

Marcelo Correia da Silva

**Desenvolvimento de sistemas interativos–responsivos para
a análise e comunicação de dados**

Recife

19 de setembro de 2023



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Desenvolvimento de sistemas interativos–responsivos para a análise e comunicação de dados

Tese apresentada como requisito exigido, conforme decisão N^o 044/2021 do CCD do Programa de Pós-Graduação em Biometria e Estatística Aplicada, para obtenção do título de Doutor em Biometria e Estatística Aplicada.

Área de concentração: Biometria e Estatística Aplicada

Orientador: Dr. Rômulo Simões Cezar Menezes

Coorientador: Dr. Antonio Samuel Alves da Silva

Recife

19 de setembro de 2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

S586d

Silva, Marcelo Correia da Silva

Desenvolvimento de sistemas interativos–responsivos para a análise e comunicação de dados / Marcelo Correia da Silva Silva. - 2023.

123 f. : il.

Orientador: Romulo Simoes Cezar Menezes.

Coorientador: Antonio Samuel Alves da Silva.

Inclui referências e apêndice(s).

Tese (Doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, 2023.

1. R Programming Language. 2. RStudio. 3. Shiny. 4. Flexdashboard -R Markdown. 5. interactive–responsive. I. Menezes, Romulo Simoes Cezar, orient. II. Silva, Antonio Samuel Alves da, coorient. III. Título

CDD 519.5

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Desenvolvimento de sistemas interativos–responsivos para a análise e comunicação de dados

Marcelo Correia da Silva

Tese apresentada como requisito exigido, conforme decisão Nº 044/2021 do CCD do Programa de Pós-Graduação em Biometria e Estatística Aplicada, para obtenção do título de Doutor em Biometria e Estatística Aplicada.

Orientador:

Dr. Rômulo Simões Cezar Menezes
Orientador

Banca examinadora:

Antonio Samuel Alves da Silva
Universidade Federal Rural de
Pernambuco

José Rodrigo Santos Silva
Universidade Federal de Sergipe

Aldo Torres Sales
Universidade Federal de Pernambuco

Moacyr Cunha Filho
Universidade Federal Rural de
Pernambuco

*Este trabalho é dedicado aos meus pais e
a minha filha Maria Sophia, que de
pequeninha hoje siga o exemplo e se
torne grandiosa em conhecimento.*

Agradecimentos

À Deus.

Ao professor Dr. Antonio Samuel Alves da Silva, pelo curso de programação em R e por me mostrar o ponto de partida para o meu desenvolvimento dentro do PPGBEA, pelas orientações, pelos ensinamentos e confiança em meu potencial.

Ao professor Dr. Rômulo Simões Cezar Menezes, pela oportunidade, pelas orientações, por todos os ensinamentos e, principalmente, pela confiança depositada em mim.

Ao professor Dr. Moacyr Cunha Filho, pelas orientações, conversas e os diversos incentivos desprendidos ao longo dessa caminhada. Pela parceria e confiança.

Ao professor Dr. Adauto José Ferreira de Souza, pelas orientações e conselhos sempre valioso, deste o processo de seleção no PPGBEA, ou melhor, desde o curso de graduação em Licenciatura em Física da UFRPE, até o presente momento.

A minha mãe, em memória, pela dedicação, apoio incondicional em tudo que seria de bom e opositora em tudo que julgava ser prejudicial a mim. Pelo direcionamento me dado, sobre a óptica de que por meio da educação eu poderia ir mais longe, mesmo ela não tendo esta oportunidade.

Ao meu pai, pelo apoio e dedicação ao longo da minha vida até hoje, pela sua competência como o formador do meu caráter e incentivador dos meus estudos, mesmo não tendo tido oportunidade para si.

Ao Colégio Dom Agostinho Ikas da UFRPE e, em particular, ao Diretor Geral, professor Michel Saturnino Barbosa, pelo apoio, confiança e parceria ao longo dessa jornada.

Aos membros da Comissão de Vestibular do Colégio Dom Agostinho Ikas da UFRPE: professoras Marcella Feitosa dos Santos, Ana Paula Costa de Lucena, Giselli Cristine Nunes de Santana e ao professor Felipe Matheus Teles de Vasconcelos, por contribuírem com as percepções sobre o sistema *Select CODAI* e, principalmente, pela confiança em o aplicar nos diversos processos seletivos.

Ao meu grande amigo, Laércio de Lima, pedagogo e professor de Português, companheiro indispensável de longas conversas e corretor deste trabalho nas horas vagas.

Aos meus grandes amigos do Intensivo Curso Preparatórios em São Lourenço da Mata: professores Edinei Novais, Fábio Nascimento, Hedymur França, Jonath Tavares, Laércio Lima e Lucas Wallemberg, os quais cederam espaço físico, sem quaisquer cerimônias, para que eu pudesse desenvolver este trabalho nas suas diversas etapas onde havia a necessidade de um escritório tranquilo para codificar ou escrever.

A minha filhinha, Maria Sophia (Buchinho), incentivando-me pelo simples fato de eu pensar em dar o exemplo, pelos momentos que me alegrou – “papai vamos brincar!?”. Pela ajuda, no meu coloco, digitando no teclado algumas palavras ditadas por mim que estão escritas neste trabalho.

Aos meus colegas da Escola João Cavalcante Petribu de Paudalho/PE, sobretudo ao professor Rogério Paulo do Monte, com os quais convivi por longos anos e recebi apoio incondicional.

À Universidade Federal Rural de Pernambuco e, em particular, ao Programa de Pós-Graduação em Biometria e Estatística Aplicada, seu corpo docente, discente e técnicos administrativos, proporcionando sempre um ambiente acolhedor e saudável.

Aos órgãos de fomento a pesquisa.

A todas as pessoas que, ao longo de minha vida, se tornaram grandes amigos.

*“Mesmo muitas vezes desacreditado
e ignorado por todos, não posso desistir,
pois para mim, vencer é nunca desistir.”
(Albert Einstein)*

Resumo

É inegável a evolução nos métodos para coleta de dados, levando-se a uma crescente e acelerada disponibilização desses nas mais variadas áreas do conhecimento. Neste sentido, é evidente a necessidade de haver uma evolução proporcional das ferramentas associadas à governança de dados, já que o processo de prospecção e análise depende inteiramente dessa. Então, considerando-se o quesito da governança de dados associada a comunicação/visualização, é necessário seguir, basicamente, as etapas de importação, organização, transformação, filtragem, modelagem e por fim a apresentação/comunicação. Evidentemente que na execução das etapas citadas há a aplicação de ferramentas estatísticas visando múltiplos interesses e, portanto, exigindo o uso de poder computacional, tanto no que diz respeito ao *hardware* como ao *software*, sendo que em geral, os *softwares* para análise de dados tem um custo significativo. Neste sentido, este trabalho demonstrará que os sistemas interativos-responsivos, voltados às etapas de governança de dados que correspondem a importação, transformação, filtragem, modelagem e comunicação/visualização, ao se implementarem as ferramentas de análises e saídas pertinentes à aplicação, constitui uma ferramentas eficientes, de fácil acesso e de baixo custo para a análise da grande demanda de dados que se apresenta na atualidade. As áreas de educação e saúde estão entre as prioridades para o fortalecimento do desenvolvimento sustentável no Brasil e em muitos outros países e, por esse motivo, foram foco da presente tese para aplicação dos sistemas desenvolvidos. Assim, no trabalho, desenvolveram-se sistemas de interesses diversos, visando elucidar os métodos de implementação, cada qual considerando-se suas particularidades. Em primeiro lugar implementou-se o código fonte para acessar, tratar e modelar os dados das inscrições do processo seletivo do Colégio Dom Agostinho Ikas da Universidade Federal Rural de Pernambuco, onde são construídas a lista de classificação em documento .pdf, bem como a apresentação dos dados de inscrição e resultados de classificação através de sistema interativo-responsivo via *web*. Além disso, desenvolveu-se o código fonte para análise dos dados da pandemia de Covid-19, disponibilizadas pelo Ministério da Saúde do Brasil para o Estado de Pernambuco, sendo esta motivada pela pertinência significativa do tema associado aos dados, bem como o fato de oportunizar a análise dos dados oficiais disponibilizados pelo Ministério da Saúde do Brasil, sejam esses de boa qualidade ou não. Em suma, as pesquisas envolvendo os sistemas em tela são complementares no que o trabalho propõe demonstrar, já que o primeiro tem relativo grau de sofisticação em relação a processamento e tratamento dos dados, enquanto o segundo versa sobre uma robusta base de dados.

Palavras-chaves: Linguagem de Programação R; RStudio; Shiny; Flexdashboard; R Markdown; interativo-responsivo

Abstract

The evolution of methods for data collection is undeniable, leading to a growing and accelerated availability of these in the most varied areas of knowledge. This context, the need for a proportional evolution of the tools associated with data governance is evident, since the prospecting and analysis process depends entirely on it. Therefore, considering the issue of data governance associated with communication/visualization, it is necessary to basically follow the steps of import, organization, transformation, filtering, modeling and finally presentation/communication. Evidently, in the execution of the mentioned steps, there is the application of statistical tools aiming at multiple interests and, therefore, demanding the use of computational power, both in terms of hardware and software, and in general, software for data analysis has a significant cost. In this sense, this work will demonstrate that interactive–responsive systems, focused on the data governance stages of import, transformation, filtering, modeling, and communication/visualization, when implementing relevant analysis tools and outputs, constitute efficient, easily accessible, and low cost tools for analyzing the significant data demand present in the present time. The fields of education and healthcare are among the priorities for strengthening sustainable development in Brazil and many other countries, and for this reason, they were the focus of the present thesis for the application of the developed systems. Thus, in this work, systems of diverse interests were developed to elucidate the implementation methods, each considering its own particularities. Firstly, the source code was implemented to access, process, and model the registration data for the selection process at Colégio Dom Agostinho Ikas of the Universidade Federal Rural de Pernambuco. This includes generating the ranking list in a .pdf document, as well as presenting the registration data and classification results through an interactive–responsive via the web. Additionally, the source code was developed for the analysis of Covid-19 pandemic data provided by the Ministry of Health of Brazil for the state of Pernambuco. This was motivated by the significant relevance of the topic associated with the data, as well as the opportunity to analyze the official data provided by the Brazilian Ministry of Health, regardless of their quality. In summary, the research involving the aforementioned systems are complementary in what the work proposes to demonstrate, since the first has a relative degree of sophistication in relation to data processing and treatment, while the second deals with a robust database.

Key-words: R Programming Language; RStudio; Shiny; Flexdashboard; R Markdown; interactive–responsive.

Lista de Figuras

Figura 1 – Criadores do R.	6
Figura 2 – Tela do RStudio indicando criação de um aplicativo Shiny.	11
Figura 3 – Janela no RStudio para inserção das informações básicas para a criação de um aplicativo Shiny.	12
Figura 4 – <i>Script</i> básico no RStudio para criação de um aplicativo Shiny.	13
Figura 5 – <i>Script</i> básico no RStudio para criação de um aplicativo Shiny.	13
Figura 6 – Processamento básico de um documento R Markdown.	14
Figura 7 – Diagrama que exhibe a estrutura de sistemas desenvolvidos em R.	18
Figura 8 – Diagrama que elucida o fluxo para a ocupação das vagas, considerando-se o sistemas de cotas. No presente diagrama, exemplificam-se os valores considerando a oferta de 36 vagas, tal como apresentado na Tabela 2. Quando há repetição na numeração das etapas, considera-se que estas são independentes e podem ocorrer em paralelo.	26
Figura 9 – Aba da Google Sheets intitulada “Gerais”, com dados de configuração do processo seletivo e do sistema. Tem a finalidade de configurar as informações básicas do processo seletivo.	31
Figura 10 – Aba da Google Sheets intitulada “Links”, com dados de configuração do sistema. Tem a finalidade conter links de todas as planilhas/abas com dados utilizados pelo sistema. Desfigura-se as informações da coluna “Information” por questões de segurança. Os dados da coluna “Information” estão ocultos por questões de segurança.	32
Figura 11 – Aba da Google Sheets intitulada “Vagas”, com dados de configuração do sistema. Tem a finalidade de configurar os nomes dos cursos oferecidos, seus turnos e suas respectivas quantidades de vagas, considerando-se cada tipo de concorrência (as cotas).	33
Figura 12 – Aba da Google Sheets intitulada “Pesos”, com dados de configuração do sistema. Tem a finalidade permitir à CPV configurar os componentes curriculares exigidos, bem como os seus pesos para o cálculo do <i>Score</i> dos candidatos.	34

Figura 13 – Diagrama elucidativo do arcabouço de variáveis envolvidas e manipuladas pelo sistema desenvolvido em R . À esquerda (1), tem-se informações associadas à configuração do processo seletivo, já à direita (2), tem-se os dados das inscrições e outros de interesse.	38
Figura 14 – Diagrama elucidativo da estrutura do sistema em R, observando-se a utilização dos diversos <i>scripts</i> (.R) e documentos (.Rmd) envolvidos no funcionamento do software.	42
Figura 15 – Aba do sistema “Dados gerais”, configura-se como a aba padrão, oferece representações gráficas e <i>dashboards</i> associados às informações gerais sobre os dados das inscrições e datas relevantes do processo seletivo. Os detalhes dos elementos enumerados se encontram na Tabela 4.	45
Figura 16 – Aba do sistema “Inscrições Diárias”, onde é possível acompanhar a dinâmica diária das inscrições no processo seletivo e em tempo real. Os detalhes dos elementos enumerados se encontram na Tabela 5.	47
Figura 17 – Janela de login para área exclusiva e privativa aos membros da Comissão Permanente de Vestibular do CODAI/UFRPE.	48
Figura 18 – Aba do sistema “Exclusivo Comissão”, destinada aos membros da Comissão Permanente de Vestibular, e que disponibiliza os resultados da classificação e informações relevantes para o Departamento de Registro Escolar. Os detalhes dos elementos enumerados se encontram na Tabela 6. Os dados das colunas “NOME”, “EMAIL”, “CELULAR” E “CPFO” estão ocultos por questões de privacidade.	49
Figura 19 – Capa e sumário empregados de forma típica a todos os relatórios contendo os resultados de classificação do processo seletivo. A capa em questão apresenta a data de 24/03/2023 por ter se efetuada o <i>download</i> da via o sistema nesta data.	51
Figura 20 – Página de notas/observações da Comissão Permanente de Vestibular empregada no relatório contendo o resultado definitivo da classificação do processo seletivo. No entanto, para todos os demais relatórios o modelo é análogo, modificando-se o título e as informações pertinentes ao tipo de divulgação.	52
Figura 21 – Início da página mostrando a composição dos resultados da classificação do processo seletivo. É composto por muitas páginas, de modo que cada curso/turno corresponde a uma secção. Os dados da coluna “NOME” estão ocultos por questões de privacidade.	53

Figura 22 – Diagrama da estrutura do sistema em R desenvolvido para os dados do MS da Covid-19. Nota-se que há um <i>scripts</i> (.R) e sete documentos R Markdown envolvidos no funcionamento do software.	58
Figura 23 – Dados originais dos casos novos de Covid-19 para a cidade de Recife de 07/04/2021 a 07/07/2021.	62
Figura 24 – Dados com filtragem suprimindo os <i>outliers</i> ($z_i > 2$) dos casos novos da Covid-19 para a cidade de Recife de 07/04/2021 a 07/07/2021.	64
Figura 25 – Dados com filtragem substituindo os <i>outliers</i> ($z_i > 2$) pelo inteiro mais próximo a média dos casos novos da Covid-19 para a cidade de Recife de 07/04/2021 a 07/07/2021.	65
Figura 26 – Detalhamento da barra de manu lateral presente em todas as abas do sistema.	67
Figura 27 – Apresentação de uma página típica do sistema. Os detalhes dos elementos enumerados se encontram na Tabela 9.	68
Figura 28 – Mapa de cores interativo–responsivo da aba casos acumulados acompanhado do <i>date input</i> para escolha da data de visualização. Observa-se que a data escolhida é 26 de abril de 2023, bem como foi clicado no mapa a cidade de São Lourenço da Mata.	69
Figura 29 – Diagrama das relações de interatividade–reatividade na aba de Óbitos Novos, configurando-se uma situação típica para o aplicativo. Os detalhes indicativos da interatividade–reatividade são apresentados na Tabela 10.	70
Figura 30 – Principais elementos da aba Casos Acumulados. Os objetos numerados são descritos na Tabela 11. A exibição corresponde ao dia 15/06/2022 para o mapa de cores e a plotagem dos gráficos de linha são para capital pernambucana de 27/03/2020 a 26/04/2023.	73
Figura 31 – Principais elementos da aba Casos Novos do sistema. Os objetos numerados são descritos na Tabela 12. A exibição corresponde ao dia 15/06/2022 no mapa de cores e a plotagem dos gráficos de linha, bem como das métricas nos <i>dashboards</i> são para a cidade de Gravata de 27/03/2020 a 31/01/2022.	75
Figura 32 – Principais elementos da aba Óbitos Acumulados do sistema. Principais elementos da aba Óbitos Acumulados. Os objetos numerados são descritos na Tabela 13. A exibição corresponde ao dia 21/09/2021 para o mapa de cores e a plotagem dos gráficos de linha são para o município do Cabo de Santo Agostinho de 27/03/2020 a 15/06/2023.	78

- Figura 33 – Principais elementos da aba Óbitos Novos do sistema. Os objetos numerados são descritos na Tabela 14. A exibição corresponde ao dia 15/06/2022 no mapa de cores e a plotagem dos gráficos de linha, bem como das métricas nos *dashboards* são para a cidade de São Lourenço da Mata de 27/03/2020 a 15/06/2022. 81
- Figura 34 – Principais elementos da aba Rank Municípios do sistema. Os objetos numerados são descritos na Tabela 15. A exibição corresponde ao intervalo de datas de 27/03/2020 a 03/06/2022 e foi selecionado a visualização do 1º ao 15º município com maior média de casos novos por 10⁵ habitantes. Aba de classificação em ordem crescente da média dos casos e óbitos novos por 100.000 habitantes. 84
- Figura 35 – Tela inicial do sistema desenvolvido pela Johns Hopkins University Center for Systems Science and Engineering para acompanhar a pandemia de *Sars-Cov-2*. Acessado em 22/08/2023. Link de acesso: <https://coronavirus.jhu.edu/map.html> 86
- Figura 36 – Tela inicial do sistema desenvolvido Oliveira (2022) expondo *dashboards* para os dados da Covid-19 do Estado do Espírito Santo. Link de acesso: GitHub: https://github.com/DaniloSI/covid-19_es 88

Lista de tabelas

Tabela 1 – Oferta das vagas dos processos seletivos do CODAI/UFRPE, para ingresso nos semestres letivos de 2022.1 e 2022.2.	23
Tabela 2 – Sistema de cotas aplicado no processo seletivo do CODAI/UFRPE considerando-se a Lei N ^o 12.711, de 29 de agosto de 2012 (BRASIL, 2012). Apresentam-se os números, como exemplo, associado a oferta de 36 vagas.	24
Tabela 3 – Pesos dos componentes curriculares para o cálculo do <i>Score</i> para os diversos cursos ofertados no processo seletivo do semestre letivo 2021.1. Onde: POR: Português, MAT: Matemática, HIS: História, GEO: Geografia, QUI: Química e BIO: Biologia	28
Tabela 4 – Detalhamento dos elementos indicados na Figura 15 para a aba “Dados Gerais”. Usa-se o símbolo  para um objeto interativo e  para responsivo.	46
Tabela 5 – Detalhamento dos elementos indicados na Figura 16 para a aba “Inscrições Diárias”.	47
Tabela 6 – Detalhamento dos elementos indicados na Figura 18.	50
Tabela 7 – Principais argumentos do função <i>Zoo::rollmean</i> para extração da média móvel simples.	61
Tabela 8 – Recorte das observações enfatizando os <i>outliers</i> evidenciados no gráfico da Figura 23, bem como os <i>z-Scores</i> (z_i) calculados para os respectivos x_i	64
Tabela 9 – Detalhamento dos elementos indicados na Figura 27.	68
Tabela 10 – Detalhamento das relações dos objetos de interatividade-reatividade presentes na Figura 29. Usa-se o símbolo  para um objeto interativo e  para reativo.	71
Tabela 11 – Detalhamento dos elementos apresentados na Figura 30, para a aba Casos Acumulados. Usa-se o símbolo  para um objeto interativo e  para reativo.	74
Tabela 12 – Detalhamento dos elementos apresentados na Figura 31, para a aba Casos Novos. Usa-se o símbolo  para um objeto interativo e  para reativo.	76

Tabela 13 – Detalhamento dos elementos apresentados na Figura 32, para a aba Óbitos Acumulados. Usa-se o símbolo  para um objeto interativo e  para reativo.	79
Tabela 14 – Detalhamento dos elementos apresentados na Figura 33, para a aba Óbitos Novos. Usa-se o símbolo  para um objeto interativo e  para reativo.	81
Tabela 15 – Detalhamento dos elementos apresentados na Figura 34, para a aba Rank Municípios. Usa-se o símbolo  para um objeto interativo e  para reativo.	84

Sumário

1	Introdução	1
2	Objetivos	5
2.1	Geral	5
2.2	Específicos	5
3	Revisão de Literatura	6
3.1	A linguagem de programação R	6
3.2	IDE's – RStudio	8
3.3	<i>Package</i> do R	9
3.3.1	O <i>package</i> Shiny	9
3.3.2	O R Markdown	14
3.3.3	O <i>package</i> Flexdashboard	15
3.3.4	Estrutura básica de sistemas em R Markdown e outros <i>packages</i>	17
4	Sistema de classificação do processo seletivo do CODAI/UFRPE	21
4.1	Introdução	21
4.2	Metodologia	22
4.2.1	Processo seletivo do CODAI e a oferta de vagas	22
4.2.2	O sistema de cotas	23
4.2.3	Fluxo para a ocupação das vagas	25
4.2.4	Métricas para classificar	27
4.2.5	Coleta e armazenamento dos dados de configuração do processo seletivo	29
4.2.6	Coleta e armazenamento dos dados das inscrições	35
4.2.7	Arquitetura do sistema em R	39
4.3	Resultados	44
4.3.1	O <i>site web</i> : Sistema interativo–responsivo <i>web</i>	44
4.3.1.1	Aba <i>Dados gerais</i>	45
4.3.1.2	Aba <i>Inscrições diárias</i>	46
4.3.1.3	Aba <i>Exclusivo Comissão</i>	48
4.3.2	Relatórios com resultado de classificação	50
4.4	Conclusões e considerações finais	53
5	Sistema dos dados da Covid-19 para o Estado de Pernambuco	55

5.1	Introdução	55
5.2	Metodologia	56
5.2.1	Dados da Covid-19 do Ministério da Saúde do Brasil	56
5.2.2	Arquitetura do sistema em R	57
5.2.3	Métricas utilizadas	58
5.2.3.1	Média móvel simples e o <i>Package Zoo</i>	60
5.2.4	Implementação do z -Score aos dados	61
5.3	Resultados e discussão	65
5.3.1	O Sistema interativo-responsivo local ou <i>web</i>	65
5.3.2	A barra lateral de menu – <i>Sidebar</i>	66
5.3.3	Página típica do sistema	67
5.3.4	Seleção de dados por município/cidade	69
5.3.5	Diagrama de interatividade-reatividade	70
5.3.6	Aba Casos Acumulados	73
5.3.7	Aba Casos Novos	75
5.3.8	Aba Óbitos Acumulados	78
5.3.9	Aba Óbitos Novos	80
5.3.10	Aba Rank Municípios	83
5.3.11	Discussão	85
5.4	Conclusões e considerações finais	89
6	Conclusão geral	92
	Referências Bibliográficas	94
	Apêndices	101
	APÊNDICE A Certificado de registro do sistema Select Codai	102
	APÊNDICE B Certificado de registro do sistema Covid-19 Pernambuco	104

1 Introdução

A oferta e coleta de dados das mais variadas áreas do conhecimento são crescentes e já apresenta elevado volume. O desenvolvimento de ferramentas autônomas e de forma remota para coletar dados com baixo custo, tal como os dispositivos desenvolvidos através de sistemas embarcados construídos com microcontroladores inspirados em Arduíno é apenas um exemplo do porquê termos uma massiva e crescente quantidade de dados disponíveis. Segundo Fávero e Belfiore (2017), a ocorrência simultânea das cinco características: volume, velocidade, variedade, variabilidade e complexidade – caracterizam a geração de dados que recebem o nome de *Big Data*, sendo a realidade vislumbrada na atualidade e que continuar-se-á a entrever pelos pesquisadores das diversas áreas de conhecimento.

Não se precisa de esforço para encontrar um exemplo de *Big Data* na atualidade e podemos elencar os dados associados à pandemia de *Sars-Cov-2* declarada em onze de março do ano de dois mil e vinte (ORGANISATION, 2022). Os dados coletados associados à Covid-19, considerando-se apenas o Brasil e disponibilizados no site do Ministério da Saúde (<<https://covid.saude.gov.br/>>) SUS (2020), em condições regulares, são arquivos acrescidos de uma observação para cada município/cidade brasileira diariamente. Assim, para se ter uma percepção minimista, considerando que o Brasil apresenta 5.570 municípios (ESTATÍSTICA, 2022) e, os dados começaram a ser disponibilizados em 27/03/2020, temos até 10/05/2023, 4.316.750 observações compostas por 17 variáveis, totalizando, portanto, uma matriz com 73.384.750 elementos.

Evidentemente que essa não é a maior base de dados para se debruçar em análise, nem tão pouco a quantidade de dados é a única dificuldade no processo de análise, já que outros fatores podem trazer maiores dificuldades a depender dos objetivos na análise e processamento. Um exemplo de outro tipo de dificuldade é a necessidade de, mesmo com uma quantidade relativamente pequena de dados, oferecer como saída a relação de classificados de um processo seletivo em conformidade com o sistema de cotas estabelecido pela Lei 12.711 de 29 de agosto de 2012, levando a um nível de complexidade associado ao processamento frente ao objetivo e sendo, neste caso, a quantidade de dados secundária.

Neste sentido, tem-se uma problemática evidente associada à gestão e governança de dados abrangente, a qual, segundo Fávero e Belfiore (2017) está relacionada com a própria prática da coleta, o armazenamento e manutenção segura, o acesso aos dados quando necessário, bem como a sua comunicação eficiente, associada aos mais variados

propósitos e áreas do conhecimento, possibilitando a sua visualização clara, rápida e intuitiva. Acrescento, que é importante considerarem-se conjunturas em que o público alvo não terá conhecimento especializado, de modo a se promover uma “tradução” das informações para uma linguagem acessível. A comunicação/visualização dos dados é uma peça fundamental no arcabouço de todo o processo da gestão de dados. É o extremo “final” dentre as etapas da governança de dados e tem um papel central na tomada de decisão em diversas atividades de natureza social e econômica, pois, é a informação filtrada, organizada e disponível com ferramentas de fácil visualização que tendem a estabelecer uma condição com peso substancial à percepção do analisador, ajudando-o na tomada de decisões importantes.

Ao se prospectarem conjuntos de dados massivos, com o intuito de proporcionar a sua visualização/comunicação, são necessárias a sua importação, organização, transformação, filtragem e modelagem e, por fim, a apresentação/comunicação. Intrinsecamente, na execução das etapas citadas, há a aplicação de ferramentas estatísticas visando múltiplos interesses, tais como a influência e efeitos provocados por certas variáveis (independente) em alguma outra variável (dependente) de interesse, cálculos de estatísticas descritivas, aplicação de modelos matemáticos–estatísticos diversos e que, muitas vezes, apresentam nível significativo de sofisticação. Percebe-se a evidente inevitabilidade do uso de poder computacional para se obter êxito na comunicação e visualização dos dados. Além disso, ter o poder computacional por si mesmo não é suficiente para prospectar os dados até a fase de sua comunicação, pois, em geral, são utilizados poderosos softwares dedicados a se obter os objetos de visualização (gráficos, tabelas com filtros, “gauges” e etc.).

Segundo [Fávero e Belfiore \(2017\)](#), a análise de *Big Data* não pode ser suportada sem que seja acompanhada do aprimoramento de software profissionais, entre os quais temos o IBM SPSS Statistics Software e o Stata Statistical Software, que oferecem capacidade enorme de processamento e são eficazes para desenvolver os mais diversos modelos e testes. Todavia, a maioria dos softwares voltados para desbravar dados, através da aplicação de ferramentas estatísticas, são pagos. O Stata Statistical Software por exemplo, cotado em 24/02/2022, para licença perpétua com suporte garantido a análise de *Big Data*, tem valor de \$795,00 (setecentos e noventa e cinco dólares) ([TIMBERLAKE, 2022](#)).

Um ponto crucial a ser considerado reside nas especificidades de como a comunicação/visualização é efetuada no tocante a dinâmica/estática e interatividade/não interatividade com o usuário interessado na prospecção dos dados. Há relevância flagrante de se implementar sistemas que possibilitem dinâmica e interatividade na implementação de gráficos e tabelas, entre outras ferramentas de visualização, que possam exibir informações de interesse na medida necessária, quando necessário e sem sobrecarregar a visualização,

de modo que mudanças ocorram em tempo real com a solicitação interativa do usuário. Visualizações que se atualizem imediatamente com a solicitação através de comandos disponibilizados para o usuário interagir com o sistema, promovendo, por exemplo, novos filtros, tais como intervalos de datas e regiões espaciais (coordenadas geográficas específicas, por exemplo) ou escolhas de informações específicas (variáveis, por exemplo).

Os sistemas interativos–responsivos podem trazer atualização em tempo real na visualização das ferramentas com as quais os dados são analisados, proporcionada simplesmente pelas modificações promovidas pelo usuário, trazendo agilidade na percepção para a tomada de decisão e até para o surgimento de *insights* sobre a fenomenologia relacionada com os dados. Segundo [Silva, Cook e Lee \(2017\)](#), a possibilidade de visualizar e explorar conjuntos de dados, dinâmica e interativamente, remete a *insights* que podem ser providos de forma mais eficiente. Uma vantagem flagrante na adoção de sistemas interativos–responsivos é a redução de material “físico” para análise, já que em uma única “página” interativa–responsiva pode-se implementar inúmeras associadas à responsividade do sistema em consequência das escolhas interativas do usuário e, portanto, tem-se um maior conforto na manipulação do material de consulta.

Um aspecto que é perceptivelmente relevante, consiste no fato de se desenvolver sistemas com uma característica singular em relação as já expostas anteriormente, a qual consiste no estabelecimento da conexão do sistema de comunicação/visualização diretamente à base de dados que está sendo alimentada pelo(s) dispositivo(s) de coleta. Isto é, os dados não são coletados por dispositivo(s), armazenados, importados, tratados e “embutidos” no sistema. Porém, os dados são coletado(s) por dispositivo(s) que o armazenam em algum servidor, inclusive em nuvem, e são importados diretamente do ponto de armazenamento para o sistema e, de forma integrada, aplicará as ferramentas de análise pertinentes a esses dados que automaticamente são comunicados com as características já expostas. Um sistema com esta característica singular traz benefícios adicionais: o primeiro consiste em, ao ocorrer atualizações e/ou adições aos dados no processo de coleta em tempo real, o sistema processará as atualizações/inserções de dados aplicando todas as ferramentas implementadas e disponibilizará a visualização/comunicação atualizada em tempo real também; Um segundo, é ter a possibilidade de se implementar um código fonte capaz de analisar como se dá a coleta dos dados pelo(s) dispositivo(s) de coleta, assim poder-se-á facilmente implementar mecanismos que apontem para as anomalias no processo de coleta.

Considerando-se o exposto nos parágrafos anteriores, este trabalho demonstrará que sistemas interativos–responsivos são ferramentas eficientes, de fácil acesso, de baixo custo e incorporam as singularidades mencionadas anteriormente. Eles são capazes de

atender às etapas de governança de dados propostas, mesmo diante das demandas de oferta atuais, tornando-se ferramentas de apoio à gestão de dados inequívocas. Neste sentido, a linguagem de programação R (R Core Team, 2022) se mostra crucial e é utilizada para o desenvolvimento e implementação dos sistemas com as características mencionadas.

Para melhor percepção da estruturação deste trabalho, elucida-se que o mesmo está organizado da seguinte forma: o [Capítulo 1](#) constitui a introdução; No [Capítulo 2](#), apresenta-se o objetivo geral e os objetivos específicos relacionado as pretensões a serem atingidas; No [Capítulo 3](#), expõe-se a revisão de literatura, na qual se apresenta na [seção 3.1](#), uma breve descrição do R e alguns *package* relevantes, bem como uma exposição da estrutura básica para a construção dos sistemas propostos; No [Capítulo 4](#), apresenta-se os detalhes de um sistema computacional com as características propostas, onde resolve-se através do mesmo, a problemática de classificação dos candidatos inscritos no processo seletivo do Colégio Dom Agostinho Ikas da Universidade Federal de Pernambuco, configurando o acesso aos dados, inclusive em nuvem, a sua organização, manipulação e comunicação/visualização dinâmica e forma interativa-responsiva; No [Capítulo 5](#), discorre-se sobre a edificação de um sistema computacional, considerando-se as características expostas, para desbravar a robusta massa de dados disponibilizada pelo Ministério da Saúde brasileiro associada à pandemia de *Sars-Cov-2*, declarada pela OMS – Organização Mundial da Saúde em 11 de março de 2020 voltando as atenções para as informações associadas ao Estado de Pernambuco; Finalmente, no [Capítulo 6](#), expõe-se conclusão geral em torno de todo o trabalho desenvolvido.

2 Objetivos

2.1 Geral

Demonstrar que sistemas interativos–responsivos, voltados às etapas de governança de dados que correspondem a importação, transformação, filtragem, modelagem e, principalmente, comunicação/visualização, implementando as ferramentas de análises e saídas pertinentes as especificidades de cada aplicação constitui uma ferramenta eficiente e singular para o desbravamento dos dados.

2.2 Específicos

1. Desenvolver e avaliar o código fonte implementado em linguagem de programação R em conjunto com *packages* pertinentes, para desenvolvimento de um *software* que permita visualizar e acompanhar em tempo real, os dados das inscrições do processo seletivo do Colégio Dom Agostinho Ikas (CODAI/UFRPE), bem como processar os referidos dados e oferecer como saída a classificação dos candidatos, considerando-se as particularidades do processo seletivo e sistemas de cotas da Lei N^o 12.711, de 29 de agosto de 2012;
2. Desenvolver e avaliar o código fonte implementado em linguagem de programação R em conjunto com *packages* pertinentes, para análise e visualização/comunicação de forma interativo–responsiva dos dados disponibilizados pelo Ministério da Saúde do Brasil para a Covid-19 com foco no Estado de PE, implementando-se diversas ferramentas de análise;

3 Revisão de Literatura

3.1 A linguagem de programação R

A linguagem R foi criada em 1993 no Departamento de Estatística da Universidade de Auckland na Nova Zelândia e seu nome é inspirado nas iniciais dos criadores, Ross Ihaka ([Figura 1a](#)) e Robert Gentleman ([Figura 1b](#)) e de um jogo figurado com a Linguagem S. Os professores de Estatística Ross Ihaka e Robert Gentleman, sensíveis ao fato de os softwares comerciais voltados à estatística serem de código fechado, o qual não permitiam aos usuários qualquer intervenção, além de custo para as instituições, pesquisadores ou estudantes, perceberam a relevância de se ter um software de código aberto acessível aos alunos. Então, em 29 de fevereiro de 2000 foi disponibilizada a primeira versão do R.

Figura 1 – Criadores do R.



(a) George Ross Ihaka (★1954).

Fonte: [Auckland](#) (2022).



(b) Robert Clifford Gentleman (★1959).

Fonte: [School](#) (2022).

Pode-se considerar que linguagem R é uma implementação da linguagem S, a qual foi desenvolvida nos laboratórios Bell Labs (Nokia Bell Labs), originalmente chamado AT&T que se destina ao desenvolvimento científico, sobretudo para pesquisas voltadas à indústria, sendo subsidiada pela empresa finlandesa Nokia. Os desenvolvedores da linguagem S, foram Rick Becker, John Chambers e Allan Wilks, formando a base para os sistemas S-Plus ([INSIGHTFUL](#), 2022), o qual é a distribuição comercial para o S.

