



BACHAREL EM CIÊNCIA DA COMPUTAÇÃO

**UM SISTEMA DE VISUALIZAÇÃO PARA IDENTIFICAÇÃO
DE ANOMALIAS EM DADOS EPIDEMIOLÓGICOS DE
ARBOVIROSES**

IGOR FERREIRA

Rio Verde, GO

2024



INSTITUTO FEDERAL GOIANO - CAMPUS RIO VERDE
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO

**UM SISTEMA DE VISUALIZAÇÃO PARA IDENTIFICAÇÃO
DE ANOMALIAS EM DADOS EPIDEMIOLÓGICOS DE
ARBOVIROSES**

IGOR FERREIRA

Trabalho de Conclusão de Curso apresentado ao Instituto Federal Goiano - Campus Rio Verde, como requisito parcial para a obtenção do Grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Douglas Cedrim Oliveira

Rio Verde, GO
Novembro, 2024

Sistema desenvolvido pelo ICMC/USP
Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - Instituto Federal Goiano

A663s Araújo, Igor Ferreira
Um sistema de visualização para
identificação de anomalia em dados
epidemiológicos de Arboviroses / Igor
Ferreira Araújo ; orientador Douglas Cedrim
Oliveira. -- Rio Verde, 2024.
36 f.

TCC (Bacharelado em Ciência da
computação) -- Instituto Federal Goiano,
Campus Rio Verde, 2024.

1. Análise visual de dados. 2. Suporte à
tomada de decisão. 3. Vigilância
epidemiológica. 4. Geolocalização. 5. Saúde

Responsável: Johnathan Pereira Alves Diniz - Bibliotecário-Documentalista CRB-1 nº2376



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÃO TÉCNICA NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO

Repositório Institucional do IF Goiano - RIIF Goiano Sistema Integrado de Bibliotecas

- Profissional de Educação do IF Goiano -

Com base no disposto na Lei Federal nº 9.610/98, e manual sobre a Produção Técnica, publicado pela DAV/CAPES/MEC*, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano, a disponibilizar gratuitamente o documento no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada eletronicamente abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

Identificação da Produção Técnica – DAV/CAPES

- | | |
|---|--|
| <input type="checkbox"/> Editoria | <input type="checkbox"/> Material Didático |
| <input type="checkbox"/> Curso de Formação Profissional | <input type="checkbox"/> Projetos de Extensão à Comunidade |
| <input type="checkbox"/> Relatório Técnico Conclusivo | <input type="checkbox"/> Atividade Técnica/Tecnológica |
| <input type="checkbox"/> Disseminação do Conhecimento Técnico/Tecnológico | <input type="checkbox"/> Produto Bibliográfico |
- Outras Produções Técnicas - Tipo: TCC (Graduação)

Nome Completo do Autor/a: Igor Ferreira Araujo

Matrícula: 2016202192010010

Título do Trabalho: Um sistema de visualização para identificação de anomalia em dados epidemiológicos de Arboviroses

Restrições de Acesso ao Documento

Documento confidencial: Não Sim

Justifique: _____

Informe a data que poderá ser disponibilizado no RIIF Goiano: 12/ 12 / 2024

O documento está sujeito a registro de patente? Sim Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a docente e/ou autor/a declara que:

1 - o documento é seu trabalho original, detém os direitos autorais da produção técnica e não infringe os direitos de qualquer outra pessoa ou entidade;

2 - obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;

3 - cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Rio Verde, 12 de dezembro de 2024.

(Assinado Eletronicamente)

Igor Ferreira Araujo (Autor)

(Assinado Eletronicamente)

Douglas Cedrim Oliveira (Orientador)

1058004

(Assinatura do Docente, Autor e/ou Detentor dos Direitos Autorais)

Documento assinado eletronicamente por:

- Douglas Cedrim Oliveira, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 12/12/2024 19:16:29.
- Igor Ferreira Araujo , 2016202192010010 - Discente, em 13/12/2024 10:11:04.

Este documento foi emitido pelo SUAP em 12/12/2024. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 661650
Código de Autenticação: 96c7725a1f





SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

Ata nº 74/2024 - GGRAD-RV/DE-RV/CMPRV/IFGOIANO

ATA DE DEFESA DE TRABALHO DE CURSO

Aos vinte e dois dias do mês de novembro de dois mil e vinte e quatro, às quatorze horas, reuniu-se a banca examinadora composta pelos docentes: Dr. Douglas Cedrim Oliveira (orientador), Dr. Danilo Pereira Barbosa (membro interno), Dr. Heverton Barros de Macêdo (membro interno), para examinar o Trabalho de Conclusão de Curso intitulado "Um sistema de visualização para Identificação de Anomalias em Dados Epidemiológicos de Arboviroses" de Igor Ferreira Araujo, estudante do Curso de Bacharelado em Ciência da Computação do IF Goiano – Campus Rio Verde, sob Matrícula nº 2016202192010010. A palavra foi concedida ao estudante para a apresentação oral do TCC, em seguida houve arguição do candidato pelos membros da Banca Examinadora. Após tal etapa, a Banca Examinadora decidiu pela APROVAÇÃO do estudante. Ao final da sessão pública de defesa foi lavrada a presente ata, que segue assinada pelos membros da Banca Examinadora.

(Assinado Eletronicamente)

Douglas Cedrim Oliveira

Orientador(a)

(Assinado Eletronicamente)

Danilo Pereira Barbosa

Membro

(Assinado Eletronicamente)

Heverton Barros de Macêdo

Membro

Observação:

() O(a) estudante não compareceu à defesa do TC.

Documento assinado eletronicamente por:

- **Douglas Cedrim Oliveira**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 26/11/2024 14:47:52.
- **Heverton Barros de Macedo**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 26/11/2024 14:52:42.
- **Danilo Pereira Barbosa**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 26/11/2024 14:53:01.

Este documento foi emitido pelo SUAP em 26/11/2024. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 655300

Código de Autenticação: 8fcadacb8a



INSTITUTO FEDERAL GOIANO

Campus Rio Verde

Rodovia Sul Goiana, Km 01, Zona Rural, 01, Zona Rural, RIO VERDE / GO, CEP 75901-970

(64) 3624-1000

RESUMO

FERREIRA, Igor. **Um sistema de Visualização para Identificação de Anomalias em Dados Epidemiológicos de Arboviroses**. Novembro, 2024. 36 f. Monografia – Bacharelado em Ciência da Computação, Instituto Federal Goiano - Campus Rio Verde. Rio Verde, GO.

Consideradas pela Organização Mundial da Saúde como doenças tropicais negligenciadas, a dengue e a chikungunya são as arboviroses com maior número de casos notificados no Brasil. Com a dificuldade existente de erradicar os vetores que as transmitem, estratégias de avaliar efetivamente as notificações de arboviroses para uma dada região surgem como ferramentas importantes. Nesse contexto, este trabalho propõe uma ferramenta interativa de visualização de dados que combina estatísticas de notificação de arboviroses e sua distribuição espacial, possibilitando representar valores atípicos de incidência de forma simples para o usuário.

Palavras-chave: Análise visual de dados; Suporte à tomada de decisão; Vigilância epidemiológica; Geolocalização; Saúde pública.

ABSTRACT

FERREIRA, Igor. Data Visualization. November, 2024. 36 f. Final Paper – Bachelor of Computer Science, Instituto Federal Goiano - Campus Rio Verde. Rio Verde, GO, November, 2024.

Considered by the World Health Organization as neglected tropical diseases, dengue and chikungunya are the arboviruses with the highest number of cases reported in Brazil. With the existing difficulty of eradicating the vectors that transmit them, strategies to effectively evaluate notifications of arboviruses for a given region emerge as important tools. In this context, this work proposes an interactive data visualization tool that combines arbovirus notification statistics and their spatial distribution, making it possible to represent atypical incidence values in a simple way for the user.

Keywords: Visual analytics; Decision-making support; Epidemiological Surveillance; Geolocation; Public health.

LISTA DE FIGURAS

Figura 1 – Distribuição geográfica dos casos de dengue notificados em todo o mundo, 2021.	5
Figura 2 – Fluxograma do Sistema de Informação da Dengue.	8
Figura 3 – Modelo da ficha de notificação de casos de dengue de chikungunya, modelo atual de 2016.	8
Figura 4 – Modelo do Boxplot	13
Figura 5 – Telas do Aplicativo Vaza Dengue.	14
Figura 6 – Ferramenta de análise temporal do DENGOSA.	15
Figura 7 – Previsão de surto de dengue usando dados climáticos.	16
Figura 8 – Dados disponíveis na plataforma do Cnes, de onde foi feita a coleta do endereço.	19
Figura 9 – <i>Arbolytics</i> ilustrando a visão combinada entre a distribuição de casos e sua distribuição espacial para o município de Rio Verde-GO, em 2016.	22
Figura 10 – Relação do gráfico de linha em paralelo ao <i>boxplot</i>	25
Figura 11 – Exibição detalhada do mapbox.	26
Figura 12 – Casos notificados por semana epidemiológica em Rio Verde-Go. Em laranja a série temporal de 2015 (a) e 2016 (b).	28
Figura 13 – Distribuição dos casos por semana epidemiológica em Rio Verde-Go nos anos de 2015 e 2016	29
Figura 14 – Casos notificados por semana epidemiológica em São Paulo-SP. Em laranja a série temporal de 2014 (a) e 2016 (b).	30
Figura 15 – Distribuição das notificações na semana epidemiológica 7 em São Paulo-SP.	30
Figura 16 – Casos notificados por semana epidemiológica em Salto de Pirapora-SP. Em laranja a série temporal de 2014 (a) e 2015 (b).	31
Figura 17 – Distribuição dos casos por semana epidemiológica em Salto de Pirapora-SP no ano de 2015.	32

LISTA DE TABELAS

Tabela 1 – Relação de tempo de processamento nas diversas etapas do sistema. . . 27

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface;</i>
CHIKV	Vírus da chikungunya;
CNES	Cadastro Nacional de Estabelecimentos de Saúde;
DENV	Vírus da dengue;
DENV-1	Sorotipo do vírus da Dengue;
DENV-2	Sorotipo do vírus da Dengue;
DENV-3	Sorotipo do vírus da Dengue;
DENV-4	Sorotipo do vírus da Dengue;
IBGE	Instituto Brasileiro de Geografia e Estatística;
ID	Código identificador;
IQR	Intervalo entre quartis;
KDD	<i>Knowledge Discovery in Databases;</i>
OGD	Open Government Data;
SINAN	Sistema de Informação de Agravos de Notificação;
UBS	Unidades Básica de Saúde;
UPA	Unidade de Pronto Atendimento.

LISTA DE ALGORITMOS

Algoritmo 1 – Processo de Agregação.	18
Algoritmo 2 – Processo de <i>web crawling</i>	20
Algoritmo 3 – Função de cálculo dos intervalos de Q1 à Q3	23
Algoritmo 4 – Cálculo do intervalo interquartil	24
Algoritmo 5 – Função de construção do Mapbox	25

SUMÁRIO

1	–	INTRODUÇÃO	1
1.1		Definição do Problema	1
1.2		Objetivos	2
1.3		Estrutura do trabalho	3
2	–	FUNDAMENTAÇÃO TEÓRICA	4
2.1		Doenças Arbovirais	4
2.1.1		Dengue	4
2.1.2		Vigilância Epidemiológica no Controle da Dengue	7
2.2		Dados Abertos Governamentais	9
2.3		Extração de conhecimento	9
2.4		Pré-Processamento de Dados	10
2.4.1		Qualidade dos Dados	10
2.4.2		Agregação	11
2.4.3		Amostragem	11
2.4.4		Redução de dimensionalidade	11
2.4.5		Seleção de subconjuntos de atributos	12
2.4.6		Criação de atributos	12
2.4.7		Discretização e Binarização	12
2.5		Boxplot	12
3	–	TRABALHOS RELACIONADOS	14
4	–	MATERIAIS E MÉTODOS	17
4.1		Preparação e pré-processamento do Dataset	17
4.2		Desenvolvimento do sistema	19
4.3		<i>Arbolytics</i>	21
4.3.1		Medidas estatísticas e o Gráfico de Linha	22
4.3.2		Mapbox	24
5	–	RESULTADOS	27
5.1		Rio Verde-GO	28
5.2		São Paulo-SP	29
5.3		Salto de Pirapora-SP	30
6	–	CONCLUSÃO	33
		REFERÊNCIAS	34

1 INTRODUÇÃO

Em um país de clima tropical, problemas relacionados à infecções causadas por mosquitos é cada vez maior. Doenças como: a dengue, febre-amarela, febre chikungunya e zika vírus, chamadas de arboviroses, os quais são um grupo de doenças virais transmitidas principalmente por artrópodes, como mosquitos e carrapatos, sempre foram um desafio para as autoridades de combate às endemias.

Nos últimos anos os casos de arboviroses vêm crescendo de maneira alarmante, colocando o Brasil na posição de país com o maior número de casos registrados da América latina. Os fatores que influenciam esse crescimento variam: desde a resistência dos vetores de transmissão até a insuficiência de medidas públicas efetivas no combate.

Sendo consideradas pela Organização Mundial da Saúde como doenças tropicais negligenciadas, a dengue e a chikungunya são as arboviroses com maior número de casos registrados no Brasil (NETO, 2022). Apesar de que a erradicação dos vetores de transmissão destas doenças seja considerada praticamente impossível, sobretudo, devido a fatores como: crescimento populacional, ocupação desordenada e a falta de infraestrutura dos centros urbanos; o combate a essas doenças é extremamente necessário para reduzir o impacto no Sistema Único de Saúde - SUS, bem como no sistema suplementar de saúde.

O controle do mosquito transmissor da doença pode ser efetivado através de ações contínuas de educação da comunidade e eliminação do foco dos criadouros. Atualmente esse controle também é feito pela inspeção dos domicílios pelos agentes comunitários de saúde e endemias e pela aplicação de medidas de caráter químico (WANG et al., 2020). Uma parte fundamental nesse processo é o acesso à informação, os agentes comunitários precisam de informações relacionadas a locais e quantidades de casos para tomar as devidas atitudes no combate aos focos de reprodução dos mosquitos.

