Clusterização e Visualização Espacial do IQE em Sergipe: Identificação de Padrões e Desigualdades no Âmbito do ICMS-Social

Iago Vieira Rocha[1,3], Juliana Sampaio Silva Ivo[1,3], João Paulo Lordelêlo Pedreira Vivas[1], Victor Flávio de Andrade Araujo[2,3], Andrea Karla Ferreira Nunes[1,3]

[1]Campus Aracaju - Universidade Tiradentes
Sergipe, Brazil.

{iago.vieira, juliana.ivo, joao.lordelo, andrea.karla}@souunit.com.br,

[2]Escola Politecnica - PUCRS
Porto Alegre, Brazil.

victor.araujo@edu.pucrs.br

[3]Grupo de Pesquisa, Docência, Avaliação,
Currículo e Contempraneidade (GPDACC)
Aracaju, Brazil.

Abstract. *Faced with quality and equity challenges in Brazilian education, the state of Sergipe established the ICMS-Social Program, which redistributes a portion of the Tax on the Circulation of Goods and Services (ICMS) based on educational indices, such as the Municipal Educational Quality Index (IQE). In this context, this study uses data clustering from the IQE calculation years, alongside choropleth maps, to identify similarities among municipalities. The clustering revealed three performance profiles, ranging from highest to lowest, which remained stable across the three years analyzed. Furthermore, their geographic distribution followed that of the IQE, reinforcing the consistency of the groups and helping identify municipalities requiring priority attention.*

Resumo. *Diante dos problemas de qualidade e equidade na educação brasileira, o estado de Sergipe criou o Programa ICMS-Social, que redistribui parte do Imposto sobre Circulação de Mercadorias e Serviços (ICMS) com base em índices educacionais, como o Índice Municipal de Qualidade Educacional (IQE). Nesse contexto, este trabalho utiliza a clusterização dos dados dos anos de apuração do IQE e mapas coropléticos para identificar similaridades entre os municípios. Os agrupamentos revelaram três perfis de desempenho, do melhor ao mais baixo, estáveis nos três anos analisados, e sua distribuição no mapa acompanhou a do próprio IQE, o que reforça a consistência dos grupos e ajuda a identificar os municípios que demandam atenção prioritária.*

1. Introdução

A Educação no Brasil enfrenta desafios significativos em termos de equidade e qualidade, que refletem no acesso díspar à educação e nos baixos índices de desempenho acadêmico, em especial em regiões estruturalmente e socialmente vulneráveis. Na edição de 2022 do

Programa Internacional de Avaliação de Alunos (PISA), o Brasil obteve uma média de 379 pontos em matemática, 410 em leitura e 403 em ciências, ficando abaixo da média da Organização para a Cooperação e Desenvolvimento Econômico (OCDE)[1]. Os resultados demonstram que a situação em que o Brasil se encontra é crítica, tornando essencial compreender os padrões de evolução educacional para embasar políticas públicas.

As reformas educacionais no Ceará têm sido amplamente estudadas e debatidas, particularmente no que diz respeito às políticas de incentivo financeiro baseadas no desempenho e às estratégias de apoio técnico para aprimorar a qualidade da educação. Dessa forma, observando o cenário do Ceará, onde essas políticas são aplicadas há décadas, e seus resultados [Oliveira et al. 2026], o Governo do Estado de Sergipe instituiu o Programa ICMS-Social[2], com a finalidade de, a partir do Imposto sobre Circulação de Mercadorias e Serviços (ICMS), criar uma redistribuição baseada nos indicadores educacionais, bem como o programa de apoio Sergipe na Idade Certa[3]. O primeiro é focado no modelo de transferência de recursos com base no desempenho educacional municipal, enquanto o segundo visa à assistência técnica e ao apoio pedagógico para fortalecer a gestão escolar e a aprendizagem dos alunos, melhorando os índices educacionais municipais.

