Classification of countries based on development indices by using K-means and grey relational analysis

Sayel Basel · K. U. Gopakumar · R. Prabhakara Rao

Abstract Clustering countries based on their development profile is important, as it helps in the efficient allocation and use of resources for institutions like the World Bank, IMF and many others. However, measuring the status of development in each country is challenging, as development encompasses several facets such as economic, social, environmental and institutional aspects. These dimensions should be captured and aggregated appropriately before attempting to classify countries based on development. In this context, this paper attempts to measure various dimensions of development through four indices namely, Economic Index (EI), Social Index (SI), Sustainability Index (SUI) and Institutional Index (II) for the period between 1996 through 2015 for 102 countries. And then we categorize the countries based on these development indices using the grey relational analysis and K-means clustering method. Our study classifies countries into four clusters with twelve countries in the first cluster, fifty in second, twenty-seven and thirteen countries in third and fourth clusters respectively. Having taken each of the dimensions of development independently, our results show that no cluster has performed poorly in all four aspects.

Keywords Development · K-means clustering · Grey relational analysis · Principal component analysis

JEL Classification C38 · C49 · O10 · O57

S. Basel (✉)
Department of Economics, Christ University, Bengaluru, Karnataka 560029, India
e-mail: sayel.basel@christuniversity.in

K. U. Gopakumar · R. P. Rao
Department of Economics, Sri Sathya Sai Institute of Higher Learning, Prasanthi Nilayam, Andhra Pradesh 515134, India
e-mail: gopakumarku@sssihl.edu.in

R. P. Rao
e-mail: rprabhakararao@sssihl.edu.in

Introduction

Categorization of the countries and their ranking as per the development status has been a major focus for the institutions like the World Bank, IMF, and UNDP as it could facilitate better policymaking and efficient use of resources. For example, the release of funds for development activities (Mylevaganam, 2017). However, this classification of countries according to the level of development has been challenging, as the process of development is multidimensional. According to Nielsen (2011), "when it comes to classifying countries according to their levels of development, there is no criterion (either grounded in theory or based on an objective benchmark) that is generally accepted". The major difficulties are on two grounds. The first is to arrive at a unique and widely

acceptable measure of development that could capture the overall development process. And the second relates to its methodological difficulties in choosing an appropriate technique to classify the countries based on the chosen measure of development.

The measurement issue could be tackled by choosing an existing measure of development and using it as the criteria for classification. For example, some of the popular indicators of development could be, GDP measuring level of income, *Level of Living Index* (Drewnowski & Scott, 1966), *Socio-Economic Development Index* (UNRISD, 1970), *Physical Quality of Life Index* (Morris, 1979), *Human Development Index* (UNDP, 1990), *Happy Planet Index* (NEF, 2006), *Multidimensional Poverty Index* (Alkire & Santos, 2010) and *Social Progress Index* (Social Progress Imperative,[1] 2010). Examples for classification based on the existing measures could be the ones by international organizations such as the World Bank, the OECD, the UNDP, and UNCTAD. Since 1978, the World Bank has been classifying the countries based on per capita Gross National Income (GNI) as it considers per capita income as the simple and appropriate indicator of economic progress and human well-being (World Bank, 2012). Accordingly, the World Development Reports by the World Bank classifies countries into four income groups. These are: (1) *Low-Income Economies* (LIE) with GNI of $1025 or less. (2) *Lower-Middle Income Economies* (LMIE) with GNI between $1026 and $3995. (3) *Upper-Middle Income Economies* (UMIE) with GNI between $3996 and $12,375. (4) *High-Income Economies* (HIE) with GNI of $12,376 or more.

The classification by OECD's Development Assistance Committee (DAC), uses the World Bank's income classification and categorizes countries into two groups namely, "developed" and "developing" where all the countries belonging to LIE, LMIE, and UMIE are classified as developing countries and HIE countries to developed countries. Though these classifications are widely accepted, it suffers from the drawbacks of using GDP as the underlying measure of development and its classification.[2] While the OECD DAC and the World Bank use income as the criteria,

UNDP takes an alternative route and classifies countries based on the HDI where the countries are classified into four groups viz., (1) *Very high human development countries*, with HDI greater than 0.800 in 2019. (2) *High human development countries*, with HDI between 0.700 and 0.799. (3) *Medium human development countries*, with HDI between 0.550 and 0.699. (4) *Low human development countries*, with HDI less than 0.550. Even though the use of HDI is a definite improvement over the earlier classifications based on income,[3] the former is criticized by many as development in recent times implies far more than just literacy, good health and quality of living could capture (Decancq & Lugo, 2009; Desai, 1991; Lind, 2004; Mylevaganam, 2017; Noorbaksh, 1998; Ravallion, 1997; Santos & Santos, 2014; Srinivasan, 1994; Streeten, 1995).

In short, a major drawback in choosing GDP or HDI as an indicator of development is that they only capture a few aspects of development and fail to represent the overall development that is a multi-dimensional process (Basel et al., 2020). As observed by Todaro and Smith (2011), development should encompass all the economic, social, institutional, and political mechanisms that could enrich human life. Moreover, in recent times with emphasis on the sustainability of development, we believe that an appropriate measure of development should also account for variables related to environment and energy use. This would require going beyond the already existing measures and arriving at a new, broader indicator of development. This is easier said than done, as there are a lot of difficulties involved right from selecting variables to choosing an appropriate methodology to aggregate the variables.

Once the appropriate measure is identified, the next difficulty is in selecting a methodology to classify the countries according to the state of development. The classifications by the World Bank, OECD, and UNDP mentioned above follow the ordinal criterion methodology while grouping the countries (Vázquez & Sumner, 2012). The major limitation in such

---

[1] Published by Social Progress Imperative can be retrieved from https://www.socialprogress.org/.

[2] Authors like Sen (1983), Goossens et al. (2007), Stiglitz et al. (2009) Costanza et al. (2009), Wilkinson et al. (2010),

---

Footnote 2 continued
Schepelmann et al. (2010) among many have pointed out the drawbacks of GDP while measuring the development.

[3] United Nations Development Programme (2015, 2016) Human Development Report 2016: Human Development for Everyone, New York, USA.

classification is that it fails to include an appropriate number of groups of countries and more importantly, the "development thresholds" that separate the groups are subjective (Nielsen, 2011). This gap can be bridged through cluster analysis, a statistical technique that provides a more accurate and objective classification of countries instead of mere ordering or ranking based on a given development indicator. Further, it also allows to include a large number of indicators that could reflect the multidimensionality of the development process. The existing panel clustering methods are based purely on multivariate statistical theory (Liu et al., 2021). For instance, Bayesian approach (Aßmann & Boysen-Hogrefe, 2011), artificial bee colony (Banharnsakun, 2018), data mining (Alizadeh et al., 2008), Density-Based Spatial Clustering of Applications with Noise (Azzalini & Menardi, 2016) and Gaussian Mixture Model (Malsiner-Walli et al., 2016). While these methods are a significant improvement over the univariate methods like the ordinal criterion techniques, a major limitation with the above-mentioned methods is that they mainly focus on the improvement and design of the clustering algorithm, rather than examining the multiple dynamic information and features present in the panel data (Liu et al., 2021). The dynamic relationship between the variables across the panel is important when one speaks about development, its measurement, and classification. Ignoring these dynamics may lead to underutilization of the available information that could turn significant while forming clusters thereby resulting in clustering errors (Cameron & Miller, 2015).

In this context, this paper has two broad objectives. The first is to capture the broader dimensions of development through various development indicators such as economic, social, sustainability, and institutional factors for 102 countries starting from the year 1996 through 2015. To this end, the paper attempts to construct four indices for measuring development namely, Economic Index (EI), Social Index (SI), Sustainability Index (SUI), and Institutional Index (II). The variables under each of the indices are aggregated using the weights obtained from the PCA analysis. Having constructed the indices, the second objective is to use them for classifying the countries through the clustering analysis. Using K-means clustering algorithms we try to categorize the countries as per the above-mentioned development indices. However,

before proceeding with K-means clustering, we apply grey relational analysis (GRA) as it will explain the partial information and uncertainty related to the development trend for the chosen countries. Using the grey relational analysis and then proceeding with K-means clustering will allow us to incorporate all the available information and panel dynamics. Apart from constructing four indices of development and integrating the GRA and K-means methodology, a major contribution of our study is its analytical framework. As noted, the clustering exercise is based on four indicators of development rather than taking one composite measure of development. This, according to us, will help capture the development profile of each cluster with a clear understanding of those factors that have led to its development or underdevelopment.

