UNIVERSIdade FEEVALE

GUSTAVO WILLRICH

PLATAFORMA PARA VISUALIZAÇÃO DOS MICRODADOS DO ENEM

Novo Hamburgo

2022

GUSTAVO WILLRICH

PLATAFORMA PARA VISUALIZAÇÃO DOS MICRODADOS DO ENEM

Trabalho de Conclusão de Curso

apresentado como requisito parcial

à obtenção do grau de Bacharel em

Ciência da Computação pela

Universidade Feevale

Orientador: Dr. Juliano Varella de Carvalho

Novo Hamburgo

2022

Resumo

A avaliação de indicadores educacionais é um importante instrumento de gestão, permitindo à rede de ensino identificar situações que necessitem de mudanças, incentivos ou aprimoramentos. O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) realiza regularmente diversas avaliações, com o objetivo de monitorar a qualidade do ensino no Brasil, disponibilizando posteriormente os resultados para o público em geral na forma de microdados, permitindo assim a realização de estudos independentes. Uma dessas avaliações é o Exame Nacional do Ensino Médio (ENEM), que visa avaliar o desempenho dos estudantes das instituições de ensino. Embora os microdados do ENEM estejam disponíveis ao público, o grande volume de dados torna o processo de visualização e análise complexo, desestimulando a utilização desses dados no âmbito do setor educacional. Desta forma este trabalho consiste no desenvolvimento da plataforma web DadosENEM, voltada à análise e visualização destes microdados. O trabalho apresenta um resumo sobre o ENEM e sua importância, tanto como indicador da qualidade da educação no Brasil, como forma de acesso ao ensino superior. Foi analisada a evolução na estrutura dos microdados do ENEM a partir de 2009 e como as notas são compostas. Também foi realizada uma pesquisa bibliográfica sobre plataformas de visualização de dados do ENEM e, por fim, foi desenvolvida a plataforma web DadosENEM.

Palavras-chave: Visualização de Dados, Plataforma Web, ENEM, Microdados, INEP.

Abstract

The evaluation of educational indicators is an important management tool, allowing the educational system to identify situations that require changes, incentives or improvements. The Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) regularly conducts various assessments, with the aim of monitoring the quality of education in Brazil, subsequently making the results available to the general public in the form of microdata, thus allowing independent studies. One of these assessments is the Exame Nacional do Ensino Médio (ENEM), which aims to assess the performance of students from educational institutions. Although ENEM's microdata are publicly available, the large volume of data makes the process of visualization and analysis complex, discouraging the use of this data within the educational sector. This work consists in the development of the DadosENEM web platform, focused on the analysis and visualization of these microdata. The paper presents a summary about the ENEM and its importance both as an indicator of the quality of education in Brazil as a form of access to higher education. The evolution in the structure of ENEM's microdata from 2009 was analyzed and a bibliographical research on platforms for visualization of ENEM data was also made. Based on this, the DadosENEM web platform was developed.

Keywords: Data Visualization, Web Platform, ENEM, Microdata, INEP.

Lista de Figuras

[Figura 1 - Número de participantes inscritos por edição do ENEM – 1998 a 2020 14](file:///D:\p_vault\Feevale\TCC%202022\TCC2%20GustavoW%20Rev2_06-06-22.docx#_Toc105497762)

[**Figura 2 - Relação entre nota e número de acertos no modelo do TRI** 16](#_Toc105497763)

[Figura 3 - Fases de seleção dos estudos 26](#_Toc105497764)

[Figura 4 - Exemplo de visualização por tipo de escola 28](#_Toc105497765)

[Figura 5 - Exemplo de análise de notas a nível Brasil 29](#_Toc105497766)

[Figura 6 - Exemplo de análise de notas por renda 29](#_Toc105497767)

[Figura 7 - Página inicial da plataforma Resultados ENEM - ranking da escola selecionada 34](#_Toc105497768)

[Figura 8 - Plataforma Resultados ENEM - avaliação individual de cada questão 35](#_Toc105497769)

[Figura 9 - Evolucional - Análise de desempenho da uma escola específica 36](#_Toc105497770)

[Figura 10 - Qedu - Exemplo de análise do ENEM por município 37](#_Toc105497771)

[Figura 11 - Painéis ENEM - Sinopse Estatística 38](#_Toc105497772)

[Figura 12 - VisDadosENEM - Comparativo de regiões com gráficos de Bolha 41](#_Toc105497773)

[Figura 13 - Protótipo de tela inicial 42](#_Toc105497774)

[Figura 14 - Protótipo de tela de comparativos 43](#_Toc105497775)

[Figura 15 - Modelo ER da plataforma DadosENEM 44](#_Toc105497776)

[Figura 16 – Exemplo de código para renomar colunas com python e pandas 45](#_Toc105497777)

[Figura 17 – Exemplo de código para substituição de valores com python e pandas 45](#_Toc105497778)

[Figura 18 – Arquitetura da plataforma DadosENEM 48](#_Toc105497779)

[Figura 19 - Exemplo de código Plotly. 52](#_Toc105497780)

[Figura 20 - DadosENEM - Tela inicial. 54](#_Toc105497781)

[Figura 21 - DadosENEM - Comparativos. 56](#_Toc105497782)

[Figura 22 - DadosENEM - Comparativos por microrregião e gênero. 58](#_Toc105497783)

[Figura 23 - DadosENEM - Comparativos por região e renda. 59](#_Toc105497784)

[Figura 24 - DadosENEM - Comparativos por região e renda, apenas região Sul 60](#_Toc105497785)

Lista de Quadros

[Quadro 1 - Relação de campos alterados nos arquivos de microdados ao longo dos anos. 19](#_Toc107955188)

[Quadro 2 - Comparativo entre estudos x questões de qualidade. 33](#_Toc107955189)

[Quadro 3 - Lista de visões materializadas 50](#_Toc107955190)

[Quadro 4 - Visão materializada view\_media\_nota\_escola\_ano. 51](#_Toc107955191)

[Quadro 5 - Comparativo entre a plataforma DadosENEM e as ferramentas existentes 61](#_Toc107955192)

Lista de Abreviaturas e Siglas

|  |  |
| --- | --- |
| INEP | Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira |
| ENADE | Exame Nacional de Desempenho de Estudantes |
| ENEM | Exame Nacional do Ensino Médio |
| MEC | Ministério da Educação |
| IBGE | Instituto Brasileiro de Geografia e Estatística |
| SISU | Sistema de Seleção Unificada |
| ProUni | Programa Universidade para Todos |
| FIES | Financiamento Estudantil |
| TRI | Teoria de Resposta ao Item |
| PRONATEC | Programa Nacional de Acesso ao Ensino Técnico e Emprego |
| PNAE | Programa Nacional de Alimentação Escolar |
| SAS | Statistical Analysis System |
| SPSS | Statistical Package for the Social Sciences |
| API | Application Programming Interface |
| IES | Instituição de Ensino Superior |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Sumário

[1 Introdução 9](#_Toc107955213)

[2 ENEM 13](#_Toc107955214)

[2.1 OBJETIVOS 15](#_Toc107955215)

[2.2 TRI 15](#_Toc107955216)

[2.3 MICRODADOS 17](#_Toc107955217)

[3 TRABALHOS RELACIONADOS 25](#_Toc107955218)

[3.1 FASE DA REVISÃO DA LITERATURA 25](#_Toc107955219)

[3.2 DISCUSSÃO SOBRE OS ARTIGOS SELECIONADOS 27](#_Toc107955220)

[3.2.1 Um Estudo Voltado À Transparência Pública Através Da Análise De Dados Abertos Brasileiros - (MELO, 2017) 27](#_Toc107955221)

[3.2.2 VisDadosEnem: Visualização De Dados Do Enem Com A Biblioteca D3 - (SANTOS e FREITAS, 2016) 28](#_Toc107955222)

[3.2.3 Visualização dos resultados das edições de 2010 a 2015 do ENEM através de um Shiny App - (NASCIMENTO, 2017) 29](#_Toc107955223)

[3.2.4 Gestão dos resultados do ENADE com uso de dados abertos e tecnologias digitais - (ALVARES e CAMPOS, 2018) 31](#_Toc107955224)

[3.2.5 Análise dos microdados do ENADE, Proposta de uma ferramenta de exploração utilizando mineração de dados – (ARAÚJO, 2019) 31](#_Toc107955225)

[3.2.6 Comparativo entre estudos 33](#_Toc107955226)

[3.3 FERRAMENTAS EXISTENTES 35](#_Toc107955227)

[3.3.1 Resultados ENEM 35](#_Toc107955228)

[3.3.2 Evolucional 36](#_Toc107955229)

[3.3.3 QEdu 38](#_Toc107955230)

[3.3.4 Painéis ENEM 39](#_Toc107955231)

[4 DESENVOLVIMENTO DA PLATAFORMA 41](#_Toc107955232)

[4.1 Prototipação das visualizações 41](#_Toc107955233)

[4.2 Modelagem da plataforma 45](#_Toc107955234)

[4.3 Importação e padronização dos microdados 46](#_Toc107955235)

[4.4 Tecnologias utilizadas 48](#_Toc107955236)

[4.4.1 Banco de dados 49](#_Toc107955237)

[4.4.2 Visões materializadas 49](#_Toc107955238)

[4.4.3 Python 51](#_Toc107955239)

[4.4.4 Pandas 52](#_Toc107955240)

[4.4.5 Plotly/Dash 52](#_Toc107955241)

[4.5 Plataforma DadosENEM 54](#_Toc107955242)

[4.6 Comparação com as ferramentas existentes 61](#_Toc107955243)

[5 CONCLUSÃO 63](#_Toc107955244)

[Referências Bibliográficas 64](#_Toc107955245)

# Introdução

A comunicação de forma visual não é um assunto novo, as imagens têm sido usadas como meio de comunicação mesmo antes da formalização da linguagem escrita e continuam presentes no cotidiano das pessoas. Exemplos são mapa da previsão do tempo demonstrando o deslocamento de tempestades, placas de trânsito informando sobre uma curva acentuada na estrada ou mesmo em campanhas de marketing incentivando a aquisição de produtos e serviços. Um dos principais motivos para isso é devido a interpretação de informações em imagens ser realizada em paralelo com o sistema de percepção humano, sendo esse muito mais rápido quando comparado com a velocidade na interpretação de informações escritas, pois estas acabam sendo limitadas pelo processo sequencial de leitura (WARD; GRINSTEIN; KEIM, 2015).

No contexto de análise de dados, ao longo dos anos foram criados diversos métodos para representar determinados conjuntos de informações de forma gráfica, alguns métodos se destacam em popularidade e ainda dominam as salas de reuniões, tais como, gráficos de pizza, barras e linhas, mesmo sendo originários do século XVIII. Porém, com o avanço da capacidade tecnológica, tanto da coleta quanto do processamento de dados, atualmente é possível analisar e exibir informações de modo mais abrangente e dinâmico (KIRK, et. al., 2016).

Segundo Knaflic (2019), "A tecnologia permitiu acumular volumes de dados cada vez maiores e há também um crescente desejo associado de entendê-los." Ward, Grinstein e Keim (2015) complementam que em praticamente qualquer domínio, a visualização de dados está se tornando uma ferramenta eficaz para auxiliar a análise e comunicação.

Segundo Kirk et. al. (2016) a visualização de dados é um processo de descoberta, permitindo ao leitor não apenas ver os dados brutos, mas compreendê-los e convertê-los em informação útil. Mazza (2009) reforça que para dar sentido a esses dados, deve-se primeiro utilizar ferramentas e métodos adequados para organizar, processar e representar os dados considerando o contexto que se deseja analisar, assim transformando-os em informação.

Um dos ambientes que mais se beneficiam do grande volume de dados existentes atualmente é o ambiente corporativo. A adoção de técnicas como *Data Analytics* e *Business Intelligence* sobre os dados históricos de uma organização permite, por exemplo, identificar o perfil de compra dos clientes, os produtos mais vendidos ou se uma campanha de marketing obteve o resultado esperado, auxiliando no planejamento e tomada de decisões futuras (SHARDA; DELEN; TURBAN, 2019).

De acordo com Alvarez (2021), atualmente, as escolas também estão passando a adotar conceitos de *Data Analytics* e *Business Intelligence*, sobre os dados educacionais para auxiliar na tomada de decisões

No ambiente escolar, essas decisões podem ocorrer em todos os níveis, desde um professor que percebe a necessidade de fazer a revisão para um grupo específico de alunos que não compreenderam certo conteúdo, chegando a um gestor que decide mudar a estratégia de marketing para atrair novos alunos da vizinhança. (Alvarez, 2021).

Uma das fontes de dados educacionais no Brasil é o próprio INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), que disponibiliza para *download* diversos arquivos de microdados. Estes arquivos constituem no menor nível de desagregação de dados recolhidos por pesquisas, avaliações e exames, como Censo Escolar, ENADE, ENEM, entre outros (INEP, 2021a).

O Exame Nacional do Ensino Médio (ENEM) consiste em uma avaliação aplicada anualmente com o objetivo de mensurar a qualidade do ensino no país e o desempenho individual, tanto de estudantes que completaram o ensino médio quanto do público em geral que esteja interessado em avaliar seus próprios conhecimentos. Atualmente o exame é constituído de uma redação e 180 questões de múltipla escolha, divididas em quatro áreas: linguagens, códigos e suas tecnologias; ciências humanas e suas tecnologias; ciências da natureza e suas tecnologias; matemática e suas tecnologias. Há também um questionário socioeconômico que permite que sejam realizados estudos levando em consideração a realidade dos estudantes e sua família (INEP, 2021b).

Um dos principais objetivos da avaliação educacional é garantir a qualidade do ensino, porém para os estudantes, a nota obtida no ENEM também colabora para o acesso à educação superior por meio de programas de bolsa de estudos, financiamento estudantil, entre outros programas do MEC (INEP, 2021b).

Os microdados relacionados ao ENEM são constituídos de arquivos texto em formato “csv”, onde cada linha representa um participante do exame. Cada participante é identificado por meio de um número de inscrição, não sendo possível identificar a identidade do estudante. Para cada participante estão disponíveis diversos dados como respostas do questionário socioeconômico, respostas das questões objetivas, nota da redação, município onde foi realizada a prova, dentre outros (INEP, 2021a).

A análise dos microdados disponibilizados pelo INEP não é um processo trivial, Almeida Filho (2014) comenta que embora os microdados estejam divulgados ao público na internet, a grande quantidade de dados e a complexidade para analisá-los impede a ampla difusão do conhecimento no âmbito do setor educacional.

Travitzki (2013) descreve sua experiência sobre a utilização dos microdados do ENEM em seu projeto de pesquisa, “No início das pesquisas tivemos muita dificuldade em acessar os dados, pois não utilizamos nenhum dos programas proprietários (SAS e SPSS) para os quais são fornecidos, nos próprios microdados, arquivos de leitura”. Como alternativa, foi necessário modificar funções em linguagem R utilizadas na análise de dados do IBGE. Alves (2018) também comenta que durante sua pesquisa não foi possível visualizar os microdados do ENEM de 2016 diretamente com editores de texto e planilhas comuns como Bloco de Notas, Microsoft Excel e Libreoffice Calc, pois devido ao grande volume de dados acabavam ocorrendo erros nos softwares. Vale ressaltar que na edição do ENEM de 2016 os microdados possuem cerca de 8 milhões de linhas armazenadas em um único arquivo com tamanho de aproximadamente 5 Gigabytes (INEP, 2021a).

Atualmente já existem plataformas para análise dos microdados do ENEM, algumas desenvolvidas por empresas privadas, outras através de trabalhos de graduação. No entanto, foi identificado que essas plataformas tendem a ser direcionadas a ajudar as escolas a localizarem sua posição em um ranking geral, comparar sua posição com outras escolas da região e identificar em quais as áreas do ensino seus alunos sentiram mais dificuldades. Desta forma, existem lacunas que podem ser aprimoradas para melhor aproveitar o volume de informações que os microdados do ENEM têm a oferecer. Por exemplo, algumas ferramentas permitem analisar apenas uma edição do ENEM, outras disponibilizam poucas variáveis para análise.

Dada a dificuldade de analisar os dados do ENEM, este trabalho, portanto, descreve o processo de desenvolvimento da plataforma DadosENEM, uma plataforma Web para visualização dos microdados do ENEM, com o objetivo de contribuir para uma melhor compreensão sobre os fatores que influenciam a qualidade do ensino no Brasil, permitindo que escolas, pesquisadores e a gestão pública possam usufruir destas informações.

Este trabalho está dividido em 4 capítulos. O primeiro é esta introdução. O segundo apresenta uma contextualização sobre o ENEM, uma análise sobre a estrutura dos microdados e as alterações feitas ao longo dos anos. O terceiro capítulo descreve uma revisão da literatura onde são identificados estudos relacionados à visualização de dados educacionais e algumas ferramentas já existentes para visualização dos dados do ENEM e suas características. O quarto capítulo apresenta o processo de desenvolvimento da plataforma proposta.

# ENEM

O Exame Nacional do Ensino Médio (ENEM) consiste em uma avaliação aplicada anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) com o intuito de mensurar a qualidade do ensino básico no país e o desempenho individual de estudantes concluintes do ensino médio, mas também é aberto ao público em geral que esteja interessado em avaliar seus próprios conhecimentos, sendo estes classificados como “treineiros” (INEP, 2021b).

Criado em 1998, o ENEM passou por uma série de mudanças ao longo dos anos, sendo a última grande mudança em 2009, quando juntamente com a criação do Sistema de Seleção Unificada (SISU) o exame foi reformulado para atender as exigências necessárias para substituir os vestibulares tradicionais e se tornar o principal critério de seleção para o acesso às instituições públicas de ensino superior (INEP, 2021c). Desde então notou-se um aumento no número de participantes, chegando a 9.490.952 inscritos em 2014.