Para promover a melhoria da qualidade da educação por meio dos recursos destinados aos municípios, o Governo de Sergipe promulgou o Decreto n.º 40.540, de 5 de março de 2020[4]. Esse decreto estabelece o critério de distribuição da parcela do ICMS, baseado no Índice Municipal de Qualidade da Educação do Município (IQE) o qual esse trabalho concentra-se e no Índice Municipal de Qualidade da Saúde (IQS). Em termos percentuais, a distribuição, que vigorará integralmente a partir de 2026[5], destina 18% ao IQE, 3% ao IQS e 4% à parcela igualitária, totalizando 25%.

Com vistas a acompanhar e avaliar os impactos dessa política, a análise de dados educacionais, estimulada pelo aumento exponencial do volume de informações escolares disponíveis, bem como pelo avanço das técnicas de visualização de dados, tem se mostrado relevante para compreender e solucionar problemas complexos, além de identificar correlações que auxiliam na tomada de decisões, principalmente no campo das políticas públicas, em que desafios sociais e econômicos requerem estratégias eficientes [Williamson 2016].

O objetivo geral deste trabalho é analisar e mapear o desempenho educacional dos municípios sergipanos, agrupando-os com o algoritmo *k-means* a partir das variáveis que compõem o IQE e representando os resultados em mapas coropléticos. Especificamente, busca-se organizar os dados públicos das apurações de 2023, 2024 e 2025, definir o número de grupos pelo método do cotovelo e pelo coeficiente de silhueta, caracterizar o perfil de cada agrupamento a partir de seus centroides e comparar espacialmente os grupos obtidos ao gradiente do próprio índice. Diante da ausência de literatura que avalie o impacto do Programa ICMS-Social na educação de Sergipe, foi desenvolvido um protótipo que oferece uma nova perspectiva para o monitoramento dessas políticas, permitindo que

[1]https://g1.globo.com/educacao/noticia/2023/12/05/ranking-da-educacao-brasil-esta-nas-ultimas-posicoes-no-pisa-2022-veja-notas-de-81-paises-em-matematica-ciencias-e-leitura.ghtml

[2]https://legislacao.se.gov.br/visualizar/lei-ordinaria/8628

[3]https://icmssocial.se.gov.br/regime-de-colaboracao-entre-estado-e-municipios/

[4]https://legislacao.se.gov.br/visualizar/decreto-numerado/40.540

[5]https://icmssocial.se.gov.br/o-que-e-o-icms-social/

gestores e pesquisadores explorem os dados de forma mais eficaz. Como contribuição, o trabalho oferece uma leitura dos 75 municípios sergipanos em perfis de desempenho e um instrumento visual que apoia a identificação dos municípios que demandam assistência técnica prioritária.

2. Trabalhos Relacionados

Estudos sobre a experiência do Ceará apontam um problema na qualidade da educação decorrente da distribuição não proporcional dos recursos federais investidos. O mesmo estudo destaca que reformas educacionais implementadas no estado contribuíram significativamente para a melhoria dos resultados educacionais dos municípios [Oliveira et al. 2026]. Entre essas reformas, destaca-se a adoção de mecanismos de incentivo que vinculam parte dos repasses financeiros aos municípios ao desempenho educacional de suas redes de ensino, medido por indicadores pré-estabelecidos, como o Índice de Qualidade Educacional (IQE), calculado com base nos resultados de avaliações aplicadas pela Secretaria da Educação do Estado aos estudantes do ensino fundamental.

Assim como [Oliveira et al. 2026], o estudo de [de Melo Silva et al. 2025] busca compreender fatores associados ao desempenho educacional dos municípios brasileiros. Contudo, sua abordagem é voltada à análise de dados educacionais, fornecidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), referentes aos índices das escolas públicas no ensino fundamental. Por meio da técnica de clusterização com o algoritmo *k-means*, o trabalho identifica grupos com características semelhantes, possibilitando a análise de estratégias eficientes para a melhoria da situação no sistema de educação brasileiro, a partir da revisão do repasse econômico entre municípios.

