

**UNIVERSIDADE ESTADUAL “JÚLIO DE MESQUITA FILHO”
FACULDADE DE ARQUITETURA, ARTES, COMUNICAÇÃO E DESIGN
PROGRAMA DE PÓS-GRADUAÇÃO MESTRADO EM MÍDIA E TECNOLOGIA**

MARCELO JOSÉ DOS SANTOS

**UM DASHBOARD DE DADOS ABERTOS DE ÓBITOS
POR COVID-19 NO BRASIL**

**BAURU
2024**

MARCELO JOSÉ DOS SANTOS

**UM DASHBOARD DE DADOS ABERTOS DE ÓBITOS
POR COVID-19 NO BRASIL**

Relatório Técnico apresentado ao
Programa de Pós-graduação em Mídia e
Tecnologia (PPGMiT) – FAAC – UNESP –
Bauru para obtenção do título de Mestre
em Mídia e Tecnologia, sob a orientação
do Prof. Associado João Pedro Albino.

BAURU

2024

S237d	<p>Santos, Marcelo José dos</p> <p>Um Dashboard de Dados Abertos de Óbitos por COVID-19 no Brasil / Marcelo José dos Santos. -- Bauru, 2024</p> <p>89 p.</p> <p>Dissertação (mestrado profissional) - Universidade Estadual Paulista (Unesp), Faculdade de Arquitetura, Artes, Comunicação e Design, Bauru</p> <p>Orientador: João Pedro Albino</p> <p>1. Dashboard. 2. COVID-19. 3. Dados Abertos. 4. Linguagem de programação R. 5. Análise de Dados. I. Título.</p>
-------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

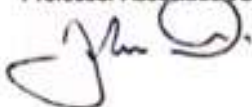
Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Arquitetura, Artes, Comunicação e Design, Bauru. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE Mestrado de MARCELO JOSE DOS SANTOS, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM MÍDIA E TECNOLOGIA, DA FACULDADE DE ARQUITETURA, ARTES, COMUNICAÇÃO E DESIGN - CÂMPUS DE BAURU.

Aos 25 dias do mês de março do ano de 2024, às 09:00 horas, no(a) Auditório da Seção Técnica de Pós-Graduação, realizou-se a defesa de DISSERTAÇÃO DE Mestrado de MARCELO JOSE DOS SANTOS, intitulada **Um Dashboard de Dados Abertos de Óbitos por COVID-19**. A Comissão Examinadora foi constituída pelos seguintes membros: Professor Associado JOÃO PEDRO ALBINO (Orientador(a) - Participação Presencial) do(a) Departamento de Computação / Faculdade de Ciências de Bauru, Professor Doutor MARCOS AMÉRICO (Participação Presencial) do(a) Departamento de Comunicação Social / Faculdade de Arquitetura, Artes, Comunicação e Design de Bauru, Professor Doutor ALAN CÉSAR BELO ANGELUCI (Participação Presencial) do(a) Escola de Comunicações e Artes / Universidade de São Paulo. Após a exposição pelo mestrando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final: APROVADO. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Professor Associado JOÃO PEDRO ALBINO



MARCELO JOSÉ DOS SANTOS

**UM DASHBOARD DE DADOS ABERTOS DE ÓBITOS
POR COVID-19 NO BRASIL**

Área de Concentração: **Ambientes Midiáticos e Tecnológicos**

Linha de Pesquisa: **Tecnologias Midiáticas**

Banca Examinadora:

Presidente/Orientador: Prof Associado João Pedro Albino

Instituição: Programa de Pós-graduação em Mídia e Tecnologia. Faculdade de Arquitetura, Artes, Comunicação e Design – FAAC/UNESP-Bauru

Prof. 1: Prof. Dr. Marcos Américo

Instituição: Programa de Pós-graduação em Mídia e Tecnologia. Faculdade de Arquitetura, Artes, Comunicação e Design – FAAC/UNESP-Bauru

Prof. 2: Prof. Dr. Alan César Belo Angeluci

Instituição: Universidade de São Paulo - USP/SP

Resultado: **Aprovado**

Bauru, 25 de Março de 2024.

AGRADECIMENTOS

À Deus, agradeço pelo dom da vida e por Sua constante presença, guiando-me em todas as etapas e ajudando-me a superar as dificuldades encontradas ao longo da realização deste trabalho.

Deixo um agradecimento especial aos meus pais, Oswaldo (*in memoriam*) e Aparecida pelos ensinamentos, exemplos de vida e por não medirem esforços para me proporcionar uma educação de qualidade.

À minha esposa e filhos, que são a base sólida na construção da minha vida e que são fundamentais ao longo da minha jornada. A razão dos meus objetivos são vocês!

Agradeço imensamente ao meu orientador Professor Associado João Pedro Albino, que fez o papel de facilitador. Seu conhecimento, entusiasmo e direcionamento transformaram obstáculos em oportunidades de aprendizado.

Aos colegas do grupo de pesquisa LInDa da Unesp de Bauru, expresso meus sinceros agradecimentos pelo companheirismo e cumplicidade demonstrados em nossas atividades. Ninguém constrói nada sozinho.

À Casa da Fraternidade Santa Rita de Cássia de Bauru-SP, sou eternamente grato pela acolhida e pela oportunidade proporcionada como professor voluntário. Encontrei um ambiente de colaboração e incentivo à educação que contribuíram para minha formação.

Ao Programa de Pós-Graduação em Mídia e Tecnologia da Faculdade de Arquitetura Artes, Comunicação e Design – FAAC Unesp Bauru-SP, principalmente ao seu corpo docente e direção que oportunizaram e contribuíram para esta etapa da minha formação acadêmica.

Por fim, nada disso seria possível sem a intercessão de Nossa Senhora iluminando meus caminhos, e sem as bênçãos dos Santos aos quais sou devoto.

“Amemos a DEUS, meus irmãos, amemos a DEUS, mas que isto seja a custa dos nossos braços, que isto seja com o suor dos nossos rostos”

São Vicente de Paulo

SANTOS, M. J. **UM DASHBOARD DE DADOS ABERTOS DE ÓBITOS POR COVID-19 NO BRASIL**, 2024, 89 f. Relatório Técnico-Científico (Mestrado em Mídia e Tecnologia) - FAAC - UNESP, sob a orientação do Prof. Associado João Pedro Albino, Bauru, 2024.

RESUMO

O avanço tecnológico atual possibilita o processamento e transformação de grandes quantidades de dados em informações precisas e relevantes em todas as áreas do conhecimento. Durante a pandemia de COVID-19, essa proliferação de dados criou um ambiente propício para a realização de estudos e a aplicação de métodos científicos, como a análise exploratória de dados, para demonstrar e gerar resultados, fornecendo apoio essencial nas tomadas de decisões no combate a esse fenômeno sanitário global. A dúvida sobre a origem e a veracidade das informações apresentadas sobre a pandemia trouxeram desinformação, preocupação e insegurança para toda a sociedade. Diante desta problemática, este trabalho propõe o desenvolvimento de um painel de informações de acesso público, chamado *dashboard*, que apresenta dados abertos sobre os óbitos ocorridos durante a pandemia de COVID-19 no Brasil. Utilizou-se uma metodologia de Ciência de Dados e uma análise exploratória de um conjunto de dados abertos existente no repositório do Ministério da Saúde do Brasil. Este *dashboard* oferece uma exibição de dados abertos consolidados por meio de artefatos estatísticos e gráficos criados utilizando a linguagem de programação *R*. Concluiu-se que, utilizando-se da análise exploratória de dados e o *dashboard* por meio de técnicas estatísticas, computacionais e visuais, foram importantes para simplificar e apresentar, de forma efetiva, as informações compiladas, permitindo visualizar a identificação de evidências, padrões e oportunidades para embasar prováveis tomadas de decisões.

Palavras-chave: *dashboard*; COVID-19; linguagem de programação *R*; análise de dados.

SANTOS, M. J. **A DASHBOARD OF OPEN DATA ON DEATHS DUE TO COVID-19 IN BRAZIL**, 2024, 89 f. Technical-Scientific Report (Master in Media and Technology) - FAAC - UNESP, under the guidance of Associate Professor João Pedro Albino, Bauru, 2024.

ABSTRACT

The current technological advancement enables the processing and transformation of large quantities of data into precise and relevant information across all areas of knowledge. During the COVID-19 pandemic, this proliferation of data created a conducive environment for conducting studies and applying scientific methods, such as exploratory data analysis, to demonstrate and generate results, providing essential support in decision-making in combating this global health phenomenon. Doubts about the origin and veracity of the information presented about the pandemic have brought misinformation, concern, and insecurity to society as a whole. Considering this problem, this work proposes the development of a publicly accessible information panel, called a dashboard, which presents open data on deaths that occurred during the COVID-19 pandemic in Brazil. A Data Science methodology and exploratory analysis of an existing open dataset in the repository of the Brazilian Ministry of Health were used. This dashboard offers a display of consolidated open data through statistical artifacts and graphics created using the R programming language. It was concluded that the use of exploratory data analysis and the dashboard, through statistical, computational, and visual techniques, were important for simplifying and effectively presenting the compiled information, allowing the visualization of the identification of evidence, patterns, and opportunities to support probable decision-making.

Keywords: dashboard; COVID-19; R programming language; data analysis.

LISTA DE FIGURAS

Figura 1 – Objetivo de Desenvolvimento Sustentável 3 da ONU.

Figura 2 – Monitoramento em tempo real dos casos e mortes pelo novo coronavírus.

Figura 3 – *RedMonk Top 20 Languages Over Time: January 2023*.

Figura 4 – Tela inicial do *RStudio Desktop*®.

Figura 5 – Sobre a *Linguagem R* e *RStudio Desktop*®.

Figura 6 – Metodologia de Ciência de Dados.

Figura 7 – Conjunto de dados do *openDataSUS*.

Figura 8 – Dados e Recursos do conjunto de dados SRAG 2021 a 2023.

Figura 9 – Detalhes da pré-visualização do arquivo SRAG 2023 – 31/07.csv.

Figura 10 – Informações Adicionais do conjunto de dados.

Figura 11 – Os três arquivos texto do conjunto de dados.

Figura 12 – Análise inicial do conjunto de dados SRAG 2021 a 2023.

Figura 13 – Filtros, classificação, evolução do caso e data de óbito.

Figura 14 – Ficha de Notificação – Sinais e sintomas.

Figura 15 – Ficha de Notificação – Fatores de Risco/comorbidades.

Figura 16 – Tela principal do *dashboard*.

Figura 17 – Cabeçalho do *dashboard*.

Figura 18 – Rodapé do *dashboard*.

Figura 19 – Filtro do *dashboard*.

Figura 20 – Informações de totais de notificações COVID-19

Figura 21 – Gráfico de Faixa Etária.

Figura 22 – Gráfico de Gênero.

Figura 23 – Gráfico de Etnia.

Figura 24 – Gráfico de Escolaridade.

Figura 25 – Gráfico de Sinais e Sintomas.

Figura 26 – Gráfico de Fatores de Risco.

Figura 27 – Ranking de Sinais e Sintomas e Fatores de Risco.

Figura 28 – Informações gerais dos óbitos.

LISTA DE ABREVIATURAS

AED	Análise Exploratória de Dados
IBGE	Instituto Brasileiro de Geografia e Estatística
IEEE	Instituto de Engenheiros Eletricistas e Eletrônicos
LInDa	Laboratório de Inteligência de Dados – Unesp Bauru-SP
ODS	Objetivos de Desenvolvimento Sustentável
OMS	Organização Mundial de Saúde
ONU	Organização das Nações Unidas
OPAS	Organização Pan-Americana da Saúde
SRAG	Síndrome Respiratória Aguda Grave
SUS	Sistema Único de Saúde
UF	Unidade federativa
UTI	Unidade de Terapia Intensiva

SUMÁRIO

1 INTRODUÇÃO	13
1.1 OBJETO	16
1.2 PROBLEMA	16
1.3 OBJETIVOS	18
1.3.1 Objetivo Geral	18
1.3.2 Objetivos Específicos	18
1.4 JUSTIFICATIVA	18
2 REFERENCIAL TEÓRICO	22
2.1 O fenômeno da COVID-19 no Brasil	22
2.2 A Ciência de Dados e a importância dos dados na tomada de decisão	28
2.3 Dados Abertos no Brasil e o <i>OpenDataSUS</i>	30
2.4 A visualização de dados e os <i>dashboards</i>	32
2.5 A linguagem de programação <i>R</i> e o ambiente de desenvolvimento integrado <i>RStudio®</i>	35
2.6 A plataforma <i>GitHub</i> e a reprodutibilidade do método científico	44
3 METODOLOGIA	46
4 DESENVOLVIMENTO	48
4.1 Identificar o problema e a abordagem para resolvê-lo	48
4.2 Definir métodos de coleta e requisitos	48
4.3 Entender e preparar os dados	53
4.4 Gerar os modelos	60
4.5 Implantar o modelo	60
5 RESULTADOS ALCANÇADOS	62
6 CONSIDERAÇÕES FINAIS	69
REFERÊNCIAS	72
ANEXO A – Ficha de Registro Individual – Casos de SRAG-Hospitalizado	77
ANEXO B – Dicionário de dados – Ficha de Registro Individual	79

1 INTRODUÇÃO

A tecnologia da informação em constante evolução e desenvolvimento transforma a sociedade e impõe oportunidades e desafios no cotidiano das pessoas. A Internet como uma infraestrutura de conectividade global apresenta um ambiente interligado e colaborativo, resultando na geração de grandes quantidades de dados e informações.

Segundo a Agência de Notícias do IBGE (2022), a Internet já era acessível em 90% dos domicílios do país em 2021 e teve um avanço considerável nos acessos tanto na zona rural quanto na zona urbana. Um cenário de descobertas e conhecimentos é gerado com este notável fluxo de informações.

O fenômeno da pandemia COVID-19 impactou mundialmente não somente nas áreas biomédicas ou epidemiológicas, mas nas demais áreas como as sociais, econômicas, políticas, culturais etc. Consequentemente, durante este período, houve uma produção significativa de dados e informações de diferentes origens e formatos.

No Brasil, a necessidade de políticas públicas, a definição de ações urgentes e os incentivos às pesquisas científicas tornaram-se essenciais para a investigação e definição das estratégias no enfrentamento à COVID-19 (FIOCRUZ, [s.d]).

A geração e a análise dos dados relacionados à COVID-19 são processos desafiadores e fundamentais para o entendimento deste fenômeno. A Agência Senado (2020) citou a importância do levantamento e análise dos dados durante a pandemia:

Quando a pandemia acabar, ou tiver arrefecido os órgãos sanitários, as universidades e os pesquisadores independentes terão ainda de fazer uma cuidadosa escavação para determinar de fato o quanto a pandemia afetou a população brasileira em seus mais variados contornos: desde a situação social até a aspectos como a cor, o gênero e as doenças pré-existentes. O que se têm como muito provável, até em razão do aumento incomum de casos de Síndrome Respiratória Aguda Grave (SRAG) é que há muita subnotificação.

Neste sentido o fenômeno da pandemia COVID-19, demonstrou claramente que para o enfrentamento desta ameaça global, a sociedade e o poder público precisaram de mais Ciência.

A extração, coleta e geração de informações com qualidade, transparência e são essenciais para que o poder público possa entender os fenômenos, monitorar e

priorizar a utilização de seus recursos escassos, decidir sobre as políticas públicas e para responder e explicar para a sociedade o que aconteceu neste período emergencial (Agência Senado, 2020).

No entanto, segundo o Instituto Butantan (2022) com este ambiente colaborativo de dados, enfrentamos outra pandemia que é a grande quantidade e rápida disseminação das informações falsas. Essas informações são comparadas a um vírus ou mesmo a uma pandemia, afetando a sociedade como um todo e gerando confusão e transtornos.

Nesta perspectiva, por meio de uma política de dados abertos no Brasil vários órgãos públicos já publicam seus dados na Internet demonstrando os resultados das ações do governo e gerando efeitos importantes nas políticas públicas como a transparência, a inclusão e a responsabilidade.

O conceito de Governo Digital no Brasil foi instituído pelo Decreto nº 8.777, de 11 de maio de 2016 e a gestão desta política cabe à Controladoria-Geral da União, por meio da Infraestrutura Nacional de Dados Abertos – INDA (Decreto nº 9.903/2019) (Governo Digital – Dados Abertos, [s.d]).

A utilização dos dados abertos na investigação de fenômenos e tomada de decisões públicas, se faz indispensável nos assuntos de extrema importância na sociedade. Além disso, é imprescindível uma maior atenção com uma política de dados que garanta não apenas quantidade, mas também qualidade visando à eficiência do Estado (Agência Senado, 2020).

Em 2020, a pesquisadora da Universidade de São Paulo (USP) e Coordenadora da Rede de Pesquisa Solidária¹ Lorena Barberia, citou a importância do poder público garantir dados abertos e detalhados para melhorar as medidas de combate a COVID-19 e salvar vidas.

Os pesquisadores da Rede de Solidária já enfatizaram na época algumas lacunas como a demora na divulgação dos dados pelo poder público, a motivação política nas evidências deste fenômeno, a falta de padronização, qualidade e confiança nas informações coletadas e demais preocupações que resultaram em impactos severos na Ciência em relação ao entendimento deste fenômeno, na

¹ A Rede de Pesquisa Solidária é uma iniciativa de pesquisadores pra calibrar o foco e aperfeiçoar a qualidade de políticas públicas dos governos, estaduais e municipais que procuram atuar em meio à crise da COVID-19 para salvar vidas. Disponível em: <<https://jornal.usp.br/ciencias/rede-de-pesquisa-solidaria-acesse-as-ultimas-noticias/>>. Acesso em: 24 jul. 2023.

administração dos recursos do poder público, na definição de medidas adotadas em caráter de urgência pelo governo.

Os pesquisadores concluíram que toda a sociedade precisa de mais Ciência e do estudo dos dados, sendo que as lições aprendidas nesta pandemia, devem se tornar respostas para a sociedade e auxiliar os gestores na definição de eficientes políticas públicas (Agência Senado, 2020).

Visando atingir as características apontadas anteriormente de transparência, integridade e disponibilidade, o governo brasileiro oferece uma plataforma denominada *openDataSUS*² que busca oferecer um repositório centralizado de dados abertos do Ministério da Saúde, com informações que podem subsidiar análises objetivas da situação sanitária, tomadas de decisão baseadas em evidências e elaboração de programas de ações de saúde. Nesta plataforma existem 30 conjuntos de dados disponíveis especificamente sobre a pandemia da COVID-19.

Dessa forma, o principal objetivo deste estudo é desenvolver um painel visual, que forneça informações atualizadas e consolidadas sobre os óbitos relacionados à COVID-19 no Brasil, utilizando o conjunto de dados abertos disponibilizado e mantido pelo Ministério da Saúde.

Esse painel visual, ou *dashboard*, oferece acesso público e gratuito às informações sobre os óbitos de COVID-19 compiladas, permitindo visualizar a identificação de evidências e padrões sobre a pandemia.

Para viabilizar este estudo, será realizada uma Análise Exploratória de Dados (AED) com abordagem quantitativa e procedimentos estatísticos, seguindo a estrutura definida na *Metodologia de Ciência de Dados* apresentada em Siddiqi (2021).

O *dashboard* será construído com métodos e procedimentos computacionais, estatísticos e gráficos existentes na linguagem de programação *R*³.

Esta ferramenta visual conterá gráficos, tabelas, e informações claras e objetivas originadas de uma AED de um conjunto de dados abertos existente no repositório de dados abertos e licenciados do Ministério da Saúde do Brasil (“SRAG

² <https://opendatasus.saude.gov.br/>. Acesso em: 27 jul. 2023.

³ É uma linguagem de programação e um ambiente para desenvolvimento de ferramentas computacionais que envolvem conceitos de estatísticas e apresentação de resultados em forma gráfica. (Silva; Peres; Boscaroli, 2021).

2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 - OPENDATASUS”, [s.d.]).

Tratando-se do período pandêmico e da constante e rápida evolução da tecnologia da informação, este estudo contribuirá na meta do Objetivo 3 da Organização Mundial de Saúde (ONU). Este objetivo trata da Saúde e Bem-Estar e busca assegurar uma vida saudável e promover o bem-estar para todas as pessoas, em todas as idades.

1.1 OBJETO

Um painel visual e dinâmico de acesso público e gratuito, composto de dados abertos⁴ consolidados.

1.2 PROBLEMA

Ao longo do tempo, a Internet e a constante evolução de tecnologias de comunicação propiciaram um ambiente colaborativo e favorável à disseminação rápida de informações. Neste contexto desafiador, o mundo foi impactado pela pandemia COVID-19.

A Organização Mundial de Saúde (OMS) declarou este fenômeno como uma emergência de saúde pública de importância internacional em 30 de janeiro de 2020, o mais alto nível desta organização (OPAS, [s.d.]).

A COVID-19 afetou o mundo inteiro de forma sem precedentes, trazendo impactos significativos na economia, na saúde pública, na sociedade em geral e no bem-estar das pessoas. Trouxe um cenário de difícil controle para todas as esferas da sociedade e principalmente para os sistemas nacionais de saúde pública (FIOCRUZ, [s.d.]).

O Ministério da Saúde do Brasil ([s.d]) define:

A COVID-19 é uma infecção respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global. O SARS-CoV-2 é um betacoronavírus descoberto em amostras de lavado broncoalveolar obtidas de pacientes com pneumonia de causa desconhecida na cidade de Wuhan, província de Hubei, China, em dezembro de 2019. Pertence ao subgênero Sarbecovírus da família

⁴ Metodologia para a publicação de dados do governo em formatos reutilizáveis e sob uma licença aberta. (Governo Digital – Dados Abertos, [s.d.]).

Coronaviridae e é o sétimo coronavírus conhecido a infectar seres humanos.

Uma das principais medidas para conter o contágio da COVID-19, foi o distanciamento social que trouxe uma desordem significativa na saúde mental das pessoas. Este impacto foi mitigado pela utilização da Internet e de softwares de comunicação que permitiram o contato das pessoas com seus amigos e família, além de possibilitarem o desenvolvimento do trabalho remoto e propiciarem reais transformações na sociedade (*World Health Organization*, 2022).

Durante o período da pandemia, devido às restrições sanitárias, em diferentes partes do planeta tivemos muitos dispositivos conectados, gerando uma grande quantidade de dados. Com um ambiente propício para compartilhamentos, tais dados desempenharam um papel essencial na geração de informações e evidências importantes para monitoramento da propagação do vírus, além da identificação de grupos de risco, avaliação de intervenções direcionadas e pesquisas científicas.

As informações assim compartilhadas serviram principalmente para apoiar as estratégias e tomada de decisões das autoridades de saúde pública e governos referentes à priorização e gerenciamento dos recursos de saúde pública em face da demanda urgente gerada pela pandemia (Agência Senado, 2020).

As informações falsas ou desinformações já existem há muito tempo, porém, durante a pandemia foram tão ou mais destrutivas quanto o próprio vírus da COVID-19. O excesso de informação dificultando a identificação das fontes, a agilidade das redes sociais, a resistência das plataformas digitais em combater as notícias falsas e as campanhas públicas de comunicação sobre a COVID-19 que caminharam a passos lentos, colaboraram para a disseminação de desinformação sobre a pandemia (Agência Senado, 2021).

As informações falsas causaram pânico e confusão em toda a sociedade. Espalharam-se informações falsas sobre a origem da pandemia, os tratamentos e curas, os métodos de prevenção, as vacinas, os contágios e muitos outros aspectos (Instituto Butantan, 2022).

Diante dessa problemática, esse trabalho propõe o desenvolvimento de uma ferramenta e/ou painel visual, de acesso público e gratuito, que conterá informações consolidadas e sistematizadas, originárias de dados abertos e licenciados, sobre

óbitos no contexto da COVID-19 no Brasil, no período de 2021 a 2023, utilizando a Ciência de Dados e métodos científicos de Análise Exploratória de Dados.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Desenvolver uma ferramenta sistematizada e visual de acesso público e gratuito, denominado *dashboard*, que disponibilizará informações consolidadas de um conjunto de dados abertos do Ministério da Saúde sobre os óbitos durante a COVID-19 no Brasil.

1.3.2 Objetivos Específicos

- Desenvolver/sistematizar uma Análise Exploratória de Dados e uma metodologia de Ciência de Dados⁵ utilizando um conjunto de dados abertos sobre a COVID-19 disponível no repositório de dados abertos do Ministério da Saúde;
- Estudar a linguagem de programação *R* e sua aplicação na Análise Exploratória de Dados e no desenvolvimento de *dashboards*;
- Desenvolver um *dashboard*, composto de filtros, gráficos e tabelas, para visualização de informações originadas da Análise Exploratória de Dados;
- Disponibilizar a documentação desse estudo e o código-fonte da linguagem de programação *R*, em um repositório online e público denominado *GitHub*⁶, para atender a reprodutibilidade da pesquisa;
- Disponibilizar o *dashboard* na *Internet* para acesso público e gratuito.

1.4 JUSTIFICATIVA

⁵ Trata-se de uma Ciência para obter informação e conhecimento de forma sistemática, bem como normalizar e organizar este conhecimento. Estuda o dado em todo o seu ciclo de vida (AMARAL, 2021).

⁶ É uma plataforma mundial de hospedagem de código-fonte e arquivos com controle de versão Git. Permite que qualquer usuário cadastrado contribua em projetos privados. Utilizada para divulgação de trabalhos ou para prover um ambiente colaborativo para projetos. Disponível em <<https://github.com/>> (GITHUB, 2022).

Os dados constituem a matéria-prima deste estudo. Segundo Amaral (2016), é fundamental compreender os conceitos de dados, informações e o conhecimento. Os dados representam fatos ou valores coletados e normalmente armazenados. As informações, por sua vez, são os dados que foram analisados, consolidados e adquiriram algum significado. Já o conhecimento é a informação entendida por meio de algum raciocínio e aplicada para alguma finalidade.

Nos dias atuais, a tecnologia da informação está transformando nossa sociedade de forma rápida e abrangente, gerando dados em grande volume e por meio de uma variedade de dispositivos existentes. Para aproveitar os benefícios desta matéria-prima bruta, os dados devem ser coletados, analisados, preparados e visualizados de maneira a resultar em informações claras, objetivas e precisas para uma compreensão certa dos fatos e para embasar tomadas de decisões eficazes (Favero *et al*, 2009).

Neste contexto, a pandemia da COVID-19 evidenciou a necessidade de uma abordagem baseada na Ciência para enfrentar essa ameaça global. A geração de informações com qualidade, transparência e com rapidez torna-se fundamental para que o poder público compreenda este fenômeno, monitore e priorize a utilização de recursos escassos, decida sobre políticas públicas e possa responder adequadamente e explicar à sociedade o que ocorreu neste fenômeno emergencial (Agência Senado, 2020).

Já uma grande oportunidade deste trabalho reside na utilização do repositório de dados do Ministério da Saúde do Brasil, para enfatizar a importância da utilização de dados abertos, certificados e públicos mantidos pelo governo e sob sua guarda. Este enfoque visa promover a transparência e a divulgação das informações, contribuindo para o combate à desinformação.