A aceitação e sucesso na iniciativa dos professores Ross Ihaka e Robert Gentleman na disponibilização do R, tal como descrito, levaram a criação de compartilhamento e colaboração no seu desenvolvimento, resultando numa explosão de uso do R no meio acadêmico, e neste meio tempo, percebeu-se a flexibilidade de uso para as mais diferentes

áreas do conhecimento. Desde então, o R vem ganhando popularidade e segundo [Alcoforado, Luciane Ferreira e Levy \(2017\)](#), em 2015, o R atingiu a marca da sexta e segundo [Peslak e Conforti \(2020\)](#), em 2019, a manteve-se no ranking com a linguagem mais popular no mundo segundo a métrica da IEEE Spectrum (revista editada pelo Institute of Electrical and Electronics Engineers) ([IEEE, 2022](#)).

O uso do R é encontrado nas mais diversas áreas do conhecimento, como exemplos, na área de saúde [Sabater et al. \(2021\)](#), desenvolve um software para realizar diagnósticos automáticos de doenças cardíacas, através dos dados de eletrocardiograma (ECG); voltado para a educação [Konrath et al. \(2018\)](#), desenvolve uma aplicação utilizando os resultados do ENEM no Estado de Santa Catarina; em Bioinformática [Konrath et al. \(2018\)](#), um sistema para fornecer uma interface gráfica simples com visualização interativa para visões de Macro-Sinenteny, Micro-Synteny e Dot em nível de genoma e cria uma visualização de zoom dinâmico desde genes homólogos de macro-sinenteny a individuais. [Slater et al. \(2019\)](#) indica que a linguagem R ganhou um lugar central nas ciências hidrológicas, enfatiza os benefícios dessa do ponto de vista computacional, mas também considera questões sociais lembrando que através da linguagem R ocorre uma democratização da ciência de dados e a alfabetização numérica, bem como o aprimoramento da pesquisa reproduzível e da ciência aberta.

Atualmente, o R é desenvolvido colaborativamente através da codificação de *Packages* para diversas finalidades e áreas do conhecimento. Porém, há cerca de três décadas, existe um grupo principal, o R Core Team, do qual John Chambers é um membro. O grupo tem acesso à gravação ao código fonte do R. A lista com os membros do R Core Team, pode ser encontrada em <https://www.r-project.org/contributors.html>.

Considerando uma visão geral do R, é perceptível que ainda é muito comum sua utilização como um sistema de análise estatística ou programa para análises estatísticas [Wang et al. \(2022\)](#), por exemplo, refere-se ao R como uma linguagem de programação usada para computação estatística e gráficos. Não obstante, a perspectiva de uso do R vem mudando, onde o mesmo é considerado uma linguagem de programação científica completa como apontou em 2012 [Hornik \(2012\)](#), relatando que R é uma linguagem de programação abrangente, que oferece uma ampla gama de técnicas estatísticas e gráficas, além de fornecer uma plataforma para o desenvolvimento de novos métodos para a análise de dados. Deste modo, é assertivo considerá-lo como um ambiente onde foram implementadas e disponibilizadas muitas ferramentas estatísticas voltadas para a estatística, sejam implementações incorporadas na base do R ou através de *packages*, pois, segundo [R Core Team \(2022\)](#), documentação oficial do R, existem cerca de vinte e cinco *packages* fornecidos na base do R, voltados à estatística. Notoriamente, o R não pode

ser considerado simplesmente um software ou aplicativo de análise estatística, mas uma linguagem de programação orientada a objeto¹, interpretada², de alto nível e apresentando um repertório amplo e ampliável de ferramentas robustas e facilitadores para a análise de dados e estatística, mas não restrito a tal. Além disso, não está preocupado com um ambiente amigável para o usuário (como é comum em softwares de estatística), mas com flexibilidade.

3.2 IDE's – RStudio

Uma IDE – *Integrated Development Environment* (Ambiente Integrado de Desenvolvimento), nada mais é que um software onde ocorre a integração de um repertório ferramental para facilitar e apoiar o desenvolvimento do programador. Programas em R, bem como em outras linguagens, podem ser desenvolvidos em um simples editor de texto, tal como o Notepad (bloco de notas) do Windows. Porém, a utilização de uma IDE apresenta inúmeras vantagens, as quais listamos algumas a seguir:

- Facilita e orienta na criação de novos projetos;
- Ajuda no gerenciamento dos projetos existentes;
- Promove a integração de projetos à ferramentas Web, tal como [gitHub](#)³ (GITHUB, 2022) e [shinyapps.io](#)⁴ (RSTUDIO, 2022);
- Simplifica a instalação, atualização, gerenciamento e carregamento *packages*;
- Oferece suporte crucial na edição dos códigos, através de ênfase textual em funções, operadores e outros elementos de codificação;
- Auxilia na manutenção da indentação adequada do código, crucial para facilitar seu entendimento;
- Proporciona o acesso à documentação de ajuda, de forma simplificada.

Percebe-se que muitos editores ou IDE's podem ser utilizados para desenvolvimento de projetos com a linguagem de programação R, tal como os IDE's Visual Studio

¹ Ocorre quando as informações são manipuladas pela memória RAM na forma de objetos.

² Em linguagens de programação interpretada, o código não tem a necessidade ser compilado, de modo que os comandos são executados imediatamente. Outro exemplo de linguagem interpretada é o Python.

³ A grosso modo, pode-se dizer que é uma plataforma Web de gerenciamento e controle de versões colaborativamente ou não para desenvolvimento de softwares e projetos: <https://github.com/>.

⁴ Plataforma desenvolvida e mantida pela equipe do RStudio para a implantação de aplicativos Shiny na Web: <https://www.shinyapps.io/>.

(RTVS) (MICROSOFT, 2022a), compatível com Windows, sendo possível (MICROSOFT, 2023b) encontrar orientações sobre seu uso com o R. Outro exemplo é o Eclipse (StatET) (ECLIPSE, 2022), multiplataforma (Linux, MacOS e Windows), sendo ambos exemplos bem conhecidos no meio dos programadores. No entanto, o IDE que oferece uma interface de usuário mais completa para o desenvolvimento no R é o RStudio, sendo esta gratuita e multiplataforma (Linux, MacOS e Windows). R Studio é um IDE para o R, que reúne diversas funcionalidades e ferramentas tornando amigável o seu uso, a importação de dados e sua visualização, bem como a inserção de comandos e funções, apresentação de gráficos e documentos (HENNING et al., 2016).

3.3 *Package* do R

Um *package* R é uma coleção de funções, dados e documentação que estende as capacidades do R base (WICKHAM; GROLEMUND, 2019). *Packages* do R são um conjunto de funções e métodos personalizados que podem ser compartilhados com outros desenvolvedores, cuja motivação está relacionada com a automatização de tarefas, resolução de problemas diversos e cálculos de interesse comum. Assim, não é raro que um *package* desenvolvido para uso pessoal se torne útil e disponível para outros desenvolvedores. A possibilidade da criação e implementação de *packages* com certo grau de simplicidade proporciona aplicação do R nas mais variadas áreas do conhecimento. Calenge (2006) indica que esta característica encoraja os usuários a embarcarem na programação sem nem perceber, e Wickham e Grolemund (2019) aponta que a utilização desses é a chave para o uso bem-sucedido da linguagem R.

Portanto, os *packages* desenvolvidos para diversas finalidades e áreas do conhecimento torna o R uma linguagem de programação versátil e, além disso, a criação de *packages* pela comunidade de desenvolvedores é crescente e dinâmica, surgindo novas possibilidades a cada dia.

Considerando-se as perspectivas deste trabalho, por serem a base do desenvolvimento dos sistemas propostos, detalharemos os *packages* Shiny, Flexdashboard e R Markdown.

3.3.1 O *package* Shiny

O Shiny é um *package* do R desenvolvido pela equipe do RStudio que proporciona a criação de aplicações (app, aplicativos ou sistemas) *web* através de linhas de código do R, no RStudio ou outras IDE's. Pode ser executado diretamente no R ou pode ser hospedado em nuvem, através do shinyapps.io (RSTUDIO, 2022), por exemplo. Sua primeira versão

foi lançada em dezembro de 2012, sendo concebido pela autoria de diversos pesquisadores e colaboradores, tendo atualmente como mantedor responsável Winston Chang, tal como pode ser evidenciado em [RStudio \(2023\)](#).

A simplicidade e poder do *package* Shiny consiste em oferecer uma estrutura de codificação na linguagem R para se desenvolver sistemas voltados às etapas de governança de dados correspondentes à importação, transformação, filtragem, modelagem e, principalmente, comunicar/visualizar com a possibilidade interativa com o usuário, bem como responsividade em tempo real. [Schmuller \(2018\)](#) defende que através do Shiny se tem a melhor forma de disponibilizar um aplicativo R em navegadores de internet, compartilhar dados e análises, sem precisar conhecer HTML ou *JavaScript* para fazê-lo. Isto se deve ao desenvolvimento acontecer através do RStudio, sem a necessidade de se ter conhecimento de linguagens como HTML ou *JavaScript*, e os desenvolvedores terem a possibilidade de utilizar as ferramentas da linguagem R para efetuar as análises dos dados, por exemplo.

Em síntese, Shiny renderiza os elementos necessários do código da linguagem R para HTML, assim, desenvolve-se os *scripts* em R e o aplicativo desenvolvido é interpretado pela máquina, após a renderização, como em HTML, com a vantagem de ser possível utilizar todas as ferramentas, *package* e versatilidade do R. Para uma iniciação no Shiny, recomenda-se uma visita no site oficial ([RSTUDIO, 2023](#)), onde há um guia de instalação do *package*, bem como exemplos básicos.

O uso do Shiny vem sendo empregado na computação científica em diversas áreas, denunciando o seu potencial aplicado à etapa de governança de dados associada à comunicação/visualização, mas de forma interativa-responsiva. [Jiang Ziren e Cao \(2023\)](#) apresentam uma ferramenta intitulada *RIMeta*, construída usando Shiny e implementa métodos que permitem ao usuários, não proficientes em R ou quaisquer outras linguagens de programação, estimar um intervalo de referência no qual se espera que uma proporção pré-especificada de medições para uma população saudável caia', a partir de uma meta-análise usando dados agregados (média, desvio padrão e tamanho da amostra) ([JIANG ZIREN E CAO, 2022](#)); [Silva, Cook e Lee \(2023\)](#) descrevem a estruturação de dados e a construção de parcelas para explorar modelos de classificação florestal de forma interativa; [Xiao e Lam \(2022b\)](#) desenvolvem um aplicativo intitulado *Shinysyn* na estrutura do McScan⁵, o qual fornece uma interface gráfica fácil de usar para análises de sinteny⁶ sem exigir habilidades de programação, reduzindo as barreiras técnicas ([XIAO; LAM, 2022a](#)). De certo, pode-se encontrar na literatura especializada de diferentes áreas do conhecimento,

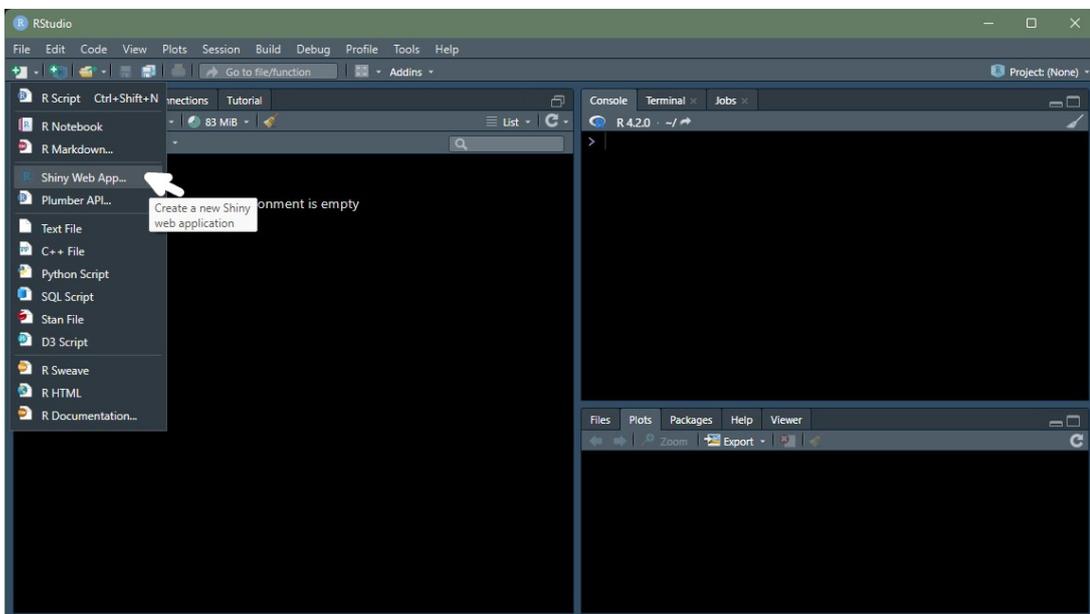
⁵ Estrutura do McScan é um método de análise de genomas que utiliza comparações de sequências de DNA para identificar genes homólogos e colinearidades em genomas relacionados.

⁶ Estrutura amplamente usada em estudos genômico comparativos. Fornecem informações para revelar a colinearidade cromossômica em intra-espécies e inter-espécies

aplicações do R combinado com o *package* Shiny para tornar a comunicação científica de dados mais acessível e prática, pois, ao se construir sistemas com as características apresentadas, é possível permitir uma análise das informações eficiente e prática.

Para estabelecer uma comparação entre o desenvolvimento de sistema *web* através do *package* Shiny diretamente e com sua implementação por meio de um documento R Markdown conjuntamente ao Flexdashboard, exibe-se a forma mais básica de se iniciar o desenvolvimento de um aplicativo em Shiny na [Figura 2](#). Essa indica que com o RStudio aberto, clica-se no botão de *New File* → *Shiny Web App*, o que levará à janela exibida na [Figura 3](#).

Figura 2 – Tela do RStudio indicando criação de um aplicativo Shiny.

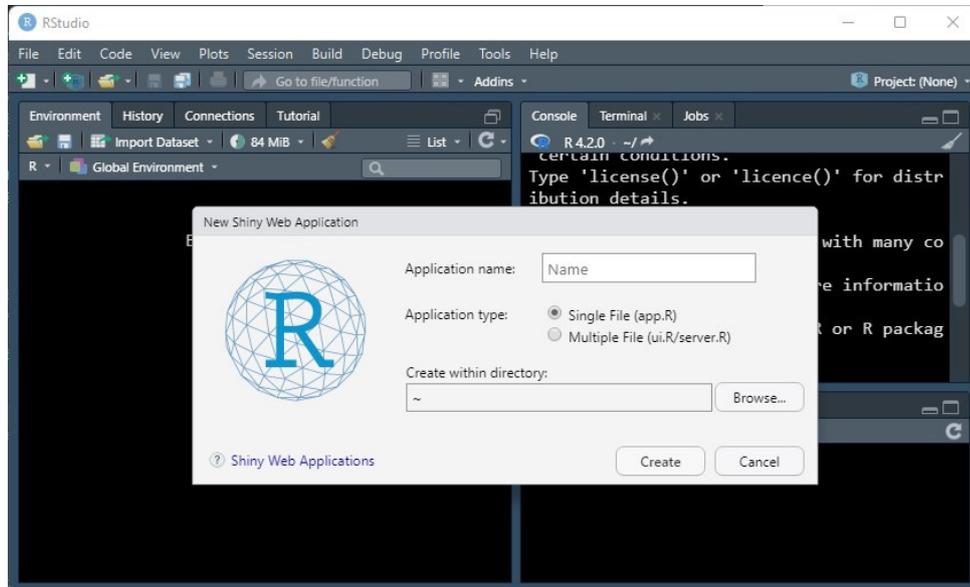


Fonte: Próprio autor.

A janela que se segue é intuitiva, guiando a criação de um aplicativo Shiny básico. Em *Application name*, insere-se um nome para o aplicativo e em *Create Within directory*, escolhe-se o diretório de armazenamento dos arquivos para funcionamento do aplicativo. Nota-se que há duas possíveis escolhas em *Application type*: *Single File (app.R)* ou *Multiple File (ui.R/Server.R)*. Na primeira, há um único *script* .R para trabalhar as funções *User Interface* e *Server*; Na segunda, são criados dois *scripts* .R separados para a *User Interface* e *Server*, sendo recomendado para projetos mais robustos.

Considerando a escolha para o tipo da aplicação com *Single File*, ocorrerá a criação o *script* *app.R*, o qual é exibido na [Figura 4](#). A estrutura básica de um *script* para a programação de um aplicativo desenvolvido com o Shiny apresenta dois componentes atribuídos a variáveis e definidos como *ui* e *server*, sendo estas detalhadas a seguir:

Figura 3 – Janela no RStudio para inserção das informações básicas para a criação de um aplicativo Shiny.



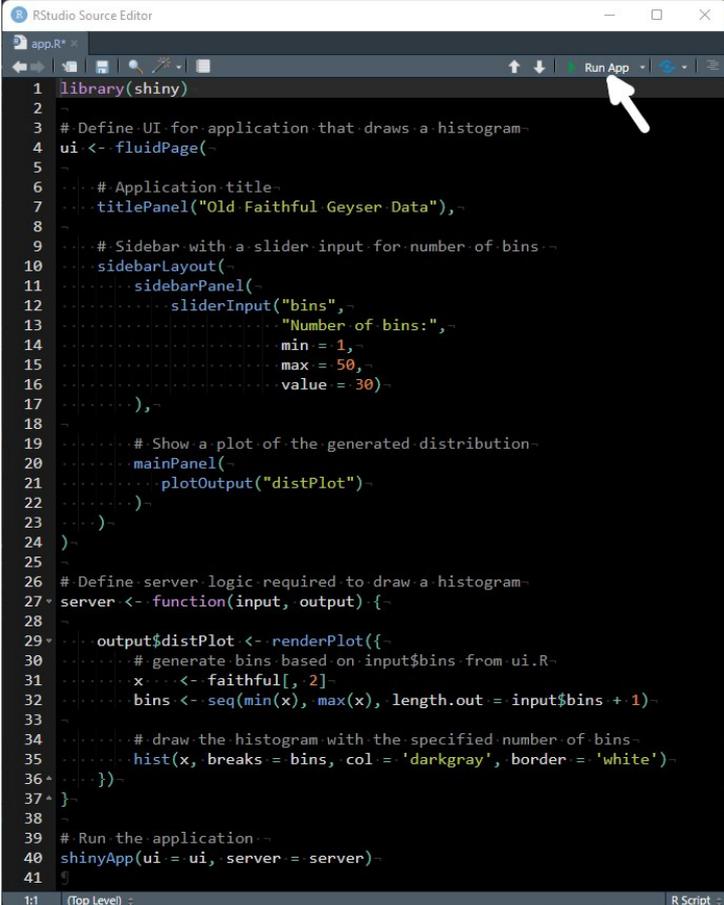
Fonte: Próprio autor.

- *ui* – *User Interface*: Responsável pelo *layout* do sistema, tal como a aparência e os elementos visuais de interação e reação. A *ui* é definida através de alguma função específica do *package* Shiny que determina o tipo de página, tais como *fluidPage()*, *fixedPage()* e *fixedRow()*.
- *server*: Responsável pelo processamento computacional executado pelo R, em que são consideradas as possibilidades de tarefas a serem executadas (análises, gráficos, filtro e outros), ficando sempre pronta para a inserção de *inputs* de interação do usuário para processamento e oferta de *outputs*. O *server* é definido através de uma função genérica que apresenta dois parâmetros, não necessariamente, mas por questões de simplicidade, o *input* e o *output*.

As duas estruturas básicas se comunicam através de componentes de *inputs* e *outputs*, os quais, basicamente correspondem, respectivamente, à interação do usuário e a reação do sistema (aplicativo).

Na Figura 4, clicando-se em *Run App*, será apresentada a janela exibida na Figura 5, a qual constitui um exemplo de aplicativo interativo e responsivo. Percebe-se a simplicidade com que é criada a aplicação. Evidentemente, este aplicativo base poderá ser editado e diferentes funcionalidades, bem como diversas ferramentas de análise implementadas.

Figura 4 – *Script* básico no RStudio para criação de um aplicativo Shiny.



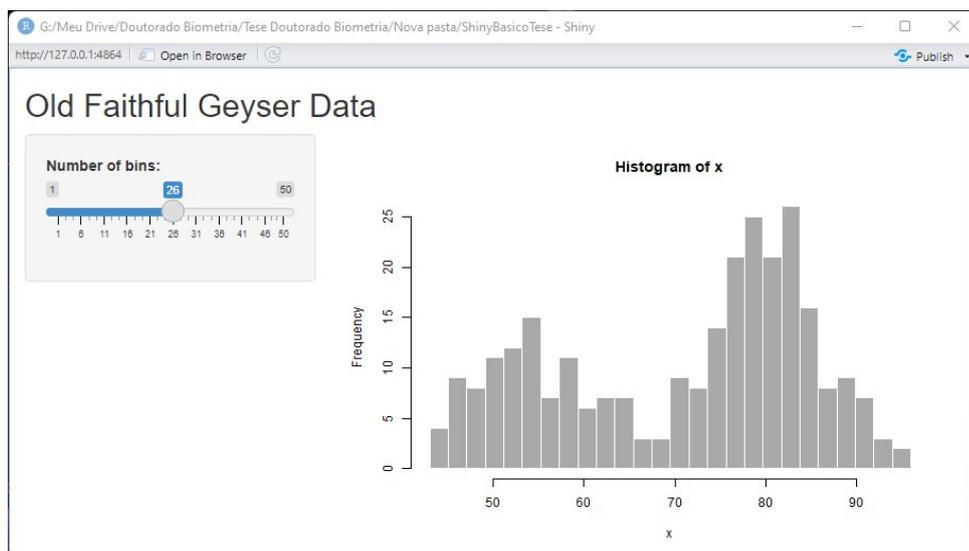
```

1 library(shiny)
2
3 # Define UI for application that draws a histogram
4 ui <- fluidPage(-
5
6   # Application title
7   titlePanel("Old Faithful Geyser Data"),
8
9   # Sidebar with a slider input for number of bins
10  sidebarLayout(-
11    sidebarPanel(-
12      sliderInput("bins",
13                  "Number of bins:",
14                  min = 1,
15                  max = 50,
16                  value = 30)
17    ),
18
19    # Show a plot of the generated distribution
20    mainPanel(-
21      plotOutput("distPlot")
22    )
23  )
24 )
25
26 # Define server logic required to draw a histogram
27 server <- function(input, output) {
28
29   output$distPlot <- renderPlot({
30     # generate bins based on input$bins from ui.R
31     x <- faithful[, 2]
32     bins <- seq(min(x), max(x), length.out = input$bins + 1)
33
34     # draw the histogram with the specified number of bins
35     hist(x, breaks = bins, col = 'darkgray', border = 'white')
36   })
37 }
38
39 # Run the application
40 shinyApp(ui = ui, server = server)
41

```

Fonte: Próprio autor.

Figura 5 – *Script* básico no RStudio para criação de um aplicativo Shiny.



Fonte: Próprio autor.

3.3.2 O R Markdown

O R Markdown (ALLAIRE et al., 2019) oferece uma estrutura de editoração unificada, envolvendo as ferramentas de ciências de dados, linha de comando, saídas de resultados, comentários e os elementos textuais em geral. Segundo Baumer et al. (2014) o R Markdown é um sistema com fácil uso, permitindo a combinação de análise estatística e a geração de documentos capazes de processar um *script* do R e gerar um documento de saída com elementos gráficos.

Documentos R Markdown tem extensão .Rmd e podem ser renderizados em muitas saídas diferentes, tais como .pdf, .doc, .ppt, html, dentre outras possibilidades (RSTUDIO-RMARKDOWN, 2023). Assim, é possível gerar relatórios de alta qualidade, livros e, considerando-se o propósito deste trabalho, desenvolver-se páginas da *web*. Segundo Xie, Allaire e Grolemund (2018), o R Markdown foi projetado para facilitar a reprodutibilidade, pois tanto o código computacional quanto as narrativas estão no mesmo documento, e os resultados são gerados automaticamente a partir do código-fonte, suportando formatos de saídas estáticos e dinâmicos/interativos.

A Figura 6 elucida como um documento R Markdown é processado. O knitr (XIE, 2022) “costura” o documento .Rmd executando os trechos de códigos e cria um documento Markdown (.md) genérico que mantém incluso os códigos e suas respectivas saídas. Então, o .md é processado por pandoc (PANDOC.ORG, 2023), sendo esse responsável por criar o arquivo de saída final.

Figura 6 – Processamento básico de um documento R Markdown.



Fonte: RStudio-RMarkdown (2022)

Para este trabalho, o relevante é perceber que o R Markdown suporta formatos renderizáveis com as saídas para html, proporcionando a visualização em navegadores de internet e a publicação na *web* com elementos dinâmicos e interativos–responsivos. Não obstante, o R Markdown por si só não possibilita a dinâmica–interatividade–responsividade com conexão ao R. Quando se utiliza saídas html simples, mesmo que através de *package* Flexdashboard (detalhado na subseção 3.3.3), há interatividade do tipo *client-side*, ocorrendo diretamente no navegador e independente do R. Para Wickham e Grolemund (2019), essa independência tem vantagem no sentido de se poder distribuir o arquivo html e esse “rodar” sem conexão com o R em qualquer navegador de internet. Contudo, frisam que é

limitante no que se pode fazer com o que foi implementado em html. Isto se deve ao código R ser executado uma vez na renderização e os resultados armazenados estaticamente, impossibilitando que o *client* possa “solicitar” uma nova execução de código R para obter novos resultados a serem apresentadas no navegador, a não ser que seja usado algo como o JavaScript (PLURALSIGHT, 2016).

Portanto, para se conseguir o resultado esperado, para este trabalho, o *package* R Markdown deve ser combinado ao Shiny, possibilitando a interatividade em conexão circular com o código em R, por meio do qual são executadas as transformações de dados, cálculos de quantidades desejadas, geração de gráficos bem como outros elementos necessários. Xie, Allaire e Grolemund (2018) indica que ao contrário do fluxo de trabalho mais tradicional de criação de relatórios estáticos, podem-se criar documentos que permitem aos leitores alterarem os parâmetros subjacentes à sua análise e obterem os resultados imediatamente.

A inserção do Shiny no R Markdown é feita se incluindo *runtime: shiny* no YAML do .Rmd. Isso quer dizer que se pode construir aplicações *web* sem o conhecimento de JavaScript. Por outro lado, vale salientar que as interações Shiny ocorrem na vertente do servidor e, portanto, precisa-se de um servidor Shiny para serem executados, seja localmente ou *online*.

Considerando a versatilidade e potencialidades do desenvolvimento de sistemas *web* com elementos interativos–responsivos, diversas ferramentas foram desenvolvidas nas mais variadas áreas de conhecimento. Loginova e Wohland (2020) mostra como criar painéis interativos de dados *on-line* usando R Markdown, exemplificando através de um rastreador Covid-19 para regiões de saúde e hospitais em Queensland, Austrália. Haymond (2022) descreve o uso do *package* Flexdashboard e suas possibilidades de personalização para o desenvolvimento de relatórios constando resultados em um laboratório clínico, mas o referido *package* é sempre usado combinado como R Markdown, tal como será apresentado na [subseção 3.3.3](#). É salutar o ressaltar que as potencialidades para o uso do R Markdown são abrangentes. No [Capítulo 4](#), usa-se o R Markdown para criar o sistema *web*, bem como para gerar um série de relatórios de interesse, evidenciando bem a versatilidade do mesmo.

3.3.3 O *package* Flexdashboard

O objetivo do *package* Flexdashboard (IANNONE; ALLAIRE; BORGES, 2020) consiste em simplificar a criação de painéis, inclusive interativos, na linguagem de programação R, através do R Markdown. Segundo Xie, Allaire e Grolemund (2018), o Flexdashboard estende a personalização e as funcionalidades da estrutura principal do R para a criação de documentos no R Markdown.

Então, utilizando-se dos *package* Shiny , Flexdashboard e R Markdown conjuntamente, pode-se construir sistemas interativos-responsivos em nível profissional, de forma relativamente simples e com baixo custo. O Flexdashboard permite que os usuários agrupem análises e visualizações para criar painéis e *dashboards* estáticos (ou seja, uma página da Web padrão) ou dinâmicos (ou seja, documentos responsivos baseados na Web usando R Shiny) (HAYMOND, 2022).

Muitos recursos comuns a documentos da Web contemporâneos estão disponíveis no *package* Flexdashboard, tais como: Gráficos, tabelas, *dashboards*, *gauge*, texto, implementação de imagens ou ícones, muitos com a possibilidade de interatividade-responsividade. Os *layouts* (RSTUDIO-FLEXDASHBOARD, 2022) disponibilizados para o Flexdashboard são personalizáveis e podem ser configurados para se adaptarem automaticamente à exibição ideal em telas de tamanhos diferentes. Ou seja, dispositivos móveis. Há, inclusive, a possibilidade de se criar páginas exclusivas para telas móveis que não são exibidas em telas de computador e vice-versa. Todavia, percebe-se limitações na adaptabilidade para dispositivos móveis, pois a adaptação não depende apenas do *layout* da página web automatizada pelo Flexdashboard, mas dos elementos que são inseridos nela e, muito embora a página se adapte às telas dos dispositivos móveis, os elementos aplicados, tais como os gráficos e tabelas, podem proporcionar uma visualização ineficiente, desconfortável ou até impossível.

É importante frisar que o Flexdashboard é um *package* desenvolvido para o R Markdown, onde através dos *layouts* pré definidos torna mais simples a criação de documentos com *dashboards*, inclusive páginas web. Tem vantagem na facilidade de uso e codificação, já que oferece a possibilidade de criar painéis interativos em apenas uma linha de código do R, podendo-se haver muitas personalizações. Se o desenvolvimento for de sistema interativos-responsivos com a aplicação do Shiny, elimina a necessidade do uso explícito das funções *ui* e *serve* descritas na subseção 3.3.1, onde se relatou que se mostraria a criação de um aplicativo Shiny apenas para estabelecer uma comparação com a criação de aplicativos via R Markdown combinado ao Flexdashboard. O ponto comparativo mais importante é estabelecido agora: não há a necessidade no Flexdashboard de se chamar as funções *ui* e *server* explicitamente, tornando as renderizações de objetos interativos-responsivos mais simples e direta, proporcionando uma significativa simplificação na codificação. Isto pelo fato do Flexdashboard ser construído para R Markdown e constituir um ecossistema de *package* para criar documentos computacionais em R (RSTUDIO-FLEXDASHBOARD, 2022).

No entanto, apesar de sua facilidade no uso, o Flexdashboard tem suas limitações. Por exemplo, a personalização e o controle dos *layout* são limitados e outras ferramentas

de criação de *dashboards* podem ser melhores neste sentido, tais como o *Google Data Studio* (GOOGLE, 2022) e o *Power BI* (MICROSOFT, 2022b), os quais, inclusive, apresentam uma curva de aprendizagem menos desafiadora. Por outro lado, os *dashboards* implementados através das ferramentas mais robustas, tal como o *JavaScript* (PLURALSIGHT, 2016), mas com uma curva de aprendizado mais desafiadora para um iniciante no universo da programação e, principalmente, para cientistas de dados, já que a integração do Shiny, R Markdown e Flexdashboard ao R, oferece tudo que é necessário em uma só linguagem para tornar a eficiente o processo de divulgação/comunicação científica.

É perceptível que sistemas desenvolvidos em R para a publicação na *web*, disponibilizados como *site* através do R Markdown, devem apresentar a incorporação do Flexdashboard no YAML, tornando o seu uso intrínseco nesses sistemas. Já para se conseguir implementar as funcionalidades de interatividade–responsividade, o *package* Shiny deverá ser incluído e o *runtime: shiny* no YAML. Neste sentido, muitos sistemas foram criados com as mais variadas finalidades utilizando dessas ferramentas. Haymond (2022), por exemplo, mostra como criar um laboratório de negócios com *dashboards* inteligentes gratuitamente usando R e o Flexdashboard. Ele destaca que, em um ambiente orientado a dados, é crucial para pesquisadores e profissionais da área desenvolverem ferramentas de visualização interativas facilmente e com custo acessível sempre que possível. Já Talagala e Shashikala (2022) utilizam o Flexdashboard e desenvolvem painéis interativos para monitorar o surto de Covid-19 e administração de vacinas no Sri Lanka.

3.3.4 Estrutura básica de sistemas em R Markdown e outros *packages*

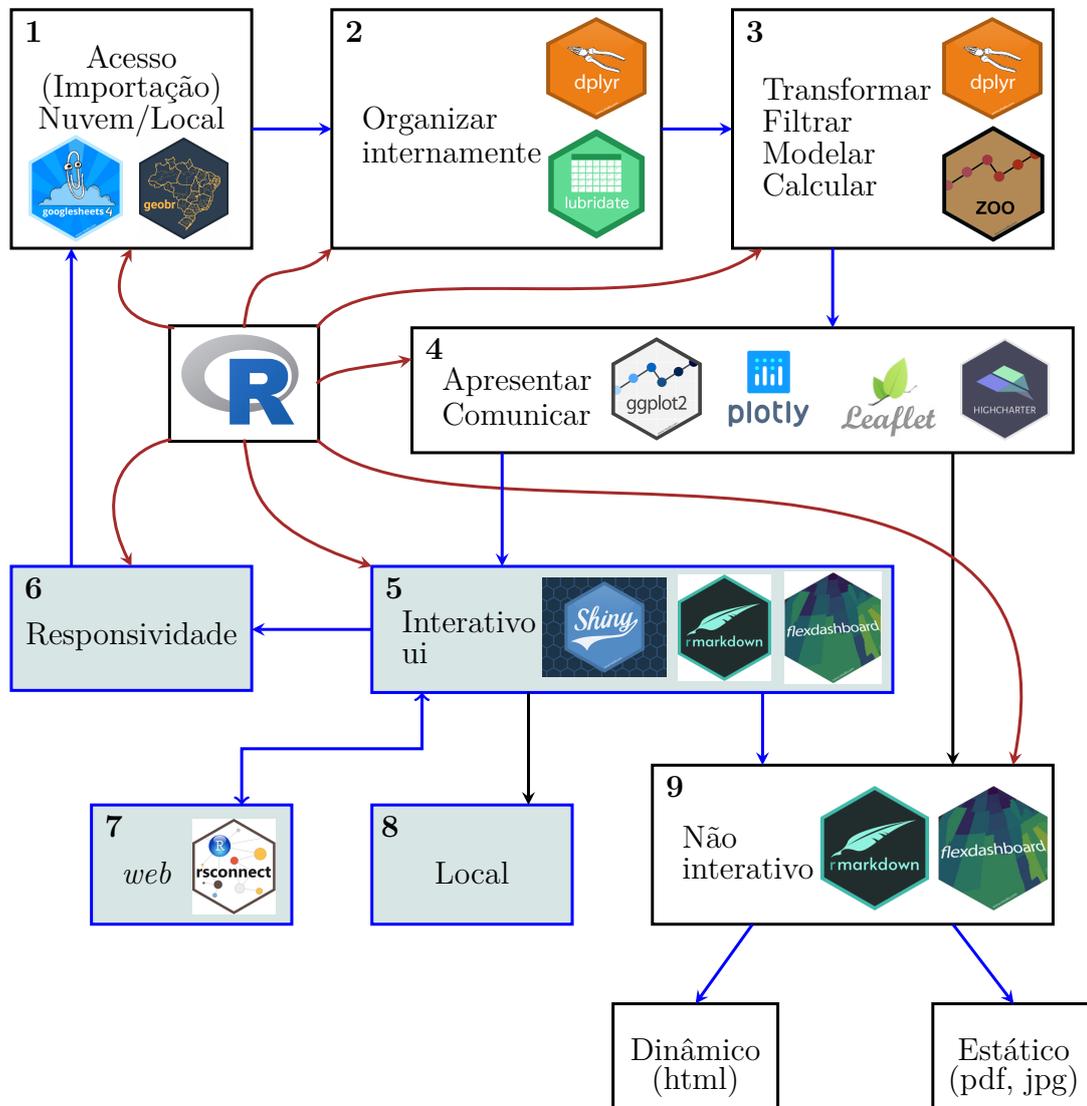
Trabalhar no desenvolvimento de sistemas interativos–responsivos com o R Markdown, conjuntamente com o Shiny e Flexdashboard, não necessariamente, mas pode contemplar várias etapas associada à governança de dados, desde a fase de acesso/importação dos dados. Para Wickham e Grolemund (2019), o R Markdown podem ser destinados à comunicação aos tomadores de decisão que querem focar nas conclusões, e não no código por trás da análise; Para colaboração com outros cientistas de dados (incluindo seu futuro eu!) que estão interessados tanto nas conclusões quanto em como essas foram construídas e; Como um ambiente no qual *fazer data science*, como um diário moderno onde pode ser registrado não só o que você fez, mas também em que você estava pensando. Evidentemente, que este último ponto destacado por Wickham e Grolemund (2019) tem haver com a possibilidade de uso do R Markdown para a construção de documentos diversos que possam ser utilizados como nossa memória de trabalho em *data science*.

