

Ezequiel de Braga Santos

Modelagem conjunta aplicada ao estudo da leishmaniose no estado de São Paulo

Rio de Janeiro, Brasil

17 de junho de 2024

Ezequiel de Braga Santos

**Modelagem conjunta aplicada ao estudo da leishmaniose
no estado de São Paulo**

Projeto de monografia apresentado para a Escola de Matemática Aplicada (FGV/EMAp) como requerimento parcial para a continuidade da monografia.

Orientador: Prof. Luiz Max Carvalho

Fundação Getúlio Vargas – FGV
Escola de Matemática Aplicada
Curso de Graduação em Matemática Aplicada

Rio de Janeiro, Brasil
17 de junho de 2024

Sumário

1	INTRODUÇÃO	3
2	OBJETIVO PRINCIPAL	4
3	METODOLOGIA	5
3.1	Dados utilizados	5
3.2	Modelo	7
4	RESULTADOS PRELIMINARES	9
	Referências	13

1 Introdução

A leishmaniose é uma doença causada por um protozoário parasita da espécie *Leishmania*, transmitida através da picada de flebotomíneos infectadas. Ela pode se manifestar através de três formas principais ([WORLD HEALTH ORGANIZATION, 2024](#)): visceral (a mais grave), cutânea (a mais comum, geralmente causando úlceras na pele) e mucocutânea (afetando boca, nariz e garganta). Nossa estudo será concentrado na leishmaniose visceral (LV), fatal em mais de 95% dos casos se não for tratada, segundo a Organização Mundial da Saúde (OMS) ([WORLD HEALTH ORGANIZATION, 2024](#)). No estado de São Paulo, os dados revelam uma baixa incidência na última década, mas há uma concentração de casos em determinadas regiões do estado que merecem uma atenção.

Devido a complexidade e a diversidade de fatores que influenciam na sua propagação, é importante explorar técnicas mais avançadas que permitam capturar o efeito desses diferentes aspectos e forneçam insights mais profundos sobre a dinâmica da doença. Nesse contexto, os “joint models” (modelos conjuntos) se destacam como ferramenta estatística poderosa para a modelagem conjunta de dados longitudinais e de tempo até um evento.

Em ([ALSEFRI et al., 2020](#)) é destacado o uso dessa metodologia em 75 artigos e a aplicação em 14, usando diferentes submodelos tanto para as medições longitudinais quanto para o tempo até o evento. Para o primeiro, é constatado o uso desde modelos lineares generalizados (do inglês, GLM) a modelos mais complexos, assumindo diferentes estruturas de associação com o segundo submodelo.

Embora essa abordagem tenha sido aplicada a várias doenças, sua aplicação à leishmaniose parece pouco explorada. Assim, seu estudo não só ampliará o escopo de aplicações da modelagem conjunta, como também permitirá uma melhor compreensão da eficácia dessa metodologia na análise da leishmaniose.

2 Objetivo principal

O objetivo deste projeto é ser capaz de capturar e identificar os fatores que influenciam na dinâmica da doença, sobretudo como as condições ambientais de cada região interferem no surgimento de vetores transmissores e, consequentemente, na infecção de cachorros, por exemplo.

3 Metodologia

3.1 Dados utilizados

Para a realização deste trabalho, foi construído um conjunto de dados que contém várias covariáveis obtidas de diferentes fontes. Para cada cidade do estado de São Paulo, foi obtido o código, região e população, através do Instituto Brasileiro de Geografia e Estatística (IBGE) ([INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE, s.d.](#)). Além disso, foram coletados dados referentes ao solo: do pacote R, **geobr** ([PEREIRA; GONCALVES, 2024](#)), foi coletada a área de cada bioma nas cidades; e as classificações de uso da terra, obtidas do MapBiomas ([\(SEEG\), 2024](#)). Por outro lado, foram adicionadas covariáveis climáticas, obtidas do European Centre for Medium-Range Weather Forecasts (ECMWF) ([ECMWF, 2024b](#)), do Copernicus Atmosphere Monitoring Service (CAMS) ([ECMWF, 2024a](#)) e do Instituto Nacional de Pesquisas Espaciais (INPE) ([INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE, s.d.](#)). Por fim, há os dados referentes aos casos de Leishmaniose entre de 2003 a 2017, obtidos de ([CASANOVA et al., 2015](#)), do Departamento de Informação e Informática do Sistema Único de Saúde (DATASUS) ([MINISTÉRIO DA SAÚDE, s.d.](#)) e da Secretaria de Estado de Saúde de São Paulo ([CENTRO DE VIGILÂNCIA EPIDEMIOLÓGICA 'PROF. ALEXANDRE VRANJAC- CVE, s.d.](#)). Abaixo está a descrição das variáveis.

