SOCIOECONOMIC ANALYTICS

Exploratory Analysis regarding educational indicators, financial resources, and socioeconomic conditions of Pernambuco: A Cluster Analysis

Joed Freire Pereira da Silva [1] *
Ana Paula Amazonas Soares [2]

[1] Department of Economics, Federal Rural University of Pernambuco, Recife, Brazil
Email: joed.freire@ufrpe.br
* corresponding author

[2] Department of Economics, Federal Rural University of Pernambuco, Recife, Brazil
Email: ana.soares@ufrpe.br

RESEARCH ARTICLE

Socioeconomic Analytics
https://periodicos.ufpe.br/revistas/SECAN/
ISSN Online: 2965-4661

Submitted on: 14.09.2025.
Accepted on: 27.12.2025.
Published on: 29.12.2025.

Abstract

This paper presents an exploratory study on disparities in educational indicators and their relationship with financial and socioeconomic factors across municipalities in Pernambuco, Brazil. Using Ward's Agglomerative Hierarchical Cluster Analysis, three estimates were conducted by gradually adding new variables to examine group formation. The methodology revealed regional patterns in learning and quality-of-life indicators, suggesting that per capita income alone does not fully explain educational outcomes and that FUNDEB resources may be insufficient to ensure improvements. Clusters are not necessarily composed of geographically neighboring municipalities, as similarities can occur irrespective of spatial proximity. The study also underscores the limitations of the method and highlights the need for complementary approaches to achieve a more comprehensive analysis.

Keywords

Educational indicators. Financial resources to education. Socioeconomic factors. Cluster analysis. Pernambuco. Public policy.

1. Introduction

The relationship between human development, income growth, social inequality reduction, and improved quality of life is widely acknowledged as intrinsically linked to education. Various studies and academic works, such as those by Cangussu et al. (2009) and Alencar et al. (2013), reiterate this fundamental connection. From an economic perspective, education is

regarded as a strategic investment capable of driving inclusive growth and generating positive externalities as highlighted by Mankiw (2013), Borjas (2012), Meier (2005), and Pires (2005).

Considering this central role of education, it becomes crucial to examine how government actions influence human capital development, especially in contexts where the public sector holds primary responsibility for educational formation, as is the case in Brazil. It is fundamental that policies implemented by governments at all levels are effective in promoting educational excellence, thereby contributing to improvements in economic and social development indicators.

In this context, indicators are used to assess the effectiveness of state action, but some of them paradoxically create incentives that harm educational quality. The pursuit of better quantitative results may compromise the intrinsic quality of education, representing a trade-off that requires careful consideration. Thus, apparent improvements in certain indicators do not always translate into significant advances in human capital quality, population well-being, productivity, or overall quality of life.

To evaluate the performance of policies that employ those indicators, robust analytical tools can be used to move beyond surface-level metrics of educational outcomes. Modern and efficient scientific instruments, which combine advanced statistical techniques with technological resources to process available data, can be employed to construct an objective and up-to-date overview of the educational landscape. These tools offer initial insights and also deepen understanding, subsidizing informed decision-making.

In this case study, to investigate the contribution of different variables to educational and socioeconomic complexity, a cluster analysis was conducted in stages, progressively adding new variables to evaluate how each enhanced the differentiation between the formed groups. The primary aim was to analyze, through an exploratory approach, the effectiveness of educational policies, with a focus on the allocation of financial resources, in transforming the socioeconomic context of Pernambuco. A secondary objective was to examine the spatial distribution of key variables across the state's municipalities and, specifically, to characterize Pernambuco’s geographic space through data visualizations that highlight regions demanding greater attention and intervention.

This article is structured into five additional sections. Section 2 presents a theoretical review of education’s role as a driver of social transformation. Section 3 discusses educational data at different territorial levels, with a focus on contextualizing the state of Pernambuco. Section 4 outlines the methodology employed for cluster analysis. Section 5 presents the results of this analysis for Pernambuco, and Section 6 offers a discussion of the conclusions drawn from the observed outcomes and their implications for the state.

2. Productivity, social welfare and regional challenges

This section delves into the multifaceted role of education as a fundamental determinant of productivity, social welfare, and the mitigation of regional disparities. It will explore the theoretical underpinnings that links education to economic growth and individual well-being, followed by an examination of the situation in the context of Brazil and Pernambuco, particularly concerning issues of inequality and development challenges.

2.1. Education as a driver of productivity and quality of life

According to Mankiw (2013), disparities between economies can largely be attributed to variations in productive efficiency, which are determined by physical capital, natural resources, human capital, and technology. Mochón (2007) complements this perspective by arguing that a country's productivity can be measured through the average output each worker contributes to the economy, highlighting that per capita income results from both labor productivity and the proportion of employed individuals in the population. From a microeconomic standpoint, Borjas (2012) emphasizes that firms hire workers based on their marginal revenue compared to its cost, and due to diminishing marginal returns, sustained employment and wage growth depend on significant gains in labor productivity.

Education plays a pivotal role in this context. Low-skilled labor faces higher barriers to labor market integration, especially as new technology is adopted (Martone, 2015). Beyond improving individual productivity, education fosters social capital by enhancing collective efficiency through coordinated actions. Pires (2005), citing Putnam, notes that educated individuals are more likely to engage civically and act responsibly, creating social environments that encourage economic growth. Romer and Lucas further underscore that education is the key to improving productivity, integrating human capital and ideas into economic development models and refining earlier growth theories (Jones, 2013). Consequently, regional economies with low-quality educational processes are likely to face substantial development challenges.

Empirical studies in Brazil support this theoretical foundation. Alencar et al. (2013) demonstrate that states with lower average per capita household income also tend to exhibit lower levels of educational attainment. Fernandes (2001), as cited by Alencar et al. (2013), highlights that investment in human capital through education is a central mechanism for increasing potential future earnings. Souza et al. (2013) examine the relationship between inequality and education, showing that higher income and educational attainment are inversely correlated with distributional inequality. This aligns with the foundational insights of Schultz (1963) and Becker (1964), who identified that the redistribution of educational opportunities contributes to reducing social inequality. Collectively, these studies suggest that both income and education are negatively correlated with the Gini Index, reinforcing the role of education in promoting a more equitable distribution of wealth.

