

Projeto Final de Visualização de Dados: O impacto da pandemia de COVID-19 no ENEM

Guilherme Moreira Castilho

Escola de Matemática Aplicada (EMAp)

Fundação Getulio Vargas (FGV)

Rio de Janeiro, Brasil

guilherme222castilho@gmail.com

Paulo César Gomes Rodrigues

Escola de Matemática Aplicada (EMAp)

Fundação Getulio Vargas (FGV)

Rio de Janeiro, Brasil

pcgomes137@gmail.com

Samuel Corrêa Lima

Escola de Matemática Aplicada (EMAp)

Fundação Getulio Vargas (FGV)

Rio de Janeiro, Brasil

samuelslmlr@gmail.com

Abstract—Este documento tem como objetivo documentar o trabalho final de Visualização de Dados do curso de Ciência de Dados e Inteligência Artificial da Escola de Matemática Aplicada da FGV (EMAp/FGV) com o tema "O impacto da Pandemia de COVID-19 no ENEM".

Index Terms—ENEM, pandemia de COVID-19, visualização de dados, desigualdade educacional, D3.js

I. INTRODUÇÃO

A pandemia de COVID-19 chegou ao Brasil no início de 2020 e, em pouco tempo, se espalhou por todo o país.

Um dos setores que mais sofreu impacto foi a educação, ampliando desigualdades já existentes e provocando mudanças estruturais no sistema de ensino.

As escolas públicas e privadas de todo o país interromperam as suas atividades presenciais e, como alternativa, foi implementado o sistema de ensino remoto de forma emergencial que evidenciou diversos problemas:

- **Infraestrutura precária:** Muitos alunos, especialmente da rede pública, não tinham acesso adequado à internet, computadores ou celulares, essenciais para o funcionamento do ensino remoto.
- **Falta de capacitação:** Muitos professores não estavam devidamente capacitados para ministrar aulas online de forma eficaz.

Com o aprendizado comprometido, houve um aumento muito grande na evasão escolar. Alunos deixaram de frequentar as aulas para trabalhar, cuidar de familiares ou até mesmo por não perceber sentido em continuar participando das aulas da maneira como estavam sendo ministradas.

O Exame Nacional do Ensino Médio (ENEM) é atualmente uma das principais formas de avaliação educacional do Brasil. Seu objetivo inicial era avaliar o desempenho dos concluintes do ensino médio, mas, com o tempo, consolidou-se como a principal porta de entrada para o ensino superior público (via SISU) e privado (via PROUNI e FIES). O exame é conhecido por avaliar as competências dos alunos em quatro grandes áreas: Linguagens, Ciências Humanas, Ciências da Natureza e Matemática, além da redação, que possui grande peso na nota final.

Dada sua relevância como instrumento de avaliação da educação brasileira, propomos fazer uma análise de como a

pandemia pode ou não ter influenciado em seus resultados, utilizando-se de dados de antes, durante e após a pandemia para desenvolver visualizações interativas.

II. METODOLOGIA

A. Dados

Para a produção das visualizações, utilizamos dos microdados do ENEM. Eles são conjuntos de dados públicos e detalhados divulgados anualmente pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), contendo informações individuais (anônimas) acerca de cada inscrito no exame de um determinado ano.

A fim de termos uma visão das estatísticas em um contexto pré-pandêmico, utilizamos os microdados do ENEM [1] do ano de 2019. Para um contexto durante a pandemia, utilizamos os microdados do ENEM dos anos de 2020, 2021 e 2022. E, por fim, para analisar a recuperação após a pandemia, utilizamos os microdados do ENEM dos anos de 2023.

Todas essas bases de dados podem ser encontradas no site do INEP.

B. Ferramentas

Para o desenvolvimento das visualizações interativas, foi utilizado o D3.js, biblioteca de JavaScript. Ela dispõe de manipulação direta do DOM com base de dados, criação dos mais diversos tipos de mapa, suporte a animações e transições. A ferramenta se destaca por suas diversas opções de customização. Para o tratamento das bases de dados, utilizou-se a biblioteca Pandas do python. Com ela, foi possível transformar arquivos naturalmente pesados — por conterem informações detalhadas de milhões de inscritos — em dataframes agregados e mais leves, garantindo maior fluidez às visualizações.

III. ESTRUTURA DO SITE

O site no qual o projeto é mantido é dividido em abas.

- **Introdução:** contém uma breve introdução acerca da motivação do projeto, seu desenvolvimento e uma explicação sobre o funcionamento de cada visualização;
- **Projeto:** contém um primeiro conjunto de visualizações interativas;

- Mapa de Bolhas: contém um segundo conjunto de visualizações interativas;
- Autores: contém informações acerca dos estudantes envolvidos no presente trabalho.

IV. VISUALIZAÇÕES

A. Mapa Seleccionável do Brasil

Consiste em um mapa do Brasil dividido por estados. O mesmo não tem como função transmitir algum dado acerca do ENEM, mas sim servir como filtro para as demais visualizações na página.

