PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

Painel Interativo para acompanhamento e análise de dados a respeito dos impactos do ensino remoto no Brasil

Ana Carolina Ferreira Junger

PROJETO FINAL DE GRADUAÇÃO

CENTRO TÉCNICO CIENTÍFICO - CTC

DEPARTAMENTO DE INFORMÁTICA

Curso de Graduação em Engenharia da Computação

Rio de Janeiro, novembro de 2021

Ana Carolina Ferreira Junger

Painel Interativo para acompanhamento e análise de dados a respeito dos impactos do ensino remoto no Brasil

Relatório de Projeto Final, apresentado ao Curso Engenharia de Computação da PUC-Rio como requisito parcial para a obtenção do título de Engenheira de Computação.

Prof.ª Simone Diniz Junqueira Barbosa
Orientadora
Departamento de Informática — PUC–Rio

Rio de Janeiro, novembro de 2021

Agradecimentos

À minha orientadora Simone Barbosa, pela oportunidade, apoio e por toda a colaboração e suporte ao longo desse ano atípico e conturbado;

À PUC-Rio, por me possibilitar experiências multidisciplinares;

A todos que responderam o questionário de testes com usuários, pois sem eles este trabalho não seria possível;

A todos os meus amigos, pelo suporte e acolhimento ao longo desses anos;

Aos meus amigos do VOA Educação, que me mostraram outras possibilidades dentro da computação com um propósito de melhorar o sistema de ensino do país;

Aos meus pais, Patricia e Paulo, por todo apoio e acolhimento e por sempre acreditarem em mim e nos meus sonhos;

Ao meu irmão, Henrique, por todos os livros emprestados e conselhos dados ao longo do curso;

Ao meu namorado, Miguel, por todo apoio e incentivo nessa reta final;

À minha avó, Lenir, que sempre me incentivou e que me possibilitou esse estudo;

Finalmente, ao meu avô, Roberto, por sempre acreditar em mim, valorizar cada conquista, incentivar todos os meus estudos e me mostrar que a educação é um dos bens mais valiosos que podemos ter.

Resumo

Junger, Ana Carolina. Barbosa, Simone. Painel Interativo para acompanhamento e análise de dados a respeito dos impactos do ensino remoto no Brasil. Rio de Janeiro, 2021. 77 p. Relatório Final de Projeto de Conclusão de Curso – Centro Técnico Científico – CTC, Departamento de Informática. Pontifícia Universidade Católica do Rio de Janeiro.

Nesse trabalho foi desenvolvido um painel interativo para acompanhamento e análise de dados a respeito dos impactos do ensino remoto no Brasil. Para isso, foram feitas análise e tratamento de dados do IBGE em R e posteriormente desenvolvida uma interface utilizando React e Ant Design Charts. Com os resultados foram feitos testes com usuários e a partir dos testes ajustes na interface.

Palavras-chaves:
Painel interativo; ensino remoto; Ant Design Charts; análise de dados;

Abstract

Junger, Ana Carolina. Barbosa, Simone. Interactive dashboard to monitor and analyze data about the impacts of remote education in Brazil. Rio de Janeiro, 2021. 77p. Relatório Final de Projeto de Conclusão de Curso – Centro Técnico Científico – CTC, Departamento de Informática. Pontifícia Universidade Católica do Rio de Janeiro.

In this paper, an interactive dashboard was developed to monitor and analyze data about the impacts of remote education in Brazil. In order to complete this work, IBGE data were used with R and lately an interface was developed using React and Ant Design Charts. With analysis results, tests were carried out with users, and from these tests, improvements were made to the interface.

Palavras-chaves:
Interactive Dashboard; Remote education; Ant Design Charts; Data analysis;

Sumário

Lista de Figuras

Lista de Tabelas

1 Introdução

No ano de 2020 a COVID-19 atingiu o Brasil e obrigou diversas áreas a se remodelarem para que pudessem se adaptar ao novo normal. Isso inclui todo o ramo da educação. Com o isolamento social vivemos um ano atípico de ensino remoto, em que os estudantes dependem do acesso à Internet, de computadores, celulares ou tablets para assistir às aulas.

Com isso, o debate a respeito da inclusão digital voltou a se popularizar. Se já existiam diferenças entre o ensino de uma escola particular de uma grande cidade e o ensino de uma escola pública da zona rural, com o ensino a distância elas ficaram cada vez maiores. De acordo com dados do IBGE, 74,4% da população brasileira tem acesso à Internet no domicílio. Comparando a zona urbana com a zona rural temos uma diferença exorbitante, de 79,4% na área urbana, e apenas 46,5% na rural [1].

Para além do acesso à Internet, a escola em muitos momentos foi refúgio dos estudantes. A dinâmica de estudar em casa é completamente diferente da presencial, principalmente para alunos que dividem equipamentos ou ambientes de estudo com outros familiares.

Dessa maneira, é de extrema importância entender como os estudantes foram impactados, não só academicamente, mas também emocionalmente com essa nova dinâmica de ensino.

Além disso, é necessário que essas informações estejam disponíveis e com livre acesso para a população, devido à sua relevância para a implantação de políticas públicas. Hoje estamos enfrentando um corte orçamentário no Censo Demográfico. Ele deveria ter sido realizado no ano de 2020, mas devido à pandemia, foi adiado para o ano de 2021. Entretanto, nesse ano sofreu cortes orçamentários, teve uma redução de mais de 95% do valor acordado e correu o risco de ser inviabilizado, mas acabou sendo adiado para 2022. O que aumenta a importância da disponibilidade desta informação [2] [3].

O objetivo deste projeto é construir um painel visual para acompanhamento e análise de dados a respeito dos impactos do ensino remoto no Brasil, no primeiro plano apenas para estudantes acima de 10 anos de idade, e fazendo um recorte regional, social e racial.

Os dados utilizados serão extraídos, principalmente, da Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), sobre o acesso à internet, equipamentos, como microcomputador, tablet e celular, utilizados para acessá-la e sobre os estudantes, recortando em instituição pública ou privada e em raça/cor.

Outros dados possíveis de serem utilizados são os do Censo Escolar, realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), sobre número de escolas, matrículas, sobre a migração para o sistema de ensino a distância, nível socioeconômico nas escolas e outros dados que vierem a ser interessantes. Ademais, podem ser cruzados com dados a respeito dos últimos ENEMs e outras pesquisas realizadas ao longo do último ano.

Ao longo do projeto será desenvolvida uma aplicação React, para acesso livre e apenas pelo computador, por se tratar de um painel interativo repleto de gráficos é de extrema importância que os acessos sejam feitos em telas maiores, para não perder nenhuma visualização.

Serão aplicados conhecimentos adquiridos ao longo do curso de interação humano–computador, análise e tratamento de dados, programação modular e banco de dados.

2 Situação Atual

Hoje, apesar de parte desses dados estarem acessíveis, não existe uma plataforma que faça o cruzamento das informações e exiba os valores em gráficos visuais ou em mapas. Temos, a respeito dos dados do Censo Escolar, uma plataforma criada pelo INEP para acessá-los de forma visual, mas sua interface é muito limitada por possuir apenas gráficos de barras, tabelas e um mapa que não transmite tanta informação.

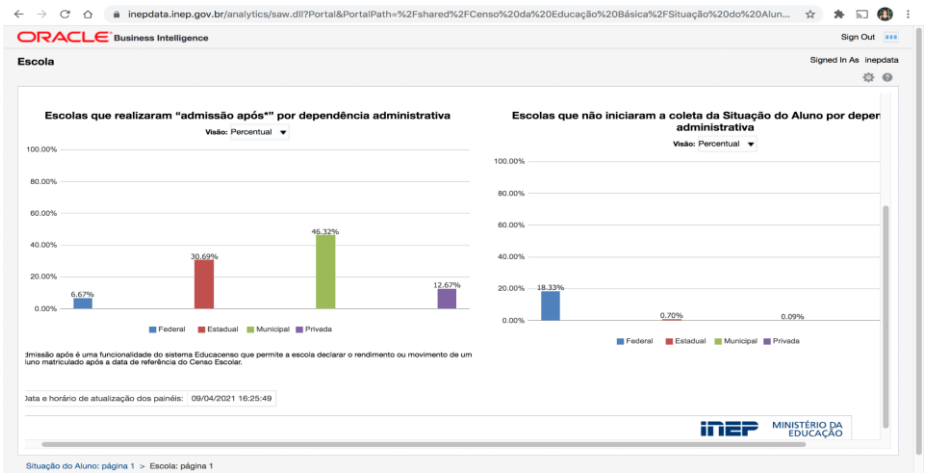

Figura 1: Tela de consulta do Censo Escolar sobre Escola

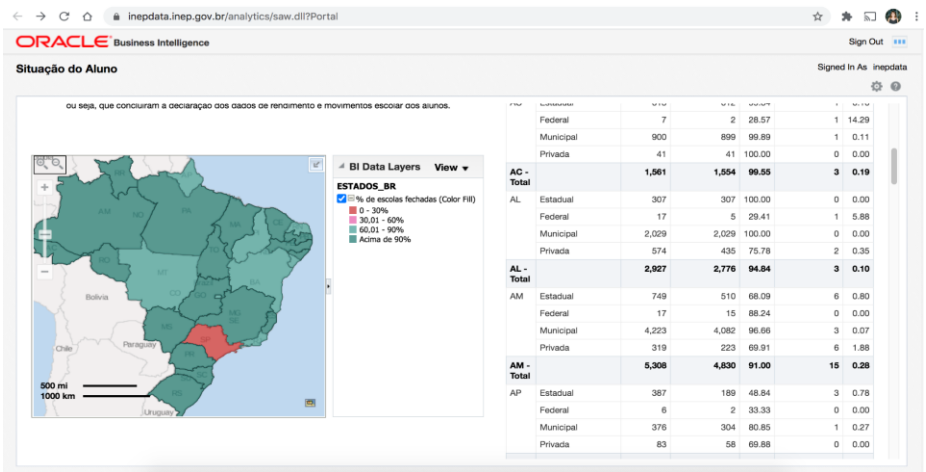

Figura 2: Tela de consulta do Censo Escolar sobre situação do aluno

Como outro exemplo, temos um banco de dados interativo criado pelo Centro de Políticas Sociais da Fundação Getúlio Vargas em parceria com a Telefônica a respeito de informações da inclusão digital que, apesar de ter uma gama de dados muito interessantes de serem avaliados, não tem muitas opções de visualização além de tabelas e gráficos de barras.

Além disso, na plataforma do PNAD, onde temos diversos dados, não temos nenhum ambiente de comparação entre eles, para visualizar algum é preciso entrar em sua respectiva página, aplicar os filtros desejados, se necessário, e partir daí é gerado, apenas, uma tabela.

3 Objetivos do trabalho

O objetivo desse projeto é carregar diferentes tipos de dados a respeito dos impactos do ensino remoto nos estudantes brasileiros e a partir daí construir um painel interativo com diferentes tipos de gráficos e visualizações. A ideia é poder aplicar diferentes filtros como de data, região que está sendo avaliada, renda per capita local, classes sociais, raça, gênero entre outros.

A intenção é que com as visualizações consigamos responder às seguintes perguntas:

- Quantos estudantes tiveram acesso a uma internet de qualidade para assistir às aulas?
    - Desses, quantos tinham acesso a equipamentos de qualidade como computador ou tablet e quantos tinham acesso apenas ao celular?
    - Esses equipamentos eram de uso único do aluno, ou compartilhados com outras pessoas?
- Quantos alunos tiveram aula ao vivo e quantos tinham que acompanhar um material gravado?
    - Desses, quantos tinham um ambiente de aprendizado online no formato *moodle* ou tinham algum ambiente que possibilitasse interação com o professor?
- Como o Índice Socioeconômico da região/estado afetou o ensino ao longo deste ano?
- A evasão escolar foi mais alta que nos anos anteriores?
    - Como a evasão neste ano se relaciona com gênero, raça/cor e nível socioeconômico?
- Como foi a adesão ao ENEM dos estudantes dos últimos anos, 2° e 3° do ensino médio?
    - Desses, o motivo pela não realização está relacionado ao não preparo adequado ou outros motivos?
    - Qual foi o desempenho comparativo?
- Houve uma mudança no olhar dos educadores para tecnologia em sala de aula?
- Entre outras perguntas que podem surgir com o desenvolvimento do trabalho.

A aplicação será desenvolvida em React, devido à sua facilidade de criação de interfaces de usuários. Ao longo do desenvolvimento serão estudados frameworks e bibliotecas para React que complementam o projeto e criam uma interface fácil, rápida e simples, para que usuários com diferentes conhecimentos de tecnologia e leitura de gráficos possam utilizar o dispositivo sem pormenores.

Será necessário, no mínimo, considerar dois tipos de usuários: um administrador, responsável por carregar as informações, com acesso direto e ilimitado ao banco de dados, e outro, que seria o usuário padrão, aquele que irá entrar no sistema para visualizar a informação desejada e explorar a aplicação.

Em comparação com as outras aplicações citadas, o sistema irá agrupar diferentes informações provenientes de diversas fontes, podendo fazer um cruzamento de dados completo e prover *insights* para o usuário. Além disso, serão estudadas bibliotecas mais completas para construção de uma interface mais agradável e diferentes tipos de visualizações para que seja entregue mais valor ao usuário.

4 Atividades Realizadas

4.1 Estudos preliminares

A aluna já tinha conhecimento prévio de JavaScript e React, devido a um estágio realizado entre 2019 e 2021 na startup VOA Educação, que foi incubada pelo Instituto Gênesis, incubadora de empresas da PUC-Rio, no qual desenvolvia, principalmente, um painel visual para acesso de escolas.

Apesar da recente experiência, por se tratar de um dos *frameworks* mais atuais do mercado, novas *toolchains (cadeia de ferramentas)* e bibliotecas foram criadas nesse meio tempo, por isso foi necessário estudar e mapear as diversas opções para definir quais caminhos seguir com o projeto.

4.2 Estudos da aplicação

Para construir uma aplicação React, o primeiro passo foi definir qual *toolchain* seria utilizada, ou se seria necessário criar uma *toolchain* do zero. Dentre as sugeridas pela documentação oficial do React, foram selecionadas duas para serem aprofundadas e avaliadas com maior detalhamento: Create-React-App e Next.js.

O Create-React-App é voltado para quem está querendo aprender React, com um ambiente simples e aprimorado para produção. Uma vantagem é que não tem restrições sobre quais bibliotecas podem ser utilizadas. É uma aplicação que utiliza Client-Side Rendering, ou seja, o projeto renderiza todo ao lado do cliente e depois é disponibilizado para visualização e interação. Uma desvantagem é que é difícil customizar a aplicação: para alterar o webpack, por exemplo, é necessário utilizar alguma ferramenta de terceiros.

Já o Next.js é muito mais completo e com funcionalidades embutidas, como o sistema de rotas e soluções para estilização. É uma aplicação Server-Side Rendering que gera sites estáticos, ou seja, a aplicação é carregada do lado do servidor, sendo muito mais rápida e eficiente. Também oferece facilidade na hora do *deployment* – a Vercel, empresa criadora do Next.js, oferece um ambiente completo para isso. E nele é possível configurar o *babel* ou webpack, ao contrário do Create-React-App. Além disso, ele fica responsável por todo o sistema de rotas, não necessitando de uma biblioteca terceira para isso [4].

Considerando as suas vantagens, decidimos usar o Next.js como *toolchain* nos testes iniciais.

Existem diversas opções de bibliotecas React para desenvolvimento de interface, cada uma delas oferece componentes pré-programados que auxiliam o desenvolvedor a construir a sua página com mais facilidade e simplificação.

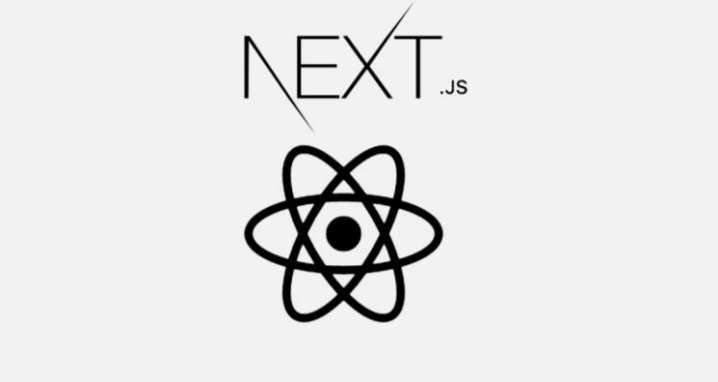

Figura 3: Arquitetura contendo apenas React e NextJs

4.3 Estudos e definição da biblioteca UI

Como, nesse projeto, o principal objetivo é a visualização de dados, é esperado que a biblioteca utilizada para desenvolver a interface tenha: disponibilidade de componentes como gráficos e mapas visuais ou fácil integração com outra biblioteca com esse foco; um alto desempenho; uma boa diversidade de componentes; documentação de qualidade; uma comunidade bem estabelecida; o reconhecimento no mercado; e compatibilidade com o NextJs. Serão avaliadas nesses aspectos quatro bibliotecas selecionadas previamente.

O Material UI é a biblioteca mais popular no GitHub, com mais de 72.000 estrelas (Dado extraído dia 12/11/2021). Tem um design simples, baseado nas especificações do Google's Material Design.[1] Tem diversos componentes disponíveis; apesar de não disponibilizar gráficos ou mapas visuais, é possível integrá-la com outras bibliotecas que cumprem esse serviço. É utilizada por empresas como: Quinto Andar, OpenClassrooms e Leroy Merlin.

Já o Ant Design possui uma vasta variedade de componentes disponíveis, totalizando mais de 60 opções, além de uma lista de opções de bibliotecas terceirizadas compatíveis para construção de gráficos e mapas. Apesar de ser de origem chinesa, sua documentação é bem completa e está disponível em inglês. Companhias como Alibaba, Tencent e Baidu a utilizam.

[1] https://material.io/design

O React Bootstrap é uma das bibliotecas para React mais antigas, construída a partir do antigo Bootstrap JavaScript. Também tem grande variedade de componentes, é facilmente customizável e responsivo. Tem um visual mais simples e não tem bibliotecas de gráficos.

Finalmente, a Evergreen é uma biblioteca bem avaliada, com mais de 11.000 estrelas no GitHub (Dado extraído dia 12/11/2021). Apesar de ter pouca variedade de componentes, é simples e intuitiva, mas também não possui bibliotecas de gráficos.

	Material UI	Ant Design	React Bootstrap	Evergreen
diversidade de componentes	alto	alto	alto	médio
componentes de visualização de dados	por terceiros	por terceiros	por terceiros	por terceiros
facilidade de implantação (código)	médio	alto	alto	alto
tamanho da comunidade	alto	médio	alto	alto
grande empresas usam	sim	sim	não	não
qualidade da documentação	alto	alto	médio	médio
estética visual	alto	alto	médio	médio
Responsividade	médio	alto	alto	baixo
conhecimento prévio	sim	sim	não	não

Tabela 1: Interfaces de UI

Dessa forma, pela maioria de pontos positivos, a biblioteca selecionada foi o Ant Design.

Apesar dos muitos pontos positivos, foi gasto um longo tempo na configuração do Ant Design com o Next.js. O Ant exige algumas configurações no *webpack* e de variáveis de ambiente, mas o Next.js faz todo esse trabalho previamente para o desenvolvedor. Sendo assim, foi necessário alterar algumas informações no *webpack*, criando o arquivo “next.config.js” na raiz do projeto [5].