Taken together, these findings suggest that productivity, education, and inequality are deeply entangle, with educational investment functioning both as a growth driver and as a mechanism of social equity. This dual role makes education particularly strategic in contexts of persistent regional disparities, such as in Brazil.

2.2. Private and Social returns

Beyond economic performance, education generates both private and social returns (Alencar, 2013). At the individual level, it enhances quality of life through better resource use, health awareness, and increased potential earning. At social level, education creates positive externalities by improving health, promoting civic engagement, and reducing dependence on state interventions (Meier, 2005; Alencar, 2013). Notably, increased education among women raises opportunity costs of domestic work, boosts labor market participation, and contributes to poverty reduction and improved public service outcomes (Ribemboim, 2019). A region's development, measured through indicators such as literacy, income, poverty, and

employment, underscores education as a central determinant of regional inequality (Mochón, 2007). Illiteracy and low education remain obstacles to development, reinforcing the regional component of inequality (Adas, 2001). Ultimately, education is inherently valuable, consistently improving individual and societal well-being (Pires, 2005).

Finally, in his work titled *Economics of Education*, Pires (2005) concludes that education is an investment whose outcome is always an improvement in the condition of citizens. Education is inherently valuable, because, regardless of the intentions behind its implementation, it consistently leads to an welfare state for the population compared to their condition before receiving education.

2.3. Human Capital Theory and state action

Education functions as an investment with measurable returns. Human capital theory posits that additional years of schooling increase individual productivity, contributing to higher personal income, per capita GDP, and reduced inequalities. The decision to pursue more education depends on the individual discount rate applied to its future consumption, with higher rates typically observed among lower-income individuals (Borjas, 2012; Fernandes, 2001; Meier, 2005). Investments made at younger ages are generally more profitable, as the accumulation of human capital over time maximizes potential returns.

School quality is another crucial determinant of human capital development. Empirical studies show that student-teacher ratios and teacher qualifications are strongly associated with student performance (Borjas, 2012; Aaronson et al., 2007). INEP (2019) data indicate that teacher adequacy varies considerably by region, with the North and Northeast of Brazil, including Pernambuco, displaying the lowest rates of teacher qualification. These findings underscore the importance of improving educational conditions to enhance student outcomes and long-term productivity.

In Pernambuco, low levels of literacy combined with early entry into the labor market pose significant challenges to human capital formation. This context highlights opportunities for targeted educational investments, particularly in municipalities with lower human capital stocks. By addressing both: quantity and quality of education, investments can increase productivity, promote social equity, and reduce regional disparities, ultimately contributing to sustainable economic and social development.

3. The importance of State intervention

Positive externalities, efficiency gains, and poverty reduction associated with raising educational levels should not be ignored. Martone (2015) shows that market failures in human capital investment could lead to under allocation relative to the social optimum, resulting in negative consequences for long-term product growth and real income. Based on this, the author argues that government promotion of education is essential. Complementing this view, Riani (2010) suggests that without public sector involvement in supporting education, the most disadvantaged would be excluded from access to this crucial means of private and social development. Therefore, the state's role and evaluation of its actions are fundamental to reducing inequalities, extreme poverty, and poor well fare states.

3.1. Disparities of educational indicators in Pernambuco

In the specific context of Pernambuco, Racena (2018) points out that, referring to programs established by the Northeast Development Working Group (GTDN) in 1958, population

training was not prioritized in past development plans. The author states that the importance of education in eradicating poverty was not recognized, although education is a necessary — albeit not sufficient — condition for overcoming poverty and integrating into a competitive economy.

On the other hand, for more than a decade, national efforts have been made to improve the quality of education. Pernambuco has experienced progress and stands out among Northeastern states through various programs such as the Integral Education Program, the Paulo Freire/Pernambuco Escolarizado Program, Public Management Modernization, and the "Ganhe" Program. These initiatives aim to reward professionals who meet targets, improve student training, and create opportunities (Araújo & Lima, 2018). These steps are important because inequalities among federative units persist, and the overall quality of education remains low throughout the country. The case of Pernambuco illustrates the complexity of this dynamic at a subnational level.

Although Pernambuco has shown notable advances and initiatives, significant challenges and regional disparities persist when compared to the states in the South and Southeast of Brazil (IBGE, 2019).

One of the fundamental indicators analyzed is the illiteracy rate. Pernambuco stands out in the Northeast for having the lowest rates, with a reduction from 14.4% to 12.6% of illiterate individuals aged 18 and older between 2017 and 2018. However, this regional improvement contrasts sharply with the South and Southeast regions, where rates are considerably lower, highlighting persistent educational inequality in the country (IBGE, 2019). This reality is reinforced by Racena (2018), who argues that high illiteracy rates significantly hinder development in an increasingly competitive world. The spatial distribution of illiteracy in Pernambuco underscores this disparity, with only 8 out of 184 municipalities exhibiting rates below 14%, mostly concentrated along the coast and in the Recife Metropolitan Region, with few exceptions in the interior (Fundação João Pinheiro, 2019).

The quality of the teaching staff, a crucial factor for educational excellence, also presents challenges in Pernambuco. In 2018, the state was among those where only 20% to 31% of early childhood education teachers had completed higher education degrees appropriate to the subjects they taught (INEP, 2019). This adequacy is vital, as Borjas (2012) suggests, pointing to teacher qualification as a key element impacting student performance. INEP (2019) also indicates that the percentage of classes taught by properly qualified teachers varies significantly across Brazil's regions, with the North and Northeast presenting the worst indices.

Another aspect of educational quality is the average number of students per class. In Pernambuco's elementary education, there has been a slight improvement: the percentage of municipalities with more than 27 students per class decreased from 7% in 2010 to 5% in 2017 (INEP, 2019). This may affect learning quality, given the relationship between class size and student performance (Borjas, 2012). Geographically, municipalities with higher student numbers in elementary education are dispersed.

School attendance rates and age-grade distortion serve as key indicators of student flow and retention within the educational system. In 2010, nearly all municipalities in Pernambuco (over 83%) provided schooling for children aged 6 to 14, which represents a positive outcome. Attendance issues are more pronounced at higher levels of education, a trend

exacerbated by the fact that the main reasons for non-attendance or school dropout — particularly among youth— are the pursuit of employment or the need to care for household chores and dependents, reflecting significant socioeconomic pressures on students. Over time, improvements in attendance rates have been observed in elementary education; on the other hand, despite these reductions in non-attendance, most municipalities still report age-grade distortion rates between 17.6% and 42.2% at the elementary level, indicating that aligning students with the appropriate grade for their age remains an area requiring further attention (BDE, 2019).