Como observado pelo Instituto Butantan (2022), a rápida disseminação das informações falsas é compatível com um vírus ou até mesmo com uma pandemia.

Esta pesquisa contribui para os Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU). Alinha-se com o Objetivo Número 3 - Saúde e Bem-Estar da ONU, que visa garantir o acesso à saúde de qualidade e promover o bem-estar para todos, em todas as idades. No item 3.d deste objetivo, consta: “3.d Reforçar a capacidade de todos os países, particularmente os países em desenvolvimento, para o alerta precoce, redução de riscos e gerenciamento de riscos nacionais e globais de saúde” (ONU, [s.d.]).

Ao utilizar dados abertos e públicos do Ministério da Saúde do Brasil para analisar e visualizar informações sobre a mortalidade relacionada à COVID-19, esta pesquisa contribui diretamente para fortalecer a capacidade de alerta precoce e gerenciamento de riscos de saúde em níveis nacionais e globais.

Para avançar nesse objetivo, a pesquisa desenvolverá uma ferramenta visual de acesso público, transparente e objetiva. Esta ferramenta divulgará informações geradas por meio de uma metodologia de Ciência de Dados utilizando dados abertos e consistentes para a descoberta de padrões e tendências.

Ao disponibilizar estas informações, este estudo visa apoiar prováveis tomadas de decisões e a formulação de políticas públicas urgentemente necessárias para a redução de riscos nacionais e globais de saúde.

Com o intuito de atrair atenção e garantir a qualidade da análise, a pesquisa será composta de uma Análise Exploratória de Dados, realizada utilizando a linguagem de programação *R*.

De acordo com a Sage Campus (2019), essa linguagem é amplamente empregada nas universidades e oferece recursos computacionais avançados de visualização de dados, habilidades em Ciência de Dados e supercompatibilidade com outras tecnologias. Um dos pontos-chave que torna a linguagem de programação *R* tão valiosa é a possibilidade de garantir a reprodutibilidade dos resultados, um dos princípios fundamentais da pesquisa científica.

Utilizando dados da Prefeitura de Bauru, cidade do interior do estado de São Paulo, no início de 2021, o Portal G1 (2021) divulgou que 47 pessoas haviam morrido na fila de espera por um leito de UTI ou enfermaria.

Posto isso, também se faz necessário elencar que esta pesquisa comporta também uma motivação pessoal, pois, neste período citado da pandemia de COVID-19, um amigo perdeu a vida devido à falta de leitos no Sistema Único de Saúde no município.

Este pesquisador mantinha próxima convivência semanal com este amigo, pois participamos de uma equipe esportiva amadora do *Sport Club Corinthians Paulista*, e frequentemente viajavamos para as competições.

Esta perda pessoal, juntamente com outras vidas perdidas, destaca a urgência e a importância de compreender os desafios enfrentados pelo sistema de saúde durante a pandemia. Motivou o empenho em contribuir com soluções para enfrentar esses desafios.

Deve-se ressaltar também que a pesquisa compõe um projeto maior chamado Impactos da COVID-19 no Ensino Superior do Brasil existente no Grupo de Pesquisa *LinDa*⁷. Ao fazer parte deste projeto, a presente pesquisa se beneficia do ambiente colaborativo e dos recursos disponíveis no grupo de pesquisa e no Programa de Pós-graduação em Mídia e Tecnologia da FAAC-UNESP/Bauru.

Portanto, essas razões serão os principais pilares que guiaram este trabalho. Ao integrar a necessidade de compreender os desafios enfrentados durante a pandemia, bem como a motivação pessoal oriunda de uma experiência impactante, a pesquisa busca contribuir significativamente para o conhecimento sobre a importância dos dados abertos na geração das informações para ações destinadas a enfrentar os desafios de saúde pública e de toda a sociedade.

⁷ O Laboratório de Inteligência de Dados – LinDa da Unesp de Bauru-SP, tem como objetivo realizar pesquisas relacionadas à Ciência de Dados, e em particular, estudar estruturas de dados onde seja possível extrair "inteligência". Disponível em < <http://dgp.cnpq.br/dgp/espelhogrupo/379670>>. Acesso em: 17 jul. 2023.

2 REFERENCIAL TEÓRICO

2.1 O fenômeno da COVID-19 no Brasil

No final de 2019, a OMS recebeu um alerta de vários casos de pneumonia na cidade de Wuhan, na República Popular da China. Uma semana após o alerta, os chineses identificaram esse novo tipo de coronavírus. Os coronavírus são a principal causa do resfriado comum e raramente causam doenças graves nos seres humanos. No entanto, esses casos foram atribuídos a um novo tipo de coronavírus, posteriormente chamado de SARS-CoV-2, responsável pela COVID-19 (OPAS, [s.d.]).

Em janeiro de 2020, a OMS declarou que o surto do novo coronavírus constituía uma emergência de saúde pública de importância internacional, o mais alto nível de alerta desta organização. No mês de março de 2020, a COVID-19 foi caracterizada como uma pandemia, este termo se refere à distribuição geográfica de uma doença e não a sua gravidade. Esta denominação refletiu na necessidade de ações coordenadas em nível internacional para enfrentar esta crise de saúde pública (OPAS, [s.d.]).

Sobre o coronavírus e a COVID-19, o Ministério da Saúde do Brasil ([s.d.]) comentou:

Os coronavírus são uma grande família de vírus comuns em muitas espécies diferentes de animais, incluindo o homem, camelos, gado, gatos e morcegos. Raramente os coronavírus de animais podem infectar pessoas e depois se espalhar entre seres humanos como já ocorreu com o MERS-CoV e o SARS-CoV-2. Até o momento, não foi definido o reservatório silvestre do SARS-CoV-2.

A Covid-19 é uma infecção respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global. O SARS-CoV-2 é um betacoronavírus descoberto em amostras de lavado broncoalveolar obtidas de pacientes com pneumonia de causa desconhecida na cidade de Wuhan, província de Hubei, China, em dezembro de 2019. Pertence ao subgênero Sarbecovírus da família Coronaviridae e é o sétimo coronavírus conhecido a infectar seres humanos.

No Brasil o fenômeno da pandemia de COVID-19 teve início em 26 de janeiro de 2020 com a confirmação do primeiro caso no Estado de São Paulo (“Pandemia de COVID-19 no Brasil”, 2022).

Assim, na primeira semana de março de 2020 os casos de infecção foram confirmados em São Paulo, Bahia, Rio de Janeiro, Distrito Federal e Espírito Santo. Quando da decretação oficial da pandemia pela Organização Mundial de Saúde, o

Brasil já registrava mais de cem casos e em poucos dias o Ministério da Saúde do Brasil anunciou a transmissão comunitária em todo o território.

A pandemia evoluiu em ondas ou fases, com o surgimento de novas variantes do vírus infeccioso que causaram um aumento nos casos. O pico da primeira onda aconteceu em meados de 2020, enquanto uma nova onda, mais grave, atingiu o Brasil em março de 2021.

Mesmo com o início das vacinações, ocorreu uma terceira escalada no início de 2022, devido a uma nova variante difundida no exterior. Após um período de diminuição da terceira onda, houve um aumento de casos e hospitalizações em outubro de 2022, relacionado a uma sub variante do vírus (“Pandemia de COVID-19 no Brasil”, 2022).

Neste período desafiador, o sistema de saúde do Brasil em 2021 passou por uma séria crise sanitária devido a mortes em filas de espera por leitos, desabastecimentos de produtos de oxigênio e equipamentos para intubação. Durante esse período a FIOCRUZ ([s.d]) comentou:

A estimativa de infectados e mortos concorre diretamente com o impacto sobre os sistemas de saúde, com a exposição de populações e grupos vulneráveis, a sustentação econômica do sistema financeiro e da população, a saúde mental das pessoas em tempos de confinamento e temor pelo risco de adoecimento e morte, acesso a bens essenciais como alimentação, medicamentos, transporte, entre outros.

No âmbito da Saúde, neste período pandêmico afetou principalmente os profissionais, auxiliares ou técnicos de enfermagem, pois estavam na linha de frente do enfrentamento com carga de trabalho intensificada, problemas de saúde mental e exposição ao vírus (“Pandemia de COVID-19 no Brasil”, 2022).

A falta de suprimentos, profissionais, leitos e equipamentos no contexto operacional gerou um caos hospitalar nacional e até mesmo impactou nos serviços funerários. Na primeira quinzena de março de 2021, 24 estados apresentaram taxas de ocupação de leitos de UTI superiores a 80% (“Pandemia de COVID-19 no Brasil”, 2022).

Todo este fenômeno atingiu também a rotina diária da área da Saúde, atrasando e cancelando atendimentos, diagnósticos, calendário de vacinas e os

tratamentos das demais doenças, além de agravar seriamente a saúde mental de todas as pessoas.

Na economia a pandemia impactou diretamente o emprego e a renda da população. As ações de combate à pandemia principalmente o distanciamento social resultaram em demissões nas empresas e reações em todo o mercado de trabalho (“Pandemia de COVID-19 no Brasil”, 2022).

O Congresso Nacional do Brasil aprovou um programa emergencial de manutenção de empregos possibilitando: a suspensão do contrato de trabalho, a redução da jornada e dos salários (“Pandemia de COVID-19 no Brasil”, 2022). Também aprovou o auxílio emergencial, um programa de transferência de renda mínima aos mais vulneráveis, para aliviar os impactos do mercado de trabalho, das medidas sanitárias, da pressão inflacionária dos alimentos e o comprometimento do orçamento das famílias menos favorecidas.

Tratando-se do governo, todas estas ações emergenciais executadas trouxeram despesas que resultaram em um aumento da dívida pública (“Pandemia de COVID-19 no Brasil”, 2022).

A pandemia também alterou a rotina dos estudantes no Brasil, afetando o sistema educacional nos mais diversos níveis de aprendizado com a paralisação das aulas em todo o território nacional. O Brasil foi o país que manteve escolas por mais tempo fechadas em 2020 (“Pandemia de COVID-19 no Brasil”, 2022).

Nas universidades públicas os calendários foram alterados e em muitas instituições privadas as férias foram antecipadas, porém com a impossibilidade de aulas presenciais as instituições educacionais adotaram a educação à distância. Esta modalidade de ensino resultou em uma continuidade da educação, porém evidenciou mais ainda a desigualdade social no Brasil. Os alunos mais prejudicados com a pandemia foram os alunos em fase de alfabetização (“Pandemia de COVID-19 no Brasil”, 2022).

Como os demais setores, a cultura e o turismo tiveram atividades paralisadas. Na cultura, o cancelamento de shows, estabelecimentos fechados e o distanciamento social resultaram em demissões, encerramento de empresas especializadas, queda de faturamento e principalmente o aumento da desigualdade social da sociedade brasileira (“Pandemia de COVID-19 no Brasil”, 2022).

Quanto ao turismo, a pandemia diminuiu o número de turistas internacionais e resultou em certo aumento no turismo interno no decorrer deste período pandêmico.

Vale citar que o setor aéreo sofreu por ter a frota parada devido ao fechamento das fronteiras (“Pandemia de COVID-19 no Brasil”, 2022).

Todos estes efeitos abalaram o setor do turismo e cultura, e principalmente repercutiram no capital humano envolvido nessas atividades (“Pandemia de COVID-19 no Brasil”, 2022).

A taxa de mortalidade ou coeficiente de mortalidade é um índice demográfico que indica o número de mortes registradas. Torna-se um indicador social importante, pois quanto piores as condições de vida, maior a taxa de mortalidade e consequentemente menor a esperança de vida das pessoas. Esta taxa é aplicada em conjunto com as causas de óbitos e o tamanho da população afetada por esta causa (Taxa de Mortalidade, 2023).

A gravidade de uma doença pode ser descrita pela Taxa de Mortalidade por Infecção, que é um percentual calculado: dividir o número de mortes pela doença dividido pelo número de indivíduos infectados, esse resultado deve ser multiplicado por cem. O monitoramento em tempo real desta tendência de gravidade precisa ser realizado com os dados disponíveis e de forma prioritária (OPAS, 2020).

Referente à COVID-19 e o excesso de mortalidade, a OPAS (2022) comenta:

O excesso de mortalidade inclui mortes associadas diretamente à COVID-19 (devido à doença) ou indiretamente (devido ao impacto da pandemia nos sistemas de saúde e na sociedade). As mortes ligadas indiretamente à doença são atribuíveis a outras condições de saúde para as quais as pessoas não tiveram acesso à prevenção e tratamento porque os sistemas de saúde foram sobrecarregados pela pandemia. O número estimado de mortes em excesso pode ser influenciado também pelas mortes evitadas durante a pandemia devido aos menores riscos de determinados eventos, como acidentes no trânsito ou acidentes de trabalho.

O atendimento adequado para os casos confirmados ou suspeitos da COVID dependem do reconhecimento precoce das condições, fatores de risco e comorbidades. Estes aspectos foram constantemente mudando o cenário epidemiológico da COVID-19. O Ministério da Saúde (2021) citou:

Em decorrência da pandemia da covid-19 e consequente aumento da demanda assistencial sobre o Sistema Único de Saúde (SUS), o MS reforça que a organização da rede de atenção e dos fluxos deve estar bem estabelecida tanto para pessoas com síndrome gripal (SG), causada ou não pela covid-19, quanto para as que necessitem de acompanhamento por outras condições e agravos de saúde, como outras doenças transmissíveis, doenças crônicas e situações de vulnerabilidade e sofrimento psíquico, garantindo a continuidade do cuidado pelos serviços de saúde.

O tema “Saúde e Bem-Estar” citado no Objetivo Sustentável número 3 dos Objetivos de Desenvolvimento Sustentável (ODS) da Organização Mundial da Saúde foi evidentemente o mais impactado (Observatório do Futuro, 2021).

Segundo o Observatório do Futuro (2021) essa crise sanitária que causou interrupções e lentidão nos serviços de saúde pode reverter décadas de avanços na qualidade de saúde da população mundial. Existe um consenso de que o Sistema Único de Saúde (SUS) atuou com maestria para o enfrentamento da trajetória desta pandemia. As consequências para este sistema demonstraram que investimentos em infraestrutura e capacitação devem ser repensados urgentemente, pois o acesso à saúde deve estar disponível para todos independente da doença e de eventuais fenômenos de emergência.

O Objetivo Desenvolvimento Sustentável número 3⁸ da ONU tem como objetivo principal: garantir o acesso à saúde de qualidade e promover o bem-estar para todos, em todas as idades e contém os seguintes itens:

- a) 3.1 Até 2030, reduzir a taxa de mortalidade materna global para menos de 70 mortes por 100.000 nascidos vivos;
- b) 3.2 Até 2030, acabar com as mortes evitáveis de recém-nascidos e crianças menores de 5 anos, com todos os países objetivando reduzir a mortalidade neonatal para pelo menos 12 por 1.000 nascidos vivos e a mortalidade de crianças menores de 5 anos para pelo menos 25 por 1.000 nascidos vivos;
- c) 3.3 Até 2030, acabar com as epidemias de AIDS, tuberculose, malária e doenças tropicais negligenciadas, e combater a hepatite, doenças transmitidas pela água, e outras doenças transmissíveis;
- d) 3.4 Até 2030, reduzir em um terço a mortalidade prematura por doenças não transmissíveis via prevenção e tratamento, e promover a saúde mental e o bem-estar;
- e) 3.5 Reforçar a prevenção e o tratamento do abuso de substâncias, incluindo o abuso de drogas entorpecentes e uso nocivo do álcool;

⁸ Disponível em: <<https://brasil.un.org/pt-br/sdgs/3>>. Acesso em: 27 jul. 2023.

- f) 3.6 Até 2020, reduzir pela metade as mortes e os ferimentos globais por acidentes em estradas;
- g) 3.7 Até 2030, assegurar o acesso universal aos serviços de saúde sexual e reprodutiva, incluindo o planejamento familiar, informação e educação, bem como a integração da saúde reprodutiva em estratégias e programas nacionais;
- h) 3.8 Atingir a cobertura universal de saúde, incluindo a proteção do risco financeiro, o acesso a serviços de saúde essenciais de qualidade e o acesso a medicamentos e vacinas essenciais seguros, eficazes, de qualidade e a preços acessíveis para todos;
- i) 3.9 Até 2030, reduzir substancialmente o número de mortes e doenças por produtos químicos perigosos, contaminação e poluição do ar e água do solo
- j) 3.a Fortalecer a implementação da Convenção-Quadro para o Controle do Tabaco em todos os países, conforme apropriado;
- k) 3.b Apoiar a pesquisa e o desenvolvimento de vacinas e medicamentos para as doenças transmissíveis e não transmissíveis, que afetam principalmente os países em desenvolvimento, proporcionar o acesso a medicamentos e vacinas essenciais a preços acessíveis, de acordo com a Declaração de Doha, que afirma o direito dos países em desenvolvimento de utilizarem plenamente as disposições do acordo TRIPS sobre flexibilidades para proteger a saúde pública e, em particular, proporcionar o acesso a medicamentos para todos;
- l) 3.c Aumentar substancialmente o financiamento da saúde e o recrutamento, desenvolvimento e formação, e retenção do pessoal de saúde nos países em desenvolvimento, especialmente nos países menos desenvolvidos e nos pequenos Estados insulares em desenvolvimento;
- m) 3.d Reforçar a capacidade de todos os países, particularmente os países em desenvolvimento, para o alerta precoce, redução de riscos e gerenciamento de riscos nacionais e globais de saúde.

Figura 1: Objetivo de Desenvolvimento Sustentável número 3 da ONU.



Fonte: Adaptação de ONU ([s.d.]).

2.2 A Ciência de Dados e a importância dos dados na tomada de decisão

Os conceitos de dados, informação e conhecimento são confundidos como sinônimos, mas no contexto de Ciência de Dados, eles têm significados distintos. De acordo com Silva, Peres e Boscarioli (2021), os dados são fatos, valores ou resultados de uma medição. Quando estes dados são interpretados e um significado é atribuído a eles, gera-se a informação. Já o conhecimento é a informação interpretada e aplicada para um determinado fim. Esses conceitos formam as matérias-primas para a Ciência de Dados.

Conforme observado por Gil (2008), os dados são de extrema importância em todas as etapas de um processo de investigação. Eles fornecem a base para a fundamentação teórica, a análise e interpretação dos resultados, a possibilidade para fazer inferências e identificar padrões. Os dados representam uma fonte relevante para tomada de decisões e, principalmente atuam como fonte de conhecimento, ao possibilitar correlações, evidências, teorias, hipóteses e direções de uma pesquisa.

A Ciência de Dados é uma Ciência ou disciplina interdisciplinar, que estuda os dados em todo o seu ciclo de vida, desde a produção até o descarte. Este ciclo de vida do dado compreende as etapas de: produção, armazenamento, transformação, análise e descarte dos dados. A Ciência de Dados busca extrair informação e conhecimento por meio de processos, modelos e tecnologias, além de organizar este conhecimento (Amaral, 2016).

Escovedo e Koshiyama (2020) comentaram:

O conceito de *Data Science* (ou Ciência de Dados), por sua vez, é mais amplo: refere-se à coleta de dados de várias fontes para fins de análise, com o objetivo de apoiar a tomada de decisões, utilizando geralmente grandes quantidades de dados, de forma sistematizada. Quase sempre, além de olhar para os dados passados para entender o comportamento dos mesmos (atividade conhecida como *Business Intelligence* - BI), deseja-se também realizar análises de forma preditiva, por exemplo, utilizando técnicas de *Data Mining* e/ou *Machine Learning*.

Já a Análise Exploratória de Dados (AED) tem como principal objetivo gerar hipóteses, aplicar métodos estatísticos e descobrir padrões, correlações e tendências preliminares nos dados (Favero *et al*, 2009).

Segundo IBM ([s.d.]), a AED possibilita um melhor entendimento dos dados, colaborando para a identificação de padrões, eventos anômalos e relações. Além disso, a AED pode garantir que os resultados produzidos sejam válidos e aplicáveis.

Na AED, várias técnicas estatísticas podem ser usadas, incluindo agrupamentos, filtros, seleção, medidas estatísticas descritivas, correlação entre variáveis, regressão linear, modelos preditivos, entre outras.

De acordo com IBM ([s.d.]) existem quatro tipos de análise exploratória de dados:

- a) Não gráfico univariado. É a forma mais simples, consiste em apenas uma variável e sem imagens;
- b) Gráfico univariado. Fornece imagens dos dados, com gráficos univariados como histograma, gráficos de caixa que representam mínimo, primeiro quartil, mediana, terceiro quartil e máximo;
- c) Não Gráfico multivariado. Utiliza mais de uma variável por meio de tabulação cruzada e estatística, não possui imagem;
- d) Gráfico Multivariado. Correlaciona dados de um ou mais conjunto de dados, utilizando gráficos de barras agrupadas ou nivelados por variáveis. Também pode conter gráficos de dispersão, gráfico de bolhas ou mapa de calor.

Segundo o Blog Sirius Educação (2022), o processo de tomada de decisão envolve a definição de um plano para resolver um problema específico. Os responsáveis por essa definição precisam ter um amplo conhecimento para elaborar

planos mais assertivos e funcionais. Essa definição é dinâmica e requer a interpretação e a junção de muitas informações.

A precipitação em um processo de tomada de decisões pode resultar em retrabalho, interrupção de projetos, prejuízo, desperdício de recursos e até mesmo afetar a motivação e a orientação das pessoas envolvidas no processo. Este processo possui cinco fases: identificação do problema; coleta de dados; análise de estratégias; definição de melhor alternativa; e a execução.

A coleta, análise e visualização dos dados são essenciais neste processo, fornecendo informações concretas e certas que possibilitam a obtenção de conhecimentos, monitoramento, previsões e, principalmente, a identificação de oportunidades ou desafios. A Ciência de Dados desempenha um papel fundamental neste processo, ajudando na definição do plano a ser seguido.

2.3 Dados Abertos no Brasil e o *OpenDataSUS*

Dados abertos é uma metodologia ou procedimento para divulgação de dados do governo em formatos digitais, permitindo que qualquer pessoa os utilize, reutilize e distribua sem a necessidade de restrições técnicas, legais ou sociais (Open Knowledge Foundation, [s.d]).

A *Open Knowledge Foundation* ([s.d]) comenta que os dados podem ser considerados abertos quando:

Qualquer pessoa pode acessar, usar, modificar e compartilhar livremente para qualquer finalidade (sujeito a, no máximo, a requisitos que preservem a proveniência e a sua abertura). Isso geralmente é satisfeito pela publicação dos dados em formato aberto e sob uma licença aberta.

A *OpenKnowledge Foundation*⁹ é uma organização sem fins lucrativos fundada em 2004, que trabalha na intersecção de ferramentas digitais e uma rede distribuída de comunidades e movimentos para promover o conhecimento aberto. Sua missão é criar um futuro, justo, sustentável e aberto, promovendo o conhecimento aberto e desenvolvendo modelos, padrões e políticas de forma ética e ágil. Assim, acredita que o conhecimento aberto deve ser adotado por todos os governos, instituições e movimentos. Essa adoção garantirá o acesso às

⁹ <https://okfn.org/>. Acesso em: 27 jul. 2023.

informações para capacitar os seres humanos a resolver problemas urgentes no futuro (*Open Knowledge Foundation*, [s.d]).

No Brasil o Poder Executivo definiu a política de Dados Abertos pelo Decreto nº 8.777, de 11 de maio de 2016¹⁰. O gestor desta política federal a Controladoria-Geral da União por meio da Infraestrutura de Dados Abertos – INDA (Decreto nº 9.903/2019¹¹).

Diversos órgãos da administração pública já adotam a metodologia de dados abertos para publicar suas informações na Internet (Governo Digital – Dados Abertos, [s.d.]).

Alguns efeitos dos dados abertos governamentais nas políticas públicas são:

- a) Inclusão. Os dados disponibilizados em formatos padronizados permitem que qualquer pessoa utilize softwares para análises e demais necessidades;
- b) Transparência. As informações das políticas públicas abertas e acessíveis contribuem para a transparência e no uso adequado das informações;
- c) Responsabilidade: Os conjuntos de dados abertos, comunicam informações e opiniões sobre o cumprimento das políticas públicas e suas metas.

De acordo com o Portal de Dados Abertos – Ministério da Saúde ([s.d]) no âmbito do Poder Executivo Federal, o órgão responsável pela organização e elaboração de planos e políticas públicas voltadas para a assistência à saúde dos brasileiros é o Ministério da Saúde.

O Ministério da Saúde tem como principal função proporcionar uma melhor qualidade de vida à sociedade brasileira, mediante proteção e recuperação da saúde da população, redução de enfermidades, controle de doenças endêmicas e parasitárias, e aprimorar a vigilância à saúde. Sua missão é promover a saúde da população por meio da integração e parcerias com as unidades da Federação, os

¹⁰ https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm. Acesso em: 27 jul. 2023.

¹¹ https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Decreto/D9903.htm#art1. Acesso em: 27 jul. 2023.

municípios, a iniciativa privada e a sociedade (Portal de Dados Abertos – MINISTÉRIO DA SAÚDE, [s.d]).

O repositório ou plataforma online de dados abertos do Ministério da Saúde, chama-se *openDataSUS*¹². Esta plataforma é uma iniciativa do Ministério da Saúde para disponibilizar informações com transparência, responsabilidade e inclusão.

Nesta plataforma existem 30 conjuntos de dados disponíveis, abrangendo vários temas como saúde, epidemiologia, serviços e indicadores de saúde, entre outros. A plataforma *openDataSUS* tem como definição:

O DATASUS disponibiliza informações que podem servir para subsidiar análises objetivas da situação sanitária, tomadas de decisão baseadas em evidências e elaboração de programas de ações de saúde.

A mensuração do estado de saúde da população é uma tradição em saúde pública. Teve seu início com o registro sistemático de dados de mortalidade e de sobrevivência (Estatísticas Vitais – Mortalidade e Nascidos Vivos). Com os avanços no controle das doenças infecciosas (informações Epidemiológicas e Morbidade) e com a melhor compreensão do conceito de saúde e de seus determinantes populacionais, a análise da situação sanitária passou a incorporar outras dimensões do estado de saúde. Dados de morbilidade, incapacidade, acesso a serviços, qualidade da atenção, condições de vida e fatores ambientais passaram a ser métricas utilizadas na construção de Indicadores de Saúde, que se traduzem em informação relevante para a quantificação e a avaliação das informações em saúde. Nesta seção também são encontradas informações sobre Assistência à Saúde da população, os cadastros (Rede Assistencial), das redes hospitalares e ambulatoriais, o cadastro dos estabelecimentos de saúde, além de informações sobre recursos financeiros e informações Demográficas e Socioeconômicas. Além disso, em Saúde Suplementar, são apresentados links para as páginas de informações da Agência Nacional de Saúde Suplementar – ANS (OPENDATASUS, [s.d]).

