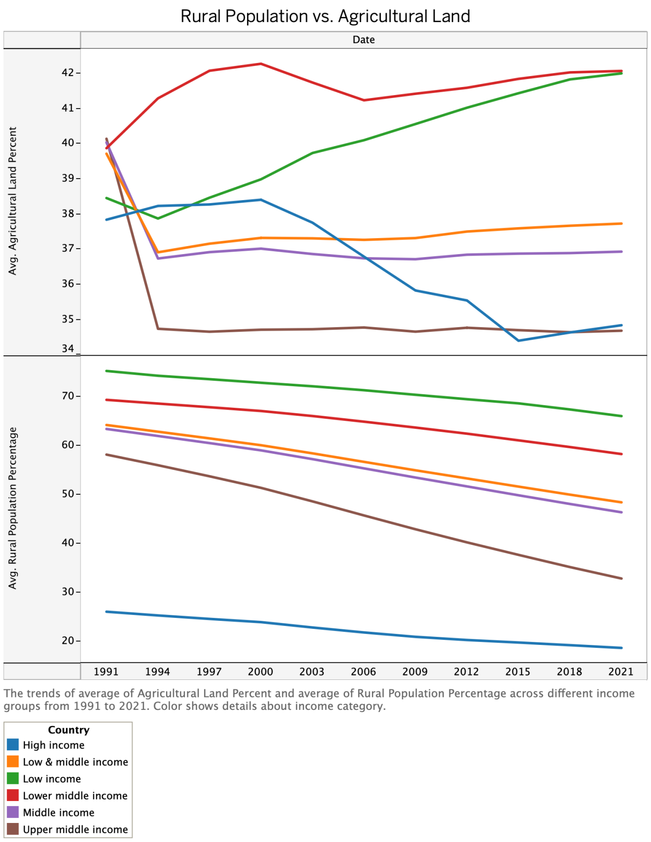


Fig 25. Line Charts showing the relationship between Rural Population Percentage and Agricultural Land Percentage across income groups (1991–2021)

Figure 25 presents 2 line charts depicting trends in percentages of population living in rural areas and percentages of agricultural land out of total land for country categories based on income levels: high income, middle income, lower middle income, and low income. The data spans from 1991 to 2021, a period chosen because agricultural land percentage prior to 1991 was either limited or non-existent in many countries.

Rural Population Trends: The rural population has decreased across all income groups due to urbanization and economic shifts.

Agricultural Land Area:

- Low-income and lower-middle-income countries: These countries have seen an increase in agricultural land area as they rely heavily on agriculture for sustenance and economic growth.
- Middle-income countries: Consistent agricultural land usage is observed due to balanced rural-to-urban migration and stable agricultural needs.
- High-income countries: Agricultural land has decreased, reflecting technological advancements, a shift towards imports, and efforts toward environmental preservation.

The sudden drop in agricultural land area around initial 1990s can be attributed to several factors. One key reason is land reclassification, where land previously categorized as agricultural was redefined for urban development or conservation purposes. Additionally, increased environmental regulations during this period, driven by international agreements and policies aimed at reducing deforestation and preserving biodiversity, led to a decrease in agricultural land. Lastly, a global shift toward industrialization and urbanization contributed to less reliance on domestic agricultural land, particularly in high-income countries, as economies moved toward service and industrial sectors and agricultural imports increased.

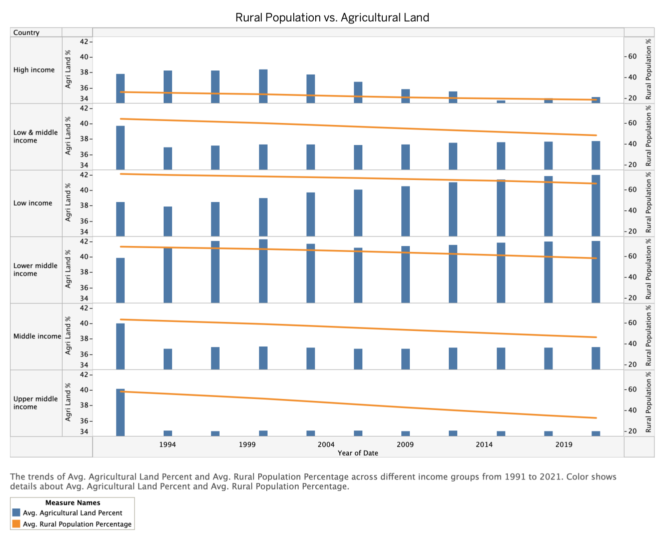


Fig 26. Dual combination showing the relationship between Rural Population Percentage and Agricultural Land Percentage across income groups over the years (1991–2021)

