

Data Descriptor

Not peer-reviewed version

---

# The Subnational Gender Development Database: Uncovering Variation in Gender Inequality Within Countries

---

[Jeroen Smits](#)\* and Iñaki Permanyer

Posted Date: 15 October 2025

doi: 10.20944/preprints202510.1214.v1

Keywords: gender inequality; subnational variation; global index; gender-specific human development; 2000-2023



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

*Data Descriptor*

# The Subnational Gender Development Database: Uncovering Variation in Gender Inequality Within Countries

Jeroen Smits <sup>1,\*</sup> and Iñaki Permanyer <sup>2</sup>

<sup>1</sup> Global Data Lab, Economics, Institute for management Research, Radboud University, Netherlands

<sup>2</sup> Centre d'Estudis Demogràfics, Universitat Autònoma de Barcelona, Spain

\* Correspondence: Jeroen.smits@ru.nl

## Abstract

In this paper we describe and present the Subnational Gender Development Database. This database contains for the period 2000-2023 the national and subnational values of the Subnational Gender Development Index (SGDI), the male and female values of the Subnational Human Development Index and the underlying gender-specific education, health and standard of living indices. The subnational values of the indicators were computed on the basis of data from statistical offices and from the Area Database of the Global Data Lab, which contains indicators aggregated from household surveys and census datasets. These subnational values were computed in such a way that their population-weighted national averages are equal to their official national values, as presented by the United Nations Development Program (UNDP) in their Human Development Reports. Subnational values for missing years have been calculated by using interpolation and extrapolation on the basis of real values. The SGDI Database contains values for the SGDI and underlying indicators for 1810 subnational regions in 166 countries.

**Keywords:** gender inequality; subnational variation; global index; gender-specific human development; 2000-2023

---

## 1. Introduction

In most societies worldwide, gender inequality is widespread, with males better positioned in social, economic, and political hierarchies (Guthridge et al., 2022; UN, 2024; Silva & Klasen, 2021; World Economic Forum, 2025; ). Gender equality is important in and on itself, and it is also instrumental in achieving other socially desirable goals, like the eradication of poverty, guaranteeing equality of opportunities for all, or fostering economic growth (Silva & Klasen, 2021; Fernández et al., 2021; Ghosh & Ramanayake, 2020). For these reasons, the goal of reducing gender inequality has held a prominent place in the international development agenda of the last decades. One of the 17 Sustainable Development Goals (Goal 5) is specifically focused on improving the situation of women and there is also broad agreement that reduction of gender inequality is an important prerequisite for realizing most of the other SDGs (Filho et al., 2023; UN, 2018).

A major instrument for measuring gender inequalities across the globe is the Gender Development Index (GDI), which since 2014 has been published yearly in the Human Development Reports of the United Nations Development Programme (UNDP, 2025a). This GDI is defined as the ratio between the female and male values of the Human Development Index (HDI) at the country level. It is currently available for 184 countries from all regions of the world (UNDP, 2025a).

Like most other indices of gender inequality, the GDI has until now only been available at the national level. This is a disadvantage, as it seems likely that gender disparities are not the same everywhere within a country. The position of women as well as gender norms are known to differ

between ethnic, religious, educational and other groups (Hossain, 2024; Wildeman, Smits & Schrijner, 2023; Forman-Rabinovici & Sommer, 2018; Bras & Smits, 2021), which might be living in different mixtures in different areas of a country. The resulting variety of within country gender patterns can obviously not be represented well by a national indicator. The first Human Development Report therefore already highlighted the importance of developing group and (subnational) region-specific indices (UNDP, 1990, p.32).

For the HDI itself, already much research into its within-country variation has been done. While subnational information on the HDI for specific countries has been available for a long time (Foster et al. 2005; Grimm et al. 2008, 2011; Harttgen and Klasen 2011; Permanyer 2013; Permanyer et al. 2015), until recently it was not possible to study these differences for more than a handful of countries. However, in 2019 a global subnational HDI (called SHDI) was introduced (Smits & Permanyer, 2019; Permanyer & Smits, 2020) which in its most recent version covers 1855 subnational regions within 168 countries.

