UNIVERSITY OF HOUSTON

EDS 6397 – INFORMATION VISUALIZATION

Fall 2023

Professor: Lucy Nwosu

Group - 4

Authors:

2205379 Akash Borigi

2250744 Siddharth Reddy Enjam

2249602 Chandra Prakash Rao Boinapally

2251153 Mukesh Reddy Mavurapu

2249009 Venkata Satya Reddy Kallam

2251590 Venkata Surya Narayana Raju Mantena

2251160 Hari Phalguna Reddy Pulla

2190756 Ravi Teja Thandra

2250576 Shiva Sai Ram Prasad Reddy Yelipeddi

INTRODUCTION

Objective:

In an increasingly interconnected world, the factors that contribute to a nation's development and the well-being of its citizens are of paramount importance to policymakers, economists, and social scientists. The quest to understand the multifaceted nature of economic development and health indicators necessitates a comprehensive exploration of global country characteristics. This report aims to dissect the interplay between a wide array of development indicators and how they correlate with economic prosperity and health outcomes across different regions.
Utilizing the " Countries_of_the_world " dataset, we embark on an empirical journey to unravel the complex relationships between economic metrics, such as Gross Domestic Product (GDP) per capita, and a spectrum of health indicators, including infant mortality, birth rates, and death rates. Our analysis extends to understanding the impact of technology and education, measured through variables such as the number of phones per 1000 people and literacy rates, on a country's economic status.

As we proceed, we will consider the historical context and current global dynamics that shape these indicators, while also pondering the influence of factors such as climate, agricultural viability, and technological advancement. The insights garnered from this investigation will not only shed light on the current state of global development but also pave the way for future research and policy formulation aimed at enhancing economic and health outcomes worldwide.

This introduction sets the stage for a detailed analysis and encourages a multidimensional approach to understanding development indicators. The subsequent sections of your report will delve into specific analyses as outlined in your study's objectives.

Dataset Overview:

The dataset contains various indicators for countries around the world, including:

- Country Name
- Region
- Population
- Area (sq. km)
- Population Density (people per sq. km)
- Coastline Ratio (coast/area)
- Net Migration (migrant(s)/1,000 population)
- Infant Mortality (per 1,000 births)
- GDP per Capita ($)
- Literacy (%)
- Phones (per 1,000 people)
- Percentage of land arable
- Percentage of land under crops
- Percentage of land other
- Climate (1-4)
- Birthrate (per 1,000 people)
- Deathrate (per 1,000 people)

- Agriculture (percentage of GDP)
- Industry (percentage of GDP)
- Service (percentage of GDP)

Data Cleaning:

- Renaming the columns with more readability
- Converting the decimal values to international system from European system.
- Converting the datatype from object to float
- Filling the Missing values according to the mean of the specific region.
- Climate is classified into 4 parts and missing values are rounded off to its nearest integer.
- Saving the Data frame to a .csv file

Data Visualization:

Data Exploration & Comparative Visualizations:

1. Regional Comparison of GDP and Literacy Rates:
    - Visualize the distribution of GDP and Literacy across different regions.
    - Identify the regions with highest, lowest GDP and Literacy rate.
2. Impact of Agricultural and Industrial Sectors:
    - Create a scatter plot across different sectors.
    - Identify the countries with rich in different sectors.
3. Health Indicators and Net Migration:
    - Visualize using choropleth map that indicates health indicators, Net migration across different countries.
    - Identify the countries with highest, lowest health indicators and Net migration.
4. Climate and Arable Percentage:
    - Visualize using tree map that indicates different climate conditions and Arable percentage of land.
    - Identify the country with highest, lowest Arable percentage of land.
5. Technology and communication:
    - Visualize using choropleth map that indicates advancement of Technology over economic prosperity, Literacy percentage.
    - Identify the countries with lowest, highest Technology usage, Literacy rate.

Axis Labelling:

Use a consistent scale for both axes to avoid distortion, ensure proper scaling to accurately represent the distribution. High-resolution graphics for legibility, Minimal use of ink with clear data representation. Clear labelling to minimize confusion, Ensure clear legends for both indicators, High-resolution map for legibility. Use distinct colours for each climate condition, Proper scaling for accurate representation. Use a proper colour scale for technology usage.

Use of Colour, Size and Shape:

Use different colours for each region to distinguish them (example Use different colours for each region to distinguish them), Use a colour gradient for health indicators, with darker colours indicating poorer health. Represent the size of each block based on the arable percentage of land, Represent the size of each point based on the country's GDP or another relevant metric. Use different shapes to represent countries from different regions, show

detailed information about climate conditions and arable percentage when hovering over each block.

