UNIVERSITY *of* WASHINGTON

Name: Baisakhi Sarkar
Course: DATA 511 Data Visualization for Data Scientists
Date: 11/04/2023
Professor: Nathan Mannheimer

Exploratory Visual Analysis – Child Labor

Introduction

Child labor is a pressing global issue that transcends cultural, economic, and social boundaries, impacting the lives of millions of children around the world. To better understand the various aspects of child labor and its prevalence, we embark on an exploratory data analysis of the World Data Indicators dataset dedicated to this crucial matter. This report serves as a preliminary investigation to gain insights into the factors and trends associated with child labor, as well as to inform future in-depth research and policymaking.
The World Data Indicators dataset is a comprehensive and reliable source of information, providing a rich repository of data from diverse countries and regions. It encompasses various socio-economic, educational, and demographic indicators that are essential in comprehending the complex dynamics of child labor. By delving into this dataset, we aim to uncover patterns, correlations, and potential determinants of child labor on a global scale.
In this report, we will embark on an analytical journey, using EDA techniques to unveil the hidden stories within the data. Our objectives are threefold:

- **Descriptive Exploration**: We will start by providing a comprehensive overview of child labor data, examining the distribution, trends, and disparities across countries and regions.
- **Identification of Correlations**: We will investigate potential relationships between child labor and various socio-economic, educational, and demographic variables. This will help us discern factors that might contribute to or mitigate child labor.
- **Geospatial Analysis**: Utilizing geographic data, we will create visualizations to pinpoint global hotspots and regions with higher prevalence of child labor, aiding in the identification of areas that require particular attention.

Data Profile

In this project, my primary focus was on a specific subset of data sourced from the World Bank's World Development Indicator dataset, which I obtained directly from The World Bank's official website (https://datacatalog.worldbank.org/dataset/world-development-indicators). The World Bank, headquartered in Washington, D.C., is a globally recognized institution established in 1944, with a core mission of offering financial and technical support to nations, with a particular emphasis on developing countries worldwide.

The World Bank stands as a trusted and up-to-date source for a comprehensive array of global development data, encompassing topics spanning from gender and health to economic growth and education. Their World Development Indicators (WDI) database receives regular updates on a quarterly basis, with new data releases occurring in April, July, September, and December. The dataset offers an extensive temporal coverage from 1960 to 2022. This substantial dataset is packaged in a zip file comprising six CSV files. However, for the purposes of this exploratory analysis, our focus will be primarily on the WDICSV.csv file. This file is quite extensive, measuring 188.5 MB in size, and it boasts around 1400+ indicators for every country. Within this dataset, we encounter a rich mixture of both qualitative and quantitative data, providing a broad spectrum of information for in-depth analysis. The dataset is divided into two main categories: categorical data and quantitative data. Categorical data encompass fundamental information like country name, country code, indicator name, and indicator code, which serve as identifiers and descriptors. On the other hand, the quantitative data primarily comprise numerical values, specifically, the indicator values themselves. These indicators cover an expansive array of topics, including but not limited to economic indicators like GDP, environmental data such as CO2 emissions, agricultural statistics, population figures, birth and mortality rates, educational metrics, school enrollment rates, income-related data, family employment ratios, and employer statistics. In essence, this dataset encompasses a diverse range of indicators that offer a comprehensive view of various aspects of global development and societal trends.
The World Development Indicators dataset, like any dataset, has its limitations. While the dataset is comprehensive, not all countries may have data available for every indicator. Some countries may have missing or incomplete data, which can limit the ability to make direct comparisons or draw comprehensive conclusions. The dataset often presents aggregated values at the country or regional level. This can obscure variations within countries or regions, potentially leading to oversimplified or misleading conclusions.

Question Exploration

“What are the key socio-economic, demographic, and geographic factors associated with the prevalence of **child labor** in different regions and countries, as revealed by the World Development Indicators dataset? We anticipate that there is a correlation between higher child labor rates and factors such as lower GDP per capita, lower educational enrollment rates, higher birth rates, compulsory education years and specific geographic regions. This hypothesis is based on existing research and common patterns observed in the context of child labor. Through exploratory data analysis, we aim to test and refine this hypothesis, shedding light on the nuanced relationships between these variables and child labor prevalence.”

