



**UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”**

Projeto Final — FCM2025



Autores

Gabriel M. Gandin

Marcio Gomes

Raquel Lima

Valdemiro Pitoro

Professor: Thomas Nogueria Vilches
Novembro 2025

1 Introdução

Nosso grupo é formado por estudantes das áreas de Agronomia, Biologia, Engenharia Civil e Zootecnia. Ao longo do projeto, buscamos aplicar na prática todos os conceitos apresentados em aula. Para isso, construímos uma base de dados sobre casos de tuberculose, incluindo resultados positivos e negativos.

A base foi filtrada para incluir dados dos anos de 2023 e 2024, referentes a municípios do estado de São Paulo, separados por sexo (masculino e feminino) e por duas faixas etárias: 20–39 anos e 40–59 anos.

Durante o desenvolvimento, aplicamos conceitos de versionamento de código utilizando o GitHub para organizar e compartilhar os arquivos do projeto. As análises e manipulações dos dados foram realizadas no RStudio, com uso dos pacotes `tidyverse` e `ggplot2`. Ao final, produzimos um resumo em formato HTML via RMarkdown e este relatório em LaTeX.

A figura 1 mostra a tela inicial do DataSUS (aba TABNET), usada para extrair os dados.

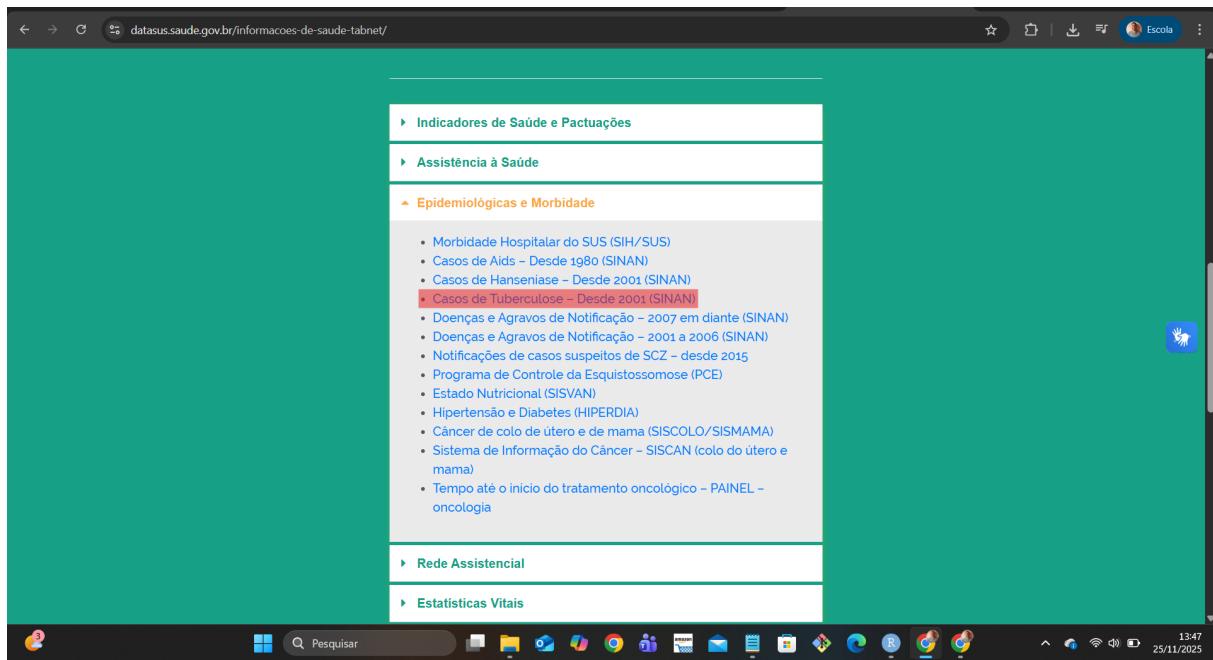


Figure 1: Tela de seleção de dados no DataSUS.

Após a definição do tema, iniciamos os filtros para selecionar os dados de interesse relacionados à tuberculose (ver Figura 2 e 3).

► TUBERCULOSE - CASOS CONFIRMADOS NOTIFICADOS NO SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO - SÃO PAULO

The screenshot shows the initial filtering stage. It includes three columns: 'Linha' (Row) with 'Município de notificação', 'Região de Saúde (CIR) de notif', 'Reg. Saúde/Município de notific', and 'Macrorreg. de Saúde de notif'; 'Coluna' (Column) with 'Não ativa', 'Ano Diagnóstico', 'Mês Diagnóstico', and 'Ano Notificação'; and 'Conteúdo' (Content) with 'Casos confirmados'. Below this is a section titled '► PERÍODOS DISPONÍVEIS' containing a dropdown menu with years from 2019 to 2024.

