Multiple Imputation in R: A Method to Bridging Data Gaps in Regional Sustainability Analysis

Antonín Hořčica[1]

Abstract: This study builds on research presented last year on the application of advanced statistical methods in R software to analyze sustainability indicators in EU regions. Using R software, a widely used programming language for statistical calculations and data analysis, the study examines the availability of data from the Eurostat database necessary to calculate the Regional Human Development Index (RHDI) at various NUTS levels. The study evaluates missing data for the period 2000–2020 and explores the potential to impute these values using Amelia software in R, a specialized R package for multiple imputations. This method is used to estimate missing values, enhancing analytical quality. Furthermore, the study validates the methodology for setting limits for the Min-Max method used in the calculation of the subindicators of health, education and income. The data processing methodology recommended by the official RHDI guidelines was verified against actual Eurostat data. The objective was to determine whether Eurostat data could be used effectively to calculate indicators to assess the effectiveness of regional sustainability strategies within the EU.

Keywords: sustainability, R statistics, RHDI, missing data, regional development, NUTS, Amelia
JEL Classification: C14, Q01, R11, R58

1 Introduction

Sustainable development is a key concept in current global politics, playing a fundamental role within the European Union (hereafter the EU), where it is embedded in several strategies. This concept has been implemented in major documents such as *The European Union Strategy for Sustainable Development* (European Commission, 2001) and *The European Green Deal* (European Commission, 2019).

The significance of the sustainability concept lies in the integration of economic development, social cohesion, and environmental protection, which together form the foundation for the long-term prosperity of regions. Within the EU, it is essential to actively monitor progress in this regard. For this purpose, the Regional Human Development Index (RHDI) was developed, initiated by the European Commission, and created by Hardeman and Dijkstra (2014) from the Joint Research Center (n.d.). This index evaluates regions based on three primary components: health, education, and income, which are further divided into six subindicators. The authors selected these components to ensure compatibility with Eurostat data and facilitate meaningful comparisons among EU regions. Although the index is not widely used in practice, it has significant potential due to its components, allowing the assessment of quality of life at the regional level and serves as an effective tool for evaluating regional sustainability. This potential was one of the reasons for selecting this indicator for this study.

Data on regional sustainability are mainly sourced from the Eurostat database, the official statistical office of the European Union (Eurostat, n.d.). Eurostat provides data at various levels of regional division, for which the Nomenclature of Territorial Units for Statistics (NUTS) was established. This system, used for the classification of territorial units within the EU, allows for the comparison of regions not only within individual countries but also across the Union as a whole (NUTS, n.d.). This capability enables the analysis of regional progress in sustainable development and the identification of areas where targeted interventions are necessary at the regional policy level.

For a precise sustainability analysis and effective use of indicators for monitoring and evaluation purposes, complete and reliable data over long periods of time. These data allow not only the comparison of regions in specific years, but also the observation of developmental trends, which is crucial for assessing the impacts of strategic decisions. However, it often happens that some values in time series are missing, necessitating advanced statistical methods to replace these missing values.

[1] University of South Bohemia in České Budějovice, Faculty of Economics, Department of Regional Management, Studentská 13, 370 05 České Budějovice, Czech Republic, ahorcica@ef.jcu.cz

https://doi.org/10.32725/978-80-7694-102-1.02

The methodology for the RHDI recommends using multiple imputation with the Amelia tool in the R software. R is a programming language and working environment specifically designed for statistical computing and data analysis. Its open source nature and wide range of packages allow researchers to handle complex data challenges, such as missing values, in a flexible and efficient way. Amelia is one of these specialized packages available in R, developed to perform multiple imputation, a statistical technique that estimates missing data by generating multiple plausible datasets. This approach preserves the relationships between variables and reduces biases in subsequent analyzes. The aim of this study is to validate this method on data for the various dimensions required to calculate the RHDI at different NUTS levels according to Eurostat. This paper follows the author's previous work on sustainability indicators and their analysis using advanced statistical methods (Hořčica, 2023).