O foco deste trabalho é voltado ao desenvolvimento de sistemas, preferencialmente que sejam disponibilizados em *sites* na *web*, de baixo custo, voltados para a etapa da

governança de dados para comunicação/visualização desses. O R Markdown , combinado com Shiny, o Flexdashboard e os *package* R que sejam pertinentes à aplicação do sistema, evidentemente, encaixam-se como ferramentas para a construção desses sistemas.

Assim, para esclarecer de uma forma eficiente a estrutura de desenvolvimento, bem como a dinâmica de funcionamento estrutural dos sistemas com as perspectivas descritas anteriormente e desenvolvidos com o R Markdown, conjuntamente aos *packages* Shiny e Flexdashboard, assim como demonstrar a importância de outros *package* do R com relevância indispensável, apresenta-se o diagrama exibido na [Figura 7](#).

Figura 7 – Diagrama que exhibe a estrutura de sistemas desenvolvidos em R.



Fonte: Próprio autor.

No diagrama supracitado, em (1) ocorre a fase de importação dos dados através de linhas de código no R . Destacam-se dois *package* no diagrama em (1), isto por

serem particularmente capazes de acessarem os dados em “nuvem”: o `googlesheets4` (BRYAN, 2021) capaz de importar, exportar e manipular dados em planilhas da Google (Google Sheets), o qual, inclusive, mostrou-se fundamental no desenvolvimento do sistema apresentado no Capítulo 4 e, o `geobr` (PEREIRA; GONCALVES, 2022) que fornece acesso ao conjunto de dados espaciais do Brasil, crucial no desenvolvimento do sistema apresentado no Capítulo 5; Em (2), tem-se a fase de organizar os dados internamente no sistema, onde são destacados dois *package*: o `lubridate` (GROLEMUND; WICKHAM, 2011), essencial para lidar com dados de datas, tais como a extração de informações (segundo, minutos, mês e dia da semana, por exemplo), transformações e operações aritméticas e o `dplyr` (WICKHAM et al., 2021), considerado um dos *package* mais úteis para realizar a transformação de dados com simplicidade e eficiência; Em (3), ocorre de fato a etapa de transformação, filtragem de dados, modelagem e criação de novas variáveis, inclusive com cálculos diversos. Muitos *package* podem ser utilizados na presente etapa, em destaque no diagrama é repetido o `dplyr`, já que é o principal *package* de transformação de dados, e o `Zoo` (ZEILEIS; GROTHENDIECK, 2005) que tem funções úteis para se tratar séries temporais, onde, por exemplo, utiliza-se a função `zoo::rollmean(x, k, fill = NA, ...)` para o cálculo das médias móveis do sistema apresentado no Capítulo 5; Em (4), tem-se a fase onde ocorre a construção de variadas ferramentas úteis para a apresentação/comunicação dos dados, tais como: gráficos de diversos tipos, inclusive através de mapas, tabelas e outras. Vários *package* podem ser convocados para a tarefa, tais como o famoso `ggplot2` (WICKHAM, 2016) – um dos mais utilizados para construir representações gráficas, o `plotly` (SIEVERT, 2020) – incorpora aos gráficos um *layout* possibilitando maior dinâmica e interatividade, o `leaflet` (CHENG; KARAMBELKAR; XIE, 2021) – possibilita a incorporação de mapas com elementos de interatividade e o `highcharter` (KUNST, 2022) – permite a criação de gráficos interativos, ativos e com uma dinâmica muito sensíveis a modificação dos dados. Evidentemente que há muitos outros *package* importantes aplicados/aplicáveis nesta etapa. Assim, citar-se-á um último por ter se apresentado crucial na construção dos relatórios implementado no sistema descrito no Capítulo 4: o `DT` (XIE; CHENG; TAN, 2021) – suscita a confecção de tabelas de dados com um grau bem significativo de customização e possibilidades de apresentação, tanto com relação a apresentações interativas–responsivas, quanto em documentos estáticos; Em (5), tem-se a ação dos *packages* responsáveis pela construção do código em si, onde são implementados os objetos de visualização/Comunicação/interatividade/responsividade. Primeiramente, há o documento R Markdown, onde havendo a implementação do Flexdashboard, consegue-se desenvolver elementos dinâmicos, porém não interativos–responsivos, bem como estáticos. Neste caso considere-se (9). Já com a incorporação do Shinyem (5), tem-se a possibilidade de interatividade do usuário em sua interface dando luz aos sistemas mais modernos e

eficientes com interatividade–responsividade. Neste caso, considera-se (6). Ao se criar um sistema que possa ser visualizado em navegadores *sites*, tem-se a possibilidade de se ter um servidor Shiny local – considera-se (8) ou na *web* – (7). Neste último caso, deve-se considerar o *package* `rsconnect` (MCPHERSON; ALLAIRE, 2021) por ser a ferramenta fundamental para se estabelecer uma conexão eficiente entre *hardware* (PC) de desenvolvimento e o local de hospedagem do sistema *web* – <https://www.shinyapps.io/> (RSTUDIO, 2022), de modo que ao se fazer atualizações/modificações do código fonte da aplicação no *hardware* de desenvolvimento com um único clique do *mouse* no RStudio haverá a publicação da nova versão do *site* na *web*; Há uma dupla seta entre (5) e (7) pelo fato de uma vez que o sistema esteja publicado na *web*, os usuários oferecerão seus comandos pelo *site*, o qual serão enviados para o código fonte que apresentará a responsividade (6) e, a depender do que foi solicitado, pode-se até mesmo retornar a fase (1) – importação dos dados, mas, evidentemente não necessariamente, já que a interação do usuário poderá efetuar solicitações que sejam necessárias, por exemplo, retornar a (3), (4) ou outra conjuntura; Por fim, percebe-se no diagrama o R, “regente” de todas as ferramentas que são aplicadas para o desenvolvimento dos sistemas.

É importante salientar que no diagrama apresentado na Figura 7 e descrito anteriormente, a numeração não constitui etapas bem definidas. Assim, por exemplo, na construção de um objeto Shiny (4 no diagrama) destinado a *ui* para interatividade do usuário, tal como uma caixa de seleção, pode-se utilizar o *package* `dplyr` para selecionar e filtrar os dados que serão as opções da caixa para o usuário selecionar (3 no diagrama). Neste caso, é como se (3) e (4) estivessem se retroalimentando. Wickham e Grolemund (2019) expõem que o desenvolvimento com R Markdown ofusca as linhas de comando entre a exploração interativa e captura o código, possibilitando integrar, dentro dos *chunks* (blocos de código R nos documentos R Markdown), os códigos e, fora destes os elementos textuais e não códigos.

4 Sistema de classificação do processo seletivo do CODAI/UFRPE

4.1 Introdução

O processo de seleção para admissão de novos estudantes nas instituições públicas do Brasil é baseado, de forma geral, nas notas obtidas em exames admissionais. Frequentemente, os exames incluem a participação de milhares de inscritos, o que demanda rigorosos sistemas de gestão para conduzir o processo de seleção. A complexidade dessas operações tornou-se ainda maior a partir das legislações que introduziram o sistema de cotas para candidatos baseados em critérios de renda, etnia, histórico educacional, entre outros. A necessidade de implantar um sistema de seleção capaz de lidar com essa complexidade de forma rápida e eficaz para executar um processo que atenda as exigências legais foi a motivação para a condução do presente estudo.

A Comissão Permanente de Vestibular (CPV) do Colégio Agrícola Dom Agostinho Ikas (CODAI) da Universidade Federal Rural de Pernambuco (UFRPE) é designada pelo diretor geral da instituição, cuja composição incorpora seis participantes, dos quais um é o presidente. A função da comissão é conduzir os processos seletivos para ocupação das vagas ofertadas semestralmente pelos diversos cursos oferecidos pelo CODAI. Deste modo, basicamente, a cada semestre são elaborados o edital público da seleção, o plano de publicidade, os métodos para realização das inscrições, o tratamento dos dados para edificação do relatório de classificação, acompanhamento do processo de matrícula e o reprocessamento dos dados, incorporando as informações das matrículas para construir e divulgar os relatórios, indicando a classificação de candidatos remanejados.

De todas as atribuições da CPV, a tarefa mais laboriosa é a concepção dos relatórios de classificação, sobremaneira, em razão de como as vagas são distribuídas (detalhada na [subseção 4.2.1](#)) levando em consideração o sistema de cotas que reserva vagas para egressos de escolas públicas, baixa renda e etnias específicas, conforme a Lei N^o 12.711, de 29 de agosto de 2012 (BRASIL, 2012). Além disso, há as vagas reservadas para Pessoas com Deficiência (PcD ou PCD) e as particularidades definidas pela CPV, tal como o método de obtenção do *Score* de classificação e os critérios de desempate, os quais, entre outros, são detalhados na [subseção 4.2.4](#).

Um ponto relevante a se considerar em torno do processo seletivo, consiste em se ter uma ferramenta eficiente que permita acompanhar a dinâmica das inscrições em tempo real, possibilitando dimensionar o número de colaboradores necessários para efetuar a conferência de autenticidade das informações declaradas pelos candidatos, confrontando-as com os documentos enviados no ato da inscrição. Bem como, promover o direcionamento publicitário aos cursos com baixo número de inscritos ao longo do período de inscrições.

Inicialmente as inscrições nos processos seletivos do CODAI eram *in loco*, através do preenchimento de formulário físico acompanhado da entrega de documentos pertinentes, inclusive o currículo escolar. Posteriormente, tais informações eram inseridas em planilhas eletrônicas do Microsoft Excel (MICROSOFT, 2023a), as quais eram manipuladas curso/turno-a-curso/turno para passo-a-passo promover, de forma extremamente laboriosa, a classificação. Posteriormente, desenvolveu-se uma automação do ranqueamento por meio de planilhas eletrônicas Calc da LibreOffice (FOUNDATION, 2010), sendo esta última escolhida por ser gratuita, já que o CODAI/UFRPE não dispõe de licenças do Microsoft Excel. No processo de desenvolvimento das planilha, precisou-se utilizar diversas camadas de fórmulas sofisticadas levando a uma lentidão exacerbada. Para cada processo seletivo, tinha-se que desenvolver inúmeras planilhas, cujo desenvolvimento se mostrou muito sofisticado, laborioso e extremamente suscetível a erros.

Considerando-se o exposto, desenvolveu-se um *software* em linguagem de programação R, cuja perspectiva é atender as demandas apresentadas pela CPV. O sistema foi batizado de *Select CODAI* e foi desenvolvido para apoiar a governança de dados, não apenas associada à comunicação/visualização de forma interativa-responsiva, mas no acesso, tratamento, transformação e processamento. É hospedado na *web*, proporciona o acompanhamento das inscrições em tempo real, promove a geração dos relatórios de classificação em caráter provisório, preliminar, definitivo e associado aos remanejamentos, sendo este último com a incorporação das informações associadas as matrículas.

4.2 Metodologia

4.2.1 Processo seletivo do CODAI e a oferta de vagas

O processo seletivo do Colégio Agrícola Dom Agostinho Ikas (CODAI) da Universidade Federal Rural de Pernambuco (UFRPE) ocorre semestralmente e, como exemplo de oferta das vagas disponibilizadas é apresentada a [Tabela 1](#), cujas informações foram obtidas nos editais dos processos seletivos do primeiro e segundo semestres letivos de 2022 e constantes nos editais (CPV 2022.1, CODAI, 2021) e (CPV 2022.2, CODAI, 2022). Vale salientar que entre os anos de 2018 e 2022, observou-se a oferta de vagas seguindo o mesmo

padrão.

Tabela 1 – Oferta das vagas dos processos seletivos do CODAI/UFRPE, para ingresso nos semestres letivos de 2022.1 e 2022.2.

Semestre	Cursos	Turno	Nº vagas
2022.1	Ensino médio	manhã	36
		tarde	36
	Técnico em agropecuária integrado ao ensino médio	integral	36
	Técnico subsequente em administração	manhã	36
		tarde	36
	Técnico subsequente em agropecuária	manhã	36
		tarde	36
	Técnico subsequente em alimentos	manhã	36
		tarde	36
	2022.2	Técnico subsequente em administração	manhã
tarde			36
Técnico subsequente em agropecuária		manhã	36
		tarde	36
Técnico subsequente em alimentos		manhã	36
		tarde	36

Fonte: Edital do processo seletivo do semestre letivo 2022.1 do CODAI/UFRPE ([CPV 2022.1](#), [CODAI, 2021](#)) e do processo seletivo do semestre letivo 2022.2 do CODAI/UFRPE ([CPV 2022.2](#), [CODAI, 2022](#)).

4.2.2 O sistema de cotas

Um dos grandes desafios para efetuar a classificação das pessoas inscritas no processo seletivo do CODAI está relacionada ao sistema de cotas, em conformidade com Lei Nº 12.711, de 29 de agosto de 2012 ([BRASIL, 2012](#)). Neste sentido, a [Tabela 2](#) resume, como exemplo, a distribuição de 36 vagas, na qual observa-se que 50% do total são reservados para pessoas inscritas oriundas de escolas públicas (CEP), das quais 50% são reservados para as com renda per capita menor ou igual a um salário mínimo e meio (CEP-), das quais 62,4% são reservadas para as declaradas com etnia preta, parda ou indígena (ETNIA) e, das quais 9,86% são reservadas para portadoras de deficiência (PCD). O mesmo princípio é aplicado aos outros 50% das vagas que ficam reservadas para às pessoas inscritas com renda per capita maior que um salário mínimo e meio (CEP+). Outras siglas na tabela são: vagas gerais (VG), demais etnias (DEMAIS), ampla concorrência (AC)

Tabela 2 – Sistema de cotas aplicado no processo seletivo do CODAI/UFRPE considerando-se a Lei N° 12.711, de 29 de agosto de 2012 (BRASIL, 2012). Apresentam-se os números, como exemplo, associado a oferta de 36 vagas.

COTA ESCOLA PÚBLICA (CEP)								AC
Mínimo: 50% total								
RP \leq 1,5 SM (CEP-)				RP $>$ 1,5 SM (CEP+)				50%
50% da CEP 25% do total				50% da CEP 25% do total				
ETNIA		DEMAIS		ETNIA		DEMAIS		18
62,4% da CEP- 15,6% do total		37,6% da CEP- 9,4% do total		62,4% da CEP+ 15,6% do total		37,6% da CEP+ 9,4% do total		
VG	PCD	VG	PCD	VG	PCD	VG	PCD	
90,14%	9,86%	90,14%	9,86%	90,14%	9,86%	90,14%	9,86%	
14,06%	1,54%	14,06%	1,54%	14,06%	1,54%	14,06%	1,54%	
5	1	2	1	5	1	2	1	

Fonte: Adaptado do edital do processo seletivo do semestre letivo 2022.1 do CODAI/UFRPE (CPV 2022.1, CODAI, 2021)

É importante salientar a ocorrência de ajustes no número de vagas com o intuito de acomodar os arredondamentos associados aos cálculos dos números de vagas a um determinado tipo de concorrência. Como exemplo, observa-se na Tabela 2 o ajuste do número de vagas para o tipo de concorrência PcD para 1 vaga, a qual seria 1,54% do total de vagas (36), correspondendo, portanto, a 0,55 vagas.

Em tempo, são descritos os acrônimos utilizados na Tabela 2, os quais surgirão futuramente. Inclusive, salienta-se que estes são aplicados nos relatórios de divulgação da classificação, bem como na obtenção dos dados no ato de inscrição e no código fonte desenvolvido para o sistema.

- **AC**: Ampla concorrência;
- **CEP**: Cota para pessoas inscritas oriundas de escola pública;
- **CEP-**: Cota para pessoas inscritas oriundas de escola pública com renda per capita menor ou igual a um salário mínimo e meio;
- **CEP+**: Cota para pessoas inscritas oriundas de escola pública com renda per capita maior que um salário mínimo e meio;
- **ETNIA**: Cota para pessoas inscritas oriundas de escola pública e que se declarem pretos, pardos ou indígenas;

- **DEMAIS:** Cota para pessoas inscritas oriundas de escola pública que se declarem de outras etnias que não pretos, pardos ou indígenas;
- **VG:** Vagas de concorrência geral;
- **PcD:** Vagas para pessoas portadoras de deficiência;
- **VNP:** Vagas não preenchidas.

4.2.3 Fluxo para a ocupação das vagas

A dificuldade imposta pelo sistema de cotas na construção dos relatórios de classificação está associada ao fluxo de ocupação das vagas ao longo da classificação, sendo de suma relevância para que o ranqueamento ocorra, conforme as prerrogativas estabelecidas pelo sistemas de cotas e as particularidades do edital do processo seletivo. Para entender, considera-se o diagrama exibido na [Figura 8](#), o qual exhibe as etapas seguidas para a ocorrência da alocação das 36 vagas dispostas na [Tabela 2](#).

- **Etapa 1:** Inicialmente (1), as 18 vagas para AC são distribuídas para todos os inscritos, inclusive os que concorrem por meio de cotas. Nesta etapa, o que importa são as métricas para classificar (detalhadas na [subseção 4.2.4](#)) para ocorrer o ranqueamento;

- **Etapa 2:** Em segundo (2), as 9 vagas de CEP⁻ e CEP⁺ serão alocadas de forma independente.¹ Todavia, estas vagas serão distribuídas efetivamente nas etapas seguintes;

- **Etapa 3:** Em terceiro (3), as vagas citadas na etapa anterior são divididas entre ETNIA (6 vagas) e DEMAIS (2 vagas), cuja alocação ocorre de forma independente e são distribuídas efetivamente na etapa seguinte;

- **Etapa 4:** Em quarta (4), considerando-se as métricas para classificar, é de fato ocupada a vaga PcD (1 vaga) e, no caso desta vaga não ser ocupada, considera-se o SIM para VNP, a vaga é direcionada para VG;

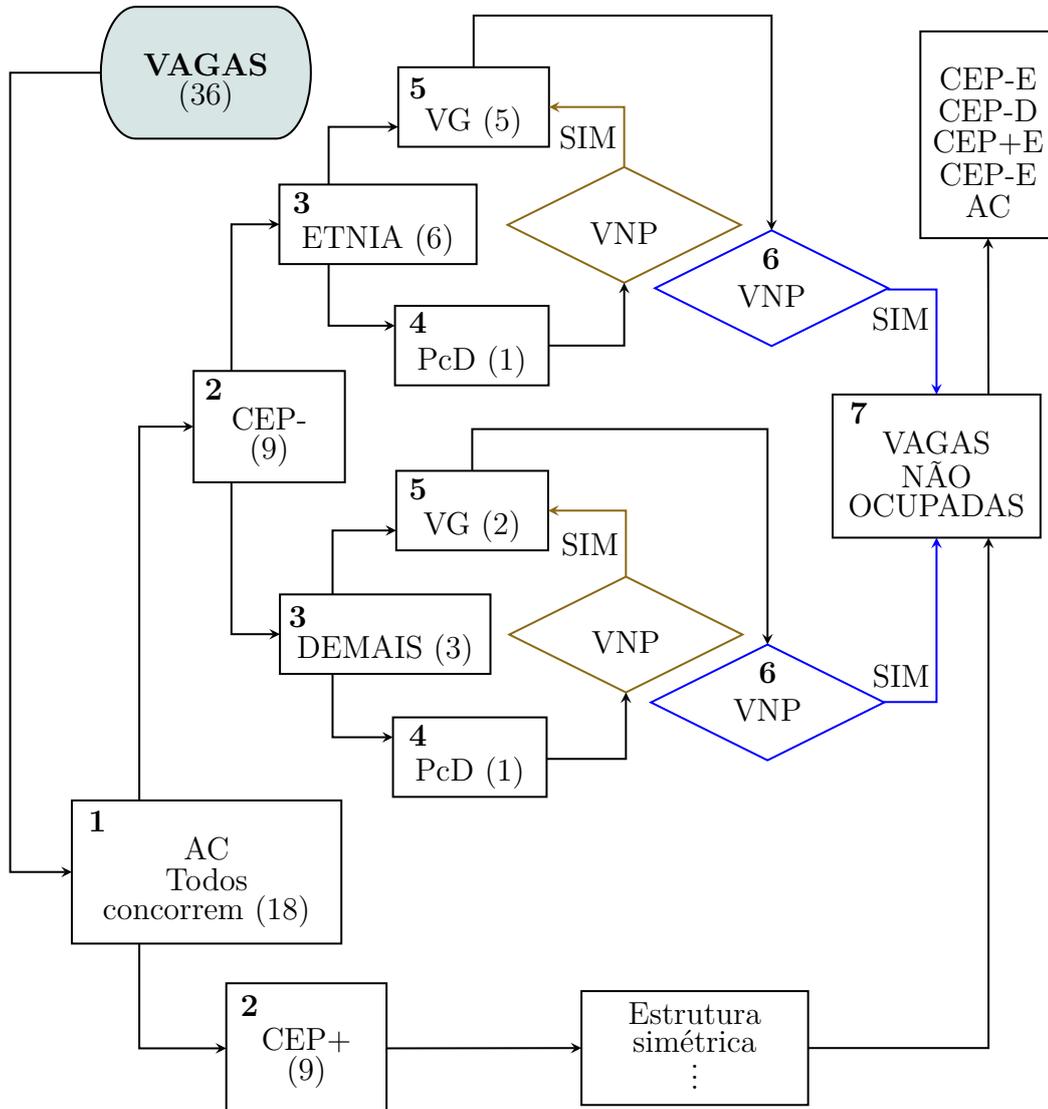
- **Etapa 5:** Em quinto (5), considerando-se as métricas para classificar, são de fato ocupadas as vagas VG (5 ou 6 vagas). Sendo todas as vagas ocupadas o processo de ocupação de vagas se encerra, mas havendo vagas não ocupadas o processo segue;

- **Etapa 6:** Em sexto (6), checando na etapa anterior a existência de vagas não ocupadas, ocorre uma contabilização destas para serem distribuídas posteriormente;

- **Etapa 7:** Em sétimo (7) e último passo, todas as vagas não preenchidas de todos os tipos de concorrência, se existirem, são distribuídas tendo em vista as métricas para

¹ Percebe-se que a distribuição em CEP⁺ tem uma conjuntura simétrica à CEP⁻. Portanto, concentra-se a descrição do que ocorre em CEP⁻.

Figura 8 – Diagrama que elucida o fluxo para a ocupação das vagas, considerando-se o sistemas de cotas. No presente diagrama, exemplificam-se os valores considerando a oferta de 36 vagas, tal como apresentado na Tabela 2. Quando há repetição na numeração das etapas, considera-se que estas são independentes e podem ocorrer em paralelo.



Fonte: Próprio autor.

classificar, incorporando mais uma específica para esta etapa, a qual se trata do próprio tipo de concorrência seguindo a ordem hierárquica CEP- e ETNIA, CEP- e DEMAIS, CEP+ e ETNIA, CEP+ e DEMAIS e AC.

Reitera-se que a descrição detalhada das etapas **1** a **6** para CEP- ocorre independente e de mesma forma para CEP+ e, apenas, na etapa **7** as vagas não ocupadas associadas a CEP- e CEP+ são reunidas para serem alocadas.

Contudo, percebe-se que o sistema de cotas, por si, já estabelece um certo nível de

complexidade para ocorrência da distribuição das vagas à medida que estas são alocadas ao longo do processo de ranqueamento levando à necessidade de um sistema computacional para efetuar a tarefa.

4.2.4 Métricas para classificar

Para ranquear os candidatos e efetuar a classificação, utiliza-se como principal métrica o *Score*, assim intitulado nos editais de seleção. Para se calcular essa quantidade, primeiramente se tomam as médias anuais (A_i) dos componentes curriculares, específicos para cada curso, e são calculadas as médias aritméticas simples dessas, conforme a [Equação 4.1](#), em que N é o número de médias do respectivo componente curricular, e M_j a média das médias anuais.

$$M_j = \frac{1}{N} \cdot \left(\sum_{i=1}^N A_i \right) \quad (4.1)$$

Em seguida, calcula-se o *Score* (S), através da [Equação 4.2.4](#), com aproximação em unidade de milésimos, o qual corresponde a média aritmética ponderada das médias anuais (M_j), obtidas pela [Equação 4.1](#).

$$S = \left(\sum_{j=1}^n p_j \right)^{-1} \cdot \left(\sum_{j=1}^n M_j \right)$$

Os pesos (p_j) para o cálculo do *Score* são associados aos componentes curriculares cujos valores são estabelecidos pela coordenação de cada curso, bem como cabe as mesmas a decisão de quais os componentes curriculares serão exigidos para compor o *Score*. Neste sentido, o curso de ensino médio e o curso técnico em agropecuária integrado ao ensino médio exigem as médias anuais de português e matemática, do sexto ao oitavo anos do ensino fundamental; O curso técnico subsequente em administração exige as médias anuais de português, matemática, história e geografia, do primeiro ao terceiro anos do ensino médio; Já os cursos técnicos subsequentes em agropecuária e alimentos exigem as médias anuais de português, matemática, química e biologia, do primeiro ao terceiro anos do ensino médio. Assim, cada curso poderá escolher pesos distintos para efetuar o cálculo. A [Tabela 3](#), exibe os pesos que foram aplicados aos componentes curriculares no processo seletivo do semestre letivo de 2022.1 ([CPV 2022.1](#), [CODAI, 2021](#)). Excetuando-se a não oferta dos cursos de ensino médio e técnico em agropecuária integrado ao ensino médio, os pesos aplicados para o processo seletivo do semestre letivo 2022.2, foram os mesmos.

Tabela 3 – Pesos dos componentes curriculares para o cálculo do *Score* para os diversos cursos ofertados no processo seletivo do semestre letivo 2021.1. Onde: POR: Português, MAT: Matemática, HIS: História, GEO: Geografia, QUI: Química e BIO: Biologia

Curso	POR	MAT	HIS	GEO	QUI	BIO
Ensino Médio	5,0	5,0	–	–	–	–
Técnico em Agropecuária Integrado ao Ensino Médio	5,0	5,0	–	–	–	–
Técnico Subsequente em Administração	3,5	3,5	1,5	1,5	–	–
Técnico Subsequente em Agropecuária	2,0	3,0	–	–	2,0	3,0
Técnico Subsequente em Alimentos	2,0	3,0	–	–	3,0	2,0

Fonte: Editais dos processos seletivos dos semestres letivos 2022.1 (CPV 2022.1, CODAI, 2021) e 2022.2 (CPV 2022.2, CODAI, 2022) do CODAI/UFRPE.

Assim, uma vez obtido o *Score*, discorre-se a classificação em ordem decrescente do mesmo, considerando-se o tipo de concorrência em cada respectivo curso e seguindo o fluxo de ocupação das vagas conforme foi detalhado na [subseção 4.2.3](#).

Uma segunda métrica utilizada remete a situação de empate por igualdade dos *Scores* arredondados até unidade de milésimo entre candidatos. Neste caso, aplicam-se os *critérios de desempate*, os quais são descritos a seguir, elencados em ordem para priorizar a classificação:

1. Maior M_j de português;
2. Maior M_j de matemática;
3. Maior idade;
4. Menor data-hora na inscrição.

Uma outra métrica utilizada, só aplicada quando vagas não são preenchidas, tal como foi detalhado na [subseção 4.2.3](#), corresponde ao próprio tipo de concorrência e passa a ser o primeiro critério na classificação, sendo seguido pelo *Score* e os *critérios de desempate*. Nestes casos, a ordem de priorização é a seguinte:

1. **CEP–E**: Cota para pessoas inscritas oriundas de escola pública com renda per capita menor ou igual a um salário mínimo e meio, tendo se declarado com etnia preta, parda ou indígena;
2. **CEP–D**: Cota para pessoas inscritas oriundas de escola pública com renda per capita menor ou igual a um salário mínimo e meio, tendo se declarado com as demais etnias;
3. **CEP+E**: Cota para pessoas inscritas oriundas de escola pública com renda per capita maior que um salário mínimo e meio, tendo se declarado com etnia preta, parda ou indígena;
4. **CEP+D**: Cota para pessoas inscritas oriundas de escola pública com renda per capita maior que um salário mínimo e meio, tendo se declarado com as demais etnias;
5. **AC**: Ampla concorrência.

4.2.5 Coleta e armazenamento dos dados de configuração do processo seletivo

O método empregado para o armazenamento e inserção das informações características do processo seletivo é o primeiro passo propriamente dito que reflete na estrutura do sistema em R, detalhado na [subseção 4.2.7](#). Isto se deve a observância da pertinência das etapas associadas à governança de dados para a coleta e armazenamento, de modo a tornar tudo mais factível na construção do sistema, planejando estas etapas e se considerando a perspectiva de que os dados serão acessados e atrelados ao desenvolvimento de um sistema computacional de processamento e comunicação/visualização em tempo real, tal como é o caso apresentado neste capítulo. Privilégio que ocorreu por se ter havido a estruturação de todo o sistema realmente do ponto zero.

Assim, construiu-se uma planilha da Google (Google Sheets) constituída por quatro abas, as quais tem função de um configurador para a inserção e armazenamento de informações relevantes do processo seletivo, sendo estas fundamentais para estrutura do sistema em R.

Construindo deste modo, tais informações podem ser editadas por membros da Comissão Permanente de Vestibular sem a necessidade de intervenção do desenvolvedor. Ao se editar a referida planilha, as novas informações são consideradas em tempo real pelo sistema, característica esta que figura como o principal fator de escolha por uma Google Sheets, já que através de codificação em linha de comando em *scripts* com a linguagem de programação R, fazendo-se uso de funções disponibilizadas no *package* `googlesheets4` (BRYAN, 2021), pode-se acessar os dados em tempo real. Há outras motivações para o uso

da planilha da Google para a finalidade em tela, tais como o armazenamento em nuvem de forma segura, bem como a facilidade de compartilhamento via *web* para colaboração e inserção das informações necessárias pelos membros da Comissão Permanente de Vestibular.

A primeira aba da referida planilha recebe o nome “Gerais” e tem a finalidade de conter as informações básicas do processo seletivo. Com essas informações na planilha, é possível construir o sistema em R as utilizando para, por exemplo, gerar o título dos relatórios de divulgação dos resultados do processo seletivo e divulgar informações relevantes no sistema *web*, tal como as datas de início e término das inscrições, entre outras. As datas citadas também são usadas para estabelecer o tipo de relatório *default* que será gerado e disponibilizado para *download*: se o provisório, preliminar, definitivo ou remanejamento. Para melhor compreensão, tem-se a relação completa de informações contidas na aba “Gerais” apresentada abaixo, já na [Figura 9](#) é possível observar os detalhes da aba.

- Nome do processo seletivo;
- Semestre letivo de entrada nas turmas;
- *Score* mínimo para aprovação (ponto de corte);
- Data de início das inscrições;
- Data de término das inscrições;
- Data de divulgação do resultado preliminar;
- Data de divulgação do resultado definitivo;
- Data do primeiro remanejamento.

Figura 9 – Aba da Google Sheets intitulada “Gerais”, com dados de configuração do processo seletivo e do sistema. Tem a finalidade de configurar as informações básicas do processo seletivo.

The screenshot shows a Google Sheets interface with the following content:

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
COLÉGIO AGRÍCOLA DOM AGOSTINHO IKAS
 Configurador processo seletivo 2023.1
CONFIGURADOR DE DADOS GERAIS DO PROCESSO SELETIVO

Observação:
 - No nome do processo seletivo não colocar acentos para evitar erros de reconhecimento
 - Formato da data xx/xx/xxxx (d/m/Y)

DADOS GERAIS DO PROCESSO SELETIVO (aparecerá no título do relatório)

Description	Information
Processo name	PROCESSO SELETIVO CODAI/UFRPE
Processo semester	2023.1
Minimum score	1
Start registration	24/10/2022
End registration	11/11/2022
Preliminary result date	29/11/2022
Final result date	05/12/2022
First relocation	13/12/2022

The spreadsheet also shows a navigation bar at the bottom with tabs: Gerais, Links, Vagas, Pesos.

Fonte: Próprio autor.

A segunda aba da referida planilha recebe o nome de “Links” já que seu objetivo é armazenar todos os links de acesso, via codificação em R, às abas/planilhas que são necessárias para o sistema acessar os dados pertinentes ao seu funcionamento. Deste modo, como a aba em questão, ao contrário de se escolher inserir vários links no sistema em R, escolhe-se referenciar apenas ao link aba “Links” e através dessa acessar todos os links que são necessários para capturar as informações/dados relevantes ao sistema.

Figura 10 – Aba da Google Sheets intitulada “Links”, com dados de configuração do sistema. Tem a finalidade conter links de todas as planilhas/abas com dados utilizados pelo sistema. Desfigura-se as informações da coluna “Information” por questões de segurança. Os dados da coluna “Information” estão ocultos por questões de segurança.

The screenshot shows a Google Sheets spreadsheet with the following content:

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
COLÉGIO AGRÍCOLA DOM AGOSTINHO IKAS
 Configurador processo seletivo 2023.1
LINKS DAS GOOGLE SHEETS UTILIZADAS PELO SISTEMA

Observação:
 No nome do processo seletivo não colocar acentos para evitar erros de reconhecimento

ENDEREÇAMENTO DE PLANILHAS EM .CSV (será a fonte de dados do sistema em R)

Description	Information	Observation
Link settings geral	[Redacted]	Aba Gerais da planilha configuradora.
Link settings vagas	[Redacted]	Aba Vagas da planilha configuradora
Link settings pesos	[Redacted]	Aba Pesos da planilha configuradora (pesos das disciplinas)
Link inscriptions	[Redacted]	Aba da planilha de coleta de inscrições
Link aba links	[Redacted]	Aba de links das google sheets utilizadas. Está aba.
Link forms	[Redacted]	Link formulário inscrição
Link Confe	[Redacted]	Link planilha conferência. Própria Google sheets

The spreadsheet interface includes a search bar, navigation icons, and a bottom toolbar with tabs for 'Gerais', 'Links', 'Vagas', and 'Pesos'.

Fonte: Próprio autor.

Na Figura 10, é possível visualizar a aba em discussão constituída pelas colunas “Description”, “Information” e “Observation”, as quais, respectivamente, apresentam uma breve descrição, os links em si e alguma observação pertinente. Para melhor clareza, lista-se a descrição dos links presentes:

- Link settings geral: Link da aba “Gerais”;

- Link settings vagas: Link da aba “Vagas”;
- Link inscriptions: Link da planilha com os dados de inscrições;
- Link forms: Link do formulário de inscrições;
- Link Confe: Link da planilha de conferência de documentos e informações de matrículas.

A terceira aba da planilha é intitulada “Vagas”, a qual pode ser visualizada na Figura 11, sendo essa destinada principalmente a configurar o número de vagas ofertadas pelos diversos cursos que estarão presentes para entrada no processo seletivo.

Figura 11 – Aba da Google Sheets intitulada “Vagas”, com dados de configuração do sistema. Tem a finalidade de configurar os nomes dos cursos oferecidos, seus turnos e suas respectivas quantidades de vagas, considerando-se cada tipo de concorrência (as cotas).