Nesse contexto, o estudo de [Catalino Piquet et al. 2024] apresenta uma análise da situação do governo do México, utilizando como base os seus indicadores educacionais. Diferentemente do método adotado por [de Melo Silva et al. 2025], o trabalho destaca a importância de encontrar uma solução focada em um melhor aproveitamento dos dados recolhidos. Argumenta-se que, embora exista atualmente uma enorme quantidade de dados, não há uma visualização adequada dessas informações; como solução, o estudo aponta para a criação de *dashboards* com o intuito de auxiliar na visualização e, consequentemente, na interpretação dos dados, no reconhecimento de possíveis falhas nas políticas educacionais atuais e na elaboração de estratégias para melhorá-las.

Em síntese, os trabalhos relacionados evidenciam diferentes abordagens para a análise e visualização de indicadores educacionais, variando quanto ao escopo geográfico, às fontes de dados, às técnicas empregadas e aos objetivos propostos. A Tabela 1 apresenta uma comparação entre esses estudos e destaca as principais diferenças em relação à proposta deste trabalho.

3. Metodologia

Esta seção tem como objetivo apresentar a base de dados utilizada neste trabalho e a metodologia adotada para sua análise. São descritas a origem dos dados referentes ao Índice de Qualidade Educacional (IQE) dos municípios sergipanos, suas principais características e as etapas de pré-processamento realizadas para viabilizar as análises. Em seguida, são detalhadas as técnicas utilizadas para a identificação de padrões de similaridade entre os

Tabela 1. Comparação entre trabalhos relacionados.

Trabalho	Escopo	Dados	Técnica	Foco
Oliveira et al. (2026)	Ceará	Indicadores de reforma educacional	Análise de eficiência / qualidade	Avaliação do modelo de incentivo financeiro
de Melo Silva et al. (2025)	Brasil (municípios)	INEP, ensino fundamental	k-means	Identificação de grupos por desempenho educacional
Catalino Piquet et al. (2024)	México	Indicadores educacionais governamentais	Dashboard	Visualização para apoio à decisão
Este trabalho	Sergipe (75 municípios)	IQE/SAESE 2023–2025	k-means + mapas coropléticos	Visualização espacial e temporal comparada ao gradiente do índice

municípios, com destaque para a aplicação do algoritmo de clusterização *k-means*. Posteriormente, são apresentados os métodos de visualização para representar os resultados e facilitar a interpretação dos indicadores educacionais.

3.1. Banco de Dados e Pré-processamento

Os resultados da apuração do Índice de Qualidade Educacional (IQE) referentes aos anos de 2023, 2024 e 2025 foram utilizados como base de dados neste estudo, uma vez que correspondem às únicas apurações disponibilizadas até o momento. Esses dados são produzidos pela Comissão Especial do Programa ICMS Social em conjunto com a Secretaria de Estado da Fazenda de Sergipe (Sefaz-SE) e encontram-se publicamente disponíveis no portal do programa[6], o que permite a replicação e a verificação independente deste estudo. Foram incluídos na análise os 75 municípios sergipanos, correspondendo à totalidade dos municípios com apuração publicada pelo programa nos três anos considerados, não havendo casos de exclusão por ausência de dados. O pré-processamento consistiu na reorganização e separação dos dados disponibilizados, alterando o nome das colunas para um padrão que facilitasse a análise com a biblioteca Pandas[7] e mantendo somente as colunas a serem utilizadas. Ao final, obtiveram-se quatro conjuntos de dados: um com os valores do IQE dos três anos de apuração e outros três, um por apuração, contendo os resultados das provas do Sistema de Avaliação da Educação Básica de Sergipe (SAESE) que fundamentam o cálculo do índice.