Clicando com o mouse sobre um estado no mapa, o mesmo passará a estar selecionado, com isso, os demais gráficos passarão a exibir dados somente dos inscritos daquela região. É permitido selecionar múltiplos estados.

Caso o usuário tenha selecionado múltiplas regiões e deseje voltar a visualizar gráficos sobre o país inteiro, basta o mesmo clicar no botão "Remover Filtros" na parte inferior do mapa.

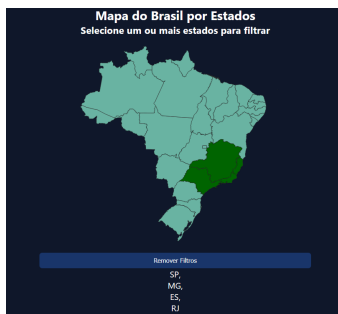


Fig. 1. Mapa do Brasil que serve como filtro

B. Gráfico de Barras Agrupadas e Gráfico de Fluxo Interativo

O gráfico de barras agrupado interativo é, a princípio, um gráfico de barras comum que representa o número de inscrições ocorridas ao longo dos 5 anos (2019 a 2023). Abaixo do título, há uma caixa com as seguintes opções:

- Cor ou raça;
- Estado civil;
- Faixa etária;
- Possui internet;
- Celulares na residência;
- Computadores na residência;
- Renda;
- Sexo;
- Tipo de Escola

Uma vez que uma dessas opções seja selecionada, o gráfico de barras será atualizado para um gráfico de barras agrupadas com as categorias da variável escolhida para cada ano.

Há também um gráfico de fluxo acoplado logo abaixo permitindo visualizar melhor a variação no número de inscritos para cada categoria da variável de forma mais fácil. O mesmo pode ser visualizado de três formas: Faixas unidas, faixas separadas com proporções entre faixas e com faixas separadas com proporções dentro das faixas.

Passando o mouse por qualquer parte do gráfico, surgirá uma caixa de texto dando informações mais detalhadas acerca da mesma.

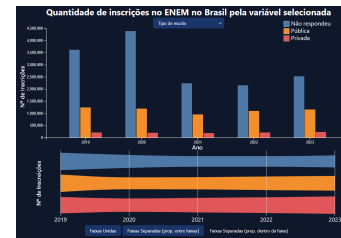


Fig. 2. Gráfico de barras agrupadas e gráfico de fluxo

Como mencionado anteriormente, ainda é possível utilizar do mapa mencionado anteriormente para filtrar os dados de ambos os gráficos de forma a representar somente as regiões selecionadas.

C. Mapa de calor interativo

Para visualizar o mapa de calor, é necessário que o usuário selecione uma variável para cada um dos eixos (X e Y). Uma vez que ambas forem selecionadas, um heatmap para cada ano surgirá. Eles mostrarão o número de inscritos no ENEM para cada combinação de variável no respectivo ano. As variáveis disponíveis para seleção são as mesmas do gráfico de barras agrupadas e gráfico de fluxo interativo.

Bem como os anteriores, selecionar regiões no mapa também influencia nos heatmaps.

Além disso, passando o mouse sobre qualquer um dos retângulos do heatmap, uma caixa de texto surgirá com informações mais detalhadas sobre a mesma.



Fig. 3. Mapa de calor

D. Mapa de Bolhas e Boxplot

Abaixo do título há um slider que varia de 2019 a 2023, permitindo a visualização de ambos os gráficos para o ano escolhido. Caso o usuário deseje é possível pressionar o botão de "Play" para que os gráficos se alterem automaticamente ao longo do tempo.

O mapa de bolhas consiste em um mapa do Brasil dividido por estados. O mesmo possui bolhas sobre cada estado, bolhas essas que representam a quantidade de participantes (pessoas que se inscreveram e compareceram às quatro provas) em cada estado.

Ao lado do mapa de bolhas há boxplots para cada estado, no qual cada ponto representa a média da taxa de presenças de cada município.

O mapa de bolhas permite que o usuário selecione um conjunto de estados para aparecer no boxplot, caso o usuário não deseje visualizar todos.



Fig. 4. Gráfico de bolhas e boxplot

V. DISCUSSÃO E RESULTADOS

Por meio das visualizações desenvolvidas, foi possível observar como a presença da pandemia de COVID-19 em determinados anos afetou os resultados da prova. O gráfico de barras e de fluxo, por exemplo, evidenciam uma clara queda no número de inscrições a partir de 2021, indicando que muitos alunos optaram por não fazer a prova durante esses anos.

No mapa de calor, é possível perceber que o padrão na proporção de inscritos para cada par de variáveis se mantém, porém, a partir de 2021, os tons do mapa de calor são mais escuros.

O mapa de bolhas e heatmap evidenciam um aumento grande no número de ausências no ano de 2020, sendo um dos anos com maior número de desistências da história do exame.

REFERENCES

- [1] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), “Microdados do Enem,” [Online]. Available: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>