Tabela 1 – Descrição das variáveis

Variável	Descrição
code_7d	Código de 7 dígitos do IBGE
name_city	Nome da cidade
name_state	Nome do estado
name_region	Nome da região
year	Ano
total_area_m2	Área total da cidade
prop_cer	Proporção do bioma Cerrado na cidade
prop_ma	Proporção do bioma Mata Atlântica na cidade
prop_sc	Proporção do bioma Sistema Costeiro na cidade
mb_i_j_k_l_m	Classificação de uso do solo do MapBiomias, onde i representa o nível 0 e m, o nível 4
co_ppb	Média anual da concentração de monóxido de carbono em ppb
no2_ppb	Média anual da concentração de dióxido de nitrogênio em ppb
o3_ppb	Média anual da concentração de ozônio em ppb
pm25_ugm3	Média anual da concentração de material particulado de até 2.5 micrômetros, medido em micrograma por metro cúbico
so2_ugm3	Média anual da concentração de dióxido de enxofre, medido em micrograma por metro cúbico
rainfall	Média anual da precipitação em mm
temperature_c	Média anual da temperatura em graus Celsius
relative_humidity_percent	Média anual da umidade relativa do ar em %
wind_direction_degree	Média anual da direção do vento em graus
wind_speed_ms	Média anual da velocidade do vento em metros por segundo
fire_outbreaks	Quantidade de focos de queimada por ano
status_dogs	(1) representa a presença do cão infectado, onde o município a partir do ano de invasão, é considerado invadido ao longo da série
status_human	Como status_dogs, representa a presença de humano infectado
status_vector	Como status_dogs, representa a presença do vetor no município
n_cases	Número de casos na cidade por ano
n_deaths	Número de mortes na cidade por ano
pop	População por ano
incidence	Incidência por 100 mil habitantes
death_rate	Taxa de mortalidade por 100 mil habitantes

Para a modelagem, serão utilizados covariáveis selecionadas daquelas que apresentam menos de 25% de dados faltantes (ou zeros, no caso da classificação de uso da terra), conforme a figura a seguir:

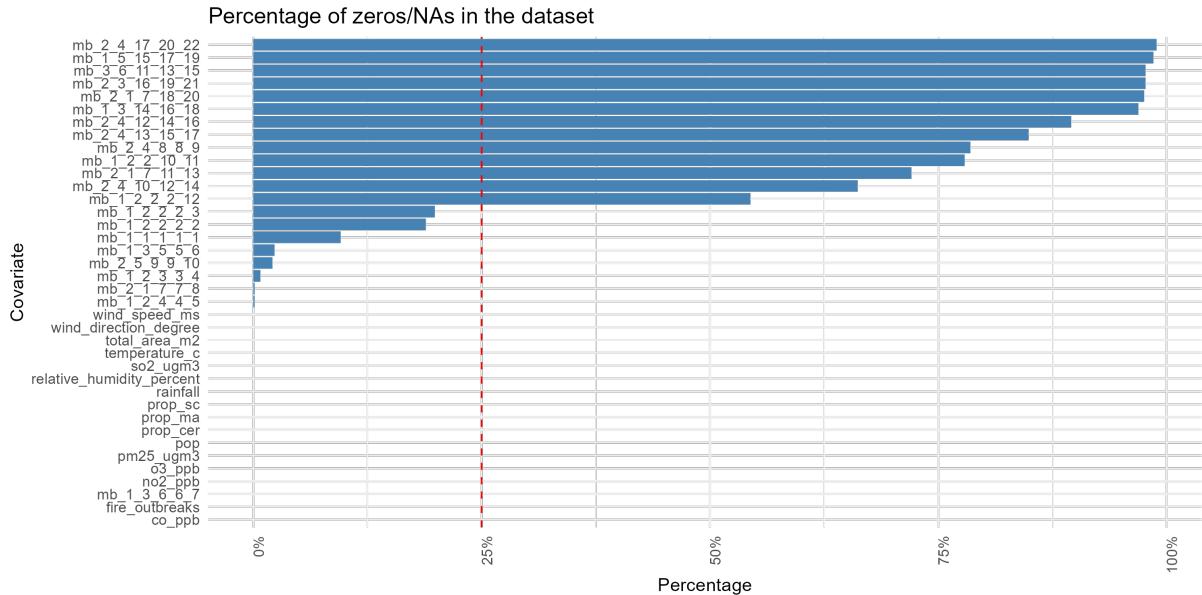


Figura 1 – Porcentagem de dados faltantes (zeros, no caso das variáveis do MapBiomas).

3.2 Modelo

A modelagem usual de um modelo conjunto, proposta em ([HENDERSON; DIGGLE; DOBSON, 2000](#)), consiste em supor que a distribuição conjunta das medições longitudinais e do evento é regida por um processo gaussiano $W_i(t) = \{W_{1i}(t), W_{2i}(t)\}$ para cada indivíduo i . Para isso, considera-se que foram observados m sujeitos ao longo de um intervalo de tempo com medições y_{ij} no tempo t_{ij} , para $j = 1, \dots, n_i$ e i -ésimo indivíduo. Além disso, considera-se uma função $H_i(u)$ que indica se o sujeito vivenciou o evento no tempo u .

O modelo para as medições longitudinais y_{ij} no tempo t_{ij} é dado por

$$Y_{ij} = \mu_i(t_{ij}) + W_{1i}(t_{ij}) + Z_{ij},$$

com $\mu_i(t) = x_{1i}^\top(t)\beta_1$ e $Z_{ij} \sim \text{Normal}(0, \sigma_Z^2)$, onde $x_{1i}^\top(t)$ são as covariáveis selecionadas no tempo t .

Para a intensidade do evento no tempo t , o modelo é dado por

$$\lambda_i(t) = H_i(t)\alpha_0(t) \exp[x_{2i}^\top(t)\beta_2 + W_{2i}(t)],$$

para alguma função de base $\alpha_0(t)$ e covariáveis selecionadas $x_{2i}^\top(t)$.

Seguindo essa estrutura, propõe-se modelar as variáveis respostas que indicam a presença do vetor/cão contaminado (*status_vector/status_dogs*) considerando covariáveis como biomarcadores.

4 Resultados preliminares

Até o momento, os esforços foram concentrados em preparar os dados e obter informações a partir de análises exploratórias. Um dos primeiros resultados trata-se da incidência da doença, onde pode-se perceber uma maior concentração de casos ao longo de uma faixa regional, conforme a figura a seguir:

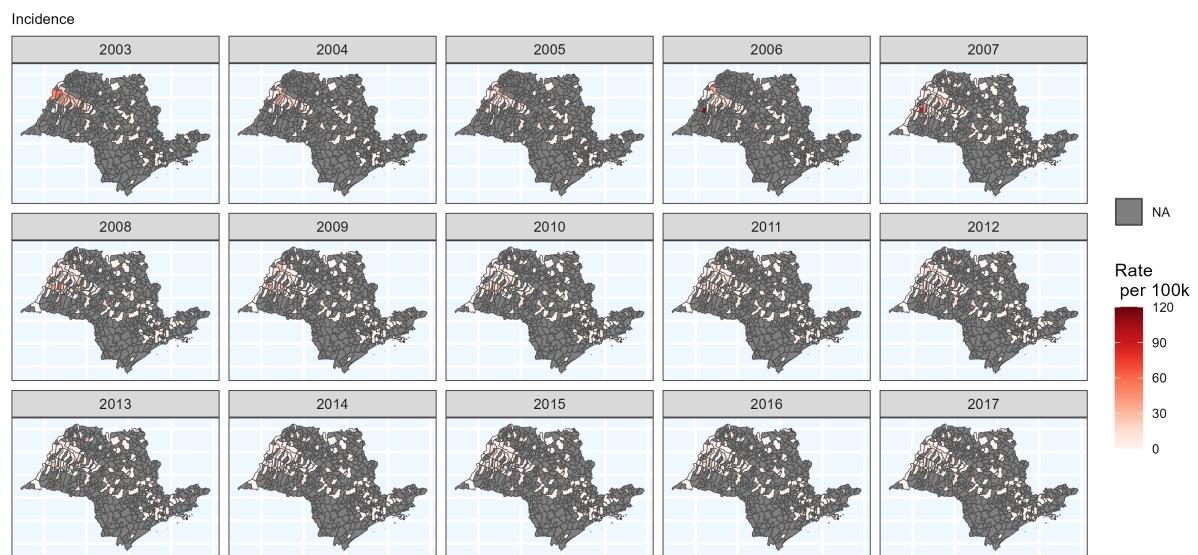


Figura 2 – Incidência por 100 mil habitantes.

Isso se deve ao fato de que há uma grande concentração do vetor transmissor (*Lutzomyia*) nessa região, como pode-se perceber na figura abaixo:

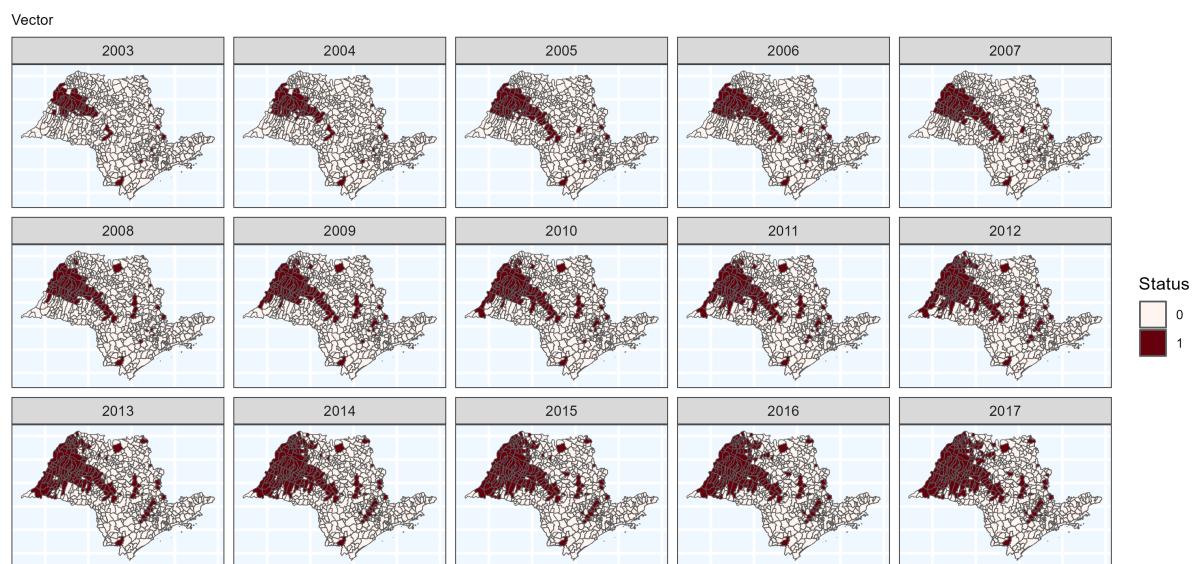


Figura 3 – Detecção do vetor na cidade: 1 para presença e 0 para ausência.