Interactive Features:
Users can select over points to see specific region details for a more detailed comparison. Create an interactive scatter plot to display the distribution of countries based on Agricultural and Industrial sectors. Utilize a choropleth map to visualize health indicators and Net migration across countries. Use an interactive tree map to represent different climate conditions and Arable percentage of land. Employ a choropleth map to visualize the relationship between Technology usage, economic prosperity, and Literacy rates.

Hypothesis and Analysis Direction:

1. Regional Comparison of GDP and Literacy Rates:

Regional Comparison of GDP and Literacy Rates: Calculate the average GDP per capita and literacy rates for each region and analyze if regions with higher GDP per capita also have higher literacy rates.

Let's start with the first analysis. We will group the data by region and calculate the average GDP per capita and literacy rates for each region. Then, we will analyze the correlation between these two indicators.

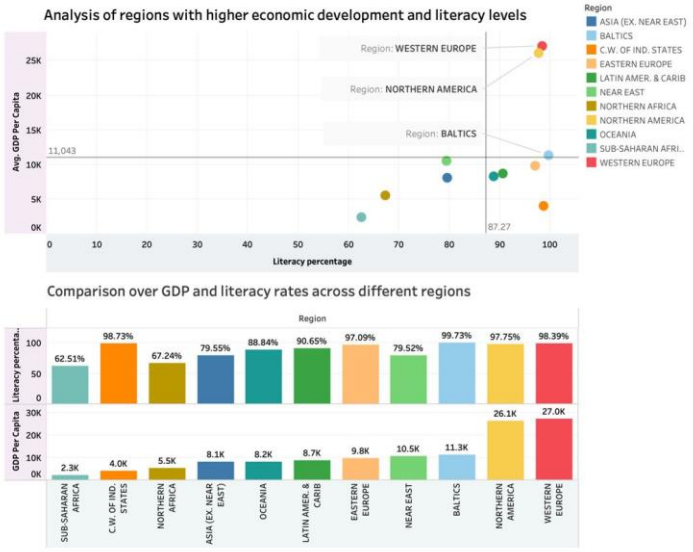
The regional comparison of GDP per capita and literacy rates reveals that Western Europe has the highest average GDP per capita, followed by Northern America and the Baltics. These regions also have high literacy rates, suggesting a positive correlation between economic development and literacy levels.

Specifically, the average GDP per capita and literacy rates for each region are as follows:

- Western Europe: $27,046.43 with an average literacy rate of 98.39%
- Northern America: $26,100.00 with an average literacy rate of 97.75%
- Baltics: $11,300.00 with an average literacy rate of 99.73%
- Near East: $10,456.25 with an average literacy rate of 79.52%
- Eastern Europe: $9,808.33 with an average literacy rate of 97.09%
- Latin America & Caribbean: $8,682.22 with an average literacy rate of 90.65%
- Oceania: $8,247.62 with an average literacy rate of 88.84%
- Asia (Excluding Near East): $8,053.57 with an average literacy rate of 79.55%
- Northern Africa: $5,460.00 with an average literacy rate of 67.24%
- Commonwealth of Independent States: $4,000.00 with an average literacy rate of 98.73%
- Sub-Saharan Africa: $2,323.53 with an average literacy rate of 62.51%

The correlation coefficient between GDP per capita and literacy rates across regions is approximately 0.544, indicating a moderate positive correlation. This suggests that regions with higher economic development tend to have higher literacy levels, but the relationship is

not perfect and there are variations within regions.



2. Impact of Agricultural and Industrial Sectors:

Examine the correlation between GDPs per capita and the contributions of the agricultural, industrial, and service sectors to the economy across countries?

