




genderstat: An R Package for Comprehensive Gender Inequality Analysis

S M Mashrur Arafin Ayon¹

¹ Department of Women and Gender Studies, University of Dhaka

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Susan Holmes](#) 

Reviewers:

- [@thegargiulian](#)
- [@joaojcorreia](#)

Submitted: 02 September 2023

Published: unpublished

License

Authors of papers retain copyright
and release the work under a
Creative Commons Attribution 4.0
International License ([CC BY 4.0](#))

Summary

The genderstat package is an exhaustive tool developed for the R ([R Core Team, 2023](#)) programming environment, explicitly devised to expedite quantitative evaluations in the field of gender studies. It encompasses an array of functions that facilitate the computation of pivotal gender disparity indicators, including the *Gender Pay Gap (GPG)*, *Gender Inequality Index (GII)*, *Gender Development Index (GDI)*, and *Gender Empowerment Measure (GEM)*. Moreover, it incorporates authentic datasets ([Arafin Ayon, 2023](#)), derived from the UNDP Human Development Reports Data Center and the World Bank Gender Data Portal, which can be leveraged for pedagogical and practical applications.

This package is fundamentally engineered to support scholars in the social sciences engaged in gender-centric research. Being a freely accessible open-source resource from R ([R Core Team, 2023](#)), it does not incur any financial burdens. Moreover, by being open-source, it supports the ethos of collaborative knowledge sharing and innovation that underpin open-source initiatives and their core philosophy. This provision enables researchers to focus more intently on their profound theoretical explorations rather than dedicating an exorbitant amount of time to craft functions for data visualization or resorting to expensive proprietary alternatives. The genderstat package is endowed with the requisite functionalities to scrutinize and visualize gender metrics, thereby serving as a specialized asset for gender analysis.

Additionally, the genderstat package relies on other R ([R Core Team, 2023](#)) packages, namely ggplot2 ([Wickham, 2016](#)), dplyr ([Wickham et al., 2020](#)), and tidyr ([Wickham et al., 2023](#)), to provide its full range of functionalities. These dependencies are crucial for the data manipulation, transformation, and visualization tasks that genderstat performs.

Statement of Need

For researchers probing into the intricacies of gender-related issues, it is imperative to meticulously dissect both the quantitative and qualitative dimensions of any manifestation of discrimination. While genderstat is not tailored for qualitative inquiries, it occupies a unique niche as the inaugural, and hitherto, *the pioneer and sole dedicated R package* accessible on **CRAN** for the analysis and visual representation of gender metrics. Born from the spirit of the foundational tenets of open-source software, this package bestows upon researchers the capability to scrutinize the contributory elements underpinning discrimination in the societal fabric. This facility not only liberates researchers from the constraints of paid proprietary software, thereby democratizing the process of gender analysis but also enables a shift in focus towards more substantive aspects of their research, minimizing the time and effort expended on the analytical component. Ultimately, genderstat serves as a pivotal tool in the researcher's arsenal, aiding in the elucidation and subsequent eradication of gender-based disparities.

Core Theoretical Concepts

Gender Pay Gap:

The Gender Pay Gap (Miller & Vagins, 2018) is a significant metric that encapsulates the differential in the mean earnings of men and women, typically expressed as a percentage. It serves as a lens through which the variance in the remunerations of men and women in the workforce is discerned. A larger percentage denotes a more pronounced incongruity in the earnings of the two genders, while a 0% pay gap symbolizes absolute parity in compensation. As a pivotal barometer of gender inequality, the gender pay gap is shaped by a multitude of determinants, encompassing not merely discrimination, but also disparities in the nature of occupations, the quantum of hours worked, and the opportunities for career advancement (Miller & Vagins, 2018).

The Gender Pay Gap is computed employing the subsequent formula:

$$\text{GPG} = \left(\frac{\text{Male Income} - \text{Female Income}}{\text{Male Income}} \right) \times 100\%$$

Gender Inequality Index:

The Gender Inequality Index (GII) is an intricate composite measure meticulously designed to encapsulate the disparities engendered by gender across three pivotal dimensions: reproductive health, empowerment, and participation in the labor market (Ferrant, 2010). Its value oscillates between 0 and 1; a score of 0 indicates perfect gender parity across the analyzed dimensions, while a score of 1 signifies the apogee of gender inequality. The computation of the GII entails taking the harmonic mean of the geometric means across the aforementioned dimensions for both genders, subsequently dividing this result by a reference standard premised on treating both genders equitably (Conceição et al., 2022). The resultant figure is then subtracted from 1.