To initiate the data preparation process for this analysis, my first step was to import the WDICSV.csv file into Tableau. Following that, I executed several data cleaning actions, including reorganizing field names to occupy the first row, restructuring the years from columns to rows using pivot techniques, and excluding any occurrences of "null" values. Additionally, I converted the Indicator Names column from a row-based structure to a column-based format, simplifying data comparison.

To commence my data exploration, my initial goal was to gain insights into global statistics related to Educational Attainment percentages. To achieve this, I began by arranging latitude and longitude (both generated fields) in the rows and columns to create a world map view (Fig 1.1).

Subsequently, I crafted a Dashboard that displays the average percentages of Educational Attainment up to the bachelor's level, as well as at least lower and post-secondary levels. The color scheme on the map is clearly explained through Tableau legends positioned at the top and bottom. My decision to prioritize the examination of Education stems is just because by analyzing this data, we can gain valuable insights into how educational levels influence various aspects, including income, employment opportunities, and other socioeconomic outcomes.
The indicators that I chose for the below visualization are "Educational attainment, at least Bachelor's or equivalent, population 25+ total (%) (cumulative)", "Educational attainment, at least completed lower secondary, population 25+ total (%) (cumulative)" and "Educational attainment, at least completed post-secondary, population 25+ total (%) (cumulative)". For the second Map, I have used Dual-axis to plot two indicators in the same map. The percentage of educational attainment till post-secondary is encoded by the size of the pink dots on the second map (larger dots means higher %) (Tableau legends shown in Fig 1.2). Color saturation serves as the visual indicator for educational attainment till bachelor's (left map) and till lower secondary (right map), with a darker color signifying a higher rate of education attainment.
A fundamental insight derived from the visualization below is that the percentage of educational attainment tends to decline as we progress from lower secondary to post-secondary and bachelor's levels. This observation suggests that, in most countries, there is a prevailing trend of individuals concluding their education shortly after reaching the secondary level.

