UNICEF Global Database on School-Age Digital Connectivity: An Analysis of Internet Access

1. Data Preparation

The data we use in this project comes from 'Global database on school-age digital connectivity-Updated.zip' which was extracted After unzipping to produce a folder 'Global database on school-age digital connectivity-Corrected'. The folder contained three csv files named: Primary, Secondary, Total School Age, and a readme.txt file with the explanations. The name of the Total School Age.csv was shortened to Total for standardization.

The first problem was that the header row could not be used for analysis, so the solution was to remove it and set the next row as the header row. The second issue was irrelevant columns, such as ISO3 and Data source, so they were dropped. The third problem was the presence of missing values, which were dealt with by dropping the corresponding rows. The fourth issue was the dataset not being in numeric form, so the solution was to remove the % signs from the numeric columns and convert them to integers. Finally, the fifth concern was values in the 'Time Period' column exceeding one year. To address this, only the earlier year was chosen, and the remaining data was removed. After these steps, we had cleaned datasets files. Each datafile was saved with 'cleaned_' added before their names. The files were saved in the same directory as the jupyter notebook to avoid having to write complete paths to these files.

2. Data Exploration

2.1 Exploratory Data Analysis - Primary

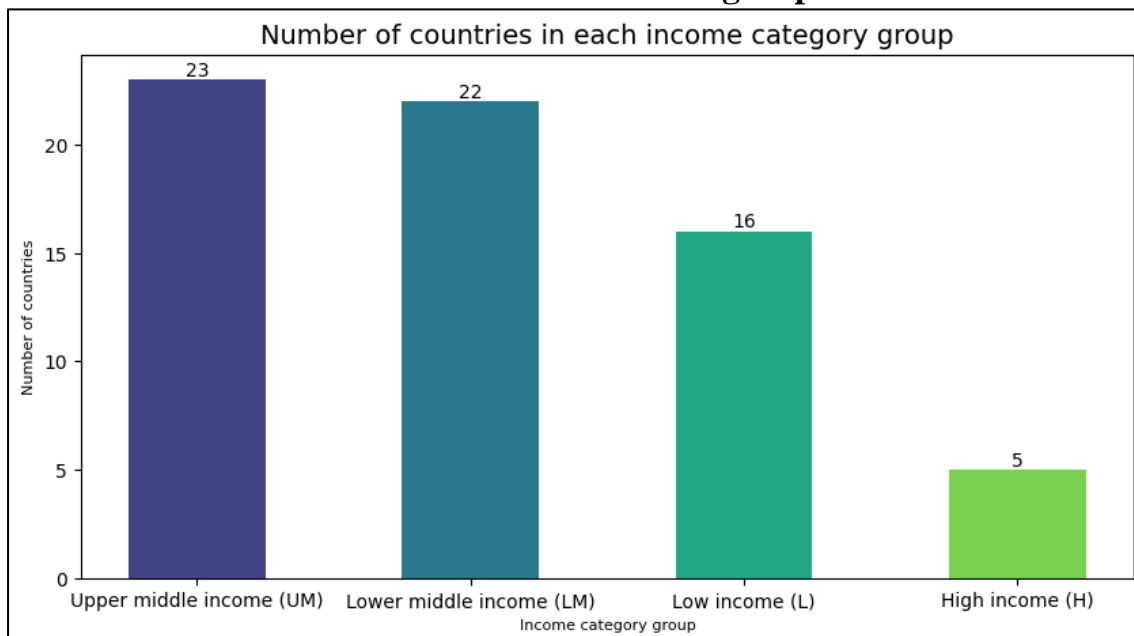**Nominal: Number of countries in each income group**