The streamlined formula for GII is delineated as follows:

$$\text{GII} = 1 - \sqrt[3]{H \cdot E \cdot L}$$

Where:

- H is the health component, calculated using the maternal mortality ratio and adolescent birth rate.
- E is the empowerment component, calculated using the proportion of parliamentary seats occupied by females and the proportion of females with secondary education.
- L is the labor market participation component, calculated using the female and male labor force participation rates.

Gender Development Index:

The Gender Development Index (GDI) is an indispensable metric engineered to elucidate gender disparities across three quintessential dimensions of human development: health, education, and mastery over economic resources. Health is evaluated by the life expectancy at birth for both genders; education is assessed by the anticipated years of schooling for children coupled with the average years of schooling for adults aged 25 and older, for both sexes; and economic dominion is appraised by the estimated earned income for females and males. This pivotal index plays a crucial role in pinpointing disparities between the sexes, thereby steering policymakers in devising strategies that foster gender equality and empower all individuals (Charmes & Wieringa, 2003).

The computation of the GDI entails a four-step procedure, encompassing the estimation of female and male earned incomes, normalization of indicators, computation of female and male

81 Human Development Index (HDI) values, and ultimately, the juxtaposition of female and male
82 HDI values(Conceição et al., 2022). The formula for GDI is delineated as follows:

83 The HDI for each gender is calculated using the formula:

$$\text{HDI} = (\text{health} \cdot \text{education} \cdot \text{standard of living})^{1/3}$$

84 Finally, the GDI in this package is calculated using the formula:

$$\text{GDI} = \frac{\text{female HDI}}{\text{female HDI} + \text{male HDI}}$$

85 **Gender Empowerment Measure:**

86 The Gender Empowerment Measure (GEM) is a pivotal metric meticulously crafted to scruti-
87 nize gender inequality across cardinal domains of participation and decision-making. It is
88 predominantly geared towards evaluating women's empowerment in political representation,
89 economic decision-making echelons, and income distribution(Charmes & Wieringa, 2003).This
90 results in a value between 0 and 1, where a higher value indicates higher gender empowerment.

91 The GEM is quantified using the subsequent formula:

$$\text{GEM} = \frac{\text{FPS} + \text{FPP} + \text{IR}}{3}$$

92 Where:

- 93 ■ FPS is the proportion of parliamentary seats occupied by females.
- 94 ■ FPP is the proportion of professional and technical job roles occupied by females.
- 95 ■ IR is the income ratio for females in comparison to males.

96 Application

97 In the forthcoming section, rudimentary instances elucidating the application of the genderstat
98 package will be tendered. Prior to that, the fundamental architecture of the package is
99 delineated in the subsequent illustration.

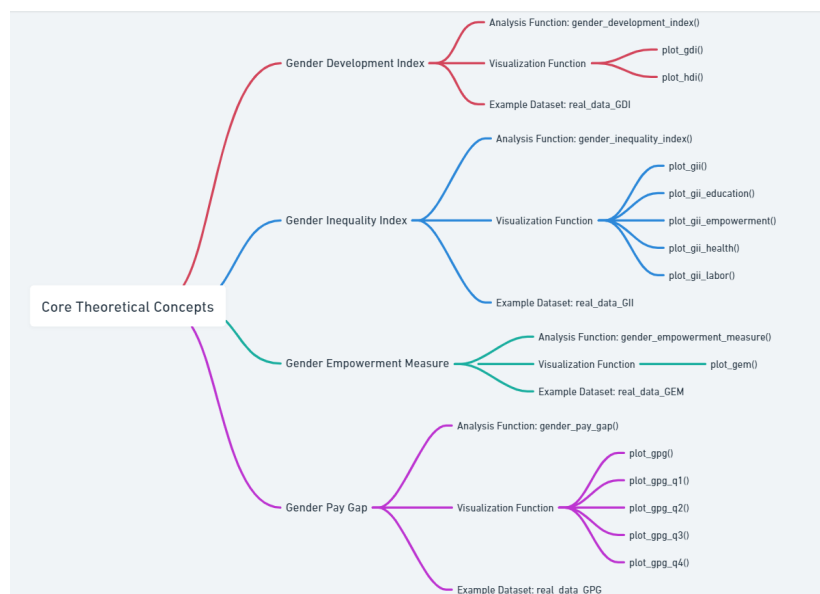


Figure 1: Basic Structure of genderstat package

Examples

In the following, five exemplar applications of the genderstat package are delineated. These examples are meticulously curated to assist users in comprehending the utilization of this package, whilst simultaneously illuminating the simplicity and efficiency facilitated by minimalistic single-line coding.