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
COLÉGIO AGRÍCOLA DOM AGOSTINHO IKAS
 Configurador processo seletivo 2023.1
CONFIGURADOR DA DISTRIBUIÇÃO DE VAGAS DO PROCESSO SELETIVO

Observações importantes:
 - Não utilizar acentos nos nomes dos cursos (suprima-os)
 - O total de vagas é obtido por: AC + RE + RD + SE + SD
 Ou seja, as vagas NE estão inseridas nestas
 - IdTurma deve ter mínimo de 3 e máximo de 4 caracteres
 - IdTurma não pode ter duplicidade

Legenda:
 AC: Ampla concorrência; ACNE: Idem para PcD
 RE: Renda menor ou igual 1,5 salários e etnia preta parda o indígena declada; RENE: idem para PcD
 RD: Renda menor ou igual 1,5 salários e demais etnias; RENE: idem para PcD
 SE: Renda maior que 1,5 salários e etnia preta parda o indígena declada; SENE: idem para PcD
 SD: Renda maior que 1,5 salários e demais etnias; SDNE: idem para PcD

CURSO	TURNO	IdTurma	AC		CEP-E		CEP-D		CEP+E		CEP+D		TOTAL
			AC	NEAC	RE	NERE	RD	NERD	SE	NESE	SD	NESD	
AGROPECUARIA INTEGRADO AO MEDIO	INTEGRAL	InAI	36	0	9	1	9	1	9	1	9	1	72
ENSINO MEDIO	INTEGRAL	EMIn	20	0	5	1	5	1	5	1	5	1	40
TECNICO SUBSEQUENTE EM ADMINISTRACAO	MANHA	AdmM	16	0	6	1	4	1	6	1	4	1	36
TECNICO SUBSEQUENTE EM ADMINISTRACAO	TARDE	AdmT	16	0	6	1	4	1	6	1	4	1	36
TECNICO SUBSEQUENTE EM AGROPECUARIA	MANHA	AgrM	16	0	6	1	4	1	6	1	4	1	36
TECNICO SUBSEQUENTE EM AGROPECUARIA	TARDE	AgrT	16	0	6	1	4	1	6	1	4	1	36
TECNICO SUBSEQUENTE EM ALIMENTOS	MANHA	AlIM	16	0	6	1	4	1	6	1	4	1	36
TECNICO SUBSEQUENTE EM ALIMENTOS	TARDE	AlIT	16	0	6	1	4	1	6	1	4	1	36

Fonte: Próprio autor.

Portanto, é nesta aba que se configuram, também, os nomes dos cursos ofertados na coluna “CURSOS”, o turno na coluna “TURNO” e um identificador único na coluna “IdTurma”.

A quarta aba da planilha, chamada “Pesos”, pode ser visualizada na [Figura 12](#). Ela é um configurador dos componentes curriculares que serão exigidos para constar suas médias anuais, bem como os pesos que tais componentes têm no cálculo do *Score*, tal como é explicado de forma detalhada na [subseção 4.2.4](#). Além disso, estão presentes as colunas “CURSO”, “TURNO” e “IdTurma”, as quais já constavam na aba destinada a inserção do número de vagas e ao inserir estas informações na aba “Vagas”, elas são automaticamente transferidas para a aba de inserção dos pesos.

Figura 12 – Aba da Google Sheets intitulada “Pesos”, com dados de configuração do sistema. Tem a finalidade permitir à CPV configurar os componentes curriculares exigidos, bem como os seus pesos para o cálculo do *Score* dos candidatos.

The screenshot shows a Google Sheets spreadsheet with the following content:

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
COLÉGIO AGRÍCOLA DOM ÁGOSTINHO IKAS
 Configurador processo seletivo 2023.1

CONFIGURADOR DO PESO DOS COMPONENTES CURRICULARES

Observações importantes:
 - Usar como separador de decimal o ponto
 - Curso, Turno e IdTurma são automáticos sendo transportados da aba de Vagas

CURSO	TURNO	IdTurma	Português PMP	Matemática PMM	História PMH	Geografia PMG	Química PMQ	Biologia PMB
AGROPECUARIA INTEGRADO AO MEDIO	INTEGRAL	InAI	5	5	0	0	0	0
ENSINO MEDIO	INTEGRAL	EMIn	5	5	0	0	0	0
TECNICO SUBSEQUENTE EM ADMINISTRACAO	MANHA	AdmM	3.5	3.5	1.5	1.5	0	0
TECNICO SUBSEQUENTE EM ADMINISTRACAO	TARDE	AdmT	3.5	3.5	1.5	1.5	0	0
TECNICO SUBSEQUENTE EM AGROPECUARIA	MANHA	AgrM	2	3	0	0	2.0	3.0
TECNICO SUBSEQUENTE EM AGROPECUARIA	TARDE	AgrT	2	3	0	0	2	3
TECNICO SUBSEQUENTE EM ALIMENTOS	MANHA	AlimM	2	3	0	0	3	2
TECNICO SUBSEQUENTE EM ALIMENTOS	TARDE	AlitT	2	3	0	0	3	2

Fonte: Próprio autor.

4.2.6 Coleta e armazenamento dos dados das inscrição

Seguindo as mesmas perspectivas expostas na [subseção 4.2.5](#), optou-se por coletar os dados das inscrições dos candidatos através de um formulário da Google, onde a inserção das informações são efetuadas pelos próprios candidatos via *web*. São dados de identificação pessoal, médias do currículo escolar pertinentes ao cálculo do *Score*, bem como *upload* de arquivos em formatos específicos com imagens dos documentos comprobatórios, os quais serão utilizados para certificação da veracidade das informações relatadas pelos candidatos. A escolha desta ferramenta de coleta se deve ao fato de as informações inseridas poderem ser direcionadas para uma planilha da Google (Google Sheets) contendo os respectivos dados armazenados em nuvem (Google Drive da CPV) de modo a possibilitar o acesso em tempo real através de código em linha de comando em *scripts* do R, fazendo-se uso de funções disponibilizadas no *package* `googlesheets4` ([BRYAN, 2021](#)).

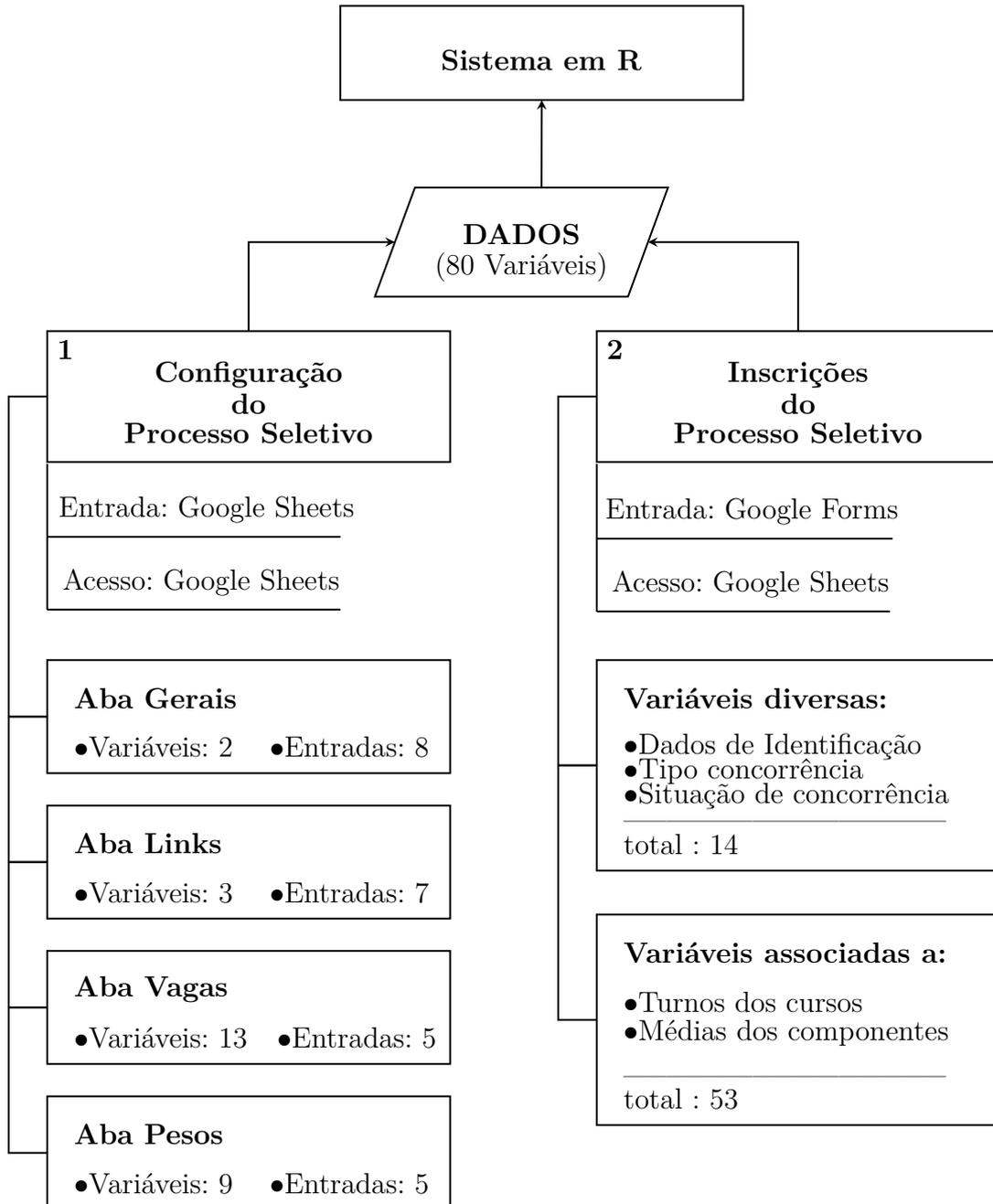
Das oitenta variáveis (colunas) presentes na planilha em questão, apenas as associadas a *upload* de arquivos e a nomeada de “OBSERVACAO”, destacadas em sublinhado, não são usadas pelo sistema em R.

- “DINSCRICAO”: data inscrição;
- “NOME”: nome do candidato;
- “NASC”: data de nascimento do candidato;
- “CPF”: CPF do candidato;
- “RG”: número do RG do candidato;
- “EMAIL”: e-mail do candidato;
- “CELULAR”: número do celular do candidato;
- “CIDADE”: cidade de residência do candidato;
- “FRG”: arquivo com a imagem frontal do RG;
- “VRG”: arquivo com a imagem do verso do RG;
- “TIPO” tipo de concorrência associada ao sistema de cotas (descrita na [subseção 4.2.2](#));
- “PCD”: se o candidato concorrerá para vagas reservadas para portadores de deficiência;
- “CURSO”: o curso que deseja concorrer;

- “TURNOEMI”: turno que deseja concorrer, em caso do curso escolhido ser o ensino médio;
- “N6PEM”, “N7PEM”, “N8PEM”, “N6MEM”, “N7MEM”, “N8MEM”: médias anuais, respectivamente, de português do 6º, 7º e 8º anos e de matemática do 6º, 7º e 8º anos do ensino fundamental, para o caso do curso escolhido ser o ensino médio;
- “ANFF18EM”: arquivo com a imagem frontal do documento comprobatório de médias dos componentes curriculares;
- “ANVF18EM”: arquivo com a imagem do verso do documento comprobatório de médias dos componentes curriculares;
- “TURNOInA”: turno que deseja concorrer, em caso do curso escolhido ser o técnico em agropecuária integrado ao ensino médio;
- “N6PI”, “N7PI”, “N8PI”, “N6MI”, “N7MI”, “N8MI”: médias anuais, respectivamente, de português do 6º, 7º e 8º anos e de matemática do 6º, 7º e 8º anos do ensino fundamental, para o caso do curso escolhido ser o técnico em agropecuária integrado ao ensino médio;
- “ANFF18IN”: arquivo com a imagem frontal do documento comprobatório de médias dos componentes curriculares;
- “ANVF18IN”: arquivo com a imagem do verso do documento comprobatório de médias dos componentes curriculares;
- “TURNOAdm”: turno que deseja concorrer, em caso do curso escolhido ser o técnico subsequente em administração;
- “N1PADM”, “N2PADM”, “N3PADM”, “N1MADM”, “N2MADM”, “N3MADM”, “N1HADM”, “N2HADM”, “N3HADM”, “N1GADM”, “N2GADM”, “N3GADM”: médias anuais, respectivamente, de português do 1º, 2º e 3º anos, de matemática do 1º, 2º e 3º, de história do 1º, 2º e 3º e de geografia do 1º, 2º e 3º anos do ensino médio, para o caso do curso escolhido ser o técnico subsequente em administração;
- “ANFF19ADM”: arquivo com a imagem frontal do documento comprobatório de médias dos componentes curriculares;
- “ANVF19ADM”: arquivo com a imagem do verso do documento comprobatório de médias dos componentes curriculares;
- “TURNOAgr”: turno que deseja concorrer, em caso do curso escolhido ser o técnico subsequente em agropecuária;

- “N1PAGRO”, “N2PAGRO”, “N3PAGRO”, “N1MAGRO”, “N2MAGRO”, “N3MAGRO”, “N1QAGRO”, “N2QAGRO”, “N3QAGRO”, “N1BAGRO”, “N2BAGRO”, “N3BAGRO”: médias anuais, respectivamente, de português do 1º, 2º e 3º anos, de matemática do 1º, 2º e 3º, de química do 1º, 2º e 3º e de biologia do 1º, 2º e 3º anos do ensino médio, para o caso do curso escolhido ser o técnico subsequente em agropecuária;
- “ANFF19AGRO”: arquivo com a imagem frontal do documento comprobatório de médias dos componentes curriculares;
- “ANVF19AGRO”: arquivo com a imagem do verso do documento comprobatório de médias dos componentes curriculares;
- “TURNOAli”: turno que deseja concorrer, em caso do curso escolhido ser o técnico subsequente em alimentos;
- “N1PALI”, “N2PALI”, “N3PALI”, “N1MALI”, “N2MALI”, “N3MALI”, “N1QALI”, “N2QALI”, “N3QALI”, “N1BALI”, “N2BALI”, “N3BALI”: médias anuais, respectivamente, de português do 1º, 2º e 3º anos, de matemática do 1º, 2º e 3º, de química do 1º, 2º e 3º e de biologia do 1º, 2º e 3º anos do ensino médio, para o caso do curso escolhido ser o técnico subsequente em alimentos;
- “ANFF19ALI”: arquivo com a imagem frontal do documento comprobatório de médias dos componentes curriculares;
- “ANVF19ALI”: arquivo com a imagem do verso do documento comprobatório de médias dos componentes curriculares;
- “SITUACAO”: situação de concorrência do candidato (CONCORRENDO, ELIMINADO), sendo este dado inserido pela Comissão de Vestibular no processo de conferência dos arquivos de documentos enviados pelos candidatos no ato da inscrição;
- “MOTIVO”: refere-se ao motivo de ter ocorrido a eliminação do candidato ou anulação de sua inscrição pela Comissão de Vestibular;
- “MATRICULA”: designa se a matrícula foi efetuada pelo candidato classificado;
- “OBSERVACAO”: Observações gerais proferidas pela CPV sobre os candidatos .

Figura 13 – Diagrama elucidativo do arcabouço de variáveis envolvidas e manipuladas pelo sistema desenvolvido em R . À esquerda (1), tem-se informações associadas à configuração do processo seletivo, já à direita (2), tem-se os dados das inscrições e outros de interesse.



Fonte: Próprio autor.

Considerando-se o exposto nas [subseção 4.2.5](#) e [subseção 4.2.6](#), sintetiza-se o arcabouço de variáveis utilizadas pelo sistema em R no diagrama exibido na [Figura 13](#). Percebe-se à esquerda (1) que os dados são associados à configuração do processo seletivo, já à direita (2) os dados são atrelados às inscrições do processo seletivo. Com relação a

(1), tem-se um total de vinte e sete variáveis com vinte e cinco entradas (observações), as quais podem mudar em poucas quantidades, já que dependem do número de cursos ofertados, cuja modificação é rara. Em (2), tem-se sessenta e sete variáveis com milhares de entradas oriundas das inscrições no processo seletivo. Assim, há uma totalidade de noventa e quatro variáveis manipuladas pelo sistema.

4.2.7 Arquitetura do sistema em R

A estrutura do sistema em R, é, basicamente, constituída por arquivos cujas extensões são `.R` (*script* do R) e `.Rmd` (documentos R Markdown). Precisamente, são quatro *scripts* do R e dez documentos R Markdown, os quais são listados a seguir, explicitando-se o seu papel no *software*:

- ***Scripts* do R (`.R`):**

1. ***DadosFixSelect.R***: Acessa os dados da Google Sheets (detalhada na [subseção 4.2.5](#)) que contém informações de configuração do processo seletivo. Esse *script* também tem a função de promover um tratamento dos dados em conformidade com as necessidades de sua utilização em etapas posteriores da arquitetura do sistema, as quantidades de vagas e pesos dos componentes curriculares, por exemplo, são organizados e gravados em objetos do tipo `list()`;
2. ***DadosInscSelect.R***: Acessa os dados da Google Sheets (detalhada na [subseção 4.2.6](#)) que contém informações das inscrições dos candidatos, situação de concorrência e matrícula. Esse *script* também promove alguns tratamentos destes dados, tal como organização do formato de data/hora e números, segrega os dados por curso/turno, cria variáveis para as médias dos componentes curriculares e para o *Score*, calculando-os. Ainda, trata a duplicidade de inscrições deixando a mais recente e, evidentemente, organiza as informações em objetos do R pertinentes para serem usadas em outras conjunturas necessárias;
3. ***DadosTratSelect.R***: Este *script* faz uso dos dois expostos anteriormente e tem a função de promover a hierarquização classificatória dos candidatos de cada curso/turno, considerando-se o sistema de cotas e fluxo de ocupação das vagas, tal como exposto na [subseção 4.2.2](#) e a [subseção 4.2.3](#). Neste, são criados objetos para armazenar as saídas de forma estratégica no intuito de atender as necessidades dos documentos R Markdown, tanto para a construção do *site*, quanto para a geração dos diversos relatórios em PDF que são disponibilizados para *download* via *web* em tempo real;

4. ***DadosTratSelectRema.R***: Este *script* tem a mesma função do anterior. Todavia, é destinado exclusivamente para quando é solicitada as informações dos processos de remanejamento, pois neste caso se deve considerar os dados associados às matrículas não efetivadas dos candidatos que foram classificados na divulgação do resultado definitivo.

- **Documentos R Markdown (.Rmd)**

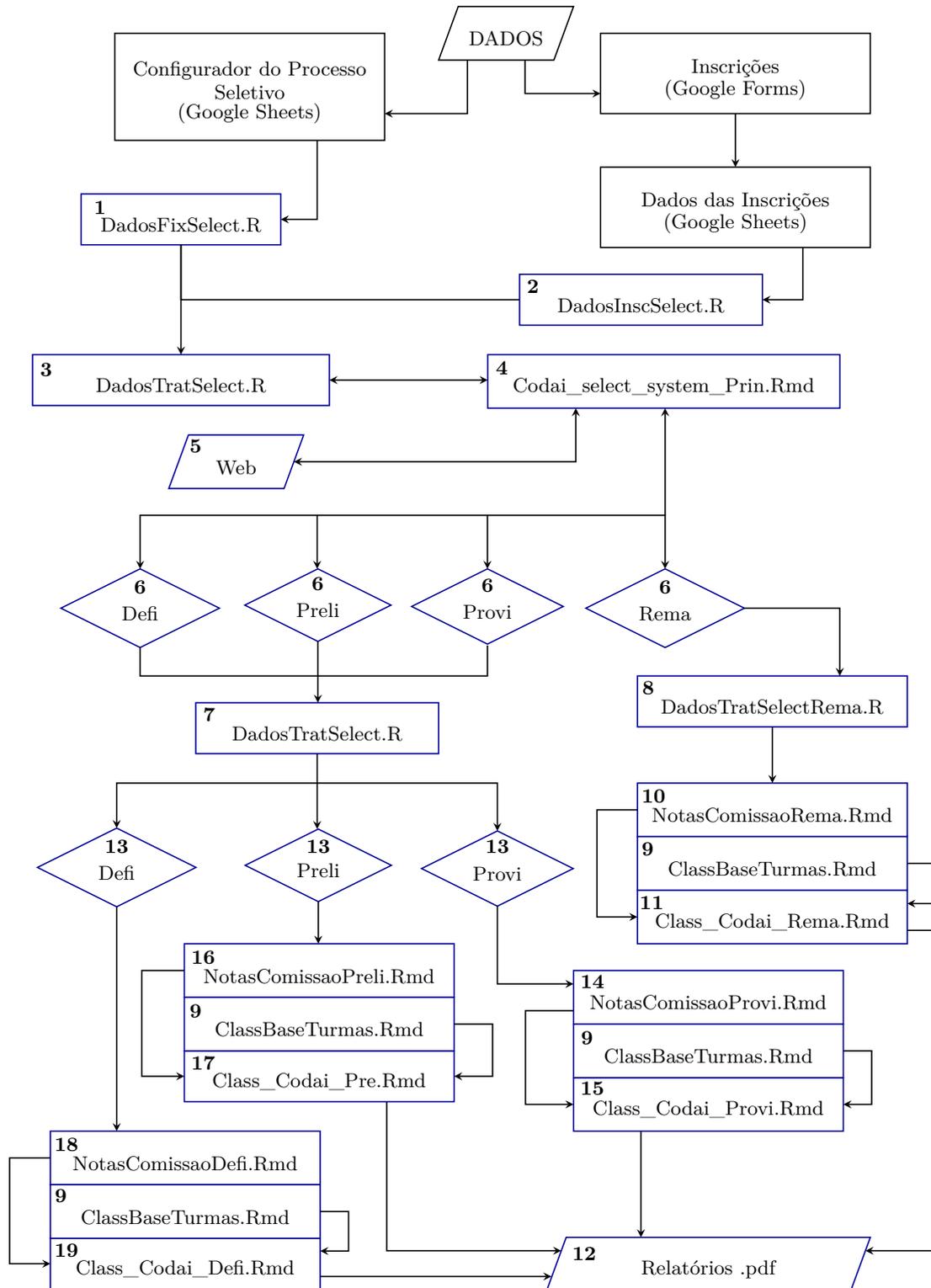
1. **NotasComissaoProvi.Rmd**: É um documento R Markdown auxiliar, pois é construído para a inserção de observações e avisos relevantes que a Comissão Permanente de Vestibular necessita implementar nos relatórios em PDF de divulgação do resultado de classificação provisório, os quais podem ser emitidos através da *web* durante a realização das inscrições dos candidatos;
2. **NotasComissaoPreli.Rmd**: Tem o mesmo papel do documento anterior, mas associado ao relatório para divulgação do resultado de classificação preliminar, o qual ocorre após o encerramento das inscrições e a conferência de documentos pela Comissão Permanente de Vestibular, marcando o início do prazo para que os candidatos possam entrar com recursos contestando a classificação apresentada;
3. **NotasComissaoDefi.Rmd**: Tem o mesmo papel dos documentos anteriores, mas associado ao relatório para divulgação do resultado de classificação definitiva, o qual ocorre após a análise e resposta dos recursos;
4. **NotasComissaoRema.Rmd**: Tem o mesmo papel dos três documentos anteriores, mas associado aos relatórios de classificação dos remanejamentos, os quais ocorrem após a realização de matrícula e constatação de ausência na mesma por parte de candidatos e/ou desistência declarada por classificados à vaga;
5. **ClassBaseTurmas.Rmd**: É um documento fundamental, pois teve-se a inquietude de ter seu *script*/texto desenvolvido para o tornar genérico no intuito de produzir tabelas com as informações de classificação dos candidatos por curso/turno e tipo de concorrência, mas, de modo a ser capaz de o fazer independentemente do curso/turno, bem como independente do relatório de resultado que será solicitado (PROVISÓRIO, PRELIMINAR, DEFINITIVO OU REMANEJAMENTO), tornando-o um R Markdown genérico e utilizado em todos os documentos .Rmd's que geram os relatórios .pdf's como saída final dos resultados de classificação. Isto foi possível através da construção um *loop* no código, de modo que, há uma identificação para o tipo de resultado, bem como para os cursos/turnos;
6. **Codai_select_Provi.Rmd**: É o documento que constrói o relatório .pdf como saída final com as informações de classificação para um resultado provisório,

tendo em conta os documentos `NotasComissaoProvi.Rmd` e `ClassBaseTurmas.Rmd`. Salienta-se que neste documento são incluídos os *scripts* `.R` pertinentes através do comando `source("name.R")`, os quais são: `DadosFixSelect.R`, `DadosInscSelect.R` e `DadosTratSelect.R`;

7. **Codai_select_Pre.Rmd**: É o documento que constrói o relatório `.pdf` como saída final com as informações de classificação para o resultado preliminar, tendo em vista os documentos `NotasComissaoPreli.Rmd` e `ClassBaseTurmas.Rmd`. Destaca-se que neste documento são incluídos os mesmos *scripts* `.R` mencionados no anterior;
8. **Codai_select_Defi.Rmd**: É o documento que constrói o relatório `.pdf` como saída final com as informações de classificação definitiva, considerando-se os documentos `NotasComissaoDefi.Rmd` e `ClassBaseTurmas.Rmd`. Neste documento são incluídos os mesmos *scripts* `.R` mencionados no anterior;
9. **Codai_select_Rema.Rmd**: É o documento que constrói o relatório `.pdf` como saída final com as informações de classificação para os resultados de remanejamento, considerando-se os documentos `NotasComissaoRema.Rmd` e `ClassBaseTurmas.Rmd`. Neste documento, como nos três anteriores são usados os *scripts* `.R` `DadosFixSelect.R`, `DadosInscSelect.R`, mas o de tratamento é o *script* `DadosTratSelectRema.R`;
10. **Codai_select_system_Prin.Rmd**: Este documento `.Rmd` é responsável por fazer a integração de todos os descritos anteriormente, bem como proporcionar que o sistema seja renderizado circularmente através dos *package* `Shiny` e `Flexdashboard`, possibilitando interatividade – responsividade através da *web*. É neste `.Rmd` que são definidos os objetos gráfico do *site* (da *web*), tal como os botões de ação, listas de seleção, gráficos, *dashboards* e outros, baseados nos dados existentes das inscrições e configuração do processo seletivo. Neste arquivo, também constam os comandos através do *package* `rsconnect`, para efetuar a publicação do sistema na *web*, hospedando-o no `shinyapp.oi` (RSTUDIO, 2022).

Para facilitar o entendimento da arquitetura do sistema, haja vista as descrições apresentadas anteriormente dos *scripts* (`.R`) e documentos R Markdown (`.Rmd`), construiu-se o diagrama exibido na [Figura 14](#), onde considerou-se simplificações, tal como não incluir as apresentações de informações no sistema *web*. Deste modo, apresentam-se as principais inter-relações da arquitetura. Para melhor entendimento, principia-se da entrada dos dados através das planilhas da Google descritas na [subseção 4.2.5](#) e [subseção 4.2.6](#).

Figura 14 – Diagrama elucidativo da estrutura do sistema em R, observando-se a utilização dos diversos *scripts* (.R) e documentos (.Rmd) envolvidos no funcionamento do software.



Fonte: Próprio autor.

Assim, no referido diagrama, (1) é o *script* que tem os códigos para acessar os dados de configuração do processo seletivo na Google Sheets descrita na [subseção 4.2.5](#) e os organizam em objetos convenientes; (2) é o *script* codificado para acessar os dados das inscrições dos candidatos na planilha da Google descrita na [subseção 4.2.6](#), organizando-os em objetos pertinentes e os tratando de forma preliminar; (3) é o *script* que leva em conta os dados organizados em (1) e (2), processando-os e promovendo a classificação do processo seletivo em si, considerando-se, evidentemente, o sistema de cotas descrito na [subseção 4.2.2](#) e na [subseção 4.2.3](#). Este código em R é um dos mais laboriosos do sistema, apresentando cerca de seiscentas e cinquenta linhas de comando; (4) é o documento R Markdown chave para gerenciar todos os *sprints* .R e demais documentos .Rmd, usando como padrão (2) para disponibilizar as informações da página *web* e, a depender das interações via *web*, utilizam-se dos .R's e .Rmd's necessários para construir as saídas como requeridas, sejam essas na própria página *web* ou através de arquivos para *download*, daí a presença de seta dupla entre (3) e (4); (5) é o *site* publicado na *web* via o shinyapp.oi (RSTUDIO, 2022), por meio do qual os usuários interagem com o sistema. Assim, é através deste *site* que ocorrem a interatividade–responsividade de acordo com as solicitações requeridas pelo usuário, as quais são gerenciadas por (4), de modo que a responsividade, dependendo do que foi solicitado, pode ocorrer para a atualização na própria página *web* ou através da geração de documentos para *download*; (6) representa uma tomada de decisão considerando a requisição do usuário através de sua interação em (5) e gerenciada por (4), especificamente para gerar o relatório de resultados da classificação e o disponibilizando para *download* no formato .pdf: Caso seja solicitado o relatório do resultado definitivo (Defi), preliminar (Preli) ou provisório (Provi) (6) levará à convocação de (7), por outro lado sendo a solicitação interativa para gerar o resultado de remanejamento (Rema) é utilizado (8); (11) se utiliza de (9) e (10) para construir em tempo real e disponibilizar o *download* de (12); Considerando-se (7), tem-se as possibilidades (13), de modo a ser possível ter havido solicitação entre os relatórios de classificação definitivo (Defi), preliminar (Preli) ou provisório (Provi), pois cada um apresenta documentos R Markdown específicos para serem processados: Caso seja solicitado provisório, (15) é executado utilizando-se de (9) e (14) para disponibilizar o *download* de (12), se preliminar for solicitado, (17) é executado utilizando-se de (9) e (16) para disponibilizar (12) e, finalmente, se o definitivo for solicitado, (19) é executado se utilizando de (9) e (18) para ocorrer a geração de (12).

De fato, há mais do que foi exposto para que o sistema exista e funcione corretamente. Todavia, foram apresentados os elementos estruturais básicos e principais, as demais são codificações para conceber elementos gráficos de interação ou não, estética do *site* para publicação na *web*, a codificação para proporcionar o acesso por senha na aba de uso registo para a Comissão Permanente de Vestibular, para a formatação dos relatórios

disponibilizados para *download*, funcionalidades de botões para convocar os *scripts* do R e documentos R Markdown .

4.3 Resultados

4.3.1 O *site web*: Sistema interativo–responsivo *web*

O sistema desenvolvido neste trabalho para processar os dados das inscrições e conceber a classificação do processo seletivo do Colégio Dom Agostinho Ikas da Universidade Federal de Pernambuco, pode ser executado localmente, onde há a necessidade de se ter o R instalado. Porém, uma das maiores vantagens está em sua publicação na *web* por meio do *site* https://marcelosilvaufupe.shinyapps.io/Codai_select_system Prin/ (SILVA RÔMULO S. C. MENEZES, 2022), possibilitando ser acessado de qualquer computador com acesso à internet através de um simples navegador, tal como o Google Chrome e Firefox.

O referido *site*, é o meio pelo qual os usuários em geral navegam, interagem e podem acompanhar, em tempo real, a evolução das inscrições no processo seletivo. Bem como, a Comissão Permanente Vestibular (CPV), em página exclusiva e protegida, pode ter acesso aos relatórios de classificação para *download*. Independentemente de ser um usuário geral ou um com acesso à aba restrita aos membros da CPV, estes podem interagir dinamicamente com o sistema, basicamente, via as seguintes formas:

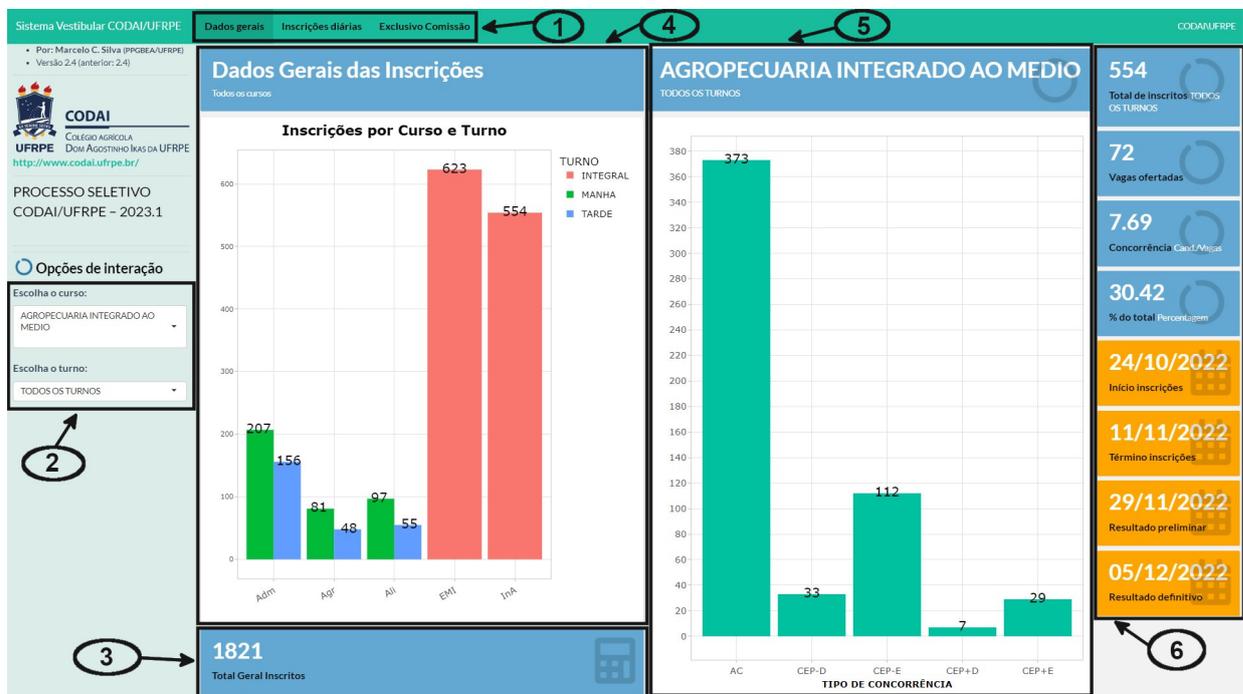
- ***mouse–click***: Interação que ocorre quando o usuário clica com o botão esquerdo do *mouse* em comandos, caixas de seleção, botões, em mapas e gráficos;
- ***mouse–over***: Interação que ocorre através da simples presença do cursor do *mouse* sobre algum elemento interativo, ocorrendo algum destaque na maioria dos elementos, tal como os botões de endereçamento de abas. Mas, o principal destaque desta interatividade são os gráficos, onde ocorre a apresentação de informações associadas ao ponto específico em relação ao posicionamento do cursor do *mouse*;
- ***brushing***: Interação onde o usuário escolhe um ponto da área de plotagem do gráfico (não é necessariamente um ponto que pertence ao gráfico) e clica no botão esquerdo do *mouse*, então “arrasta” o mouse de modo a desenhar um retângulo na área do gráfico que deseja visualizar com ênfase.

Quanto a estrutura do *site*, têm-se as abas intituladas Dados Gerais, Inscrições Diárias e Exclusivo Comissão, as quais serão detalhadas a seguir.

4.3.1.1 Aba *Dados gerais*

A aba “Dados gerais” é configurada como a página *web* inicial e surge ao se carregar o sistema através endereço do *site*. Nestas constam, gráficos e *dashboards* com resumo dos dados das inscrições dos diversos cursos/turnos ofertados, bem como informações relevantes do processo seletivo. É composta, basicamente, por seis áreas, as quais podem ser visualizadas na [Figura 15](#).

Figura 15 – Aba do sistema “Dados gerais”, configura-se como a aba padrão, oferece representações gráficas e *dashboards* associados às informações gerais sobre os dados das inscrições e datas relevantes do processo seletivo. Os detalhes dos elementos enumerados se encontram na [Tabela 4](#).



Fonte: Próprio autor.

Atendendo-se às seis áreas destacadas na [Figura 15](#), apresenta-se a [Tabela 4](#) com uma breve descrição de seus elementos:

Tabela 4 – Detalhamento dos elementos indicados na [Figura 15](#) para a aba “Dados Gerais”.

Usa-se o símbolo  para um objeto interativo e  para responsivo.