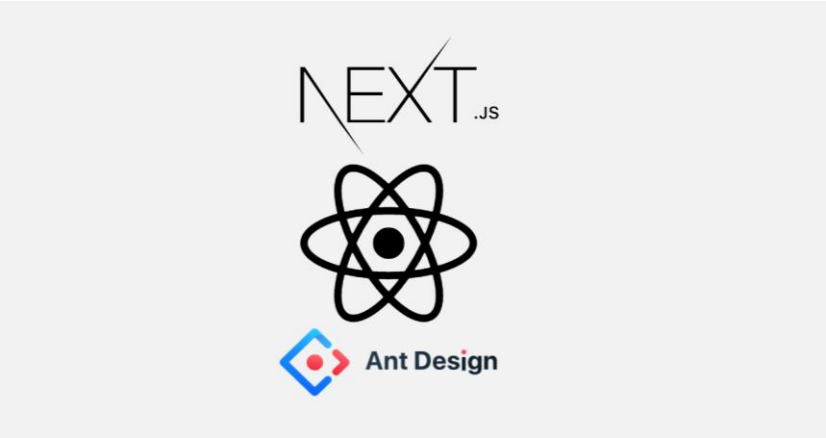

Figura 4: Arquitetura com a inclusão do Ant Design

4.4 Estudos e definição da biblioteca de visualização de dados

O próximo passo foi selecionar a biblioteca de visualização de dados. O AntD sugere algumas opções que interagem bem com ele, são elas: AntV Data Visualization, AntV Charting Library BizCharts, Recharts e Viser. Dessas, apenas AntV Charting Library e Recharts têm uma documentação completa em inglês; as outras são majoritariamente em chinês.

Na realidade o AntV Data Visualization não é apenas uma biblioteca de visualização, mas sim um time que reúne diversos frameworks com diversos modelos visuais, como gráficos, mapas físicos e de palavras, diagramas, portabilidade para *mobile*, com bibliotecas React Native, etc. Nele é sugerido o uso da Ant Design Charts, uma evolução da AntV Charting Library (g2 plot), baseada no React, facilitando o estilo de código. Ele possui uma gama enorme de gráficos, mas por ser uma biblioteca ainda nova, tem apenas 779 usuários, 565 estrelas no GitHub e +22.000 downloads semanais (Dados extraídos dia 24/04/2021).

A outra opção seria usar o Recharts, biblioteca famosa e consagrada, com +590.000 downloads semanais, mais de 16 mil estrelas no GitHub e mais de 45 mil usuários (dados extraídos dia 24/04/2021). Apesar de ter muitas opções de

gráficos, não é tão diverso quanto o Ant Design Charts. Além disso, a sua estrutura de código pode ser um pouco confusa nos primeiros usos.

Para testar o desempenho dessas duas bibliotecas foi renderizado um gráfico de linhas com dados definidos arbitrariamente. A Ant Design Charts renderizou em 19.5ms e a Recharts que levou 29.4ms. Dessa forma, foi decidido utilizar nos primeiros testes a Ant Design Charts por seu desempenho elevado e facilidade de código.

4.5 Estudos e definição do banco de dados

A respeito do *back-end*, o primeiro passo a se pensar é na estruturação do banco de dados. Como as informações presentes nele serão extraídas de diferentes plataformas, que não necessariamente têm uma relação direta entre os elementos, foi difícil pensar num modelo relacional de dados. Hoje, já temos no mercado modelos de bancos não relacionais, em que as informações são guardadas de uma forma não tabular, sendo muito mais flexível do que o modelo tradicional.

Além disso, bancos não relacionais são muito recomendados para uma quantidade grande e diversa de dados, como é nosso caso. E normalmente é mais performático, pois não tem que checar diversas tabelas para entregar uma informação; mais escalável; fácil de expansão por não se tratar de algo estático; coleta diferentes tipos de estrutura e podemos instalá-lo num sistema de nuvem.

Logo, decidimos que o protótipo iria utilizar um banco não relacional para os primeiros testes: o MongoDB.

O primeiro passo é criar um cluster no MongoDB Atlas, que é uma plataforma em nuvem para desenvolvimento de banco de dados. Nele podemos fazer todo o controle do banco, incluindo gerenciamento, como a construção de bancos de dados e coleções, todo o processo de inserção de dados e controle de permissão, entre outros recursos.

É necessário instanciar o banco em algum provedor, como Amazon Web Services, Azure ou Google Cloud Platform. O Atlas disponibiliza uma versão gratuita em que utiliza um cluster com RAM compartilhada e com baixo tráfego de dados. Como o projeto não se trata de uma aplicação pesada, foi escolhido para o protótipo o modelo mais simples, M0 SandBox, que dispõe de 100 conexões, 100 bancos e 500 coleções ao máximo. O banco está hospedado em

São Paulo pela Google Cloud Platform. Caso seja necessário um upgrade, o processo é simples e está descrito na documentação oficial do Atlas[2].

Para conectar a aplicação ao banco é necessário definir a *string* de conexão e o(s) nome(s) do(s) banco(s) nas variáveis de ambiente do projeto, importar o MongoClient e estabelecer a conexão. Neste código é definida uma variável global como a cache da conexão, para prevenir que as conexões cresçam exponencialmente, principalmente com os *hot reloads* de desenvolvimento. Para fazer uma consulta ou inserir um dado no banco é muito simples, basta instanciar a conexão com o banco, definir a coleção em questão e aplicar o método de *find* ou *insert*. O código abaixo define a conexão e foi extraído de um vídeo tutorial oficial da MongoDB [6].

```
import { MongoClient } from 'mongodb'

const { MONGODB_URI, MONGODB_DB } = process.env

if (!MONGODB_URI) {
  throw new Error(
    'Please define the MONGODB_URI environment variable inside .env.local'
  )
}

if (!MONGODB_DB) {
  throw new Error(
    'Please define the MONGODB_DB environment variable inside .env.local'
  )
}

/**
 * Global is used here to maintain a cached connection across hot reloads
 * in development. This prevents connections growing exponentially
 * during API Route usage.
 */
let cached = global.mongo

if (!cached) {
  cached = global.mongo = { conn: null, promise: null }
}

export async function connectToDatabase() {
  if (cached.conn) {
    return cached.conn
  }

  if (!cached.promise) {
    const opts = {
      useNewUrlParser: true,
      useUnifiedTopology: true,
    }

    cached.promise = MongoClient.connect(MONGODB_URI, opts).then((client) => {
      return {
        client,
        db: client.db(MONGODB_DB),
      }
    })
  }
  cached.conn = await cached.promise
  return cached.conn
}
```

Figura 5: Trecho de código relativo a conexão no banco de dados

---

[2] https://docs.atlas.mongodb.com/scale-cluster/#std-label-scale-cluster-instance Último acesso em: 30/06/2021

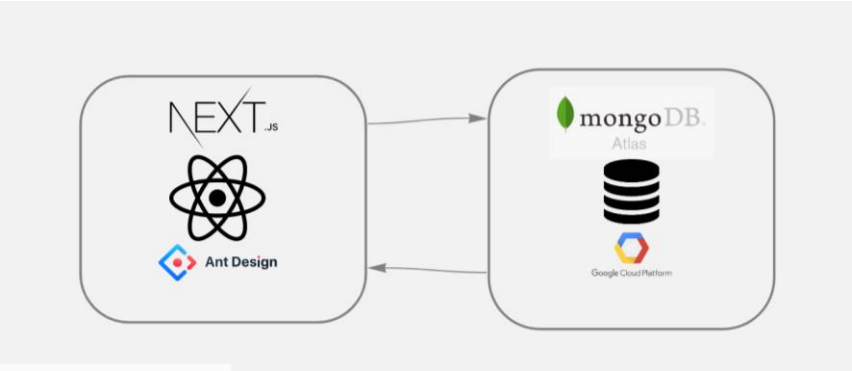

Figura 6: Arquitetura com a inclusão do Banco de Dados

4.6 Testes e Protótipos para aprendizado e demonstração

Com a definição das bibliotecas de *front-end* e *back-end*, o próximo passo foi definir quais dados serão exibidos no primeiro protótipo, e qual informação seria transmitida através deles. Como o objetivo é fazer um recorte atual, não faz muito sentido extrair informações do Censo Demográfico, visto que o último é de 2010, então as informações utilizadas, majoritariamente, serão extraídas de pesquisas realizadas ao longo do último ano.

Nesse sentido, foi estudado o "*Retratos da Educação no Contexto da Pandemia do Coronavírus - Um olhar sobre múltiplas desigualdades*" (LIMA, 2021) [7], publicação da Conhecimento Social, uma consultoria especializada na produção de conhecimento no campo social, que reúne informações de pesquisas da Fundação Carlos Chagas, Fundação Roberto Marinho, Fundação Lemann, Itaú Social, Instituto Península e Iede, feitos no ano de 2020.

Alguns dos dados interessantes selecionados para exibição foram:

- Acesso a equipamentos tecnológicos e internet por região do país, para avaliar o quantitativo de alunos que não tiveram acesso ou tiveram um acesso muito precário às aulas;
- Evasão escolar nos últimos anos, para entender se a educação remota comprometeu para a evasão escolar, fazendo um comparativo entre rede pública e privada de ensino;
- Níveis de aprendizagem adequada entre educação pública e privada num recorte de cor/raça;
- O que nesse novo modelo afetou os professores, baseado numa pesquisa do Instituto Península;

- Grau de contato que os professores mantiveram com os alunos durante esse período, fazendo um recorte por região e rede de ensino;
- Equipamentos usados para acessar a internet em casa durante o isolamento social, pesquisa feita pelo Conselho Nacional de Juventude (CONJUVE);
- Proporção de estudantes do Ensino Médio que informam ter atividades ou materiais oferecidos ou indicados pela escola, pesquisa do CONJUVE;
- Intenção de prestar o ENEM por quantos de fato fizeram e comparar com os anos anteriores;
- Comparativo entre as visões dos educadores sobre tecnologia em sala de aula, antes e depois da pandemia, pesquisa do Instituto Península;
- Matrículas no Ensino médio por região no ano de 2020, no recorte de gênero e localidade (urbana ou rural), para entender quantos estudantes foram impactados com o ensino remoto neste ano;
- Indicador de Nível Socioeconômico das Escolas de Educação Básica por região no ano de 2019, para poder comparar o nível socioeconômico entre as regiões do país;

4.6.1 Seleção dos dados

Decidimos selecionar os dois últimos para serem exemplificados no protótipo. Para isso, fizemos estudos nas principais bases do país. Esses dois últimos estão disponíveis no site do INEP.

4.6.2 Importação dos dados

Para a importação desses dados, construímos uma interface, bem básica: o modelo tem apenas duas caixas de input, uma para definir o nome da coleção e outra para o dado no formato JSON.

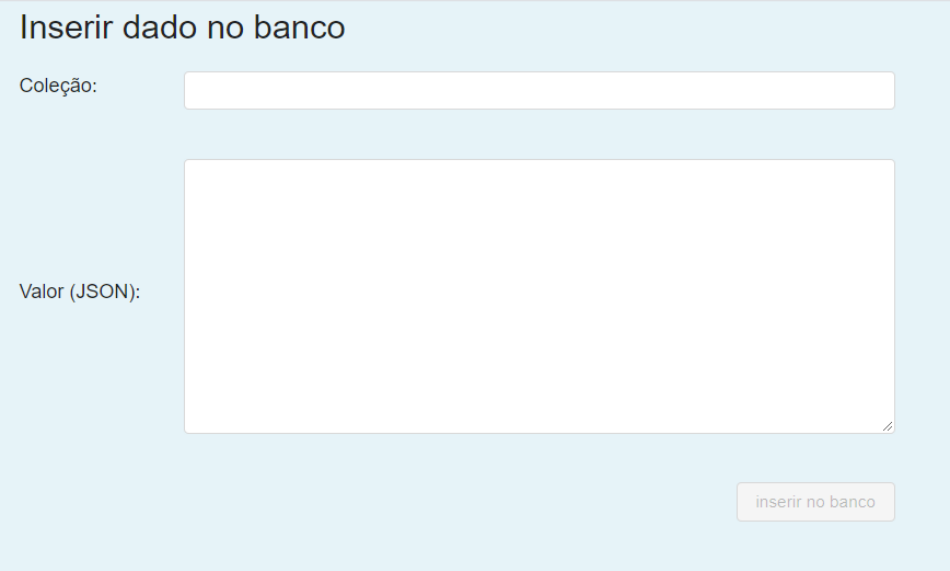

Figura 7: Interface de inserção de dados no banco

O *input* da “Coleção” é um *input* simples, uma única linha de texto, e o do “Valor (JSON)” é uma *TextArea,* que permite a inserção de textos com mais linhas. O botão “inserir no banco" é habilitado quando as duas caixas de textos são preenchidas corretamente, isto é, se a “Coleção” contém algum texto e se o “Valor (JSON)” contém uma string no formato JSON. O código abaixo realiza essa verificação.

```
function IsJsonString(str) {
  try {
      JSON.parse(str);
  } catch (e) {
      return false;
  }
  return true;
}

function validate(collection, document){
  if(collection && collection !== '' && IsJsonString(document)){
    return true
  }
  return false
}
```

Figura 8: Código de validação de uma string no formato JSON

Quando o usuário preencher as duas caixas de texto corretamente e clicar no botão de inserção de dados, é feita uma chamada para o banco para que o documento seja inserido corretamente. Depois disso, o usuário recebe um *feedback* no formato de mensagem de sucesso ou erro, caso tenha ocorrido algum erro no processo.

Num segundo momento, para obter a visualização descrita na próxima seção, foi necessário importar mais de 5.500 dados, e portanto não era viável fazer isso por esta interface. Para realizar a importação, como os dados estavam numa tabela de Excel e o MongoDB recebe um JSON, precisaríamos apenas converter a tabela para JSON e realizar uma inserção de múltiplos dados. Para isso, foi utilizada a biblioteca *convert-excel-to-json*[3] , que recebe um arquivo no formato xlsx e retorna no formato JSON.

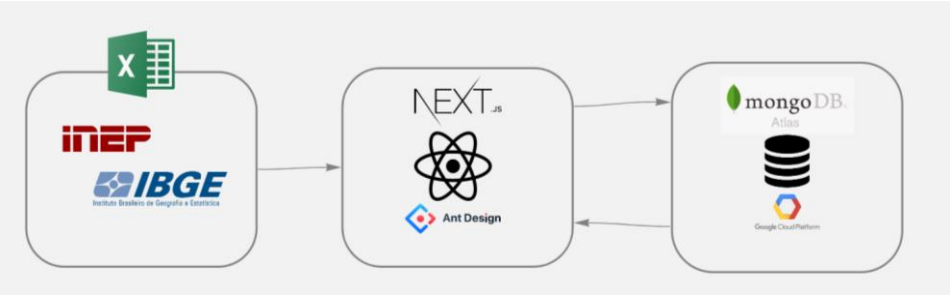

Figura 9: Arquitetura do protótipo

4.6.3 Visualização em Colunas

Para exemplificar os dados relativos às matrículas no Ensino médio por região no ano de 2020, foi utilizado um gráfico de colunas para entendimento da distribuição das informações por região do país. Um fator interessante nesta visualização é que podemos ver os dados brutos na forma de *tooltip*, basta passar o mouse pela região que deseja aprofundar.

Nesta visualização, foram feitos dois recortes: raça/cor dos alunos e a localidade onde residiam, urbana ou rural. No primeiro momento, as duas informações foram exibidas em um único gráfico, em duas colunas separadas para cada região, mas essa visualização foi revista pensando no caso de o leitor tender a fazer associações indevidas entre as colunas que, apesar de se tratarem do mesmo tema, não se relacionam entre si. Além disso, foi adicionado o total de estudantes para cada região no topo da coluna. Na imagem abaixo podemos ver o resultado final da visualização.

[3] https://www.npmjs.com/package/convert-excel-to-json Acesso em: 29/06/2021

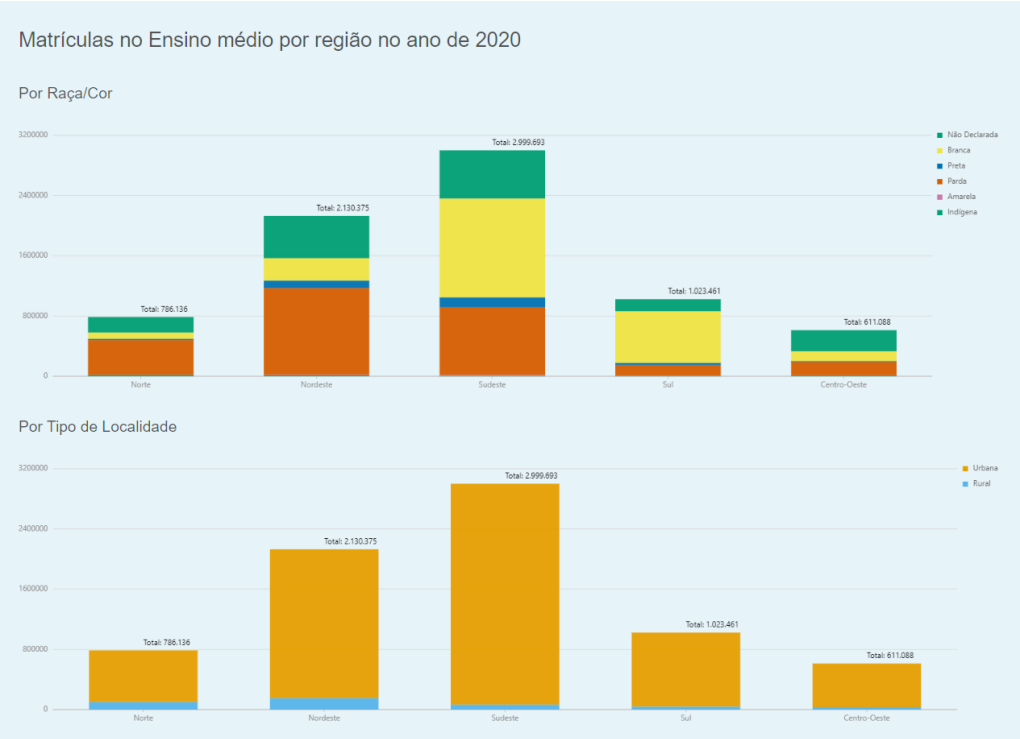

Figura 10: Gráfico de colunas relativo às matrículas no Ensino médio em 2020

Analisando os gráficos, é possível observar que o número de estudantes Indígenas e Amarelos é muito menor do que as outras raças/cor, e que temos uma maioria de estudantes Brancos, Pardos ou com raça/cor não declarada para todas as regiões. Além disso, vemos uma predominância de Pardos na região Norte, em comparação com a predominância de Brancos no Sul. E sobre as localidades, urbana e rural, podemos perceber que a maioria dos estudantes é da zona urbana.

4.6.4 Visualização *Violin-Plot*

O Indicador de Nível Socioeconômico das Escolas (INSE) é uma medida cujo objetivo é situar os alunos em um grupo, definido pela posse de bens domésticos, renda e contratação de serviços pela família dos alunos e pelo nível de escolaridade de seus pais [8]. Os dados coletados foram do INSE de 2019 e tratados para termos a visão por região do país.

O *Violin-plot* utilizado é uma combinação do *Box-plot* com o traço de densidade. Esse modelo de gráfico foi escolhido, pois temos uma gama muito grande de pontos a serem visualizados (aproximadamente 5.500) e com ele conseguimos extrair informações preciosas como o máximo, mínimo, mediana e

quartis, mas, além disso, temos a visão de como a distribuição dos valores se dá [9].