This encompassing index made it possible to study within and between country inequality in human development and its underlying indicators on a scale that has never been possible before. Smits and Permanyer (2020) found within country variation in human development to be particularly strong in low and middle income countries. While in high income countries observed inequality in human development increased by about 20% when moving from national to subnational information, in low and middle income countries observed inequality doubled. Hence in the poorer regions of our globe, not only the level of human development is lower, but the available human development resources are also divided more unequal among regions.

While thus already much is known about how Human Development itself varies between regions within countries, much less is known about the way in which gender inequality in human development varies within countries. The central aim of this paper is to shed light on this issue by presenting a subnational version of the UNDP's Gender Development Index. This Subnational Gender Development Index (SGDI) and its underlying indicators are currently available for 1810 subnational regions in 166 countries, covering all regions and development levels of our world.

With this new index it becomes possible to study inequalities in human development between women and men on a global scale with a degree of detail that is ten times more precise than what has been possible before. This is highly important as for the realization of SDG 5 "Achieving gender equality and empowering all women and girls" it is essential that we are able to study and monitor the position of women versus men as detailed as possible.

The indicators in the SGDI Database are scaled in such a way that their population weighted national averages equal their official national values as presented by the Human Development Report Office (HDRO) of the UNDP (2025c). This procedure, which has been used before to create subnational versions of other established indicators, like the Human Development Index and its underlying subindices (Smits & Permanyer, 2019), the corruption indices of World Bank and Transparency International (Crombach & Smits, 2024) and the GDL Vulnerability Index (Smits & Huisman, 2025), guarantees that the indicators that are produced can be seen as subnational versions of these established indicators.

The new SGDI database provides researchers worldwide with high-detailed contextual variables that help improve our understanding of gender inequality across the globe and to study its role in a broad range of fields (including family formation, migration, health and mortality, epidemiology, cultural/ideational/normative change, religion, socio-economic change, or environmental sustainability). It may play an essential role in developing tailor-made policy measures for reducing gender inequalities in specific situations and in this way help realizing the global development agenda. With subnational data it is possible to direct resources much more precisely to the places where they are needed than with national data. As such, the SGDI and its underlying indices have great potential to put gender equality more prominent on the development agenda and to help articulate national and international gender policies into a more coherent whole.

## 2. Methods

The Subnational Gender Development Index (SGDI) which is introduced here translates the UNDP's official Gender Development Index (GDI, <https://hdr.undp.org>) to the subnational level. The GDI provides a comparison of the level of human development of men and women within a specific country. A value of one means complete gender equality. A value below one indicates a higher level of human development for men and a value above one a higher level of human development for women. The SGDI works in the same way, but provides information on gender inequality within subnational areas of countries.

Starting point for the construction of the GDI are gender-specific versions of the Human Development Index (HDI) of the UNDP, which is one of the most important indicators of the level of development of countries across the globe (UNDP, 2025; Myrskylä et al. 2009; Bray et al. 2012; Permanyer & Smits, 2020). The HDI shows a country's performance with regard to three major dimensions of development: education, health and standard of living. Since its introduction in 1990, the HDI has become the key reference indicator for assessing the socio-economic performance of countries, for policy-makers, academics, development workers, and the public at large (Klugman et al., 2011).

The procedure for computing the GDI is for a large part equal to the procedure for computing the HDI (UNDP, 2025b). The major difference is that this procedure is followed separately for women and men. To measure the three dimensions of human development – both for country's as a whole and for women and men separately – gender-specific information on four indicators is needed. For the health dimension the underlying indicator is life expectancy at birth; for standard of living it is gross national income per capita (PPP, 2021 US\$) and for the educational dimension it is mean years of schooling of adults aged 25+ and expected years of schooling of children aged 6.

To construct their subnational version of the HDI, Smits and Permanyer (2019) collected subnational and national data for these four indicators and for the population sizes of the areas for which the index was constructed. This data was derived from three separate sources: statistical offices, the Area Database of the Global Data Lab (Smits, 2016) and the HDI Database of the UNDP (2025c). The last source was used to obtain national data needed to make the SHDI at the national level equal to the UNDP's HDI.

