




Sumário Executivo

Novo Ensino Médio no Brasil: Clusterização de municípios

Disciplina: Laboratório de Estatística



Pesquisadores:

Bruno Gondim Toledo,
João Carlos Doretto Nascimento,
Luiza Carolina Martins Servo

Professores:

Ana Maria Nogales Vasconcelos,
Jhames Matos Sampaio

Mentores:

Valeria Maria Rodrigues Fachine,
Robson dos Santos

Sumário Executivo

Novo Ensino Médio no Brasil: Clusterização de municípios

1

Contextualização

No Brasil está sendo proposta uma reforma do ensino médio, ou como será chamado neste trabalho, o Novo Ensino Médio. Esta proposta considera três grandes frentes: o desenvolvimento do protagonismo dos estudantes e de seu projeto de vida, por meio da escolha orientada do que querem estudar; a valorização da aprendizagem, com a ampliação da carga horária de estudos; e a garantia de direitos de aprendizagem comuns a todos os jovens, com a definição do que é essencial nos currículos a partir da BNCC.

A partir da garantia de aprendizagens essenciais e comuns a todos os estudantes, referenciadas na BNCC, e da oferta de itinerários formativos organizados e estruturados pedagogicamente, o jovem brasileiro poderá escolher, entre diferentes percursos, a formação que mais se ajusta às suas aspirações e aptidões e ao seu projeto de vida.

 BASE NACIONAL COMUM CURRICULAR (BNCC) Para que as alterações curriculares do Ensino Médio tenham os efeitos positivos esperados, outras políticas e ações se fazem necessárias. Uma delas é a (re)elaboração dos currículos a partir da BNCC - essencial para colocar em prática a proposta de flexibilização curricular.	 A ESCOLHA POR ITINERÁRIOS FORMATIVOS Os currículos do Novo Ensino Médio terão uma parte referenciada na BNCC (formação geral básica) e os itinerários formativos, que oferecem caminhos distintos aos estudantes, ajustados às suas preferências e ao seu projeto de vida, cuja oferta considera as possibilidades de escolas e redes. É principalmente na escolha do itinerário, portanto, que se materializa o protagonismo juvenil.	 FORMAÇÃO TÉCNICA E PROFISSIONAL NO ENSINO MÉDIO REGULAR Os estudantes matriculados no Ensino Médio regular terão a possibilidade de cursar integralmente um itinerário técnico, fazer um curso técnico junto com cursos de Formação Inicial e Continuada (FIC), ou até mesmo um conjunto de FICs articuladas entre si. Existe ainda a oportunidade de os jovens percorrerem itinerários voltados para uma ou mais áreas do conhecimento complementados por cursos FIC.	 AMPLIAÇÃO E DISTRIBUIÇÃO DA CARGA HORÁRIA O Novo Ensino Médio amplia a carga das escolas de 2.400 horas para pelo menos 3.000 horas totais, garantindo até 1.800 horas para a formação geral básica, com os conhecimentos previstos na BNCC, e o restante da jornada para os itinerários formativos. As escolas têm até março de 2022 para se adaptar a essa mudança.
--	---	---	--

Fonte: <https://anec.org.br/wp-content/uploads/2021/04/Guia-de-implantacao-do-Novo-Ensino-Medio.pdf>

O objetivo neste trabalho foi agrupar os municípios brasileiros segundo seus indicadores relacionados a educação, a fim de fazer um estudo de caso sobre a possibilidade de implementação da reforma do ensino médio no Brasil.

Foi utilizado o método de clusterização K-means, o qual será explicado mais à frente. Ele tem como objetivo efetuar a classificação dos municípios em grupos, com estes grupos tendo os indicadores escolhidos de forma mais homogênea.

