

# **Arbolytics: Monitoramento de dados de arboviroses utilizando estatísticas de notificações de casos e distribuição espacial**

**Igor Ferreira Araujo<sup>1</sup>, Douglas Cedrim<sup>1</sup>**

<sup>1</sup>Instituto Federal Goiano - Campus Rio Verde  
Rio Verde – Goiás – Brasil

igor.ferreira@estudante.ifgoiano.edu.br, douglas.cedrim@ifgoiano.edu.br

**Abstract.** *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.*

**Resumo.** *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 de forma efetiva 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.*

## **1. Introdução**

No Brasil, o aumento das infecções causadas por mosquitos tem se tornado uma preocupação crescente. Doenças como a dengue, febre-amarela, febre chikungunya e zika, que possuem parte do seu ciclo de replicativo em artrópodes, sempre foram um desafio para as autoridades de combate às endemias. No contexto brasileiro, a dengue e a chikungunya são as infecções com maior número absoluto de casos registrados, colocando o país como líder no número de casos registrados da América Latina [Silva Neto et al., 2022]. Apesar de que a erradicação dos vetores de transmissão destas doenças seja considerado praticamente impossível, sobretudo, devido ao crescimento populacional, ocupação desordenada e a falta de infraestrutura dos centros urbanos, o combate a essas doenças é extremamente necessário para impedir uma crise no sistema de saúde brasileiro. Em áreas urbanas e semi-urbanizadas, o vírus da dengue e da chikungunya são transmitidos aos humanos pela picada de uma fêmea do mosquito *Aedes aegypti*, vetor da doença.

Em Brady e Hay [2020] é apresentado que ao se estabelecer em áreas sombrias, úmidas e quentes, dentro das residências, o *Aedes aegypti* reduz a sua exposição a condições ambientais variáveis e extremas, garantindo a sobrevivência em momentos e

locais nos quais espécies menos propensas à convivência com humanos não conseguem fazê-lo. Conseqüentemente, as fêmeas de *Aedes aegypti* raramente precisam se aventurar para longe de seu ambiente de reprodução. Essas características fazem com que o mapeamento de surto de casos identifique regiões em que existe uma abundância de vetores de transmissão e criadouros.

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 torna possível efetuar análises epidemiológicas atualizadas.

Esse trabalho propõe a utilização de medidas estatísticas, juntamente com técnicas de visualização interativa, para possibilitar uma análise exploratória de dados abertos de vigilância epidemiológica de arboviroses, mais especificamente dengue e chikungunya, compilados no período de 2013-2020 [Silva Neto et al., 2022]. Uma análise importante de ser feita tem como objetivo entender se a quantidade de casos notificados em uma determinada semana epidemiológica está dentro do esperado historicamente, ou se enquadra-se como uma anomalia a ser melhor observada.

A estratégia adotada aqui é apresentar uma visão geral sobre a notificação de casos em um determinado município e ano detectando possíveis anomalias, considerando a seu histórico no período 2013-2020, e possibilitar ao usuário, uma exploração de detalhes sobre sua distribuição geográfica ao longo das semanas epidemiológicas para um determinado ano, dentro do limite de um município.