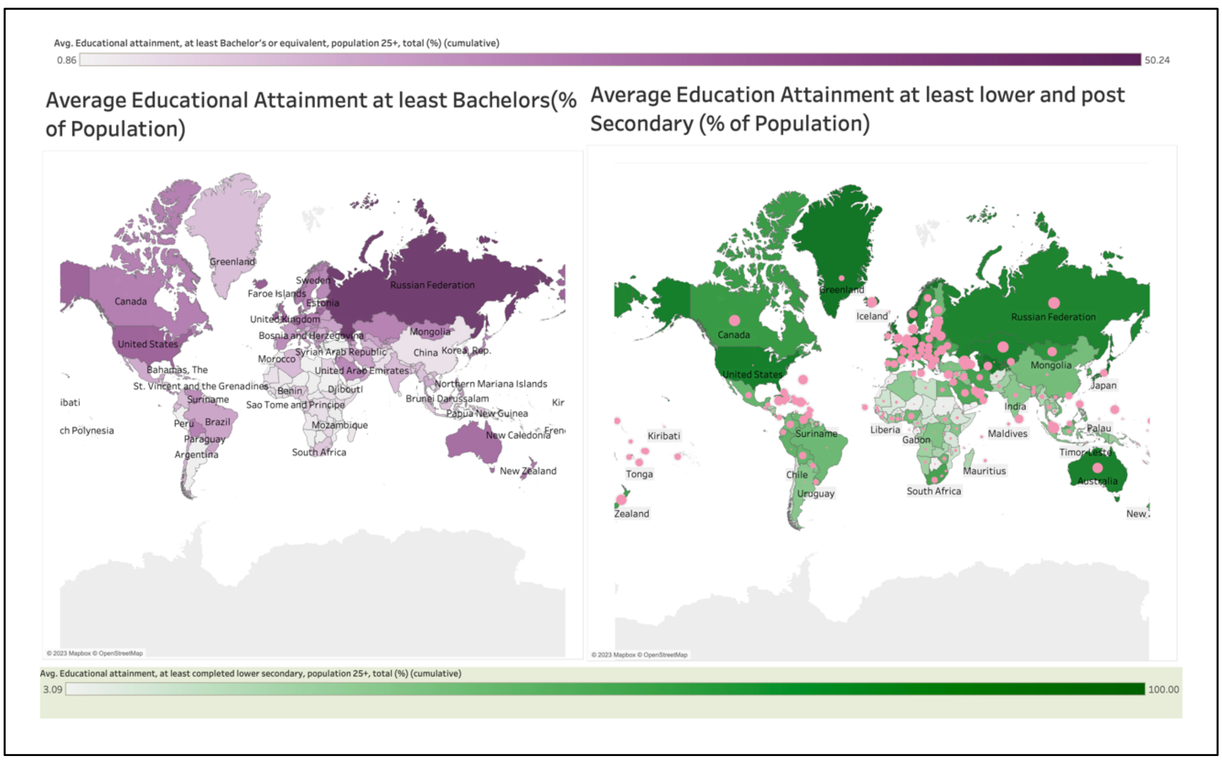

Fig1.1: Worldwide View of Average Educational Attainment till lower secondary, post-secondary and Bachelor's

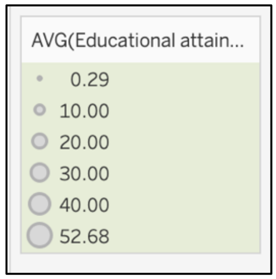

Fig1.2: Tableau Legends for Fig1.1 (right map, pink circle indicators) for Average educational attainment till post-secondary

In our subsequent visualization (Fig 1.3), by using Tableau's mapping functionality, we've created a detailed representation of the Average Poverty Headcount Ratio at National Poverty Lines (% of the population) and the Average Adjusted Net National Income per capita (current US$). We have once more depicted this data on a world map to enhance visual comprehension. The correlation between these two factors appears to be quite pronounced in this representation. Both these factors play a crucial role in shaping our ultimate objective.