The sections in the paper following this discussion are organized as follows: "Literature survey: a review of the methodologies" section presents the literature review on the construction of development index, grey relational analysis, and K-means clustering; "Data and methodology" section covers the explanation of the data and outlines the methodologies; "Empirical analysis" section discusses the empirical findings; and, finally, "Summary and conclusion" section concludes the paper.

Literature survey: a review of the methodologies

In this section, we discuss various studies undertaken related to the construction of development indices, grey relational analysis, and K-means clustering.

Construction of development index[4]

In the recent past, an argument has been that GDP as a measure of development suffers from many limitations as this narrow definition of development fails to encompass the broader aspects of development which have a considerable effect on individual and social well-being (Goossens et al., 2007; Schepelmann et al., 2010; Sen, 1983; Stiglitz et al., 2009; Wilkinson et al., 2010). There exists extensive research on the limitation of and criticism for the use of income alone as a parameter of the overall development of the country.

---

[4] A comprehensive review of literature on construction of development index can be found in Basel et al. (2020).

The general opinion is that the use of GDP as an indicator of development may be misleading because it merely gives a monetary measure of the level of production (Costanza et al., 2009; McGranahan et al., 1972; Victor, 2010). This argument is valid as the growth in GDP fails to account for the satisfaction of basic needs and also the distribution pattern of income in the society (Van den Bergh & Antal, 2014). It is in this context that Sen (1983) argues, "supplementing data on GNP per capita by income distributional information is quite inadequate to meet the challenges of development analysis".

These limitations and criticisms with income as a general measure of development led to a paradigm shift in the way by which economists viewed the development process. The popular view was that economic development is a much broader concept than just the growth of income (Goossens et al., 2007; Nordhaus & Tobin, 1973). This led to redefining the concept of economic development in a way that emphasized the inclusion of economic, social, political, and institutional mechanisms that could bring about rapid and large-scale improvement in the standard of living of the people (Todaro, 1989). Thus, in the post-1970s the emphasis was on measuring development as a multidimensional concept involving various indicators of development which goes beyond just income and its related measures (Booysen, 2002; Greco et al., 2016). The initial steps towards the construction of a comprehensive measure for development were led by institutions such as UNESCO (1974, 1976), UN (1975), UNRISD (1978, 1979), OECD (1973, 1977). The pioneer in this field was the UN Research Institute for Social Development (UNRISD) which had taken several initiatives for the formulation of composite development indices. For example, 'Level of Living Index' (Drewnowski & Scott, 1966), 'Socio-Economic Development Index' (McGranahan et al., 1972) and Physical Quality of Life Index (Morris, 1979).

Later the World Bank (World Development Report, 1991) presented a more comprehensive view of development which asserted that the development should encompass better education, good health, less poverty, cleaner environment e.t.c. Around the same time, Mahbub ul Haq devised the Human Development Index (HDI) which has been recognized as the measure of economic development by the United Nations Development Programme (UNDP). However, there are several criticisms concerning the construction and components of HDI. Among others relating to construction of the index, a major conceptual criticism was that it fails to capture broader dimensions of the development (Decancq & Lugo, 2009; Desai, 1991; Lind, 2004; Noorbaksh, 1998; Ravallion, 1997; Santos & Santos, 2014; Srinivasan, 1994; Streeten, 1995). Post-2000's witnessed a surge in the construction of various measures of societal well-being comprising wider indicators of development. Some of the major indices of this period are the Happy Planet Index (2006), Multidimensional Poverty Index (2010), Social Progress Index (2010), and Gross Happiness Index (2012).

The major challenge and greater difficulty in constructing a broader measure of development are in choosing an appropriate methodology to aggregate them into a single composite index. Common practice is to aggregate the variables into a composite index by assigning equal weights. However, assigning equal weights would lead to biasness as it fails to distinguish between more important and less important variables (Greco et al., 2018). Hence, for our study, we have considered PCA to derive weights. PCA is advantageous while dealing with 'double counting' by correcting overlapping information in two or more variables (OECD, 2008). Among others, Ram (1982), Noorbakhsh (1996), Lai (2003), Guptha and Rao (2018) and Basel et al. (2020) have used PCA for construction of various developmental indices.

Grey relational analysis (GRA) and K-means clustering

Grey System Theory was developed by Deng in 1982 for understanding the process where information is partial, unclear, and uncertain (Deng, 1989). This system is mainly used for modeling uncertainty, analyze system relations, establish models and make forecasts and decisions (Tsai et al., 2005). This system has been widely used in various fields of social science and natural sciences (Huang et al., 2006; Shi et al., 2008; Wang et al., 2010). In economics, this methodology has been used in different areas such as supply chain management, decision-making process, financial performance evaluation, credit risk, energy consumption, forecasting, etc.

For instance, based on the Grey system theory, Sahu et al. (2014) have developed a multi-criteria

group decision-making method and multi-objective optimization ratio analysis, which facilitate multi-criteria group decision-making (MCGDM) problems under uncertain environments. His analysis also provides an appropriate compromise ranking order for available possible alternatives. For the evaluation of financial performance, Wu et al. (2010) have used the grey relational analysis for a performance evaluation of one of Taiwan's banks through a combined analytical hierarchy process (AHP) and GRA. Similarly, using the GRA, Jin et al. (2012) have analyzed the credit risk of a commercial bank in China and have recognized the implementation of differentiated credit policy to promote the development of the banking and financial sector. For understanding the relationship between energy consumption and economic growth, Kose et al. (2013) have analyzed the development pattern of Turkey through GRA and concluded that oil, renewable energy sources, and GDP are the major drivers of development in Turkey. Likewise, Huang and Wang (2013) constructed and evaluated an index system for the industrial economy-ecology-coordinated development in China.

The studies covering the grey forecasting models include that of Zhang and Chen (2014) where the model GM (1, 1) and the metabolic GM (1, 1) have been used for predicting the monetization ratio to determine the facts that are responsible for the financing difficulty of the small and medium-sized enterprises in China. On similar lines, Tai et al. (2011) propose both a grey decision and prediction model based on the GM (1, 1) model for easing the decision-making process and enabling an effective forecasting approach. A study by Huang (2014) used a grey forecasting method to predict the development of the logistics industry in the Henan province. It forecasts the trends and factors leading to industrial development in Henan. Further, the studies that have also used the grey systems theory in different fields of economics include maximizing bidder's profit in online auctions (Lim et al., 2012), dynamics and chaos control (Hu & Xia, 2012), analysis of consumer income elasticity and predicting its trend (Luo & Song, 2012), evaluation of innovation competency in the aviation industry (Zhu et al., 2012), ranking the advanced manufacturing systems (Goyal & Grover, 2012), examining the qualitative relationship among the innovative capability of the manufacturing

industry and local government behavior (Bi et al., 2011) and a real estate marketing investment model (Shui, 2014).

While GRA was used for analyzing a variety of economic problems, the use of K-means in economics was largely confined to clustering analysis. An example of such an exercise in context to our study could be one by Mylevaganam (2017) where he clustered countries into various groups using the K-means clustering algorithm with a modified HDI through PCA as a measure of development. Recently, Mallikarjuna and Rao (2020) have used K-means clustering to classify stock markets based on financial factors. Their study considered forty-five stock markets for classification and they clustered the financial markets by using data mining techniques, viz. K-Means, Hierarchical, and Fuzzy C-Means. Apart from economics, K-means has been used in other fields of studies too. Some examples could be, Zhao et al. (2006), where they applied a K-means Clustering algorithm using data detection and symbol-timing recovery for a burst-mode optical receiver. Barca et al. (2008), using a modified K-means algorithm, have removed noise in multicolor motion Capture Image Sequences. K-means clustering was used by Roy and Sharma (2010) to identify several gene-expression benchmarks. However, despite its use in analyzing various issues across the fields, studies using the K-means cluster algorithm have remained subdued.

Considering studies that incorporate the fusion of GRA and K-means clustering, despite its advantage over individual exercises using either K-means or GRA, we notice that studies of such nature are rather rare. Using the mix of GRA and K-means would allow us to integrate and efficiently use the available information while accounting for panel dynamics. Emami and Razi (2014) applied a hybrid grey relational analysis and K-means to cluster and measure the performance of the banking system for Semnan, Iran. They also attempts to re-examine the implementation of this method for a newly established bank in the city of Tehran, Iran. A similar study was conducted by Zarandi et al. (2014) measuring the relative efficiency of different banks using a hybrid of K-means and grey relational analysis.