2 Goals

The aim of this study is to verify the availability of data required to calculate the Regional Human Development Index (RHDI) in the Eurostat database for various territorial levels, specifically NUTS 0, NUTS 1, and NUTS 2. This requires analyzing data availability over extended time periods, ideally covering the past two to three seven-year programming periods relevant to EU regional policy. Analysis involves not only assessing the extent of missing data but also identifying the causes of these data gaps, which is essential for finding effective solutions to address these deficiencies.

The primary goal is to propose methods for efficiently imputing missing data, with particular emphasis on validating the methodology recommended by the authors of the RHDI framework, using the Amelia software tool in the R environment. This multiple imputation method will be tested on a data set that includes various NUTS levels to assess its efficacy in filling missing values and thus improving the quality of the analysis.

Based on the obtained and imputed data, the RHDI indicator will be calculated for these different NUTS levels. Additionally, a basic statistical analysis, including functional analysis, to evaluate regional trajectories and identify potential trends and disparities in their development. This functional analysis will serve as a key tool to understand regional differences and the effectiveness of regional EU policies aimed at promoting sustainable development.

3 Methods

At the beginning of the chapter, the methodology for calculating the Regional Human Development Index (RHDI) is introduced and applied to updated Eurostat datasets, ensuring compatibility with current data sources and facilitating the assessment of its applicability in the present context. This is followed by a brief description of Eurostat data sources and their availability according to the NUTS regional classification. In addition, the chapter outlines the procedure for addressing missing data through multiple imputation using the Amelia software package in R. Finally, it explores the possibilities of visualizing the results using functional analysis.

3.1 Calculation methodology of RHDI

The methodology for calculating the Regional Human Development Index (EU RHDI) was determined by Hardeman and Dijkstra (2014). This index is structured around three dimensions, each comprising two subindicators:

- Health: life expectancy at birth and infant mortality.
- Education: young people not in education, employment or training (NEET) and tertiary education.
- Income: household disposable income per capita and employment rate.

To ensure data comparability, each indicator is normalized using the Min-Max method, which scales the original values from 0 to 1. The minimum and maximum values are determined from the entire dataset, extended by a linear interpolation projected 5 years forward.

For indicators with a positive trend, normalization is applied according to formula (1), while for those with a negative trend, formula (2) is used. This ensures compatibility across all indicators.

Positive trend: $x_t = \frac{x_i - \mathrm{Min}(x_n)}{\mathrm{Max}(x_n) - \mathrm{Min}(x_n)}$ (1)   Negative trend: $x_t = \frac{x_i - \mathrm{Max}(x_n)}{\mathrm{Min}(x_n) - \mathrm{Max}(x_n)}$(2),

where x_t represents the value of a subindex component of one of the two components within the relevant dimension. The variable x_i denotes the specific value of the component in original units (years, percentages, or monetary terms), with $\mathrm{Min}(x_n)$ as the minimum and $\mathrm{Max}(x_n)$ as the maximum values of the variable.

The resulting RHDI is then calculated as the geometric mean of the averages of the two indicators for each of the three dimensions:

$$\text{RHDI} = \sqrt[3]{D_{\text{Health}} \cdot D_{\text{Education}} \cdot D_{\text{Income}}} \ (3),$$

where RHDI represents the final regional index, and the variables D_{Health}, DEducation, and D_{Income} refer to the indices for dimensions of health, education, and income, respectively.

3.2 Data sources

3.2.1 Eurostat

Eurostat (n.d.) is the statistical office of the EU, responsible for collecting and harmonizing data from Member States and ensuring their comparability for analyses at the EU level. These data are essential for the development and evaluation of policies targeting economic growth, employment, social cohesion, and regional development. Eurostat provides statistics through a public database that covers areas such as demographics, economy, environment, and education. Eurostat also plays a critical role in publishing data at the regional level according to the NUTS nomenclature, which enables the monitoring of regional disparities and developments across the EU.