2.4 A visualização de dados e os *dashboards*

Sobre visualização de dados, a Oracle (2014) comentou:

Peneirar informações para entender o que importa e o que não importa está se tornando mais difícil. O visual torna a análise muito mais fácil e rápida e oferece a capacidade de ver de relance o que é importante. Além disso, a maioria das pessoas responde muito melhor aos recursos visuais do que ao texto - 90% das informações enviadas ao cérebro são visuais, e o cérebro processa os recursos visuais a 60.000 vezes a velocidade do texto. Esses pontos são um forte argumento para o uso da visualização de dados para analisar e transmitir informações.

¹² <https://opendatasus.saude.gov.br/>. Acesso em: 27 jul. 2023.

A visualização de dados é uma metodologia essencial para análises avançadas, conceituada pela Oracle (2014) como fundamental na inteligência de negócios. Para que as pessoas ou organizações compreendam todas as informações geradas, os dados podem ser representados de forma gráfica ou por meio de outras apresentações visuais. Uma visualização de dados bem elaborada não apenas comunica as informações de forma eficaz, mas também desperta a atenção e o interesse em um determinado assunto. Além disso, é fundamental para descobrir percepções que embasam um plano de tomada de decisão.

A visualização de dados não é um conceito novo, mas tem evoluído com o advento de novas tecnologias, diversidade de dispositivos conectados, integração com grandes conjuntos de dados e com uma enorme variedade de interfaces. Essa evolução apresenta novos métodos e funcionalidades, tais como gráficos dinâmicos, mapas em tempo real e outras ferramentas interativas e personalizáveis.

De acordo com Peng (2016), a visualização dos dados por meio de gráficos e demais funcionalidades visuais é muito importante nas etapas iniciais da análise dos dados, ajudando a compreender as propriedades básicas dos dados, identificar padrões e apontar possíveis estratégias para um modelo de tratamento dos dados. Além disso, durante as fases da análise, os métodos de visualização dos dados são úteis para depurar problemas e definir ajustes na preparação e tratamento dos resultados finais da análise de dados.

Amaral (2015) discutiu a importância da visualização de dados por meio de um *dashboard*, destacando-a como um modelo analítico apropriado para a análise de dados:

Dashboards são painéis visuais que mostram indicadores de um mesmo assunto. Trazem informação resumida, normalmente de cunho estratégico ou geral, mas também têm aplicações nas áreas operacionais. Oferece características de navegação de dados, como filtros, drill downs e drill ups. Embora não deva conter detalhes, pode trazer os melhores ou piores. Também pode conter indicadores de performance.

Segundo Few (2013), *dashboard* é o novo nome dos sistemas de informações executivas desenvolvidos nos anos 1980. Esses sistemas eram baseados em painéis digitais que apresentavam indicadores ou medidas financeiras de fácil acesso para o entendimento dos executivos.

Desta forma, os *dashboards* oferecem a possibilidade de contar história por meio de dados gerados de diversas fontes e em grandes quantidades. Eles desempenham um papel fundamental na descoberta de padrões, informações anômalas ou discrepantes e até mesmo na previsão de tendências, conforme afirmado por Amaral (2015).

Nos anos 1990, com o advento e popularização da Internet e o desenvolvimento de novas tecnologias, esses painéis digitais foram aperfeiçoados e amplamente utilizados no ambiente corporativo.

A XP Educação (2022) descreve quatro tipos de *dashboards*:

- a) ***Dashboard Operacional*** que apresenta operações técnicas e processos e serve para demonstrar métricas das rotinas produtivas. Tem como principal objetivo diminuir os erros e aumentar a produtividade;
- b) ***Dashboard Analítico*** que fornece informações importantes para o estudo do cenário com um todo, onde a gestão é capaz de monitorar e identificar tendências e padrões;
- c) ***Dashboard Estratégico*** é usado pela diretoria na tomada de decisões de longo prazo. Também pode ser apresentado para a empresa toda em forma de ajudar a envolver todos os colaboradores nas decisões;
- d) ***Dashboard Tático*** usado pelos cargos de gestão e gerência, para a tomada de decisões no médio prazo na obtenção de vantagens competitivas.

Resumidamente, os principais benefícios na utilização de um *dashboard* incluem: apoio na tomada de decisões com base em dados; transparência das informações apresentadas; melhoria na comunicação com a utilização de diferentes visualizações, cores e objetos gráficos; layout moderno e claro; otimização de tempo e recursos; alinhamento de estratégias; e integração dos dados relacionados a um determinado assunto.

Durante a pandemia, o site mais visitado para monitoramento em tempo real dos casos e mortes pelo novo coronavírus foi um *dashboard* denominado *COVID-19 Dashboard*¹³.

Este site foi desenvolvido pelo *Coronavirus Resource Center* da Universidade John Hopkins¹⁴ em Maryland, Estados Unidos.

As atualizações diárias deste mapa foram utilizadas pelos principais veículos de mídia, departamentos de saúde pública e agências governamentais no mundo durante a fase crítica da COVID-19 (CNN Brasil, 2020).

Figura 2: Monitoramento em tempo real dos casos e mortes pelo novo coronavírus.



Fonte: Adaptação de JOHN HOPKINS UNIVERSITY (2022).

2.5 A linguagem de programação R e o ambiente de desenvolvimento integrado RStudio®

A linguagem de programação R é uma linguagem de programação computacional com um ambiente de *software* integrado para métodos estatísticos e apresentação de resultados em formas de gráficos. Seu ambiente de desenvolvimento permite coletar, processar, consolidar e visualizar os dados para atender as etapas de uma análise de dados (Matos, 2015).

¹³ <https://coronavirus.jhu.edu/map.html>. Acesso em: 27 jul. 2023.

¹⁴ <https://coronavirus.jhu.edu/>. Acesso em: 27 jul. 2023.

A linguagem de programação *R* foi criada originalmente por *Ross Ihaka* e *Robert Gentleman* na década de 1990 na Universidade de *Auckland* na *Nova Zelândia*. O nome da linguagem deriva das iniciais dos criadores e de um jogo configurado com a *Linguagem S* (Escovedo; Koshiyama, 2020).

A linguagem é distribuída sob uma licença de *software* livre e de código-aberto chamada *GNU General Public License* (GPL)¹⁵. Esta licença garante aos usuários o direito de usar, modificar e distribuir o *software R*, bem como o acesso ao código-fonte. Segundo Amaral (2016), desde sua primeira versão em 1995, a linguagem tem evoluído nas comunidades de estatísticos e pesquisadores, tornando-se amplamente utilizada em Ciência de Dados e métodos científicos.

O site oficial para instalação do *software R* é o *CRAN-R* (*The Comprehensive R Archive Network* - Rede abrangente de arquivos *R*)¹⁶, cujo endereço eletrônico é mantido por uma comunidade de desenvolvedores em linguagem de programação *R*. Neste site, estão centralizadas informações importantes como eventos, manuais, guias de referência, versões, bibliotecas e outros links para fontes de suporte para a linguagem (Código Fonte TV, 2020).

Segundo *R (linguagem de programação)* (2023), a linguagem de programação *R* possui como principais características:

- a) *Software* gratuito, executado nos principais sistemas operacionais: *Linux*, *Windows®* e *macOS®*;
- b) Fornece uma extensa variedade de técnicas computacionais, matemáticas, e estatísticas tais como: testes estatísticos clássicos, análise de séries temporais, agrupamento, classificação, regressão, agrupamento e outras;
- c) Contém uma ampla coleção de pacotes, que são bibliotecas de funções, o ambiente de programação pode ser estendido, oferecendo uma variedade de recursos específicos para várias áreas de pesquisa;

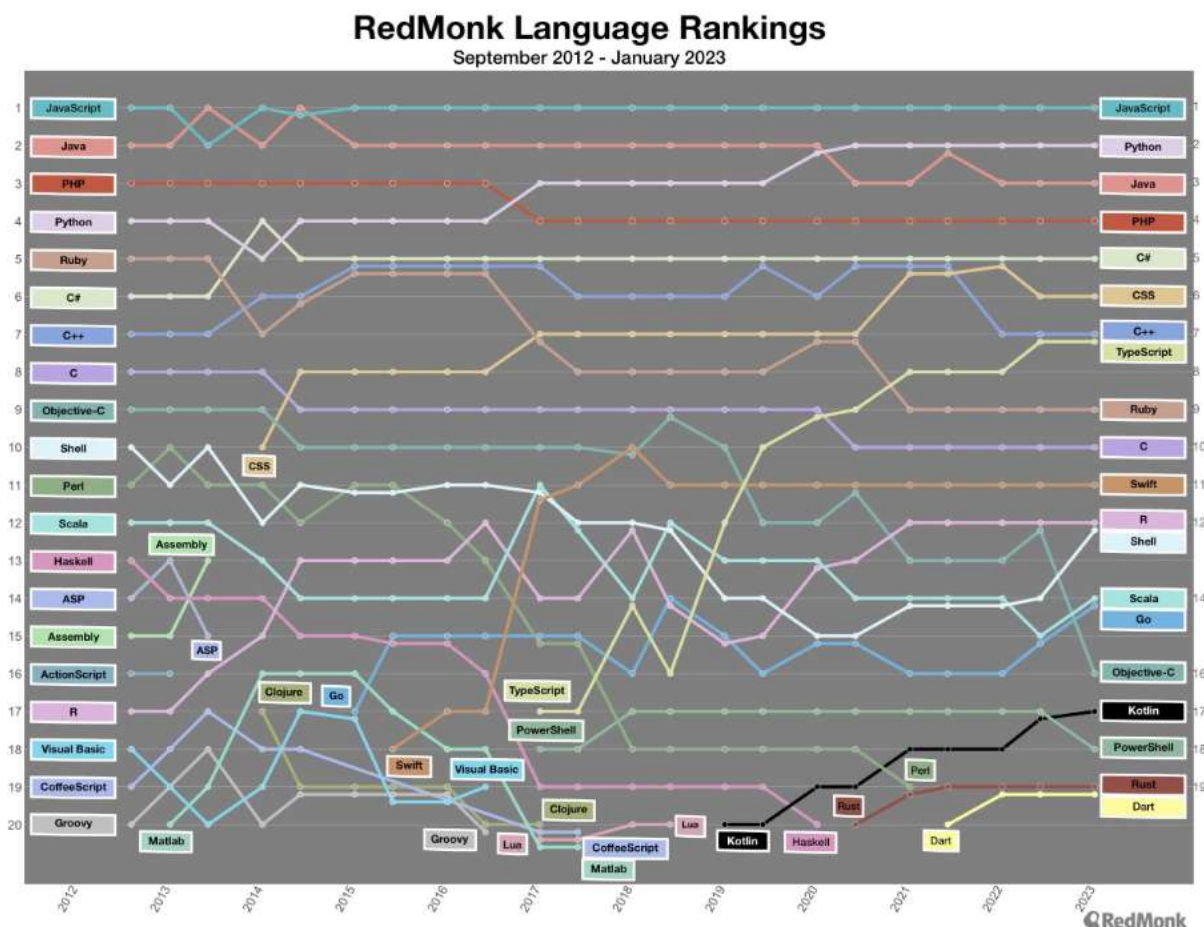
¹⁵ Representa uma designação da licença para *software* idealizada para o Projeto GNU de acordo com as definições de *software* livre da *Free Software Foundation*. Por ser uma licença copyleft (uma forma de usar a legislação de proteção dos direitos autorais com o objetivo de retirar barreiras à utilização, difusão e modificação de uma obra criativa), trabalhos derivados de um produto originalmente licenciado pela GPL só podem ser distribuídos se utilizarem a mesma licença. Fonte: *GNU General Public License* (2023).

¹⁶ O *CRAN-R* é uma rede de servidores espalhada pelo mundo que armazena versões idênticas e atualizadas de códigos e documentações para a linguagem de programação *R* e seu sítio está localizado em: <https://cran.r-project.org/>. Acesso em: 27 jul. 2023.

- d) Ferramenta que oferece a capacidade de produzir gráficos estáticos, dinâmicos e interativos. Além disso, possibilita a geração de imagens com alta qualidade para publicações profissionais e científicas;
- e) Linguagem muito utilizada nos estudos estatísticos e na área de análise de dados;

De acordo Stephens (2023), a empresa Redmonk¹⁷ elabora um ranking das 20 linguagens mais utilizadas no mundo, correlacionando o uso da linguagem e a quantidade de discussões e informações sobre elas na Internet. A linguagem de programação *R* ocupa a 12ª posição neste conceituado ranking, com informações atualizadas em janeiro de 2023. A Figura 3 demonstra o histórico do ranking de setembro de 2012 até janeiro de 2023.

Figura 3: RedMonk Top 20 Languages Over Time: January 2023.



Fonte: <https://redmonk.com/rstephens/2023/05/16/top20-jan2023/>. Acesso em: 28 jul. 2023.

¹⁷ <https://redmonk.com/>. Acesso em: 27 jul. 2023.

A linguagem de programação *R* tornou-se uma escolha comum para as pesquisas acadêmicas devido a uma série de razões. Principalmente, por ser gratuita, possibilitar funcionalidades para análise de dados e métodos estatísticos, permitir uma pesquisa reprodutível e produzir gráficos de qualidade. Por fim, sua ampla compatibilidade com outras ferramentas e tecnologias a torna uma escolha conveniente e versátil para pesquisadores (Sage Campus, 2019).

Posto isso, Matos (2015) citou as principais vantagens para a utilização da linguagem de programação *R*:

- a) Linguagem gratuita com sintaxe intuitiva;
- b) Possui uma grande variedade de bibliotecas e pacotes estatísticos disponíveis para facilitar a análise de dados de diversas áreas;
- c) Recursos de alta qualidade na geração de gráficos estáticos e interativos;
- d) Ambiente colaborativo de procedimentos e métodos, garantidos por uma comunidade mundial ativa de pesquisadores e estatísticos;
- e) Disponibiliza funções para execução de uma análise exploratória de dados e também para análises estatísticas complexas;
- f) Ótimo desempenho na execução de relatórios, tabelas, mapas e gráficos;
- g) Garante o princípio básico de reprodutibilidade dos métodos e procedimentos aplicados em uma pesquisa científica;
- h) Possibilita a integração com outros softwares: de controle de versão, de renderização de arquivos *HTML/PDF* e até mesmo com outras linguagens.

No entanto, de acordo com Matos (2015), existem aspectos importantes a serem analisados ao adotar a linguagem como solução:

- a) Falta de suporte comercial por ser um *software* livre;
- b) Curva de aprendizagem significativa, pois existem muitos métodos e funções possíveis a serem utilizados;

- c) Pré-processamento dos dados diretamente na memória *RAM* do computador, o que gera uma limitação em relação aos recursos do computador local;
- d) Necessidade de conhecimentos básicos de lógica de programação para se utilizar a linguagem.

As funções, dados ou coleções de códigos da linguagem de programação *R* estão disponíveis e organizadas nos chamados pacotes ou bibliotecas (*Packages*) e podem ser acessados e utilizados para estender a capacidade da linguagem.

A partir do momento em que o pacote for instalado no ambiente e posteriormente carregado para a memória do computador, estas funções do pacote ficam disponíveis para uso.

Os pacotes são desenvolvidos e mantidos pela Comunidade *R* e estão disponíveis para download no repositório *The Comprehensive R Archive Network* (CRAN)¹⁸ existente na *Internet* (Silva *et al.*, 2021).

Deste modo, existem muitos pacotes poderosos e populares na linguagem de programação *R* que podem ser utilizados nas etapas de uma AED. Neste trabalho foram utilizadas as bibliotecas *readr*¹⁹, *dplyr*²⁰, *ggplot2*²¹, *flexdashboard*²² e *shiny*²³, com suas principais funções resumidas nos parágrafos seguintes.

Na etapa de coleta de dados, o pacote *readr* pode ser utilizado para importar dados de arquivos texto de diferentes formatos, principalmente arquivos texto separados por vírgula chamados: *comma separated value* (CSV).

Uma das etapas mais importantes de uma análise de dados é a preparação dos dados, que consiste em selecionar, criar, alterar e resumir dados de forma eficiente. O pacote *dplyr*, chamado de gramática da preparação de dados, é amplamente usado nesta etapa.

O pacote *ggplot2* é um dos muitos pacotes utilizados para a criação de objetos gráficos na linguagem de programação *R*, sendo o mais versátil e elegante. A visualização dos dados torna-se importante, pois é a comunicação visual e

¹⁸ <https://cran.r-project.org/>. Acesso em: 27 jul. 2023.

¹⁹ <https://cran.r-project.org/web/packages/readr/index.html>. Acesso em: 27 jul. 2023.

²⁰ <https://cran.r-project.org/web/packages/dplyr/index.html>. Acesso em: 27 jul. 2023.

²¹ <https://cran.r-project.org/web/packages/ggplot2/index.html>. Acesso em: 27 jul. 2023.

²² <https://cran.r-project.org/web/packages/flexdashboard/index.html>. Acesso em: 27 jul. 2023.

²³ <https://cran.r-project.org/web/packages/shiny/index.html>. Acesso em: 01 jan.2024.

apresenta descobertas, novas interpretações e questionamentos sobre os dados demonstrados (Wickham; Grolemund, 2017).

Como discutido nos tópicos anteriores, o *dashboard* é uma maneira eficaz para comunicar e demonstrar uma grande quantidade de dados consolidados, de forma rápida e visual. O pacote *flexdashboard*, por sua vez, permite a criação de painéis por meio de elementos visuais como barras de rolagem, caixas de texto, objetos indicadores, abas e divisões para apresentação de gráficos, mapas, tabelas, imagens e demais objetos gráficos, resultando em uma experiência de visualização interativa e eficiente.

O pacote *shiny* disponibiliza instruções para a criação de aplicativos dinâmicos na Internet, com muitas possibilidades e interfaces customizáveis. Aplicações dinâmicas criadas com este pacote podem ser disponibilizadas na Internet, através de um serviço integrado ao *RStudio*® denominado *shinyapps.io*²⁴

Dentre os ambientes integrados existentes para desenvolvimento das rotinas da linguagem de programação *R*, o *RStudio*® é o que possui uma interface amigável, prática, simples e organizada para a utilização dos procedimentos e funções da linguagem afirma Noleto (2022).

O *RStudio*® é uma IDE - *Integrated Development Environment* - ou um ambiente integrado de desenvolvimento de programas, utilizado para o uso da linguagem de programação *R*. Este software livre foi desenvolvido e é mantido pela *Posit Software*²⁵.

O *RStudio*® possui uma interface gráfica amigável e tem como principal objetivo facilitar uma melhor utilização da linguagem de programação *R* por meio de uma console para execução de programas e demais funcionalidades para monitoramento, visualização de resultados e debug de prováveis erros (Noleto, 2022).

Na versão *Desktop*, isto é, para execução em computadores locais, este software está disponível gratuitamente e é chamado de *RStudio Desktop* ®. Também existe uma versão paga denominada *RStudio Desktop Pro*® (RSTUDIO, 2023).

²⁴ Plataforma integrada ao *RStudio*®, que possibilita hospedar aplicações interativas na Internet. <https://www.shinyapps.io/>. Acesso em: 20 fev. 2024

²⁵ <https://posit.co/>. Acesso em: 27 jul. 2023.

A ferramenta *RStudio Desktop*²⁶ pode ser executada nos sistemas operacionais: *Windows*®, *macOS*® e *Linux*, sendo projetada para facilitar o desenvolvimento de programas, a análise e a visualização de dados. Também, permite a integração com outros *softwares* como, por exemplo, o sistema de controle de alterações realizadas nos códigos, os *softwares* de renderização de documentos²⁷ PDF e HTML, dentre outros.

Além de ser o ambiente integrado mais utilizado e recomendado, Noletto (2022) cita:

O *RStudio*® é uma das versões de software mais desenvolvidas para a utilização de uma linguagem de programação como o *R*, pois ele oferece uma série de funcionalidades e painéis que podem contribuir grandemente para a análise de Big Data²⁸.

Conforme mostrado na Figura 4, o *RStudio Desktop*® possui em sua tela inicial 4 divisões, denominadas quadrantes ou painéis, que podem ser organizadas conforme a preferência dos usuários.

O painel **EDITOR** é onde são escritos os códigos da linguagem de programação *R*, também chamados de *scripts*. O *RStudio Desktop*® colore as palavras e símbolos para facilitar a leitura do código.

Já no painel **CONSOLE** são exibidos os resultados da execução dos scripts dos programas criados em linguagem de programação *R*. Qualquer erro, na execução destes códigos, é mostrado em forma de destaque neste painel.

O **OUTPUT** é um painel auxiliar que possui várias funções tais como: acessar pasta de arquivos (*Files*); mostrar objetos gráficos (*Plot*); exibir os pacotes instalados no ambiente e suas versões (*Packages*); acesso a uma importante ajuda da linguagem (*Help*) e a exibição de relatórios de dados (*Viewer*).

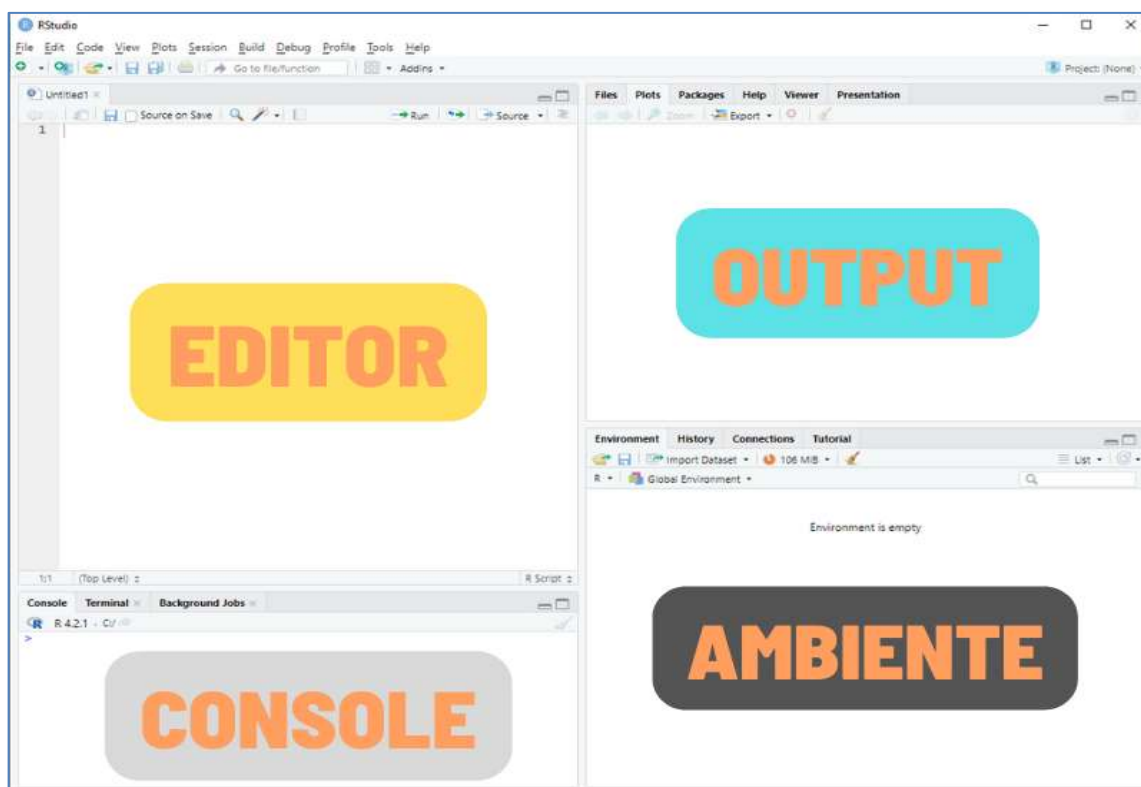
Outro painel auxiliar é o **AMBIENTE ou Environment**, muito importante na visualização de todos os objetos criados na execução do código da linguagem.

Figura 4: Tela inicial do *RStudio Desktop*®.

²⁶ O link para baixar a versão desktop é: <https://posit.co/download/rstudio-desktop/>. Acesso em: 27 jul. 2023.

²⁷ Renderização é o processamento para combinação de um material bruto digitalizado como imagens, vídeos ou áudio num único resultado final, unificando esses elementos com objetivo de melhorar a experiência do usuário. Fonte: O que é renderização ou render?, ControleNet, s/d. Disponível em: <<https://www.controle.net/faq/renderizacao-ou-render-de-video-audio-e-imagens-3d>>. Acesso em: 27 jul. 2023.

²⁸ Segundo Amaral (2016), trata-se de um fenômeno em que os dados são produzidos em vários formatos, grandes quantidades e armazenados por um grande número de dispositivos ou equipamentos.



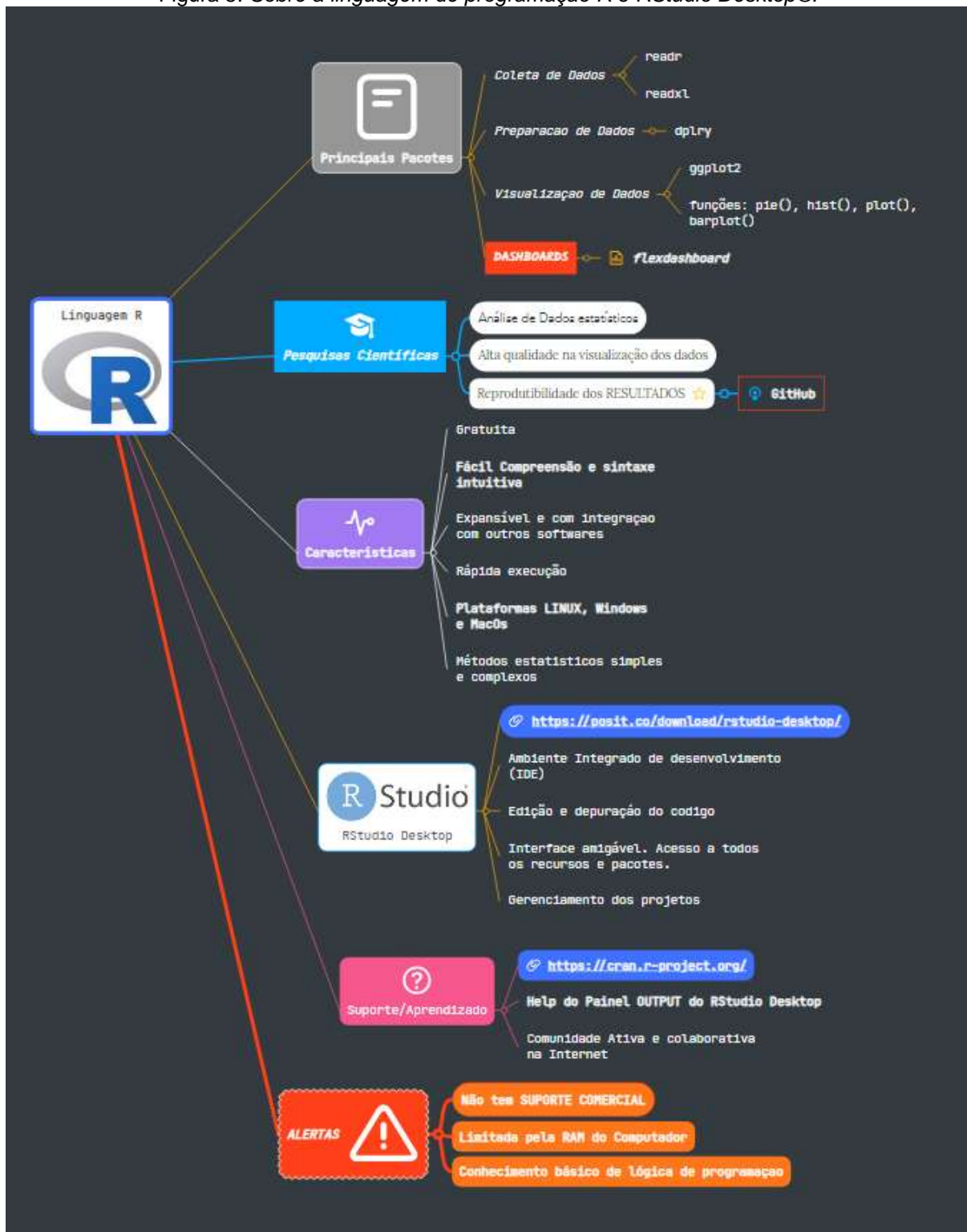
Fonte: Elaborada pelo autor.