1.1 Definição do Problema

Dado que o Brasil é hiperendêmico para arboviroses, a quantidade de dados coletados de atendimento a pacientes é enorme. Quase 1 milhão e meio de casos de Dengue e Chikungunya foram notificados ao Sistema de Informação de Agravos de Notificação (SINAN) em 2020, incluindo suspeitos e confirmações da doença. Mesmo com esse grande volume de dados, atualmente, ainda é extraída pouca informação desses recursos, e menos ainda ações são tomadas a respeito ao combate de endemias.

Em 2021, para combater o descontrole dos casos a secretaria da saúde permanece intensificando campanhas de combate à doença. Nessas campanhas foram realizados mutirões de limpeza e aplicações de inseticidas, sendo reconhecido, pela Secretaria de Saúde, que grande parte dos focos estão localizados em quintais e lotes baldios. Mesmo

com todas essas ações de combate à dengue, no ano de 2021, foram registrados um total de 46.394 casos confirmados em Goiás da semana 1 à 47. Em 2022 esse número subiu para 187.864 casos no mesmo período (SES-GO, 2024).

Até a 47^a semana epidemiológica de 2023, foram notificados 109.462 casos de dengue entre indivíduos residentes no estado de Goiás, dos quais 62.559 foram confirmados. No primeiro semestre de 2024, até a semana epidemiológica 47, o número de casos notificados subiu para 419.697, com 308.296 casos confirmados. Esse volume representa um aumento de 283,53% nos casos confirmados, ainda faltando 6 semanas para o fechamento do ano (SES-GO, 2024).

Para monitorar e responder com eficácia ao avanço das arboviroses, torna-se essencial uma análise comparativa dos dados ao longo dos anos, especialmente considerando a distribuição semanal dos casos. A comparação entre períodos anteriores e atuais por semana epidemiológica permite identificar variações significativas e valores atípicos, ajudando a planejar intervenções mais direcionadas. Isso é crucial para o enfrentamento dos surtos em municípios, onde os esforços de combate precisam considerar tanto os dados históricos dos últimos anos quanto a evolução semanal recente, permitindo uma resposta mais ágil e baseada em estatísticas sólidas.

1.2 Objetivos

Dessa forma, o objetivo geral desse trabalho consiste em propor uma ferramenta que permita uma análise comparativa das notificações de arboviroses cadastradas no SINAN ao longo de um determinado período, tanto em relação ao total de notificações quanto em relação à sua distribuição espacial ao longo de um município.

Mais especificamente, a ferramenta deve permitir a visualização interativa dos dados ao longo das semanas epidemiológicas e possibilitar a comparação estatística entre diferentes períodos de tempo, o que possibilita encontrar valores atípicos nas notificações de casos, para uma determinada semana epidemiológica. A partir disso, deve possibilitar também ao usuário uma visualização geoespacial das notificações ao longo das unidades de saúde de um determinado município, que tenham tido notificações cadastradas no SINAN.

Serão apresentados uma contextualização e o processo utilizado para desenvolver essa ferramenta de visualização, que poderá ser utilizada por gestores municipais e pesquisadores no monitoramento e no combate a Dengue e a febre Chikungunya. Partindo da hipótese de que é possível utilizar ferramentas interativas para explorar anomalias nas notificações de casos, a ferramenta proposta visa permitir uma análise mais dinâmica e visual das irregularidades nos dados, o que pode contribuir para a compreensão de padrões incomuns e apoiar intervenções mais rápidas e assertivas no combate às arboviroses.

1.3 Estrutura do trabalho

Esse trabalho irá conceituar os temas relacionados às arboviroses, cobrindo desde sua taxonomia até às políticas governamentais utilizadas no combate a essas doenças. Esse documento também descreverá o atual cenário de disponibilização de dados públicos governamentais e também sobre os temas relacionados a pré-processamento e mineração de dados.

O restante desse trabalho foi dividido nos seguintes capítulos:

- **Capítulo 2:** Neste capítulo, realiza-se uma revisão de literatura sobre arboviroses, com foco na dengue, discutindo aspectos como sua transmissão, a vigilância epidemiológica, e contextualizando a relevância do estudo no combate à doença;
- **Capítulo 3:** É descrito nesse capítulo a Fundamentação teórica onde são introduzidos conceitos fundamentais como dados abertos governamentais, extração de conhecimento, pré-processamento de dados e mineração de dados, oferecendo a base teórica necessária para o desenvolvimento e compreensão deste estudo;
- **Capítulo 4:** Descrevem-se nesse capítulo os materiais utilizados e os métodos aplicados durante o desenvolvimento do trabalho, detalhando as etapas, ferramentas e técnicas empregadas;
- **Capítulo 5:** Apresenta a análise detalhada dos resultados obtidos, discutindo a eficácia da implementação proposta;
- **Capítulo 6:** Conclusão onde são apresentados os resultados alcançados durante o desenvolvimento do trabalho, destacando as principais contribuições do estudo e sugerindo possíveis direções para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Doenças Arbovirais

São denominados como Arbovírus, vírus em que parte de seu ciclo replicativo ocorre em artrópodes, como os mosquitos. Sua transmissão aos seres humanos vem da picada de um desses artrópodes hematófagos infectados. Os autores Lopes, Nozawa e Linhares (2014) destacam que as principais famílias de arbovírus que podem causar doenças em humanos e outros animais de sangue quente são *Bunyaviridae*, *Togaviridae*, *Flaviviridae*, *Reoviridae*, *Rhabdoviridae* e *Iridoviridae*, as duas últimas causando infecções apenas em animais.

A família de arbovírus responsável por causar dengue é a *Flaviviridae*, especificamente pelo vírus da dengue (DENV), que pertence ao gênero *Flavivirus*. Já a chikungunya é causada por um vírus da família *Togaviridae*, mais especificamente pelo vírus chikungunya (CHIKV), que pertence ao gênero *Alphavirus*.

Por conta de seu clima, com condições ecológicas e ambientais favoráveis a reprodução de mosquitos transmissores de arboviroses, a anos o Brasil vem enfrentando problemas relacionados a doenças como a Dengue(DENV) e chikungunya(CHIKV). Tudo isso, combinado com fatores socioeconômicos que facilitam a reprodução do mosquito(políticas de armazenamento e distribuição de água e políticas de controle de coleta e descarte de lixo) o Brasil acabou se tornando o país com a maior quantidade de casos confirmados de dengue no mundo (EUROGEOGRAPHICS, 2022).

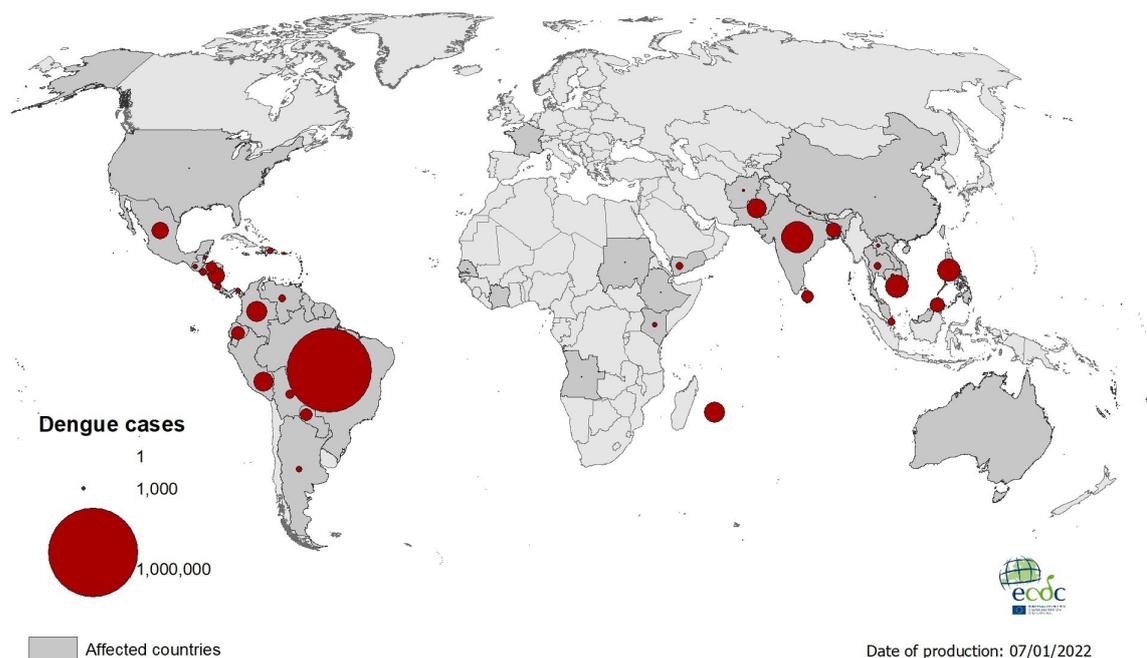
Doenças arbovirais são um constante desafio para as autoridades da saúde brasileira, o diagnóstico é complexo, já que entre si podem ter sintomas parecidos, além de também ter similaridades com outras doenças, na maioria das vezes variando a severidade e a duração. O vírus da Dengue possui quatro sorotipos(DENV-1, DENV-2, DENV-3 e DENV-4), cada um desses sorotipos possuem padrões de sintomas distintos que podem se diferir na duração e na intensidade (RAMOS et al., 2021).

Em áreas urbanas e semi-urbanizadas, o vírus da dengue e da chikungunya são os arbovírus com o maior número de casos registrados, ambos sendo transmitidos aos humanos pela picada de uma fêmea do mosquito *Aedes aegypti*. Sendo considerado o principal vetor da doença no mundo, o mosquito *Aedes aegypti*, pertencente a família *Culicidae*, gênero *Aedes* e subgênero *Stegomya*.

2.1.1 Dengue

Conhecido popularmente como "mosquito-da-dengue" ou "pernilongo-rajado", esses insetos medem cerca de 1 cm e possuem manchas brancas pelo corpo e pernas que tornam fácil sua identificação sem o auxílio de lentes ou de instrumentos ópticos (SOUZA-NETO;

Figura 1 – Distribuição geográfica dos casos de dengue notificados em todo o mundo, 2021.



Fonte: (PREVENTION; CONTROL, 2021).

POWELL; BONIZZONI, 2019). A fêmea do *Aedes aegypti* pode voar por mais de 2 km em busca de alimento, tendo seu maior período de atividade durante o dia, contudo pode atacar o homem a qualquer hora. O ciclo de reprodução desse mosquito requer um clima quente e úmido, características de áreas tropicais (MAGALHAES et al., 2020). A fêmea tende a depositar seus ovos em qualquer meio, que tenha disponível, água parada, limpa e com pouca quantidade de material orgânico em decomposição, nesse meio as larvas podem iniciar seu desenvolvimento que continua até a fase adulta. Os ovos do *Aedes aegypti* podem ficar até um ano sem entrar em contato com água para iniciar seu período larvário.

Em países com condições favoráveis, como pode ser observado na Figura 1, doenças infecciosas como a dengue podem criar proporções assustadoras. Segundo o “Doenças infecciosas e parasitárias: Guia de Bolso do Ministério da Saúde de 2010” (EPIDEMIOLOGICA, 2010), desde o ano de 1986 existem casos registrados de doenças causadas pelo *Aedes aegypti*, nessa data em específico houve a introdução do sorotipo DENV1. Entre os anos de 1990 e 2000, pela primeira vez, foram registrados casos dos sorotipos DENV2 e DENV3 no Rio de Janeiro. No período dessas duas décadas aconteceram as maiores epidemias de dengue registradas até então, com cerca de 530 mil e 800 mil casos respectivamente.

Ainda segundo o livro do Departamento de Vigilância Epidemiológica (2010), pacientes que sofre a primeira infecção pelo vírus transmitido pelo *Aedes aegypti* desenvolvem imunidade homologa ao sorotipo viral, dessa forma esse paciente não pode ser reinfectado pelo mesmo subtipo, contudo, essa pessoa ainda pode ser igualmente afetada pelos outros sorotipos que não adquiriu imunidade. Vale ressaltar, que no contexto epidemiológico, tanto o mosquito transmissor quando a dengue vem passando por evoluções ao longo do

tempo, evoluções essas que alteram sua resistência, mutação, proliferação e adaptação a condições populacionais (ARAUJO, 2018).

Sendo assim, atualmente o foco das autoridades de saúde quanto ao combate a dengue é a prevenção de mortalidade e evitar o surgimento de números incontáveis de casos, já que a erradicação da doença é considerado algo muito difícil e distante da nossa realidade. Segundo Wang et al. (2020), o controle do mosquito transmissor da doença pode ser efetivado através de ações contínuas de educação da comunidade e eliminação do foco dos criadouros. Atualmente esse controle também é feito pela inspeção dos domicílios pelos agentes comunitários de saúde e endemias e pela aplicação de medidas de caráter químico.

Fatores socioeconômicos também são fortes influenciadores na dificuldade de controle de reprodução do mosquito e crescimento desenfreado do número de casos. A urbanização desgovernada, as condições precárias de habitação e de saneamento básico combinado com as mudanças climáticas constantes, são elementos que se unem para favorecer a manutenção dos criadouros do mosquito *Aedes*. Araujo (2018) enfoca que, tanto a implementação de ações de ordem intersetorial, quanto o alerta para a responsabilidade coletiva (e isso inclui, efetiva e ativamente a comunidade), são ordens primárias para que se alcance o efetivo controle da dengue, através de aplicações das ações e políticas públicas de combate.

Clinicamente a manifestação dos sintomas pode variar de intensidade, indo de casos assintomáticos e sutis, até a entrada de quadros febris e complicações da doença (dengue hemorrágica). Os sintomas apresentados por pacientes diagnosticados com dengue são dores nos músculos e nas juntas, dor de cabeça e atrás dos olhos, cansaço, perda do apetite e paladar, leucopenia, exantema e erupções cutâneas.

Para reduzir a mortalidade e controlar a intensidade dos sintomas, diagnosticar a doença de forma rápida é de extrema importância. Atualmente as medicações recomendadas para o tratamento da infecção agem sobre os sintomas com foco em reduzir a severidade. Algumas das recomendações aos pacientes com a doença são: analgésicos para redução de dores, antipirético para o controle da febre, hidratação e descanso. Em 21 de dezembro de 2023, a vacina contra dengue foi incorporada no Sistema Único de Saúde (SUS). A inclusão da vacina da dengue é uma importante ferramenta no SUS para que a dengue seja classificada como mais uma doença imunoprevenível (SAUDE, 2024).