## **2. Trabalhos relacionados**

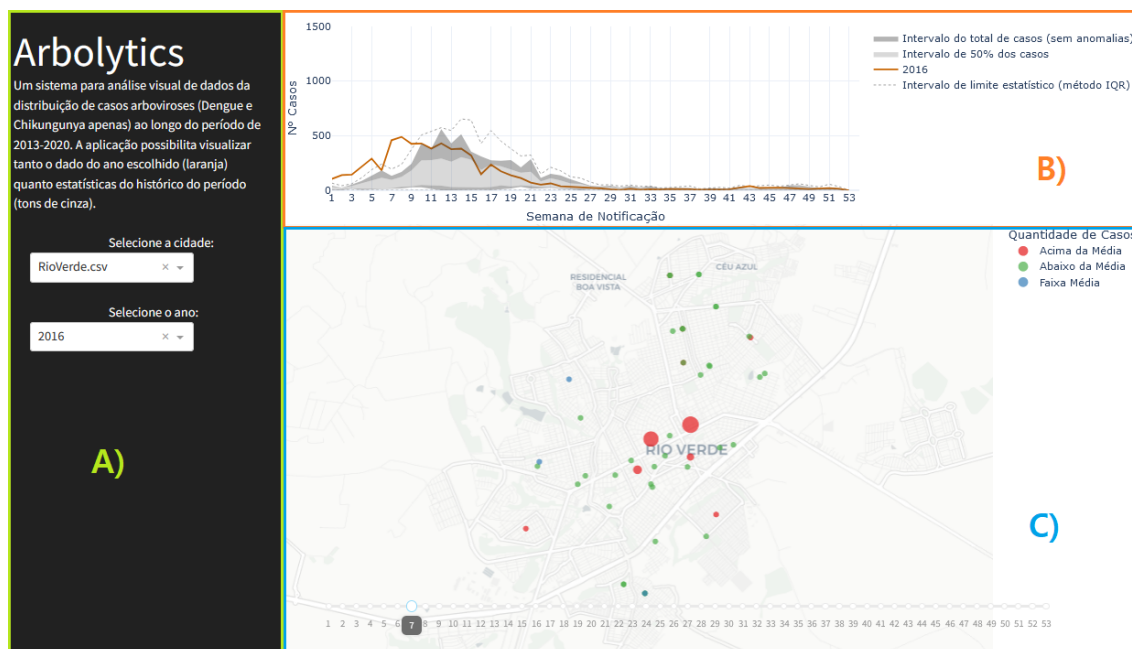
Diversos trabalhos utilizam de ferramentas computacionais no contexto de arboviroses, dentre os quais: incidência de casos através de mineração em redes sociais [Fiorini et al., 2016; Euzebio et al., 2020], simulação de distribuição de armadilhas para os mosquitos [Baldí et al., 2017], tendência de termos em máquinas de busca [Izbicki e Cruz, 2018], previsão de surtos usando séries temporais de dados climáticos [Souza et al., 2022] e até mesmo processamento de imagens [Sá et al., 2022].

A secretarias de saúde de alguns estados possuem sistemas para monitoramento restritos ao seus territórios. O trabalho proposto por Bezerra et al. [2017] possibilita que a vigilância epidemiológica analise a distribuição geográfica de casos de dengue utilizando um *dashboard* que integra a visualização da serie temporal de casos com um mapa de densidade de casos por região. Ele utiliza os dados da cidade de Aracari-CE como base para análise. O sistema permite comparação entre as semanas epidemiológicas apenas por valores brutos, deixando de lado outras informações estatísticas que podem ser relevantes.

## **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 (Fig. 1A); utiliza histórico de dados de anos

anteriores para gerar estatísticas comparativas e detectar possíveis anomalias (Fig. 1B). Adicionalmente, pode explorar os detalhes da distribuição de notificações no município filtrado por semana epidemiológica (Fig. 1C).



**Figura 1. Arbolytics ilustrando a visão combinada entre a distribuição de casos e sua distribuição espacial para o município de Rio Verde-GO, em 2016.**

### 3.1. Pré-processamento

O trabalho foi desenvolvido em Python, v3.11, juntamente com as bibliotecas Numpy v1.23.4, Pandas v1.5.1 e Plotly Dash v5.15. 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 Silva Neto et al. [2022] da plataforma do SINAN, no período de 2013 a 2020.

Após extração dos dados brutos Silva Neto et al. [2022] efetuam uma etapa de pré-processamento, onde foram removidos tanto informações irrelevantes (ex.: sintomas, dados pessoais dos pacientes) como dados incompletos, ficando ao final com 56 atributos e aproximadamente 6.7 milhões de entradas. De todos esses atributos, apenas três foram necessários para a metodologia proposta: 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).

Para um determinado município, foi extraído o nome de cada uma das suas unidades com o seu endereço através da coleta no site do CNES, utilizando a biblioteca Selenium Webdriver v4.18.1. Foram consideradas apenas as unidades identificados na base de dados. A partir disso foi feita a conversão dos endereços de cada UBS em coordenadas geográficas, usando Geopy v2.3.0. Em seguida, os casos individuais foram agrupados em um conjunto de informações que apresentam a quantidade de casos por unidade de saúde registradas em uma semana epidemiológica. Nesse ponto, a base de dados reduzida possuía os campos de: latitude, longitude e ano, concatenados à semana epidemiológica, identificador da unidade e quantidade de casos registrados na semana.