Item	Objeto	Descrição
	 interativo  responsivo	
1	 Menu de Abas	Acesso às abas do sistema, apresenta-se em todas as abas.
2	  Curso e turma	Controles de <i>input</i> do menu que permite a escolha do curso e turno para apresentação do gráfico presente no elementos (5), bem como nos <i>dashboards</i> que apresentam informações específicos por curso.
3	 <i>Dashboard</i>	Total geral de inscritos no processo seletivo.
4	  Área de Gráficos	Área do gráfico para os dados de inscrições de todos os cursos permitindo fácil comparação.
5	  Área de Gráficos	Área do gráfico específico para os dados do curso/turno escolhido pelo usuário permitindo analisar as inscrições por tipo de concorrência.
6	 <i>Dashboard</i>	<i>Dashboards</i> em azul apresentam resumo sobre os dados das inscrições do curso/turno selecionado pelo usuário constando o total de inscritos, o número de vagas ofertadas, a concorrência (razão candidato – vaga) e o percentual de inscritos no curso em relação ao total. Já em laranja, são apresentadas datas relevantes do processo seletivo.

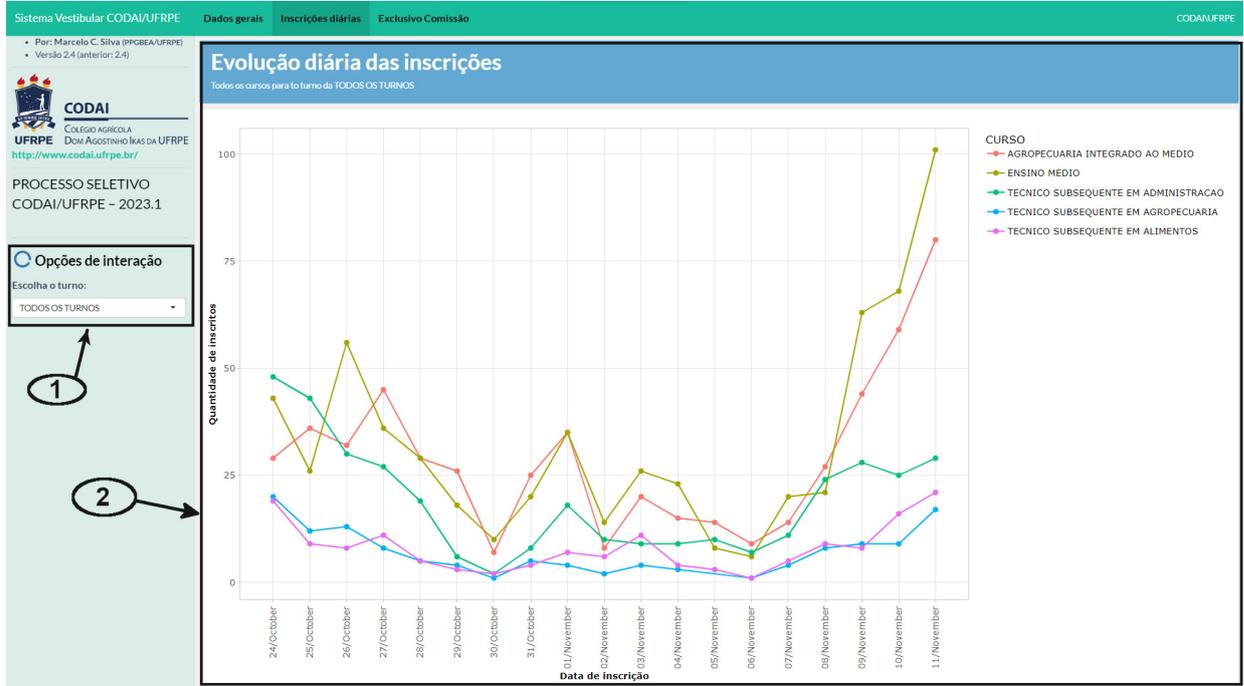
Fim da Tabela 4

Fonte: Próprio autor

4.3.1.2 Aba *Inscrições diárias*

A aba “Inscrições Diárias” é destinada para o acompanhamento da evolução das inscrições diariamente em todos os cursos ofertados. Apresenta um gráfico de linha para cada curso e o usuário pode escolher o turno de oferta dos cursos, possibilitando efetuar a comparação entre os mesmos. A aba é composta, basicamente, por duas áreas, as quais podem ser visualizadas na [Figura 16](#).

Figura 16 – Aba do sistema “Inscrições Diárias”, onde é possível acompanhar a dinâmica diária das inscrições no processo seletivo e em tempo real. Os detalhes dos elementos enumerados se encontram na Tabela 5.



Fonte: Próprio autor.

Considerando-se as áreas destacadas na Figura 16, apresenta-se a Tabela 5 com uma breve descrição de seus elementos:

Tabela 5 – Detalhamento dos elementos indicados na Figura 16 para a aba “Inscrições Diárias”.

Item	Objeto interativo responsivo	Descrição
1	Turno	Controle de <i>input</i> no menu que permite escolher o turno para comparação gráfica do número de inscrições. Inclui-se a opção “Todos os turnos”, a qual considera todos os turnos de um mesmo curso.
2	Área do gráfico	Área de apresentação do gráfico de linhas indicando o número de inscrições diárias com atualização, em tempo real, a realização de inscrições.

Continua na próxima página...

... Continuação da Tabela 5

Item	Objeto	Descrição
	 interativo  responsivo	

Fim da Tabela 5

Fonte: Próprio autor

4.3.1.3 Aba *Exclusivo Comissão*

A aba “Exclusivo Comissão” é destinada exclusivamente aos membros da Comissão Permanente de Vestibular, onde são disponibilizados os diversos relatórios dos resultados para *download* e informações relevantes para que o Departamento de Registro Escolar (responsável pelas matrículas) possa contactar os classificados e gerar planilhas com tais informações afim de proporcionar agilidade no processo. Ao clicar para acessar a aba, surge a tela apresentada na Figura 17, na qual são solicitados o *Username* e *Password* ao usuário. As credenciais de acesso são pré-definidas no escopo do sistema e não há a possibilidade do auto-cadastramento.

Figura 17 – Janela de login para área exclusiva e privativa aos membros da Comissão Permanente de Vestibular do CODAI/UFRPE.



Fonte: Próprio autor.

Uma vez que são inseridas as credenciais válidas na tela de autenticação descrita anteriormente, haverá, de fato, acesso à aba “Exclusivo Comissão”, a qual pode ser visualizada na Figura 18. Nota-se nesta, o destaque de quatro áreas de interatividade/responsividade. A com identificação (1) é constituída por dois objetos interativos: a caixa de seleção (*selectInput*) e o botão de comando (*downloadButtonRmd*), sendo esse último uma função criada com características específicas e baseado num botão genérico para atender às necessidades do sistema. A caixa de seleção proporciona ao usuário a escolha do tipo de relatório da

classificação que irá se fazer o *download*, com as opções: PROVISÓRIO, PRELIMINAR, DEFINITIVO ou REMANEJAMENTO. Todavia, o tipo de relatório padrão será definido com base nas datas características do processo seletivo, de modo que, se a data atual é menor que a de divulgação do resultado preliminar o padrão será PROVISÓRIO, se a data atual for maior ou igual a de divulgação do resultado preliminar e menor que a de divulgação do resultado definitivo será configurado PRELIMINAR e assim por diante.

Figura 18 – Aba do sistema “Exclusivo Comissão”, destinada aos membros da Comissão Permanente de Vestibular, e que disponibiliza os resultados da classificação e informações relevantes para o Departamento de Registro Escolar. Os detalhes dos elementos enumerados se encontram na Tabela 6. Os dados das colunas “NOME”, “EMAIL”, “CELULAR” E “CPF” estão ocultos por questões de privacidade.

The screenshot displays the 'Consulta rápida da classificação' page. The top navigation bar includes 'Sistema Vestibular CODAI/UFRPE' and 'Exclusivo Comissão'. The sidebar on the left contains the CODAI logo and navigation options. The main content area features a table with the following data:

Ord	NOME	EMAIL	CELULAR	NASC	CPF	TIPO	PoD	TURNO	Score	SIT
1				19/10/2007	77	AC	NAO	INTEGRAL	9.833	CLASSIFICADO
2				14/09/2008	17	AC	NAO	INTEGRAL	9.666	CLASSIFICADO
3				03/11/2007	00	AC	NAO	INTEGRAL	9.584	CLASSIFICADO
4				12/06/2007	24	AC	NAO	INTEGRAL	9.583	CLASSIFICADO
5				28/08/2007	00	AC	NAO	INTEGRAL	9.5	CLASSIFICADO
6				10/04/2008	54	AC	NAO	INTEGRAL	9.417	CLASSIFICADO
7				28/08/2007	19	AC	NAO	INTEGRAL	9.334	CLASSIFICADO
8				07/11/2007	80	AC	NAO	INTEGRAL	9.334	CLASSIFICADO
9				03/06/2008	39	AC	NAO	INTEGRAL	9.316	CLASSIFICADO
10				13/08/2007	03	AC	NAO	INTEGRAL	9.2	CLASSIFICADO
11				04/09/2007	76	CEPE	NAO	INTEGRAL	9.167	CLASSIFICADO
12				27/07/2007	54	AC	NAO	INTEGRAL	9.084	CLASSIFICADO
13				05/12/2007	04	AC	NAO	INTEGRAL	9	CLASSIFICADO
14				26/08/2007	93	AC	NAO	INTEGRAL	9	CLASSIFICADO
15				04/05/2008	32	AC	NAO	INTEGRAL	8.917	CLASSIFICADO

Fonte: Próprio autor.

Uma vez expostas as particularidades da área com identificação (1), levou-se em conta todas as áreas destacadas na Figura 18, por meio da Tabela 6, a qual apresenta uma breve descrição de seus elementos:

Tabela 6 – Detalhamento dos elementos indicados na Figura 18.

Item	Objeto	Descrição
	 interativo  responsivo	
1	  Controles	Controle de <i>input</i> do menu que permite ao usuário escolher que tipo de resultado deseja fazer <i>download</i> do relatório ao clicar no botão <i>Download.pdf</i> . As possibilidades são: PROVISÓRIO, PRELIMINAR, DEFINITIVO ou REMANEJAMENTO. Os relatórios são gerados em tempo real e contém o resultado de todos os cursos/turnos num único arquivo .pdf.
2	 Controles	Controles de <i>input</i> para escolher o curso, turno e a situação de classificação dos candidatos para apresentação dos resultados na área (3) da aba.
3	  Área de tabela	Área de apresentação do resultados conforme selecionados em (2). Contém informações com dados pessoais que não podem figurar no relatório .pdf de divulgação do resultado. Estas informações são importantes para uso do Departamento do Registro Escolar.
4	 Botões	Botões que oferecem opções para <i>download</i> dos resultados apresentados na área (3) em formatos distintos com extensões .csv e suportadas por planilhas com o Microsoft Excel ou Google Sheets.

Fim da Tabela 6

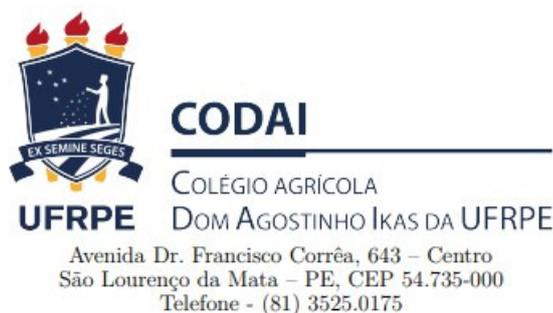
Fonte: Próprio autor

4.3.2 Relatórios com resultado de classificação

Os relatórios no formato PDF (Formato Portátil de Documento) são o meio principal pelo qual o sistema disponibiliza os resultados de classificação e que a Comissão Permanente de Vestibular divulga os resultados no *site* oficial do Colégio Agrícola Dom Agostinho Ikas (<http://www.codai.ufrpe.br/>). Os relatórios são constituídos de quatro partes: a capa, o sumário, as notas/observações da comissão e a relação de classificação

para cada curso/turno. Como exposto em seções anteriores, há as opções e relatórios para o resultado: PROVISÓRIO, PRELIMINAR, DEFINITIVO ou REMANEJAMENTO e, assim, uma capa e nota/observação da comissão para cada tipo de resultado. A Figura 19 exibe a capa do relatório para o caso de resultado definitivo, mas para os demais são mantidos os mesmos elementos.

Figura 19 – Capa e sumário empregados de forma típica a todos os relatórios contendo os resultados de classificação do processo seletivo. A capa em questão apresenta a data de 24/03/2023 por ter se efetuado o *download* da via o sistema nesta data.



RESULTADO DEFINITIVO

24/Mar/2023

Contents

1	INFORMAÇÕES DA COMISSÃO DE SELEÇÃO	3
2	AGROPECUARIA INTEGRADO AO MEDIO - INTEGRAL	4
2.1	Classificados - Ocupam vagas	4
2.2	Candidatos que podem ser remanejados - Poderão vir a ocupar vagas	6
2.3	Candidatos não classificáveis - Não podem ocupar vagas	10
3	ENSINO MEDIO - INTEGRAL	16
3.1	Classificados - Ocupam vagas	16
3.2	Candidatos que podem ser remanejados - Poderão vir a ocupar vagas	18
3.3	Candidatos não classificáveis - Não podem ocupar vagas	27
4	TECNICO SUBSEQUENTE EM ADMINISTRACAO - MANHA	29

Fonte: Próprio autor.

É importante ressaltar que muitas informações constantes nestes documentos no formato PDF, são implementados com codificação em R no documento R Markdown, de modo que suas informações são requeridas dos dados obtidos da planilha de configuração. Dados como datas, o título e semestre letivo do processo seletivo, entre outros, não precisam ser inseridos via linha de código no sistema.

Em seguida, na Figura 20, exibe-se a página que contém as notas/observações da Comissão Permanente de Vestibular para o caso do resultado definitivo. No entanto, nos demais tipos de relatórios contendo os resultados as notas/observações seguem o mesmo

modelo, mas, evidentemente, contendo as informações relevantes para o tipo de resultado em questão.

Figura 20 – Página de notas/observações da Comissão Permanente de Vestibular empregada no relatório contendo o resultado definitivo da classificação do processo seletivo. No entanto, para todos os demais relatórios o modelo é análogo, modificando-se o título e as informações pertinentes ao tipo de divulgação.

RESULTADO DEFINITIVO

1 INFORMAÇÕES DA COMISSÃO DE SELEÇÃO

A Comissão de Seleção do CODAI/UFRPE confere a documentação enviada pelos candidatos classificados, inclusive, em momentos posteriores a divulgação do resultado definitivo. Nas listas, as palavras estão sem os acentos para evitar problemas no sistema computacional de classificação bem como simplificar a codificação.

Salientamos que na divulgação do Resultado Final foram analisados os Pedidos de Recurso e que o sistema de classificação leva em conta o sistema de cotas.

A inscrição foi considerada **ANULADA** quando:

- a) Faltou documentação obrigatória;
- b) Houve divergências de informações declaradas pelo(a) candidato(a), no ato da inscrição;
- c) Apresentou outras irregularidades, previstas no Edital de Seleção 2023.1.

Nestes casos, novos(as) candidatos(as) ocuparam as vagas, respeitando a lista dos(as) que poderiam ser remanejados(as).

OBSERVAÇÃO: Candidatos(as) remanejáveis somente serão chamados(as) para a matrícula se houver vaga e se as informações e documentos, inseridos no ato da inscrição, estiverem corretos, conforme Edital do PROCESSO SELETIVO CODAI/UFRPE – 2023.1.

Os(as) candidatos(as) aprovados(as), que não comparecerem para a matrícula, no CODAI TIÚMA, com todos os documentos indicados no Edital, serão **DECLASSIFICADOS(AS)**.

Para mais informações, consulte o edital deste processo seletivo no site: www.codai.ufrpe.br

- Contato: selecoes.codai@ufrpe.br

Fonte: Próprio autor.

A última parte dos relatórios PDF com o resultado da classificação são os resultados em si, dispostos em tabelas, onde cada seção (iniciando da 2) corresponde a um curso/turno e as subseções são os tipos de correntes, tal como pode ser observado de forma parcial na [Figura 21](#). Para o processo seletivo referente ao semestre letivo de 2023.1, o referido relatório apresentou cinquenta e seis páginas, das quais cinquenta e três foram com as tabelas de ranqueamento dos candidatos classificados, dos remanejáveis (que podem vir a serem classificados nos remanejamentos) e com os candidatos que foram desclassificados ou tiveram a sua inscrição anulada.

Figura 21 – Início da página mostrando a composição dos resultados da classificação do processo seletivo. É composto por muitas páginas, de modo que cada curso/turno corresponde a uma seção. Os dados da coluna “NOME” estão ocultos por questões de privacidade.

2 AGROPECUARIA INTEGRADO AO MEDIO - INTEGRAL

2.1 Classificados - Ocupam vagas

2.1.1 Classificados Pessoas com Deficiência(PcD)

[1] "NÃO HÁ CANDIDATOS CLASSIFICADOS NESTE TIPO DE CONCORRÊNCIA OU NÃO HÁ VAGAS."

2.1.2 Classificados Ampla Concorrência(AC)

Ord	NOME	NASC	CPF	TIPO	Score	SIT
1		19/10/2007	***.176.524.**	AC	9.833	CLASSIFICADO
2		16/09/2008	***.906.894.**	AC	9.666	CLASSIFICADO
3		03/11/2007	***.324.994.**	AC	9.584	CLASSIFICADO
4		12/06/2007	***.759.834.**	AC	9.583	CLASSIFICADO
5		28/08/2007	***.350.684.**	AC	9.500	CLASSIFICADO
6		10/04/2008	***.732.544.**	AC	9.417	CLASSIFICADO
7		28/08/2007	***.386.494.**	AC	9.334	CLASSIFICADO
8		07/11/2007	***.730.024.**	AC	9.334	CLASSIFICADO
9		03/06/2008	***.381.184.**	AC	9.316	CLASSIFICADO
10		13/08/2007	***.796.234.**	AC	9.200	CLASSIFICADO
11		04/09/2007	***.846.334.**	CEP-E	9.167	CLASSIFICADO
12		27/07/2007	***.573.264.**	AC	9.084	CLASSIFICADO
13		05/12/2007	***.074.914.**	AC	9.000	CLASSIFICADO
14		26/08/2007	***.457.174.**	AC	9.000	CLASSIFICADO
15		06/05/2008	***.490.714.**	AC	8.917	CLASSIFICADO

Fonte: Próprio autor.

4.4 Conclusões e considerações finais

O sistema *web* desenvolvido utilizando a linguagem de programação R e os *packages* Shiny, Flexdashboard e R Markdown, foi projetado com o objetivo de promover a classificação dos candidatos do processo seletivo do Colégio Dom Agostinho Ikas da UFRPE, em conformidade com o sistema de cotas estabelecido pela Lei 12.711 de 29 de agosto de 2012, intitulado *Select CODAI*. Constituiu uma aplicação prática e factual, atendendo às demandas da governança de dados, desde o acesso aos dados até o tratamento, transformação e comunicação/visualização. O sistema comprovou sua eficiência e vantagens em comparação aos métodos tradicionais, em suas aplicações nos processos seletivos realizados nos semestres letivos 2021.1, 2021.2, 2022.1 e 2022.2. Fato corroborado pelo certificado de registro obtido por intermédio do Núcleo de Empreendedorismo e Inovação – IPÊ da UFRPE e presente no [Apêndice A](#).

Com interatividade–responsividade, o sistema obteve acesso aos dados por meio de planilhas da Google, fornecendo informações em tempo real sobre as inscrições por meio

de gráficos e *dashboards* em duas abas. Além disso, o sistema possui uma aba dedicada à comissão de seleção, acessível por senha, que permite o *download* dos relatórios em formato PDF com a classificação dos candidatos. Ao adotar abordagens interativo-responsiva, constatou-se uma experiência mais dinâmica e eficiente.

Em contraste com os métodos tradicionais, que muitas vezes são estáticos e demandam um processo manual demorado, o sistema interativo-responsivo trouxe inúmeras vantagens. A agilidade na visualização/comunicação das informações e resultados desejados, aliada à possibilidade de interação direta com as informações, permitiu uma resposta mais rápida às demandas da comissão de seleção. Isso resultou em maior eficiência, otimização do tempo e melhoria na tomada de decisões.

O sistema desenvolvido na linguagem de programação R e os *packages* Shiny, Flexdashboard e R Markdown, não só demonstrou ser uma solução eficiente no desenvolvimento de aplicativos voltados para as etapas de governança de dados focadas neste trabalho, atendendo às demandas de acesso, tratamento, transformação e comunicação/visualização, como usou o poder do R como linguagem de programação para resolver um problema específico e complexo.

Um passo futuro em relação ao *Select CODAI*, já em fase de desenvolvimento, é a implementação da classificação, em conformidade com o sistema de cotas estabelecido pela Lei 12.711 de 29 de agosto de 2012, através do sorteio. Promover a distribuição das vagas por meio randômico já é aplicado em diversas instituições, tal como o Colégio de Aplicação da UFPE e o Instituto Federal de Pernambuco. O Colégio Agostinho Dom Ikas da UFRPE já vislumbra aplicar o sorteio em seus processos seletivos. Desta forma, espera-se que a ferramenta passe a ser mais completa, ofertando as duas opções de métodos.

5 Sistema dos dados da Covid-19 para o Estado de Pernambuco

5.1 Introdução

Em 11 de março de 2020, a OMS – Organização Mundial da Saúde ([ORGANIZATION, 2022](#)), declarou a pandemia de *Sars-Cov-2*. Antes mesmo da decretação do período pandêmico, foi óbvia a necessidade de se acompanhar a evolução, sobretudo, do número de casos e óbitos diariamente, de modo a se ter condições de desenvolver políticas públicas no sentido de contenção da evolução da pandemia. Neste sentido, esforços foram despendidos por vários pesquisadores e instituições para se desenvolver estratégias de análises e apresentação dos dados associados a Covid-19, no intuito de auxiliar na tomada de decisão, principalmente, por parte do poder público. Dada a magnitude dos problemas observados na gestão dos sistemas de apoio à saúde coletiva, é evidente que todos os setores da sociedade carecem de informações claras e atualizadas sobre a pandemia. [Talagala e Shashikala \(2022\)](#), analisa quinze aplicativos com *dashboards*, onde há a aplicação de diversas técnicas para a visualização de dados da Covid-19, tais como o [Zoho \(2022\)](#) e [Shifa \(2022\)](#).

Acompanhando este propósito, mas focando o desenvolvimento de sistemas interativos-responsivos em R com a aplicação *packages* R Markdown/Flexdashboard, voltados para a etapa de governança de dados à comunicação/visualização, desenvolveu-se um sistema para a análise dos dados disponibilizados pelo Ministério da Saúde do Brasil associados à pandemia de Covid-19 para o Estado de Pernambuco, podendo este ser facilmente estendido para todo o Brasil.

Um fator determinante na utilização dos dados em questão é evidenciar a aplicação da linguagem de programação R na construção de sistemas interativos-responsivos, voltados para as etapas de governança de dados associadas à importação, organização, transformação, filtragem, modelagem e apresentação/comunicação. Considerando-se que os dados são disponibilizados diariamente desde 27/03/2020, temos até 26/04/2023 6.271.820 observações compostas por 17 variáveis, totalizando, portanto, uma matriz com 106.620.940 elementos, constituindo uma base de dados robusta para demonstrar a aplicação do R, no desenvolvimento dos sistemas propostos. Um segundo fator considerado consiste na pertinência do tema associado aos dados, bem como o fato de oportunizar a análise dos

dados oficiais disponibilizados pelo Ministério da Saúde do Brasil, sejam esses de boa qualidade ou não, pois o sistema permitirá ao usuário, de forma interativa–responsiva, ter suas próprias percepções. Ressalta-se ainda que a abordagem adotada serve como base para possíveis desafios futuros que demandem o desenvolvimento de ferramentas de apoio à gestão de dados públicos complexos, de maneira a oferecer soluções eficientes e de baixo custo.

5.2 Metodologia

5.2.1 Dados da Covid-19 do Ministério da Saúde do Brasil

Os dados utilizados podem ser obtidos no site do Ministério da Saúde do Brasil (SUS, 2020) e ao se fazer o *download* é obtido um arquivo no formato compactado com nome: “**HIST_PAINEL_COVIDBR_DDMMMYYYY.rar**”, onde **DD** é o dia, **MMM** o mês (jan, fev, mar, mai, ...) e **YYYY** é o ano até onde os dados são disponibilizados. Ao se prospectar os dados, observa-se que são viabilizados com data de início em 27/03/2020 e com a perspectiva de ocorrer atualização diária. Ao se efetuar a descompactação do arquivo *.rar*, descrito anteriormente, obtém-se uma pasta contendo diversos arquivos com extensão *.csv*, os quais são muito comuns no fornecimento de dados em geral cujo nome genérico é: “**HIST_PAINEL_COVIDBR_YYYY_ParteX_DDMMMYYYY.csv**”, onde **X** é a parte do arquivo associado ao ano **YYYY**. Já **DD** e **MMM** correspondem ao último dia e mês de observação disponível, respectivamente. Nota-se que cada ano (2020, 2021) apresentam duas partes.

Com relação às variáveis disponibilizadas, têm-se:

- “**regiao**”: Região associada a observação – Norte, Nordeste, Sul, Sudeste e Centro Oeste;
- “**estado**”: UF do Estado brasileiro – PE, BA, etc.;
- “**municipio**”: Nome do município da observação;
- “**coduf**”: Código da UF – Unidade da Federação ;
- “**codmun**”: Código do município;
- “**codRegiaoSaude**”: Código da região saúde;
- “**nomeRegiaoSaude**”: Nome da região saúde;
- “**data**”: Data da observação;

- “**semanaEpi**”: Semana epidemiológica;
- “**populacaoTCU2019**”: População do município associado à observação;
- “**casosAcumulado**”: Número de casos acumulados até o dia;
- “**casosNovos**”: Número de casos novos do dia;
- “**obitosAcumulado**”: Número de óbitos acumulados até o dia;
- “**obitosNovos**”: Numero de óbitos novos do dia;
- “**RecuperadosNovos**”: Número de pacientes recuperados;
- “**emAcompanhamentoNovos**”: Número de casos em acompanhamento;
- “**interior/metropolitana**”: Se é interior ou região metropolitana (0 ou 1).

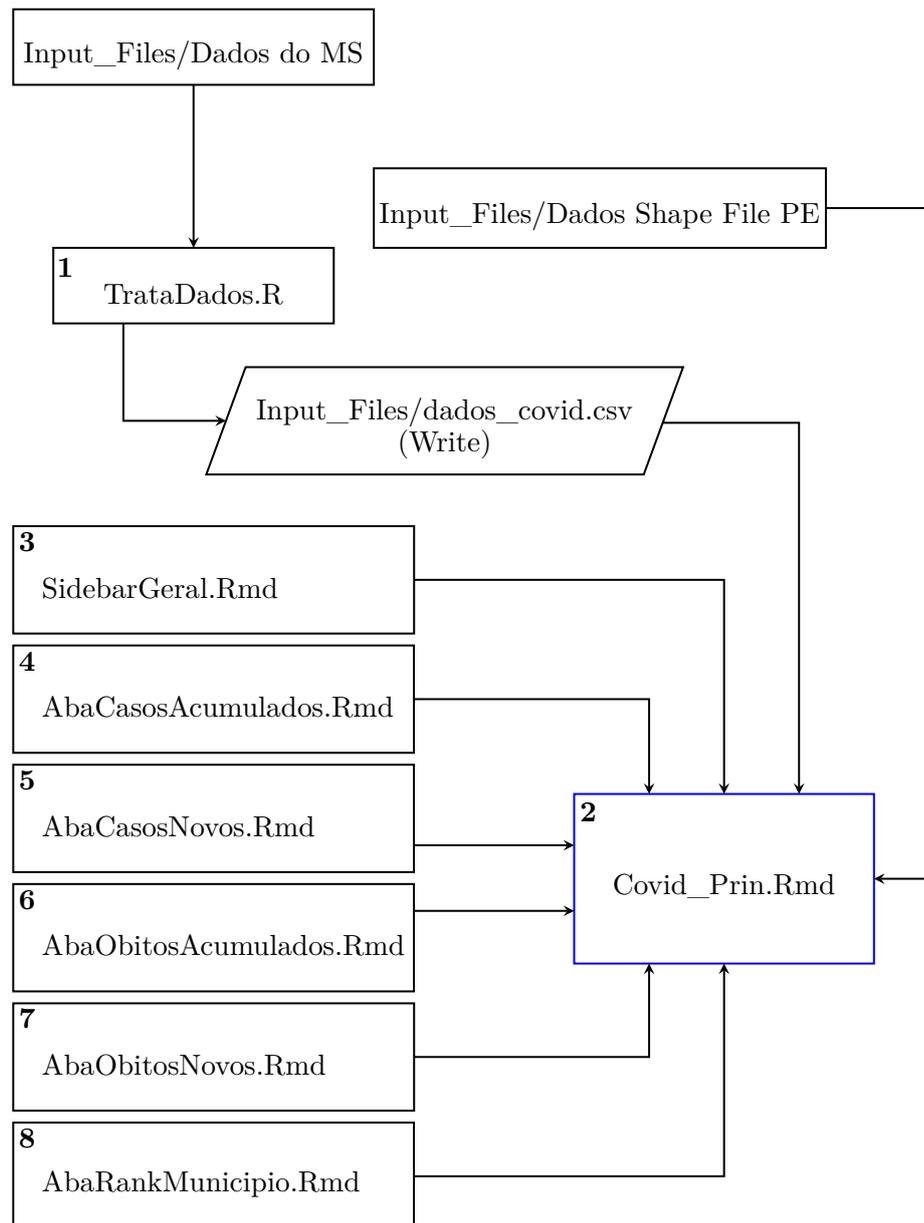
5.2.2 Arquitetura do sistema em R

A estrutura do sistema em R pode ser visualizada, de uma forma básica, através do diagrama, exibido na [Figura 22](#), que apresenta um único arquivo com extensão `.R` (*script* do R) e sete documentos `.Rmd` (R Markdown). O *script* do R (1) – `TrataDados.R` executa uma série de comandos, onde, em primeiro lugar, realiza uma busca na pasta `input_Files`, por meio da existência de parte do nome (“`HIST_PAINEL`”)¹, acessa os arquivos `.csv` com as partes separadas dos dados, armazenando-os em um objeto do tipo `list()`. Em seguida, promove a união das partes em um único objeto do tipo `data.frame()` e, por fim, escreve-os (`write_csv()`) em um arquivo com o nome `dados_covid.csv` gravado na mesma; O documento R Markdown (2) – `Covid Prin.Rmd` contém todos *packages* utilizados pelo sistema, apresenta um *chunk* R global que ler o arquivo `dados_covid.csv` gerado por (1), acessa os dados espaciais georreferenciados, unindo-os com os do `dador_covid.csv` usando como chave a variável `codmun` (código do município), bem como efetua mais uma série de tratamentos necessários. Neste `.Rmd` consta toda a estrutura que cria o sistema, incluindo o `YAML`. Todos os demais arquivos R Markdown são utilizados pelo `Covid Prin.Rmd` para construir o sistema; O arquivo (3) – `SidebarGeral.Rmd` é dedicado a criação da barra de menu lateral, chamada na documentação do shiny *Sidebar* ([RSTUDIO, 2022](#)); Os próximos `.Rmd`'s, (4) – `AbaCasosAcumados.Rmd`, (5) – `AbaCasosNovosdos.Rmd`, (6) – `AbaObitosAcumados.Rmd`, (7) – `AbaObitosNovos.Rmd` e (8) – `AbaRankMunicipio.Rmd` são, respectivamente, utilizados para gerar as abas do sistema *web*, cujos nomes são,

¹ Note-se que na [subseção 5.2.1](#), foi exposto que sempre os nomes dos arquivos disponibilizados tem esta sequência de caracteres.

respectivamente, Casos Acumulados, Casos Novos, Óbitos Acumulados, Óbitos Novos e Rank Municípios.

Figura 22 – Diagrama da estrutura do sistema em R desenvolvido para os dados do MS da Covid-19. Nota-se que há um *scripts* (.R) e sete documentos R Markdown envolvidos no funcionamento do software.



Fonte: Próprio autor.

5.2.3 Métricas utilizadas

É importante ressaltar que este trabalho não tem como propósito aprofundar-se nas grandezas e métricas aplicadas na epidemiologia, muito embora haja preocupação com as mesmas. A Prevalência Pontual e a Lápsica, por exemplo, são consideradas medidas

importantes em epidemiologia. A Prevalência Pontual é a proporção da população portadora da doença em determinado ponto do tempo, já a Prevalência Lápsica descreve os casos que estavam presentes em qualquer momento durante determinado período (FILHO, 2022). Nota-se que as prevalências são proporções em relação à população, do número de casos conhecidos da doença, retirando-se os que já foram curados, morreram e que migraram. Portanto, observando a descrição dos dados na [subseção 5.2.1](#), pode-se concluir que não há informações sobre as migrações e, principalmente, ao se inspecionar os dados em si, nota-se que a variável “RecuperadosNovos”, a qual conteria o número de curados (essencial para o cálculo das prevalências), muito embora exista, não contém dados e o que figura em todas as observações é NA (*Not Available*). Deste modo, percebe-se a inviabilidade em seguir o rigor da Epidemiologia e Bioestatística com os dados em questão. Por outro lado, muitas métricas com um grau tão refinado do ponto de vista da epidemiologia são relevantes para análise estes dados. Filho (2022), aponta que medidas matemáticas básicas, tal como razões, proporções e taxas tem papel fundamental nas análises epidemiológicas e, neste sentido, segue-se o caminho da aplicação de métricas genéricas.

Evidentemente que ao se deparar com uma massa de dados, à primeira vista, já se pensa em aplicar as métricas para descrevê-los sumarizadamente através das medidas de tendência ou localização e variabilidade. Segundo Bruce e Bruce (2019) as medidas de tendência é um passo fundamental na exploração de dados associando um “valor típico” para cada variável. Porém, também elucida que as métricas de localização correspondem a apenas uma dimensão na sumarização de uma variável, sendo extremamente importantes as medidas de variabilidade ou dispersão, as quais correspondem a uma segunda dimensão que indicam se os dados estão compactados ou espalhados.

Assim, implementou-se no sistema métricas associadas à tendência e dispersão, além de informações relevantes relacionadas a essas métricas. Contudo, ressalta-se que o grande diferencial associado são as características de interatividade–responsividade, as quais permitem ao usuário interagir e selecionar, por exemplo, um município específico, uma data específica ou um intervalo de datas associados aos seus interesses de observação. Por sua vez, o sistema atualizará, praticamente em tempo real, as medidas conforme a escolha do usuário. Por exemplo: escolhendo-se datas entre 01/05/2021 e 30/05/2021, o sistema conseguirá obter as métricas associadas ao mês de maio do ano de 2021. Dentre as quantidades presentes, destaca-se o número diário de casos novos e de óbitos, o número acumulado de casos e óbitos, o número diário de casos novos e de óbitos por 10^5 habitantes, o número acumulado de casos e óbitos por 10^5 , bem como as médias, medianas, modas e desvios-padrão dos casos e óbitos e, ainda, a média móvel (detalhada na [subseção 5.2.3.1](#)) dos casos e óbitos.

5.2.3.1 Média móvel simples e o *Package Zoo*

Uma das medidas mais utilizadas nos meios de comunicação como um indicador de tendência no acompanhamento e análise dos dados de casos e óbitos novos na pandemia da Covid-19 a partir do ano de 2020 foi a média móvel. Segundo [Venables e Smith \(2022\)](#), até então a média móvel tinha aplicação mais visível restritivamente ao mercado financeiro. Como o nome sugere, a média móvel é uma média que “se move” ao longo da sequência de dados. Deste modo, pode-se considerar que os dados existentes são “substituídos”, no período de extração da média, de modo que a média “se move” ao longo do tempo. Assim, além de se ter uma média do período considerado de extração, observa-se que a média configura uma medida de tendência, mas com um atraso que dependerá do tempo escolhido para o seu cálculo. Quando se trata de média, sabemos que há uma variedade destas, tais como a média aritmética simples, a média aritmética ponderada, a média harmônica e outras. No caso da média móvel, também temos essa variedade: média móvel aritmética, média móvel ponderada, média móvel exponencial, como pode ser evidenciado em [Machado \(2018\)](#). Para a análise dos dados no contexto da Covid-19, a média móvel aplicada é a média móvel aritmética simples e, portanto, vamos nos ater a mesma.