In these instances, the sample datasets: *real_data_GPG*, *real_data_GII*, and *real_data_GDI* are employed to maintain simplicity. However, users wishing to load data from CSV files are encouraged to utilize the *readr* (Wickham et al., 2021) package, while those aiming to modify data are advised to employ the *dplyr* (Wickham et al., 2020) package as a comprehensive solution.

With genderstat, you can perform a range of visualizations and analyses:

Analytical Functions

Gender Development Index (GDI):

```
# Install the packages if you haven't already
install.packages("ggplot2")
install.packages("dplyr")
install.packages("readr")
install.packages("genderstat")
```

```
# If you want to obtain the latest version of genderstat package from GitHub.
# Use this code instead
install.packages("devtools")
devtools::install_github("mashrur-ayon/genderstat", subdir = "genderstat")
```

```
# Load the packages
```

```
library(ggplot2)
```

```
library(dplyr)
library(genderstat)
library(readr)
```

```
# In this case, I am loading example data from the package
# If you want to load your CSV then follow: data <- read_csv("/path/to/your/data.csv")
```

```
gender_development_index(real_data_GDI)
```

113 *Output shown as table:*

No.	Country	GDI	Female HDI	Male HDI
1	Afghanistan	0.455578857	0.403890975	0.482653625
2	Albania	0.508329304	0.8462988	0.818564494
3	Algeria	0.481143309	0.70003075	0.754901153
4	Andorra			
5	Angola	0.463289727	0.54259953	0.628588819
6	Antigua and Barbuda			
7	Argentina	0.506390402	0.845486531	0.824147269

114 **Gender Pay Gap:**

```
# Install the packages if you haven't already
install.packages("ggplot2")
install.packages("dplyr")
install.packages("genderstat")
install.packages("readr")

# Load the packages

library(ggplot2)
library(dplyr)
library(genderstat)
library(reader)

# Finding Gender Pay Gap from the example dataset

gender_pay_gap(real_data_GPG)
```

115 *Output shown as table:*

No.	Country	Gender Pay Gap
1	Afghanistan	82.74522499
2	Albania	30.02405292
3	Algeria	80.04160342
4	Andorra	
5	Angola	23.33387123
6	Antigua and Barbuda	
7	Argentina	40.9273582

116 **Data Visualization and Plot packages**

117 **Plotting Gender Development Index**

```
# Install the packages if you haven't already
install.packages("ggplot2")
install.packages("dplyr")
install.packages("genderstat")
install.packages("readr")
```

```
# Load the packages
```

```
library(ggplot2)
library(dplyr)
library(genderstat)
library(readr)
```

```
# Visualizing the gender development index from the example data
plot_gdi(real_data_GDI)
```

118 *Plot output shown as image:*

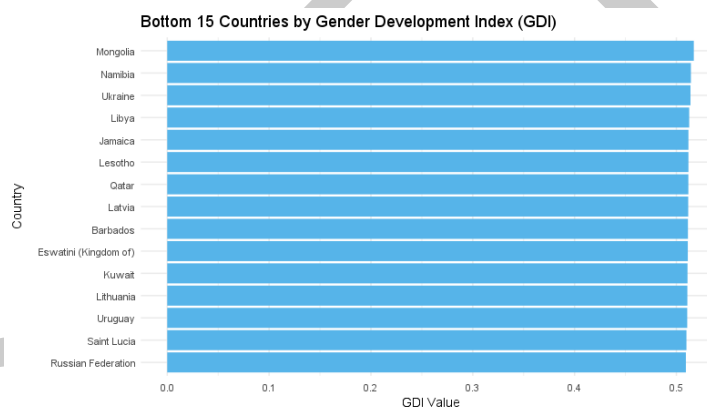


Figure 2: Gender Development Index Plot

119 **Plotting Human Development Index**

120 Comparative Analysis of HDIs: Present a side-by-side comparison of the Male and Female
121 HDIs.

```
# Install the packages if you haven't already
install.packages("ggplot2")
install.packages("dplyr")
install.packages("genderstat")
install.packages("readr")
```

```
# Load the packages
```

```
library(ggplot2)
library(dplyr)
library(genderstat)
library(readr)
```

```
# Plotting Human Development Index from example dataset
plot_hdi(real_data_GDI)
```