Os Indicadores escolhidos para a análise serão os:

- Índice de Desenvolvimento da Educação Básica - IDEB (Fonte: SAEB 2019/INEP)
- Índice de Vulnerabilidade Social – Capital Humano – IVS-CH (Fonte: Atlas de Vulnerabilidade Social/IPEA, com base na malha municipal do Censo Demográfico de 2010/IBGE)
- Formação superior dos docentes* (Fonte: Censo Escolar 2019/INEP)
- Quantidade de escolas no município (Fonte: Censo Escolar 2019/INEP)

Tais indicadores foram escolhidos considerando as sugestões e orientações dos mentores do projeto, tendo em vista que o indicador IDEB mede qualidade da educação, com o fluxo escolar e as médias de desempenho dos alunos nos municípios; o indicador sintético IVS-CH aborda os aspectos de inclusão social dos indivíduos em componentes de saúde e educação, como: mortalidade infantil; da presença, nos domicílios, de crianças e jovens que não frequentam a escola; da presença, nos

domicílios, de mães precoces, e de mães chefes de família, com baixa escolaridade e filhos menores; da ocorrência de baixa escolaridade entre os adultos do domicílio; e da presença de jovens que não trabalham e não estudam. O indicador que aborda formação dos professores reflete a capacidade dos docentes em assumir desafios do Novo Ensino Médio, e por fim a quantidade de escolas no município aborda a capacidade dos municípios em acolher os alunos nesta nova jornada.

*indicador construído de maneira independente para realização desse estudo

A ideia central da Análise de Cluster é a possibilidade de efetuar a classificação dos objetos em grupos, de forma que os objetos dentro do mesmo grupo sejam mais similares quanto possível e, de forma análoga, que os diversos grupos (clusters) sejam mais diferentes o possível em sua constituição (Kassambara 2017a).

O método de clusterização K-means classifica os objetos dentro de múltiplos grupos, de forma que a variação intra-cluster seja minimizada pela soma dos quadrados das distâncias Euclidianas entre os itens e seus centroides.

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

Nessa técnica, deve-se selecionar o número de clusters desejado para que sejam criadas as classificações que precisar ou, executar o comando que definirá o número ótimos de clusters para a amostra carregada. O método de Elbow foi utilizado para a determinação do número de clusters. O processo de seleção das variáveis, por padrão passa por (Kassambara 2017a):

- (a) Determinação do número de clusters;
- (b) Selecionar randomicamente objetos para determinar os valores centrais;
- (c) Assinar as observações pela distância Euclidiana em relação aos seus centróides;
- (d) Efetuar atualizações calculando a nova média dos valores dentro de seu cluster definido;
- (e) Minimizar a soma dos quadrados intra-cluster (o R utilizou 25 repetições dos passos d-e).

Iniciando a análise foi obtido o gráfico abaixo para definir a quantidade de clusters que serão utilizados.