### 3.2. Medidas estatísticas

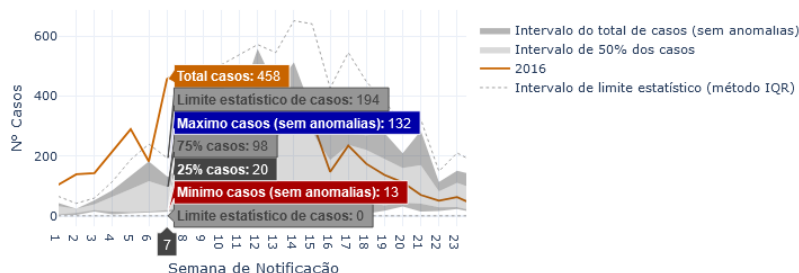
Para a criação das medidas estatísticas são considerados os dados de notificações de uma determinada semana epidemiológica utilizando todos os dados no período de 7 anos disponível. A partir disso, foi utilizado o método proposto por Tukey et al. [1977]. Todas as medidas calculadas podem ser visualizadas no exemplos ilustrados nas Figuras 2 e 4.

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. 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.

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.

### 3.3. Visualização interativa

A Figura 1 ilustra o *dashboard* web que compõe a ferramenta. O usuário deve escolher o ano que deseja avaliar através do menu lateral esquerdo (Fig. 1A). A partir disso os dois componentes interativos são atualizados com os dados correspondentes: gráfico de linhas (Fig. 1B) e mapa (Fig. 1C).



**Figura 2. Casos notificados em Rio Verde-GO, 7ª semana epidemiológica de 2016.**

O gráfico de linhas é construído como descrito na Seção 3.2. O usuário pode interagir de diversas formas com ele: ao posicionar o mouse sobre uma determinada semana serão exibidos detalhes daquela semana, como ilustrado na Figura 2; O usuário pode filtrar um conjunto de semanas selecionando diretamente no gráfico; e por fim pode dar um *zoom* em uma região específica para entender melhor um o comportamento que esteja mais difícil de visualizar.

Para evidenciar a distribuição espacial dos casos notificados, foi utilizado um *scatterplot* integrado ao componente *mapbox* do Plotly. Para cada unidade de saúde onde há notificação de casos foi criado um disco centrado em suas coordenadas geográficas com

um raio variando linearmente entre 4 e 15 *pixels*, codificando a quantidade de casos na semana epidemiológica anterior à analisada. O canal de cor do disco codifica a quantidade de casos da semana epidemiológica, ele pode assumir três valores distintos indicados na Figura 3. 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. Ao clicar no disco, o sistema permite visualizar as informações de cada unidade de forma individual e aprofundada, como apresentado na Figura 3.



Figura 3. Notificações em Rio Verde-GO, na 7ª semana epidemiológica de 2016.

#### 4. Discussão dos resultados obtidos

Analisando a cidade de Rio Verde-GO, ilustrada na Figura 1, pode-se observar que: No período compreendido entre a 7ª e 9ª semana vê-se nitidamente um comportamento anômalo, com mais casos que o esperado, que não deveria passar de 194, mas foram notificados 458 casos, ilustrado na Figura 2. Ainda sobre esse experimento, analisando melhor essa semana anômala, observa-se na Figura 3 que a UBS selecionada possui tanto casos acima da média na semana analisada (7ª), codificada com a cor, como também possui casos acima da média na semana anterior (6ª), codificado no raio. Nota-se também que a maioria das outras unidades mostravam número de casos abaixo da média, indicando a concentração de casos em algumas UBS apenas.