Em busca de uma maior facilidade e apoio ao realizar uma análise de dados utilizando a linguagem de programação *R* e o *RStudio*®, Noletto (2022) destaca as seguintes vantagens:

- a) O *RStudio*® é compatível com os principais sistemas operacionais: *Windows*®, *macOS*® e *Linux*;
- b) A plataforma é um software gratuito e de código aberto, oferecendo muitas funcionalidades;
- c) Colabora com a facilidade de utilização da linguagem de programação *R* e possui interface amigável para interação com os objetos e métodos da linguagem;
- d) A comunidade que utiliza o *RStudio*® e a linguagem de programação *R*, é bastante ativa e colaborativa, disponibilizando documentações e muitas informações na *Internet*;
- e) Possui grande integração com demais softwares ou linguagens de programação no apoio às funcionalidades necessárias para as etapas de análise de dados e Ciência de dados.

A Figura 5 contém uma síntese das características, vantagens e alertas sobre a linguagem de programação *R* e o *RStudio Desktop*®.

Figura 5: Sobre a linguagem de programação *R* e *RStudio Desktop*®.



Fonte: Elaborada pelo autor

2.6 A plataforma *GitHub* e a reprodutibilidade do método científico

O *Git* é um software livre para controle de versões de arquivos. Seu principal objetivo é manter um registro das alterações feitas nestes arquivos ao longo do tempo, permitindo que os usuários acompanhem o histórico de alterações. Este software é um sistema de controle de versões distribuído e colaborativo, onde cada usuário que trabalha em um projeto no *Git* terá uma cópia completa do repositório desse projeto, incluindo os arquivos e os históricos de alterações (GIT, 2022).

O *Git* oferece uma série de vantagens para quem o utiliza: possibilita o controle das alterações, como desfazer ou recuperar uma versão do arquivo, permite a demonstração de um histórico das alterações, torna fácil documentar as alterações realizadas e traz confiança nas alterações desejadas. Essas vantagens tornam o *Git* uma ferramenta indispensável para o desenvolvimento de software moderno.

A utilização do repositório em *Git* propicia um ambiente colaborativo com a possibilidade de dividir partes do projeto entre os integrantes, com a finalidade de atuar em funcionalidades distintas que serão combinadas quando estiverem concluídas.

A interface online ou o site na *Internet* que disponibiliza um repositório *Git* é o *GitHub*²⁹. Este local centralizado permite carregar uma cópia do repositório *Git* e disponibiliza demais funcionalidades que possibilitam o monitoramento das alterações, as atividades de uma equipe e a colaboração ou compartilhamento de arquivos de um projeto.

Esta interface é utilizada mundialmente e chega a ter mais de 36 milhões de usuários ativos e milhões de projetos armazenados. As principais funcionalidades deste site são gratuitas (GitHub, 2022).

Além de ser um local centralizado para armazenar os repositórios *Git*, o *GitHub* apresenta demais vantagens como: documentar requisitos do projeto, criar linhas de desenvolvimento do mesmo projeto visando colaborar ou compartilhar o mesmo projeto para equipes distintas na implementação de diferentes funcionalidades e principalmente monitorar por meio da interface web, as alterações dos arquivos e o andamento dos projetos (Bell; Beer, 2015).

O *GitHub* é uma ferramenta que possibilita o armazenamento e organização de toda a documentação de um projeto científico em um repositório online *Git*, desta

²⁹ <https://github.com/>. Acesso em: 27 jul. 2023.

forma atendendo a reprodutibilidade como um princípio importante do método científico.

O *Git* (2022) comentou:

Cada diretório de trabalho do *Git* é um repositório com um histórico completo e habilidade total de acompanhamento das revisões, não dependente de acesso a uma rede ou a um servidor central. O *Git* também facilita a reprodutibilidade científica em uma ampla gama de disciplinas, da ecologia à bioinformática, arqueologia à zoologia.

Um dos princípios básicos do método científico é a reprodutibilidade da pesquisa. Neste sentido, a execução do estudo ou pesquisa deve ser passível de reprodução em diferentes momentos e por diversos pesquisadores, resultando em conclusões semelhantes. Para assegurar a reprodutibilidade, é fundamental a disponibilização de dados brutos e registros detalhados dos passos realizados no estudo, assim como os protocolos da pesquisa em repositórios de acesso público (Ciência Aberta USP, [s.d.]).

A reprodução dos resultados das pesquisas científicas e a maneira pelo qual esses resultados foram gerados são temas fundamentais para a produção das evidências mais robustas possíveis (Pilatti, 2019).

3 METODOLOGIA

Este capítulo apresenta a metodologia utilizada neste estudo, visando atingir os objetivos propostos. Cita as etapas e processos executados, desde a pesquisa teórica seguida por uma Análise Exploratória de Dados (AED), embasada por uma metodologia de Ciência de Dados aplicada para a apuração dos resultados e culminando no desenvolvimento de um painel visual denominado *dashboard*.

A IBM ([s.d]) destaca que a AED é principalmente utilizada para descobrir o que os dados podem revelar: seja uma característica, um teste de hipótese, um modelo ou um padrão. Seu resultado proporciona uma compreensão das variáveis e das relações dentro de um conjunto de dados. Essa análise é composta por funções técnicas estatísticas, computacionais, gráficas e matemáticas que auxiliam na solução de problemas em geral.

Escovedo e Koshiyama (2020) definem *Data Science* ou Ciência de Dados como um conjunto de métodos com o propósito de apoiar decisões de negócio baseadas em dados. De maneira mais abrangente, essa Ciência diz respeito à coleta de dados em grande quantidade, provenientes de diversos formatos e fontes, com o propósito de análise, com o objetivo de descobrir padrões, hipóteses ou comportamentos relevantes para análises estatísticas e preditivas, planejamentos e tomadas de decisão.

Este estudo utilizará como fonte de dados, um conjunto de dados do repositório de dados abertos do Ministério da Saúde *openDataSUS*. Além disso, empregará a linguagem de programação *R* e a plataforma *GitHub* para alcançar os resultados esperados.

Para realizar a AED, será seguida a estrutura definida na metodologia de Ciência de Dados apresentada por Siddiqi (2021).

Essa metodologia possui cinco etapas:

- a) A identificação do problema e a abordagem para resolvê-lo;
- b) A definição do método de coleta dos dados;
- c) A preparação e o processamento dos dados;
- d) A geração de modelos e finalmente a implantação.

A Figura 6 apresenta os principais elementos da metodologia definida por Siddiqi (2021).

Figura 6: Metodologia de Ciência de Dados.



Fonte: Elaborada pelo autor.

4 DESENVOLVIMENTO

Este estudo disponibiliza os códigos-fonte da linguagem de programação *R* para análise de dados abertos e o desenvolvimento do *dashboard*, acessíveis no repositório público *GitHub* : <https://github.com/DaInLab/PPGMiT-openDataSUS>.

A fonte de dados utilizada, juntamente com toda documentação auxiliar, está disponível no repositório do Ministério da Saúde do Brasil, *openDataSUS*, acessível no link: <https://opendatasus.saude.gov.br/dataset/srag-2021-a-2023>.

Para acessar o painel visual e dinâmico na *Internet*, o *dashboard*, segue o link : https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/.

4.1 Identificar o problema e a abordagem para resolvê-lo

O passo inicial se torna importante no entendimento do problema a resolver. Sendo importante responder as perguntas: o que vai resolver?; o que deu errado neste cenário?; o que vai acontecer a seguir?; o que eu preciso fazer? (Siddiqi, 2021).

Essa fase inicial, que representa o primeiro passo, é importante para a compreensão do problema em questão e para determinar o rumo a ser tomado.

Considerando a problemática abordada por este estudo, que envolve o cenário pandêmico no Brasil e a análise de dados para descobrir informações e conhecimentos, foi utilizado uma Análise Exploratória de Dados (AED) utilizando dados abertos juntamente com um *dashboard*. Toda essa escolha foi fundamentada na Metodologia de Ciência de Dados mencionada em Siddiqi (2021).

Sobre a AED foi utilizada a linguagem de programação *R* aplicada com funções de objetos visuais, análise de dados e métodos matemáticos e estatísticos como seleção, filtro, contagem, médias, operações aritméticas, classificação e *rankings*.

Quanto à análise dos dados nas atividades de coleta, preparação, processamento, visualização e desenvolvimento do *dashboard* foram utilizados os seguintes pacotes da linguagem: *readr*, *dplyr*, *ggplot2*, *flexdashboard*, *shiny* e demais pacotes auxiliares.

4.2 Definir métodos de coleta e requisitos

Nesta segunda etapa ocorre a definição sobre os dados que serão utilizados na resolução do problema, são os dados brutos. Sendo que se deve analisar o formato dos dados a serem coletados e quais dados específicos serão necessários. Torna-se importante também a forma de coleta e a tempestividade dos dados (Siddiqi, 2021).

O conjunto de dados “SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19” foi utilizado, conforme a Figura 7, como fonte de dados principal deste estudo.

Figura 7: Conjunto de dados do *openDataSUS*.



Fonte: Adaptação de SRAG 2021 A 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 – OPENDATASUS ([s.d.]).

Portanto, a temporalidade deste estudo, abrange o período pandêmico no Brasil, compreendendo os anos de 2021 a 2023.

SRAG 2021 A 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 – OPENDATASUS ([s.d.]) cita:

Esta página tem como finalidade disponibilizar o legado dos bancos de dados (BD) epidemiológicos de SRAG, da rede de vigilância da Influenza e outros vírus respiratórios, desde o início da sua implantação (2009) até os dias atuais (2023), com a incorporação da vigilância da covid-19. Atualmente, o sistema oficial para o registro dos casos e óbitos por SRAG é o Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe). Ressaltamos que os dados da vigilância de SRAG no Brasil disponibilizados nesta página, estão sujeitos a alterações decorrentes da investigação, ou mesmo correções de erros de digitação, pelas equipes de

vigilância epidemiológica que desenvolvem o serviço nas três esferas de gestão. Esclarece-se que as bases de dados de SRAG disponibilizadas neste portal passam por tratamento que envolve a anonimização, em cumprimento a Lei 13.709/2018.

Os recursos e requisitos para este estudo incluíram os conceitos da Ciência de Dados, da Análise Exploratória de Dados, do uso da linguagem de programação R e dos documentos auxiliares disponibilizados no site do *openDataSUS* conforme a Figura 8 na área de “Dados e Recursos”.

Figura 8: Dados e Recursos do conjunto de dados SRAG 2021 a 2023.

The screenshot displays the web interface for the SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19. The page is structured with a left sidebar and a main content area.

Left Sidebar:

- SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19**
- Seguidores:** 3
- Organização:**
 - Ministério da Saúde** (Logo: SUS Sistema Único de Saúde)
 - Não há descrição para essa organização
 - Redes Sociais:**
 - Instagram
 - Twitter
 - Facebook
 - Licença:** Creative Commons Attribution (CC BY)

Main Content Area:

- Tabs:** Conjunto de dados (selected), Grupos
- Section Header:** SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19
- Text:**

O Ministério da Saúde (MS), por meio da Secretaria de Vigilância em Saúde (SVS), desenvolve a vigilância da Síndrome Respiratória Aguda Grave (SRAG) no Brasil, desde a pandemia de Influenza A(H1N1) pdm09. A partir disso, a vigilância da SRAG foi implantada na rede de vigilância da Influenza e outros vírus respiratórios, que anteriormente atuava exclusivamente com a vigilância sentinela de Síndrome Gripal (SG).

Em 2020, a vigilância da COVID-19, a infecção humana causada pelo novo Coronavírus, que vem causando uma pandemia, foi incorporada na rede de vigilância da Influenza e outros vírus respiratórios.

Esta página tem como finalidade disponibilizar o legado dos bancos de dados (BD) epidemiológicos da SRAG, da rede de vigilância da Influenza e outros vírus respiratórios, desde o início da sua implantação (2009) até os dias atuais (2023), com a incorporação da vigilância da covid-19. Atualmente, o sistema oficial para o registro dos casos e óbitos por SRAG é o Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe).

O MS publicou o Guia de Vigilância Epidemiológica Emergência de Saúde Pública de Importância Nacional pela Doença pelo Coronavírus 2019 ainda estão disponíveis informações sobre definições de casos, critérios de confirmação e encerramento dos casos, dentre outros.

Resaltamos que os dados da vigilância de SRAG no Brasil disponibilizados nesta página, estão sujeitos a alterações decorrentes da investigação, ou mesmo correções de erros de digitação, pelas equipes de vigilância epidemiológica que desenvolvem o serviço nas três esferas de gestão.

Esclarece-se que as bases de dados de SRAG disponibilizadas neste portal passam por tratamento que envolve a anonimização, em cumprimento a Lei 13.709/2018.

Os dados de 2020 a 2023 são disponibilizados semanalmente, às quartas-feiras. Excepcionalmente, a atualização pode ocorrer em outro dia, de acordo com a funcionalidade do Sistema de Informação.
- Para mais informações, acesse:**
 - Gripe/Influenza - <https://www.gov.br/saude/pt-br/saude/saude-de-a-a-z/gripe-influenza>
 - COVID-19 - <https://www.gov.br/saude/pt-br/coronavirus>
 - Guia Nacional de Vigilância da COVID-19 - <https://www.gov.br/saude/pt-br/coronavirus/publicacoes-tecnicas/guias-e-planos/guia-de-vigilancia-epidemiologica-covid-19/view>
- Dados e recursos**
 - Folha de Notificação** [Explorar]
 - Dicionário de Dados** [Explorar]
 - SRAG 2021 - 01/06** [Explorar]
 - SRAG 2022 - 01/04** [Explorar]
 - SRAG 2023 - 28/01** [Explorar]

Fonte: Adaptação de SRAG 2021 A 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 – OPENDATASUS ([s.d.]).

A **Ficha de Notificação**³⁰ é o formulário principal de entrada de dados, deste conjunto de dados “SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19”, e consta no **ANEXO A** deste estudo. Este formulário é preenchido pelos funcionários do SUS para registro de cada paciente e contém 83 itens para preenchimento.

O **ANEXO B** contém o documento **Dicionário de Dados**³¹, onde são descritas e detalhadas todas variáveis existentes no conjunto de dados. Torna-se um requisito muito importante no entendimento dos dados, pois contém todo o detalhamento dos itens da Ficha de Registro Individual, existente no **ANEXO A**, consequentemente detalham os dados e as variáveis que são os dados brutos do conjunto de dados.

Os demais arquivos disponibilizados na área de “Dados e recursos” no repositório *openDataSUS* são os arquivos textos separados por vírgula (CSV) que contém os dados brutos anuais respectivos aos anos 2021, 2022 e 2023. Cada arquivo contém 173 colunas ou variáveis descritas no **ANEXO B** deste estudo. A Figura 9, exibe um trecho de um arquivo texto deste conjunto de dados.

Figura 9: Detalhes da pré-visualização do arquivo SRAG 2023 – 31/07.csv.

1	A	B	C	D	E	F	G	H	I	J	K	L
1	DT_NOTIFIC	SEM_NOT	DT_SIN_PRI	SEM_PRI	SG_UF	NCID_REGIOICO	REGIONAL	ID_MUNICCO_MUN	ID_UNIDADE		CO_UNI	PCS_SEXO
2	05/01/2023	1	02/01/2023	1	MS	NRS JARDI	5546 BELA VIST	500210	HOSPITAL SAO VICENTE DE PAULA		2376458	F
3	10/01/2023	2	07/01/2023	1	MG	DIVINOPC	1472 DIVINOPC	312230	UPA 24 HORAS PADRE ROBERTO CORDEIRO MARTINS		7469144	M
4	12/01/2023	2	07/01/2023	1	SC	FLORIANC	1476 FLORIANC	420540	HOSPITAL INFANTIL JOANA DE GUSMÃO		2091808	F
5	04/01/2023	1	02/01/2023	1	MG	BELO HOR	1449 BELO HOR	310620	HOSPITAL RISOLETA TOLENTINO NEVES		27863	M
6	05/01/2023	1	03/01/2023	1	PE	1	1497 RECIFE	201100	HOSPITAL BARÃO DE LUCENA		2427427	F
7	06/01/2023	1	04/01/2023	1	CE	1 CRES FO	1519 FORTALEZ	230440	HIAS HOSPITAL INFANTIL ALBERT SABIN		2563681	F

Fonte: Adaptação de SRAG 2021 A 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 – OPENDATASUS ([s.d.]).

As informações adicionais indicam características importantes da fonte de dados como a granularidade geográfica por município, a granularidade temporal por dia, a cobertura geográfica e a última atualização conforme mostra a Figura 10.

³⁰ Disponível em: <https://opendatasus.saude.gov.br/dataset/39a4995f-4a6e-440f-8c8f-b00c81fae0d0/resource/9f0edb83-f8c2-4b53-99c1-099425ab634c/download/ficha_srag_hospitalizado_19.09.2022.pdf>. Acesso em: 18 Fev. 2024.

³¹ Disponível em: <https://s3.sa-east-1.amazonaws.com/ckan.saude.gov.br/SRAG/pdfs/Dicionario_de_Dados_SRAG_Hospitalizado_19.09.2022.pdf>. Acesso em: 18 Fev. 2024.

Figura 10: Informações Adicionais do conjunto de dados SRAG 2021 a 2023.



Informações Adicionais

Campo	Valor
Autor	Datasus
Última Atualização	31 de janeiro de 2024, 18:00 (UTC-03:00)
Criado	11 de janeiro de 2022, 15:30 (UTC-03:00)
Cobertura Geográfica	Nacional
Cobertura Temporal	A partir da Semana Epidemiológica 01 de cada ano.
Contato	dadosabertos@saude.gov.br
Frequência Atualização	Semanal
Granularidade Geográfica	Município
Granularidade temporal	Dia
Referências	Classificação Brasileira de Ocupações (MTE), Tabela de Código de Municípios (IBGE), Cadastro Nacional de Estabelecimentos de Saúde (CNES).
Área Responsável	Coordenação-Geral do Programa Nacional de Imunizações (CGPNI/DEIDT/SVS/MS)

Fonte: Adaptação de SRAG 2021 A 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 – OPENDATASUS ([s.d.]).

A presença de três arquivos neste conjunto de dados, como ilustrado na Figura 11, totalizam 1,48 *Gigabytes* de tamanho. Essa grandeza representa um alerta ou uma preocupação para as próximas etapas, que envolveram as operações de preparação e processamento dos dados.

Figura 11: Os três arquivos texto do conjunto de dados.

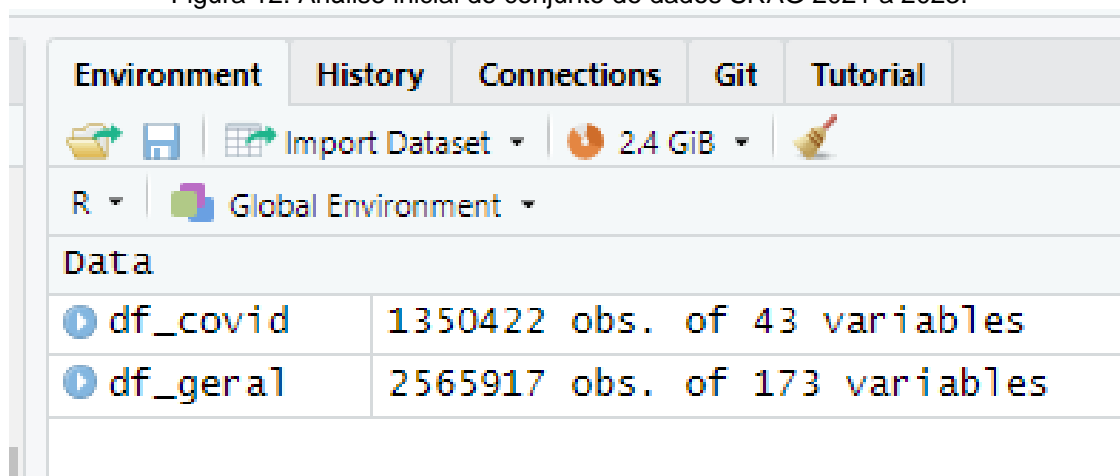
Nome	Data de modifi...	Tipo	Tamanho
 INFLUD21-01-05-2023.csv	28/01/2024 10:23	Arquivo de Valores Separados por Vírgulas do Microsoft Office Excel	1.000.462 KB
 INFLUD22-03-04-2023.csv	28/01/2024 10:22	Arquivo de Valores Separados por Vírgulas do Microsoft Office Excel	370.662 KB
 INFLUD23-29-01-2024.csv	18/02/2024 10:16	Arquivo de Valores Separados por Vírgulas do Microsoft Office Excel	190.693 KB

Fonte: Elaborada pelo autor.

Os arquivos foram obtidos manualmente a partir do *link* do repositório online e posteriormente copiados para o diretório “*dados*” no repositório do projeto no ambiente da linguagem de programação *R*. Para esta etapa foram utilizadas as funções nativas da linguagem e principalmente um pacote para importação de dados para o ambiente, o pacote *readr*.

Após uma análise inicial (Figura 12) utilizando a linguagem de programação *R* devido à vasta quantidade de dados contidos nos arquivos em questão, que totalizam aproximadamente 2,6 milhões de registros gerais com 173 variáveis cada e por volta de 1,4 milhões de registros com 43 variáveis relacionadas ao COVID-19, optou-se por realizar a granularidade e o agrupamento com base na localização geográfica da notificação e no ano correspondente, os quais serão obtidos a partir da data de notificação.

Figura 12: Análise inicial do conjunto de dados SRAG 2021 a 2023.



Data	
df_covid	1350422 obs. of 43 variables
df_geral	2565917 obs. of 173 variables

Fonte: Elaborada pelo autor.

4.3 Entender e preparar os dados

A compreensão dos dados gerados podem responder as perguntas elencadas no passo inicial desta metodologia (Siddiqi, 2021).

Esta etapa demanda mais tempo e esforço dentro da metodologia adotada, uma vez que envolve a compreensão e a delimitação conforme a granularidade e o entendimento inicial dos dados existentes.

Durante esta fase, são realizadas análises e operações como totalizações, seleção, filtragens, agrupamentos, tratamento de dados ausentes, além da criação de novos dados ou variáveis significativas a partir dos dados brutos. Todas estas atividades foram realizadas utilizando principalmente o pacote *dplyr*.

A primeira atividade realizada neste conjunto de dados, após a importação citada no tópico anterior, foi a seleção de registros por meio de condições ou filtros para delimitar e identificar os registros correspondentes ao estudo: os óbitos no contexto da COVID-19 no período de 2021 a 2023.

Os itens do **ANEXO A** que compõem o filtro necessário para este estudo, estão na área Conclusão, conforme a Figura 13.

Figura 13: Filtros, classificação final, evolução do caso e data de óbito.

Conclusão	75	Classificação final do caso: __ 1-SRAG por influenza 2-SRAG por outro vírus respiratório 3-SRAG por outro agente etiológico, qual 4-SRAG não especificado 5-SRAG por covid-19		
	77	Evolução do Caso: __ 1- Cura 2-Óbito 3-Óbito por outras Causas 9-Ignorado	78	Data da alta ou óbito: __ __ __

Fonte: Adaptação do **ANEXO A**.

De acordo com o **ANEXO A**, as condições que compõe o filtro aplicado neste conjunto de dados são:

- Para os casos da COVID-19. **75 Classificação do Caso** : valor 5-SRAG por COVID-19.
- Óbitos. **77 Evolução do Caso** : valor 2-Óbito.
- Período. **78 Data da alta ou óbito** : valor Anos 2021 a 2023

Ainda referente ao filtro, as variáveis necessárias para atender a principal condição, e que constam no **ANEXO B**,:

- CLASSI_FIN. 78-Classificação final do caso. Diagnóstico final do caso. Opções: 1-SRAG por influenza; 2-SRAG por outro vírus respiratório; 3-SRAG por outro agente etiológico, qual; 4-SRAG não especificado ou **5-SRAG por covid-19**. Tipo e tamanho da variável: um caractere;
- EVOLUCAO. 80-Evolução do caso. Evolução do caso. Opções: 1-Cura; **2-Óbito**; 3- Óbito por outras causas ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- DT_EVOLUCA. 81-Data da alta ou óbito. Tipo e tamanho da variável: data e formato DD/MM/AAA³².

³² Formado de data. Dia com dois dígitos: DD. Mês com dois dígitos: MM. Ano com quatro dígitos: AAAA.

Após a seleção dos dados para atender este estudo, segue a escolha das variáveis para compor os métodos de análise.

Foram selecionadas as seguintes variáveis para análise e validação dos dados: data de notificação, data de internação, data de evolução do óbito, data de encerramento, Unidade Federativa da unidade de saúde, informações gerais do paciente, informações do atendimento ou internação, sinais e sintomas, e fatores de risco ou comorbidades detalhados no atendimento.

Portanto, as variáveis que foram utilizadas neste estudo constam no **ANEXO B**, e são referentes a datas e localização:

- a) DT_NOTIFIC. 1-Data do preenchimento da ficha de notificação. Data de preenchimento da ficha de notificação. Tipo e formato da variável: data e formato DD/MM/AAA;
- b) DT_EVOLUCA. 81-Data da alta ou óbito. Data da alta ou óbito. Tipo e formato da variável: data e formato DD/MM/AAAA;
- c) SG_UF_NOT. 2-UF. Unidade Federativa onde está localizada a Unidade que realizou a notificação. Tipo e tamanho da variável: dois caracteres;
- d) ID_MUNICIP. 4-Município. Município onde está localizada a Unidade que realizou a notificação. Tipo e tamanho da variável: caracteres.