In summary, Pernambuco has made progress in various educational indicators, standing out regionally in reducing illiteracy and improving age-grade distortion in secondary education. However, the state still faces substantial disparities compared to more developed regions of the country, especially regarding teacher adequacy and the persistence of high illiteracy and distortion rates in more vulnerable areas. The analysis of these indicators reveals a complex scenario for this case study, where advances coexist with structural and regional challenges that require ongoing public policies and investments to ensure educational excellence and equity.

4. Methodology

Mingoti (2005) defines multivariate statistics as a set of statistical methods used when a large number of variables—usually correlated—are measured simultaneously and need to be analyzed together. In such cases, univariate statistical analysis becomes more complex and limited. For the present case study in Pernambuco, the complexity of educational and socioeconomic indicators necessitates a multivariate approach that allows capturing and understanding the interrelationships among these variables.

The technique employed in this research falls under the category of multivariate statistics and is known as cluster analysis. According to Mingoti (2005), cluster analysis is applicable in various contexts, including Geography, where it is used to classify cities, states, or countries based on demographic, physical, or economic characteristics. Its application in this case study allows for the identification of municipality profiles within Pernambuco.

Clustering is a long-standing practice employed to detect patterns that may initially be hidden from the analyst. More specifically, clustering is a set of machine learning procedures that, based on a definition of similarity or dissimilarity among elements, allow for the grouping of data into subsets that are internally homogeneous but significantly different from each other. For the Pernambuco study, this technique enables the grouping of municipalities with similar educational, financial, and socioeconomic characteristics, revealing the state's internal heterogeneity.

The clustering process can be classified into supervised and unsupervised learning. As noted by James et al. (2013), in unsupervised learning — under which cluster models like the one adopted here are categorized — there is no response variable to verify correlations. Instead, the learning algorithm simply searches for patterns within the input data and groups them without prior assumptions or predefined labels. This approach is ideal for exploring the natural structure of data from Pernambuco's municipalities, without imposing prior assumptions.

Clustering typically involves three main stages. First, it is necessary to define which variables will serve as criteria for forming groups. Second, a measure of similarity or dissimilarity must

be selected to guide the grouping process. Finally, the analyst chooses a clustering procedure, determining how the data will be grouped and how the number of clusters will be established.

The similarity measure used in this study is the minimum variance method proposed by Ward in 1963, also known as Ward's method. This approach ensures that the variance within each partition is minimized, correcting the tendency — observed in some other methods — for intergroup variance to decrease and intragroup variance to increase as the algorithm progresses. The choice of Ward's method for this case study aims to ensure the formation of cohesive and distinctive groups of municipalities.

Once the variables and similarity measure are defined, it is possible to either (1) assume an ideal number of groups and assign the remaining data based on representative elements (non-hierarchical methods), or (2) allow the algorithm to explore the data without predefining the number of clusters (hierarchical methods). Furthermore, based on the research objectives, the analyst must decide whether to begin by treating all data as initially distinct (an agglomerative process) or initially identical (a divisive process). For the Pernambuco case study, an agglomerative hierarchical approach was chosen to explore the similarity relationships among municipalities.

In this research, the clValid package (in R) was used to compare the performance of hierarchical, PAM, and K-means clustering methods. The results indicated that the hierarchical method performed best. This package also helped infer the optimal number of clusters, suggesting four or five clusters. However, in the hierarchical method itself, no specific number of clusters is imposed a priori.

4.1. Pernambuco's Characteristics

The data used in this case study are divided into three main dimensions for a comprehensive characterization of Pernambuco's municipalities: Educational, Financial, and Socioeconomic. Educational data were collected from the 2010 School Census, selecting only those indicators available for all municipalities.

The second group of data focused on FUNDEB (Fund for the Maintenance and Development of Basic Education and the Valorization of Education Professionals), which reflects municipal expenditure per student for the year 2010. For the purposes of this study, the calculation was made by dividing the total amount received through FUNDEB transfers in 2010 by the number of students in fact enrolled in that same year. This approach is used since FUNDEB's transfers are based on student enrollment figures from the previous year, not the current one.

The socioeconomic indicators reflect characteristics such as mortality and life expectancy, the Human Development Index (HDI) in its dimensions of longevity and income, income concentration, per capita income, income of the poorest population segment, the percentage of people living in poverty, and access to healthcare. It is important to emphasize that the data used in this case study represents a snapshot of Pernambuco' scenario in the year 2010, serving as a basis for understanding the municipal dynamics and typologies of that period. See Table 1, listing all variables.