Figure 26 presents a similar analysis of the relationship between rural population percentage and agricultural land area across different income groups, extending the insights observed in Figure 25. The trends depicted in Figure 26 reinforce the conclusions drawn earlier: while rural populations depicted by the orange line across the years continue to decline across all income groups, the changes in agricultural land area depicted by the bars follow comparable patterns. Low-income and lower-middle-income countries continue to see an increase in agricultural land, reflecting their reliance on agriculture. In contrast, middle-income countries show stable land use, while high-income countries experience a further decrease in agricultural land. These trends reinforce the global patterns in rural land use and agricultural dependence. Also, the general global drop in agricultural land area percentage in early 1990s can also be seen here.

Hypothesis 2: Analyzing Per Capita CO_2 Emissions and Power Consumption Trends Across Different Income Groups

In this hypothesis, we explore the differences in resource utilization and the environmental impact of individuals across various income groups. By examining these differences, we aim to understand how varying levels of income influence the per capita consumption of resources and the associated CO_2 emissions. The analysis also investigates whether increased electric power consumption correlates with higher CO_2 emissions and how this relationship varies based on the economic status of countries.

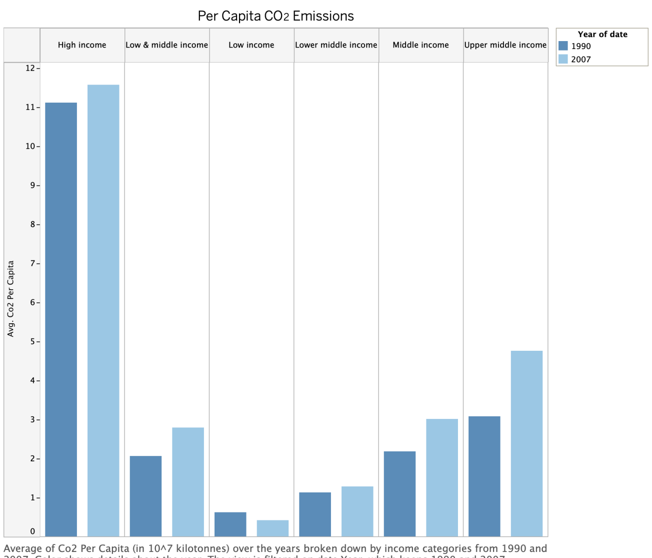


Fig 27. Double bar chart representing the per capita CO_2 emissions for individuals across different income groups (1990 and 2007)

Figure 27 and 28 present two double bar charts illustrating per capita CO_2 emissions and electrical power consumption across different income groups for the years 1990 and 2007. The first chart shows CO_2 emissions per capita for individuals categorized by income levels: high income, middle income, lower middle income, and low income. The second chart depicts per capita electrical power consumption for the same income categories and years. The year 1990 was selected as emission data was not able for years preceding 1990 and 2007 was chosen to provide a comparative view of trends over a significant period, highlighting changes and patterns in resource utilization and environmental impact across different income levels.

It can be seen that per capita CO_2 emissions and electric power consumption have generally increased across all income groups from 1990 to 2007. This trend indicates a strong relationship between higher per capita power consumption and increased CO_2 emissions across all groups apart from the exception of low-income countries experiencing a decrease in per capita CO_2 emissions despite some increase in power consumption.

One possible reason for this exception could be rapid population growth in these countries, which may have mitigated the per capita impact of increased power consumption. Additionally, the limited industrial development in low-income countries may result in lower CO_2 emissions despite rising power consumption, as fewer industrial activities typically contribute less to overall emissions.