Os sintomas da dengue hemorrágica, podem incluir febre, dores abdominais fortes e contínuas, hemorragia, hepatomegalia, dificuldade respiratória e em casos mais graves é possível que o paciente apresente falha circulatória. Na dengue hemorrágica, o quadro clínico se agrava rapidamente, apresentando insuficiência circulatória e choque distributivo, podendo levar a pessoa a morte em até 24 horas.

Apesar dos pacientes com dengue manifestarem sintomas similares a outras doenças infecciosas, é possível validar e realizar a confirmação de dengue por meio de um diagnóstico laboratorial. O método tradicionalmente usado para a detecção do vírus

DENV é o isolamento do vírus e o diagnóstico dos seus biomarcadores proteicos (WANG et al., 2020). Apesar de o isolamento e análise viral apresentar um resultado definitivo, esse processo não é prático, por conta de que essa análise precisa de muito tempo para ser realizada, levando vários dias até que o resultado do exame esteja pronto.

Além da análise de biomarcadores, a dengue também pode ser confirmada analisando a resposta imune natural do paciente. O exame sorológico é baseado na sorologia, um conceito utilizado especialmente no campo da medicina que se refere a um exame de laboratório efetuado para comprovar a presença de anticorpos no sangue, ou seja, determinar concretamente sua presença (OHST et al., 2018). Existem dois anticorpos diferentes que podem confirmar uma infecção de dengue, os anticorpos IgM são produzidos entre um período de 3 a 5 dias após a infecção, essa sorologia é feita com maior frequência para o diagnóstico de dengue. Já o anticorpo IgG é produzido bem no fim do quadro da infecção e permanece no corpo durante meses. O exame de sangue para IgG é normalmente utilizado para identificar uma infecção passada.

2.1.2 Vigilância Epidemiológica no Controle da Dengue

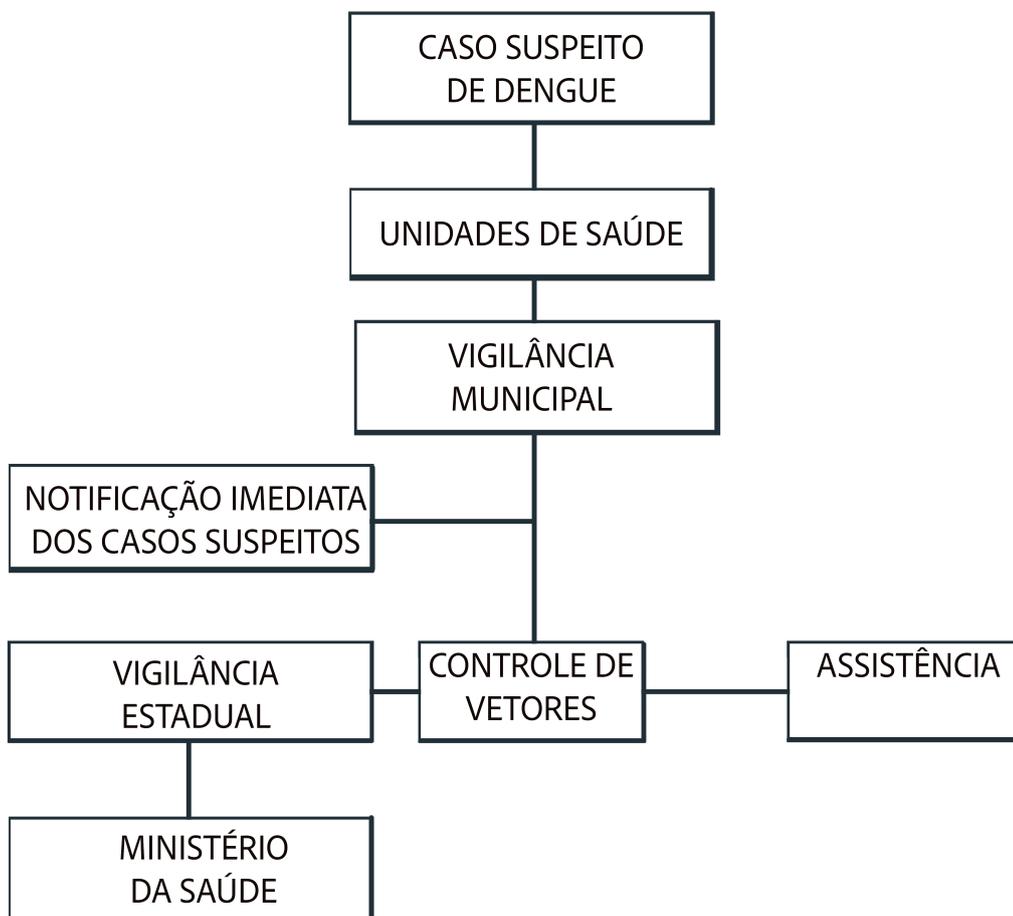
A legislação brasileira define, na lei 8080/90, a vigilância epidemiológica como sendo: "Um conjunto de ações que proporciona o conhecimento, a detecção ou a prevenção de qualquer mudança nos fatores determinantes e condicionantes de saúde individual e coletiva, com a finalidade de recomendar e adotar medidas de prevenção e controle das doenças ou agravos"(Lei 8080/90, Art. 6º, § 2º).

O fluxograma ilustrado na Figura 2 ilustra o processo em que todas as notificações de dengue devem seguir para serem devidamente registradas e assim auxiliem na tomada de medidas estratégicas para o monitoramento e controle de casos de dengue e do mosquito *Aedes aegypti* (ARAUJO, 2018).

As notificações de casos suspeitos de dengue devem seguir um determinado processo para que possam ser devidamente registradas e encaminhadas para o sistema gestor, e em seguida ao Ministério da Saúde, para serem computadas. Essas notificações são feitas a partir de *fichas de notificação* padronizadas, preenchidas por profissionais das unidades básicas de saúde(UBS). Esses dados são lançados no Sistema de Informação de Agravos de Notificação(SINAN) em até uma semana após a realização do diagnóstico, o que possibilita efetuar análises epidemiológicas atualizadas.

O modelo da ficha (Figura 3), usada atualmente, possui um total de 71 campos não descritivos. A partir dessa coleta, a vigilância epidemiológica pode analisar a evolução das notificações no município e adotar ações de combate às arboviroses.

Figura 2 – Fluxograma do Sistema de Informação da Dengue.



Fonte: Autoria própria.

Figura 3 – Modelo da ficha de notificação de casos de dengue de chikungunya, modelo atual de 2016.

Local Provável de Infecção (no período de 15 dias)			
56 O caso é autóctone do município de residência? 1-Sim 2-Não 3-Indeterminado		57 UF	58 País
59 Município	Código (IBGE)	60 Distrito	61 Bairro
62 Classificação 5- Descartado 10- Dengue 11- Dengue com Sinais de Alarme 12- Dengue Grave 13- Chikungunya		63 Critério de Confirmação/Descarte 1- Laboratório 2- Clínico-Epidemiológico 3-Em investigação	64 Apresentação clínica 1- Aguda 2- Crônica
65 Evolução do Caso 1-Cura 2- Óbito pelo agravamento 3- Óbito por outras causas 4-Óbito em investigação 9-Ignorado		66 Data do Óbito	67 Data do Encerramento
Preencher os sinais clínicos para Dengue com Sinais de Alarme e Dengue Grave			
68 Dengue com sinais de alarme <input type="checkbox"/> Vômitos persistentes <input type="checkbox"/> Aumento progressivo <input type="checkbox"/>			

Fonte: Sinan (2016).

2.2 Dados Abertos Governamentais

O termo "Dados Abertos Governamentais", ou *Open Government Data* (OGD) em inglês, se refere a uma prática de disposição de dados oficiais, de vários setores, que vem ganhando cada vez mais espaço na agenda da gestão pública brasileira (POSSAMAI; SOUZA, 2020). Em 2011, o Brasil e outros sete países assumiram um compromisso de colaboração no combate à corrupção e transparência. Por conta desse acordo foi aprovada, no Brasil, a Lei de Acesso à Informação - LAI (Lei n.º 12.527/2011) - que regulou o direito de acesso à informação previsto na Constituição Federal de 1988.

Em 2012, após essa regularização e para promover essa transparência de informações, foi criado o Portal brasileiro de Dados Abertos, uma plataforma que conta com mais de seis mil conjuntos de dados de várias áreas de interesse, que pode ser reutilizado e redistribuído por qualquer pessoa e disponível de modo gratuito a partir de fontes governamentais (KLEIN; KLEIN; LUCIANO, 2018).

Por dados abertos governamentais entendem-se os dados públicos, produzidos, coletados ou custodiados por órgãos estatais e disponibilizados na internet em formato aberto (não proprietário), estruturado e compreensível logicamente por uma máquina, de modo que qualquer um possa livremente acessar, reutilizar, modificar e redistribuir para quaisquer finalidades, estando sujeito a, no máximo, exigência de creditar a sua autoria e compartilhar sob a mesma licença (POSSAMAI; SOUZA, 2020).

Os autores OLIVEIRA et al. (2021) e Klein, Klein e Luciano (2018) destacam que uma boa governança de dados abertos, na área da saúde e em outros campos da esfera pública, permitem o acesso público e publicizável a certos grupos sem comprometer a integridade dos mesmos, esse acesso também estimula pesquisas secundárias e fomentam o desenvolvimento científico e tecnológico. A disposição desses dados governamentais de forma livre, democratiza o acesso às informações, facilita o processo de desenvolvimento de inovações e também proporciona um ambiente favorável a pesquisas e análises de dados.

2.3 Extração de conhecimento

Extração de conhecimento, ou *Knowledge Discovery in Databases* (KDD) em inglês, é o processo de transformar dados em informações. Essa área de estudo é geral e de interesse comum de diversas ciências, sua aplicação usa de fundamentos presentes na matemática, estatística, banco de dados, inteligência artificial, reconhecimento de padrões e visualização de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). O processo de extração de Conhecimento gera como produto a informação relevante que pode ser utilizada pelos tomadores de decisão, Schmitt (2005) destaca que esse conhecimento descoberto não precisa ser necessariamente incorporado a um sistema de apoio a decisão. O processo de Extração de Conhecimento é reconhecido por possuir várias etapas operacionais em sua execução. A complexidade deste processo está na necessidade de entender

e decifrar da forma correta os fatos visualizados e relacionar dinamicamente essas interpretações para decidir quais ações devem ser executadas em cada caso. Cabe ao analista humano a difícil tarefa de guiar a execução desse processo.

Com base no modelo proposto de Fayyad, Piatetsky-Shapiro e Smyth (1996) o processo de Extração de Conhecimento possui 9 fases:

A primeira fase é desenvolver e compreender o domínio da aplicação, essa etapa é o momento em que se identifica o objetivo do *KDD* do ponto de vista de um cliente.

A segunda é a criação de um conjunto de dados alvo, nesse estágio é realizada a seleção da base de dados ou do subconjunto que será trabalhado.

Na terceira fase é realizada a limpeza e o pré-processamento dos dados, agora é removido os ruídos e o dados do *dataset* que não são relevantes ou apropriados ao trabalho proposto, também são tratados os dados faltantes e a padronização dos campos e valores.

A etapa de Transformação de Dados é a quarta fase onde se busca representar os dados com base no objetivo da análise. Realizando redução do número de variáveis.

A quinta fase é a escolha do método de mineração de dados com base nos objetivos da aplicação (Por exemplo: sumarização, classificação, regressão, agrupamento, etc.).

A análise exploratória e a seleção de modelos e hipóteses é a sexta fase, fase em que se escolhe os algoritmos de *datamining* e os métodos a serem utilizados na busca de padrões de dados

A sétima é a mineração de dados: busca por padrões de interesse em uma forma representacional específica ou um conjunto de tais representações, isso inclui regras de agrupamento, classificação e regressão.

A oitava fase é o momento de se visualizar e interpretar os padrões extraídos.

E a nona fase é a etapa de ação e tomada de decisões com base no conhecimento descoberto, é nesse momento que se usa diretamente as informações descobertas e/ou incorpora o conhecimento em outro sistema de ação.

2.4 Pré-Processamento de Dados

Nesta seção será detalhado os processos envolvidos na etapa de Pré-Processamento de dados. Esta etapa tem a função de aprimorar a qualidade dos dados fazendo com que os processos de mineração de dados sejam mais eficientes (SILVA et al., 2018).

2.4.1 Qualidade dos Dados

Três elementos definem qualidade dos dados: precisão, completude e consistência. Normalmente, não se pode contar com uma base de dados inicial perfeita, é comum que elas apresentem erros humanos, falhas na coleta dos dados ou limitações nos dispositivos de medição. Também, muita das vezes a mineração de dados é aplicada em dados que foram coletados e organizados para outra finalidade, desse modo tendem a apresentar formatos de

armazenamento diferentes e conter inconsistências (TAN; STEINBACH; KUMAR, 2009). As inconsistências mais comuns são:

- Ruídos: São erros aleatórios que podem implicar na distorção de um valor ou adicionar objetos falsos. Eles podem ser gerados devido a falhas na coleta de dados, erros de entrada de dados, entre outras situações difíceis de prever.
- *Outliers*: São objetos com valores que destoam muito dos outros objetos da mesma classe, eles tendem a ter um comportamento anormal se comparado aos demais e podem influenciar a visualização dando resultados incorretos.
- Valores Ausentes: valores que são ignorados ou perdidos deixando seus campos sem informações.
- Dados duplicados: um conjunto pode conter objetos que estão duplicados, ou seja, todos os campos possuem os mesmos valores.

Para a mineração de dados, uma baixa qualidade dos dados levam a uma baixa qualidade dos resultados. Desse modo, o pré-processamento constitui-se de diversas técnicas diferentes usadas para melhorar a qualidade dos dados(SILVA et al., 2018).

2.4.2 Agregação

Segundo Silva et al. (2018), quando o volume de dados é muito grande o processamento dos dados pode se tornar mais lento e caro. A Agregação é a união de duas ou mais transações em uma única transação. Atributos quantitativos, normalmente, são agregados atribuindo uma soma dos valores ou a média, já atributos qualitativos podem ser omitidos ou agrupados em um conjunto com todos os objetos semelhantes.

