

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

**Instituto de Geociências**

**Programa de Pós-Graduação em Análise e Modelagem de Sistemas Ambientais**

Murilo Ribeiro Spala

**DOENÇAS NEGLIGENCIADAS NO BRASIL: mapeamento de uma década de  
ocorrências**

Belo Horizonte

2023

Murilo Ribeiro Spala

**DOENÇAS NEGLIGENCIADAS NO BRASIL: mapeamento de uma década de  
ocorrências**

Dissertação apresentada ao Programa de Pós Graduação em Análise e Modelagem de Sistemas Ambientais da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Análise e Modelagem de Sistemas Ambientais.

Orientadora: Sónia Maria Carvalho Ribeiro

Belo Horizonte  
2023

S734d  
2023

Spala, Murilo Ribeiro.

Doenças negligenciadas no Brasil [manuscrito] : mapeamento de uma década de ocorrências / Murilo Ribeiro Spala. – 2023.

84 f., enc.: il. (principalmente color.)

Orientadora: Sônia Maria Carvalho Ribeiro.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Geociências, 2023.

Área de concentração: Análise, Modelagem e Gestão de Sistemas Ambientais.

Bibliografia: f. 43-46.

Inclui anexos.

1. Modelagem de dados – Aspectos ambientais – Teses. 2. Doenças – Brasil – Teses. 3. Saúde pública – Brasil) – Teses. I. Carvalho- Ribeiro, Sônia Maria. II. Universidade Federal de Minas Gerais. Instituto de Geociências. III. Título.

CDU: 911.2:519.6(81)



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Instituto de Geociências  
Programa de Pós-Graduação em Análise e Modelagem de Sistemas Ambientais

## FOLHA DE APROVAÇÃO

### **Doenças negligenciadas no Brasil: Mapeamento de uma década de ocorrências**

**Murilo Ribeiro Spala**

Dissertação de Mestrado defendida e aprovada, no dia 31 de maio de 2023, pela Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Análise e Modelagem de Sistemas Ambientais da Universidade Federal de Minas Gerais constituída pelos seguintes professores:

Sônia Maria Carvalho Ribeiro - UFMG - (Doutor)  
Maurício Roberto Viana Sant Anna - UFMG - (Doutor)  
Camila Palhares Teixeira - UEMG - (Doutora)

Belo Horizonte, 31 de maio de 2023.



Documento assinado eletronicamente por **Sônia Maria Carvalho Ribeiro, Professora do Magistério Superior**, em 13/06/2023, às 11:08, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Camila Palhares Teixeira, Usuário Externo**, em 22/06/2023, às 10:24, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maurício Roberto Viana Sant Anna, Professor do Magistério Superior**, em 23/06/2023, às 15:03, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador 2341861 e o código CRC2983DA77.

Dedico esse trabalho ao Amigo Thiago  
Felisbino Fernandes, que ainda permanece vivo  
em minhas lembranças.

## **AGRADECIMENTOS**

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento que possibilitou o desenvolvimento da pesquisa.

Ao Programa de Pós-Graduação em Análise e Modelagem de Sistemas Ambientais, especialmente na figura dos professores, por me apresentarem ferramentas e caminhos para a conclusão da pesquisa.

No plano mais pessoal, devo destacar a colaboração recebida da minha família e da professora Sónia Maria Carvalho Ribeiro.

## RESUMO

As doenças negligenciadas são enfermidades que afetam as populações com acesso limitado aos serviços de saúde. Este estudo foca em oito doenças negligenciadas no Brasil, como a Doença de Chagas, Hanseníase, Leishmaniose Tegumentar, Leishmaniose Visceral, Malária, Tuberculose, Dengue e Esquistossomose. Essas doenças representam problemas de saúde pública e são endêmicas em diferentes regiões do país. O trabalho é inovador ao analisar as doenças negligenciadas no Brasil, considerando sua dimensão espacial e identificando os determinantes espaciais associados à sua propagação e persistência ao longo das últimas décadas. O estudo mapeia as áreas mais afetadas por essas doenças, identifica variáveis preditoras significativas e fornece subsídios para orientar a implementação de medidas preventivas e de controle mais eficazes em saúde pública. Além disso, busca contribuir para o desenvolvimento de metodologias inovadoras no mapeamento dessas doenças e para o cumprimento das metas globais estabelecidas pela Organização Mundial da Saúde. A proposta metodológica do trabalho foca no modelo linear generalizado (GLM) em duas escalas (municipal e pixel a pixel), foram analisadas variáveis socioeconômicas e ambientais como fatores explicativos e preditores da ocorrência dessas doenças. Foram registrados 14.514.304 casos acumulados no período de 2010 a 2020. As variáveis preditoras que mais influenciam o número de casos são: Altitude, densidade demográfica, alfabetização, atenção básica, lixo coletado, ambiente rural, saneamento, energia elétrica, precipitação e temperatura. Foram identificadas áreas mais vulneráveis (GLM) e agrupamentos de municípios com números semelhantes de casos. Como implicações práticas, tais resultados permitem identificar as regiões mais afetadas e vulneráveis pelas doenças negligenciadas, dando subsídios para implementação de medidas preventivas e de controle mais eficazes da saúde pública no Brasil.

Palavras-chave: doenças negligenciadas; Brasil; espacialização; determinantes de saúde; modelo linear generalizado.

## ABSTRACT

Neglected diseases are illnesses that affect populations with limited access to health services. This study focuses on eight neglected diseases in Brazil, such as Chagas Disease, Leprosy, Tegumentary Leishmaniasis, Visceral Leishmaniasis, Malaria, Tuberculosis, Dengue, and Schistosomiasis. These diseases represent public health problems and are endemic in different regions of the country. The work is innovative in analyzing neglected diseases in Brazil, considering their spatial dimension, identifying the spatial determinants associated with their spread and persistence over the last few decades. The study maps the areas most affected by these diseases, identifies significant predictive variables, and provides subsidies to guide the implementation of more effective preventive and control measures in public health. In addition, it seeks to contribute to the development of innovative methodologies for mapping these diseases and meeting the global goals established by the World Health Organization. The methodological proposal of the work focuses on the generalized linear model (GLM) at two scales (municipal and pixel by pixel). Socioeconomic and environmental variables were analyzed as explanatory factors and predictors of the occurrence of these diseases. A total of 14,514,304 cases were registered in the period from 2010 to 2020. The predictive variables that most influence the number of cases are: Altitude, population density, literacy, primary care, garbage collection, rural environment, sanitation, electricity, precipitation, and temperature. The most vulnerable areas (GLM) and clusters of municipalities with similar numbers of cases were identified. As practical implications, these results make it possible to identify the regions most affected and vulnerable by neglected diseases, providing subsidies for the implementation of more effective preventive measures and control of public health in Brazil.

Keywords: neglected diseases; Brazil; spatialization; health determinants; generalized linear model



## LISTA DE ILUSTRAÇÕES

Figura 1: Determinantes de Saúde, dispostos dos mais específicos aos mais gerais.....	13
Figura 2: Infográfico da área de estudo, mostrando áreas com maior número de casos de doenças negligenciadas no território nacional brasileiro. A maior intensidade de cores escuras, representa um maior número de casos de doenças negligenciadas.....	21
Figura 3: Etapas de coleta e processamento de dados .....	25
Figura 4: Infográfico e mapa de calor (estimativa de densidade Kernel) das doenças negligenciadas no Brasil – Acumulado de 2010 a 2020, mostrando áreas com maior número de casos para cada doença negligenciada estudada no território nacional brasileiro. A maior intensidade cores do azul claro ao vermelho, representa um maior número de casos de doenças negligenciadas.....	26
Figura 5: Mapa de agrupamentos (k-means) das doenças negligenciadas no Brasil – Acumulado de 2010 a 2020.. .....	30
Figura 6: Mapa de GLM pixel a pixel e por municípios normalizados de 0 a 100 para as doenças negligenciadas no Brasil, mostrando áreas com maior predição (maior vulnerabilidade) com cores escuras, e com menor predição (menor vulnerabilidade) com cores claras. ....	32
Figura 7: Mapa de GLM pixel a pixel e por municípios normalizados para a Dengue em Belo Horizonte, mostrando áreas com maior predição (maior vulnerabilidade) com cores escuras, e com menor predição (menor vulnerabilidade) com cores claras.....	33

## LISTA DE TABELAS

Tabela 1 – Fatores associados às doenças negligenciadas .....	22
Tabela 2 – Sistemas de Informação da Saúde – Origem dos dados consultados para cada uma das doenças negligenciadas incluídas no estudo .....	23
Tabela 3 – Incidência acumulada (2010 a 2020) por 100.000 habitantes (2010) por Unidade da Federação – Ordenado de forma decrescente pela ‘Incidência Total Negligenciadas’.....	29
Tabela 4 – Resultados da análise da base de dados que justificam a utilização do GLM ‘hurdle’ – casos de 2010 a 2020 Brasil .....	34
Tabela 5 – Efeito de variáveis independentes (efeito negativo ou positivo) sobre os municípios com casos de Doenças Negligenciadas .....	35

## LISTA DE ABREVIATURAS E SIGLAS

AM: Amazonas

BA: Bahia

CE: Ceará

CO: Centro-Oeste

DS: Determinantes de Saúde

DE: Determinantes Espaciais

DES: Determinantes Estruturais de Saúde

DSS: Determinantes Sociais da Saúde

GLM: *Generalized Linear Model* (Modelo Linear Generalizado)

IBGE: Instituto Brasileiro de Geografia e Estatística

MA: Maranhão

MAPBIOMAS: Projeto de Mapeamento da Cobertura e Uso do Solo do Brasil

MG: Minas Gerais

N: Norte

ODS: Objetivos de Desenvolvimento Sustentável

OMS: Organização Mundial da Saúde

ONU: Organização das Nações Unidas

OPAS: Organização Pan-Americana da Saúde

PA: Pará

PCE: Programa de Controle da Esquistossomose

PE: Pernambuco

RJ: Rio de Janeiro

RS: Rio Grande do Sul

SCNES: Sistema de Cadastro Nacional de Estabelecimentos de Saúde

SE: Sudeste

SINAN: Sistema de Informação de Agravos de Notificação

SIPNI: Sistema de Informação do Programa Nacional de Imunizações

SIS: Sistemas de Informação da Saúde

SIVEP: Sistema de Informação de Vigilância Epidemiológica

SP: São Paulo

SRTM: *Shuttle Radar Topography Mission* (Missão de Topografia por Radar do Ônibus Espacial)

UFMG: Universidade Federal de Minas Gerais

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	12
<b>1.1 Objetivos gerais e específicos da dissertação</b> .....	16
<b>2 ARTIGO</b> .....	17
<b>2.1 Introdução</b> .....	19
<b>2.2 Métodos</b> .....	21
2.2.1 Área de estudo .....	21
2.2.2 Revisão da literatura .....	22
2.2.3 Coleta de dados .....	23
2.2.4 Processamento de dados .....	24
<b>2.3 Resultados</b> .....	26
2.3.1 Densidade espacial dos casos de doenças negligenciadas .....	26
2.3.2 Agrupamento municipal da morbidade .....	30
2.3.3 Modelo linear generalizado .....	32
<b>2.4 Discussão</b> .....	39
<b>2.5 Conclusão</b> .....	41
<b>REFERÊNCIAS</b> .....	43
<b>ANEXOS</b> .....	47
<b>CONSIDERAÇÕES FINAIS</b> .....	83

## 1 INTRODUÇÃO

As doenças negligenciadas são aquelas que afetam principalmente as populações mais pobres e marginalizadas, com acesso limitado aos serviços de saúde, e que não recebem a devida atenção dos governos e da indústria farmacêutica. Entre as doenças negligenciadas incluídas no presente estudo estão: (1) Doença de Chagas, (2) Hanseníase, (3) Leishmaniose Tegumentar, (4) Leishmaniose Visceral, (5) Malária, (6) Tuberculose, (7) Dengue e (8) Esquistossomose. Todas essas doenças podem ser consideradas endêmicas em diferentes regiões do Brasil além causar graves problemas de saúde pública. E foram elencadas pois constam na lista de doenças negligenciadas em documento elaborado pela Organização Mundial da Saúde, e são notificadas nacionalmente e sistematicamente pelos Sistemas de Informação da Saúde (WORLD HEALTH ORGANIZATION, 2020).

Para essa pesquisa, foi utilizado o modelo linear generalizado (GLM) em duas escalas distintas, incluindo a escala municipal e pixel a pixel, usando uma resolução espacial de 1.000 metros. Além disso, foram utilizadas variáveis socioeconômicas e ambientais consideradas como explicativas e preditoras para ocorrência de número casos para essas doenças, a fim de identificar os fatores que influenciam a propagação e manutenção dessas.

Os resultados do estudo mostraram a média estimada pelo GLM para os municípios, bem como as variáveis preditoras mais significativas para cada uma das doenças negligenciadas. Esses determinantes espaciais indicam se o efeito é negativo ou positivo na quantidade de casos, o que é fundamental para a implementação de medidas preventivas e curativas adequadas. As análises envolveram dados à escala municipal com apenas um valor para cada um dos 5.570 municípios. E também foi estimada de forma espacialmente explícita, ou seja, para cada um dos 20.522.368 pixels de 1.000 por 1.000 metros que compõe o Brasil, o que resultou em uma melhor compreensão dos fatores que influenciam a ocorrência das doenças negligenciadas.

Além disso, o estudo identificou áreas mais vulneráveis e agupamentos de municípios similares em número de casos, por meio da Densidade Kernel e do K-means. Esses resultados foram fundamentais para a validação do modelo linear generalizado - GLM - de cada doença negligenciada, feita por meio da correlação entre os rasters de Densidade Kernel e do resultado do GLM com a mesma resolução espacial.

Em suma, o presente estudo contribui para o desenvolvimento de metodologias inovadoras para o mapeamento das doenças negligenciadas no Brasil.

Dada a importância do contexto espacial para estudar as doenças negligenciadas, em modelagem espacial o conceito de Determinantes Espaciais (DE) referem-se a fatores ou variáveis que influenciam a organização e distribuição de elementos ou eventos em um espaço geográfico. Eles desempenham um papel crucial na compreensão das interações entre a sociedade e o ambiente físico. Muito utilizado na epidemiologia, existe o modelo de Determinantes de Saúde (DS), que muito conversa com o conceito anterior. Alguns modelos conceituais foram elaborados para esquematizar os DS, o modelo de *Dahlgren e Whitehead* inclui os DS dispostos em diferentes camadas, desde uma camada mais próxima, os determinantes individuais até uma camada distal, onde se situam os macro determinantes (Figura 1).



Figura 1: Determinantes de Saúde, dispostos dos mais específicos aos mais gerais. Fonte: Dahlgren e Whitehead, 1991.

Os Determinantes de Saúde, tendo como termos sinônimos, Determinantes Estruturais de Saúde (DES) ou Determinantes Sociais da Saúde (DSS), configuram-se como circunstâncias em que pessoas nascem, crescem, vivem, trabalham e envelhecem, bem como os sistemas disponíveis para lidar com doenças. Tais circunstâncias são, por sua vez, moldadas por um conjunto mais amplo de forças: econômicas, sociais e políticas. Forças essas que compõem o espaço.

Embora o termo “Determinantes de Saúde” tenha sido cunhado e sistematizado apenas no final do século XX (BUSS; PELLEGRINI FILHO, 2007) , estudos sobre a dimensão espacial associada às doenças compõem a história da medicina ocidental e da saúde pública - como na obra de Hipócrates, ‘Sobre o ar, a água e a terra’ - por mais de dois mil anos (BUCK, 1989); a história da dimensão espacial sistematizada nos estudos sobre a saúde pública tem como marco temporal inicial a segunda metade do século XVIII - com a obra de Finke, onde busca a gênese

das doenças em uma única causa - classificada como 'Da polícia médica à geografia médica' (BOUSQUAT; COHN, 2004).

Posteriormente a esse marco temporal inicial, já no século XIX, 'A era dos atlas e das doenças infecciosas' e o espaço dissociado à dimensão humana. Ainda no mesmo século o 'Determinismo, ecologia humana e gênero de vida' onde pressupõe que o espaço físico dita a dinâmica das sociedades. 'Os complexos patogênicos' na primeira metade do século XX proposto por Sorré, mostra que os padrões observados são mutáveis no tempo, consonantes com as mudanças das sociedades. A 'Teoria da nidalidade' no mesmo período, onde as doenças apresentariam focos naturais, e ações pragmáticas deveriam ser adotadas. Seguindo a cronologia 'A comissão de Geografia Médica de Saúde e Doença da União Geográfica Internacional', um movimento que visava estudar a relação saúde-espaço; e por fim 'A nova geografia médica' marcada pela base quantitativa e objetiva, fomentada na segunda metade do século XX, consolidando-se como corrente influente nos estudos sobre análise espacial da saúde até os dias atuais (BOUSQUAT; COHN, 2004).

As doenças negligenciadas - para a pesquisa, Doença de Chagas, Hanseníase, Leishmaniose Tegumentar, Leishmaniose Visceral, Malária, Tuberculose, Dengue e Esquistossomose -, também conhecidas como doenças esquecidas ou doenças tropicais negligenciadas são caracterizadas por serem subfinanciadas e possuem baixo reconhecimento, mas são as principais ameaças em países pouco desenvolvidos.

As crises humanitárias, as mudanças climáticas podem impactar todos os espectros da sociedade, incluindo a saúde; e praticamente todas as doenças negligenciadas sofrem influência de fatores ambientais, tais como ar, água e solo, ou seja, os padrões geográficos dessas doenças poderão ser afetados pelas mudanças climáticas de curto, médio e longo prazo. Com destaque para redistribuição de vetores e efeitos das mudanças climáticas sobre os sistemas de saneamento sanitário (AL-DELAIFY, 2022; BOOTH, 2018; GAINOR; HARRIS; LABEAUD, 2022; HALL; WALL; STEVENS, 2016; NAU et al., 2019)

A Organização Mundial da Saúde (OMS) elaborou em 2020 um documento com metas para combater 20 doenças negligenciadas que afetam mais de um bilhão de pessoas nas regiões mais vulneráveis do planeta até 2030. As metas globais descritas pelo roteiro são: a) reduzir em 90% o número de pessoas que precisam de tratamento para doenças tropicais negligenciadas, b) pelo menos 100 países eliminando ao menos uma dessas 20 doenças, c) erradicar duas doenças



tropicais negligenciadas (dracunculíase e boubá) e d) reduzir em 75% os anos de vida perdidos por incapacidade relacionados a essas enfermidades (OMS, 2021; OPAS, 2021).

Além das metas citadas, o documento define outras ações estratégicas, tais como, diminuir em “75% no número de mortes por doenças negligenciadas transmitidas por vetores, como dengue, leishmaniose e outros; promover o acesso total ao abastecimento básico de água, saneamento e higiene em áreas endêmicas [...]” (OPAS, 2021).

O documento reforça a importância de uma vigilância em saúde robusta pós conclusão das metas, para isso é necessário o fortalecimento dos sistemas nacionais de saúde (HATHERELL et al., 2021).

O estudo dos determinantes espaciais para as doenças negligenciadas é de fundamental importância, pois sistematiza e evidencia, com maior ou menor nível de detalhe, as forças associadas na dinâmica espacial desse grupo de doenças, seu reflexo nas populações e se configura como importante ferramenta de gestão das paisagens que sofrem com a Doença de Chagas, Hanseníase, Leishmaniose Tegumentar, Leishmaniose Visceral, Malária, Tuberculose, Dengue e Esquistossomose.

Para alcançar os resultados da revisão bibliográfica optou-se pela base que reúne diversos tipos de conteúdo científico, o ‘Scopus’, acessado por meio do *login* UFMG e via periódicos capes. Na aba ‘*Advanced document search*’ utilizou-se a *string*:

*TITLE-ABS-KEY("neglected diseases" OR "doenças negligenciadas" OR "enfermedades desatendidas" OR "maladies négligées" OR "doença negligenciada" OR "doenças esquecidas" OR "doenças tropicais negligenciadas") AND ALL("spatial analysis" OR "spatial modeling" OR "análise espacial" OR "modelagem espacial").*

A revisão de estudos sobre doenças negligenciadas que de alguma forma utilizaram análise e/ou modelagem espacial, seja com o cálculo de incidência municipal, por exemplo (vetorial), localização pontual (vetorial) ou modelos espaciais matriciais, apontou que de 2007 a 2022, 27% dos estudos tiveram como objeto de pesquisa um grupo de doenças negligenciadas, 24% apresentaram como objeto de estudo apenas a Leishmaniose, 11% a Hanseníase, 8% a Esquistossomose, 4% Doença de Chagas, e outras doenças como Dengue, cisticercose e verminoses transmitida pelo solo, variaram de 3 a 1% de representatividade nos estudos.

A depender da natureza da pesquisa, uma área de estudo não se faz necessária; dentre os estudos que tiveram uma localidade, seja ela município, país ou região; 58% tiveram como continente a América do Sul, 24% África, 13% Ásia, 4% Europa e 1% América do Norte.

A partir da revisão dos estudos, evidencia-se as características interseccional e de retroalimentação dos fatores associados às doenças negligenciadas, com a renda explicando às condições de moradia e vice-versa, por exemplo. Os estudos relatam em maiores ou menores detalhes os fatores associados à problemática, desde os resultados do NDVI (*Normalized Difference Vegetation Index*) - uma medida que usa informações espectrais para avaliar a densidade e a saúde da vegetação em uma determinada área, permitindo monitorar e compreender melhor os ecossistemas terrestres - como fator de risco para esquistossomose, passando pelo aumento da poluição plástica e seu impacto na dengue, até chegar em fatores associados mais genéricos, como ‘populações vulneráveis’ (CLARK et al., 2019; GAINOR; HARRIS; LABEAUD, 2022).

### **1.1 Objetivos gerais e específicos da dissertação**

A pesquisa aborda a importância da dimensão espacial para análise das doenças negligenciadas integradas com os determinantes espaciais (variáveis preditoras) no Brasil, contribuindo como diagnóstico do que aconteceu, e o que acontece.

Os objetivos da pesquisa baseiam-se em mapear passado e presente e caracterizar a distribuição espacial e áreas vulneráveis para doenças negligenciadas no país. Para alcançar os objetivos as perguntas norteadoras são:

Quais regiões apresentaram e apresentam elevados casos de doenças negligenciadas?

Quais fatores espacialmente explícitos podem ser associados às doenças negligenciadas e como eles influenciam em sua dinâmica espacial?

O que pode mitigar e potencializar a problemática das doenças negligenciadas no Brasil?

Este trabalho foi desenvolvido no formato de artigo científico visando publicação em periódico, cumprindo os requisitos para a obtenção do título de Mestre em Análise e Modelagem de Sistemas Ambientais.

Desta forma após a revisão bibliográfica e definição das questões, será apresentado o artigo científico que inclui: Introdução, Materiais e Métodos, Resultados e Discussão e Conclusão.