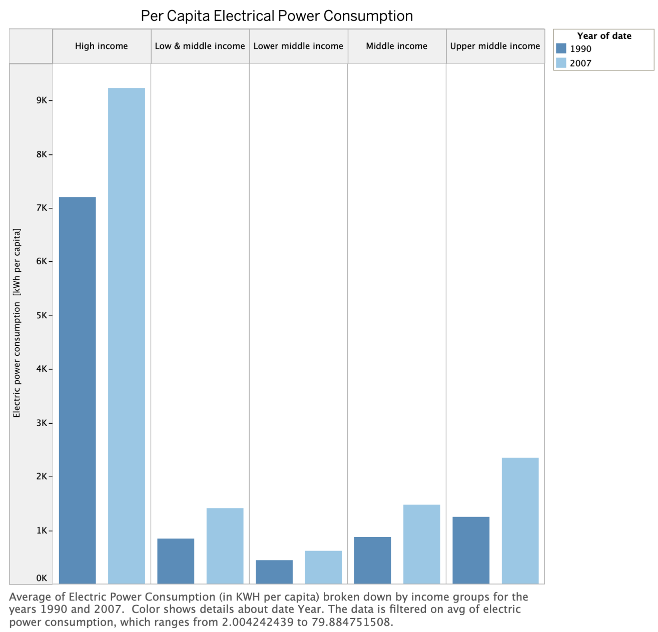


Fig 28. Double bar chart representing the per capita electrical power consumption for individuals across different income groups (1990 and 2007)

A comparative study shows that individuals in high-income countries consume significantly more electricity per capita compared to those in low-income countries. However, this increased consumption is associated with higher CO_2 emissions. In contrast, while individuals in low-income countries consume less electricity, their CO_2 emissions per capita have decreased, reflecting differing levels of environmental impact and resource utilization between high-income and low-income individuals.In high-income countries, urbanization often leads to increased energy use per capita, with more energy-

consuming infrastructure and higher living standards. In contrast, low-income countries may still be in stages of development where energy consumption is lower and spread across less energy-intensive activities.
A person in a high-income economy uses more than 12 times as much energy on average as a person in a low-income economy

Hypothesis 3: Geospatial Analysis of CO_2 and Greenhouse Gas Emissions and Electrical Power Consumption Across Countries

In this hypothesis, we conduct a geospatial analysis to explore global patterns in CO_2 emissions, greenhouse gas emissions, and electrical power consumption. By visualizing these variables on symbol maps, we aim to examine potential correlations between different types of environmental emissions, such as CO_2 and greenhouse gases, and to investigate any direct relationships with electrical power consumption. This analysis will reveal how these environmental and energy factors vary across regions and highlight any notable geographic trends.

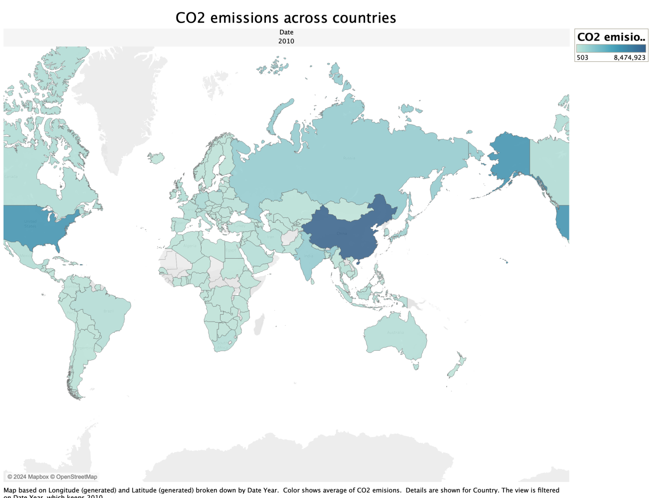


Fig 29. Map representing CO_2 emissions across countries (2010)

Figure 29 and 30 present two maps illustrating CO2 emissions and other greenhouse gas emissions across different countries for the year 2010. The year 2010 was selected arbitrarily. The CO2 emissions represented on the map are measured in kilotonnes (kt). By visualizing these emissions in kilotonnes, the map provides a clear perspective on the scale and distribution of CO2 emissions across different regions, making it easier to compare the environmental impact of various countries.From the symbol maps depicting CO2 emissions and greenhouse gas emissions, we observe a similar trend in both datasets.

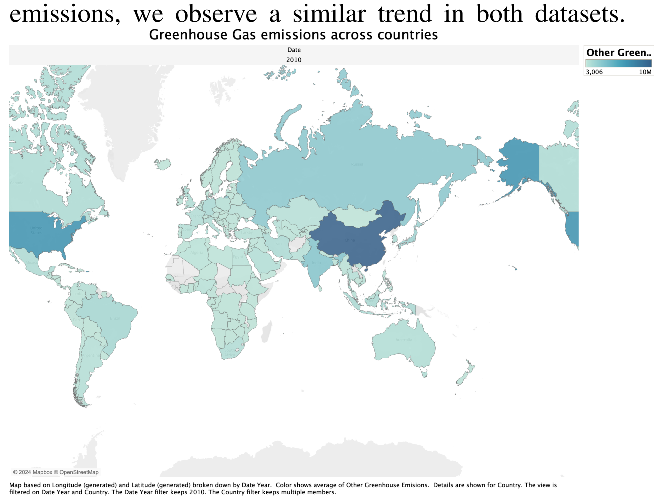


Fig 30. Map representing other greenhouse gas emissions across countries (2010)