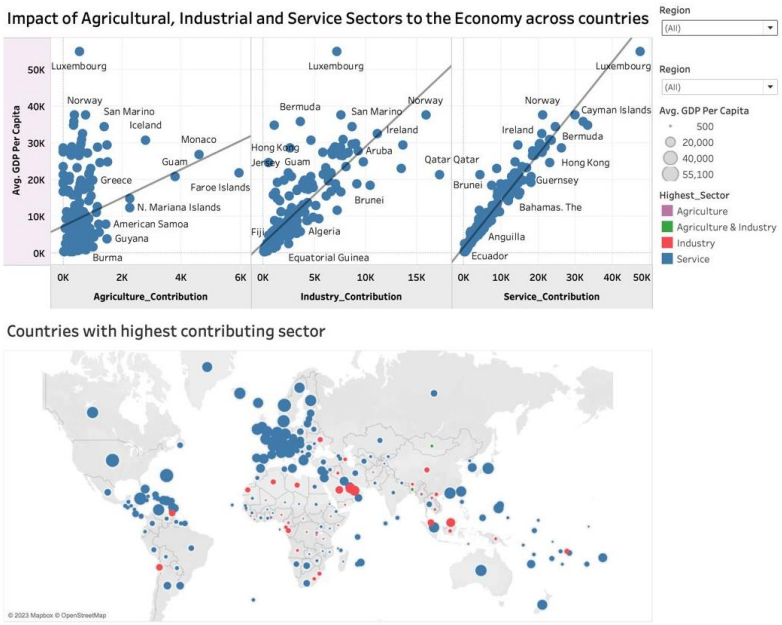
Now, let's analyze the contribution of the agricultural, industrial, and service sectors to the economy across countries and their correlation with the country's economic development (GDP per capita).

The analysis of the contributions of the agricultural, industrial, and service sectors to the economy in correlation with GDP per capita yields the following insights:

- There is a moderate negative correlation between the agricultural sector's contribution to GDP and GDP per capita, with a correlation coefficient of approximately -0.585. This suggests that as countries become more economically developed, the relative contribution of agriculture to their GDP tends to decrease.
- The industrial sector shows a very weak negative correlation with GDP per capita, with a correlation coefficient of about -0.039. This indicates that the contribution of the industrial sector to the economy does not have a clear relationship with the economic development of a country.
- The service sector has a moderate positive correlation with GDP per capita, with a correlation coefficient of approximately 0.552. This suggests that in more economically developed countries, the service sector tends to make up a larger portion of the GDP.

These findings indicate that there is a trend towards a lower reliance on agriculture and a higher reliance on services in more economically developed countries. This data supports the notion that as nations develop economically, there is a shift from agricultural and industrial production

to service-based economies.



3. Health Indicators and Net Migration:

Explore the relationship between net migration and health indicators like infant mortality, birth rates, and death rates. Are countries with high net migration rates associated with better or worse health indicators?

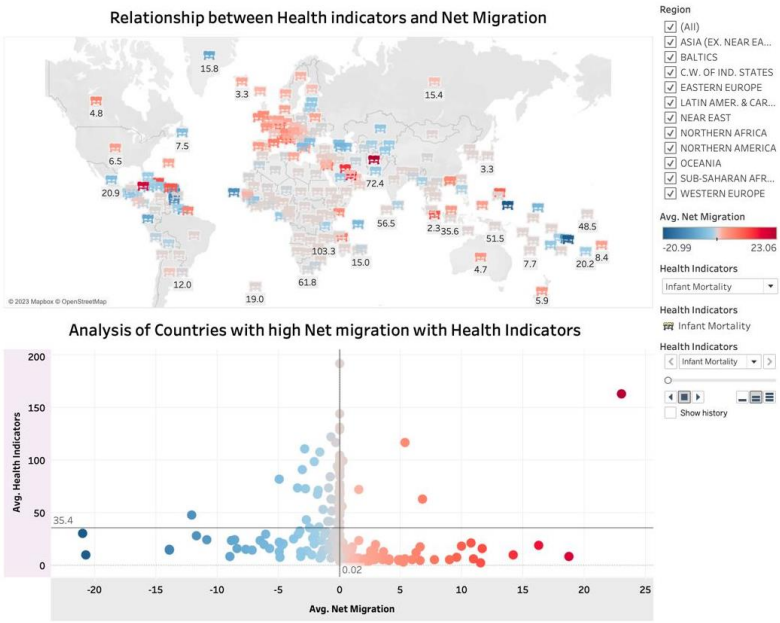
This report explores the correlations between net migration and key health indicators, namely infant mortality, birth rates, and death rates, across different countries. Our findings reveal that there is no strong correlation between net migration rates and the selected health indicators, suggesting that other factors may be more influential in a country's net migration trends.

We calculated the Pearson correlation coefficients between net migration and the three health indicators to assess the strength and direction of the relationships.

The correlation coefficients between net migration and the health indicators are as follows:

- Infant Mortality: -0.0233 (very weak negative correlation)
- Birthrate: -0.0648 (very weak negative correlation)
- Deathrate: 0.0362 (very weak positive correlation)

The analysis indicates that there is no substantial correlation between net migration rates and the health indicators studied. It implies that the quality of health, as indicated by infant mortality, birth rates, and death rates, is not a primary factor in the net migration rates observed across countries.



4.Climate and Agricultural Viability:

Investigate the relationship between climate attributes and arable land. How do different climates affect a country's agricultural potential?