Table 1: Variables, descriptions and dimensions

Code	Name	Dimension
MATI	Average number of students per class in early childhood education	Education
MATF	Average number of students per class in elementary school	Education
MATM	Average number of students per class in high school	Education
ADFIT1	Teacher qualification adequacy (initial years, group 1)	Education
ADFIT2	Teacher qualification adequacy (initial years, group 2)	Education
ADFIT3	Teacher qualification adequacy (initial years, group 3)	Education
ADFIT4	Teacher qualification adequacy (initial years, group 4)	Education
ADFIT5	Teacher qualification adequacy (initial years, group 5)	Education
ADFFT1	Teacher qualification adequacy (final years, group 1)	Education
ADFFT2	Teacher qualification adequacy (final years, group 2)	Education
ADFFT3	Teacher qualification adequacy (final years, group 3)	Education
ADFFT4	Teacher qualification adequacy (final years, group 4)	Education
ADFFT5	Teacher qualification adequacy (final years, group 5)	Education
ADMT1	Teacher qualification adequacy (high school, group 1)	Education
ADMT2	Teacher qualification adequacy (high school, group 2)	Education
ADMT3	Teacher qualification adequacy (high school, group 3)	Education
ADMT4	Teacher qualification adequacy (high school, group 4)	Education
ADMT5	Teacher qualification adequacy (high school, group 5)	Education
RI	IDEB Performance Index (initial years)	Education
RF	IDEB Performance Index (final years)	Education
ABDI	Dropout rate in the initial years of elementary school	Education
ABDF	Dropout rate in the final years of elementary school	Education
API	Approval rate for the initial years of elementary school	Education
APF	Approval rate for the final years of elementary school	Education
DISI	Age-grade distortion rate in the initial years of elementary school	Education
DISF	Age-grade distortion rate in the final years of elementary school	Education
REPI	Repetition rate for the initial years of elementary school	Education
REPF	Repetition rate for the final years of elementary school	Education
DISM	Age-grade distortion rate in high school	Education
ABDM	Dropout rate in high school	Education
REPM	Repetition rate in high school	Education
IDHME	Human Development Index - Education	Education
TXALF	Literacy rate	Education
FPM	FUNDEB (Financial resource to Development of Basic Education)	Economic
IDHMR	Human Development Index - Income	Social
IDHML	Human Development Index - Longevity	Social
Igi	Gini Index	Social
EV	Life expectancy at birth (2010)	Social
RPC	Per capita income	Economic
RPCPB	Per capita income of the poorest 40%	Economic
PFS	Health professionals per 1000 inhabitants	Health
ES	Healthcare facilities per 1000 inhabitants	Health
MI	Infant mortality rate (2010)	Social
MI5	Under-5 mortality rate (2010)	Social
% PB	Poverty rate (2010)	Social

P60	Probability of survival until age 60 (2010)	Social

4.2. Estimations

Unsupervised hierarchical estimations using Ward's linkage method and agglomerative clustering were carried out for all municipalities in the state of Pernambuco, except for two municipalities: Maraial and the State District of Fernando de Noronha, due to the unavailability of complete data for all variables for these locations.

The first estimation considers only educational indicators and therefore forms clusters of municipalities with similar educational performance. This analysis is important as it lays the groundwork for identifying which municipalities in Pernambuco had the best educational outcomes in isolation.

The second estimation introduces financial resource data, explicitly the effective expenditure per student available to each municipality in 2010. This value was calculated based on the total amount received through FUNDEB transfers for that year, divided by the actual number of enrolled students. Since the transfers are calculated based on enrollment numbers from the previous year, this results in an effective, not estimated, measure of expenditure. By adding this new variable, the second estimation incorporates both performance and expenditure into the grouping criteria, allowing municipalities to be clustered based on similar educational results and funding levels.

The third estimation expands the analysis by including socioeconomic indicators in addition to educational performance and expenditure. This results in a new division of municipalities into clusters that reflect similarities across three dimensions: education, financial investment, and socioeconomic context. This stage is crucial for a more complete understanding of Pernambuco’s case.

As the estimations progressively introduce more complex variables, they reveal the interconnectedness of education with broader social and economic indicators which affect both students and their families. With the inclusion of these dimensions, it becomes possible to characterize the clusters not only by educational performance, but also by developmental characteristics, in line with the approach proposed by Écio (2008) and the indicators discussed by Mochón (2010), providing a multifaceted portrait of the reality of Pernambuco's municipalities.

5. Results

The initial estimation, which establishes the basis for the grouping structure, identified five clusters of municipalities. The relative importance of these clusters was determined by calculating the average values of their indicators, with variables classified into two categories: (i) those for which higher values are desirable; and (ii) those for which lower values are preferable. On this basis, a hierarchy of groups was constructed according to their average indicator values. The results are presented in Table 2. It is important to note that this table, as well as the others throughout this paper, present in column 1 of 'Groups' a resource of analysis that identifies the ranking of the best-performing clusters.

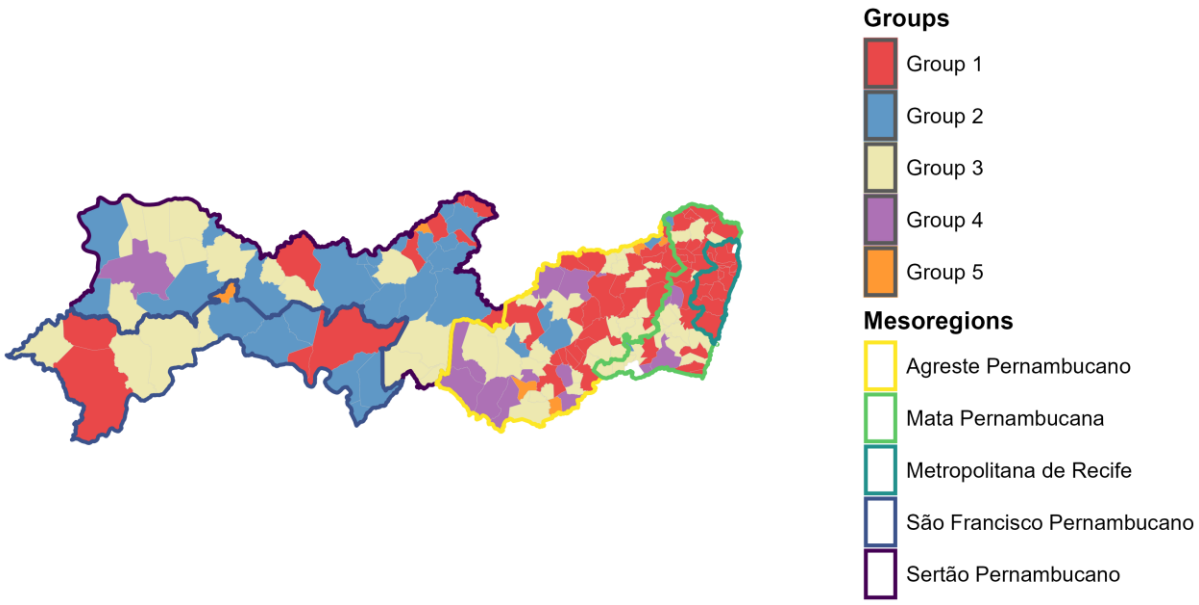
In Table 2, it is evident that Group 2 exhibits the highest average for the higher is better variables and the lowest average for the lower is better variables. This superior performance positions it as the reference group in terms of pure educational indicators.