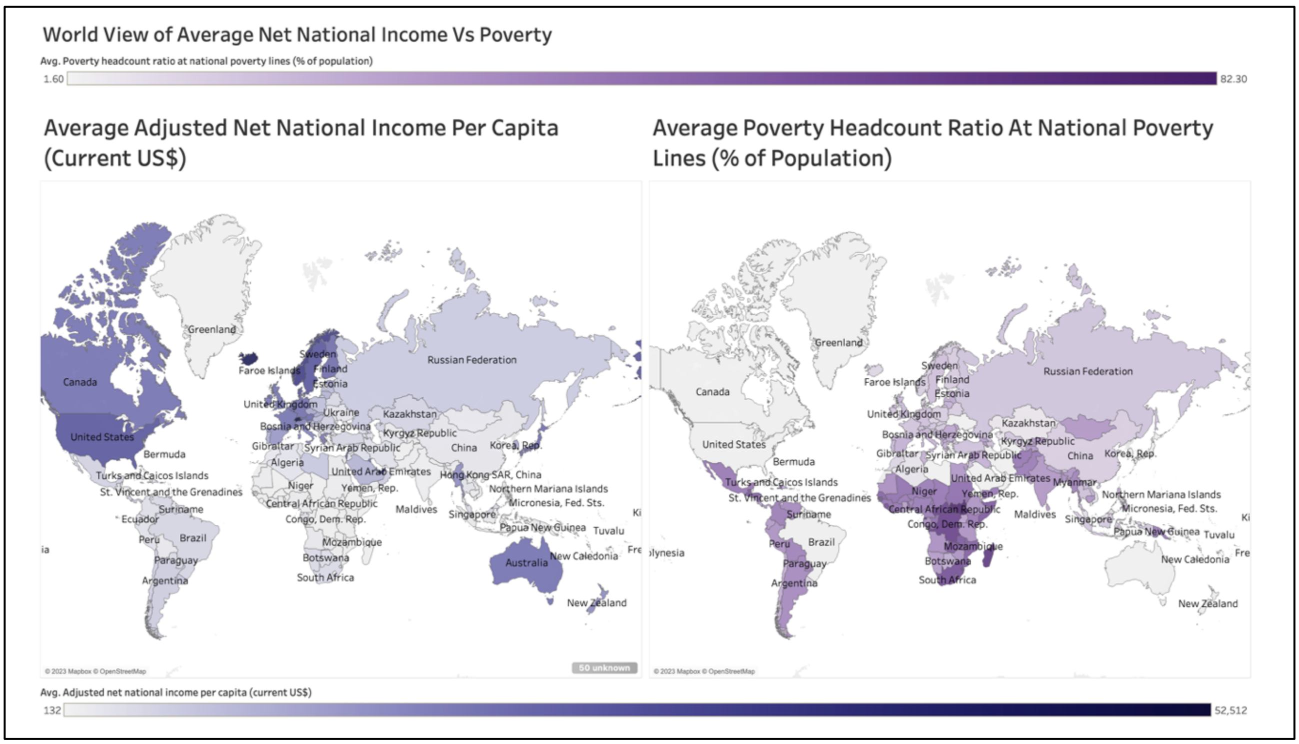

Fig1.3: Worldwide View of Average Net Income per capita and Poverty headcount ratio

By harnessing the remarkable mapping capabilities provided by Tableau, we were able to delve into the intricate details of countries worldwide seamlessly and visually, particularly focusing on those nations where a significant proportion of their population resides below the critical threshold of 50% of the median income level. This robust feature empowers us to gain unique insights into global income disparities and poverty, shedding light on regions that require targeted interventions and policy considerations. (Fig 1.4)

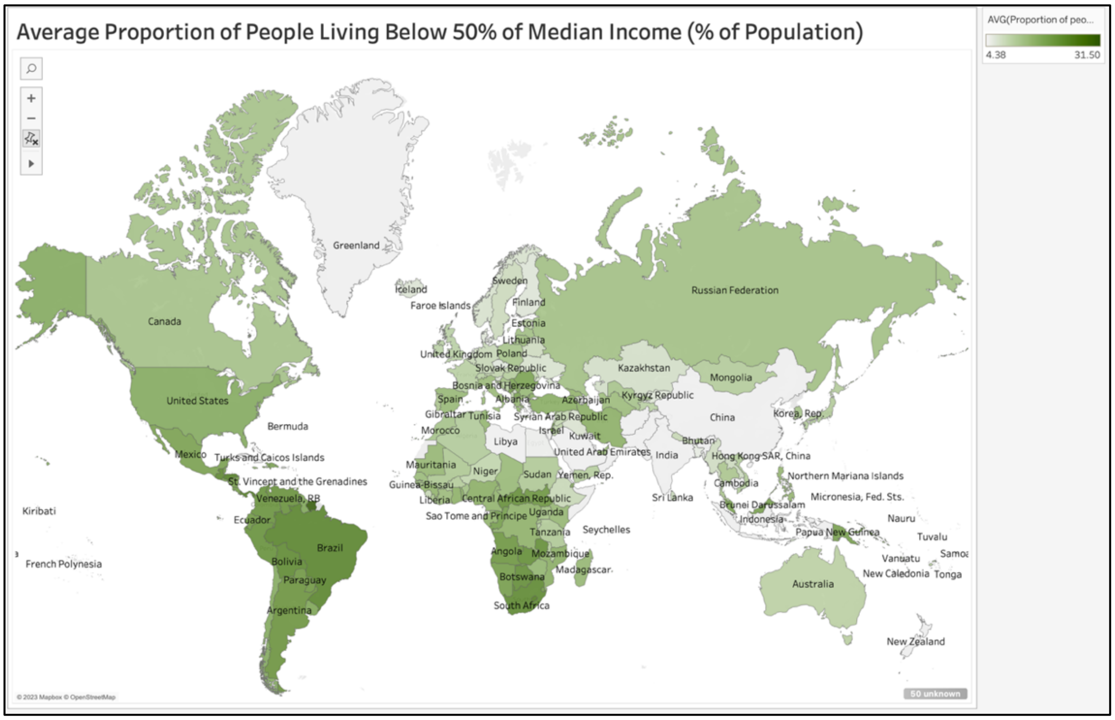

Fig1.4: Worldwide View of Average Proportion of people living below 50% of Median Income

In our subsequent world map visualizations (Fig 1.5), we aimed to visually represent our primary focus, which is child labor. By examining the patterns in the previous map visualizations featuring various socio-economic and educational parameters, we can intuitively discern the correlations between each of these factors and child labor. Moving forward, our approach will involve conducting a thorough correlation analysis to substantiate and validate our initial visual observations.