## 2 ARTIGO

**Resumo:** As doenças negligenciadas são um grande problema de saúde pública em todo o mundo, especialmente em países com dimensões continentais como o Brasil. Para atingir as metas globais no combate a essas doenças, é fundamental conhecer o seu contexto espacial. O conhecimento e a espacialização dos casos das doenças permitem identificar as áreas mais afetadas e implementar medidas preventivas e de controle eficazes. De forma a enfrentar tal problema, uma modelagem espacial nacional para oito doenças negligenciadas, permite identificar os territórios que mais sofrem com esse flagelo, apontando as regiões mais vulneráveis do país. O trabalho pôde também identificar variáveis que foram comuns a diferentes doenças para que intervenções e políticas públicas possam ser direcionadas para mitiga-las de forma efetiva. Como resultado, este estudo visa oferecer um conjunto de modelos espacialmente explícitos para permitir que isso aconteça. A proposta metodológica do trabalho foca no mapeamento do número de casos acumulados para oito doenças negligenciadas: Doença de Chagas, Hanseníase, Leishmaniose Tegumentar, Leishmaniose Visceral, Malária, Tuberculose, Dengue e Esquistossomose durante o período 2010 a 2020 nos 5.570 municípios brasileiros. Por meio de modelos lineares generalizados (GLM) (obstáculos “hurdle” e poisson) à escala do município e também pixel a pixel (1.000 metros de resolução espacial), utilizaram-se 23 variáveis socioeconômicas e ambientais como explicativas e preditoras do número de casos das doenças. Os resultados da indicam 14.514.304 casos de doenças negligenciadas entre 2010 e 2020, onde cada doença negligenciada estudada apresenta uma espacialização nacional específica (densidade kernel), assim como os municípios similares em número de casos (agupamentos). Os modelos GLM indicaram as áreas mais vulneráveis para cada doença negligenciada no Brasil, assim como as variáveis preditoras que mais influenciam o número de casos dessas: Altitude, densidade demográfica, alfabetização, atenção básica, lixo coletado, ambiente rural, saneamento, energia elétrica, precipitação e temperatura.

**Palavras-chave:** doenças negligenciadas; Brasil; espacialização; determinantes de saúde; modelo linear generalizado.

**Abstract:** Neglected diseases pose a significant global public health challenge, particularly in countries with large territories like Brazil. Understanding the spatial context of these diseases is crucial for achieving global goals in combating them. Knowledge of disease cases and their spatial distribution enables the identification of the most affected areas and facilitates the implementation of effective preventive and control measures. This study aims to address this issue by developing a national spatial modeling approach for eight neglected diseases, which allows for the identification of the most affected regions and the formulation of targeted interventions and public policies. The research also aims to identify common variables associated with multiple diseases to support effective mitigation strategies. Consequently, the study presents a collection of spatially explicit models to facilitate these objectives. The proposed methodology focuses on mapping the cumulative number of cases for eight neglected diseases, namely Chagas Disease, Leprosy, Tegumentary Leishmaniasis, Visceral Leishmaniasis, Malaria, Tuberculosis, Dengue, and Schistosomiasis, between 2010 and 2020 across 5,570 municipalities in Brazil. Generalized linear models (GLM), specifically hurdle and Poisson models, were employed at both the municipal scale and pixel-by-pixel resolution (with a spatial resolution of 1,000 meters). Twenty-three socioeconomic and environmental variables were utilized as explanatory and predictive factors for disease case numbers. The results revealed a total of 14,514,304 neglected disease cases recorded between 2010 and 2020. Each of the studied neglected diseases exhibited a specific national spatial distribution (kernel density), while certain municipalities demonstrated similar case numbers (clusters). The GLM models identified the most vulnerable areas for each neglected disease in Brazil and highlighted predictor variables that exerted the greatest influence on the number of neglected disease cases. These influential variables included altitude, population density, literacy rates, primary healthcare availability, waste management, rural environment, sanitation infrastructure, access to electricity, precipitation, and temperature.

**Keywords:** neglected diseases; Brazil; spatialization; health determinants; generalized linear model.

## 2.1 Introdução

Todos os anos, milhões de pessoas nas regiões mais vulneráveis do planeta são afetadas por doenças negligenciadas. Combater a ocorrência de doenças negligenciadas é imprescindível para atingir Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU) (ODS 3º: Saúde e bem-estar). As metas Globais da Organização Mundial de Saúde (OMS) para combater 20 doenças negligenciadas<sup>1</sup> - que devem afetar mais de um bilhão de pessoas em todo planeta até 2030 - são: a) reduzir em 90% o número de pessoas que precisam de tratamento para doenças tropicais negligenciadas, b) eliminar, em pelo menos 100 países, ao menos uma das doenças, c) erradicar dracunculíase e boubas e d) reduzir em 75% os anos de vida perdidos por incapacidade relacionados a essas enfermidades (WORLD HEALTH ORGANIZATION, 2020).

Os desafios são ainda maiores nos países em desenvolvimento, onde há uma distribuição de renda desigual, carência sociais e condições socio ambientais favoráveis para o fomento de doenças negligenciadas (WORLD HEALTH ORGANIZATION, 2020). Para planejar estratégias em saúde pública e formulação de políticas para minimizar o impacto das doenças negligenciadas é necessário conhecer sua distribuição espacial e analisar suas associações com variáveis socioeconômicas e ambientais.

As doenças negligenciadas podem ser conceituadas como um grupo de doenças infecciosas que afetam predominantemente as populações mais pobres e marginalizadas em regiões tropicais e subtropicais do mundo (MITRA; MAWSON, 2017).

Neste trabalho usamos como caso de estudo o território Brasileiro, para o qual foram mapeadas oito (8) doenças negligenciadas, a saber: Doença de Chagas, Hanseníase, Leishmaniose Tegumentar, Leishmaniose Visceral, Malária, Tuberculose, Dengue e Esquistossomose; também conhecidas como doenças esquecidas ou doenças tropicais negligenciadas (ver Material Suplementar 2). Dessas apenas Malária e Tuberculose não estão presentes na lista global da OMS, no entanto, assim como as demais elencadas, configuram como doenças recorrentes em populações vulneráveis e sistematicamente registradas no Brasil.

Até então, apenas uma minoria dos estudos no âmbito das doenças negligenciadas foi realizada em continentes desenvolvidos, como Europa e América do Norte. Porém, recentemente conceitos como *'Blue Marble Health'* relatam o paradoxo do aumento das doenças

---

<sup>1</sup> Úlcera de Buruli; Doença de Chagas; Dengue e Chikungunya; Dracunculíase; Equinococose; Trematodíases transmitidas por alimentos; Tripanossomíase africana humana; Leishmaniose; Hanseníase; Filariose linfática; Micetoma, cromoblastomicose e outras micoses profundas; Oncocercose; Raiva; Escabiose e outras ectoparasitoses; Esquistossomose; Helmintíases transmitidas pelo solo; Envenenamento por picada de cobra; Teníase e cisticercose; Tracoma; Boubas.

negligenciadas em países ricos, como observado no Texas, nos Estados Unidos (HOTEZ, 2016, 2018). A atenção e vigilância em saúde estão cada vez mais nas agendas internacionais, sobretudo em contextos de crises humanitárias (epidemias, pandemias e guerras), migrações, taxas de vacinação em declínio e um novo paradigma de pobreza.

A literatura internacional vem mostrando que variáveis como baixa renda (HOTEZ; KAMATH, 2009), tempo/clima (EBI; NEALON, 2016), doenças concomitantes (MURRAY; LOPEZ, 2013), baixa escolaridade (STEINMANN et al., 2006) entre outras variáveis estão associadas às doenças negligenciadas. Outros autores destacam o papel do aumento da carga viral em mamíferos silvestres, redistribuição de vetores e efeitos das mudanças climáticas nos sistemas de saneamento sanitário (AL-DELAIFY, 2022; BOECKMANN; JOYNER, 2014; BOOTH, 2018; CARLSON et al., 2022; GAINOR; HARRIS; LABEAUD, 2022; HALL; WALL; STEVENS, 2016; NAU et al., 2019).

Alguns autores detalham as estratégias a serem seguidas, como vigilância em saúde (MALECELA; DUCKER, 2021), melhoria no saneamento (WORLD HEALTH ORGANIZATION, 2004), educação (KANG et al., 2018), redes de atenção à saúde, especialmente a atenção primária, distribuição e transferência de renda e controle biológico (ALMEIDA et al., 2020; DOS SANTOS et al., 2012; MONTEIRO et al., 2017; SALEEM et al., 2020).

Os objetivos deste trabalho são mapear e quantificar a distribuição espacial dos casos de doenças negligenciadas registradas à escala nacional. No mapeamento à nível nacional, além da análise de densidade dos casos de doenças no Brasil, foi utilizada uma análise de agrupamentos para identificar similaridades entre municípios, para as diferentes doenças negligenciadas. O trabalho permite quantificar o efeito negativo ou positivo de variáveis socioeconômicas e ambientais estatisticamente significativas, na quantidade de casos para as doenças negligenciadas estudadas. Ou seja, entender a relação entre uma variável resposta e uma ou mais variáveis explicativas (ou preditoras), identificando a importância relativa dos preditores no processo de explicação da variabilidade da variável resposta (número de casos de doenças negligenciadas). Para que políticas em saúde pública possam ser direcionadas às variáveis que estão explicando um maior número de doenças negligenciadas.

Ao longo do trabalho foi utilizado a palavra “casos” para expressar o número de ocorrências de doenças negligenciadas registradas pelos Sistemas de Informação da Saúde (SIS), visto que muitas pessoas adoecem, mas os sistemas não registram. Também o conceito de “morbidade” é usado para descrever como e quanto a população adoecer.

## 2.2 Métodos

### 2.2.1 Área de estudo

As doenças negligenciadas afetam mais as populações de países tropicais em desenvolvimento como é o caso do Brasil (BOOTH, 2018). O Brasil possui mais de 210 milhões de habitantes, e uma das maiores extensões territoriais do planeta, com uma área de 8,5 milhões de km<sup>2</sup> (Figura2).

O elevado número de casos doenças negligenciadas no país - mais de 14 milhões de em uma década - tem sido relacionado, entre outros fatores, ao nível de desenvolvimento econômico, social e ambiental (TELES; FONSECA; GONÇALVES, 2019), como a concentração de populações vulneráveis em centros urbanos.

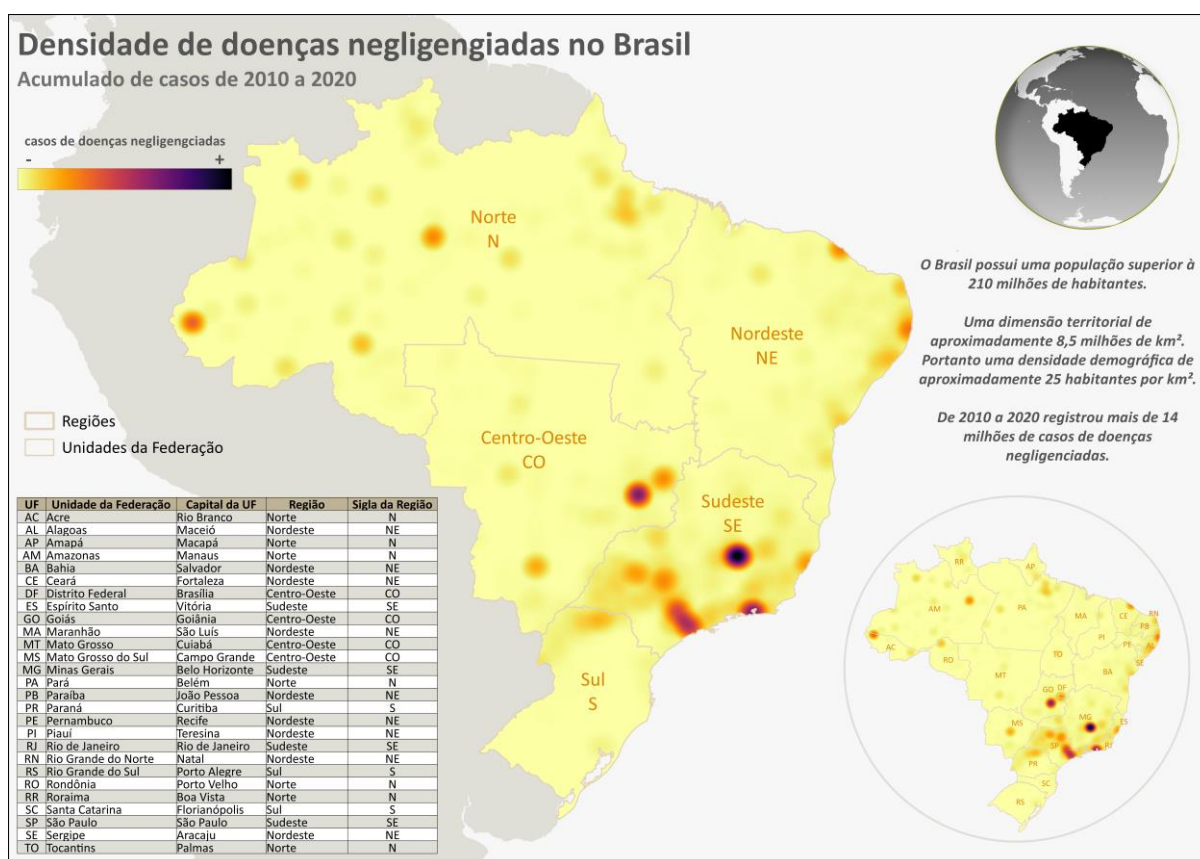


Figura 2: Infográfico da área de estudo, mostrando áreas com maior número de casos de doenças negligenciadas no território nacional brasileiro. A maior intensidade de cores escuras, representa um maior número de casos de doenças negligenciadas. Fonte: Elaborado pelo autor, 2023.

### 2.2.2 Revisão da literatura

A revisão da literatura mostra que são múltiplas as variáveis que estão associadas à ocorrência de doenças negligenciadas. Por exemplo, baixa renda é associada à ocorrência de doenças negligenciadas por 25% dos artigos revisados. Variáveis associadas a tempo e clima são apontadas em 20% dos trabalhos revisados (Tabela 1).

Foram revisados 134 trabalhos, tabulados e analisados de acordo com as perguntas da pesquisa. Por meio da leitura de todos os resumos, 13 trabalhos não possuíam nenhuma aderência com a pesquisa, e foram excluídos. Os 121 trabalhos restante, foram tabulados e classificados em três níveis de aderência, ‘1’ baixa aderência, ‘2’ média e ‘3’ alta aderência. Tal classificação dependia do número de perguntas que o trabalho respondia.

**Tabela 1 – Fatores associados às doenças negligenciadas**

<b>TEMA</b>	<b>%</b>
<b>Baixa renda</b>	25%
<b>Tempo/Clima</b>	20%
<b>Presença de outras doenças concomitantes</b>	13%
<b>Baixa escolaridade</b>	6%
<b>Rural</b>	6%
<b>Saneamento ruim ou ausente</b>	6%
<b>Moradia precária</b>	3%
<b>Sistema de Saúde fragilizado</b>	3%
<b>Urbanização</b>	3%
<b>Poluição</b>	2%
<b>Vacinação em declínio</b>	2%
<b>Cor</b>	2%
<b>Cultura</b>	2%
<b>Faixa etária</b>	2%
<b>Sexo</b>	2%
<b>Interfaces Florestais</b>	2%
<b>Migrações</b>	2%
<b>Política</b>	2%
<b>TOTAL</b>	100%

Fonte: Elaborado pelo autor 2023, baseado em pesquisas de artigos científicos utilizando o banco de dados “Scopus”, com base nos anos entre 2007 a 2022.



### 2.2.3 Coleta de dados

Os casos de doenças negligenciadas no Brasil de 2010 a 2020 por municípios brasileiros (5.570 municípios) foram coletados dos Sistemas de Informação da Saúde (SIS) (Tabela 2).

**Tabela 2 – Sistemas de Informação da Saúde – Origem dos dados consultados para cada uma das doenças negligenciadas incluídas no estudo**

<b>DOENÇA NEGLIGENCIADA</b>	<b>Sigla</b>	<b>Extenso</b>
<b>Doença de Chagas</b>	SINAN	Sistema de Informação de Agravos de Notificação
<b>Hanseníase</b>	SINAN	Sistema de Informação de Agravos de Notificação
<b>Leishmaniose Tegumentar</b>	SINAN	Sistema de Informação de Agravos de Notificação
<b>Leishmaniose Visceral</b>	SINAN	Sistema de Informação de Agravos de Notificação
<b>Malária</b>	SIVEP/SINAN	Sistema de Informações de Vigilância Epidemiológica -Malária / Sistema de Informação de Agravos de Notificação
<b>Tuberculose</b>	SINAN	Sistema de Informação de Agravos de Notificação
<b>Dengue</b>	SINAN	Sistema de Informação de Agravos de Notificação
<b>Esquistossomose</b>	SINAN/PCE	Sistema de Informação de Agravos de Notificação / Programa de Controle da Esquistossomose

Fonte: Elaborado pelo autor, 2023.

Os dados socioeconômicos e ambientais (23 variáveis estudadas/testadas), foram originados de diversas fontes: Instituto Brasileiro de Geografia e Estatística (IBGE), Sistema de Cadastro Nacional de Estabelecimentos de Saúde (SCNES), Sistema de Informação do Programa Nacional de Imunizações (SIPNI), WorldClim 2.1, MapBiomas coleção 7, missão SRTM e criadouros de mosquitos do gênero *Anopheles* cadastrados, e adquiridos por meio da Ouvidoria Integrada e Plataforma de Acesso à Informação "Fala.BR" (ver Material Suplementar 1).

Os dados socioeconômicos e ambientais tiveram como recorte temporal o ano de 2010, para que se pudesse testar a relação com o desfecho (acumulado de casos de doenças 2010 a 2020). Quando não disponíveis ou devido as características metodológicas das variáveis socioeconômica e ambientais importantes, o recorte temporal para 2010, optou-se por utilizar o recorte temporal disponível, como no caso de variáveis datadas para 2019, média de 1970 a 2000, soma de 2001 a 2021 e diferença de 2000 para 2010.

No início da pesquisa, realizou-se um estudo das possíveis metodologias e uma análise exploratória robusta dos dados. A revisão da literatura e a consulta a especialistas levaram a

selecionar a metodologia mais adequada. Utilizou-se técnicas para identificar tendências e relacionamentos relevantes. A análise exploratória dos dados foi fundamental para obter *insights* preliminares e direcionar a pesquisa. Essa abordagem sólida permitiu uma compreensão mais profunda do fenômeno estudado e o desenvolvimento de estratégias fundamentadas.

#### 2.2.4 Processamento de dados

Com os dados de doenças negligenciadas de 2010 a 2020, organizados por tipo de doença e segundo municípios, consolidou-se a base de variáveis dependentes do estudo. A mesma consolidação foi feita para as variáveis independentes testadas, prioritariamente para o ano de 2010 e também segundo município.

O primeiro modelo elaborado foi a estimativa de densidade Kernel dos casos de doenças negligenciadas (cada tipo), georreferenciadas nas sedes dos municípios. A densidade foi calculada com base no número de pontos mapeados, considerando como peso a coluna com o total de casos no município. Grandes números de pontos/peso agrupados resultam em maiores valores (“25.1.5. Interpolação — documentação QGIS Documentation”, [s.d.]).

O segundo modelo, diz respeito à elaboração de agrupamentos, por meio do algoritmo K-means (“scipy.cluster.vq.kmeans — SciPy v1.9.3 Manual”, [s.d.]); por exemplo, um município pertenceu ao agrupamento “verde” se estivesse mais próximo (quantitativo de casos de doenças) do agrupamento “verde”, do que qualquer outro agrupamento. Para cada agrupamento relatou-se a média dos casos de doenças e desvio padrão, atribuindo categorias qualitativas para cada doença negligenciada estudada.

Por fim, elaborou-se a partir dos dados, o modelo linear generalizado GLM (*hurdle*), que é um modelo estatístico utilizado para analisar dados de contagem com excesso de zeros (FENG, 2021) , O modelo *hurdle* que incluiu duas partes: a primeira parte é uma regressão logística binária que modela a probabilidade de que a variável dependente seja zero versus diferente de zero, em cada município. A segunda parte é uma regressão binomial que modela a contagem da variável dependente condicional à primeira parte do modelo. Ou seja, quais variáveis independentes são relevantes para o número de casos de doenças negligenciadas nos municípios com registros e sua predição para todos os municípios. Devido algumas limitações, para a Malária e a Esquistossomose o modelo linear generalizado, foi do tipo *family* poisson, onde o interesse é modelar dados de contagem, em que a variável resposta é o número de ocorrências de um evento.

O GLM foi gerado por município e também pixel a pixel. No caso do modelo pixel a pixel a validação seguiu os seguintes passos:

1. Retirou-se 1.000 pontos amostrais do raster de "densidade Kernel" de cada doença estudada com os maiores valores.
2. Retirou-se 1.000 pontos amostrais do raster de "GLM" com base na mesma localização dos pontos amostrais do raster de "densidade Kernel" de cada doença estudada.
3. E por fim, elaborou-se teste de correlação (Kernel x GLM) para cada doença e avaliou-se os resultados (ver Material Suplementar 3).

Todas as etapas de coleta e processamento de dados estão esquematizadas na Figura 3.

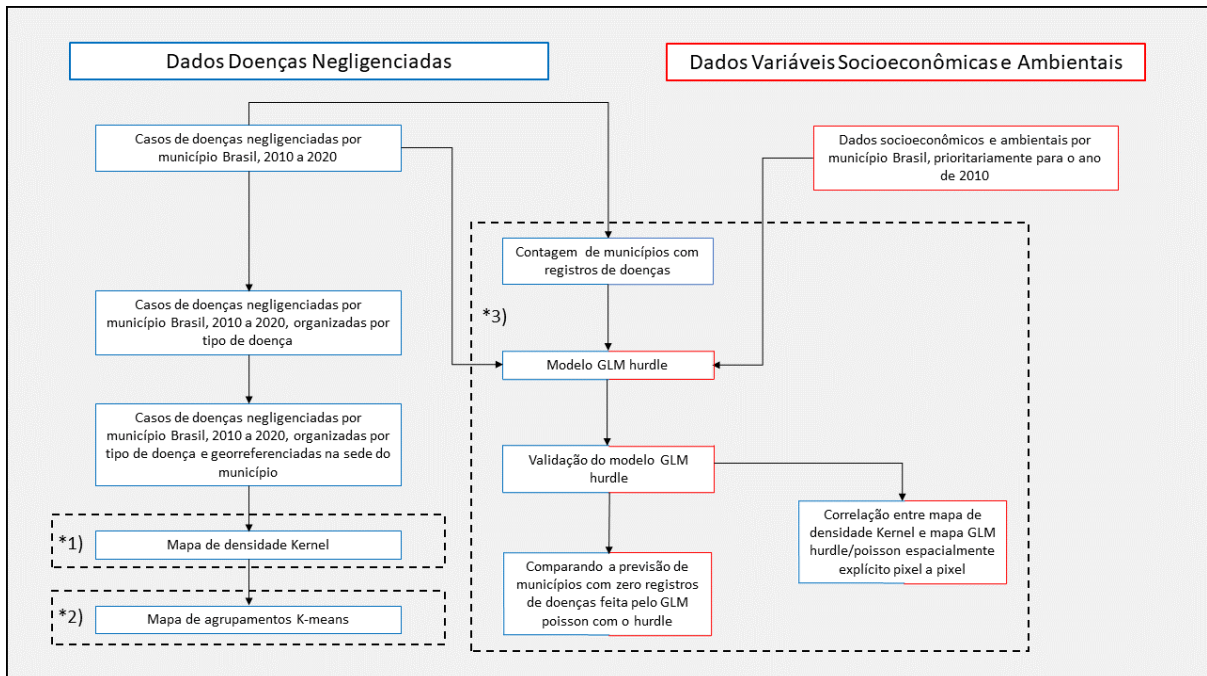


Figura 3: Etapas de coleta e processamento de dados. \*1) Modelo Densidade Kernel, 2) Modelo Agrupamento K-means, 3) Modelo Linear Generalizado. Fonte: Elaborado pelo autor, 2023.