2.4.3 Amostragem

Essa técnica é usada para montar subconjuntos de dados a serem examinados. Esse subconjunto é montado a partir do *dataset* analisado e deve ser representativo, isto é, ele deve conter praticamente a mesma propriedade de interesse da base de dados original. O resultado do processamento de um desses subconjuntos tende a ser mais rápido e mais barato do que o mesmo processamento do conjunto de dados completo (TAN; STEINBACH; KUMAR, 2009).

2.4.4 Redução de dimensionalidade

Quando a base de dados possui um número muito grande de atributos a análise desses dados se torna difícil, esse número de atributos também pode dificultar a visualização das informações já que essa quantidade de valores está diretamente relacionada à distribuição espacial dos objetos, logo, tendo um grande número de atributos os dados tendem a se espalhar.

O processo de redução de dimensionalidade busca reduzir as dimensões da base de dados removendo as características desnecessárias e os ruídos (BATISTA et al., 2003).

2.4.5 Seleção de subconjuntos de atributos

Uma base pode contar com atributos redundantes ou irrelevantes, uma maneira de reduzir a dimensão desse *dataset* é a atualização de subconjuntos, desse modo apenas se trabalha com os atributos de interesse (TAN; STEINBACH; KUMAR, 2009).

2.4.6 Criação de atributos

A partir dos atributos iniciais de uma base de dados, é possível criar um novo conjunto que possui todas as informações importantes presentes no *dataset* e que tenha um número menor de atributos.

2.4.7 Discretização e Binarização

A discretização é uma técnica que converte atributos contínuos em categóricos. Para realizar essa alteração é preciso estabelecer a quantidade de categorias que serão utilizadas e também como converter os valores contínuos deste atributo para essas categorias (BATISTA et al., 2003).

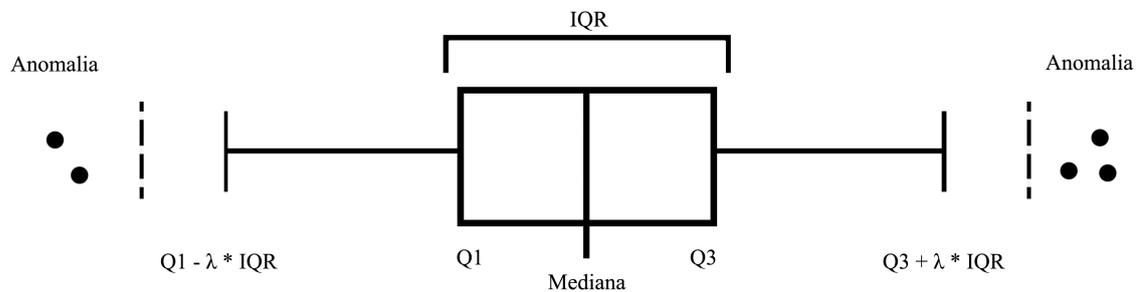
2.5 Boxplot

Boxplot é uma representação gráfica que resume a distribuição de um conjunto de dados numéricos. Ele fornece uma visão clara dos principais elementos de uma distribuição, como a mediana, os quartis, o intervalo interquartil e os valores anormais ao conjunto (outliers). Essa visualização permite identificar rapidamente a distribuição dos dados e a presença de valores atípicos.

Um *boxplot*, como ilustrado na Figura 4, é composto por uma caixa que se estende do primeiro quartil (Q1) até o terceiro quartil (Q3), com uma linha interna que marca a mediana. As "extensões", ou *whiskers*, representam geralmente o intervalo de 1,5 vezes o intervalo interquartil (IQR), que é a diferença entre Q3 e Q1. Qualquer valor que esteja fora deste intervalo é considerado um *outlier*.

O cálculo de *outliers*, no contexto de um *boxplot*, utiliza o intervalo interquartil para definir esses valores atípicos. Para determinar se um ponto de dados é um *outlier*, ele deve ser maior que $Q3 + 1,5 * IQR$ ou menor que $Q1 - 1,5 * IQR$. Valores fora desses limites são classificados como *outliers* e podem indicar dados anômalos, erros ou eventos incomuns que merecem atenção especial.

Figura 4 – Modelo do Boxplot

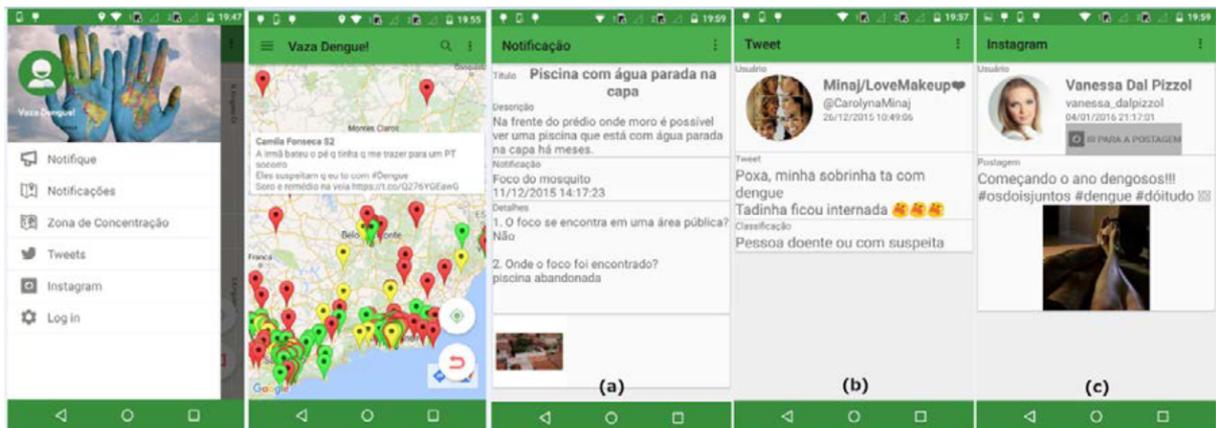


Fonte: Autoria própria.

3 TRABALHOS RELACIONADOS

Diversos trabalhos utilizam ferramentas computacionais no contexto de arboviroses, como, por exemplo, a incidência de casos através de mineração em redes sociais como descrito nos trabalhos de Fiorini et al. (2016) representado na Figura 5 e também Euzébio et al. (2020). Nestes trabalhos, é proposto um sistema que realiza a mineração de dados em uma rede social e alimenta uma plataforma de vigilância. Estes sistemas conseguem classificar e identificar esses registros; contudo, exigem uma padronização dos *posts* na rede social. Além disso, como essas notificações são feitas pelos próprios usuários, os dados apresentam uma confiabilidade reduzida, comprometendo a precisão e a abrangência das análises realizadas.

Figura 5 – Telas do Aplicativo Vaza Dengue.



Fonte: Fiorini et al. (2016)

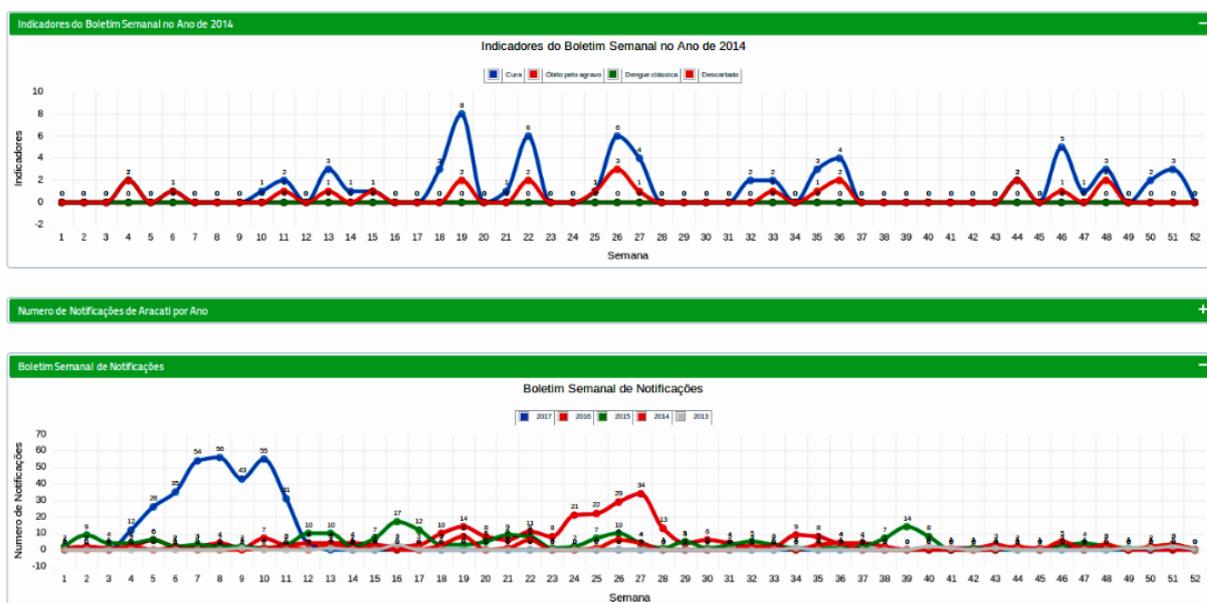
O trabalho de Baldi et al. (2017), que simula o espalhamento do mosquito e planeja a distribuição de armadilhas, realiza uma investigação detalhada sobre o comportamento do vetor de transmissão. A simulação, baseada nas construções registradas na API de mapeamento, identifica as possíveis áreas para onde os vetores podem se deslocar. Embora o estudo apresente um bom modelo preditivo para o deslocamento do mosquito, ele carece de um sistema de alimentação que indique os locais onde os casos de dengue foram efetivamente registrados. Essa limitação pode reduzir a eficácia das estratégias de combate ao vetor, já que a distribuição das armadilhas não é diretamente informada pelos dados reais de incidência da doença.

Nos trabalhos que lidam com a tendência de termos em máquinas de busca, como o de Izbicki e Cruz (2018), é possível prever surtos de infecção em determinados municípios, relacionando os possíveis casos à quantidade de pesquisas no site de busca Google. Embora esse trabalho consiga gerar indicadores para determinadas regiões, o sistema analisa apenas uma pequena fração dos infectados e não disponibiliza dados geoespaciais e temporais,

que são de extrema importância para os agentes de combate a endemias.

Bezerra et al. (2017) desenvolveu um sistema de auxílio na tomada de decisão denominado 'Dengosa', ilustrado na Figura 6. O sistema permite delimitar áreas de cobertura para unidades de saúde e, com dados alimentados, consegue indicar o comportamento do número de casos nessas áreas. No entanto, o trabalho exige que um usuário mapeie e divida as regiões de cada município, o que inicialmente limita sua cobertura. Embora o sistema possa ser adaptado para qualquer município, em grandes metrópoles, seria necessária a atuação de várias equipes para realizar essa tarefa de população das regiões no mapa. Por meio de um gráfico de linha, este sistema também permite ao usuário visualizar dados temporais e acompanhar a progressão do número de casos ao longo dos anos. No entanto, a eficácia dessa visualização é altamente dependente da escala dos dados registrados ao longo do período analisado, o que pode dificultar a detecção de outliers. Além disso, a ausência de medidas estatísticas claras, como médias ou desvios padrão, compromete a compreensão da situação real em cada intervalo de tempo, tornando a análise menos precisa e dificultando a tomada de decisões baseada em evidências.

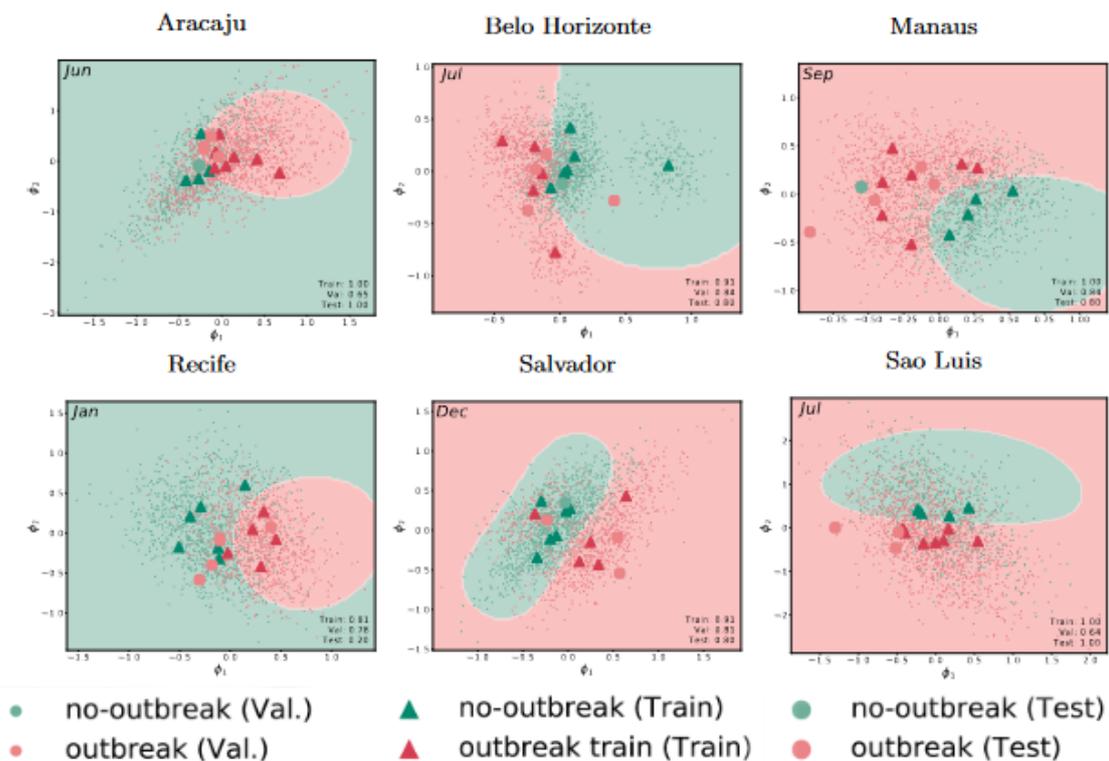
Figura 6 – Ferramenta de análise temporal do DENGOSA.



Fonte: Bezerra et al. (2017).