Como resultado, há uma maior presença de animais infectados nas regiões próximas, como é o caso dos cachorros:

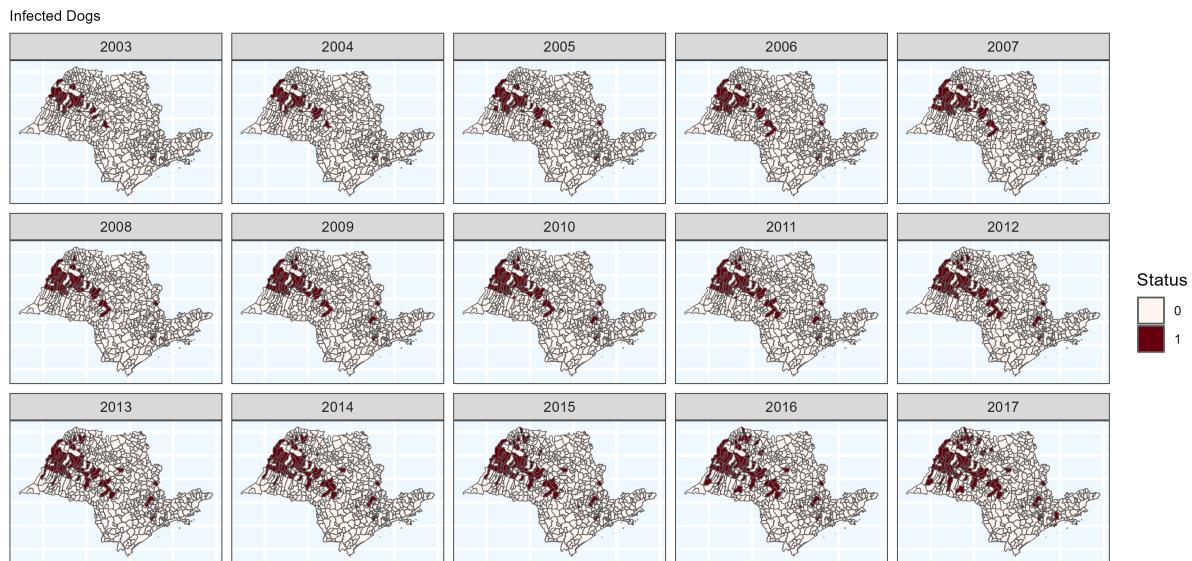


Figura 4 – Detecção da doença em cachorros: 1 para presença e 0 para ausência.

Outro fator importante analisado foi a concentração de moléculas indicadoras de poluição do ar como monóxido de carbono, por exemplo. Pode-se perceber uma maior presença dessas variáveis na região metropolitana de São Paulo, conforme a figura abaixo:

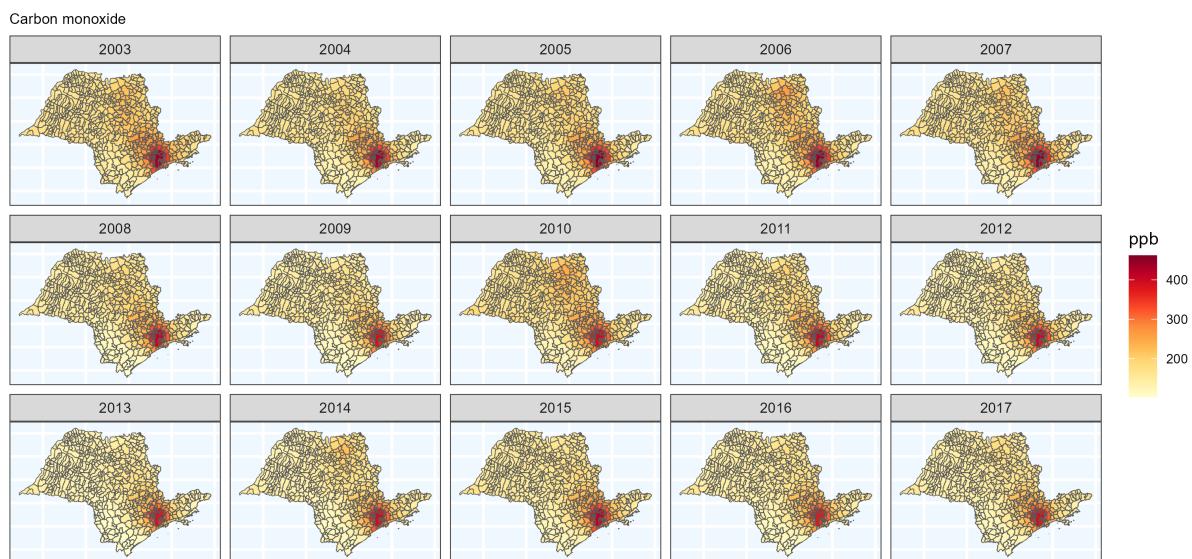


Figura 5 – Concentração de moléculas de monóxido de carbono.