### 2.3 Resultados

Entre 2010 e 2020 foram registrados 14.514.304 casos de doenças negligenciadas. Os casos de Dengue representam um total de 10.517.018 (72,5%), Malária 2.067.827 (14,2%), Tuberculose 793.630 (5,5%), Esquistossomose 476.914 (3,3%), Hanseníase 398.359 (2,7%), Leishmaniose Tegumentar 219.476 (1,5%), Leishmaniose Visceral 38.608 (0,3%) e Doença de Chagas com 2.472 casos, o que representa apenas 0,02% do total de doenças negligenciadas registradas no Brasil de 2010 a 2020.

A região Sudeste apresentou o maior registro percentual do total de doenças negligenciadas, com 42,9% (6.229.214), seguida pela região Norte com 19,6% (2.838.400), Nordeste com 19,1% (2.769.039), Centro-Oeste com 13,4% (1.941.572) e região Sul com 5,1% (736.079).

#### 2.3.1 Densidade espacial dos casos de doenças negligenciadas

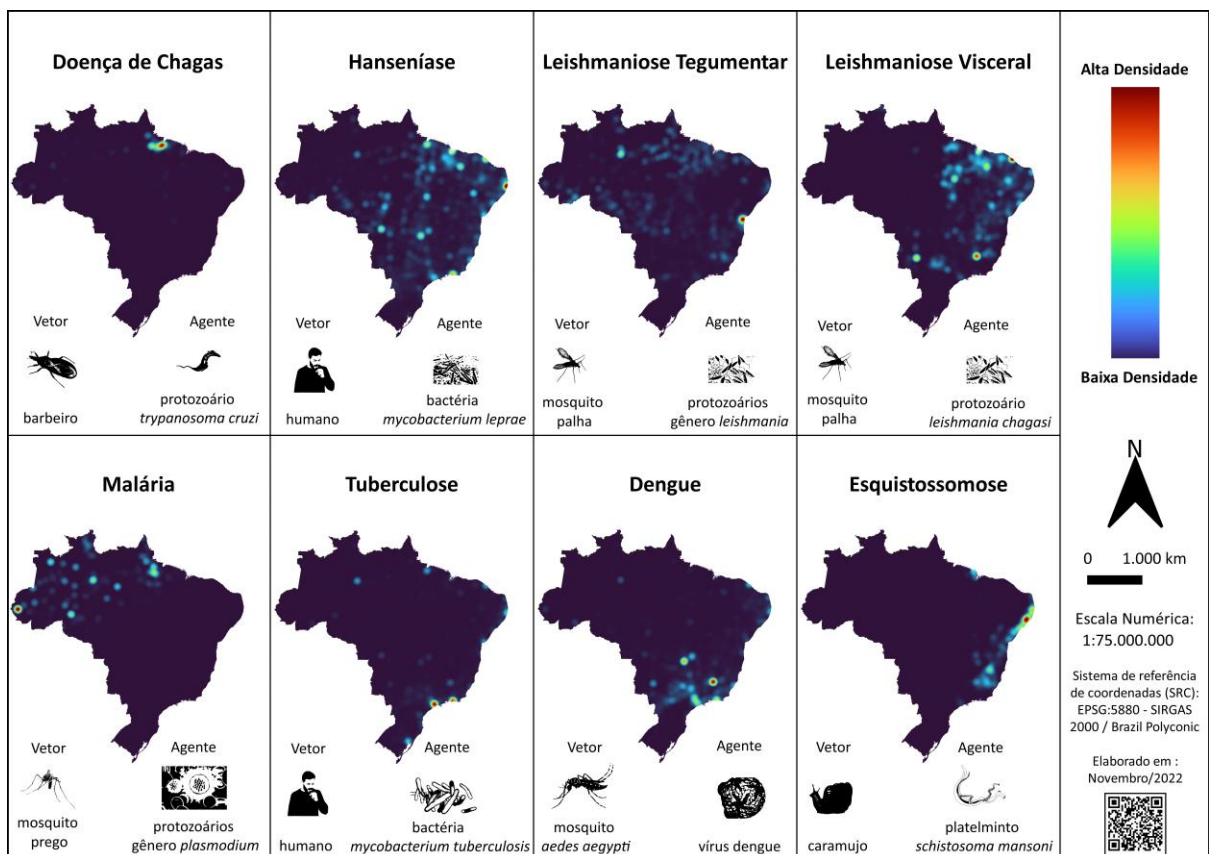


Figura 4: Infográfico e mapa de calor (estimativa de densidade Kernel) das doenças negligenciadas no Brasil – Acumulado de 2010 a 2020, mostrando áreas com maior número de casos para cada doença negligenciada estudada no território nacional brasileiro. A maior intensidade cores do azul claro ao vermelho, representa um maior número de casos de doenças negligenciadas. Fonte: Elaborado pelo autor, 2023.

Como mostra a Figura 4, o modelo de densidade Kernel a partir do número de casos (2010 a 2020) em cada município brasileiro e georreferenciados na sede municipal, indicou espacializações distintas para cada tipo de doença negligenciada estudada.

Embora tenham sido registrados casos de Doença de Chagas em 21 das 27 unidades da federação do Brasil (26 estados mais o distrito federal), a doença se concentra na região Norte do país, com 95% dos casos, tendo o estado do Pará apresentado 81% dos casos. Com atenção para os pontos quentes, os municípios de Abaetetuba-PA, Belém-PA e Breves-PA, com 11%, 10% e 10% do total de casos respectivamente.

Para a Hanseníase no Brasil, o mapa apresenta um caráter mais difuso, no entanto a região Nordeste foi a que apresentou 42% dos casos. O estado do Maranhão apresentou 12% dos casos. Os casos de Leishmaniose Tegumentar no país, apresentam dois polos de elevada densidade, um na região Norte com 44% dos casos e outro na região Nordeste com 29% dos casos. Especificamente os estados do Pará e Bahia, com 17% e 14% respectivamente.

A Leishmaniose Visceral exibe uma maior densidade na região Nordeste, com 55% dos casos. No entanto, existem pontos quentes nas regiões Sudeste, Norte e Centro-Oeste, como por exemplo os estados de Minas Gerais - SE com 16%, Pará - N com 10% e Mato Grosso do Sul - CO com 5% dos casos.

A Malária é uma doença regionalizada no Norte do país, concentrando 99% dos casos observados. Os pontos quentes se concentram nos estados do Amazonas, Pará e Acre, com 37%, 26% e 14% do total de casos. Chama atenção um ponto quente no município de Cruzeiro do Sul no Acre, que apresentou 8% dos casos totais.

A Tuberculose concentrou-se na região Sudeste, com 46% dos casos. E nos estados de São Paulo e Rio de Janeiro com 24% e 25% dos casos respectivamente. O mapa de densidade também evidencia que os pontos quentes estão concentrados nas capitais dos estados, onde dos 10 municípios que apresentaram maiores casos de Tuberculose, nove são capitais de seus estados, em ordem decrescente de casos: São Paulo - SP, Rio de Janeiro - RJ, Manaus - AM, Salvador - BA, Fortaleza - CE, Recife - PE, Belém - PA, Porto Alegre - RS, Duque de Caxias - RJ e São Luís - MA.

A Dengue também teve o número de casos concentrados na região Sudeste, com 54% dos casos, seguido do Nordeste com 19%, Centro-Oeste 17% Sul 6% e Norte com 5% dos casos. Os estados de São Paulo, Minas Gerais e Goiás, apresentaram os maiores números de casos, com 23%, 20% e 10% do total de casos de Dengue, respectivamente. O município de Belo Horizonte - MG foi o ponto mais intenso no mapa, com 4,3% dos casos de todo o país.

O perfil da Esquistossomose no Brasil é bem delineado, com 68% dos casos na região Nordeste e 31% no Sudeste. O mapa evidencia os três pontos principais da Esquistossomose no país, um no Nordeste, na faixa leste dos estados da Paraíba, Pernambuco, Alagoas e Sergipe, que juntos concentram 45% dos casos. Ainda no Nordeste, um segundo ponto no noroeste do estado do Maranhão, que no total representa 6% dos casos. E por fim, a faixa leste do estado de Minas Gerais mais a faixa oeste do estado do Espírito Santo, que juntos representam 30% dos casos. Na Tabela 3 observa-se o *ranking* de incidência de doenças negligenciadas por 100.000 habitantes na segunda coluna, ordenado da maior incidência para a menor. No período de estudo, para o total de doenças negligenciadas, o AC apresentou maior incidência, seguido por RR, depois AM, esses dois últimos estados dividem a terra indígena Yanomami, que foi cenário da crise humanitária de saúde no início do ano de 2023, ocasionada por atividades de mineração ilegal.

Observa-se que a distribuição para cada tipo de doenças negligenciada tem um padrão espacial diferente, variando a depender da região, demonstrado na Figura 4.

**Tabela 3 – Incidência acumulada (2010 a 2020) por 100.000 habitantes (2010) por Unidade da Federação – Ordenado de forma decrescente pela ‘Incidência Total Negligenciadas’**

UF	Incidência Total Negligenciadas	Incidência Chagas	Incidência Hanseníase	Incidência Leishmaniose T	Incidência Leishmaniose V	Incidência Malária	Incidência Tuberculose	Incidência Dengue	Incidência Esquistossomose
AC	59590,2	9,1	256,0	1596,1	0,0	40623,9	589,2	16515,9	0,0
RR	36169,1	0,7	329,9	904,8	49,5	30794,1	446,4	3643,5	0,2
AM	27226,3	3,4	203,0	542,2	0,3	21962,9	859,9	3654,6	0,1
AP	24821,9	16,7	239,6	1130,4	0,6	20296,4	361,2	2776,6	0,4
GO	17682,0	0,3	386,6	86,4	9,2	1,0	171,4	17025,4	1,6
RO	16084,4	0,3	544,2	726,6	0,6	10814,5	387,1	3589,2	22,0
MS	15795,2	0,0	353,4	55,6	80,9	1,8	423,0	14877,8	2,6
MG	11702,4	0,0	82,9	88,0	26,4	0,2	199,7	10659,1	646,2
MT	10622,8	0,1	1427,3	837,4	10,9	535,7	454,0	7349,9	7,6
ES	10238,8	0,0	219,7	46,2	2,3	12,9	379,0	9144,6	434,1
PA	10063,6	26,3	554,3	488,3	51,1	7050,2	547,6	1333,0	12,8
AL	9380,2	0,0	141,8	20,8	17,1	0,1	362,1	5754,7	3083,7
RN	8090,8	0,3	100,3	4,9	31,8	0,4	365,4	7346,4	241,3
TO	7786,9	3,1	1147,9	375,6	207,6	7,4	144,7	5898,9	1,6
DF	6608,4	0,0	109,1	17,1	6,4	0,3	144,6	6328,5	2,4
SP	6375,5	0,0	47,7	9,2	4,7	0,3	452,9	5844,3	16,2
PR	5935,1	0,0	96,1	34,0	0,5	0,7	233,8	5566,1	3,8
RJ	5911,9	0,0	98,2	3,9	0,6	0,7	766,9	5036,9	4,7
PE	5683,3	0,4	371,8	42,0	16,4	0,1	556,6	3971,9	724,1
CE	5448,4	0,0	277,0	82,1	57,5	0,1	443,9	4576,4	11,3
BA	4913,4	0,0	220,7	219,3	26,4	0,7	361,3	3553,3	531,7
PB	4714,9	0,1	201,6	17,8	12,3	0,6	314,8	3839,8	327,8
SE	4139,2	0,0	234,2	3,7	34,1	0,0	346,2	1551,1	1969,9
MA	3208,6	0,7	721,0	307,2	95,2	120,1	345,4	1175,7	443,2
PI	3101,2	0,0	444,8	29,5	70,3	3,2	241,6	2310,9	0,9
SC	745,2	0,0	36,2	3,5	0,3	0,2	310,1	394,1	0,9
RS	649,4	0,0	16,6	0,8	0,4	0,1	486,1	145,2	0,2
<b>BRASIL</b>	<b>7607,0</b>	<b>1,3</b>	<b>208,8</b>	<b>115,0</b>	<b>20,2</b>	<b>1083,8</b>	<b>415,9</b>	<b>5512,0</b>	<b>250,0</b>

Fonte: Elaborado pelo autor 2023, baseado nos dados dos Sistemas de Informação da Saúde e IBGE.

### 2.3.2 Agrupamento municipal da morbidade

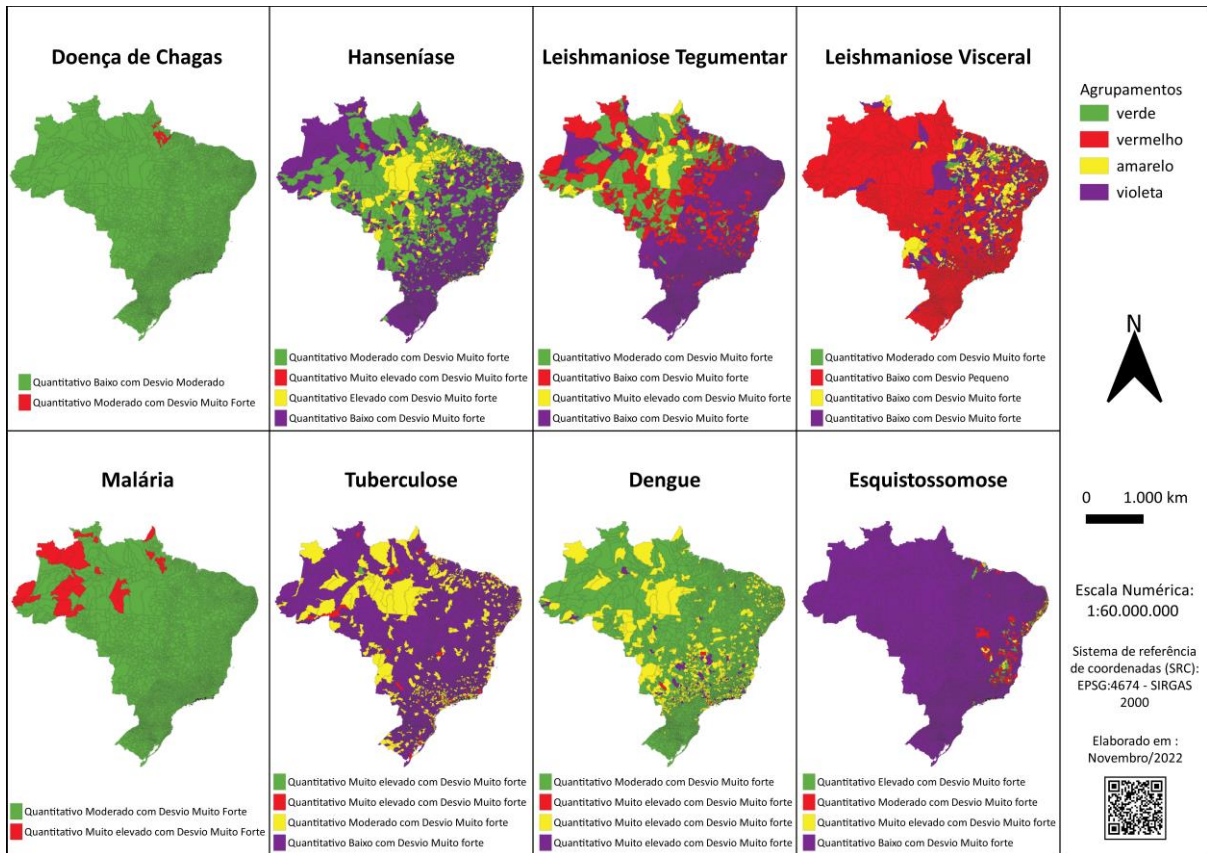


Figura 5: Mapa de agrupamentos (k-means) das doenças negligenciadas no Brasil – Acumulado de 2010 a 2020. Fonte: Elaborado pelo autor, 2023.

O algoritmo “k-means” ajustou a categorização dos casos de doenças negligenciadas em agrupamentos, segundo similaridade, considerando a coluna com o número de casos.

Dois tipos de doenças (Doença de Chagas e Malária) apresentaram dois agrupamentos “verde” (menos presença) e “vermelho” (mais presença). Enquanto as demais doenças apresentaram quatro agrupamentos: “verde”, “vermelho”, “amarelo” e “violeta” com diferentes médias e desvios padrões para o número de casos.

Para fins analítico descritivos, a média de casos entre os municípios de um mesmo agrupamento seguiu a seguinte classificação: 0 a 100 casos (Quantitativo de casos Baixo), 100 a 500 (Quantitativo de casos Moderado), 500 a 1.500 (Quantitativo de casos Elevado) e > 1.500 (Quantitativo de casos Muito elevado). Podendo ser com Desvio Padrão Pequeno quando o resultado entre os municípios do mesmo agrupamento for < 1; com Desvio Padrão Moderado quando entre 1 e 2; e com Desvio Padrão Muito Forte quando > 3.

A Doença de Chagas apresentou característica bi polarizada no agrupamento em escala nacional, com 13 municípios da região Norte: 12 do estado do Pará e um do Amapá, compondo um polo com 121,2 casos de média e desvio padrão de 81,7 entre os municípios. E os outros



5.557 municípios do país compondo o outro grupo de 0,2 casos de média e desvio padrão de 1,8.

A Hanseníase dividiu os municípios brasileiros em quatro agrupamentos equivalentes, “*verde*” com 850 municípios (147,3 casos e desvio padrão de 75), “*vermelho*” com 20 (3.947,3 casos e desvio padrão de 1.692,2), “*amarelo*” com 174 (739,7 casos e desvio padrão de 362,6) e “*violeta*” com 4.526 municípios (14,5 casos e desvio padrão de 16,2).

Em relação a Leishmaniose Tegumentar, observou-se 200 municípios em “*verde*” (415,6 casos e desvio padrão de 166), 591 em “*vermelho*” (98,7 casos e desvio padrão de 48,7), 32 em “*amarelo*” (1.645,4 casos e desvio padrão de 970,16) e 4.746 municípios em “*violeta*” (5,3 casos e desvio padrão de 8,6). Para Leishmaniose Visceral, “*verde*” (382,7 casos e desvio padrão de 317,5), “*vermelho*” (0,6 casos e desvio padrão de 1), “*amarelo*” (56,6 casos e desvio padrão de 33,8) e “*violeta*” (11,2 casos e desvio padrão de 5,6), apresentaram 27, 4.433, 292 e 818 municípios respectivamente.

A Malária também exibiu uma configuração bipolar, com 5.540 municípios em um agrupamento de 126,9 casos de média e desvio padrão de 1.060,6 (“*verde*”) e 30 municípios em outro com 45.485,7 casos de média e desvio padrão de 33.988,7 (“*vermelho*”), todos municípios de “*vermelho*” na região Norte do país.

Para a Tuberculose, três municípios no grupo “*verde*” (50.596,7 casos e desvio padrão de 25.194,9): Manaus, capital do estado do Amazonas-N; assim como Rio de Janeiro e São Paulo, capitais dos estados homônimos, na região Sudeste do país. Outros 86 municípios no agrupamento “*vermelho*” (3.446,6 casos e desvio padrão de 3.523,9), 608 em “*amarelo*” (359,8 casos e desvio padrão de 262,3) e 4.873 municípios no grupo “*violeta*” (26 casos e desvio padrão de 28,1).

A Dengue agrupa 4.779 municípios em “*verde*” (361,7 casos e desvio padrão de 426,1), 12 em “*vermelho*” (212.966,5 casos e desvio padrão de 114.594,4), 687 em “*amarelo*” (4.613,5 casos e desvio padrão de 2.904,4) e 92 em “*violeta*” (33.297,1 casos e desvio padrão de 19.192,4). Dos municípios que compõem “*vermelho*”, sete são capitais de estados, dois fazem parte da região metropolitana de capitais de estados e outros três são grandes cidades do interior do estado de São Paulo.

E por fim, a Esquistossomose, agrupa em “*verde*” (1.194,9 casos e desvio padrão de 423,7) 137 municípios, “*vermelho*” (331,7 casos e desvio padrão de 148,6) com 326, “*amarelo*” (3.810,8 casos e desvio padrão de 1.522,4) com 46 e “*violeta*” (5,9 casos e desvio padrão de 18,5) com 5.061 municípios.

### 2.3.3 Modelo linear generalizado

Os resultados do modelo linear generalizado foram gerados para todos os municípios e pixel a pixel. Os resultados por municípios tiveram um valor único para cada município, uniformizando o resultado para todo o limite municipal, já pixel a pixel, é possível observar as diferenças na vulnerabilidade para cada doença em um mesmo município (Figura 6).