No trabalho Souza et al. (2022), que prevê surtos usando séries temporais de dados climáticos, é realizada uma análise que relaciona os dados climáticos ao longo do ano com os períodos de surtos de dengue em cada semana epidemiológica, estimando os possíveis momentos de explosão no número de vetores de transmissão, representado na Figura 7. Essa ferramenta é excelente para prever quando o aumento no número de infecções pode ocorrer; no entanto, não trabalha diretamente com as regiões específicas do município e não oferece uma visão detalhada das unidades e do número de casos registrados.

Figura 7 – Previsão de surto de dengue usando dados climáticos.



Fonte: Souza et al. (2022).

4 MATERIAIS E MÉTODOS

4.1 Preparação e pré-processamento do Dataset

Para o desenvolvimento desse projeto, é de extrema importância ter uma base de dados que tenha passado por todas as etapas de pré-processamento e tenha a melhor qualidade de dados possível. A qualidade desse dataset influencia bastante o desenvolvimento, e definir os parâmetros desse pré-processamento logo de início é a pedra angular para evitar problemas durante as várias etapas do desenvolvimento do sistema, e com isso perda de tempo útil.

O trabalho foi desenvolvido em Python, v3.11, juntamente com as bibliotecas Numpy v1.23.4, Pandas v1.5.1, Geopy v2.3.0, Selenium v4.14.0, Plotly v5.15, Dash v2.10.2, dash_bootstrap_components v1.4.1 e dash_bootstrap_templates v1.0.8. No desenvolvimento desse trabalho foram utilizados dados de notificações de casos suspeitos e confirmados de Dengue e Chikungunya, de todos os 26 estados do Brasil, incluindo o Distrito Federal, coletados por Neto (2022) da plataforma do Sistema de Informação de Agravos de Notificação(SINAN), no período de 2013 a 2020. O dataset original possui dados clínicos, resultados de testes laboratoriais e dados sócio-demográficos.

Os dados de todos os estados foram agregados, resultando em 13.421.230 notificações e 118 atributos. No pré-processamento inicial de Neto (2022), os registros foram agrupados em três grupos distintos com base no atributo CLASSI_FIN:

- Dengue: pacientes com Dengue confirmada.
- Chikungunya: pacientes com Chikungunya confirmada.
- Descartados/Inconclusivos: pacientes com resultados negativos ou inconclusivos para Dengue ou Chikungunya.

Foram selecionadas, apenas notificações confirmadas ou descartadas/inconclusivas após diagnóstico clínico. O critério de confirmação seguiu as definições do Ministério da Saúde do Brasil. A seguir foi realizado um processo de limpeza e Transformação de Dados. Atributos com mais de 60% de dados nulos ou que não estavam no dicionário de dados original foram removidos. Campos nulos restantes foram preenchidos com o valor padrão “não informado”. Também foi realizado um processo de transformação de dados categóricos para numéricos. Após o pré-processamento, o conjunto de dados consistia em:

- 4.307.513 registros para Dengue.
- 325.000 registros para Chikungunya.
- 2.100.029 registros para a categoria Descartados/Inconclusivos.

Com essa base de dados inicialmente pré-processada, nesse trabalho, foi preciso selecionar os registros pertinentes e reduzir a dimensão dos campos, para o melhor uso pela ferramenta proposta. A primeira etapa foi selecionar apenas os casos registrados

com a confirmação da infecção, para visualizar os casos de modo gráfico é preciso que só sejam exibidos os casos confirmados e que não seja possível que o usuário faça uma análise incorreta, nessa etapa o número de registro caiu para 6.732.542. O segundo passo desse pré-processamento é a remoção dos campos com valores nulos e à atributos desconhecidos. A consistência da base de dados é de extrema importância para a montagem da ferramenta gráfica e a distribuição geográfica dos valores. Nesse ponto, todos os casos preenchidos de forma incompleta e os que possuíam valores não reconhecidos, mesmo sendo casos confirmados das infecções, foram removidos.

Em seguida, foi realizada a seleção dos dados que vão ser utilizados no trabalho, reduzindo o número de registros para os casos reportados em determinado município escolhido e onde os casos da infecção também são desse município. De todos esses atributos da base de dados, apenas três foram necessários para a metodologia desenvolvida: Id do município no IBGE (ID_MUNIC); semana epidemiológica de notificação do caso (SEM_NOT); identificação da UBS (ID_UNID) no Cadastro Nacional de Estabelecimentos de Saúde (CNES). Em seguida, todos os registros foram agregados em uma tabela dinâmica. A função utilizada agrupa os dados do DataFrame com base nas colunas 'ID_UNIDADE' e 'SEM_NOT'. Cada combinação única desses valores se tornará uma linha na nova tabela. A função de agregação 'aggfunc='size' conta o número de ocorrências de cada combinação dos valores de 'ID_UNIDADE' e 'SEM_NOT' e salva essa quantidade em uma nova coluna nomeada como 'COUNT' (Algoritmo 1).

Algoritmo 1: Processo de Agregação.

```
1 def db_count(nome_cidade_conc):
2     dfc = pd.read_csv(os.path.join("MontagemDF\ArquivoGerado", "dataset_" +
3         nome_cidade_conc + ".csv"), encoding="ISO-8859-1", sep=',',
4         index_col=None)
5
6     dfc2 = dfc.pivot_table(index = ['ID_UNIDADE', 'SEM_NOT'], aggfunc = 'size')
```

Fonte: Autoria própria.

Logo, todo o desenvolvimento do trabalho será com o foco em apresentar uma visualização gráfica de um município por vez.

O sistema possibilita essa preparação da base de dados para qualquer município brasileiro que possui pelo menos um registro no SINAN, basta que o usuário tenha o código do IBGE da cidade em questão.

4.2 Desenvolvimento do sistema

O desenvolvimento da ferramenta proposta requer a integração precisa de dados geográficos das unidades de saúde à base de dados de casos registrados. Inicialmente, a base de dados foi construída a partir de casos isolados de Dengue e Chikungunya, cada um associado a uma unidade específica por meio do código identificador de unidade de saúde. Contudo, uma limitação evidente foi a ausência das coordenadas geográficas (latitude e longitude) dessas unidades.

Para superar essa lacuna, foi implementado um processo de obtenção das coordenadas geográficas das unidades de saúde diretamente do portal do CNES, recorte da página representado na Figura 8.

Figura 8 – Dados disponíveis na plataforma do Cnes, de onde foi feita a coleta do endereço.

Identificação							
Nome	UPA DR PAULO CESAR DE CARVALHO TELLES	CNES	2997045	CNPJ	---		
Nome Empresarial	MUNICIPIO DE RIO VERDE	Natureza Jurídica(Grupo)			ADMINISTRAÇÃO PÚBLICA		
Logradouro	RUA 28	Número	S/N	Complemento	Q17 AL02		
Bairro	VILA SANTA CRUZ I	Município	521880 - RIO VERDE	UF	GO		
CEP	75907-100	Telefone	--	Dependência	MANTIDA	Regional de Saúde	001
Tipo de Estabelecimento	PRONTO ATENDIMENTO	Subtipo de Estabelecimento	UPA	Gestão	MUNICIPAL		
Diretor Clínico/Gerente/Administrador						MARCOS VINICIUS MEIRA VAZ	
Cadastrado em	14/08/2022	Atualização na Base Local	07/11/2023	Última atualização Nacional	13/06/2024		
Horário de funcionamento						Sempre aberto	
Data Desativação		Motivo Desativação					

Fonte: CNES (2024)

Foi desenvolvido um *script* em Python que realiza um processo de *web crawling* para coletar endereços de unidades de saúde no site do CNES (Cadastro Nacional de Estabelecimentos de Saúde). Utilizando a biblioteca Selenium, o *script* automatiza a navegação na web e extrai informações específicas das páginas de cada unidade de saúde, identificadas por meio de um ID único. O processo começa configurando o navegador e acessando a URL específica da unidade de saúde. Em seguida, o *script* aguarda até que o elemento da página contendo o endereço esteja visível e então extrai os valores dos campos de logradouro, número e bairro, seguindo o código do Algoritmo 2.

Algoritmo 2: Processo de *web crawling*.

```
1 #Construcao do Url utilizando os parametros fornecidos
2 def get_nome(ID_MUNICIP, ID_UNIDADE):
3     url = "https://cnes.datasus.gov.br/pages/estabelecimentos/ficha/
4     index.jsp?coUnidade=" + str(ID_MUNICIP) + str(ID_UNIDADE)
5     options = Options()
6     browser = webdriver.Chrome()
7     browser.implicitly_wait(2)
8     browser.get(url)
9
10 #Procura aos elementos na pagina
11     try:
12         element = WebDriverWait(browser, 10).until(
13             EC.visibility_of_element_located((By.XPATH,
14                 '/html/body/div[2]/main/div/div[3]/div[1]/div/
15                 section/div[3]/div/div[2]/div[1]/div/form/div[3]
16                 /div[1]/div/input')))
17         logradouro = browser.find_elements (By.XPATH, 'Caminho do elemento 01')
18         numero = browser.find_elements (By.XPATH, 'Caminho do elemento 01')
19         bairro = browser.find_elements (By.XPATH, 'Caminho do elemento 01')
20
21     except Exception as e:
22         print("Elemento nao encontrado ou nao visivel. Ocorreu um erro:", str(e))
23
24 #Atribuicao dos valores retornados
25     if len(logradouro):
26         for x in logradouro:
27             aux1 = x.get_attribute('value')
28         for y in numero:
29             aux2 = y.get_attribute('value')
30         for z in bairro:
31             aux3 = z.get_attribute('value')
32         aux = aux1 + ', ' + aux2 + ', ' + aux3
33         print('Encontrada: ' + aux + ' Unidade: ' + str(ID_UNIDADE))
34         return(aux)
35     else:
36         print('Unidade nao encontrada:' + str(ID_MUNICIP) + ', ' + str(ID_UNIDADE)
37             + ' link: https://cnes.datasus.gov.br/pages/estabelecimentos/ficha/
38             index.jsp?coUnidade=' + str(ID_MUNICIP) + str(ID_UNIDADE))
```

Caso o endereço seja encontrado, ele é concatenado e exibido juntamente com o ID da unidade. Se o elemento não for encontrado ou não estiver visível, uma mensagem de erro é gerada. O *script* é eficaz em automatizar a coleta de dados de múltiplas páginas, o que é essencial para a construção de um banco de dados com informações atualizadas das unidades de saúde em diferentes municípios.

Com os dados de endereço disponíveis, a biblioteca Geopy foi aplicada para transformar de forma eficiente essas informações em coordenadas geográficas. Esse passo crucial adicionou uma camada valiosa à base de dados, fornecendo a localização espacial exata de cada unidade de saúde. A abordagem integrada de *web crawling*, automação e geocodificação possibilitou a construção de uma base de dados rica e geograficamente informada, essencial para a montagem dos gráficos.

O modelo de visualização gráfica escolhido inicialmente foi o *heatmap* que é uma representação visual de dados em que os valores em uma matriz são representados por cores. Ele é frequentemente utilizado para mostrar a distribuição de valores em uma grade bidimensional, onde cores diferentes representam diferentes níveis de intensidade ou magnitude. O *heatmap* é uma forma eficaz de destacar padrões e variações em grandes conjuntos de dados, facilitando a identificação de áreas de concentração ou tendências.

Por conta da limitação dos dados, e, que nos registros dos casos das doenças virais não constavam os dados do local de infecção, a montagem da matriz para o *heatmap* se tornou muito difícil. Logo, para uma base de dados com valores relacionados diretamente à unidade de saúde, a proposta de visualização foi alterada para o modelo Scattermapbox. Scattermapbox é um tipo específico de gráfico de dispersão (*scatter plot*) que utiliza a tecnologia Mapbox para criar gráficos interativos em um contexto geoespacial. O Scattermapbox combina a flexibilidade de um gráfico de dispersão tradicional com a capacidade de representar pontos em um mapa interativo. O Scattermapbox é especialmente útil quando os dados a serem visualizados têm uma componente geoespacial. Ele permite que os pontos sejam representados em um mapa utilizando as coordenadas de latitude e longitude.

No desenvolvimento desse trabalho foi utilizado o Plotly, que é uma biblioteca de visualização de dados em Python que suporta uma variedade de gráficos interativos, como gráficos de dispersão, barras, linhas e mapas. É integrado ao Dash para criar visualizações dinâmicas e interativas em aplicativos web.

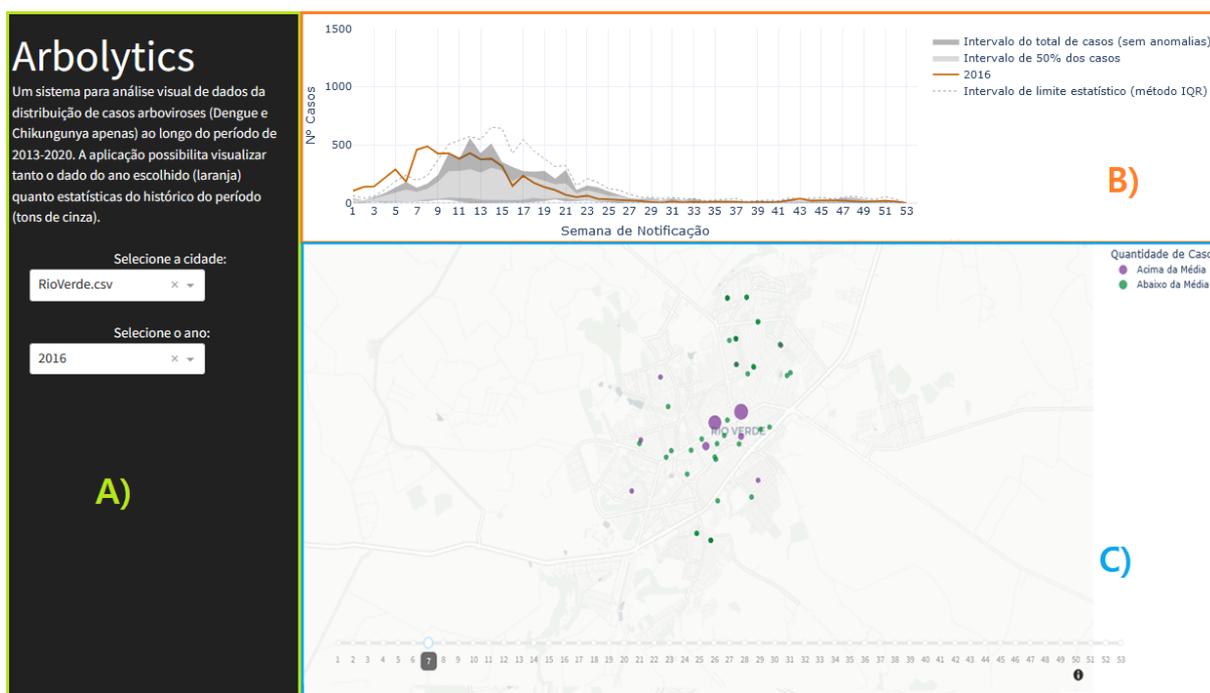
4.3 *Arbolytics*

A ferramenta proposta é denominada *Arbolytics*. Ela difere de Bezerra et al. (2017) em dois aspectos: possibilita a visualização de qualquer município que possui casos notificados no SINAN no período de 2013-2020 (Figura 9A); utiliza histórico de dados

de anos anteriores para gerar estatísticas comparativas e detectar possíveis anomalias (Figura 9B). Adicionalmente, pode explorar os detalhes da distribuição de notificações no município filtrado por semana epidemiológica (Figura 9C).