Because not all required data was available for all indicators and all countries in all years, Smits and Permanyer used several estimation procedures to fill in data gaps and performed test analyses to determine the size of the error introduced by these procedures. Here we make use of the same sources and procedures to build a new database with gender-specific data on the underlying four indicators for the period 2000-2023.

In the following section, the data sources that were used to build the SGDI database are discussed. Thereafter we elaborate on the measurement of the dimension indicators for the three dimensions on which the HDI is based. In the "Data Processing" section we subsequently discuss how the indicators were scaled, missing values estimated, the male and female SHDI were computed and connected together in the SGDI.

### 2.1. Data sources

Three major data sources were used to create our SGDI database. (1) We obtained data from statistical offices -- including Eurostat, the statistical office of the European Union -- by downloading indicators from their website or through email communication with them. (2) We derived data from the Area Database of the Global Data Lab (Smits, 2016). (3) we downloaded data from the HDI website of the United Nations Development Program (UNDP, 2025c). Below we discuss these three sources.

### 2.1.1. Statistical Offices

For most EU member states, data were obtained from Eurostat. Eurostat's definition of subnational regions follows the NUTS classification (*Nomenclature of Territorial Units for Statistics*; <https://ec.europa.eu/eurostat/web/nuts>), which provides a hierarchical framework for dividing the economic territory of the European Union. Within this system, NUTS1 represents the major socio-economic regions, while NUTS2 denotes the intermediate regions commonly used for the implementation of regional policies. In most cases, data were used at the NUTS2 level. However, for Germany and the United Kingdom, this level of disaggregation was considered too detailed, and NUTS1 data were therefore used instead. For several smaller EU countries, subnational data were not available from Eurostat and were therefore obtained from national statistical offices. These include Estonia, Ireland, Lithuania, Latvia, and Slovenia. For Cyprus, Malta and Luxembourg, no subnational statistics could be retrieved.

Eurostat provides time series for different indicators: mean years of schooling (2000–2022), expected years of schooling (2013–2022), GDP in PPP-adjusted Euros (2004–2022), and life expectancy at birth (2004–2022). For the United Kingdom and Hungary, Eurostat does not report mean years of schooling. For these countries, estimates were derived from the European Social Survey (ESS); for Hungary (2010, 2012, 2015) and for the UK (2010, 2012, 2014, 2016). Data on life expectancy for the French overseas departments—Guadeloupe, Martinique, French Guiana, Réunion, and Mayotte—were obtained from Worldstat (2018) for the year 2010.

For other high-income countries—Australia, Canada, China, Croatia, Japan, New Zealand, South Korea, Russia, and the United States—data were taken from national statistical offices. For Russia and South Korea, educational data could not be obtained from this source. Instead, survey-based estimates were used: the European Social Survey (2012, 2017) for Russia and the World Values Survey (2010) for South Korea.

### 2.1.2. GDL Area Database

Since 2016, the Global Data Lab (GDL) ([www.globaldatalab.org](http://www.globaldatalab.org)) has provided open-access subnational development indicators for low- and middle-income countries (LMICs) through its Area Database. These indicators are constructed by aggregating microdata from nationally representative household surveys and population censuses. The principal data sources are the Demographic and Health Surveys (DHS), UNICEF Multiple Indicator Cluster Surveys (MICS), and census datasets distributed by IPUMS International. These datasets typically include between 50,000 and 100,000 or more respondents and contain detailed information on all household members.

Where none of these sources were available, GDL relied on alternative country-specific surveys or, in some cases, on less comprehensive datasets such as Afrobarometer surveys ([www.afrobarometer.org](http://www.afrobarometer.org)) or Americas Barometer surveys ([www.americasbarometer.org](http://www.americasbarometer.org)), which cover only adult (18+) respondents rather than entire households.