Table 2: First estimate with educational varibles

Groups	Number of municipalities	Mean (higher is better)	Mean (lower is better)
2	31	76.35	17.35
1	71	75.58	19.73
3	60	70.97	21.36
5	6	69.63	20.96
4	15	66.00	26.09
Total	**183**	**73.22**	**20.42**

Furthermore, it can be seen in Figure 1 that Group 2 has 80% of its members located in the Sertão region, with the exception of a few municipalities in the Agreste (Sanharó, Venturosa, São Bento do Una, Orobó) and Zona da Mata (Macaparana and Belém de Maria). Group 1 has 37% of its municipalities in the Agreste and 28% in the Zona da Mata. Group 3 has 45% of its members in the Agreste. Both Groups 4 and 5 have 67% of their members in the Agreste. In this case, despite the high regional shares, Groups 4 and 5 comprise relatively few municipalities, with Group 4 containing 15 and Group 5 only 6 members.

Figure 1: Spatial distribution of educational indicators



From the beginning, the level of financial resources draws attention. At this initial stage of the analysis, Groups 3, 5, and 4 — those with the poorest educational outcomes — presented higher average FUNDEB per-enrollment values R$ 1,248.14; R$ 1,292.12; and R$ 1,320.04, respectively.

Considering the second estimate, which incorporates both educational and FUNDEB financial data in the clustering process, approximately 27% of municipalities change groups, with the main effect observed in the reallocation of municipalities that belonged to Group 1 in the first estimate. The new averages for the educational variables can be seen in Table 3. In this case, it is observed that Groups 1 and 2 exhibit the highest averages for the higher is better variables and the lowest averages for the lower is better variables.

Table 3: Second estimate including financial resources

Groups	Number of Municipalities	Mean (Higher is Better)	Mean (Lower is Better)
1	30	77.59	19.51
2	45	75.75	17.70
3	87	75.06	21.19
5	6	69.63	20.96
4	15	65.03	25.74
Total	**183**	**73.22**	**20.42**

In Table 4, it should be noted that the rows represent Estimate 1 and the columns represent Estimate 2. Therefore, the diagonal shows the municipalities that maintained their relative ranking based on educational indicators after the introduction of FUNDEB. As can be seen, the main shift in ranking therefore occurred between Group 2 of Estimate 1 and Group 1 of Estimate 2. Furthermore, municipalities belonging to Group 5 in the previous estimate remain consolidated.

Table 4: Reallocation across groups

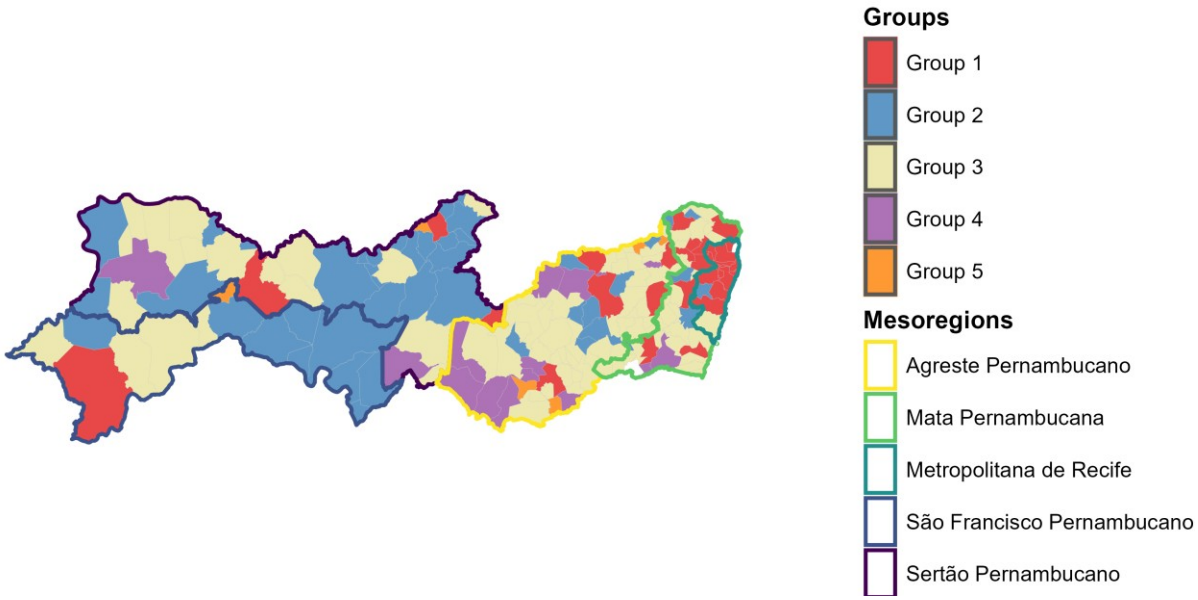
Estimate 1	Estimate 2					Total
	1	**2**	**3**	**5**	**4**	
2	1	28	2			31
1	29	17	25			71
3			57		3	60
5				6		6
4			3		12	15

Positions above and to the right of the main diagonal in Table 4 indicate municipalities whose average educational indicators worsened relative to their previous ranking, while positions below the main diagonal indicate municipalities that improved. For instance, the municipality of Salgueiro migrates from Cluster 2 to Cluster 1, remaining grouped with municipalities that exhibit better educational indicators. Similarly, São João, Chã Grande, and Pombos are grouped with municipalities in a better learning situation, migrating from Cluster 4 to Cluster 3. In contrast, Sanharó and São Bento do Una are among the municipalities that, in Estimate 2, are grouped with those showing poorer indicators. Additionally, Inajá, Caetés, and Salgadinho also join a cluster with lower educational performance. As previously mentioned, the municipalities in Cluster 5 (Solidão, Terra Nova, Casinhas, Lagoa do Ouro, Machados, and Saloá) remain consolidated, maintaining the same relative condition regarding their averages, indicating greater homogeneity.

The most significant implication of this transition concerns the relationship between educational performance and FUNDEB allocations across the leading clusters. In Estimate 1, the highest-performing cluster reported an average per-student FUNDEB allocation of R$ 1,047.90. Once financial variables were incorporated into the clustering procedure (Estimate 2), however, the new top-performing cluster emerged with a substantially lower average allocation of only R$ 768.03 per student. Once again, the clusters positioned lower in the hierarchy of Estimate 2 generally exhibit higher average FUNDEB allocations — Clusters 3 (R$ 1,208.29), 5 (R$ 1,292.12), and 4 (R$ 1,305.46). In contrast, the leading Clusters (1 and 2) receive, on average, approximately R$ 300 less per student.