Porém, a partir de 2014 o número de inscritos vem caindo. Segundo o INEP (2019), um dos motivos pode estar relacionado à mudança da prova para dois domingos consecutivos, não mais apenas em um fim de semana e também por mudanças no processo de inscrição do ENEM, que a partir de 2017, adotou medidas para promover a inscrição consciente e evitar o desperdício da verba pública com inscritos que não comparecem para realizar a prova. Dentre estas medidas estão o aumento na taxa de inscrição e a adoção de controles mais rigorosos para a concessão da isenção da taxa de inscrição. Em 2018 o INEP também criou a obrigatoriedade de justificar a ausência para aqueles que buscavam novamente a gratuidade.

Em 2020 houve um aumento no número de inscritos em comparação aos dois anos anteriores, porém em função da pandemia da COVID-19 55,3% dos 5.783.357 inscritos não compareceram para realizar a prova, sendo este o maior número de abstenções na história do ENEM (OLIVEIRA, 2021).

A Figura 1 demonstra a evolução no número de participantes inscritos confirmados ao longo das edições.

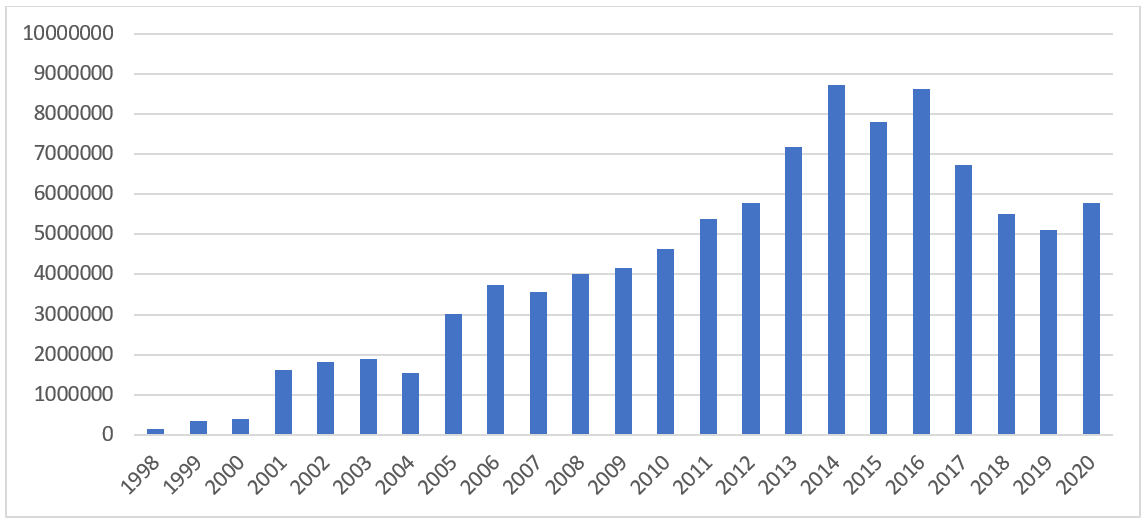


Figura 1 - Número de participantes inscritos por edição do ENEM – 1998 a 2020

Fonte: Elaborado pelo autor

Segundo o INEP (2021b) atualmente o exame é constituído de uma redação onde os participantes devem elaborar um texto dissertativo-argumentativo a partir de uma situação-problema e 180 questões de múltipla escolha, divididas em quatro áreas do conhecimento, que procuram abranger as disciplinas do ensino médio, são elas:

• Linguagens, códigos e suas tecnologias: engloba conteúdos como língua portuguesa (gramática e interpretação de texto), língua estrangeira (espanhol ou inglês), literatura e artes.

• Ciências humanas e suas tecnologias: engloba os conteúdos de geografia, história, filosofia, sociologia e conhecimentos gerais.

• Ciências da natureza e suas tecnologias: engloba os conteúdos de biologia, física e química.

• Matemática e suas tecnologias: engloba os conteúdos relacionados à matemática.

Também é aplicado um questionário socioeconômico que permite que sejam realizados estudos levando em consideração a realidade dos estudantes e sua família (INEP, 2021b).

Todas as questões são desenvolvidas com base em um documento chamado Matriz de Referência do ENEM, este documento aponta com detalhes quais os conhecimentos e habilidades esperadas dos estudantes que estão concluindo o ensino médio. As questões também procuram avaliar o eixo cognitivo do estudante, ou seja, a capacidade de utilizar o conhecimento que possuem para analisar e resolver situações adversas. No ENEM, cinco eixos são avaliados: dominar linguagens, compreender fenômenos, enfrentar situações-problema, construir argumentação e elaborar propostas. Desta forma, considera-se que o estudante esteja apto a avançar para níveis mais complexos de estudos e ingressar no mercado de trabalho (INEP, 2021d).

## OBJETIVOS

Para os estudantes concluintes do ensino médio, desde 2009, a nota do ENEM é essencial para o acesso ao ensino superior através de programas de apoio aos estudantes criados pelo MEC, como: o Sistema de Seleção Unificada (SISU), que permite concorrer a vagas para o ingresso em universidades públicas de forma gratuita; o Programa Universidade para Todos (ProUni), que permite concorrer a bolsas de estudo parciais ou integrais em instituições de ensino privadas; e o Programa de Financiamento Estudantil (FIES), que permite o financiamento da mensalidade em instituições de ensino privadas (INEP, 2021b).

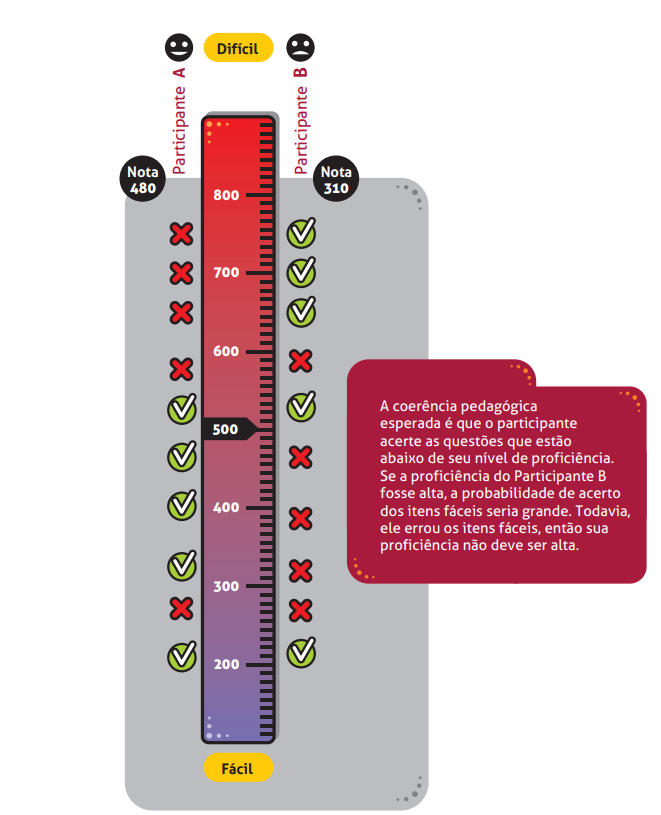
Além disso, por se tratar de um exame padronizado, com alto nível de segurança e confiabilidade e aplicado simultaneamente em todo o país, o ENEM é utilizado como referência nacional para direcionar a construção e revisão dos currículos do ensino médio. O ENEM também é referência para pesquisas independentes sobre a educação no Brasil, e isso é possível pois o INEP disponibiliza, de forma pública, os arquivos de microdados, contendo informações detalhadas sobre cada edição do ENEM.

## TRI

A partir de 2009 o ENEM passou a utilizar o método de correção baseado na Teoria da Resposta ao Item (TRI), que diferentemente do método clássico de correção onde a nota está diretamente ligada ao número de acertos e erros, o TRI busca levar em consideração a coerência entre acertos e erros conforme o grau de dificuldade das questões. O TRI baseia-se na premissa que os participantes tendem a acertar questões fáceis e médias e errar questões difíceis, pois entende-se que a aquisição de conhecimento ocorre de forma cumulativa, onde habilidades mais complexas dependem do domínio de habilidades mais simples. Quando um participante erra questões dadas como fáceis e acerta questões difíceis, pressupõe-se que o participante não domine essa área do conhecimento por completo e tenha acertado a questão por acaso (INEP, 2012).

Desta forma, dois participantes podem obter notas diferentes acertando o mesmo número de questões, como pode ser observado na Figura 2. Nela o Participante A obteve a nota 480 por ter acertado questões fáceis e errado as difíceis; já o Participante B obteve a nota 310 por ter acertado questões difíceis e errado questões fáceis.

**Figura 2 - Relação entre nota e número de acertos no modelo do TRI**



Fonte: INEP (2012)

A utilização do TRI também permite a comparação dos resultados entre diferentes edições do ENEM em uma mesma escala, mesmo com questões e participantes diferentes. Isso é possível pois todas as questões são construídas a partir da mesma matriz de competências. Desta forma, em uma edição onde a prova tenha questões muito fáceis, a nota mínima e máxima também será mais baixa, já em uma prova com questões mais difíceis as notas serão mais altas. O INEP (2012) descreve que, “em outras palavras, o mínimo e o máximo em cada prova dependem somente das questões que compõem a prova e não de quem as responde”. Sendo assim, a nota de um participante do ENEM nunca será 0 ou 1000 nas questões objetivas (isso é possível na redação pois se utilizam outros critérios de avaliação) devido às notas serem calculadas de acordo com o desempenho dos demais participantes e a avaliação dos profissionais que analisam as dificuldades das questões.

## MICRODADOS

Segundo Winkelmann (2006), o termo microdados está associado a diversos tipos de dados e aplicações. No entanto, a definição mais comum é um conjunto de dados geralmente armazenados em forma de tabela com o mínimo de tratamento ou agregação, ou seja, cada linha representa um indivíduo do estudo, podendo ser pessoas, empresas, produtos, etc, e cada coluna representa as características desse indivíduo relacionadas ao estudo. Essas informações podem ter sido coletadas propositadamente em avaliações e pesquisas ou podem ser provenientes de outras fontes, como registros automáticos gerados por sistemas de gestão. Desta forma, os microdados podem ser utilizados como base para estudos posteriores, permitindo obter informações relevantes ao interagir e correlacionar os indivíduos, analisando características em comum.

O INEP disponibiliza publicamente os microdados relacionados a diversas avaliações e pesquisas relacionadas à educação no Brasil. Os microdados disponibilizados pelo INEP representam o menor nível de desagregação dos dados obtidos de avaliações, exames e pesquisas como o Censo Escolar, ENEM, ENADE e diversos outros. Os microdados do INEP são disponibilizados em arquivo-texto, no formato “csv”, juntamente com um dicionário de dados e os questionários ou avaliações que foram aplicadas (INEP, 202la).

Os microdados do ENEM são utilizados de diversas formas, seja por pesquisadores para embasar pesquisas acadêmicas como: a pesquisa “Renda familiar e escolaridade dos pais: reflexões a partir dos microdados do ENEM 2012 do Estado de São Paulo” (PIRES, 2015), realizada pelo Professor André Pires, da PUC Campinas; por escolas, que ao analisar o desempenho de seus alunos pode auxiliar nas decisões pedagógicas; para o jornalismo, como a pesquisa “Os eliminados na redação do Enem não são quem você imagina” (SOARES, 2017) escrita pelo jornalista Marcelo Soares que aponta o perfil dos participantes desclassificados da redação do ENEM em 2015 por desrespeito aos direitos humanos. Além da gestão pública que utiliza diversas fontes de dados, incluindo o ENEM, para decidir quais áreas necessitam de mais ou menos investimentos.

Os arquivos de microdados estão disponíveis na página do INEP[[1]](#footnote-2), esses arquivos estão estruturados de forma semelhante, cada registro representa um participante, e cada coluna representa uma atributo relacionado a este participante, como escola, município de aplicação da prova, resposta do questionário socioeconômico, gabarito com a resposta em cada questão, nota, entre outros, porém, não é possível identificar a identidade do participante, pois seu número de identificação é gerado no momento da inscrição e não é relacionado às edições anteriores ou a qualquer outro registro. (INEP, 2021a)

Os microdados de 2019, por exemplo, são compostos por 136 atributos divididos nas seguintes seções:

* **Dados do participante**, informações como número de inscrição, município de residência, estado civil, idade, etc.
* **Dados da escola**, informações como código da escola, município, dependência administrativa (estadual, federal, municipal).
* **Dados dos pedidos de atendimento especializado**, informativo de baixa visão, cegueira, surdez, dislexia, etc.
* **Dados dos pedidos de atendimento específico**, indicador sobre inscritos idosos, gestantes, lactantes ou estudantes de classe hospitalar.
* **Dados dos pedidos de recursos especializados e específicos para realização das provas**, indicador de assistência para a realização da prova como, braile, ledor, cadeira especial, etc.
* **Dados do local de aplicação da prova,** informações sobre o município onde será realizado.
* **Dados da prova objetiva**, informações como código da prova que cada inscrito realizou, vetor com a gabarito da prova, vetor com as respostas do inscrito, indicador de presença, etc.
* **Dados da redação**, contém a situação sobre a redação do participante e a nota em cada uma das 5 competências.
* **Dados do questionário socioeconômico**, 25 questões que buscam compreender a realidade do participante e sua família, como grau de instrução dos pais, renda familiar, entre outros.

Vale ressaltar que o ENEM está em constante evolução, por este motivo a estrutura dos arquivos de microdados passou por mudanças ao longo dos anos. A última grande mudança ocorreu em 2009 quando o ENEM foi reformulado, desde então, pequenas mudanças vêm sendo feitas a cada nova edição para se ajustar às novas necessidades.

Durante a elaboração deste trabalho foi realizado a identificação de todas as mudanças na estrutura dos arquivos de microdados entre 2009 e 2020, isso foi necessário para auxiliar na criação do *script* de importação dos dados, a fim de se ter um banco de dados consistente e poder realizar comparações entre diversas edições do ENEM.

Para identificação das mudanças nos microdados foi elaborado um *software* em Delphi (escolhido apenas devido à familiaridade do autor com o mesmo). Este *software* recebe dois arquivos de dicionários de dados do ENEM no formato “.xls” e compara as colunas “nome da variável”, “descrição”, “variáveis categóricas”, “tamanho” e “tipo”. Ao final do processamento é retornado uma lista com as diferenças encontradas. Por se tratar de um *software* específico para resolver um problema pontual, não foi dado atenção à interface e à usabilidade. Foi feito uma conferência manual nas diferenças encontradas para remover falsos positivos como campos descritivos com quebra de linha no arquivo A e sem quebra de linha no arquivo B ou palavras com acentuação no arquivo A e sem acentuação no arquivo B. O Quadro 1 apresenta a relação com as mudanças na estrutura dos microdados do ENEM ao longo dos anos.