A formulação do IQE é feita a partir de 95% do Índice Municipal de Melhoria da Aprendizagem (IMA) e 5% do Índice de Aumento de Equidade (IAE). O IMA, por sua vez, é composto por 50% do Índice de Qualidade da Alfabetização (IQA), formulado pelo resultado dos educandos do 2º ano do Ensino Fundamental (EF) em Alfabetização na prova do SAESE; 45% do Índice de Qualidade do Ensino Fundamental (IQF), formulado pelos resultados no SAESE dos educandos do 5º ano do EF em Português e Matemática; e 5% da Média da Taxa de Aprovação dos educandos do 1º ao 5º ano do EF. Por fim, o IAE é calculado usando os resultados dos estudantes do 5º ano do EF no Indicador de Nível

[6]https://icmssocial.se.gov.br/apuracao-dos-resultados-em-2023/

[7]https://pandas.pydata.org/

Socioeconômico (INSE), o qual tem como base as respostas do questionário de contexto socioeconômico do SAESE.

Cada conjunto anual de dados do SAESE reúne 27 variáveis por município, utilizadas como entrada da clusterização. Para a apuração de 2023, que considera as edições de 2021 e 2022 da avaliação, as variáveis correspondem à Taxa de Aprovação dos educandos do 1º ao 5º ano do EF e, para cada edição das provas de Matemática (MAT) e Língua Portuguesa (LP), aplicadas ao 5º ano, e de Alfabetização (ALFB), aplicada ao 2º ano, à taxa de participação, à média de proficiência e aos percentuais de estudantes por padrão de desempenho: Muito Crítico e Adequado, em MAT e LP, e Não Alfabetizados, Alfabetização Incompleta e Alfabetização Desejável, em ALFB. Essas variáveis abrangem, portanto, indicadores de rendimento, participação e níveis de proficiência. As apurações de 2024 e 2025 seguem a mesma estrutura, alterando apenas os anos das edições consideradas, conforme estabelecido pela legislação do programa[8].

3.2. Processo de Visualização

Clusterização: A utilização da clusterização neste estudo foca em identificar padrões de similaridade entre os municípios sergipanos a partir dos indicadores educacionais utilizados na composição do IQE. Diferentemente das análises baseadas exclusivamente em *rankings*, o método de agrupamento permite identificar conjuntos de municípios que apresentam características semelhantes, mesmo quando ocupam posições distintas na classificação geral.

O agrupamento foi conduzido separadamente para cada ano de apuração, já que cada uma constitui um retrato independente. A escolha do algoritmo partiu de três requisitos impostos pelos dados e pelo objetivo do trabalho. O método precisa operar sobre variáveis numéricas e contínuas, natureza das 27 variáveis do SAESE, para as quais não há registros ausentes entre os 75 municípios. Precisa também atribuir um grupo a todos os municípios, pois o mapa coroplético não admite unidades sem classificação. Por fim, precisa fornecer um perfil médio por grupo, insumo da etapa seguinte, em que os agrupamentos são ordenados em melhor, intermediário e baixo desempenho. O *k-means* atende aos três, é um método particional baseado em distância euclidiana, classifica todas as observações e representa cada grupo por um centroide, que corresponde ao perfil médio dos municípios nele contidos. Soma-se a isso a comparabilidade metodológica com [de Melo Silva et al. 2025], que aplica o mesmo algoritmo a indicadores educacionais de municípios brasileiros.