The inclusion of additional FUNDEB resources reveals that funding alone does not guarantee comparative improvements in educational outcomes: some municipalities were unable to convert extra resources into relative gains, whereas others effectively managed funds and achieved notable comparative improvements, highlighting the importance of both resource availability and efficient use. Redistribution of resources is not sufficient to guarantee strong educational outcomes. Municipalities such as Ouricuri, Betânia, Brejo da Madre de Deus, Saloá, Iatí, and Solidão, among others, despite receiving allocations above R$ 1,100 per student, rank among the lowest in educational indicators in both previous estimates. Similarly, the municipalities of Dormentes, Santa Filomena, and Santa Cruz receive allocations above R$ 1,400 and have comparable income levels (low income); however, their educational outcomes differ. Dormentes and Santa Filomena are grouped in the first and second highest-performing clusters in Estimate 1, while Santa Cruz falls into the third cluster. The spatial distribution for Estimate 2 can be seen in the map in Figura 3 below.

Figura 2: Spatial distribution of educational indicators and FUNDEB



The map above shows that Clusters 2 and 3 are concentrated in the interior of the state, with 80% of the municipalities located in the Sertão. The Agreste region accounts for 48% of the municipalities in Cluster 3. Similarly, the members of Clusters 4 and 5 are also concentrated in the Agreste, with 73% and 66% of the municipalities in these clusters, respectively, located in this region.

The third estimate introduces variables reflecting the social and economic realities of the municipalities, such as the income of the poorest residents, the percentage of poor individuals, and other indicators. The inclusion of socioeconomic variables leads to changes in the clustering, most notably in Cluster 1, and in this case also in Cluster 3. The ranking of clusters is also significantly altered. The new averages for the clusters and their ranking can be seen in Table 5. It is observed that clusters 1 and 5 achieved the highest positions in the ranking, obtaining the highest averages for the higher is better variables. Together, clusters 1 and 5 represent approximately 30% of the municipalities in the state, with 20 and 34 members, respectively.

Table 5: Third estimate with financial resources and socioeconomic variables

Groups	Number of Municipalities	Mean (Higher is Better)	Mean (Lower is Better)
1	20	79.32	18.53
5	34	74.94	20.15
2	47	74.29	17.89
3	50	72.09	20.95
4	32	67.75	24.77
Total	**183**	**73.22**	**20.42**

Cluster 1 emerges as the top-performing group, with an even lower average FUNDEB allocation (R$ 661.41). In this case, it can also be observed that municipalities that stand out in a more comprehensive scenario — considering education, financial resources, and social/economic context — either demonstrate high efficiency in the use of resources or already possess favorable socioeconomic characteristics (such as per capita income, access to healthcare and others).

Table 6: Reallocation following the inclusion of socioeconomic variables

Estimate 1	Estimate 3					Total
	1	5	2	3	4	
2	4		23	4		31
1	16	31	12	12		71
3		3	10	31	16	60
5			2	3	1	6
4					15	15

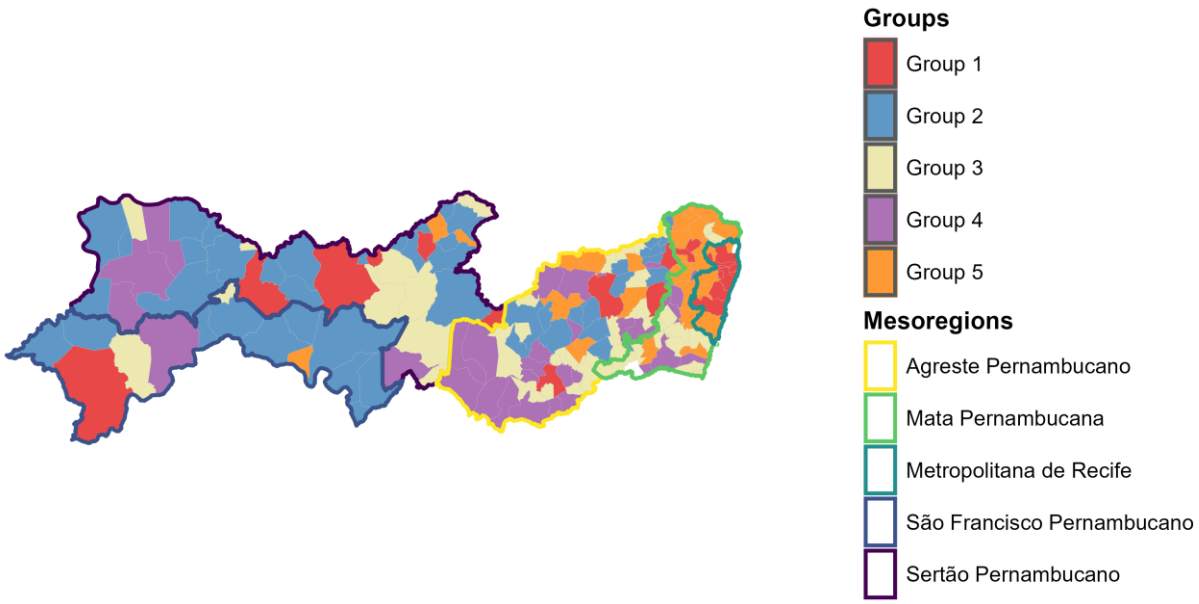
Table 6 illustrates the changes in cluster membership between Estimate 1 and Estimate 3. The last column, as before, indicates the initial number of municipalities in each group from Estimate 1. The table shows that 16 municipalities that belonged to the second-best group in Estimate 1 moved to the top position in the new ranking (Estimate 3). Only four cities from the highest-performing group in Estimate 1 — Salgueiro, Serra Talhada, Triunfo, and Afogados da Ingazeira — maintain their status. From group 3, Vicência, Aliança, and Belo Jardim join the second-best group. From group 5, Solidão and Machados are the two municipalities that show improvement. Finally, Table 6 confirms that all municipalities in group 4 remain in the same position.