Figure 1: Number of countries in each income group

Per the dataset, a preponderance of primary school pupils hail from Upper-Middle Income Countries, 31 and Lower-Middle Income countries, 30 in number, with the latter cohort coming in second. Low-income countries trail not far behind, numbering 18. A comparatively diminutive fraction of students originates from High-Income countries, with only 8 entries in the dataset.
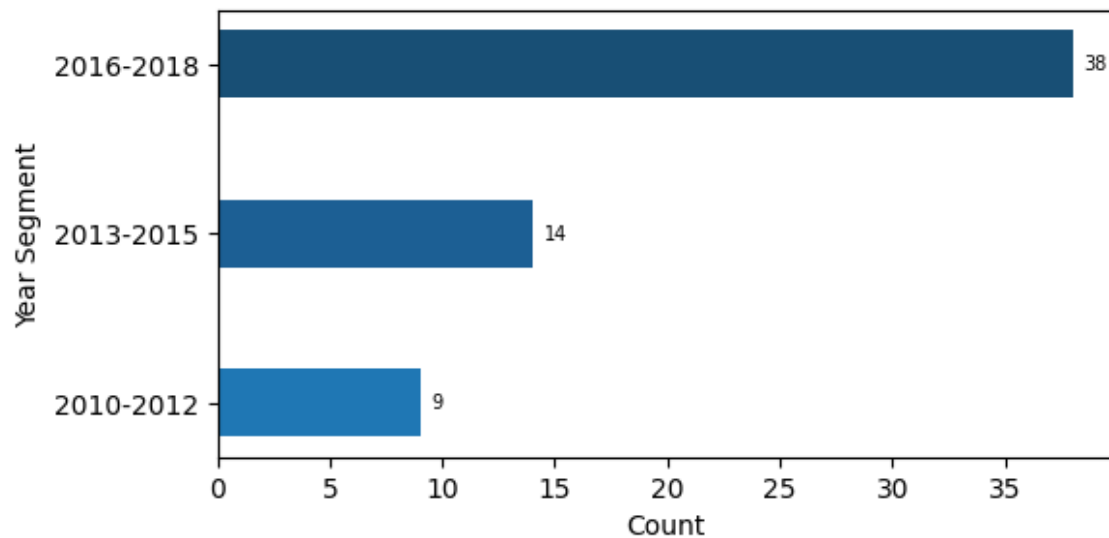
Ordinal: Period wise data collection

Figure 2: Number of countries from which data was collected by certain period

In order to assess the currency of our datasets, we employed a horizontal bar chart to visualize the "Time period" category. This chart categorized the years of data collection pertaining to digital connectivity in countries into three major groups spanning from 2010 to 2018. The findings indicate that the majority of the data was gathered between 2016 and 2018 (38), followed by the period spanning from 2013 to 2015 (14), while the smallest amount of data was collected between 2010 and 2012 (9).

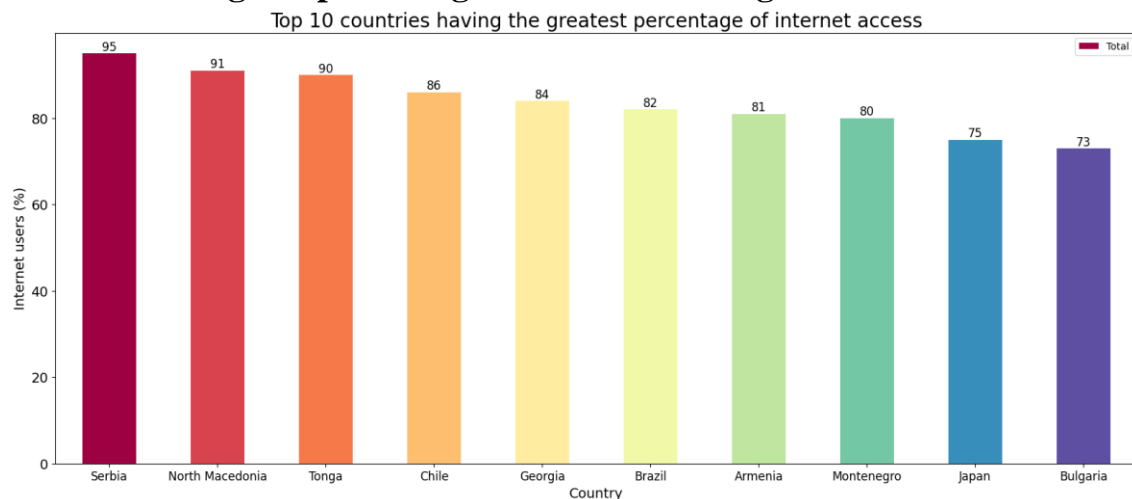## Numerical: Highest percentage of students having access to the Internet

Figure 3: Top 10 countries with the most percentage of students digitally connected

The data visualization shows that most of the top 10 countries and areas with the highest percentage of primary school children with internet connection at home are located in the northern hemisphere, such as Serbia, Armenia, Bulgaria and so on. Additionally, it is important to note that only 9 countries in the cleaned dataset have achieved at least 75% connectivity.