A média móvel aritmética simples consiste em uma média extraída de subconjuntos de uma sequência, considerando-se um certo número de observações, desde a primeira observação até a última. Deste modo, tem-se uma sequência de médias que “substituirá” a sequência de dados original. Formalizando, considere uma sequência $S = (p_i)_{i=1}^m = (p_1, p_2, \dots, p_m)$, com $i, m \in \mathbb{N}^*$ e as subsequências de S , $S'_j = (p_t)_{t=j}^{n+j-1}$, onde $\{j \in \mathbb{N}/1 \leq j \leq m - n + 1\}$ e $\{n \in \mathbb{N}/1 < n \leq m\}$ é número de termos para o cálculo da média móvel. Note que se $n = m$, tem-se a média da sequência original S . Então, nota-se que cada subsequência S'_j terá média móvel simples m_j , nas n observações dada pela [Equação 5.1](#).

$$m_j = \frac{\sum_{t=j}^{n+j-1} p_t}{n} \quad (5.1)$$

Deste modo, constrói-se a sequência de médias $M = (m_j)_{t=j}^{n+j-1}$, a qual é a média móvel simples da sequência S de n em n observações que oferecem uma suavização da curva representativa dos dados. Note-se que para uma sequência com m observações, as sequências de médias móveis extraídas de n em n observações, terá $m - n + 1$ termos. Assim, quanto maior n mais suavização se terá ao longo das observações.

Na linguagem de programação R, pode-se obter a média móvel aplicando a função *rollmean* do *package zoo* ([ZEILEIS; GROTHENDIECK, 2005](#)). Os principais parâmetros da referida função são apresentados na [Tabela 7](#).

Tabela 7 – Principais argumentos do função *Zoo::rollmean* para extração da média móvel simples.

Argumento	Descrição
x	Vetor, lista ou data.frame contendo a sequência numérica para extração da média móvel
k	Inteiro correspondente ao comprimento das sub-sequências para cálculo das médias
$fill$	Específica o que é inserido onde não há valor de média. Pode ser um valor ou NA
$align$	Específica onde será colocado o valor escolhido em $fill$. Tem-se as opções: “right”, “left” e colocando-se NULL será inserida no local a esquerda e direita, deixando os valores das médias no centro.
\vdots	

Fonte: Próprio autor

É importante ressaltar, com relação à apresentação da média móvel, que há implementação de interatividade–reatividade no sistema, de modo a oferecer ao usuário a liberdade de escolher o n (de quantos em quantos dias) na [Equação 5.1](#) para a construção da representação gráfica da referida média.

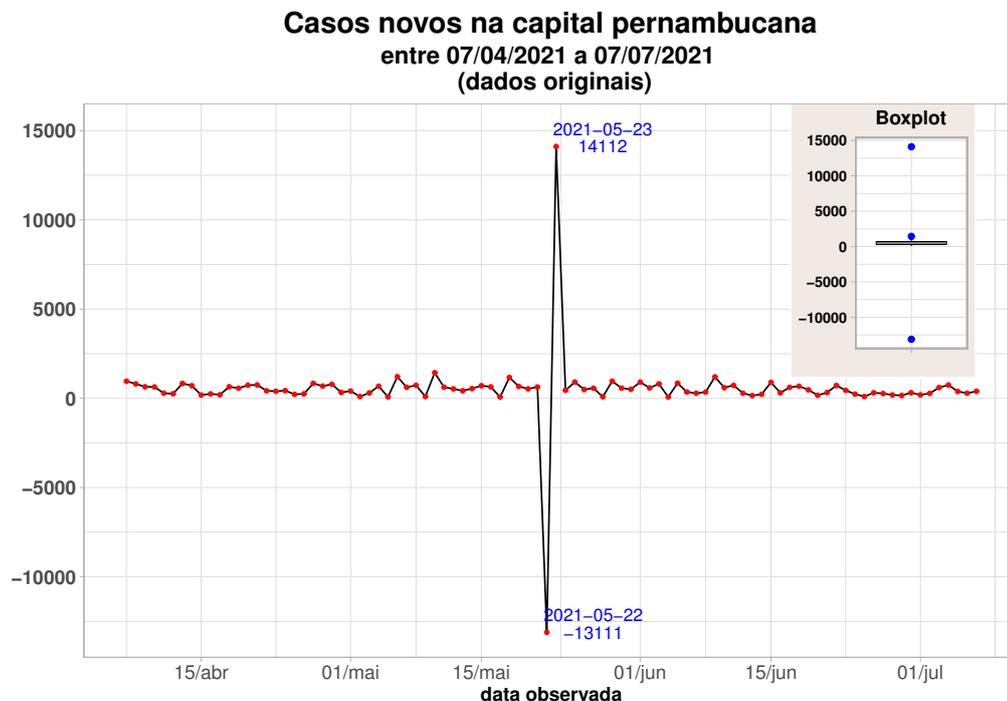
5.2.4 Implementação do z -Score aos dados

Dados associados a fenômenos diversos estudados no mundo real podem apresentar *outliers*, o que segundo [Bruce e Bruce \(2019\)](#) é qualquer valor que seja muito distante dos outros valores em um conjunto de dados. Porém, segundo os mesmos, a definição de *outliers* é muito subjetiva. Já [Saleem, Aslam e Shaukat \(2021\)](#), indicam que os *outliers* são valores discrepantes com observações extremas e que desviam significativamente dos valores “convencionais” das outras observações. Todavia, também preconiza que cada conjunto de dados tem sua própria definição exata do que é um *outlier*. Percebe-se que *outliers* surgem devido a diversos fatores possíveis, tais como: erro na coleta através de sensores defeituosos ou informação equivocada, erro de sistemas, erro humano e até mesmo comportamento fraudulento. A questão é que a presença destes valores discrepantes em conjuntos de dados pode proporcionar resultados inválidos, levar a análises estatísticas (ANOVA, Regressão Linear e outras) espúrias e prejudicar fortemente na tomada de decisão, quando for o caso. [Seo \(2006\)](#), avalia que os *outliers* podem levar a erros de especificação de modelos estatísticos e promover estimativas tendenciosas.

Deste modo, é muito importante perceber se há observações que desviam significativamente frente as demais e, neste sentido, [Saleem, Aslam e Shaukat \(2021\)](#) fazem um estudo focado na revisão e aplicação de métodos de detecção de *outliers* univariados, envolvendo métodos formais e informais. Dentre os informais, pode-se citar o Método SD, método do z-Score, método do z-Score modificado, regra da mediana, método de Tukey e método boxplot.

Os dados descritos na [subseção 5.2.1](#) apresentam valores que “explodem” e desviam significativamente das observações em geral, tal como pode ser evidenciado no gráfico apresentado na [Figura 23](#). Ele foi construído com os casos novos confirmados de infectados pelo vírus da Covid-19, em Recife do dia 07 de abril de 2021 a 07 de julho de 2021. Nele, observam-se dois pontos que de forma expressiva é fulgaz aos demais, os quais ocorrem em 22/05/2021 com indicação de -13.111 casos novos e 23/05/2021 com registro de 14.112 casos novos. Mesmo considerando, conjecturalmente, a explicação de que o valor referente ao dia 23/05/2021 seja uma correção do valor apresentado no dia 22/05/2021 (note-se que é negativo), há uma dificuldade e grande desconforto para uma visualização e análise gráfica da série, afetando a percepção do avaliador. Portanto, fica evidente que se deve considerar o tratamento de *outliers*.

Figura 23 – Dados originais dos casos novos de Covid-19 para a cidade de Recife de 07/04/2021 a 07/07/2021.



Fonte: Próprio autor.

Há muitas técnicas para efetuar a identificação de *outliers* nos dados e, neste

sentido, recorremos ao estudo de [Saleem, Aslam e Shaukat \(2021\)](#) citado anteriormente. Porém, considerando-se que este trabalho está focado em demonstrar o quão importante é um sistema de apresentação/comunicação de dados interativo-responsivo, não se entra em detalhes sobre o tratamento de *outliers* e foi escolhido o método do *z-Score* para tratar os valores discrepantes dos dados em questão, o qual segundo [Iglewicz e Hoaglin \(1993\)](#) não é apropriado para rotular *outliers* em conjuntos de dados pequenos o que, obviamente, não o caso tratado no presente capítulo deste trabalho.

O *z-Score* é um método que utiliza dois estimadores para detectar *outliers*, sendo estes a média (\bar{x}) e o desvio padrão ($sd = \sigma^2$), onde σ é a variância. A [Equação 5.2](#) mostra o *z-Score* (z_i) associado à observação x_i .

$$z_i = \frac{x_i - \bar{x}}{sd} \quad (5.2)$$

de forma que, considera-se $x_i \sim N(\mu, \sigma^2)$. A ideia básica é que se x_i segue uma distribuição normal com média μ e variância σ^2 , então z_i segue uma distribuição normal padrão, ou seja, $\{z_i \sim N(0, 1)\}$ e observações x_i , com *z-Score* em valor absoluto $|z_i|$ maiores que 3 são considerados *outliers* ([KALIYAPERUMAL et al., 2015](#)). Porém, [Seo \(2006\)](#) considerando o trabalho de [Shiffler \(1988\)](#), mostra que o *z-Score* máximo para se rotular *outliers*, depende do tamanho da amostra, de modo que o *z-Score* máximo (z_{max}) para uma amostra de tamanho n é dado pela equação a seguir:

$$z_{max} = \frac{n - 1}{\sqrt{n}} \quad (5.3)$$

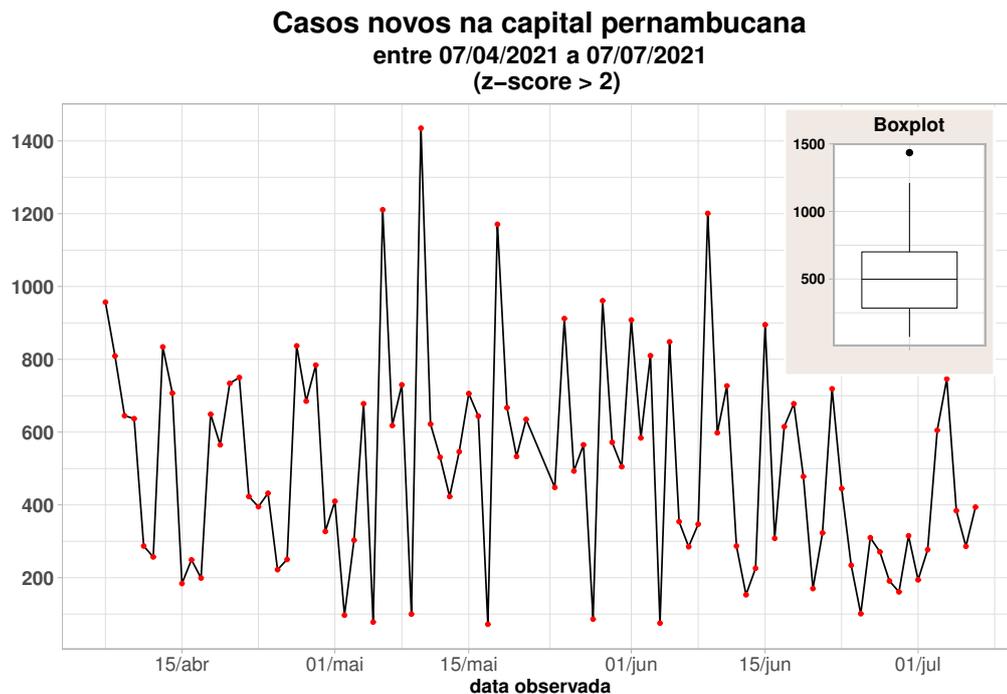
Para exemplificar, utilizam-se os dados já empregados na construção do gráfico exibido na [Figura 23](#), em que 92 observações são apresentadas. Porém, para os cálculos do *z-Score* e de seu máximo, consideram-se todas as observações para o município em questão (Recife) e como há diariamente inserção de novas medidas, conseqüente o *z-Score* e seu máximo é atualizado. Deste modo, para construir o exemplo, concebem-se os dados até 03/06/2022, correspondentes a $n = 799$ observações, com média $\mu = 297,14$ e desvio padrão $\sigma^2 = 750,91$. Daí, através da [Equação 5.3](#) obtemos que $z_{max} = 28,23$ e calcula-se os z_i através da [Equação 5.2](#). A [Tabela 8](#) mostra um recorte das observações em questão, onde se destacam os *outliers* evidenciados claramente no gráfico da [Figura 23](#), bem como os *z-Scores* (z_i) calculados para os respectivos x_i , considerando-se as observações até a data limite estipulada. Já a [Figura 24](#) exhibe uma plotagem do gráfico apresentado na [Figura 23](#), porém com um filtro excluindo as observações que foram consideradas *outliers*. É perceptível como a ausência dos pontos destacados anteriormente proporciona uma melhor visualização e percepção dos dados por meio do gráfico. Além disso, construiu-se o gráfico apresentado na [Figura 25](#), o qual difere do anterior pelo fato de os *outliers* terem sido substituídos pela média em detrimento de serem simplesmente suprimidos.

Tabela 8 – Recorte das observações enfatizando os *outliers* evidenciados no gráfico da Figura 23, bem como os *z-Scores* (z_i) calculados para os respectivos x_i .

$n = 799$; $\mu = 297,14$; $\sigma^2 = 750,91$; $z_{max} = 28,23$ Considera-se <i>outliers</i> se $ z_i > 2$				
data	x_i	z_i	check	Outlier
07/04/2021	957	0,88	$ z < 2$	NÃO
⋮				
19/05/2021	667	0,49	$ z < 2$	NÃO
20/05/2021	533	0,31	$ z < 2$	NÃO
21/05/2021	635	0,45	$ z < 2$	NÃO
22/05/2021	-13111	-17,86	$ z > 2$	✓ SIM
23/05/2021	14112	18,40	$ z > 2$	✓ SIM
24/05/2021	448	0,20	$ z < 2$	NÃO
25/05/2021	912	0,82	$ z < 2$	NÃO
26/05/2021	493	0,26	$ z < 2$	NÃO
⋮				
07/07/2021	394	0,13	$ z < 2$	NÃO

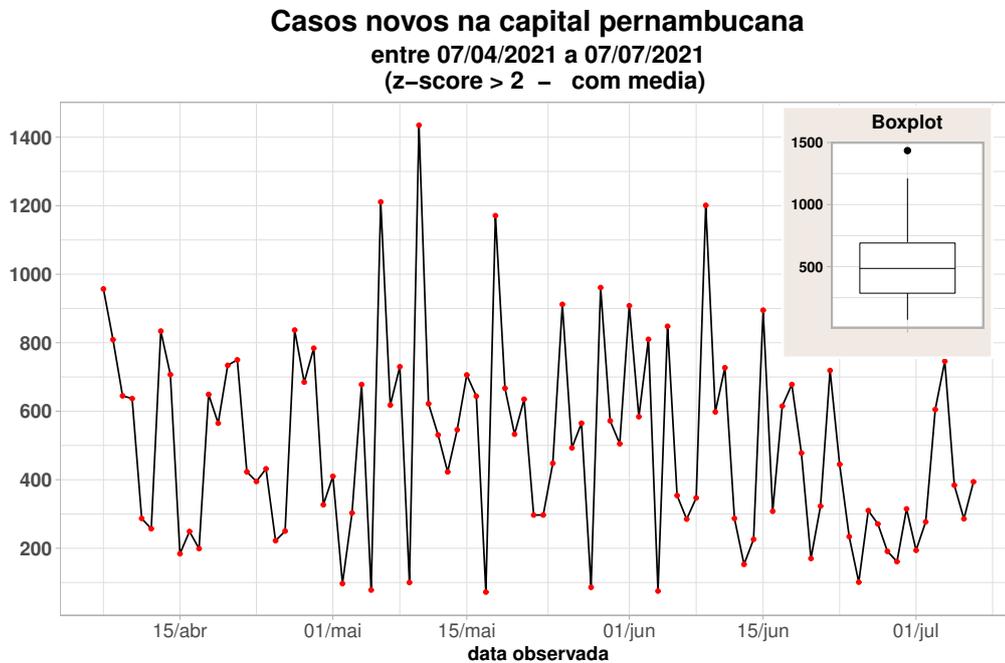
Fonte: Próprio autor

Figura 24 – Dados com filtragem suprimindo os *outliers* ($z_i > 2$) dos casos novos da Covid-19 para a cidade de Recife de 07/04/2021 a 07/07/2021.



Fonte: Próprio autor.

Figura 25 – Dados com filtragem substituindo os *outliers* ($z_i > 2$) pelo inteiro mais próximo a média dos casos novos da Covid-19 para a cidade de Recife de 07/04/2021 a 07/07/2021.



Fonte: Próprio autor.

É significativo para este trabalho, o fato do sistema de visualização/comunicação dos dados oferecer ao usuário a opção de manter os dados originais ou aplicar os filtros, acionando um botão de comando intitulado “Usar z-Score” e, ainda, considerando valores diferentes para o *z-Score*, por meio de um *slider input* intitulado “Escolher |z-score|”. Além disso, o usuário decidirá se deseja que os valores considerados *outliers* sejam simplesmente suprimidos ou substituídos pelo inteiro mais próximo à média, mediante um botão de comando.

É importante ressaltar que a aplicação do *z-Score* considera todos os dados de Pernambuco para fazer o tratamento de *outliers*. Portanto, não considera os municípios individualmente e nem o intervalo de datas escolhido pelo usuário no processo de interação.

5.3 Resultados e discussão

5.3.1 O Sistema interativo–responsivo local ou *web*

O sistema apresentado neste capítulo pode ser executado localmente, havendo a necessidade de se ter o R instalado, bem como os *packages* necessários e, preferencialmente, o RStudio. No entanto, pode-se acessar e usar o sistema em sua publicação na *web* por

meio do *site* https://appsmarcelo.shinyapps.io/Covid19_V3_Prin/ (SILVA MARCELO C. SILVA, 2022) e, neste caso, não há a necessidade de ter conhecimento do R, seus *packages* e, muito menos, ter os mesmos instalados no computador, pois, o aplicativo será executado em um servidor na “nuvem” e acessado por meio de navegadores convencionais, para a navegação na internet. O Sistema apresenta em suas cinco abas diversos controles de interação, plotagem de gráficos interativos, *dashboards* e outros elementos de interesse, onde o usuário poderá interagir, basicamente, através das seguintes formas:

- ***mouse-click***: Interação que ocorre quando o usuário clica com o botão esquerdo do *mouse* em comandos, caixas de seleção, botões, em mapas e gráficos;
- ***mouse-over***: Interação que ocorre através da simples presença do cursor do *mouse* sobre algum elemento interativo, ocorrendo algum destaque na maioria dos elementos, tal como os botões de endereçamento de abas. Mas, o principal destaque desta interatividade são os gráficos, nos quais ocorrem a apresentação de informações associadas ao ponto específico atrelado ao posicionamento do cursor do *mouse*;
- ***brushing***: Interação onde o usuário escolhe um ponto da área de plotagem do gráfico (não é necessariamente um ponto que pertence ao gráfico) e clica no botão esquerdo do *mouse*, então “arrasta” o mouse de modo a desenhar um retângulo na área do gráfico que deseja visualizar com ênfase.
- ***scroll***: Interação onde o usuário pode usar o *scroll* (roda do *mouse*) na área do mapa para promover a sua ampliação e redução. Porém, caso o *mouse* do usuário não apresente o *scroll* há dois botões com os símbolos + e – que são destinados a efetuar a ampliação e redução o mapa, respectivamente.

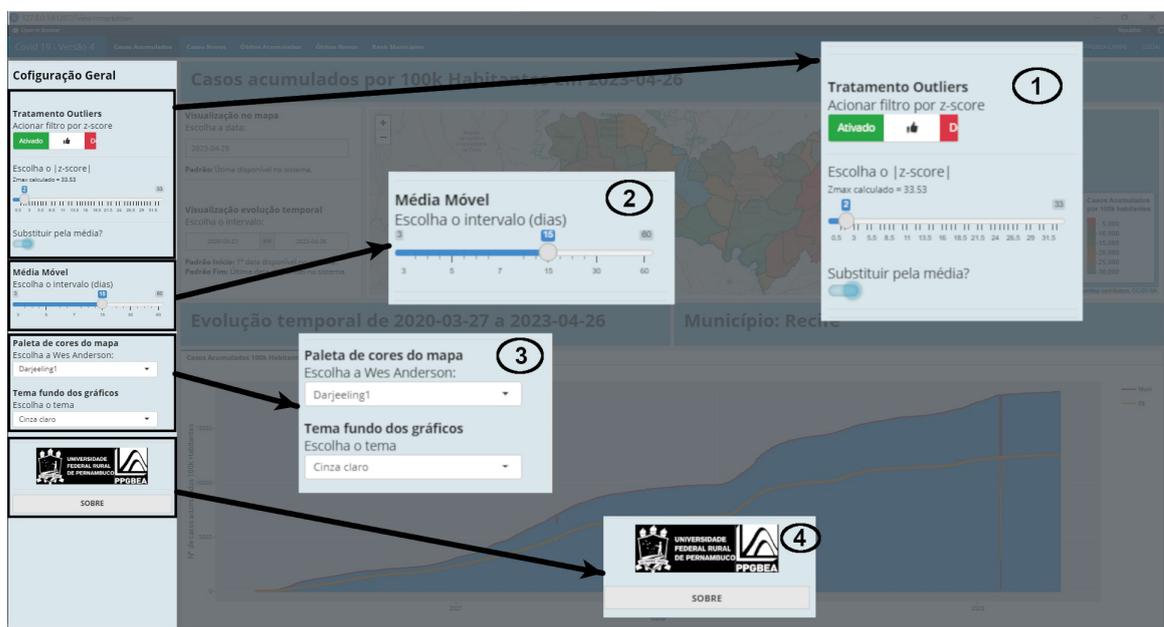
Quanto à estrutura do aplicativo, há em todas as abas, a barra de menu lateral (*Sidebar*) e as cinco abas presentes são intituladas de Casos Acumulados, Casos Novos, Óbitos Acumulados, Óbitos Novos e Rank Municípios, as quais serão detalhadas a seguir.

5.3.2 A barra lateral de menu – *Sidebar*

A barra de menu lateral do sistema é global, intitulada Configuração Geral, e presente em todas as abas, é uma ferramenta importante, por meio da qual ocorrem diversas interações e escolhas de configuração do usuário. A Figura 26 exibe os detalhes dos elementos que estão presentes na barra. A parte 1 apresenta um botão de comando para ativar/desativar a aplicação do filtro de *outliers* através do z-Score, um *slider input* que permite a escolha do nível de aplicação do filtro, e um segundo botão de comando que permite a escolha da substituição dos *outliers* pela média; A parte 2 permite ao usuário,

através de um *sliders input*, escolher o intervalo diário para o cálculo das médias móveis que são apresentadas em gráficos e oferece as opções de 3, 5, 7, 15, 30 ou 60 dias; A parte 3 permite ao usuário escolher a paleta de cores, baseada no *package* Wesanderson (RAM; WICKHAM, 2018), para a representação gráfica nos mapas de cores presentes nas quatro primeiras abas do sistema, além da escolha da cor para fundo na área de plotagem dos gráficos de linhas; Por fim, em 4, tem-se um botão de comando para acessar uma janela com informações da autoria do *software*.

Figura 26 – Detalhamento da barra de manu lateral presente em todas as abas do sistema.

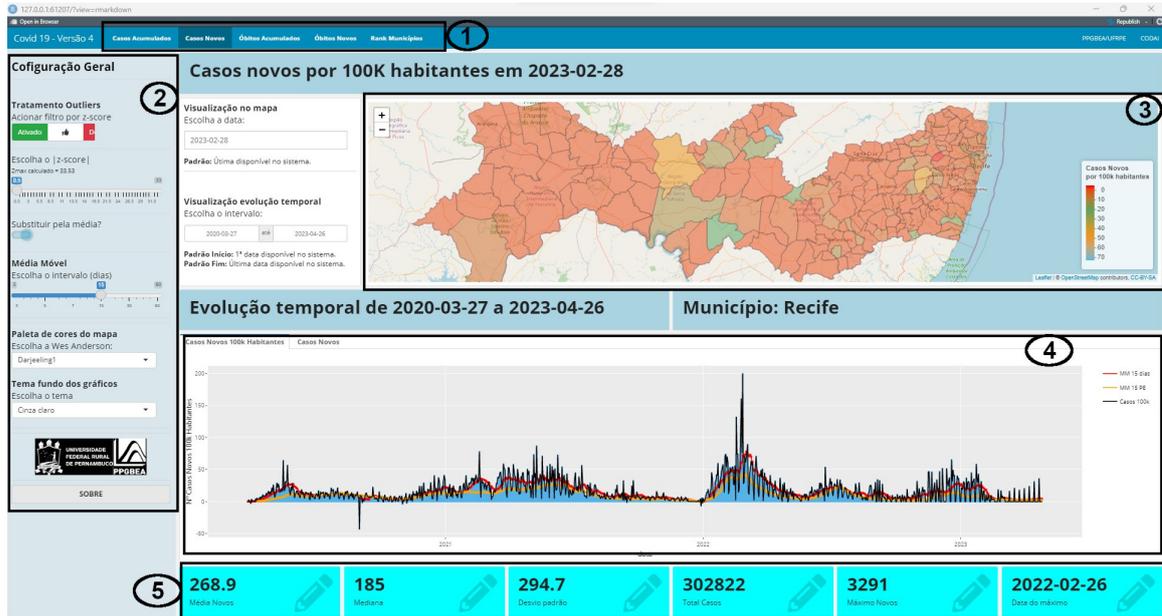


Fonte: Próprio autor.

5.3.3 Página típica do sistema

O aplicativo, como já mencionado, é composto por cinco abas. As denominadas Casos Acumulados, Casos Novos, Óbitos Acumulados e Óbitos Novos apresentam basicamente a mesma estrutura, havendo uma diferenciação entre as dos dados acumulados em relação aos novos, pelo fato das abas de dados novos apresentarem uma área para *dashboards* e a presença da média móvel na área dos gráficos. A quinta aba, intitulada Rank Municípios, tendo uma estrutura diferente. Considerando a semelhança entre as quatro primeiras abas, utiliza-se da aba de Casos Novos, exibida na Figura 27, para apresentar a estrutura típica das abas do sistema. Neste sentido, é possível observar o destaque de cinco áreas na figura: a primeira corresponde à barra que permite acesso às diversas abas, a segunda é *SideBar* de configuração geral, a terceira corresponde ao mapa de Pernambuco, a quarta é uma área dedicada para plotagem de gráficos e a quinta são os *dashboards*.

Figura 27 – Apresentação de uma página típica do sistema. Os detalhes dos elementos enumerados se encontram na Tabela 9.



Fonte: Próprio autor.

As páginas oferecem navegação intuitiva, e ao se mover o cursor do *mouse* nos diversos pontos das abas, ocorre interatividade do tipo *mouse-over*, indicando ao usuário, através da mudança do aspecto do cursor, a existência de elemento interativo. Para melhor clareza, a Tabela 9 resume as funções dos principais elementos.

Tabela 9 – Detalhamento dos elementos indicados na Figura 27.

Item	Descrição
1	Navegação pelas abas do sistema: Casos Acumulados, Casos Novos, Óbitos Acumulados, Óbitos Novo e Rank Municípios.
2	Menu de configuração geral, o qual reflete em todo o sistema (veja a subseção 5.3.2).
3	Área de plotagem do mapa de Pernambuco interativo-responsivo, detalhado na subseção 5.3.4.
4	Área de plotagem dos gráficos interativo-responsivo, podendo haver mais de uma guia (mais de um conjunto de gráficos).
5	Área de apresentação dos <i>dashboards</i> com métricas de tendência, variabilidade e outras quantidades de interesse.

Fonte: Próprio autor.

5.3.4 Seleção de dados por município/cidade

Ao se apresentar as diversas visualizações associados aos dados descritos na [subseção 5.2.1](#), construiu-se uma forma intuitiva e simples para o usuário interagir com o sistema no que diz respeito à seleção do município de interesse. Neste sentido, nas abas Casos Acumulados, Casos Novos, Óbitos Acumulados e Óbitos Novos, tem-se um mapa de Pernambuco construído através dos dados do *package geobr* (PEREIRA; GONCALVES, 2022), na qual o usuário poderá clicar no polígono associado à cidade desejada, levando à apresentação das visualizações para os dados do tema da aba em questão, do município selecionado. Ao se clicar numa cidade, o aplicativo exibe uma guia suspensa contendo um resumo informacional para o município em questão.

No mapa de Pernambuco, há uma plotagem das cidades em escala de cores, considerando-se o número de casos acumulados, casos novos, óbitos acumulados ou óbitos novos, respectivamente à aba que se está navegando, com todas as quantidades mencionadas calculadas para 100.000 habitantes, objetivando percepção comparativa entre as cidades em uma determinada data, cuja configuração padrão é a última disponível nos dados.

A [Figura 28](#) exibe o mapa anunciado para a aba de casos acumulados como exemplo. À esquerda do mesmo, observa-se o objeto de interação do tipo *date input*, através do qual o usuário escolhe a data de visualização e a construção da escala de cores. Ao se escolher uma nova dada, há atualização da mesma no título de forma responsiva. O menu suspenso aparece à direita, onde foi clicado no município de São Lourenço da Mata. Porém, a cidade padrão, ao se abrir o sistema, é a capital pernambucana.

Figura 28 – Mapa de cores interativo–responsivo da aba casos acumulados acompanhado do *date input* para escolha da data de visualização. Observa-se que a data escolhida é 26 de abril de 2023, bem como foi clicado no mapa a cidade de São Lourenço da Mata.

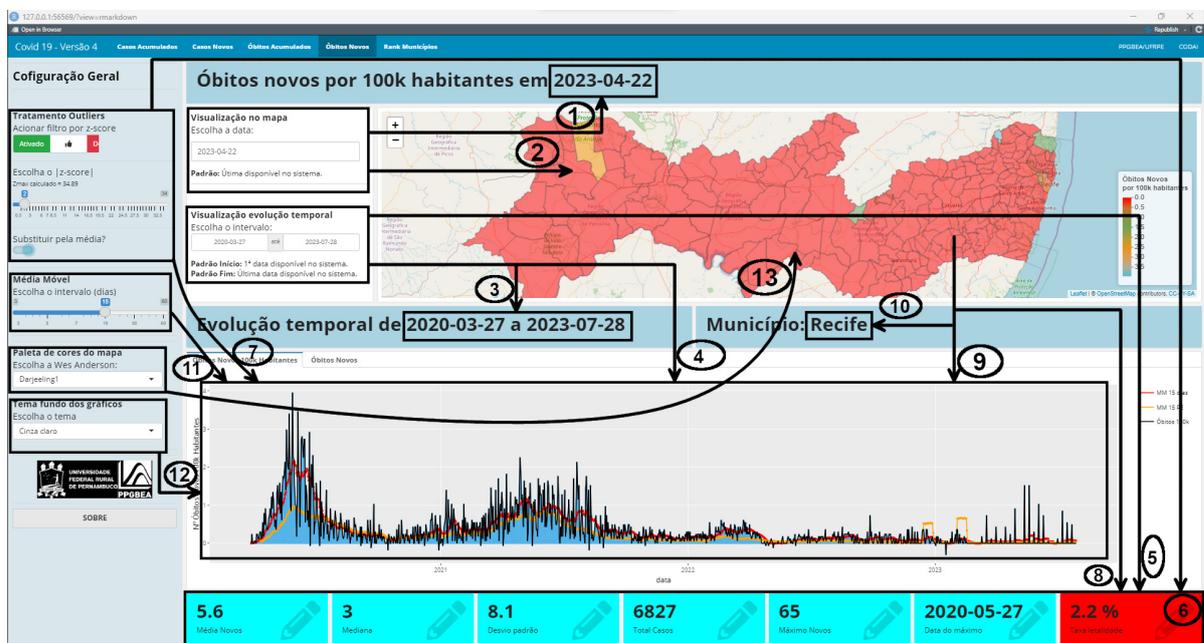


Fonte: Próprio autor.

5.3.5 Diagrama de interatividade–reatividade

Para demonstrar a complexa trama de interatividade–responsividade do sistema, apresenta-se a [Figura 29](#), na qual é exibido um diagrama com as relações interatividade–responsividade entre os objetos de uma página típica. Todas as abas do sistema apresentam interatividade–responsividade, mas para construir o diagrama em questão, foi escolhida a aba Óbitos Novos, como um exemplo, que ajudará a perceber o resultado da trama de interatividade–responsividade.

Figura 29 – Diagrama das relações de interatividade–reatividade na aba de Óbitos Novos, configurando-se uma situação típica para o aplicativo. Os detalhes indicativos da interatividade–reatividade são apresentados na [Tabela 10](#).



Fonte: Próprio autor.

Para efetuar a apresentação dos elementos de uma forma simples, elaborou-se a [Tabela 10](#), descrevendo a interatividade–responsividade das setas presentes no diagrama. As setas são traçadas dos objetos de interatividade para os de respectiva reatividade e, neste sentido, ressalta-se que há objetos de interatividade que podem afetar mais de um para a reatividade e vice-versa. É salutar perceber que um mesmo objeto pode ser interativo e responsivo, tal como ocorre com o mapa de Pernambuco, que é utilizado para selecionar a cidade e ao mesmo tempo sofre modificações baseada na escolha da data e da paleta de cores para sua plotagem.

Tabela 10 – Detalhamento das relações dos objetos de interatividade–reatividade presentes na Figura 29. Usa-se o símbolo  para um objeto interativo e  para reativo.

Item	Objeto		Descrição
	 interativo	 responsivo	
1	 Visualização no mapa.	 Data.	Modificando-se a <u>Visualização no mapa</u> , o sistema reage atualizando no título do mapa a <u>Data</u> proporcionando ao usuário claramente percepção da data selecionada.
2	 Visualização no mapa.	 Mapa de cores.	Modificando-se a <u>Visualização no mapa</u> , ocorrerá uma reação do sistema para a distribuição espacial plotada no <u>Mapa de cores</u> .
3	 Visualização evolução temporal.	 Intervalo de datas.	Atualizando-se a <u>Visualização evolução temporal</u> , haverá reação atualizando o <u>Intervalo de datas</u> apresentado no cabeçalho da área de plotagem dos gráficos que constam a evolução temporal.
4	 Visualização evolução temporal.	 Gráficos.	Atualizando-se a <u>Visualização evolução temporal</u> , ocorrerá a atualização dos <u>Gráficos</u> , considerando-se as datas de início e fim selecionadas.
5	 Visualização evolução temporal.	 <i>Dashboard</i> de métricas.	Atualizando-se a <u>Visualização evolução temporal</u> , ocorrerá a atualização das métricas de tendência, variabilidade e outras, caso exista, na área de <u><i>Dashboard</i> de métricas</u> .
6	 Tratamento <i>Outliers</i> .	 <i>Dashboard</i> de métricas.	Em <u>Tratamento <i>Outliers</i></u> , ao ativar/desativar o botão (Acionar filtro por z-score), o sistema reagirá nos <u><i>Dashboard</i> de métricas</u> , considerando-se a escolha para o valor absoluto do <i>z-Score</i> no <i>slider input</i> , bem como se haverá a substituição pela média.
7	 Tratamento <i>Outliers</i> .	 Gráficos.	Em <u>Tratamento <i>Outliers</i></u> , ao ativar/desativar o botão (Acionar filtro por z-score), o sistema reagirá atualizando os <u>Gráficos</u> na área de plotagem, considerando-se a escolha do $ z\text{-Score} $ no <i>slider input</i> , bem como se haverá a substituição pela média.

Continua na próxima página...