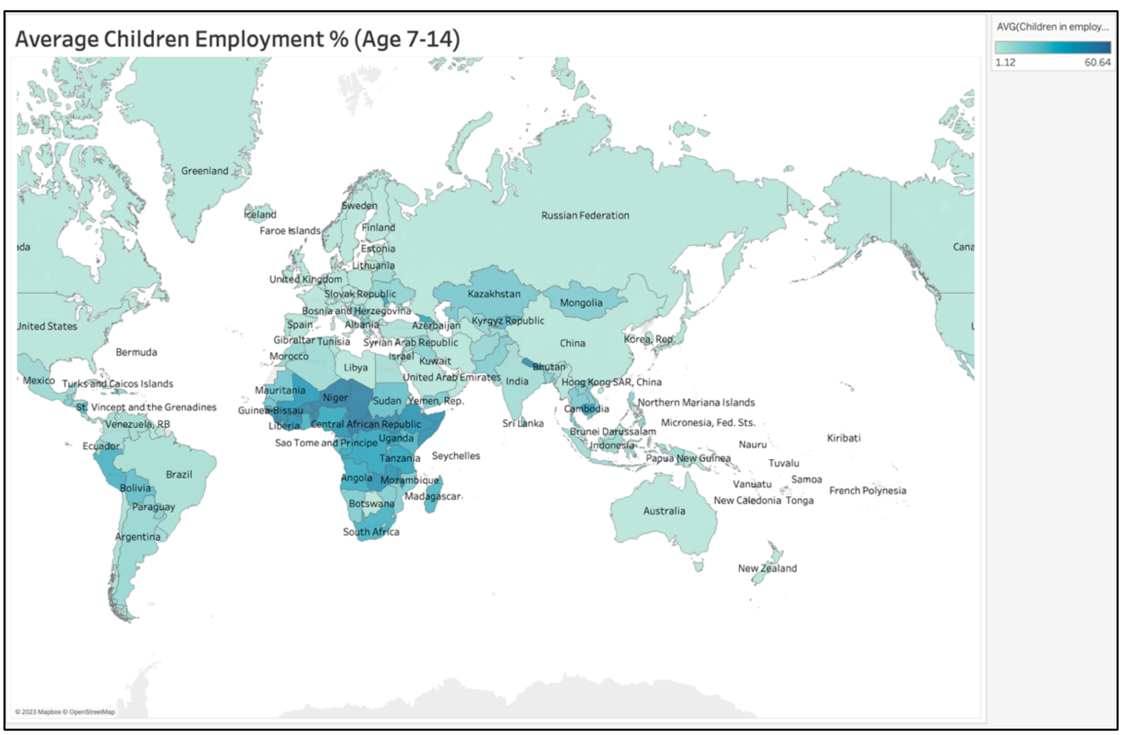

Fig1.5: Worldwide View of Average Child Employment

Now, we proceed to explore the correlations among the factors we've observed on the map view and visually analyzed to discern prevailing trends. Correlation scatterplots are particularly effective in providing a visually accessible means of grasping these trends, enabling individuals to quickly understand the relationships merely by glancing at the plots. (A Tableau scatterplot is a data visualization that uses points on a graph to represent the values of two different variables. Each point on the scatterplot represents a single data point, and the x and y coordinates of the point correspond to the values of the two variables being compared.)

The correlation between educational attainment percentage (at different levels like Bachelors, post-secondary, lower-secondary, primary) and poverty, child labor, and the proportion of people living below 50% of the median income is a complex relationship that can be influenced by numerous factors (Fig 1.6). Here's an overview of how these variables correlate based on our dataset:

- Educational Attainment and Poverty:
  Negative Correlation: it is clear from the visualization that there is a negative correlation between educational attainment and poverty (for all levels of education). As the level of education in a population increases, individuals are more likely to access better employment opportunities, which can lead to higher incomes and reduced poverty rates.