3.2.2 Nomenclature of territorial units for statistics (NUTS)

The Nomenclature of Territorial Units for Statistics (NUTS, n.d.) is a system used within the EU for standardized territorial division of Member States, facilitating international and regional analysis. The NUTS system is organized into several levels, from national to regional, encompassing the following categories:

- NUTS 0: Includes Member States such as the Czech Republic or Germany.
- NUTS 1: Comprise large regions, which may correspond to federal states (e.g., Bavaria in Germany), autonomous communities (e.g., Catalonia in Spain), macroregions (e.g., the South of Italy), or other types of larger territorial units. Some countries have only one region at this level, such as the Czech Republic, which is considered a single region at the NUTS 1 level.
- NUTS 2: Corresponds to medium-sized territorial units often utilized for European regional programs and funding. In the Czech Republic, this includes grouped regions like the Southeast or Northwest. Certain smaller countries correspond to the NUTS 2 level as a whole, such as Malta, Cyprus, and Lithuania.
- NUTS 3: Refers to smaller regional units. In the Czech Republic, this corresponds to regions, such as the South Bohemian Region or the South Moravian Region.

The NUTS system enables precise regional comparisons, the allocation of resources from EU funds, and the evaluation of the impacts of regional policies.

3.2.3 Programming periods

EU programming periods are seven-year time frames designated for implementing and financing EU policies in the areas of structural funds, regional development, and cohesion. Each period has specific budgetary priorities and includes operational programs at both the national and regional levels. Examples of programming periods include 2007-2013, 2014-2020, and 2021-2027. Long-term analysis in multiple programming periods allows tracking of policy impacts and the analysis of regional disparities and their development over time.

3.3 Data availability analysis

Analysis of data availability for calculating the RHDI focuses on different NUTS levels, specifically NUTS 0, NUTS1 and NUTS 2. For each of these levels, the data required for each dimension of the index (health, education, income) and their subindicators are examined. The analysis is focused on the programming periods 2000-2006, 2007-2013, and 2014-2020.

3.3.1 Identification of available data for each NUTS level

In the first phase of the analysis, the available data for each of the RHDI indicators collected by Eurostat are identified. For each NUTS level, an overview is provided to determine whether relevant data are available for the selected programming periods and to assess their completeness. The focus is on key indicators such as life expectancy, infant mortality, the NEET rate, tertiary education, disposable income and employment rate.

3.3.2 Causes of missing data

Missing data may have various causes, which are analyzed in this section. The main reasons for missing data can include methodological factors, where certain regions do not meet the criteria for tracking a given indicator or are not systematically measured. Political reasons may involve disparities in national statistics or insufficient harmonization of data collection among EU member states. Administrative reasons can be related to the lack of resources at regional or national statistical offices, leading to improper or insufficient data recording.

3.4 Methods for imputing missing data

Data imputation methods are generally divided into simple and advanced methods. Simple methods include directly filling in missing data based on values from reliable external sources or using average values, for example, from higher territorial classifications, or utilizing historical data even after structural changes to regions, such as historical adjustments in NUTS regions. However, it is essential to consider whether these methods introduce bias, as they may not account for intervariable relationships.

To preserve the structure of the data, multiple imputation is recommended, employing advanced algorithms that maintain the variance and relational structure of the missing data. The R software documentation (R Project, n.d.) describes and provides access to the Amelia package (R Project. Amelia, n.d.), which is also recommended by the authors of the RHDI methodology (Hardeman and Dijkstra, 2014). An objective of this study was to verify its practical applicability to current Eurostat data.

The Amelia package, developed by Honaker, King and Blackwell (2011), requires specification of parameters and distributions for individual variables, allowing it to retain intervariable relationships and estimate likely values based on existing data. Amelia is suitable for temporally ordered data, making it applicable to Eurostat data. Through multiple imputation, Amelia generates several versions of the dataset with varied imputed values, which enhances the estimation of missing data and reduces bias. Amelia employs the Expectation-Maximization (EM) algorithm combined with bootstrapping for iterative filling of missing values, minimizing noise and producing more reliable results than simple imputation. In practice, the imputed datasets are typically averaged. The algorithm further considers temporal sequence and/or geographic location, essential for accurately imputing missing values in complex data structures like economic and social data.

