



# Institute for Health Metrics and Evaluation

## Data Release Information Sheet

### ***Data Summary***

Dataset name: Low- and Middle-Income Country Educational Attainment Geospatial Estimates 2000-2017

Date of release: December 24, 2019

Summary:

This dataset contains estimates produced for educational attainment for adults ages 15-49, and the 20–24 subgroup, by sex at the 5x5 km-level for 105 low- and middle-income countries for 2000-2017. It provides years of education and proportion of the population attaining key levels. These estimates were produced using individual records from 528 geo-referenced household sample survey and census sources.

The dataset includes the following:

- GeoTIFF raster files for pixel-level estimates of mean educational attainment, and proportion of the population achieving zero, less than primary, primary, and secondary schooling for adults ages 15-49 and 20-24, divided by sex
- CSV files of aggregated estimates for each country at the zero, first, and second administrative divisions
- Code files used to generate the estimates

[Get Data Files](#)

Relevant publications and visualizations:

- Local Burden of Disease Educational Attainment Collaborators. Mapping disparities in education across low- and middle-income countries. *Nature*. 24 December 2019.
- [Local Burden of Disease - Education](#)

### **Acknowledgements**

Contributing organizations:

- Institute for Health Metrics and Evaluation (IHME)

#### Funders:

- Bill and Melinda Gates Foundation (BMGF)
- National Institutes of Health (NIH)

#### Suggested Citation:

Institute for Health Metrics and Evaluation (IHME). Low- and Middle-Income Country Educational Attainment Geospatial Estimates 2000-2017. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2019.

### ***Data Files Information***

#### **CSV files of Aggregated Estimates of Educational Attainment**

Stored in files named <INDICATOR>\_<AGE>\_<SEX>\_<LEVEL\_OF\_AGGREGATION>.CSV

(Example: IHME\_LMIC\_EDU\_2000\_2017\_ZEROPROP\_15\_49\_FEMALE\_ADO\_Y2019M12D24.CSV)

- **Indicator:**
  - [MEAN]: Mean years of attainment
  - [ZEROPROP]: Proportion of population achieving zero years of education
  - [NOPRIMARYPROP]: Proportion of population achieving 1-6 years education
  - [PRIMARYPROP]: Proportion of population achieving 6-11 years education
  - [SECONDARYPROP]: Proportion of population achieving 12+ years education
- **Level of aggregation:** admin0, admin1, or admin2, corresponding to first and second administrative level areas as defined in the Database of Global Administrative Areas (GADM) shapefiles, with minor adjustments made where names were missing in the original shapefile. Each row in each table is unique by administrative unit and year
- **Age:** [15\_49] for 15-49 or [20\_24] for 20-24
- **Sex:** [FEMALE] or [MALE]

Variable	Variable Label	Variable Definition
ADM0_CODE	GADM Admin 0 Code	GADM code identifying the administrative unit
ADM0_NAME	Admin 0 Name	Zero level administrative unit (Country) name as found in the GADM shapefile
ADM1_CODE	GADM Admin 1 Code	GADM code identifying the administrative unit (Only in the admin1 files)
ADM1_NAME	Admin 1 Name	First level administrative unit name as found in the GADM shapefile
ADM2_CODE	GADM Admin 2 Code	GADM code identifying the administrative unit (Only in the admin2 files)
ADM2_NAME	Admin 2 Name	Second level administrative unit name as found in the GADM shapefile (Only in the admin2 files)
year	Year	Time period of estimate. Possible values: years in the range 2000-2017

Variable	Variable Label	Variable Definition
age_group_id	Age Group ID	Unique numeric identifier for the age group generated and stored in an IHME database of data dimensions. Possible values: 9, 24
age_group_name	Age Group Name	Age group estimated. Possible values: 15-49, 20-24
measure	Measure	The measure (indicator) estimated. Possible values: <ul style="list-style-type: none"> <li>• Mean years of attainment</li> <li>• Proportion of population achieving 0 years education</li> <li>• Proportion of population achieving 1-6 years education</li> <li>• Proportion of population achieving 6-11 years education</li> <li>• Proportion of population achieving 12+ years education</li> </ul>
mean	Mean	Mean posterior population-weighted estimate for the administrative unit
lower	Lower Confidence Interval	2.5% population-weighted posterior quantile estimate for the administrative unit
upper	Upper Confidence Interval	97.5% population-weighted posterior quantile estimate for the administrative unit

## Codebooks

Variable names, labels, and value encoding for admin 0 files can be found in the machine-actionable codebook file [IHME\\_LMIC\\_EDU\\_2000\\_2017\\_CODEBOOK\\_ADO\\_Y2019M12D24.CSV](#)

Variable names, labels, and value encoding for admin 1 files can be found in the machine-actionable codebook file [IHME\\_LMIC\\_EDU\\_2000\\_2017\\_CODEBOOK\\_AD1\\_Y2019M12D24.CSV](#)

Variable names, labels, and value encoding for admin 2 files can be found in the machine-actionable [IHME\\_LMIC\\_EDU\\_2000\\_2017\\_CODEBOOK\\_AD2\\_Y2019M12D24.CSV](#)

## GeoTIFF Raster Files for Educational Attainment

Stored in files named EDU\_<INDICATOR>\_<AGE>\_<SEX>\_<YEAR>\_<MEASURE>.TIF

(Example: IHME\_LMIC\_EDU\_2000\_2017\_MEAN\_15\_49\_FEMALE\_LOWER\_Y2019M12D24.TIF)

- **Indicator:** Educational attainment indicator
  - [MEAN]: Mean years of attainment
  - [ZEROPROP]: Proportion of population achieving zero years of education
  - [NOPRIMARYPROP]: Proportion of population achieving 1-6 years education
  - [PRIMARYPROP]: Proportion of population achieving 6-11 years education
  - [SECONDARYPROP]: Proportion of population achieving 12+ years education
- **Age:** [15\_49] for 15-49 or [20\_24] for 20-24
- **Sex:** [FEMALE] or [MALE]

- **Measure:** [MEAN], [UPPER], or [LOWER] summary statistics from the predictive posterior distribution at each pixel. Lower and upper correspond to 2.5% and 97.5% quantiles
- **Year:** From 2000 to 2017, corresponding to the time period of the estimate

Note that rasters mask (i.e., have NA values) for lakes and areas with low population (10 people per 1km and classified as barren/sparsely vegetated).

## **Data Input Sources**

This CSV file contains relevant metadata about the input sources as suggested in the [Guidelines for Accurate and Transparent Health Estimates Reporting \(GATHER\)](#), a statement that promotes best practices in reporting health estimates.

## ***Additional Information***

### **Terms and Conditions**

<http://www.healthdata.org/about/terms-and-conditions>

### **Contact information**

To request further information about this dataset, please contact IHME:

Institute for Health Metrics and Evaluation

2301 Fifth Ave., Suite 600

Seattle, WA 98121

USA

Telephone: +1-206-897-2800

Fax: +1-206-897-2899

Email: [data@healthdata.org](mailto:data@healthdata.org)

[www.healthdata.org](http://www.healthdata.org)

These files may be updated periodically, so we appreciate hearing feedback or additional information about how these data are being used.