Quadro 1 - Relação de campos alterados nos arquivos de microdados ao longo dos anos.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variável** | **09** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** |
| CO\_MUNICIPIO\_RESIDENCIA |  |  | R |  | R |  | R |  |  |  |  | E |
| NO\_MUNICIPIO\_RESIDENCIA |  |  | R |  | R |  |  |  |  |  |  | E |
| SG\_UF\_RESIDENCIA |  |  | R |  | R |  | R |  |  |  |  | E |
| CO\_UF\_RESIDENCIA |  |  | E |  | IR |  | R |  |  |  |  | E |
| NU\_IDADE |  |  | R |  |  |  | R |  |  |  |  | E |
| TP\_SEXO |  |  | A |  | A |  |  |  |  |  |  |  |
| TP\_ESTADO\_CIVIL |  | I | A |  |  |  |  |  |  |  | A |  |
| TP\_ST\_CONCLUSAO |  |  | RA | A | A | A | RA | A | A | A | A | A |
| TP\_ENSINO |  |  | R |  | A |  | RA |  |  |  |  |  |
| CO\_ESCOLA |  |  | R |  | R |  | R |  |  |  |  | E |
| CO\_MUNICIPIO\_ESC |  |  | R |  |  |  | R |  |  |  |  |  |
| CO\_UF\_ESC |  |  | E |  | IR |  | R |  |  |  |  |  |
| SG\_UF\_ESC |  |  | R |  |  |  | R |  |  |  |  |  |
| TP\_DEPENDENCIA\_ADM\_ESC |  |  | R |  | R |  | R |  |  |  |  |  |
| TP\_LOCALIZACAO\_ESC |  |  | R |  | R |  | R |  |  |  |  |  |
| TP\_SIT\_FUNC\_ESC |  |  | R |  | R |  | R |  | A |  |  |  |
| CO\_MUNICIPIO\_PROVA |  |  | R |  |  |  | R |  |  |  |  |  |
| CO\_UF\_PROVA |  |  | E |  | IR |  | R |  |  |  |  |  |
| SG\_UF\_PROVA |  |  | R |  | R |  | R |  |  |  |  |  |
| TP\_PRESENCA\_CN |  |  | R |  |  |  | R |  |  |  |  |  |
| TP\_PRESENCA\_CH |  |  | R |  |  |  | R |  |  |  |  |  |
| TP\_PRESENCA\_LC |  |  | R |  |  |  | R |  |  |  |  |  |
| TP\_PRESENCA\_MT |  |  | R |  |  |  | R |  |  |  |  |  |
| CO\_PROVA\_CN |  |  | R |  |  |  | R |  |  |  |  |  |
| CO\_PROVA\_CH |  |  | R |  |  |  | R |  |  |  |  |  |
| CO\_PROVA\_LC |  |  | R |  |  |  | R |  |  |  |  |  |
| CO\_PROVA\_MT |  |  | R |  |  |  | R |  |  |  |  |  |
| NU\_NOTA\_CN |  |  | R |  | R |  | R |  |  |  |  |  |
| NU\_NOTA\_CH |  |  | R |  | R |  | R |  |  |  |  |  |
| NU\_NOTA\_LC |  |  | R |  | R |  | R |  |  |  |  |  |
| NU\_NOTA\_MT |  |  | R |  | R |  | R |  |  |  |  |  |
| TX\_GABARITO\_CN |  |  | R |  | R |  | R |  |  |  |  |  |
| TX\_GABARITO\_CH |  |  | R |  | R |  | R |  |  |  |  |  |
| TX\_GABARITO\_LC |  |  | R |  | R |  | R |  |  |  |  |  |
| TX\_GABARITO\_MT |  |  | R |  | R |  | R |  |  |  |  |  |
| TP\_STATUS\_REDACAO |  |  | RA | A | A |  | R | A | A |  |  |  |
| IN\_CONCLUINTE\_CENSO |  |  | I | E |  |  |  |  |  |  |  |  |
| COD\_ETAPA\_ENSINO\_CENSO |  |  | I | E |  |  |  |  |  |  |  |  |
| COD\_ENTIDADE\_CENSO |  |  | I | E |  |  |  |  |  |  |  |  |
| COD\_MUNICIPIO\_ESC\_CENSO |  |  | I | E |  |  |  |  |  |  |  |  |
| NO\_MUNICIPIO\_ESC\_CENSO |  |  | I | E |  |  |  |  |  |  |  |  |
| UF\_ESC\_CENSO |  |  | I | E |  |  |  |  |  |  |  |  |
| ID\_DEPENDENCIA\_ADM\_CENSO |  |  | I | E |  |  |  |  |  |  |  |  |
| ID\_LOCALIZACAO\_CENSO |  |  | I | E |  |  |  |  |  |  |  |  |
| SIT\_FUNC\_CENSO |  |  | I | E |  |  |  |  |  |  |  |  |
| IN\_UNIDADE\_PRISIONAL |  |  |  | E |  |  |  |  |  |  |  |  |
| TP\_ESCOLA |  |  |  | I |  |  | A |  |  |  |  |  |
| ANO\_CONCLUIU |  |  |  | I |  |  | RA | A | A | A | A | A |
| IN\_UNIDADE\_HOSPITALAR |  |  |  | I | R |  |  |  |  |  |  | E |
| IN\_SURDO\_CEGUEIRA |  |  |  | I |  |  |  |  |  |  |  | E |
| IN\_IDOSO |  |  |  | I |  |  |  |  |  |  |  | E |
| IN\_AUTISMO |  |  |  | I |  |  |  |  |  |  |  | E |
| IN\_MESA\_CADEIRA\_RODAS |  |  |  | I |  |  |  |  |  |  |  | E |
| IN\_MESA\_CADEIRA\_SEPARADA |  |  |  | I |  |  |  |  |  |  |  | E |
| IN\_APOIO\_PERNA |  |  |  | I |  |  |  |  |  |  |  | E |
| IN\_GUIA\_INTERPRETE |  |  |  | I |  |  |  |  |  |  |  | E |
| NO\_ENTIDADE\_CERTIFICACAO |  |  |  | I |  |  |  |  | E |  |  |  |
| UF\_ENTIDADE\_CERTIFICACAO |  |  |  | I |  |  | R |  | E |  |  |  |
| COD\_UF\_ENTIDADE\_CERTIFICACAO |  |  |  |  | I |  | R |  | E |  |  |  |
| IN\_AMPLIADA |  |  |  |  | R |  |  |  |  |  |  | E |
| IN\_AMPLIADA\_24 |  |  |  |  | I |  |  |  |  |  |  | E |
| IN\_AMPLIADA\_18 |  |  |  |  | I |  |  |  |  |  |  | E |
| NACIONALIDADE |  |  |  |  | I |  | RA |  |  |  |  |  |
| COD\_MUNICIPIO\_NASCIMENTO |  |  |  |  | I |  | R |  |  |  |  | E |
| NO\_MUNICIPIO\_NASCIMENTO |  |  |  |  | I |  |  |  |  |  |  | E |
| COD\_UF\_NASCIMENTO |  |  |  |  | I |  | R |  |  |  |  | E |
| UF\_NASCIMENTO |  |  |  |  | I |  | R |  |  |  |  | E |
| IN\_TREINEIRO |  |  |  |  |  |  | I |  |  |  |  |  |
| TP\_COR\_RACA |  |  |  |  |  |  | A |  | A |  |  |  |
| IN\_DISCALCULIA |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_VISAO\_MONOCULAR |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_OUTRA\_DEF |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_SEM\_RECURSO |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_NOME\_SOCIAL |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_MACA |  |  |  |  |  |  | I |  | E |  |  |  |
| IN\_COMPUTADOR |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_CADEIRA\_ESPECIAL |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_CADEIRA\_CANHOTO |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_CADEIRA\_ACOLCHOADA |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_PROVA\_DEITADO |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_MOBILIARIO\_OBESO |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_LAMINA\_OVERLAY |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_PROTETOR\_AURICULAR |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_MEDIDOR\_GLICOSE |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_MAQUINA\_BRAILE |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_SOROBAN |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_MARCA\_PASSO |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_SONDA |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_MEDICAMENTOS |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_SALA\_INDIVIDUAL |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_SALA\_ESPECIAL |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_SALA\_ACOMPANHANTE |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_MOBILIARIO\_ESPECIFICO |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_MATERIAL\_ESPECIFICO |  |  |  |  |  |  | I |  |  |  |  | E |
| IN\_SABATISTA |  | I |  |  |  |  |  |  | E |  |  |  |
| IN\_CERTIFICADO |  | I |  |  |  |  |  |  | E |  |  |  |
| IN\_TEMPO\_ADICIONAL |  |  |  |  |  |  |  |  |  |  | I | E |
| IN\_BAIXA\_VISAO |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_CEGUEIRA |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_SURDEZ |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_DEFICIENCIA\_AUDITIVA |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_DEFICIENCIA\_FISICA |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_DEFICIENCIA\_MENTAL |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_DEFICIT\_ATENCAO |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_DISLEXIA |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_GESTANTE |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_LACTANTE |  | I |  |  |  |  |  |  |  |  |  | E |
| IN\_LEITURA\_LABIAL |  | I |  |  |  |  |  |  |  |  |  | E |
| TP\_FAIXA\_ETARIA |  |  |  |  |  |  |  |  |  |  |  | I |

Legenda: R = Renomeado / I = Incluído / A= Opções alteradas / E = Excluído

Fonte: Elaborado pelo autor a partir dos microdados do ENEM

Percebe-se a grande quantidade de campos excluídos nos microdados de 2020, isso ocorreu devido à Lei Geral de Proteção de Dados (LGPD) que estabelece princípios norteadores da coleta, do compartilhamento e do tratamento de dados pessoais, além de um conjunto de obrigações aos controladores e responsáveis. Mesmo nunca tendo disponibilizado nos microdados do ENEM quaisquer variáveis que permitissem a identificação direta do participante, o INEP optou pela cautela e decidiu adotar um modelo simplificado de microdados, a ser utilizado para a edição 2020 do ENEM e posteriores. O objetivo é eliminar da base pública variáveis que facilitem a identificação indevida do participante (INEP, 2022).

Devido às mudanças na estrutura dos microdados em cada edição do ENEM, faz-se necessário realizar um pré-processamento dos microdados, a fim de organizá-los de forma que seja possível realizar a comparação entre diferentes edições. Isso torna o processo de análise dos microdados custoso, desincentivando o estudo entre as edições do ENEM ao longo dos anos.

Este capítulo apresentou um resumo sobre o ENEM e seu histórico, com o objetivo de contextualizar a relevância dessa avaliação para o monitoramento e melhoria da educação no Brasil. Também foram apresentados como os dados coletados pela instituição podem ser utilizados, tanto como apoio na tomada de decisões educacionais, quanto para entender como a sociedade em geral está estruturada através de informações retiradas dos questionários socioeconômicos. Além disso, foi realizado o mapeamento das mudanças na estrutura dos microdados a partir de 2009, demonstrando que os microdados disponibilizados pelo INEP requerem um esforço adicional por parte dos pesquisadores ao analisar múltiplas edições. Vale ressaltar que não foram considerados os microdados anteriores a 2009 devido a reformulação do ENEM, já que até 2008 era formado por apenas 63 questões objetivas, inviabilizando assim a comparação e a consistência das análises.

# TRABALHOS RELACIONADOS

Este capítulo descreve algumas das pesquisas e ferramentas existentes relacionadas à análise e visualização dos microdados do ENEM, tendo como objetivo identificar como os microdados estão sendo utilizados, quais análises estão sendo feitas e pontuar possíveis limitações e/ou melhorias que possam ser aplicadas à ferramenta que este trabalho se propõe a desenvolver. Também foram analisadas algumas ferramentas comerciais existentes, as quais não possuem uma pesquisa acadêmica relacionada.

## FASE DA REVISÃO DA LITERATURA

Esta revisão da literatura tem como objetivo identificar artigos, dissertações e teses relacionadas às ferramentas para visualização e análise dos dados de avaliações escolares de larga escala. Durante a elaboração do protocolo desta revisão foi definido que serão identificados estudos relacionados ao ENEM, ENADE e PISA. Desta forma, para a identificação dos estudos foi elaborada a seguinte *string* de busca: “(visualização OR plataforma web OR ferramenta OR *visualization* OR web *platform* OR *tool*) AND (*enem* OR *enade* OR *pisa*) AND (microdados OR base de dados OR dataset OR *microdata* OR *database*)”, no período de 2016 a 2021, com idiomas português e inglês, aplicada no portal de periódicos da CAPES e Google Acadêmico.

Foram localizados 98 estudos no Google Acadêmico e 35 estudos no portal de periódicos da CAPES. O processo de classificação dos estudos foi dividido em quatro fases, descritas abaixo.

Fase 1

Na primeira etapa foram analisados apenas o título e a data dos estudos, a fim de desconsiderar estudos duplicados, anteriores a 2016 ou que não tenham relação com o tema deste trabalho. Ao final desta etapa, foram selecionados 22 estudos do Google Acadêmico e um estudo do portal Capes, porém este foi desconsiderado já que também foi localizado no Google Acadêmico.

Fase 2

Nesta etapa foram analisados o resumo e as palavras-chave dos estudos selecionados na Fase 1, com o objetivo de escolher apenas os estudos relacionados às ferramentas de visualização e análise de dados. Foram desconsiderados estudos que não estavam relacionados aos dados educacionais, que eram focados em mineração de dados e não possuíam interface para interação, ou ainda estudos que analisaram uma situação específica sem desenvolver uma ferramenta. Ao final desta etapa, foram elencados 8 estudos.

Fase 3

Na terceira etapa foram analisadas a introdução e conclusão, a fim de identificar a proposta e os resultados alcançados. Nesta fase foram desconsiderados três estudos, sendo um com o objetivo de analisar os dados do censo escolar, outro que busca centralizar os dados de diversas fontes e o terceiro que propôs uma metodologia para análise de dados sobre o ensino superior, mas não desenvolveu uma ferramenta. Ao final desta etapa foram selecionados 5 estudos.

Fase 4

Na quarta etapa os estudos foram analisados com base nas questões de qualidade definidas durante a elaboração do protocolo desta revisão de literatura, a fim de identificar semelhanças entre os estudos selecionados e características que possam ser úteis para a elaboração da ferramenta de visualização.

As questões de qualidade consideradas foram:

* Qual prova (ENEM, ENADE, PISA) pode ser visualizada?
* Quais tipos de análise as plataformas disponibilizam ao usuário?
* Quais tecnologias foram utilizadas na plataforma?
* Permite comparação entre diferentes edições das provas?
* Quais edições das provas estão disponíveis para análise?
* É possível realizar consultas personalizadas ou apenas consultas já pré-programadas?
* Quantos atributos existem nas análises?
* Qual o objetivo da plataforma e público-alvo esperado?
* O acesso à ferramenta está disponível?

A Figura 3 demonstra cada fase do processo de filtragem.

Figura 3 - Fases de seleção dos estudos

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Fonte: Elaborado pelo autor

## DISCUSSÃO SOBRE OS ARTIGOS SELECIONADOS

A seguir serão apresentados os cinco artigos selecionados durante a etapa de revisão da literatura, cada uma foi analisado conforme as questões de qualidade.

### Um Estudo Voltado À Transparência Pública Através Da Análise De Dados Abertos Brasileiros - (MELO, 2017)

Este artigo teve como objetivo compilar os dados abertos sobre o PROUNI, PRONATEC, ENEM e PNAE, no período de 2014 a 2016 e desenvolver uma ferramenta baseada em *Business Intelligence*. Foi verificado que mesmo os dados estando acessíveis na forma de microdados, o público em geral nem sempre tem o conhecimento ou ferramentas necessárias para processá-los e extrair as informações desejadas. Além disso, devido às mudanças na estrutura dos microdados ao longo do tempo, tais análises requerem um pré-processamento antes de se comparar diferentes edições.

Não foi desenvolvida nenhuma ferramenta, foi utilizado o software QlikView para tratar os dados e gerar as visualizações. Por analisar diferentes avaliações, o estudo focou em analisar pontos comuns entre elas, como o número de escolas participantes em cada região. Sobre o ENEM foram realizadas análises baseando-se nas variáveis disponíveis nos microdados: Ano de aplicação do ENEM, UF da escola, Município da escola, Dependência Administrativa da escola, Localização (Rural / Urbana) da escola, Código da escola, Média de notas da escola em cada área do conhecimento e Número de participantes por escola.

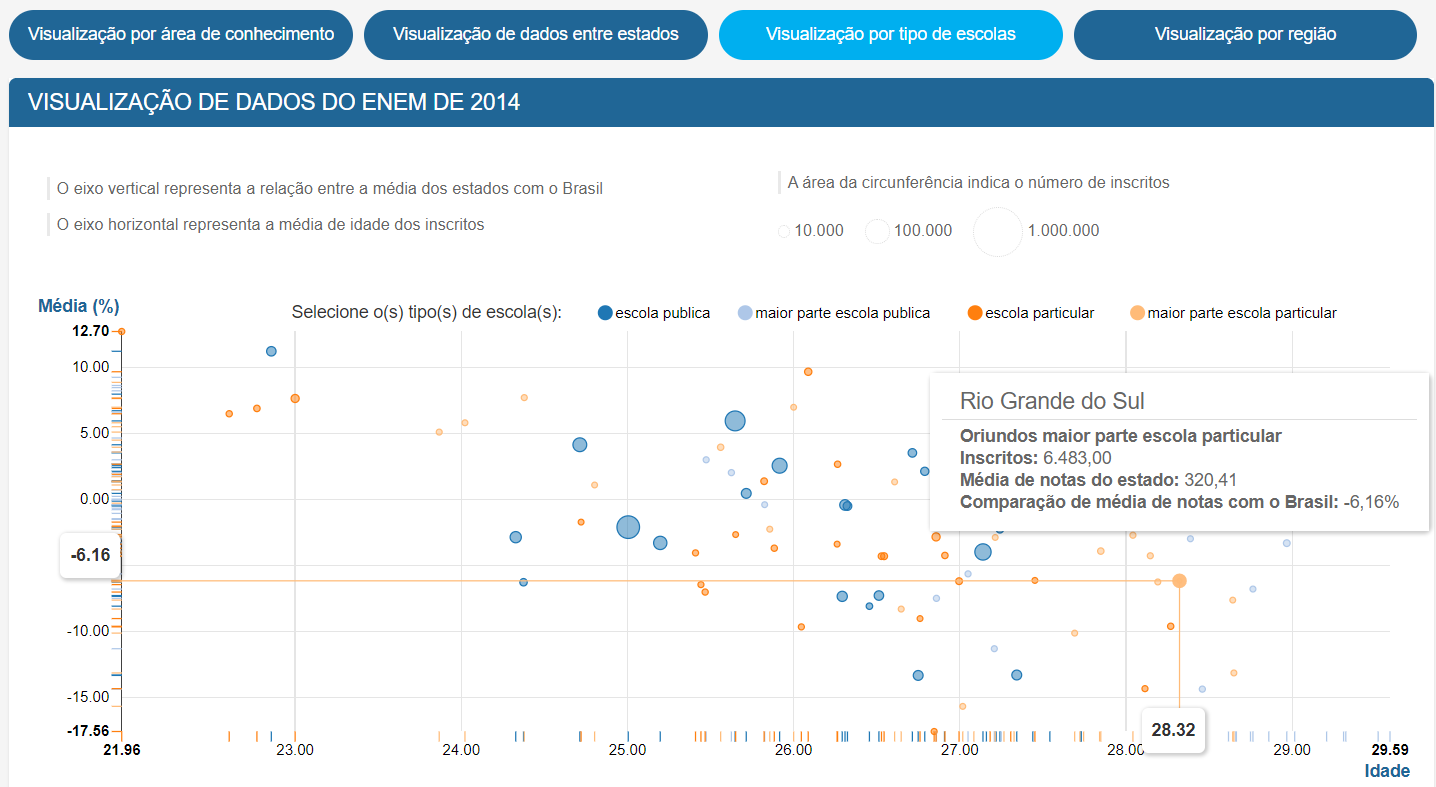
Não foram disponibilizadas formas de acesso a ferramenta ou aos dados processados.