2.2 Analysis by Income Group and Residence

High Income:

High-income countries exhibited a comparatively smaller average disparity in internet connectivity percentages between their urban and rural areas. Japan, in particular, stood out as a country where rural internet connectivity surpassed that of urban areas. Surprisingly, it was also the only country whose 80% of the students had access to the internet, across all residential areas. Conversely, Panama, despite being a high-income country, faced challenges in providing internet facilities to a significant number of its students, and especially in the rural areas of the country.
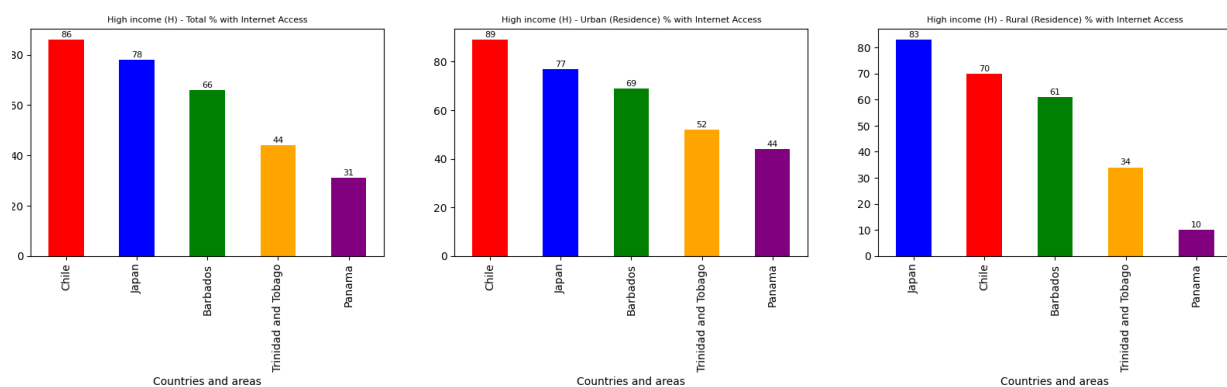

Figure 4: High Income Countries with Internet Connectivity

Upper Middle Income:

The data indicates that internet connectivity varies for students in rural and urban areas across most countries in this group. While smaller countries such as Serbia, North Macedonia and Tonga have achieved almost equal internet connectivity for students in urban and rural areas, countries such as Brazil, Georgia and Armenia have a substantial gap of percentage points for the students who have access to the internet. Likewise, countries such as Bulgaria, Thailand and Costa Rica do not appear in all three categories together, indicating potential disparity of internet connectivity between the students of rural and urban areas.
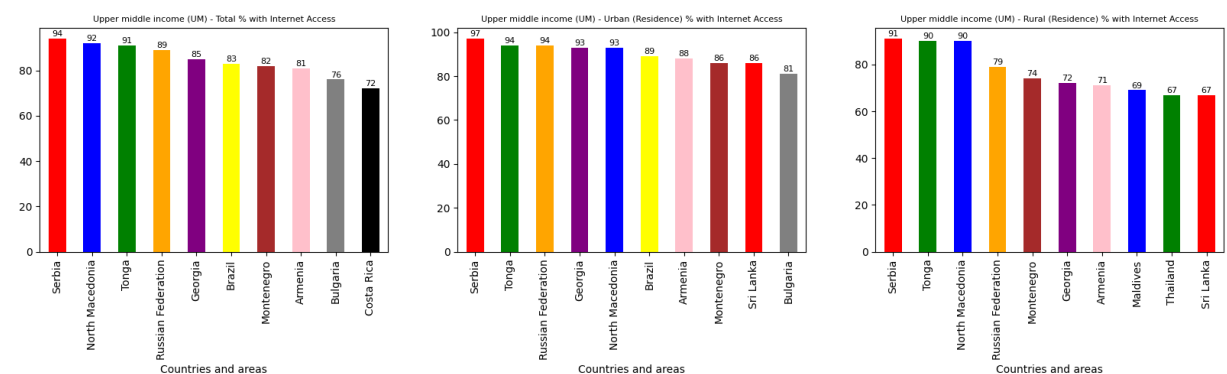