For most LMICs, the GDL Area Database provides both indicators required for the education dimension index: mean years of schooling and expected years of schooling. However, data for the health and income dimensions are usually unavailable in directly suitable form. Consequently, subnational estimates for these dimensions were derived using child mortality and household wealth data from the GDL Area Database.

### 2.1.3. UNDP Database

The third data source used for constructing the Subnational Gender Development Index (SGDI) database is the United Nations Development Programme (UNDP) database of national development indicators (UNDP, 2025c). This dataset, compiled as part of the UNDP's Human Development Programme, provides annual time series from 1990 to 2023 for the Gender Development Index (GDI), the gender-specific versions of the Human Development Index (HDI), and of their respective dimension indices and underlying indicators. It also contains an extensive set of socio-economic,

health, education, demographic, environmental, and related variables. From this database, we derived the national reference values, that were used to standardize the SGDI values around the values presented by the UNDP. Because UNDP coverage for the GDI and its components before the year 2000 is limited to a relatively small number of countries, our SGDI database begins in 2000.

## 2.2. Indicators

### 2.2.1. Education dimension

For the education dimension, gender-disaggregated indicators are required for both the mean years of schooling among adults aged 25 and older, and the expected years of schooling for children entering school at age six. The first indicator reflects the current educational attainment of the adult population, while the second represents the anticipated educational attainment for the next generation.

For most low- and middle-income countries (LMICs), both variables were directly obtained from the GDL Area Database. The expected years of schooling for boys and girls were derived using data on school attendance among children and youth aged 6–24 in each region. For every single-year age group, the proportion of girls attending school was computed and subsequently summed across ages 6 to 24; the same procedure was applied to boys. The total of these percentages indicates the number of years of schooling that a child entering school at age six can be expected to complete, if current enrolment patterns across age groups remain unchanged.

The mean years of schooling indicator in the GDL Area Database is calculated as the average number of completed years of formal education among women and men aged 25 years and above, using data from census and survey microdata. In the majority of these datasets, education is measured in years completed, allowing a straightforward computation of the mean for the 25+ population. In a limited number of cases, however, educational attainment was recorded as highest completed level of education. For such datasets, these categories were converted to equivalent years of schooling based on the duration of each educational level within the specific national education system—typically six years for primary education, nine years for lower or junior secondary, twelve years for upper or senior secondary, fifteen years for a bachelor's degree, and sixteen or seventeen years for a master's degree.

In countries where only adult samples were available, expected years of schooling could not be directly calculated. In such cases, this indicator was estimated by applying the observed regional variation in mean years of education for men and women to the corresponding gender-specific national expected schooling values reported by the UNDP. For this, the following formula was used:

$$E_i = E_n \frac{M_i}{M_n} \quad (1)$$

whereby  $M_i$  and  $E_i$  are the mean and expected years of schooling of region  $i$  and  $M_n$  and  $E_n$  the national values of mean and expected years of schooling in the UNDP database.

For high-income countries (HICs), educational data were primarily drawn from national statistical offices. In many of these cases, however, figures on expected years of schooling were not available, as national offices often do not publish this indicator. Eurostat provides an exception, as it reports subnational data for many EU member states, including the number of girls and boys enrolled in school by age and total population by age, allowing the computation of age-specific enrolment rates and thus expected years of schooling. Consequently, this indicator could be included for most EU countries.

Data on mean years of schooling from Eurostat and national statistical offices are generally presented as distributions of girls and boys across educational levels. These distributions were converted into years of education using information on the typical duration of each level, following the same conversion scheme as described above.

For 23 countries—including Australia, Chile, Cape Verde, Ecuador, Ireland, Canada, China, Cuba, Estonia, Croatia, Japan, South Korea, Kuwait, Lebanon, Libya, Lithuania, Malta, Mauritius, New Zealand, Russia, Saudi Arabia, Slovenia, and the United States—data on expected years of schooling were not available. For these countries, regional variation in mean years of schooling was applied to the national UNDP values for expected years of schooling. In Latvia, neither mean nor expected years of schooling were available; therefore, national UNDP values for both indicators were assigned to all subnational regions.