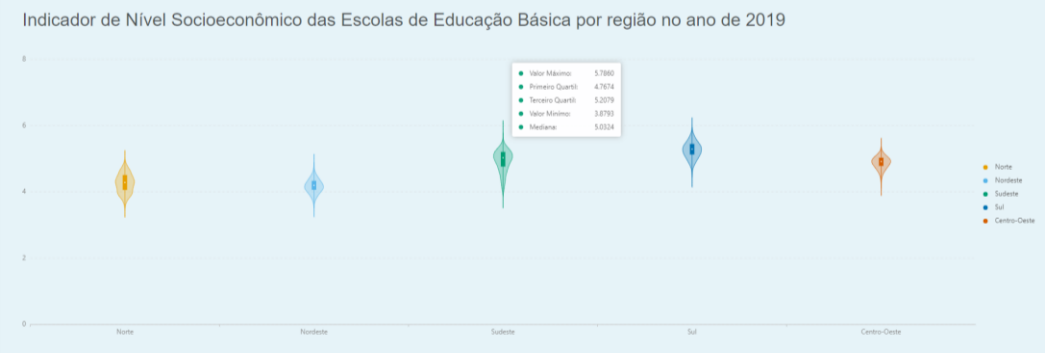

Figura 11: Violin-Plot relativo ao INSE por região em 2019

Analisando o gráfico, na região Norte, temos os menores índices, mas também temos uma distribuição mais uniforme entre os valores. Na região Nordeste, os valores estão mais concentrados em torno da média, entre o maior e menor valor. Já na região Sudeste, temos uma maior concentração de valores mais elevados, mas ainda temos valores baixos do indicador, podemos perceber pelo seu afunilamento. E, por fim, o Sul e o Centro-Oeste têm um comportamento similar ao do Nordeste, mas com valores mais elevados.

4.6.5 Acesso às visualizações

Para acessar as visualizações no protótipo, como ainda não temos um layout definido, foram colocados na página inicial dois botões, que fazem o redirecionamento para a visualização em questão. Posteriormente seria criado um layout possibilitando uma navegação mais interessante.

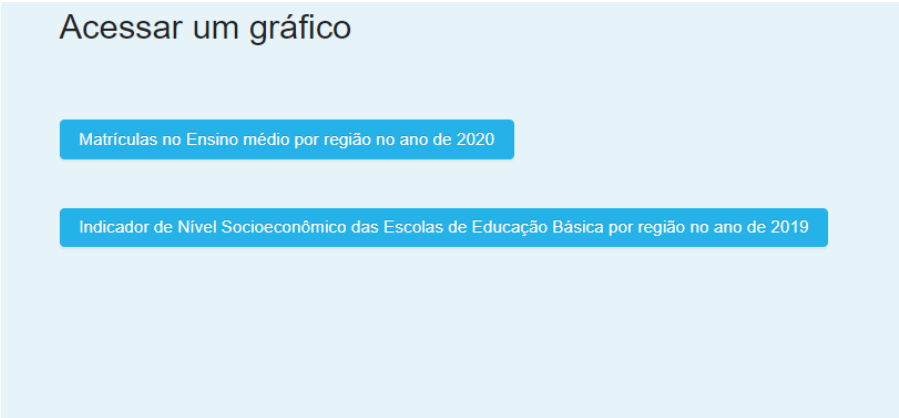

Figura 12: Acesso aos gráficos

4.7 Método

A principal tarefa, e talvez a mais difícil, é encontrar os dados que respondam as perguntas determinadas para as visualizações. Para atingir esse objetivo foi criada uma tabela no *Notion* com as seguintes colunas: ‘Pergunta’, a ser respondida com a visualização, ‘Dados necessários’ para responder a essa pergunta, ‘Fonte de dados encontrada?’, ‘Dado’, lista com o título do dado na tabela de ‘Documentos’, ‘Dados tratados?', ‘Tipo de dado’, numérico, categórico, temporal e ‘Tipo de visualização’. Já a tabela documentos contém o ‘Título’ do dado, ‘Fonte’ e o caminho onde o arquivo se encontra no computador, para garantir um *backup* caso haja alterações na fonte.

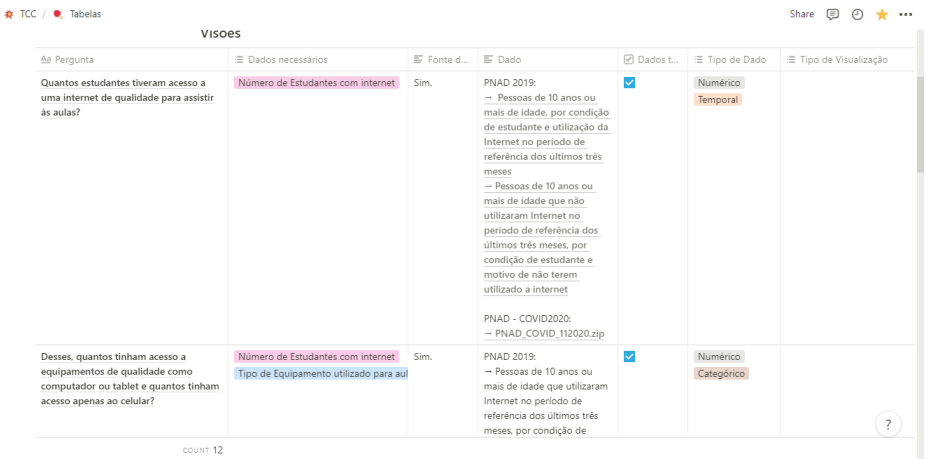

Figura 13: Organização dos dados

A principal fonte dos dados foi a Pesquisa Nacional por Amostra de Domicílios Contínua - IBGE (PNAD - Contínua), que visa acompanhar as flutuações trimestrais e a evolução, no curto, médio e longo prazos de informações necessárias para o estudo do desenvolvimento socioeconômico do país, produzindo indicadores anuais sobre temas como tecnologia da informação. A partir daí, é possível fazer um recorte regional e filtrar pela condição de estudante. Os dados de 2016 a 2019 já estão tratados e reduzidos a indicadores e os de 2020, edição especial devido a COVID-19, estão sem qualquer tratamento nos dando mais possibilidades de extrair informações interessantes.

Depois de encontrarmos os dados, iniciamos um processo de tratamento e análise para chegarmos aos resultados desejados. Essa parte foi iniciada por uma fase de estudos, visto o pouco conhecimento em Data Science, antes do desenvolvimento desse projeto, seguido por, de fato, os tratamentos e análises.

Por fim, começamos o processo de estudo de visualizações e o desenvolvimento da interface. Para as visualizações o primeiro passo era definir qual pergunta queríamos responder com ela e, a partir, daí buscar o melhor gráfico para passarmos essa informação. E, também, a melhor maneira de posicionar os dados no gráfico.

4.7.1 Cronograma

O cronograma do Projeto Final 1, apresentado na tabela abaixo, foi todo realizado; incluindo as tarefas que ficaram pendentes dele para o projeto dois.

Tarefa	Data de início oficial	Data de fim oficial	Status
Configurar AntD	10/04	17/04	
Estudo e Definição da Biblioteca de Visualização de dados	23/04	24/04	
Pesquisa sobre o melhor banco de dados	9/5	04/06	
Configuração do Banco de Dados e Backend	20/05	04/06	
Estudo de análise de dados	12/06	27/06	
Estudo da importação de dados para o banco	29/06	-	Finalizada no projeto final 2
Finalização do relatório final 1	21/06	30/06	

Tabela 2: Cronograma Projeto Final 1

Durante o Projeto Final 1, as tarefas para o Projeto Final 2 foram definidas por meses em vez de prazos diários, por na época não termos uma definição exata do escopo de cada tarefa e muitas delas poderem ser executadas em paralelo, como a de Estudo de análise de dados e Definição dos dados a serem usados.

Tarefa	JUL	AGO	SET	OUT	NOV
Estudo de análise de dados	X	X			

Tarefa	JUL	AGO	SET	OUT	NOV
Definição dos dados a serem usados	X	X			
Criação de um script de importação		X			
Desenvolvimento da interface			X	X	
Preparações para o deployment				X	
Ajustes finais				X	X
Finalização do Relatório Final 2					X

Tabela 3: Cronograma do Projeto Final 2

Como muitas tarefas foram realizadas paralelamente, seguimos esse modelo de meses ao invés de prazos diários. Segue abaixo o cronograma realizado. Foram incluídas tarefas como “Buscar dados a serem usados” e “Testes com usuários”, que não estavam previstos no fim do desenvolvimento do projeto um, mas que foram extremamente necessárias para o Projeto Final 2.

Tarefa	JUL	AGO	SET	OUT	NOV
Estudo de análise de dados	X	X	X	X	
Definição dos dados a serem usados	X	X			
Buscar dados a serem usados		X	X	X	
Criação de um script de importação			X		
Desenvolvimento da interface			X	X	
Preparações para o deployment				X	
Testes com usuários				X	X

Ajustes finais					X
Finalização do Relatório Final 2					X

Tabela 4: Cronograma realizado

Os prazos também foram alterados: a tarefa de “Estudo de análise de dados” perdurou praticamente todo o tempo do projeto, pois com as construções e ajustes nas visualizações, tínhamos muitas vezes que voltar a estudar os dados e mudar alguma coisa em sua análise, sendo uma tarefa retroativa.

Já a tarefa de “Buscar dados a serem usados” foi a mais desafiadora deste projeto. Como estávamos tratando de um tema bastante atual, e hoje há um baixo investimento em pesquisas no Brasil, tivemos acesso a poucas informações relevantes, pois a maioria dos dados era de anos muito anteriores ou não tínhamos acesso aos microdados, apenas a indicadores já tratados, o que também limitava a nossa possibilidade de explorá-los da maneira que gostaríamos.

A tarefa de “Criação de um script de importação” foi ligeiramente adaptada. Ao invés de criar um script de fato, como os nossos dados finais do tratamento em R vinham no formato JSON, apenas adaptamos a função de inserir dados no banco para aceitar vários dados de uma vez só. Essa função atua de forma recursiva, dividindo os dados em blocos de 500 por requisição, para não tomarmos *time out*, e a cada inserção de um bloco, retorna uma mensagem ao usuário, até que todos os dados estejam no banco.

O desenvolvimento da interface durou realmente o tempo planejado, pois a aluna já tinha experiência prévia nessa área, logo o desenvolvimento ocorreu de forma fluída. E a tarefa do deployment foi bem rápida, pois já havíamos feito *builds* locais quando inserimos uma nova dependência no projeto, garantindo seu funcionamento.

Os testes com usuários, apesar de não terem sido previstos, foram essenciais para avaliar o que havia sido desenvolvido. Esta tarefa foi dividida em três partes: o desenvolvimento do questionário, a divulgação do mesmo e os ajustes a partir dos resultados.

5 Projeto e especificações dos sistemas

5.1 Análise e tratamento de dados

Para tratar os dados, no primeiro momento, foram feitas análises em Python, devido ao conhecimento prévio da linguagem, com a biblioteca Numpy, devido ao objeto próprio *narray* e as suas diversas funções para tratamento e processamento do mesmo. E Pandas, pelas vastas alternativas de manipular e visualizar *data frames*. Tudo foi feito no Jupyter Notebook em função da sua facilidade e agilidade de testes e visualizações. Com essas análises foram obtidos *insights* iniciais.

A partir daí foram inseridos os dados brutos no banco de dados e construída as primeiras visualizações. A página fazia um *request* para o banco que devolvia um *array* com todos os dados e os filtros foram aplicados diretamente no *front-end*, o que causava lentidão na renderização da página e acabava deixando o código muito desorganizado. Por isso, foi definido que nas próximas visualizações os dados seriam inseridos já tratados no banco de dados.

Por conta de certas dificuldades em extrair o dado da maneira desejada foi decidido fazer alguns experimentos em R, devido à sua grande popularidade no meio de Data Science, e por causa da sua grande facilidade de manipulação de dados foi decidido seguir com essa linguagem para tratar as próximas visualizações.

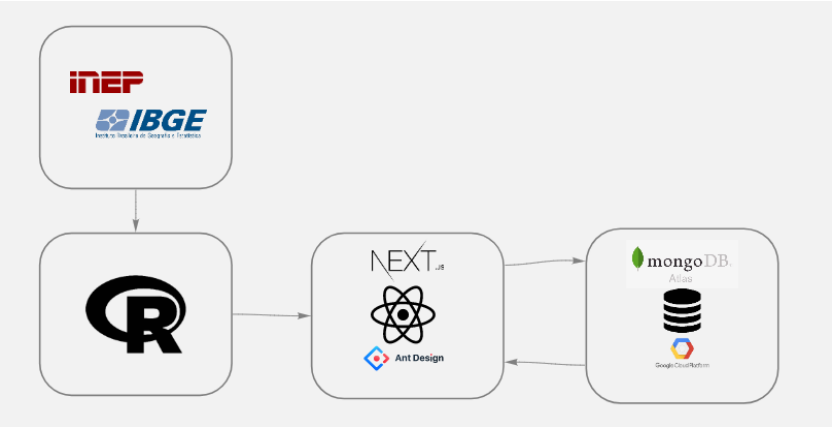

Figura 14: Arquitetura do Projeto

Os primeiros scripts em R foram feitos para tratar os dados da PNAD - contínua entre 2016 e 2019. Por se tratarem de dados numéricos e categóricos

com variância no tempo, ao invés de apenas organizar os dados e aplicar os filtros, os resultados brutos foram transformados em dados percentuais para poder fazer uma comparação fiel no tempo.

Como essas tabelas seguiam o mesmo modelo, os scripts acabaram sendo muito parecidos. Os passos gerais foram os seguintes:

1. Ler o arquivo csv no *data frame* a partir da terceira linha, primeira linha útil do arquivo;
2. Utilizar algumas funções básicas para entender o documento como *names*, que exibe o nome das colunas, *head*, que exibe as primeiras cinco linhas e *summary* que faz um resumo do *data frame*. Para dados numéricos conseguimos analisar o valor mínimo e máximo, mediana, média e o primeiro e terceiro quartil;
3. Converter todos os “-” (zero absoluto) para zero de fato;
4. Remover todas as colunas sem valor (N/A);
5. Verificar se realmente não tem N/A, valores negativos (para atributos numéricos) ou linhas incompletas;
6. Renomear as colunas com 'Ano', 'Região', categoria numérica em questão e 'Estudantes' (valor total);
7. Criar uma função que retorna a percentagem de estudantes de acordo com o Ano e a Região para determinada categoria e incluir esses valores na coluna ‘Percentual’;
8. Executar um filtro final para tirar as linhas equivalentes ao total da região naquele ano e os dados relativos ao Brasil como um todo;
9. Converter o *data frame* em JSON e salvar o resultado localmente para inserir no banco de dados.

Nesse sentido, ao rodar os *scripts* gerais podemos encontrar certas inconsistências no *data frame* e fazer os ajustes necessários para corrigi-las. Esse foi o caso da tabela relativa ao número de pessoas que tinham um celular para uso pessoal com acesso à Internet. Quando analisamos os valores finais, vimos um resultado muito estranho, em todas as regiões. Nos anos avaliados mais de 90% tinham esse acesso, mas na região norte, por exemplo, em 2019, apenas 74,2% tinham acesso à internet. Então, checamos os totais brutos, em cada tabela, e vimos que tinham algumas divergências: muito menos pessoas foram avaliadas a respeito do celular móvel, o que causava divergência quando calculada a percentagem. Para solucionar esse problema e trazer um valor mais

próximo do real, utilizamos como total o número de pessoas que utilizaram Internet nesse período, para contexto também de visualização que será descrito mais à frente.

Depois de inserir os dados já tratados do PNAD - Contínua 2016-2019, divididos por visualizações no banco, o passo seguinte foi fazer o mesmo com os dados do PNAD - Contínua 2020. A primeira tarefa foi fazer os filtros iniciais básicos, para todas as visualizações, que também já havia sido feito quando esse arquivo foi tratado no Jupyter Notebook: por idade, devido ao fato que apenas algumas respondiam às perguntas relativas à educação e aplicando idade mínima de 10 anos, pois os dados dos anos anteriores eram a partir dessa idade; condição de estudante e ensino médio incompleto ou menos.

Além disso, como nesta tabela não tínhamos a região do país definida, apenas o código da Unidade da Federação, foi necessário adicionar essa coluna ao *data frame*. Para isso foi criado um vetor em R, exemplificado abaixo, relacionando os códigos das UFs e as regiões. Dessa maneira, com a divisão inteira por 10 do código da UF temos a região. A partir daí o processo de tratamento foi bastante similar ao descrito anteriormente.

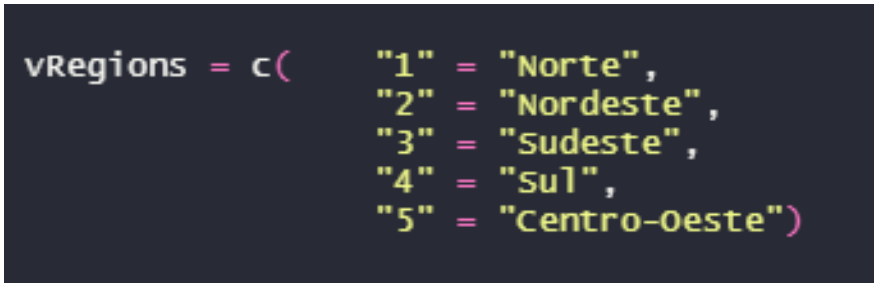

Figura 15: Vetor de regiões

Num segundo momento, fizemos um recorte por raça/cor e por tipo de instituição de ensino (pública ou privada). Para isso, seguimos o exemplo de tratamento dos dados por região, mas ao invés de fazer o agrupamento dos dados por região, fizemos por essa nova categoria. Depois, seguimos o mesmo processo.

5.2 Definições de Layout

Para o layout principal foi utilizado um Header contendo o título do projeto, um Footer descrevendo o tipo de projeto e o nome da aluna, e um menu lateral para navegação. A ideia é que na página inicial tenha *insights* das visualizações, para que o usuário possa se aprofundar mais em cada, mas também existe um

menu lateral, onde o usuário pode ir diretamente para uma página específica. Além disso, foi desenvolvido um ícone para a página. A ideia foi que remetesse a um HeatMap, utilizado em algumas visualizações.

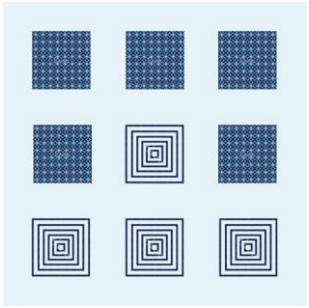

Figura 16: Ícone da página

Na página inicial foram exibidos os *insights* citados nas seções anteriores como um breve resumo do que a página irá exibir. A ideia foi utilizar elementos chamativos, como uma barra de progresso para o usuário conseguir comparar visualmente uma informação, ou, percentuais com tamanho de fonte maior do que o texto, para evidenciar a diferença, ou círculo de percentual para trazer uma ideia de progresso e também poder variar os componentes visuais. Além disso, foi inserido um botão primário como *call to action* para o usuário depois que ler os *insights* entrar na página determinada e ter acesso às visualizações completas.