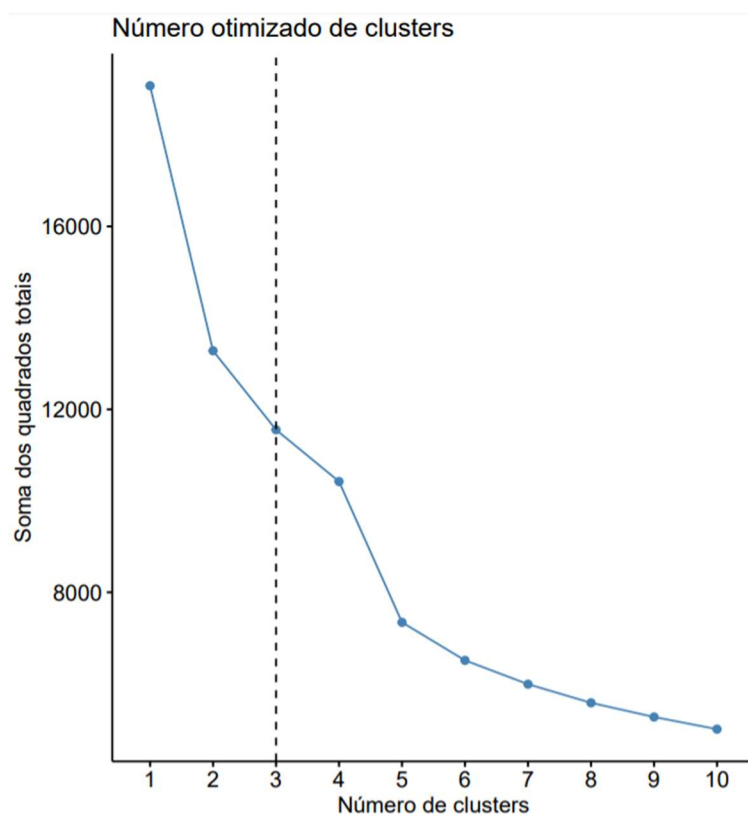


Figura 1

Na figura 1, utilizando o método de Elbow, é possível observar que optou-se pela utilização 3 clusters para o modelo, tendo em vista que a soma dos quadrados começa a diminuir, levando a pequenas diferenças na inclusão de mais um agrupamento. Porém, observando o banco de dados percebe-se a necessidade de imputar um quarto grupo, levando em consideração os municípios com ausência de valores (N/A) na nota do IDEB e nos municípios sem informação do IVS-CH, tendo em vista que no Censo de 2010 totalizava 5.565 municípios.

O quarto cluster será um cluster artificial e, por este motivo, ele não tem qualquer tipo de análise dos outros indicadores, simplesmente tem essa função de agregar os municípios. Estes municípios recebem a nota NA geralmente quando contestam a nota do IDEB, assim permanecendo NA até a resolução do processo na justiça.

Após a escolha da quantidade de clusters, serão definidos os significados de cada um. Abaixo estão os cluster definidos e algumas observações sobre eles, como as cores, as quais serão observadas na figura 2.

- **Cluster 1 (Azul)** – Número de municípios 1619.
 - Esse Cluster representa os municípios com melhor estrutura, com indicador do capital humano na faixa média de vulnerabilidade, alta formação específica na área que leciona dos professores, maior número de escolas de ensino médio no município e IDEB próximo a meta do PNE.
- **Cluster 2 (Amarelo)** – Número de municípios 1472.
 - Representa os municípios com estrutura aceitável, com indicador do capital humano na faixa média de vulnerabilidade, aceitável formação específica na área que leciona dos professores, médio número de escolas de ensino médio no município e IDEB na meta do PNE.
- **Cluster 3 (Rosa)** – Número de municípios 1679.
 - Representa os municípios com estrutura precária, com indicador do capital humano na faixa muito alta de vulnerabilidade, aceitável formação específica na área que leciona dos professores, baixo número de escolas de ensino médio no município e IDEB bem abaixo da meta do PNE.
- **Cluster 4 (Cinza)**- Número de municípios 793.
 - Sem informações em pelo menos um indicador, por isso não foi classificado no algoritmo.

Na figura 2 abaixo está a representação gráfica destes clusters:

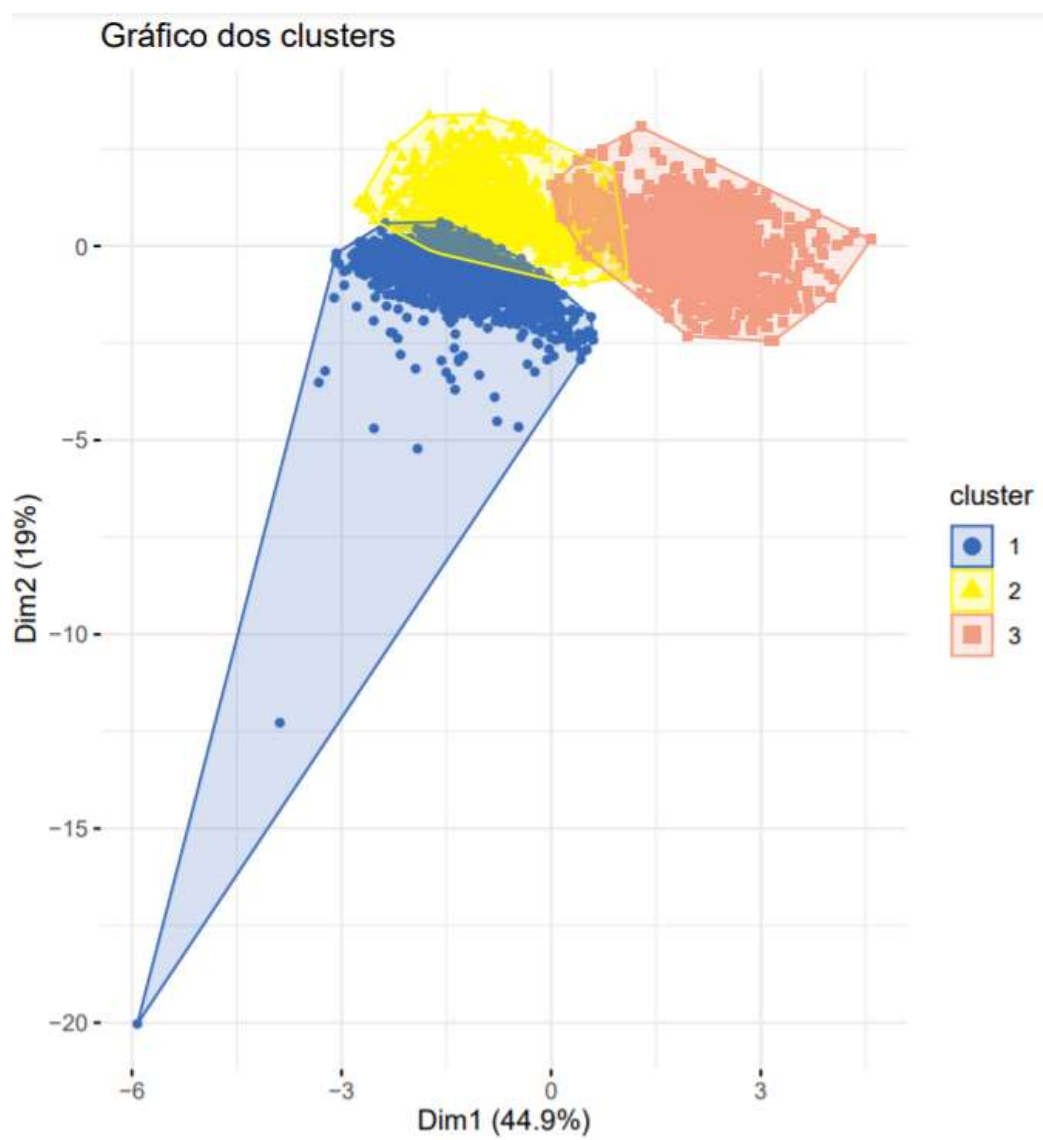


Figura 2

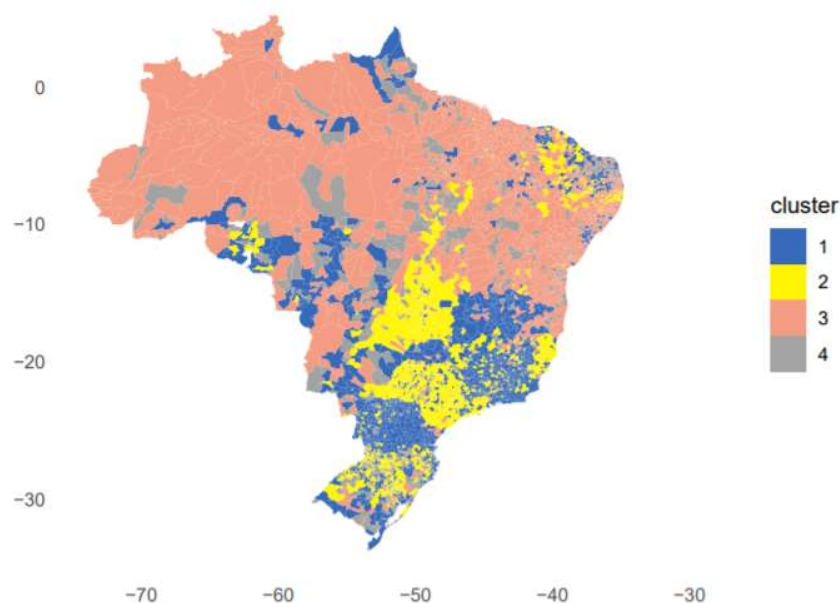


Figura 3

A partir do mapa é possível perceber uma diferença regional representada pelos clusters, as regiões Norte e Nordeste foram classificadas dentro do cluster 3, que conforme visto acima, indica uma estrutura mais precária e o indicador do capital humano em uma alta faixa de vulnerabilidade. Enquanto as regiões Sul, Sudeste e Centro-Oeste não tiveram grande presença do cluster 3, os cluster 1 e 2 prevaleceram nessas regiões, mostrando que existem indicativos de que há uma melhora na estrutura e no indicador do capital humano, como também o número de escolas e na formação específica lecionada pelos professores.