### VisDadosEnem: Visualização De Dados Do Enem Com A Biblioteca D3 - (SANTOS e FREITAS, 2016)

Os autores desenvolveram uma aplicação Web para a visualização dos dados do ENEM apenas da edição de 2014, utilizando PHP, base de dados MySQL e a Biblioteca D3 (Javascript) para geração dos gráficos. Foram utilizadas nove variáveis extraídas dos microdados do ENEM, são elas: Número Inscrição, UF do Participante, Nota em Ciências da Natureza, Nota em Ciências Humanas, Nota em Linguagens e Códigos, Nota em Matemática, Nota da Redação, Tipo de escola e Idade do participante. A partir destes dados foram criadas quatro formas de visualização, todas baseadas em gráficos de bolhas (*Bubble Map*).

* Visualização por área de conhecimento: é apresentado um gráfico para cada estado onde é possível visualizar a média de notas por tipo de escola (Pública, Particular, etc). Nesta tela é possível selecionar quais áreas do conhecimento serão analisadas.
* Visualização de dados entre estados: permite a comparação da média de notas por tipo de escola (Pública, Particular, etc) entre dois estados. Ou a comparação de um estado com a média nacional.
* Visualização por tipos de escola: permite visualizar o panorama nacional das escolas em um único gráfico. Como demonstrado na Figura 4, o tamanho da bolha representa a quantidade de estudantes, a cor da bolha representa o tipo de escola, o eixo vertical representa a relação entre a média dos estados com o Brasil e o eixo horizontal representa a média de idade dos inscritos. Ao passar o mouse sobre o gráfico é possível visualizar detalhes sobre cada estado.
* Visualização por região: representa as mesmas informações da tela “Visualização por tipos de escola”, porém agrupando os tipos de escolas por região.

Figura 4 - Exemplo de visualização por tipo de escola



Fonte: VisDadosEnem (2019)

A plataforma VisDadosEnem possui acesso gratuito e está disponibilizada no endereço: http://labdata.sites.ufsc.br/visdadosenem

### Visualização dos resultados das edições de 2010 a 2015 do ENEM através de um Shiny App - (NASCIMENTO, 2017)

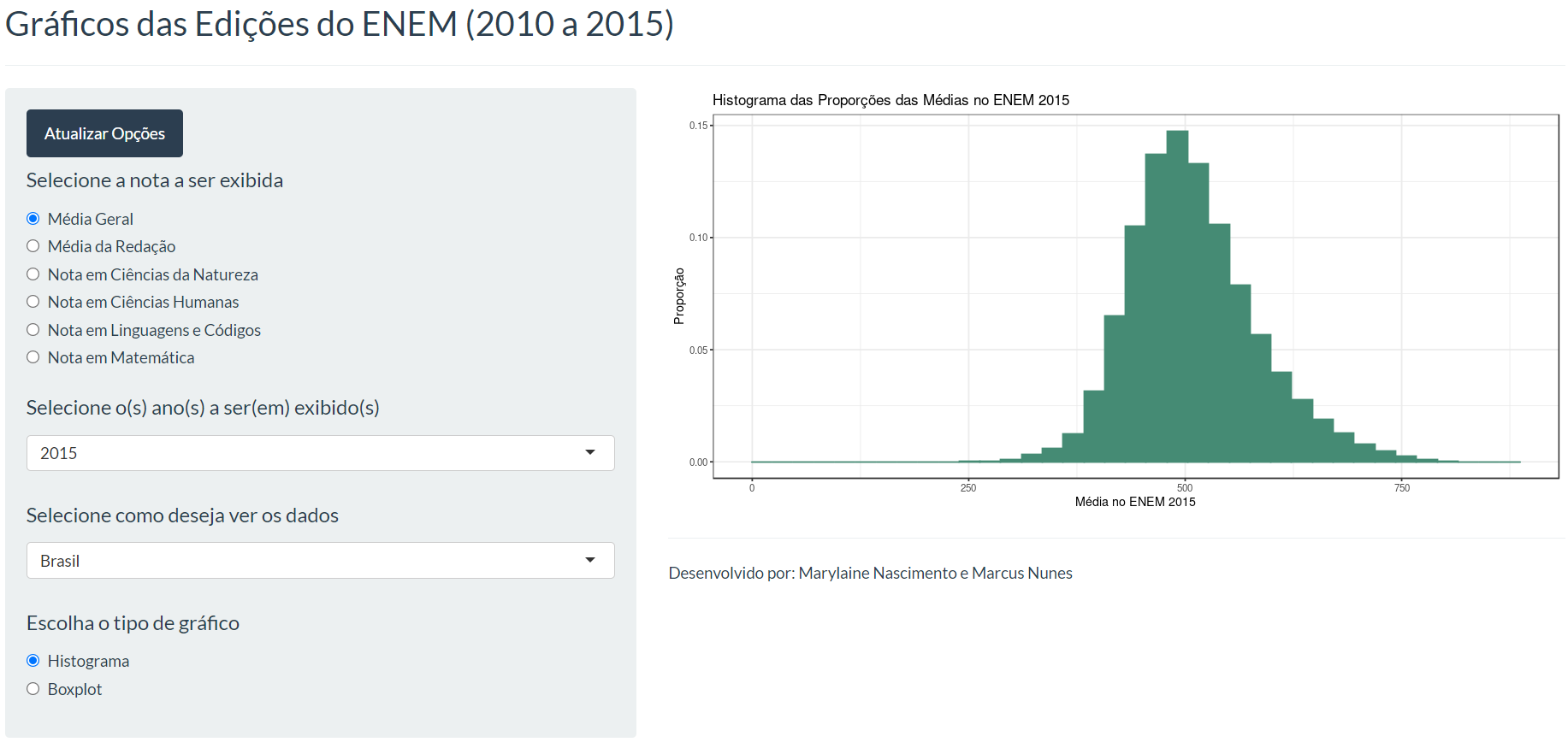
Este estudo foi elaborado como trabalho de conclusão de um curso de Bacharelado em Estatística e teve como objetivo criar um aplicativo WEB, interativo, permitindo que leigos em estatística possam explorar os dados como desejarem e tirar suas próprias conclusões.

O aplicativo foi desenvolvido utilizando a linguagem R, juntamente com o pacote *Shiny*, para desenvolver a interface Web e o pacote ggplot2 para elaboração dos gráficos, permitindo analisar as notas do ENEM entre 2010 e 2015.

As variáveis utilizadas na ferramenta são: Nota de cada área do conhecimento, Sexo, Renda e Região. A partir delas foram construídas cinco formas de visualização:

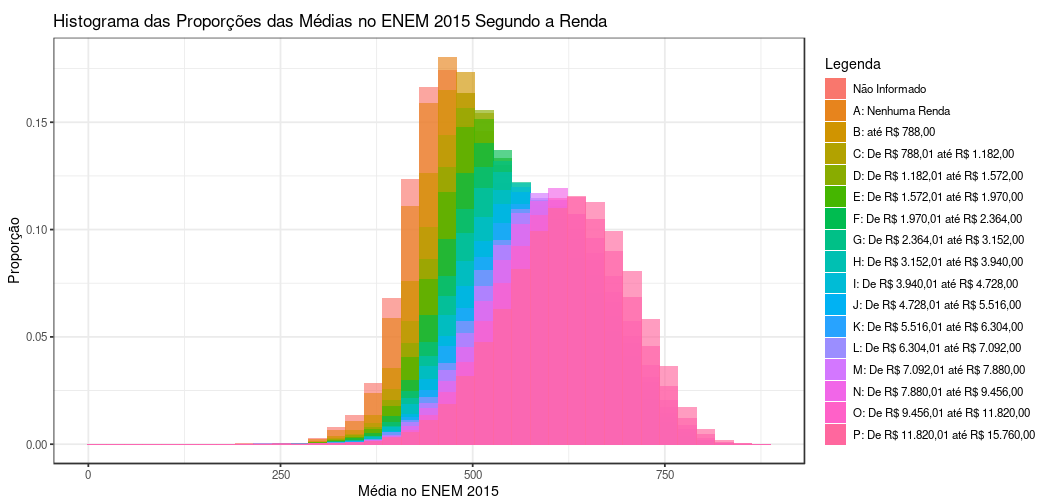
* Brasil: análise de distribuição das notas a nível nacional. Vide exemplo da Figura 5.
* Por sexo: análise da distribuição das notas conforme o sexo masculino e feminino
* Por região: média de notas em cada região (noroeste, nordeste, norte, sudoeste, sudeste e sul)
* Por sexo e região: divide a visualização por sexo em cada região
* Por renda: análise da média de notas de acordo com a renda. Vide exemplo da Figura 6.

Figura 5 - Exemplo de análise de notas a nível Brasil



Fonte: Nascimento (2017)

Figura 6 - Exemplo de análise de notas por renda



Fonte: Nascimento (2017)

A análise pode ser feita selecionando cada edição do ENEM individualmente (entre 2010 e 2015) ou a média geral de todos as edições. Não é possível comparar duas edições. Além disso, é possível filtrar por uma área do conhecimento específica ou a média de nota de todas as áreas do conhecimento.

Foram utilizados dois tipos de gráficos para visualizar as informações: histograma ou *Boxplot*. Como o público-alvo descrito pelo autor são de pessoas leigas em estatística, foi feita uma página com informações de como interpretar os gráficos.

A plataforma está disponível no endereço <http://shiny.estatistica.ccet.ufrn.br/enem/>

### Gestão dos resultados do ENADE com uso de dados abertos e tecnologias digitais - (ALVARES e CAMPOS, 2018)

Neste estudo os autores desenvolveram duas aplicações com o objetivo de prover aos gestores de Instituições de Ensino Superior uma ferramenta para auxiliar na gestão baseada nos dados do ENADE.

As aplicações foram desenvolvidas com o software QlikSense, versão *desktop*. A primeira aplicação baseou-se em planilhas do ENADE e CPC do INEP e segundo os autores “tem por objetivo recuperar o resultado de qualquer IES, participante do Enade, bem como prover comparação entre IES que ofertam o mesmo curso em determinada edição” (Alvares e Campos, 2018). A segunda aplicação utiliza os microdados do ENADE de 2011 a 2016 e “permite que a IES tenha acesso a diversos painéis de dados com o detalhamento dos resultados dos seus concluintes no ENADE” (Alvares e Campos, 2018).

Para a elaboração das telas foram feitas entrevistas com docentes e coordenadores de cursos de graduação para entendimento das necessidades e das estratégias utilizadas ao analisar os dados do ENADE. Foi percebido que a maioria dos entrevistados utiliza as planilhas divulgadas pelo INEP e que 90% não conheciam os microdados.

Não foram disponibilizadas formas de acesso a ferramenta ou os dados processados.

### Análise dos microdados do ENADE, Proposta de uma ferramenta de exploração utilizando mineração de dados – (ARAÚJO, 2019)

Este estudo se propôs a desenvolver uma ferramenta WEB de análise exploratória dos microdados do ENADE de 2017, incluindo uma etapa onde foram aplicadas técnicas de mineração de dados para avaliar se o desempenho dos alunos é alto ou baixo. Para elaboração da ferramenta foram utilizadas a linguagem R e o pacote Shiny para desenvolver a visualização Web. Na etapa de mineração de dados foi utilizado o software Weka. A fim de visualizar os dados foram utilizados gráficos de barras, gráficos de pizza, histogramas, gráficos de dispersão e diagramas de caixa (*boxplot*).

Para avaliar a ferramenta, um questionário com 10 questões foi enviado para professores (38,5%), alunos (15,4%) e coordenadores (46,2%) de IES. Como resultado, foi avaliado que a ferramenta cumpre o que promete de forma integral (65,6%) ou parcial (31,3%), já sobre a chance de utilizar a ferramenta no dia a dia, em uma escala de 0 a 5, a maioria dos entrevistados avaliaram como 3 (37,5%) e 4 (31,3%).

Os atributos disponíveis para análise não foram listados por se tratar de um estudo voltado exclusivamente ao ENADE, sendo dados diferentes do ENEM. O endereço para acessar a ferramenta não foi localizado.

### Comparativo entre estudos

O Quadro 2 apresenta a relação dos estudos analisados de acordo com cada questão de qualidade.

Quadro 2 - Comparativo entre estudos x questões de qualidade.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **VisDadosEnem: Visualização De Dados Do Enem Com A Biblioteca D3** | **Visualização dos resultados das edições de 2010 a 2015 do ENEM através de um Shiny App** | **Análise dos microdados do Enade, Proposta de uma ferramenta de exploração utilizando mineração de dados** | **Um Estudo Voltado À Transparência Pública Através Da Análise De Dados Abertos Brasileiros** | **Gestão dos resultados do ENADE com uso de dados abertos e tecnologias digitais** |
| **Qual prova pode ser visualizada?** | ENEM | ENEM | ENADE | PROUNI, PRONATEC, ENEM e PNAE | ENADE |
| **Quais tipos de análise as plataformas disponibilizam ao usuário?** | Visualização de notas por área de conhecimento, dados entre estados, tipo de escolas e por região | Notas das edições do ENEM de 2010 a 2015 em seis grupos e dois tipos de gráfico | Mineração de dados com Weka, visualização com diagrama de caixa, gráfico de pizza, histograma | Plataforma para concentrar os dados de várias bases de dados abertas, análise de números de alunos e média de notas com gráficos de barras e pizza | Média de notas por curso, número de participantes em cada curso por instituição com gráfico de barras, pizza e bolhas |
| **Quais tecnologias foram utilizadas na plataforma?** | PHP + Mysql (WAMP), Biblioteca D3 e Bootstrap | R + Pacote Shiny + Pacote ggplot2 | Weka + R + Shiny | QlikView® | QlikSense Desktop |
| **Permite comparação entre diferentes edições?** | Não | Sim | Não | Sim | Não |
| **Quais edições do Enem (ano) estão disponíveis para análise?** | 2014 | 2010 a 2015 | 2017 | 2013 a 2015 | 2016 |
| **É possível realizar consultas personalizadas?** | Consultas estáticas, permite apenas classificar por escola, por região, etc. | Consultas estáticas, com intervalos pré-definidos, com 2 tipos de gráficos | Consultas prontas, permite apenas filtrar | Consultas prontas, permite apenas filtrar | Consultas prontas, permite apenas filtrar |
| **Quantos atributos do ENEM existem nas análises?** | 9 | 5 | 0 | 12 | 0 |
| **Qual o público-alvo esperado? / Objetivo da plataforma?** | Instituições de ensino, mas não exclusivamente | Leigos em informática interessados nos dados do ENEM / Facilitar para trabalhos estatísticos | Professores, alunos e coordenadores | Público em geral interessado em dados abertos educacionais | Gestores de Instituições de Ensino Superior |
| **Acesso disponível?** | Sim | Sim | Não | Não | Não |

Fonte: Elaborado pelo autor

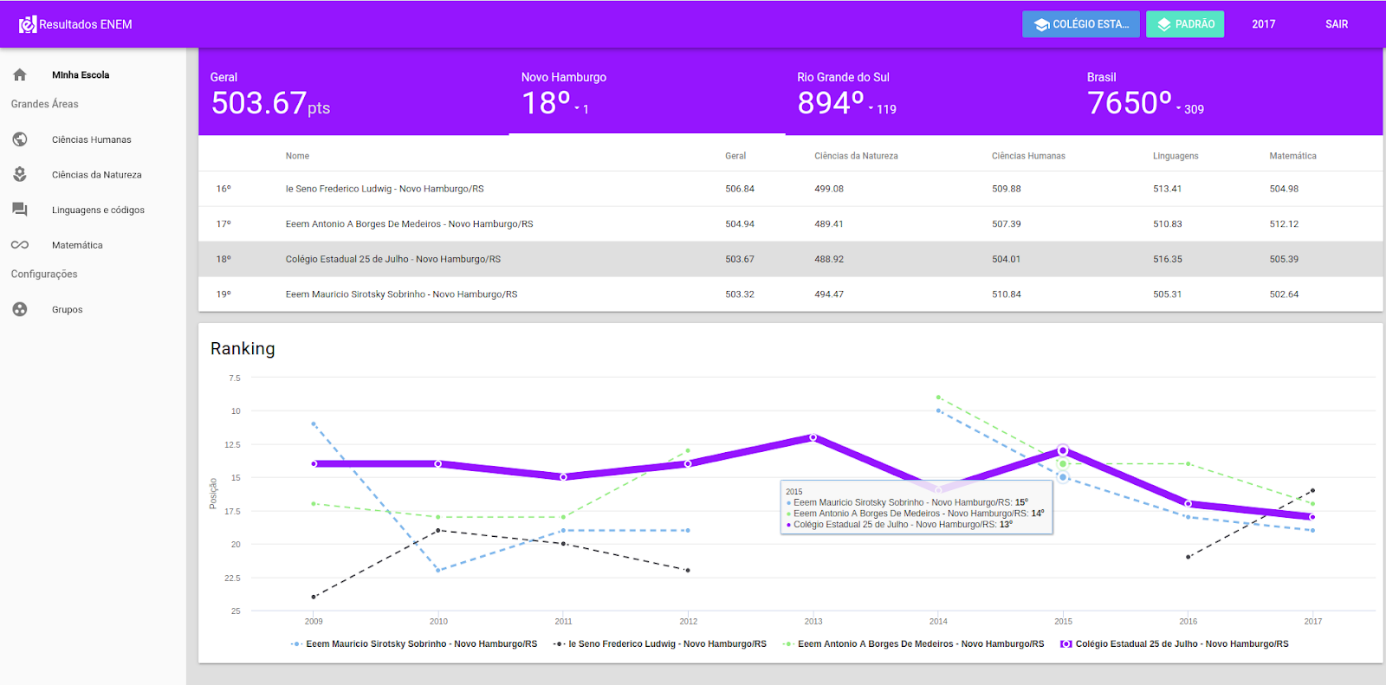
## FERRAMENTAS EXISTENTES

Durante a construção desta pesquisa, foram encontradas ferramentas desenvolvidas por empresas privadas com o objetivo de auxiliar instituições de ensino a identificarem as áreas que os alunos têm mais dificuldades, oferecendo uma comparação com outras instituições da mesma região.