Para informações gerais, de acordo com o **ANEXO B**, as variáveis utilizadas:

- a) NU_IDADE_N. 13-(ou) Idade. Idade informada pelo paciente quando não se sabe a data de nascimento. Se digitado a data de nascimento, a idade é calculada e preenchida automaticamente pelo sistema: considerando o intervalo entre a data de nascimento e a data dos primeiros sintomas. Tipo e tamanho da variável: três caracteres;
- b) CS_SEXO. 11-Sexo. Sexo do paciente. Opções: 1- Masculino, 2- Feminino ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- c) CS_GESTANT. 14-Gestante. Idade gestacional do paciente. Opções: 1-1º Trimestre, 2-2º Trimestre, 3-3º Trimestre, 4-Idade Gestacional Ignorada, 5-Não, 6-Não se aplica ou 9-Ignorado. Tipo e tamanho da variável: um caractere;

- d) CS_ESCOL_N. 19-Escolaridade. Nível de escolaridade do paciente. Para os níveis fundamental e médio deve ser considerada a última série ou ano concluído. Opções: 0-Sem escolaridade/Analfabeto, 1-Fundamental 1º ciclo (1ª a 5ª série), 2-Fundamental 2º ciclo (6ª a 9ª série), 3- Médio (1º ao 3º ano) 4-Superior, 5-Não se aplica ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- e) CS_RACA. 15-Raça/Cor. Cor ou raça. Opções: 1-Branca, 2-Preta, 3-Amarela, 4-Parda, 5-Indígena ou 9-Ignorado. Tipo e tamanho da variável: um caractere.

A Figura 14 demonstra o item 34 do **ANEXO A**, que corresponde aos sinais e sintomas do paciente relatados na Ficha de Notificação.

Figura 14: Ficha de Notificação – Sinais e sintomas.

34	Sinais e Sintomas:	1-Sim	2-Não	9-Ignorado	<input type="checkbox"/> Febre	<input type="checkbox"/> Tosse	<input type="checkbox"/> Dor de Garganta	<input type="checkbox"/> Dispneia	<input type="checkbox"/>	
	Desconforto Respiratório	<input type="checkbox"/>	Saturação O ₂ < 95%	<input type="checkbox"/>	Diarreia	<input type="checkbox"/>	Vômito	<input type="checkbox"/>	Dor abdominal	<input type="checkbox"/>
	Fadiga	<input type="checkbox"/>	Perda do olfato	<input type="checkbox"/>	Perda do paladar	<input type="checkbox"/>	Outros			

Fonte: Adaptação do **ANEXO A**.

As variáveis correspondentes aos sinais e sintomas utilizadas, constam no **ANEXO B**, e são:

- a) FEBRE. 34-Sinais e Sintomas/Febre. Paciente apresentou febre?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- b) TOSSE. 34-Sinais e Sintomas/Tosse. Paciente apresentou tosse?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- c) GARGANTA. 34-Sinais e Sintomas/Dor de Garganta. Paciente apresentou dor de garganta?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- d) DISPNEIA. 34-Sinais e Sintomas/Dispneia. Paciente apresentou dispneia?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;

- e) DESC_RESP. 34-Sinais e Sintomas/Desconforto Respiratório. Paciente apresentou desconforto respiratório?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- f) SATURACAO. 34-Sinais e Sintomas/Saturação O₂< 95%. Paciente apresentou saturação O₂< 95%?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- g) DIARREIA. 34-Sinais e Sintomas/Diarreia. Paciente apresentou diarreia?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- h) VOMITO. 34-Sinais e Sintomas/Vômito. Paciente apresentou vômito?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- i) DOR_ABD. 34-Sinais e Sintomas/Dor abdominal. Paciente apresentou dor abdominal?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- j) FADIGA. 34-Sinais e Sintomas/Fadiga. Paciente apresentou fadiga?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- k) PERD_OLFT. 34-Sinais e Sintomas/Perda do Olfato. Paciente apresentou perda do olfato?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- l) PERD_PALA. 34-Sinais e Sintomas/Perda do Paladar. Paciente apresentou perda do paladar?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- m) OUTRO_SIN. 34-Sinais e Sintomas/Outros. Paciente apresentou outro(s) sintoma(s)?. Opções: 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- n) OUTRO_DES. 34-Sinais e Sintomas/Outros. Listar outros sinais. Tipo e tamanho da variável: trinta caracteres;

Conforme pode ser observado na Figura 15, o item 35 do **ANEXO A**, corresponde aos fatores de risco ou comorbidades do paciente relatados na Ficha de Notificação.

Figura 15: Ficha de Notificação – Fatores de Risco/comorbidades.

35	Possui fatores de risco/comorbidades? <input type="checkbox"/> 1-Sim 2-Não 9-Ignorado			Se sim, qual(is)? (Marcar X)
	<input type="checkbox"/> Puérpera (até 45 dias do parto)	<input type="checkbox"/> Doença Cardiovascular Crônica	<input type="checkbox"/> Doença Hematológica Crônica	
	<input type="checkbox"/> Síndrome de Down	<input type="checkbox"/> Doença Hepática Crônica	<input type="checkbox"/> Asma	
	<input type="checkbox"/> Diabetes mellitus	<input type="checkbox"/> Doença Neurológica Crônica	<input type="checkbox"/> Outra Pneumopatia Crônica	
	<input type="checkbox"/> Imunodeficiência/Imunodepressão	<input type="checkbox"/> Doença Renal Crônica	<input type="checkbox"/> Obesidade, IMC <input type="text"/>	
	<input type="checkbox"/> Outros <input type="text"/>			

Fonte: Adaptação do **ANEXO A**.

No **ANEXO B** constam as seguintes variáveis, as quais correspondem aos fatores de risco ou comorbidades:

- FATOR_RISC.** 35-Fatores de risco. Paciente apresenta algum fator de risco. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- PUERPERA.** 35-Fatores de risco/ Puérpera. Paciente é puérpera ou parturiente (mulher que pariu recentemente – até 45 dias do parto)?. Habilitado se selecionado no campo 8- Sexo Feminino. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- CARDIOPATI.** 35-Fatores de risco/ Doença Cardiovascular Crônica. Paciente possui Doença Cardiovascular Crônica?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- HEMATOLOGI.** 35-Fatores de risco/ Doença Hematológica Crônica. Paciente possui Doença Hematológica Crônica?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- SIND_DOWN.** 35-Fatores de risco/ Síndrome de Down. Paciente possui Síndrome de Down?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- HEPATICA.** 35-Fatores de risco/ Doença Hepática Crônica. Paciente possui Doença Hepática Crônica?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- ASMA.** 35-Fatores de risco/ Asma. Paciente possui Asma?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- DIABETES.** 35-Fatores de risco/ Diabetes mellitus. Paciente possui Diabetes mellitus?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;

- i) NEUROLOGIC. 35-Fatores de risco/ Doença Neurológica Crônica. Paciente possui Doença Neurológica?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- j) PNEUMOPATI. 35-Fatores de risco/ Outra Pneumopatia Crônica. Paciente possui outra pneumopatia crônica? Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- k) IMUNODEPRE. 35-Fatores de risco/ Imunodeficiência ou Imunodepressão. Paciente possui Imunodeficiência ou Imunodepressão (diminuição da função do sistema imunológico)?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- l) RENAL. 35-Fatores de risco/ Doença Renal Crônica. Paciente possui Doença Renal Crônica?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- m) OBESIDADE. 35-Fatores de risco/ Obesidade. Paciente possui obesidade?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- n) OBES_IMC. 35-Fatores de risco/ Obesidade (Descrição IMC). Valor do IMC (Índice de Massa Corporal) do paciente calculado pelo profissional de saúde. Habilitado se selecionado categoria 1-Sim em Fatores de risco/Obesidade. Tipo e tamanho da variável: três caracteres;
- o) OUT_MORBI. 35-Fatores de risco/ Outros. Paciente possui outro(s) fator(es) de risco?. Opções; 1-Sim, 2-Não ou 9-Ignorado. Tipo e tamanho da variável: um caractere;
- p) MORB_DESC. 35-Fatores de risco/ Outros (Descrição). Listar outro(s) fator(es) de risco do paciente. Tipo e tamanho da variável: trinta caracteres;

Finalmente, para o agrupar e sustentar os filtros deste estudo, são utilizadas as variáveis:

- a) DT_EVOLUC como data referência para o óbito, utilizado o ano deste dado no filtro do *dashboard*;

- b) Para a UF e município do óbito, foram utilizadas as variáveis de localização da notificação ou registro do formulário **ANEXO A** : SG_UF_NOT e ID_MUNICIP.

Concluindo esta etapa, os dados brutos foram selecionados e agrupados conforme o objetivo do estudo, e também contemplou a conferência dos dados em conformidade com os documentos auxiliares, que constam como anexos.

4.4 Gerar os modelos

Esta etapa consiste na criação de um modelo, que pode ser *descritivo*, onde demonstra um interesse, ou *preditivo*, demonstrando valores futuros com base em entradas. É importante avaliar se o modelo atingiu seus objetivos e permite também um refinamento (Siddiqi, 2021).

Para preparação e tratamento dos dados foi utilizado o pacote *dplyr*, para agrupar, ordenar, filtrar, calcular somatórios, rankings e percentuais dos dados.

Também foram utilizados os pacotes *flexdashboard*, *ggplot2*, *shiny* juntamente com outros pacotes fundamentais em métodos de visualização de dados, a fim de criar a estrutura das áreas ou setores do *dashboard* e as funcionalidades dinâmicas de filtros, gráficos, tabelas e totalizações dos dados.

Esses elementos compõem as informações apresentadas no *dashboard* sobre os óbitos no contexto da COVID-19 no período citado neste estudo.

4.5 Implantar o modelo

O modelo proposto é implantado para o uso restrito ou público nesta etapa (Siddiqi, 2021).

Esta etapa de implantação foi executada com duas atividades.

A primeira consistiu na geração de um documento utilizando principalmente os pacotes *flexdashboard* e *shiny* disponíveis na linguagem *R*, nos quais foram incluídos todas as informações, códigos-fonte desenvolvidos e objetos gráficos preparados e processados nas etapas anteriores.

Este arquivo contém o painel interativo, o *dashboard* acessível por meio de um navegador (*browser*) de *Internet*.

A segunda atividade envolveu a configuração e publicação deste documento na plataforma *shinyapps.io* utilizando o *RStudio*®. Isso permitiu que o objeto deste estudo fosse acessível publicamente na *Internet* por meio deste link³³.

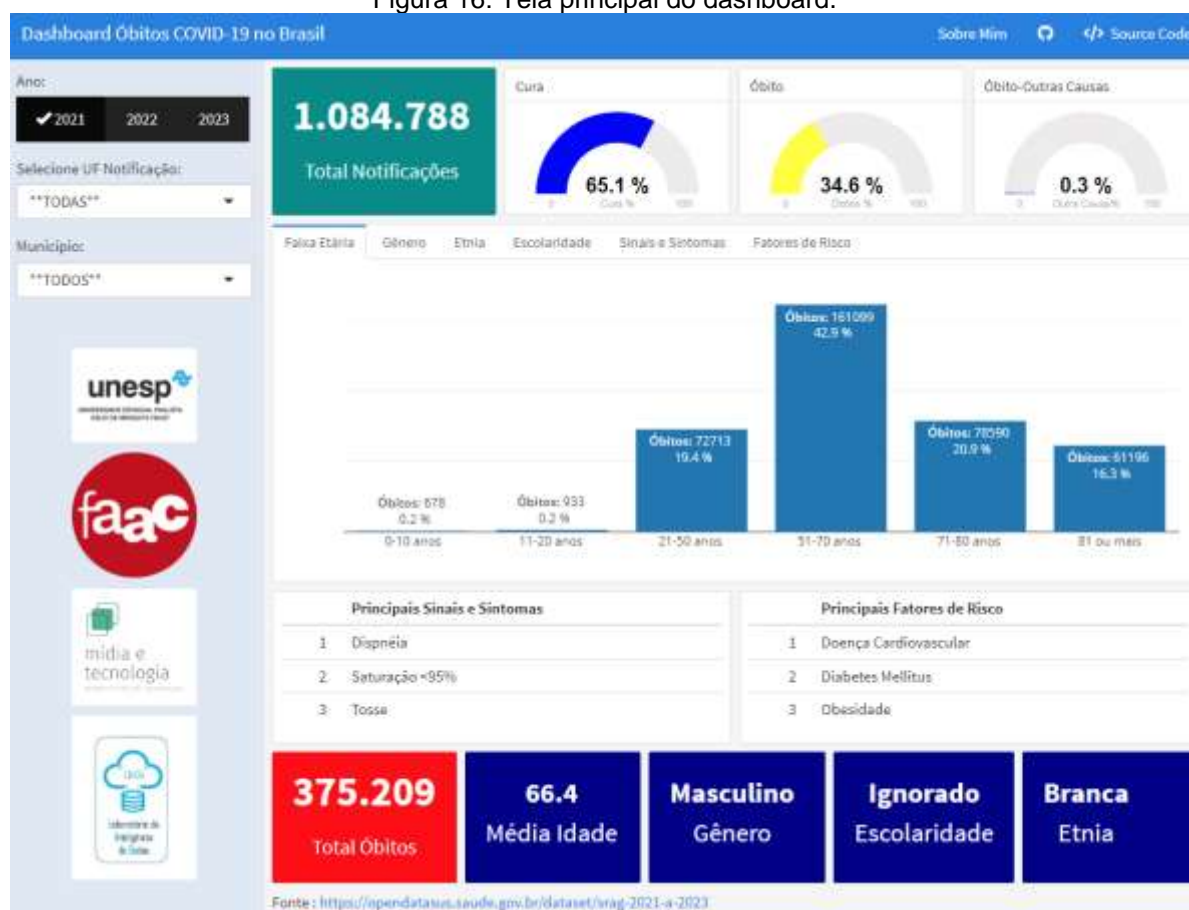
³³ Link de acesso público ao *dashboard*.
https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em: 18 Fev. 2024.

5 RESULTADOS ALCANÇADOS

O objetivo principal deste estudo é desenvolver um painel visual e dinâmico de acesso público com informações originadas de dados abertos. Neste capítulo, serão descritos os resultados obtidos ao longo do desenvolvimento do referido painel.

A Figura 16 ilustra a tela principal do *dashboard*, acessada de um link³⁴ na *Internet*. Esta tela é composta por várias áreas ou setores, incluindo cabeçalho, rodapé, filtro, informações de totais de notificações, gráficos, informações totais de óbitos e ranking dos principais fatores de riscos e principais sinais e sintomas.

Figura 16: Tela principal do dashboard.



Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

No **cabeçalho** (Figura 17), situado na parte superior do *dashboard*, são exibidos o título e três botões para acessos rápidos:

³⁴ https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em: 18 Feb. 2024.

- a) Sobre mim: link para informações sobre o autor do estudo;
- b) Ícone *Git*: link³⁵ para o acesso a toda documentação deste estudo no repositório GitHub;
- c) `</>` Source Code: exibe o código-fonte em linguagem *R* do *dashboard* em execução.

Figura 17: Cabeçalho do dashboard.



Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

Na parte inferior do painel, no **rodapé** (Figura 18), foi incluído um link para acessar a fonte de dados abertos utilizada neste estudo.

Figura 18: Rodapé do *dashboard*.



Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

À esquerda do *dashboard*, encontra-se uma área destinada ao **filtro** (Figura 19) das informações. Nessa área, são apresentadas opções visuais para selecionar o ano do óbito, o estado ou unidade federativa e o município da notificação.

Trata-se da área mais importante deste painel, pois conforme são alteradas estas opções as informações são exibidas automaticamente em todo o *dashboard*. Vale ressaltar que é obrigatório escolher um ano específico, enquanto que para as unidades federativas e municípios da notificação existe a opção de selecionar “TODOS”.

Também nesta seção são exibidas as identidades visuais da Universidade Estadual Paulista, da Faculdade de Arquitetura, Artes, Comunicação e Design – Câmpus de Bauru, do Programa de Pós-Graduação em Mídia e Tecnologia e do Grupo de Pesquisa LInDa.

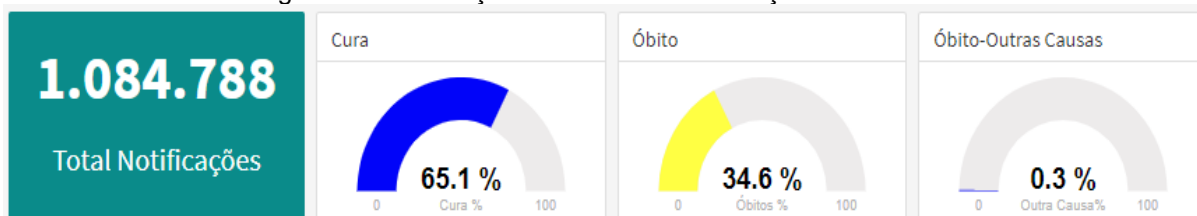
³⁵ <https://github.com/DaInLab/PPGMIT-openDataSUS.git>. Acesso em: 15 fev.2024.

Figura 19: Filtro do *dashboard*.

Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

Na área de **informações de totais de notificações** (Figura 20) são exibidos o total de notificações e os percentuais de classificação final dos casos COVID-19 como cura, óbito ou óbitos por outras causas para o ano, a unidade federativa e o município selecionados.

Figura 20: Informações de totais de notificações COVID-19



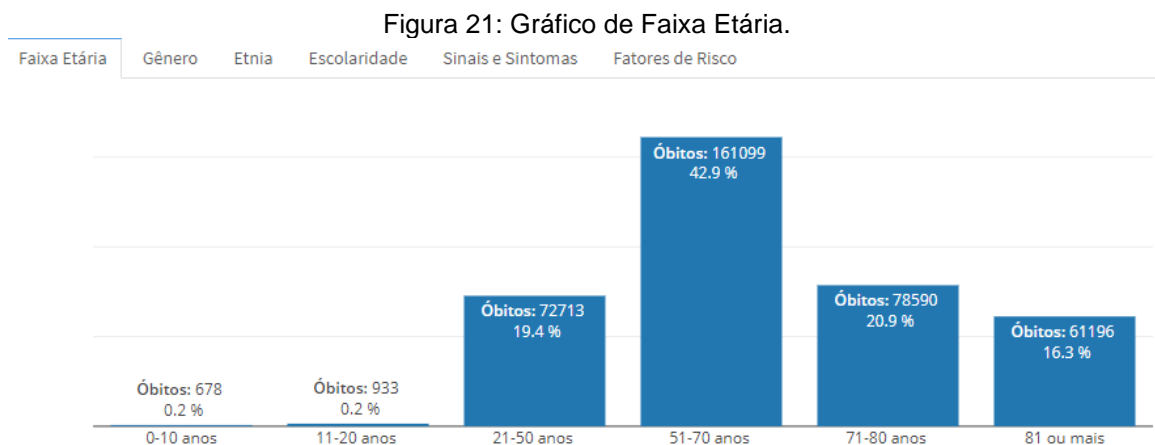
Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

Logo abaixo da área de informações de totais de notificações COVID-19, encontra-se a área de **gráficos** (Figura 21), que oferece acesso a seis gráficos distintos, organizados e acessados por meio de “abas”.

Nesta seção, em cada gráfico, na parte superior direita existe uma barra de ferramentas onde é possível ajustar o gráfico para visualização, excluir ou incluir alternativas da legenda e demais funcionalidades dinâmicas.

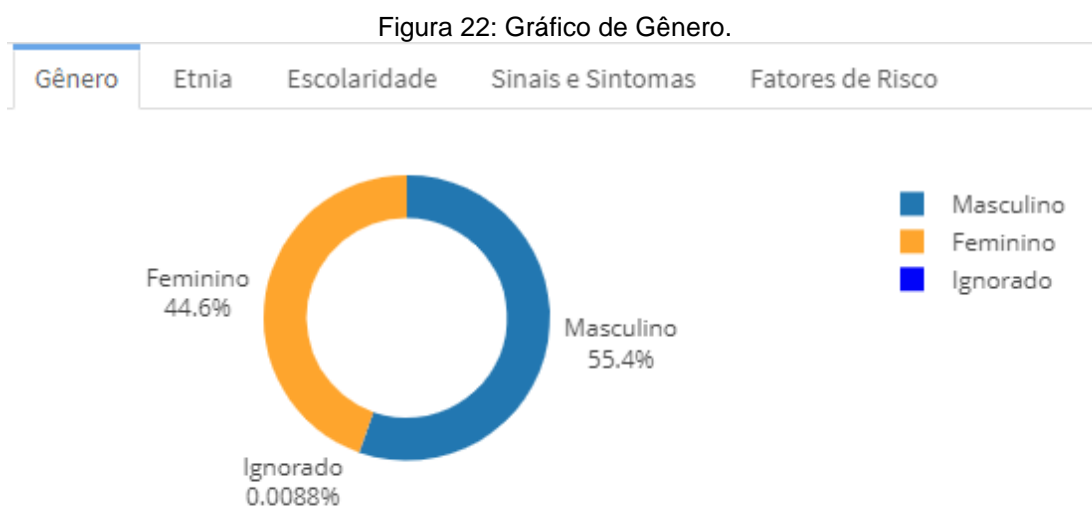
Esses gráficos exibem totalizadores e percentuais relacionados aos óbitos causados pela COVID-19, variando de acordo com a aba selecionada e as opções selecionadas nos filtros.

Ao iniciar o *dashboard*, é exibido um **gráfico de faixa etária** (Figura 21). Este gráfico representa a distribuição dos óbitos de acordo com as faixas etárias definidas.



Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

Na segunda aba, é exibido um **gráfico de gênero** (Figura 22). Este gráfico ilustra a distribuição dos casos de acordo com os gêneros masculino, feminino e ignorado, permitindo uma análise na disparidade de casos.



Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

Na terceira aba, é exibido um **gráfico de etnia** (Figura 23). Este gráfico apresenta a distribuição dos casos de acordo com diferentes grupos étnicos, demonstrando a diversidade étnica dos indivíduos afetados pela COVID-19.

Figura 23: Gráfico de Etnia.

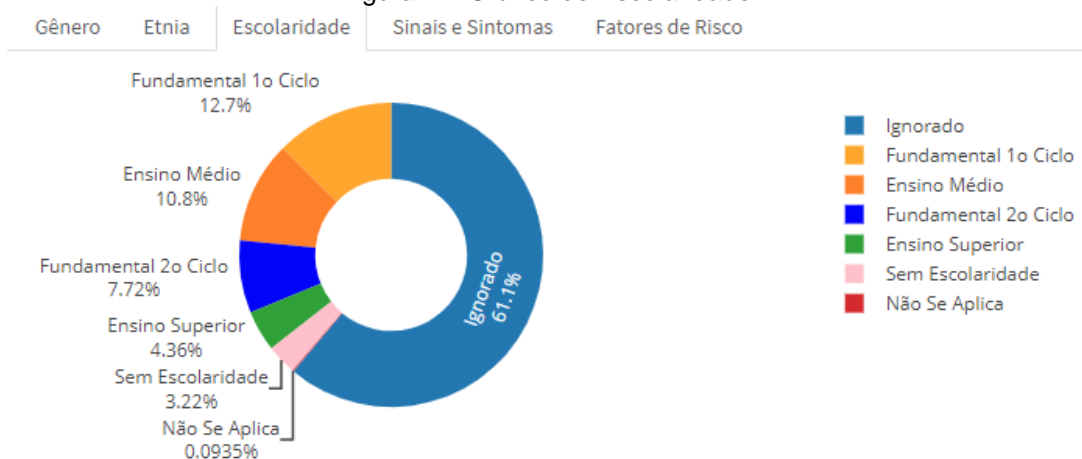


Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

É exibido um **gráfico de escolaridade** (Figura 24) na quarta aba da área de gráficos. Fornece informações sobre a relação entre a escolaridade e a incidência de óbitos COVID-19.

Ao analisar os dados encontrados neste gráfico, notou-se uma grande quantidade de dados “ignorado” para o item 19 da seção de “Dados do Paciente” que consta do **ANEXO A**, principalmente na totalização em determinadas unidades federativas.

Figura 24: Gráfico de Escolaridade.



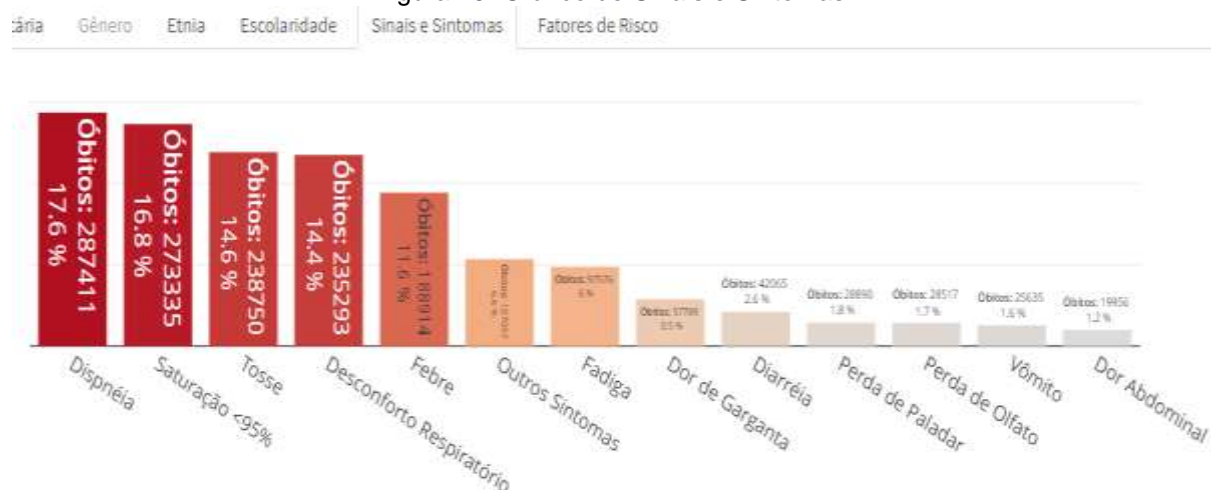
Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

Na quinta aba da área de gráficos, um gráfico importante é exibido, apresentando informações sobre óbitos e os **sinais e sintomas** (Figura 25) registrados nas notificações relacionadas à COVID-19.

Este gráfico é essencial para entender os padrões e as características clínicas associadas aos casos fatais da doença.

Dentre as alternativas para esta informação, consta no **ANEXO A** na seção de “Dados Clínicos e Epidemiológicos” no item 34 a alternativa "Outros".

Figura 25: Gráfico de Sinais e Sintomas.



Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 04 mar.2023.

Na sexta aba, encontra-se um gráfico que exhibe os **fatores de risco** (Figura 26) registrados nas notificações presentes no estudo. Esse gráfico oferece informações sobre os fatores que podem aumentar a mortalidade associada à COVID-19. Dentre as alternativas para esta informação, consta no ANEXO A na seção de “Dados Clínicos e Epidemiológicos” no item 35 a alternativa "Outros".

Figura 26: Gráfico de Fatores de Risco.



Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 04 mar.2023.

Nesta área de **ranking** (Figura 27), são exibidos os três principais sinais e sintomas, bem como os três principais fatores de risco totalizados neste estudo.

Esta seção destaca os principais elementos relevantes identificados na análise de dados. Fornece uma visão consolidada dos aspectos clínicos e epidemiológicos da COVID-19 consolidados.