O sistema proposto foi dividido em duas partes: a primeira é responsável pelo pré-processamento e preparação dos arquivos .csv, enquanto a segunda cuida da construção dos gráficos e da aplicação das medidas estatísticas.

Figura 9 – *Arbolytics* ilustrando a visão combinada entre a distribuição de casos e sua distribuição espacial para o município de Rio Verde-GO, em 2016.



Fonte: Autoria própria.

4.3.1 Medidas estatísticas e o Gráfico de Linha

Para a criação das medidas estatísticas são considerados os dados de notificações de uma determinada semana epidemiológica utilizando todos os dados no período de 7 anos disponível. A partir disso, foi utilizado o método proposto por Tukey et al. (1977).

A primeira estatística calculada refere-se à distribuição de casos ao longo de uma medida de tendência central (mediana). Nesse contexto, foi calculado o intervalo interquartil (IQR) entre o primeiro e terceiro quartis (Q_1 e Q_3), indicando em quais valores de incidência se distribuem 50% dos casos daquela semana, Algoritmo 3. A segunda medida estatística calcula o intervalo onde estão variando todos os casos no período (desconsiderando as anomalias), ou seja, os quartis Q_0 e Q_4 nos sete anos de dados.

Algoritmo 3: Função de cálculo dos intervalos de Q1 à Q3

```
1 #Funcao de Calculo dos Intervalos
2
3 def calcular_q0(valores):
4     return np.quantile(valores, 0)
5
6 def calcular_q1(valores):
7     v = valores.values
8     median = np.median(v)
9     q1 = np.median(np.array(v[np.where(v <= median)]))
10    return q1
11
12 def calcular_q3(valores):
13    v = valores.values
14    median = np.median(v)
15    q3 = np.median(np.array(v[np.where(v >= median)]))
16    return q3
```

Fonte: Autoria própria.

A terceira medida estatística permite encontrar anomalias (*outliers*) na quantidade de casos na semana. Uma forma de defini-las considera *outliers* as medidas que se distanciem de um determinado fator ($\lambda \cdot IQR$) dos primeiro e terceiro quartis (Q_1 e Q_3), isto é, será uma anomalia se o valor estiver fora do intervalo $[Q_1 - \lambda \cdot IQR, Q_3 + \lambda \cdot IQR]$, definindo o que chamamos de *intervalo de limite estatístico*. O valor tipicamente utilizado na literatura é de $\lambda = 1,5$, sendo este o utilizado neste trabalho. Nos casos onde isso implicar valores negativos, o mínimo de 0 casos é utilizado, como descrito no Algoritmo 4.

Algoritmo 4: Cálculo do intervalo interquartil

```
1 # Funcao de Calculo dos Intervalos
2 # Calculo do limite inferior
3 def calcular_min_fence_gauss(valores):
4     k = 1.5
5     q1 = calcular_q1(valores)
6     q3 = calcular_q3(valores)
7     iqr = q3 - q1
8     return max(0, q1 - k * iqr)
9
10 # Calculo do limite superior
11 def calcular_max_fence_gauss(valores):
12     k = 1.5
13     q1 = calcular_q1(valores)
14     q3 = calcular_q3(valores)
15     iqr = q3 - q1
16     return q3 + k * iqr
```

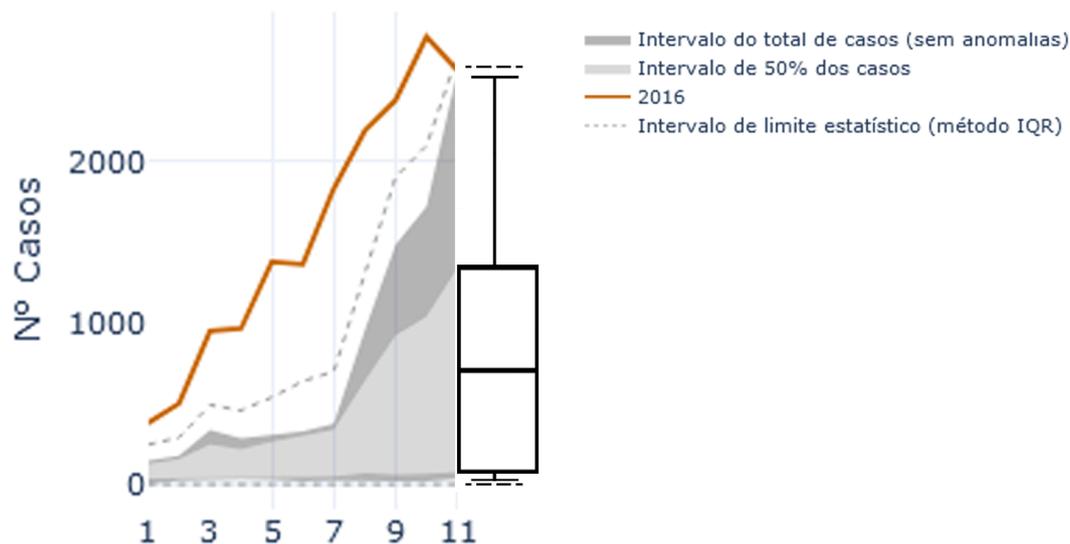
Fonte: Autoria própria.

Ao combinar essas medidas estáticas, que podem ser resumidas em um *boxplot*, e exibir esses dados em um gráfico de linha nos temos a região de acompanhamento de evolução do número de casos ao longo do período do *dataset* (Figura 9B.)

A relação do *mapbox* com as informações temporais do gráfico de linha pode ser visualizada na Figura 10.

4.3.2 Mapbox

Para evidenciar a distribuição espacial dos casos notificados, foi utilizado um *scatterplot* integrado ao componente *mapbox* do Plotly. A função utilizada na geração desse elemento foi o `scatter_mapbox()` e, da base de dados, foram usadas como parâmetro as variáveis `LATITUDE` e `LONGITUDE` para definir a posição do marcador no mapa, `NUMERO_DE_CASOS` e a variável com o número de casos na semana anterior para definir os elementos quantitativos, como mostrado no Algoritmo 5.

Figura 10 – Relação do gráfico de linha em paralelo ao *boxplot*.

Fonte: Autoria própria.

Algoritmo 5: Função de construção do Mapbox

```
1 figura_mapa = px.scatter_mapbox(filtered_df, lat="LATI", lon="LONG",
    color="QUALI", size="SEMANA_ANTERIOR", mapbox_style="carto-positron",
    color_discrete_map=color_discrete_mapx, zoom=12, size_max= 15,
    hover_data=['COUNT', 'SEMANA_ANTERIOR', 'MEDIA_SEMANAL'])
```

Fonte: Autoria própria.

Para cada unidade de saúde onde há notificação de casos foi criado um disco centrado em suas coordenadas geográficas com um raio variando entre duas categorias, codificando a quantidade de casos na semana epidemiológica anterior à analisada, como pode ser visto na Figura 11.

O canal de cor do disco codifica a quantidade de casos da semana epidemiológica atual, ele pode assumir dois valores distintos, sendo roxo para número de casos acima da média semanal e verde quando o número de casos está abaixo ou na média semanal. Foi considerado que a unidade está *na média* se sua quantidade de casos não ultrapassa 15% do valor médio de cada semana desses sete anos calculado por cada unidade de saúde do município.

O mapbox também conta com um seletor de semana epidemiológica onde o usuário consegue escolher a semana a ser analisada.

Figura 11 – Exibição detalhada do mapbox.



Fonte: Autoria própria.

5 Resultados

A divisão do sistema em duas parte foi estrategicamente planejada para reduzir o tempo de processamento e melhorar sua usabilidade. A etapa de coleta de dados geoespaciais é demorada e aumenta proporcionalmente à dimensão do município e ao número de unidades de saúde. Assim, após a criação do arquivo .csv contendo os registros de casos e as informações geoespaciais das UBSs, a segunda parte do sistema foca apenas na montagem das telas. Com o arquivo pronto, o sistema pode, após uma verificação rápida, avançar diretamente para a construção dos gráficos, tornando a aplicação mais eficiente e prática.

Para obter uma noção do custo de tempo da etapa de montagem do arquivo .csv, que será utilizado na construção das ferramentas de visualização, a Tabela 1 apresenta uma tabela com a quantidade de unidades de saúde do município selecionado para análise, o tempo para a conclusão do pré-processamento básico e a seleção da cidade(Préproc), o tempo total necessário até a conclusão da etapa dedicado à coleta das coordenadas de latitude e longitude de todas as unidades de saúde do município(Coleta), além do tempo total de execução para a montagem completa do arquivo .csv, pronto para a construção das ferramentas gráficas.

Tabela 1 – Relação de tempo de processamento nas diversas etapas do sistema.

Município	Qtd de UBSs	Préproc (s)	Coleta (s)	Total (s)
Salto de Pirapora-SP	11	65.65	82.14	147.79
Rio Verde-GO	52	73.85	365.75	439.60
São Paulo-SP	943	122.36	6634.85	6757.21

Fonte: Autoria própria.

Com base nos tempos de execução observados para os três municípios, pode-se notar uma escalabilidade diretamente proporcional ao número de Unidades Básicas de Saúde. Em Salto de Pirapora, com 11 UBSs, o tempo total de processamento foi de 147,79 segundos, resultando em um tempo médio de 7,46 segundos por UBS. Enquanto em Rio Verde, com 52 UBSs, esse valor aumentou para 439,60 segundos, com o em um tempo médio de 7,03 segundos por UBS. Já em São Paulo, com 943 UBSs, o tempo total de execução chegou a 6757,21 segundos, resultando em um tempo médio de 7,03 segundos por unidade. A etapa de coleta de coordenadas geoespaciais representou a maioria do tempo total em todos os casos, sendo responsável por cerca de 55% do tempo total em Salto de Pirapora, 83% em Rio Verde, e 98% em São Paulo. Isso evidencia que o tempo de processamento escala consideravelmente à medida que a quantidade de UBSs aumenta, especialmente na fase de coleta de dados geoespaciais. No entanto, a etapa de pré-processamento, que envolve a preparação inicial dos dados, demonstrou uma variação menos acentuada entre

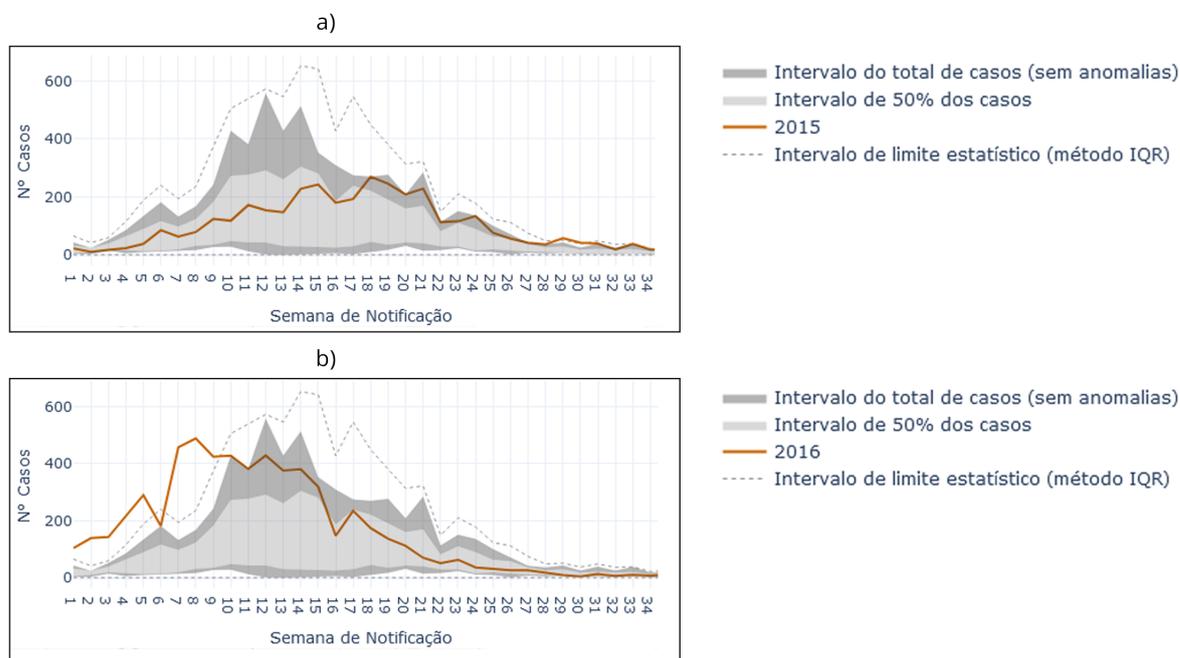
os municípios, indicando que essa parte do sistema é mais eficiente independentemente do tamanho do conjunto de dados. Em resumo, o sistema mostra um comportamento escalável, e a divisão em duas partes garante que, uma vez que os dados estejam coletados, o restante do processo seja executado de forma otimizada, mesmo para municípios maiores como São Paulo.

5.1 Rio Verde-GO

Nas Figuras 12 e 13, geradas para o município de Rio Verde-GO referentes aos anos de 2015 e 2016, é possível identificar certos padrões no comportamento e na distribuição dos casos de dengue na região.