This parallel trend suggests that countries exhibiting higher CO2 emissions also tend to have higher levels of greenhouse gas emissions. The analysis indicates that both types of emissions are closely correlated, likely reflecting common underlying factors such as industrial activity, energy consumption patterns, and economic development. Furthermore, the maps highlight that larger economies and more industrialized nations tend to be the largest contributors to global emissions. Countries with lower emissions, on the other hand, are often those with less industrialization and smaller populations, pointing to the ongoing global challenge of balancing development with sustainability..

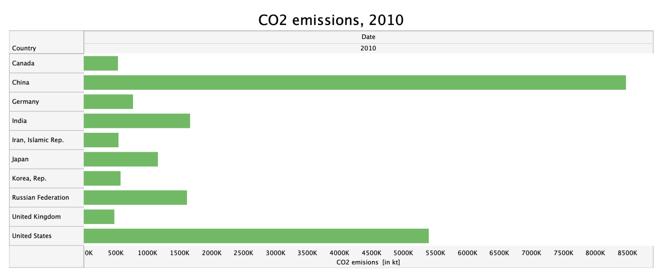


Fig 31. Top 10 largest contributing countries to CO_2 emissions (2010)

Figure 31 presents a horizontal bar chart illustrating top 10 countries with the highest CO_2 emissions for the year 2010. This visual representation highlights the dominant contributors to global carbon dioxide emissions, providing a clear comparison among the top emitters. The symbol map depicting CO_2 emissions across countries for the year 2010 and the bar chart showing the top 10 largest contributing countries to CO_2 emissions provide a complementary view of global emission patterns. The map reveals that darker countries, which signify higher levels

of CO_2 emissions, align closely with those highlighted in the bar chart as the top emitters.
The top 4 darkest countries in the map which are also represented by top 4 longest bars in the horizontal bar chart in their decreasing order of emission for the year 2010 are :
China, United States of America, India and Russian Federation.

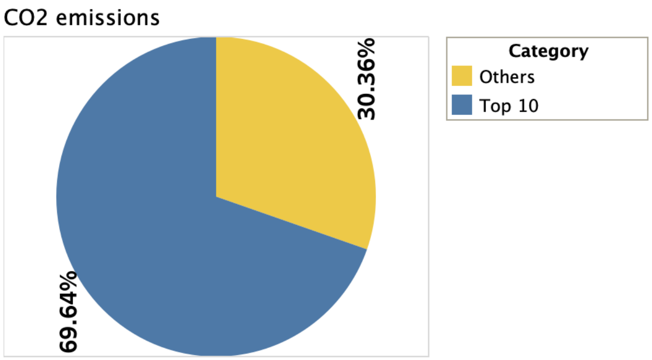


Fig 32. Share of global CO emissions by the top 10 emitting countries compared to the rest of the world (2010)

Figure 32 presents a pie chart showing the percentage share of top 10 CO_2 emitting countries as a part of total CO_2 emissions for the year 2010.These countries are China, USA, India, Russia, Japan, Germany, Canada, Korea, Iran and UK.

The 10 largest contributors to CO emissions account for approximately 70% of the global total, underscoring the significant concentration of emissions within a relatively small number of countries. This disproportionate share of global emissions highlights the crucial role these nations play in shaping the trajectory of global climate change.

By emitting the majority of CO_2, these countries are pivotal in driving global warming and environmental degradation. Addressing climate issues, therefore, hinges on the ability of these top emitters to implement more sustainable practices, adopt cleaner technologies, and commit to substantial emissions reductions. Their actions will be critical in the global effort to mitigate the adverse effects of climate change.

Figure 33 presents a pie chart depicting the percentage share of other greenhouse gas emissions contributed by the top 10 CO_2 emitting countries as part of the total global emissions for 2010. These same 10 countries also account for approximately 64% of total other greenhouse gas emissions, further reinforcing the patterns observed in the maps showing CO_2 and greenhouse gas emissions across all countries. This correlation highlights the significant contribution of the top CO_2 emitters to overall global greenhouse gas emissions, extending the insights gained from the map visualizations.