### Resultados ENEM

Resultados Enem é uma plataforma gratuita que permite avaliar o desempenho médio das escolas baseado no desempenho de seus alunos no ENEM, no momento do cadastro é preciso escolher uma escola base, que será utilizada na comparação. No momento da elaboração deste trabalho a plataforma permite a análise dos dados entre as edições de 2009 a 2017. Na tela inicial é possível ver o *ranking* da escola comparado em relação a cidade, estado e Brasil, além de um gráfico demonstrando a evolução da escola ao longo dos anos conforme demonstrado na Figura 7.

Figura 7 - Página inicial da plataforma Resultados ENEM - ranking da escola selecionada

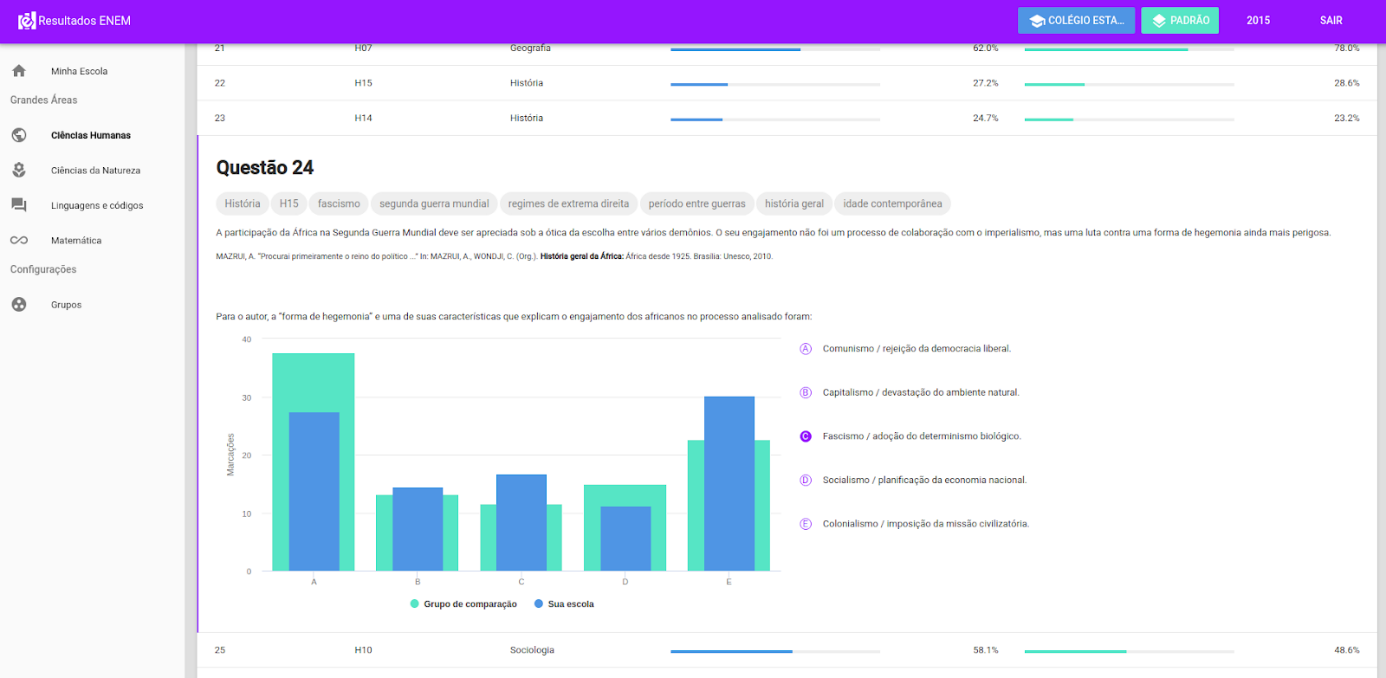


Fonte: Resultados Enem, 2021

O site não possui menu de ajuda, nem informações sobre o desenvolvedor ou sobre suporte. Todas as análises são relacionadas à escola base, não sendo possível, por exemplo, verificar qual é a escola com a primeira colocação no *ranking* em cada categoria.

Um ponto interessante nesta plataforma, como pode ser visto na Figura 8, é a possibilidade de analisar individualmente cada questão das provas do ENEM. Assim, é possível visualizar qual opção os alunos escolheram, assim como o que os alunos de outras escolas da região escolheram. Isso permite identificar deficiências no ensino, e quais áreas do conhecimento precisam de mais apoio.

Figura 8 - Plataforma Resultados ENEM - avaliação individual de cada questão



Fonte: Resultados Enem, 2021

Desta forma, a plataforma parece ser interessante para os gestores de escolas avaliarem o desempenho da instituição, mas carece de possibilidades para que a administração pública e pesquisadores independentes possam fazer análises de uma região ou outra classificação sem estar relacionado a uma escola específica.

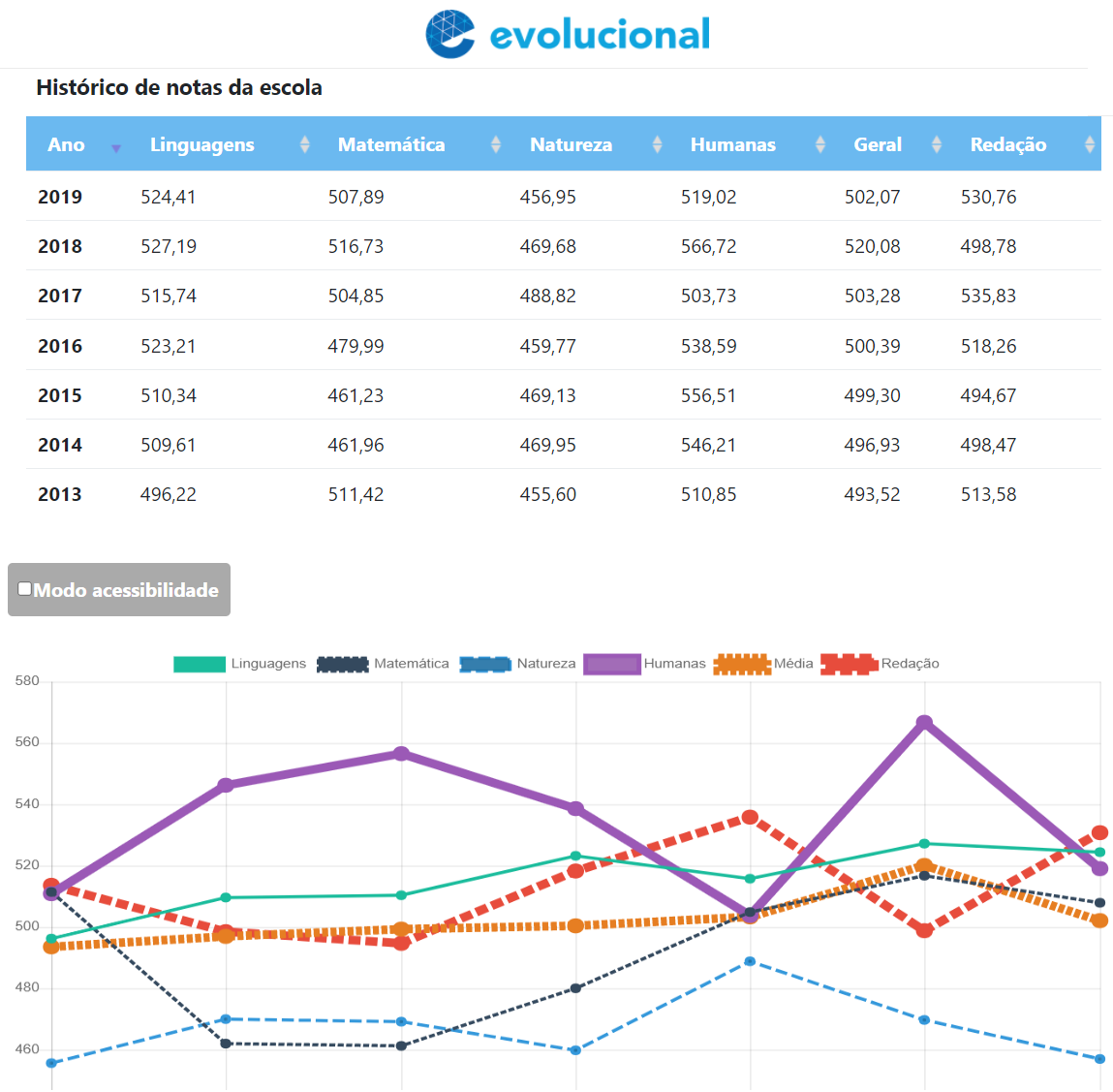
A plataforma Resultados ENEM está acessível no endereço: <https://www.resultadosenem.com.br/>

### Evolucional

Evolucional é uma empresa focada em simulados. Ela aplica simulados presenciais, semelhantes ao ENEM oficial, com o objetivo de preparar os alunos e aumentar o *score* das escolas. A plataforma fornece “diagnósticos precisos para escolas e alunos, permitindo a intervenção estratégica e pedagógica e contribuindo para a melhora do desempenho no ENEM” (Evolucional, 2021).

Está disponibilizada de forma gratuita uma ferramenta para analisar a evolução do desempenho de uma escola específica, como pode ser visualizado na Figura 9. Além disso, a ferramenta faz comparações com outras escolas da região. No momento do desenvolvimento deste trabalho estão disponíveis os dados entre as edições do ENEM de 2016 a 2019.

Figura 9 - Evolucional - Análise de desempenho da uma escola específica



Fonte: Evolucional, 2021

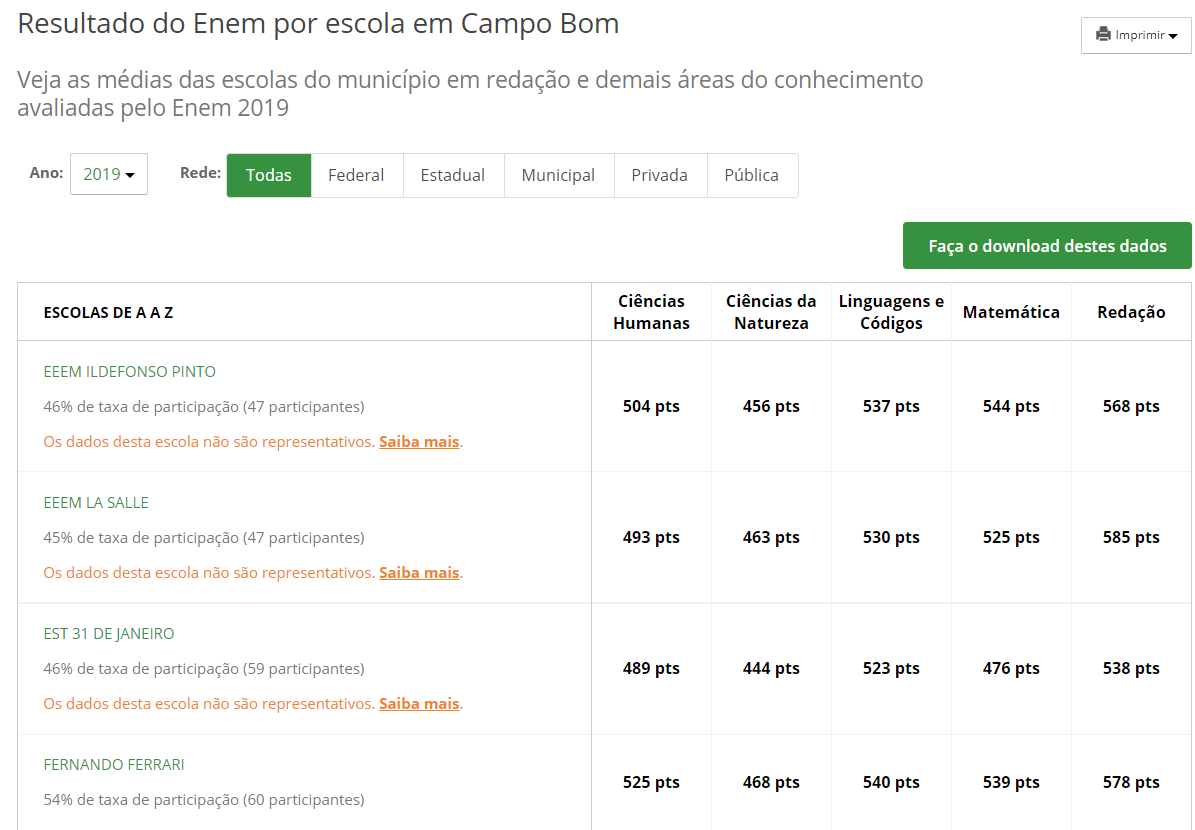
A plataforma Evolucional está acessível no endereço: <https://microdados.evolucional.com.br/microdados/43092462-D68BF9D>

### QEdu

O QEdu é uma plataforma que contém os principais dados do nosso ensino básico proveniente de fontes oficiais (INEP, Censo Escolar, Saeb, Prova Brasil) com o propósito de auxiliar gestores, diretores, professores e todos os interessados fornecendo informações acerca da qualidade do aprendizado dos alunos em cada escola, município e estado do Brasil.

A plataforma permite analisar tanto as informações de escolas especificas quanto por região, municípios e estados. Um exemplo pode ser visto na Figura 10, onde é possível verificar o *ranking* das escolas na cidade de Campo Bom de acordo com cada área de conhecimento.

Figura 10 - Qedu - Exemplo de análise do ENEM por município



Fonte: QEdu, 2021

Por ser uma plataforma que concentra várias fontes de dados, nem todas as informações continas nos microdados do ENEM estão disponíveis para análise, como os questionários socioeconômicos. A plataforma QEdu pode ser acessada no endereço: <https://www.qedu.org.br/>

### Painéis ENEM

A plataforma Painéis ENEM foi desenvolvida pelo INEP e conta com dados estatísticos sobre o ENEM das edições de 2015 a 2020, sendo apresentado em duas páginas: Sinopse Estatística, que contém um gráfico de barras e um mapa do Brasil, conforme demonstrado na Figura 11, onde são apresentadas informações de um ano específico selecionado pelo usuário. Também contém a página Sinopse Série Histórica, que apresenta dois gráficos de linha com informações sobre a evolução do ENEM ao longo os anos.

Figura 11 - Painéis ENEM - Sinopse Estatística

Interface gráfica do usuário, Aplicativo, Tabela, Excel

Descrição gerada automaticamente

Fonte: INEP, 2022

Independente das variáveis ou opções selecionadas, o gráfico de barras apresentado na Figura 11 sempre é gerado com base nas regiões Centro-Oeste, Nordeste, Norte, Sudeste e Sul. Desta forma, não é possível gerar gráficos comparando dois municípios ou dois estados de forma visual, já que será exibido apenas a região a que pertencem.

Não estão disponíveis variáveis como nota, renda e escolaridade dos pais. As variáveis disponíveis para análise/comparação são: número de participantes presentes, sexo, cor/raça, faixa etária, estado civil, atendimento especializado, necessidade especial, atendimento específico, ano de conclusão do ensino médio, nome social, isenção de taxa e dependência administrativa da escola.

A página Painéis ENEM foi desenvolvida com base na plataforma Microsoft PowerBI, que, segundo a Microsoft (2022) é uma coleção de serviços de software e aplicativos que permitem de forma facilitada gerar gráficos iterativos, relatórios e *dashboards* a partir de informações obtidas em diversas fontes de dados como banco de dados, arquivos CSV e APIs.

A plataforma Painéis ENEM pode ser acessada no endereço: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/inep-data/paineis-enem>

Este capítulo apresentou cinco estudos relacionados a visualização de dados educacionais, onde foram identificados suas características e limitações. Foram identificados diversos pontos que podem ser aprimorados como, por exemplo, a maioria dos estudos avaliados permitem analisar poucas edições de cada prova, também disponibilizam poucas variáveis para análise, se comparado ao número total de variáveis presentes nos microdados do ENEM.

Também foram apresentadas quatro ferramentas relacionadas à visualização de dados do ENEM, onde foi identificado que elas têm como objetivo auxiliar escolas especificas, dificultando análises mais abrangentes, por região por exemplo. Com exceção das plataformas QEdu e Painéis ENEM, que se mostram mais completas, porém não utilizam muitas das variáveis que estão presentes nos microdados.

# DESENVOLVIMENTO DA PLATAFORMA

Este capítulo descreve os procedimentos adotados no desenvolvimento da plataforma DadosENEM, contendo as etapas de pré-processamento dos microdados do ENEM, normalização e importação dos dados, prototipação e desenvolvimento das telas. Conta também com a definição dos gráficos e comparativos que foram disponibilizados.

## Prototipação das visualizações

Inicialmente foram definidos os tipos de visualizações que a ferramenta pretende disponibilizar, quais variáveis serão utilizadas e as edições do ENEM que estarão disponíveis para análise. Estas definições basearam-se no levantamento bibliográfico feito no Capítulo 3, onde foram identificados pontos que podem ser aprimorados em cada plataforma existente.

Também foi feito o levantamento das variáveis presentes nos arquivos de microdados e que não foram utilizadas pelas plataformas existentes. A partir disto foram criadas visualizações adicionais que podem ser úteis aos usuários, além de serem utilizadas no *dashboard* com informações gerais sobre a evolução do ENEM ao longo dos anos.