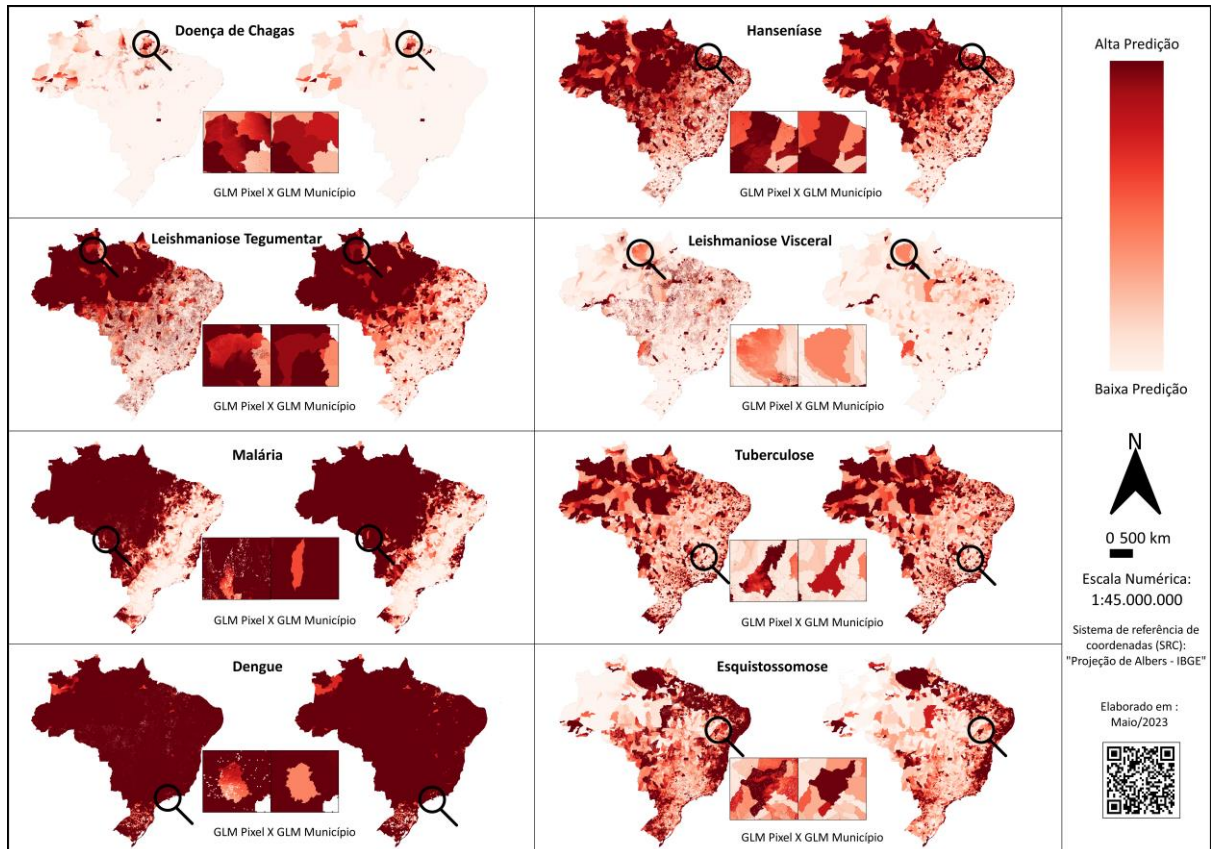


Figura 6: Mapa de GLM pixel a pixel e por municípios normalizados de 0 a 100 para as doenças negligenciadas no Brasil, mostrando áreas com maior predição (maior vulnerabilidade) com cores escuras, e com menor predição (menor vulnerabilidade) com cores claras. Fonte: Elaborado pelo autor 2023.

A Figura 6 traz um compilado gráfico dos resultados sobre as diferenças de valores estimados pelo GLM por município e pixel a pixel. Para Exemplificar essa diferença numericamente em um município brasileiro, trouxe-se o caso da Dengue (Figura 7), a doença com maior ocorrência no período de estudo, mais especificamente o caso de Belo Horizonte MG, que foi o município brasileiro que mais registrou casos de Dengue de 2010 a 2020 (452.746 casos). Nesse município o GLM por município teve valor único estimado de 536.233.728. Agora quando considera-se cada pixel de Belo Horizonte, o valor estimado varia de 148 (áreas claras) a 1.610.007.936 (áreas escuras). Assim é possível observar as áreas mais vulneráveis para Dengue dentro da cidade.

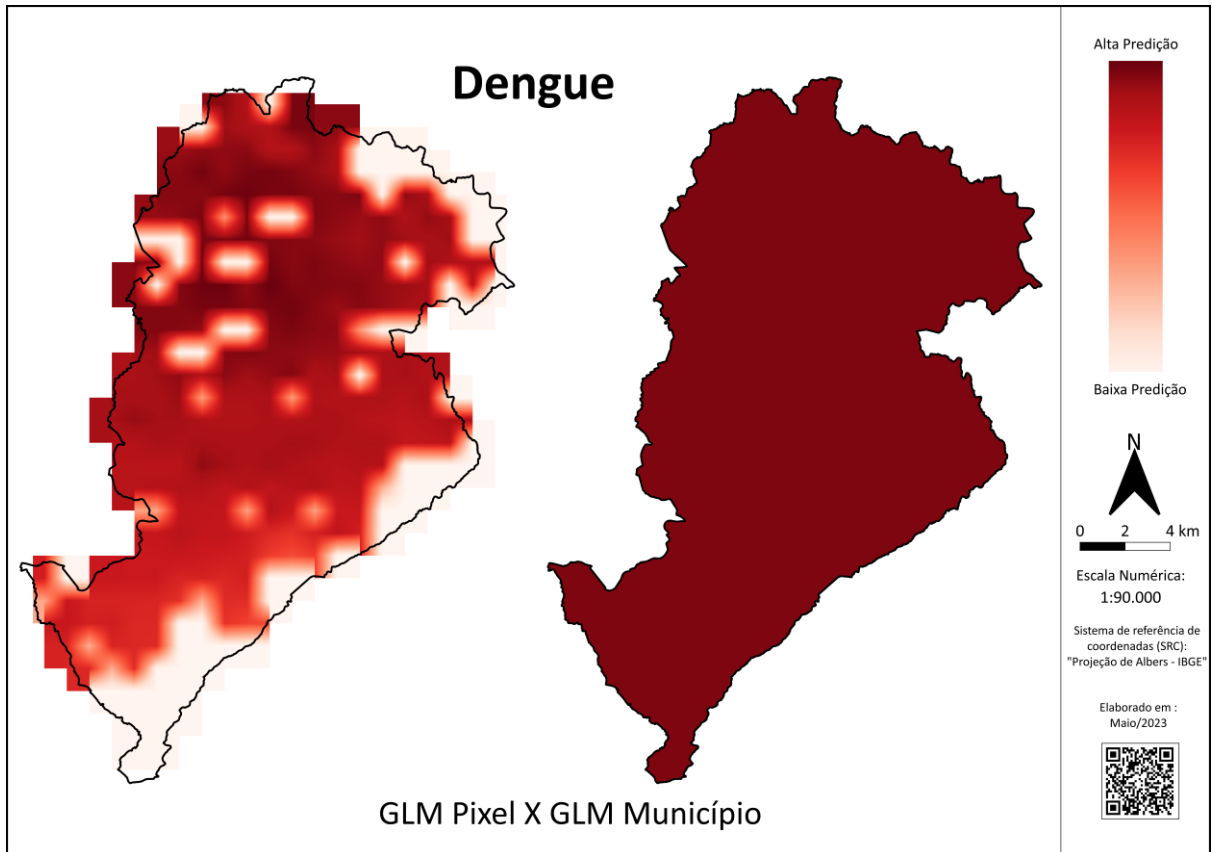


Figura 7: Mapa de GLM pixel a pixel e por municípios normalizados para a Dengue em Belo Horizonte, mostrando áreas com maior predição (maior vulnerabilidade) com cores escuras, e com menor predição (menor vulnerabilidade) com cores claras. Fonte: Elaborado pelo autor 2023.

A validação da modelagem linear generalizada pixel a pixel, correlacionando-a com a modelagem de densidade Kernel pixel a pixel, resultou em praticamente todas as doenças estudadas com correlações fracas. Doença de Chagas apresentou correlação 0,0 (Fraca), Hanseníase correlação 0,2 (Fraca), Leishmaniose Tegumentar correlação -0,1 (Fraca), Leishmaniose Visceral correlação -0,2 (Fraca), Malária correlação 0,3 (Moderada), Tuberculose correlação 0,1 (Fraca), Dengue correlação 0,3 (Moderada) e Esquistossomose correlação -0,1 (Fraca).

**Tabela 4 – Resultados da análise da base de dados que justificam a utilização do GLM ‘hurdle’ – casos de 2010 a 2020 Brasil**

<b>DOENÇAS</b>	<b>municípios sem casos de doenças no período</b>	<b>municípios com casos de doenças no período</b>	<b>municípios sem casos de doenças no período (previsão glm_poisson)</b>	<b>média de casos</b>	<b>variância dos casos</b>	<b>municípios sem casos de doenças no período (previsão glm_hurdle)</b>
<b>Doença de Chagas</b>	5406	164	5208	0,4	51,8	5406
<b>Hanseníase</b>	562	5008	57	71,5	86345,2	562
<b>Leishmaniose Tegumentar</b>	1802	3768	17	39,4	27839,3	1802
<b>Leishmaniose Visceral</b>	3061	2509	1292	6,9	1384,4	3061
<b>Malária</b>	4751	819	1145	371,2	18158067,0	X
<b>Tuberculose</b>	113	5457	13	142,5	1982640,0	112,998
<b>Dengue</b>	185	5385	0	1888,2	148369839,0	185
<b>Esquistossomose</b>	3425	2145	88	85,6	178926,7	X

Fonte: Elaborado pelo autor com base nos dados dos Sistemas de Informação da Saúde e Variáveis socioeconômicas e ambientais elencadas. \*X: script apresentou problemas, optou-se em utilizar GLM ‘poisson’ para essas duas doenças, então para malária e esquistossomose não foram testados o efeito negativo ou positivo das variáveis socioeconômicas e ambientais no número de casos.

A Tabela 4 com os seus resultados, evidencia que a melhor escolha é adotar o GLM-hurdle, por ter municípios com zero casos de doenças e variância maior que a média de casos (FENG, 2021).

Por meio do GLM ‘hurdle’, os resultados podem ser interpretados em termos das estimativas de parâmetros e dos efeitos das variáveis independentes na variável dependente. Assim sendo, as variáveis independentes significativas na modelagem, seu efeito positivo ou negativo na quantidade de casos nos municípios com casos, são compilados Tabela 5.

**Tabela 5 – Efeito de variáveis independentes (efeito negativo ou positivo) sobre os municípios com casos de Doenças Negligenciadas**

<b>VARIÁVEIS</b>	<b>Efeito negativo na quantidade de casos (-)</b>	<b>Efeito positivo na quantidade de casos (+)</b>
<b>Altitude</b>	Tuberculose	Hanseníase, Leishmaniose T e Leishmaniose V
<b>Número de doses de vacinas aplicadas 2010</b>	x	Tuberculose
<b>Número de estabelecimentos de saúde da atenção primária 2019</b>	x	Doença de Chagas e Leishmaniose V
<b>Número de habitantes por km² 2010</b>	Doença de Chagas	x
<b>Percentual da área desmatada de 2000 para 2010</b>	Hanseníase, Leishmaniose T	x
<b>Percentual da população alfabetizada 2010</b>	Leishmaniose T	x
<b>Percentual de cobertura populacional na atenção básica 2010</b>	Doença de Chagas e Leishmaniose V	x
<b>Percentual de domicílios abastecidos por rede geral de água 2010</b>	x	Leishmaniose V
<b>Percentual de domicílios com esgotamento sanitário para vala 2010</b>	Leishmaniose V	x
<b>Percentual de domicílios com lixo coletado 2010</b>	Leishmaniose T e Leishmaniose V	x
<b>Percentual de domicílios com renda mensal de até 0,25 salário mínimo 2010</b>	Hanseníase, Leishmaniose T	x
<b>Percentual de domicílios rurais em 2010</b>	Leishmaniose T e Leishmaniose V	x
<b>Percentual de domicílios sem banheiro 2010</b>	x	Leishmaniose V
<b>Percentual de domicílios sem energia elétrica 2010</b>	x	Doença de Chagas e Leishmaniose T
<b>Precipitação média de 1970 a 2020</b>	Leishmaniose V	Doença de Chagas, Hanseníase e Leishmaniose T
<b>Temperatura média de 1970 a 2000</b>	x	Doença de Chagas, Hanseníase, Leishmaniose T e Leishmaniose V

Fonte: Elaborado pelo autor com base nos dados dos Sistemas de Informação da Saúde e Variáveis socioeconômicas e ambientais elencadas.

A Tabela 5 apresenta as variáveis estatisticamente - valor-p significativo (ver Material Suplementar 3) - significativas para o efeito negativo ou positivo nos casos de algumas doenças negligenciadas, obtidas por meio do GLM-hurdle. Esses resultados foram gerados a partir de um único valor para cada variável preditora do município, por isso a importância de pensar nessa modelagem pixel a pixel. Neste trabalho as variáveis que não têm um valor único (para o GLM Pixel a Pixel) para o município, são precipitação, temperatura, altitude, área urbanizada

e desmatada, sendo as restantes um valor único à escala do município. No entanto a totalidade dos casos das doenças são disponibilizados à escala municipal. Os resultados que corroboram com a literatura revisada foram as seguintes:

- Quanto maior a Altitude, menor o número de casos de Tuberculose com o valor estimado pelo modelo de -2,118.
- Quanto maior o Número de habitantes por km<sup>2</sup> 2010, menor o número de casos de Doença de Chagas com o valor estimado pelo modelo de -1,906.
- Quanto maior o Percentual da população alfabetizada 2010, menor o número de casos de Leishmaniose Tegumentar com o valor estimado pelo modelo de -37,06.
- Quanto maior o Percentual de cobertura populacional na atenção básica 2010, menor o número de casos de Doença de Chagas e Leishmaniose Visceral, com os valores estimados pelo modelo de -17,77 e -8,11 respectivamente.
- Quanto maior o Percentual de domicílios com lixo coletado 2010, menor o número de casos de Leishmaniose Tegumentar e Visceral, com os valores estimados pelo modelo de -29,14 e -28,12 respectivamente.
- Quanto maior o Percentual de domicílios rurais em 2010, menor o número de casos de Leishmaniose Visceral com o valor estimado pelo modelo de -17,12.
- Quanto maior a Altitude, maior o número de casos de Hanseníase, Leishmaniose Tegumentar e Visceral com os valores estimados pelo modelo de 1,34; 2,53 e 2,047 respectivamente.
- Quanto maior o percentual de Percentual de domicílios sem banheiro 2010, maior o número de casos de Leishmaniose Visceral, com valor estimado pelo modelo de 39,99.
- Quanto maior o Percentual de domicílios sem energia elétrica 2010, maior o número de casos de Doença de Chagas e Leishmaniose Tegumentar com os valores estimados pelo modelo de 68,4 e 46,5.
- Quanto maior a Precipitação média, maior o número de casos de Doença de Chagas, Hanseníase e Leishmaniose Tegumentar com os valores estimados pelo modelo de 27,23; 12,19 e 17,92 respectivamente.
- Quanto maior a Temperatura média, maior o número de casos de Doença de Chagas, Hanseníase, Leishmaniose Tegumentar e Visceral com os valores estimado pelo modelo de 372,4; 522,5; 391,5 e 496,2 respectivamente.

A Tabela 5 apresenta as variáveis preditoras estatisticamente significativas - valor-p significativo - para o efeito negativo ou positivo nos casos de algumas doenças negligenciadas, obtidas por meio do GLM-hurdle. Os resultados quantificam e corroboram as expectativas. No entanto, outras variáveis preditoras, exibiram resultados mais difíceis de explicar o que pode ser interpretado pela forma de gerar esses dados. Um dos problemas é que os casos de doenças refletem a capacidade dos Sistemas de Informação da Saúde (SIS) em registrar esses adoecimentos, visto que muitas pessoas adoecem, mas os sistemas não registram (falha na notificação). Também alguns casos como Leishmaniose, o número de casos é relativamente baixo o que pode ser uma limitação para a sua estimativa. Também a dificuldade da maioria das variáveis ter apenas um valor para explicar os diferentes contextos dos municípios, ou seja, por municípios e não completamente (pixel a pixel), mas o que já indica um caminho futuro para verificação. Ou mesmo pela complexidade do tema, metodologia, quebra de paradigmas e limitações dos dados que podem gerar esses resultados controversos. Predições fora do padrão esperado:

- Quanto maior o Percentual da área desmatada de 2000 para 2010, menor o número de casos de Hanseníase e Leishmaniose Tegumentar com os valores estimados pelo modelo de -56,73 e -44,86 respectivamente.
- Quanto maior o Percentual de domicílios com esgotamento sanitário para 2010, menor o número de casos de Leishmaniose Visceral com o valor estimado pelo modelo de -34,85.
- Quanto maior o Percentual de domicílios com renda mensal de até 0,25 salário mínimo 2010, menor o número de casos de Hanseníase e Leishmaniose Tegumentar com os valores estimados pelo modelo de -134,3 e -188,2 respectivamente.
- Quanto maior o Percentual de domicílios rurais em 2010, menor o número de casos de Leishmaniose Tegumentar com o valor estimado pelo modelo de -25,02.
- Quanto maior a Precipitação média, menor o número de casos de Leishmaniose Visceral com o valor estimado pelo modelo de -11,81.
- Quanto maior o Número de doses de vacinas aplicadas 2010, maior o número de casos de Tuberculose com o valor estimado de 1,608.
- Quanto maior o Número de estabelecimentos de saúde da atenção primária 2019, maior o número de casos de Doença de Chagas e Leishmaniose Visceral com os valores estimados pelo modelo de 52,74 e 73,97 respectivamente

- Quanto maior o Percentual de domicílios abastecidos por rede geral de água 2010, maior o número de casos de Leishmaniose Visceral com o valor estimado pelo modelo de 17,77.



## 2.4 Discussão

A distribuição de doenças negligenciadas em uma sociedade é afetada por determinantes da saúde, tais como as circunstâncias que as pessoas vivem e condições de trabalho (BUSS; PELLEGRINI FILHO, 2007). A Doença de Chagas que do seu total, 81% dos casos se concentraram no estado do Para, merece atenção sobre hábitos e condições que as pessoas vivem, sabe-se que em na maioria dos casos a transmissão provável aconteça por via oral, quando uma pessoa ingere alimentos ou líquidos contaminados com fezes ou urina contendo o parasita *Trypanosoma cruzi* (SILVA; AVIZ; MONTEIRO, 2019).

A complexidade envolvida nesse processo, dificulta os isolamentos do desfecho e uma variável associada, no entanto, tal isolamento, pode se configurar como norteador para as ações de políticas públicas pragmáticas. Considerar um número maior de variáveis determinantes ao mesmo tempo para diferentes doenças, é um caminho para o entendimento dessa complexidade. Os resultados das análises e modelos do presente estudo para as observações de 2010 a 2020, exibiram espacializações segundo cada tipo de doenças negligenciada no Brasil. Além disso, foi possível identificar agrupamentos de municípios com números de casos similares – e suas “intradiferenças”, por meio do cálculo do desvio padrão –, mesmo que não fossem fronteiriços, propiciando identificar dentro de um mesmo agrupamento, quais municípios compartilharam das mesmas características em relação ao número de casos de doenças.

Em termos percentuais a Dengue foi a doença que mais assolou os brasileiros dentre as doenças estudadas, o que representa quase cinco vezes mais do que a segunda, a malária. Essas duas doenças possuem nichos regionais diferentes, evidenciado pela ausência de associação entre casos de dengue e número de criadouros do mosquito *Anopheles* (ver material suplementar 4), assim como os mapas de calor. No Brasil, a Malária (transmissão de plasmodium por mosquitos anofelinos) ocorre principalmente na região Norte, com 99% dos casos, enquanto a transmissão dos arbovírus por *Aedes aegypti* ocorrem em grandes centros urbanos devido a antropofilia e sinantropia do mosquito vetor, exibido na Figura 7, inclusive indicando áreas de mata no município de Belo Horizonte – MG, como menos vulneráveis para a Dengue.

Um estudo elaborado na Tailândia, indicou aumento da incidência da Dengue e redução da Malária, devido a deficiência do controle vetorial da Dengue, além disso ressalta a importância de esforços nacionais e internacionais para fortalecer a capacidade de treinamento e pesquisa (CORBEL et al., 2013).

Com os achados do trabalho é possível concluir que o mapeamento apenas dos casos de doenças, não exprimem as áreas vulneráveis para as doenças negligenciadas. Ainda que seja

evidente que são regiões que merecem atenção e ações de combate, no entanto pode-se afirmar que são regiões de elevada sensibilidade dos Sistemas de Informação da Saúde, ou seja, onde o Estado está conseguindo registrar os adoecimentos. Tal conclusão, é embasada pela baixa correlação entre o mapa de casos doenças negligenciadas (kernel) e o mapa de doenças negligenciadas mais as variáveis preditoras (GLM), o que suscita uma contribuição, a indicação de áreas geográficas vulneráveis para cada doença negligenciada para além das áreas dos registros atuais. Também o município não é uma unidade homogênea, possui diferenças culturais, econômicas, geofísicas e sanitárias em seu território e com o modelo adotado foi possível identificar essas diferenças segundo a ótica das doenças negligenciadas.

As variáveis preditoras que influenciam diferentes doenças negligenciadas, e que corrobora o que foi revisado na literatura, são: altitude, densidade demográfica, escolaridade, saúde básica, coleta de lixo, falta de saneamento, falta de acesso à energia elétrica e mudanças climáticas. O trabalho inova, na quantificação dessas influências e a indicação das doenças negligenciadas mais afetadas por esses determinantes da saúde.

A maior quantidade e qualidade de serviços de saúde, são comumente concentradas em espaços urbanos (GARNELO et al., 2018), devido à uma maior demanda oriunda da concentração de agravos e doenças, o que pode auxiliar na explicação das variáveis preditoras que influenciam algumas doenças negligenciadas. No entanto em alguns casos variáveis benéficas (vacinas aplicadas) estão associados um aumento na quantidade de casos e vice-versa. Esse fato pode ser explicado porque nessas áreas possuem melhores infraestruturas de saúde e por isso mais registro de ocorrências de doenças. No entanto futuras pesquisas poderão avaliar se existe essa associação na escala “intramunicipal”?

As questões sobre mudanças climáticas também devem ser destacadas, quanto maior a precipitação média e temperatura média maior foi a estimativa de Doença de Chagas, Hanseníase, Leishmaniose Tegumentar e Leishmaniose Visceral, quantificado e indicado pela pesquisa. Esses resultados corroboram de algum modo, estudos que apontam para evidente influência de fatores ambientais sobre as doenças negligenciadas, e como essas poderão ser afetadas pelas mudanças climáticas de curto, médio e longo prazo. Com destaque para o aumento da carga viral em mamíferos silvestres, redistribuição de vetores e efeitos das mudanças climáticas nos sistemas de saneamento sanitário (AL-DELAIMY, 2022; BOECKMANN; JOYNER, 2014; BOOTH, 2018; CARLSON et al., 2022; GAINOR; HARRIS; LABEAUD, 2022; HALL; WALL; STEVENS, 2016; NAU et al., 2019).

## 2.5 Conclusão

A morbidade por doenças negligenciadas é um estudo complexo no que se refere a sua associação às variáveis socioeconômicas e ambientais (WHITEHEAD et al., 2016), indo desde uma escala micro/indivíduo, passando pela meso/comunidade e chegando em uma escala macro/sociedade; portanto, dificilmente pode ser explicada por modelos lineares simples. No entanto, esses podem indicar e isolar algumas variáveis independentes estratégicas para compreensão de sua dinâmica. Os modelos lineares generalizados, possuem limitações frente à complexidade do tema, mas podem somar esforços na investigação científica.

As doenças negligenciadas são endêmicas em países tropicais e em desenvolvimento, embora políticas públicas estejam sendo pensadas e executadas para a redução da morbidade, muito ainda pode-se avançar para promover o bem-estar e saúde dessas populações vulneráveis.

Esta pesquisa apresenta numérica e espacialmente o perfil endêmico de doenças negligenciadas em um país tropical emergente. Considerando a escala nacional, há seis grandes contribuições potenciais desta pesquisa.