- Educational Attainment and Child Labor:
  Negative Correlation: A higher level of educational attainment is typically associated with a decreased likelihood of child labor. Education often provides alternatives to child labor by enabling children to access formal schooling and acquire skills for future employment.

- Educational Attainment and Proportion of People Living Below 50% of Median Income:
  Negative Correlation: In general, there is a negative correlation between educational attainment and the proportion of people living below 50% of the median income. Higher education levels are linked to increased earning potential, which, in turn, reduces the likelihood of individuals living below this income threshold. Several other factors play a role in influencing the proportion of people living below 50% of the median income, which is why the negative correlation is not as pronounced in this scenario as it is in others. Initially, I anticipated a robust correlation between this parameter and child labor. However, upon visualizing the data, the narrative took an unexpected turn. Therefore, in our iterative exploratory data analysis (EDA) approach, it may be prudent to exclude this parameter from consideration.

Fig1.6: Educational Attainment Analysis (Multi-level) with Poverty, Children Employment and % of people living below 50% of median income.

In our final visualization (Fig 1.7), we aimed to identify various additional indicators that exert a significant influence on child labor. The correlation between child labor and various socio-economic and educational factors is a multifaceted relationship influenced by a range of interconnected variables.

- Compulsory Education Years:
  Negative Correlation: Generally, there is a negative correlation between the number of compulsory education years and child labor. Countries with longer compulsory education periods often experience lower rates of child labor. Compulsory education laws are intended to keep children in school and out of the workforce.

- Coverage of Unemployment Benefits:
  Negative Correlation: A higher coverage of unemployment benefits is often associated with a reduction in child labor. This is because social safety nets can provide economic support to families, reducing the economic pressure to engage children in labor. Nonetheless, it's important to recognize that this correlation can fluctuate based on a multitude of socio-economic and cultural factors.

- Average Literacy in Youth:
  Negative Correlation: Increased youth literacy rates are typically correlated with a reduction in child labor. Education and literacy provide opportunities for older youth to access better employment options, reducing the need for child labor.

- School Enrollment Percentage:
  Negative Correlation: Higher school enrollment percentages among children usually correlate with lower child labor rates. When more children are enrolled in schools, fewer are available for labor.

- GDP Per Capita:
  Negative Correlation: There is often a negative correlation between GDP per capita and child labor. As the per capita income of a country increases, child labor rates tend to decrease. Higher income levels can reduce the economic necessity for child labor.

- Crude Birth Rate:
  Positive Correlation: The relationship between child labor and the crude birth rate is positive yet complex. Higher birth rates can sometimes lead to more child labor, as families may seek additional income sources. However, this correlation can vary depending on other socio-economic and cultural factors.

In scatterplots with a high density of points (here the points determine countries), individual data points may overlap, making it hard to discern their actual distribution. Density encoding helped here by representing the concentration of points in specific areas, providing a clearer picture of the overall data distribution. I have also used color gradients or intensity variations to encode the density pointers which helped in highlighting patterns and trends in the data. This is particularly useful when we want to emphasize areas of high or low concentration. In this visualization, it's evident that regions of greater concentration stand out, thanks to the distinct green-blue coloration.