Na Figura 12, podemos comparar esses dois anos. Em 2015, observa-se um ano sem anomalias, com a distribuição dos casos dentro do intervalo considerado normal ou esperado para cada semana (indicado pela linha laranja). Por outro lado, em 2016, nota-se que, da semana 01 até a semana 09, exceto na semana 06, o número de casos está consistentemente acima do esperado. Essas semanas são classificadas como outliers, indicando surtos de dengue, com registros acima do previsto para esse período do ano.

Figura 12 – Casos notificados por semana epidemiológica em Rio Verde-Go. Em laranja a série temporal de 2015 (a) e 2016 (b).

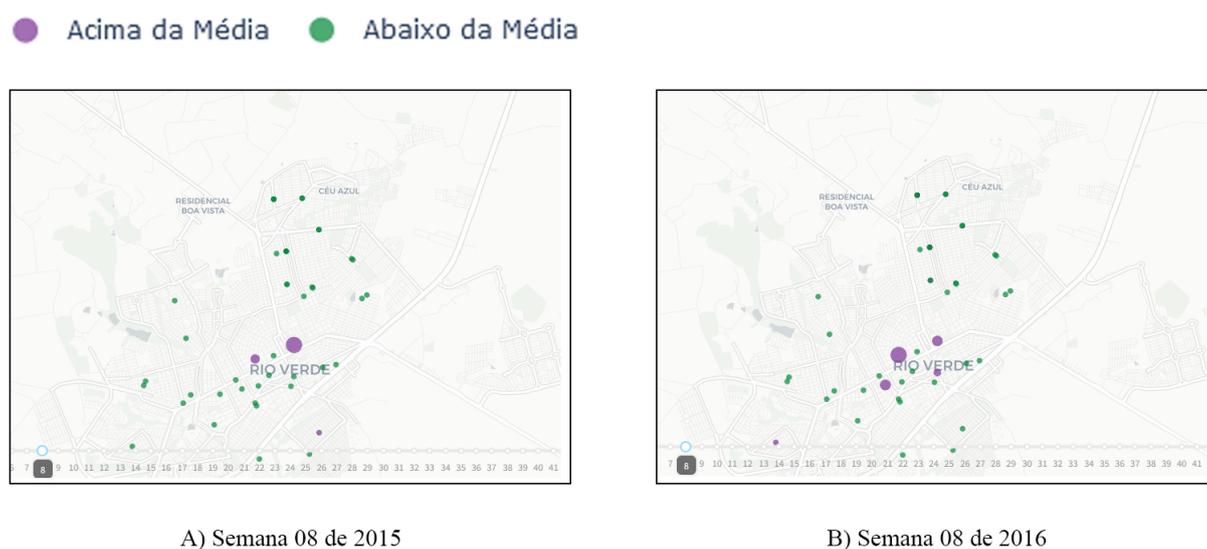


Fonte: Autoria própria.

Continuando a análise do município de Rio Verde-GO, a Figura 13 apresenta a distribuição espacial dos casos de dengue na semana 08 dos anos de 2015 e 2016. O Mapbox destaca as unidades que registram o maior número de casos, assim como aquelas

que mostram um aumento nas notificações em comparação com a semana anterior. É importante notar que essa visualização do Mapbox é focada exclusivamente no ano em questão e não se relaciona diretamente com outros períodos do conjunto de dados. Uma observação relevante é que a Unidade de Pronto Atendimento (UPA) da cidade, que atende pacientes de toda a região, constantemente registra um número de casos acima da média. Isso ocorre devido à alta demanda na unidade, que atende pacientes de diversas partes da cidade.

Figura 13 – Distribuição dos casos por semana epidemiológica em Rio Verde-Go nos anos de 2015 e 2016



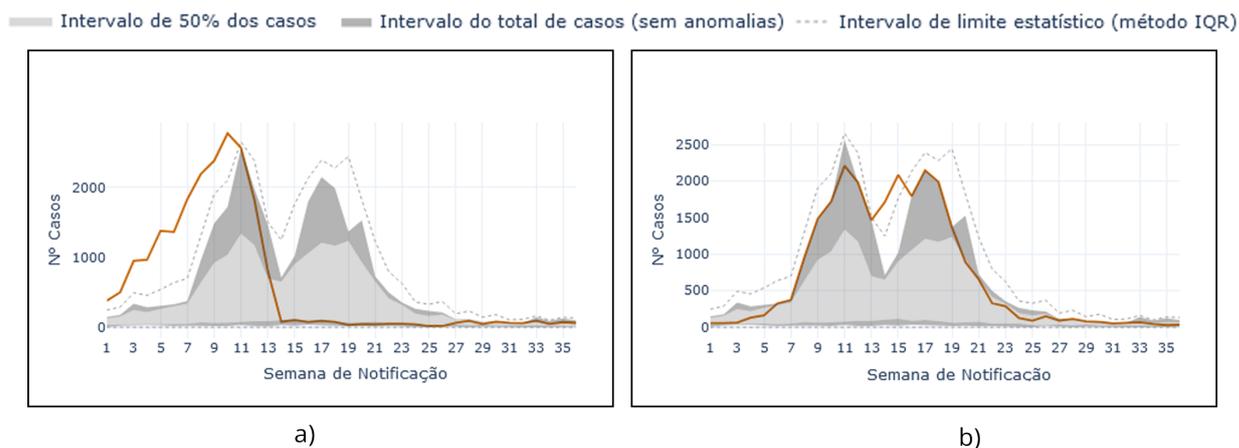
Fonte: Autoria própria.

5.2 São Paulo-SP

Analisando a cidade de São Paulo-SP, ilustrado na Figura 14, pode-se observar nitidamente que no ano de 2014 houve um comportamento anômalo entre a primeira e nona semana epidemiológicas, indicando uma quantidade de notificações bastante acima da distribuição de casos no período 2013-2020. Ao passo que no ano de 2016, o comportamento segue o esperado para a maioria das semanas, a menos da semana 15.

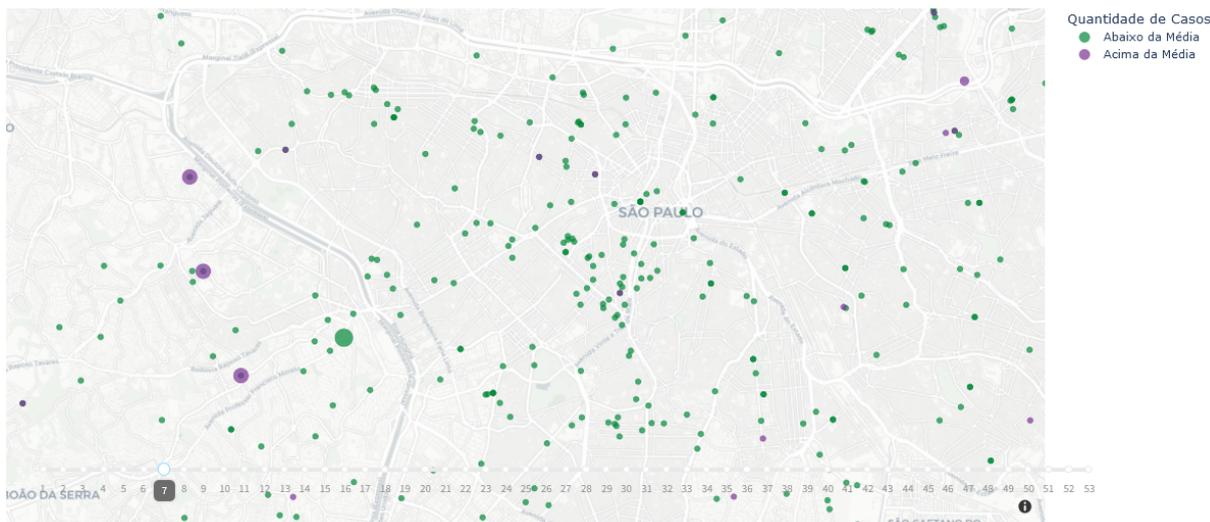
Na Figura 15, observa-se a distribuição das unidades de saúde, destacando-se a quantidade de casos registrados por unidade, indicando se estão acima ou abaixo da média semanal, além de mostrar como o número de casos se relaciona com a semana anterior. Um exemplo a ser observado é quando o tamanho do círculo aumenta, indicando um crescimento no número de casos, mas a cor permanece como abaixo da média. Esse cenário sugere uma variação nos registros, embora o aumento ainda esteja dentro da faixa esperada para a unidade.

Figura 14 – Casos notificados por semana epidemiológica em São Paulo-SP. Em laranja a série temporal de 2014 (a) e 2016 (b).



Fonte: Autoria própria.

Figura 15 – Distribuição das notificações na semana epidemiológica 7 em São Paulo-SP.



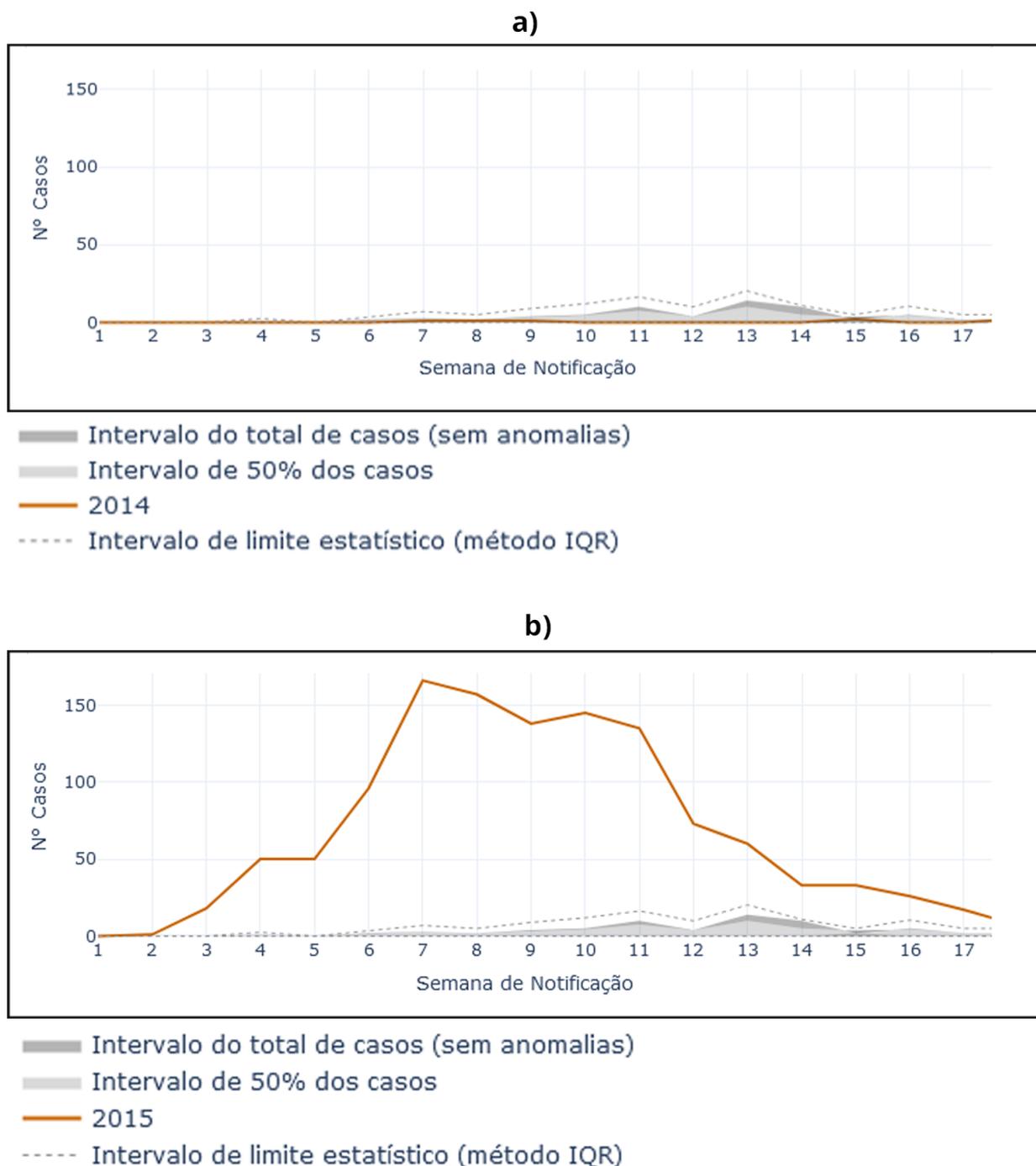
Fonte: Autoria própria.

5.3 Salto de Pirapora-SP

Um exemplo notável, que pode ser acompanhado através da nossa ferramenta, é o surto de dengue no município de Salto de Pirapora, São Paulo, entre as semanas 1 e 9 do ano de 2015, representado na Figura 16. Conforme a notícia do Diário do Grande ABC, um levantamento feito pelo jornal O Estado de S. Paulo, baseado em dados da Secretaria Estadual da Saúde, revelou que em apenas dois meses de 2015, um terço das cidades paulistas registrou mais casos de dengue do que em todo o ano de 2014. Enquanto em 2014 foram registrados apenas dois casos, entre janeiro e fevereiro de 2015 os números saltaram para 559, resultando em uma taxa de incidência que subiu de 4,8 para 1.322

casos por 100 mil habitantes, um aumento de 27.555%.

Figura 16 – Casos notificados por semana epidemiológica em Salto de Pirapora-SP. Em laranja a série temporal de 2014 (a) e 2015 (b).