Em primeiro lugar, os casos de doenças negligenciadas exibem uma espacialização (densidade kernel) segundo o número de casos e para cada tipo de doença em uma década de observações, indicando as áreas prioritárias para uma política nacional de combate à essas doenças.

Em segundo lugar, os casos de doenças negligenciadas mais as variáveis preditoras exibem uma regionalização - por meio do GLM - e para cada tipo de doença em uma década de observações; o que indica as áreas mais vulneráveis, não necessariamente com elevado número de casos, mas com variáveis socioeconômicas e ambientais que indicam áreas propícias para manutenção ou surgimento dessas doenças, para o município e “intramunicipal”.

Em terceiro lugar, dentro da espacialização de cada doença estudada, identifica-se agrupamentos segundo equivalência de número de casos ao nível municipal (mas de modo geral há uma grande variação ou dispersão em torno da média de cada agrupamento), diferenciando-se dentro de uma região de elevada endemicidade os municípios similares.

Importante destacar que as variáveis socioeconômicas e ambientais, que no campo da epidemiologia é chamada como determinantes de saúde, e que na pesquisa chamou-se também de determinantes espaciais, variáveis preditoras ou independentes das doenças negligenciadas, foram elencadas a partir da revisão da literatura, e que eram possíveis de serem levantadas a partir de uma escala nacional. O que proporcionou a quarta grande contribuição, identificar variáveis preditoras relevantes para a dinâmica das doenças negligenciadas no Brasil. E considerando o método de pesquisa utilizado, novas variáveis preditoras podem ser testadas, o

que evidencia mais duas contribuições da pesquisa: Sua replicabilidade e adaptabilidade, ou seja, aperfeiçoamento e melhoria, tanto dos dados de entrada como como as saídas dos modelos.

## REFERÊNCIAS

- 25.1.5. Interpolação — documentação QGIS Documentation.** Disponível em: <[https://docs.qgis.org/3.22/pt\\_BR/docs/user\\_manual/processing\\_algs/qgis/interpolation.html?highlight=kernel](https://docs.qgis.org/3.22/pt_BR/docs/user_manual/processing_algs/qgis/interpolation.html?highlight=kernel)>. Acesso em: 17 nov. 2022.
- AL-DELAIMY, A. K. The Prospective Effects of Climate Change on Neglected Tropical Diseases in the Eastern Mediterranean Region: a Review. **Current Environmental Health Reports**, 2022.
- ALMEIDA, Â. S. et al. Spatial and spatiotemporal dynamics of visceral leishmaniasis in an endemic north-eastern region of Brazil. **Geospatial Health**, v. 15, n. 2, p. 285–292, 2020.
- ALVES, B. / O. / O.-M. **DeCS.** , [s.d.]. Disponível em: <[https://decs.bvsalud.org/th/s/resource/?id=54010&filter=ths\\_termall&q=doen%C3%A7as%20negligenciadas](https://decs.bvsalud.org/th/s/resource/?id=54010&filter=ths_termall&q=doen%C3%A7as%20negligenciadas)>. Acesso em: 8 maio. 2022
- BOECKMANN, M.; JOYNER, T. A. Old health risks in new places? An ecological niche model for *I. ricinus* tick distribution in Europe under a changing climate. **Health & Place**, v. 30, p. 70–77, 1 nov. 2014.
- BOOTH, M. **Climate Change and the Neglected Tropical Diseases.** [s.l.] Academic Press, 2018. v. 100p. 126
- BOUSQUAT, A.; COHN, A. A dimensão espacial nos estudos sobre saúde: uma trajetória histórica. **História, Ciências, Saúde-Manguinhos**, v. 11, n. 3, p. 549–568, dez. 2004.
- BUCK, C. (ED.). **The challenge of epidemiology: issues and selected readings.** Repr ed. Washington, DC: Pan American Health Organization, Pan American Sanitary Bureau, Regional Off. of the World Health Organization, 1989.
- BUSS, P. M.; PELLEGRINI FILHO, A. A saúde e seus determinantes sociais. **Physis: Revista de Saúde Coletiva**, v. 17, n. 1, p. 77–93, abr. 2007.
- CARLSON, C. J. et al. Climate change increases cross-species viral transmission risk. **Nature**, v. 607, n. 7919, p. 555–562, 21 jul. 2022.
- CLARK, N. J. et al. Mapping *Schistosoma mansoni* endemicity in Rwanda: A critical assessment of geographical disparities arising from circulating cathodic antigen versus Kato-Katz diagnostics. **PLoS Neglected Tropical Diseases**, v. 13, n. 9, 2019.
- CORBEL, V. et al. Challenges and prospects for dengue and malaria control in Thailand, Southeast Asia. **Trends in Parasitology**, v. 29, n. 12, p. 623–633, 1 dez. 2013.
- DOS SANTOS, F. L. A. et al. Research, development and innovation for the control of neglected diseases. **Revista de Ciências Farmaceuticas Basica e Aplicada**, v. 33, n. 1, p. 37–47, 2012.
- EBI, K. L.; NEALON, J. Dengue in a changing climate. **Environmental Research**, v. 151, p. 115–123, 1 nov. 2016.

- FENG, C. X. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. **Journal of Statistical Distributions and Applications**, v. 8, n. 1, p. 8, 24 jun. 2021.
- FOTSO, J. C. et al. Child growth in urban deprived settings: Does household poverty status matter? At which stage of child development? **Health & Place**, v. 18, n. 2, p. 375–384, 1 mar. 2012.
- GAINOR, E. M.; HARRIS, E.; LABEAUD, A. D. Uncovering the Burden of Dengue in Africa: Considerations on Magnitude, Misdiagnosis, and Ancestry. **Viruses**, v. 14, n. 2, 2022.
- GARNELO, L. et al. Acesso e cobertura da Atenção Primária à Saúde para populações rurais e urbanas na região norte do Brasil. **Saúde em Debate**, v. 42, p. 81–99, set. 2018.
- GILGEN, D. et al. Impact of migration on illness experience and help-seeking strategies of patients from Turkey and Bosnia in primary health care in Basel. **Health & Place**, v. 11, n. 3, p. 261–273, set. 2005.
- GUBLER, D. J. Dengue, Urbanization and Globalization: The Unholy Trinity of the 21st Century. **Tropical Medicine and Health**, v. 39, n. 4SUPPLEMENT, p. S3–S11, 2011.
- HALL, M. J. R.; WALL, R. L.; STEVENS, J. R. **Traumatic Myiasis: A Neglected Disease in a Changing World**. [s.l.] Annual Reviews Inc., 2016. v. 61p. 176
- HARPHAM, T. Urban health in developing countries: What do we know and where do we go? **Health & Place**, v. 15, n. 1, p. 107–116, 1 mar. 2009.
- HATHERELL, H.-A. et al. Sustainable Surveillance of Neglected Tropical Diseases for the Post-Elimination Era. **Clinical Infectious Diseases**, v. 72, p. S210–S216, 2021.
- HOTEZ, P. J. **Blue marble health: An innovative plan to fight diseases of the poor amid wealth**. [s.l.] Johns Hopkins University Press, 2016. p. 205
- HOTEZ, P. J. The rise of neglected tropical diseases in the “new Texas”. **PLoS Neglected Tropical Diseases**, v. 12, n. 1, 2018.
- HOTEZ, P. J.; KAMATH, A. Neglected Tropical Diseases in Sub-Saharan Africa: Review of Their Prevalence, Distribution, and Disease Burden. **PLOS Neglected Tropical Diseases**, v. 3, n. 8, p. e412, 25 ago. 2009.
- KANG, S. et al. Extending the global worm index and its links to human development and child education. **PLOS Neglected Tropical Diseases**, v. 12, n. 6, p. e0006322, 21 jun. 2018.
- MALECELA, M. N.; DUCKER, C. A road map for neglected tropical diseases 2021–2030. **Transactions of The Royal Society of Tropical Medicine and Hygiene**, v. 115, n. 2, p. 121–123, 28 jan. 2021.
- MENDONÇA, F. DE A.; SOUZA, A. V. E; DUTRA, D. DE A. Saúde pública, urbanização e dengue no Brasil. **Sociedade & Natureza**, v. 21, p. 257–269, dez. 2009.
- MITRA, A. K.; MAWSON, A. R. Neglected Tropical Diseases: Epidemiology and Global Burden. **Tropical Medicine and Infectious Disease**, v. 2, n. 3, p. 36, 5 ago. 2017.

MONTEIRO, L. D. et al. Social Determinants Of Leprosy In A Hyperendemic State In North Brazil. **Revista de Saude Publica**, v. 51, 2017.

MULLIGAN, K.; ELLIOTT, S. J.; SCHUSTER-WALLACE, C. The place of health and the health of place: Dengue fever and urban governance in Putrajaya, Malaysia. **Health & Place**, v. 18, n. 3, p. 613–620, maio 2012.

MURRAY, C. J. L.; LOPEZ, A. D. Measuring the Global Burden of Disease. **New England Journal of Medicine**, v. 369, n. 5, p. 448–457, 1 ago. 2013.

NAU, L. H. et al. Leptospirosis in Germany: current knowledge on pathogen species, reservoir hosts, and disease in humans and animals. **Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz**, v. 62, n. 12, p. 1510–1521, 2019.

OMS. Weekly Epidemiological Record, 2021, vol. 96, 48 [full issue]. **Weekly Epidemiological Record = Relevé épidémiologique hebdomadaire**, v. 96, n. 48, p. 585–596, 3 dez. 2021.

OPAS. **OMS lança plano de 10 anos para acabar com sofrimento causado por doenças tropicais negligenciadas - OPAS/OMS | Organização Pan-Americana da Saúde.**

Disponível em: <<https://www.paho.org/pt/noticias/29-1-2021-oms-lanca-plano-10-anos-para-acabar-com-sofrimento-causado-por-doencas-tropicais>>. Acesso em: 6 maio. 2022.

SALEEM, S. et al. Reduce disease burden of human schistosomiasis in Asia through biological control. **Mini-Reviews in Medicinal Chemistry**, v. 20, n. 12, p. 1118–1132, 2020.

**scipy.cluster.vq.kmeans — SciPy v1.9.3 Manual.** Disponível em:

<<https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.vq.kmeans.html#scipy.cluster.vq.kmeans>>. Acesso em: 17 nov. 2022.

SILVA, G. G. E.; AVIZ, G. B. D.; MONTEIRO, R. C. Perfil epidemiológico da Doença de Chagas aguda no Pará entre 2010 e 2017. **Pará Research Medical Journal**, v. 4, p. e29, 2019.

STEINMANN, P. et al. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. **The Lancet Infectious Diseases**, v. 6, n. 7, p. 411–425, 1 jul. 2006.

TELES, G. C.; FONSECA, F. R.; GONÇALVES, M. J. F. American tegumentary Leishmaniasis in the Brazilian Amazon from 2010 to 2014. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 61, 2019.

TONG, M. X. et al. Infectious Diseases, Urbanization and Climate Change: Challenges in Future China. **International Journal of Environmental Research and Public Health**, v. 12, n. 9, p. 11025–11036, set. 2015.

WHITEHEAD, M. et al. How could differences in ‘control over destiny’ lead to socio-economic inequalities in health? A synthesis of theories and pathways in the living environment. **Health & Place**, v. 39, p. 51–61, 1 maio 2016.

WORLD HEALTH ORGANIZATION. **The sanitation challenge : turning commitment into reality.** [s.l: s.n.].

WORLD HEALTH ORGANIZATION. **Ending the neglect to attain the sustainable development goals: a road map for neglected tropical diseases 2021–2030**. Geneva: World Health Organization, 2020.



## ANEXOS

### ANEXO A - Metadados e processamento das variáveis socioeconômicas e ambientais

Material Suplementar (1) para

**Título: Doenças negligenciadas no Brasil: Mapeamento de uma década de ocorrências**

***Texto Suplementar:***

Número de estabelecimentos de saúde da atenção primária 2019

1. Aquisição da malha de pontos, em formato shapefile (.shp), que representa a localização dos estabelecimentos de saúde de atenção primária (Centros de Apoio à Saúde da Família, Unidades Básicas de Saúde, Consultórios de Saúde Pública, Postos de Saúde e Unidades de Atenção à Saúde Indígena) do Brasil. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística (IBGE). (arquivo: 'Esauade\_Atencao\_Primary').  
<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/tipologias-do-territorio/15788-aglomerados-subnormais.html?=&t=acesso-ao-produto>
2. Aquisição do limite dos municípios brasileiros, em formato shapefile (.shp), Sistema de Referência de Coordenadas, SIRGAS 2000'. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística (IBGE). (arquivo: 'BR\_Municipios\_2020').  
<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=downloads>
3. Os dados são referentes às coordenadas e endereços utilizados para espacialização dos Estabelecimentos de Saúde e provenientes do Cadastro Nacional dos Estabelecimentos de Saúde (CNES), cadastro oficial do Ministério da Saúde.
4. Por meio de *software* de Sistema de Informações Geográficas (SIG) - QGIS - elaborou-se a contagem de pontos (estabelecimentos de saúde) para cada um dos 5.570 municípios brasileiros e foi salvo consolidado em arquivo .csv.

Número de estabelecimentos com suporte de observação e internação 2019

1. Aquisição da malha de pontos, em formato shapefile (.shp), que representa a localização dos estabelecimentos de saúde com suporte de observação e internação (Hospital Especializado, Hospital Geral, Pronto Atendimento, Pronto Socorro Especializado, Pronto Socorro Geral e Unidade Mista) do Brasil. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística (IBGE). (arquivo: 'ESaude\_Internacao\_observacao').  
<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/tipologias-do-territorio/15788-aglomerados-subnormais.html?=&t=acesso-ao-produto>
2. Aquisição do limite dos municípios brasileiros, em formato shapefile (.shp), Sistema de Referência de Coordenadas, SIRGAS 2000'. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística (IBGE). (arquivo: 'BR\_Municipios\_2020').  
<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=downloads>
3. Os dados são referentes às coordenadas e endereços utilizados para espacialização dos Estabelecimentos de Saúde e provenientes do Cadastro Nacional dos Estabelecimentos de Saúde (CNES), cadastro oficial do Ministério da Saúde.
4. Por meio de *software* de Sistema de Informações Geográficas (SIG) - QGIS - elaborou-se a contagem de pontos (estabelecimentos de saúde) para cada um dos 5.570 municípios brasileiros e foi salvo consolidado em arquivo .csv.

#### Percentual de cobertura populacional na atenção básica 2010

1. Aquisição do indicador, em formato .xlsx, da cobertura populacional estimada na Atenção Básica em 2010. Indicador selecionado considera a centralidade da Atenção Básica no Sistema Único de Saúde (SUS), com a proposta de constituir-se como ordenadora do cuidado nos sistemas locais de Saúde e eixo estruturante de programas e projetos; além de favorecer a capacidade resolutiva e os processos de territorialização e regionalização em saúde (Ministério da Saúde). Coletado no segundo semestre de 2022 no site do Ministério da Saúde (arquivo: 'Historico-AB-MUNICIPIOS-2007-202012').

<https://egestorab.saude.gov.br/paginas/acessoPublico/relatorios/relHistoricoCoberturaConsolidado.xhtml>

2. A cobertura populacional estimada na Atenção Básica, dada pelo percentual da população coberta por equipes da Estratégia Saúde da Família e por equipes de Atenção Básica tradicional equivalentes e parametrizadas em relação à estimativa populacional.
3. As fontes de informações para elaboração do indicador, foram, segundo o Ministério da Saúde: Sistema de Cadastro Nacional de Estabelecimentos de Saúde (SCNES) Instituto Brasileiro de Geografia e Estatística (IBGE) – Estimativas populacionais anuais de população, com referência em 1º de julho.
4. O indicador tem referência mensal, para o artigo, foi considerado o mês 06 de 2010, para cada um dos 5.570 municípios brasileiros e foi salvo consolidado em arquivo .csv.

#### Percentual de cobertura populacional por agentes comunitários de saúde 2010

1. Aquisição do indicador, em formato .xlsx, da cobertura populacional estimada por Agentes Comunitários de Saúde em 2010. Coletado no segundo semestre de 2022 no site do Ministério da Saúde (arquivo: ‘Historico-ACS-MUNICIPIOS-2007-2018’).  
<https://egestorab.saude.gov.br/paginas/acessoPublico/relatorios/relHistoricoCoberturaConsolidado.xhtml>
2. A cobertura populacional estimada por Agentes Comunitários de Saúde vinculados a equipes da Estratégia Saúde da Família, Equipe de Atenção Básica parametrizada e Equipe de Agentes Comunitários de Saúde, dada pelo percentual da população coberta por esses agentes em relação à estimativa populacional.
3. As fontes de informações para elaboração do indicador, foram, segundo o Ministério da Saúde: Sistema de Cadastro Nacional de Estabelecimentos de Saúde (SCNES) Instituto Brasileiro de Geografia e Estatística (IBGE) – Estimativas populacionais anuais de população, com referência em 1º de julho.
4. O indicador tem referência mensal, para o artigo, foi considerado o mês 06 de 2010, para

cada um dos 5.570 municípios brasileiros e foi salvo consolidado em arquivo .csv.

#### Número de imunizações 2010

1. Aquisição dos dados, em formato .csv, do número de imunizações ('Doses Cálculos CV') em 2010. Coletado no segundo semestre de 2022 no site do Ministério da Saúde.  
[http://tabnet.datasus.gov.br/cgi/dhdat.exe?bd\\_pni/cpnibr.def](http://tabnet.datasus.gov.br/cgi/dhdat.exe?bd_pni/cpnibr.def)
2. A fonte dos dados, foi, segundo o Ministério da Saúde: Sistema de Informação do Programa Nacional de Imunizações (SI-PNI).
3. Os dados foram consolidados para cada um dos 5.570 municípios brasileiros e salvos em arquivo .csv.

#### Percentual da população alfabetizada 2010

1. Aquisição do indicador, em formato .xlsx, taxa de alfabetização das pessoas de 10 anos ou mais de idade em 2010. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística (arquivo: 'tabela1383').  
<https://sidra.ibge.gov.br/Tabela/1383>
2. O indicador, segundo o IBGE, foi elaborado a partir dos Dados do Universo do Censo Demográfico 2010.
3. O indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Número de habitantes por km<sup>2</sup> 2010

1. A elaboração do indicador, de densidade demográfica por município brasileiro em 2010, seguiu os seguintes passos:
2. Aquisição dos dados, em formato .xlsx, população residente 2010. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística (arquivo: 'tabela608').  
<https://sidra.ibge.gov.br/Tabela/608>

3. Os dados, segundo o IBGE, são da Sinopse do Censo Demográfico 2010.
4. Os dados da população residente total, foram salvos e consolidados para cada um dos 5.570 municípios em arquivo .csv.
5. Aquisição do limite dos municípios brasileiros, em formato shapefile (.shp), reprojetoado para Sistema de Referência de Coordenadas, com parâmetros repassados pelo IBGE e definido pelo autor do artigo: “Projeção de Albers – IBGE”. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE. (arquivo: ‘BR\_Municipios\_2020’).  
<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=downloads>
6. Por meio de *software* de Sistema de Informações Geográficas (SIG) - QGIS - recalculou-se a área dos municípios brasileiros “Projeção de Albers – IBGE” em km<sup>2</sup>, salvo consolidado em arquivo .csv.
7. Com os dados de população e área dos municípios organizados em um único arquivo .xlsx, construiu-se o indicador no *software* Excel, dividindo a população do município por sua área. Salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual da área com aglomerados subnormais 2010

1. Aquisição do limite dos municípios brasileiros, em formato shapefile (.shp), reprojetoado para Sistema de Referência de Coordenadas, com parâmetros repassados pelo IBGE e definido pelo autor do artigo: “Projeção de Albers – IBGE”. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE. (arquivo: ‘BR\_Municipios\_2020’).  
<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=downloads>

2. Aquisição do limite dos aglomerados subnormais em 2010, em formato shapefile (.shp), reprojetoado para Sistema de Referência de Coordenadas, com parâmetros repassados pelo IBGE e definido pelo autor do artigo: “Projeção de Albers – IBGE”. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE. (arquivo: ‘arquivos\_SHP’).  
<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/tipologias-do-territorio/15788-aglomerados-subnormais.html?=&t=downloads>
3. Por meio de *software* de Sistema de Informações Geográficas (SIG) – QGIS - calculou-se percentual da área municipal coberta por aglomerados subnormais, depois, salvo consolidado em arquivo .csv.

#### Saldo do número de habitantes de 2000 para 2010

1. A elaboração do indicador, saldo populacional 2000 para 2010, seguiu os seguintes passos:
2. Aquisição dos dados, em formato .xlsx, população residente e localização da área 2000 e 2010. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE (arquivo: ‘tabela1309’).  
<https://sidra.ibge.gov.br/Tabela/1309>
3. Os dados, segundo o IBGE, são da Sinopse do Censo Demográfico 2010 e Sinopse Preliminar do Censo Demográfico 2000.
4. Com os dados de população 2000 e 2010 em um único arquivo .xlsx, construiu-se o indicador no *software* Excel, subtraindo a população do município em 2000 pela população do município em 2010. Salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual de domicílios rurais em 2010

1. Aquisição do indicador, em formato .xlsx, percentual domicílios particulares permanentes, por situação do domicílio ‘Rural’ 2010. Coletado no segundo semestre de

2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE (arquivo: 'tabela1395').

<https://sidra.ibge.gov.br/Tabela/1395>

2. O indicador, segundo o IBGE, foi elaborado a partir dos Dados do Universo do Censo Demográfico 2010.
3. O indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual de domicílios com superlotação 2010

1. Aquisição do indicador, em formato .xlsx, percentual domicílios particulares permanentes, por número de moradores '14 ou mais'. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE (arquivo: 'tabela3219').

<https://sidra.ibge.gov.br/Tabela/3219>

2. O indicador, segundo o IBGE, foi elaborado a partir dos Dados do Universo do Censo Demográfico 2010.
3. O indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual de domicílios com lixo coletado 2010

1. Aquisição do indicador, em formato .xlsx, percentual domicílios particulares permanentes, destino do lixo 'Coletado' 2010. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE (arquivo: 'tabela1395').