Figura 27: Ranking de Sinais e Sintomas e Fatores de Risco.

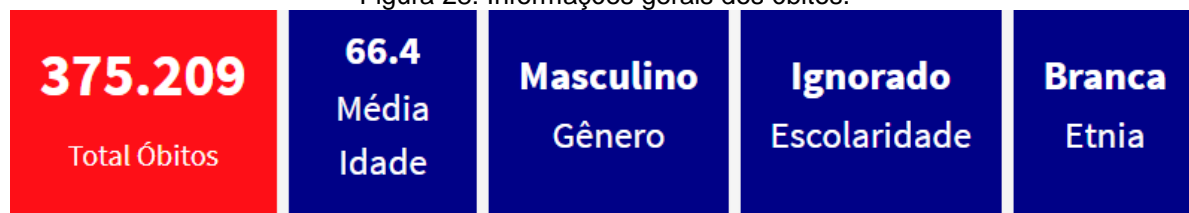
Principais Sinais e Sintomas		Principais Fatores de Risco	
1	Dispnéia	1	Doença Cardiovascular
2	Saturação <95%	2	Diabetes Mellitus
3	Tosse	3	Obesidade

Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 04 mar.2023.

Na parte inferior do *dashboard*, são exibidas as **informações gerais dos óbitos** (Figura 28) totalizadas e classificadas pela maioria de idade, gênero, escolaridade e etnia que resultaram no perfil dos óbitos por COVID-19, de acordo com o conjunto de dados estudado.

Estas estatísticas oferecem uma compreensão estratificada, permitindo a identificação de padrões associados aos impactos da pandemia na população nesta análise.

Figura 28: Informações gerais dos óbitos.



Fonte: https://ppgmitdashboardcovid.shinyapps.io/dashboard_covid_novo/. Acesso em 01 mar.2023.

6 CONSIDERAÇÕES FINAIS

A transformação de dados em informações relevantes oferece uma gama de oportunidades para promover a eficiência e a qualidade dos serviços públicos, além de estimular a inovação e o progresso social.

As estratégias mais eficazes para abordar fenômenos globais de saúde devem ser fundamentadas em Ciência, onde a apresentação de evidências baseadas em dados desempenha um elemento crucial na definição e planejamento. Essa abordagem é essencial para enfrentar de forma eficaz os desafios nesta área, como a pandemia e seus impactos, como é o caso da COVID-19.

A qualidade, a documentação, a atualização e o licenciamento dos dados abertos provenientes do repositório de dados abertos do Ministério da Saúde, *OpenDataSUS*, foram elementos fundamentais neste estudo, para assegurar a consistência das informações obtidas sobre a COVID-19. Isso resultou na valorização do conhecimento, proporcionando uma base sólida e confiável para embasar prováveis tomadas de decisões e promover a transparência.

O incentivo do governo brasileiro por meio de legislação específica sobre dados abertos, a promoção da dessa cultura entre os órgãos governamentais e o suporte para manter um portal com estes dados abertos, são alicerces que garantem transparência, precisão, integridade, acessibilidade e usabilidade. Por sua vez, tais ações governamentais contribuem para o desenvolvimento do país.

Uma das principais contribuições deste estudo reside no emprego da linguagem de programação *R* para auxiliar a metodologia de Ciência de Dados empregada, além de conduzir a Análise Exploratória de Dados.

A linguagem aqui utilizada oferece uma variedade de funções ou recursos computacionais, estatísticos, gráficos e matemáticos que facilitam o processamento, o tratamento e a visualização dos dados, desempenhando um papel essencial em operações complexas durante a análise.

Também destaca-se ainda a quantidade de funcionalidades disponíveis na linguagem para o desenvolvimento de gráficos de alta qualidade e *dashboards*.

Neste estudo, deve-se ressaltar o uso da plataforma *GitHub* utilizada como uma ferramenta fundamental para promover a documentação do projeto.

Ao proporcionar o armazenamento, o compartilhamento de dados e códigos-fonte e facilitar a documentação, a plataforma contribuiu para elevar a confiabilidade e reprodutibilidade do estudo.

Acredita-se que esta abordagem adotada possibilitará que outros pesquisadores reproduzam e validem os resultados obtidos.

O desenvolvimento do *dashboard* resultou em impactos notáveis, especialmente pela possibilidade de apresentação, visualização e simplificação do grande volume de dados em uma única tela. Essa tecnologia computacional e gráfica se revelou como um canal eficiente, potencializando não apenas a comunicação efetiva, mas também o acompanhamento contínuo dos indicadores ou métricas.

Esse estudo integrou tecnologias e métodos para disponibilizar informações públicas, confiáveis e acessíveis relacionadas à pandemia, com o objetivo de compreender e enfrentar melhor a crise sanitária por meio de dados assertivos.

Além disso, o painel visual pode colaborar para a compreensão das informações, ao mesmo tempo em que possibilita uma identificação mais clara e evidente de tendências, padrões e oportunidades para prováveis tomadas de decisões.

Como resultado desse estudo, foi demonstrado que a metodologia proposta contribui para a utilização de dados abertos e possui potencial para embasar possíveis processos decisórios de forma mais eficaz. É importante ressaltar que o painel e suas informações podem beneficiar pesquisadores e profissionais da área da saúde, bem como pode ser útil para gestores governamentais e o público em geral.

Além desse painel criado, o estudo, por meio da utilização da linguagem de programação *R* demonstrou uma gama de funcionalidades para o desenvolvimento computacional, a visualização de dados e os métodos estatísticos.

Durante o processo de análise dos dados, tornou-se evidente que o conjunto de dados contém uma quantidade de outras informações que podem ser integradas a outras fontes de dados abertos de diferentes áreas para enriquecer o painel em questão.

Deve-se salientar que, apesar da metodologia, da linguagem e da acessibilidade dos dados abertos, algumas dificuldades foram encontradas ao longo do estudo. Desde a limitação de recursos computacionais para processar um volume

grande de dados até a curva de aprendizado da linguagem *R* durante a análise exploratória dos dados e o desenvolvimento do *dashboard*.

Ao superar estes desafios, conseguiu-se adequar a metodologia e a análise dos dados, reconhecendo o valor do aprendizado obtido e sua contribuição para o progresso do conhecimento.

REFERÊNCIAS

- AGÊNCIA SENADO. **Covid desafia o Brasil a melhorar a qualidade e a transparência de dados**, 2020. Disponível em: <https://www12.senado.leg.br/noticias/infomaterias/2020/08/covid-desafia-o-brasil-a-melhorar-a-qualidade-e-a-transparencia-de-dados>. Acesso em: 19 jul. 2023.
- AGÊNCIA SENADO. **Desinformação e fake news são entraves no combate à pandemia, aponta debate**, 2021. Disponível em: <https://www12.senado.leg.br/noticias/materias/2021/07/05/desinformacao-e-fake-news-sao-entreve-no-combate-a-pandemia-aponta-debate>. Acesso em: 16 jun. 2023.
- AMARAL, F. **Introdução à Ciência de Dados: mineração de dados e big data**. Rio de Janeiro: Alta Books Editora, 2016.
- BELL, P.; BEER, B. **Introdução ao GitHub**. São Paulo: Novatec Editora, 2015.
- CASS, S. Top **Programming Languages 2022**. Disponível em: <https://spectrum.ieee.org/top-programming-languages-2022>. Acesso em: 23 jul. 2023.
- CIÊNCIA ABERTA USP. **Reprodutibilidade na ciência**. Disponível em: <https://cienciaaberta.usp.br/reprodutibilidade-na-ciencia/>. Acesso em: 12 jan. 2024.
- CNN BRASIL. **O painel da Johns Hopkins: A equipe envolvida no site mais visitado da pandemia**, 2020. Disponível em: <https://www.cnnbrasil.com.br/saude/o-painel-da-johns-hopkins-a-equipe-envolvida-no-site-mais-visitado-da-pandemia>. Acesso em: 30 jul. 2023.
- CÓDIGO FONTE TV. **R (Linguagem de Programação para Data Science) // Dicionário do Programador**, 2020. Disponível em: <https://www.youtube.com/watch?v=x4S9rIJHqxU>. Acesso em: 05 jun. 2023.
- ESCOVEDO, T.; KOSHIYAMA, A. **Introdução a Data Science: Algoritmos de Machine Learning e Métodos de Análise**. São Paulo: Casa do Código, 2020.
- FAVERO, L.P. et al. **Análise de Dados - Modelagem Multivariada para Tomada de Decisões**. Rio de Janeiro: Elsevier, 2009.
- FEW, Stephen. **Information dashboard design: displaying data for at-a-glance monitoring**. Burlingame, Califórnia: Analytics Press, 2013.
- FIOCRUZ. **Impactos sociais, econômicos, culturais e políticos da pandemia**. Disponível em: <https://portal.fiocruz.br/impactos-sociais-economicos-culturais-e-politicos-da-pandemia>. Acesso em: 16 jun. 2023.
- GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2008.

Git. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2022. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Git>. Acesso em: 27 jul. 2022.

GitHub. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2022. Disponível em: <https://pt.wikipedia.org/wiki/GitHub>. Acesso em: 17 jul. 2023.

GNU General Public License. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2023. Disponível em: https://pt.wikipedia.org/w/index.php?title=GNU_General_Public_License. Acesso em: 29 abr. 2023.

GOVERNO DIGITAL – DADOS ABERTOS. Disponível em: <https://www.gov.br/governodigital/pt-br/dados-abertos>. Acesso em: 17 jul. 2023.

GRUPO DE PESQUISA LINDA. Disponível em: <http://dgp.cnpq.br/dgp/espelhogrupo/379670>. Acesso em: 17 jul. 2023.

IBGE. Internet já é acessível em 90,0% dos domicílios do país em 2021 | Agência de Notícias, 2022. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/34954-internet-ja-e-acessivel-em-90-0-dos-domicilios-do-pais-em-2021>. Acesso em 23 jul. 2023.

IBM. O que é a análise exploratória de dados? Disponível em: <https://www.ibm.com/br-pt/topics/exploratory-data-analysis#:~:text=Ela%20ajuda%20a%20determinar%20a>. Acesso em: 17 jul. 2023.

INSTITUTO BUTANTAN. Além da Covid-19, enfrentamos outra epidemia: a de fakenews; saiba como se proteger desse “vírus”, 2022. Disponível em: <https://butantan.gov.br/bubutantan/alem-da-covid-19-enfrentamos-outra-epidemia-a-de-fake-news--saiba-como-se-proteger-desse-%E2%80%9Cvirus%E2%80%9D>. Acesso em: 16 jun. 2023.

JOHN HOPKINS UNIVERSITY. Johns Hopkins Coronavirus Resource Center, 2022. Disponível em: <https://coronavirus.jhu.edu/map.html>. Acesso em 23 jul. 2023.

MATOS, D. Uma Breve Introdução ao R. 2015. Disponível em: <https://www.cienciaedados.com/uma-breve-introducao-ao-r/>. Acesso em: 6 jun. 2023.

MINISTÉRIO DA SAÚDE. Informações Covid-19. Disponível em: <https://www.gov.br/saude/pt-br/coronavirus>. Acesso em: 16 jun. 2023.

MINISTÉRIO DA SAÚDE. Atendimento e fatores de risco, 2021. Disponível em: <https://www.gov.br/saude/pt-br/coronavirus/atendimento-tratamento-e-fatores-de-risco>. Acesso em: 01 ago. 2023.

NOLETO, Cairo. R Studio: o que é, como baixar, instalar e usar esse ambiente? TRYBE, 2022. Disponível em: <https://blog.betrybe.com/ferramentas/r-studio/>. Acesso em: 28 jul. 2023.

OBSERVATÓRIO DO FUTURO. **O impacto da pandemia nos ODS | Tribunal de Contas do Estado de São Paulo**, 2021. Disponível em: <https://www.tce.sp.gov.br/observatorio/impacto-pandemia-ods>. Acesso em: 27 jul. 2023.

ONU. **Sustainable Development Goal 3: Saúde e Bem-Estar | As Nações Unidas no Brasil**. Disponível em: <https://brasil.un.org/pt-br/sdgs/3>. Acesso em: 16 jun. 2023.

OPAS. **Histórico da pandemia de COVID-19 - OPAS/OMS | Organização Pan-Americana da Saúde**. Disponível em: <https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>. Acesso em: 16 jun. 2023.

OPAS. **Como estimar a mortalidade pela doença do novo coronavírus (COVID-19). Informe científico**, 2020. Disponível em: <https://iris.paho.org/handle/10665.2/52650#:~:text=Existem%20duas%20medidas%20usadas%20para,a%20todos%20os%20indiv%C3%ADduos%20infectados>. Acesso em: 01 ago. 2023.

OPAS. **Excesso de mortalidade associado à pandemia de COVID-19 foi de 14,9 milhões em 2020 e 2021 - OPAS/OMS | Organização Pan-Americana da Saúde**, 2022. Disponível em: <https://www.paho.org/pt/noticias/5-5-2022-excesso-mortalidade-associado-pandemia-covid-19-foi-149-milhoes-em-2020-e-2021>. Acesso em: 01 ago. 2023.

OPEN KNOWLEDGE FOUNDATION. Disponível em: <https://okfn.org/>. Acesso em: 30 jul. 2023.

OPENDATASUS. Disponível em: <https://opendatasus.saude.gov.br/>. Acesso em: 29 jul. 2023.

ORACLE. **O Que é uma Visualização de Dados?** Disponível em: <https://www.oracle.com/br/business-analytics/what-is-data-visualization/>. Acesso em: 30 jul. 2023.

PANDEMIA DE COVID-19 NO BRASIL. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2023. Disponível em: https://pt.wikipedia.org/w/index.php?title=Pandemia_de_COVID-19_no_Brasil. Acesso em: 25 jun. 2023.

PENG, Roger D. **Exploratory data analysis with R**. United States: Leanpub, 2016.

PILATI, Ricardo. **Reprodutibilidade: quando a ciência olha para dentro**. Disponível em: <https://revistaquestaodeciencia.com.br/artigo/2019/01/09/crise-de-reprodutibilidade-ciencia-olha-para-si-mesma>. 2019. Acesso em: 17 jul. 2023.

PORTAL DE DADOS ABERTOS – MINISTÉRIO DA SAÚDE. Disponível em: <https://dados.gov.br/dados/organizacoes/visualizar/ministerio-da-saude?idOrganizacao=5b283f30-ced3-4ccc-b44a-406e8a92e1ad&pagina=1>. Acesso em: 30 jul. 2023.

PORTAL G1. **Bauru tem 47 mortes de pacientes à espera de leito Covid em março, diz prefeitura**, 2021. Disponível em: <https://g1.globo.com/sp/bauru->

marilia/noticia/2021/03/30/bauru-registra-morte-de-47-pessoas-a-espera-de-leito-hospital-em-marco-diz-prefeitura.ghtml. Acesso em: 25 jul. 2023..

R (LINGUAGEM DE PROGRAMAÇÃO). In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2023. Disponível em: [https://pt.wikipedia.org/wiki/R_\(linguagem_de_programa%C3%A7%C3%A3o\)](https://pt.wikipedia.org/wiki/R_(linguagem_de_programa%C3%A7%C3%A3o)). Acesso em: 29 abr. 2023.

Rede de Pesquisa Solidária estuda aspectos sociais da pandemia, 2020. Disponível em: <https://jornal.usp.br/ciencias/rede-de-pesquisa-solidaria-acesse-as-ultimas-noticias/>. Acesso em: 25 jul. 2023.

RSTUDIO TEAM. **RStudio / Open source & professional software for data science teams**. Disponível em: <http://www.rstudio.com/>. Acesso em: 05 jul. 2023.

RSTUDIO. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2023. Disponível em: <https://pt.wikipedia.org/w/index.php?title=RStudio&oldid=65254805>. Acesso em: 26 jul. 2023.

SAGE CAMPUS. **Why universities are switching to R for teaching social science**, 2019. Disponível em: <https://campus.sagepub.com/blog/why-universities-are-switching-to-r-for-social-science>. Acesso em: 16 jun. 2023.

SIDDIQI, M. F. **Following the data science methodology**. IBM Developer Blog, 2021. Disponível em: <https://developer.ibm.com/blogs/following-the-data-science-methodology/>. Acesso em: 16 jul. 2023.

SILVA, Leandro Augusto da; PERES, Sarajane Marques; BOSCARIOLI, Clodis. **Introdução à Mineração de Dados Com aplicações em R**. 1. ed. Rio de Janeiro: LTC, 2021.

SIRIUS EDUCAÇÃO. **Tomada de decisão e como a análise de dados pode ajudar**. 2022. Disponível em: <https://blog.sirius.education/tomada-de-decisao/>. Acesso em: 31 jul. 2023.

SRAG 2021 A 2023 - BANCO DE DADOS DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE - INCLUINDO DADOS DA COVID-19 - OPENDATASUS. Disponível em: <https://opendatasus.saude.gov.br/dataset/srag-2021-a-2023>. Acesso em: 17 jan. 2024.

STEPHENS, R. **RedMonk Top 20 Languages Over Time: January 2023**, 2023. Disponível em: <https://redmonk.com/rstephens/2023/05/16/top20-jan2023/>. Acesso em: 29 jul. 2023.

TAXA DE MORTALIDADE. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2023. Disponível em: https://pt.wikipedia.org/w/index.php?title=Taxa_de_mortalidade&oldid=65621453. Acesso em: 2 ago. 2023.


The Comprehensive R Archive Network. Disponível em: <https://cran.r-project.org/>. Acesso em: 6 jun. 2023.

WICKHAM, Hadley; GROLEMUND, Garrett. ***R for Data Science: Import, Tidy, Transform, Visualize, and Model Data***. Beijing: O'Reilly, 2017.

WORLD HEALTH ORGANIZATION. ***World mental health report: Transforming mental health for all***, 2022. Disponível em: <https://www.who.int/publications/i/item/9789240049338>. Acesso em: 17 jul. 2023.

XP EDUCAÇÃO. **O que é dashboard? Conheça todos os tipos e saiba como fazer o seu!**, 2022. Disponível em: <https://blog.xpeducacao.com.br/o-que-e-dashboard/>. Acesso em: 30 jul. 2023.

ANEXO A – Ficha de Registro Individual – Casos de SRAG-Hospitalizado

Nº	
 <div> <div>MINISTÉRIO DA SAÚDE</div> <div>SECRETARIA DE VIGILÂNCIA EM SAÚDE</div> </div> <div> <div>SIVIG-Gripe - SISTEMA DE INFORMAÇÃO DE VIGILÂNCIA EPIDEMIOLÓGICA DA GRIPE</div> <div>FICHA DE REGISTRO INDIVIDUAL - CASOS DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE HOSPITALIZADO - 23/01/2021</div> </div>	
CASO DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE (SRAG-HOSPITALIZADO): indivíduo com *SG que apresenta: diagnóstico/desconforto respiratório OU pressão persistente no tórax OU saturação de O₂ menor que 95% em ar ambiente OU coloração azulada dos lábios ou rosto. (*SG: indivíduo com quadro respiratório agudo, caracterizado por pelo menos dois (2) dos seguintes sinais e sintomas: febre (mesmo que referida), calafrios, dor de garganta, dor de cabeça, tosse, coriza, distúrbios olfativos ou gustativos). Para efeito de notificação no Sivep-Gripe, devem ser considerados os casos de SRAG hospitalizados ou os óbitos por SRAG independente de hospitalização.	
1	Data do preenchimento da ficha de notificação:
2	Data de 1ºs sintomas
3	UF: 4 Município: Código (IBGE):
5	Unidade de Saúde: Código (CNES):
6	Tem CPF? (Marcar X) Sim Não
7	CPF:
8	Estrangeiro (Marcar X) Sim Não
9	Cartão Nacional de Saúde (CNS):
10	Nome:
11	SEXO: 1- Masc. 2- Fem. 9- Ign
12	Data de nascimento:
13	(Ou) Idade: 1-Dia 2-Mês 3-Ano
14	Gestante: 1-1º Trimestre 2-2º Trimestre 3-3º Trimestre 4-Idade Gestacional ignorada 5-Não 6-Não se aplica 9-Ignorado
15	Raça/Cor: 1-Branca 2-Preta 3-Amarela 4-Parda 5-Indígena 9-Ignorado
16	Se indígena, qual etnia?
17	É membro de povo ou comunidade tradicional? (Marcar X) Sim Não
18	Se sim, qual?
19	Escolaridade: 0-Sem escolaridade/Analfabeto 1-Fundamental 1º ciclo (1ª a 5ª série) 2-Fundamental 2º ciclo (6ª a 8ª série) 3-Médio (1ª ao 3º ano) 4-Superior 5-Não se aplica 9-Ignorado
20	Ocupação:
21	Nome da mãe:
22	CEP:
23	UF: 24 Município:
25	Bairro:
26	Logradouro (Rua, Avenida, etc.):
27	Nº:
28	Complemento (apto, casa, etc...):
29	(DDD) Telefone:
30	Zona: 1-Urbana 2-Rural 3-Perurbana 9-Ignorado
31	País: (se residente fora do Brasil) 1-Sim 2-Não 9-Ignorado
32	Trata-se de caso nosocomial (infecção adquirida no hospital)? 1-Sim 2-Não 9-Ignorado
33	Paciente trabalha ou tem contato direto com aves, suínos, ou outro animal? 1-Sim 2-Não 9-Ignorado
34	Sinais e Sintomas: 1-Sim 2-Não 9-Ignorado Febre Tosse Dor de Garganta Dispneia Desconforto Respiratório Saturação O ₂ < 95% Diarreia Vômito Dor abdominal Fadiga Perda do olfato Perda do paladar Outros
35	Possui fatores de risco/comorbidades? 1-Sim 2-Não 9-Ignorado Se sim, qual(is)? (Marcar X) Doença Cardiovascular Crônica Doença Hematológica Crônica Síndrome de Down Doença Hepática Crônica Asma Diabetes mellitus Doença Neurológica Crônica Outra Pneumopatia Crônica Imunodeficiência/Imunodepressão Doença Renal Crônica Obesidade, IMC Outros
36	Recebeu vacina COVID-19? 1-Sim 2-Não 9-Ignorado
37	Se recebeu vacina COVID-19, informar: Data da 1ª dose: Data da 2ª dose:
38	Laboratório Produtor vacina COVID-19:
39	Lote da vacina COVID-19: Lote 1ª Dose: Lote 2ª Dose:
40	Recebeu vacina contra Gripe na última campanha? 1-Sim 2-Não 9-Ignorado
41	Data da vacinação:
Se < 6 meses: a mãe recebeu a vacina? 1-Sim 2-Não 9-Ignorado Se sim, data: a mãe amamenta a criança? 1-Sim 2-Não 9-Ignorado Se >= 6 meses e <= 8 anos: Data da dose única 1/1: (dose única para crianças vacinadas em campanhas de anos anteriores) Data da 1ª dose: (1ª dose para crianças vacinadas pela primeira vez) Data da 2ª dose: (2ª dose para crianças vacinadas pela primeira vez)	

Dados de Atendimento	42	Usou antiviral para gripe? <input type="checkbox"/> 1-Sim 2-Não 9-Ignorado	43	Qual antiviral? <input type="checkbox"/> 1-Osetamivir 2-Zanamivir 3-Outro, especifique: _____	44	Data início do tratamento: _____
	45	Houve internação? <input type="checkbox"/> 1-Sim 2-Não 9-Ignorado	46	Data da internação por SRAG: _____	47	UF de internação: _____
	48	Município de internação: _____	Código (IBGE): _____			
	49	Unidade de Saúde de internação: _____	Código (CNES): _____			
	50	Internado em UTI? <input type="checkbox"/> 1-Sim 2-Não 9-Ignorado	51	Data da entrada na UTI: _____	52	Data da saída da UTI: _____
	53	Uso de suporte ventilatório: <input type="checkbox"/> 1-Sim, invasivo 2-Sim, não invasivo 3-Não 9-Ignorado	54	Raio X de Tórax: <input type="checkbox"/> 1-Normal 2-Infiltrado intersticial 3-Consolidação 4-Misto 5-Outro: _____ 6-Não realizado 9-Ignorado	55	Data do Raio X: _____
	56	Aspecto Tomografia: <input type="checkbox"/> 1-Típico covid-19 2-Indeterminado covid-19 3-Atípico covid-19 4-Negativo para Pneumonia 5-Outro: _____ 6-Não realizado 9-Ignorado	57	Data da tomografia: _____		
	58	Coletou amostra <input type="checkbox"/> 1-Sim 2-Não 9-Ignorado	59	Data da coleta: _____	60	Tipo de amostra: <input type="checkbox"/> 1-Secreção de Nariz-orofaringe 2-Livado Bronco-alveolar 3-Tecido post-mortem 4-Outra, qual? _____ 5-ICR 9-Ignorado
Dados Laboratoriais	61	Nº Requisição do GAL: _____		62	Tipo do teste para pesquisa de antígenos virais: <input type="checkbox"/> 1-Imunofluorescência (IF) 2-Teste rápido antígeno	
	63	Data do resultado da pesquisa de antígenos: _____		64	Resultado da Teste antígeno: <input type="checkbox"/> 1-positivo 2-Negativo 3-Inconclusivo 4-Não realizado 5-Aguardando resultado 9-Ignorado	
	65	Laboratório que realizou o Teste antígeno: _____		Código (CNES): _____		
	66	Agente Etiológico - Teste antígeno: Positivo para Influenza? <input type="checkbox"/> 1-Sim 2-Não 9-Ignorado Se sim, qual influenza? <input type="checkbox"/> 1-Influenza A 2-Influenza B Positivo para outros vírus? <input type="checkbox"/> 1-Sim 2-Não 9-Ignorado Se outros vírus respiratórios qual(is)? (marcar X) <input type="checkbox"/> SARS-CoV-2 <input type="checkbox"/> Vírus Sincicial Respiratório <input type="checkbox"/> Parainfluenza 1 <input type="checkbox"/> Parainfluenza 2 <input type="checkbox"/> Parainfluenza 3 <input type="checkbox"/> Adenovírus <input type="checkbox"/> Outro vírus respiratório, especifique: _____				
	67	Resultado da RT-PCR/outro método por Biologia Molecular: <input type="checkbox"/> 1-Detectável 2-Não Detectável 3-Inconclusivo 4-Não realizado 5-Aguardando resultado 9-Ignorado		68	Data do resultado RT-PCR/outro método por Biologia Molecular: _____	
	69	Agente Etiológico - RT-PCR/outro método por Biologia Molecular: Positivo para Influenza? <input type="checkbox"/> 1-Sim 2-Não 9-Ignorado Se sim, qual influenza? <input type="checkbox"/> 1-Influenza A 2-Influenza B Influenza A, qual subtipo? <input type="checkbox"/> 1-Influenza A(H1N1)pdm09 2-Influenza A(H3N2) 3-Influenza A não subtipado 4-Influenza A não subtipável 5-Inconclusivo 6-Outro, especifique: _____ Influenza B, qual linhagem? <input type="checkbox"/> 1-Victoria 2-Yamagata 3-Não realizado 4-Inconclusivo 5-Outro, especifique: _____ Positivo para outros vírus? <input type="checkbox"/> 1-Sim 2-Não 9-Ignorado Se outros vírus respiratórios, qual(is)? (marcar X) <input type="checkbox"/> SARS-CoV-2 <input type="checkbox"/> Vírus Sincicial Respiratório <input type="checkbox"/> Parainfluenza 1 <input type="checkbox"/> Parainfluenza 2 <input type="checkbox"/> Parainfluenza 3 <input type="checkbox"/> Parainfluenza 4 <input type="checkbox"/> Adenovírus <input type="checkbox"/> Metapneumovírus <input type="checkbox"/> Bocavírus <input type="checkbox"/> Rinovírus <input type="checkbox"/> Outro vírus respiratório, especifique: _____				
	70	Laboratório que realizou RT-PCR/outro método por Biologia Molecular: _____		Código (CNES): _____		
	71	Tipo de amostra sorológica para SARS-CoV-2: <input type="checkbox"/> 1-Sangue/plasma/soro 2-Outra, qual? _____ 9-Ignorado		72	Data da coleta: _____	
	73	Tipo de Sorologia para SARS-CoV-2: <input type="checkbox"/> 1-Teste rápido 2-Elisa 3-Quimioluminescência 4-Outro, qual? _____ Resultado do Teste Sorológico para SARS-CoV-2: <input type="checkbox"/> IgG <input type="checkbox"/> IgM <input type="checkbox"/> IgA 1-Positivo 2-Negativo 3-Inconclusivo 4-Não realizado 5-Aguardando resultado 9-Ignorado		74	Data do resultado: _____	
	Conclusão	75	Classificação final do caso: <input type="checkbox"/> 1-SRAG por Influenza 2-SRAG por outro vírus respiratório 3-SRAG por outro agente etiológico, qual _____ 4-SRAG não especificado 5-SRAG por covid-19			76
77		Evolução do Caso: <input type="checkbox"/> 1-Cura 2-Óbito 3-Óbito por outras causas 9-Ignorado			78	Data da alta ou óbito: _____
80 Número D.O: _____						
81 OBSERVAÇÕES: _____						
82 Profissional de Saúde Responsável: _____				83 Registro Conselho/Matricula: _____		

Fonte: Ficha de Notificação SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 – OPENDATASUS ([s.d.]).