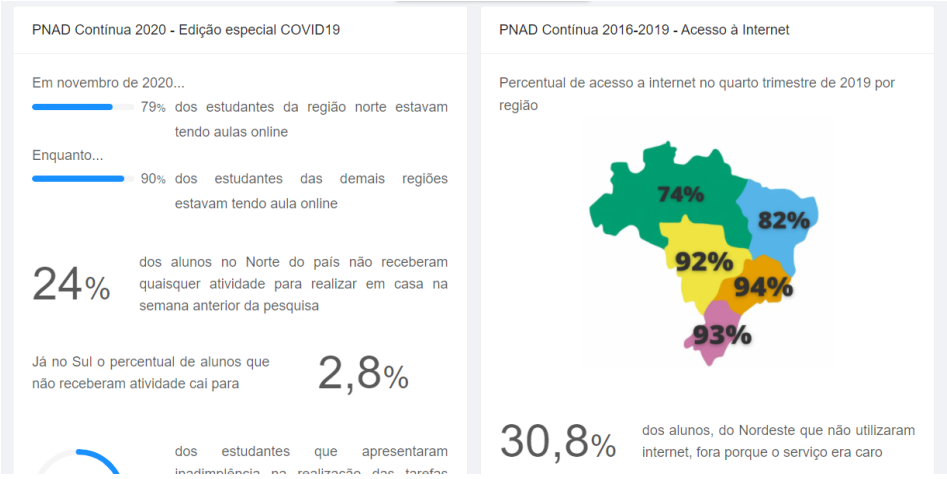

Figura 17: Parte da página inicial

Por fim, outro elemento inserido em um dos *cards* foi um mapa do Brasil dividido por regiões, com percentuais relativos ao acesso à internet, *insight* a ser destacado. Esse mapa foi renderizado como imagem e desenhado na plataforma Canva. As cores utilizadas são da mesma paleta do resto do projeto e o tamanho

da fonte dos percentuais está em perspectiva de acordo com seu valor. A ideia de trazer o tamanho da fonte variado foi trazer mais um elemento visual para destacar os percentuais. No momento de inserir essa imagem, foi necessário fazer alterações no arquivo de configurações do Next, inserindo a biblioteca *next-images* para ser possível ler uma imagem diretamente do projeto.

Sobre as cores dos gráficos foi decidido seguir a paleta de cores definida por Masataka Okabe e Kei Ito, por ser uma paleta segura para as pessoas com daltonismos que acessem a plataforma [10].

5.3 Visualizações

O objetivo do projeto foi construir visualizações que respondessem às perguntas previamente definidas na seção 3, porém no desenvolvimento fez-se necessário dividir os dados da PNAD - contínua entre 2016 e 2019 com os dados de 2020, visto que nesse ano foi feita uma edição especial devido à pandemia do COVID - 19, em que se incluía diversas perguntas sobre o ensino remoto que se fizeram necessárias destacar neste trabalho.

5.3.1 PNAD - Contínua 2020 - Edição especial COVID19

A primeira visualização é sobre os Modelos de aula por região: nela podemos observar que a grande maioria dos estudantes brasileiros estava, na época da pesquisa (novembro de 2020), num sistema de ensino remoto, mesmo os cursos sendo normalmente presenciais. Na maioria das regiões do país temos mais de 90% dos alunos com aulas remotas, salvo a região Norte, que já possuía mais de 20% dos alunos com aulas presenciais e/ou semipresenciais.

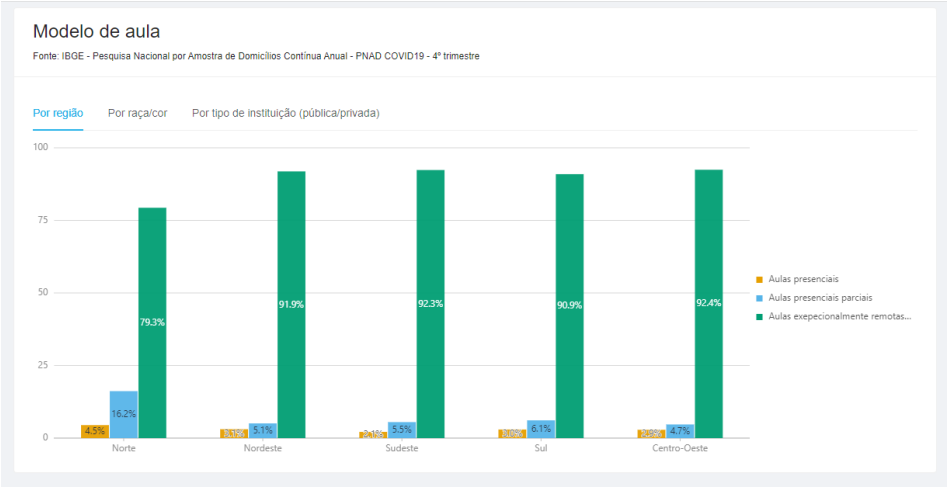

Figura 18: Modelo de aula por região

A ideia foi construir um gráfico de colunas, visto que temos três variáveis sendo tratadas: uma numérica, percentual de estudantes; e duas categóricas, região e tipo de ensino. Por se tratar de um dado numérico o percentual de estudantes ficou no eixo y do gráfico e para o eixo x ficaram as regiões, pois o objetivo desta visualização é fazer uma comparação entre as regiões [11].

A segunda visualização diz respeito à Distribuição da realização de tarefas em casa, relativo a alunos que receberam alguma atividade (aulas online, deveres, estudo dirigido etc.) para realizar em casa na semana anterior à coleta de dados. Nela podemos observar uma discrepância entre as regiões do país, visto que 24% dos alunos do Norte do país não receberam nenhuma atividade para ser realizada em casa, em comparação ao Sul, onde apenas 2,8% não receberam. Vale destacar que a região Norte era a região com mais alunos com aulas presenciais, mas mesmo assim apenas 4,5% tinham aulas totalmente presenciais.

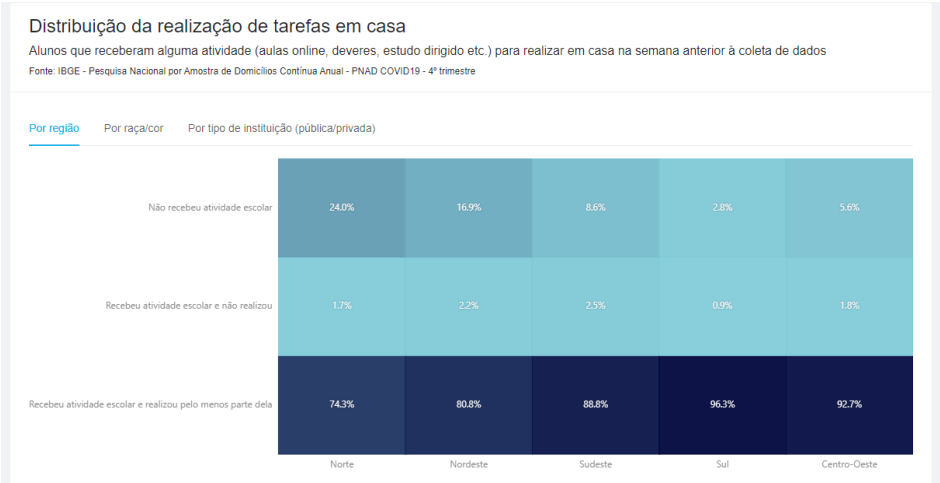

Figura 19: Distribuição da realização de tarefas em casa por região

Nesta visualização foi utilizado um HeatMap, pois apesar de termos novamente duas variáveis categóricas e uma numérica, nele conseguimos fazer comparações muito rápidas devido a sua escala de cores [12]. Nesse caso, o percentual é relativo a cada categoria e o total corresponde a cada região. No primeiro olhar já vemos que a última categoria é onde se encontra a maior parte dos alunos, devido à cor azul mais escura, com algumas divergências por região. Na categoria do meio temos um comportamento parecido por região, pois temos pouca variação na tonalidade do azul, e na primeira ficam claras as diferenças, visto que o azul do Norte é bem mais escuro do que os outros.

Apesar de ser uma porcentagem pequena, os alunos que receberam e não realizaram alguma atividade, está entre 2,5% e 0,8%. Como tínhamos acesso à informação do motivo pelo qual eles não realizaram a atividade, decidimos trazer esse dado como um complemento ao gráfico acima. Como o número de estudantes nessa categoria já era muito pequeno, não faz sentido dividir por região, então foi decidido usar um gráfico de pizza.

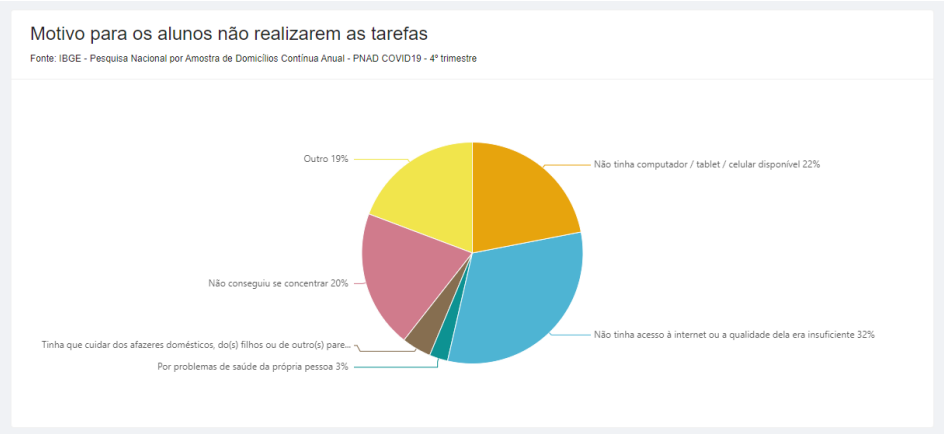

Figura 20: Motivo para os alunos não realizarem as tarefas

Esse gráfico foi escolhido para a visualização, pois ele é a melhor maneira para mostrar relacionamentos percentuais, e como temos apenas 6 categorias essa visualização é a mais sugerida [12]. Nela conseguimos, de imediato, ver que os motivos de “Não tinha computador / tablet / celular disponível” e “Não tinha acesso à internet ou a qualidade dela era insuficiente” compõem mais de 50% do gráfico, por exemplo.

Outro dado importante que a pesquisa trouxe para nós é sobre como os alunos distribuem suas tarefas ao longo da semana, entre dias e horas de dedicação. O objetivo desse gráfico era entender quanto tempo os alunos estavam se dedicando a atividades escolares durante a semana, visto que num modelo presencial os alunos têm aproximadamente cinco horas de aula por dia em cinco dias na semana.

No primeiro momento foi feito um *Heatmap*, mas depois de analisar o gráfico foi possível perceber que com ele não entendemos a distribuição por completo. O dado que a visualização nos retorna é sobre a distribuição de horas para cada tipo de dedicação semanal, o que não era a ideia inicial.

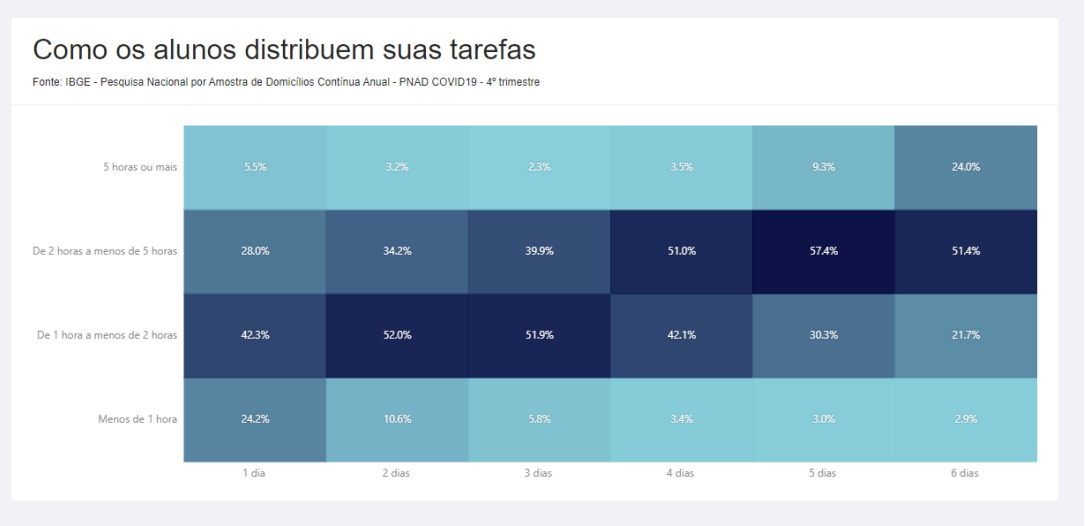

Figura 21: Como os alunos distribuem suas tarefas primeira versão

A partir daí voltamos para o R, reescrevemos o código e tiramos uma nova visualização no modelo *Treemap*. Essa visualização funcionou mais que a primeira, pois estamos descrevendo relacionamentos que fazem parte de um todo e não de um grupo [12]. Além disso, permite uma visualização completa para uma grande quantidade de dados. Mas, por se tratar de muitas categorias, 24 no total, a visualização ainda poderia ficar confusa para o usuário.

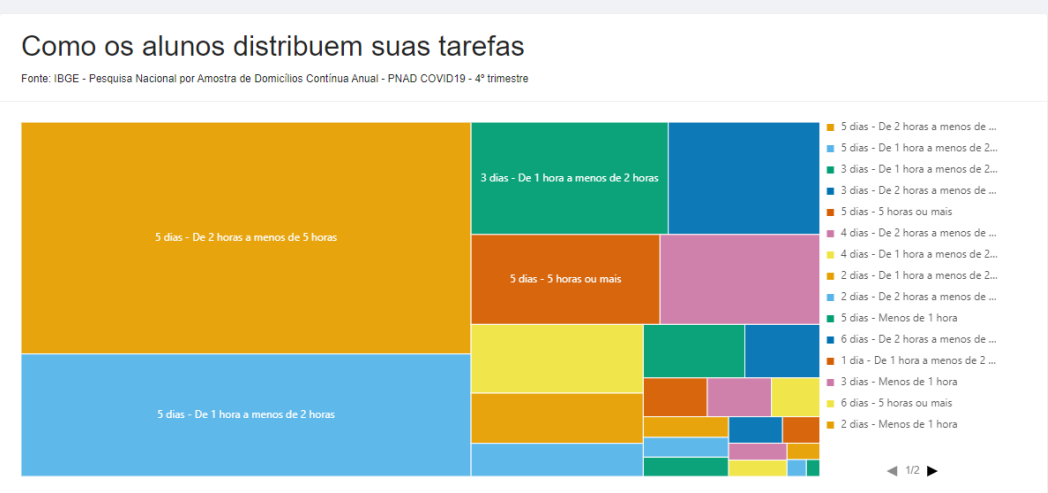

Figura 22: : Como os alunos distribuem suas tarefas segunda versão

Então, a ideia foi fazer uma junção dos pontos positivos de cada visualização testada anteriormente e decidimos construir um *bubble chart*, que nada mais é do que um Heatmap com bolhas proporcionais. Com o Heatmap conseguimos fazer melhores comparações, pois conseguimos visualizar todos os dados e com as bolhas trazemos a proporção, dando destaque ao maior

valor, e colocamos as cores nos dias de dedicação para facilitar comparações. Além disso, é possível diferenciar dos Heatmaps das outras visualizações, pois nesse caso o total percentual é relativo ao todo e não a cada categoria exposta na coluna.

Figura 23: Como os alunos distribuem suas tarefas ao longo da semana versão final

Com essa visualização conseguimos reparar que mais da metade dos alunos se dedica para as atividades em cinco dias na semana, entre uma e cinco horas diárias, sendo a maioria mais de duas horas. Além disso, vemos que temos mais estudantes que dedicam três dias do que quatro dias. Raras exceções, nas bordas do gráfico, se dedicam em apenas um ou seis dias, e também menos de uma hora ou mais de cinco horas.

5.3.2 PNAD - Contínua 2016 - 2019

Nestas visualizações temos, em adição aos dados categóricos e números, a escala temporal, visto que a pesquisa reúne dados evolutivos de 2016 a 2019. Sendo assim, foi definido em alguns componentes inserir essa escala temporal para ver o comportamento evolutivo até 2019 e a partir daí poder entender um pouco do comportamento em 2020. E, em outros, exibir apenas o dado de 2019, visto que não houve um crescimento significativo entre os anos anteriores, sendo possível usar esse valor como base de comparação.

A primeira visualização é relativa ao percentual de estudantes com 10 anos ou mais de idade que utilizaram Internet no período de referência dos últimos três meses, entre os anos de 2016 e 2019. Nela podemos observar que mesmo com o passar dos anos o gráfico ainda é dividido em dois grupos: as regiões Sudeste, Sul e Centro-Oeste, que variam entre 87,3% (Centro-Oeste em 2016) e 94,1% (Sudeste em 2019) e outro grupo que varia entre 64,5% (Norte

em 2016) e 81,7 (Nordeste em 2019). Apesar das regiões Norte e Nordeste crescerem de maneira mais rápida que as do primeiro grupo, conseguimos perceber que mesmo assim a diferença entre o acesso a internet nas regiões é exorbitante.

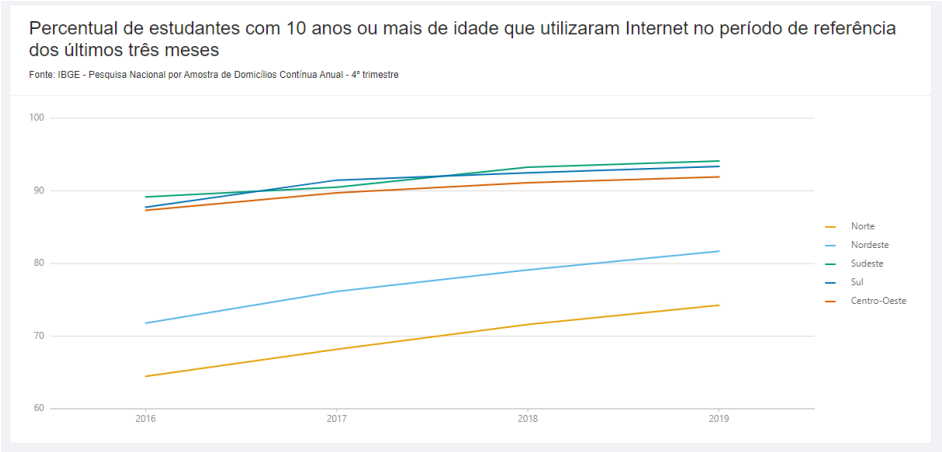

Figura 24: Percentual de estudantes com 10 anos ou mais de idade que utilizaram Internet no período de referência dos últimos três meses

Para esse dado foi escolhido um gráfico de linhas, pois é o gráfico clássico para exibir a evolução de certo aspecto com o tempo. Como não temos muitos dados, apenas a informação de cinco regiões em quatro diferentes anos, podemos ter uma visualização clara e rápida das informações [12].

A segunda foi a que trouxe *insights* mais interessantes. Nela podemos reparar a distribuição dos motivos pelos quais os entrevistados não fizeram uso da internet, no período de referência de três meses, em cada região. O Norte se destaca por ter um comportamento diferente das outras regiões, os motivos acabaram sendo bem distribuídos, podemos perceber isso pelos tons de azul no gráfico que variam pouco. Já no Nordeste temos um destaque no “Serviço de internet era caro” que corresponde a 30,8% dos entrevistados que não acessaram a internet no período de referência. O comportamento do Sul e do Sudeste é bem parecido com aproximadamente um quarto das pessoas que não acessam a internet por falta de interesse. E por fim, no Centro-Oeste os principais motivos eram que o equipamento ou o serviço eram caros.