<https://sidra.ibge.gov.br/Tabela/1395>

2. O indicador, segundo o IBGE, foi elaborado a partir dos Dados do Universo do Censo Demográfico 2010.
3. O indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual de domicílios sem banheiro 2010

1. Aquisição do indicador, em formato .xlsx, percentual domicílios particulares permanentes, por situação do domicílio e existência de banheiro ou sanitário ‘Não tinham banheiro nem sanitário’ 2010. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE (arquivo: ‘tabela1394’).  
<https://sidra.ibge.gov.br/Tabela/1394>
2. O indicador, segundo o IBGE, foi elaborado a partir dos Dados do Universo do Censo Demográfico 2010.
3. O indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual de domicílios sem energia elétrica 2010

1. Aquisição do indicador, em formato .xlsx, percentual domicílios particulares permanentes, existência de energia elétrica ‘Não tinham’ 2010. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE (arquivo: ‘tabela1395’).  
<https://sidra.ibge.gov.br/Tabela/1395>
2. O indicador, segundo o IBGE, foi elaborado a partir dos Dados do Universo do Censo Demográfico 2010.
3. O indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual de domicílios com esgotamento sanitário para vala 2010

1. Aquisição do indicador, em formato .xlsx, percentual domicílios particulares permanentes, segundo o tipo de esgotamento sanitário ‘Vala’ em 2010. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE (arquivo: ‘tabela1394’).  
<https://sidra.ibge.gov.br/Tabela/1394>



2. O indicador, segundo o IBGE, foi elaborado a partir dos Dados do Universo do Censo Demográfico 2010.
3. O indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual de domicílios abastecidos por rede geral de água 2010

1. Aquisição do indicador, em formato .xlsx, percentual domicílios particulares permanentes, por situação do domicílio e forma de abastecimento de água 'Rede Geral' em 2010. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE (arquivo: 'tabela1395').  
<https://sidra.ibge.gov.br/Tabela/1395>
2. O indicador, segundo o IBGE, foi elaborado a partir dos Dados do Universo do Censo Demográfico 2010.
3. O indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual de domicílios com renda mensal de até 0,25 salário mínimo 2010

1. Aquisição do indicador, em formato .xlsx, percentual domicílios particulares permanentes e moradores em domicílios particulares permanentes, cuja condição no domicílio não era pensionista, nem empregado(a) doméstico(a) ou seu parente, por classes de rendimento nominal mensal domiciliar 'Até 1/4 de salário mínimo' em 2010. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística - IBGE (arquivo: 'tabela1397').  
<https://sidra.ibge.gov.br/Tabela/1395>
2. O indicador, segundo o IBGE, foi elaborado a partir dos Dados do Universo do Censo Demográfico 2010.
3. O indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

### Temperatura média de 1970 a 2000

1. Aquisição das camadas rasters, em formato .tif, com 1km de resolução, de dados climáticos da temperatura média (°C) mensal de 1970 a 2000 de todo o planeta, Sistema de Referência de Coordenadas, WGS 84. Coletado no segundo semestre de 2022 no site da WorldClim versão 2.1 lançado em janeiro de 2020. (arquivos para cada mês do ano: 'wc2.1\_30s\_tavg\_xx').  
<https://www.worldclim.org/data/worldclim21.html>
2. Com os arquivos .tif com a média de cada mês global, recortou-se, em *software* de Sistema de Informações Geográficas (SIG) – QGIS -, apenas para os limites do Brasil, reprojetoado para Sistema de Referência de Coordenadas, com parâmetros repassados pelo IBGE e definido pelo autor: “Projeção de Albers – IBGE”.
3. Com os arquivos .tif com a média de cada mês Brasil, calculou-se, em *software* de Sistema de Informações Geográficas (SIG) - QGIS -, a média anual para o Brasil.
4. Com o arquivo .tif com a média anual para o Brasil, calculou-se, em *software* de Sistema de Informações Geográficas (SIG) - QGIS -, a média dos valores dos pixels dentro dos limites do polígono de cada município Brasileiro. Assim o indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

### Precipitação média de 1970 a 2020

1. Aquisição das camadas rasters, em formato .tif, com 1km de resolução, de dados climáticos de precipitação média (mm) mensal de 1970 a 2000 de todo o planeta, Sistema de Referência de Coordenadas, WGS 84. Coletado no segundo semestre de 2022 no site da WorldClim versão 2.1 lançado em janeiro de 2020. (arquivos para cada mês do ano: 'wc2.1\_30s\_prec\_xx').  
<https://www.worldclim.org/data/worldclim21.html>
2. Com os arquivos .tif com a média de cada mês global, recortou-se, em *software* de Sistema de Informações Geográficas (SIG) - QGIS -, apenas para os limites do Brasil, reprojetoado para Sistema de Referência de Coordenadas, com parâmetros repassados pelo IBGE e definido pelo autor: “Projeção de Albers – IBGE”.

3. Com os arquivos .tif com a média de cada mês Brasil, calculou-se, em *software* de Sistema de Informações Geográficas – SIG (QGIS), a média anual para o Brasil.
4. Com o arquivo .tif com a média anual para o Brasil, calculou-se, em *software* de Sistema de Informações Geográficas (SIG) - QGIS -, a média dos valores dos pixels dentro dos limites do polígono de cada município Brasileiro. Assim o indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

### Altitude

1. Aquisição da malha de pontos, em formato shapefile (.shp), que representa a localização das Localidades do Brasil. Coletado no segundo semestre de 2022 no site do Instituto Brasileiro de Geografia e Estatística (IBGE). (arquivo: ‘BR\_Localidades\_2010\_v1’).  
<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/27385-localidades.html>
2. Os dados são referentes às coordenadas em Graus Decimais, referidas ao datum SIRGAS 2000, e com altitude proveniente dos dados do SRTM (*Shuttle Radar Topography Mission*) com resolução de 90 metros (versão 4.1).
3. As localidades utilizadas para o trabalho correspondem as ‘cidades’, que é a localidade onde está sediada a Prefeitura Municipal. É constituída pela área urbana do distrito sede e delimitada pelo perímetro urbano estabelecido por lei municipal. Ou seja, os pontos da sede dos 5.570 municípios brasileiros.
4. Após essa filtragem por meio de *software* de Sistema de Informações Geográficas (SIG) - QGIS - salvou-se para cada um dos 5.570 municípios brasileiros a altitude em arquivo .csv.

### Percentual da área desmatada de 2000 para 2010

1. Aquisição das camadas rasters, em formato .tif, com 30m de resolução, de dados de uso e cobertura da terra anual 2000 e 2010 do Brasil, Sistema de Referência de Coordenadas, WGS 84. Coletado no segundo semestre de 2022 no site do MapBiomas coleção 7. (arquivos para cada ano: ‘brasil\_coverage\_20xx’).

[https://mapbiomas.org/colecoes-mapbiomas-1?cama\\_set\\_language=pt-BR](https://mapbiomas.org/colecoes-mapbiomas-1?cama_set_language=pt-BR)

2. Reclassificação das camadas rasters, com várias categorias de uso e cobertura da terra, para apenas duas: Formação Florestal/Natural (1) e Antrópico (2). Reclassificadas por meio *software* de Sistema de Informações Geográficas (SIG) – QGIS -, considerando os seguintes códigos das categorias (MapBiomias) para os dois novos códigos (Autor):

1 3 4 5 49 10 11 12 32 29 50 13 23 26 33 = 1

14 15 18 19 39 20 40 62 41 36 46 47 48 9 21 22 24 30 25 31 27 = 2

<https://mapbiomas-br->

[site.s3.amazonaws.com/downloads/\\_EN\\_C%C3%B3digos\\_da\\_legenda\\_Cole%C3%A7%C3%A3o\\_7.pdf](https://mapbiomas-br-site.s3.amazonaws.com/downloads/_EN_C%C3%B3digos_da_legenda_Cole%C3%A7%C3%A3o_7.pdf)

3. Elaboração do mapa de diferenças - o que era Formação Florestal/Natural em 2000 e se tornou Antrópico em 2010 - em formato .tif, por meio do *software* de modelagem ambiental Dinamica EGO.
4. Com o mapa de áreas antropizadas de 2000 para 2010 (raster) em mãos, reprojeteu-se para Sistema de Referência de Coordenadas, com parâmetros repassados pelo IBGE e definido pelo autor: “Projeção de Albers – IBGE”. E converteu para o formato vetorial .gpkg por meio do *software* de Sistema de Informações Geográficas (SIG) - QGIS.
5. Com a camada vetorial de áreas antropizadas, mais a camada vetorial de municípios brasileiros, calculou-se por meio de análise de sobreposição o percentual da área de cada município coberta por áreas antropizadas, com o auxílio do *software* de Sistema de Informações Geográficas (SIG) - QGIS. Assim o indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Percentual de área com urbanização 2010

1. Aquisição da camada raster, em formato .tif, com 30m de resolução, de dados de uso e cobertura da terra 2010 do Brasil, Sistema de Referência de Coordenadas, WGS 84. Coletado no segundo semestre de 2022 no site do MapBiomias coleção 7. (arquivo: ‘brasil\_coverage\_2010’).

[https://mapbiomas.org/colecoes-mapbiomas-1?cama\\_set\\_language=pt-BR](https://mapbiomas.org/colecoes-mapbiomas-1?cama_set_language=pt-BR)

2. Com o mapa de uso e cobertura da terra 2010 (raster) em mãos, reprojeteu-se para Sistema de Referência de Coordenadas, com parâmetros repassados pelo IBGE e definido pelo autor: “Projeção de Albers – IBGE”. E converteu ‘Área Urbanizada (24)’ para o formato vetorial .gpkg, por meio do *software* de Sistema de Informações Geográficas (SIG) - QGIS.

[https://mapbiomas-br-site.s3.amazonaws.com/downloads/\\_EN\\_C%C3%B3digos\\_da\\_legenda\\_Cole%C3%A7%C3%A3o\\_7.pdf](https://mapbiomas-br-site.s3.amazonaws.com/downloads/_EN_C%C3%B3digos_da_legenda_Cole%C3%A7%C3%A3o_7.pdf)

3. Com a camada vetorial de áreas urbanizadas, mais a camada vetorial de municípios brasileiros, calculou-se por meio de análise de sobreposição o percentual da área de cada município coberta por áreas urbanizadas, com o auxílio do *software* de Sistema de Informações Geográficas (SIG) - QGIS. Assim o indicador, foi salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

#### Número de criadouros do Anopheles 2001 a 2021

1. Aquisição dos dados, em formato .csv, com todos os criadouros cadastrados no “Vetores-Malária” desde sua implementação (2001 a 2021). Coletado no segundo semestre de 2021 via site da Plataforma Integrada de Ouvidoria e Acesso à Informação.  
<https://falabr.cgu.gov.br/publico/Manifestacao/SelecionarTipoManifestacao.aspx?ReturnUrl=%2f>
2. Os dados de criadouros foram processados, e apresentaram um total de 12.092 registros 2001 a 2021. Salvo e consolidado para cada um dos 5.570 municípios em arquivo .csv.

\* Balneário Rincão – SC, Mojuí dos Campos – PA, Paraíso das Águas – MS, Pescaria Brava – SC e Pinto Bandeira - RS, são municípios que só foram completamente emancipados após 2010, então para as variáveis que esses não possuíam informação, foi considerado a média dos municípios brasileiros.

## ANEXO B - Metadados e processamento das variáveis doenças

Material Suplementar (2) para

**Título: Doenças negligenciadas no Brasil: Mapeamento de uma década de ocorrências**

***Texto Suplementar:***

### Doença de Chagas Aguda 2010 a 2020

1. Aquisição dos dados acumulados 2010 a 2020, em formato .csv, do número de casos confirmados notificados no Sistema de Informação de Agravos de Notificação, tendo como ano de início de sintomas (casos novos) entre 2010 e 2020. Coletado no segundo semestre de 2022 no site do Ministério da Saúde.  
<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinannet/cnv/chagasbr.def>
2. Os casos notificados foram por municípios de infecção, os registros que não tiveram o campo município preenchido (município ignorado), foram excluídos da base de dados.
3. Os dados foram consolidados para cada um dos 5.570 municípios brasileiros e salvos em arquivo .csv.
4. Os dados consolidados também foram salvos em arquivos .shp, uma versão em formato de polígono - representando o perímetro dos municípios - e o número de casos vinculado à cada município. E outra versão em formato de ponto - representando a sede administrativa de cada município - e o número de casos vinculado à cada município.

### Hanseníase 2010 a 2020

1. Aquisição dos dados acumulados 2010 a 2020, em formato .csv, do número de casos confirmados notificados no Sistema de Informação de Agravos de Notificação, tendo como ano de diagnóstico (casos novos) entre 2010 e 2020. Coletado no segundo semestre de 2022 no site do Ministério da Saúde.  
<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/hanswbr.def>
2. Os casos notificados foram por municípios de residência, os registros que não tiveram o campo município preenchido (município ignorado), foram excluídos da base de dados.

3. Os dados foram consolidados para cada um dos 5.570 municípios brasileiros e salvos em arquivo .csv.
4. Os dados consolidados também foram salvos em arquivos .shp, uma versão em formato de polígono - representando o perímetro dos municípios - e o número de casos vinculado à cada município. E outra versão em formato de ponto - representando a sede administrativa de cada município - e o número de casos vinculado à cada município.

#### Leishmaniose Tegumentar Americana 2010 a 2020

1. Aquisição dos dados acumulados 2010 a 2020, em formato .csv, do número de casos confirmados notificados no Sistema de Informação de Agravos de Notificação, tendo como ano de notificação entre 2010 e 2020. Coletado no segundo semestre de 2022 no site do Ministério da Saúde.  
<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinannet/cnv/ltabr.def>
2. Os casos notificados foram por municípios de residência, os registros que não tiveram o campo município preenchido (município ignorado), foram excluídos da base de dados.
3. Os dados foram consolidados para cada um dos 5.570 municípios brasileiros e salvos em arquivo .csv.
4. Os dados consolidados também foram salvos em arquivos .shp, uma versão em formato de polígono - representando o perímetro dos municípios - e o número de casos vinculado à cada município. E outra versão em formato de ponto - representando a sede administrativa de cada município - e o número de casos vinculado à cada município.

#### Leishmaniose Visceral 2010 a 2020

1. Aquisição dos dados acumulados 2010 a 2020, em formato .csv, do número de casos confirmados notificados no Sistema de Informação de Agravos de Notificação, tendo como ano de notificação entre 2010 e 2020. Coletado no segundo semestre de 2022 no site do Ministério da Saúde.  
<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinannet/cnv/leishvbr.def>

2. Os casos notificados foram por municípios de residência, os registros que não tiveram o campo município preenchido (município ignorado), foram excluídos da base de dados.
3. Os dados foram consolidados para cada um dos 5.570 municípios brasileiros e salvos em arquivo .csv.
4. Os dados consolidados também foram salvos em arquivos .shp, uma versão em formato de polígono - representando o perímetro dos municípios - e o número de casos vinculado à cada município. E outra versão em formato de ponto - representando a sede administrativa de cada município - e o número de casos vinculado à cada município.

#### Malária 2010 a 2020

1. Aquisição dos dados acumulados 2010 a 2020, em formato .csv, do número de casos confirmados notificados no Sistema de Informação de Agravos de Notificação (unidades federativas extra-amazônica) e Sivep-Malária (unidades federativas que compõem a região Amazônica), tendo como ano de notificação entre 2010 e 2020. Coletado no segundo semestre de 2022 no site do Ministério da Saúde.  
[https://public.tableau.com/app/profile/mal.ria.brasil/viz/Dadosparacidado\\_201925\\_03\\_2020/Inicio](https://public.tableau.com/app/profile/mal.ria.brasil/viz/Dadosparacidado_201925_03_2020/Inicio)
2. Os casos notificados foram por municípios de provável infecção, os registros que não tiveram o campo município preenchido (município ignorado) e importados de outros países foram excluídos da base de dados.
3. Os dados foram consolidados para cada um dos 5.570 municípios brasileiros e salvos em arquivo .csv.
4. Os dados consolidados também foram salvos em arquivos .shp, uma versão em formato de polígono - representando o perímetro dos municípios - e o número de casos vinculado à cada município. E outra versão em formato de ponto - representando a sede administrativa de cada município - e o número de casos vinculado à cada município.

#### Tuberculose 2010 a 2020



1. Aquisição dos dados acumulados 2010 a 2020, em formato .csv, do número de casos confirmados notificados no Sistema de Informação de Agravos de Notificação, tendo como ano de notificação entre 2010 e 2020. Coletado no segundo semestre de 2022 no site do Ministério da Saúde.  
<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/tubercbr.def>
2. Os casos notificados foram por municípios de residência e as seguintes categorias da variável “Tipo de entrada”: caso novo, não sabe e pós óbito. Os registros que não tiveram o campo município preenchido (município ignorado), foram excluídos da base de dados.
3. Os dados foram consolidados para cada um dos 5.570 municípios brasileiros e salvos em arquivo .csv.
4. Os dados consolidados também foram salvos em arquivos .shp, uma versão em formato de polígono - representando o perímetro dos municípios - e o número de casos vinculado à cada município. E outra versão em formato de ponto - representando a sede administrativa de cada município - e o número de casos vinculado à cada município.

#### Dengue 2010 a 2020

1. Aquisição dos dados acumulados 2010 a 2020, em formato .csv, do número de casos prováveis notificados no Sistema de Informação de Agravos de Notificação, tendo como ano de notificação entre 2010 e 2020. Coletado no segundo semestre de 2022 no site do Ministério da Saúde.  
<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinannet/cnv/denguebr.def>
2. Os casos notificados foram por municípios de residência, os registros que não tiveram o campo município preenchido (município ignorado), foram excluídos da base de dados.
3. Os dados foram consolidados para cada um dos 5.570 municípios brasileiros e salvos em arquivo .csv.

4. Os dados consolidados também foram salvos em arquivos .shp, uma versão em formato de polígono - representando o perímetro dos municípios - e o número de casos vinculado à cada município. E outra versão em formato de ponto - representando a sede administrativa de cada município - e o número de casos vinculado à cada município.

#### Esquistossomose 2010 a 2020

1. Aquisição dos dados acumulados 2010 a 2020, em formato .csv, do número de casos confirmados notificados no Sistema de Informação de Agravos de Notificação (dados de vigilância passiva em áreas não endêmicas) e Programa de Controle da Esquistossomose (dados de vigilância ativa em áreas endêmicas), tendo como ano de notificação entre 2010 e 2020. Coletado no segundo semestre de 2022 no site do Ministério da Saúde.  
<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinannet/cnv/esquistobr.def>  
<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinan/pce/cnv/pcebr.def>
2. Os casos notificados foram por municípios de residência, os registros que não tiveram o campo município preenchido (município ignorado), foram excluídos da base de dados.
3. Os dados foram consolidados para cada um dos 5.570 municípios brasileiros e salvos em arquivo .csv.
4. Os dados consolidados também foram salvos em arquivos .shp, uma versão em formato de polígono - representando o perímetro dos municípios - e o número de casos vinculado à cada município. E outra versão em formato de ponto - representando a sede administrativa de cada município - e o número de casos vinculado à cada município.

## ANEXO C - Etapas das Análises e modelagens

Material Suplementar (3) para

**Título: Doenças negligenciadas no Brasil: Mapeamento de uma década de ocorrências**

***Texto Suplementar:***

### Estimativa de densidade Kernel

4. A estimativa de densidade kernel, também conhecida como mapa de calor, é uma modelagem espacial, que utiliza a interpolação, uma técnica matemática para estimar valores intermediários entre pontos conhecidos de um conjunto de dados, com objetivo de melhor visualizar os dados.
5. Para essa modelagem, utilizou-se o *software* de Sistema de Informações Geográficas (SIG) - QGIS, por meio de sua ferramenta nativa de interpolar ‘Mapa de calor (Estimativa de densidade Kernel)’.
6. Para isso foi necessário organizar os casos de cada tipo de doença negligenciada em arquivo vetorial do tipo ponto (.shp), no sistema de coordenadas planas. Como os dados são referentes a todo o Brasil, o sistema de referência de coordenadas mais adequado, portanto, o adotado, foi o: EPSG: 5880 – SIRGAS 2000 / Brazil Polyconic.
7. A densidade é comumente calculada com base no número de pontos em um lugar. Grandes números de pontos agrupados resultam em maiores valores. Na modelagem do trabalho, ao invés de considerar o número de pontos que representaria os registros de doenças, utilizou-se o peso da coluna ‘número de registros’ de cada município (representado como um ponto na sede administrativa municipal).
8. Outros parâmetros utilizados para a estimativa, foi um raio de influência do ponto que representa a sede do município, de 100.000 metros, pois mostrou-se mais adequado para área de estudo. E para a resolução espacial do modelo de saída (arquivo raster .tif), considerou-se 1.000 metros.

### Clusterings K-means

1. O algoritmo K-means, também conhecido como algoritmo hierárquico, ajusta a classificação das observações em clusters e atualiza os centroides do cluster até que a posição deles fique estável ao longo de iterações sucessivas. Nesta implementação do algoritmo, a estabilidade dos centroides é determinada comparando o valor absoluto da mudança na distância euclidiana média entre as observações e seus centroides correspondentes contra um limiar.

2. Para essa modelagem, utilizou-se o *software* de Sistema de Informações Geográficas (SIG) - QGIS, por meio de uma ferramenta complementar '*Attribute based clustering*'.
3. Para isso foi necessário organizar os casos de cada tipo de doença negligenciada em arquivo vetorial do tipo polígono (.shp), no sistema de referência de coordenadas EPSG: 4674 – SIRGAS 2000.
4. Os agrupamentos são comumente calculados com base na distância dos centroides das feições em um lugar. Grupos de centroides próximos e estáveis formam um *cluster*. Na modelagem do trabalho, ao invés de considerar apenas a distância dos centroides das feições representadas pelo perímetro dos municípios, considerou-se também o peso da coluna 'número de registros' de cada município.
5. O *plugin* '*Attribute based clustering*' então gerou grupos que melhor se ajustava a realidade espacial/alfanuméricos dos dados de cada tipo de doença negligenciada no Brasil.

#### Modelo linear generalizado

1. As variáveis independentes testadas (variáveis socioeconômicas e ambientais) compuseram um arquivo .csv, no qual continha os registros de cada tipo de doença negligenciada estudada (variáveis dependentes) para todos os municípios brasileiros.
2. Com os dados organizados, deu-se início as análises e modelagens. Para isso utilizou-se a versão 4.2.2 da linguagem R, por meio do *software* de Ambiente de Desenvolvimento Integrado 'RStudio'.
3. Para o arquivo, foi rodado o seguinte *script*:
  - a) Calcular quantos municípios com zero registro para cada tipo de doença negligenciada na década de pesquisa.
  - b) Calcular quantos municípios com pelo menos um registro para cada tipo de doença negligenciada na década de pesquisa.
  - c) Testar para cada tipo de doença negligenciada um GLM Family Poisson.
  - d) Rodar a previsão de quantos municípios com zero registro com base no GLM Family Poisson.
  - e) Calcular a média e a variância de cada tipo de doença negligenciada para justificar a utilização da modelagem hurdle.
  - f) Rodar a previsão de quantos municípios com zero registro com base no modelo hurdle.
  - g) Rodar teste gráfico para ver o quanto o modelo hurdle prevê os nossos dados, testar até 150 observações.

- h) Interpretar primeira parte da modelagem: Probabilidade de um município ter ou não determinada doença negligenciada, a partir de determinadas variáveis socioeconômicas e ambientais.
- i) Interpretar segunda parte da modelagem: Probabilidade da quantidade de casos de determinada doença negligenciada apenas nos municípios com registros dessa doença, assim observar quais variáveis socioeconômicas e ambientais tem efeito negativo ou positivo na quantidade de casos.
4. Assim organizou-se os resultados de forma tabular (ver Tabela 5 do Artigo). A fórmula da modelagem hurdle pode ser escrita da seguinte maneira:

$$P(Y=0) = \varphi$$

$$P(Y=y) = (1-\varphi) * (e^{-\lambda} * \lambda^y)/y!$$

$$\mu = (1-\varphi) * \lambda$$

onde:

Y é a variável de resposta contagem

$\varphi$  é a probabilidade de a contagem ser igual a zero

$\lambda$  é o parâmetro de taxa da distribuição Poisson na segunda parte do modelo

$\mu$  é a média da distribuição Poisson na segunda parte do modelo.