The spatial distribution of the data can be observed in Figura 3 below. It is noted that the groups with the best outcomes (clusters 1 and 5) are concentrated in the Metropolitan Region and Zona da Mata, with the exception of a few municipalities such as Itacuruba, Tabira, Petrolina, and Tuparetama, which are located in the São Francisco and Sertão regions. In this regard, 40% of the members of cluster 1 are in the Metropolitan Region, while cluster 5 has 50% of its municipalities in the Zona da Mata. The remaining municipalities in these clusters are distributed throughout the rest of Pernambuco.

With reference to the map in Figura 3, clusters 2 and 3 collectively encompass approximately 53% of the municipalities in the state and are entirely situated in the interior, beginning from the Zona da Mata. Cluster 2 includes 43% of its members in the Sertão, whereas cluster 3 has 48% of its municipalities located in the Agreste. Cluster 4, in turn, comprises 62% of its

participants in the Agreste. Cluster 1 includes some municipalities in the Recife Metropolitan Region (RMR) and throughout all regions of Pernambuco.

Figura 3: Spatial distribution of educational indicators, FUNDEB and socioeconomic variables



Moreover, the spatial distribution of income across the state can be analized. The highest per capita income group consists of seven municipalities (Caruaru, Jaboatão dos Guararapes, Olinda, Paulista, Petrolina, Recife, and Santa Cruz do Capibaribe), approximately 57% of which are situated in the Metropolitan Region, with the remainder located in the São Francisco and Agreste regions.

The intermediate-income group (350–500 per capita) has around 27% of its members in the Sertão. Municipalities with per capita income below 350 are predominantly concentrated in the Agreste and Zona da Mata. Notably, the data reveal that certain municipalities achieve relatively strong educational outcomes despite per capita incomes below 280, underscoring that income alone does not fully account for disparities in learning performance.

The third estimate, which introduces variables reflecting the social and economic realities of the municipalities — such as the income of the poorest residents, the percentage of poor individuals, and other previously mentioned indicators — reveals a substantially greater potential for distinguishing the groups than that observed with the inclusion of FUNDEB alone. While the second estimate primarily caused a reshuffling of the ranking among the groups, the addition of socioeconomic indicators in the third estimate results in a substantial reconstruction of the clusters. This suggests that socioeconomic variables have a stronger capacity to characterize and differentiate municipalities, reorganizing them into clusters with more distinct identities.

In the transition from Estimate 1 to Estimate 2 (with FUNDEB), and again to Estimate 3 (with socioeconomic variables), the groups that initially stood out for their pure educational indicators tend to be demoted or blend with lower-performing groups when other variables are considered. For instance, municipalities in the Sertão and Agreste regions that, in Estimate 1, grouped with the status of a better group in terms of educational indicators, successively

migrated to worse groups from Estimate 2 onwards, confirming this tendency for relative ranking loss.

The Human Capital Theory and the authors cited argue that education, by improving productivity, income, and individual well-being, generates positive externalities that affect society, impacting social indicators such as Gini, poverty, and health. Therefore, FUNDEB, by financing basic education, is expected to contribute to the improvement of the social environment in the long run.

However, the fact that the best performing groups (in terms of education/FUNDEB) are reconfigured when social indicators are added and often are not the same groups that stand out socially, indicates that the relationship is not simple or immediate, contrary to what some might expect. One could argue that "the improvement in educational and, consequently, social indicators, does not appear to be a linear or direct function of the volume of FUNDEB resources in the short term, given the complexity of the factors involved. Furthermore, socioeconomic variables possess a deeper discriminatory power.

6. Conclusion

The methodology applied in this study made it possible to identify clusters of municipalities with similar characteristics across the state's territory. In addition, it allowed for the assessment of how municipal characteristics change after the introduction of new variables. Furthermore, the study revealed the regional distribution of groups with favorable educational and socioeconomic indicators, supporting the conclusion that the research objectives were successfully met.

One of the central findings of this study is that financial resources, including those from FUNDEB, are not sufficient on their own to explain municipal educational outcomes. While municipalities in the Recife Metropolitan Region (RMR) have higher income levels and better living conditions when socioeconomic variables are considered, they do not rank among the top performers in pure educational indicators (Estimate 1, Group 2). Similarly, municipalities in the interior of the state — particularly in the Sertão — achieve strong educational results despite limited financial input. This suggests that effective outcomes may stem either from resource efficiency or from other structural advantages.

The three estimates demonstrated that the inclusion of new variables gradually altered the composition of municipal clusters. Municipalities in the Sertão and Agreste regions, which initially ranked highly in educational indicators, progressively migrated to lower-performing groups after the introduction of financial and socioeconomic variables (Estimates 2 and 3). In Estimate 3, only 30% of the top-performing municipalities (Group 1) remain in the Sertão, while just 9% of Group 5—the second-best performing group—are located in this region.

Meanwhile, higher-income municipalities (with per capita income above R$ 351) are largely concentrated in the RMR. However, strong educational outcomes are also observed in Sertão municipalities such as Carnaíba, Tabira, Santa Terezinha, São José do Belmonte, Cedro, and Parnamirim. This underscores that income alone does not determine educational performance, and that some low-income municipalities outperform wealthier ones.

When assessing Estimate 2, it became evident that FUNDEB transfers did not significantly alter the positions of municipalities within the clusters. Only 32 municipalities moved to groups with higher averages, 58 experienced a decline, and the rest remained unchanged.

These findings suggest that FUNDEB interacts only moderately with educational indicators, as evidenced by the limited changes in group configurations following the introduction of this variable. Specific cases include Jataúba, Tupanatinga, Itaíba, Águas Belas, Serrita, and Saloá — all of which receive over R$ 1,200 in federal financial support, yet remain in clusters with lower learning outcomes.

These results confirm that there is no direct, linear relationship between the volume of FUNDEB resources and positive educational or social indicators. In fact, groups with better learning outcomes often receive less FUNDEB funding, while more challenged groups receive more, suggesting that although the fund fulfills its redistributive purpose, resource volume alone does not ensure educational success. It is intuitively recognized that a larger volume of financial resources can expand the possibilities for intervention and improvement in education. Thus, the present analysis does not aim to recommend a reduction in FUNDEB resources but rather emphasizes the pressing need for more efficient allocation. Understanding that greater funding alone does not guarantee better outcomes opens the door to discussions about the importance of other factors. Elements such as the quality of municipal management, engagement of the school community, teacher training and valorization, and the municipalities' own socioeconomic and environmental characteristics (income, access to services, etc.) may be the true drivers of educational performance. These factors, in turn, can be influenced by performance over time, creating a virtuous cycle of development.