The procedure for applying multiple imputation to subindicators for the RHDI calculation includes data preparation, variable selection, and missing value identification. The data must be converted to a long format, where each combination of variable and time period (e.g., year and region) occupies a separate row, allowing the algorithm to leverage temporal and spatial relationships during imputation. Setting key parameters, such as the number of iterations (typically 5 to 10), is followed by applying the EM algorithm in the Amelia package, which iteratively fills missing values until convergence. Finally, the results can be validated through visualization, which confirms whether the imputed data align with variable trends in the time series.

3.5 Initial functional analysis

The imputed data were used in the R environment to calculate the values for individual dimensions and RHDI at NUTS 0, NUTS 1, and NUTS 2 levels. Basic functional analysis was conducted using the envelope method (Mrkvička, T., 2017), which identifies trends and outliers among regions. This method, included in the GET package by Myllymäki, M., & Mrkvička, T. (2020), provided data visualization and helped identify regions with different trajectories, potentially serving as a basis for regional interventions and sustainability strategies (Hořčica, 2023).

4 Research results

This chapter presents the results of the verification of the data sources and the availability of key indicators. Additionally, the scope and causes of missing data are analyzed, including their imputation using the Amelia package. Finally, a description of the calculation and values of the RHDI index for Czech regions is provided, along with visualization through functional analysis.

4.1 Data sources

Data were downloaded from the Eurostat database, as shown in Table 1, which lists the available periods. The data are current as of the last download in February 2024. Additional periods may be available, although they have not yet been analyzed. The analysis selected the period 2000–2021, corresponding to three programming periods.

Table 1 Overview of key indicators for the calculation of RHDI

No.	Indicator name	Eurostat code	Abbrevia-tion	Period of avail.*
1	Life expectancy	demo_r_mlifexp	LifeExp	1990 - 2020
2	Infant mortality rates	demo_r_minfind	InfMor	1990 - 2020
3	Young people neither in employment nor in education and training (NEET rates)	edat_lfse_22	NEET	2000 - 2021
4	Population by educational attainment level	edat_lfse_04	TertEdu	2000 - 2022
5	Income of households	nama_10r_2hhinc	Income	2000 - 2020
6	Employment rates	lfst_r_lfe2emprt	EmpRat	1999 - 2021

Source: own processing based on EUROSTAT data.

4.2 Analysis of missing data

In the initial review, inconsistent data were identified for certain regions due to historical changes in their geography. This was observed, for example, in Croatian regions, where, in 2016 (before Croatia's EU accession), the original HR04 Kontinentalna Hrvatska was divided into new regions: HR02 Panonska Hrvatska, HR05 Grad Zagreb, and HR06 Sjeverna Hrvatska, which lacked historical data. Therefore, for calculation purposes, the original regions HR03 Jadranska Hrvatska and HR04 Kontinentalna Hrvatska were retained.

Similar adjustments were necessary for the Irish regions IE04 Northern and Western, IE05 Southern and IE06 Eastern and Midland, as the well as Slovenian regions SI03 Vzhodna Slovenija and SI04 Zahodna Slovenija, where geographic changes had also occurred.

For some regions, such as Malta for the Income indicator, FRY5 Mayotte, and FI20 Åland for NEET, data were missing for the entire period. For Malta, which has the same data across all three observed levels, household income data in USD were available in the CEIC database (n.d.) for the years 2011 to 2020 per capita, which could be converted using Eurostat coefficients to match the value format used for other regions in PPS 2020. These data were added to all three datasets. For the region FRY5 Mayotte, NEET and TertEdu data were unavailable from 2000 to 2020, so these data were excluded from the NUTS 2 dataset. For the Finnish region FI20 Åland, data were unavailable for the NEET indicator; thus, this region was excluded from the NUTS 1 and NUTS 2 datasets.