Durante o processo de revisão sistemática e de pesquisa de plataformas de visualização foram identificadas quais informações as plataformas disponibilizam para os usuários e quais tipos de gráficos são utilizados em cada situação, com destaque para a plataforma VisDadosENEM. Esta utilizou gráficos de bolha para relacionar até cinco variáveis simultaneamente. Conforme o exemplo da Figura 12, cada ponto representa um estado, a cor representa a região, o tamanho dos elementos representa o número de inscritos para o ENEM, o eixo X representa a idade média dos inscritos em cada estado e o eixo Y representa uma relação da nota média do estado em relação à média nacional. Ao mover o *mouse* sobre o gráfico é possível visualizar mais detalhes sobre cada ponto.

Ao analisar a Figura 12 pode-se facilmente deduzir que o estado de São Paulo, na edição do ENEM de 2014, foi o estado que mais possuiu inscritos no ENEM, a média de idade dos inscritos foi de 25 anos e a nota média foi 2,10% inferior à média nacional. Também é possível visualizar no mesmo gráfico que, em 2014, a média de idade dos inscritos foi superior a 26 anos na maioria dos estados do nordeste.

Figura 12 - VisDadosENEM - Comparativo de regiões com gráficos de Bolha

Uma imagem contendo Gráfico de dispersão

Descrição gerada automaticamente

Fonte: VisDadosENEM (2021)

Segundo Microsoft (2021) “um gráfico de bolhas é uma variação de um gráfico de dispersão, no qual os pontos de dados são substituídos por bolhas, e uma dimensão adicional dos dados é representada no tamanho das bolhas”. Desta forma, optou-se por utilizar gráficos de bolha como principal forma de visualização na plataforma DadosENEM, devido a possibilidade de relacionar diversas variáveis em um mesmo gráfico.

Após analisar as plataformas existentes percebeu-se duas limitações importantes. Não é possível analisar o desempenho individual de escolas, cidades ou microrregiões, apenas o estado ou região como um todo. Também não é permitido analisar várias edições do ENEM simultaneamente, não sendo possível visualizar a evolução do desempenho ao longo dos anos em uma mesma escala.

A partir das visualizações das plataformas existentes e suas limitações, e baseado nas variáveis presentes nos arquivos de microdados que não foram utilizados por elas, foi elaborado um protótipo da plataforma. Este protótipo conta com um *dashboard* que possui gráficos sobre a evolução do ENEM ao longo das edições, conforme a Figura 13, e uma tela para análise mais detalhada dos dados, vide Figura 14, onde é possível definir o escopo e o tipo de comparação a ser feito.

Figura 13 - Protótipo de tela inicial

**Diagrama

Descrição gerada automaticamente**

Fonte: Elaborado pelo autor

Na tela inicial, conforme demonstrado na Figura 13, pretende-se disponibilizar diversos gráficos com informações relacionadas à evolução do ENEM ao longo dos anos, isso permite ao usuário ter um contexto geral e ajudar na interpretação dos valores quando estiver explorando os dados mais específicos na tela de comparativos. Algumas das informações disponíveis são:

**Número de participantes:**  
- Total de participantes inscritos no ENEM em cada ano

**Número de participantes por situação:**

- Relação no número de presentes, ausentes e eliminados, considerando todos os participantes inscritos no ENEM.

**Nota média nacional por área do conhecimento:**

- Média por área de conhecimento, considerando apenas participantes presentes nos 2 dias do ENEM, que não tenham sido eliminados e participantes de estejam cursando ensino médio e que estejam concluindo no ano vigente ou concluirão posteriormente (participantes que já concluíram o ensino médio em anos anteriores e treineiros não são considerados)

**Nota média nacional:**

- Média geral, considerando a média aritmética de todas as áreas do conhecimento e a redação.

Figura 14 - Protótipo de tela de comparativos

Interface gráfica do usuário

Descrição gerada automaticamente

Fonte: Elaborado pelo autor

Na tela Comparativos, demonstrada na Figura 14, o intuito é permitir aos usuários localizar informações mais especificais. O escopo delimita o que será comparado, desde escolas individuais até regiões do Brasil. O campo seleção traz informações referentes ao que foi selecionado no escopo, onde o usuário pode selecionar um ou vários itens para comparação. A Referência é o tipo de comparação a ser feito e quais edições serão retornadas. Um gráfico será gerado a partir das informações solicitadas, sendo o eixo X a média de idade em cada grupo, o eixo Y o percentual da nota média do grupo em relação com a nota média nacional, os demais atributos serão apresentados utilizando características possíveis em gráficos de bolha, como tamanho, cor e formato, dependendo do que for selecionado no campo Referência.

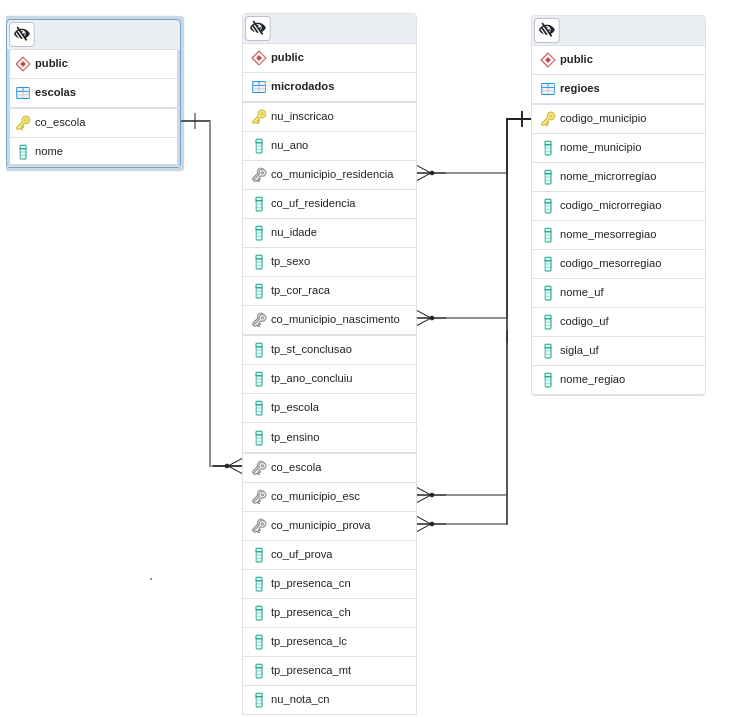
Este modelo atende aos objetivos da plataforma, permitido desde gestores de escolas verificarem a evolução de uma única escola ou comparar com demais escolas da cidade, até a gestão pública, que pode avaliar o desempenho geral das escolas nas cidades ao se utilizar o escopo baseado em cidade ou microrregiões. Além de atender também aos pesquisadores em geral que podem avaliar o desemprenho nacional ao utilizar o escopo por região ou estado.

## Modelagem da plataforma

Para armazenamento dos dados foi utilizado o banco de dados relacional PostgreSQL, onde foram criadas três tabelas. A primeira tabela denomina-se “regiões”, com dados obtidos na página do IBGE (2021) contendo a listagem de todos os municípios brasileiros, juntamente com o estado, microrregião e mesorregião a que o município pertence. A segunda tabela, “escolas”, que contém o código e o nome obtidos a partir dos dados do censo escolar, disponibilizado pelo INEP. Por fim, tem-se a tabela “microdados”, onde estão armazenadas as informações importadas dos arquivos de microdados do ENEM de todas as edições, a partir de 2013. A estrutura desta última tabela foi baseada na estrutura dos microdados do ENEM de 2019; esta edição foi escolhida porque o INEP tem seguido esse padrão desde 2016. Foram adaptadas apenas as colunas referentes aos questionários socioeconômicos.

A Figura 15 exibe o modelo ER das tabelas. A tabela microdados possui 115 colunas, a imagem apresenta apenas algumas delas. As tabelas auxiliares “escolas” e “regiões” são necessárias, pois nos microdados do ENEM são disponibilizados apenas o código das escolas / municípios.

Figura 15 - Modelo ER da plataforma DadosENEM



Fonte: Elaborado pelo autor

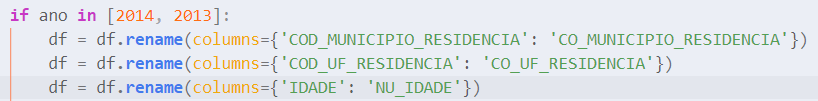
## Importação e padronização dos microdados

Os microdados do ENEM estão disponíveis em arquivo texto de formato CSV (*Comma Separated Values*), sendo que há um arquivo para cada edição (ano). Porém, devido ao tamanho dos arquivos (entre 2,97GB e 5,68GB) e as diferenças existentes entre cada edição, foi necessário criar um *script* para padronizar e importar os dados.

O *script* “importador-microdados-enem.py” foi desenvolvido em *python,* versão 3.8, e pode ser executado informando apenas o nome do arquivo de microdados como parâmetro. Ao iniciar a importação o *script* identifica a edição que está sendo importada através do campo NU\_ANO e faz os ajustes necessários para deixar os microdados de acordo com a estrutura do banco de dados previamente criado.

Com o objetivo de melhorar a performance e não carregar o arquivo completo em memória, foi utilizada a biblioteca Pandas, que permitiu importar os dados em blocos de 50.000 registros (esse conjunto de registros são chamados de *dataframes*), por meio da função “read\_csv” e o parâmetro “chunksize”. Também foi necessário padronizar as diferenças dos atributos entre as edições utilizadas. Tais diferenças foram exibidas no Quadro 1, disponível no Capítulo 2. Por exemplo, o atributo “NU\_IDADE”, em algumas edições, passou a ser chamado apenas “IDADE”. Isso foi ajustado utilizando o método *rename,* disponibilizado pelo Pandas, para renomear uma coluna do *dataframe*, conforme a Figura 16:

Figura 16 – Exemplo de código para renomar colunas com python e pandas



Fonte: Elaborado pelo autor

Há também casos em que as opções foram alteradas, como o campo “TP\_STATUS\_REDACAO”. Este campo possui valor numérico, porém ao longo das edições a representação deste número muda. Em 2014, “1” está relacionado a “Redação em Branco”, mas em 2019, “1” representa “Redação sem problemas”. Neste caso foi utilizado o método *replace* do *dataframe* para converter as opções para o padrão utilizado em 2019, destacado na Figura 17.

Figura 17 – Exemplo de código para substituição de valores com python e pandas



Fonte: Elaborado pelo autor

A Figura 17 também demostra a utilização do método *replace* para converter campos numéricos para binários, como a identificação dos participantes “treineiros”, a fim de tornar as consultas no banco de dados mais eficientes.

Existem ainda as questões socioeconômicas, que nos arquivos de microdados são identificadas da seguinte maneira: “Q01”, “Q02”, “Q03”, etc. No entanto, ao longo das edições, uma mesma questão pode ser identificada por outro número. Para contornar isso, foi criada uma coluna fixa no banco de dados para cada questão relevante, por exemplo, “qs\_escolaridade\_pai” e “qs\_renda” e durante a importação dos dados foi feito o tratamento para carregar o valor correto conforme a edição do ENEM que está sendo processada.

Uma das questões socioeconômicas precisou de um tratamento adicional, a questão socioeconômica relacionada à quantidade de veículos dos participantes. Nas edições recentes existe a informação sobre quantidade de “Carros” e quantidade de “Motocicletas”, porém nas edições mais antigas estes dois atributos eram identificados apenas como “Automóvel”, impedindo assim a utilização deste valor ao realizar comparações entre diferentes edições. Para contornar isso foram criados três campos no banco de dados, são eles: QS\_QUANTIDADE\_MOTO, QS\_QUANTIDADE\_CARRO, QS\_QUANTIDADE\_AUTOMOVEL. O campo relacionado a carro e moto é preenchido apenas nas edições que possuem a informação discriminada, já o campo “QS\_QUANTIDADE\_AUTOMOVEL” é preenchido com o valor original disponível nos microdados (em edições mais antigas) ou com a soma de carros e motos nas edições mais novas.

Foram importados apenas os microdados entre 2013 e 2020, pois os microdados anteriores possuem muitas diferenças e ausência de informações em alguns campos. Tais diferenças acabam gerando distorções ao realizar comparativos com os anos posteriores.

O total de registros entre 2013 e 2020 é de 55.393.072 registros, o equivalente a aproximadamente 31GB de dados, somando todos os arquivos de microdados brutos. Cada arquivo possui entre 2,97GB (de 2019) e 5,68GB (de 2014).

Alguns campos não foram importados por serem considerados desnecessários para os propósitos desta plataforma, como o gabarito de cada prova; ou por serem valores duplicados como é o caso do nome dos municípios, sendo que também existe um campo relacionando o código do município. Após a importação, a tabela de microdados de 2013 a 2020 ficou com tamanho de 12GB e 9GB de índices. Os índices foram aplicados sobre os nove campos mais acessados que são: co\_escola, co\_municipio\_esc, co\_municipio\_resiencia, co\_uf\_esc, co\_uf\_residencia, nu\_ano, qs\_internet, qs\_numero\_computador e qs\_renda\_familia.

Além dos microdados do ENEM também foi importada a listagem dos municípios brasileiros, juntamente com a região e microrregião que cada um pertence. Esta lista foi obtida na página do IBGE[[2]](#footnote-3), criada em 2017 e é a mais atualizada disponível durante a elaboração deste trabalho. Esta relação contém a lista de todos os municípios brasileiros, também em formato CSV, e foi importada após a elaboração de um *script* em linguagem *python*. Durante a importação foi acrescentada a coluna UF, que não está presente no arquivo disponibilizado pelo IBGE, contendo a sigla de cada estado, para facilitar a relação com os microdados do ENEM.

Também foi importada a listagem de escolas e seus respectivos nomes. Isto permitiu identificar as escolas, já que nos arquivos de microdados estão presentes apenas o código da instituição junto ao INEP. Esta listagem foi obtida a partir dos arquivos de microdados do Censo Escolar de 2019.

Todos os códigos-fonte da plataforma e os *scritps* utilizados na importação dos dados estão disponíveis no endereço: <https://github.com/gustavowillrich/DadosEnem> .

## Tecnologias utilizadas

A plataforma DadosENEM foi estruturada com o banco de dados relacional PostgreSQL, a linguagem de programação Python e as bibliotecas Pandas e Plotly. A plataforma está sendo executada em uma máquina virtual com sistema operacional Ubuntu Server 20.04 LTS sobre um servidor com sistema operacional VMWare ESXi, conforme demonstrado na Figura 18.

Figura 18 – Arquitetura da plataforma DadosENEM

Interface gráfica do usuário

Descrição gerada automaticamenteFonte: Elaborado pelo autor

### Banco de dados

Decidiu-se utilizar o PostgreSQL, versão 12.9, por ser um banco de dados robusto, grátis e de código aberto. O PostgreSQL é um banco de dados relacional que está presente em projetos como Skype e Instagram, mesmo sendo independente e *open source,* conta com o apoio de diversas empresas de tecnologia como VMWare, Google e Microsoft (POSTGRESQL, 2022).

### Visões materializadas

Um dos grandes desafios desta plataforma é lidar com o volume de dados a ser analisado, pois são cerca de 55 milhões de registros, o que torna certas consultas lentas. A solução adotada foi a utilização de visões materializadas para processar e armazenar os dados que serão disponibilizados para as consultas.

Ao realizar consultas em bancos de dados é muito comum a utilização de visões (*views*). Elas permitem auxiliar na simplificação e reutilização de código, ou mesmo por segurança, ao se utilizar uma *view* com regras e filtros específicos para ocultar registros e/ou atributos confidenciais. Já as visões materializadas (*materialized views*) são similares às visões, a grande vantagem é que o resultado fica armazenado em formato de tabela, ou seja, não sendo necessário realizar uma consulta aos dados originais a cada chamada. Desta forma, obtém-se melhor desemprenho principalmente quando a consulta em questão envolve cálculos e relacionamentos entre várias tabelas (POSTGRESQL, 2022a).

Embora o acesso aos dados armazenados em uma visão materializada seja frequentemente muito mais rápido do que acessar as tabelas subjacentes diretamente ou por meio de uma visão, os dados nem sempre são atuais. No entanto, às vezes os dados atuais não são necessários, como é o caso dos microdados do ENEM, que são constituídos apenas de dados históricos e não sofrem alterações.

Durante o desenvolvimento da plataforma foram criadas 22 visões materializadas, conforme listadas no Quadro 3, cada uma com os agrupamentos e cálculos necessários para a geração de gráficos que estão disponíveis na plataforma.