Figure 5: Upper-Middle Income Countries Internet Connectivity

Lower-Middle Income:

High-income countries exhibited a comparatively smaller average disparity in internet connectivity percentages between their urban and rural areas. Japan, in particular, stood out as a country where rural internet connectivity surpassed that of urban areas. Surprisingly, it was also the only country where 80% of the students had access to the internet, across all residential areas. Conversely, Panama, despite being a high-income country, faced challenges in providing internet facilities to a significant number of its students, especially in the rural areas of the country.
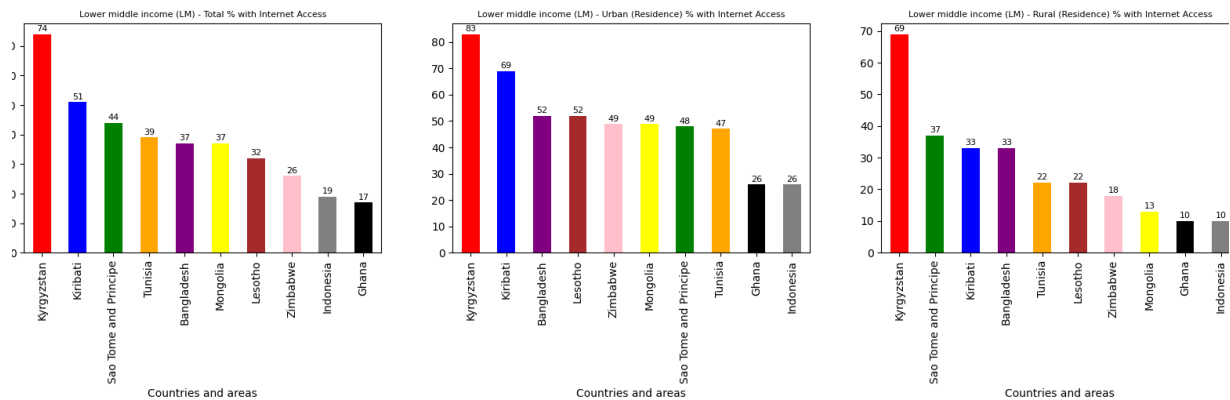

Figure 6: Lower-Middle Income Countries Internet Connectivity

Low Income:

Countries in the low-income group are mainly underdeveloped. As shown by the visualization, these countries have a large disparity in connectivity between the rural and urban areas, which is expected. Most of these countries also have half of their student population without internet access. This can be verified by looking at the chart: Tajikistan, Togo and Haiti all have low internet connectivity for their school-age children. It also implies that there is a significant gap in internet access for students in the rural and urban areas. Gambia is the only country that has performed better than the other countries in the list by achieving higher connectivity for its students.
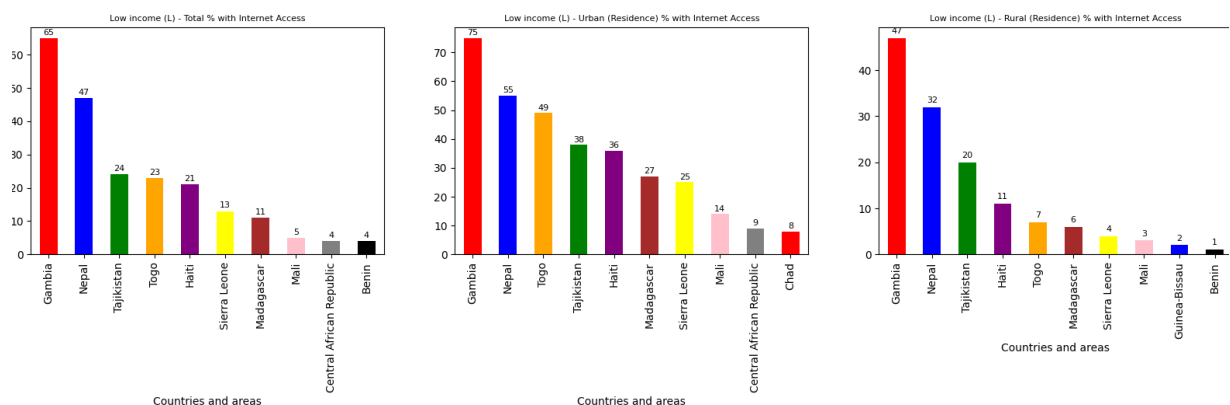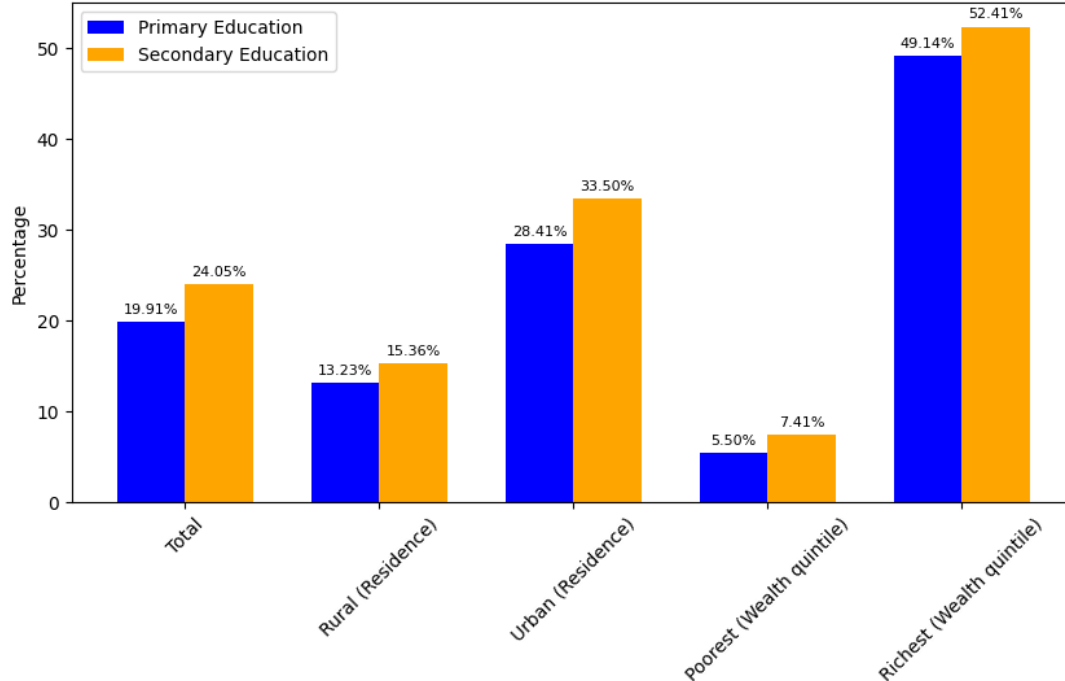

Figure 7: Low Income Countries with Internet Connectivity

2.3 Comparison of Internet Access between the students of Primary and Secondary in Lower-Middle Income Countries

Internet Access between the students of Primary and Secondary Education in Lower Middle Income Countries

The visualization chart highlights significant points about the average comparison of digital connectivity between the primary and secondary level students of lower-middle income countries.

Overall, a higher percentage of secondary level students (24.05%) have access to the internet than primary level students (19.91%). This can be explained by the fact that higher level students are more likely to need connectivity for education.

The gap in connectivity for the students residing in rural areas is smaller, where secondary students (15.36%) have access to the internet, compared to primary students (13.23%). However, the gap in connectivity for the students residing in urban areas is larger: The primary level students (28.41%) have a lower percentage of digital connectivity than the secondary students (33.50%).

In comparison of wealth quantiles, the poorest primary and secondary students are similar. Both categories of students have less than 10% of students who are able to connect to the internet. In this case, the connectivity percentage for primary students (5.50%) is a touch low of secondary students (7.41%). Not surprisingly, the students in the richest quantile categories of their countries have a very high ratio of internet connectivity: nearly half of the secondary students (52.41%) and primary students (49.14%) are connected to the internet.