Para determinar o número de *cluster* (k), combinaram-se o método do cotovelo e o coeficiente de silhueta. O primeiro analisa a inércia, soma das distâncias entre cada observação e o centroide de seu *cluster*, que decresce à medida que k aumenta o ponto de maior inflexão da curva, o ”cotovelo”, indica o número adequado de grupos [Thorndike 1953] [Marutho et al. 2018] [Cui et al. 2020]. O segundo varia entre -1 e 1 e mede simultaneamente a coesão interna e a separação entre grupos, sendo valores próximos de 1 indicativos de agrupamentos bem delimitados [Rousseeuw 1987] [Bholowalia and Kumar 2014]. Os dois são complementares o cotovelo considera apenas a compacidade interna dos grupos e depende da inspeção visual da curva, enquanto a

[8]https://api.legislacao.se.gov.br/uploads/atos/23942/DN-40540-2020-atualizado-DN-1181-2025.pdf

silhueta incorpora também a separação entre grupos e fornece um valor numérico. O uso conjunto reduz a arbitrariedade da leitura isolada de cada um.

O algoritmo *k-means* foi executado com os valores de k variando de 2 a 5, calculando-se a inércia em cada *cluster*. Posteriormente, o coeficiente de silhueta qualificou a distância entre os grupos, considerando-se como indicador ideal para o número de *clusters* o valor de k cujos coeficientes se mostraram mais equilibrados entre os *clusters* e com menor proporção de valores negativos.

Em sequência, a clusterização permite resumir um conjunto multidimensional de indicadores educacionais em perfis semelhantes, permitindo uma análise comparativa dos municípios mais facilitada e uma contribuição para a identificação de padrões relevantes para a formulação e avaliação de políticas públicas educacionais.

Análise dos perfis dos *clusters* e visualização: Após a delimitação de cada agrupamento, foram realizadas a análise e a categorização dos perfis de cada *cluster* de acordo com os valores médios das variáveis que formam os centroides. Para facilitar a interpretação, esses valores médios foram organizados segundo critérios crescentes ou decrescentes, conforme a natureza de cada variável, de modo a preservar a coerência interpretativa. Essa ordenação permitiu a comparação entre os *clusters* e a delimitação de perfis representativos, viabilizando a classificação global em grupos de melhor desempenho, desempenho intermediário e baixo desempenho educacional.

As técnicas empregadas de visualização de dados têm o objetivo de tornar os resultados mais acessíveis e evidenciar sua distribuição espacial, tornando a visualização uma ferramenta de suporte crucial [Lan et al. 2026]T. Para isso, a visualização foi estruturada por meio da biblioteca Plotly[9] na criação de mapas coropléticos dos municípios sergipanos. Nesses mapas, Figura 1, diferentes escalas de cores foram utilizadas para representar tanto os valores do IQE quanto os *clusters* atribuídos a cada município, permitindo uma comparação direta entre os mapas e assegurando a comparabilidade entre os agrupamentos de forma transparente e versátil como afirma [Sarigai et al. 2025].

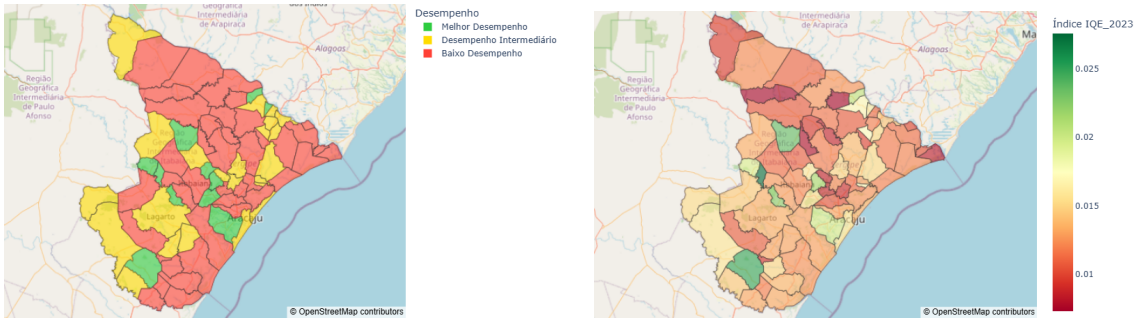
4. Resultados

O método do cotovelo indicou como quantidade ideal de *clusters* o valor 3 para os anos de 2023, 2024 e 2025. Em sequência, foi aplicado o método da silhueta para os diferentes anos, utilizando valores de k variando de 2 a 5, com a intenção de validar o k escolhido. Dessa forma, selecionou-se o valor de k que apresentou coeficientes de silhueta mais equilibrados entre os *clusters* e com menor proporção de valores negativos.