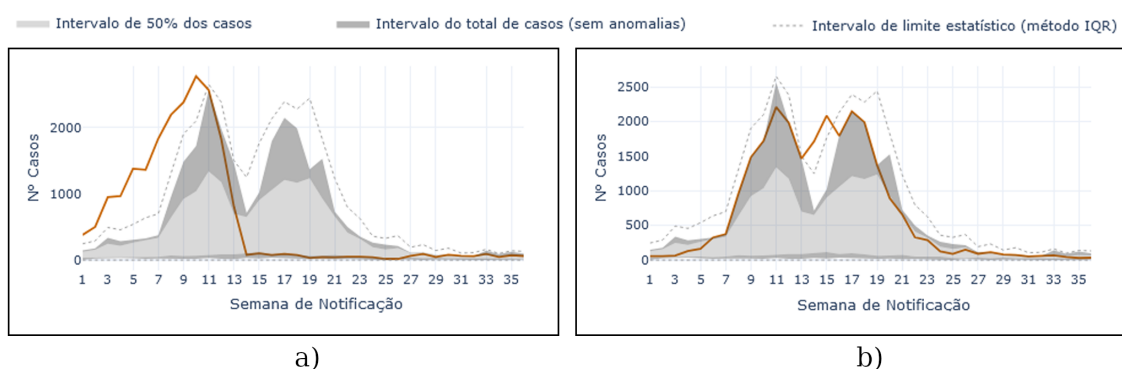


Figura 4. Casos notificados por semana epidemiológica em São Paulo-SP. Em laranja a série temporal de 2014 (a) e 2016 (b).

Analisando a cidade de São Paulo-SP, ilustrado na Figura 4, 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.

## 5. Conclusão e trabalhos futuros

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.

Como trabalhos futuros, pode-se pensar na integração de dados em tempo real poderia oferecer um monitoramento mais dinâmico e a detecção precoce de surtos, adicionando uma camada valiosa de informação. Adicionalmente, outras variáveis epidemiológicas, sociais e ambientais podem ser incorporadas, com apoio de especialistas. Por fim, outras estratégias de estimação de *outliers* podem ser exploradas.

## Referências

- Baldi, A., Zambon, E., Costa, P., e Montiel, E. (2017). Simulação de aplicação de armadilhas no combate ao *Aedes aegypti*. In *Simpósio brasileiro de computação aplicada à saúde (SBCAS)*, Porto Alegre, RS, Brasil. SBC.
- Bezerra, A., Bento Filho, J. J., Braga, R., Oliveira, C., e Oliveira, M. (2017). Dengosa: Um sistema de informação geográfica para apoio à decisão no controle de epidemias. In *XXIII Simpósio Brasileiro de Sistemas Multimídia e Web*, pages 179–183. SBC.
- Brady, O. J. e Hay, S. I. (2020). The global expansion of dengue: How *Aedes aegypti* mosquitoes enabled the first pandemic. *Annual review of entomology*.
- Euzebio, C., Agy, S., Jr., C. B., Porto, L., Alcarás, J. R., Martinez, A., e Ruiz, E. (2020). Statistical analysis of small twitter data collection to identify dengue outbreaks. In *VIII Symposium on Knowledge Discovery, Mining and Learning*, pages 17–24, Porto Alegre, RS, Brasil. SBC.
- Fiorini, S., Sousa, L., Cedrim, D., Garcia, A., Saade, D., Moraes, I., e Frajhof, L. (2016). 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)*, pages 2567–2576, Porto Alegre, RS, Brasil. SBC.
- Izbicki, R. e Cruz, L. O. d. (2018). Monitoramento online da dengue: Usando o google para prever epidemias. *Brazilian Journal of Biometrics*, 36(3):512–526.
- Silva Neto, S. R., Tabosa de Oliveira, T., Teixeira, I. V., Medeiros Neto, L., Souza Sampaio, V., Lynn, T., e Endo, P. T. (2022). Arboviral disease record data-dengue and chikungunya, brazil, 2013–2020. *Nature: Scientific Data*, 9(1):198.
- Souza, C., Maia, P., Stolerman, L. M., Rolla, V., e Velho, L. (2022). Predicting dengue outbreaks in Brazil with manifold learning on climate data. *Expert Systems with Applications*, 192.
- Sá, L., Rizzi, C., Rizzi, R., e Brun, A. (2022). Aplicação de aprendizagem de máquina na estimação do número de ovos de aedes em imagens de ovitrampas. In *Congresso Lat-Am de Soft. Livre e Tecnologias Abertas*, pages 36–42. SBC.
- Tukey, J. W. et al. (1977). *Exploratory data analysis*. Pearson.