

Data Documentation: Youth Population and Education Projections (2025-2035)

Overview

For this data dive challenge there are four datasets at your disposal that combine demographic projections from the United Nations World Population Prospects 2024 Revision with education projections from the Wittgenstein Centre's Shared Socioeconomic Pathways (SSPs). The data covers nearly all countries globally for the period 2025-2035, focusing on young people aged 15-24. While high-income countries (HIC) are included for comparison purposes, the challenge emphasizes analysis of non-high-income countries where demographic and educational transitions are most dynamic.

All datasets share common identifiers (ISO3 country codes, years, age groups, sex, and scenarios) that enable seamless integration across files. Each country is classified into multiple regional and income groupings (e.g., SSA for Sub-Saharan Africa, LIC for Low Income Countries, EAP for East Asia & Pacific), allowing for flexible aggregation and comparison. The projection scenarios reflect different assumptions about future fertility, mortality, migration, and educational investment—these are explained in detail in the accompanying PDF documentation files.

The Four Datasets

1. UN_Pop_Projections.csv

This file contains population projections from the UN for ages 15-19 and 20-24, disaggregated by sex (Male, Female, Both). The data includes 14 different projection scenarios that vary assumptions about fertility, mortality, and migration (e.g., medium_variant, high_fertility, low_fertility, zero_migration, constant_fertility). Each row represents the estimated population for a specific country-year-age-sex-scenario combination. The boolean columns (EAP, ECA, LAC, LIC, LMC, MNA, SAS, SSA, UMC, HIC, LMY, AFE, AFW) indicate which regional or income groupings each country belongs to, with countries often belonging to multiple groups simultaneously.

This dataset serves as the demographic foundation for weighting and contextualizing educational outcomes. For example, a country with rapid population growth among 15-24 year-olds will face different educational challenges than one with a shrinking youth cohort.

2. WD_MYS_Projections.csv

Mean Years of Schooling (MYS) projections from the Wittgenstein Centre provide a single summary measure of educational attainment for each country-year-age-sex combination across five SSP scenarios (1-5, where SSP2 is the "Middle of the Road" baseline). MYS represents the average number of years of formal education completed by individuals in

each demographic group. The file structure mirrors the population projections with the same country grouping indicators.

This metric is particularly useful for tracking aggregate educational progress and comparing countries or regions. For instance, you might examine which regions are expected to see the fastest improvements in MYS between 2025 and 2035, or how different SSP scenarios lead to divergent educational outcomes.

3. WD_PRP_Projections.csv

Education distribution (proportion) projections break down the population by specific educational attainment levels (e.g., No Education, Primary, Lower Secondary, Upper Secondary, Post Secondary) for each country-year-age-sex-scenario combination across the five SSP scenarios. Unlike MYS which provides a single average, this file shows the full distribution of educational achievement within each demographic group, with proportions summing to 1.0 for each unique combination.

This granular view enables analysis of educational inequality and composition. For example, you could track how the share of youth with post-secondary education evolves, identify countries where a large proportion remains without formal education, or examine how SSP scenarios differ in their assumptions about educational expansion at different levels.

4. UN_Shr_Projections.csv

Population shares show what proportion of each regional or income group's total youth population is contributed by each country, calculated separately for each age group (15-19, 20-24, and 15-24 combined), sex, and scenario combination. For instance, a row showing Burundi's share in the SSA group indicates what percentage of Sub-Saharan Africa's 15-19 year-olds (in that scenario and sex) live in Burundi.

This dataset is essential for creating properly weighted regional or group-level aggregates. When combined with the education data, it allows you to calculate statements like "the average MYS for youth in Sub-Saharan Africa" or "the proportion of 15-24 year-olds in Low Income Countries with post-secondary education," where each country's contribution is weighted by its share of the group's youth population.

Combining the Datasets

The real analytical power comes from integrating these datasets. For example, you might:

- Join UN_Shr_Projections.csv with WD_MYS_Projections.csv to calculate population-weighted mean years of schooling for specific regions or income groups

- Merge UN_Pop_Projections.csv with WD_PRP_Projections.csv to understand how absolute numbers of youth at different education levels will change (not just proportions)
- Compare scenarios by analyzing how SSP1 (Sustainability) versus SSP3 (Fragmentation) lead to different educational futures when combined with different UN demographic scenarios

All files can be linked using the common keys: iso3, year, age, sex, and scenario (noting that UN projections use different scenario names than Wittgenstein SSPs). The source code provided demonstrates how these datasets were created and can serve as a reference for additional data manipulation or validation.

Getting Started

Some potential starting points for exploration:

- Compare SSP2 (Middle Road) projections across different income groups to identify where youth populations are growing fastest and whether educational investments are keeping pace
- Examine specific countries or regions experiencing demographic transitions to understand the interplay between population change and educational expansion
- Analyze gender disparities in educational attainment across different scenarios and regions

The accompanying PDF files provide detailed explanations of the projection methodologies and scenario assumptions. We encourage you to explore the data creatively while keeping the core challenge in mind: understanding how the educational landscape of today's non-high-income countries will transform over the next decade.