Figure 2: Direcionamento dos dados relevantes para análise.

This screenshot shows the detailed refinement of the data source. It lists various categories with their respective sub-options. The visible categories include:

- Zona Residência**: Todas as categorias, Em branco/IGN, <1 Ano, 1-4, 5-9, 10-14, 15-19, 20-39, 40-59, 60-64.
- Fx Etária**: Todas as categorias, Em Branco, Ignorado, Masculino, Feminino.
- Fx Etária 7**: Todas as categorias, Em Branco, Ignorado, Masculino, Feminino.
- Fx Etária 13**: Todas as categorias, Em Branco, Ignorado, Masculino, Feminino.
- Raça**: Todas as categorias, Branca, Negra, Parda, Amarela, Indígena.
- Sexo**: Todas as categorias, Masculino, Feminino.
- Tipo de entrada**: Todas as categorias, Institucionalizado, PPL, Pop. Sit. Rua, Prof. saúde, Imigrante, Benefic. governo, Forma, Se Extrapulm.1, Se Extrapulm.2, Aids, Alcoolismo, Diabetes, Doença mental, Drogas ilícitas, Tabagismo, Outra doença.
- Confirmação laboratorial**: Todas as categorias, Com confirmação laboratorial, Sem confirmação laboratorial.
- 1ºBac Escarro**: Todas as categorias, Com confirmação laboratorial, Sem confirmação laboratorial.

Figure 3: Refinamento da fonte dos dados no DataSUS.

Após extraímos os dados, cada filtro gerou um arquivo Excel individual (marcados em verde na imagem). Esses arquivos foram reunidos em uma planilha consolidada (marcada em vermelho). No entanto, ao importar essa planilha no RStudio, enfrentamos dificuldades, o que tornou necessário converter os dados para o formato de *coluna longa* (arquivo destacado em azul). (ver Figura 4).

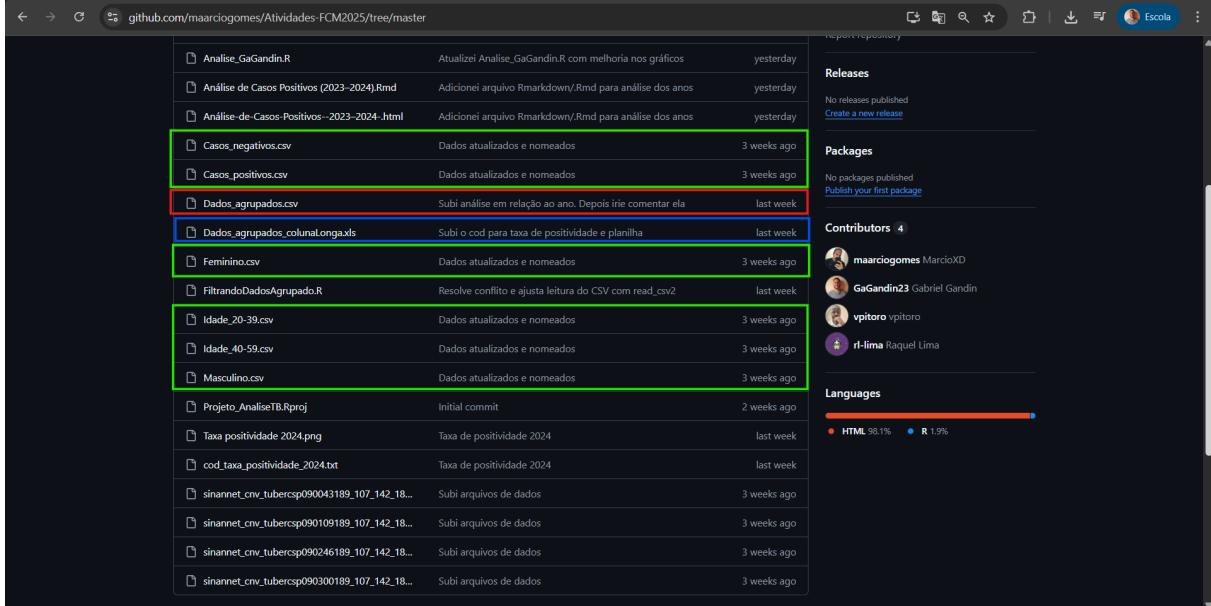


Figure 4: Arquivos do projeto organizados no GitHub.