Quadro 3 - Lista de visões materializadas

|  |  |
| --- | --- |
| **Nome da visão materializada** | **Descrição** |
| view\_media\_nota\_escola\_ano | Nota média por escola |
| view\_media\_nota\_escola\_raca | Nota média por escola agrupado por raça/cor |
| view\_media\_nota\_escola\_renda | Nota média por escola agrupado pela renda familiar |
| view\_media\_nota\_escola\_sexo | Nota média por escola agrupado por gênero |
| view\_media\_nota\_estado\_ano | Nota média por estado |
| view\_media\_nota\_estado\_raca | Nota média por estado agrupado por raça/cor |
| view\_media\_nota\_estado\_renda | Nota média por estado agrupado pela renda familiar |
| view\_media\_nota\_estado\_sexo | Nota média por estado agrupado por gênero |
| view\_media\_nota\_microrregiao\_ano | Nota média por microrregião |
| view\_media\_nota\_microrregiao\_raca | Nota média por microrregião agrupado por raça/cor |
| view\_media\_nota\_microrregiao\_renda | Nota média por microrregião agrupado pela renda familiar |
| view\_media\_nota\_microrregiao\_sexo | Nota média por microrregião agrupado por gênero |
| view\_media\_nota\_municipio\_ano | Nota média por município |
| view\_media\_nota\_municipio\_raca | Nota média por município agrupado por raça/cor |
| view\_media\_nota\_municipio\_renda | Nota média por município agrupado pela renda familiar |
| view\_media\_nota\_municipio\_sexo | Nota média por município agrupado por gênero |
| view\_media\_nota\_regiao\_ano | Nota média por região |
| view\_media\_nota\_regiao\_raca | Nota média por região agrupado por raça/cor |
| view\_media\_nota\_regiao\_renda | Nota média por região agrupado pela renda familiar |
| view\_media\_nota\_regiao\_sexo | Nota média por região agrupado por gênero |
| view\_numero\_participantes\_estado | Relação de participantes por estado agrupados por presença |
| view\_numero\_participantes\_nacional | Relação de participantes nacional agrupados por presença |

Fonte: Elaborado pelo autor

Um exemplo é a view\_media\_nota\_escola\_ano, responsável por calcular a média por escola, agrupado por código da escola e ano de aplicação do ENEM, considerando apenas os alunos concluintes do ensino médio, presentes nos dois dias de aplicação do exame e que não foram eliminados. Os campos disponíveis podem ser conferidos no Quadro 4.

Quadro 4 - Visão materializada view\_media\_nota\_escola\_ano.

|  |  |
| --- | --- |
| **Nome do campo** | **Descrição** |
| co\_escola | Código da escola |
| nome\_escola | Nome da escola |
| codigo\_municipio | Código do município |
| nome\_municipio | Nome do município |
| sigla\_uf | Sigla do estado |
| nome\_uf | Nome do estado |
| nome\_municipio\_uf | Concatenação de nome do município e Sigla do estado |
| nu\_ano | Ano da aplicação do ENEM |
| med\_nota\_cn | Média aritmética da área de Ciências da Natureza |
| med\_nota\_ch | Média aritmética da área de Ciências Humanas |
| med\_nota\_lc | Média aritmética da área de Linguagens e Códigos |
| med\_nota\_mt | Média aritmética da área de Matemática |
| med\_nota\_redacao | Média aritmética da Redação |
| num\_participantes | Número de participantes |
| med\_geral | Nota média geral da escola contando a redação |
| med\_objetiva | Nota média geral da escola sem a redação |
| med\_idade | Média de idade dos participantes |

Fonte: Elaborado pelo autor

Outro exemplo é a view\_media\_nota\_escola\_renda, retorna os mesmos campos, porém os valores são agrupados pelo tipo de renda. As demais visões materializadas são semelhantes, alterando apenas o tipo de agrupamento feito. Desta forma quando o usuário seleciona o Escopo e a Referência na tela de comparativos, a plataforma utiliza a visão materializada correspondente.

### Python[[3]](#footnote-4)

Decidiu-se utilizar a linguagem de programação Python como base para o desenvolvimento da plataforma devido à grande quantidade de bibliotecas e recursos disponíveis para a manipulação e exibição de grandes volumes de dados. As principais bibliotecas utilizadas na plataforma foram Pandas e Plotly/Dash.

### Pandas[[4]](#footnote-5)

Pandas é uma biblioteca Python de código aberto para análise e manipulação de dados de forma eficiente. Ela utiliza o conceito de *dataframes*, que consiste em uma estrutura bidimensional com elementos rotulados, semelhante às tabelas em banco de dados relacional. Sobre esta estrutura são disponibilizadas diversas funções para manipulação das informações, como: ordenação, filtro, combinação, substituição condicional, dentre outros. Além de facilitar a integração com fontes de dados externas como banco de dados e arquivos CSV.

### Plotly/Dash[[5]](#footnote-6)

Dash é uma biblioteca *open source* desenvolvida pela empresa Plotly com base em outras bibliotecas como Flask, Plotly.js e React.js e tem como objetivo facilitar a criação de páginas WEB para visualização de dados através de diversos gráficos iterativos. Ao utilizar Dash, todo o desenvolvimento é feito em Python, sem necessidade de criar páginas HTML, Javascript ou APIs para integração entre *back-end* e *front-end*, ao invés disso utiliza-se métodos do próprio plotly para estruturação das páginas conforme exemplo da Figura 19 (PLOTLY, 2022).

Figura 19 - Exemplo de código Plotly.



Fonte: Elaborado pelo autor

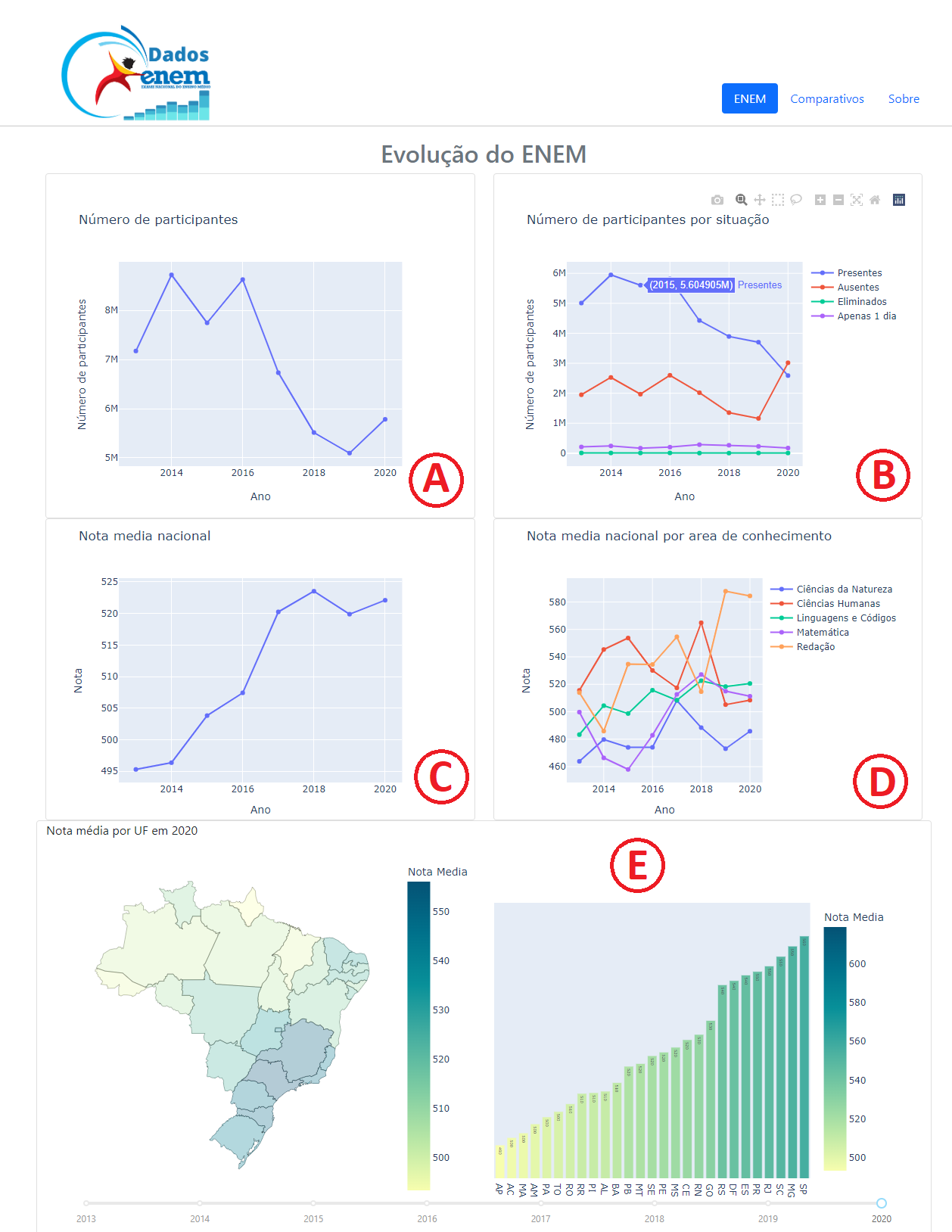
## Plataforma DadosENEM

A plataforma DadosENEM está disponível no endereço <https://vps38717.publiccloud.com.br/>. Nesta seção serão apresentadas algumas telas da plataforma desenvolvida, com descrição das funções disponíveis.

Na tela inicial, conforme demostrado na Figura 20, são apresentados alguns gráficos sobre a evolução do ENEM entre os anos de 2013 e 2020, sendo eles:

1. Número de participantes, que relaciona o total de participantes inscritos em cada edição do ENEM.
2. Número de participantes por situação, que relaciona o total de participantes presentes nos dois dias de avaliação, ausentes nos dois dias, parciais (que participaram apenas em um dia de prova) e eliminados.
3. Nota média nacional, que relaciona a média aritmética simples das questões objetivas de todas as áreas do conhecimento e redação de todos os participantes, em cada edição do ENEM. São considerados apenas participantes “válidos”, ou seja, participantes presentes nos dois dias de avaliação, que não foram eliminados e que estão concluindo o ensino médio no ano de realização da prova ou concluirão nos anos seguintes.
4. Nota média nacional por área de conhecimento, que relaciona a nota média simples separada por cada área do conhecimento. São considerados apenas participantes “válidos”, ou seja, participantes presentes nos dois dias de avaliação, que não foram eliminados e que estão concluindo o ensino médio no ano de realização da prova ou concluirão nos anos seguintes.
5. Mapa e gráfico de barras, onde é possível selecionar um ano entre 2013 e 2020 e visualizar a nota média por estado.

Figura 20 - DadosENEM - Tela inicial.



Fonte: Elaborado pelo autor

A guia COMPARATIVOS, quando clicada, abre uma tela que permite aos usuários realizar análises e comparações mais especificas. As comparações podem ser realizadas de acordo com os filtros desejados, que são formados por três campos, são eles:

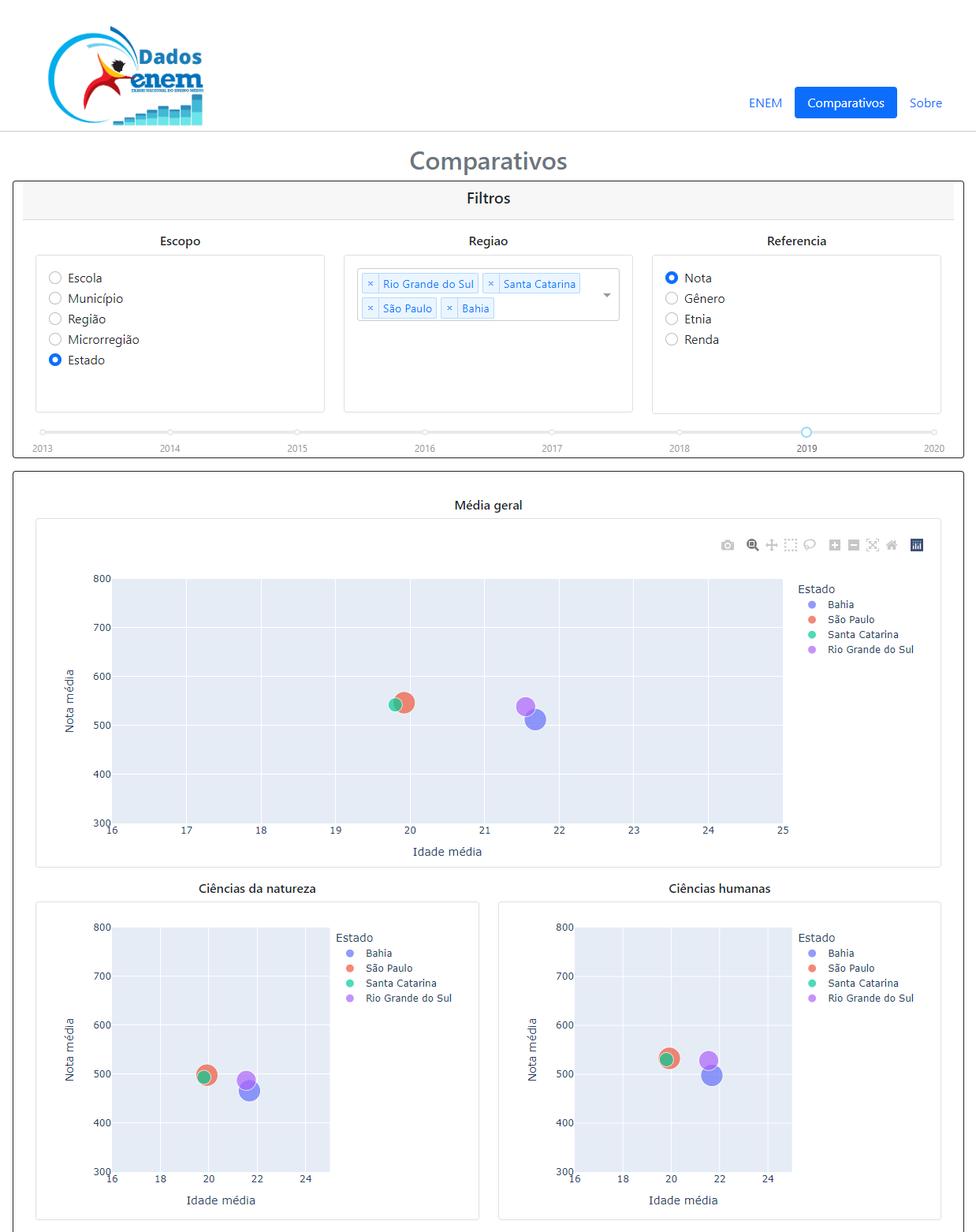
**Escopo**, define o escopo da comparação, podendo ser por Escola, Município, Microrregião, Região e Estado.

**Região**, consiste em um campo do tipo “*dropdow*” para seleção do que será comparado. A lista de opções é montada dinamicamente conforme a opção no Escopo.

**Referência**, define o que será comparado, e suas opções são: nota média, gênero, etnia e renda.

A Figura 21 exibe a tela de Comparativos completa. Ela é constituída dos filtros de comparação, um gráfico com a média aritmética geral e cinco gráficos menores com a média aritmética para cada área do conhecimento. Abaixo do gráfico de média geral existe um componente *slider* que permite alternar entre as edições do ENEM.

Figura 21 - DadosENEM - Comparativos.



Fonte: Elaborado pelo autor

Os gráficos ilustrados na Figura 21 são do tipo bolha. Eles são formados por:

Eixo Y: Nota média.

Eixo X: idade média dos participantes.

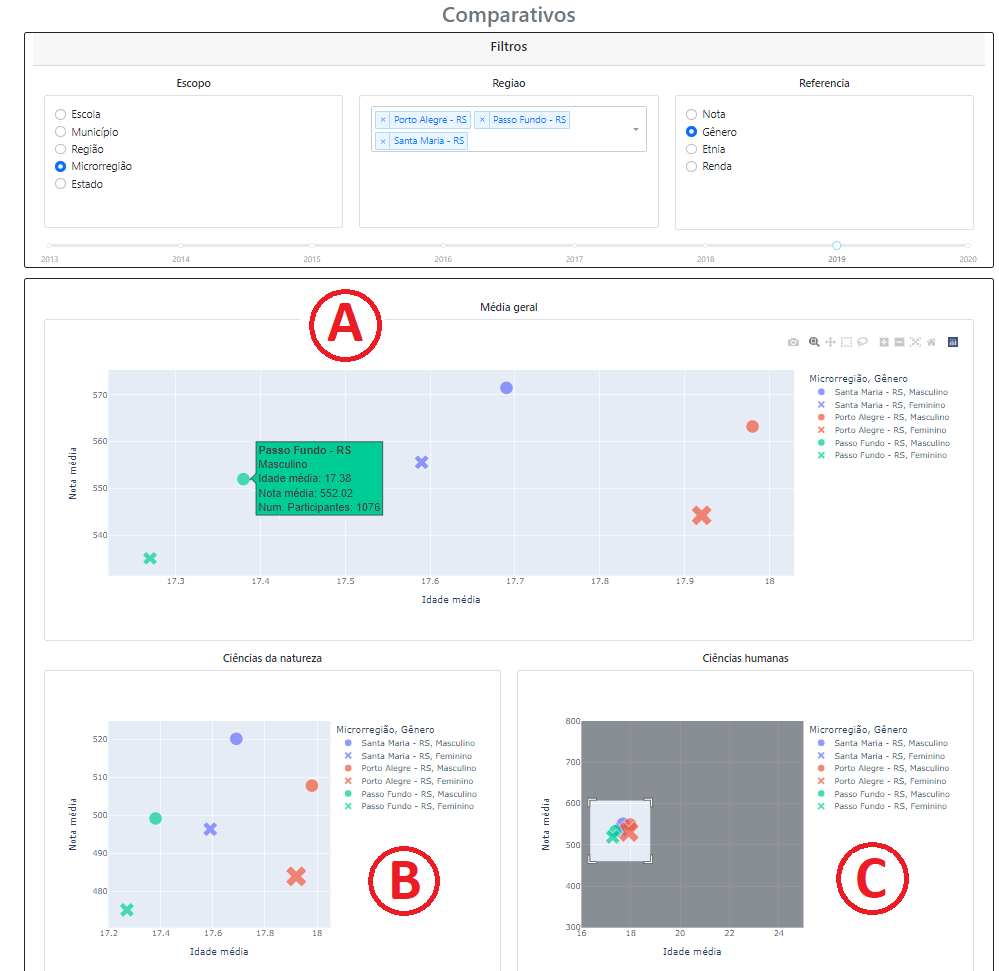
Cor: valores selecionados no escopo.

Tamanho da bolha: número de participantes na prova.

Formato: identificação de cada tipo de referência (ex. Gênero, quadrado representa masculino e círculo representa feminino).