Nessa mesma linha, observa-se que a trajetória dessas covariáveis são parecidas ao longo do tempo para diferentes cidades, como pode-se perceber no caso do dióxido de nitrogênio, por exemplo:

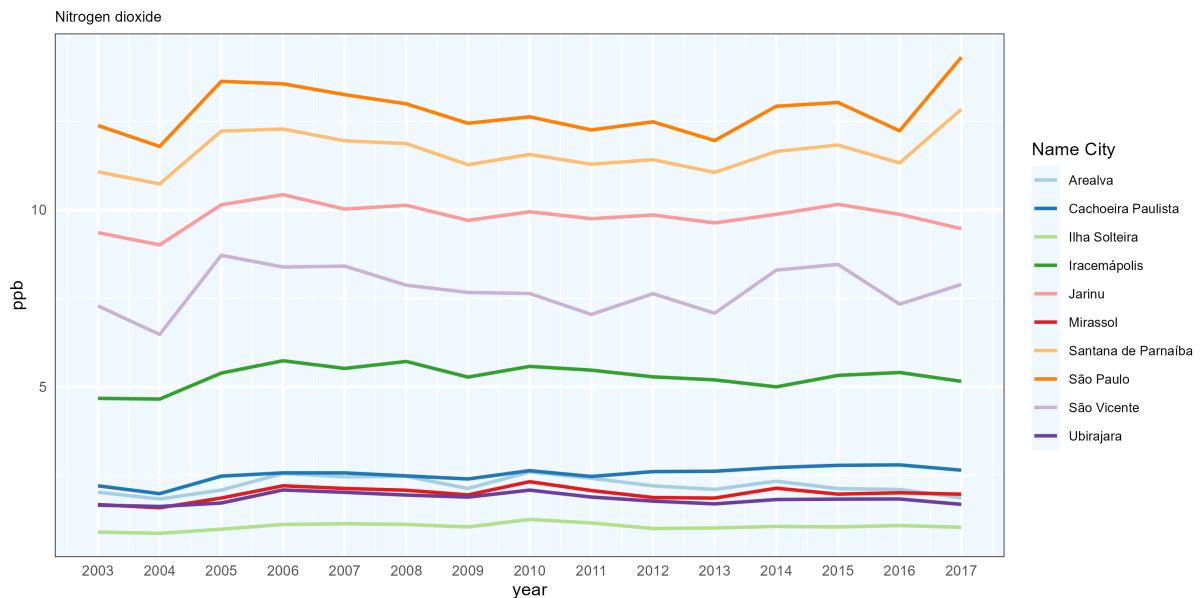
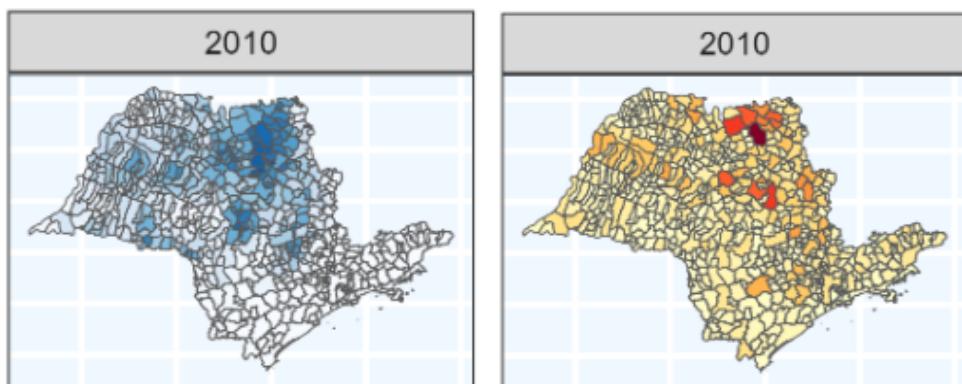


Figura 6 – Série temporal de dióxido de nitrogênio.