2 Objetivos

Aprender a manipular, analisar e visualizar dados no R, utilizando o GitHub para versionamento e organização do projeto, aplicando esses conhecimentos a dados reais de tuberculose do DataSUS.

3 Resultados

A seguir apresentamos os gráficos e tabelas gerados pelo script desenvolvido e interpretado por cada autor.

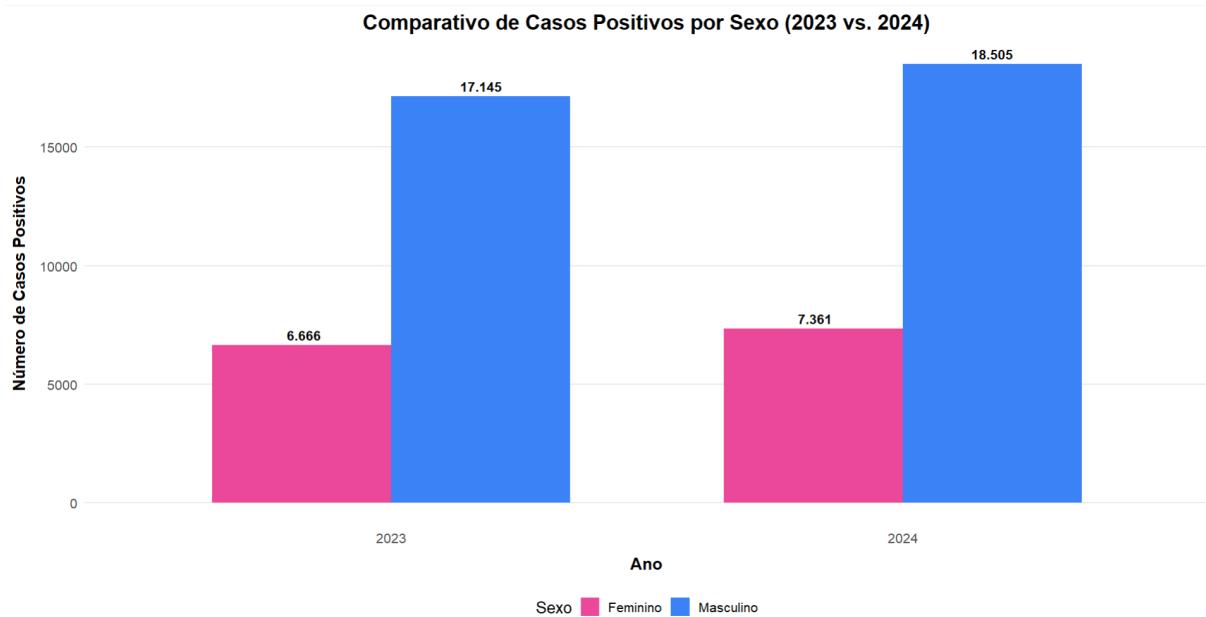


Figure 5: Casos positivos e negativos de tuberculose — filtro por sexo.

Como pode ser visto na Figura 5, o sexo masculino apresenta maior número de casos positivos.

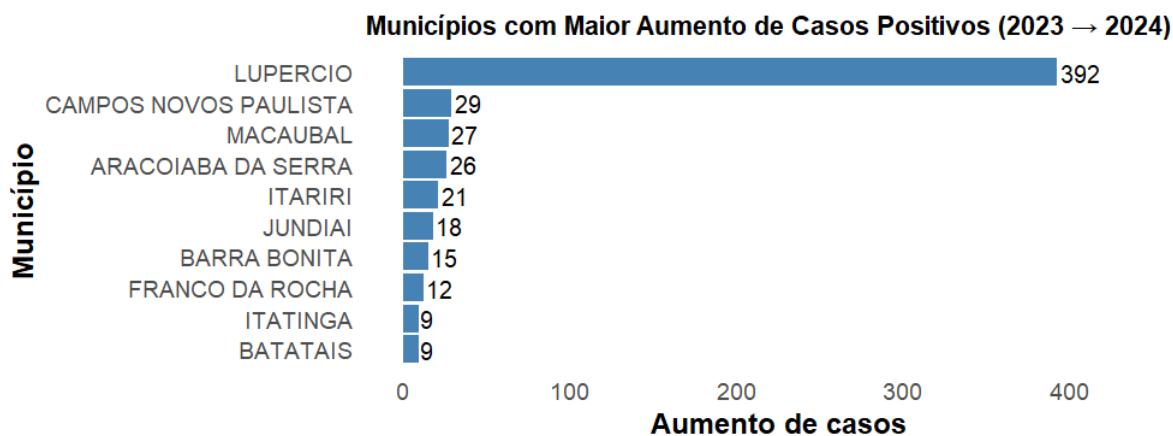


Figure 6: Maiores aumentos de casos positivos entre 2023 e 2024.