### 2.2.2. Standard of Living Dimension

For the standard of living dimension of the Human Development Index (HDI), the indicator used is the natural logarithm of Gross National Income per capita (LGNIc), expressed in 2021 US dollars adjusted for purchasing power parity (PPP)(UNDP, 2025b). For high-income countries (HICs) and several middle-income countries (MICs), subnational LGNIc values were derived from Eurostat and national statistical offices.

In many cases, the available data differed from the required specification — for example, data referred to Gross Domestic Product per capita (GDPc) instead of GNIC, were expressed in national currencies rather than 2021 PPP-adjusted US dollars, or lacked PPP corrections altogether. These inconsistencies are not problematic for comparability, as all subnational data were normalized using national LGNIc values that adhered to the UNDP definition.

For most low- and middle-income countries (LMICs), estimates of living standards were obtained from the GDL Area Database. Since household surveys and censuses in LMICs typically do not contain reliable income data -- and where such data exist, their quality tends to be low in poorer regions -- subnational LGNIc values were estimated indirectly using household wealth information. This was achieved through the International Wealth Index (IWI; Smits & Steendijk, 2015), which measures household economic status based on asset ownership, housing characteristics, and access to essential services. The IWI is scaled from 0 to 100, with 0 representing ownership of none of the listed assets and very poor housing and service access, and 100 representing ownership of all assets and access to high-quality housing and services.

To convert IWI values into LGNIc estimates, a regression model developed by Smits and Permanyer (2018) was applied. These authors tested several models predicting national LGNIc (as reported by the UNDP HDI database) from national-level IWI scores (derived from the GDL Area Database). The best-performing model was linear in IWI, and included controls for year and world region, achieving an explained variance of 82.6%, which indicates an excellent fit for a predictive model (Studenmund, 2017). The regions distinguished in this model are: Central America and the Caribbean, South America, West Africa, Central Africa, Southern Africa, East Africa, the Middle East and North Africa, Central Asia, South Asia, East and South-East Asia and the Pacific, and Eastern Europe. Additionally, a dummy variable was included for oil-exporting countries—namely East Timor, Equatorial Guinea, Gabon, Kazakhstan, Kuwait, Saudi Arabia, and Turkmenistan—to account for their disproportionately high GNIC levels.

Because statistical offices and Eurostat do not present gender-specific information on GDPc and GNIC and the IWI is only available at household level, it was not possible to obtain separate female and male versions of GNIC at the level of subnational regions. We therefore used the gender-specific national GNIC values from the UNDP HDI database to create gender-specific subnational versions. This was done by applying the percentage differences between the UNDPs national female and male GNIC versions and the UNDPs national GNIC level to the subnational GNIC versions, as explained in the next formula:

$$SGNIC_x = SGNIC * \frac{GNIC_x}{GNIC} \quad (2)$$

where by  $SGNIC_x$  are the female ( $x=1$ ) and male ( $x=2$ ) subnational versions of GNIC,  $SGNIC$  is the average subnational version of GNIC,  $GNIC_x$  are the female ( $x=1$ ) and male ( $x=2$ ) national versions of GNIC and  $GNIC$  is

the average national version of GNIC. The national versions are derived from the UNDP-HDI Database and SGNIC is estimated following the approach used by Smits & Permanyer (2019).

### 2.2.3. Health Dimension

For the health dimension of the Human Development Index (HDI), the indicator employed is life expectancy at birth (LEXP). In high-income countries (HICs), subnational estimates of LEXP were obtained from national statistical offices and Eurostat, which provide gender-disaggregated life expectancy data. For low- and middle-income countries (LMICs), estimates were drawn from the GDL Area Database.