Outro resultado a ser destacado é a intensidade de queimadas em regiões de plantação de cana-de-açúcar, como pode-se perceber mais claramente no ano de 2010:



(a) Área cultivada por cana-de-açúcar normalizada pela área total.

(b) Quantidade de queimadas.

Além disso, foi analisada a correlação entre as covariáveis, onde pode-se destacar forte correlação entre aquelas indicadoras de poluição ambiental como também entre a área urbana e essas mesmas covariáveis. Outros resultados esperados foram a correlação de umidade relativa e direção do vento com formação florestal e também com temperatura. Todas as correlações estão sumarizadas na figura a seguir.

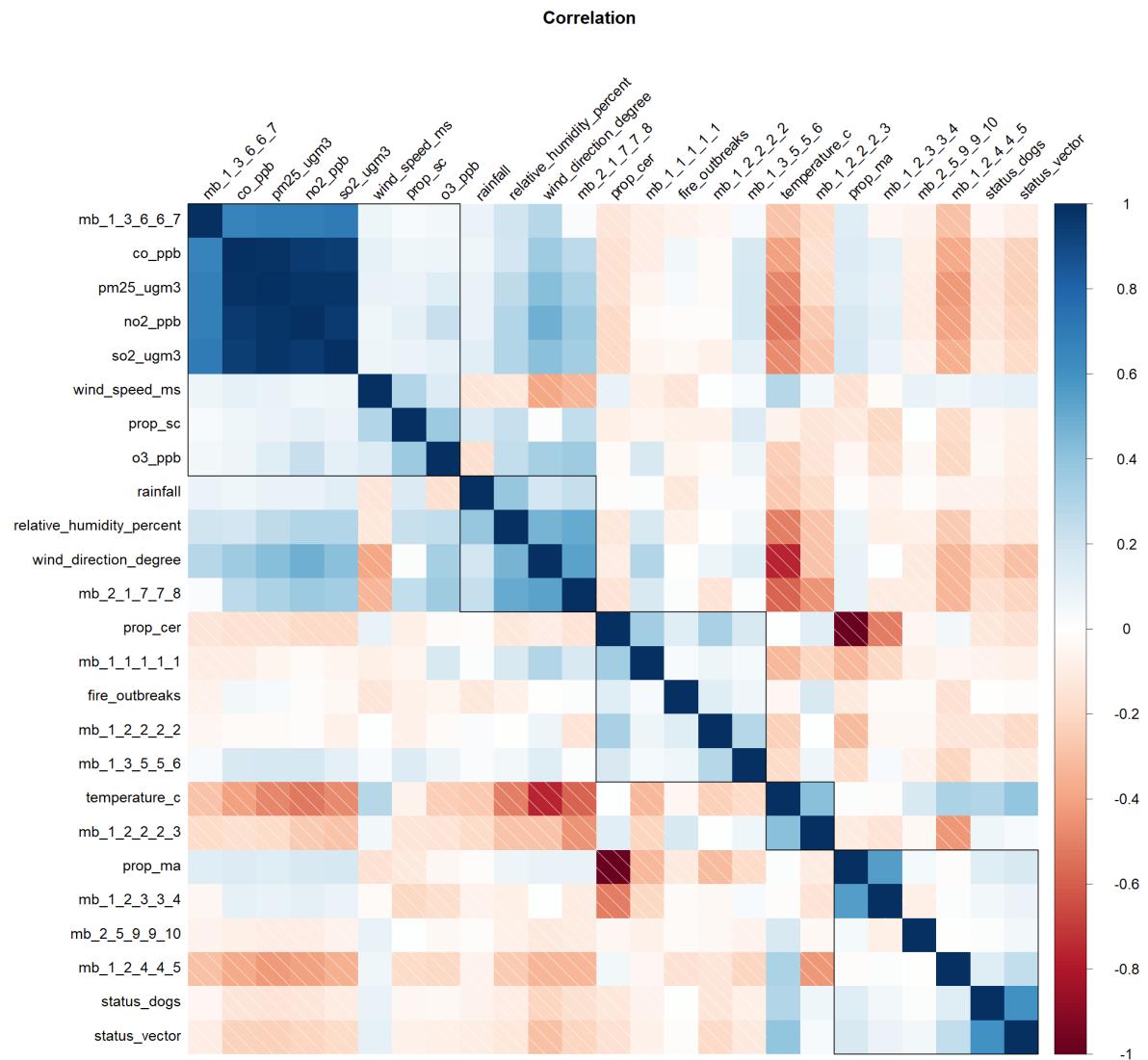


Figura 8 – Correlação das covariáveis

Referências

- (SEEG), Greenhouse Gas Emissions Estimation System. **Projeto MapBiomas – Coleção 5.0 da Série Anual de Mapas de Cobertura e Uso de Solo do Brasil.** [S.l.: s.n.], 2024. Disponível em: <<https://brasil.mapbiomas.org/en/>>.
- ALSEFRI, Maha et al. Bayesian joint modelling of longitudinal and time to event data: a methodological review. **BMC Medical Research Methodology**, v. 20, p. 465–480, abr. 2020. ISSN 1471-2288. DOI: [10.1186/s12874-020-00976-2](https://doi.org/10.1186/s12874-020-00976-2). Disponível em: <<https://doi.org/10.1186/s12874-020-00976-2>>.
- CASANOVA, Claudio et al. Distribution of Lutzomyia longipalpis Chemotype Populations in São Paulo State, Brazil. **PLOS Neglected Tropical Diseases**, Public Library of Science, v. 9, n. 3, p. 1–14, mar. 2015. DOI: [10.1371/journal.pntd.0003620](https://doi.org/10.1371/journal.pntd.0003620). Disponível em: <<https://doi.org/10.1371/journal.pntd.0003620>>.
- CENTRO DE VIGILÂNCIA EPIDEMIOLÓGICA "PROF. ALEXANDRE VRANJAC- CVE. Leishmaniose Visceral. [S.l.: s.n.]. CVE website. Disponível em: <<https://saude.sp.gov.br/cve-centro-de-vigilancia-epidemiologica-prof.-alexandre-vranjac/areas-de-vigilancia/doencas-de-transmissao-por-vetores-e-zoonoses/agravos/leishmaniose-visceral/>>.