With this, we conclude our literature survey on the various issues that are important while we attempt to cluster countries based on their development status. As seen in the aforesaid discussion, (1) the need for an

appropriate measure for development and (2) the methodology to cluster the countries without losing the information and dynamics involved in the panel are two important issues that need to be addressed while one attempts to group countries.

Data and methodology

In this section, at first, we present details of the data and variables that are used in the construction of the development indices that are used in our study. Then we proceed with explaining the methodologies, PCA, grey relational analysis, and K-means clustering.

Data and variables

The data has been obtained from official sources such as UNESCO and the World Bank covering 102 countries from 1996 to 2015. The variables used to construct social, economic, sustainability, and institutional development indices are given along with notation in Table 1.

The social index comprises the variables which encompass the aspect of education, health, and access to drinking water and sanitation facilities. Better education and good health contribute to better human capital and enhances productivity. These factors play a crucial role in improving the standard of living as well as the way of living and thus contribute to the overall development of the country. The variables for the construction of the sustainability index include consumption of renewable and non-renewable energy sources, forest cover, and emission of CO_2. Forest cover along with the use of energy sources are taken as a proxy for the availability of natural resources. Thus, a higher availability of these resources at present could imply sustainable development in the later phases of the development process. The variables used for the construction of the economic index show the overall macro-economic performance of the country. High levels of GDP per capita, productivity, and capital stock can help to achieve higher levels of economic growth. Whereas, lower inflation is desirable for a stable economy. Finally, the governance attributes captured by the rule of law and voice and accountability are proxies for the institutional index.

Methodology: construction and classification

The methodologies employed for the construction and classification of the development indices are discussed in the following section.

Principal component analysis

In the construction of Economic Index (EI), Social Index (SI), Sustainability Index (SUI), and Institutional Index (II), we employed the statistical technique of factor analysis to obtain the weights of these indices. We identify the factor weights by using the principal components analysis (PCA). PCA is a technique of analyzing and identifying patterns in data and expressing the data in condensed form to highlight their similarities and differences. It transforms a large number of correlated variables into a smaller number of uncorrelated variables but retains the information in a large set. These uncorrelated variables which are extracted from original set's variables using their correlation matrix are called principal components.

Let X_i (i = 1, 2, 3, …, n) be the set of n variables that are transformed into a new set of m variables.

Let P_j (j = 1, 2, 3, …, m) (m << n) be the principal component which is represented as the linear combination of the Xs i.e.,

$$PC_k = a_{k1}X_1 + a_{k2}X_2 + \ldots + a_{kn}X_n \qquad (1)$$

where PC_k is the kth principal component. $X_1 + X_2 + \ldots + X_n$ are the original set of variables. $a_{k1} + a_{k2} + \ldots + a_{kn}$ are loading or scores of respective X_i in the kth principal component. The component loadings are the weights showing the variance contribution of principal components to variables. The variance of the principal components is given by eigenvector (x) and eigenvalues (λ) of the variance–covariance matrix (A) as,

$$Ax = \lambda \qquad (2)$$

$$|A - \lambda I| = 0 \qquad (3)$$

where 'I' is an identity matrix.

Grey relational analysis

The categorization of the countries based on development is challenging as it often encounters uncertainty,

Table 1 Variables, notations, and sources

	Variable	Notation	Source
Social	Expected year of schooling (years)	EYS	UNESCO Institute of Statistics
	Mean year of schooling (years)	MYS	
	Life expectancy at birth, total (years)	LEX	World Bank database
	Infant mortality rate (per 1000, live birth)	IMR	
	Improved sanitation facilities (percentage of the population with access)	ISF	
	Improved water facilities (percentage of the population with access)	IWF	
Sustainability	Non-renewable energy consumption (kg of oil equivalent per capita)	NEG	
	Renewable energy consumption (percentage of total final energy consumption)	REG	
	Forest cover (sq. km)	FRT	
	CO2 emissions (Kt)	CO2	
Economic	y-o-y GDP per capita growth (in percentages)	GDP	
	y-o-y Inflation, consumer prices (in percentages)	INF	
	Gross capital formation (percentage of GDP)	GCF	
Institution	Rule of law (in percentiles)	ROL	
	Voice and accountability (in percentiles)	VAA	

lack of information, and complexity of data. To overcome this, we choose to employ grey relational analysis. The steps involved in this method are as follows:

1. The formation of Decision Matrix: The first step is to transform the panel data into the decision matrix or behavior matrix. For this, we first remove the dimension effect of different indices by averaging the original dataset. The decision matrix is given as,

$$X_i = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(n) \\ x_2(1) & x_2(2) & \ldots & x_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(1) & x_n(2) & \cdots & x_n(n) \end{bmatrix} \quad (4)$$

where $i = 1, 2, \ldots, n$

2. The formation of reference series: Let X_0 be the reference series which is taken as the maximum values within the dataset.

$$X_0 = \{x_0(1), x_0(2), \ldots, x_0(n)\} \quad (5)$$

3. The formation of Absolute value: The absolute value is obtained as,

$$\Delta X_{0i}(j) = |x_i(j) - x_0(j)|$$
$$= \begin{bmatrix} \Delta x_{01}(1) & \Delta x_{01}(2) & \cdots & \Delta x_{01}(n) \\ \Delta x_{02}(1) & \Delta x_{02}(2) & \ldots & \Delta x_{02}(n) \\ \vdots & \vdots & \ddots & \vdots \\ \Delta x_{0n}(1) & \Delta x_{0n}(2) & \cdots & \Delta x_{0n}(n) \end{bmatrix}$$
$$(6)$$

4. The formation of grey relational coefficient matrix: The grey relational coefficient is computed as follows,

$$\gamma_{0i}(j) = \frac{\Delta min + \xi \Delta max}{\Delta X_{0i}(j) + \xi \Delta max} \quad (7)$$

where $\Delta min = min_i min_j \Delta X_{0i}(j)$ and $\Delta max = max_i max_j \Delta X_{0i}(j)$.

5. The calculation of grey relational degree: The grey relational degree is calculated through the formula as,

$$\theta_{0i} = \frac{1}{n} \sum_{j=1}^{n} \gamma_{0i}(j) \quad (8)$$

If the weights (W_i) of the data are to be taken into consideration then we have,

$$\theta_{0i} = \sum_{j=1}^{n} [W_i(j)\gamma_{0i}(j)]. \tag{9}$$

K-means clustering

K means clustering, developed by MacQueen (1967) is the most popular and effective clustering technique which is widely used for dividing data sets into clusters (Aggarwal, 2003; Han et al., 2011; Jain & Dubes, 1988; Jain & Flynn, 1996). K means clustering algorithms aim at minimizing the variation in the inter-clustering data point while maximizing the variation between the clusters. This is done by minimizing the objective function J, which is defined as:

$$J = \sum_{i=1}^{m} \sum_{k=1}^{n} \gamma_{ik} \|x_i - \mu_k\|^2 \tag{10}$$

where $\gamma_{ik} = 1$ for the data point x_i if it belongs to cluster k; otherwise $\gamma_{ik} = 0$ and μ_k is the centroid of x_i cluster.

Now, the objective function J is minimized into two ways, first w.r.t γ_{ik} treating μ_k as constant and second w.r.t μ_k treating γ_{ik} as constant. Thus, we have,

$$\frac{\partial J}{\partial \gamma_{ik}} = \sum_{i=1}^{m} \sum_{k=1}^{n} \|x_i - \mu_k\|^2 \tag{11}$$

$$\Rightarrow \gamma_{ik} = \begin{cases} 1, & if \ k = argmin_j \|x_i - \mu_k\|^2 \\ 0, & otherwise \end{cases} \tag{12}$$

And,

$$\frac{\partial J}{\partial \mu_k} = 2 \sum_{i=1}^{m} \gamma_{ik}(x_i - \mu_k) = 0 \tag{13}$$

$$\Rightarrow \mu_k = \frac{\sum_{i=1}^{m} \gamma_{ik}\mu_k}{\sum_{i=1}^{m} \gamma_{ik}} \tag{14}$$

The steps involved in the k-means algorithm steps are:

1. Select the number of clusters (k).
2. Select the cluster centers, μ_k.
3. Assign the data points x_i to the nearest cluster.
4. Calculate the centroid or mean of all objects in each cluster.
5. Repeat steps 2, 3, and 4 until J is invariant (variance $< \varepsilon$).

Empirical analysis

Having explained the data and methodology, in this section we turn into empirical analysis.