ANEXO B – Dicionário de dados – Ficha de Registro Individual



Dicionário de Dados

FICHA DE REGISTRO INDIVIDUAL – CASOS DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE HOSPITALIZADOS

Este documento tem como finalidade descrever as variáveis exportadas para o banco de dados em DBF.

CAMPO OBRIGATÓRIO é aquele cuja ausência de dado impossibilita a inclusão do registro no sistema;
CAMPO ESSENCIAL é aquele que, apesar de não ser obrigatório, registra dado necessário à investigação do caso ou ao cálculo de indicador epidemiológico ou operacional;
CAMPO INTERNO é aquele que apesar de não constar na ficha e não aparecer no display da tela, é preenchido automaticamente pelo sistema;
CAMPO OPCIONAL é aquele que só deve ser preenchido caso seja necessário, aparece no display da tela e consta no banco de dados.

Nome do campo	Tipo	Categoria	Descrição	Características	DBF
NR	Varchar(20)		Número do registro	Campo Interno Número sequencial gerado automaticamente pelo sistema. Utilizar o padrão: 3202.200002.22 Dígito 1: caracteriza o tipo da ficha (3=SG, 2=SRAG-UTI e 5=SRAG Hospitalizado). Dígitos 2 a 12: número sequencial gerado automaticamente pelo sistema.	NU_NOTIFIC
3-Data de preenchimento da ficha de notificação	Date DD/MM/AAAA		Data de preenchimento da ficha de notificação.	Campo Obrigatório Data deve ser <= a data de digitação.	DT_NOTIFIC
Semana Epidemiológica do preenchimento da ficha de notificação	Varchar(20)		Semana Epidemiológica do preenchimento da ficha de	Campo Interno Calculado a partir da data dos Primeiros Sintomas. (SS)	SEM_NOT

SIVEP-Gripe: Sistema de Informação da Vigilância Epidemiológica da Gripe. Página 1

2-Data de 1ºs sintomas	Date DD/MM/AAAA		notificação.	Campo Obrigatório Data deve ser <= a data de digitação e data do preenchimento da ficha de notificação.	DT_1º_SIN_PRI
Semana Epidemiológica dos Primeiros Sintomas	Varchar(20)		Semana Epidemiológica do início dos sintomas.	Campo Interno Calculado a partir da data dos Primeiros Sintomas. (SS)	SEM_PRI
3-UF	Varchar(2)	Tabela com código e siglas das UF padronizadas pelo IBGE.	Unidade Federativa onde está localizada a unidade que realizou a notificação.	Campo Obrigatório Se usuário que está digitando a ficha for de nível: • Unidade – o campo é preenchido automaticamente pelo sistema com a UF, município e unidade onde está cadastrado o usuário. • Municipal – o campo é preenchido automaticamente pelo sistema com a UF e município onde está cadastrado o usuário. • Estadual – o campo é preenchido automaticamente pelo sistema com a UF do usuário. • Federal – abre tabela com todas as UF que possuem unidades cadastradas no sistema.	SG_UF_NOT
4-Município (Código (IBGE))	Varchar(20)	Tabela com código e nomes dos Municípios padronizados pelo IBGE.	Município onde está localizada a unidade que realizou a notificação.	Campo Obrigatório Preenchendo o nome do município de notificação, o código é preenchido automaticamente, e vice-versa; Se usuário que está digitando a ficha for de nível: • Unidade – o campo é preenchido automaticamente pelo sistema com o Município onde está localizada a unidade de notificação. • Municipal – o campo é preenchido automaticamente pelo sistema com o município do usuário. • Estadual ou Federal – abre tabela com todos os municípios da UF selecionada no campo 3 que possuem unidades cadastradas no sistema.	ID_MUNICIP OU CO_MUN_NOT
Regional de Saúde de Notificação (Código (IBGE))	Varchar(20)	Tabela com código e nomes das Regionais de Saúde dos municípios de notificação padronizados pelo IBGE.	Regional de Saúde onde está localizada o Município realizou a notificação.	Campo Interno Preenchendo o nome da regional de saúde de notificação, o código é preenchido automaticamente, e vice-versa; Se usuário que está digitando a ficha for de nível:	ID_REGIONAL OU CO_REGIONAL

SIVEP-Gripe: Sistema de Informação da Vigilância Epidemiológica da Gripe. Página 2

				<ul style="list-style-type: none"> Unidade – o campo é preenchido automaticamente pelo sistema com a Regional do Município onde está localizada a unidade de notificação. Município – o campo é preenchido automaticamente pelo sistema com a Regional do município do usuário. 	
5- Unidade de Saúde (Código CNES)	Varchar(27)	Tabela com códigos CNES e nomes das Unidades cadastradas no sistema.	Unidade que realizou o atendimento, coleta de amostra e registro do caso.	Campo Obrigatório Preenchendo o nome da unidade, o código é preenchido automaticamente, e vice-versa. Se usuário que está digitando a ficha for de nível: <ul style="list-style-type: none"> Unidade – o campo é preenchido automaticamente pelo sistema. Município – abre tabela apenas com as unidades do município. Estado do Brasil – abre tabela com as unidades do município selecionado o campo 4. 	ID_UNIDADE OU CD_UNE_NOT
6- Tem CPF?	Varchar(1)	1- Sim 2- Não	Informar se o paciente notificado possui o número do Cadastro de Pessoa Física (CPF).	Campo Obrigatório Se selecionado "Sim", preencher campo "CPF". Se selecionado "Não" preencher CNES. Se o paciente não tiver de CPF é obrigatório o preenchimento do CNES. No caso de pacientes raça/etnia indígenas, somente o CNES é considerado como campo obrigatório.	TEM_CPF
7- CPF do paciente	Varchar(15)	Número (11 dígitos)	Número do Cadastro de Pessoa Física (CPF) do paciente notificado.	Campo Obrigatório Quando preenchido o número do CPF o sistema deverá preencher o Nome, Sexo, Data de Nascimento, Idade, Raça/Cor e o nome da mãe do paciente.	NU_CPF
8- Estrangeiro	Varchar(1)	1- Sim 2- Não	Informar se o paciente é estrangeiro.	Campo Obrigatório Se selecionado "Sim", o campo CPF e CNES, deixa de ser obrigatório.	ESTRANG
9- Cartão Nacional de Saúde (CNS)	Varchar(15)	Número (14 dígitos)	Preencher com o número do Cartão Nacional de Saúde do paciente.	Campo Obrigatório	NU_CNS
10- Nome	Varchar(70)		Nome completo do paciente (sem abreviações).	Campo Obrigatório	NM_PACIENT
11- Sexo	Varchar(1)	1- Masculino	Sexo do paciente.	Campo Obrigatório	CS_SEXO

INVSF (Sipre - Sistema de Informação de Vigilância Epidemiológica da Dengue) – Página 3

		2- Feminino 3- Ignorado			
12- Data de nascimento	Varchar(10)	DD/MM/AAAA	Data de nascimento do paciente.	Campo Essencial Data deve ser <= a data dos primeiros sintomas.	DT_NASC
13- (ou) Idade	Varchar(3)		Idade informada pelo paciente quando não se sabe a data de nascimento. Na falta dessa idade é registrada a idade aparente.	Campo Obrigatório Se digitado a data de nascimento, a idade é calculada e preenchida automaticamente pelo sistema, considerando o intervalo entre a data de nascimento e a data dos primeiros sintomas. Idade deve ser <= 150.	NU_IDADE_R
13- (ou) Tipo/Idade	Varchar(2)	1- Dia 2- Mês 3- Ano		Campo Obrigatório Se digitado a data de nascimento, o campo idade/tipo é calculado e preenchido automaticamente pelo sistema, considerando o intervalo entre a data de nascimento e a data dos primeiros sintomas. Se a diferença for de 0 a 30 dias, o sistema grava em idade = (int dias) e em Tipo = 1- Dia. Por exemplo: se Data de nascimento = 05/12/2012 e Data dos 1ºs sintomas = 15/12/2012, então Idade = 6 e Tipo = 1- Dia. Se a diferença for de 3 a 11 meses, o sistema grava em idade = (int meses) e em Tipo = 2- Mês. Por exemplo: se Data de nascimento = 05/10/2012 e Data dos 2ºs sintomas = 15/12/2012, então Idade = 2 e Tipo = 2- Mês. Se a diferença for maior ou igual a 12 meses, o sistema grava em idade = (int anos) e em Tipo = 3- Ano. Por exemplo: se Data de nascimento = 05/10/2011 e Data dos 3ºs sintomas = 15/12/2012, então Idade = 1 e Tipo = 3- Ano.	TP_IDADE
14- Gestante	Varchar(1)	1- 1º Trimestre 2- 2º Trimestre 3- 3º Trimestre 4- Idade Gestacional 5- Não 6- Não se aplica 7- Ignorado	Idade gestacional da paciente.	Campo Obrigatório Se selecionada categoria 2- Feminino no campo Sexo. Se selecionado sexo igual a Masculino ou a idade for menor ou igual a 9 anos o campo é preenchido automaticamente com 6- Não se aplica. Se selecionado sexo igual a Ignorado e idade for maior que 9 anos, o campo não pode ser preenchido com 6- Não se aplica.	CS_GESTANT
15- Raça/Cor	Varchar(2)	1- Branco	Cor da raça	Campo Obrigatório	CS_RACA

INVSF (Sipre - Sistema de Informação de Vigilância Epidemiológica da Dengue) – Página 4

		2-Preto 3-Amarelo 4-Pardo 5-Indígena 9-Ignorado	declarado pelo paciente: Branco, Preto, Amarelo, Pardo (pessoa que se declarou mulato, caboclo, cabula, mestiça ou mestiça de preto com pessoa de outra cor ou raça); e, indígena.		
16- Se indígena, qual etnia?	Varchar(20)	Tabela de SIA/S com código e nomes das etnias indígenas.	Nome e código da etnia do paciente, quando indígena.	Campo Essencial Habilitado se campo 15-Raça/Cor for igual a 5-Indígena.	CS_ETHNA
17- É membro de povo ou comunidade tradicional?	Varchar(20)	1-Sim 2-Não	Informar se o paciente foi membro de algum povo ou comunidade tradicional	Campo Obrigatório	POV_CT
18- Se sim, qual?	Varchar(100)	Tabela de Povo e Comunidades Tradicionais	Informar o povo ou comunidade tradicional	Campo Obrigatório - Habilitado se campo 17- É membro de povo ou comunidade tradicional? for igual a 1- Sim	TP_POV_CT
19- Escolaridade	Varchar(3)	0-Sem escolaridade/ Alfabeto 1-Fundamental 1ª série (2ª a 3ª série) 2-Fundamental 2ª série (3ª a 4ª série) 3-Médio (3ª ao 5ª ano) 4-Superior 5-Não se aplica 9-Ignorado	Nível de escolaridade do paciente. Para os níveis fundamental e médio deve ser considerada a última série do ano concluído.	Campo Essencial Preenchido automaticamente com a categoria "Não se aplica" quando idade for menor que 7 anos. Quando idade for maior que 7 anos, o campo não pode ser preenchido com "Não se aplica".	CS_ESCOL_R
20- Ocupação	Varchar(25)	Tabela com código da Ocupação da Classificação Brasileira de Ocupações (CBO)	Ocupação profissional do paciente	Campo Essencial	PAC_COCBO ou PAC_DISCBO

INVSF (Sipre - Sistema de Informação de Vigilância Epidemiológica da Dengue) - Página 6

21- Nome de mãe	Varchar(70)		Nome completo da mãe do paciente (sem abreviações).	Campo Essencial	NM_MAE_PAC
22- CEP	Varchar(20)		CEP de residência do paciente.	Campo Essencial Validado a partir de tabela de CEP dos Correios.	RM_CEP
23-UF	Varchar(2)	Tabela com código e sigla das UF padronizadas pelo IBGE.	Unidade Federativa de residência do paciente.	Campo Obrigatório Se campo 15-País for Brasil. Se preenchido o campo CEP, a UF é preenchida automaticamente pelo sistema e desabilitada para edição.	SI_UF
Regional de Saúde de Residência (RSG)	Varchar(20)	Tabela com código e nomes das Regionais de Saúde dos municípios de residência padronizadas pelo IBGE.	Regional de Saúde onde está localizada o Município de residência do paciente.	Campo Interno Preenchendo o nome da regional de saúde de residência, o código é preenchido automaticamente, e vice-versa;	ID_RG_RES OU CO_RG_RES
24-Município (RSG)	Varchar(20)	Tabela com código e nome dos Municípios padronizados pelo IBGE.	Município de residência do paciente.	Campo Obrigatório Se campo 15-País for Brasil. Se preenchido o campo CEP, o Município e seu respectivo código IBGE são preenchidos automaticamente pelo sistema e desabilitados para edição. Se a CEP não for preenchido, o campo é habilitado depois de selecionada uma UF no campo 23. Nesse caso, o sistema abre tabela com os municípios da UF. Preenchendo o nome do município, o código é preenchido automaticamente, ou vice-versa.	ID_MN_RES OU CO_MN_RES
25-Bairro	Varchar(17)	Tabela com código e nome dos Bairros padronizados pelos Correios.	Bairro de residência do paciente.	Campo Essencial Se preenchido o campo CEP, o Bairro é preenchido automaticamente pelo sistema.	RM_BAIRRO
26-Logradouro (Rua, Avenida, etc.)	Varchar(150)	Tabela com código e nome dos logradouros padronizados pelos Correios.	Logradouro (Rua, Avenida, quadra, travessa, etc.) de residência do paciente.	Campo Essencial Se preenchido o campo CEP, o logradouro é preenchido automaticamente pelo sistema.	RM_LOGRADO

INVSF (Sipre - Sistema de Informação de Vigilância Epidemiológica da Dengue) - Página 6

			paciente.		
27-Alt	Varchar(20)		Nr do Signatário (nr da casa ou de edifício).	Campo Essencial	NU_NUMERO
28-Complemento (apto, casa, etc.)	Varchar(133)		Complemento do logradouro (bloco, apto, casa, etc.).	Campo Essencial	NM_COMPLEM
29-DDD(Telefone	Varchar(4) Varchar(133)		Código DDD e número do telefone para contato do paciente.	Campo Essencial	NU_DDD_TEL OU NU_TELEFON
30-Zona	Varchar(3)	1-Urbano 2-Rural 3-Periurbana 9-Ignorado	Zona geográfica do endereço de residência do paciente.	Campo Essencial	CS_ZONA
31-Pais (se residente fora do Brasil)	Varchar(23)	Tabela com código e nome dos Países.	País de residência do paciente.	Campo Obrigatório Se preenchido CEP, ou for selecionada uma UF, o campo País é preenchido automaticamente pelo sistema e desabilitado para edição. Se selecionado País diferente de Brasil, os campos 22 a 26 são desabilitados.	ID_PAIS OU CO_PAIS
32-Trata-se de caso nosocomial (infecção adquirida no hospital)?	Varchar(2 (1))	1-Sim 2-Não 9-Ignorado	Caso de SRAG com infecção adquirida após internação.	Campo Essencial Quando o campo 32 for igual a 1, é permitido digitar data de início das sintomas posterior a data de internação.	NOSCOMIAL
33- Paciente trabalha ou tem contato direto com aves, suínos, ou outro animal?	Varchar(2)	1-Sim 2-Não 9-Ignorado	Caso com contato direto com aves ou suínos.	Campo Essencial	AVE_SUINO
33-Paciente trabalha ou tem contato direto com aves, suínos/Outro animal (especificar)	Varchar(100)		Informar o animal que o paciente teve contato se selecionado a opção 3.	Campo Essencial Habilitado de Campo 33- Contato com outro animal = 3 (Outro).	OUT_ANIM
34-Sinais e Sintomas/Febre	Varchar(2 (1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou febre?	Campo Essencial	FEBRE
34-Sinais e Sintomas/Tosse	Varchar(2 (1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou tosse?	Campo Essencial	TOSE

INVSF (Sipre - Sistema de Informação de Vigilância Epidemiológica da Sipe) - Página 7

		9-Ignorado			
34-Sinais e Sintomas/Dor de Garganta	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou dor de garganta?	Campo Essencial	GARGANTA
34-Sinais e Sintomas/Diápnéia	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou diápnéia?	Campo Essencial	DISPNEIA
34-Sinais e Sintomas/Desconforto Respiratório	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou desconforto respiratório?	Campo Essencial	DESC_RESP
34-Sinais e Sintomas/Saturação O ₂ < 95%	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou saturação O ₂ < 95%?	Campo Essencial	SATURACAO
34-Sinais e Sintomas/Diarreia	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou diarreia?	Campo Essencial	DIARREIA
34-Sinais e Sintomas/Vômito	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou vômito?	Campo Essencial	VOMITO
34-Sinais e Sintomas/Dor abdominal	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou dor abdominal?	Campo Essencial	DOR_ABD
34-Sinais e Sintomas/Fadiga	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou fadiga?	Campo Essencial	FADIGA
34-Sinais e Sintomas/Perda do Olfato	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou perda do olfato?	Campo Essencial	PERD_OLFT
34-Sinais e Sintomas/Perda do Paladar	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou perda do paladar?	Campo Essencial	PERD_PALA
34-Sinais e Sintomas/Outros	Varchar(2(1))	1-Sim 2-Não 9-Ignorado	Paciente apresentou outro(s) sintoma(s)?	Campo Essencial	OUTRO_SIN
34-Sinais e Sintomas/Outros	Varchar(100)		Listar outros sinais	Campo Essencial	OUTRO_DES

INVSF (Sipre - Sistema de Informação de Vigilância Epidemiológica da Sipe) - Página 8

(Descrição)			e sintomas apresentados pelo paciente.	Validado se selecionado categoria 3-Sim em Sinais e Sintomas/Outros.	
35-Fatores de risco	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente apresenta algum fator de risco	Campo Essencial	FATOR_RISC
35-Fatores de risco/ Puberdade	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente é puérpera ou puérpera (mulher que pariu recentemente - até 45 dias do parto)?	Campo Essencial Validado se selecionado no campo 8- Sexo Feminino.	PUBERPA
35-Fatores de risco/ Doença Cardiovascular Crônica	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui Doença Cardiovascular Crônica?	Campo Essencial	CARDIOPATI
35-Fatores de risco/ Doença Hematológica Crônica	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui Doença Hematológica Crônica?	Campo Essencial	HEMATOLOGI
35-Fatores de risco/ Síndrome de Down	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui Síndrome de Down?	Campo Essencial	SND_DOWNI
35-Fatores de risco/ Doença Hepática Crônica	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui Doença Hepática Crônica?	Campo Essencial	HEPATICA
35-Fatores de risco/ Asma	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui Asma?	Campo Essencial	ASMA
35-Fatores de risco/ Diabetes mellitus	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui Diabetes mellitus?	Campo Essencial	DIABETES
35-Fatores de risco/ Doença Neurológica Crônica	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui Doença Neurológica?	Campo Essencial	NEUROLOGIC
35-Fatores de risco/ Outra Pneumopatia Crônica	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui outra pneumopatia crônica?	Campo Essencial	PNEUMOPATI

INVSF Orque. Sistema de Informação da Vigilância Epidemiológica da Orque. Página 9

35-Fatores de risco/ Insuficiência ou Imunodepressão	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui insuficiência ou imunodepressão (alteração da função do sistema imunológico)?	Campo Essencial	IMUNODEPRE
35-Fatores de risco/ Doença Renal Crônica	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui Doença Renal Crônica?	Campo Essencial	RENAL
35-Fatores de risco/ Obesidade	Varchar2 (3)	1-Sim 2-Não 9-Igorado	Paciente possui obesidade?	Campo Essencial	OBSIDADE
35-Fatores de risco/ Obesidade (Exemplo IMC)	Varchar2(3)		Valor do IMC (Índice de Massa Corporal) do paciente calculado pelo profissional de saúde.	Campo Essencial Validado se selecionado categoria 3-Sim em Fatores de risco/Obesidade.	OBS_IMC
35-Fatores de risco/ Outros	Varchar2(3)	1-Sim 2-Não 9-Igorado	Paciente possui outro(s) fator(es) de risco?	Campo Essencial	OUT_MORBI
35-Fatores de risco/ Outros (Exemplo)	Varchar2(30)		Listar outro(s) fator(es) de risco do paciente.	Campo Essencial Validado se selecionado categoria 3-Sim em Fatores de risco/Outros.	MORBI_DESC
36- Recebeu vacina COVID-19?	Varchar(1)	1-Sim 2-Não 9-Igorado	Informar se o paciente recebeu vacina COVID-19, após verificar a documentação / cadastro.	Campo Obrigatório *Integração com a Base Nacional de Vacinação	VACINA_COV
37- Data 1ª dose de vacina COVID-19	Varchar(10)	Data DD/MM/AAAA	Informar a data em que o paciente recebeu a 1ª dose da vacina COVID-19	Campo essencial *Integração com a Base Nacional de Vacinação	DOSE_1_COV
37- Data 2ª dose de vacina COVID-19	Varchar(10)	Data DD/MM/AAAA	Informar a data em que o paciente recebeu a 2ª dose	Campo essencial *Integração com a Base Nacional de Vacinação	DOSE_2_COV

INVSF Orque. Sistema de Informação da Vigilância Epidemiológica da Orque. Página 10

			da vacina COVID-19	Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
37- Data da dose reforço de vacina COVID-19	Varchar(10)	Date	DD/MM/AAAA	Informe a data em que o paciente recebeu a dose reforço	DOSE_REF
				Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
37- Data da 2ª dose reforço de vacina COVID-19	Varchar(10)	Date	DD/MM/AAAA	Informe a data em que o paciente recebeu a 2ª dose reforço	DOSE_2REF
				Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
38- Fabricante 1ª dose de vacina COVID-19	Varchar(80)			Informe o fabricante da vacina, que o paciente recebeu na primeira dose	FAB_COV1
				Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
38- Fabricante 2ª dose de vacina COVID-19	Varchar(80)			Informe o fabricante da vacina, que o paciente recebeu na segunda dose	FAB_COV2
				Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
38- Fabricante dose reforço de vacina COVID-19	Varchar(80)			Informe o fabricante da vacina, que o paciente recebeu na dose reforço	FAB_COVR1
				Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
38- Fabricante 2ª dose reforço de vacina COVID-19	Varchar(80)			Informe o fabricante da vacina, que o paciente recebeu na 2ª dose reforço	FAB_COVR2
				Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
39- Lote da vacina COVID-19: Lote 1ª Dose	Varchar(20)			Informe o Lote da 1ª dose da vacina COVID-19, que o paciente recebeu	LOTE_1_COV
				Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
39- Lote da vacina COVID-19: Lote 2ª Dose	Varchar(20)			Informe o Lote da 2ª dose da vacina COVID-19, que o paciente recebeu	LOTE_2_COV

SINEP Orque - Sistema de Informação da Vigilância Epidemiológica do Orque - Página 11

			COVID-19, que o paciente recebeu	Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
39- Lote da vacina COVID-19: Lote dose reforço	Varchar(20)			Informe o Lote da dose reforço da vacina COVID-19, que o paciente recebeu	LOTE_REF
				Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
39- Lote da vacina COVID-19: Lote 2ª dose reforço	Varchar(20)			Informe o Lote da 2ª dose reforço da vacina COVID-19, que o paciente recebeu	LOTE_REF2
				Validado se campo 36- Recebeu vacina COVID-19? for igual a 1.	
39- Fonte dos dados/informação sobre a vacina COVID-19	Varchar(2)	1- Manual 2- Integração		Campo Interno Número gerado automaticamente pelo sistema. Campo preenchido de acordo com a fonte dos dados/informação sobre a vacina COVID-19, se foi digitada manualmente ou recuperada via integração com a Base Nacional de Vacinação.	FNT_IN_COV
40- Recebeu vacina contra Gripe na última campanha?	Varchar(2) (1)	1-Sim 2-Não 3-Ignorado	Informe se o paciente foi vacinado contra gripe na última campanha, após verificar a documentação / caderneta. Caso o paciente não tenha o caderneta, direcionar a pergunta para o responsável e preencher o campo com o código correspondente a resposta.	Campo Essencial	VACINA
41- Data da vacinação	Date		Data da última dose	Campo Essencial	DT_VT_DOSE

SINEP Orque - Sistema de Informação da Vigilância Epidemiológica do Orque - Página 12