A Figura 22 demonstra um exemplo de filtro: em escopo foi escolhido microrregião e em referência foi escolhido gênero. Cada microrregião é exibida duas vezes, mas com símbolos diferentes para representar cada gênero. Também é possível observar na Figura 22 destacado pela letra (A) que é possível interagir com os gráficos, pois ao mover o mouse sobre os elementos são exibidos mais detalhes sobre ele. Ao clicar na legenda lateral pode-se exibir ou ocultar elementos para tornar a visualização mais limpa. Também é possível contar com funções de zoom e movimentação como pode ser observado nos gráficos destacados pelas letras (B) e (C).

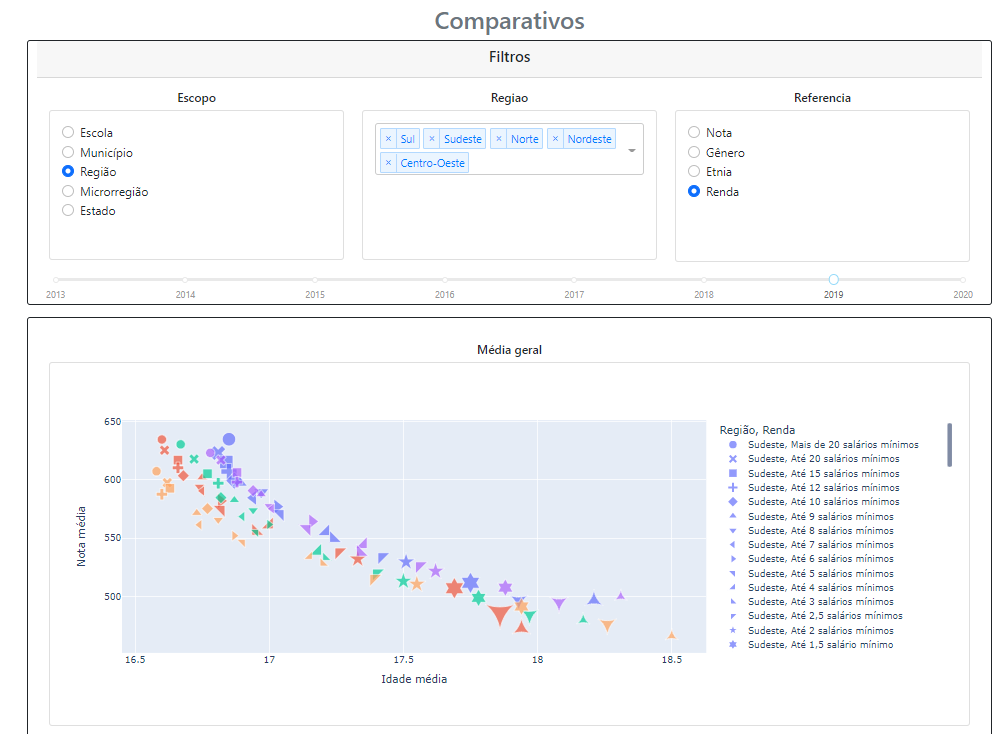
Figura 22 - DadosENEM - Comparativos por microrregião e gênero.



Fonte: Elaborado pelo autor

A Figura 23 exibe um exemplo de análise onde foi escolhido no escopo a região e na referência a renda familiar. A renda familiar é classificada em número de salários-mínimos e segue a mesma escala desde 2009. É importante salientar que a renda familiar é informada pelo participante no questionário socioeconômico, mesmo o INEP reforçando a responsabilidade do participante informar sua situação real, não é possível validar a veracidade das informações.

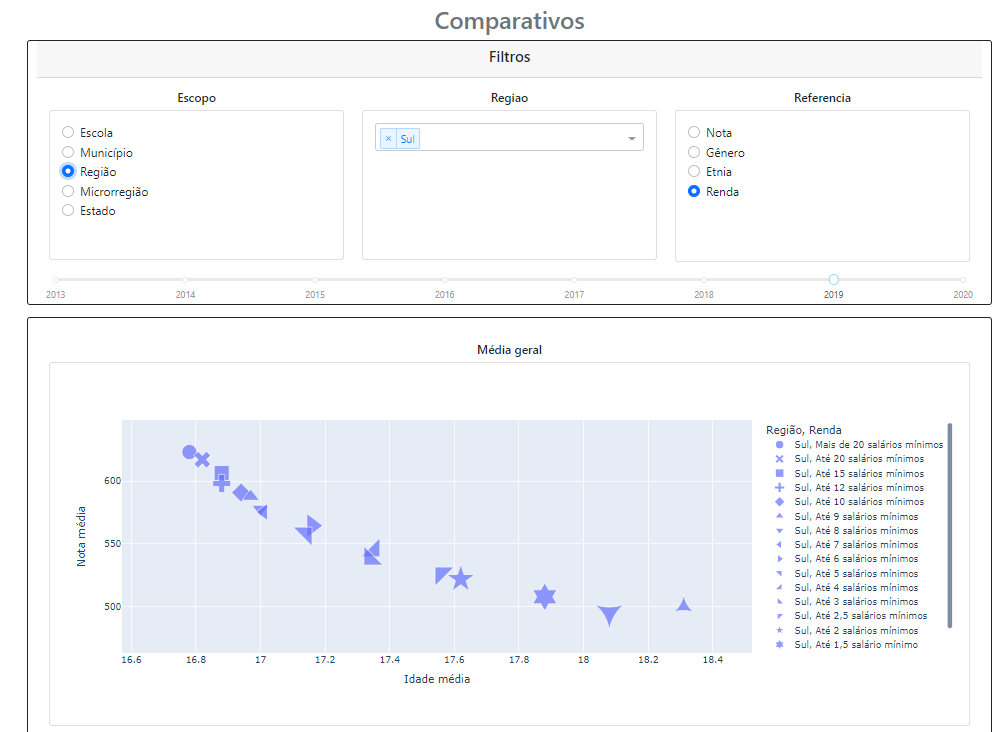
Figura 23 - DadosENEM - Comparativos por região e renda.



Fonte: Elaborado pelo autor

Nota-se na Figura 23 que o gráfico pode ficar confuso, a depender das opções selecionadas pelo usuário, já que o número de elementos selecionados no filtro Região é multiplicado pelo número de categorias existentes no tipo de Referência selecionado. Neste exemplo, a renda familiar possui 17 classificações diferentes. A Figura 24 demonstra o mesmo gráfico, porém analisando apenas a região sul, o que facilita o entendimento.

Figura 24 - DadosENEM - Comparativos por região e renda, apenas região Sul



Fonte: Elaborado pelo autor

Por fim, a tela “Sobre” contém informações sobre a plataforma, observações sobre os microdados e endereços de contato.

## Comparação com as ferramentas existentes

O Quadro 5 relaciona as características da plataforma DadosENEM, desenvolvida neste

trabalho, em comparação com as demais plataformas existentes conforme relacionado no Capítulo 3. Foram comparados o número de edições do ENEM disponível para análise, os tipos de gráficos, se é possível realizar consultas personalizadas ou se é apenas filtrar os dados, se é possível comparar múltiplos atributos (como número de participantes, nota média, gênero etc.), se é possível fazer análises com as questões do questionário socioeconômico, qual o escopo das análises (abrangência) e se existem gráficos que demostrem a evolução do ENEM ao longo dos anos.

Quadro 5 - Comparativo entre a plataforma DadosENEM e as ferramentas existentes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **DadosENEM** | **Resultados ENEM** | **Evolucional** | **QEdu** | **Painéis ENEM** |
| **Edições do ENEM** | 2013 a 2020 | 2009 a 2017 | 2016 a 2019 | 2014,2015,  2017,2018,  2019,2020 | 2015 a 2020 |
| **Tipos de gráficos** | Bolhas, Linhas, Barras, Mapa | Linha,  Barras | Linhas | Tabela | Barras, Linhas, Mapa |
| **Consultas personalizadas** | Sim | Não | Não | Sim | Não |
| **Comparações entre múltiplas atributos** | Sim | Não | Não | Sim | Não |
| **Comparações com questões do questionário socioeconômico** | Sim | Não | Não | Não | Não |
| **Escopo** | Escolas, Regiões, Microrregiões, Municípios e Estados | Escola | Escola | Escolas, Regiões, Municípios e Estados | Regiões |
| **Gráficos com a evolução do ENEM à nível nacional** | Sim | Não | Não | Não | Sim |

Fonte: Elaborado pelo autor

Percebe-se que a plataforma DadosENEM busca agregar as melhores características de cada ferramenta, desde o número de edições que estão disponíveis, o escopo amplo e a possibilidade de realizar consultas personalizadas conforme a necessidade dos usuários.

Este capítulo apresentou todo o processo de desenvolvimento da plataforma DadosENEM, desde a estruturação do banco de dados, o tratamento e importação dos microdados, prototipação das telas e tecnologias utilizadas. Também foram apresentadas imagens da plataforma finalizada descrevendo o funcionamento e os recursos disponíveis aos usuários.

# CONCLUSÃO

Neste trabalho foi realizada uma contextualização sobre o ENEM e sua evolução ao longo dos anos. Além disso, foi realizado o mapeamento das mudanças na estrutura dos microdados a partir de 2009, demonstrando que os microdados disponibilizados pelo INEP requerem um esforço adicional por parte dos pesquisadores ao analisar múltiplas edições.

Também foi feito o levantamento bibliográfico sobre estudos relacionados à visualização de dados do ENEM, ENADE e PISA, onde foram identificados pontos positivos e algumas limitações. Por exemplo, poucas das variáveis presentes nos microdados estão disponíveis para análise e há a possibilidade de investigar poucas ou apenas uma única edição do ENEM.

Com base nisso foi desenvolvido uma plataforma web para visualização dos microdados do ENEM, descrevendo o processo de desenvolvimento desde a estruturação do banco de dados, o tratamento e importação dos microdados, prototipação das telas e tecnologias utilizadas. A plataforma permite visualizar a evolução do ENEM entre 2013 e 2020 em reação a notas e números de participantes, além de permitir a comparações entre escolas, município, microrregiões e estados baseadas em nota, nota, gênero, etnia e renda.

Apesar de ter sido importado diversas variáveis, muitas delas não foram utilizadas. Como sugestão de trabalhos futuros, pode-se expandir o número de variáveis analisadas e os tipos de gráficos. Também verificar a viabilidade de um estudo mais detalhado sobre as questões objetivas e uma análise da relação entre número de acertos e erros em determinados tipos de questão, já que o INEP disponibiliza a prova completa, a resposta de cada participante em cada questão e o gabarito da prova.

Referências Bibliográficas

ALMEIDA FILHO, Álvaro Cavalcanti de. **Modelo de mensuração do desempenho dos institutos federais: uma análise a partir de microdados,** 2014.

ALVARES, Reinaldo Viana; DE SOUZA CAMPOS, Nathielly. **Gestão dos resultados do ENADE com uso de dados abertos e tecnologias digitais**. 2018. Revista EDUC-Faculdade de Duque de Caxias Vol. 05- Nº 1 Jan-Jun 2018

ALVAREZ, Luciana. **Análise de dados ganha espaço na gestão escolar**. Revista Educação, v. 242, 2017. Disponível em: <https://revistaeducacao.com.br/2017/09/19/analise-de-dados-ganha-espaco-na-gestao-escolar/>. Acesso em: 12 mar. 2021.

ALVES, Rafael Damiani. **Predição do desempenho da redação do ENEM utilizando técnicas de mineração de dados**. Trabalho de Conclusão de Curso (Graduação) - Curso de Tecnologias da Informação e Comunicação, Universidade Federal de Santa Catarina, Araranguá. 2018.

ARAÚJO, Rodrigo. **Análise dos microdados do Enade: proposta de uma ferramenta de exploração utilizando mineração de dados**. 2019. Dissertação de Mestrado em Ciência da Computação - Universidade Federal de Goiás.

EVOLUCIONAL. **Soluções educacionais para escolas de alta-performance.** Disponível em: <https://www.evolucional.com.br/>. Acesso em: 15 mai. 2021.

INEP. **ENEM - Microdados**. 2021a. Disponível em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados>. Acesso em: 13 mar. 2021

INEP. **ENEM**. 2021b. Disponível em: <http://portal.inep.gov.br/enem>. Acesso em: 13 mar. 2021

INEP. Microdados ENEM 2020 Leia-me. Fevereiro 2022. Disponível em: <site>. Acesso em: Abril/2022

INEP. **Painéis ENEM**. 2022. Disponível em: < <https://app.powerbi.com/view?r=eyJrIjoiYTdlOGQ3ZTgtMzc1Ny00ZDFkLTk4NjQtZDBkNTUyNjVhNmQ1IiwidCI6IjI2ZjczODk3LWM4YWMtNGIxZS05NzhmLWVhNGMwNzc0MzRiZiJ9>. Acesso em: 07 mai. 2022

INEP. **ENEM - Histórico.** 2021c. Disponível em: <http://portal.inep.gov.br/enem/historico >. Acesso em: 14 mar. 2021

INEP. **Matrizes de Referência**. 2021d. Disponível em: <http://portal.inep.gov.br/matriz-de-referencia >. Acesso em: 20 mar. 2021

INEP. **Entenda a sua nota no ENEM, Guia do Participante.** 2012. Disponível em: <https://download.inep.gov.br/educacao\_basica/enem/guia\_participante/2013/guia\_do\_participante\_notas.pdf> Acesso em: 22 mar. 2021.

KIRK, Andy; TIMMS, Simon; RININSLAND, Andrew; TELLER, Swizec. **Data Visualization: Representing Information on Modern Web**. Livery Place: Packt Publishing Ltd., 2016.

KNAFLIC, Cole Nussbaumer. **Storytelling com dados: Um guia sobre visualização**. Rio de Janeiro, RJ: Alta Books. 2019.

MAZZA, Riccardo. **Introduction to information visualization**. Springer Science & Business Media, 2009.

MELO, Katrine. **Um estudo voltado à transparência pública através da análise de dados abertos brasileiros**. 2017. Trabalho de Conclusão de Curso. Universidade Federal de Santa Catarina.

MICROSOFT, **O que é Power BI?.** 2022. Disponível em: <https://docs.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>. Acesso em: 20 mar. 2022.

NASCIMENTO, Marylaine Pereira do. **Visualização dos resultados das edições de 2010 a 2015 do ENEM através de um Shiny App**. 2017. Trabalho de Conclusão de Curso. Universidade Federal do Rio Grande do Norte.

OLIVEIRA, Elida. **Abstenção do Enem 2020 é de 55,3%**. 2021. Disponível em: <https://g1.globo.com/educacao/enem/2020/noticia/2021/01/24/abstencao-do-enem-2020-e-de-553percent-24-milhoes-foram-aos-locais-de-prova-neste-domingo.ghtml>. Acesso em: 11 jun. 2021.

PIRES, André. **Renda familiar e escolaridade dos pais: reflexões a partir dos microdados do ENEM 2012 do Estado de São Paulo**. ETD - Educação Temática Digital, Campinas, SP, v. 17, n. 3, nov. 2015. Disponível em: <http://periodicos.sbu.unicamp.br/ojs/index.php/etd/article/view/8638262>. Acesso em: 15 mai. 2021.

PLOTLY, **Dash Open Source**. 2022. Disponível em: <https://plotly.com/dash/open-source/>Acesso em: 25 mar. 2022.

POSTGRESQL, **PostgreSQL: About.** 2022. Disponível em: <<https://www.postgresql.org/about/>>. Acesso em: 06 abr. 2022.

POSTGRESQL, **Materialized Views.** 2021. Disponível em: <https://www.postgresql.org/docs/current/rules-materializedviews.html>. Acesso em: 25 mar. 2022a.

QEDU. **QEdu.** Disponível em: <https://www.qedu.org.br/> Acesso em: 18 mai. 2021.

RESULTADOSENEM. **ResultadosEnem.** Disponível em: <https://www.resultadosenem.com.br/>. Acesso em: 15 mai. 2021.

SANTOS, César Henrique Cardoso dos.; FREITAS, Maykon Carlos de. **VisdadosENEM: visualização de dados do ENEM com a biblioteca D3**. 2016. Trabalho de Curso. Universidade Federal de Santa Catarina

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business Intelligence e Análise de Dados para Gestão do Negócio-4**. Porto Alegre, RS: Bookman Editora. 2019.

SOARES, Marcelo. **Os eliminados na redação do Enem não são quem você imagina**. Disponível em: < https://medium.com/numeralha/os-eliminados-na-reda%C3%A7%C3%A3o-do-enem-n%C3%A3o-s%C3%A3o-quem-voc%C3%AA-imagina-47bc2acaa9e>. Acesso em: 12 mar. 2021.

TRAVITZKI, Rodrigo. **ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar**. 2013. Tese de Doutorado. Universidade de São Paulo.

VISDADOSENEM. **VisDadosEnem.** Disponível em: <http://labdata.sites.ufsc.br/visdadosenem/> Acesso em: 18 mai. 2021.

WARD, Matthew O.; GRINSTEIN, Georges; KEIM, Daniel. **Interactive Data Visualization Foundations Techniques and Applications**. 2nd Edition. Boca Raton, Florida: CRC Press, 2015

WINKELMANN, Rainer; BOES, Stefan. **Analysis of microd**ata. Springer Science & Business Media, 2006.

1. https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem [↑](#footnote-ref-2)
2. https://www.ibge.gov.br/explica/codigos-dos-municipios.php [↑](#footnote-ref-3)
3. https://www.python.org/ [↑](#footnote-ref-4)
4. https://pandas.pydata.org/ [↑](#footnote-ref-5)
5. https://plotly.com/dash/ [↑](#footnote-ref-6)