Since household surveys and population censuses typically lack direct information on life expectancy, subnational LEXP values were indirectly estimated from under-five mortality rates (U5M) for girls and boys. To derive life expectancy from under-five mortality, a regression model developed by Smits and Permanyer (2019) was applied. These authors evaluated a series of models predicting national life expectancy (as reported by the UNDP HDI database) based on national averages of U5M derived from the GDL Area Database. The optimal model included both linear and quadratic terms for U5M (U5M and U5M<sup>2</sup>), along with controls for year, world region, and a dummy variable capturing Rwanda's exceptional mortality pattern. This specification achieved an adjusted R<sup>2</sup> of 89.1%, indicating a very strong predictive fit (Studenmund, 2017).

To estimate the gender-specific versions, the prediction model was estimated for men and women separately and the gender-specific coefficients used to estimate the gender-specific versions of LEXP. The explained variance in terms of adjusted R<sup>2</sup> was 87.3 for men and 90.4 for women.

For 16 countries (Argentina, Bosnia and Herzegovina, Barbados, Cape Verde, Costa Rica, Ecuador, Croatia, Kuwait, Lebanon, Libya, Malta, Mauritius, Malaysia, Panama, Saudi Arabia, Uruguay) no data on life expectancy was available. For these countries, the national UNDP values were used for the subnational regions.

## 2.3. Data Processing

### 2.3.1. Scaling the Indicators

To obtain the best possible estimates for the eight indicators given the data limitations, we have taken their national values from the UNDP-HDI database and scaled the subnational values in such a way that their population weighted mean for a given year equals the national UNDP value for that year. In this way we have created indicators that on the one hand display as well as possible the male and female subnational variation of the our data, while on the other hand their population weighted national averages equal the gender-specific values used by the UNDP in constructing the GDI.

For each country-year-gender-SGDI indicator combination, we use a multiplicative scaling coefficient that inflates/deflates the subnational estimates in such a way that their population-weighted averages coincide with the corresponding national UNDP value. By definition, these scaling coefficients take the value of one when no re-scaling is necessary.

### 2.3.2. Addressing Missing Years

Given that household surveys and censuses are not held every year, for many countries the indicators are only available for a restricted number of years. To obtain their values for the whole period 1995-2015, the missing information had to be estimated by interpolation or extrapolation. This estimation process was facilitated by the fact that the UNDP Database contains the national values for all eight indicators for each year in this period, which means that only the subnational variation had to be interpolated or extrapolated.

If information on the indicator value for both a preceding and succeeding year was available, we interpolated the national and subnational values linearly and subsequently applied the obtained variation to the national UNDP value of the indicator for the given year. As no other information for the missing years was available, presuming that in each of the in-between years the indicator values

changed in a similar way does seem a reasonable assumption, which has also been used in other data projects (Smits & Permanyer, 2019; Crombach & Smits, 2024).

If only a preceding or succeeding value was available, we applied the subnational variation of the nearest year for which information was available to the national UNDP value of the indicator for the given year. In this way indicators were obtained that at the national level exactly follow the variation in the UNDP indicators, whereas the subnational variation is in line with the closest year(s) for which real subnational data is available.

#### 2.4. Dimension Indices and SGDI

After having constructed the database with the eight gender-specific indicators, the dimension indices could be computed. This was done in the same way as the dimension indices of the regular HDI are computed by the UNDP (2025b), by using the following formula:

$$\text{Dimension index} = \frac{\text{actual value} - \text{min value}}{\text{max value} - \text{min value}} \quad (2)$$

The minimum and maximum reference values, known as “goalposts”, are applied to ensure that all dimension indices remain standardized within the 0–1 range. For the standard of living dimension, the UNDP defines the lower and upper goalposts as 100 and 75,000, respectively. In the case of life expectancy at birth, the reference values are 22.5 and 87.5 years for women, and 17.5 and 82.5 years for men. The slightly higher goalposts for women reflect their generally longer life expectancy.

For the education dimension, the goalposts are set at the same values for males and females, with 0 and 18 years for expected years of schooling and 0 and 15 years for mean years of schooling. The overall education index is then derived as the geometric mean of the two separate indices for expected and mean years of schooling. For a few regions, the value of one of the education indicators was higher than the maximum goalpost. In these cases, the goalpost levels were taken.