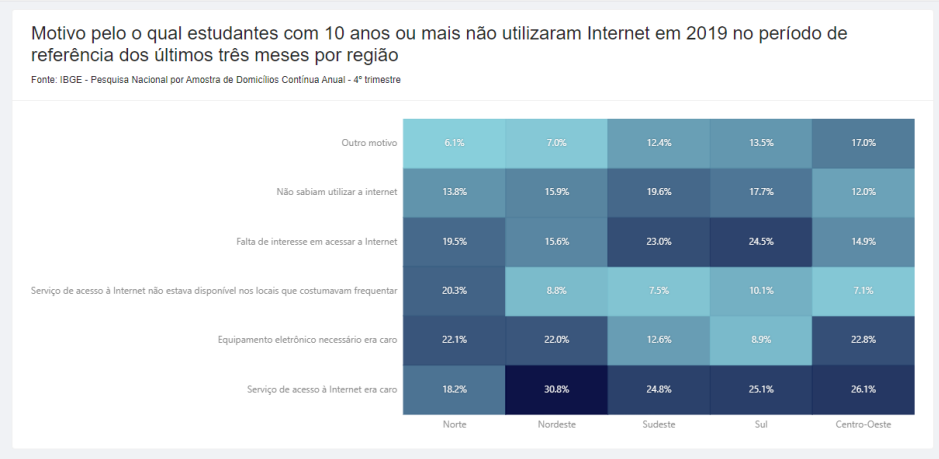

Figura 25: Motivo pelo o qual estudantes com 10 anos ou mais não utilizaram Internet em 2019 no período de referência dos últimos três meses por região

O gráfico utilizado para essa visualização foi um *HeatMap*, pois temos, novamente, três variáveis, duas categóricas: uma com cinco valores no eixo x, região do país, e outra com seis possíveis valores no eixo y; e para o campo de cor temos a porcentagem de estudantes que se encaixa em cada categoria. Poderíamos utilizar outros gráficos como em colunas empilhadas ou agrupadas, mas por se tratar de muitas categorias no eixo y, os valores de menores porcentagens poderiam se perder. Além disso, com a escala de cores conseguimos comparar de maneira mais rápida os motivos em cada região [12].

Na visualização seguinte, o principal objetivo era entender quais equipamentos eram utilizados pelos estudantes para acessar a internet. Um desafio foi que nos exemplos anteriores as opções categóricas eram únicas e nesse múltiplas, ou seja, somando os percentuais de cada região para cada equipamento temos mais que 100%, um único entrevistado poderia ter utilizado internet em múltiplos aparelhos diferentes.

Podemos reparar pelo gráfico que o uso de internet pelo celular é uma quase unanimidade para todas as regiões do país, todas em mais de 97%. Fora isso, reparamos que o uso de microcomputador ou tablet é muito mais forte nas regiões Sudeste, Sul e Centro-Oeste que fica em torno dos 60%. Isso foi um grande diferencial para o Nordeste e o Norte, que ficam em torno de 40%.

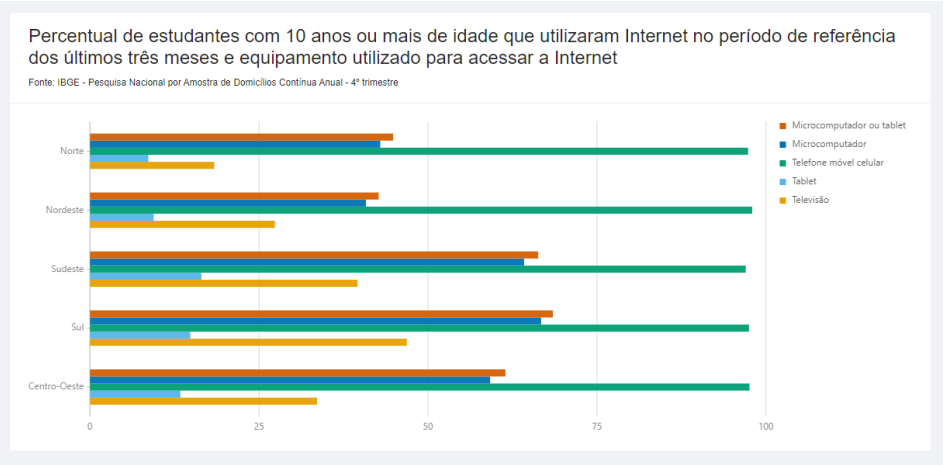

Figura 26: Percentual de estudantes com 10 anos ou mais de idade que utilizaram Internet no período de referência dos últimos três meses e equipamento utilizado para acessar a Internet

A escolha desse gráfico deve-se ao fato de mostrar uma comparação lado a lado dos dados principais. A ideia é que ele seja como uma barra de progresso, podendo chegar até 100%, por isso em barras horizontais e não colunas verticais [12]. Além dos dados exibidos, temos o inverso deles, por exemplo: “Não utilizam microcomputador”, mas por a tela já conter muitas informações foi decidido não exibir esse dado, pois ele já está implícito. Se 42,9% dos estudantes utilizaram o microcomputador, 47,1% não o utilizaram. Além disso, mantemos as regiões no eixo categórico e não variando nas cores para manter a mesma lógica das outras visualizações.

Por fim, no último gráfico dessa página tivemos um desafio, já que o total de estudantes que respondeu essa pergunta foi diferente dos gráficos acima. Para resolver esse problema, ao tratar os dados, usamos como total o número de estudantes que disseram utilizar a internet no período em questão. Mas, ainda assim o número de respostas era menor do que isso.

Para tratar essa questão, resolvemos manter o valor total dos estudantes que utilizaram a internet como valor total, mas exibimos o percentual relativo ao total de estudantes que responderam essa pergunta e o percentual que afirmou que tem um telefone móvel para uso pessoal. Num primeiro momento, mostramos esse percentual como uma linha auxiliar, mas por já possuirmos cinco linhas, relativas às regiões, resolvemos apenas acrescentar esse valor na tooltip e fazer uma observação no subtítulo do gráfico.

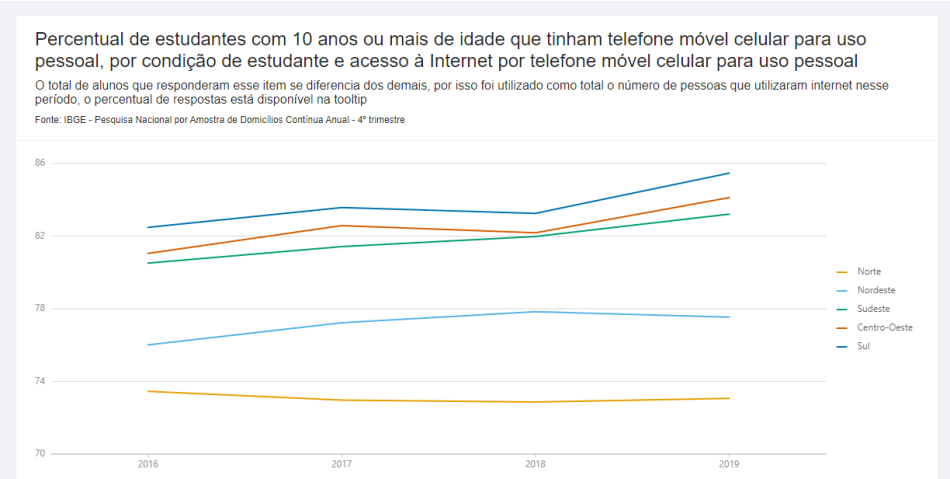

Figura 27: Percentual de estudantes com 10 anos ou mais de idade que tinham telefone móvel celular para uso pessoal, por condição de estudante e acesso à Internet por telefone móvel celular para uso pessoal.

5.4 Opções de design

No primeiro momento os gráficos foram plotados diretamente na tela, mas como possuem transparência, era possível que o usuário não conseguisse interpretar tão bem os dados, pois parecia que eles estavam voando sobre a tela. Dessa forma, decidimos colocar os gráficos em *cards* com fundo branco, para poder dar maior destaque e evitar possíveis problemas na interpretação. Além disso, foram inseridas margens para definir melhor os espaços de cada gráfico. Segue abaixo o resultado.

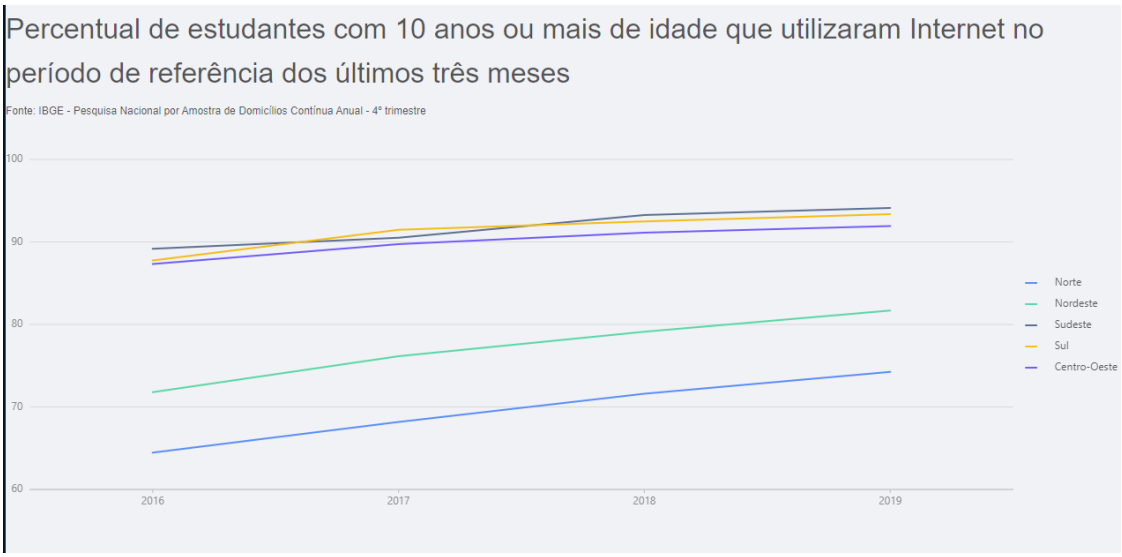

Figura 28: Gráfico fora do card

Outra mudança foi no subtítulo do gráfico: a ideia inicial era que realmente existisse um subtítulo e que a fonte dos dados entrasse em outra posição, mas como apenas poucas visualizações necessitam de subtítulo, a fonte acabou entrando no lugar dele para maioria dos gráficos. Para não causar uma inconsistência, foi decidido manter o subtítulo, com uma fonte um pouco maior do que a anterior, e adicionar também a fonte no cabeçalho do *card*. Assim, a fonte fica presente em todas as visualizações e o subtítulo naquelas em que for necessário. Abaixo segue como era antes e depois como ficou.

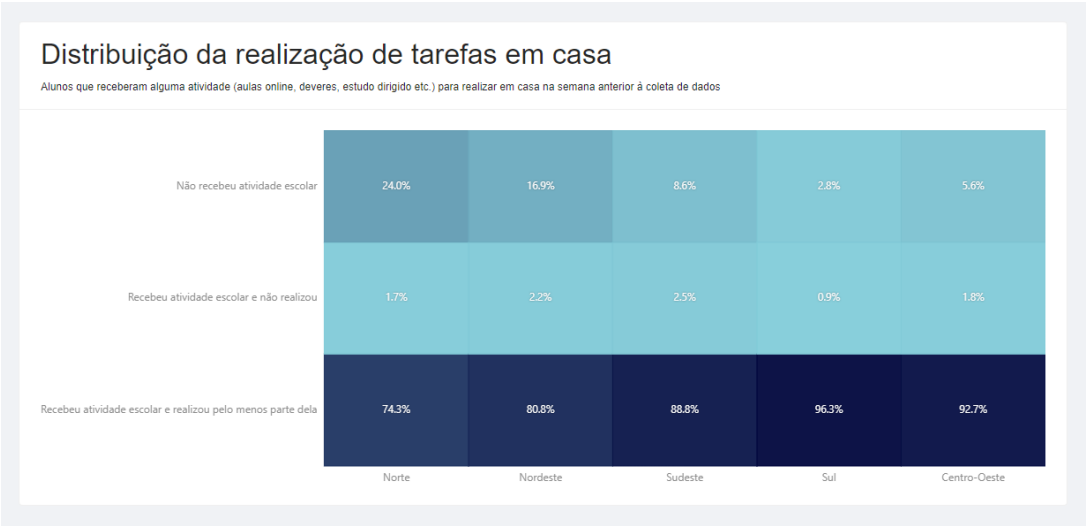

Figura 29: Gráfico com subtítulo e sem fonte

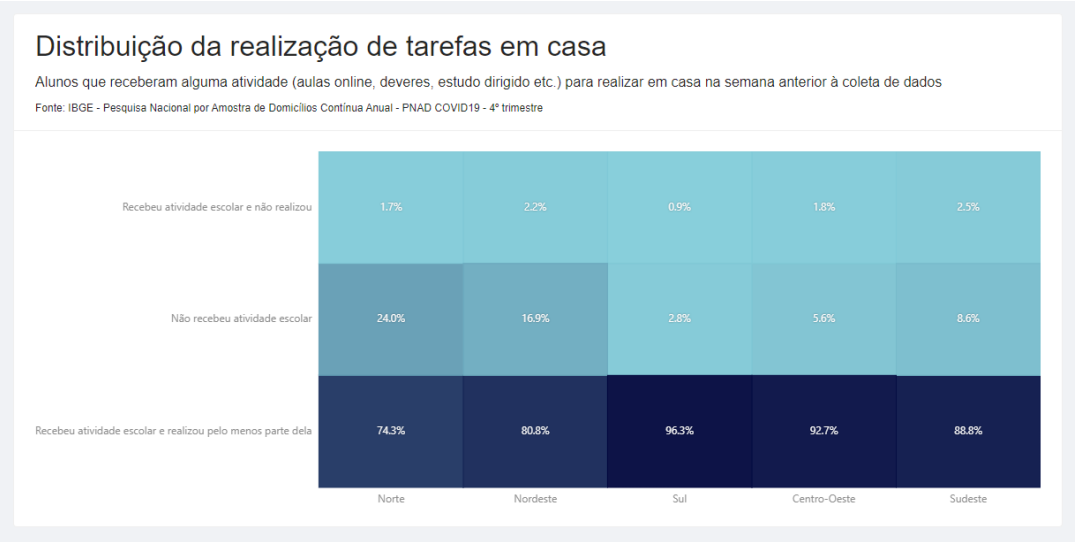

Figura 30: Gráfico com subtítulo e fonte

Além disso, foram definidas cores para cada região, para facilitar as comparações, principalmente nos gráficos de linhas.

5.4.1 Novas visualizações

Com as análises realizadas num segundo momento, tínhamos mais dois recortes a serem exibidos para as visualizações de Modelo de aula e Distribuição da realização de tarefas em casa, ambas relativas ao PNAD – 2020 edição COVID19. Nesse caso, inserimos *Tabs* para cada recorte e mantemos o mesmo modelo de gráfico, visto que a única diferença foi a variável categórica no eixo X.

Figura 31: Modelo de aula por raça/cor

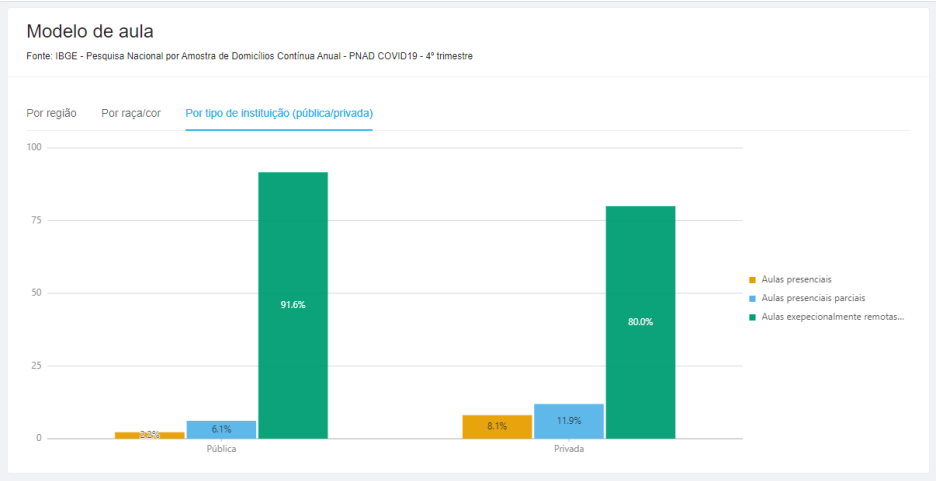

Figura 32: Modelo de aula por tipo de instituição (pública/privada)

Nessa visualização, podemos perceber que não temos muita variação entre as raças, apenas a raça Amarela se destaca das outras por ter mais aulas presenciais parciais. No entanto, quando vamos comparar por tipo de instituição as diferenças se destacam, na rede pública temos mais de 90% dos alunos em aulas remotas, para 80% da rede particular.

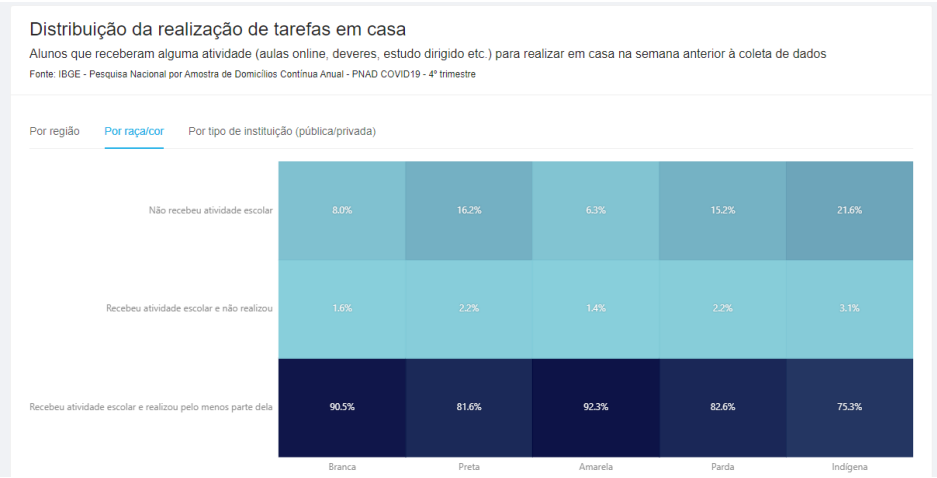

Figura 33: Distribuição da realização de tarefas em casa por raça/cor

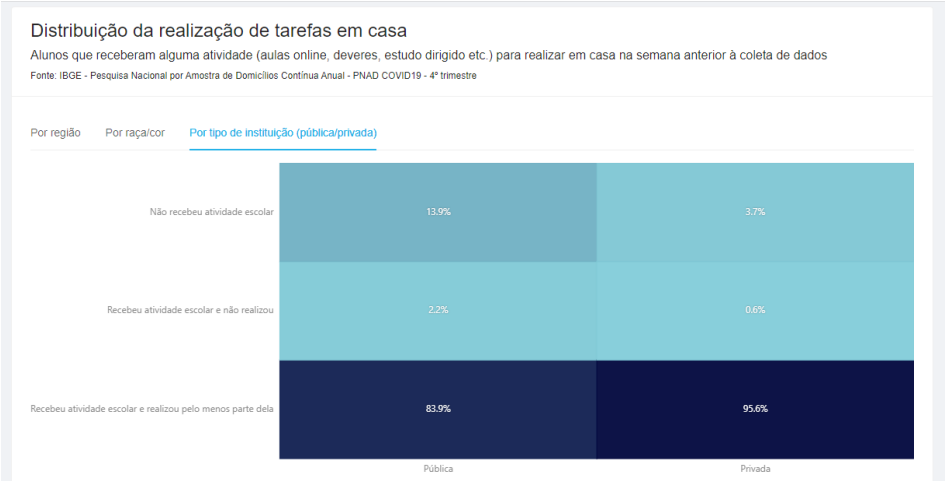

Figura 34: Distribuição da realização de tarefas em casa por tipo de instituição (pública/privada)

Curiosamente, comparando este gráfico com o acima, no recorte por raça/cor, vimos que apesar dos alunos terem o mesmo modelo de aula, alunos Pretos e Pardos recebem menos atividades do que Brancos e Amarelos e, Indígenas, menos ainda, onde 21,6% dos estudantes não receberam nenhuma tarefa. A diferença nesse caso também é evidente quando analisamos o recorte por instituição, alunos de instituições privadas recebem, e realizam mais tarefas que os de instituições públicas.