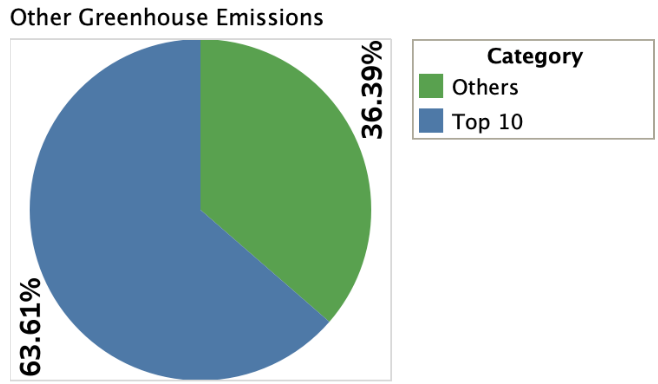


Fig 33. Share of global other greenhouse gas emissions by the top 10 CO_2 emitting countries compared to the rest of the world (2010)

Building on the general increase in per capita CO_2 emissions and per capita electrical power consumption observed across different income groups in Hypothesis 2, we now aim to examine this relationship across individual countries. This will be achieved through geospatial mapping of electrical power consumption, allowing us to explore the correlation between power consumption and CO_2 emissions on a global scale.

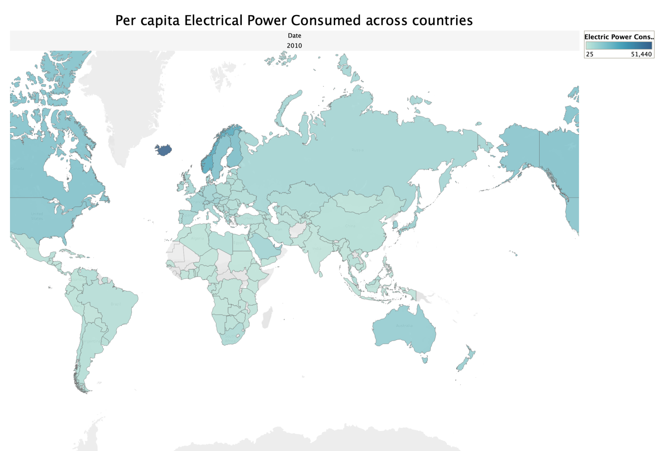


Fig 34. Map representing per capita electrical power consumed across countries (2010)

Figure 34 presents a map illustrating per capita electrical power consumption across countries for the year 2010. Contrary to initial expectations, this map does not directly correlate with the maps depicted in Figures 29 and 30, which show CO_2 and greenhouse gas emissions. This suggests that countries with higher per capita electrical power consumption do not necessarily contribute proportionally to higher per capita greenhouse gas emissions. Notably, regions like Iceland, Norway, and Canada are shaded darker, indicating much higher per capita power consumption. This trend suggests that countries with abundant energy resources, such as geothermal energy in Iceland or hydroelectric power in Norway and Canada, consume much more electricity on an individual basis compared to others. These nations often rely on clean and renewable energy sources, which explains their high energy usage but relatively lower greenhouse gas emissions, a trend that contrasts with the expected correlation between higher energy consumption and emissions.

In contrast, countries in Africa and parts of South Asia, such as Nigeria and India, display much lighter shades, indicating lower per capita power consumption. This reflects the limited access to electricity or lower overall energy usage in these regions due to slower industrial development or less extensive energy infrastructure. However, while these countries consume less electricity per person, some, like India, still contribute significantly to global CO_2 and greenhouse gas emissions due to their reliance on coal and other non-renewable energy sources for power generation.

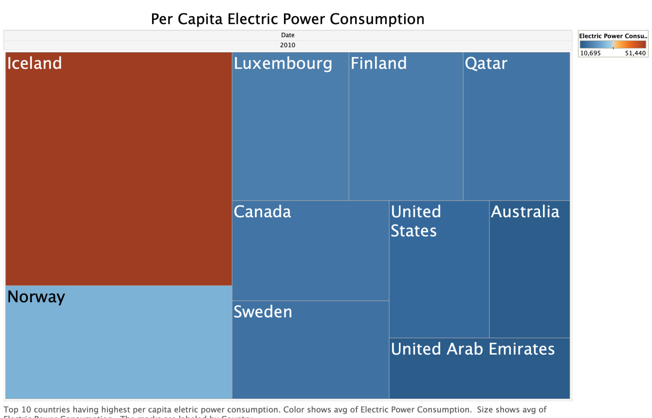


Fig 35. Tree Map displaying the top 10 countries with the highest per capita electrical power consumption (2010)