Data for each indicator were available for EU27 Member States and for regions at NUTS 1 and NUTS 2 levels. The processed data were analyzed in terms of the frequency of missing values, as shown in Table 2, which presents the data after the dataset adjustments described above. Percentages for EU27 were calculated based on the total number of values for each variable. To calculate the overall percentage of missing values, it was necessary to consider all six variables for each region.

Table 2 Summary of missing data counts and percentages for various regional levels

Variable	EU27		NUTS 1		NUTS 2	
	Count	%	Count	%	Count	%
LifeExp	3	0,53	114	5,97	378	7,56
InfMor	2	0,35	96	5,02	324	6,48
NEET	13	2,29	69	3,61	450	9,00
TertEdu	2	0,35	22	1,15	270	5,40
Income	11	1,94	11	0,57	13	0,26
EmpRat	2	0,35	9	0,47	213	4,26
Total	33	0,97	342	2,98	1759	5,86
Indicator dataset	567		1911		4 998	
Complete dataset	3 402		11 466		29 988	
Countries/regions	27		91		238	

Source: own processing based on EUROSTAT data.

4.3 Imputation procedure

The data for all six indicators were converted to long format and uploaded to R, where basic statistical analysis was performed; these statistics, along with the raw and processed data, are available from the author of this article. The imputation calculations were conducted using the amelia function. Below is a sample of the code from the Amelia package in R software:

```
a.out.time_NUTS2 <- amelia(NUTS2_df, ts = "Year", cs = "NUTS2", polytime = 1, in-
tercs = TRUE, p2s = 2, m=10)
```

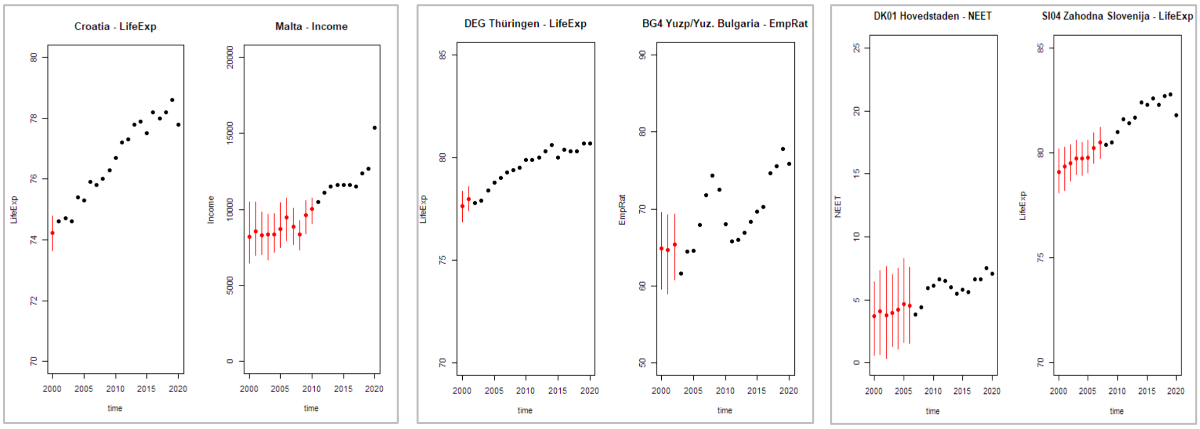
The `amelia()` function is the core function for multiple imputation of missing data. `NUTS2_df` is a time series with missing values subject to imputation, where `ts = "Year"` specifies the time dimension, and a `cs = "NUTS2"` specifies the cross-sectional dimension, containing region names. The parameter `polytime = 1` is used for linear polynomial transformation to model the time trend; other parameters allow for modification, such as setting the parameter to 2 for a quadratic polynomial, with higher values increasing the polynomial trend degree. The setting `intercs = TRUE` enables interaction between temporal and regional variables, and `p2s = 2` sets the Bayesian adjustment of variance to none, weak, or strong. The parameter `m=10` determines the number of imputations. For large datasets, such as NUTS2, the number of iterations can reach hundreds or thousands, with computation times extending to several hours.