Principal component analysis

In the construction of the indices, we employed the statistical technique of factor analysis to obtain the weights for each of the variables that are mentioned in Table 1 earlier. We identify the factor weights by using the principal components analysis (PCA). To ensure the suitability of PCA, the Kaiser–Meyer–Olkin (KMO) test is conducted to measure the sampling adequacy. If the K–M–O measure is between 0.5 and 1.0, then the factorability is assumed. The result of the adequacy test is given in Table 2.

From the above table, we can see that all the proposed indices satisfy the condition for KMO. In Table 3, we present factor loading or scores for social, sustainablility, economic, and institution indices.

The selection of principal components is based on Eigenvalues, where the minimum value is expected to be more than one. In our case, principal component one satisfies this criterion for all the indices and thus we have used only the first components for calculating the weights.

The weights are obtained by squaring each of the factor loadings representing the proportion of the total unit variance of the indicator explained by the factors.[5] For instance, the weight of EYS is 0.1669 which is calculated as (0.9964/(0.9964 + 0.9988 + 0.9919 + 0.9963 + 0.9974 + 0.9900)). Similarly, the weights of the other variables are calculated using explained variation and are presented in Table 4. These weights are then aggregated to determine the various indices for each of the countries using the following formula:

$$SI_t^{(j)} = \frac{\sum W_i z_{ti}^{(j)}}{\sum W_i} \tag{15}$$

$$SUI_t^{(j)} = \frac{\sum W_i z_{ti}^{(j)}}{\sum W_i} \tag{16}$$

---

[5] Handbook on Constructing Composite Indicator—methodology and user guide, OECD (2008).

Table 2 Results of sample adequacy test

	Social Index	Sustainablilty Index	Economic Index	Institution Index
KMO value	0.859	0.661	0.613	0.500

Table 3 Results of rotated factor loadings

Index	Variables	Factor 1	Eigenvalue	Cumulative percent of variance
Social	EYS	0.998	5.971	99.515
	MYS	0.999		
	LEX	0.996		
	IMR	− 0.998		
	ISF	0.999		
	IWF	0.995		
Sustainablility	NEG	− 0.991	3.024	75.605
	REG	− 0.345		
	FRT	0.973		
	CO2	0.988		
Economic	GDP	0.708	1.721	57.378
	INF	0.735		
	GCF	0.824		
Institution	ROL	0.726	1.053	52.636
	VAA	0.726		

Table 4 Squared factor loadings and weights of the variables (scaled to unity sum)

Index	Variables	Squared factor	Weight (W_i)
Social	EYS	0.9964	0.1669
	MYS	0.9988	0.1673
	LEX	0.9919	0.1661
	IMR	0.9963	0.1669
	ISF	0.9974	0.1670
	IWF	0.9900	0.1658
Sustainability	NEG	0.1054	0.0348
	REG	0.0016	0.0005
	FRT	0.0980	0.0324
	CO2	0.1043	0.0345
Economic	GDP	0.5017	0.2914
	INF	0.5404	0.3139
	GCF	0.6793	0.3946
Institution	VAA	0.5264	0.5000
	ROL	0.5264	0.5000

$$EI_t^{(j)} = \frac{\sum W_i z_{ti}^{(j)}}{\sum W_i} \tag{17}$$

$$II_t^{(j)} = \frac{\sum W_i z_{ti}^{(j)}}{\sum W_i} \tag{18}$$

where $SI_t^{(j)}, SUI_t^{(j)}, EI_t^{(j)}$ and $II_t^{(j)}$ are the social, sustainability, economic, and institution indices for j^{th} country at period t. $z_{ti}^{(j)}$ are the normalized variables taken for the study.

Grey relational analysis

Considering the spatio-temporal feature of panel data, and the uncertainty and variation of the development process across the countries during the period of study we employ grey relational analysis before proceeding further for cluster analysis. At first we average the panel data indices over the period and then form a decision matrix. We then select the reference series as the maximum value of each index across the 102 countries as

$$X_0 = \{0.618, 0.622, 0.649, 0.668\} \tag{19}$$

After selecting the reference series, we calculated the absolute value of each index as,

$$\Delta X_{0i}(SI) = |x_i(SI) - 0.618| \tag{20}$$

$$\Delta X_{0i}(SUI) = |x_i(SI) - 0.622| \tag{21}$$

$$\Delta X_{0i}(EI) = |x_i(SI) - 0.649| \tag{22}$$

$$\Delta X_{0i}(II) = |x_i(SI) - 0.668| \tag{23}$$

Now, for the calculation of the grey relational coefficient we have,

$$\Delta min = max_i max_j \Delta X_{0i}(j) = 0.405 \tag{24}$$

$$\Delta max = min_i min_j \Delta X_{0i}(j) = 0 \tag{25}$$

Thus,

$$\gamma_{0i}(j) = \frac{\Delta min + \xi \Delta max}{\Delta X_{0i}(j) + \xi \Delta max} = \frac{0 + 0.5 * 0.405}{\Delta X_{0i}(j) + 0.5 * 0.405} \tag{26}$$

The calculation result of the grey relational coefficient for all the 102 countries is presented in "Appendix 1".

K-means clustering

Based on the grey relational coefficient obtained through Eq. (26) we now proceed to cluster the countries on the basis of various development indices. For this purpose, we use K-means clustering and the cluster visualization plot that is presented in Fig. 1. For clear understanding, the countries are presented in Table 5 and also in pictorial form in Fig. 2.

Before analyzing the characteristics of the four clusters obtained in the analysis, at first, we should examine if the four indices of development are influential in discriminating these four groups of countries. This is important as it allows us to recognize whether the groups of developing countries are statistically distinguishable. In other words, it explains whether the cluster exhibits significantly different means in the development indices. For this, we perform a one-way ANOVA analysis to calculate the cluster centroids and see if there exists any significant difference between the clusters. The result of the ANOVA analysis is given in Table 6. The higher and significant value of the F statistic indicates that the relationship between the overall between-cluster variation and the overall within-cluster variation is good. It reflects that the designed indices for identifying groups of countries are significant.

With this result, we now proceed to discuss the characteristics of the four clusters obtained in the study. We compare the classification of the countries made by the World Bank in terms of per-capita GNI, classification of UNDP based on HDI, and the clusters that we obtained from the four indices of development. Table 7 provides an overview of the clusters along with the development taxonomies used by the World Bank and UNDP.

We observe that the first cluster comprises 12 countries, the second cluster consists of 50 countries, while there are 27 and 13 countries each in the third and fourth clusters respectively. Out of 12 countries in the first cluster (C1), 11 countries belong to high-income economies (HIE) and a very high HDI group. In this cluster Thailand belongs to upper-middle-income economies with high HDI. The second cluster (C2) consists of a diverse group of countries where 6 countries belong to low-income economies (LIE), 16 to lower-middle-income economies (LMIE), 21 to upper-middle-income economies (UMIE) and 7 countries are from high-income economies. Comparing with the UNDP classification of countries, the second cluster has 4 countries with low HDI, 14 with middle HDI, 18 with high HDI, and 14 having very high HDI. The third cluster (C3) comprises 2 LIE, 6 LMIE, 9 UMIE, and 10 HIE. The HDI classification for this cluster turns out to be 3 with low HDI, 2 with medium HDI, 11 having high HDI, and 11 with very high HDI. Finally, the fourth cluster (C4) consists of 10 HIE, 2 LMIE, and 1 LIE. However, it has 10 very high HDI countries, 1 high, 1 medium, and 1 low HDI country.

From our clusters, C1 and C4 include a relatively major share of "developed" countries as per the World Economic Situation and Progress (WESP, 2019), except four countries namely, El Salvador, Thailand, Tunisia, and Yemen. However, Thailand and Tunisia are considered as high HDI countries by UNDP. Moreover, all the G7 countries (except Germany) are found in C1 and C4.