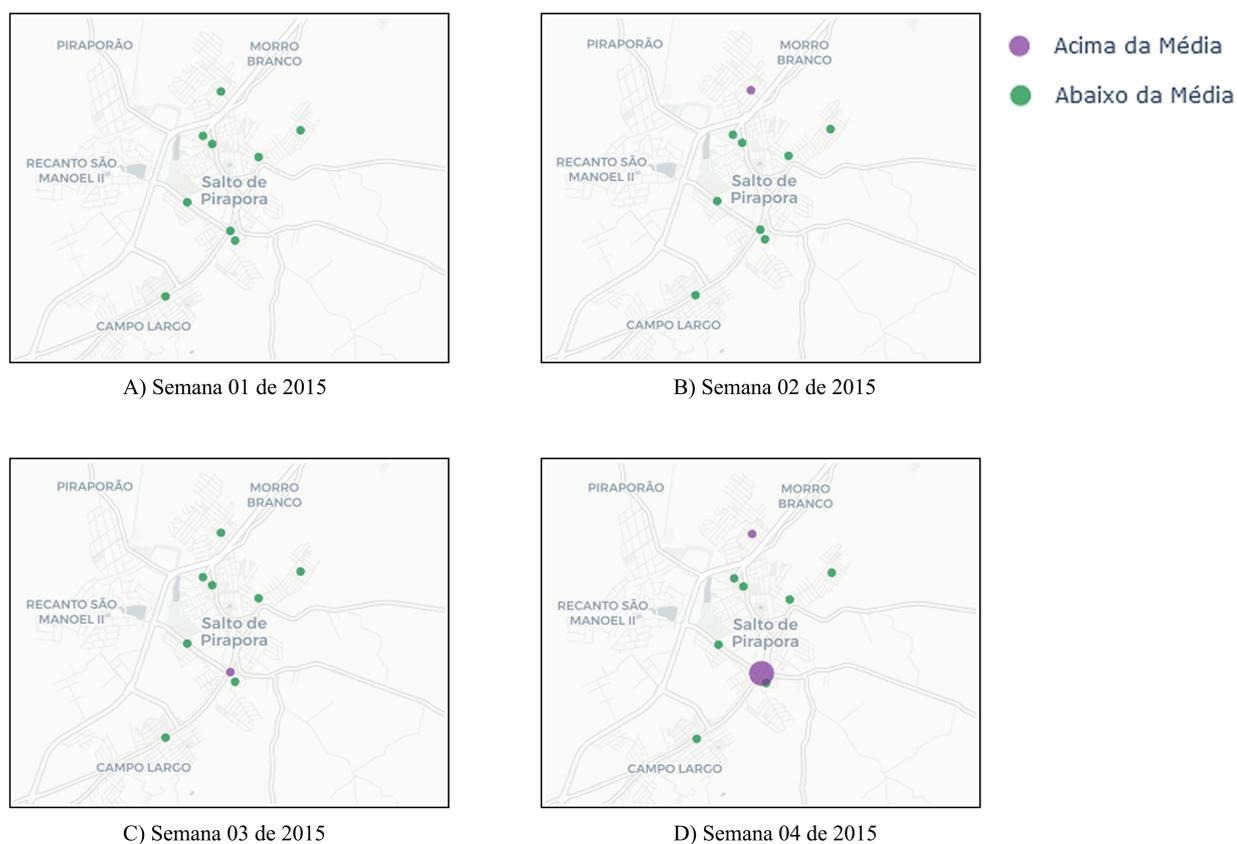
Fonte: Autoria própria.

Analisando a Figura 17 podemos acompanhar a progressão no número de casos registrados no município de Salto de Pirapora ao longo das quatro primeiras semanas do ano de 2015. Por ser uma cidade pequena e com um baixo histórico de número de registros,

logo na semana 02, um único registro já aparece no mapbox como um valor acima da média geral.

Na progressão da semana 03 para a semana 04 é possível notar uma nova unidade com registros acima da média e na semana 04 podemos observar que a unidade se mantém com um número alto e, por conta do tamanho do ponto que representa a unidade, que esse valor se manteve desde a última semana.

Figura 17 – Distribuição dos casos por semana epidemiológica em Salto de Pirapora-SP no ano de 2015.



Fonte: Autoria própria.

6 CONCLUSÃO

A combinação de informações geográficas, temporais e quantitativas na ferramenta oferece uma visão holística da situação epidemiológica, fortalecendo a capacidade de resposta rápida e a tomada de decisões no combate as arboviroses. A ferramenta fornece uma interface intuitiva e interativa tanto para os profissionais de saúde quanto gestores visualizarem os dados de arboviroses, que pode facilitar a tomada de decisões informadas e o planejamento estratégico para o combate e prevenção dessas doenças.

A primeira versão deste trabalho foi publicada e premiada com o título 'Arbolytics: Monitoramento de dados de arboviroses utilizando estatísticas de notificações de casos e distribuição espacial' no concurso de Trabalhos de Iniciação Científica do Simpósio Brasileiro de Computação Aplicada à Saúde (ARAUJO; CEDRIM, 2024). A partir dessa publicação inicial, realizamos várias iterações para refinar a ferramenta, como, por exemplo, a classificação dos pontos no Mapbox em apenas duas categorias, de três, acima e abaixo da média do número de casos por unidade de saúde.

Como trabalhos futuros, pode-se considerar a integração de dados em tempo real, o que poderia oferecer um monitoramento mais dinâmico e a detecção precoce de surtos, adicionando uma camada valiosa de informação. Além disso, seria interessante incorporar as características do atendimento das unidades de saúde, já que algumas unidades, como a UPA, atendem pacientes de qualquer região do município, enquanto as UBS atendem apenas os pacientes de sua área de cobertura, e algumas unidades prestam atendimento privado. A inclusão dessas informações tornaria a análise mais precisa, refletindo melhor o impacto do atendimento em diferentes regiões. Adicionalmente, outras variáveis epidemiológicas, sociais e ambientais poderiam ser incluídas, com o apoio de especialistas. Por fim, estratégias adicionais de estimação de outliers podem ser exploradas para aumentar a robustez das análises.

Como próximo passo, propõe-se a validação do programa por um especialista em saúde pública. Para essa avaliação, a escala de Likert é uma opção possível para medir a usabilidade e a eficácia da ferramenta no contexto epidemiológico. Com uma estrutura simples e organizada, a escala permite que o profissional avalie seu grau de concordância ou discordância em relação a diferentes aspectos do sistema, como clareza, eficiência e aplicabilidade prática. Essa metodologia facilita a obtenção de um feedback objetivo e mensurável, destacando tanto os pontos fortes quanto as oportunidades de melhoria (SOUTH et al., 2022).

Referências

ARAUJO, A. B. de. Uma análise das políticas de controle e combate à dengue no brasil. 2018. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/33931>>. Citado 2 vezes nas páginas 6 e 7.

ARAUJO, I.; CEDRIM, D. Arbolytics: Monitoramento de dados de arboviroses utilizando estatísticas de notificações de casos e distribuição espacial. In: *Anais Estendidos do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2024. p. 25–30. ISSN 2763-8987. Disponível em: <https://sol.sbc.org.br/index.php/sbcas_estendido/article/view/28697>. Citado na página 33.

BALDI, A. et al. Simulação de aplicação de armadilhas no combate ao *Aedes aegypti*. In: *Simpósio brasileiro de computação aplicada à saúde (SBCAS)*. Porto Alegre, RS, Brasil: SBC, 2017. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/3710>>. Citado na página 14.

BATISTA, G. E. d. A. P. et al. *Pré-processamento de dados em aprendizado de máquina supervisionado*. Tese (Doutorado) — Universidade de São Paulo, 2003. Citado na página 12.

BEZERRA, A. et al. Dengosa: Um sistema de informação geográfica para apoio à decisão no controle de epidemias. In: SBC. *XXIII Simpósio Brasileiro de Sistemas Multimídia e Web*. [S.l.], 2017. p. 179–183. Citado 2 vezes nas páginas 15 e 21.

CNES. *Plataforma Cnes*. 2024. [Online; accessed August 27, 2024]. Disponível em: <<https://cnes.datasus.gov.br/pages/estabelecimentos/consulta.jsp?search=2042967>>. Citado na página 19.

EPIDEMIOLOGICA, D. de V. Doenças infecciosas e parasitárias : Guia de bolso. *Ministério da Saúde*, 2010. Citado na página 5.

EUROGEOGRAPHICS. *Dengue worldwide overview Geographical distribution of dengue cases reported worldwide, March to May 2022*. 2022. Disponível em: <<https://www.ecdc.europa.eu/en/dengue-monthly>>. Citado na página 4.

EUZEBIO, C. et al. Statistical analysis of small twitter data collection to identify dengue outbreaks. In: *VIII Symposium on Knowledge Discovery, Mining and Learning*. Porto Alegre, RS, Brasil: SBC, 2020. p. 17–24. ISSN 2763-8944. Disponível em: <<https://sol.sbc.org.br/index.php/kdmile/article/view/11954>>. Citado na página 14.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37–37, 1996. Citado 2 vezes nas páginas 9 e 10.

FIORINI, S. et al. Vigilância entomológica da dengue, zika e chikungunya: Uma solução baseada em redes sociais e dispositivos móveis. In: *Simpósio brasileiro de computação aplicada à saúde (SBCAS)*. Porto Alegre, RS, Brasil: SBC, 2016. p. 2567–2576. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/9903>>. Citado na página 14.

- IZBICKI, R.; CRUZ, L. O. d. Monitoramento online da dengue: Usando o google para prever epidemias. *Brazilian Journal of Biometrics*, v. 36, n. 3, p. 512–526, Sep. 2018. Disponível em: <<https://www.cerne.ufla.br/index.php/BBJ/article/view/227>>. Citado na página 14.
- KLEIN, R. H.; KLEIN, D. C. B.; LUCIANO, E. M. Identificação de mecanismos para a ampliação da transparência em portais de dados abertos: uma análise no contexto brasileiro. *Cadernos Ebape. br*, SciELO Brasil, v. 16, p. 692–715, 2018. Citado na página 9.
- LOPES, N.; NOZAWA, C.; LINHARES, R. E. C. Características gerais e epidemiologia dos arbovírus emergentes no Brasil. *Revista Pan-Amazônica de Saúde*, scielo, v. 5, p. 55 – 64, 09 2014. ISSN 2176-6223. Disponível em: <http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S2176-62232014000300007&nrm=iso>. Citado na página 4.
- MAGALHAES, T. et al. The endless challenges of arboviral diseases in brazil. *Tropical Medicine and Infectious Disease*, v. 5, n. 2, 2020. ISSN 2414-6366. Disponível em: <<https://www.mdpi.com/2414-6366/5/2/75>>. Citado na página 5.
- NETO, S. e. a. da S. Arboviral disease record data - dengue and chikungunya, brazil, 2013–2020. *Nature*, Nature, v. 1, n. 1, p. 2–9, 2022. Citado 2 vezes nas páginas 1 e 17.
- OHST, C. et al. Reliable serological testing for the diagnosis of emerging infectious diseases. In: _____. *Dengue and Zika: Control and Antiviral Treatment Strategies*. Singapore: Springer Singapore, 2018. p. 19–43. ISBN 978-981-10-8727-1. Disponível em: <https://doi.org/10.1007/978-981-10-8727-1_3>. Citado na página 7.
- OLIVEIRA, T. B. D. et al. Governança de dados abertos da atenção primária de sistemas de saúde públicos e universais: Um comparativo entre brasil, canadá e reino unido. In: *Congresso Internacional em Saúde*. [S.l.: s.n.], 2021. Citado na página 9.
- POSSAMAI, A. J.; SOUZA, V. G. de. Transparência e dados abertos governamentais: possibilidades e desafios a partir da lei de acesso à informação. *Administração Pública e Gestão Social*, 2020. Citado na página 9.
- PREVENTION, E. C. for D.; CONTROL. *Geographical distribution of dengue cases reported worldwide, 2020*. 2021. [Online; accessed August 27, 2024]. Disponível em: <<https://www.ecdc.europa.eu/en/publications-data/geographical-distribution-dengue-cases-reported-worldwide-2020>>. Citado na página 5.
- RAMOS, A. L. B. M. et al. A eficiência das ações de combate à dengue na atenção primária à saúde no brasil. *Brazilian Journal of Health Review*, v. 4, n. 3, p. 10575–10595, 2021. Citado na página 4.
- SAUDE, M. da. *Dengue - Ministério da Saúde*. 2024. [Online; accessed December 12, 2024]. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/dengue>>. Citado na página 6.
- SCHMITT, J. Pré-processamento para a mineração de dados: uso da análise de componentes principais com escalonamento ótimo. *AI magazine*, v. 1, n. 5, p. 12, 2005. Citado na página 9.

- SES-GO, G. de Vigilância Epidemiológica/ Superintendência de Vigilância em Saúde/ Secretaria de Estado da Saúde de G. G. S. *Boletim Epidemiológico 2024*. 2024. [Online; accessed November 25, 2024]. Disponível em: <<https://goias.gov.br/saude/wp-content/uploads/sites/34/boletins/epidemiologicos/arboviroses/2024/boletim%20arboviroses-volume3-numero9.pdf>>. Citado na página 2.
- SILVA, D. A. d. et al. Aplicação de técnicas de pré-processamento e agrupamento na base de dados de benefícios previdenciários do ministério público do trabalho. Universidade Federal de Uberlândia, 2018. Citado 2 vezes nas páginas 10 e 11.
- SINAN. *Modelo da ficha de notificação de casos de dengue de chikungunya, 2016*. 2016. [Online; accessed August 27, 2024]. Disponível em: <https://portalsinan.saude.gov.br/images/documentos/Agravos/Dengue/Ficha_DENGCHIK_FINAL.pdf>. Citado na página 8.
- SOUTH, L. et al. Effective use of likert scales in visualization evaluations: A systematic review. *Computer Graphics Forum*, v. 41, n. 3, p. 43–55, 2022. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14521>>. Citado na página 33.
- SOUZA, C. et al. Predicting dengue outbreaks in Brazil with manifold learning on climate data. *Expert Systems with Applications*, Elsevier, v. 192, 2022. Citado 2 vezes nas páginas 15 e 16.
- SOUZA-NETO, J. A.; POWELL, J. R.; BONIZZONI, M. Aedes aegypti vector competence studies: A review. *Infection, Genetics and Evolution*, v. 67, p. 191–209, 2019. ISSN 1567-1348. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1567134818307159>>. Citado na página 5.
- TAN, P.; STEINBACH, M.; KUMAR, V. Introdução ao datamining: mineração de dados. Ciência Moderna, 2009. Citado 2 vezes nas páginas 11 e 12.
- TUKEY, J. W. et al. *Exploratory data analysis*. [S.l.]: Pearson, 1977. Citado na página 22.
- WANG, W.-H. et al. Dengue hemorrhagic fever – a systemic literature review of current perspectives on pathogenesis, prevention and control. *Journal of Microbiology, Immunology and Infection*, v. 53, n. 6, p. 963–978, 2020. ISSN 1684-1182. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1684118220300670>>. Citado 3 vezes nas páginas 1, 6 e 7.