Como essas visualizações foram feitas em um segundo momento, elas não foram avaliadas nos testes com usuários descritos na seção 6.

5.5 *Deployment*

O processo de *deployment* foi bem simples, pois a Vercel disponibiliza um ambiente gratuito para usuários com sites não comerciais. Ele se conecta com a conta no *GitHub* para reconhecer o projeto, depois é só adicionar as variáveis de ambiente e fazer o processo de *build*. Além disso, ele disponibiliza algumas URLs; a que está sendo usada é https://projeto-final-carolfjunger.vercel.app/ .

Depois do primeiro *deploy*, foram feitas algumas alterações no modo de requisição ao banco de dados, pois as chamadas são um *fetch* para uma URL de API dentro do projeto que se conecta com o banco. Essas chamadas estavam com a URL de desenvolvimento, http://localhost:3000/, mas o desejável é que elas fizessem a requisição de acordo com o domínio da aplicação. No entanto, por se tratar de uma aplicação *Server-Side rendering*, nem sempre temos acesso a qual é o domínio durante a chamada. Dessa maneira, foi

construída a função abaixo que retorna o *endpoint* de acordo com a necessidade do projeto, basta comentar/descomentar a linha antes de subir para produção.

```
1  export default function getEndPoint () {
2      return "https://projeto-final-carolfjunger.vercel.app/"
3      // return "http://localhost:3000"
4  }
```

Figura 35: Função getEndPoint

6 Avaliação

6.1 Planejamento e execução de testes

Ao finalizar cada versão é necessário validar com usuários se as visualizações estão, de fato, cumprindo com o objetivo proposto. A ideia foi criar um questionário pelo Google Forms. As primeiras perguntas foram relacionadas ao perfil do usuário, como informações básicas pessoais e dados a respeito do seu conhecimento em visualização de gráficos. As seguintes foram relacionadas a cada página de visualização, de uma a três perguntas por gráfico, algumas de múltipla escolha e, para gráficos que possam ser mais complexos, espaço de texto livre. E, ao final, uma caixa para dúvidas, sugestões e/ou comentários.

Para validar o questionário, a aluna respondeu como primeiro usuário testando, logo depois fez alguns ajustes e em seguida, realizou um teste, com uma usuária que ainda não tinha contato com a ferramenta a partir daí realizou os ajustes finais antes de disparar o questionário. Esses dois testes foram descartados para fins de avaliação.

O questionário se encontra no apêndice A.

6.2 Comentários sobre a implementação

O questionário foi enviado para diversos usuários por meio de redes sociais e ficou ativo por 12 dias, visto os prazos necessários para os ajustes e finalização do projeto. No final, foram obtidas 20 respostas, os resultados foram analisados em R, para descobrimos possíveis problemas com a interface. Ao longo das análises, percebemos que um dos entrevistados havia utilizado a plataforma pelo celular, e como o sistema atualmente não suporta telas pequenas, esse usuário foi eliminado da pesquisa, tendo assim 19 respostas válidas. Para cada pergunta fizemos uma análise de acordo com o perfil do usuário. Segue abaixo os resultados.

6.2.1 Perfil dos usuários

Todos os usuários que responderam ao questionário deram seu consentimento. A maior parte tem entre 18 e 24 anos mas, ainda assim, temos no mínimo um usuário em cada faixa etária, como podemos perceber pelo gráfico abaixo.

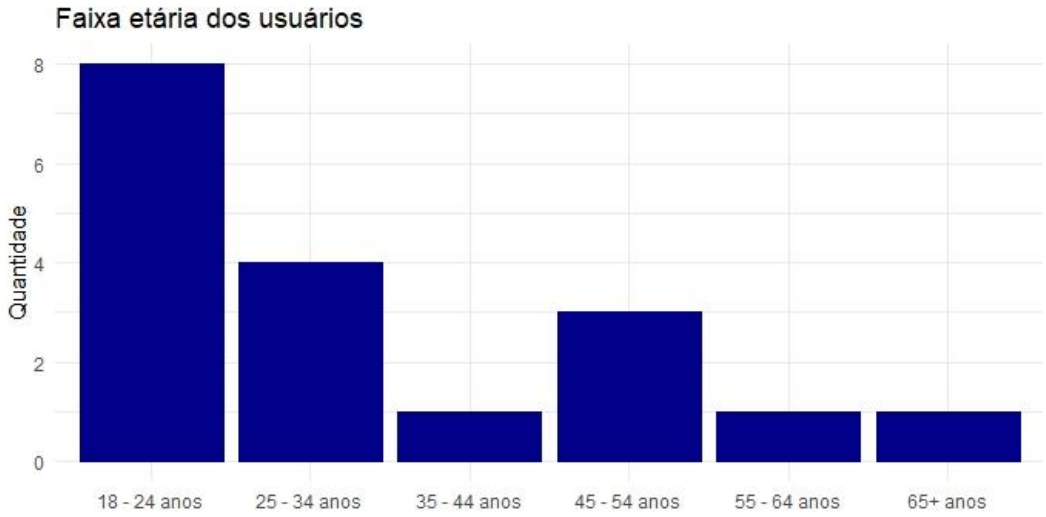

Figura 36: Faixa etária dos usuários respondentes

A respeito da escolaridade dos avaliados, a grande maioria tem graduação completa, e uma boa parte pós-graduação também.

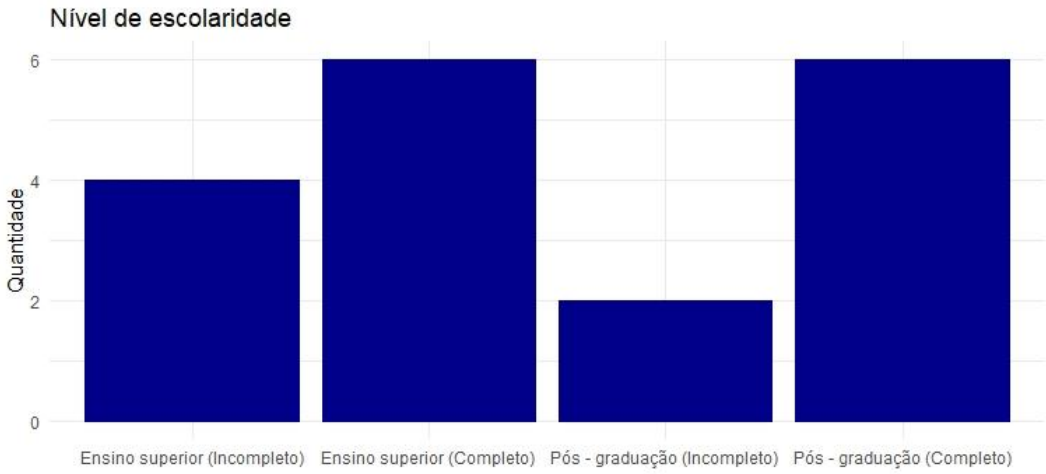

Figura 37: Nível de escolaridade

Para a principal área de formação do usuário, resolvemos dividir em dois grupos, as áreas *STEM*, sigla em inglês que engloba as áreas de ciência, tecnologia, engenharia e matemática e outro grupo com as restantes. Por essa categoria ser de texto livre e não termos um número elevado de respostas, apenas criamos uma lista com as áreas enquadradas no *STEM* e filtramos por quais estavam nessa lista.

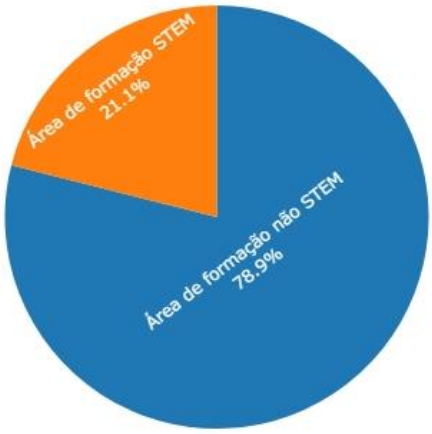

Figura 38: Divisão de área de formação

Apesar de menos de um quarto dos usuários se encaixarem neste grupo, o recorte foi mantido, pois partimos de uma premissa que estes entrevistados teriam mais costume em utilizar tecnologias, interpretar gráficos e mexer em painéis visuais. Essa baixa porcentagem de avaliados da área de tecnologia deve-se ao fato que o questionário foi enviado para o núcleo de contato da aluna que é majoritariamente de pessoas formadas em áreas fora desse grupo.

Sobre a frequência que os usuários interpretam gráficos, temos que para os STEM todos interpretam no mínimo todo trimestre, a maioria toda semana, e para os não STEM a maioria interpreta anualmente ou nunca interpreta.

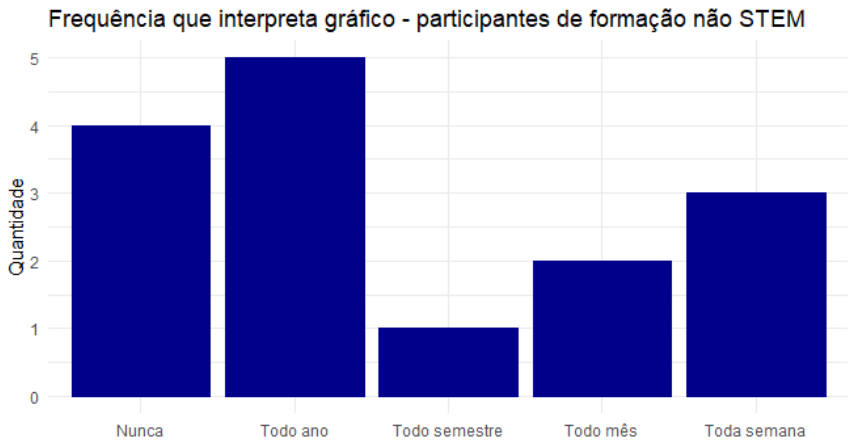

Figura 39: Frequência que interpreta gráfico – participantes sem formação em STEM

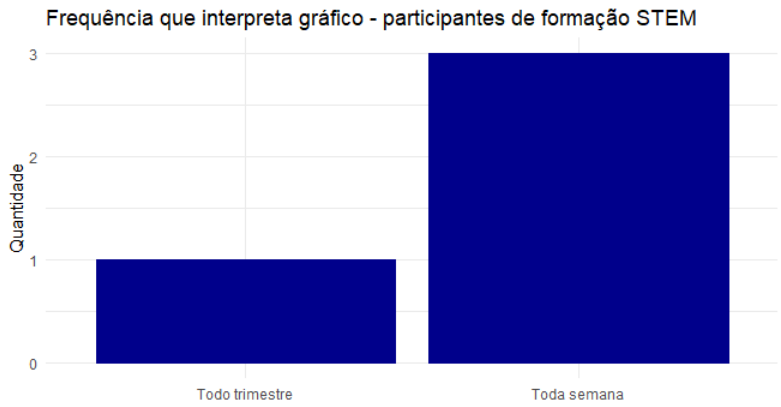

Figura 40: Frequência que interpreta gráfico participantes com formação em STEM

Sobre o contato que cada usuário teve com cada gráfico, tivemos os seguintes resultados: em geral, os usuários têm um contato de moderado a alto com os gráficos de pizza, gráfico de linhas e gráfico de colunas. E a maioria não tem nenhum contato com gráficos de bolhas, diagrama de caixa e mapa de calor. Quando fazemos o recorte para STEM, houve alguns usuários com baixo contato nesses últimos gráficos citados. Um fato curioso é que nenhum dos entrevistados marcou a opção especialista para qualquer tipo de gráfico.

Figura 41: Contato com gráfico

6.2.2 PNAD - 2020 edição COVID-19

Nessa seção foram feitas perguntas relativas às visualizações da página “PNAD 2020”. As primeiras foram sobre o primeiro gráfico, a respeito do modelo de aula mais adotado de acordo com a visualização, todos os usuários responderam corretamente a pergunta, mas sobre qual região tem algum modelo de aula presencial tivemos algumas respostas diferentes.

Figura 42: Modelo de aula por região

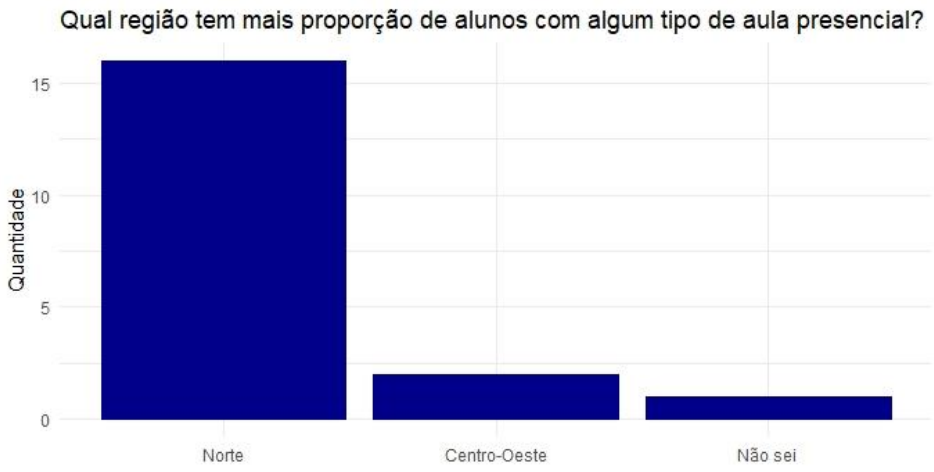

Figura 43: Qual região tem mais proporção de alunos com algum tipo de aula presencial?

Neste caso, vimos que a grande maioria respondeu “Norte”, como era esperado, mas ainda assim, tivemos outras respostas. Um usuário respondeu “Não sei”. Como se trata de uma pesquisa quantitativa e anônima, não conseguimos entrar em contato com esse usuário para entender quais foram as possíveis dificuldades na análise, visto que ele não deixou outro feedback. As outras duas respostas foram “Centro-Oeste”, o que foi curioso, visto que o Centro-oeste é a região que está com mais aulas online. Dessa forma, resolvemos mudar a redação da legenda de “Sem aulas presenciais, mas curso

presencial”, que poderia ter causado alguma ambiguidade, para “Aulas excepcionalmente remotas em curso presencial”.

O segundo gráfico é relativo à distribuição da realização de tarefas em casa. Sobre esse fizemos uma pergunta para entender o que os usuários interpretam desse gráfico. Das 19 respostas válidas, 11 conseguiram interpretar o gráfico da maneira esperada e parte desses fez destaque para a região Norte receber menos atividades do que as outras. Dentre as redações mais interessantes estão: “A região sul tem mais alunos que receberam atividade escolar e realizaram pelo menos uma parte dela (96,3%), enquanto na região norte está a maior porcentagem de alunos que não receberam atividade escolar (24%). Além disso, de maneira geral, os alunos receberam consideravelmente mais atividades escolares das quais realizaram pelo menos uma parte dela, enquanto que foi muito baixa a quantidade de estudantes que afirma não ter realizado as atividades designadas” e “Que no Norte quase um quarto dos alunos não recebeu a atividade escolar - bem acima da média das demais regiões - mas de forma geral todos que recebem fazem ao menos uma parte dela”. Outras 5 pessoas declararam não entender o gráfico, porém como puseram respostas muito vagas seria importante um novo estudo para avaliar o que exatamente os entrevistados não entenderam para avaliar mudanças.

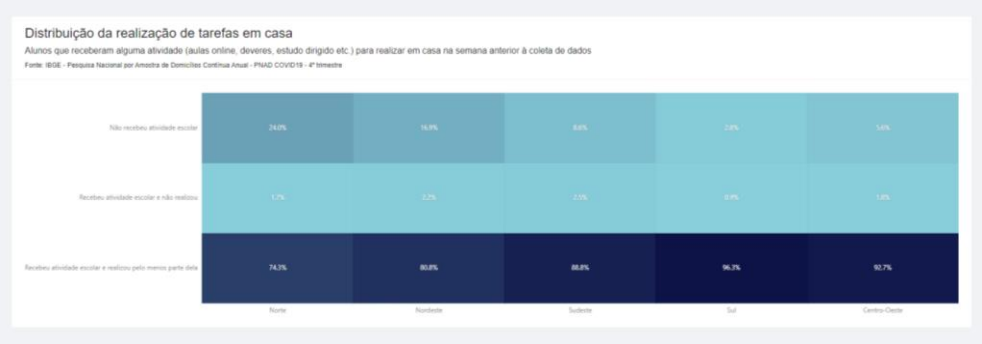

Figura 44: Distribuição das tarefas de casa

A seguinte pergunta era sobre se os alunos do Centro-oeste receberam mais atividades do que os do sul. Recebemos respostas diversas e ao analisar vimos que essa pergunta pode ter causado alguma ambiguidade para o entrevistado, visto que os dados relativos estão em porcentagem, logo ele poderia fazer uma comparação entre as porcentagens ou assumir que não sabe/não era possível responder por não termos os valores brutos.

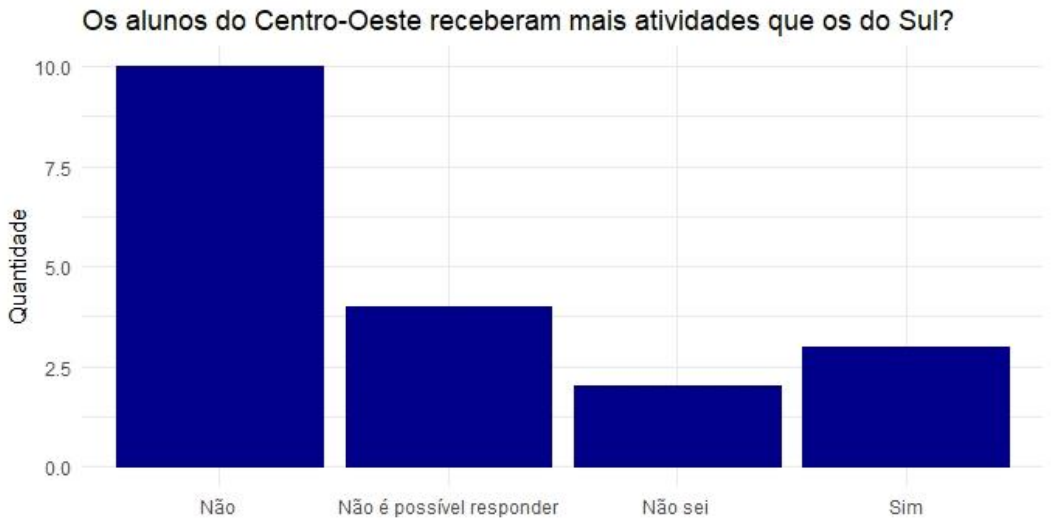

Figura 45: Os alunos do Centro-Oeste receberam mais atividades que os do Sul?