Cluster 2 contains almost all the "developing" economies of the world as per the classification made
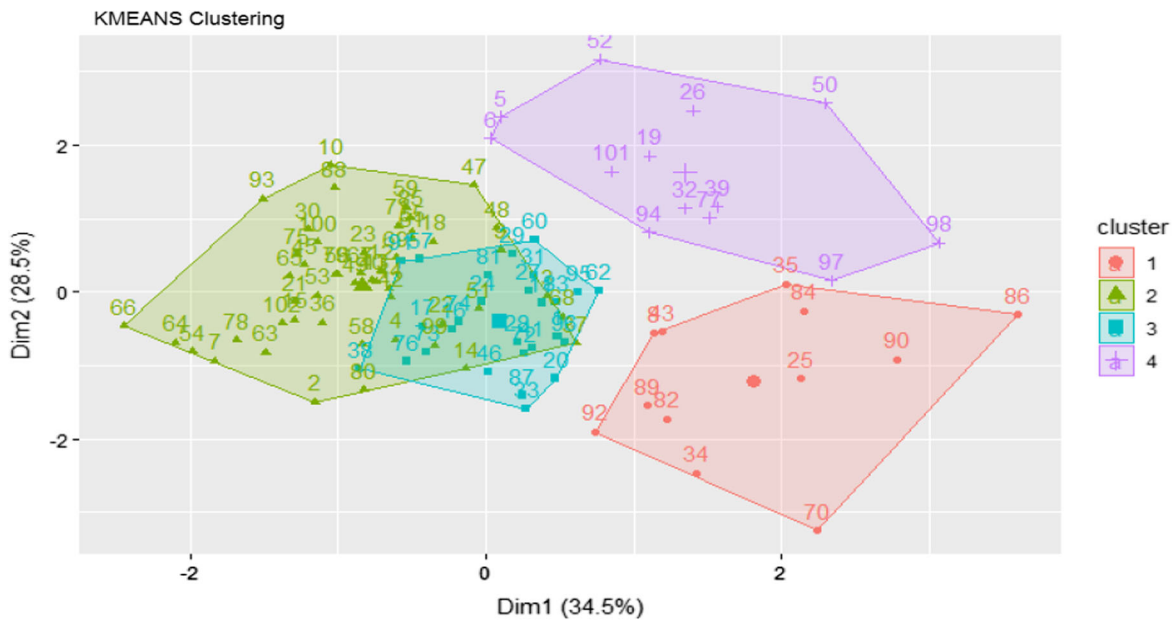
Fig. 1 Cluster plot with K-means method

Table 5 Countries with respective clusters

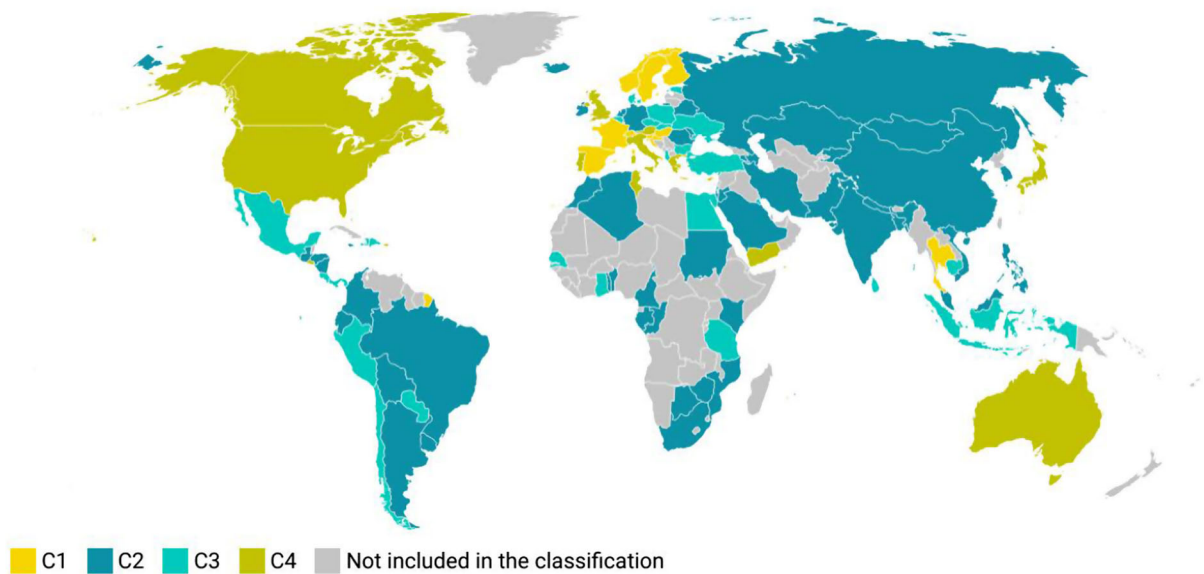| Cluster	Countries
Cluster 1 (C1)	Bahrain Croatia Finland France Hungary Norway Singapore Slovenia Spain Sweden Switzerland Thailand
Cluster 2 (C2)	Algeria Argentina Armenia Azerbaijan Bangladesh Belarus Benin Bolivia Botswana Brazil Cameroon China Colombia Congo, Dem. Rep. Ecuador Gabon Germany Guatemala Honduras Iceland India Iran, Islamic Rep. Ireland Israel Jamaica Jordan Kazakhstan Kenya Korea, Rep. Kyrgyz Republic Malaysia Mauritius Moldova Mongolia Morocco Mozambique Nepal Netherlands Nicaragua Pakistan Philippines Romania Russian Federation Saudi Arabia South Africa Sudan Togo Uruguay Vietnam Zimbabwe
Cluster 3 (C3)	Albania Belgium Bulgaria Cambodia Chile Costa Rica Czech Republic Denmark Dominican Republic Egypt, Arab Rep. Estonia Ghana Haiti Indonesia Kuwait Malta Mexico Panama Paraguay Peru Poland Senegal Sri Lanka Tanzania Turkey Ukraine Slovak Republic
Cluster 4 (C4)	Australia Austria Canada Cyprus El Salvador Greece Italy Japan Portugal Tunisia United Kingdom United States Yemen, Rep.

by WESP.[6] This cluster also includes nine emerging market economies.[7] Most importantly, the BRICS countries are present in this cluster. Cluster 3 consists

of both developed and developing economies. Even though C3 consists of developing countries in terms of income, the developing countries in this cluster are characterized by high and very high HDI. As explained above, 22 out of 27 belong to the very high and high HDI group. This shows that the countries belonging to cluster 3 are more developed in terms of HDI in comparison to the developing countries present in cluster 2.

However, we observe that to some extent our development taxonomy differs from the usual income

---

[6] As per the UN reports on World Economic Situation and Prospects (WESP), 2019. Can be accessed from https://unctad.org/en/pages/publications/World-Economic-Situation-and-Prospects-(Series).aspx.

[7] As per the Morgan Stanley Capital International Emerging Market Index 24 developing economies qualify as emerging markets, out of these 9 are present in C2 and remaining 9 are in C3, 6 are not included in this study.

C1 C2 C3 C4 Not included in the classification

Created with Datawrapper

Fig. 2 Cluster plot of the countries

Table 6 ANOVA output

Index	Variation	Sum of squares	Degree of freedom	Mean square	F-Statistic	p value
SI	Between	0.033	3	0.011	4.896	0.003
	Within	0.220	98	0.002		
	Total	0.253	102	0.013		
SUI	Between	0.104	3	0.035	16.401	0.000
	Within	0.207	98	0.002		
	Total	0.311	102	0.037		
EI	Between	0.186	3	0.062	19.066	0.000
	Within	0.319	98	0.003		
	Total	0.506	102	0.065		
II	Between	0.508	3	0.169	103.464	0.000
	Within	0.160	98	0.002		
	Total	0.668	102	0.171		

and HDI classifications as some of the countries labeled as "developed" are not in the developed economies cluster while some developing countries are found to be clustered with developed economies. The basic rationale behind such diverse result is that the clusters in this study are based not only on one dimension of development but includes various facets of well-being that leads to the overall development of the country. We also note that each cluster in the study has its own and characteristic development issues. There is no group of countries with the best (or worst)

indicators in all development dimensions. Therefore, it would not be appropriate to classify one cluster as the developed countries cluster and the other as the least developed economies cluster. Needless to say, it is the framework of our analysis considering each of the dimensions of development independently that allowed us to make such conclusions.