Mapa de Clusters da região Norte

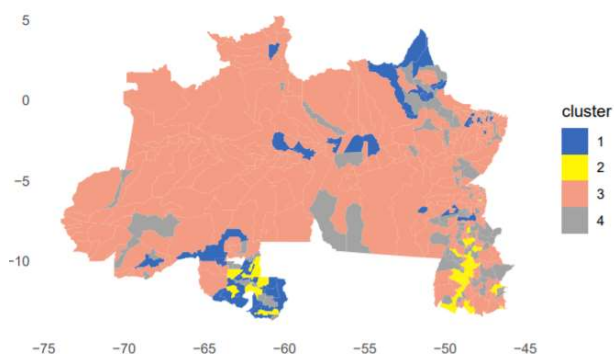


Figura 4

Analisando a figura 4 é notável a maior representatividade do cluster 3, além disso percebemos pontos espalhados do cluster 4 pela região. Os municípios dos estados do Amapá e de Rondônia foram classificados com os clusters 1 e 2, representando uma diferença em comparação com a sua região pois aparenta ter melhor estrutura e mais escolas, o que pode ser

um bom indicativo para a aplicação do Novo Ensino Médio, porém o mapa demonstra que a maior parte da região não possui a estrutura.

Mapa de Clusters da região Nordeste

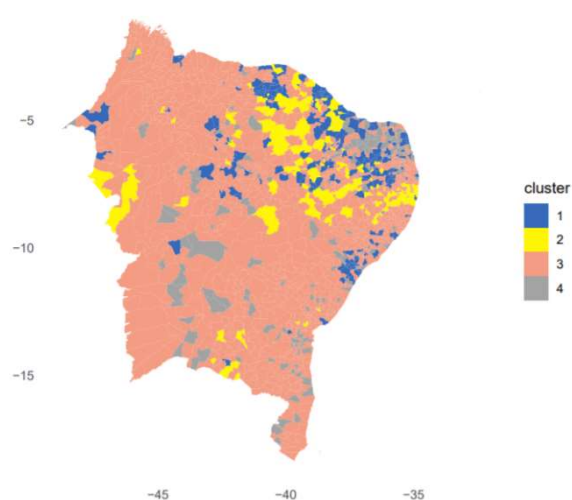


Figura 5

Na região Nordeste os clusters 3 e 4 foram prevalentes no interior dos estados, contudo é possível observar diversas ilhas de excelência ao nos aproximarmos das áreas litorâneas, onde começamos a observar a ocorrência dos clusters 1 e 2. Representando uma melhora em estrutura e possível implementação do novo ensino médio.

Mapa de Clusters da região Sul

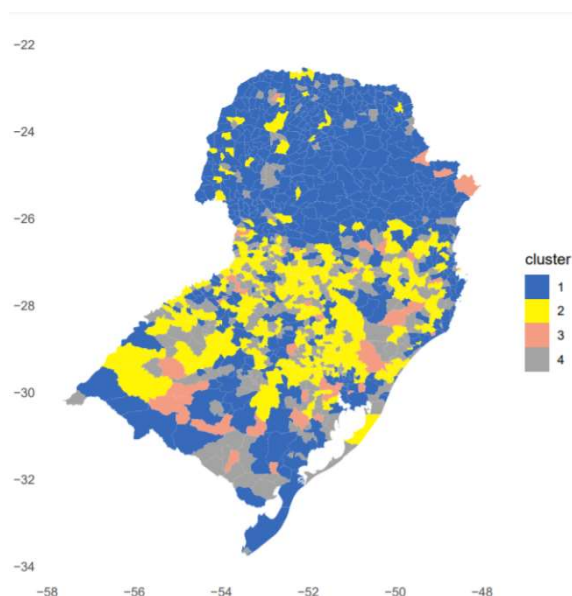


Figura 6

Ao contrário do analisado nas regiões anteriores, a região Sul apresenta um cenário mais favorável na clusterização. Aqui, a dominância é dos clusters 1 e 2 principalmente no estado do Paraná.