Na sequência, sobre o gráfico de pizza, relativo aos motivos dos alunos não utilizarem a Internet, perguntamos o mais frequente. Praticamente todos os entrevistados responderam o motivo: "Não tinha acesso à internet ou a qualidade dela era insuficiente", que, de fato, é o mais frequente. Mas, um usuário respondeu "Não sei". Analisando suas outras respostas, vimos que ao final ele deu esse feedback: "O "Não sei" representa que não sabia responder ou que a informação questionada não era possível de ser obtida apenas pela leitura do gráfico (que, em alguns casos, possuía alternativa de resposta própria nesse sentido)? Não ficou muito claro isso. Mas muito boa a pesquisa!", logo assumimos que essa resposta vem, justamente, dessa dúvida citada.

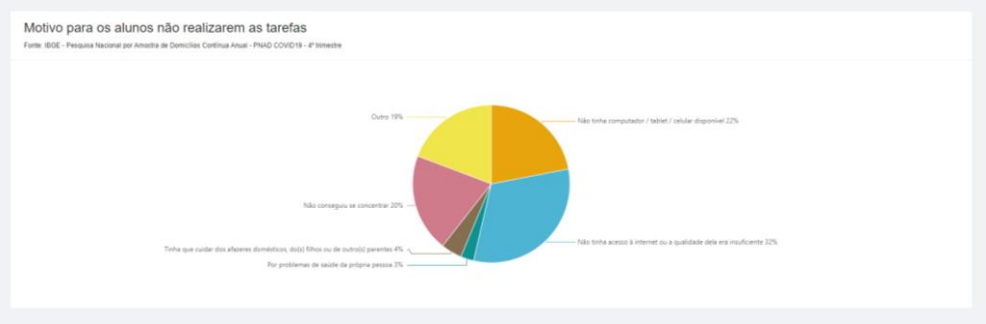

Figura 46: Motivo para os alunos não realizarem as tarefas

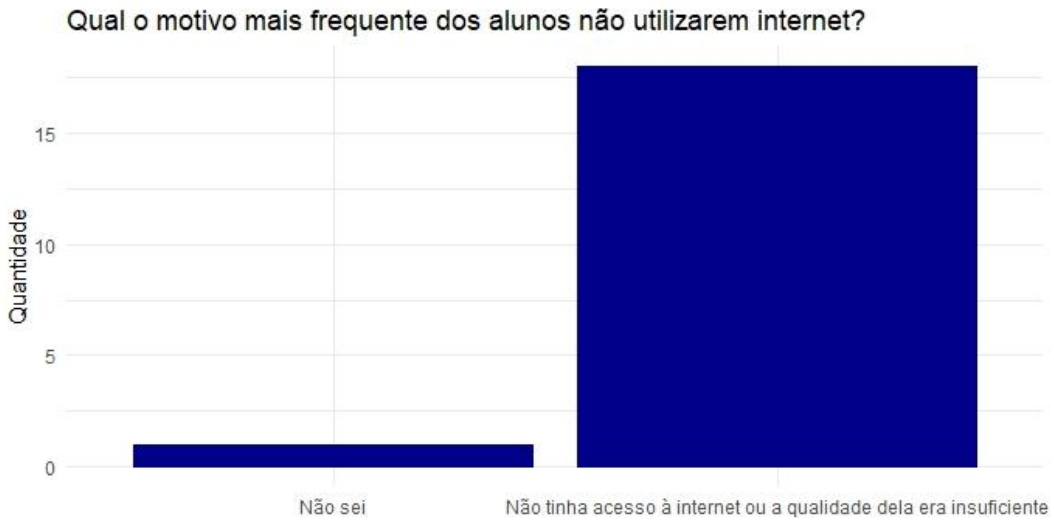

Figura 47: Qual o motivo mais frequente dos alunos não utilizarem internet?

Sobre o último gráfico desta página fizemos três perguntas. A primeira era uma de texto livre, perguntando como o usuário descreve as informações desse gráfico, nela a grande maioria destaca entender o gráfico e o interpreta de forma correta, mas alguns citam a dificuldade de relacionar as horas com os dias da semana. Para evitar esses problemas, trocamos o título de “Como os alunos distribuem suas tarefas” por “Como os alunos distribuem suas tarefas ao longo da semana” e incluímos “por dia” em cada categoria das horas.

Das perguntas de múltipla escolha dessa visualização a grande maioria dos usuários respondeu corretamente e, dos que responderam fora do esperado, não conseguimos tirar muitos *insights*, pois eles não descrevem muita coisa na resposta de texto livre.

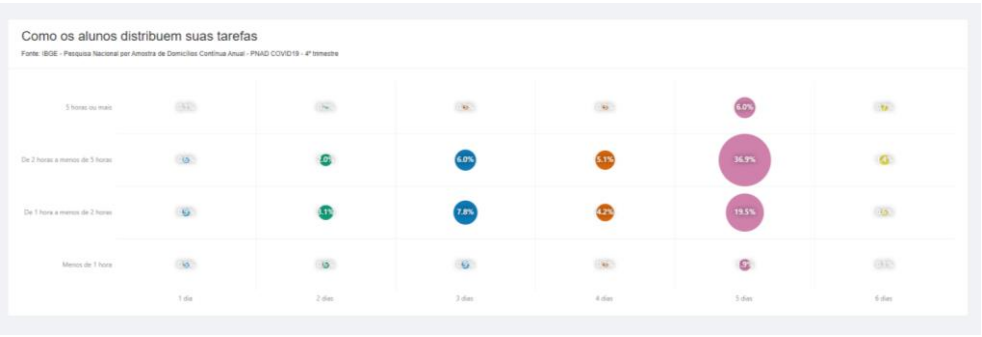

Figura 48: Como os alunos distribuem suas tarefas

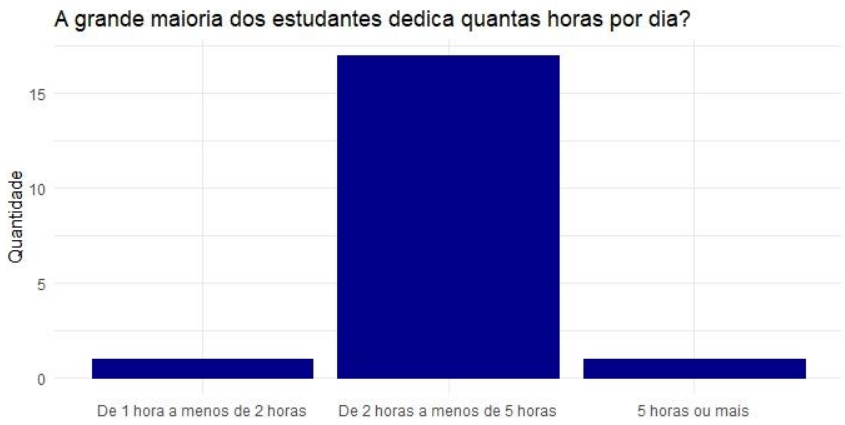

Figura 49: A grande maioria dos estudantes dedica quantas horas por dia?

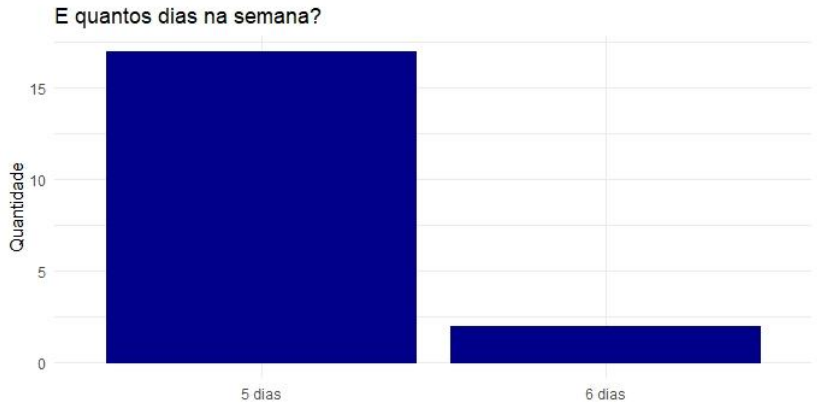

Figura 50: E quantos dias na semana?

6.2.3 Acesso à internet PNAD Contínua 2016-2019

Em relação a esse grupo de visualizações, fizemos mais perguntas objetivas, pois como é a quinta seção do questionário e, pelo Google Forms não podemos randomizar a ordem das perguntas, o usuário provavelmente já estaria mais cansado nesse ponto e poderia apresentar fadiga. Logo, com as perguntas mais objetivas deixamos o processo mais leve, fácil e rápido, a fim de evitar esse problema.

A primeira questão era relacionada ao primeiro gráfico de linhas, relativo ao percentual de estudantes com 10 anos ou mais de idade que utilizaram Internet no período de referência dos últimos três meses. Perguntamos qual região teve mais acesso à Internet em 2017 e em 2019. Na primeira, apenas um entrevistado respondeu fora do esperado, possivelmente uma fadiga e na segunda todos responderam dentro do esperado.

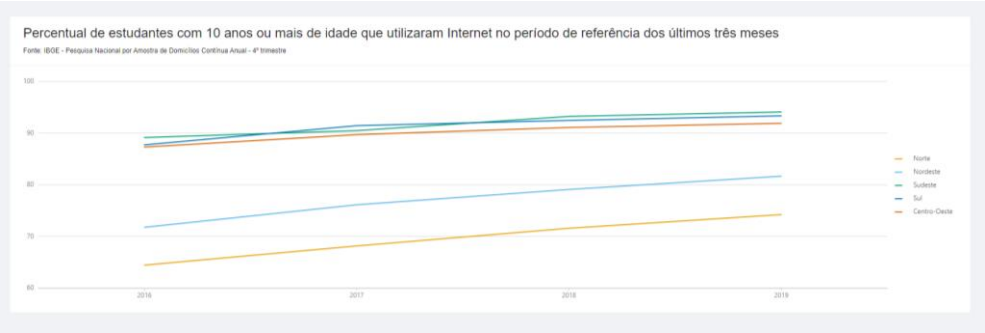

Figura 51: Percentual de estudantes com 10 anos ou mais de idade que utilizaram Internet no período de referência dos últimos três meses

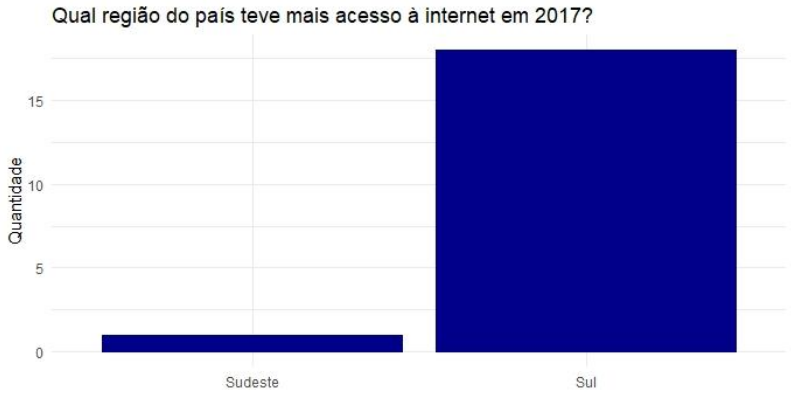

Figura 52: Qual região do país teve mais acesso à Internet em 2017?

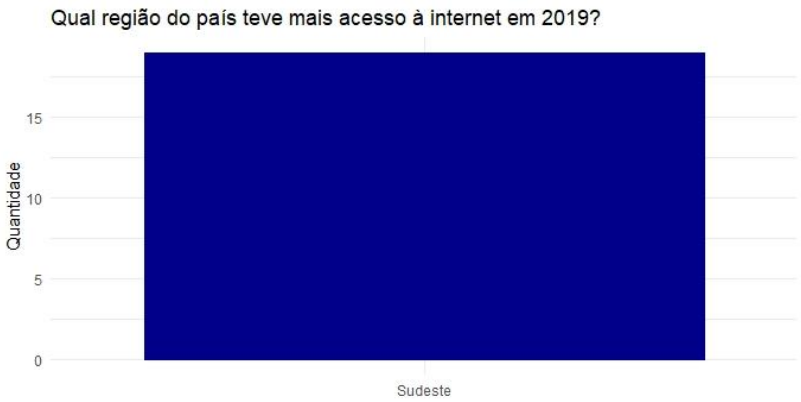

Figura 53: Qual região do país teve mais acesso à internet em 2019?

No item seguinte, perguntamos qual a diferença percentual (aproximada) entre as regiões Norte e Centro-Oeste em 2019. Nele, a maioria também respondeu dentro do esperado, com exceção de dois usuários. Analisando, um deles é o mesmo da questão anterior e o outro declarou que interpreta gráfico apenas anualmente, logo não vimos correções a serem feitas nesse gráfico.

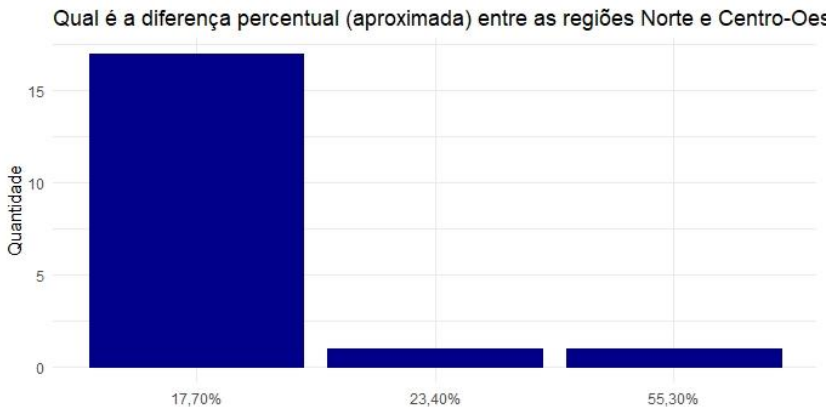

Figura 54: Qual é a diferença percentual (aproximada) entre as regiões Norte e Centro-Oeste em 2019?

Seguindo, temos as perguntas relativas aos motivos pelos quais estudantes com 10 anos ou mais não utilizaram Internet em 2019. Quando indagamos qual era o motivo, num geral, mais frequente, temos o mesmo entrevistado anterior respondendo fora do esperado e outro respondendo “Não sei”, mas este relatou ter baixo conhecimento com mapa de calor.

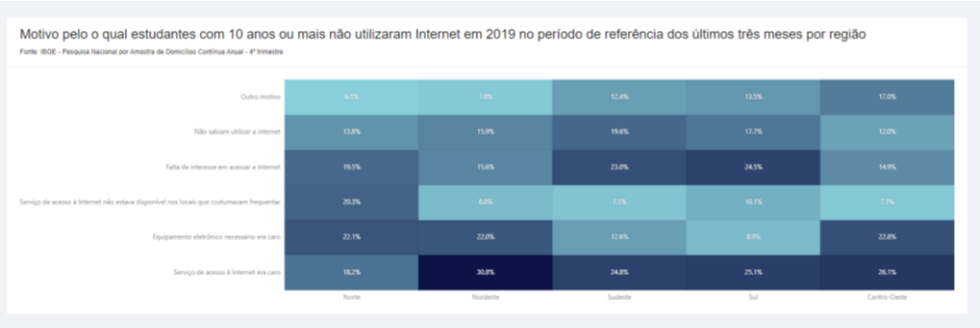

Figura 55: Motivo pelo o qual estudantes com 10 anos ou mais não utilizaram internet em 2019 no período de referência dos últimos três meses por região

A questão seguinte era, apesar de objetiva, mais aberta e já assumimos duas possíveis respostas. Como perguntamos qual região se destaca das outras pelo “Serviço era caro demais”, o entrevistado poderia interpretar como destaque para mais ou para menos em comparação às outras e isso também é um *insight* interessante que tivemos. Nela, reparamos que a maioria vê o destaque do Nordeste em relação às outras, mas uma pequena parte faz destaque ao Norte.

Figura 56: Qual o motivo mais frequente, num geral, para a não utilização da internet?

Na próxima visualização, a respeito do equipamento usado pelos estudantes para acessar a internet, trazemos duas questões. Uma é sobre qual região mais utilizou “Microcomputador ou tablet”, em sua maioria os entrevistados responderam de acordo com o esperado, com exceção, dos mesmos dois que responderam fora do esperado nas primeiras perguntas desta seção.

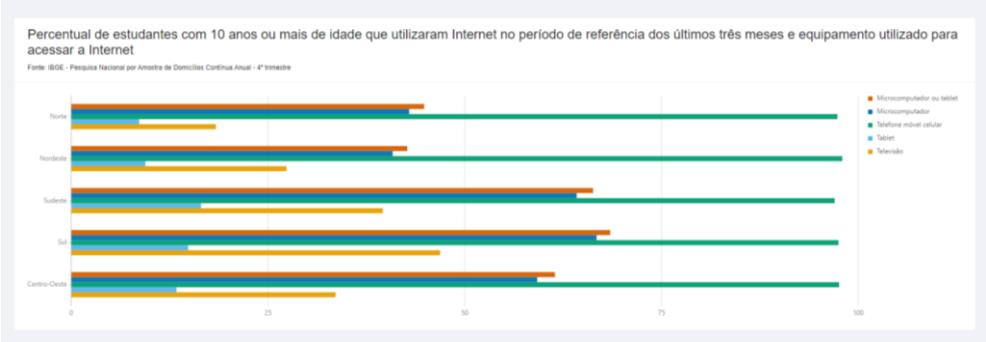

Figura 57: Percentual de estudantes com 10 anos ou mais de idade que utilizaram Internet no período de referência dos últimos três meses e equipamento utilizado para acessar a internet

Figura 58: Qual região utilizou mais “Microcomputador ou tablet”?

Já na segunda pergunta desta visualização, apenas um entrevistado respondeu fora do esperado, assinalando a opção “Não sei”, esse foi o mesmo que fez a observação sobre não saber se o “não sei” representava “não sabia” ou que não era possível responder. Logo, não vimos necessidade de mudar algo nessa visualização.

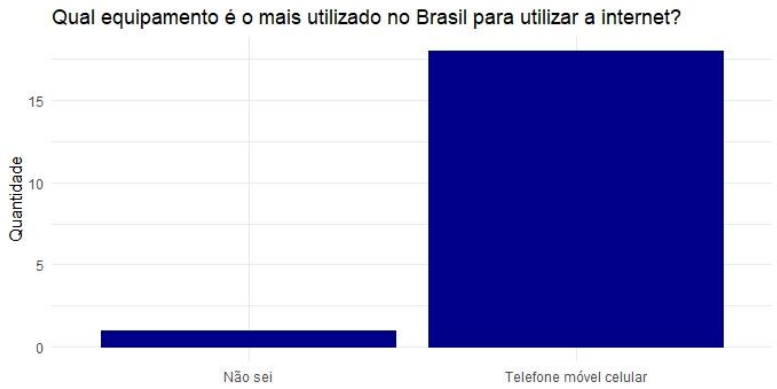

Figura 59: Qual equipamento é o mais utilizado no Brasil para utilizar a internet?