Further, assessing the cluster centroids i.e. the average value of each index of all countries in a certain cluster, we outline a more precise interpretation of the four clusters obtained in the analysis. This

Table 7 Classification of countries

Countries	Income group	HDI group	Cluster	Countries	Income group	HDI group	Cluster
Bahrain	HIE	Very high	C1	Singapore	HIE	Very high	C1
Croatia	HIE	Very high	C1	Slovenia	HIE	Very high	C1
Finland	HIE	Very high	C1	Spain	HIE	Very high	C1
France	HIE	Very high	C1	Sweden	HIE	Very high	C1
Hungary	HIE	Very high	C1	Switzerland	HIE	Very high	C1
Norway	HIE	Very high	C1	Thailand	UMIE	High	C1
Algeria	UMIE	High	C2	Jordan	UMIE	High	C2
Argentina	UMIE	Very high	C2	Kazakhstan	UMIE	Very high	C2
Armenia	UMIE	High	C2	Kenya	LMIE	Medium	C2
Azerbaijan	UMIE	High	C2	Korea, Rep.	LIE	Very high	C2
Bangladesh	LMIE	Medium	C2	Kyrgyz Republic	LMIE	Medium	C2
Belarus	UMIE	Very high	C2	Malaysia	UMIE	Very high	C2
Benin	LIE	low	C2	Mauritius	UMIE	High	C2
Bolivia	LMIE	High	C2	Moldova	LMIE	High	C2
Botswana	UMIE	High	C2	Mongolia	LMIE	High	C2
Brazil	UMIE	High	C2	Morocco	LMIE	Medium	C2
Cameroon	LMIE	Medium	C2	Mozambique	LIE	low	C2
China	UMIE	High	C2	Nepal	LIE	Medium	C2
Colombia	UMIE	High	C2	Netherlands	HIE	Very high	C2
Congo, Dem. Rep.	LMIE	Medium	C2	Nicaragua	LMIE	Medium	C2
Ecuador	UMIE	High	C2	Pakistan	LMIE	Medium	C2
Gabon	UMIE	High	C2	Philippines	LMIE	High	C2
Germany	HIE	Very high	C2	Romania	UMIE	Very high	C2
Guatemala	UMIE	Medium	C2	Russian Federation	UMIE	Very high	C2
Honduras	LMIE	Medium	C2	Saudi Arabia	HIE	Very high	C2
Iceland	HIE	Very high	C2	South Africa	UMIE	High	C2
India	LMIE	Medium	C2	Sudan	LIE	Low	C2
Iran, Islamic Rep.	UMIE	High	C2	Togo	LIE	Low	C2
Ireland	HIE	Very high	C2	Uruguay	HIE	Very high	C2
Israel	HIE	Very high	C2	Vietnam	LMIE	Medium	C2
Jamaica	UMIE	High	C2	Zimbabwe	LMIE	Medium	C2
Albania	UMIE	High	C3	Indonesia	LMIE	High	C3
Belgium	HIE	Very high	C3	Kuwait	HIE	Very high	C3
Bulgaria	UMIE	Very high	C3	Malta	HIE	Very high	C3
Cambodia	LMIE	Medium	C3	Mexico	UMIE	High	C3
Chile	HIE	Very high	C3	Panama	HIE	High	C3
Costa Rica	UMIE	High	C3	Paraguay	UMIE	High	C3
Czech Republic	HIE	Very high	C3	Peru	UMIE	High	C3
Denmark	HIE	Very high	C3	Poland	HIE	Very high	C3
Dominican Republic	UMIE	High	C3	Senegal	LMIE	low	C3
Egypt, Arab Rep.	LMIE	High	C3	Sri Lanka	UMIE	High	C3
Estonia	HIE	Very high	C3	Tanzania	LIE	Low	C3
Ghana	LMIE	Medium	C3	Turkey	UMIE	Very high	C3
Haiti	LIE	low	C3	Ukraine	LMIE	High	C3
Australia	HIE	Very high	C4	Slovak Republic	HIE	Very high	C3

Table 7 continued

Countries	Income group	HDI group	Cluster	Countries	Income group	HDI group	Cluster
Austria	HIE	Very high	C4	Japan	HIE	Very high	C4
Canada	HIE	Very high	C4	Portugal	HIE	Very high	C4
Cyprus	HIE	Very high	C4	Tunisia	LMIE	High	C4
El Salvador	LMIE	Medium	C4	United Kingdom	HIE	Very high	C4
Greece	HIE	Very high	C4	United States	HIE	Very high	C4
Italy	HIE	Very high	C4	Yemen, Rep.	LIE	Low	C4

comparative procedure enables us to have a clear understanding of each cluster as per the development indices. The outcome is presented in Table 8.

From Table 8, the four development clusters can be described as follows:

Cluster 1 (C1): *Developed with High Social and Institutional development.* Based on the average values of the indices, C1 stands first among the four clusters. Besides this, the average value of the social index and institution index for the members in C1 is also higher than the other clusters. This reflects that the countries belonging to cluster 1 have excelled in their social and institution dimension of development when compared to C2, C3, and C4. Moreover, the other dimensions of development namely sustainablility and economic in this cluster are also better than C2 and C3.

Cluster 2 (C2): *Moderate development:* The average value of the indices in this cluster is lower than all the other clusters and thus could be taken as moderately developed when compared to other clusters. However, as mentioned earlier, a lesser overall average value cannot tag them as underdeveloped. The social development in this cluster is found to be better compared to C4 and on par with C3.

Cluster 3 (C3): *Moderately high development:* C3 appears to be more developed than C2 as all the indices values of C3 are higher than that of C2, yet lower than C1. This cluster also excels in social and institutional development when compared to C2 and C4. Moreover, C3 also stands second in the average institution index value, next to cluster 1. Finally, based on the average of indices, C3 stands third, marginally lesser than C4.

Cluster 4 (C4): *Developed with High Economic and Sustainable development but low Social development.* Cluster 4 exhibits a higher value of economic and sustainable development but shows a lower average value in social development in comparison with the other three clusters. Besides this, the institutional development in this cluster is also found lacking when compared to C3. However, based on the average of indices, C4 stands second. As one could see, C4 is driven by excellent economic performance backed with the sustainability of development.

Summary and conclusion

The concept of development has grown multi-dimensional which includes several facets such as economic, social, sustainability, and institutional aspects. However, the traditional way of looking at economic development stressed only on the growth of national income as its indicator is inadequate. Therefore, have we attempted to construct four indices of development namely social, sustainability, economic, and institution index which could capture a wider aspect of development covering 102 countries during the period 1996 to 2015. Further, using these indices the study

Table 8 Cluster centroid

Index	Cluster 1	Cluster 2	Cluster 3	Cluster 4
SI	0.566	0.502	0.498	0.463
SUI	0.507	0.465	0.478	0.544
EI	0.522	0.442	0.453	0.569
II	0.604	0.446	0.577	0.497

categorises countries using grey relational analysis and K-means clustering.

The results show that the countries taken for the study can be classified into four clusters with twelve countries in the first cluster, fifty in the second, twenty-seven and thirteen countries in the third and fourth clusters respectively. Further, based on the average value of the four indices, we have also labeled the four clusters (C1, C2, C3, and C4) as developed with high social and institutional development, moderately developed, moderately high developed, and developed with high economic and sustainable development respectively. We observe that the average value of the social and institution index is high for the countries belonging to the first cluster and the average value of the economic and sustainability index is more for the countries in cluster four. While cluster 3 flairs fairly well in social and institutional development, cluster 2 is marginally better in social development.

Declaration

Conflict of interest The authors declares that they have no conflict of interests.

Appendix 1: Grey relational coefficient

Country	Social	Sustainability	Economic	Institution
Albania	0.649	0.625	0.496	0.778
Algeria	0.728	0.484	0.406	0.590
Argentina	0.733	0.574	0.688	0.472
Armenia	0.739	0.566	0.495	0.498

Country	Social	Sustainability	Economic	Institution
Australia	0.424	0.691	0.668	0.515
Austria	0.461	0.659	0.682	0.485
Azerbaijan	0.646	0.454	0.397	0.520
Bahrain	0.734	0.610	0.653	0.670
Bangladesh	0.639	0.610	0.641	0.490
Belarus	0.531	0.727	0.456	0.412
Belgium	0.652	0.609	0.574	0.650
Benin	0.590	0.573	0.521	0.536
Bolivia	0.635	0.565	0.535	0.483
Botswana	0.793	0.500	0.612	0.483
Brazil	0.622	0.499	0.470	0.521
Bulgaria	0.626	0.557	0.502	0.680
Cambodia	0.639	0.518	0.543	0.602
Cameroon	0.625	0.611	0.593	0.451
Canada	0.545	0.662	0.821	0.495
Chile	0.676	0.505	0.584	0.761
China	0.664	0.544	0.475	0.432
Colombia	0.661	0.482	0.625	0.539
Congo, Dem. Rep.	0.598	0.596	0.519	0.477
Costa Rica	0.598	0.605	0.502	0.698
Croatia	0.733	0.711	0.542	0.974
Cyprus	0.553	0.835	0.726	0.508
Czech Republic	0.592	0.642	0.509	0.732
Denmark	0.686	0.487	0.655	0.602
Dominican Republic	0.629	0.725	0.483	0.620
Ecuador	0.564	0.599	0.497	0.433
Egypt, Arab Rep.	0.660	0.704	0.503	0.628
El Salvador	0.684	0.862	0.567	0.600
Estonia	0.729	0.496	0.544	0.740
Finland	0.807	0.481	0.615	0.889
France	0.686	0.687	0.724	0.743
Gabon	0.642	0.462	0.548	0.481
Germany	0.787	0.614	0.598	0.572
Ghana	0.622	0.533	0.379	0.730
Greece	0.668	0.708	0.787	0.517
Guatemala	0.627	0.524	0.582	0.459
Haiti	0.630	0.619	0.448	0.826
Honduras	0.627	0.553	0.524	0.544
Hungary	0.692	0.618	0.624	0.747
Iceland	0.692	0.640	0.466	0.480
India	0.618	0.570	0.494	0.425
Indonesia	0.722	0.524	0.546	0.644
	0.624	0.798	0.494	0.457