Utilizing a dataset with climate classifications and arable land percentages, we analysed how different climates correlate with a country's agricultural potential. Our findings suggest that continental and temperate climates, which tend to have more significant seasonal variations and moderate weather patterns, are associated with higher percentages of arable land. In contrast, tropical and dry climates, which may face challenges such as excessive rainfall or lack of it, show a lower percentage of arable land.

Methodology: The dataset, " Countries_of_the_world "was used for the analysis. Climate types were categorized, and the average arable land percentages were computed for each climate category. A correlation analysis was performed to understand the strength of the relationship between climate types and arable land.

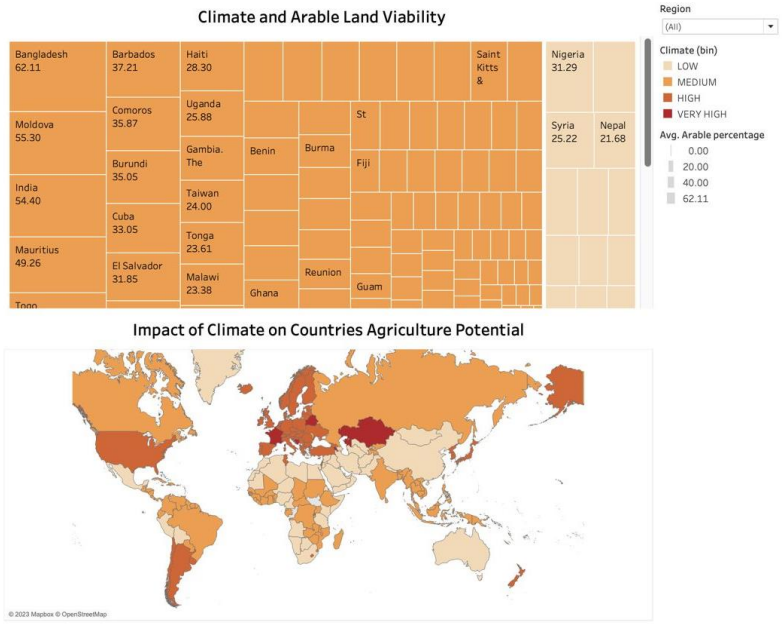
The analysis yielded the following average percentages of arable land for each climate type:

1. Tropical: 7.66%
2. Dry: 12.37%
3. Temperate: 20.50%
4. Continental: 21.20%

The correlation coefficient between climate types and the percentage of arable land is approximately 0.347, indicating a moderate positive correlation.

Continental climates, characterized by a considerable temperature range between summer and winter, show the highest average arable land, which may suggest a favorable condition for diverse crop production during the growing seasons. Temperate climates, with their moderate

conditions, also exhibit a high percentage of arable land conducive to agriculture. However, the lower percentages in tropical and dry climates could be due to factors such as excessive precipitation, forest cover, soil quality, and water scarcity, which may limit the extent of usable agricultural land.



The study indicates that climate attributes significantly influence a country's agricultural land viability, with continental and temperate climates showing a higher potential for arable land. While a moderate positive correlation exists, it also highlights that other factors, perhaps socio-economic or geo-specific agricultural practices, contribute to agricultural viability.

5. Technology and Communication:

Examine the correlation between the number of phones (per 1000 people) and literacy rates or GDP. Does technological advancement correspond to economic prosperity or literacy rates?

Our analysis indicates a significant positive correlation between the number of phones per 100 people and both GDP per capita and literacy rates. This suggests that countries with greater access to technology tend to have higher economic prosperity and literacy rates.

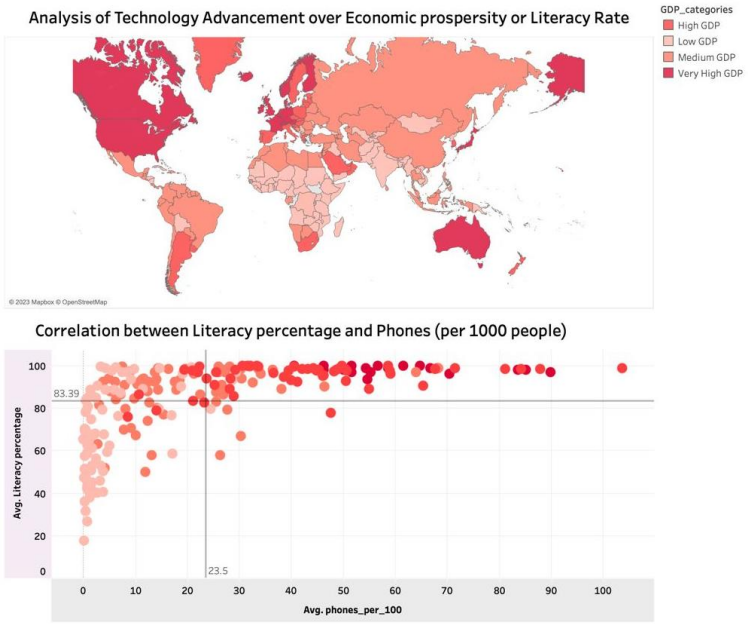
Pearson correlation coefficients were computed to assess the relationships between the number of phones per 100 people and the two key indicators: GDP per capita and literacy rates.