- ECMWF. **Copernicus Atmosphere Monitoring Service (CAMS)**. [S.l.: s.n.], 2024. Disponível em: <<https://atmosphere.copernicus.eu/>>.
- _____. **ECMWF Reanalysis v5 (ERA5)**. [S.l.: s.n.], 2024. Disponível em: <<https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5>>.
- HENDERSON, Robin; DIGGLE, Peter; DOBSON, Angela. Joint modelling of longitudinal measurements and event time data. **Biostatistics**, v. 1, n. 4, p. 465–480, dez. 2000. ISSN 1465-4644. DOI: [10.1093/biostatistics/1.4.465](https://doi.org/10.1093/biostatistics/1.4.465). eprint: <https://academic.oup.com/biostatistics/article-pdf/1/4/465/655029/010465.pdf>. Disponível em: <<https://doi.org/10.1093/biostatistics/1.4.465>>.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Downloads**. [S.l.: s.n.]. IBGE website. Disponível em: <<https://www.ibge.gov.br/geociencias/downloads-geociencias.html>>.
- INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **BDQUEIMADAS**. [S.l.: s.n.]. INPE website. Disponível em: <<https://terrabrasilis.dpi.inpe.br/queimadas/bdqueimadas/>>.

MINISTÉRIO DA SAÚDE. LEISHMANIOSE VISCERAL - CASOS CONFIRMADOS NOTIFICADOS NO SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO - SÃO PAULO. [S.l.: s.n.]. DATASUS website.

Disponível em: <<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinanwin/cnv/leishvSP.def>>.

PEREIRA, Rafael H. M.; GONCALVES, Caio Nogueira. **geobr: Download Official Spatial Data Sets of Brazil**. [S.l.: s.n.], 2024. R package version 1.8.2. Disponível em: <<https://CRAN.R-project.org/package=geobr>>.

WORLD HEALTH ORGANIZATION. **Leishmaniasis**. [S.l.: s.n.], 2024. World Health Organization website. Disponível em:

<<https://www.who.int/news-room/fact-sheets/detail/leishmaniasis>>. Acesso em: 11 jun. 2024.