Em seguida, temos mais um gráfico de linhas com o evolutivo do tempo para a utilização de celular para acessar a internet nas regiões. Na primeira pergunta, relativa a qual região apresentou queda na utilização de celular entre 2016 e 2017, a maioria respondeu de acordo com o esperado. Já a segunda pergunta, foi mais interessante, pois, apesar de objetiva, ela é mais aberta, dando ao usuário margem de interpretação. Nela vimos que a maioria respondeu Norte, que é o que mantêm mais estabilidade no sentido de não variar muito, mas uma parte colocou Sudeste, que também apresenta uma estabilidade no sentido de crescimento estável.

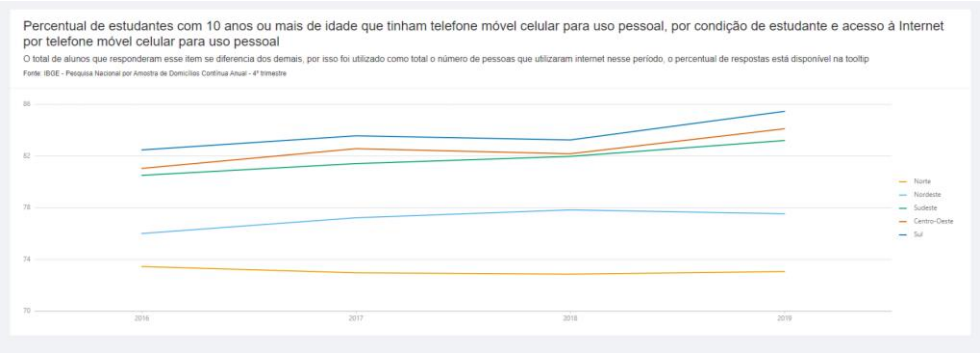

Figura 60: Percentual de estudantes com 10 anos ou mais que tinham telefone móvel celular para uso pessoal, por condição de estudante e acesso à internet por telefone móvel para uso pessoal

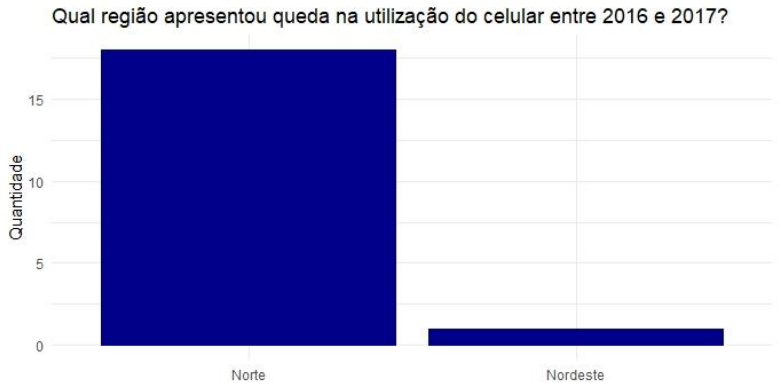

Figura 61: Qual região apresentou queda na utilização do celular entre 2016 e 2017?

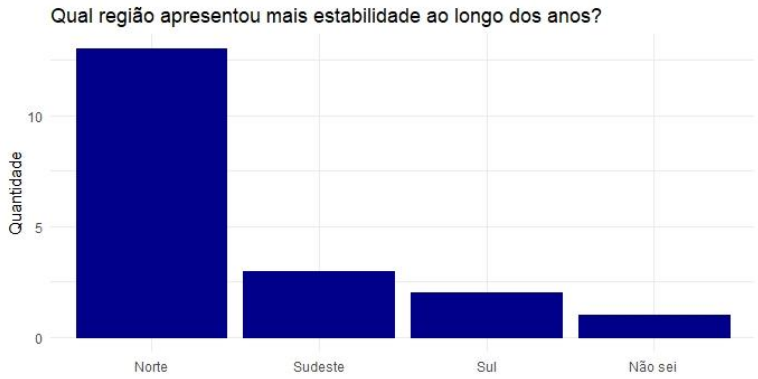

Figura 62: Qual região apresentou mais estabilidade ao longo dos anos?

7 Considerações Finais

Em conclusão, creio que esse trabalho trouxe contribuições para a comunidade acadêmica e para outros possíveis usuários do sistema, pois ele tornou públicos dados importantíssimos para educação do país. Infelizmente não conseguimos responder todas as perguntas que gostaríamos sobre os impactos do ensino remoto nos últimos anos, mas, como comentamos nas seções acima, isso se deve ao fato do baixo investimento em pesquisas que tivemos nos últimos anos e da falta de divulgação dos dados que as instituições já têm hoje. Além disso, esse relatório também evidencia a importância da transparência desses tipos de dados, não só para o público, mas também para instituições que possam agir para diminuir as desigualdades no país.

Acredito que esse trabalho me desafiou a explorar novas áreas de conhecimento e me aprofundar nas áreas que eu já tinha alguma experiência. Aprendi a criar uma aplicação do zero, incluindo todas as configurações iniciais de webpack e dependências, até as últimas finalizações, fazendo o *deploy* e botando o sistema no ar. E, também, tive meu primeiro contato com banco de dados não relacionais. Com essa experiência adquirida, pude crescer profissionalmente e trazer todo o conhecimento que adquiri com Nextjs para o meu trabalho fora da universidade.

Juntamente com o projeto final, cursei com a minha orientadora a matéria de projeto orientado para me aprofundar no universo de *Data Science* e pude trazer esse conhecimento nas análises dos gráficos e visualizações. Além disso, tive minha primeira experiência com a linguagem R, que é mundialmente reconhecida, e pude ter mais contato com questões de estatísticas, que não havia me aprofundado muito durante o curso.

Se eu tivesse que começar o trabalho hoje, com todo o conhecimento que eu adquiri nesses últimos meses, eu daria mais importância para o momento de buscar os dados. Essa parte foi bastante trabalhosa e tomou um tempo bem maior do que eu havia planejado. E, tentaria entrar em contato com instituições que fazem pesquisas privadas, para tentar acesso a mais dados visto a carência de base dados públicos que temos hoje. Investiria, também, mais tempo analisando os dados e tentando construir previsões caso não conseguisse, novamente, dados atuais.

Para quem deseja continuar com essa pesquisa, eu vejo a oportunidade de inserir dados do ENEM de 2020 e 2021, que ainda não foram disponibilizados, e explorar esses dados vendo as possíveis relações com dados de acesso à internet e educacionais já contidos no sistema. Além disso, a página que exibe os resultados do PNAD nos últimos anos pode ser continuada, inserindo os dados quando o IBGE liberar as pesquisas reduzidas nos indicadores selecionados. Os scripts em R para tratamento dos dados estão disponíveis no GitHub do projeto.

Além disso, podem ser inseridos outros dados que não estão diretamente ligados à educação, mas que impactam indiretamente, como dados socioeconômicos. Caso um próximo aluno ou aluna tenha acesso a microdados onde seja possível fazer recortes mais específicos como de gênero, raça/cor e zona de residência (urbana/rural), seria interessante adicionar essas informações também.

Acredito que a informação é o primeiro caminho para a mudança, e a tecnologia pode ajudar tornando a informação acessível ao maior número de pessoas.

8 Referências bibliográficas

1. IBGE. **USO DE INTERNET, TELEVISÃO E CELULAR NO BRASIL**. Disponível em: https://educa.ibge.gov.br/jovens/materias-especiais/20787-uso-de-internet-televisao-e-celular-no-brasil.html. Acesso em: 2 abr. 2020.
2. GUERRA, Susana Cordeiro; RIOS-NETO, Eduardo L. G.. **IBGE sai em defesa do orçamento do Censo 2021**. Disponível em: https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/30350-ibge-sai-em-defesa-do-orcamento-do-censo-2021. Acesso em: 2 abr. 2020.
3. **Corte do Orçamento pode inviabilizar Censo 2021; entenda a importância da pesquisa**. Disponível em: https://g1.globo.com/economia/noticia/2021/03/26/corte-do-orcamento-pode-inviabilizar-censo-2021-entenda-a-importancia-da-pesquisa.ghtml. Acesso em: 2 abr. 2020.
4. L, Malcolm. **What's The Difference Between NextJS and Create-React-App?** Disponível em: https://frontend-digest.com/whats-the-difference-between-nextjs-and-create-react-app-11b55650a612. Acesso em: 4 abr. 2020.
5. UDAIPURWALA, Burhanuddin. **Using Ant Design with NextJS (custom variables for Ant Design).** [S. l.], 18 jun. 2020. Disponível em: https://dev.to/burhanuday/using-ant-design-with-nextjs-custom-variables-for-ant-design-57m5. Acesso em: 17 abr. 2021.
6. **INTEGRATING MongoDB Into Your NextJS App**. [*S. l.*]: MongoDB, 2021. Disponível em: https://www.youtube.com/watch?v=aAupumVpqcE. Acesso em: 1 jun. 2021.
7. **RETRATOS da Educação no Contexto da Pandemia do Coronavírus: Um olhar sobre múltiplas desigualdades**. [S. l.], Outubro 2021. Disponível em: https://www.fcc.org.br/fcc/wp-content/uploads/2021/02/Retratos-da-Educacao-na-Pandemia_digital-_outubro20.pdf. Acesso em: 23 jun. 2021.
8. Brasil. Instituto Nacional de Estudos e Pesquisas Anísio Teixeira. **Saeb 2019: indicador de nível socioeconômico do Saeb 2019**: nota técnica. Brasília, DF: Inep, 2021

9. STATISTICAL Computing and Graphics: Violin Plots: A Box Plot-Density Trace Synergis. In: HINTZE, Jerry L.; NELSON, Ray D. **The American Statistician.** [S. l.: s. n.], 1998.
10. MOL, Mike. **A Color-Safe Palette**. [S. l.], 11 fev. 2018. Disponível em: https://mikemol.github.io/technique/colorblind/2018/02/11/color-safe-palette.html. Acesso em: 13 nov. 2021.
11. HOLTZ, Yan; HEALY, Connor**. FROM Data to Viz**: From Data to Viz leads you to the most appropriate graph for your data. It links to the code to build it and lists common caveats you should avoid.. [S. l.], 2018. Disponível em: https://www.data-to-viz.com/. Acesso em: 21 nov. 2021.
12. A Reader on Data Visualization. 2019. Santa Clara University. Disponível em: https://mschermann.github.io/data_viz_reader/ Acesso em: 21 nov. 2021.

Apêndices

A. Questionário de testes com usuários

Seção 1

Título: Testes com usuário - Projeto Final

Descrição:

Título do Projeto: Painel Interativo para acompanhamento e análise de dados a respeito dos impactos do ensino remoto no Brasil

Professora Responsável: Simone DJ Barbosa

Aluno responsável: Ana Carolina Junger

Instituição a que pertencem a professora e o aluno responsável: PUC-Rio

Telefones para contato: (21) 98225-8479

E-mail: carolfjunger@gmail.com

Você está sendo convidado a participar da pesquisa para o projeto “Painel Interativo para acompanhamento e análise de dados a respeito dos impactos do ensino remoto no Brasil", de responsabilidade da aluna Ana Carolina Junger. Este trabalho faz parte da disciplina ENG1133 - PROJ GRAD EM ENG COMPUTACAO II do Departamento de Informática da PUC-Rio.

Todas as perguntas e todos os assuntos tratados no questionário levarão em conta sua experiência com o tema. O estudo terá seus dados coletados estatisticamente para os fins da pesquisa.

Durante todo o processo, se você se sentir desconfortável por qualquer motivo que seja, poderá não responder ou até mesmo cancelar a sua participação no mesmo momento sem nenhuma necessidade de explicação a este pesquisador ou a qualquer outra parte relacionada a esta pesquisa. Caso ocorra a interrupção dos questionários, todos os dados coletados até então serão completamente apagados e não entrarão para a análise dos dados da pesquisa.

Por fim, nós garantimos a sua confidencialidade e privacidade. Não iremos incluir, sob nenhuma hipótese ou circunstância, o nome ou outras informações pessoais, de qualquer participante ao qual tivemos contato para a realização

deste trabalho, mesmo que o participante tenha recusado ou desistido da entrevista.

Como esta pesquisa é de participação voluntária, sem nenhum custo para o participante, seu consentimento poderá ser retirado a qualquer tempo, sem nenhuma espécie de prejuízo ou qualquer outra penalização. Além disso, esta pesquisa também não irá fornecer nenhum pagamento, em nenhuma forma, para aqueles que participarem do estudo. Para sanar qualquer dúvida referente aos procedimentos, riscos, benefícios e outros assuntos relacionados com a pesquisa, basta entrar em contato com os pesquisadores responsáveis pela forma desejada presente no topo deste termo.

Ao prosseguir neste formulário, será considerado que você consentiu com a coleta e uso dos dados conforme o termo acima.

Perguntas:

1. Você afirma o seu consentimento de acordo com os termos descritos acima?
    a. Dou meu consentimento para o uso dos dados fornecidos conforme os termos descritos acima
    b. NÃO dou meu consentimento para o uso dos dados fornecidos conforme os termos descritos acima

Seção 2

Título: Informações Gerais do participante

Descrição: Antes de começar o estudo, temos algumas perguntas para lhe fazer sobre seu perfil, para nos ajudar na análise de suas respostas. Por favor, responda as perguntas a seguir:

Perguntas:

1. Qual é a sua faixa etária?
    a. 18 - 24 anos
    b. 25 - 34 anos
    c. 35 - 44 anos
    d. 45 - 54 anos
    e. 55 - 64 anos
    f. 65+ anos
2. Qual o seu nível de escolaridade?

a. Ensino médio (Completo)
b. Ensino superior (Incompleto)
c. Ensino superior (Completo)
d. Pós - graduação (Incompleto)
e. Pós - graduação (Completo)

3. Qual a sua área de graduação? (Ex: Informática, Design)
    a. Texto de resposta curta
4. Em média, com que frequência você lê e interpreta gráficos?
    a. Nunca
    b. Todo ano
    c. Todo semestre
    d. Todo trimestre
    e. Todo mês
    f. Toda semana
5. Qual contato você tem com os modelos de gráfico abaixo? Opções: [Nenhum, Baixo, Moderado, Alto, Especialista]
    a. Gráfico de linhas
    b. Gráfico de colunas
    c. Gráfico de pizza
    d. Mapa de calor
    e. Gráfico de bolhas
    f. Diagrama de caixa (Box plot)

Seção 3

Título: Analisando a plataforma

Descrição: A seguir você deve entrar na plataforma pelos links determinados na descrição das próximas sessões e responder as perguntas subsequentes. É de extrema importância que você faça esse teste em um computador, pois a plataforma ainda não foi projetada para uma boa visualização no celular e/ou tablet.

Seção 4

Título: Página PNAD - 2020 edição COVID-19

Descrição: Entre em https://projeto-final-carolfjunger.vercel.app/pnad2020 e responda as perguntas abaixo em relação a cada gráfico.

Perguntas:

Modelo de aula por região. A respeito desta visualização responda:

1. Qual o modelo mais adotado?
    a. Aulas presenciais
    b. Aulas presenciais parciais
    c. Sem aulas presenciais, mas curso presencial
    d. Não sei
2. Qual região tem mais proporção de alunos com algum tipo de aula presencial?
    a. Norte
    b. Nordeste
    c. Sudeste
    d. Sul
    e. Centro-oeste
    f. Não sei

Distribuição da realização de tarefas em casa. A respeito desta visualização responda:

1. O que você consegue interpretar nesse gráfico?
    a. Texto de resposta longa
2. Os alunos do Centro-Oeste receberam mais atividades que os do Sul?
    a. Sim
    b. Não
    c. Não sei
    d. Não é possível responder

Motivo para os alunos não realizarem as tarefas. A respeito desta visualização responda:

1. Qual o motivo mais frequente dos alunos não utilizarem internet?
    a. Não tinha computador / tablet / celular disponível
    b. Não tinha acesso à internet ou a qualidade dela era insuficiente
    c. Por problemas de saúde da própria pessoa
    d. Tinha que cuidar dos afazeres domésticos, do(s) filhos ou de outro(s) parentes
    e. Não conseguiu se concentrar
    f. Outro
    g. Não sei

Como os alunos distribuem suas tarefas. A respeito desta visualização responda:

1. Como você descreveria as informações desse gráfico?
    a. Texto de resposta longa
2. A grande maioria dos estudantes dedica quantas horas por dia?
    a. 5 horas ou mais
    b. De 2 horas a menos de 5 horas
    c. De 1 hora a menos de 2 horas
    d. Menos de 1 hora
    e. Não sei
3. E quantos dias na semana?
    a. 1 dia
    b. 2 dias
    c. 3 dias
    d. 4 dias
    e. 5 dias
    f. 6 dias
    g. Não sei

Seção 5

Título: Acesso à internet PNAD Contínua 2016-2019

Descrição:

Entre em https://projeto-final-carolfjunger.vercel.app/acessoAInternet e responda as perguntas abaixo em relação a cada gráfico.

Perguntas:

Percentual de estudantes com 10 anos ou mais de idade que utilizaram Internet no período de referência dos últimos três meses. A respeito dessa visualização responda

1. Qual região do país teve mais acesso à internet em 2017?
    a. Norte
    b. Nordeste
    c. Sudeste
    d. Sul
    e. Centro-oeste
    f. Não sei
2. e em 2019?
    a. Norte
    b. Nordeste

c. Sudeste
d. Sul
e. Centro-oeste
f. Não sei

3. Qual é a diferença percentual (aproximada) entre as regiões Norte e Centro-Oeste em 2019?
    a. 17,7%
    b. 23,4%
    c. 42,2%
    d. 55,3%
    e. 34,7%
    f. Não sei

Motivo pelo o qual estudantes com 10 anos ou mais não utilizaram Internet em 2019 no período de referência dos últimos três meses por região. A respeito dessa visualização responda.

1. Qual o motivo mais frequente, num geral, para a não utilização da internet?
    a. Não sabiam utilizar a internet
    b. Falta de interesse em acessar a internet
    c. Serviço de acesso à internet não estava disponível nos locais que costuma frequentar
    d. Equipamento eletrônico era caro
    e. Serviço de acesso a internet era caro
    f. Não sei
2. Qual região se destaca das outras para o “Serviço era caro demais”?
    a. Norte
    b. Nordeste
    c. Sudeste
    d. Sul
    e. Centro-oeste
    f. Não sei

Percentual de estudantes com 10 anos ou mais de idade que utilizaram Internet no período de referência dos últimos três meses e equipamento utilizado para acessar a Internet. A respeito dessa visualização responda

1. Qual região utilizou mais “Microcomputador ou tablet”?
    a. Norte
    b. Nordeste

c. Sudeste
d. Sul
e. Centro-oeste
f. Não sei

2. Qual equipamento é o mais utilizado no Brasil para utilizar a internet?
    a. Microcomputador ou tablet
    b. Microcomputador
    c. Telefone móvel celular
    d. Tablet
    e. Televisão
    f. Não sei

Percentual de estudantes com 10 anos ou mais de idade que tinham telefone móvel celular para uso pessoal, por condição de estudante e acesso à Internet por telefone móvel celular para uso pessoal. A respeito dessa visualização responda

1. Qual região apresentou queda na utilização do celular entre 2016 e 2017?
    a. Norte
    b. Nordeste
    c. Sudeste
    d. Sul
    e. Centro-oeste
    f. Não sei
2. Qual região apresentou mais estabilidade ao longo dos anos?
    a. Norte
    b. Nordeste
    c. Sudeste
    d. Sul
    e. Centro-oeste
    f. Não sei

Seção 6

Título: Fim

Descrição: Obrigada por participar!

Perguntas:

1. Dúvidas, sugestões e/ou comentários?
    a. Texto de resposta longa