... Continuação da Tabela 10

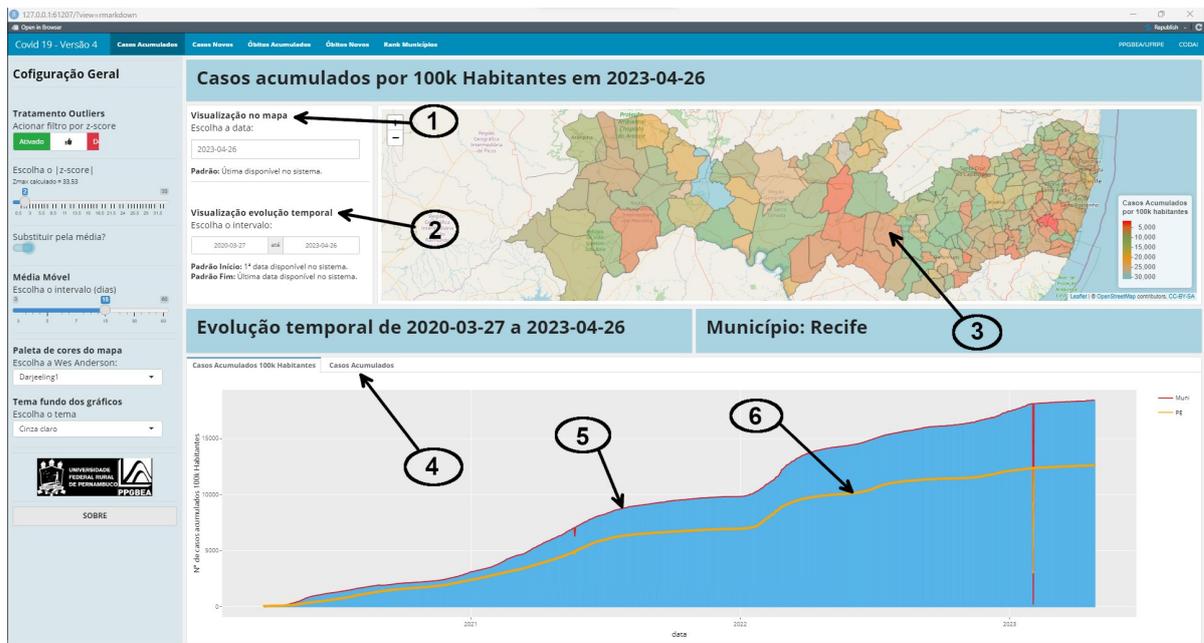
Item	Objeto		Descrição
	 interativo	 responsivo	
8	 Mapa.	 <i>Dashboard</i> de métricas.	Ao se clicar em um município do <u>Mapa</u> , o sistema reagirá atualizando os <u><i>Dashboard</i> de métricas</u> , considerando-se os dados para o município escolhido pelo usuário.
9	 Mapa.	 Gráficos.	Ao se clicar em um município do <u>Mapa</u> , o sistema reagirá atualizando os <u>Gráficos</u> , na área de plotagem de gráficos, considerando-se os dados para o município selecionado pelo usuário.
10	 Mapa.	 Municípios.	Ao se clicar em um município do <u>Mapa</u> , o sistema reagirá atualizando os <u>Município</u> , na barra de título que apresenta o município associado a plotagem dos gráficos.
11	 Média Móvel.	 Gráficos.	Ao se escolher o intervalo de cálculo para a média móvel no <i>slider inpu</i> em <u>Média Móvel</u> , o sistema reagirá atualizando a área dos gráficos os <u>Gráficos</u> da média móvel, considerando-se o intervalo de cálculo selecionado pelo usuário.
12	 Tema fundo dos gráficos.	 Gráficos.	Ao modificar o tema, em <u>Tema fundo dos gráficos</u> , o sistema reagirá atualizando na área de plotagem dos gráficos o tema do <u>Gráficos</u> , promovendo uma visualização confortável conforme as preferências do usuário.
13	 Paleta de cores do mapa.	 Mapa de cores.	Ao modificar a paleta de cores em <u>Paleta de cores do mapa</u> , o sistema reagirá atualizando a mesma na plotagem do <u>Mapa de cores</u> , promovendo uma visualização confortável conforme as preferências do usuário.

Fim da Tabela 10

5.3.6 Aba Casos Acumulados

Na aba Casos Acumulados, apresentada na [Figura 30](#), são exibidas três áreas: A de configuração geral, o mapa de Pernambuco e a área de plotagem dos gráficos. Na área para a apresentação dos gráficos, há duas sub-abas: Casos Acumulados 100k Habitantes e Casos Acumulados. Na apresentação do número de casos acumulados por 100.000 habitantes, são traçadas duas curvas: uma para o município escolhido (traçado em vermelho) e outra para o Estado e Pernambuco (traçado em laranja), permitindo a comparação da cidade selecionado como Estado.

Figura 30 – Principais elementos da aba Casos Acumulados. Os objetos numerados são descritos na [Tabela 11](#). A exibição corresponde ao dia 15/06/2022 para o mapa de cores e a plotagem dos gráficos de linha são para capital pernambucana de 27/03/2020 a 26/04/2023.



Fonte: Próprio autor.

Obviamente, a interação em objetos da barra lateral Configuração Geral afeta as visualizações da aba Casos Acumulados, mas como foi descrito na [subseção 5.3.2](#), omite-se aqui. Para uma melhor compreensão dos elementos enumerados na figura anterior, considere a descrição apresentada de forma resumida na [Tabela 11](#). Percebe-se que são usados os símbolos  e , respectivamente, para destacar que o objeto em questão é interativo ou responsivo. Os objetos que figuram ambos símbolos, são, simultaneamente, interativo e responsivo.

Tabela 11 – Detalhamento dos elementos apresentados na [Figura 30](#), para a aba Casos Acumulados. Usa-se o símbolo  para um objeto interativo e  para reativo.

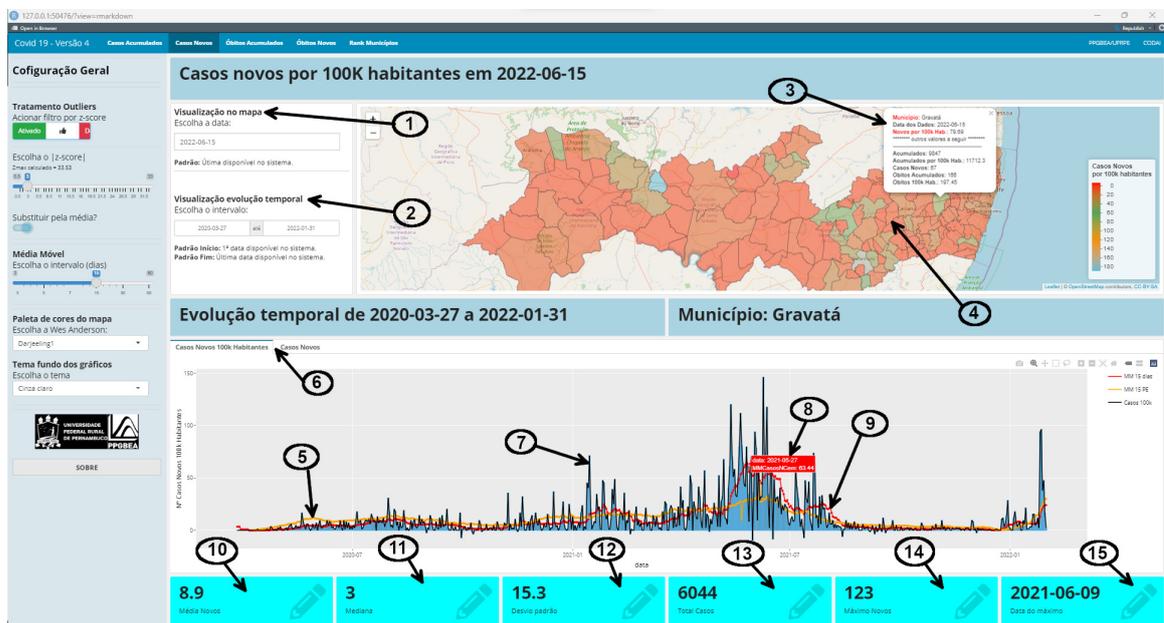
Item	Objeto		Descrição
	 interativo	 responsivo	
1		Visualização no mapa.	Apresenta um objeto do tipo <i>date input</i> , o qual permite escolher o dia (data) para a construção da distribuição espacial por cidade, do número de casos acumulados por 10^5 habitantes, plotado através da escala de cores no mapa de Pernambuco.
2		Visualização evolução temporal.	Figura um objeto do tipo <i>date range input</i> , que permite escolher o intervalo de datas para visualização dos gráficos. Ao se interagir com este controle, ocorrerá uma atualização dos gráficos, considerando-se os intervalo de datas escolhido.
3	 	Mapa.	Permite escolher o município para a visualização dos gráficos. Deve-se clicar com o botão esquerdo do <i>mouse</i> sobre o município desejado.
4		Sub-abas.	As sub-abas Casos Acumulados 100k Habitantes e Casos Acumulados permitem escolher entre a visualização para 10^5 habitantes ou real.
5	 	Gráfico para a cidade escolhida.	Gráfico que mostra a evolução do número de casos acumulados por 10^5 habitantes ou real, considerando-se a aba escolhida. É interativo e ao se passar o cursor do <i>mouse</i> sobre o mesmo é apresentado uma guia suspensa com resumo dos dados do ponto/dia sobre o qual se encontra o cursor.
6	 	Gráfico para Pernambuco.	Gráfico que mostra a evolução do número de casos acumulados por 10^5 habitantes de Pernambuco, sendo presente apenas na aba que corresponde a 10^5 habitantes.

Fim da Tabela 11

5.3.7 Aba Casos Novos

Na aba Casos Novos, apresentada na [Figura 31](#), há quatro áreas: A barra lateral Configuração Geral, a para o mapa de Pernambuco, a de plotagem dos gráficos e a área dos *dashboards* para a apresentação das métricas de sumarização. Na área para exibição dos gráficos há duas sub-abas: Casos Novos 100k Habitantes e Casos Novos. Na apresentação do número de casos novos diário por 100.000 habitantes, são traçadas três curvas: uma para os casos novos diário por 100.000 habitantes para a cidade escolhida (em preto), uma para a média móvel do número de casos novos diário por 100.000 habitantes para a cidade escolhida (em vermelho) e uma para expressar a média móvel do número de casos novos diário por 100.000 habitantes do Estado de Pernambuco (em laranja). Na aba de Casos Novos, são apresentados os valores reais do número de casos novos diário e sua média móvel, mas sem o gráfico para Pernambuco, já que não seria pertinente para se efetuar uma comparação.

Figura 31 – Principais elementos da aba Casos Novos do sistema. Os objetos numerados são descritos na [Tabela 12](#). A exibição corresponde ao dia 15/06/2022 no mapa de cores e a plotagem dos gráficos de linha, bem como das métricas nos *dashboards* são para a cidade de Gravata de 27/03/2020 a 31/01/2022.



Fonte: Próprio autor.

Já na área dos *dashboards*, tem-se a Média, Mediana, Desvio padrão, Total Casos, Máximo Novos e Data do máximo, descritos na [Tabela 12](#), juntamente com outros elementos presentes na [Figura 31](#).

Tabela 12 – Detalhamento dos elementos apresentados na Figura 31, para a aba Casos Novos. Usa-se o símbolo  para um objeto interativo e  para reativo.

Item	Objeto		Descrição
	 interativo	 responsivo	
1		Visualização no mapa.	Apresenta um objeto do tipo <i>date input</i> , o qual permite escolher o dia (data) para a construção da distribuição espacial por cidade, do número de casos novos por 10^5 habitantes, plotado através da escala de cores no mapa de Pernambuco.
2		Visualização evolução temporal.	Figura um objeto do tipo <i>date range input</i> , que permite escolher o intervalo de datas para visualização dos gráficos e para o cálculo das métricas apresentadas nos <i>dashboards</i> .
3		Guia suspensa – Mapa.	Ao se clicar em um município do mapa, surgirá a guia suspensa apresentando um resumo dos dados do município em questão.
4	 	Mapa.	Permite escolher o município para a visualizar dos gráficos e exibe o mapa de cores dos casos novos por 10^5 habitantes. Deve-se clicar com o botão esquerdo do <i>mouse</i> sobre a cidade desejado.
5	 	Gráfico média móvel dos casos novos – Pernambuco.	Gráfico traçado na cor laranja que mostra a evolução do número de casos acumulados por 10^5 habitantes ou real, considerando-se a aba escolhida. O gráfico é interativo.
6	 	As sub-abas.	As Sub-abas Casos Novos 100k Habitantes e Casos Novos permitem escolher entre a visualização para 10^5 habitantes ou real.
7	 	Gráfico casos novos – município.	Gráfico traçado na cor preta que mostra a evolução do número de casos acumulados por 10^5 habitantes para a cidade selecionada. O gráfico é interativo.

Continua na próxima página...

... Continuação da Tabela 12

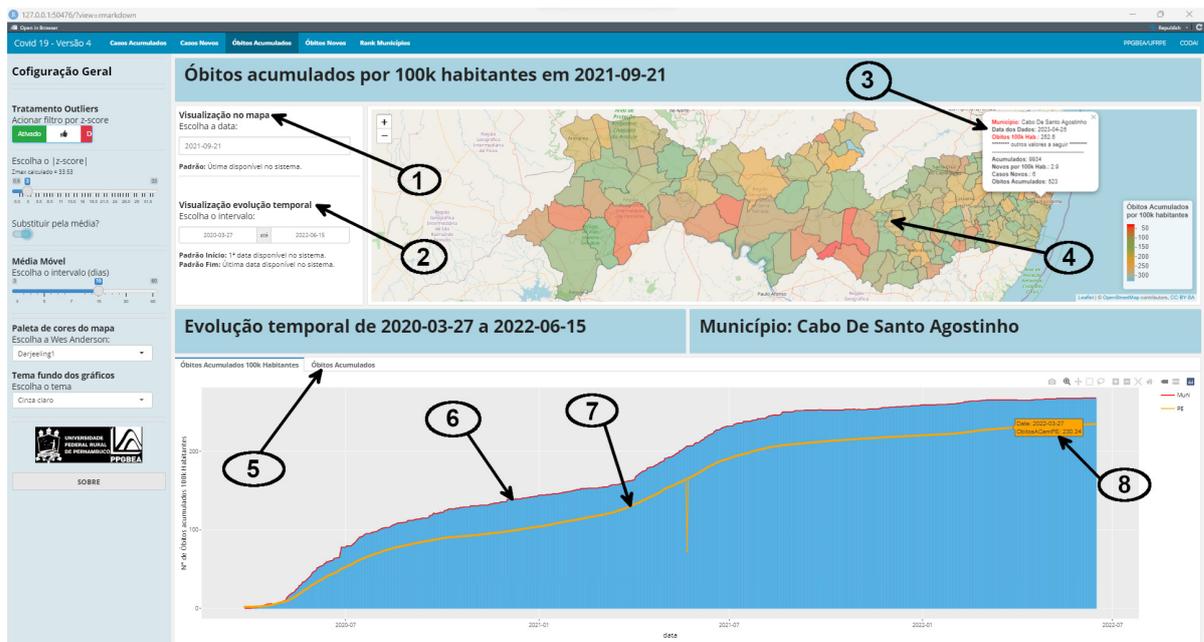
Item	Objeto		Descrição
	 interativo  responsivo		
8	 Guia suspensa Gráficos.	–	Os gráficos são interativos e ao se posicionar o cursor do <i>mouse</i> em um ponto do gráfico surgirá a guia suspensa com as informações dos dados associados ao ponto sobre o qual está o cursor.
9	  Gráfico média mó- vel dos casos novos – Município.	–	Gráfico traçado na cor vermelha que mostra a evolução da média móvel dos casos novos por 10^5 habitantes para município escolhido. O gráfico em questão é interativo.
10	 Média Novos.		Média do número de casos novos para o município escolhido, considerando-se o intervalo de datas anteposto.
11	 Mediana.		Mediana do número de casos novos para o município escolhido, considerando-se o intervalo de datas configurado.
12	 Desvio padrão.		Desvio padrão do número de casos novos para o município escolhido, considerando-se o intervalo de datas elencado.
13	 Total Casos.		Número de casos novos para o município escolhido, considerando-se o intervalo de datas designado.
14	 Máximo Novos.		Maior valor dentro o número de casos novos diários para o município escolhido, considerando-se o intervalo de datas configurado.
15	 Data do máximo.		Data em que ocorreu o máximo do número de casos novos para o município escolhido, considerando-se o intervalo de datas anteposto.

Fim da Tabela 12

5.3.8 Aba Óbitos Acumulados

A aba Óbitos Acumulados, apresentada na [Figura 32](#), é similar a aba Casos Acumulados e contém as mesmas duas áreas. Na área para apresentação dos gráficos, há duas sub-abas: Óbitos Acumulados 100k Habitantes e Óbitos Acumulados. Na apresentação do número de óbitos acumulados por 100.000 habitantes, são traçadas duas curvas: uma para o município escolhido (traçado em vermelho) e outra para o Estado e Pernambuco (traçado em laranja), permitindo a comparação da cidade selecionado como Estado.

Figura 32 – Principais elementos da aba Óbitos Acumulados do sistema. Principais elementos da aba Óbitos Acumulados. Os objetos numerados são descritos na [Tabela 13](#). A exibição corresponde ao dia 21/09/2021 para o mapa de cores e a plotagem dos gráficos de linha são para o município do Cabo de Santo Agostinho de 27/03/2020 a 15/06/2023.



Fonte: Próprio autor.

Obviamente, a interação em objetos da barra lateral Configuração Geral afeta as visualizações da aba Óbitos Acumulados, mas como foi descrito na [subseção 5.3.2](#), omite-se aqui. Para uma melhor compreensão dos elementos enumerados na figura anterior, considere a descrição apresentada de forma resumida na [Tabela 13](#). Percebe-se que são usados os símbolos  e , respectivamente, para destacar que o objeto em questão é interativo ou responsivo. Os objetos que figuram ambos símbolos, são, simultaneamente, interativo e responsivo.

Tabela 13 – Detalhamento dos elementos apresentados na Figura 32, para a aba Óbitos Acumulados. Usa-se o símbolo  para um objeto interativo e  para reativo.

Item	Objeto		Descrição
	 interativo	 responsivo	
1		Visualização no mapa.	Apresenta um objeto do tipo <i>date input</i> , o qual permite escolher o dia (data) para a construção da distribuição espacial por cidade, do número de óbitos acumulados por 10^5 habitantes, plotado através da escala de cores no mapa de Pernambuco.
2		Visualização evolução temporal.	Figura um objeto do tipo <i>date range input</i> , que permite escolher o intervalo de datas para visualização dos gráficos. Ao se interagir com este controle, ocorrerá uma atualização dos gráficos, considerando-se os intervalo de datas escolhido.
3		Guia suspensa – Mapa	Ao se clicar em um município do mapa, surgirá uma guia suspensa apresentando um resumo dos dados do município em questão.
4	 	Mapa.	Permite escolher o município para a visualizar dos gráficos. Deve-se clicar com o botão esquerdo do <i>mouse</i> sobre o município desejado.
5		Sub-abas.	Sub-abas Óbitos Acumulados 100k Habitantes e Óbitos Acumulados, permitem escolher entre a visualização para 10^5 habitantes ou real.
6	 	Gráfico para a cidade escolhida.	Gráfico que mostra a evolução do número de óbitos acumulados por 10^5 habitantes ou real, considerando-se a aba escolhida. É interativo e ao se passar o cursor do <i>mouse</i> sobre o mesmo é apresentado uma guia suspensa com resumo dos dados do ponto/dia sobre o qual se encontra o cursor.

Continua na próxima página...

... Continuação da Tabela 13

Item	Objeto	Descrição
	<p>👉 interativo</p> <p>🔄 responsivo</p>	
7	👉 🔄 Gráfico para Pernambuco.	Gráfico que mostra a evolução do número de óbitos acumulados por 10^5 habitantes de Pernambuco, sendo presente apenas na aba que corresponde a 10^5 habitantes.
8	🔄 Guia suspensa Gráficos	— Ao se posicionar o cursor do <i>mouse</i> em um ponto do gráfico surgirá a guia suspensa com as informações dos dados associados ao ponto sobre o qual o cursor se encontra. Esta funcionalidade está presente em todos os gráficos.

Fim da Tabela 13

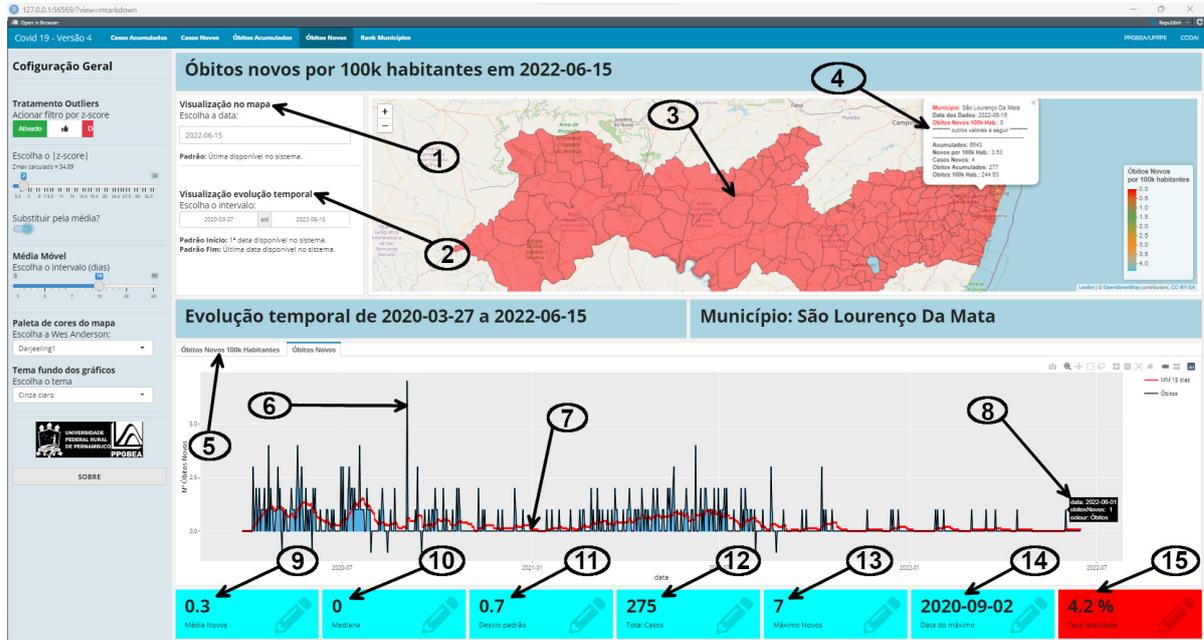
Fonte: Próprio autor

5.3.9 Aba Óbitos Novos

A aba Óbitos Novos, apresentada na [Figura 33](#), é similar a aba Casos Novos e apresenta quatro áreas: A barra lateral Configuração Geral, a do mapa de Pernambuco, a de plotagem dos gráficos e a área dos *dashboards* para apresentação das métricas. Na área para exibição dos gráficos há duas sub-abas: Óbitos Novos e Óbitos Novos 100k Habitantes. Na apresentação do número de óbitos novos diário, são traçadas duas curvas: uma para o número de óbitos novos diário para a cidade escolhida (em preto) e uma para a média móvel do número de óbitos diário para a cidade escolhida (em vermelho). Já na sub-aba Óbitos Novos 100k Habitantes estão presentes três curvas: uma para os óbitos novos diário por 100.000 habitantes para a cidade escolhida (em preto), uma para a média móvel do número de óbitos novos diário por 100.000 habitantes para a cidade escolhida (em vermelho) e uma para expressar a média móvel do número de óbitos novos diário por 100.000 habitantes do Estado de Pernambuco (em laranja).

Já na área dos *dashboards*, tem-se a Média, Mediana, Desvio padrão, Total Óbitos, Máximo Óbitos e Data do máximo, os quais são descritos na [Tabela 14](#) juntamente com outros elementos presentes na [Figura 33](#).

Figura 33 – Principais elementos da aba Óbitos Novos do sistema. Os objetos numerados são descritos na Tabela 14. A exibição corresponde ao dia 15/06/2022 no mapa de cores e a plotagem dos gráficos de linha, bem como das métricas nos *dashboards* são para a cidade de São Lourenço da Mata de 27/03/2020 a 15/06/2022.



Fonte: Próprio autor.

Tabela 14 – Detalhamento dos elementos apresentados na Figura 33, para a aba Óbitos Novos. Usa-se o símbolo  para um objeto interativo e  para reativo.

Item	Objeto  interativo  reativo	Descrição
1	 Visualização no mapa.	Apresenta um objeto do tipo <i>date input</i> , o qual permite escolher o dia (data) para a construção da distribuição espacial por cidade, do número de óbitos novos por 10^5 habitantes, plotado através da escala de cores no mapa de Pernambuco.
2	 Visualização evolução temporal.	Figura um objeto do tipo <i>date range input</i> , que permite escolher o intervalo de datas para visualização dos gráficos e para o cálculo das métricas apresentadas nos <i>dashboards</i> .

Continua na próxima página...

... Continuação da Tabela 14

Item	Objeto		Descrição
	 interativo	 responsivo	
3	 	Mapa.	Permite escolher o município para a visualizar dos gráficos e exibe o mapa de cores dos óbitos novos por 10^5 habitantes. Deve-se clicar com o botão esquerdo do <i>mouse</i> sobre a cidade desejado.
4		Guia suspensa – Mapa.	Ao se clicar em um município do mapa, surgirá a guia suspensa apresentando um resumo dos dados do município em questão.
5	 	Sub-abas.	As Sub-abas Óbitos Novos 100k Habitantes e Óbitos Novos, permitem escolher entre a visualização para 10^5 habitantes ou real.
6	 	Gráfico óbitos novos – município.	Gráfico traçado na cor preta que mostra a evolução do número de óbitos novos. O gráfico em questão é interativo.
7	 	Gráfico média móvel dos óbitos novos – Município	Gráfico traçado na cor vermelha que mostra a evolução do número de óbitos novos por 10^5 habitantes para município escolhido. O gráfico em questão é interativo.
8		Guia suspensa – Gráficos.	Ao se posicionar o cursor do <i>mouse</i> em um ponto do gráfico surgirá a guia suspensa com as informações dos dados associados ao ponto sobre o qual o cursor se encontra.
9		Média Novos	Média do número de óbitos novos para o município escolhido, considerando-se o intervalo de datas anteposto.
10		Mediana	Mediana do número de óbitos novos para o município escolhido, considerando-se o intervalo de datas configurado.

Continua na próxima página...

... Continuação da Tabela 14

Item	Objeto		Descrição
	👉 interativo	🔄 responsivo	
11	🔄	Desvio padrão	Desvio padrão do número de óbitos novos para o município escolhido, considerando-se o intervalo de datas selecionado.
12	🔄	Total Casos	Número de óbitos novos para o município escolhido, considerando-se o intervalo de datas anteposto.
13	🔄	Máximo Novos	Maior valor dentro o número de óbitos novos diários para o município escolhido, considerando-se o intervalo de datas configurado.
14	🔄	Data do máximo	Data em que ocorreu o máximo do número de óbitos novos para o município escolhido, considerando-se o intervalo de datas selecionado.
15	🔄	Taxa letalidade	Indica o percentual do total de óbitos em relação o número de casos para o município escolhido, considerando-se o intervalo de datas configurado.

Fim da Tabela 14

Fonte: Próprio autor

5.3.10 Aba Rank Municípios

Na aba Rank Municípios, apresentada na [Figura 34](#), são exibidas quatro áreas: A barra lateral Configuração Geral, a área de plotagem dos gráficos (6), a área para configuração do intervalo de visualização (1) e a área para efetuar a configuração do intervalo do número de municípios que se deseja fazer o ranqueamento (3). Na área para exibição dos gráficos há duas sub-abas (5): Médias de casos novos 100k Habitantes e Média de óbitos novos 100k Habitantes, onde, na primeira o ranqueamento se dar pela média do número de casos novos diários por 10^5 habitantes e no segundo pela média do número de óbitos novos diários por 10^5 habitantes.

Figura 34 – Principais elementos da aba Rank Municípios do sistema. Os objetos numerados são descritos na Tabela 15. A exibição corresponde ao intervalo de datas de 27/03/2020 a 03/06/2022 e foi selecionado a visualização do 1º ao 15º município com maior média de casos novos por 10⁵ habitantes. Aba de classificação em ordem crescente da média dos casos e óbitos novos por 100.000 habitantes.



Fonte: Próprio autor.

Evidentemente, a interação em objetos da barra lateral Configuração Geral afeta as visualizações da aba Rank Municípios no tocante aos cálculos das médias, já que a supressão ou substituição de *outliers* poderá afetar o cálculo das médias. Para uma melhor compreensão dos elementos enumerados na figura anterior, considere a descrição apresentada de forma resumida na Tabela 15.

Tabela 15 – Detalhamento dos elementos apresentados na Figura 34, para a aba Rank Municípios. Usa-se o símbolo  para um objeto interativo e  para reativo.

Item	Objeto	
	 interativo	Descrição
1	 Intervalo de visualização.	Apresenta um objeto do tipo <i>date range input</i> , que permite escolher o intervalo de datas para visualização dos gráficos e para o ranqueamento.
2	 Período de observação.	Exibi o intervalo de datas selecionado pelo usuário.

Continua na próxima página...

... Continuação da Tabela 15

Item	Objeto		Descrição
	👉 interativo	🔄 responsivo	
3	👉	Intervalo para o rank.	Figura um objeto do tipo <i>slider input</i> que permitir escolher o intervalo de ranqueamento das cidades para plotagem na área do gráfico de barras.
4	🔄	Rank escolhido.	Exibe o início e fim de ranqueamento escolhido no controle descrito anteriormente.
5	👉	sub-abas.	Sub-abas para selecionar se visualizará o ranqueamento é através do número de casos novos ou óbitos novos por 10^5 habitantes.
6	👉	🔄 Área do gráfico.	Área para exibição do gráfico de barras ranqueado em ordem decrescente da média, para os casos novos por 10^5 habitantes ou óbitos novos por 10^5 habitantes, conforme a sub-aba selecionada.

Fim da Tabela 15

Fonte: Próprio autor

5.3.11 Discussão

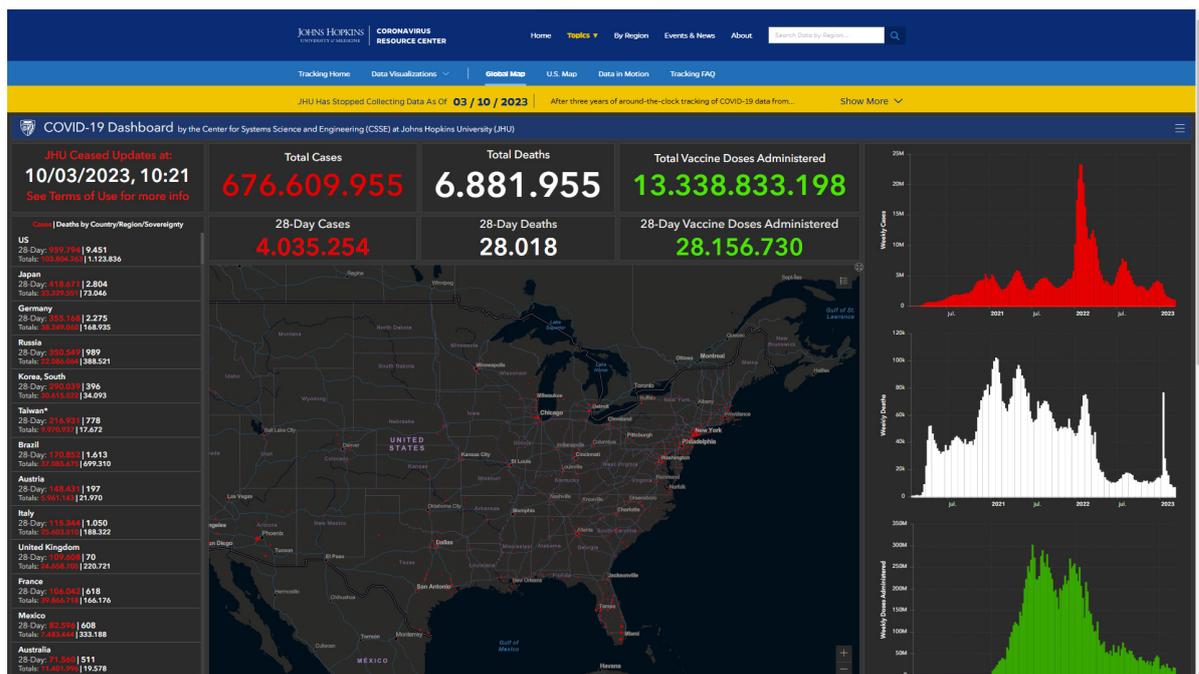
A pandemia de *Sars-Cov-2* e os dados associados, dada a sua gravidade, dentre outros fatores, desperta sobremaneira o interesse de pesquisadores em todo mundo, sejam da área de saúde ou das ciências de dados. Deste modo, percebe-se que ocorreram o desenvolvimento de diversos sistemas, inclusive *web*, para a comunicação/visualização dos dados da Covid-19. Tais sistemas são cruciais para acompanhar a evolução da pandemia, bem como nas tomadas de decisões associadas as diversas ações necessárias para a sua mitigação. Evidentemente que o exposto se configura como evidência corroboradora da relevância deste trabalho ao trazer as perspectivas do desenvolvimento de sistemas interativos-responsivos, enfocando as etapas de governança de dados já mencionadas, através de ferramentas computacionais de baixo custo. Neste sentido, apresenta-se a seguir alguns sistemas com perspectivas análogas, preiteando-se ter parâmetro comparador para o sistema da Covid-19 apresentado neste trabalho.

Sob esse contexto se considera um dos primeiros sistemas mundiais para acompanhar

os dados da Covid-19, lançado um dia após os EUA relatarem seu primeiro caso de Covid-19, segundo [Dong et al. \(2022\)](#), desenvolvido pela Johns Hopkins University Center for Systems Science and Engineering (Centro de Ciências e Engenharia de Sistemas da Universidade Johns Hopkins) ([HOPKINS, 2020](#)) e cuja tela inicial pode ser observada na [Figura 35](#). [Dong et al. \(2022\)](#) expõe que o sistema em questão foi construído por meio do *ArcGIS Dashboard* ([ESRI, 2023a](#)), desenvolvido pela empresa [Esri \(2023b\)](#) e considerada líder no mercado global em softwares de sistemas de informações geográficas (GIS). Relatam também ser necessário de oito a dez profissionais dedicados em tempo integral para manutenção do sistema, dentre cientistas de dados e engenheiros de software, além de especialistas em saúde pública e sistemas de informações geográficas.

Figura 35 – Tela inicial do sistema desenvolvido pela Johns Hopkins University Center for Systems Science and Engineering para acompanhar a pandemia de *Sars-Cov-2*. Acessado em 22/08/2023.

Link de acesso: [https:// coronavirus.jhu.edu/map.html](https://coronavirus.jhu.edu/map.html)



Fonte: ([HOPKINS, 2020](#)).

Navegando-se pelo sistema do [Hopkins \(2020\)](#), pode-se observar interatividade-responsividade ao selecionar na barra lateral esquerda o país que se deseja visualizar os detalhes associados aos dados, onde se tem o total de casos, total de vacinas aplicadas, dentre outras quantidades de interesse. Ao se clicar em U.S Map, tem-se as informações detalhadas dos Estados dos EUA, de modo que a navegação nessa aba proporciona informações mais detalhadas, bem como níveis de interatividade-responsividade bem maiores.