Para cada ano, realizou-se a análise dos perfis dos *clusters* a partir dos valores médios das variáveis presentes nos centroides, permitindo a caracterização estatística de cada agrupamento. Essa análise possibilitou a ordenação global dos *clusters*, classificando-os em melhor desempenho, desempenho intermediário e baixo desempenho, conforme apresentado na Figura 1a. Tal classificação garantiu coerência interpretativa, considerando simultaneamente indicadores de rendimento, participação e níveis de proficiência.

A comparação espacial entre o mapa da clusterização, Figura 1a, e o mapa do IQE, Figura 1b, é um dos pontos centrais deste estudo, funcionando como ferramenta de

[9]https://plotly.com/python/



(a) Mapa de Sergipe com os municípios clusterizados em relação a 2023.

(b) Mapa de Sergipe em relação ao valor do IQE 2023.

Figura 1. Comparação entre o mapa dos municípios clusterizados (a) e o mapa do gradiente do IQE (b), ambos referentes à apuração de 2023.

validação e diagnóstico. Enquanto o mapa do IQE distribui os municípios em um gradiente contínuo baseado no índice bruto, o mapa de clusterização sintetiza essa complexidade ao evidenciar similaridades espaciais e identificar padrões de desigualdade regional de forma intuitiva. A elevada correspondência visual entre os dois mapas indica que os agrupamentos obtidos pelo algoritmo *k-means* refletem de forma consistente os valores do índice, confirmando a capacidade do modelo de capturar a estrutura subjacente dos dados e validando o uso da clusterização sobre as variáveis que compõem o IQE.

Cabe destacar que a síntese do gradiente contínuo em três grupos envolve um compromisso ao agregar os municípios em perfis, perdem-se as variações individuais de magnitude do índice, de modo que municípios situados nos limites de cada grupo tornam-se menos distinguíveis entre si. Por essa razão, o mapa de clusterização não substitui o mapa do IQE, mas o complementa a leitura em três perfis facilita a identificação imediata dos conjuntos de municípios que demandam intervenção ou assistência técnica prioritária, enquanto o gradiente permanece disponível para a análise pormenorizada de cada caso. A proximidade entre as duas representações, evidenciada pelos mapas, reforça a validade do *k-means* como instrumento de apoio à decisão nesse contexto.

Resultados análogos foram observados para os anos de 2024 e 2025, mantendo-se a estabilidade do padrão de separação em três grupos bem definidos e permitindo monitorar a evolução das políticas públicas no estado.

5. Conclusão

Os resultados obtidos demonstram que a aplicação de técnicas de clusterização aos dados do IQE constitui uma nova ferramenta de análise, capaz de sintetizar múltiplos indicadores educacionais em perfis interpretáveis. A segmentação dos municípios em grupos de melhor, médio e baixo desempenho permitiu identificar padrões espaciais, desigualdades regionais e estabilidade temporal, oferecendo subsídios objetivos para o monitoramento e avaliação de políticas públicas educacionais.

Adicionalmente, a integração dos resultados em um *dashboard* para visualização potencializa a interpretação dos dados, proporcionando uma análise interativa, intuitiva e comparativa entre os anos, o que favorece a tomada de decisão baseada em evidências, a transparência dos resultados e o apoio ao planejamento estratégico das ações educacionais

no âmbito estadual, como afirma [Ladeira et al. 2022] em sua pesquisa que trata a ideia de *dashboards* ao apoio à gestão educacional.