A Figura 6 apresenta os dez municípios com maior aumento.

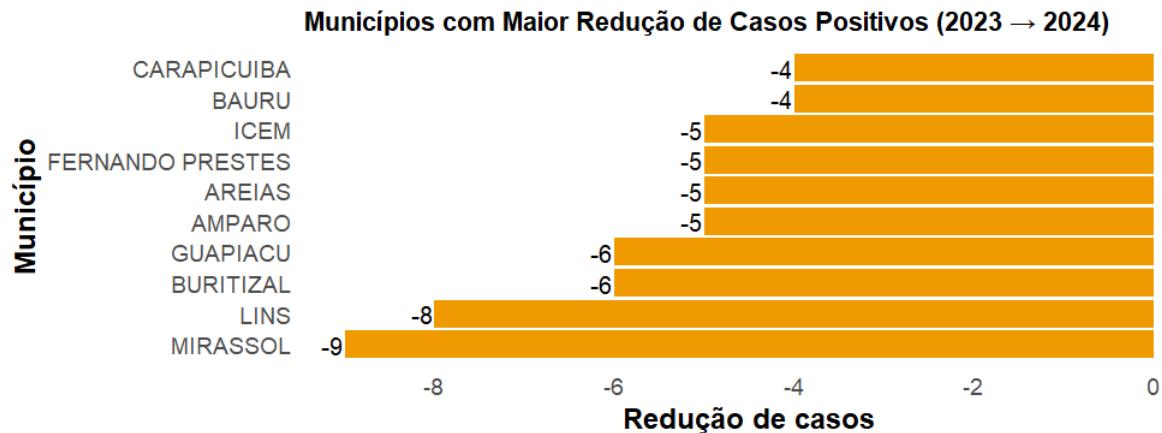


Figure 7: Maiores reduções de casos positivos entre 2023 e 2024.

A Figura 7 apresenta os dez municípios com maior redução.

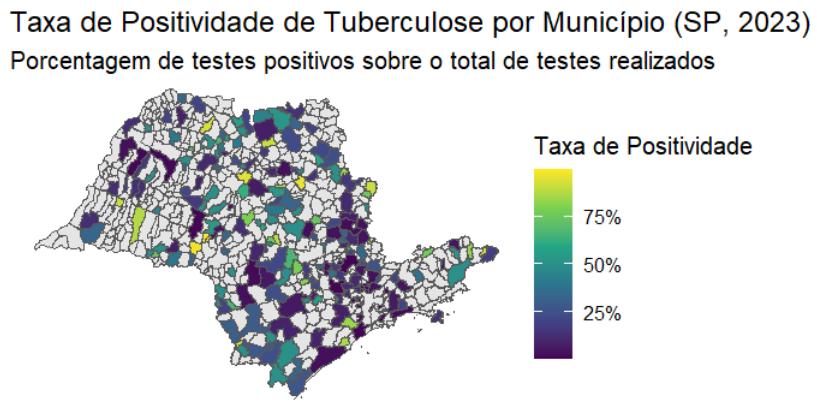


Figure 8: Mapa de calor — casos positivos de tuberculose em 2023.