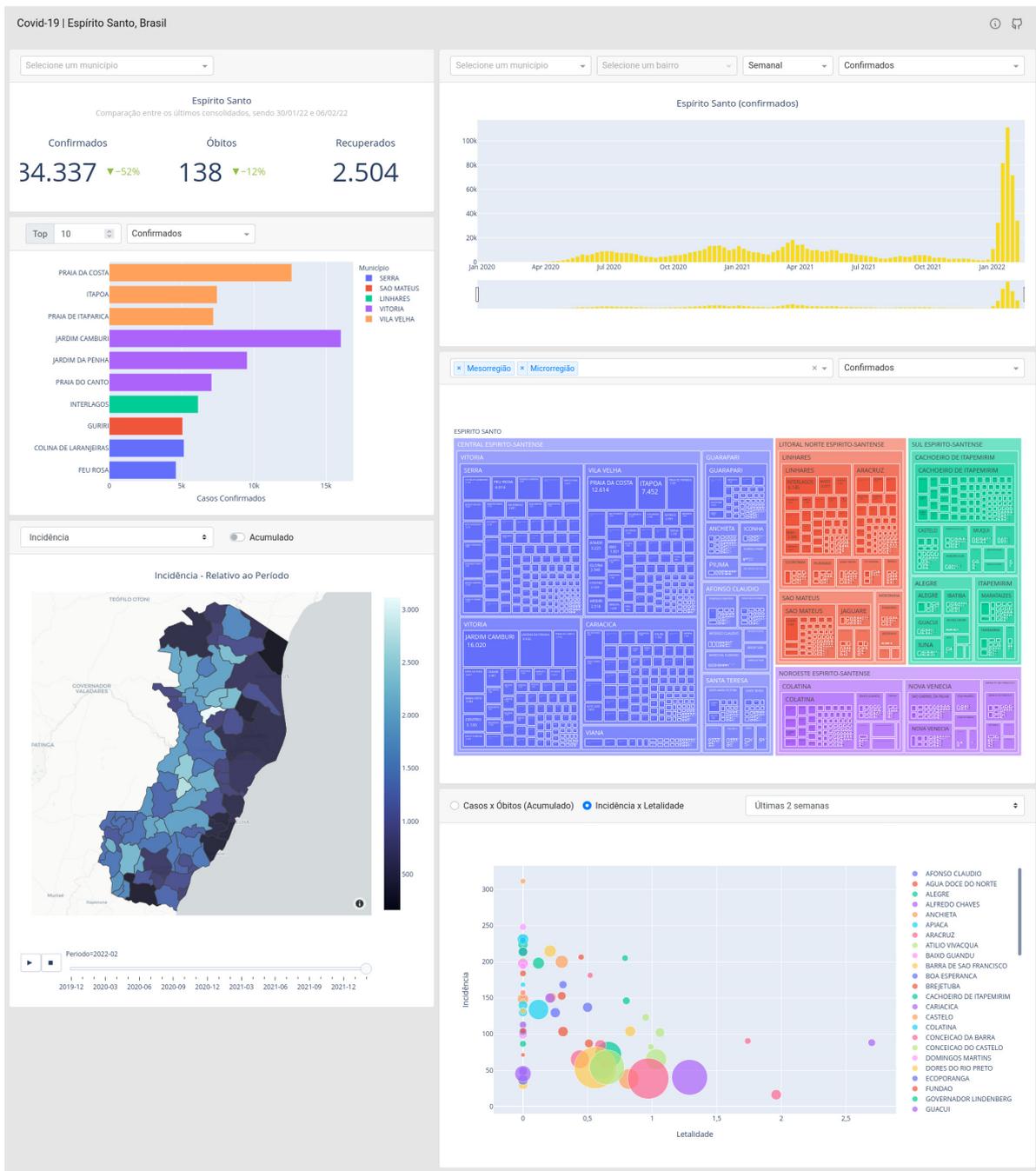
Analisando a descrição do sistema em tela através dos relatos de [Dong et al. \(2022\)](#), pode-se discorrer uma breve comparação com os resultados apresentados neste trabalho. Inicialmente constata-se a prática da existência de pré-processamento/tratamento dos dados e antes de serem acessados pelo sistema do [Hopkins \(2020\)](#) ocorre uma série de correções, como por exemplo, associadas a dados faltantes. Em oposição, o sistema apresentado neste trabalho não considera tal preocupação, mas, apenas, a união dos dados que são disponibilizados pelo MS-BR em arquivos distintos. Todavia, como já exposto, percebe-se a existência de tratamento pelo próprio sistema em si e através da interatividade com o usuário, como por exemplo, o filtro de *outliers* através do *Z-score* e uma outra série de filtros possíveis de aplicar, bem como quantidades descritivas, sempre através da escolha interativa do usuário. Portanto, com relação a nível de interatividade-responsividade, o sistema desenvolvido neste trabalho apresenta um nível significativo de recursos, já o sistema do [Hopkins \(2020\)](#) tem uma maior e mais eficiente construção dinâmica quando se trata da aba que reporta as informações dos EUA. Um ponto crucial a ser comparado diz respeito à equipe de desenvolvimento e manutenção dos sistemas, pois enquanto o sistema deste trabalho conta com uma equipe de poucos profissionais (máximo três), o da [Hopkins \(2020\)](#) requer uma equipe robusta de profissionais. Como ponto final de comparação, considera-se o custo. Embora [Dong et al. \(2022\)](#) não revele valores, percebe-se, como já exposto, que as ferramentas de desenvolvimento não são gratuitas, já o sistema apresentado neste trabalho foi desenvolvido através de ferramentas computacionais sem custos, inclusive associado à hospedagem e acesso na *web*.

Na sequência, considera-se um segundo sistema desenvolvido para monitoramento da Covid-19 no Estado do Espírito Santo utilizando software livre. O sistema em questão foi desenvolvido por [Oliveira \(2022\)](#), onde em seu trabalho é possível constatar a escolha da linguagem de programação Python ([FOUNDATION, P.S., 2000–present](#)) combinada com APIs disponibilizadas pelo IBGE ([IBGE, 1934–present](#)) ([APIS-IBGE, 1934–present](#)), as bibliotecas Plotly ([INC., 2012–present](#)) e framework Dash ([INC., 2017–present](#)) para a exibição de *Dashboards*. Quanto aos dados, [Oliveira \(2022\)](#) revela que são obtidos através do *web site* intitulado “Painel Covid-19 – Estado do Espírito Santo” ([PRODEST, 2023](#)) no formato .csv e compactados no formato .zip, os quais apresentam atualização diária. Então, utiliza-se da biblioteca Pandas ([TEAM, 2008–present](#)) para um processo ETL (Extract–Extração, Transform–Transformação, Load–Carga) dos dados, onde nos detalhes expostos sobre a etapa de transformação é clara a existência de um pré-processamento que considera desde a incompletude/inconsistência dos dados a criação de novas variáveis de interesse, diga-se que análoga a estratégia utilizada pelo [Hopkins \(2020\)](#) e já fica aqui uma comparação análogo em relação ao sistema desenvolvido neste trabalho. Já em relação à etapa da carga, os dados são carregados (armazenados) em um banco de dados MongoDB

através da biblioteca pymongo (MongoDB, Inc., 2009–present). Não encontra-se indicação de um *link web* para navegação, mas pode ser observado o resultado do sistema através da Figura 36, além disso há indicação da disponibilização no GitHub para sua instalação local com orientações detalhadas.

Figura 36 – Tela inicial do sistema desenvolvido Oliveira (2022) expondo dashboards para os dados da Covid-19 do Estado do Espírito Santo.

Link de acesso: GitHub: https://github.com/DaniloSI/covid-19_es



Fonte: (OLIVEIRA, 2022).

Um ponto importante indicado pelo Oliveira (2022) é a utilização da plataforma Heroku (SALESFORCE, 2007–present) para disponibilizar os *Dashboards* na *web*, sendo suficiente uma licença gratuita no início do desenvolvimento, mas com o passar do tempo, uma quantidade de processamento de dados grande passou a exigir a utilização de serviços pagos, tal como o Google Cloud Platform (Google LLC, 2008–present), sendo um diferencial relevante do sistema apresentado neste trabalho por ser hospedado e acessado via *web* sem qualquer custo. Uma última diferenciação apresentada em relação ao sistema desenvolvido neste trabalho, é que Oliveira (2022) indica orientações concisas e claras para uma instalação do Git, Python 2.8.5 e Pip para acessar o seu sistema, como pode ser verificado no GitHub publicado (https://github.com/DaniSI/covid-19_es), exigindo, portanto, do usuário um substancial conhecimento computacional. Em oposição, o sistema apresentado neste trabalho não exige qualquer conhecimento computacional mais significativo que a utilização de um navegador *web*.

Poder-se-ia manter a comparação com outros sistemas desenvolvidos para a análise e comunicação/visualização dos dados da Covid-19, inclusive o próprio “Painel Covid-19 – Estado do Espírito Santo” (PRODEST, 2023), desenvolvido por meio do PowerBI (CORPORATION, 2003–present) da Microsoft e não gratuito. Todavia, o maior objetivo deste trabalho é elucidar a grande relevância de sistemas interativos-responsivos para as etapas de governança de dados elencadas neste trabalho com desenvolvimento através de ferramentas gratuitas. Sendo os dados da Covid-19 um exemplo que trouxe a urgência evidente dos sistemas interativo-responsivos como ferramenta crucial na tomada de decisão eficiente. Por fim, há ainda o olhar do desenvolvedor, já que um sistema desenvolvido com a mesma finalidade poderá ter um conjunto de ferramentas e forma de acesso totalmente diferentes a esmo das escolhas da equipe de desenvolvimento.

5.4 Conclusões e considerações finais

O sistema apresentado neste capítulo para desbravar os dados da Covid-19 disponibilizados pelo Ministério da Saúde do Brasil e desenvolvido com a linguagem de programação R e os *packages* Shiny, R Markdown e Flexdashboard, mostrou-se eficiente para a comunicação/visualização dos dados em tela através de diversas ferramentas úteis e facilitadoras. Constituiu-se um instrumento de grande valia para entender os dados, bem como a própria pandemia em si, mesmo os dados apresentando alguns aspectos estranhos, tal como os *outliers* flagrantes, expostos na Figura 23. Todavia, a ferramenta de tratamento implementada através do z – *Score* permitiu uma visualização clara das informações.

A marca principal para o sistema é a interatividade–responsividade, a qual permitiu

em poucas abas, explorar uma massa de dados robusta com simplicidade e eficiência, já que o usuário pôde escolher quais informações iam ser exibidas e de que forma. A personalização do intervalo de dias para o cálculo das médias móveis e suas consequentes representações gráficas permitem a suavização da curva conforme a concepção de necessidade para cada analisador, bem como a personalização no nível de tratamento dos *outliers*, via seleção do valor absoluto do *z-Score*, acompanhado de sua substituição por uma média ou simples supressão. Assim, a quantidade de informações que podem ser apresentadas no aplicativo, quando é considerada a interatividade–responsividade, deixou evidente que sistemas desenvolvidos com as características em tela, constituem uma ferramenta singular na comunicação/visualização de dados, agregando em um mesmo objeto de comunicação inúmeras possibilidades de escolha para visualização, proporcionando agilidade em uma possível tomada de decisão. Caso fosse em um método não interativo–responsivo, certamente levaria, no mínimo, a centenas de páginas de um relatório, proporcionando uma busca laboriosa por qualquer informação desejada.

O acesso ao aplicativo na *web*, por meio do *site* appsmarcelo.shinyapps.io/Covid19_V3 Prin (SILVA, 2022), demonstra que os sistemas desenvolvidos com esta perspectiva podem levar à democratização do acesso aos dados, com baixo custo e de forma simples. A hospedagem do aplicativo é através do shinyapps.io (RSTUDIO, 2022) do RStudio em uma inscrição gratuita (*FREE*), levando a uma limitação no desempenho do aplicativo, tal como, a restrição para um máximo da memória RAM de processamento em 1 Giga Byte. Não obstante, mesmo considerando as limitações, pode-se perceber o quanto um sistema desenvolvido na linguagem de programação R, com muitas ferramentas de interatividade–responsividade, pode ser eficiente e relevante. Corroborando, portanto, que o desenvolvimento de aplicativos desta natureza são ferramentas indispensáveis no “mar” de dados da atualidade, quanto o R é uma linguagem de programação suficientemente robusta para ser empregada nesta tarefa. Evidentemente, as limitações impostas pela hospedagem gratuita pode ser facilmente resolvida mediante a contratação de um plano anual ao custo de \$ 440,00 (quatrocentos e quarenta dólares), viável no caso de se ter a pretensão de implementar diversos sistemas, já que o plano concede suporte para uma quantidade ilimitada de aplicativos, com 500 horas de atividade e aumento de memória RAM para processamento. Há outras soluções mais trabalhosas, tal como a instalação de um servidor Shiny em uma máquina virtual no AWS Virtual Cloud Servers.

Considerando-se sistemas análogos abordando dados da Covid-19, primeiramente é crucial ter em mente que muito embora haja zelo no que diz respeito a natureza epidemiológica dos dados, o objetivo central deste trabalho é voltado para a relevância dos sistemas interativos-responsivos nas etapas de governança de dados já mencionadas e através de ferramentas computacionais de baixo custo. Neste sentido, dois pontos relevantes

pode ser considerados sobre os sistemas utilizados como parâmetro na [subseção 5.3.11](#), o primeiro é que em alguma etapa do desenvolvimento recorrem a ferramentas computacionais não gratuitas e o segundo é a ocorrência de uma segmentação no desenvolvimento desde o pré-processamento dos dados até a disponibilização do sistema. O software do presente trabalho, através do R e os *packages* pertinentes, apenas de alimentou de ferramentas computacionais gratuitas e pode ser evidenciado uma integração maior nas etapas de desenvolvimento, trazendo, provavelmente, mais conforto para o desenvolvedor.

Um passo futuro, considerando-se a flexibilidade do sistema de permitir adaptações e atualizações, será a implementação para outras áreas geográficas, inclusive para o Brasil, onde seja possível escolher o Estado e em seguida as cidades, como já está atualmente, além da criação de interfaces específicas para dispositivos móveis, já que constitui-se uma limitação atual o fato de as interfaces serem eficientes para visualização em telas de computador. Desta forma, espera-se que a ferramenta passe a ser mais completa, tanto no que diz respeito ao acesso de informações, quanto na disponibilidade satisfatória abarcando dispositivos mobile, tal como os *smartphones*.

6 Conclusão geral

Os sistemas/aplicativos *web* desenvolvidos com a linguagem de programação R e os *packages* Shiny, Flexdashboard e R Markdown conseguiram resultados satisfatórios e seguiram a linha pretendida de sistemas interativos–responsivos aplicados à comunicação/visualização de dados. O desenvolvimento desses sistemas teve êxito em mostrar a eficiência da linguagem de programação R para o desenvolvimento de sistemas deste tipo, bem como, demonstrou a importância desses para as etapas de governança de dados, abrangendo a importação, transformação, filtragem, modelagem e, principalmente, a comunicação/visualização.

O primeiro sistema apresentado, *Select CODAI*, voltado para o processo de classificação do Colégio Dom Agostinho Ikas da UFRPE, em conformidade com o sistema de cotas estabelecido pela Lei 12.711 de 29 de agosto de 2012, permitiu o acompanhamento em tempo real do progresso das inscrições e, ao término dessas, com simplicidade permitiu o *download* do relatório em PDF com a classificação dos candidatos. Essa aplicação prática comprovou a eficácia de sistemas interativos–responsivos no tratamento das diversas etapas de governança de dados focadas neste trabalho, desde o acesso aos dados até a visualização/comunicação, neste caso, de forma interativa e estática (relatórios em PDF), bem como, confirmou ser a linguagem de programação R uma ferramenta eficiente para o desenvolvimento desses modelos de *softwares*.

O segundo sistema enfocou a visualização/comunicação dos dados oficiais disponibilizados pelo Ministério da Saúde do Brasil da Covid-19 para o Estado de Pernambuco. Com recursos interativo–responsivo, mantendo a perspectiva dos sistemas propostos e diversas ferramentas/métricas de análise dos dados, constituiu mais uma aplicação prática que comprovou a eficácia de sistemas com tais características, no trato das diversas etapas de governança de dados, como corroborou ser a linguagem de programação R, uma ferramenta crucial no desenvolvimento de sistemas interativos–responsivos, inclusive voltados à publicação na *web*.

Por fim, considerando-se a flexibilidade dos sistemas de permitir atualizações, há projetos para a melhoria de ambos, tornando-os mais robustos e atendendo uma gama maior de solução para os suas respectivas finalidades. Todavia, o mais relevante neste trabalho é a demonstração clara de que sistemas interativos–responsivos é uma forma eficaz de lidar com as etapas de governança de dados consideradas e que a linguagem de

programação R é uma ferramenta eficiente, robusta e de baixo custo para o desenvolvimento dos sistemas em questão.

Referências Bibliográficas

- Alcoforado, Luciane Ferreira; LEVY, A. **Visualização de dados com o software R**. 1ª. edição. ed. Niterói-RJ: 2017, 2017. 125 p. ISBN 9788592293208.
- ALLAIRE, J. et al. **markdown: Render Markdown with the C Library 'Sundown'**. [S.l.], 2019. R package version 1.1. Disponível em: <<https://CRAN.R-project.org/package=markdown>>.
- APIS-IBGE, I. B. de Geografia e E. **Serviço de dados e api do IBGE**. 1934–present. <<https://servicodados.ibge.gov.br/api/docs/>>. Serviços de dados e api do IBGE.
- AUCKLAND, U. of. **University of Auckland: Ua – new zealand**. 2022. Accessed = 2022-03-10. Disponível em: <<https://www.auckland.ac.nz/en/science/about-the-faculty/department-of-statistics/ihaka-lecture-series.html>>.
- BAUMER, B. et al. R markdown: Integrating a reproducible analysis tool into introductory statistics. **arXiv preprint arXiv:1402.1894**, 2014.
- BRASIL. Lei nº 12.711, de 29 de agosto de 2012. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 2012. ISSN 1677-7042. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12711.htm>.
- BRUCE, A.; BRUCE, P. **Estatística Prática para Cientistas de Dados**. [S.l.]: Alta Books, 2019.
- BRYAN, J. **googlesheets4: Access Google Sheets using the Sheets API V4**. [S.l.], 2021. R package version 0.3.0. Disponível em: <<https://CRAN.R-project.org/package=googlesheets4>>.
- CALENGE, C. The package “adehabitat” for the r software: a tool for the analysis of space and habitat use by animals. **Ecological modelling**, Elsevier, v. 197, n. 3-4, p. 516–519, 2006.
- CHENG, J.; KARAMBELKAR, B.; XIE, Y. **leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library**. [S.l.], 2021. R package version 2.0.4.1. Disponível em: <<https://CRAN.R-project.org/package=leaflet>>.
- CORPORATION, M. **Microsoft Power BI**. 2003–present. Website oficial do Microsoft Power BI. Disponível em: <<https://powerbi.microsoft.com/>>.
- CPV 2022.1, CODAI, U. **Edital processo seletivo CODAI 2022.1**. São Lournenço da Mata-PE: [s.n.], 2021. 1–17 p. Disponível em: <<https://www.ufrpe.br/br/content/col{é}gio-da-ufrpe-codai-oferta-312-vagas-para-ensino-m{é}dio-e-cur>>.
- CPV 2022.2, CODAI, U. **Edital processo seletivo CODAI 2022.2**. São Lournenço da Mata-PE: [s.n.], 2022. 1–22 p.

DONG, E. et al. The Johns Hopkins University Center for Systems Science and Engineering COVID-19 dashboard: data collection process, challenges faced, and lessons learned. **The Lancet Infectious Diseases**, Elsevier, v. 22, n. 12, p. e370–e376, 2022.

ECLIPSE, F. **IDE Eclipse**: Eclipse. 2022. Accessed = 2022-03-13. Disponível em: <<https://www.eclipse.org/>>.

ESRI. **ArcGIS Dashboards**. 2023. <<https://www.esri.com/en-us/arcgis/products/arcgis-dashboards/overview#easy-to-understand>>. (Accessed on 17/08/2023).

ESRI. **Esri**. 2023. <<https://www.esri.com/en-us/home>>. (Accessed on 17/08/2023).

ESTATÍSTICA, I. B. de Geografia e. **IBGE**: Ibge cidades do Brasil. 2022. Accessed = 2022-02-24. Disponível em: <<https://cidades.ibge.gov.br/>>.

FÁVERO, L. P.; BELFIORE, P. **Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®**. [S.l.]: Elsevier Brasil, 2017.

FILHO, P. F. de O. **Epidemiologia e Bioestatística—Fundamentos para a Leitura Crítica**. 2ª. edição. ed. [S.l.]: Editora Rubio, 2022.

FOUNDATION, P. S. **Python**. P.S., 2000–present. <<https://www.python.org/>>. A linguagem de programação Python é amplamente utilizada para análise de dados, desenvolvimento de software e muito mais.

FOUNDATION, T. D. **Home | LibreOffice - A melhor suite office livre**. 2010. <<https://pt-br.libreoffice.org/>>. (Accessed on 06/09/2023).

GITHUB. **gitHub**: github. 2022. Accessed = 2022-03-13. Disponível em: <<https://github.com/>>.

GOOGLE. **Google Analytics Solutions | Data Studio**. 2022. <<https://datastudio.withgoogle.com/>>. (Accessed on 04/14/2023).

Google LLC. **Google Cloud Platform**. 2008–present. <<https://cloud.google.com/>>. A Google Cloud Platform é uma plataforma de computação em nuvem que oferece serviços de infraestrutura, análise de dados, machine learning e muito mais.

GROLEMUND, G.; WICKHAM, H. Dates and times made easy with lubridate. **Journal of Statistical Software**, v. 40, n. 3, p. 1–25, 2011. Disponível em: <<https://www.jstatsoft.org/v40/i03/>>.

HAYMOND, S. Create laboratory business intelligence dashboards for free using R: A tutorial using the flexdashboard package. **Journal of Mass Spectrometry and Advances in the Clinical Lab**, Elsevier, v. 23, p. 39–43, 2022.

HENNING, E. et al. Para além da computação estatística: O uso do ambiente R para o ensino de métodos numéricos. **RENOTE**, v. 14, n. 1, 2016.

HOPKINS, J. COVID-19 map—Johns Hopkins Coronavirus Resource Center. **Johns Hopkins Coronavirus Resource Center**, v. 1, 2020.

HORNİK, K. The comprehensive r archive network. **Wiley interdisciplinary reviews: Computational statistics**, Wiley Online Library, v. 4, n. 4, p. 394–398, 2012.

IANNONE, R.; ALLAIRE, J.; BORGES, B. **flexdashboard: R Markdown Format for Flexible Dashboards**. [S.l.], 2020. R package version 0.5.2. Disponível em: <<https://CRAN.R-project.org/package=flexdashboard>>.

IBGE, I. B. de Geografia e E. **Instituto Brasileiro de Geografia e Estatística (IBGE)**. 1934–present. <<https://www.ibge.gov.br/>>. O IBGE é o órgão responsável pela produção e divulgação de informações estatísticas, geográficas e cartográficas no Brasil.

IEEE. **IEEE Spectrum**: Ieee – usa. 2022. Accessed = 2022-03-10. Disponível em: <<https://spectrum.ieee.org/>>.

IGLEWICZ, B.; HOAGLIN, D. **How to Detect and Handle Outliers (The ASQC Basic Reference in Quality Control)**. [S.l.]: Milwaukee, WI: American Society for Quality, Statistics Division, 1993.

INC., P. T. **Plotly for Python**. 2012–present. <<https://plotly.com/python/>>. Plotly é uma biblioteca de visualização interativa para Python e outras linguagens de programação.

INC., P. T. **Dash - A Python web application framework for interactive web analytics**. 2017–present. <<https://dash.plotly.com/>>. Dash é um framework em Python para criar aplicativos web interativos para análise de dados e visualização.

INSIGHTFUL. **S-PLUS**: Autores. 2022. Accessed = 2022-08-17. Disponível em: <<http://www.insightful.com/products/splus/default.asp>>.

JIANG ZIREN E CAO, W. e. C. H. e. B. F. e. S. L. **RIMeta: Meta-Analysis for Reference Interval**. 2022. <<https://cers.shinyapps.io/RIMeta/>>. (Accessed on 04/08/2023).

JIANG ZIREN E CAO, W. e. C. H. e. B. F. e. S. L. Rimeta: Uma ferramenta r shiny para estimar o intervalo de referência a partir de uma meta-análise. **Métodos de Síntese de Pesquisa**, Wiley Online Library, 2023.

KALIYAPERUMAL, S. et al. Labeling methods for identifying outliers. **International Journal of Statistics and Systems**, v. 10, n. 2, p. 231–238, 2015.

KONRATH, A. C. et al. Desenvolvimento de aplicativos web com r e shiny: inovações no ensino de estatística. **Abakós**, v. 6, n. 2, p. 55–71, 2018.

KUNST, J. **highcharter: A Wrapper for the 'Highcharts' Library**. [S.l.], 2022. R package version 0.9.4. Disponível em: <<https://CRAN.R-project.org/package=highcharter>>.

LOGINOVA, J.; WOHLAND, P. How to create an interactive dashboard using r: the example of the queensland covid-19 tracker. **Australian Population Studies**, v. 4, n. 2, p. 39–47, 2020.

MACHADO, M. V. Análise da utilização da técnica de médias móveis no processo de tomada de decisão de investimentos no mercado acionário. 2018.

MCPHERSON, J.; ALLAIRE, J. **rsconnect: Deployment Interface for R Markdown Documents and Shiny Applications**. [S.l.], 2021. R package version 0.8.18. Disponível em: <<https://CRAN.R-project.org/package=rsconnect>>.

MICROSOFT. **Visual Studio**: Visual studio. 2022. (Accessed on 04/08/2023). Disponível em: <<https://code.visualstudio.com/>>.

MICROSOFT. **Visualização de Dados | Microsoft Power BI**. 2022. <<https://powerbi.microsoft.com/pt-br/>>. (Accessed on 04/14/2023).

MICROSOFT. **Microsoft Excel | Microsoft 365**. 2023. Software. (Accessed on 06/09/2023). Disponível em: <<https://www.microsoft.com/pt-br/microsoft-365/excel>>.

MICROSOFT. **R no Visual Studio Code**. 2023. <<https://code.visualstudio.com/docs/languages/r>>. (Accessed on 04/08/2023).

MongoDB, Inc. **PyMongo**. 2009–present. <<https://pymongo.readthedocs.io/>>. PyMongo é uma biblioteca Python que permite interagir com o MongoDB a partir de código Python.

OLIVEIRA, D. d. Desenvolvimento de um sistema para monitoramento da covid-19 no espírito santo utilizando software livre. Serra, 2022.

ORGANISATION, W. H. **Vision impairment and blindness**: Fact sheet n°282. 2022. Accessed = 2022-02-24. Disponível em: <<https://www.who.int/pt/home>>.

PANDOC.ORG. **Pandoc**. 2023. <<https://pandoc.org/>>. (Accessed on 04/09/2023).

PEREIRA, R. H. M.; GONCALVES, C. N. **geobr: Download Official Spatial Data Sets of Brazil**. [S.l.], 2022. R package version 1.6.5999. Disponível em: <<https://github.com/ipeaGIT/geobr>>.

PESLAK, A.; CONFORTI, M. Computer programming languages in 2020: What we use, who uses them, and how do they impact job satisfaction. **Issues in Information Systems**, v. 21, n. 2, 2020.

PLURALSIGHT. **JavaScript.com | Sobre**. 2016. <<https://www.javascript.com/about>>. (Accessed on 04/14/2023).

PRODEST, T. da I. **Painel da COVID-19 do Espírito Santo**. 2023. <<https://coronavirus.es.gov.br/painel-covid-19-es>>. Painel online com informações atualizadas sobre a situação da COVID-19 no estado do Espírito Santo.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>.

RAM, K.; WICKHAM, H. **wesanderson: A Wes Anderson Palette Generator**. [S.l.], 2018. R package version 0.3.6. Disponível em: <<https://CRAN.R-project.org/package=wesanderson>>.

RSTUDIO. **Shiny from RStudio**: Shiny. 2023. Accessed = 2023-04-06. Disponível em: <<https://shiny.rstudio.com/>>.

RSTUDIO, E. **Shinyapps.io**: Shinyapps.io. 2022. Accessed = 2022-03-13. Disponível em: <<https://www.shinyapps.io/>>.

RSTUDIO-FLEXDASHBOARD. **R Markdown Format for Flexible Dashboards – flexdashboard**: Flexdashboard. 2022. <<https://pkgs.rstudio.com/flexdashboard/>>. (Accessed on 04/12/2023).

RSTUDIO-RMARKDOWN. **R Markdown from RStudio**: R markdown. 2022. Accessed = 2023-04-07. Disponível em: <<https://rmarkdown.rstudio.com/>>.

RSTUDIO-RMARKDOWN. **R Markdown Formats**: Formats. 2023. <<https://rmarkdown.rstudio.com/formats.html>>. (Accessed on 04/09/2023).

SABATER, C. R. et al. Compelling new electrocardiographic markers for automatic diagnosis. **arXiv preprint arXiv:2112.12196**, 2021.

SALEEM, S.; ASLAM, M.; SHAUKAT, M. R. A review and empirical comparison of univariate outlier detection methods. **Pakistan Journal of Statistics**, v. 37, n. 4, 2021.

SALESFORCE. **Heroku**. 2007–present. <<https://www.heroku.com/>>. Heroku é uma plataforma em nuvem que oferece serviços de hospedagem e implantação de aplicativos da web.

SCHMULLER, J. **Projetos em R para leigos**. Rio de Janeiro, BR: "Alta Books.", 2018.

SCHOOL, H. M. **Harvard Medical School**: Harvard – boston/usa. 2022. Accessed = 2022-03-10. Disponível em: <<https://hms.harvard.edu/news/new-leader-computational-biomedicine>>.

SEO, S. **A review and comparison of methods for detecting outliers in univariate data sets**. Tese (Doutorado) — University of Pittsburgh, 2006.

SHIFA, I. H. L. **Painel Oficial COVID-19 do Paquistão | Hospital Internacional de Shifa**. 2022. <<https://www.shifa.com.pk/covid-19-pakistan/>>. (Accessed on 05/01/2023).

SHIFFLER, R. E. Maximum z scores and outliers. **The American Statistician**, Taylor & Francis, v. 42, n. 1, p. 79–80, 1988.

SIEVERT, C. **Interactive Web-Based Data Visualization with R, plotly, and shiny**. Chapman and Hall/CRC, 2020. ISBN 9781138331457. Disponível em: <<https://plotly-r.com>>.

SILVA, M. C. **Covid–19**: Autores. 2022. Accessed = 2022-02-24. Disponível em: <https://appsmarcelo.shinyapps.io/Covid19_V3_Prin/>.

SILVA MARCELO C. SILVA, R. S. C. M. Antonio Samuel Alves da. **Covid 19 - Versão 4**. 2022. Accessed = 2023-05-05. Disponível em: <https://appsmarcelo.shinyapps.io/Covid19_V3_Prin/>.

- SILVA, N. da; COOK, D.; LEE, E.-K. Interactive Graphics for Visually Diagnosing Forest Classifiers in R. v. 2, n. Breiman 1996, 2017. ISSN 2331-8422. Disponível em: <<http://arxiv.org/abs/1704.02502>>.
- SILVA, N. da; COOK, D.; LEE, E.-K. Interactive graphics for visually diagnosing forest classifiers in r. **Computational Statistics**, Springer, p. 1–21, 2023.
- SILVA RÔMULO S. C. MENEZES, M. C. S. A. S. A. **Processo seletivo Codai**: Autores. 2022. Accessed = 2022-02-24. Disponível em: <https://marcelosilvaufupe.shinyapps.io/App_Inscricoes_2022_2_Prin/>.
- SLATER, L. J. et al. Using r in hydrology: a review of recent developments and future directions. **Hydrology and Earth System Sciences**, Copernicus GmbH, v. 23, n. 7, p. 2939–2963, 2019.
- SUS, D. de Informática do. **CORONAVÍRUS – BRASIL**. 2020. Disponível em: <<https://covid.saude.gov.br/>>.
- TALAGALA, T. S.; SHASHIKALA, R. Interactive Dashboard to Monitor the COVID-19 Outbreak and Vaccine Administration. p. 1–17, 2022. Disponível em: <<http://arxiv.org/abs/2205.07286>>.
- TEAM, P. D. **Pandas Documentation**. 2008–present. <<https://pandas.pydata.org/>>. Website oficial da documentação do Pandas - Python Data Analysis Library.
- TIMBERLAKE. **Stata**: Autores. 2022. Accessed = 2022-02-24. Disponível em: <<https://www.stata-brasil.com/>>.
- VENABLES, W. N.; SMITH, D. M. An Introduction to R. 2022.
- WANG, H. et al. Visprodom: an interactive shiny/r application for displaying protein domains with transcriptional features. **BMC genomics**, Springer, v. 23, n. 1, p. 472, 2022.
- WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <<https://ggplot2.tidyverse.org>>.
- WICKHAM, H. et al. **dplyr: A Grammar of Data Manipulation**. [S.l.], 2021. R package version 1.0.6. Disponível em: <<https://CRAN.R-project.org/package=dplyr>>.
- WICKHAM, H.; GROLEMUND, G. **R para data science: importe, arrume, transforme, visualize, and modele dados**. Rio de Janeiro, BR: "Alta Books.", 2019.
- XIAO, Z.; LAM, H.-M. **GitHub - obenno/ShinySyn: A shiny application to visualize MCscan result**. 2022. <<https://github.com/obenno/shinysyn>>. (Accessed on 04/08/2023).
- XIAO, Z.; LAM, H.-M. ShinySyn: a shiny/r application for the interactive visualization and integration of macro-and micro-synteny data. **Bioinformatics**, Oxford Academic, v. 38, n. 18, p. 4406–4408, 2022.

XIE, Y. **knitr: A General-Purpose Package for Dynamic Report Generation in R**. [S.l.], 2022. R package version 1.39. Disponível em: <<https://yihui.org/knitr/>>.

XIE, Y.; ALLAIRE, J. J.; GROLEMUND, G. **R markdown: The definitive guide**. [S.l.]: Chapman and Hall/CRC, 2018.

XIE, Y.; CHENG, J.; TAN, X. **DT: A Wrapper of the JavaScript Library 'DataTables'**. [S.l.], 2021. R package version 0.18. Disponível em: <<https://CRAN.R-project.org/package=DT>>.

ZEILEIS, A.; GROTHENDIECK, G. zoo: S3 infrastructure for regular and irregular time series. **Journal of Statistical Software**, v. 14, n. 6, p. 1–27, 2005.

ZOHO, A. **COVID 19 Dashboard India | Coronavirus Dashboard (Live Update) - Zoho Analytics**. 2022. <<https://www.zoho.com/covid/india/>>. (Accessed on 05/01/2023).

Apêndices

APÊNDICE A – Certificado de registro do sistema Select Codai



REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA, COMÉRCIO E SERVIÇOS
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512023001379-0**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 15/11/2021, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: Codai selection system in R

Data de criação: 15/11/2021

Titular(es): UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

Autor(es): ANTONIO SAMUEL ALVES DA SILVA; ROMULO SIMOES CEZAR MENEZES; MARCELO CORREIA DA SILVA

Linguagem: R

Campo de aplicação: AD-01

Tipo de programa: AT-01; CD-01; FA-01

Algoritmo hash: SHA-256

Resumo digital hash: 1213CBD5260A779C8697F12ADC7E9462535ED8560D65F912F8BFE327DE827B62

Expedido em: 23/05/2023



Aprovado por:
Carlos Alexandre Fernandes Silva
Chefe da DIPTO

APÊNDICE B – Certificado de registro do sistema Covid-19 Pernambuco



REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA, COMÉRCIO E SERVIÇOS
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512023001824-4**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 05/11/2020, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: Covid-19 Pernambuco

Data de criação: 05/11/2020

Titular(es): UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

Autor(es): ANTONIO SAMUEL ALVES DA SILVA; ROMULO SIMOES CEZAR MENEZES; MARCELO CORREIA DA SILVA

Linguagem: R

Campo de aplicação: BL-04; IF-10

Tipo de programa: AP-03; CD-01; FA-01; TC-01

Algoritmo hash: SHA-256

Resumo digital hash: 70979789AAD06BF6015E5DD13300E235601D7CBB2A519F521AE5BAB5AAF8F2DF

Expedido em: 04/07/2023



Aprovado por:
Carlos Alexandre Fernandes Silva
Chefe da DIPTO