Next, the gender-specific versions of the SHDI were computed by taking the geometric mean of the three gender-specific indices using the following formula:

$$\text{SHDI}_g = (\text{Education}_g \cdot \text{Health}_g \cdot \text{Income}_g)^{1/3} \quad (3)$$

In this equation,  $g$  indicates whether the indicators are for females ( $f$ ) or males ( $m$ ). The geometric mean was taken instead of the arithmetic mean to take care that a high score can only be obtained if a region performs well on all three indices (UNDP, 2025b).

To compute the SGDI on the basis of the male and female SHDI, the quotient of the gender-specific indices is taken, thus:

$$\text{SGDI} = \frac{\text{SHDI}_f}{\text{SHDI}_m} \quad (4)$$

If the male and female SHDIs are equal, the SGDI has value 1. A value below 1 indicates a higher level of human development for males in the region and a value above 1 a higher value for females in the region.

### 3. SGDI Database

The SGDI computed as described above, together with the gender-specific SHDIs, the gender-specific dimension indices and the gender specific indicators are brought together in the Subnational Gender Development Index Database (SGDI Database), which can be downloaded as an Excel file connected as Supplementary Material to this paper.

The SGDI Database contains the subnational and national values of the these indexes and indices for each year between 2000 and 2023 for 1810 subnational regions in 166 countries. Given that the

subnational indicators were standardized around the corresponding national UNDP values, their national values are equal to those in the UNDP HDI database (UNDP, 2025c).

The SGDI Database is also available at the Global Data Lab website ([hdi.globaldatalab.org](http://hdi.globaldatalab.org)), where also visualizations can be made and the data can be connected with other indicators.

**Acknowledgments:** We are grateful to Eurostat and many national statistical offices and the Global Data Lab for making the subnational indicators available for the construction of this database.

## References

- Bras, H & J. Smits (2021). Contexts of Reproduction. Gender Dynamics and Unintended Pregnancy in sub-Saharan Africa. *Journal of Marriage and Family*, 84(2), 438-456.
- Bray, Freddie, Ahmedin Jemal, Nathan Grey, Jacques Ferlay & David Forman. (2012). "Global Cancer Transitions According to the Human Development Index (2008-2030): A Population-Based Study." *The Lancet Oncology* 13(8): 790–801.
- Crombach, L & J. Smits (2024). The Subnational Corruption Database: Grand and petty corruption in 1,473 regions of 178 countries, 1995-2022. (*Nature*) *Scientific Data*, 11, 686.
- Eurostat (2024). Eurostat database. Available at <http://ec.europa.eu/eurostat/data/database> (Statistical office of the European Union).
- Fernández, Raquel, Asel Isakova, Francesco Luna & Barbara Rambousek (2021). Gender Equality and Inclusive Growth. IMF Working Paper WP21/59.  
<https://www.elibrary.imf.org/downloadpdf/view/journals/001/2021/059/article-A001-en.pdf>
- Filho, W., Kovaleva, M., Tsani, S. et al. (2023). Promoting gender equality across the sustainable development goals. *Environ Dev Sustain* 25, 14177–14198. <https://doi.org/10.1007/s10668-022-02656-1>
- Forman-Rabinovici, Alizia & Udi Sommer (2018). An impediment to gender Equality? Religion's influence on development and reproductive policy. *World Development*, 105, 48-58,
- Foster, James, Luis Lopez-Calva & Miguel Szekely (2005). "Measuring the Distribution of Human Development: Methodology and an Application to Mexico." *Journal of Human Development* 6(1): 5–25.
- Ghosh, Taniya & Sanika S. Ramanayake (2020). The macroeconomics of gender equality. *International Journal of Financial Economics*, 26, 1955–1977.
- Grimm, M., Harttgen, K., Klasen, S. & Misselhorn, M. (2008). A Human Development Index by Income Groups. *World Development*. 36, 2527-2546.
- Grimm, M., Harttgen K., Klasen, S., Misselhorn, M., Munzi, T. & Smeeding, T. (2011). Inequality in Human Development: An Empirical Assessment of 32 Countries. *Social Indicators Research*. 97, 191-211.
- Guthridge, M., Kirkman, M., Penovic, T. et al. (2022). Promoting Gender Equality: A Systematic Review of Interventions. *Soc Just Res* 35, 318–343. <https://doi.org/10.1007/s11211-022-00398-z>
- Harttgen, Kenneth, and Stephan Klasen. (2011). "Household-Based Human Development Index." *World Development* 40(5): 878–899.
- Hossain, D., Islam, R. (2024). A systematic literature review on the relationship between education and gender ideology . *Discov glob soc* 2, 39. <https://doi.org/10.1007/s44282-024-00071-7>.
- Klugman, J., Rodríguez, F. & Choi, H. J. (2011). The HDI 2010: New Controversies, Old Critiques. *Journal of Economic Inequality* 9, 249–288.
- Kummu, M., Taka, M. & Guillaume, J.H.A. (2018). Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015. *Scientific Data*. 5:180004, 1-15.
- Myrskylä, Mikko, Hans-Peter Kohler, and Francesco Billari (2009). "Advances in Development Reverse Fertility Declines." *Nature* 460: 741–743.