5. Essa formulação assume que a probabilidade de a contagem ser zero é diferente da probabilidade de ser maior do que zero, e que a distribuição Poisson é adequada para modelar a contagem de valores não zero.
6. Além do GLM hurdle indicar as variáveis preditoras estatisticamente significativas e seu efeito negativo e positivo em algumas doenças negligenciadas. Juntamente com o GLM poisson (para malária e esquistossomose), gerou-se predições por município e pixel a pixel de 1000 metros de resolução espacial (SRC: “Projeção de Albers – IBGE”).
7. Para o mapeamento espacialmente explícito, pixel a pixel, elaborou-se validação conforme exposto no capítulo “Métodos” do artigo no subcapítulo “Processamento de dados”, para cada doença negligenciada estudada.

### Script em linguagem R

```
##### GLM #####
##### LEGENDA DAS VARIÁVEIS #####
#doenca_chagas_sinan_2010a2020_      = Casos/registros Doença de Chagas 2010 a 2020
#doenca_hanseníase_sinan_2010a2020_  = Casos/registros Hanseníase 2010 a 2020
#doenca_ltegumentar_sinan_2010a2020_ = Casos/registros Leishmaniose tegumentar 2010 a 2020
#doenca_lvisceral_sinan_2010a2020_   = Casos/registros Leishmaniose visceral 2010 a 2020
```

```

#doenca_malaria_sivep_sinan_2010a2020_ = Casos/registros Malária 2010 a 2020
#doenca_tuberculose_sinan_2010a2020_ = Casos/registros Tuberculose 2010 a 2020
#doenca_dengue_sinan_2010a2020_ = Casos/registros Dengue 2010 a 2020
#doenca_esquis_sinan_pce_2010a2020_ = Casos/registros Esquistossomose 2010 a 2020
#altitude_mun = Altitude
#criadouros_anophe = Número de criadouros do Anopheles 2001 a 2021
#doses_vacina_2010 = Número de doses de vacinas aplicadas 2010
#atencao_complexa_internação_2019 = Número de estabelecimentos com suporte de observação e
#atencao_primaria = Número de estabelecimentos de saúde da atenção primária 2019
#dens_dem_2010 = Número de habitantes por km² 2010
#muni_aglomerados_pc_area = Percentual da área com aglomerados subnormais 2010
#desf_2000to2010_pc_area = Percentual da área desmatada de 2000 para 2010
#taxa_alfabe = Percentual da população alfabetizada 2010
#urb_pc_2010 = Percentual de área com urbanização 2010
#cobertura_ab = Percentual de cobertura populacional na atenção básica 2010
#cobertura_acs_comunitários_de_saúde_2010 = Percentual de cobertura populacional por agentes
#agua_rede_geral_2010 = Percentual de domicílios abastecidos por rede geral de água
#pc_vala_2010 = Percentual de domicílios com esgotamento sanitário para vala
#lixo_coletado = Percentual de domicílios com lixo coletado 2010
#renda_salário_mínimo_2010 = Percentual de domicílios com renda mensal de até 0,25
#moradores_mais = Percentual de domicílios com superlotação 2010
#dom_rural = Percentual de domicílios rurais em 2010
#sem_banheiro = Percentual de domicílios sem banheiro 2010
#nao_tinham_energia = Percentual de domicílios sem energia elétrica 2010
#precip_media = Precipitação média de 1970 a 2020
#saldo_pop = Saldo do número de habitantes de 2000 para 2010
#temp_media = Temperatura média de 1970 a 2000
#####
install.packages("tidyverse")
# ----- tidyverse includes ggplot2 --- dplyr ---- tidyr--- and others
install.packages("readxl")
install.packages("writexl")
install.packages("ggbeeswarm")
install.packages("beeswarm")
install.packages("ggthemes")
install.packages("scales")

```

```
install.packages("viridis")
install.packages("viridisLite")
install.packages("ggpubr")
install.packages("fitdistrplus")
install.packages("MASS")
install.packages("pscl")
install.packages("countreg", repos="http://R-Forge.R-project.org")
install.packages("rgdal")
install.packages("raster")
library(tidyverse)
library(readxl)
library(writexl)
library(ggbeeswarm)
library(beeswarm)
library(ggthemes)
library(scales)
library(viridis)
library(viridisLite)
library(ggpubr)
library(fitdistrplus)
library(MASS)
library(pscl)
library(countreg)
library(rgdal)
library(raster)
#importar a tabela com os dados
base_dados <- read.csv('base_dados.csv', sep = ";")
#visualizar a distribuição da doença_x
plotdist(base_dados$doenca_x, histo=TRUE, demp=TRUE)
#calcular quantos zeros têm para a doença_x
sum(base_dados$doenca_x < 1)
#calcular quantos maior que zero para a doença_x
sum(base_dados$doenca_x > 0)
#caso tenha muitos muitos zeros, não pode ser poisson, tem de ser zeroinflated models
#testar glm com family poisson
doenca_x_glm_poisson<- glm(doenca_x ~ altitude_mun + criadouros_anophe + doses_vacina_2010
                           + atencao_complexa + atencao_primaria + dens_dem_2010 + moradores_mais
                           + muni_aglomerados_pc_area + desf_2000to2010_pc_area + taxa_alfabe
```

```

+ urb_pc_2010 + cobertura_ab + cobertura_acs + agua_rede_geral
+ pc_vala

+ lixo_coletado + renda + dom_rural + sem_banheiro

+ nao_tinham_energia + precip_media + saldo_pop + temp_media, data =
base_dados, family = "poisson")

doenca_x_glm_poisson

#prever as contagens da média
objeto1 <- predict(doenca_x_glm_poisson, type = "response")

objeto1

#soma das probabilidades de quantidade de zeros para cada média
exp <- sum(dpois(x = 0, lambda = objeto1))

#prever o número de zeros com base no glm com family poisson
round(exp)

#observar quantos zeros o modelo previu

#calcular o número realmente observado
sum(base_dados$doenca_x < 1)

#comparar os zeros da base de dados com o que o modelo glm com family poisson previu

#aqui é onde entra o modelo de hurdle (obstáculo) para uma base de dados com muitos zeros

#outro motivo para não usar glm com family poisson é quando a variância é muito maior que a média
mean.doenca_x = mean(base_dados$doenca_x)

mean.doenca_x

variance_doenca_x = var(base_dados$doenca_x)

variance_doenca_x

hurdle_doenca_x <- hurdle(doenca_x ~ altitude_mun + criadouros_anophe + doses_vacina_2010
+ atencao_complexa + atencao_primaria + dens_dem_2010 + moradores_mais
+ muni_aglomerados_pc_area + desf_2000to2010_pc_area + taxa_alfabe
+ urb_pc_2010 + cobertura_ab + cobertura_acs + agua_rede_geral +
pc_vala
+ lixo_coletado + renda + dom_rural + sem_banheiro
+ nao_tinham_energia + precip_media + saldo_pop + temp_media, data =
base_dados, dist = "negbin")

summary(hurdle_doenca_x)

#prever as contagens da média
objeto2 <- predict(hurdle_doenca_x, type = "response")

objeto2

write.csv(objeto2, "previsao_media_hurdle_municipios_doenca_x.csv", row.names = FALSE)

#ver o número de zeros previsto pelo modelo hurdle
sum(predict(hurdle_doenca_x, type = "prob")[,1])

#para ver o quanto o modelo hurdle prevê os dados

#testar até 150 observações

```



```

#primeiro ver os valores para doenca_x da nossa base de dados
plot(table(base_dados$doenca_x))

#ver o que o modelo prevê
chagas_fit<- rootogram(hurdle_doenca_x, max = 150)

#ver se a linha vermelha acompanha os dados

#para assim aceitar o modelo

# ter salvo em arquivos RASTER cada variável preditora em uma pasta
rasters<-list.files(path="C:/preditoras", pattern='tif$', full.names=TRUE)
image<-stack(rasters, quick=TRUE)
pred<- predict(object=image,hurdle_doenca_x, type="response")
writeRaster(pred, "C:/output_raster/hurdle_doenca_x.tif" ,"GTiff")

#validação: correlação raster kernel/GLM

# 1) Abrir arquivo raster de "densidade kernel"
densidade_raster_doenca_x <-raster("C:/doenca_x_kernel.tif")

# 2) Retirar pontos amostrais do raster de "densidade kernel" com os maiores valores
densidade_points_doenca_x <- rasterToPoints(densidade_raster_doenca_x)
densidade_points_sorted_doenca_x <- densidade_points_doenca_x[order(densidade_points_doenca_x[,3],
decreasing = TRUE),]

amostra_doenca_x <- densidade_points_sorted_doenca_x[1:1000,]

# 3) Salvar amostra em um arquivo
write.csv(amostra_doenca_x, "C:/amostra_doenca_x_kernel.csv")

# 4) Abrir arquivo raster de "GLM"
glm_raster_doenca_x <- raster("C:/output_raster/hurdle_doenca_x.tif")

# 5) Retirar pontos amostrais do raster de "GLM" com base na mesma localização dos pontos amostrais do
raster de "densidade kernel"
amostra_doenca_x_df <- data.frame(amostra_doenca_x)
glm_points_doenca_x <- extract(glm_raster_doenca_x, amostra_doenca_x_df[,1:2])

# 6) Salvar amostra em um arquivo
write.csv(glm_points_doenca_x, "C:/amostra_doenca_x_glm.csv ")

# 7) Fazer teste correlação entre esses dois arquivos com pontos amostrais
correlation_doenca_x <- cor(amostra_doenca_x[,3], glm_points_doenca_x)

# 8) Ver resultado
print(correlation_doenca_x)

##### FIM #####

```

### Efeito de variáveis independentes sobre os municípios com casos de Doenças Negligenciadas

*Doença de Chagas*

```

Zero hurdle model coefficients (binomial with logit link):
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.370e+01  2.241e+00 -6.110 9.95e-10 ***
altitude_mun -3.439e-04  5.291e-04 -0.650 0.515651
criadouros_anophe -5.087e-04  1.022e-03 -0.498 0.618654
doses_vacina_2010 -2.123e-06      NaN      NaN      NaN
atencao_complexa -9.754e-03      NaN      NaN      NaN
atencao_primaria  5.274e-02  7.224e-03  7.300 2.88e-13 ***
dens_dem_2010 -1.906e-03  5.012e-04 -3.803 0.000143 ***
moradores_mais -1.203e-01  1.750e-01 -0.687 0.491971
muni_aglomerados_pc_area  8.500e-02  6.214e-02  1.368 0.171310
desf_2000to2010_pc_area  2.639e-02  1.751e-02  1.507 0.131790
taxa_alfabe -3.984e-02  1.934e-02 -2.060 0.039432 *
urb_pc_2010  7.397e-02  4.115e-02  1.798 0.072249 .
cobertura_ab -1.777e-02  4.224e-03 -4.207 2.58e-05 ***
cobertura_acs  6.304e-03  2.714e-03  2.323 0.020193 *
agua_rede_geral -2.290e-03  5.710e-03 -0.401 0.688460
pc_vala  4.709e-02  1.588e-02  2.966 0.003017 **
lixo_coletado  1.783e-02  9.700e-03  1.838 0.066040 .
renda  7.261e-03  5.018e-02  0.145 0.884959
dom_rural -9.727e-03  8.764e-03 -1.110 0.267024
sem_banheiro -1.915e-02  1.528e-02 -1.254 0.209856
nao_tinham_energia  6.840e-02  1.490e-02  4.590 4.43e-06 ***
precip_media  2.723e-02  3.242e-03  8.399 < 2e-16 ***
saldo_pop  1.757e-08      NaN      NaN      NaN
temp_media  3.724e-01  4.325e-02  8.610 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Theta: count = 0.0025

Number of iterations in BFGS optimization: 72

Log-likelihood: -825.5 on 49 Df

*Hanseníase*

```

Zero hurdle model coefficients (binomial with logit link):
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.794e+00  2.130e+00 -4.599 4.25e-06 ***
altitude_mun  1.340e-03  2.534e-04  5.286 1.25e-07 ***
criadouros_anophe  8.638e+00      NaN      NaN      NaN
doses_vacina_2010  5.830e-04      NaN      NaN      NaN
atencao_complexa  3.052e-01  1.131e-01  2.698 0.006972 **
atencao_primaria  1.076e-02  3.159e-02  0.341 0.733434
dens_dem_2010  5.428e-03  2.752e-03  1.972 0.048557 *
moradores_mais -1.778e-02  1.040e+00 -0.017 0.986364
muni_aglomerados_pc_area  9.492e+00      NaN      NaN      NaN
desf_2000to2010_pc_area -5.673e-02  1.645e-02 -3.449 0.000563 ***
taxa_alfabe -3.664e-02  1.705e-02 -2.148 0.031681 *
urb_pc_2010 -1.079e-01  4.534e-02 -2.379 0.017338 *
cobertura_ab  2.244e-03  3.225e-03  0.696 0.486520
cobertura_acs  3.186e-03  2.918e-03  1.092 0.274914
agua_rede_geral -6.977e-03  3.822e-03 -1.826 0.067899 .
pc_vala -1.328e-03  1.563e-02 -0.085 0.932280
lixo_coletado  7.951e-03  4.672e-03  1.702 0.088752 .
renda -1.343e-01  3.470e-02 -3.871 0.000108 ***
dom_rural -7.994e-04  4.612e-03 -0.173 0.862394
sem_banheiro  1.843e-02  1.624e-02  1.135 0.256525
nao_tinham_energia  7.677e-02  3.223e-02  2.382 0.017232 *
precip_media  1.219e-02  3.107e-03  3.924 8.70e-05 ***
saldo_pop -4.371e-05      NaN      NaN      NaN
temp_media  5.225e-01  4.119e-02  12.686 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Theta: count = 0.7831

Number of iterations in BFGS optimization: 42

Log-likelihood: -2.302e+04 on 49 Df

*Leishmaniose Tegumentar*

```

Zero hurdle model coefficients (binomial with logit link):
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.153e+00 1.124e+00 -4.584 4.56e-06 ***
altitude_mun 2.530e-03 1.804e-04 14.025 < 2e-16 ***
criadouros_anophe 2.144e-01 8.776e-02 2.442 0.0146 *
doses_vacina_2010 1.476e-04      NaN      NaN      NaN
atencao_complexa 4.110e-02 6.912e-02 0.595 0.5521
atencao_primaria 1.737e-02 1.083e-02 1.603 0.1088
dens_dem_2010 -1.037e-03 4.286e-04 -2.419 0.0155 *
moradores_mais -3.227e-01 2.686e-01 -1.201 0.2296
muni_aglomerados_pc_area -1.407e-01 2.080e-01 -0.676 0.4988
desf_2000to2010_pc_area -4.486e-02 9.500e-03 -4.722 2.34e-06 ***
taxa_alfabe -3.706e-02 8.703e-03 -4.258 2.07e-05 ***
urb_pc_2010 1.777e-02 2.825e-02 0.629 0.5294
cobertura_ab 1.104e-03 2.145e-03 0.515 0.6068
cobertura_acs 4.886e-03 2.044e-03 2.390 0.0168 *
agua_rede_geral -5.119e-03 2.596e-03 -1.972 0.0486 *
pc_vala -4.107e-03 9.725e-03 -0.422 0.6728
lixo_coletado -2.914e-02 3.572e-03 -8.159 3.39e-16 ***
renda -1.882e-01 1.858e-02 -10.129 < 2e-16 ***
dom_rural -2.502e-02 3.314e-03 -7.550 4.36e-14 ***
sem_banheiro 1.055e-04 6.909e-03 0.015 0.9878
nao_tinham_energia 4.650e-02 1.050e-02 4.430 9.44e-06 ***
precip_media 1.792e-02 1.541e-03 11.631 < 2e-16 ***
saldo_pop -2.556e-05      NaN      NaN      NaN
temp_media 3.915e-01 2.301e-02 17.012 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Theta: count = 0.1183
Number of iterations in BFGS optimization: 80
Log-likelihood: -1.773e+04 on 49 Df

```

#### *Leishmaniose Visceral*

```

Zero hurdle model coefficients (binomial with logit link):
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.279e+00 1.136e+00 -8.165 3.21e-16 ***
altitude_mun 2.047e-03 1.963e-04 10.424 < 2e-16 ***
criadouros_anophe -7.722e-03 3.592e-03 -2.150 0.031581 *
doses_vacina_2010 1.272e-05      NaN      NaN      NaN
atencao_complexa -5.315e-02 3.455e-02 -1.538 0.123989
atencao_primaria 7.397e-02 8.165e-03 9.059 < 2e-16 ***
dens_dem_2010 5.636e-06 3.298e-04 0.017 0.986368
moradores_mais -5.426e-01 1.657e-01 -3.275 0.001055 **
muni_aglomerados_pc_area 5.911e-02 9.196e-02 0.643 0.520344
desf_2000to2010_pc_area 1.706e-02 8.900e-03 1.916 0.055315 .
taxa_alfabe -1.107e-02 8.493e-03 -1.304 0.192311
urb_pc_2010 1.123e-02 2.407e-02 0.466 0.640948
cobertura_ab -8.110e-03 2.096e-03 -3.869 0.000109 ***
cobertura_acs 5.082e-03 2.197e-03 2.313 0.020718 *
agua_rede_geral 1.777e-02 2.730e-03 6.509 7.58e-11 ***
pc_vala -3.485e-02 9.265e-03 -3.762 0.000168 ***
lixo_coletado -2.812e-02 4.308e-03 -6.527 6.71e-11 ***
renda 6.707e-03 1.949e-02 0.344 0.730735
dom_rural -1.712e-02 3.810e-03 -4.494 6.98e-06 ***
sem_banheiro 3.999e-02 7.462e-03 5.358 8.40e-08 ***
nao_tinham_energia 4.897e-03 9.254e-03 0.529 0.596673
precip_media -1.181e-02 1.391e-03 -8.490 < 2e-16 ***
saldo_pop 3.499e-05 1.729e-05 2.024 0.042926 *
temp_media 4.962e-01 2.522e-02 19.673 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Theta: count = 0.0576
Number of iterations in BFGS optimization: 56
Log-likelihood: -1.006e+04 on 49 Df

```

*Tuberculose*

```

Zero hurdle model coefficients (binomial with logit link):
      Estimate Std. Error z value Pr(>|z|)
(Intercept)      8.653e+00  4.288e+00   2.018 0.043613 *
altitude_mun     -2.118e-03  5.526e-04  -3.834 0.000126 ***
criadouros_anophe  9.793e+00  6.549e+02   0.015 0.988070
doses_vacina_2010  1.608e-03  1.325e-04  12.136 < 2e-16 ***
atencao_complexa  2.665e-01  2.795e-01   0.953 0.340502
atencao_primaria  -3.032e-03  1.066e-01  -0.028 0.977315
dens_dem_2010    -2.487e-03  6.621e-03  -0.376 0.707224
moradores_mais   -2.796e-01  8.789e-01  -0.318 0.750398
muni_aglomerados_pc_area  5.064e-02  4.594e+02   0.000 0.999912
desf_2000to2010_pc_area  4.579e-02  3.404e-02   1.345 0.178597
taxa_alfabe      -7.740e-02  3.466e-02  -2.233 0.025528 *
urb_pc_2010      2.091e-01  3.653e-01   0.572 0.567106
cobertura_ab      5.852e-03  6.165e-03   0.949 0.342478
cobertura_acs    -5.059e-03  6.614e-03  -0.765 0.444341
agua_rede_geral  -7.117e-04  7.246e-03  -0.098 0.921763
pc_vala          7.910e-02  4.037e-02   1.959 0.050061 .
lixo_coletado     1.133e-02  8.498e-03   1.333 0.182438
renda            -1.498e-01  6.006e-02  -2.494 0.012631 *
dom_rural        4.532e-03  9.922e-03   0.457 0.647867
sem_banheiro     -5.283e-03  2.605e-02  -0.203 0.839271
nao_tinham_energia -4.090e-02  2.429e-02  -1.683 0.092285 .
precip_media     -9.784e-03  5.770e-03  -1.696 0.089922 .
saldo_pop        4.439e-04  1.477e-04   3.006 0.002651 ***
temp_media       1.633e-02  7.163e-02   0.228 0.819647
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Theta: count = 1.2442

Number of iterations in BFGS optimization: 37

Log-likelihood: -2.488e+04 on 49 Df

*Dengue*