The imputed data can be displayed graphically using the `tscPlot()`, unction, as illustrated in Figure 1. Selected examples of imputed data for various regions and indicators are shown here.

Data validation was conducted using basic statistical methods to identify unrealistic values and assess outliers, skewness, and kurtosis according to methodological recommendations. The analysis was performed on the most extensive dataset for the NUTS 2 region, with basic statistics presented in Table 3. Skewness and kurtosis values for the NEET and InfMor indicators reveal asymmetry and the presence of extreme values. This indicates that while most regions have lower values, some regions show significantly higher values, suggesting notable regional differences, likely due to socioeconomic disparities. In addition, histograms were generated to visualize the data distribution, which showed slight skewness; however, the values remain within an acceptable realistic range.

Figure 1 Examples of data imputation for selected regions and indicators

Country NUTS 1 NUTS 2



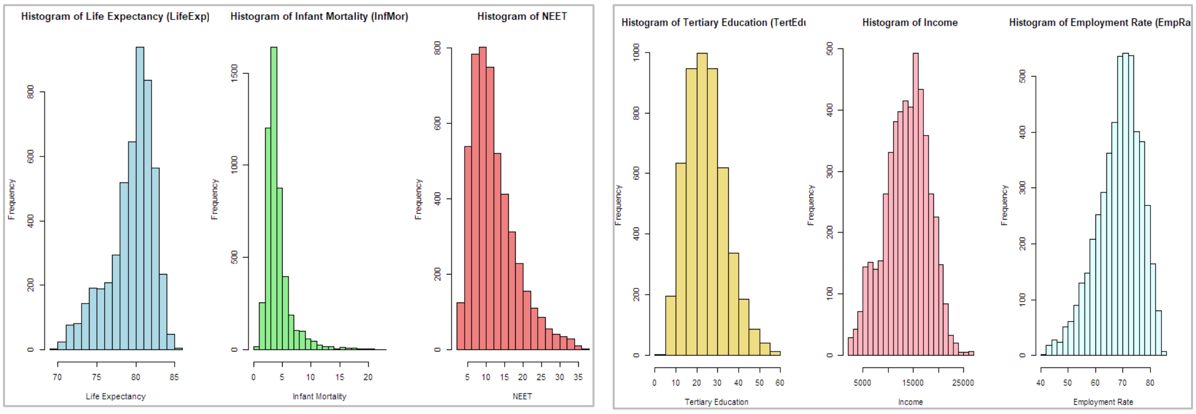
Source: own processing based on EUROSTAT data.

Table 3 Summary statistics for key variables for NUTS2 region after imputation

Variable	Mean	Std_Dev	Min	Max	Skewness	Kurtosis
LifeExp	79,5014	2,893741	69,4	85,8	-0.9266112	3.390583
InfMor	4,2278	2,374283	0,0	23,0	2.8737808	14.797107
NEET	12,0035	6,049099	2,0	36,6	1.1524171	4.320024
TertEdu	24,4049	9,437044	3,6	59,7	0.5450497	3.101606
Income	13 694	4 338	2 100	26 700	-0.236727	2.662249
EmpRat	68,5898	7,9871	40,3	85,6	-0.6169516	3.108149

Source: own processing based on EUROSTAT data.

Figure 2 Histograms of variables for RHDI calculation for NUTS2 regions



Source: own processing based on EUROSTAT data.

4.4 Calculation of the RHDI index

To calculate the RHDI, minimum and maximum values were determined from the imputed datasets, supplemented with data projected five years forward using linear extrapolation. Statistics before and after linearization, together with the final min and max values, are presented in Table 4.

Table 4 Summary statistics for key variables for the NUTS2 region, linearized

Variable	Original dataset		Dataset after linearization		Final min and max values	
	Min	Max	Min	Max	Min	Max
LifeExp	69,39	85,8	69,39	86,84	69,39	86,84
InfMor	0,0	23,00	-1,154	23,00	0	23,00
NEET	2,0	36,64	1,147	36,65	2,0	36,65
TertEdu	3,6	59,70	3,6	67,90	3,6	67,9
Income	2 100	26 700	2 100	28 597	2 100	28 597
EmpRat	40,30	85,60	40,30	90,48	40,30	90,48

Source: own processing based on EUROSTAT data.