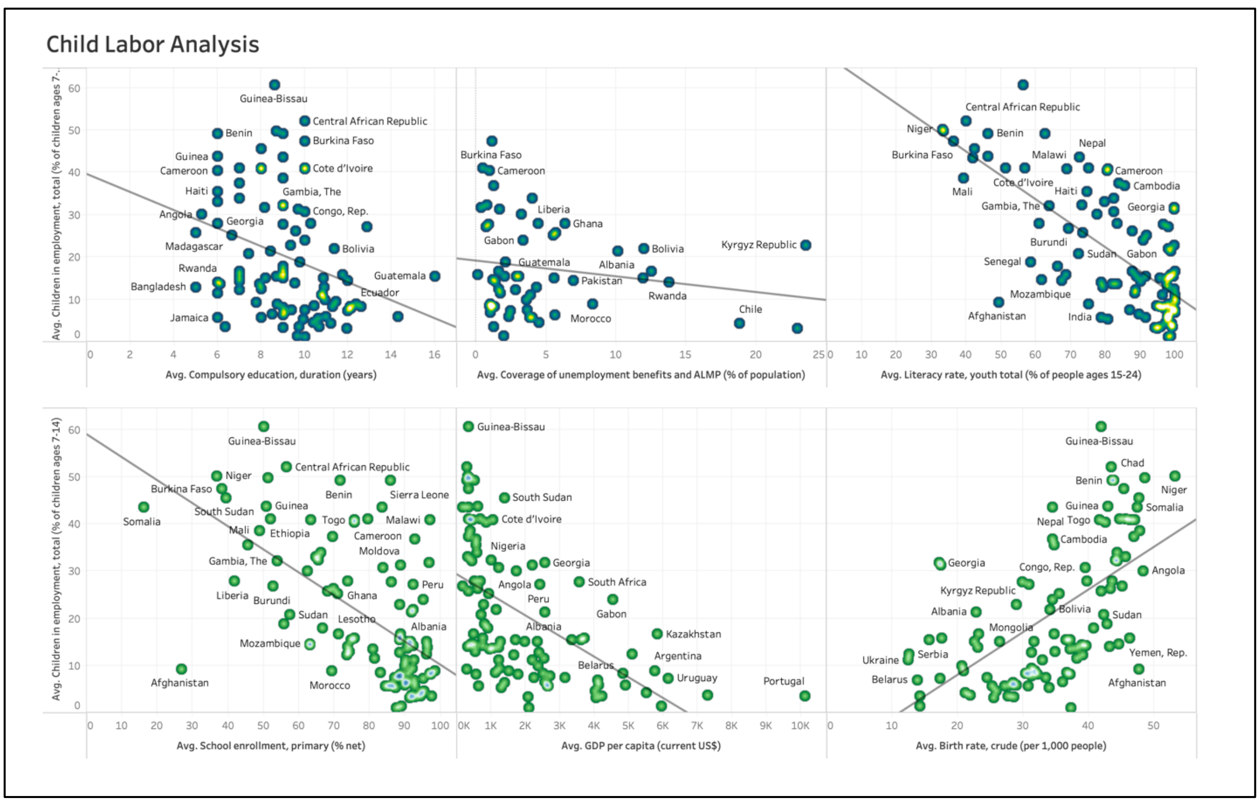

Fig1.7: Child Labor Analysis based on 6 parameters

Discussion/Conclusion

In wrapping up our exploratory visual analysis on child labor using the World Data Indicators dataset, several compelling insights have come to the forefront. The examination of key variables such as educational attainment, compulsory education years, youth literacy rates, average school enrollment, GDP per capita, crude birth rates and poverty levels has unveiled a nuanced and intricate landscape. Firstly, while initial expectations hinted at straightforward correlations, the data has woven a more complex narrative. The interplay between these factors and child labor is multifaceted, with certain surprising relationships and disparities challenging preconceived notions.

Notably, the visualizations have highlighted the crucial role of education-related metrics in influencing child labor rates. Countries with higher educational attainment and better literacy rates tend to exhibit lower instances of child labor, underscoring the importance of investing in education as a potent tool in combating this societal challenge. Additionally, economic indicators, such as GDP per capita, have shown intriguing correlations with child labor. The nuances in these relationships emphasize the need for targeted policies that address both educational and economic dimensions to effectively curb child labor.

While our exploratory visual analysis provides valuable insights, it is essential to acknowledge the limitations inherent in the dataset and visualization techniques employed. Further research could delve deeper into regional variations, cultural influences, and specific policy interventions to refine our understanding of the factors influencing child labor. In conclusion, this EDA serves as a steppingstone, offering a comprehensive overview of the complex dynamics surrounding child labor on a global scale. The insights gained pave the way for informed discussions, policy considerations, and future research endeavors aimed at creating a world where every child enjoys the right to a childhood free from exploitative labor practices.