```

Zero hurdle model coefficients (binomial with logit link):
      Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.395e+01    NA      NA      NA
altitude_mun     3.600e-03    NA      NA      NA
criadouros_anophe  7.537e+00    NA      NA      NA
doses_vacina_2010  4.953e-04    NA      NA      NA
atencao_complexa  -5.710e-02    NA      NA      NA
atencao_primaria  3.158e-02    NA      NA      NA
dens_dem_2010     3.618e-03    NA      NA      NA
moradores_mais    1.949e+01    NA      NA      NA
muni_aglomerados_pc_area  9.608e+00    NA      NA      NA
desf_2000to2010_pc_area -1.183e-01    NA      NA      NA
taxa_alfabe      -4.843e-02    NA      NA      NA
urb_pc_2010      1.154e-02    NA      NA      NA
cobertura_ab      3.479e-03    NA      NA      NA
cobertura_acs    -3.974e-04    NA      NA      NA
agua_rede_geral  -1.244e-02    NA      NA      NA
pc_vala          -2.954e-02    NA      NA      NA
lixo_coletado     1.717e-02    NA      NA      NA
renda            8.745e-03    NA      NA      NA
dom_rural        -1.074e-02    NA      NA      NA
sem_banheiro     -8.149e-02    NA      NA      NA
nao_tinham_energia -3.386e-02    NA      NA      NA
precip_media      1.380e-03    NA      NA      NA
saldo_pop        -3.952e-05    NA      NA      NA
temp_media       9.709e-01    NA      NA      NA

```

Theta: count = 0.5985

Number of iterations in BFGS optimization: 45

Log-likelihood: -3.951e+04 on 49 Df

## ANEXO D - Análise da modelagem regressiva linear simples

Material Suplementar (4) para

### **Título: Doenças negligenciadas no Brasil: Mapeamento de uma década de ocorrências**

#### ***Texto Suplementar:***

Em uma análise inicial exploratória, para nosso conjunto de dados utilizou-se modelagem regressiva linear simples.

Os resultados indicam associação positiva entre registros de casos de Dengue e Tuberculose, em municípios com maior oferta de infraestruturas de saúde, concentrada nos centros urbanos nacionais.

Foi identificada associação positiva entre Dengue e Tuberculose em municípios que registraram elevado o aumento populacional nomeadamente no caso de migrações.

A maior ocorrência de quantidade de casos doenças negligenciadas pode estar associada às variáveis: Altitude( estatística para cada doença) , Número de habitantes por km<sup>2</sup> 2010, área desmatada de 2000 para 2010, população alfabetizada 2010, cobertura populacional na atenção básica 2010, esgotamento sanitário para vala nos domicílios2010, lixo coletado nos domicílios2010, domicílios com renda mensal de até 0,25 salário mínimo 2010, domicílios rurais em 2010 e Precipitação média de 1970 a 2020. E por fim o efeito positivo na quantidade de casos de algumas doenças negligenciadas considerando as variáveis: Altitude, Número de doses de vacinas aplicadas 2010, Número de estabelecimentos de saúde da atenção primária 2019, domicílios abastecidos por rede geral de água 2010, domicílios sem banheiro 2010, domicílios sem energia elétrica 2010, Precipitação média de 1970 a 2020 e Temperatura média de 1970 a 2000.

As análises focaram em explorar as variáveis dependentes (cada tipo de doença negligenciada) com o objetivo de testar e medir a associação dessas variáveis com variáveis independentes (socioeconômicas e ambientais). Foi usada modelagem regressiva linear simples - posteriormente ao cálculo do coeficiente de correlação de Kendall ( $\tau$ ) -, que é um modelo matemático linear que permite analisar se o valor de uma variável aumenta ou diminui conforme o valor da outra variável aumenta ou diminui. Como resultado, os valores de  $\tau$  vão de -1 a +1. Quanto mais próximo dos extremos (-1 ou 1), maior é a força da correlação, enquanto valores próximos de 0 implicam em correlações mais fracas ou inexistentes.

Assim, concentrou-se a análise no coeficiente de determinação ( $r^2$ ), ou seja, o quanto cada modelo regressivo explica a porcentagem da variação de  $y$  (variável dependente, que pode ser explicada pela relação de  $x$  e  $y$  (os dados)).

A análise da associação entre cada tipo de doença negligenciada e variáveis socioeconômicas e ambientais, resultou na análise dos coeficientes de correlação e de determinação conforme as Tabelas Suplementares 1 e 2 a baixo. Enquanto o coeficiente de correlação ( $\tau$ ) é usado para identificar padrões nas variáveis, o coeficiente de determinação ( $r^2$ ) é usado para identificar a força de um modelo.

Entre os resultados obtidos na avaliação da associação entre doenças negligenciadas e variáveis socioeconômicas e ambientais (Tabela Suplementar 1), encontrou-se correlação linear positiva moderada ( $0,5 \leq \tau < 0,8$ ), para “Hanseníase/Número de doses de vacinas aplicadas 2010”, “Tuberculose/Número de estabelecimentos com suporte de observação e internação 2019”, “Tuberculose/Número de estabelecimentos de saúde da atenção primária 2019”, “Tuberculose/Número de doses de vacinas aplicadas 2010”, respectivamente com 0,51; 0,53 ; 0,59 e 0,73.

Para a correlação linear negativa, encontrou-se apenas coeficiente correlação baixo, ou seja,  $0 \leq \tau < -0,5$ . Com atenção para “Tuberculose/Percentual de cobertura populacional na atenção básica 2010” e “Dengue/Percentual de domicílios rurais em 2010”, respectivamente -0,31 e -0,34.

Quando é considerado o coeficiente de determinação ( $r^2$ ) (Tabela Suplementar 2), salienta-se os resultados entre 0,3 e 0,5:

- “Hanseníase/Número de estabelecimentos de saúde da atenção primária 2019”.
- “Hanseníase/Número de estabelecimentos com suporte de observação e internação 2019”.
- “Hanseníase/Saldo do número de habitantes de 2000 para 2010”.
- “Malária/Número de criadouros do Anopheles 2001 a 2021”.
- “Dengue/Número de estabelecimentos de saúde da atenção primária 2019”.
- “Dengue/Número de doses de vacinas aplicadas 2010”.
- “Dengue/Saldo do número de habitantes de 2000 para 2010”.

Todos os resultados indicando associação positiva.

Para os resultados de coeficiente de determinação ( $r^2$ ) maiores que 0,5; têm-se:

- “Dengue/Número de estabelecimentos com suporte de observação e internação 2019”.
- “Tuberculose/Número de estabelecimentos de saúde da atenção primária 2019”.
- “Tuberculose/Saldo do número de habitantes de 2000 para 2010”.
- “Tuberculose/Número de estabelecimentos com suporte de observação e internação 2019”.
- “Tuberculose/Número de doses de vacinas aplicadas 2010”.

Todos os resultados indicando associação positiva.

Importante destacar alguns resultados de coeficientes de determinação ( $r^2$ ) baixos (entre 0,15 e 0,30), mas que podem indicar um caminho frente à complexidade.

- “Dengue/ Percentual de área com urbanização 2010”.
- “Tuberculose/ Número de habitantes por km<sup>2</sup> 2010”.
- “Hanseníase/ Percentual da área com aglomerados subnormais 2010”.
- “Leishmaniose visceral/Dengue” e vice-versa.
- “Hanseníase/Dengue” e vice-versa.
- “Hanseníase/Tuberculose” e vice-versa.
- “Hanseníase/ Número de doses de vacinas aplicadas 2010”.
- “Leishmaniose visceral/Hanseníase” e vice-versa.
- “Dengue/Tuberculose” e vice-versa.

Todos os resultados indicando associação positivo.

Tabela Suplementar1 – Coeficiente de correlação ( $\tau$ ) das variáveis dependentes/independentes

VARIÁVEIS	Doença de Chagas	Hansení.	Leish. Tegume.	Leish. visceral	Malária	Tuberc.	Dengue	Esquisto.
Doença de Chagas	1,00	0,16	0,17	0,07	0,35	0,14	0,03	-0,04
Hanseníase	0,16	1,00	0,44	0,38	0,33	0,47	0,40	0,17
Leishmaniose tegumentar	0,17	0,44	1,00	0,25	0,39	0,29	0,31	0,22
Leishmaniose visceral	0,07	0,38	0,25	1,00	0,12	0,27	0,26	0,17
Malária	0,35	0,33	0,39	0,12	1,00	0,21	0,08	-0,05
Tuberculose	0,14	0,47	0,29	0,27	0,21	1,00	0,38	0,26
Dengue	0,03	0,40	0,31	0,26	0,08	0,38	1,00	0,27
Esquistossomose	-0,04	0,17	0,22	0,17	-0,05	0,26	0,27	1,00
Número de estabelecimentos de saúde da atenção primária 2019	0,11	0,41	0,24	0,27	0,15	0,59	0,34	0,24
Número de estabelecimentos com suporte de observação e internação 2019	0,09	0,43	0,25	0,22	0,16	0,53	0,38	0,16
Percentual de cobertura populacional na atenção básica 2010	NCL	0,08	0,08	0,07	NCL	NCL	0,05	0,04
Percentual de cobertura populacional por agentes comunitários de saúde 2010	NCL	-0,05	NCL	0,07	0,07	-0,25	-0,16	-0,06
Número de imunizações 2010	0,14	0,51	0,32	0,31	0,21	0,73	0,39	0,28
Percentual da população alfabetizada 2010	NCL	-0,12	-0,07	-0,28	-0,06	0,04	0,03	-0,08
Número de habitantes por km <sup>2</sup> 2010	-0,06	0,04	-0,07	NCL	-0,19	0,30	0,15	0,26
Percentual da área com aglomerados subnormais 2010	0,03	0,16	0,09	0,11	0,04	0,20	0,15	0,16
Saldo do número de habitantes de 2000 para 2010	0,14	0,34	0,21	0,22	0,18	0,44	0,28	0,15
Percentual de domicílios rurais em 2010	0,03	-0,15	-0,08	NCL	0,06	-0,28	-0,34	-0,10
Percentual de domicílios com superlotação 2010	0,32	0,09	0,15	NCL	0,31	0,07	-0,03	-0,08
Percentual de domicílios com lixo coletado 2010	-0,08	NCL	-0,07	-0,16	-0,15	0,18	0,21	0,06
Percentual de domicílios sem banheiro 2010	0,10	0,16	0,13	0,35	0,15	NCL	-0,08	0,03
Percentual de domicílios sem energia elétrica 2010	0,16	0,19	0,22	0,27	0,27	NCL	-0,08	NCL
Percentual de domicílios com esgotamento sanitário para vala 2010	0,16	0,04	0,03	0,04	0,18	0,07	-0,18	0,03
Percentual de domicílios abastecidos por rede geral de água 2010	-0,10	0,03	-0,05	NCL	-0,17	0,15	0,22	0,06
Percentual de domicílios com renda mensal de até 0,25 salário mínimo 2010	0,07	0,10	0,04	0,31	0,08	NCL	-0,10	0,06
Temperatura média de 1970 a 2000	0,17	0,39	0,24	0,40	0,31	0,13	0,17	NCL
Precipitação média de 1970 a 2020	0,17	0,03	0,15	-0,21	0,32	NCL	-0,15	-0,17
Altitude	-0,16	-0,20	-0,07	-0,15	-0,26	-0,20	NCL	NCL
Percentual da área desmatada de 2000 para 2010	NCL	0,08	0,06	0,17	0,12	NCL	-0,12	NCL
Percentual de área com urbanização 2010	-0,04	0,12	NCL	NCL	-0,11	0,35	0,25	0,17
Número de criadouros do Anopheles 2001 a 2021	0,34	0,16	0,20	0,05	0,35	0,09	NCL	-0,08

\*NCL: Não há correlação linear.



**Tabela Suplementar2 – Coeficiente de determinação (r<sup>2</sup>) das variáveis dependentes/independentes**

<b>VARIÁVEIS</b>	<b>Doença de Chagas</b>	<b>Hansení.</b>	<b>Leish. Tegume.</b>	<b>Leish. visceral</b>	<b>Malária</b>	<b>Tuberc.</b>	<b>Dengue</b>	<b>Esquistoso.</b>
<b>Doença de Chagas</b>	RMRL	0,02	0,00	0,01	0,03	0,01	RMRL	RMRL
<b>Hanseníase</b>	0,02	RMRL	0,07	0,28	0,01	0,26	0,20	0,00
<b>Leishmaniose tegumentar</b>	0,00	0,07	RMRL	0,01	0,11	0,02	0,02	RMRL
<b>Leishmaniose visceral</b>	0,01	0,28	0,01	RMRL	RMRL	0,04	0,19	0,00
<b>Malária</b>	0,03	0,01	0,11	RMRL	RMRL	0,01	0,00	RMRL
<b>Tuberculose</b>	0,01	0,26	0,02	0,04	0,01	RMRL	0,29	0,00
<b>Dengue</b>	RMRL	0,20	0,02	0,19	0,00	0,29	RMRL	0,01
<b>Esquistossomose</b>	RMRL	0,00	RMRL	0,00	RMRL	0,00	0,01	RMRL
<b>Número de estabelecimentos de saúde da atenção primária 2019</b>	0,01	0,33	0,05	0,11	0,01	0,60	0,37	0,02
<b>Número de estabelecimentos com suporte de observação e internação 2019</b>	0,00	0,37	0,02	0,12	0,00	0,78	0,52	0,01
<b>Percentual de cobertura populacional na atenção básica 2010</b>	0,01	0,02	0,01	0,00	0,01	0,01	0,01	RMRL
<b>Percentual de cobertura populacional por agentes comunitários de saúde 2010</b>	RMRL	0,01	RMRL	RMRL	RMRL	0,02	0,03	0,00
<b>Número de imunizações 2010</b>	0,01	0,27	0,02	0,09	0,00	0,87	0,37	0,01
<b>Percentual da população alfabetizada 2010</b>	RMRL	0,00	RMRL	RMRL	0,00	0,01	0,01	0,02
<b>Número de habitantes por km<sup>2</sup> 2010</b>	RMRL	0,10	RMRL	RMRL	RMRL	0,16	0,12	0,00
<b>Percentual da área com aglomerados subnormais 2010</b>	0,01	0,17	RMRL	0,04	RMRL	0,10	0,05	0,00
<b>Saldo do número de habitantes de 2000 para 2010</b>	0,01	0,34	0,06	0,12	0,01	0,65	0,42	0,01
<b>Percentual de domicílios rurais em 2010</b>	RMRL	0,03	RMRL	RMRL	RMRL	0,01	0,03	RMRL
<b>Percentual de domicílios com superlotação 2010</b>	0,01	RMRL	0,01	RMRL	0,04	RMRL	RMRL	RMRL
<b>Percentual de domicílios com lixo coletado 2010</b>	0,00	RMRL	0,00	RMRL	0,00	0,01	0,02	RMRL
<b>Percentual de domicílios sem banheiro 2010</b>	RMRL	0,00	0,00	0,00	0,00	RMRL	0,00	RMRL
<b>Percentual de domicílios sem energia elétrica 2010</b>	0,01	RMRL	0,03	RMRL	0,04	RMRL	0,00	RMRL
<b>Percentual de domicílios com esgotamento sanitário para vala 2010</b>	0,01	RMRL	0,01	RMRL	0,02	RMRL	0,00	0,00
<b>Percentual de domicílios abastecidos por rede geral de água 2010</b>	0,01	0,00	0,02	RMRL	0,01	0,01	0,02	RMRL
<b>Percentual de domicílios com renda mensal de até 0,25 salário mínimo 2010</b>	RMRL	0,00	RMRL	RMRL	RMRL	RMRL	0,01	RMRL
<b>Temperatura média de 1970 a 2000</b>	0,01	0,03	0,03	0,02	0,01	RMRL	RMRL	RMRL
<b>Precipitação média de 1970 a 2020</b>	0,02	0,01	0,05	RMRL	0,04	RMRL	RMRL	0,00
<b>Altitude</b>	0,01	0,02	0,02	0,01	0,01	0,00	RMRL	RMRL
<b>Percentual da área desmatada de 2000 para 2010</b>	RMRL	0,00	0,01	0,00	0,00	RMRL	0,00	RMRL
<b>Percentual de área com urbanização 2010</b>	RMRL	0,09	RMRL	RMRL	RMRL	0,12	0,16	0,00
<b>Número de criadouros do Anopheles 2001 a 2021</b>	0,00	0,00	0,15	RMRL	0,31	0,01	RMRL	RMRL

\*RMRL: Reprovado modelo de regressão linear.

O primeiro principal achado desse material suplementar (análise exploratória), foi a identificação da elevada associação positiva entre as doenças Dengue e Tuberculose, com variáveis benéficas para uma sociedade, como uma maior oferta de serviços e infraestruturas de saúde (Número de estabelecimentos com suporte de observação e internação 2019; Número de estabelecimentos de saúde da atenção primária 2019; Número de doses de vacinas aplicadas 2010), O que não faz sentido prático, mas que pode ser explicado em parte, pelo fato de tais doenças se concentrarem médias e grandes cidades onde os sistemas de controles e registros são maiores. Além de concentrarem a maior parte da população nacional, o que foi identificado na elevada associação positiva entre a Tuberculose e aumento populacional (Saldo do número de habitantes de 2000 para 2010).

Fato que deve ser investigado em próximos estudos. Há uma relação de interferência ou apenas uma associação entre Dengue e Tuberculose? Sendo apenas associação há um terceiro elemento causando as duas?

Além disso, importante destacar, mesmo que seja uma associação negativa baixa, observou-se alguma; entre população coberta pela atenção básica e por agentes comunitários de saúde com casos de Tuberculose, indicando que quanto maior é a cobertura da “atenção primária a saúde” menor são os casos de “tuberculose”.

Registra-se também, associação negativa baixa, entre número de domicílios rurais e casos de Tuberculose e Dengue. Como apontado em trabalho da Dengue no Brasil, sobre a relevância do socioambiente urbano, na medida que se torna espaço de maior concentração populacional e por consequência, maior incidência de doenças (MENDONÇA; SOUZA; DUTRA, 2009). Nosso trabalho aponta para importância e desafios da migração - aumento populacional - e urbanização na dinâmica de surgimento e ressurgimento de doenças, como em outros trabalhos da literatura (FOTSO et al., 2012; GILGEN et al., 2005; GUBLER, 2011; HARPHAM, 2009; MULLIGAN; ELLIOTT; SCHUSTER-WALLACE, 2012; TONG et al., 2015).

Importante salientar, mesmo que sejam associações positivas baixas, o presente material suplementar permite concluir que algumas doenças estudadas são associadas entre si. Tais achados, permitem auxiliar no planejamento e ações de combate às doenças de forma conjunta e otimizada, ou seja, direcionando recursos financeiros e humanos compatíveis com os tipos de doenças negligenciadas mais comumente observadas, além de educação em saúde para as comunidades afetadas.

As questões sobre mudanças climáticas também novamente devem ser consideradas nesse material suplementar, embora seja uma associação positiva baixa, há relação entre maior temperatura média com Hanseníase, Leishmaniose Tegumentar, Leishmaniose Visceral e

Malária; assim como associação positiva entre maior precipitação média e Malária (modelagem linear simples).

Etapas da modelagem linear simples:

1. Cada variável independente testada (variáveis socioeconômicas e ambientais) compôs um arquivo .csv, no qual continha os registros de cada tipo de doença negligenciada estudada (variáveis dependentes) para todos os municípios brasileiros.

2. Com os dados organizados, deu-se início as análises e modelagens de cada arquivo. Para isso utilizou-se a versão 4.2.2 da linguagem R, por meio do *software* de Ambiente de Desenvolvimento Integrado ‘RStudio’.

3. Para cada arquivo, foi rodado o seguinte *script*:

- a) Verificação de valores *missing*.
- b) Análise preliminar visual gráfica da correlação, ou seja, foram plotados gráficos de dispersão com cada tipo de doença negligenciada (y) e cada variável socioeconômica e ambiental testada (x).
- c) Análise da normalidade da variável x e y, para identificar qual o coeficiente de correlação adequado, ‘Pearson’ (distribuição normal) e ‘Kendall’ (distribuição diferente de normal). Para isso adotou-se os testes numéricos de normalidade ‘Anderson-Darling’, ‘Lilliefors’ e ‘Cramer-Von Mises’, onde H0 (hipótese nula) p-value é  $> 0,05$  que indica distribuição normal, e Ha (hipótese alternativa) p-value é  $\leq 0,05$  que indica distribuição diferente de normal.
- d) Após identificar a distribuição dos dados, calculou-se a correlação linear, utilizando o método mais adequado: ‘Coeficiente de correlação de Kendall’. Cálculo do Coeficiente de Kendall:

$$\tau = \frac{x_i > x_j \text{ e } y_i > y_j \text{ ou se } x_i < x_j \text{ e } y_i < y_j - (x_i > x_j \text{ e } y_i < y_j \text{ ou se } x_i < x_j \text{ e } y_i > y_j)}{n(n-1)/2}$$

(quantidade de pares concordantes) – (quantidade de pares discordantes)

Coeficiente de correlação ( $\tau$ )	Correlação Positiva	Coeficiente de correlação ( $\tau$ )	Correlação Negativa
$\tau = 1$	Perfeita	$\tau = -1$	Perfeita
$0,95 \leq \tau < 1$	Muito forte	$-0,95 \leq \tau < -1$	Muito forte
$0,8 \leq \tau < 0,95$	Forte	$-0,8 \leq \tau < -0,95$	Forte
$0,5 \leq \tau < 0,8$	Moderada	$-0,5 \leq \tau < -0,8$	Moderada
$0 \leq \tau < 0,5$	Fraca	$0 \leq \tau < -0,5$	Fraca

- e) Após calcular o coeficiente de correlação linear, elaborou-se teste estatístico para validar a correlação linear na mesma função do cálculo do coeficiente de correlação, onde H0 (hipótese nula) p-value é  $> 0,05$  que indica que não há correlação linear, e Ha (hipótese alternativa) p-value é  $\leq 0,05$  que indica que existe correlação linear (ver Tabela 3 do Artigo).
- f) Análise dos resíduos: 1. Linearidade, 2. Normalidade dos resíduos, 3. Homocedasticidade e 4. Pontos de alavancagem.
- g) Resumo e Validação do Modelo:
1. Estatística t (variáveis independentes): Ho = coeficiente igual a zero:  $p > 0,05$  (coeficiente não validado); Ha = coeficiente diferente de zero:  $p \leq 0,05$  (coeficiente validado).
  2. Estatística F: Ho = Modelo predictor é igual com ou sem variável independente:  $p > 0,05$  (reprovado modelo de regressão); Ha = Modelo predictor é diferente com ou sem variável independente:  $p \leq 0,05$  (aprovado modelo de regressão). (ver Tabela 4 do Artigo).
3. Assim organizou-se o ‘Coeficiente de determinação’ de forma tabular (ver Tabela 5 do Artigo). Cálculo do Coeficiente de determinação:

$$r^2 = \frac{\textit{Variação encontrada}}{\textit{Variação total}}$$

$$r^2 = \tau^2$$

## CONSIDERAÇÕES FINAIS

Este estudo de mapeamento das doenças negligenciadas no Brasil trouxe importantes contribuições para a compreensão do contexto espacial dessas doenças, bem como para a identificação das variáveis socioeconômicas e ambientais que mais afetam a ocorrência dessas enfermidades.

A modelagem linear generalizada, utilizada tanto por municípios quanto pixel a pixel, permitiu não só a previsão da contagem média, como também a identificação de variáveis preditoras mais significativas considerando as oito doenças estudadas. A análise dos resultados permitiu a identificação de áreas mais vulneráveis e agrupamentos de municípios com características semelhantes em termos de número de casos de cada doença.

Entre as variáveis socioeconômicas e ambientais estudadas, destacam-se o Percentual de cobertura populacional na atenção básica, Percentual de domicílios com lixo coletado, Percentual de domicílios sem banheiro, Percentual de domicílios sem energia elétrica, Percentual de cobertura populacional na atenção básica, Percentual da população alfabetizada, Densidade demográfica, Precipitação e Temperatura média. Esses fatores tiveram grande impacto na ocorrência de doenças negligenciadas no país, e sua identificação é fundamental para a elaboração de políticas públicas eficazes de combate a essas enfermidades.

É importante ressaltar, no entanto, que alguns resultados foram ambíguos e exigem uma análise mais detalhada. Por exemplo, o maior número de doses de vacina aplicadas teve um efeito positivo na quantidade de casos de tuberculose. Esses resultados ressaltam a importância de estudos mais detalhados para uma compreensão mais precisa das dinâmicas dessas doenças.

Além disso, é importante destacar que o estudo teve como foco o contexto espacial das doenças negligenciadas no Brasil, mas outros fatores, como a disponibilidade e o acesso aos serviços de saúde, também são fundamentais para o controle e prevenção dessas enfermidades. É necessário, portanto, uma abordagem ampla e integrada para o combate a essas doenças, que inclua a melhoria das condições socioeconômicas e ambientais.

Em suma, este estudo traz importantes contribuições para a compreensão do contexto espacial das doenças negligenciadas no Brasil, bem como para a identificação de variáveis preditoras mais significativas para cada uma dessas enfermidades. Os resultados obtidos podem auxiliar na elaboração de políticas públicas mais eficazes para o controle e prevenção dessas doenças, contribuindo para a promoção da saúde e o bem-estar da população brasileira; como por

exemplo, distribuição de médicos, priorizar ações, desenvolver intervenções direcionadas, promover a educação e a conscientização e melhorar o monitoramento e a avaliação da saúde da população local.