Como limitações, destaca-se que a clusterização foi realizada sobre variáveis em escala original, sem padronização prévia, o que pode ter conferido peso desigual a indicadores de diferentes magnitudes. Ademais, a correspondência entre os mapas de cluster e de IQE foi analisada de forma predominantemente qualitativa, por meio da comparação visual dos mapas coropléticos. Tais limitações motivam diretamente a evolução metodológica descrita a seguir.

Como trabalhos futuros, encontra-se em desenvolvimento uma evolução metodológica do processo de clusterização, na qual as variáveis passam a ser padronizadas e submetidas à Análise de Componentes Principais (PCA), com retenção de componentes por variância acumulada, antes da aplicação do *k-means*. Nessa nova abordagem, a seleção do número de *clusters* combina o método do cotovelo e o coeficiente de silhueta em duas trilhas complementares, um k ótimo definido por ano e um k fixo entre os anos, de modo a garantir a comparabilidade temporal dos agrupamentos. Espera-se, com isso, reduzir o efeito das diferenças de escala entre as variáveis, aprofundar a análise da evolução dos municípios entre as apurações e fortalecer o monitoramento contínuo do Programa ICMS-Social.

6. Agradecimentos

Os autores gostariam de agradecer à FAPITEC/SE (Fundação de Apoio à Pesquisa e à Inovação Tecnológica do Estado de Sergipe) e ao CNPq (Conselho Nacional de Desenvolvimento e Tecnológica) pelo apoio financeiro concedido por meio do programa PIBIC (Programa Institucional de Bolsas de Iniciação Científica), Edital nº 05/2024.

Referências

Bholowalia, P. and Kumar, A. (2014). Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105(9).

Catalino Piquet, B. U., Hernandez Cabrera, H. A., Hernandez-Calderon, J. G., Montane-Jimenez, L. G., and Soto-Mendoza, V. (2024). Dashboard design for key performance indicators visualization of steam government initiatives: A case study. In *Proceedings of the XI Latin American Conference on Human Computer Interaction*, CLIHC '23, New York, NY, USA. Association for Computing Machinery.

Cui, M. et al. (2020). Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, 1(1):5–8.

de Melo Silva, M. L., Cruz, L., Magalhães, R., Pires, T., Macedo, J., and Andrade, R. M. C. (2025). Desigualdades educacionais no brasil: Uma análise por clusterização de indicadores educacionais e de desempenho escolar. In *Anais do XL Simpósio Brasileiro de Bancos de Dados*, pages 154–167, Porto Alegre, RS, Brasil. SBC.

Ladeira, P. P., de Lima, L. M., and Krohling, R. A. (2022). A visualization tool for data analysis on higher education dropout: a case study at ufes. *arXiv preprint arXiv:2201.12637*.

Lan, X., Xu, X., and Wang, Y. (2026). Data visualization for teenage teaching: A review of current applications, pedagogical strategies, and impact on learning. *IEEE Transactions on Visualization and Computer Graphics*.

Marutho, D., Handaka, S. H., Wijaya, E., et al. (2018). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In *2018 international seminar on application for technology of information and communication*, pages 533–538. IEEE.

Oliveira, G., Carneiro, D., Veloso, P., Shirasu, M., and Irffi, G. (2026). Efficiency and quality of management in the context of ceará's educational reform. *SciELO Preprints*.

Rousseeuw, P. (1987). Rousseeuw, p.j.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. comput. appl. math. 20, 53-65. *Journal of Computational and Applied Mathematics*, 20:53–65.

Sarigai, S., Yang, L., Slack, K., Fish, C., Buenemann, M., Wu, Q., Lin, Y., Cook, J. A., and Jacobs, D. (2025). dciwebmapper2: Enhancing the dciwebmapper framework toward integrated, interactive visualization of linked multi-type maps, charts, and spatial statistics and analysis. *arXiv preprint arXiv:2509.07897*.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4):267–276.

Williamson, B. (2016). Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy*, 31(2):123–141.