- Permanyer, I. (2013). Using Census Data to Explore the Spatial Distribution of Human Development. *World Development*, **46**, 1-13.
- Permanyer, I., García, J., Esteve, A. & McCaa, R. (2015). Human Development Index-like Small Area Estimates for Africa compute from IPUMS-International integrated census microdata. *Journal of Human Development and Capabilities*, **16**, 245-271.
- Permanyer, I. & J. Smits. (2020). Inequality in human development across the globe. *Population and Development Review*, **46**(3), 583-601.
- Santos Silva & M., Klasen, S. (2021). Gender inequality as a barrier to economic growth: a review of the theoretical literature. *Rev Econ Household* **19**, 581–614. <https://doi.org/10.1007/s11150-020-09535-6>
- Smits, J. (2016). GDL Area Database. Sub-national development indicators for research and policy-making. GDL Working paper 16-101.
- Smits, J. & Huisman, J. (2025). From National Averages to Local Realities: A Subnational Vulnerability Index to Guide Sustainable Development in Low- and Middle-Income Countries. *Sustainability*, **17**.
- Smits, J. & Permanyer, I. The Subnational Human Development Database. *Sci Data* **6**, 190038 (2019). <https://doi.org/10.1038/sdata.2019.38>.
- Smits, J. & Steendijk, R. The International Wealth Index (IWI). *Social Indicators Research*, **122**(1), 65-85 (2015).
- Studenmund, A. H. (2017). *Using Econometrics: A Practical Guide*, Pearson.
- UN (2018). Turning promises into action. Gender Equality in the 2030 agenda for sustainable development. [www.unwomen.org/en/digital-library/sdg-report](http://www.unwomen.org/en/digital-library/sdg-report)
- UN (2024). Progress on the Sustainable Development Goals: The gender snapshot 2024. United Nations. <https://www.unwomen.org/sites/default/files/2024-09/progress-on-the-sustainable-development-goals-the-gender-snapshot-2024-en.pdf>
- UNDP (2025a). Human Development Report 2025: A matter of choice: People and possibilities in the age of AI. <https://hdr.undp.org/content/human-development-report-2025>
- UNDP (2025b). Technical notes. Human Development Report 2025. [https://hdr.undp.org/sites/default/files/2025\\_HDR/HDR25\\_Technical\\_Notes.pdf](https://hdr.undp.org/sites/default/files/2025_HDR/HDR25_Technical_Notes.pdf)
- UNDP (2025c). Human Development Reports Data Center. Available at <http://hdr.undp.org/en/data>
- Wildeman, J., J. Smits & S. Schrijner (2023). Ethnic variation in fertility preferences in sub-Saharan Africa. *Population Research and Policy Review*, **42**.
- World Economic Forum (2025). Global Gender Gap Report 2025. <https://www.weforum.org/publications/global-gender-gap-report-2025/>
- Worldstat (2018). *Worldstat database*. Available at <http://en.worldstat.info>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.