Os estados de Santa Catarina e Rio Grande do Sul apresentam maior diversidade na classificação, com diversos municípios inseridos no cluster 4, apresentando uma maior taxa de litígio judicial para a nota do IDEB nesses estados.

Mapa de Clusters da região Sudeste

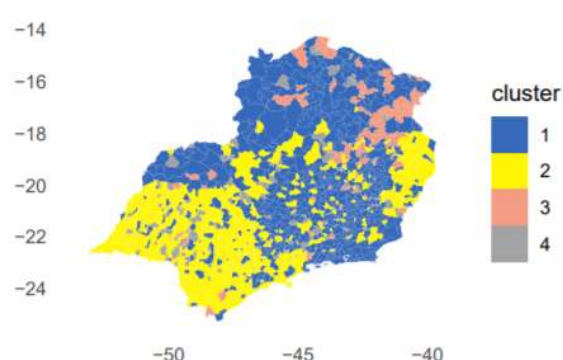


Figura 7

Assim como a região sul analisada, os clusters 1 e 2 aparecem como os mais influentes, nos estados de São Paulo e Espírito Santo o cluster 2 foi mais significativo no interior, o que representa uma democratização do conhecimento no que se refere a indicadores educacionais.

No estado do RJ o cluster 1 prevaleceu sobre todo o estado. No caso de Minas Gerais, observamos uma distribuição mais

heterogênea nos municípios. Enquanto na parte sudeste do estado a dominância é do cluster 1, no nordeste do estado observamos algumas ocorrências do cluster 3, o qual começa a seguir a tendência da região nordeste.

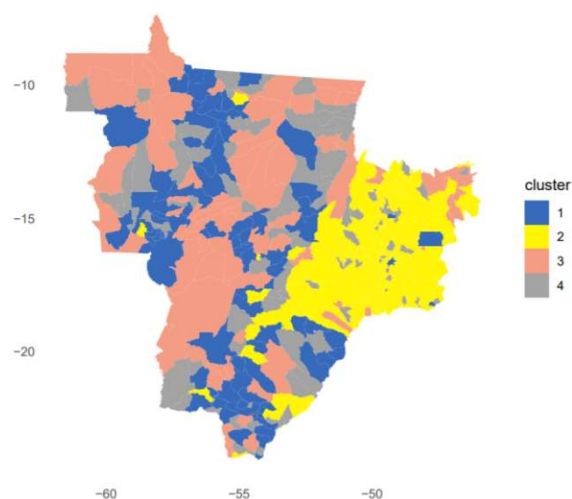
Mapa de Clusters da região Centro-Oeste

Figura 8

A região se mostra como uma região de transição, apresentando o perfil mais heterogêneo das regiões analisadas. Enquanto no estado de Goiás percebemos uma boa predominância do cluster 2, indicando uma boa democratização do ensino, o DF foi classificado dentro do cluster 1, demonstrando uma forte capacidade de aplicar o novo ensino médio.

Os estados de Mato Grosso e Mato Grosso do Sul apresentam uma grande diversificação de resultados nos municípios, o que pode representar uma grande amplitude nos indicadores dos municípios.

A partir da utilização do método K-Means foi possível alcançar o objetivo de pesquisa de agrupar os municípios brasileiros de acordo com os indicadores do IDEB, IVSCH, formação superior dos docentes e quantidade de escolas por município. Com base nisso, foi essencial a comparação entre os dados dos municípios para estabelecer os clusters e o método Elbow para definir a quantidade necessária para uma análise otimizada dos dados.

Ao analisar os resultados atingidos, percebe-se a prevalência de melhores indicadores nas regiões sul e sudeste, incluindo também algumas regiões do centro oeste e do litoral nordestino. A maior parte das capitais apresentaram um melhor desempenho nos indicadores, porém, ao se tratar do interior do norte, nordeste e centro oeste, nota-se a disparidade negativa dos indicadores.

Em suma, considerando o município como a unidade de estudo, os mapas de cluster quando analisados demonstraram que existe uma dificuldade na implementação de uma política estudantil nacional em grande parte dos municípios, portanto seria impraticável que esta fosse implementada em todo o Brasil, pelo fato dessa grande diversidade territorial e pelos resultados obtidos que indicam um despreparo para essa implementação em uma parcela significativa dos municípios e estados do país.