**CEMA 2025 Internship program task**

HIV continues to be a major public health issue in Africa, with a high number of people living with the virus, especially in sub-Saharan regions. While prevention and treatment efforts have improved, HIV is a lifelong condition, making it crucial to track data on this affected. This data helps us identify infection trends, prepare healthcare services to accommodate these patients, and address challenges like stigma and inequality. Beyond health, HIV also impacts jobs, healthcare costs, and poverty levels, making it important to use data to create targeted interventions. By analyzing this information, governments and organizations can better allocate resources and create policies to help reduce transmission and improve the quality of life for those affected.

You are provided with a dataset from the World Health Organization (WHO) Global Observatory, containing data on people living with HIV at the country level from 2000 to 2023.

Using this dataset, we would like you to:

* Create a visualization that shows the trend of HIV cases in the countries that contribute to 75% of the global burden
* Generate a visualization that displays the trend of HIV cases in the countries contributing to 75% of the burden within each WHO region (column called ParentLocationCode contains the WHO regions)

You have also been provided with World Bank data on the multidimensional poverty headcount ratio, which includes factors such as income, educational attainment, school enrolment, electricity access, sanitation and drinking water.

We would like you to merge this dataset with the HIV data above and analyze the relationship between people living with HIV and multidimensional poverty, and the individual factors that contribute to the ratio. Remember to account for the random effects (country, year).

Write a paragraph on your findings.

Question two

You have been provided with data on the under-five mortality rate and neonatal mortality rate for the African region, which has been downloaded from the UN Inter-agency Group for Child Mortality Estimation. Your task is to:

* Filter data for the eight countries belonging to the East African Community (list here: <https://www.eac.int/overview-of-eac>)
* Visualize the latest estimate of each indicator at the country level using shapefiles, which can be downloaded from [www.gadm.org](http://www.gadm.org).
* Show the average trends in the mortality rates over time (plot the average trend line and add the points in the graphic for the country level estimates for each indicator. Expectation: two plots).
* Based on your visualizations, identify the countries with the highest under-five mortality rates in East Africa and the highest neonatal mortality.

You should work on an RMD document that is saved in your name, e.g juma\_fulani.rmd.

Kindly ensure you show the code in your data.

The rmd should be uploaded on this link (<https://docs.google.com/forms/d/e/1FAIpQLSd2gFkhTt1KP9vb74pYCsvFgrO0QZb_1js5fK2Aak-phAYxiA/viewform?usp=sharing> ) by this 27th April 2025 at 8pm EAT. Kindly note that documents uploaded past the deadline will not be considered.