Across the three estimates, it is observed that geographic proximity is not a necessary condition for the formation of similar clusters. While there is some degree of spatial concentration suggesting potential regional interrelations, similarities were also identified among geographically distant municipalities. For example, municipalities such as Saloá, Casinhas, Solidão, Lagoa do Ouro, Terra Nova, and Machados remained grouped together throughout all estimates, indicating the presence of common structural or contextual characteristics not tied to territorial location.

Finally, it is important to reiterate that this is an exploratory study. Unsupervised clustering techniques do not allow for statistical correlation analysis and lack predictive capabilities. As highlighted in the introduction, spatial econometrics offers methodologies better suited for studies that consider regional dependencies and enable examination of the relationships between internal municipal variables and those of their neighbors. Thus, future research employing spatial econometric techniques would provide a valuable complement to this work. Furthermore, analyzing efficiency in the use of public funds represents another promising line of inquiry. The limitations of the methods used here are acknowledged, which underlines the incomplete nature of the present study and opens opportunities for future research to expand upon its findings.

Acknowledgements

We are deeply grateful to God for guiding our paths and giving us strength through challenges. We thank our families and friends for their sacrifices, support, and encouragement. We also thank the Federal Rural University of Pernambuco for providing a learning experience that fostered and continues to foster both knowledge and personal growth.

Conflict of Interest Declaration

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

References

1. Adas, M. (1998). *Panorama geográfico do Brasil: contradições, impasses e desafios socioespaciais*. Moderna.
2. Alencar, D. A., de Abreu Souza, M. I., Taques, F. H., & de Oliveira, J. D. C. (2013). Relação entre a desigualdade e educação no Brasil: uma tentativa de dados em painel (1995-2009). *Textos de Economia*, *16*(2), 111-142.
3. Araújo, T. P. de; Lima, R. A. (2018). Educação em Pernambuco: trajetória recente. In: ARAÚJO, T. B. de; ARAÚJO, T. P. de (orgs.). *Socioeconomia pernambucana: mudanças e desafios*. Recife: CEPE.
4. Borjas, G. J. (2012). *Economia do trabalho*. AMGH.
5. Botler, A., Marques, L., Nascimento, A., & Silva, K. (2014). Qualidade na educação municipal: um estudo comparativo entre o IDEB e o IDHM–Educação no nordeste brasileiro. *Revista de Administração Educacional*, *1*(1), 1-13.
6. Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R package for cluster validation. *Journal of Statistical Software*, *25*, 1-22.
7. Cangussu, R. C., Salvato, M. A., & Nakabashi, L. (2010). Uma análise do capital humano sobre o nível de renda dos estados brasileiros: MRW versus Mincer. *Estudos Econômicos (São Paulo)*, *40*, 153-183.
8. Desgraupes, B. (2013). *Clustering indices*. University of Paris Ouest-Lab Modal'X.
9. Fernandes, R. (2007). *Índice de desenvolvimento da educação básica (IDEB)* (Vol. 26). MEC; INEP.
10. IBGE - Intituto Brasilero de Geografia e Estatística. (2019). *Agência IBGE notícias*.
11. IBGE - Intituto Brasilero de Geografia e Estatística. (2019). *Pesquisa Nacional por Amostra de Domicílio*.
12. INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2019). *DADOS ABERTOS - INEP*. Ministério da Educação.
13. INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2015). *Resultados*. Ministério da Educação.
14. INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018). *Resultados*. Ministério da Educação.
15. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R* (Vol. 103). Springer.
16. Lopes, L. M., Vasconcellos, M. A. S. D., Gremaud, A. P., Luque, C. A., Martone, C. L., Braga, M. B., & Vasconcellos, M. S. D. (1997). *Manual de macroeconomia: nível básico e nível intermediário*.
17. Mankiw, N. G. (2013). *Introdução à economia*. Cengage Learning.

18. Martone, C. L. (2015). Extensões do Modelo de Crescimento. In: LOPES, L. M.; Vasconcellos, M. A. S. de (orgs.). *Manual de macroeconomia*. São Paulo: Atlas.

19. MEC - Ministério da Educação. (2019). *Fundeb - Apresentação*.

20. MEC - Ministério da Educação. (2019). *IDEB – Apresentação*.

21. Meier, G. M., & Rauch, E. J. (2005). *Leading issues in economic development*. Oxford University.

22. Mingoti, S. A. (2005). *Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada*. UFMG.

23. OECD - Organização de Cooperação e de Desenvolvimento Económico. (2015). *Educação em Revista*.

24. OECD - Organização de Cooperação e de Desenvolvimento Económico. (2015). *Programa Internacional de Avaliação de Estudantes (PISA)*.

25. Pindyck, R., & Rubinfeld, D. (2005). *Microeconomia*. Pearson.

26. Pires, V. (2005). *Economia da educação: para além do capital humano*. Cortez.

27. Racena, J. (2018). Pernambuco no início do século XXI. In: ARAÚJO, T. B. de.; ARAÚJO, T. P. de. (orgs.). *Socioeconomia pernambucana: mudanças e desafios*. Recife: CEPE.

28. Riani, F. (2009). *Economia do setor público: uma abordagem introdutória*. LTC.

29. Ribemboim, J. A. (2019). *Ensaios econômicos*. Babecco.

30. Sobel, T. F., Muniz, A. L. P., & de Farias Costa, E. (2009). Divisão regional do desenvolvimento humano em Pernambuco: Uma aplicação da análise de cluster. *Revista Teoria e Evidência Econômica*, *15*(33).

31. Sousa, U. D., Santos, J. S., Calado, J. E., Júnior, F. C., & SOUSA, W. D. (2018). Gastos públicos com educação e meta do IDEB no ensino fundamental: uma análise dos municípios do sertão de Pernambuco. *Brazilian Journal of Development*, *4*(5), 2578-2599.

32. STN - Secretaria do Tesouro Nacional. (2019). *Transferências constitucionais*.

33. Ward, J. H. J. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*, 236–244.