Figure 35 presents a tree map representing the top 10 countries by per capita electrical power consumption in 2010. The size of each block reflects the relative contribution of each country, with larger blocks signifying higher per capita consumption. Notably, the color gradient difference between Iceland, represented in brown, and Norway, shaded in blue, underscores a significant disparity — Iceland's per capita consumption is nearly double that of Norway, the second-highest consumer. This visualization highlights stark contrasts in energy usage among the top consumers, emphasizing the global imbalance in power consumption levels.

Several factors may contribute to Iceland's exceptionally high per capita electricity consumption:

- Abundant renewable energy sources: Iceland's electricity production is almost entirely powered by renewable sources, specifically geothermal and hydropower. This abundance of low-cost, sustainable energy likely encourages higher consumption levels across various sectors.
- Presence of energy-intensive industries: The country's economic structure includes several energy-intensive industries. Aluminum smelting, in particular, is a major electricity consumer. Additionally, Iceland hosts other sectors such as silica mining, pumice extraction, and extensive geothermal resource utilization. These industries benefit from the country's low-cost renewable energy, driving up overall electricity demand.
- Small population base: Iceland's relatively small population may also inflate its per capita electricity consumption figures. A large total consumption, driven by industrial use, when divided by a small population, results in a disproportionately high per capita consumption.
- Heating requirements: Due to Iceland's cold climate, electricity is heavily utilized for heating, including space heating and hot water systems. This further contributes to the high electricity consumption per person.
- Economic development and structure: In addition to aluminum smelting, other sectors such as data centers, which require vast amounts of electricity, take advantage of Iceland's inexpensive energy supply. This combination of economic activities focused on energy-intensive industries amplifies the overall electricity demand.

Major Challenges: A major challenge was the absence of data for many critical attributes, with several key columns filled with Null or missing values. This incomplete data, especially across various regions, complicated efforts to draw meaningful conclusions. Analyzing trends often required multiple variables, but for many countries, essential data points were lacking, making it even more difficult to identify clear patterns.

III. AUTHORS' CONTRIBUTIONS

Ayush Arya Kashyap: (Task 1: Analyzing Trends and Impacts of major economic events across the globe over the years)

- Conducted data analysis using Jupyter Notebook, Python, and libraries like pandas, datetime, os, among others to identify trends in economic activities in and around the world over the years.

- Visualizations Used : bar chart, stacked bar chart, treemaps, heatmaps area chart, line chart, and bubble diagram to represent the data effectively.
- Interpreted the visualizations and drew conclusions about the overall trends in economic activities.
- Wrote the narrative and descriptions for Task 1, providing context and insights.

Uttam Hamsaraj: (T2: Analyzing key social indicators across countries to uncover trends and insights into global social development)

- Preprocessed the data using tableau prep and analyzed data on social indicators.
- Analyzed and visualized 4 hypotheses with inferences.
- Visualizations Used : Stacked bar chart , Bubble chart , Bar chart , Scatter plot, Density plot, Treemap , Line plot , Heatmap , Dual-axis line graphs.
- Wrote the narrative and descriptions for Task 2, providing context and insights.

Pranav Laddhad: (Task 3: Analyzing trends and relations of environmental and sustainability indicators to evaluate global progress in addressing climate change)

- Preprocessed and analyzed data on environmental indicators using python libraries and tabeau prep.
- Analyzed and visualized 3 hypotheses with inferences.
- Visualizations Used : lines charts, dual combinations chart, double bar charts, cartographic charts, horizontal bar chart, pie charts and tree map to represent the data effectively.
- Interpreted the visualizations and drew conclusions
- Wrote the narrative and descriptions for Task 3, providing context and insights.

IV. References

1) https://www.kaggle.com/datasets/nicolasgonzalezmunoz/world-bank-world-development-indicators
2) https://databank.worldbank.org/source/world-development-indicators
3) https://www.unodc.org/documents/data-and-analysis/Crime-statistics/Global-Burden-of-Armed-Violence-full-report.pdf
4) https://tradingeconomics.com/indicators
5) https://documents.worldbank.org/en/publication/documents-reports