The determined min and max values were used for normalization to a range of 0 to 1 using formulas (1) and (2). Subsequently, calculations were performed for the dimensions of health, education, and income. RHDI values were derived from the geometric mean of the dimension values. As an example, Table 5 presents the indicator values, dimension values, and the RHDI index for the Czech region CZ03 Jihozápad.

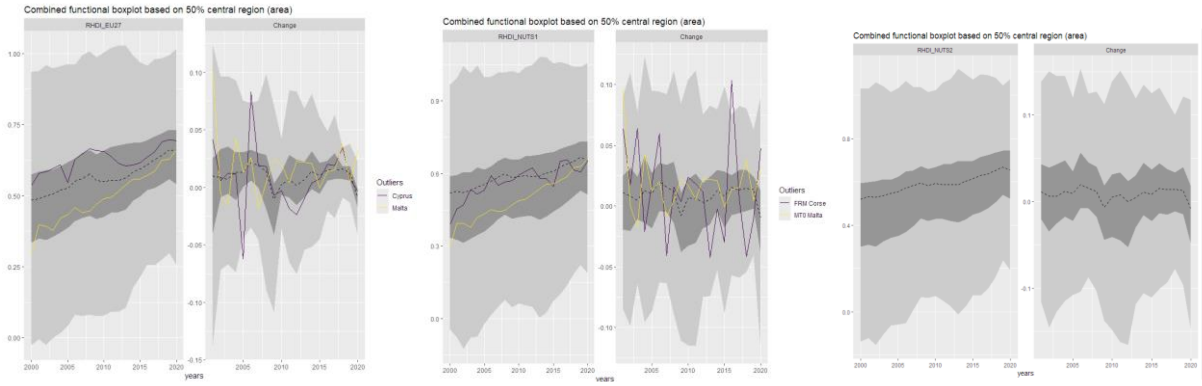
To illustrate visualization options, functional boxplots were generated using the envelope method. The resulting graphs are presented in Figure 3. In Figure 3, it can be observed that there is some convergence in RHDI values at the country level; however, this trend stopped toward the end of the observed period, probably due to the COVID-19 pandemic. Cyprus and Malta deviate slightly from the general trend, although these deviations are not significant, and similar patterns are noted for the NUTS1 and NUTS2 regions.

Table 5 Normalized indicators, dimensions, and RHDI index values for the Czech region CZ03 Jihozápad

Year	LifeExp	InfMor	NEET	TertEdu	Income	EmpRat	D_{health}	$D_{education}$	D_{income}	RHDI
2014	0,5536	0,9087	0,8423	0,2177	0,3472	0,6873	0,7312	0,5300	0,5172	0,5852
2015	0,5190	0,8826	0,8676	0,2348	0,3623	0,7211	0,7008	0,5512	0,5417	0,5937
2016	0,5652	0,8739	0,8338	0,2411	0,3623	0,7430	0,7195	0,5374	0,5527	0,5979
2017	0,5594	0,8739	0,8845	0,2551	0,3887	0,7769	0,7167	0,5698	0,5828	0,6197
2018	0,5594	0,8957	0,9070	0,2566	0,4038	0,7968	0,7275	0,5818	0,6003	0,6334
2019	0,5767	0,8957	0,8901	0,2597	0,4265	0,8028	0,7362	0,5749	0,6146	0,6384
2020	0,5190	0,9000	0,8507	0,2519	0,4189	0,7908	0,7095	0,5513	0,6049	0,6185

Source: own processing based on EUROSTAT data.

Figure 3 Functional boxplot of RHDI for EU27 countries, NUTS1 and NUTS2 regions



Source: own processing based on EUROSTAT data.