122 *Plot output shown as image:*

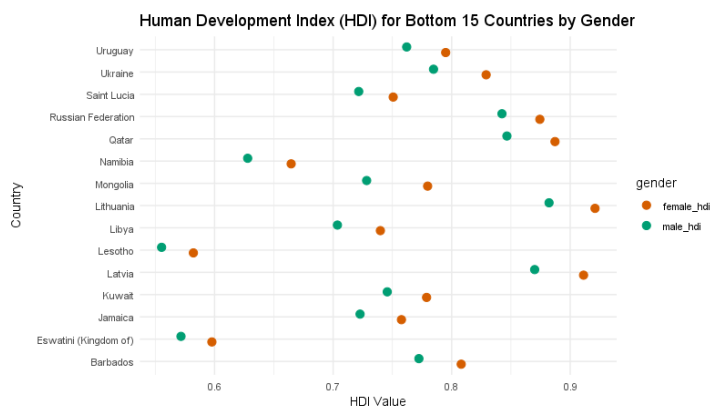


Figure 3: Human Development Index Plot

123 Plotting Health Component of Gender Inequality Index

124 Maternal Mortality Ratio and Adolescent Birth Rate: Use stacked bar plots to visualize these
125 critical health metrics.

```
# Install the packages if you haven't already
install.packages("ggplot2")
install.packages("dplyr")
install.packages("genderstat")
install.packages("readr")
```

```
# Load the packages
```

```
library(ggplot2)
library(dplyr)
library(genderstat)
library(readr)
```

```
#Plotting the health component from the example dataset
plot_gii_health(real_data_GII)
```

126 *Plot output shown as image:*

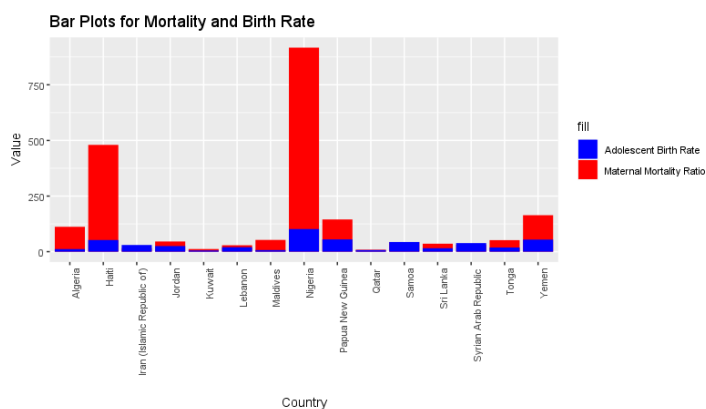


Figure 4: Health Component of Gender Inequality Index

References

- Arafin Ayon, S. M. M. (2023). *Example Dataset for R Package 'genderstat': Gender Disparities Across the Globe - Socio-Economic and Health Metrics (Version V1)* [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/GZLYLI>
- Charmes, J., & Wieringa, S. (2003). Measuring women's empowerment: An assessment of the gender-related development index and the gender empowerment measure. *Journal of Human Development*, 4, 419–435. <https://doi.org/10.1080/1464988032000125773>
- Conceição, P., Calderón, C., Esbry, F. P., Ghorai, M., Hsu, Y.-C., Ismail, G., Lengfelder, C., Lutz, B., Mirza, T., Mohammed, R., Pasanen, J., Shrestha, S. K., Tapia, H., Vázquez, C. R., Yokoi, Y., & Zhang, Y. (2022). *Human development report 2021/2022-technical notes*. UNDP.
- Ferrant, G. (2010). The gender inequalities index (GII) as a new way to measure gender inequalities in developing countries. *Shs.hal.science*. <https://shs.hal.science/halshs-00462463/>
- Miller, K., & Vagins, D. (2018). *The simple truth about the gender pay gap. Fall 2018 edition*. American Association of University Women; American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036. Tel: 800-326-2289; Tel: 202-728-7602; Fax: 202-463-7169; e-mail: foundation@aaup.org; Web site: <http://www.aauw.org>. <https://eric.ed.gov/?id=ED596219>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4
- Wickham, H., Francois, R., Bryan, J., Bearrows, S., Jylänki, J., & Jørgensen, M. (2021). *Read rectangular text data [r package readr version 2.0.0]*. cran.r-project.org. <https://cran.r-project.org/web/packages/readr/index.html>
- Wickham, H., François, R., Henry, L., Müller, K., & RStudio. (2020). *Dplyr: A grammar of data manipulation*. R-Packages. <https://cran.r-project.org/web/packages/dplyr/index.html>
- Wickham, H., Vaughan, D., & Girlich, M. (2023). *Tidyr: Tidy messy data*. <https://tidyr.tidyverse.org>