Country	Social	Sustainability	Economic	Institution
Iran, Islamic Rep.				
Ireland	0.647	0.722	0.540	0.500
Israel	0.676	0.641	0.440	0.469
Italy	0.718	1.000	0.732	0.422
Jamaica	0.699	0.636	0.506	0.576
Japan	0.437	0.965	0.512	0.600
Jordan	0.696	0.615	0.431	0.431
Kazakhstan	0.718	0.458	0.439	0.363
Kenya	0.632	0.708	0.472	0.477
Korea, Rep.	0.554	0.566	0.467	0.574
Kuwait	0.504	0.654	0.416	0.729
Kyrgyz Republic	0.698	0.538	0.470	0.538
Malaysia	0.640	0.713	0.517	0.390
Malta	0.559	0.656	0.578	0.649
Mauritius	0.540	0.588	0.561	0.551
Mexico	0.651	0.703	0.520	0.725
Moldova	0.606	0.441	0.447	0.581
Mongolia	0.641	0.458	0.396	0.454
Morocco	0.651	0.579	0.468	0.402
Mozambique	0.581	0.461	0.348	0.478
Nepal	0.530	0.455	0.617	0.538
Netherlands	0.792	0.618	0.638	0.480
Nicaragua	0.596	0.578	0.557	0.499
Norway	1.000	0.542	0.635	0.861
Pakistan	0.579	0.664	0.500	0.502
Panama	0.697	0.580	0.525	0.692
Paraguay	0.725	0.615	0.422	0.613
Peru	0.607	0.557	0.511	0.691
Philippines	0.539	0.454	0.611	0.429
Poland	0.565	0.476	0.469	0.784
Portugal	0.554	0.669	0.740	0.703
Romania	0.656	0.456	0.462	0.455
Russian Federation	0.588	0.530	0.533	0.488
Saudi Arabia	0.838	0.574	0.419	0.472
Senegal	0.517	0.525	0.616	0.691
Singapore	0.873	0.677	0.486	0.759
Slovak Republic	0.708	0.676	0.535	0.616
Slovenia	0.853	0.768	0.682	0.620
South Africa	0.575	0.635	0.562	0.467
Spain	0.711	0.782	0.737	1.000
Sri Lanka	0.774	0.589	0.474	0.697
Sudan	0.529	0.726	0.412	0.487
Sweden	0.764	0.524	0.632	0.765

Country	Social	Sustainability	Economic	Institution
Switzerland	0.863	0.757	0.677	0.794
Tanzania	0.476	0.581	0.475	0.714
Thailand	0.992	0.660	0.500	0.545
Togo	0.468	0.623	0.431	0.495
Tunisia	0.699	0.669	0.752	0.472
Turkey	0.643	0.627	0.594	0.671
Ukraine	0.644	0.613	0.503	0.797
United Kingdom	0.781	0.695	0.815	0.607
United States	0.553	0.570	1.000	0.832
Uruguay	0.705	0.533	0.537	0.568
Vietnam	0.641	0.666	0.438	0.404
Yemen, Rep.	0.552	0.659	0.768	0.504
Zimbabwe	0.784	0.597	0.426	0.333

References

Aggarwal, C. C. (2003). A framework for diagnosing changes in evolving data streams. In *Proceedings of the 2003 ACM SIGMOD international conference on management of data* (pp. 575–586).

Alizadeh, S., Ghazanfari, M., & Fathian, M. (2008). Using data mining for learning and clustering FCM. *International Journal of Computational Intelligence, 4*(2), 118–125.

Alkire, S., & Santos, M. E. (2010). *Acute multidimensional poverty: A new index for developing countries*. United Nations development programme human development report office background paper. Working Paper No. 2010/11.

Aßmann, C., & Boysen-Hogrefe, J. (2011). A Bayesian approach to model-based clustering for binary panel probit models. *Computational Statistics & Data Analysis, 55*(1), 261–279.

Azzalini, A., & Menardi, G. (2016). Density-based clustering with non-continuous data. *Computational Statistics, 31*(2), 771–798.

Banharnsakun, A. (2018). Artificial bee colony approach for enhancing LSB based image steganography. *Multimedia Tools and Applications, 77*(20), 27491–27504.

Barca, J. C., Rumantir, G., & Li, R. (2008). Noise filtering of new motion capture markers using modified K-means. *Computational Intelligence in Multimedia Processing; Recent Advances, 96*, 167–189.

Basel, S., Gopakumar, K. U., & Rao, P. (2020). Broad-based index for measurement of development. *Journal of Social and Economic Development, 21*(1), 182–206.

Bi, K. X., Zheng, Z. W., & Yang, S. F. (2011). The grey relational analysis of manufacturing industry innovation and local administration. In *The 2011 international conference on management science and engineering—18th Annual conference proceedings,*1, 595–602.

Booysen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research, 59*(2), 115–151.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372.

Costanza, R., Hart, M., Posner, S., & Talberth, J. (2009). *Beyond GDP: The need for new measures of progress.* Pardee Working Paper No.4. Boston University.

Decancq, K., & Lugo, M. A. (2009). Weights in multidimensional indices of wellbeing: An Overview. *Econometric Reviews, 32*(1), 7–34.

Deng, J. (1989). Introduction to grey system theory. *The Journal of Grey System, 1*(1), 1–24.

Desai, M. (1991). Human development: Concepts and measurement. *European Economic Review, 35*(1), 350–357.

Drewnowski, J., & Scott, W. (1966). *The level of Living Index, report 4.* United Nations Research Institute for Social Development.

Emami, M., & Razi, F. (2014). A hybrid grey based K-means and feature selection for bank evaluation. *Decision Science Letters, 3*(3), 269–274.

Goossens, Y., Mäkipää, A., Schepelmann, P., Van de Sand, I., Kuhndt, M., & Herrndorf, M. (2007). *Alternative progress indicators to gross domestic product (GDP) as a means towards sustainable development.* Brussels: European Parliament, Policy Department A: Economic and Scientific Policy.

Goyal, S., & Grover, S. (2012). Applying fuzzy grey relational analysis for ranking the advanced manufacturing systems. *Grey Systems: Theory and Application, 2*(2), 284–298.

Greco, S., Ehrgott, M., & Figueira, J. (2016). *Multiple criteria decision analysis* (2nd ed.). Springer.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2018). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research, 141*(1), 61–94.

Guptha, K. S. K., & Rao, R. P. (2018). The causal relationship between financial development and economic growth: An experience with BRICS economies. *Journal of Social and Economic Development, 20*(2), 308–326.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques.* Elsevier.

Hu, R., & Xia, H. (2012). Chaotic dynamics and chaos control in differentiated bertrand model with heterogeneous players. *Grey Systems: Theory and Application, 2*(2), 129–140.

Huang, K. C., Huang, L. F., Lin, H. Y., Lin, S. C., & Kuo, J. C. (2006). Fight evaluation of NBA teams—Application of grey relational analysis. *Journal of Information and Optimization Sciences, 27*(2), 481–498.

Huang, X., & Wang, W. (2013). Evaluating industrial economy-ecology-coordinated development level by grey target theory. *Grey Systems: Theory and Application, 3*(3), 338–348.

Huang, Y. (2014). Development prediction of logistics industry in Henan Province and its dynamic analysis. *Grey Systems: Theory and Application, 4*(2), 14–26.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data.* Prentice Hall.

Jain, A. K., & Flynn, P. (1996). *Image segmentation using clustering* (pp. 65–83). IEEE Computer Society Press.

Jin, J., Yu, Z., & Mi, C. (2012). Commercial bank credit risk management based on grey incidence analysis. *Grey Systems: Theory and Application, 2*(385), 394.