The correlation analysis revealed the following:

- A moderate positive correlation between the number of phones per 100 people and literacy rates, with a coefficient of 0.590.
- A strong positive correlation between the number of phones per 100 people and GDP per capita, with a coefficient of 0.831.

The moderate correlation with literacy rates suggests that technological accessibility is associated with higher literacy, potentially due to better access to educational resources. The strong correlation with GDP per capita implies that technological infrastructure is closely linked to economic health, supporting the idea that technology is a key factor in economic

growth and development.



Ethical Considerations:

We made sure our charts are honest and clear so you can trust the information. We didn't try to trick anyone by changing things or using weird scales. In each chart, we explain if there are any limits or possible mistakes in the data, so you know what to keep in mind. We also give you tips on how to read the charts with clear labels and notes. The interactive parts are there to help you explore the data, but we made sure it won't confuse you. We want to give you the power to think for yourself by being open about where we got the data and any issues it might have. That way, you can make your own smart decisions based on the information.

User Experience and Engagement:

From our Tableau dashboard story, here are the instances where we have included to enhance user-friendliness and interactivity, following Colin Ware's principles of engagement:

Regional Comparison of GDP and Literacy rate:

The size and colour of each region on the map represent the average literacy rate, facilitating a quick visual assessment of the distribution of literacy levels. Larger and more vibrant regions indicate higher average literacy rates, contributing to an intuitive understanding of the data.

Impact of agricultural, service, and industrial sector:

The inclusion of interactive parameters, such as a dropdown for regions and distinct colours for sectors, enhances the exploration of the dataset. A dropdown parameter for regions enhances interactivity, allowing users to select specific regions for a focused analysis. Choropleth plots use distinct colours for each economic sector, providing a clear visual representation of the contributions of Agriculture, Industry, and Service.

Health Indicator and net migration:

A parameter is created for the three health indicators—birthrate, deathrate, and infant mortality—allowing users to dynamically switch between them. A dropdown menu complements the parameter, enabling users to select their preferred health indicator for a more customized exploration. Whenever users select a health indicator from the dropdown, both scatter and choropleth plots dynamically change, providing an interactive and personalized experience.

Climate and arable land variability:

This geospatial representation focuses on the relationship between arable land and climate classes, utilizing both choropleth and tree map visualizations. The interactive nature of the representation is enhanced through the incorporation of a dropdown filter based on regions, and the use of colour saturation to distinguish between different climate categories. The inclusion of a dropdown filter based on regions allows users to focus on specific geographical areas, providing a more tailored exploration of arable land and climate relationships.

Technology and Communication:

The visualization employs a choropleth map to categorize GDP into four distinct categories, utilizing colour saturation for differentiation. Additionally, a scatter plot is included to showcase the correlation between literacy percentage and the prevalence of phones. Mean indicators for both axes are displayed, providing additional insights. Different levels of colour saturation are used to differentiate between the GDP categories, providing users with a clear and intuitive understanding of the economic landscape.

CONCLUSION:

Our investigative analysis within Tableau has offered a panoramic view of the factors influencing global development. We've uncovered a positive correlation between economic development and literacy, signaling the transformative power of education. The shift from agricultural reliance to service-oriented economies marks a transition in economic maturity. Health metrics and net migration painted a more complex picture, suggesting multifaceted influences beyond mere economics. The interplay between climate and agriculture, as well as technology's reflection of economic health, were particularly telling. This report not only sheds light on the state of global development but also sets the stage for future policy and research to foster growth and improve health outcomes worldwide. With interactive Tableau visualizations, we've provided a canvas for readers to draw their insights, ensuring that our findings resonate with a diverse audience and inspire action.

https://public.tableau.com/app/profile/siddharth.reddy.enjam8608/viz/Group-4-Worksheets/Group-4Countries-Story?publish=yes