5. Conclusion

The data required to calculate the Regional Human Development Index (RHDI) was verified to be available to a sufficient extent; however, due to various factors, primarily frequent changes in the structure of NUTS regions, the data are not always complete. For NUTS 0, nearly 1% of the data are missing, for NUTS 1 nearly 3%, and for NUTS 2 almost 6% of the dataset. Two regions were excluded due to missing data for the entire period, data for one region were supplemented from alternative sources, and for some Croatian, Irish and Slovenian regions, data from the original regions were used due to geographic changes. This consideration should be taken into account when interpreting the results.

Under these circumstances, data imputation methods become relevant. The Amelia package for multiple imputation, recommended by the authors of the RHDI methodology, can be considered an effective solution. The results indicate that despite partial data gaps, reliable estimates of missing values can be obtained, providing a foundation for calculating the key health, education, and income indicators needed for RHDI.

Other steps in calculating the RHDI were also verified, such as determining the min and max values based on the linearized trend. While this is a suitable method, due to temporal changes, it may need recalculating after some time, unlike the HDI according to UNDP methodology (n.d.), which uses fixed values. The subsequent steps in calculating the RHDI proceeded smoothly according to the established methodology.

Visualization through functional analysis enables the identification of regional differences and developmental trends, providing valuable input for strategic planning in regional policies focused on sustainability. Future research thus opens avenues for applying both advanced multiple imputation methods and functional time series analysis to more detailed mapping of differences between countries and regions. This could facilitate comparisons with other indicators, not only for sustainability but also for economic, social, and environmental parameters. The ability to compare changes across programming periods is also significant.

This approach contributes to a deeper understanding of regional development trends within the EU and their progress towards sustainable development by allowing more accurate tracking of regional differences, identifying key factors influencing their growth, and evaluating the effectiveness of regional policies. Through a comprehensive time series analysis, it is possible to address current challenges and adjust policies to support the long-term economic, social and environmental development of regions.

Acknowledgement:

This research was supported by GA JU project no. 129/2022/S.

References

CEIC. (n.d.). *Global Database*. Retrieved from https://www.ceicdata.com/en

European Commission. (2001). *A sustainable Europe for a better world: A European Union strategy for sustainable development.* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52001DC0264

European Commission. (2019). *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions: The European Green Deal.* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52019DC0640

Eurostat. (n.d.). *Eurostat: European Statistical Office.* Retrieved from https://ec.europa.eu/eurostat

Hardeman, S., & Dijkstra, L. (2014). The EU Regional Human Development Index. Luxembourg: Publications Office of the European Union. Retrieved from: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC90538/online%20version%20a4.pdf

Honaker, J., King, G., & Blackwell, M. (2011). *Amelia II: A program for missing data*. Journal of Statistical Software, 45(7), 1–47. doi: 10.18637/jss.v045.i07

Hořčica, A. (2023). *Functional analysis in R: A new approach to sustainability indicators in the context of regional development.* In Proceedings of the 17th International Scientific Conference Inproforum: Challenges and Opportunities in the Digital World (p. 29). Faculty of Economics, University of South Bohemia in České Budějovice. Retrieved from https://omp.ef.jcu.cz/index.php/EF/catalog/book/88.

Joint Research Centre. (n.d.). Science hub. European Commission. Retrieved from https://joint-research-centre.ec.europa.eu/index_en

Mrkvička, T. (2017). *Globální obálkové testy aneb jak otestovat vhodnost statistického modelu na základě funkcionální charakteristiky*. Pokroky matematiky, fyziky a astronomie, 62(1), 17–23.

Myllymäki, M., & Mrkvička, T. (2020). GET: Global envelopes in R (arXiv:1911.06583). arXiv. [Online]. Retrieved from: http://arxiv.org/abs/1911.06583

NUTS. Eurostat. (n.d.). *Nomenclature of Territorial Units for Statistics (NUTS).* Retrieved from https://ec.europa.eu/eurostat/web/nuts/overview

R Project. (n.d.). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org

R Project. Amelia. (n.d.). *Amelia: A program for missing data*. Retrieved from https://cran.r-project.org/web/packages/Amelia/index.html