	DD/MM/AAAA		de vacina contra gripe que o paciente tomou.	habilitado no campo 40- Recebeu vacina contra Gripe na última campanha? for igual a 1. Data deve ser <= a data de digitação (data atual).	
Se < 6 meses: a mãe recebeu a vacina?	Verchar2 (1)	1-Sim 2-Não 3-Igorado	Se paciente < 6 meses, a mãe recebeu vacina?	Campo Essencial habilitar campo Se a idade do caso for < 6 meses.	M_MAE_VAC
Se sim, data	Date DD/MM/AAAA		Se a mãe recebeu vacina, qual a data?	Campo Essencial habilitado no campo Se < 6 meses: a mãe recebeu a vacina for igual a 1. Data deve ser <= a data de digitação (data atual).	DT_MAE_VAC
Se < 6 meses: a mãe amamenta a criança?	Verchar2 (1)	1-Sim 2-Não 3-Igorado	Se paciente < 6 meses, a mãe amamenta a criança?	Campo Essencial habilitar campo se Se a idade do caso for < 6 meses.	M_AMAMENTA
Se >= 6 meses e <= 8 anos: Data da dose única 1/1	Date DD/MM/AAAA		Se >= 6 meses e <= 8 anos, data da dose única para crianças vacinadas em campanhas de anos anteriores	Campo Essencial habilitar campo Se a idade do caso for >= 6 meses e <= 8 anos	DT_DOSEUN
Se >= 6 meses e <= 8 anos: Data da 2ª dose	Date DD/MM/AAAA		Se >= 6 meses e <= 8 anos, data da 2ª dose para crianças vacinadas pela primeira vez	Campo Essencial habilitar campo Se a idade do caso for >= 6 meses e <= 8 anos	DT_2_DOSE
Se >= 6 meses e <= 8 anos: Data da 3ª dose	Date DD/MM/AAAA		Se >= 6 meses e <= 8 anos, data da 3ª dose para crianças vacinadas pela primeira vez	Campo Essencial habilitar campo Se a idade do caso for >= 6 meses e <= 8 anos	DT_3_DOSE
42-Usou antiviral para gripe?	Verchar2 (1)	1-Sim 2-Não 3-Igorado	Foi uso de antiviral para tratamento da doença?	Campo Essencial	ANTIVIRAL

SIMP-Gripe: Sistema de Informação de Vigilância Epidemiológica da Gripe. Página 13

43-Qual antiviral?	Verchar2 (3)	1- Oseltamivir 2- Zanamivir 3- Outros, especifique	Qual antiviral utilizado?	Campo Essencial habilitado no campo 42-Usou antiviral para gripe? for igual a 1.	TP_ANTIVIR
Qual antiviral/Outro, especifique	Verchar2(30)		Se o antiviral utilizado não foi Oseltamivir ou Zanamivir, informar qual antiviral foi utilizado.	Campo Essencial habilitado no campo 43- Qual antiviral? for igual a 3.	OUT_ANTIV
44-Data do início do tratamento	Date DD/MM/AAAA		Data em que foi iniciado o tratamento com o antiviral.	Campo Essencial habilitado no campo 42-Usou antiviral para gripe? for igual a 1. Data deve ser <= a data de digitação (data atual).	DT_ANTIVIR
45- Recebeu tratamento antiviral para covid-19?	Verchar2(1)	1-Sim 2-Não 3-Igorado	Foi uso de antiviral para tratamento de covid-19?	Campo Essencial	TRAT_COV
46- Qual antiviral?	Verchar2(1)	1- Remdesivir (ritonavir) (Padroniz*) 2- Molnupiravir (Lagevri*) 3- Baricitinib (Osimert*) 4- Outros, especifique	Se foi feito uso de antiviral para tratamento de covid-19, informar qual, conforme relação disponível.	Habilitado no campo 45-Recebeu tratamento antiviral para covid-19? for igual a 1.	TIPO_TRAT
Qual antiviral/Outro, especifique	Verchar2(30)		Se o antiviral utilizado não foi, 1- Remdesivir (ritonavir) (Padroniz*) 2- Molnupiravir (Lagevri*) 3- Baricitinib (Osimert*), informar qual antiviral foi utilizado.	Campo Essencial habilitado no campo 46- Qual antiviral? for igual a 4.	OUT_TRAT
47- Data do início do tratamento	Date DD/MM/AAAA		Data em que foi iniciado o tratamento com o antiviral, para tratamento de covid-19.	Campo Essencial habilitado no campo 45-Recebeu tratamento antiviral para covid-19? for igual a 1. Data deve ser <= a data de digitação (data atual).	DT_TRAT_COV

SIMP-Gripe: Sistema de Informação de Vigilância Epidemiológica da Gripe. Página 14

48 Houve internação?	Varchar(3)	1-Sim 2-Não 3-Ignorado	O paciente foi internado?	Campo Essencial Caso o campo não seja igual a 1 – Sim o sistema emitirá um aviso indicando que não atende a definição de caso.	HOSPITAL
49 Data da internação por SRAG	Data (DD/MM/AAAA)		Data em que o paciente foi hospitalizado.	Campo Obrigatório Data deve ser maior ou igual a 2-Data de 1ºs sintomas e menor ou igual a data de altação (atual).	DT_INTERNA
50 UF de internação	Varchar(2)	Tabela com código e sigla das UF padronizadas pelo IBGE.	Unidade Federativa de internação do paciente.	Campo Essencial habilitado se campo 48 Houve internação? for igual a 1.	SG_UF_INTE
Regional de Saúde de internação (RSG)	Varchar(3)	Tabela com código e nome das Regionais de Saúde dos municípios de internação padronizados pelo IBGE.	Regional de Saúde onde está localizado o Município de internação do paciente.	Campo Interno Preenchendo o nome da regional de saúde de internação, o código é preenchido automaticamente, e vice-versa.	ID_RG_INTE OU CO_RG_INTE
51 Município de internação/ Código (RSG)	Varchar(30)	Tabela com código e nome dos Municípios padronizados pelo IBGE.	Município onde está localizado a Unidade de Saúde onde o paciente internou.	Campo Essencial habilitado se campo 48 Houve internação? for igual a 1.	ID_MN_INTE OU CO_MN_INTE
52 Unidade de Saúde de internação/ Código CNES	Varchar(20)	Tabela com códigos CNES e nomes das Unidades de Saúde.	Unidade que realizou a internação do paciente.	Campo Essencial habilitado se campo 48 Houve internação? for igual a 1.	ID_UN_INTE OU CO_UN_INTE
53 Internado em UTI?	Varchar(3)	1-Sim 2-Não 3-Ignorado	O paciente foi internado em UTI?	Campo Essencial	UTI
54 Data da entrada na UTI	Data (DD/MM/AAAA)		Data de entrada do paciente na unidade de Terapia Intensiva (UTI).	Campo Essencial habilitado se campo 53 Internado em UTI? for igual a 1. Data deve ser maior ou igual a 2-Data de 1ºs sintomas de SRAG e menor ou igual a data de altação (atual).	DT_ENTUTI
55 Data da saída de UTI	Data (DD/MM/AAAA)		Data em que o paciente saiu da	Campo Essencial	DT_SAIDUTI

SIMP-Cipe: Sistema de Informação da Vigilância Epidemiológica do Cipe – Página 16

			Unidade de Terapia Intensiva (UTI).	habilitado se campo 53 Internado em UTI? for igual a 1. Data deve ser maior ou igual a 54 Data da entrada na UTI e menor ou igual a data de altação (atual).	
56 Uso de suporte ventilatório?	Varchar(3)	1-Sim, invasivo 2-Sim, não invasivo 3-Não 3-Ignorado	O paciente fez uso de suporte ventilatório?	Campo Essencial	SUPPORT_VEN
57-Razo X de Tórax	Varchar(3)	1-Normal 2-Subtórax intercostal 3-Circunscrito 4-Misto 5-Outro 6-Não realizado 3-Ignorado	Informar nos Achados de Razo X de Tórax.	Campo Essencial	RAIOX_RES
Razo X de Tórax/ Outro (especificar)	Varchar(30)		Informar o resultado do RX de Tórax se selecionado a opção 5-Outro.	Campo Essencial habilitado se campo 57- Razo X de Tórax = 5 (Outro).	RAIOX_OUT
58 Data do Razo X	Data (DD/MM/AAAA)		Se realizou RX de Tórax, especificar a data do exame.	Campo Essencial habilitado se campo 57- Razo X de Tórax for igual a 1, 2, 3, 4 ou 5.	DT_RAIOX
59- Aspecto Tomografia	Number(3)	1-Típico covid-19 2-Indeterminado covid-19 3-Aspecto covid-19 4-Negativo para Pneumonia 5-Outro 6-Não realizado 3-Ignorado	Informar o resultado de tomografia.	Campo Essencial	TOMO_RES
Aspecto Tomografia/Outro (especificar)	Varchar(200)		Informar o resultado de tomografia se selecionado a opção 5-Outro.	Campo Essencial habilitado se campo 59- Aspecto Tomografia = 5 (Outro).	TOMO_OUT
60-Data de Tomografia	Data (DD/MM/AAAA)		Se realizou tomografia, especificar a data do exame.	Campo Essencial habilitado se campo 59- Aspecto Tomografia for igual a 1, 2, 3, 4 ou 5.	DT_TOMO

SIMP-Cipe: Sistema de Informação da Vigilância Epidemiológica do Cipe – Página 16

61- Coletou amostra?	Varchar(3)	1- Sim 2- Não 9- Ignorado	Foi realizado coleta de amostra para realização de teste diagnóstico?	Campo Essencial	AMOSTRA
62- Data da Coleta	Data (DD/MM/AAAA)		Data da coleta de amostra para realização de teste diagnóstico.	Campo Essencial Habilitado se campo 50-Coletou amostra? = 1. Data deve ser maior ou igual a 2-Data de 14o sistema e menor ou igual a data da digitação (atual).	DT_COLETA
63- Tipo de amostra	Varchar(30)	1- Secção de Arroz-ervilha 2- Lavado Bruto-alveolar 3- Secção post-mortem 4- Outra, qual? 5- OUT 9- Ignorado	Tipo de amostra clínica coletada para o teste diagnóstico.	Campo Essencial Habilitado se campo 61-Coletou amostra? = 1.	TP_AMOSTRA
Tipo de amostra/Outra	Varchar(30)		Descrição do tipo de amostra clínica, caso diferente das listadas nas categorias do campo.	Campo Essencial Campo habilitado se selecionado categoria 4-Outra, qual em Tipo de amostra.	OUT_AMOST
64- Nº de Requisição do GAL			Numero de requisição de exames gerado pelo sistema GAL.	Campo Essencial	REQUI_GAL
65- Tipo de Teste antígeno	Number(3)	1-Imunofluorescência (IF) 2- Teste rápido antígeno	Tipo de teste antígeno que foi realizado.	Campo Essencial	TP_TES_AN
66- Data do resultado teste Antígeno	Data (DD/MM/AAAA)		Data do resultado do teste antígeno.	Campo Essencial Data deve ser maior ou igual a 62- Data da Coleta.	DT_RES_AN
67- Resultado do Teste Antígeno	Varchar(3)	1-positivo 2-negativo 3- inconclusivo 4-Não realizado 5-Aguardando resultado 9- Ignorado	Resultado do Teste Antígeno	Campo Essencial Este campo será marcado com 5-Aguardando Resultado e estará habilitado se o campo 61-Coletou amostra? = 1.	RES_AN

SIMP-Orça: Sistema de Informação da Vigilância Epidemiológica da Orça - Página 17

68-Laboratório que realizou o Teste antígeno	Varchar(30)	Nome dos Laboratórios cadastrados no sistema	Laboratório responsável pela liberação do resultado do teste antígeno.	Campo Essencial Habilitado se campo 67- Resultado do teste antígeno: estiver selecionado como 1-Positivo, 2- Negativo, 3- inconclusivo ou 5- Aguardando resultado. Preenchendo o nome do Laboratório, o código é preenchido automaticamente, ou não-será.	LAB_AN
69-Laboratório que realizou o Teste antígeno	Varchar(30)	Tabela com código CNES		Campo Essencial	CD_LAB_AN
69-Agente etiológico - Teste Antígeno. Positivo para Influenza?	Varchar(3)	1-Sim 2-Não 9- Ignorado	Resultado do Teste Antígeno que foi positivo para Influenza	Campo Essencial	POS_AN_FLU
69-Agente etiológico - Teste Antígeno. Se sim, qual Influenza?	Varchar(3)	1-Influenza A 2-Influenza B	Resultado do Teste Antígeno, para o tipo de Influenza.	Campo Essencial Habilitado se campo 69-Agente etiológico - Teste Antígeno: Positivo para Influenza? = 1.	TP_FLU_AN
69-Agente etiológico - Teste Antígeno. Positivo para outros vírus?	Varchar(3)	1-Sim 2-Não 9- Ignorado	Resultado do Teste Antígeno, que foi positivo para outro vírus respiratório.	Campo Essencial	POS_AN_OUT
69-Agente etiológico - Teste Antígeno. SARS-CoV-2	Varchar(3)	1-marcado pelo usuário Vazio - não marcado	Resultado do Teste Antígeno, para SARS-CoV-2.	Campo Essencial Habilitado se campo 69-Agente etiológico, Teste Antígeno, Positivo para outros vírus? = 1.	AN_SARS2
69-Agente etiológico - Teste Antígeno. VSR	Varchar(3)	1-marcado pelo usuário Vazio - não marcado	Resultado do Teste Antígeno, para VSR.	Campo Essencial Habilitado se campo 69-Agente etiológico, Teste Antígeno, Positivo para outros vírus? = 1.	AN_VSR
69-Agente etiológico - Teste Antígeno. Parainfluenza 1	Varchar(3)	1-marcado pelo usuário Vazio - não marcado	Resultado do Teste Antígeno, para Parainfluenza 1.	Campo Essencial Habilitado se campo 69-Agente etiológico, Teste Antígeno, Positivo para outros vírus? = 1.	AN_PARA1
69-Agente etiológico - Teste Antígeno. Parainfluenza 2	Varchar(3)	1-marcado pelo usuário Vazio - não marcado	Resultado do Teste Antígeno, para Parainfluenza 2.	Campo Essencial Habilitado se campo 69-Agente etiológico, Teste Antígeno, Positivo para outros vírus? = 1.	AN_PARA2
69-Agente etiológico - Teste Antígeno. Parainfluenza 3	Varchar(3)	1-marcado pelo usuário Vazio - não marcado	Resultado do Teste Antígeno, para Parainfluenza 3.	Campo Essencial Habilitado se campo 69-Agente etiológico, Teste Antígeno, Positivo para outros vírus? = 1.	AN_PARA3

SIMP-Orça: Sistema de Informação da Vigilância Epidemiológica da Orça - Página 18

69- Agente etiológico – Teste Antígeno: Adenovírus	Varchar2(1)	1-marcado pelo usuário Vazio – não marcado	Resultado do Teste Antígeno: Adenovírus.	Campo Essencial habilitado se campo 69-Agente etiológico, Teste Antígeno. Positivo para outros vírus? = 1.	AR_ADENO
69- Agente etiológico – Teste Antígeno: Outro vírus respiratório	Varchar2 (1)	1-marcado pelo usuário Vazio – não marcado	Resultado do Teste Antígeno: Outro vírus respiratório.	Campo Essencial habilitado se campo 69-Agente etiológico, Teste Antígeno. Positivo para outros vírus? = 1.	AR_OUTRO
69- Agente etiológico – Teste Antígeno: Outro vírus respiratório (Descrição)	Varchar2(30)		Nome do outro vírus respiratório identificado – pela Teste Antígeno.	Campo Essencial habilitado se campo 69-Agente etiológico, Teste Antígeno. Positivo para outros vírus? = 1.	OR_AR_OUT
70-Resultado de RT-PCR/outra método por Biologia Molecular	Varchar2 (1)	1-Detectável 2-Não Detectável 3-Inconclusivo 4-Não Realizado 5-Aguardando Resultado 6-gravado	Resultado do teste de RT-PCR/outra método por Biologia Molecular.	Campo Essencial Este campo será marcado com 5-Aguardando Resultado e estará habilitado se o campo 63-Colélio amostra? = 1.	PCR_RESUL
71-Data do Resultado RT-PCR/outra método por Biologia Molecular	Date DD/MM/AAAA		Data do Resultado RT-PCR/outra método por Biologia Molecular	Campo Essencial Campo habilitado se selecionado categoria 1-Detectável, 2-Não Detectável ou 3-Inconclusivo em Resultado da RT-PCR/outra método por Biologia Molecular. Data deve ser >= a data da coleta- campo 62.	DT_PCR
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para Influenza?	Varchar2 (1)	1-Sim 2-Não 3-gravado	Resultado da RT-PCR foi positivo para Influenza	Campo Essencial	POS_PCRFLU
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Se sim, qual Influenza?	Varchar2(1)	1-Influenza A 2-Influenza B	Resultado diagnóstico do RT-PCR para o tipo de Influenza.	Campo Essencial habilitado se campo 72-Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para Influenza? = 1.	TP_FLU_PCR
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Se Influenza A, qual subtipo?	Varchar2(1)	1-Influenza A(H1N1)pdm09 2-Influenza A (H3N2) 3-Influenza A não subtipado 4-Influenza A não subtipável 5-Inconclusivo 6-Outro, especifique	Subtipo para Influenza A.	Campo Essencial habilitado se campo 72-Agente etiológico – RT-PCR/outra método por Biologia Molecular: Se sim, qual Influenza? = 1.	PCR_FLUASU

SINEP Onip: Sistema de Informação de Vigilância Epidemiológica do Onip. Página 18

72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Se Influenza A, qual subtipo? Outro, especifique	Varchar2 (30)		Outro subtipo para Influenza A.	Campo Essencial habilitado se campo 72-Agente etiológico – RT-PCR/outra método por Biologia Molecular: Se Influenza A, qual subtipo? = 6.	FLUASU_OUT
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Se Influenza B, qual linhagem?	Varchar2(1)	1-Victoria 2-Yamaguchi 3-Não realizado 4-Inconclusivo 5-Outro, especifique	Linhagem para Influenza B.	Campo Essencial habilitado se campo 72-Agente etiológico – RT-PCR/outra método por Biologia Molecular: Se sim, qual Influenza? = 2.	PCR_FLUBU
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Se Influenza B, qual linhagem? Outro, especifique	Varchar2 (30)		Outra linhagem para Influenza B.	Campo Essencial habilitado se 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Se Influenza B, qual linhagem? = 3.	FLUBU_OUT
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus?	Varchar2 (1)	1-Sim 2-Não 3-gravado	Resultado da RT-PCR foi positivo para outro vírus respiratório	Campo Essencial	POS_PCROUT
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: SARS-CoV-2	Varchar2 (1)	1-marcado pelo usuário Vazio – não marcado	Resultado diagnóstico do RT-PCR para (SARS-CoV-2).	Campo Essencial habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1.	PCR_SARS2
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: VSR	Varchar2 (1)	1-marcado pelo usuário Vazio – não marcado	Resultado diagnóstico do RT-PCR para (VSR).	Campo Essencial habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1.	PCR_VSR
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Parainfluenza 1	Varchar2 (1)	1-marcado pelo usuário Vazio – não marcado	Resultado diagnóstico do RT-PCR para Parainfluenza 1.	Campo Essencial habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1.	PCR_PARA1
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Parainfluenza 2	Varchar2 (1)	1-marcado pelo usuário Vazio – não marcado	Resultado diagnóstico do RT-PCR para Parainfluenza 2.	Campo Essencial habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1.	PCR_PARA2

SINEP Onip: Sistema de Informação de Vigilância Epidemiológica do Onip. Página 19

72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Parainfluenza 3	Varchar2 (3)	1-marcado pelo usuário Vazio - não marcado	Resultado diagnóstico do RT-PCR para Parainfluenza 3.	Campo Essencial Habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_PAR3
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Parainfluenza 4	Varchar2 (3)	1-marcado pelo usuário Vazio - não marcado	Resultado diagnóstico do RT-PCR para Parainfluenza 4.	Campo Essencial Habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_PAR4
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Adenovirus	Varchar2 (3)	1-marcado pelo usuário Vazio - não marcado	Resultado diagnóstico do RT-PCR para Adenovirus.	Campo Essencial Habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_ADEN
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Metapneumovirus	Varchar2 (3)	1-marcado pelo usuário Vazio - não marcado	Resultado diagnóstico do RT-PCR para Metapneumovirus.	Campo Essencial Habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_METAP
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Bocavirus	Varchar2 (3)	1-marcado pelo usuário Vazio - não marcado	Resultado diagnóstico do RT-PCR para Bocavirus.	Campo Essencial Habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_BOCA
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Rotavirus	Varchar2 (3)	1-marcado pelo usuário Vazio - não marcado	Resultado diagnóstico do RT-PCR para Rotavirus.	Campo Essencial Habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_ROT
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Outro vírus respiratório, especifique:	Varchar2 (3)	1-marcado pelo usuário Vazio - não marcado	Resultado diagnóstico do RT-PCR para Outro vírus respiratório.	Campo Essencial Habilitado se campo 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_OUTRO
72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Outro vírus respiratório (Descrição)	Varchar2 (30)		Nome de outro vírus respiratório identificado pelo RT-PCR.	Campo Essencial Habilitado se 72- Agente etiológico – RT-PCR/outra método por Biologia Molecular: Outro vírus respiratório, especifique:	DS_PCR_OUT

SIMP-Cepi: Sistema de Informação da Vigilância Epidemiológica de Cepi. Página 21

73-Laboratório que realizou RT-PCR/outra método por Biologia Molecular Código (CNE)	Varchar2 (3)	Tabela com códigos CNEs e nomes dos laboratórios cadastrados no sistema.	Laboratório responsável pela geração do resultado de teste diagnóstico (RT-PCR) da amostra do paciente.	Campo Essencial Habilitado se selecionado categoria 1-Detectável, 2-Não Detectável ou 3-Inconclusivo em 70-Resultado da RT-PCR/outra método por Biologia Molecular: Preenchendo o nome do Laboratório, o código é preenchido automaticamente, ou vice-versa.	LAB_PCR OU CO_LAB_PCR
74- Tipo de Amostra Sorológica para SARS-Cov-2	Number(3)	1- Sangue/plasma/urina 2- Outra, qual? 3- Ignorado	Tipo de amostra sorológica que foi coletada.	Campo Essencial	TP_AM_SOR
75- Tipo de Amostra Sorológica para SARS-Cov-2/Outra, qual?			Descrição do tipo de amostra clínica, caso diferente das listadas na categoria em (3) do campo.	Campo Essencial Campo habilitado se selecionado categoria 2- Outra, qual? em Tipo de Amostra Sorológica.	SOR_OUT
76- Data da coleta	Data DD/MM/AAAA		Data da coleta da amostra para diagnóstico por sorologia.	Campo Essencial Habilitado se campo 61-Coletou amostra? = 1. Data deve ser maior ou igual a 2-Data de 1hs sistemas e menor ou igual a data da digitação (atual).	DT_CO_SOR
76- Tipo de Sorologia para SARS-Cov-2	Number(3)	1- Teste rápido 2- ELISA 3- Quimioluminescência 4- Outro, qual?	Tipo de Teste Sorológico que foi realizado	Campo Essencial	TP_SOR
76- Tipo de Sorologia para SARS-Cov-2	Varchar2 (100)		Descrição do tipo de Teste Sorológico	Campo Essencial Campo habilitado se selecionado categoria 4-Outro, qual? em Tipo de Sorologia.	OUT_SOR
76- Tipo de Sorologia para SARS-Cov-2/Outra, qual?	Varchar2 (100)		Outro tipo de amostra Sorológica		SOR_OUT
76- Resultado do Teste Sorológico para SARS-Cov-2:	Varchar2(3)	igG	Resultado da sorologia para SARS-Cov-2	Campo Essencial	RES_IGG
76- Resultado do Teste Sorológico para SARS-Cov-2:	Varchar2(3)	igM	Resultado da sorologia para SARS-Cov-2	Campo Essencial	RES_IGM

SIMP-Cepi: Sistema de Informação da Vigilância Epidemiológica de Cepi. Página 22

76-Resultado do Teste Sorológico para SARS-CoV-2:	Varchar(3)	igA	Resultado da sorologia para SARS-CoV-2	Campo Essencial	RES_LGA
77-Data do Resultado	Date	DD/MM/AAAA	Data do Resultado do Teste Sorológico	Campo Essencial Data deve ser maior ou igual a 75- Data da Coleta	DT_RES
78-Classificação final do caso	Varchar(3)	1-SRAG por influenza 2-SRAG por outro vírus respiratório 3-SRAG por outro agente etiológico, qual: 4-SRAG não especificado 5-SRAG por causa 29	Diagnóstico final do caso. Se tiver resultados divergentes entre as metodologias laboratoriais, priorizar o resultado do RT-PCR.	Campo Obrigatório	CLASS_FIN
79-Classificação final do caso 3-SRAG por outra causa, qual	Varchar(30)		Descrição de qual outro agente etiológico foi identificado	Campo Obrigatório Se campo 78-Classificação final do caso = 3, substituir o campo 78-Classificação final do caso = 5.	CLASS_OUT
79-Critério de Encerramento	Varchar(3)	1. Laboratorial 2. Clínico Epidemiológico 3. Clínico 4. Clínica Imagem	Indicar qual o critério de confirmação	Campo Essencial	CRTERIO
80-Evolução do caso	Varchar(3)	1-Cura 2-Óbito 3-Óbito por outras causas 9-Ignoto	Evolução do caso	Campo Essencial	EVDUCAO
81-Data de alta ou óbito	Date	DD/MM/AAAA	Data da alta ou óbito	Campo Essencial Data de alta ou do óbito deve ser > ou = a data dos primeiros sintomas e < ou = a data da digitação (atual). Substituir o campo 80-Evolução do caso = 3 ou 2.	DT_EVOLUCA

SIMP-Orpe- Sistema de Informação da Vigilância Epidemiológica do Orpe - Página 23

82-Data do Encerramento	Date	DD/MM/AAAA	Data do encerramento do caso.	Campo Obrigatório Se o campo 78- Classificação final do caso estiver preenchido, Data do encerramento deve ser > ou = a data do preenchimento. Data do encerramento deve ser < ou = a data da digitação (atual).	DT_ENCERRA
83- Número D.O.			Número de Declaração de Óbito	Campo Essencial substituir se o campo 80- Evolução do caso = 2 ou 3	NU_DO
84-Observações	Varchar(2000)		Outras observações sobre o paciente consultadas pertinentes.	Campo Opcional	OBSERVA
85-Profissional de Saúde Responsável	Varchar(100)		Nome completo do profissional de saúde (com abreviação) responsável pela notificação.	Campo Essencial	NOME_PROF
86-Registro Conselho/Matricula	Varchar(113)		Número do conselho ou matrícula do profissional de saúde responsável pela notificação (Ex: CRM/R 1234)	Campo Essencial	REG_PROF
Data da digitação	Date	DD/MM/AAAA	Data de inclusão do registro no sistema.	Campo Interno Preenchido automaticamente pelo sistema com a data de digitação da ficha. Não é a data de preenchimento da ficha manualmente e sim a data em que é digitado no sistema. Não é atualizada se houver alterações posteriores de dados.	DT_REGISTA

SIMP-Orpe- Sistema de Informação da Vigilância Epidemiológica do Orpe - Página 24

Fonte: Dicionário de Dados SRAG 2021 a 2023 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 – OPENDATASUS ([s.d.]).