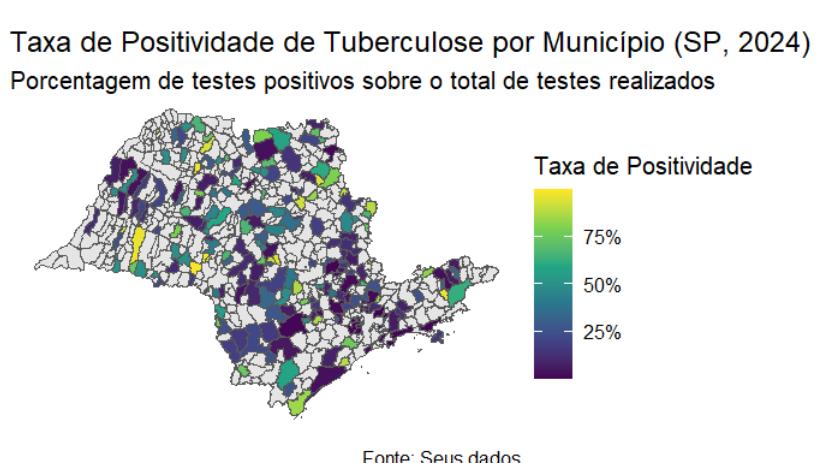
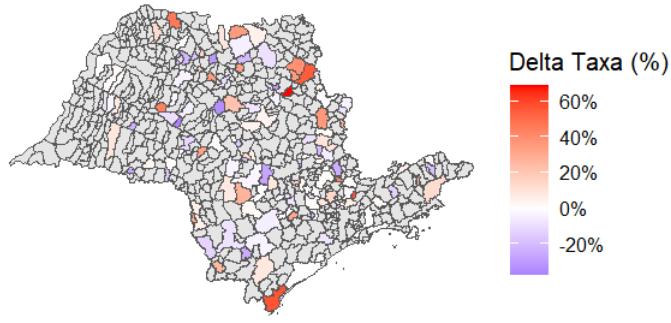


Figure 9: Mapa de calor — casos positivos de tuberculose em 2024.

Comparação da Taxa de Positividade: 2024 vs. 2023
 Variação Percentual (2024 menos 2023): Vermelho = Aumento, Azul = Redução



Fonte: Seus dados

Figure 10: Variação dos casos positivos entre 2023 e 2024 no estado de São Paulo.

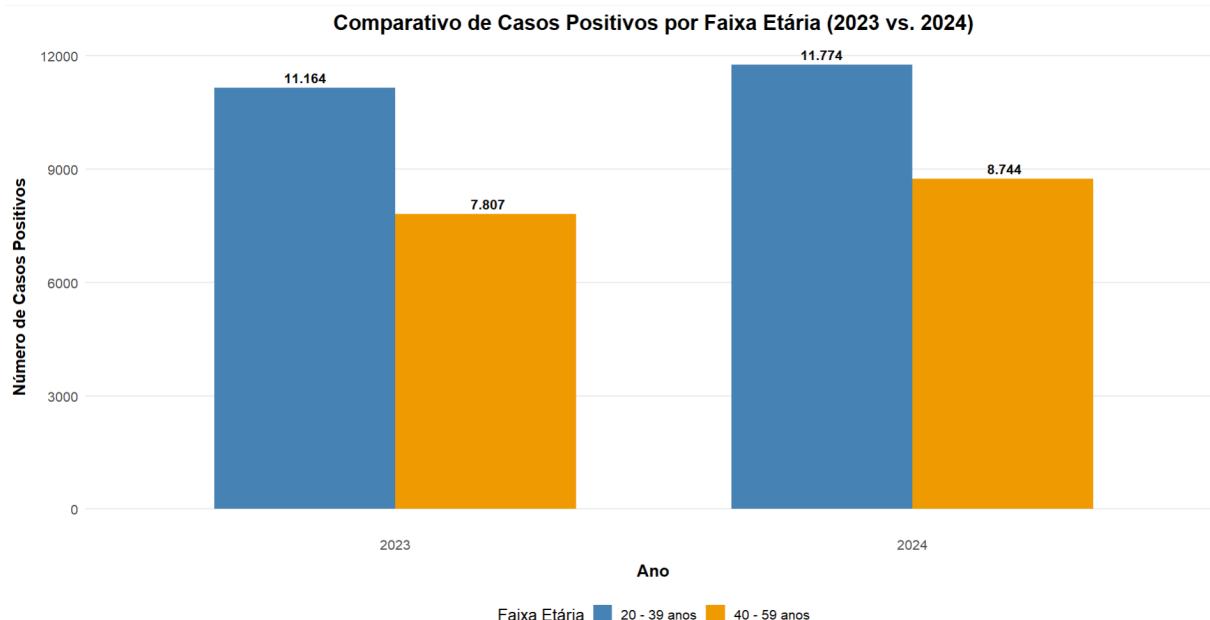


Figure 11: Casos positivos entre duas faixas etárias.

Como podemos ver a Figura 6, a faixa etária de menor idade apresenta maiores casos positivos nos dois anos de coleta de dados.

4 Conclusão

- Os dados mostram aumento dos casos de tuberculose de 2023 para 2024, com o sexo masculino apresentando valores significativamente maiores.
- Lúpercio registrou o maior aumento de casos positivos no período (392 novos casos), enquanto Mirassol apresentou a maior redução (queda de 9 casos).
- A faixa etária de 20–39 anos apresentou o maior número total de casos, com 22.938 registros.