Kose, E., Burmaoglu, S., & Kabak, M. (2013). Grey relational analysis between energy consumption and economic growth. *Grey Systems: Theory and Application, 3*(3), 291–304.

Lai, D. (2003). Principal component analysis on human development indicators of China. *Social Indicators Research, 61*(3), 319–330.

Lim, D., Anthony, P., & Chong Mun, H. (2012). Maximizin bidder's profit in online auctions using grey system theory's predictor agent. *Grey Systems: Theory and Application, 2*(2), 105–128.

Lind, N. (2004). Values reflected in the Human Development Index. *Social Indicators Research, 66*(3), 283–293.

Liu, Y., Du, J., & Li, J. (2021). A novel grey object matrix incidence clustering model for panel data and its application. *Scientia Iranica, 28*(1), 371–385.

Luo, D., & Song, B. (2012). Analysis of the income elasticity of the consumer demand of Chinese rural residents and prediction of its trend. *Grey Systems: Theory and Application, 2*(2), 148–156.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281–297).

Mallikarjuna, M., & Rao, P. (2020). Application of data mining techniques to classify world stock markets. *International Journal of Emerging Trends in Engineering Research, 8*(1), 46–53.

Malsiner-Walli, G., Frühwirth-Schnatter, S., & Grün, B. (2016). Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing, 26*(1–2), 303–324.

McGranahan, D. V., Richard-Proust, C., Sovani, N. V., & Subramanian, M. (1972). *Contents and measurement of socioeconomic development. A staff study of the United Nations Research Institute for Social Development (UNRISD)* (pp. 3–136). Praeger.

Morris, M. D. (1979). *Measuring the conditions of the world's poor: The physical quality of life index.* Pergamon.

Mylevaganam, S. (2017). The analysis of Human Development Index (HDI) for categorizing the member states of the United Nations (UN). *Open Journal of Applied Sciences, 7*(12), 661–690.

New Economic Foundation (NEF). (2006). *New economic foundation report.* New Economic Foundation.

Nielsen, L. (2011). *Classifications of countries based on their level of development: How it is done and how it could be done.* IMF Working Paper, WP11/31. International Monetary Fund.

Noorbakhsh, F. (1996). *The human development indices: Are they redundant?* Occasional Papers No. 20. Centre for Development Studies, University of Glasgow.

Noorbaksh, F. (1998). A modified Human Development Index. *World Development, 26*(3), 517–528.

Nordhaus, W. D., & Tobin, J. (1973). Is growth obsolete? In M. Moss (Ed.), *The measurement of economic and social performance* (pp. 509–564). Nber.

OECD. (1973). *List of social concerns common to most OECD countries, OECD social indicator development programme*. OECD.

OECD. (1977). *Socio-economic typologies*. OECD.

OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing.

Ram, R. (1982). International inequality in the basic needs indicators. *Journal of Development Economics, 10*(1), 113–117.

Ravallion, M. (1997). Good and bad growth: The human development reports. *World Development, 25*(5), 631–638.

Roy, D. K., & Sharma, L. K. (2010). Genetic k-means clustering algorithm for mixed numeric and categorical data sets. *International Journal of Artificial Intelligence & Applications, 1*(2), 23–28.

Sahu, A. K., Datta, S., & Mahapatra, S. S. (2014). Supply chain performance benchmarking using grey-MOORA approach. *Grey Systems: Theory and Application, 4*(1), 24–55.

Santos, M. E., & Santos, G. (2014). Composite indices of development. In B. Currie-Alder, R. Kanbur, D. M. Malone, & R. Medhora (Eds.), *International development: Ideas, experience and prospects* (pp. 133–150). Oxford University Press.

Schepelmann, P., Goossens, Y., & Makipaa, A. (2010). *Towards sustainable development: Alternatives to GDP for measuring progress.* (No. 42). Wuppertal Spezial.

Sen, A. (1983). Development which way now? *The Economic Journal, 93*(372), 745–762.

Shi, H. X., Liu, S. F., Fang, Z. G., & Zhang, A. (2008). The model of grey periodic incidence and their rehabilitation. *Chinese Journal of Management Science, 16*(3), 131–136.

Shui, L. (2014). Centipede game network model of investment in real estate market based on grey integration and forwards induction. *Grey Systems: Theory and Application, 4*(2), 321–327.

Srinivasan, T. N. (1994). Human development: A new paradigm or reinvention of the wheel? *American Economic Review, Papers and Proceedings, 84*(2), 238–243.

Stiglitz, J., Sen, A. K., & Fitoussi, J. P. (2009). *The measurement of economic performance and social progress revisited: reflections and overview.* OFCE Working Paper. Columbia University.

Streeten, P. (1995). Human development: The debate about the index. *International Social Science Journal, 47*(143), 25–37.

Tai, Y. Y., Lin, J. Y., Chen, M. S., & Lin, M. C. (2011). A grey decision and prediction model for investment in the core competitiveness of product development. *Technological Forecasting Social Change, 78*(7), 1254–1267.

Todaro, M. P. (1989). *Economic development in the third world* (4th ed.). Longman.

Todaro, M. P., & Smith, S. C. (2011). *Economic development* (11th ed.). Pearson Education Limited.

Tsai, M. T., Hsiao, S. W., & Liang, W. K. (2005). Using grey theory to develop a model for forecasting the demand for telecommunications. *Journal of Information and Optimization Sciences, 26*(3), 535–547.

UN. (1975). Developing countries and levels of development, Economic and Social Council, Committee for Development Planning, 15 October.

UNESCO. (1974). Social indicators: problems of definition and selection, Report No. 30 (Paris).

UNESCO. (1976). *The use of socio-economic indicators in development planning*. UNESCO Press.

United Nations Development Programme (UNDP). (1990). *Human development report*. Oxford University Press.

United Nations Research Institute for Social Development (UNRISD). (1970). *Contents and measurements of socioeconomic development*. UNRISD.

UNRISD. (1978, 1979). *Measurement and analysis of progress at the local level* (Vols. I, II, III).

United Nations Development Programme. (2015). Human Development Report 2016: Human Development for Everyone, New York, USA.

United Nations Development Programme. (2016). Human Development Report 2016: Human Development for Everyone, New York, USA.

Van den Bergh, J., & Antal, M. (2014). *Evaluating alternatives to GDP as measures of social welfare/progress* (No. 56). WWWforEurope Working Paper.

Vázquez, S. T., & Sumner, A. (2012). *Beyond low and middle income countries: What if there were five clusters of developing countries?*. IDS Working Papers, 2012 (404), pp. 1–40.

Victor, P. (2010). Questioning economic growth. *Nature, 468*(7322), 370–371.

Wang, Q., Zhang, D., & Hu, H. (2010). Grey relational analysis method of linguistic information and its application in group decision. In S. Liu & J. Y. Forrest (Eds.), *Advances in grey systems research* (pp. 449–459). Springer.

Wilkinson, R. G., Pickett, K. E., & De Vogli, R. (2010). Equality, sustainability, and quality of life. *BMJ(Overseas and retired doctor ed.), 341*(7783), 1138–1140.

World Bank. (1991). *The World Bank annual report 1991*. Retrieved from http://documents.worldbank.org/curated/en/262711468313513741/The-World-Bank-annual-report-1991

World Bank. (2012). *How we classify countries: A short history*. Retrieved from http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:204870

World Economic Situation and Progress (WESP). (2019). Retrieved from https://www.un.org/development/desa/dpad/publication/world-economic-situation-and-prospects-2019/

Wu, C. R., Lin, C. T., & Tsai, P. H. (2010). Evaluating business performance of wealth management banks. *European Journal of Operational Research, 207*(2), 971–979.

Zarandi, S., Shahabi, V., Emami, M., Rezaee, A., & Hassanloo, M. (2014). Ranking banks using k-means and grey relational method. *Management Science Letters, 4*(10), 2319–2324.

Zhang, Q., & Chen, R. (2014). Application of metabolic GM (1, 1) model in financial repression approach to the financing difficulty of the small and medium-sized enterprises. *Grey Systems: Theory and Application, 4*(2), 311–320.

Zhao, T., Nehorai, A., & Porat, B. (2006). K-means clustering-based data detection and symbol-timing recovery for burst-mode optical receiver. *IEEE Transactions on Communications, 54*(8), 1492–1501.

Zhu, Y., Wang, R., & Hipel, K. (2012). Grey relational evaluation of innovation competency in an aviation industry cluster. *Grey Systems: Theory and Application, 2*(2), 272–283.